

# Boundary height fields in the Abelian sandpile model

Geoffroy Piroux\* and Philippe Ruelle†

Institut de Physique Théorique  
Université Catholique de Louvain  
B-1348 Louvain-La-Neuve, Belgium

## Abstract

We study the abelian sandpile model on the upper half plane, and reconsider the correlations of the four height variables lying on the boundary. For more convenience, we carry out the analysis in the dissipative (massive) extension of the model and identify the boundary scaling fields corresponding to the four heights. We find that they all can be accounted for by the massive perturbation of a  $c = -2$  logarithmic conformal field theory.

## 1 Introduction

The description of equilibrium critical phenomena has been one of the greatest success of two-dimensional conformal theories in the past twenty years [1]. More recently, attention has focused on new types of observables in otherwise well-known models, and also on new kinds of critical systems. In both cases, non local features often play an important role, because either the observables one is interested in are themselves non local in terms of the natural microscopic variables, or else because the statistical model possesses intrinsic non local properties. These studies often lead to a description in terms of conformal theories with peculiar properties. A class of systems with such properties is provided by sandpile models. Some of them, and in particular the one we consider here, are believed to have a faithful description in terms of logarithmic conformal field theories.

Our motivation to study these models is two-fold. First, one wishes to see to what extent they lend themselves to a conformal field theoretic approach, and if the adequacy of the conformal description is as good as for the equilibrium systems. Second, the logarithmic theories have been developed for themselves, but are complex and some of their aspects are not fully understood yet. It should therefore be profitable to have concrete realizations in order to have a better understanding of the most peculiar features.

The sandpile model we consider here is the isotropic Abelian sandpile model (ASM), as originally defined in [2]. The most natural variables to consider in a conformal context are the four height variables. In the bulk, correlations of height 1 variables can be handled by local calculations [3], but height 2, 3 and 4 variables are much more complicated, and only their one-site probabilities are known [4]. For

---

\*piroux@fyoma.ucl.ac.be

†Chercheur Qualifié FNRS; ruelle@fyoma.ucl.ac.be

sites on a boundary, closed or open, Ivashkevich showed, by using suitable identities, that the non local configurations needed to handle the heights bigger than 1 could be reduced to local computations [5]. He was then able to compute the two-site joint probabilities of all height variables. He found that all correlations decay like  $r^{-4}$ , and inferred, wrongly as we will see, that all boundary height variables scale to the same conformal field.

Our aim in this paper is to revisit this problem and to identify the fields corresponding to the boundary height variables, but in a dissipative extension of the model, known to be described by a massive perturbation of the  $c = -2$  conformal theory [6]. The main advantage for doing this is that it allows for an unambiguous identification of the fields from a few 2-point correlators, off, and therefore also at, criticality. The so-obtained identification can then be checked from other 2-point and from 3-point functions. In contrast, when one considers the critical, non dissipative model, the 2-point functions yield ambiguous field identifications, which can only be fixed by using 3-point functions, and then checked from higher correlators.

The article is organized as follows. The next section defines the model and sets our conventions. Section 3 deals with the boundary unit height variable and what we call supercritical height variables. They are much easier than the other ones, and illustrate the way the identification with concrete fields is obtained. The identification of the height 1 and supercritical height fields also facilitates that of the other heights.

Sections 4 and 5 form the hardcore of the article. In Section 4, we explain our prescription—a two-step burning algorithm—to associate recurrent configurations of the sandpile to spanning trees, from which a clear characterization of the local height constraints follows. We use it to compute 2-site probabilities for having a height 1 or a supercritical height at one site, and any other height at the other site, from which we deduce the field identification of all height variables. Section 5 checks these results by computing explicitly all 2-site height probabilities in the massive extension of the sandpile model, and certain 3-site probabilities involving a height 1 or a supercritical height.

Since the bulk of the calculations reported here was done, an article by Jeng [7] has appeared, where precisely the same problem is addressed. The two works were carried out independently and differ in two ways. The first one is that we study the dissipative model, while Jeng considers it at criticality. This fact enables us to deduce the field identifications for the boundary height variables  $h > 1$  from the spanning tree characterizations of a single insertion of such variables, technically much simpler to solve than the 2-site insertions. The latter are only used as cross-checks, in contrast to the approach at criticality which needs them as inputs. So working off criticality offers a simpler and more reliable access to the fields. Secondly, we use a different characterization of the height variables bigger or equal to 2 in terms of spanning trees, which is based on a modified, two-step burning algorithm. This, we believe, leads to a more transparent formalism which generalizes to multisite probabilities. Our results and conclusions however fully agree with those of Jeng.

## 2 The dissipative sandpile model

Consider a finite portion  $\mathcal{L}$  of a square lattice and define at each site  $i$  a (sand) height variable  $h_i$  which can take the integer values  $1, 2, 3, \dots$ . A configuration  $\mathcal{C}$  of the sandpile is the set of height values  $\{h_i\}$  for all sites. The dynamics is defined in terms of a symmetric toppling matrix  $\Delta$ . Its entries are all integers, positive on the diagonal, negative off the diagonal, and it has row sums which are non negative. A configuration is called stable if all heights satisfy  $h_i \leq \Delta_{ii}$ .

The system evolves in discrete time as follows. To the stable configuration  $\mathcal{C}_t$  at time  $t$  we add a sand grain at a random site  $i$  (chosen with uniform distribution say), namely we set  $h_i \rightarrow h_i + 1$ . This new configuration, if stable, defines  $\mathcal{C}_{t+1}$ . If it is not stable, the unstable site  $i$  topples: it loses  $\Delta_{ii}$  grains,

every other site  $j$  receives  $-\Delta_{ij}$  grains, whereas  $\sum_{j \in \mathcal{L}} \Delta_{ij}$  sand grains are dissipated (they fall off the pile, to a sink). That is, when a site  $i$  topples, we update the heights according to

$$h_j \longrightarrow h_j - \Delta_{ij}, \quad \forall j \in \mathcal{L}. \quad (1)$$

If other sites become unstable after the toppling of the site  $i$ , they topple following the same rule. All unstable sites are then toppled until the configuration becomes stable again. This configuration is then taken as  $\mathcal{C}_{t+1}$ . In this way, the toppling at the seeded site can trigger a potentially large avalanche, resulting in a configuration  $\mathcal{C}_{t+1}$  which can be completely different from  $\mathcal{C}_t$ .

Provided there are dissipative sites, *i.e.* sites  $k$  for which  $\sum_{j \in \mathcal{L}} \Delta_{kj} > 0$ , the dynamics is well defined: it does not depend on the order in which the sites are toppled (the model is Abelian), and the relaxation of the seeded configuration to  $\mathcal{C}_{t+1}$  requires a finite number of topplings.

One is generally interested in the long time behaviour of the sandpile. As shown by Dhar [8], this behaviour is characterized by a unique time invariant probability measure  $P_{\mathcal{L}}^*$ , which specifies the probabilities of occurrence of all stable configurations, independently of the initial configuration. The moments of this measure, in the thermodynamic limit  $|\mathcal{L}| \rightarrow \infty$ , are what we want to put in correspondence with the correlators of a conformal field theory.

When the dynamics is started from a certain initial configuration, it produces at later times two kinds of configurations, called transient and recurrent in the terminology of Markov processes. The transient configurations are those which occur a finite number of times only (they may not occur at all, depending on the initial configuration). In the long run, they are not in the image of the dynamics, and have a zero measure with respect to  $P_{\mathcal{L}}^*$ . A simple example is the configuration with all  $h_i = 1$ , but more generally, any configuration with two 1's at neighbour sites (sites  $i, j$  with  $\Delta_{ij} \neq 0$ ) is transient.

The non transient configurations are recurrent. Their number is equal to  $\det \Delta$ , the determinant of the toppling matrix, and asymptotically, they occur with equal probability, so that the measure  $P_{\mathcal{L}}^*$  is uniform on them [8]. A criterion to decide whether a given configuration is recurrent or transient is based on the notion of forbidden sub-configuration (FSC): a sub-configuration, with support  $\mathcal{K} \subset \mathcal{L}$ , is said to be forbidden if  $h_i \leq -\sum_{j \in \mathcal{K} \setminus \{i\}} \Delta_{ji}$ , for all sites  $i$  of  $\mathcal{K}$ . For instance two neighbour sites with heights 1 form an FSC. Then a configuration is recurrent if and only if it contains no FSC [9].

A practical way to test a configuration is to use the burning algorithm [8]. At time 0, all sites in  $\mathcal{L}$  are unburnt and we define  $\mathcal{K}_0 = \mathcal{L}$  to be the set of unburnt sites at time 0. The sites  $i$  of  $\mathcal{K}_0$  such that  $h_i > -\sum_{j \in \mathcal{K}_0 \setminus \{i\}} \Delta_{ji}$  are burnable at time 1. So we burn them, obtaining a smaller set  $\mathcal{K}_1 \subset \mathcal{K}_0$  of unburnt sites at time 1. Then the sites of  $\mathcal{K}_1$  which are burnable at time 2, *i.e.* those satisfying  $h_i > -\sum_{j \in \mathcal{K}_1 \setminus \{i\}} \Delta_{ji}$ , are burnt. This leaves a smaller set  $\mathcal{K}_2 \subset \mathcal{K}_1$  of unburnt sites at time 2. This burning process is carried on until no more site is burnable, which means that  $\mathcal{K}_{T+1} = \mathcal{K}_T$  for a certain  $T$ . Then the configuration is recurrent if and only if all sites of  $\mathcal{L}$  have been burnt ( $\mathcal{K}_T = \emptyset$ ). Otherwise  $\mathcal{K}_T$  is an FSC.

The burning algorithm allows to define a unique rooted spanning tree on a graph  $\mathcal{L}^*$ , from the path followed by the fire in the lattice [9]. The graph  $\mathcal{L}^*$  has the sites of  $\mathcal{L}$  and the sink as vertices, and has links defined by  $\Delta$ : an off-diagonal entry  $\Delta_{ij} = -n$  means there are  $n$  bonds connecting the sites  $i$  and  $j$ , and each site  $i$  is connected to the sink by a number of bonds equal to  $\sum_{j \in \mathcal{L}} \Delta_{ij} \geq 0$ , the number of grains dissipated when  $i$  topples. At time 0, the sink is the only burnt site and forms the root of the tree. In the next steps, the fire propagates from the sink to those sites which are burnable at time 1, then from the sites which have been burnt at time 1 to those which are burnable at time 2, and so on. If a site burns at time  $t$ , one says that it catches fire from one among its neighbours that were burnt at time  $t - 1$  (or from the sink site at time 1). In case there are more than one of these, a fixed ordering prescription is used to decide along which bond the fire actually propagates. (The precise prescription will not be needed in what follows; the interested reader is referred to [9] for an example of such a prescription.) The collection of all bonds forming the fire path defines a spanning tree, rooted in

the sink, and growing towards the interior of the lattice  $\mathcal{L}$ .

This improved algorithm establishes a correspondence between the set of recurrent configurations on  $\mathcal{L}$  and the set of rooted spanning trees on  $\mathcal{L}^*$ . The precise mapping, although one-to-one, however depends on the prescription used. The prescription we will use below differs slightly from the one defined in [9], but is equally valid. The specific sandpile model we consider in the next sections is defined on the discrete upper half plane  $\mathcal{L} = \mathbb{Z}_{>} \times \mathbb{Z}$ , and has the massive discrete Laplacian as toppling matrix subjected to the two different boundary conditions “open” and “closed” on the boundary, which we take to be the line  $y = 1$ . The two toppling matrices are almost identical and differ only along the boundary. They both depend on a positive parameter  $t$ , controlling the rate at which sand is dissipated when sites topple. They read explicitly

$$\Delta_{ij}^{\text{op}} = \begin{cases} 4+t & \text{if } i=j, \\ -1 & \text{if } i \text{ and } j \text{ are n.n.}, \\ 0 & \text{otherwise,} \end{cases} \quad \Delta_{ij}^{\text{cl}} = \begin{cases} 4+t & \text{if } i=j \text{ are off boundary,} \\ 3+t & \text{if } i=j \text{ are on boundary,} \\ -1 & \text{if } i \text{ and } j \text{ are n.n.}, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

One easily sees that  $t$  grains of sand are dissipated (transferred to the sink) each time a site topples, or  $t+1$  if it is an open boundary site that topples. This model will be called the massive Abelian sandpile model (MASM) in reference to the massive discrete laplacian where  $\sqrt{t}$  plays the role of a mass. The usual, critical model originally defined in [2], is recovered at  $t = 0$ .

In terms of the graph  $\mathcal{L}^*$ , bulk sites and closed boundary sites have a  $t$ -fold connection to the sink, while open boundary sites have a  $(t+1)$ -fold connection to it. In addition, all nearest neighbour (n.n.) sites on  $\mathcal{L}$  are connected by a single bond.

An easy corollary of the above burning algorithm is that a site with height smaller or equal to the number of its neighbours on  $\mathcal{L}$  (3 or 4) is never burnt at time 1, and therefore catches fire from one of its neighbours and not from the sink. Conversely, a site with height strictly larger than  $\Delta_{ii}(t=0)$ —which we call a *supercritical height*—, or an open boundary site with  $h = 4$  is set afire by the sink. Supercritical height values are those which exist only when  $t$  is non zero.

According to the definition of  $\Delta$ ,  $t$  should take integer values. However the MASM correlations decay exponentially, with a correlation length that diverges only when  $t$  goes to 0, like  $1/\sqrt{t}$  [10, 6]. The large distance limit of the lattice correlations must therefore be accompanied by a small  $t$  limit, in such a way that their scaling limit  $\sqrt{t} = Ma \rightarrow 0, |i-j| = |z|/a \rightarrow \infty$  be well defined when  $a \rightarrow 0$ . So in practice, one expands the lattice MASM correlations in powers of  $t$ , and selects the dominant terms. These define correlators of a massive field theory, which, in this case, turns out to be a massive perturbation [6] of  $c = -2$  logarithmic conformal theory [11, 12, 13],

$$S = \frac{1}{\pi} \int (\partial\theta\bar{\partial}\bar{\theta} + M^2\theta\bar{\theta}/4), \quad (3)$$

where  $\theta, \bar{\theta}$  are anticommuting scalar fields.

In the course of the calculations, we will make an extensive use of the inverse toppling matrix  $\Delta_{ij}^{-1}$ . As is well-known, the inverse of  $\Delta$  on the upper half plane can be obtained in terms of the inverse massive Laplacian on the full plane  $\Delta^{-1}$ , via the image method. For  $i = (m_1, n_1)$  and  $j = (m_2, n_2)$ , the explicit formulae read

$$\begin{aligned} (\Delta^{\text{op}})_{ij}^{-1} &= \Delta_{ij}^{-1} - \Delta_{ij^*}^{-1} = \Delta_{ij}^{-1} - \Delta_{i^*j}^{-1}, & j^* &= (m_2, -n_2), \\ (\Delta^{\text{cl}})_{ij}^{-1} &= \Delta_{ij}^{-1} + \Delta_{ij^\vee}^{-1} = \Delta_{ij}^{-1} + \Delta_{i^\vee j}^{-1}, & j^\vee &= (m_2, 1-n_2). \end{aligned} \quad (4)$$

The horizontal translation invariance is preserved in both cases, so that the entries of the inverse matrices depend on  $|m_1 - m_2|, n_1$  and  $n_2$ . A short review on values of the inverse massive Laplacian on the plane can be found in [6].

The lattice open boundary condition is identified with the Dirichlet condition in the continuum ( $\theta = \bar{\theta} = 0$  on  $\mathbb{R}$ ), whereas the closed boundary condition corresponds to the Neumann condition

( $\partial\theta - \bar{\partial}\theta = \partial\tilde{\theta} - \bar{\partial}\tilde{\theta} = 0$  on  $\mathbb{R}$ ). The Lagrangian (3) then implies the following Green functions on the upper half plane

$$\langle \theta(z)\theta(w) \rangle = \langle \tilde{\theta}(z)\tilde{\theta}(w) \rangle = 0, \quad (5)$$

$$\langle \theta(z)\tilde{\theta}(w) \rangle_{\text{op}} = K_0(M|z-w|) - K_0(M|z-\bar{w}|), \quad (6)$$

$$\langle \theta(z)\tilde{\theta}(w) \rangle_{\text{cl}} = K_0(M|z-w|) + K_0(M|z-\bar{w}|), \quad (7)$$

where  $K_0$  is the modified Bessel function.

### 3 Unit height and supercritical height variables

Multisite probabilities for a number of sites to have height equal to 1 or supercritical height values ( $h_i > \Delta_{ii} - t$ ) is fairly easy if one uses the Bombay trick, a beautiful technique designed by Majumdar and Dhar [3]. It can be formulated in terms of height configurations or in terms of spanning trees. In this section, we will use it in terms of heights, the formulation with trees being a particular case of the general characterization given in the next section.

Suppose that we first want to compute the probability  $P[h_{i_0} = 1]$  that a certain site  $i_0$  has a height equal to 1. That probability is simply equal to the number of recurrent configurations with a height 1 at  $i_0$  divided by the total number of recurrent configurations, which we know equals  $\det \Delta$ .

The idea of [3] is to define a new sandpile model in which the height at  $i_0$  is always 1, and such that any recurrent configuration of this new model is in correspondence with a recurrent configuration of the original model where the height at  $i_0$  is 1. To freeze the height at  $i_0$  to the value 1, one simply reduces the diagonal entry of the toppling matrix to 1. So the toppling matrix  $\Delta'$  of the new model will have  $\Delta'_{i_0 i_0} = 1$ . Then in the new model, the site  $i_0$  will topple whenever its height exceeds 1, and each time it topples, it will lose a single grain which will go to one the neighbours or to the sink. Consequently,  $i_0$  will have a single connection, either to the sink or to one its neighbours in  $\mathcal{L}$ . Finally, the neighbours of  $i_0$  cannot have a height equal to 1 in a recurrent configuration, so that they assume only  $\Delta_{ii} - 1$  values. This can also be enforced in the new model by decreasing the diagonal entries of  $\Delta$  by 1 for those neighbours of  $i_0$  which are no longer connected to  $i_0$ . As the connections fix the off-diagonal part of the toppling matrix, and the height ranges fix its diagonal part, this will determine  $\Delta'$ .

Thus the number of recurrent configurations with a height 1 at  $i_0$  is equal to the total number of recurrent configurations of the new model, itself equal to the determinant of the new toppling matrix  $\Delta'$ . Setting  $\Delta' = \Delta + B^{(i_0)}$ , one obtains [3]

$$P[h_{i_0} = 1] = \frac{\det \Delta'}{\det \Delta} = \det(\mathbb{I} + \Delta^{-1} B^{(i_0)}), \quad (8)$$

where  $\Delta$  is the toppling matrix appropriate to the boundary condition one considers. Because the difference  $\Delta' - \Delta \equiv B^{(i_0)}$  is non zero only on sites around  $i_0$ , the previous formula reduces to the calculation of a finite determinant, even on an unbounded lattice  $\mathcal{L}$ .

On the discrete upper half plane, the defect matrix  $B^{(i_0)}$  depends on the location of  $i_0$ . If  $i_0$  is off the boundary, and if one keeps it connected to one of its four neighbours, then  $B^{(i_0)}$  is equal to

$$B^{(i_0)} = \begin{pmatrix} -3-t & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix} \quad (9)$$

on  $i_0$  (first label) and any three neighbours of  $i_0$ , and is identically zero elsewhere (if the only connection of  $i_0$  is to the sink, the  $B$  matrix is 5-by-5). In this case, the probability is given by a 4-by-4 determinant,

and depends on the distance  $m$  of  $i_0$  to the boundary. At the critical point ( $t = 0$ ) and for large values of  $m$ , it is equal to [14]

$$P[h_{i_0} = 1] = P_1 \left[ 1 \pm \frac{1}{4m^2} + \dots \right], \quad (10)$$

where the + (resp. -) sign refers to the open (resp. closed) boundary condition.  $P_1 = 2(\pi - 2)/\pi^3 = 0.0736$  is the probability that a site deep inside the lattice (equivalently, on the infinite plane) has height 1 in the critical ASM.

If  $i_0$  lies on the boundary, the matrix  $B^{(i_0)}$  depends on the boundary condition and takes one of the two forms

$$B_{\text{op}}^{(i_0)} = \begin{pmatrix} -3-t & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \quad B_{\text{cl}}^{(i_0)} = \begin{pmatrix} -2-t & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}. \quad (11)$$

The corresponding critical probability  $P[h_{i_0} = 1]$  is then a constant, which only depends on the type of boundary the site  $i_0$  is on [5],

$$P_1^{\text{op}} = \frac{9}{2} - \frac{42}{\pi} + \frac{320}{3\pi^2} - \frac{512}{9\pi^3}, \quad P_1^{\text{cl}} = \frac{3}{4} - \frac{2}{\pi}. \quad (12)$$

The one-site probability  $P[h_{i_0} = 1]$  can easily be computed for  $t \neq 0$ , but will not be needed in what follows.

The probability that a site be supercritical can be treated in a similar way, and is actually simpler. One now takes  $t \neq 0$ , since the probability does not make sense at  $t = 0$ .

Any site  $i_0$ , whatever its location and whatever the boundary condition, has  $t$  possible supercritical height values, namely  $h = \Delta_{i_0 i_0} - t + 1, \dots, \Delta_{i_0 i_0}$ . The probability that a site  $i_0$  has a fixed supercritical height  $h$  does not depend on  $h$ , because a recurrent configuration remains recurrent if one replaces a supercritical height at  $i_0$  by another one. Therefore one has

$$P[i_0 \text{ is supercritical}] = P[h_{i_0} > \Delta_{i_0 i_0} - t] = t P[h_{i_0} = h], \quad (13)$$

where  $h$  is any fixed supercritical value.

It is actually easier to compute the probability that  $i_0$  is not supercritical. To do that, one has to count the recurrent configurations with  $h_{i_0} \leq \Delta_{i_0 i_0} - t$  ( $= 4$ , or  $3$  on a closed boundary). In a new model defined by the new toppling matrix  $\Delta'_{ij} = \Delta_{ij} - t \delta_{i,i_0} \delta_{j,i_0}$  on the same lattice, all recurrent configurations have  $i_0$  not supercritical. One obtains

$$P[i_0 \text{ is supercritical}] = 1 - \frac{\det \Delta'}{\det \Delta} = t \Delta_{i_0 i_0}^{-1}, \quad (14)$$

and

$$P[h_{i_0} = h] = \Delta_{i_0 i_0}^{-1}. \quad (15)$$

The defect matrix method works here too, and is simpler because the appropriate matrix  $\bar{S}_{ij}^{(i_0)} = -t \delta_{i,i_0} \delta_{j,i_0}$  has rank 1. The corresponding one-site probability is then given by a 1-by-1 determinant.

Let us note that the probabilities (15) are well-defined for any strictly positive value of  $t$ , but behave badly in the critical limit  $t \rightarrow 0$ . For a closed boundary condition, they have a logarithmic singularity at  $t = 0$ . For an open boundary condition, they have a finite limit at  $t = 0$ , but which is not a probability in general:  $\Delta_{i_0 i_0}^{-1} = 1 - 2/\pi = 0.3634$  for  $i_0$  on the boundary, and then grows logarithmically with the distance of  $i_0$  to the boundary.

So instead we will consider the probability (14) for a site or a collection of sites to be supercritical without specifying the actual heights. As we will see below, that observable has well-defined correlations in the massive scaling limit, and corresponds to a field that vanishes in the critical limit.

After the one-site probabilities, multisite probabilities and correlations can be computed by the same method almost routinely. The observables we consider in this section are the two boundary

random variables  $\delta(h_i - 1)$  and  $\delta(i \text{ is supercritical})$  corresponding to the events “ $i$  has height 1” or “ $i$  is supercritical” for  $i$  a site on the boundary of the upper half plane. In order to get fields whose expectation value vanishes infinitely far from the boundary, one considers the random variables subtracted by their average value. Anticipating the scaling dimension 2 or 4 of the above random variables, we define their scaling fields by

$$\phi_1^{\text{op,cl}}(x) = \lim_{a \rightarrow 0} \frac{1}{a^2} [\delta(h_{x/a} - 1) - P_1], \quad \phi_{>}^{\text{op,cl}}(x) = \lim_{a \rightarrow 0} \frac{1}{a^{4,2}} [\delta(h_{x/a} \text{ is supercritical}) - \Delta_{\frac{x}{a}}^{-1}], \quad (16)$$

subjected to the scaling relations  $t = a^2 M^2$  and  $i\sqrt{t} = Mx$ .

To compute  $n$ -site probabilities, one simply inserts the proper defect matrices at the locations of the observables, so that the full defect matrix is a direct sum of  $n$  matrices  $B^{(i)}$  or  $\bar{S}^{(i)}$ . One should however remember that  $\bar{S}^{(i)}$  is not the defect matrix for  $i$  being supercritical but for  $i$  not being supercritical, the complementary event. The scaling limit of the latter gives rise to a field  $\phi_{\leq}$ , from which  $\phi_{>} = -\phi_{\leq}$  is recovered. The probability then reduces to the calculation of a finite determinant whose entries are combinations of entries of the inverse toppling matrix. As the scaling limit takes  $t$  to zero, one expands these entries in power series of  $t$ , keeping only the dominant term. From (16) the latter yields the field theoretic correlation of fields  $\phi_1$  and  $\phi_{>}$ , which are then identified with explicit fields of the Lagrangian theory (3). This is a main advantage of working with the massive theory that this identification is essentially unambiguous.

The simplest way to proceed to the identification of the boundary fields  $\phi_1$  and  $\phi_{>}$  is to use other lattice observables with already known field identifications. Examples of such observables are precisely the bulk version of the above two random variables. The corresponding bulk fields have been identified in [6] (see also [15] for a proof that these identifications are consistent with a broad class of multisite correlations),

$$\phi_1(z) = -P_1 [\partial\theta\bar{\partial}\tilde{\theta} + \bar{\partial}\theta\partial\tilde{\theta} + \frac{M^2}{2\pi} : \theta\tilde{\theta} :], \quad (17)$$

$$\phi_{>}(z) = \frac{M^2}{2\pi} : \theta\tilde{\theta} : . \quad (18)$$

One first computes the 2-point correlations involving one boundary observable and one bulk observable. From them, one may infer what the boundary fields must be, and then cross-check their form from other correlations.

We do not give much detail as the calculations are fairly straightforward, but simply illustrate the method in a simple case, namely the identification of  $\phi_{>}^{\text{op}}(x)$  on an open boundary.

We take two reference sites,  $i_0$  on the boundary and  $j_0$  in the bulk, far from the boundary. The probability that they both be non supercritical reduces to a rank 2 determinant

$$P[i_0, j_0 \text{ non supercrit.}] = \frac{\det[\Delta^{\text{op}} + \bar{S}^{(i_0)} + \bar{S}^{(j_0)}]}{\det \Delta^{\text{op}}} = \det \begin{pmatrix} 1 - t (\Delta^{\text{op}})_{i_0 i_0}^{-1} & -t (\Delta^{\text{op}})_{i_0 j_0}^{-1} \\ -t (\Delta^{\text{op}})_{j_0 i_0}^{-1} & 1 - t (\Delta^{\text{op}})_{j_0 j_0}^{-1} \end{pmatrix}. \quad (19)$$

Subtracting the product of disconnected probabilities obtained from (14), one has

$$P[i_0, j_0 \text{ non supercritical}]_{\text{connected}} = -t^2 (\Delta^{\text{op}})_{i_0 j_0}^{-2}. \quad (20)$$

If one chooses the two sites on a vertical line  $i_0 = (0, 1)$  and  $j_0 = (0, m)$ , then (see for instance Appendix A of [6])

$$(\Delta^{\text{op}})_{i_0 j_0}^{-1} = \Delta_{(0,1)(0,m)}^{-1} - \Delta_{(0,-1)(0,m)}^{-1} = -\frac{\sqrt{t}}{\pi} K_0'(m\sqrt{t}) + \dots \quad (21)$$

where the dots stand for subdominant terms in  $t$ . The dominant term in the connected 2-site probability is thus

$$P[i_0, j_0 \text{ non supercritical}]_{\text{connected}} = -\frac{t^3}{\pi^2} K_0'^2(m\sqrt{t}) + \dots \quad (22)$$

Using the scaling relation (16), valid also for bulk fields (with a power  $a^{-2}$  for  $\phi_1$  and  $\phi_{\leq}, \phi_{>}$ ), one finds the boundary/bulk 2-point function

$$\langle \phi_{\leq}^{\text{op}}(x) \phi_{\leq}(x + iy) \rangle = -\frac{M^6}{\pi^2} K_0'^2(My). \quad (23)$$

From the explicit form of  $\phi_{\leq}(z) = -\frac{M^2}{2\pi} : \theta \tilde{\theta} :$  given above, one eventually arrives at

$$\langle \phi_{\leq}^{\text{op}}(x) : \theta \tilde{\theta} : (x + iy) \rangle = \frac{2M^4}{\pi} K_0'^2(My). \quad (24)$$

Using the Green function (6) on the upper half plane with an open boundary, one sees that the only possible field assignment is  $\phi_{\leq}^{\text{op}}(x) = \frac{2M^2}{\pi} : \partial \theta \partial \tilde{\theta} :$ , and therefore  $\phi_{>}^{\text{op}}(x) = -\frac{2M^2}{\pi} : \partial \theta \partial \tilde{\theta} :$ .

Proceeding in the same way for the other observables for the two boundary conditions, we find the following scaling fields

$$\phi_1^{\text{op}} = \left( \frac{6}{\pi} - \frac{160}{3\pi^2} + \frac{1024}{9\pi^3} \right) : \partial \theta \partial \tilde{\theta} :, \quad \phi_{>}^{\text{op}} = -\frac{2M^2}{\pi} : \partial \theta \partial \tilde{\theta} :, \quad (25)$$

$$\phi_1^{\text{cl}} = -\frac{8}{\pi} \left( \frac{3}{4} - \frac{2}{\pi} \right) [ : \partial \theta \partial \tilde{\theta} : + \frac{1}{16} M^2 : \theta \tilde{\theta} : ], \quad \phi_{>}^{\text{cl}} = \frac{M^2}{2\pi} : \theta \tilde{\theta} : . \quad (26)$$

These field identifications have been checked to be consistent with many multisite probabilities: 2-site and 3-site boundary/boundary probabilities as well as 2-site and 3-site mixed boundary/bulk probabilities.

The massless limit is simply given by the limit  $t \rightarrow 0$  in the MASM and  $M \rightarrow 0$  in the field theory. In this limit, the fields  $\phi_{>}^{\text{op}}$  and  $\phi_{>}^{\text{cl}}$  are obviously null and the unit height fields for the two boundaries are identical up to a numerical factor. One may note that the latter are proportional to the holomorphic stress-energy tensor, and being descendants of the identity, they belong to a chiral representation  $\mathcal{V}_0$  [13]. This is consistent with the fact that the only fields living on an open boundary are fields of  $\mathcal{V}_0$ , and that those living on a closed boundary belong to an  $\mathcal{R}_0$  representation, which contains  $\mathcal{V}_0$  as subrepresentation [16, 17].

## 4 Spanning tree representation of recurrent configurations

For the other height variables, the situation is not as easy. Although having a height 1 or a height 2 at a given site does not seem to make much difference, the counting of the corresponding recurrent configurations is technically much more complicated for a height 2 (or 3 or 4) than for a height 1. The defect matrix method no longer works<sup>1</sup>, and the only practical alternative seems to be the use of spanning trees. One then clearly sees the difference: in terms of spanning trees, a height 1 is characterized by a local property of the tree around the reference site, while the other heights are characterized by non local properties of trees [4].

As mentioned earlier, the rooted spanning tree is defined on  $\mathcal{L}^*$ , the lattice  $\mathcal{L}$  augmented by the sink site, at which the tree is rooted. All sites  $i$  of  $\mathcal{L}$  are connected by  $\Delta_{ij}$  bonds to other sites:  $-\Delta_{ij}$  bond(s) connecting  $i$  to  $j$  and  $\sum_{j \in \mathcal{L}} \Delta_{ij}$  bond(s) connecting  $i$  to the root. With these definitions, the Kirchhoff theorem asserts that the number of rooted spanning trees on the graph  $\mathcal{L}^*$  defined by the matrix  $\Delta$  is equal to  $\mathcal{N} = \det \Delta$ , precisely the number of recurrent configurations.

As a rooted spanning tree is a connected graph containing no loop, every site  $i$  is connected to the root by a unique path. A site  $j$  is said to be a predecessor of the site  $i$  if the path from  $j$  to the root passes through  $i$ , or equivalently, if  $j$  lies on a branch growing from  $i$ . A site  $i$  which has no predecessor is called a leaf (the end of a branch).

Priezzhev first and then Ivashkevich used the correspondence between recurrent configurations and spanning trees to compute respectively the 1-site probabilities in the plane, and the 1- and 2-site

<sup>1</sup>Except a height 4 on an open boundary, which can be handled like a supercritical height.

probabilities on the boundary, open or closed, of the upper half plane. For the 1-site probabilities at  $i_0$ , they decomposed the set of recurrent configurations into subsets  $S_a$ , where  $S_a$  contain the recurrent configurations which remain allowed for any heights  $h_{i_0} \geq a$  and which are forbidden otherwise. These subsets  $S_a$  can be characterized in terms of rooted spanning trees and their cardinal can be computed using classical results in graph theory. As we will see, they decomposed local tree diagrams as sums of non local diagrams. This system is invertible without further input for  $i_0$  on a boundary, but is not for  $i_0$  in the bulk. So the calculation of probabilities for sites in the bulk is more complicated.

For the 2-site probabilities, Ivashkevich used a similar decomposition of the recurrent configurations into subsets  $S_{ab}$ . This decomposition however raises certain questions, and will not be used here. Instead we set up a particular one-to-one map between recurrent configurations and the rooted spanning trees, based on the burning algorithm. In the case of single site probabilities, the mapping yields the same characterization in terms of trees as the  $S_a$  decomposition but is much more transparent in the case of multisite probabilities.

The burning algorithm, as we described it Section 2 complemented with an ordering prescription, establishes a one-to-one mapping but with no clear correspondence between the height values at the reference sites and the bond arrangements of the trees around those sites (except for supercritical heights which are directly connected to the root of the tree). For example, depending on the recurrent configuration, a site with a height 4 can be a leaf on the tree or can support 1, 2 or 3 branches. To avoid this problem, we proceed in two steps as follows, assuming that none of the reference sites is supercritical.

First, we run the burning algorithm and let the fire propagate through the lattice until no more site is burnable but *preventing the reference sites from burning*. When this is done, one is left with a sublattice  $\mathcal{L}_b$  of burnt sites and a complementary sublattice  $\mathcal{L}_u$  of unburnt sites. The algorithm, using for example the ordering prescription of [9], yields at this stage the part of the spanning tree on  $\mathcal{L}_b$ . The other part  $\mathcal{L}_u$  is eventually burnable too and is actually burnt in the second step. By definition, none of the reference sites is burnt yet, and a certain number of them, at least one, are burnable. Those which are burnable are burnt simultaneously, and trigger the fire propagation through  $\mathcal{L}_u$ , thereby completing the spanning tree to the whole lattice. So the complete tree is made of two pieces, a subtree  $\mathcal{T}_b$  on  $\mathcal{L}_b$ , and another  $\mathcal{T}_u$  on  $\mathcal{L}_u$ . The subtree  $\mathcal{T}_u$  itself may have several roots which are among the reference sites those which were burnable and which set fire to the whole of  $\mathcal{L}_u$ . It is at those sites that  $\mathcal{T}_u$  is grafted to  $\mathcal{T}_b$  to make the full tree  $\mathcal{T}$ . As we will see, only the shape of the unburnt sublattice  $\mathcal{L}_u$  is used to characterize the height of the reference sites.

This slightly modified burning algorithm establishes a well defined correspondence between spanning trees and heights of the reference sites in the critical as well as in the massive Abelian sandpile. Let us see how this works for the single-site probabilities, and how it allows to compute the 2-site height correlations where one the two heights is equal to 1 or is supercritical. At this stage we will be able to identify the boundary fields corresponding to all heights. In the following section, we will compute other 2-site and 3-site correlations to confirm these identifications.

Let us consider a configuration of the MASM on a square lattice  $\mathcal{L}$ , and let us focus on a fixed site  $i_0$ , the reference site. We will take  $\mathcal{L}$  to be the upper half plane, but what follows applies to any sort of portion of  $\mathbb{Z}^2$ , bounded or unbounded.

- If the height at  $i_0$  is supercritical, then  $i_0$  is set afire by the root (is burnt at time 1). Thus all trees corresponding to those configurations have a bond connecting the root and  $i_0$ . The probability that  $i_0$  be supercritical is thus

$$P[i_0 \text{ is supercritical}] = \frac{\mathcal{N}_{\star, i_0}}{\mathcal{N}}, \quad (27)$$

where  $\mathcal{N} = \det \Delta$  and  $\mathcal{N}_{\star, i_0}$  is the number of different spanning trees which “use” one of the  $t$  bonds between the root and  $i_0$ . One way to compute  $\mathcal{N}_{\star, i_0}$  is to modify the toppling matrix by

removing the connections between  $i_0$  and its nearest neighbors on  $\mathcal{L}$  so that  $i_0$  has connections only to the root. Then  $\mathcal{N}_{\star, i_0} = \det \Delta'$  with  $\Delta' = \Delta + S^{(i_0)}$  and the finite-dimensional defect matrix given by  $S_{i_0 i_0}^{(i_0)} = t - \Delta_{i_0 i_0}$ ,  $S_{i_0 i_\ell}^{(i_0)} = S_{i_\ell i_0}^{(i_0)} = 1$  for  $i_\ell$  the nearest neighbors of  $i_0$ , and zero elsewhere. A simpler way is however to compute  $\mathcal{N} - \mathcal{N}_{\star, i_0}$ , the number of trees which do not use the bonds between  $i_0$  and the root. This can be done by removing precisely these bonds and leads to the 1-dimensional defect matrix  $\bar{S}^{(i_0)}$  of the previous section, with the result given in (14).

The same arguments apply to a height 4 at an open boundary site. The only difference is that such sites have  $t + 1$  connections to the root, so that

$$P_4 = \frac{1}{t} \frac{\mathcal{N}_{\star, i_0}}{\mathcal{N}} = \Delta_{i_0 i_0}^{-1}, \quad \text{for } i_0 \text{ on open boundary.} \quad (28)$$

- If the height at  $i_0$  is less or equal to the number of nearest neighbours on  $\mathcal{L}$ , we use the burning algorithm to define a partition  $\mathcal{L} = \mathcal{L}_b \cup \mathcal{L}_u$  as explained above. As one looks here at 1-site probabilities, there is only one reference site, so that in the sublattice  $\mathcal{L}_u = \mathcal{L}_{i_0}$ ,  $i_0$  is the only burnable site after all sites of  $\mathcal{L}_b$  have been burnt. It is therefore the root of the subtree  $\mathcal{T}_u = \mathcal{T}_{i_0}$ . The height at  $i_0$  can now be related to the properties of the subtree  $\mathcal{T}_{i_0}$ .

If  $\mathcal{L}_{i_0}$  contains only the site  $i_0$ , then  $h_{i_0}$  can take any of the values  $1, 2, \dots, n_{i_0}$ , where  $n_{i_0} = 3$  or  $4$  is the number of nearest neighbours of  $i_0$ . The full tree  $\mathcal{T}$  is simply obtained by connecting  $i_0$  to  $\mathcal{T}_b$  through one of the  $n_{i_0}$  bonds connecting  $i_0$  to its nearest neighbours, so that  $i_0$  is a leaf on  $\mathcal{T}$  ( $i_0$  must be connected to a nearest neighbour in  $\mathcal{L}$  and not to the root, since it catches fire from one of them). If one denotes by  $\mathcal{N}_1$  the number of spanning trees on  $\mathcal{L}$  where  $i_0$  is a leaf grown on one its neighbours, then

$$P_1 = \frac{\mathcal{N}_1}{n_{i_0} \mathcal{N}}. \quad (29)$$

If  $\mathcal{L}_{i_0}$  contains one nearest neighbor of the site  $i_0$ , the value of the height at  $i_0$  must be in the set  $\{2, \dots, n_{i_0}\}$  (it cannot be 1 since otherwise  $i_0$  would not be burnable and could not set  $\mathcal{L}_{i_0}$  afire). There are now  $n_{i_0} - 1$  possibilities to connect  $\mathcal{T}_{i_0}$  to  $\mathcal{T}_b$ , one for each nearest neighbour of  $i_0$  in  $\mathcal{L}_b$ . They correspond to the height values compatible with the burning algorithm and thus the height 2 probability reads

$$P_2 = \frac{\mathcal{N}_1}{n_{i_0} \mathcal{N}} + \frac{\mathcal{N}_2}{(n_{i_0} - 1) \mathcal{N}}, \quad (30)$$

where  $\mathcal{N}_2$  is the number of spanning trees where  $i_0$  has exactly one predecessor among its nearest neighbours.

The higher height probabilities can be determined by the same arguments,

$$P_k = P_{k-1} + \frac{\mathcal{N}_k}{(n_{i_0} - k + 1) \mathcal{N}}, \quad P_0 = 0, \quad k = 1, \dots, n_{i_0}, \quad (31)$$

where  $\mathcal{N}_k$  is the number of spanning trees  $\mathcal{T}$  on  $\mathcal{L}$  in which  $i_0$  has exactly  $k - 1$  predecessors among his neighbours.

One sees that the computation of the various 1-site probabilities requires the calculation of the numbers  $\mathcal{N}_k$ . If the lattice  $\mathcal{L}$  is the discrete upper half plane, and for  $i_0$  a site on the boundary, the Figure 1 describes the types of trees which contribute to the different  $\mathcal{N}_k$ 's.

These diagrams represent the restriction of trees to four sites, namely the reference site  $i_0$  and its three neighbours (the dashed line represents the border of  $\mathcal{L}$ , pictured as the lower half plane !). The arrows indicate the direction of the path towards the root (opposite to the fire propagation line). The black dots are the nearest neighbours which are predecessors of  $i_0$ , the white dots are those which are not. Those diagrams labelled by identical greek letters contribute the same amount to the corresponding  $\mathcal{N}_k$ . The tilded letters refer to diagrams which are the mirror images of the diagrams with untilded letters.

$$\begin{aligned}
\mathcal{N}_1 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \alpha_1 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \alpha_2 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \alpha_3 \end{array} \\
\mathcal{N}_2 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \beta_1 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \beta_2 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\beta}_1 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\beta}_2 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \gamma_1 \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \gamma_2 \end{array} \\
\mathcal{N}_3 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \delta \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \epsilon \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\epsilon} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \phi_a \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \phi_b \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\phi}_a \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\phi}_b \end{array}
\end{aligned}$$

Figure 1: Non local diagrams contributing to the 1-site probabilities.

The diagrams in Figure 1 represent non local constraints on the compatible trees. The presence of an arrow between  $i_0$  and one or more neighbours poses no computational problem, because it only means that the tree has to use specific bonds, and the counting of those trees can be achieved by modifying *locally* the toppling matrix by a finite rank defect matrix. But for a neighbour to be a predecessor or not is a non local property, and enumerating the relevant trees is trickier. For the trees contributing to  $\mathcal{N}_3$  for instance, one sees that a nearest neighbour, call it  $i_1$ , of  $i_0$  can be a predecessor of  $i_0$  because  $i_1$  is connected to  $i_0$  through the nearest neighbour bond ( $i_1$  catches fire directly from  $i_0$ , like in  $\delta, \epsilon$  and  $\tilde{\epsilon}$ ), or through a long path around the lattice ( $i_1$  catches fire after a long sequence of burnings, long after  $i_0$  got burnt, like in  $\phi$  and  $\tilde{\phi}$ ).

Various classes of non local diagrams in Figure 1, with their constraints on predecessorships, sum up to local diagrams, where these constraint are relaxed and only local arrow configurations are imposed. As the local diagrams are easily calculable by toppling matrix adjustments, this yields linear relations among the non local contributions. For a generic position of the reference site  $i_0$ , the linear system is not invertible and is not sufficient to compute the non local contributions. The crucial observation made by Ivashkevich in [5] was to note that it actually becomes invertible if  $i_0$  is on a boundary, which allows to reduce the non local contributions to local calculations (this statement will have to be qualified in the case of 2-site insertions). Indeed the relations read explicitly [5] (as we consider the heights 1, 2 and 3 only, we do not need to distinguish the diagrams for closed and open boundaries, which only make a difference through the proper toppling matrix to be used in the explicit local computations; at the same time, that also makes the number of equations smaller)

$$\begin{aligned}
Q_0 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \alpha \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \alpha \end{array} \\
Q_1 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \beta \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \beta \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \phi \end{array} & \tilde{Q}_1 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\beta} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\beta} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\phi} \end{array} \\
Q_2 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \beta \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \beta \end{array} & \tilde{Q}_2 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\beta} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\beta} \end{array} \\
Q_3 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \gamma \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \gamma \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \phi \end{array} & \tilde{Q}_3 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\gamma} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\gamma} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\phi} \end{array} \\
Q_4 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \delta \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \delta \end{array} \\
Q_5 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \epsilon \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \epsilon \end{array} & \tilde{Q}_5 &= \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\epsilon} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \\ \tilde{\epsilon} \end{array}
\end{aligned} \tag{32}$$

The four equations on the right are clearly obtained from the corresponding four on the left by a mirror symmetry about the reference site  $i_0$ . As the toppling matrix is invariant under that symmetry,

the three tilded diagrams contribute the same amount as the untilded ones,  $\tilde{\beta} = \beta, \tilde{\varepsilon} = \varepsilon, \tilde{\phi} = \phi$ , as do the tilded and the untilded local diagrams. Thus the four equations on the right are redundant, and one is left with the linear system on the left. This system is manifestly invertible for the six non local contributions, noted  $\alpha, \beta, \gamma, \delta, \varepsilon$  and  $\phi$ .

Let us note that the non local diagrams  $\alpha, \beta, \delta$  and  $\varepsilon$  turn out to be entirely local, because the arrow configurations make the predecessorship properties redundant. As the height 1 probability  $P_1$  is given solely in terms of  $\alpha$ , its computation is purely local. This remains true for any multisite height 1 probabilities and for arbitrary positions, in the bulk or on boundaries.

The Kirchoff theorem allows the local diagrams to be computed by the defect matrix method. The presence resp. the absence of an arrow from  $i$  to  $j$  means that one counts all trees which contain resp. do not contain that oriented bond. One defines a new toppling matrix  $\Delta'$  by setting to  $-\epsilon$  the  $i, j$  entry if the  $i \rightarrow j$  bond is to be used in the tree, and to 0 if that bond is not to be used; moreover the diagonal entries of  $\Delta'_{ii} = \Delta_{ii}$  should remain equal to the number of bonds going out from  $i$ . Then if  $n$  bonds are to be used, the determinant of  $\Delta'$  will contain a highest degree term  $\epsilon^n$  whose coefficient is the number of trees which precisely use the given  $n$  bonds in the prescribed direction (see for instance [4]).

Writing as before  $\Delta' = \Delta + B$ , one finds for  $Q_4$  for example

$$B = \begin{pmatrix} -3 + \epsilon & 1 & 1 - \epsilon & 1 \\ 1 - \epsilon & -1 + \epsilon & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 - \epsilon & 0 & 0 & -1 + \epsilon \end{pmatrix}, \quad (33)$$

and, in the critical limit,

$$\frac{Q_4}{\mathcal{N}} = \frac{\delta}{\mathcal{N}} = \lim_{\epsilon \rightarrow \infty} \frac{1}{\epsilon^3} \det[\mathbb{I} + \Delta^{-1}B] = \begin{cases} \frac{1}{\pi} - \frac{1}{4} & \text{on closed boundary,} \\ \frac{(3\pi-8)^3}{9\pi^3} & \text{on open boundary.} \end{cases} \quad (34)$$

The calculation of the other five local diagrams and then the inversion of the linear system yields the values of  $\mathcal{N}_1, \mathcal{N}_2$  and  $\mathcal{N}_3$ , and in turn of  $P_1, P_2$  and  $P_3$ . In the critical limit, one recovers the numbers given in [5].

In order to identify the height boundary fields, we need 2-site correlations involving the boundary heights 2 and 3. Again the simplest is to look at the correlations of a boundary height 2 or 3 with a known boundary variable, namely a height 1 or a supercritical height value. The advantage is that the latters are already known from the previous section, but more importantly, they correspond to local defect matrix insertions. This makes the above formalism, useful to compute 1-site probabilities, essentially valid.

Because one can force a site  $i_0$  to have height 1 or to be supercritical by modifying the toppling matrix by  $\Delta \rightarrow \Delta(i_0) = \Delta + B^{(i_0)}$  or  $\Delta + S^{(i_0)}$  (or better  $\Delta + \tilde{S}^{(i_0)}$ ), the 2-site probabilities  $P[h_{i_0} = 1 \text{ or supercr.}, h_{j_0} = 2 \text{ or } 3]$  can be viewed as 1-site probabilities for the height at  $j_0$  but with the toppling matrix  $\Delta(i_0)$  to account for the constraint at  $i_0$ . Then the above method remains completely valid provided we replace  $\Delta$  by the appropriate  $\Delta(i_0)$ , itself to be modified by matrices like in (33) in order to compute local diagrams<sup>2</sup>. If one does that and use  $\Delta(i_0)$  as the normalizing toppling matrix, one is actually computing the conditional probability for having a 2 or a 3 at  $j_0$  conditioned on having a height 1 or a supercritical height at  $i_0$ . To get the joint probabilities, one simply multiplies the final answers by  $P[h_{i_0} = 1]$  or  $P[h_{i_0} = \text{supercr.}]$ .

The non local diagrams contributing to the numbers  $\mathcal{N}_k$  remain as in Figure 1. However the tilded and the untilded diagrams no longer contribute equally because the mirror image about  $j_0$  spoils the constraint at  $i_0$ , and does not leave the toppling matrix  $\Delta(i_0)$  invariant. Therefore it is the full system

---

<sup>2</sup>To keep the decompositions of the  $\mathcal{N}_k$  in terms of the non local diagrams as in Figure 1, the site  $i_0$  should not be too close to  $j_0$ .

(32) that needs be solved. It is overdetermined as it involves 10 equations for only 9 unknowns, but the number of equations is reduced by one due to the following identity

$$\tilde{P}(1) + P(2) + P(3) = P(1) + \tilde{P}(2) + \tilde{P}(3), \quad (35)$$

satisfied for all values of  $t$  as a simple consequence of the fact that the inverse of  $\Delta(i_0)$  satisfies a discrete Poisson equation. The procedure is otherwise identical to that for the 1-site probabilities.

For the open boundary, the boundary joint probabilities of a site with height 2, 3 or 4 and a site with height 1 or with a supercritical height all have the same form as two unit heights on the boundary. It means the same field identification up to a numerical factor:

$$\phi_2^{\text{op}} = \left( -\frac{18}{\pi} + \frac{400}{3\pi^2} - \frac{2048}{9\pi^3} \right) : \partial\theta\partial\tilde{\theta} :, \quad (36)$$

$$\phi_3^{\text{op}} = \left( \frac{14}{\pi} - \frac{80}{\pi^2} + \frac{1024}{9\pi^3} \right) : \partial\theta\partial\tilde{\theta} :, \quad (37)$$

$$\phi_4^{\text{op}} = -\frac{2}{\pi} : \partial\theta\partial\tilde{\theta} :. \quad (38)$$

The last line can most easily obtained from (28), which implies  $\phi_4^{\text{op}} = \frac{1}{M^2} \phi_>^{\text{op}}$ .

For the closed boundary, the correlations involving a height 2 or 3 have a more complicated structure. For example, one finds (with  $m = |i_0 - j_0|$ )

$$P[h_{i_0} > 3, h_{j_0} = 2] = \frac{t^2}{\pi^2} \left[ \frac{2}{\pi} K_0^2(m\sqrt{t}) - \left(3 - \frac{12}{\pi}\right) K_0'^2(m\sqrt{t}) - \frac{1}{2} K_0(m\sqrt{t}) K_0''(m\sqrt{t}) \right] + \dots \quad (39)$$

$$P[h_{i_0} > 3, h_{j_0} = 3] = \frac{t^2}{\pi^2} \left[ \frac{1}{4} K_0^2(m\sqrt{t}) - \frac{4}{\pi} K_0'^2(m\sqrt{t}) + \frac{1}{2} K_0(m\sqrt{t}) K_0''(m\sqrt{t}) \right] + \dots \quad (40)$$

These results and the corresponding ones for  $h_{i_0} = 1$  are compatible with the following field assignments for the height 2 and 3

$$\phi_2^{\text{cl}} = \left( \frac{6}{\pi} - \frac{24}{\pi^2} \right) : \partial\theta\partial\tilde{\theta} : + \frac{1}{2\pi} : \theta\partial\partial\tilde{\theta} : + \left( \frac{1}{8\pi} - \frac{1}{\pi^2} \right) M^2 : \theta\tilde{\theta} :, \quad (41)$$

$$\phi_3^{\text{cl}} = \frac{8}{\pi^2} : \partial\theta\partial\tilde{\theta} : - \frac{1}{2\pi} : \theta\partial\partial\tilde{\theta} : - \frac{1}{4\pi} M^2 : \theta\tilde{\theta} :. \quad (42)$$

This identification is in this case not unique, since the field theory is invariant under  $\theta \rightarrow \tilde{\theta}$  and  $\tilde{\theta} \rightarrow -\theta$ . One could in particular change  $: \theta(\partial\partial\tilde{\theta}) :$  for  $: (\partial\partial\theta)\tilde{\theta} :$ , which are different fields since their correlation contains a logarithm while their self-correlations do not. If one requires that the sum  $\phi_1^{\text{cl}} + \phi_2^{\text{cl}} + \phi_3^{\text{cl}}$  be zero in the massless limit, then the choice for one the two fields must be made for both  $\phi_2^{\text{cl}}$  and  $\phi_3^{\text{cl}}$ . The correlations computed in the next section confirm this. Note that the sum of fields  $\phi_1^{\text{cl}} + \phi_2^{\text{cl}} + \phi_3^{\text{cl}} + \phi_>^{\text{cl}}$  vanishes identically. The similar sum in the open case,  $\phi_1^{\text{op}} + \phi_2^{\text{op}} + \phi_3^{\text{op}} + \phi_4^{\text{op}} + \phi_>^{\text{op}}$ , vanishes at the critical point only, because the dimension of  $\phi_>^{\text{op}}$  does not match the dimension of the universal terms of the other fields.

At this stage, all boundary height fields for the massive Abelian sandpile model have been determined. To have more checks on the field identifications, we compute in the following section, all 2-site and some 3-site height correlations.

## 5 Higher boundary correlations

In the previous section we have seen that the multisite probabilities where only one reference site  $i_0$  has a height value in  $\{2, 3\}$  can be computed from the diagrams listed in Figure 1 by using a toppling matrix properly decorated by defect matrices to account for height constraints (height 1 or supercritical) at the other sites. The calculation of multisite probabilities where two reference sites  $i_0$  and  $j_0$  have a height value in  $\{2, 3\}$  leads naturally to pairs of such diagrams, one at  $i_0$ , the other at  $j_0$ . However the situation becomes technically more complex because sites in the diagram at  $i_0$  can be predecessors of  $j_0$  and/or the other way round. So the topology of the spanning trees can be more complicated and their counting more difficult.

Let us first consider the 2-site probabilities  $P_{ab} = P[h_{i_0} = a, h_{j_0} = b]$  for  $a, b$  in  $\{2, 3\}$  and where  $i_0$  and  $j_0$  are on the boundary of the upper half plane. We start the burning algorithm as explained in Section 4 without ever burning the sites  $i_0$  and  $j_0$ , and until no other sites than those two are burnable. This yields a sublattice  $\mathcal{L}_u = \mathcal{L}_{i_0 j_0}$  of unburnt sites, which subsequently catches fire either from  $i_0$  or from  $j_0$ , or from both if they are both burnable. In turn the fire propagation on  $\mathcal{L}_{i_0 j_0}$  defines a subtree  $\mathcal{T}_u = \mathcal{T}_{i_0 j_0}$ , rooted at  $i_0$ , or at  $j_0$ , or at both sites. The full tree  $\mathcal{T}$  is made up of the subtree  $\mathcal{T}_b$  living on the sublattice of burnt sites, to which  $\mathcal{T}_{i_0 j_0}$  is grafted at  $i_0$  and/or  $j_0$ .

The restriction of any tree to the neighbourhood of a reference site looks like one of the non local diagrams shown in Figure 1. So one can visualize the restriction to the two neighbourhoods by a pair of such diagrams. Using the same labelling as in Figure 1, we will denote the pairs of diagrams by pairs of greek letters (with indices), the first one for the diagram around  $i_0$ , the other for the diagram at  $j_0$ . In an obvious notation, a pair of greek letters belongs to a certain set  $\mathcal{N}_k \times \mathcal{N}_l$ . As we did in Section 4 for the one-site probabilities, we have to compute which probabilities  $P_{ab}$  a pair of diagrams contributes to.

For 1-site probabilities, we know from Section 4 that the diagrams in  $\mathcal{N}_k$  contribute equally to the probabilities  $P_a$  for  $k \leq a \leq 3$ . Indeed the three diagrams  $\alpha_1, \alpha_2, \alpha_3$  of  $\mathcal{N}_1$  are obtained from each other by changing the arrow around the reference site. The change converts a tree which is compatible with a diagram  $\alpha_i$  into a tree which is compatible with another diagram  $\alpha_j$ , and this fact shows that the number of trees compatible with a diagram  $\alpha_i$  does not depend on  $i$ , namely  $\alpha_1 = \alpha_2 = \alpha_3$  or  $\mathcal{N}_1 = 3\alpha$ . As the position of the arrow determines univoquely the height value, the three probabilities  $P_1, P_2, P_3$  get an equal contribution  $\mathcal{N}_1/3\mathcal{N}$  from the diagrams in  $\mathcal{N}_1$ . The same is true of the six diagrams in  $\mathcal{N}_2$ . They come in pairs  $(\beta_1, \beta_2), (\tilde{\beta}_1, \tilde{\beta}_2), (\gamma_1, \gamma_2)$ , where the diagrams within a pair are related by changing the direction of the arrow coming out from the reference site. The same arguments as above show that  $\beta_1 = \beta_2, \tilde{\beta}_1 = \tilde{\beta}_2, \gamma_1 = \gamma_2$ , and that  $P_2, P_3$  receive an identical contribution  $\beta + \tilde{\beta} + \gamma = \mathcal{N}_2/2\mathcal{N}$  from the diagrams in  $\mathcal{N}_2$ . For  $\mathcal{N}_3$ , each diagram is on its own and contributes to  $P_3$ .

In the case of 2-site probabilities, the same arguments would show that the diagrams in  $\mathcal{N}_k \times \mathcal{N}_l$  contribute equally to the probabilities  $P_{ab}$  for  $k \leq a \leq 3, l \leq b \leq 3$ , provided one can prove that changing the direction of an arrow in the way recalled above in either diagram, or in both diagrams, turns a compatible tree into a compatible tree of the same class. Because the two diagrams can now be linked by fire paths, this is no longer guaranteed, and actually fails in a few cases, pictured in Figure 2.

On the first line of Figure 2, one sees for instance that the diagram denoted by  $A_3$  is a pair  $\gamma_2\beta_1$ . It is linked in such a way that when one changes the arrow in  $\beta_1$ , one obtains a well-defined tree (noted  $A_2$ ) compatible with a pair  $\tilde{\phi}_b\beta_2$ . If one changes in  $A_3$  the arrow of  $\gamma_2$ , one obtains the diagram  $A_1$ , of the type  $\gamma_1\phi_a$ . Changing the arrow of  $\gamma_2$  and of  $\beta_1$  introduces a loop, and so cannot contribute to a 2-site probability. The trees compatible with the diagrams  $A_1, A_2$  and  $A_3$  are related by local changes of arrow, but belong to different classes, namely  $\mathcal{N}_2 \times \mathcal{N}_3, \mathcal{N}_3 \times \mathcal{N}_2$  and  $\mathcal{N}_2 \times \mathcal{N}_2$ . There should normally be a fourth diagram, in  $\mathcal{N}_3 \times \mathcal{N}_3$ , but which does not exist as a tree.

It is not difficult to see that the misbehaviours with respect to arrow changes can only be of the type shown by the triplet  $(A_1, A_2, A_3)$ . When the two diagrams are tied in a special way by the fire paths, one change of arrow in a diagram in  $\mathcal{N}_2 \times \mathcal{N}_2$  sends it to a diagram in  $\mathcal{N}_2 \times \mathcal{N}_3$  or  $\mathcal{N}_3 \times \mathcal{N}_2$ , and two arrow changes introduce a loop.

Figure 2 shows four triplets of diagrams where this peculiar behaviour occurs. Diagrams labelled by the same capital letter are in equal number, since the numbers of compatible trees are equal. The twelve diagrams shown in Figure 2 and the mirror diagrams (not shown in Figure 2), obtained by exchanging the diagram at  $i_0$  with the reflected one at  $j_0$  and vice-versa, make the complete list of misbehaved diagrams. We will denote the mirror diagrams with tildes.

The two-step burning algorithm allows to determine which probability each diagram contributes to. In the diagram  $A_1$  for instance, the subtree  $\mathcal{T}_{i_0 j_0}$  catches fire from the eastern neighbour of  $j_0$ . It is thus

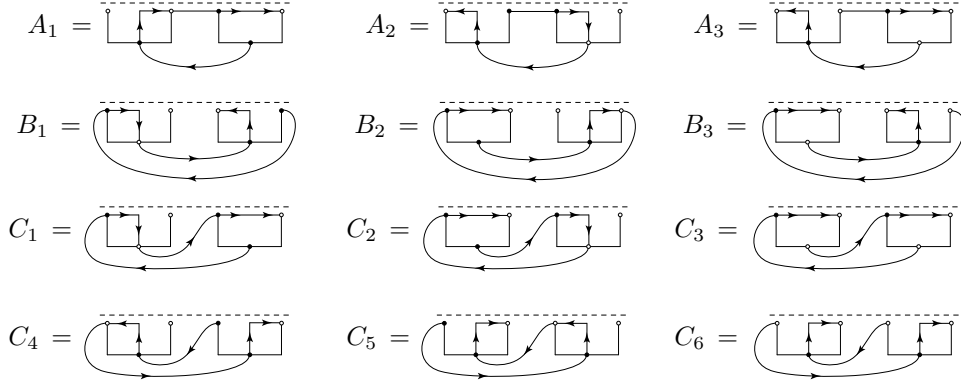


Figure 2: Non local diagrams representing spanning trees which have an anomalous behaviour under a local change of arrow around  $i_0$  and/or  $j_0$ . The mirror diagrams must be added to have the complete list of such diagrams.

burnable at a time where only one of its neighbours is burnt, and so must have a height 3. The other reference site  $i_0$  is not burnable at the time  $\mathcal{T}_{i_0 j_0}$  catches fire despite the fact that its western neighbour was burnt, which implies that its height is at most 2. When  $i_0$  is burnable, it has two burnt neighbours and one southern unburnt neighbour, meaning that its height must be 2. Thus  $A_1$  contributes to  $P_{23}$ . One finds similarly that the first column in Figure 2 contribute to  $P_{23}$ , the second column to  $P_{32}$ , and the last column to  $P_{33}$ .

We define  $[\mathcal{N}_k \times \mathcal{N}_l]$  to be the set of trees in  $\mathcal{N}_k \times \mathcal{N}_l$  which do not have this sort of misbehaviour under a local change of arrow. The set  $\mathcal{N}_k \times \mathcal{N}_l$  is equal to  $[\mathcal{N}_k \times \mathcal{N}_l]$  except in the following three cases,

$$\mathcal{N}_2 \times \mathcal{N}_2 = [\mathcal{N}_2 \times \mathcal{N}_2] + A_3 + B_3 + C_3 + C_6 + \tilde{A}_3 + \tilde{B}_3 + \tilde{C}_3 + \tilde{C}_6, \quad (43)$$

$$\mathcal{N}_2 \times \mathcal{N}_3 = [\mathcal{N}_2 \times \mathcal{N}_3] + A_1 + B_1 + C_1 + C_4 + \tilde{A}_2 + \tilde{B}_2 + \tilde{C}_2 + \tilde{C}_5, \quad (44)$$

$$\mathcal{N}_3 \times \mathcal{N}_2 = [\mathcal{N}_3 \times \mathcal{N}_2] + A_2 + B_2 + C_2 + C_5 + \tilde{A}_1 + \tilde{B}_1 + \tilde{C}_1 + \tilde{C}_4. \quad (45)$$

The trees in  $[\mathcal{N}_k \times \mathcal{N}_l]$  contribute equally to the probabilities  $P_{ab}$ ,  $k \leq a \leq 3$  and  $l \leq b \leq 3$ , while those compatible with the diagrams of Figure 2 must be handled separately. One obtains

$$P_{22} = P_{12} + P_{21} - P_{11} + \frac{[\mathcal{N}_2 \times \mathcal{N}_2]}{4\mathcal{N}}, \quad (46)$$

$$P_{23} = P_{13} + P_{22} - P_{12} + \frac{[\mathcal{N}_2 \times \mathcal{N}_3]}{2\mathcal{N}} + \frac{1}{\mathcal{N}}[A_1 + B_1 + C_1 + C_4 + \tilde{A}_2 + \tilde{B}_2 + \tilde{C}_2 + \tilde{C}_5], \quad (47)$$

$$P_{32} = P_{22} + P_{31} - P_{21} + \frac{[\mathcal{N}_3 \times \mathcal{N}_2]}{2\mathcal{N}} + \frac{1}{\mathcal{N}}[A_2 + B_2 + C_2 + C_5 + \tilde{A}_1 + \tilde{B}_1 + \tilde{C}_1 + \tilde{C}_4], \quad (48)$$

$$P_{33} = P_{23} + P_{32} - P_{22} + \frac{\mathcal{N}_3 \times \mathcal{N}_3}{\mathcal{N}} + \frac{1}{\mathcal{N}}[A_3 + B_3 + C_3 + C_6 + \tilde{A}_3 + \tilde{B}_3 + \tilde{C}_3 + \tilde{C}_6] - \frac{1}{\mathcal{N}}[A_1 + A_2 + B_1 + B_2 + C_1 + C_2 + C_4 + C_5 + \text{mirrors}]. \quad (49)$$

The subtracted term in  $P_{33}$  is due to the fact that the part of  $P_{23}, P_{32}$  related to the misbehaved diagrams in Figure 2 (first and second columns) do not contribute to  $P_{33}$ .

The sets  $[\mathcal{N}_k \times \mathcal{N}_l]$  will be further partitioned in classes labelled by a pair of diagrams, *f.i.*  $[\mathcal{N}_2 \times \mathcal{N}_2] = [\beta_1 \beta_1] + [\beta_1 \beta_2] + \dots$ . One will remember that the cardinal of a class does not depend on the *numerical* indices attached to diagrams, so that  $||[\beta_1 \beta_1]|| = ||[\beta_1 \beta_2]|| = ||[\beta_2 \beta_1]|| = ||[\beta_2 \beta_2]||$ , and so on. Replacing however  $\phi_a$  by  $\phi_b$  in a class does not necessarily conserve the cardinal of that class, *f.i.*  $||[\beta \phi_a]|| \neq ||[\beta \phi_b]||$ . The 2-site probabilities can be computed if the numbers of trees in these subclasses and of those compatible with the non local diagrams of Figure 2 can be calculated.

As for the 1-site probabilities, we can decompose each pair of local diagrams as a sum of non local



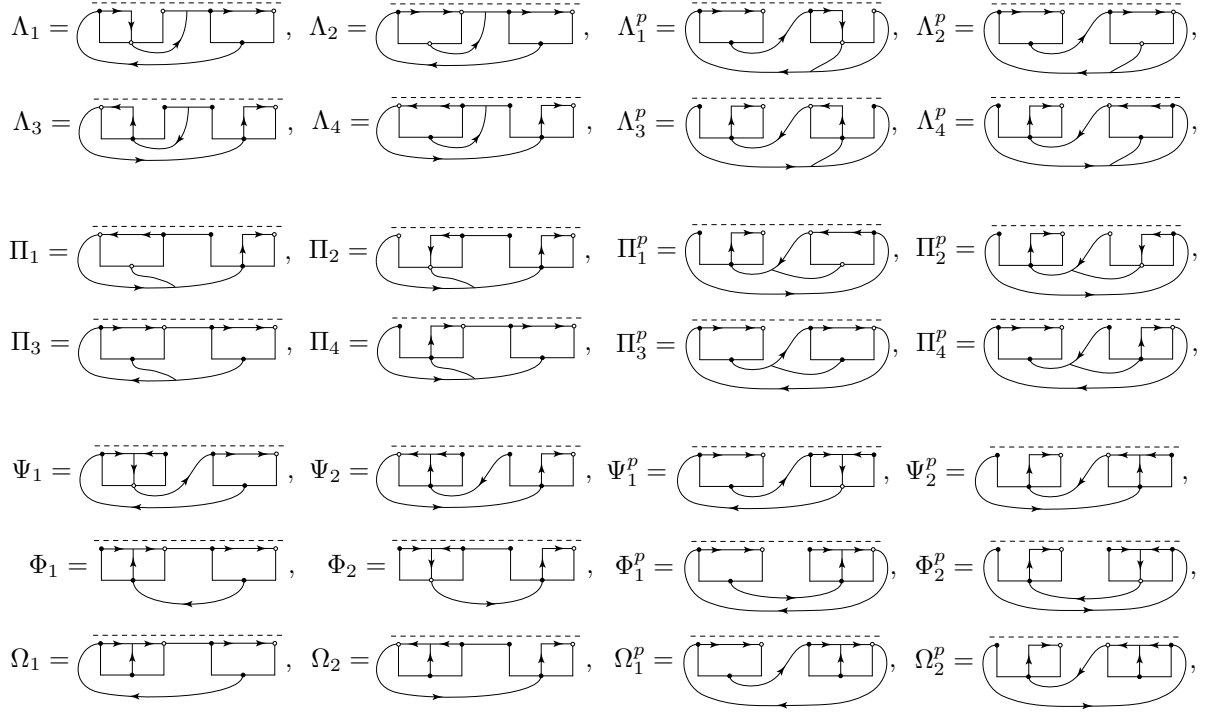


Figure 3: Non local diagrams representing spanning trees which have an anomalous behaviour under the substitution of  $\phi_a$  by  $\phi_b$ , or of  $\phi_b$  by  $\phi_a$ . The superscript  $p$  indicates that the two diagrams at the reference sites have been permuted. All mirror diagrams must be added.

for the 2-site probabilities become

$$P_{22} = P_{12} + P_{21} - P_{11} + \frac{[\mathcal{N}_2 \times \mathcal{N}_2]}{4\mathcal{N}}, \quad (55)$$

$$P_{23} = P_{13} + P_{22} - P_{12} + \frac{\{[\mathcal{N}_2 \times \mathcal{N}_3]\}}{2\mathcal{N}} + \frac{1}{2\mathcal{N}}[2A_1 + 2B_1 + 2C_1 + 2C_4 + 2\tilde{A}_2 + 2\tilde{B}_2 + 2\tilde{C}_2 + 2\tilde{C}_5 + \Lambda_1 + \Lambda_2 + \Pi_1 + \Pi_2 + \tilde{\Lambda}_1^p + \tilde{\Lambda}_2^p + \tilde{\Pi}_1^p + \tilde{\Pi}_2^p], \quad (56)$$

$$P_{33} = P_{23} + P_{32} - P_{22} + \frac{\{\mathcal{N}_3 \times \mathcal{N}_3\}}{\mathcal{N}} + \frac{1}{\mathcal{N}}[A_3 + B_3 + C_3 + C_6 - A_1 - A_2 - B_1 - B_2 - C_1 - C_2 - C_4 - C_5 + \text{mirrors}] + \frac{1}{\mathcal{N}}[\Lambda_3 + \Lambda_4 + \Pi_3 + \Pi_4 + \Psi_1 + \Psi_2 + \Phi_1 + \Phi_2 + \Omega_1 + \Omega_2 + \text{mirrors and permuted}]. \quad (57)$$

The variables entering these expressions can be determined from the same linear system as above, expressed in terms of the new variables. For instance, the first equation becomes

$$Q_{1,1} = \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} = \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ [\beta\beta] \end{array} + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ C_3 \end{array} + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \Lambda_2 \end{array} \\ + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \Lambda_2^p \end{array} + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \Pi_3 \end{array} + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \Pi_3^p \end{array} \\ + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \{[\beta\phi]\} \end{array} + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \{[\phi\beta]\} \end{array} + \begin{array}{c} \begin{array}{cc} \boxed{\phantom{a}} & \boxed{\phantom{a}} \\ \hline \end{array} \\ \{[\phi\phi]\} \end{array} \end{array} \quad (58)$$

The number of equations is the same, but we have fewer variables. The new sets  $\{[\mathcal{N}_2 \times \mathcal{N}_3]\}$ ,  $\{[\mathcal{N}_3 \times \mathcal{N}_2]\}$  and  $\{\mathcal{N}_3 \times \mathcal{N}_3\}$  have respectively 15, 15 and 25 variables, to which 20 extra variables are added for the diagrams in Figure 3 and their mirror images. So there are 88 variables, constrained by

81 linear equations, of which 73 only are linearly independent. The system is still underdetermined but unexpectedly allows to determine enough variables to compute the probabilities without further work.

First one can show, by suitably combining the independent equations, that  $[A + B + C + \tilde{C}]/\mathcal{N}$  is equal to a combination of local diagrams which turns out to be subdominant, of order  $t^3$  for both an open or a closed boundary (order  $m^{-6}$  at the critical point, where  $m$  is the distance between  $i_0$  and  $j_0$ ). As the four quantities are positive by construction, it means that each of them is at least of order  $t^3$ , and can be neglected. Thus one relation determines four variables. Once these four variables are eliminated, one is left with a system of 72 independent equations for 84 variables.

Being independent, the 72 equations allow to determine 72 combinations of variables. The point is that one can choose these 72 combinations in such a way that the probabilities can be fully expressed in terms of them only, thereby making the knowledge of the other 12 combinations useless. Alternatively, one may choose to solve the linear system for 72 variables, which then become functions of the remaining 12. When inserted in the probabilities, all dependences in the 12 unknowns drop out completely. The set of the 12 variables that the system cannot determine is not unique, but a possible choice is  $\{\Psi, \Psi^p, \Lambda^p, \Pi^p, \Phi^p, \Omega^p + \text{mirrors}\}$ .

The counting of variables and equations is different when there are no insertions at other places than  $i_0$  and  $j_0$  since the mirrored equations are redundant. One finds that the linear system is again not invertible, but is nonetheless sufficient to compute all 2-site boundary probabilities. They have been computed in the massive model to the dominant order  $t^2$ , which yields the universal terms.

For an open boundary, we found that none of the diagrams in Figure 3 contribute to the dominant order, being at least of order  $t^3$ . The probabilities  $P_{22}, P_{23}$  and  $P_{33}$  have the same form  $t^2[K_0''(m\sqrt{t}) - K_0(m\sqrt{t})]^2$  at dominant order, and only differ by their normalizations. These have been checked to be in agreement with the identifications obtained in Section 4.

The case of a closed boundary is a bit more complicated. In this case the diagrams of Figure 3 contribute to order  $t^2$  (as we have seen above, none of the diagrams of Figure 2 contribute, irrespective of the boundary condition), and the probabilities read

$$P_{22} = t^2 \left( -\frac{4}{\pi^4} K_0(m\sqrt{t})^2 - \left(\frac{48}{\pi^4} - \frac{12}{\pi^3}\right) K_0'(m\sqrt{t})^2 + \frac{2}{\pi^3} K_0(m\sqrt{t}) K_0''(m\sqrt{t}) \right. \\ \left. - \left(\frac{144}{\pi^4} - \frac{72}{\pi^3} + \frac{37}{4\pi^2}\right) K_0''(m\sqrt{t})^2 + \left(\frac{12}{\pi^3} - \frac{3}{\pi^2}\right) K_0'(m\sqrt{t}) K_0'''(m\sqrt{t}) \right) + \dots \quad (59)$$

$$P_{23} = t^2 \left( -\frac{1}{2\pi^3} K_0(m\sqrt{t})^2 + \left(\frac{8}{\pi^4} - \frac{3}{\pi^3} + \frac{3}{4\pi^2}\right) K_0'(m\sqrt{t})^2 - \left(\frac{1}{\pi^3} - \frac{1}{8\pi^2}\right) K_0(m\sqrt{t}) K_0''(m\sqrt{t}) \right. \\ \left. + \left(\frac{48}{\pi^4} - \frac{12}{\pi^3} + \frac{1}{4\pi^2}\right) K_0''(m\sqrt{t})^2 - \left(\frac{8}{\pi^3} - \frac{3}{2\pi^2}\right) K_0'(m\sqrt{t}) K_0'''(m\sqrt{t}) \right) + \dots \quad (60)$$

$$P_{33} = t^2 \left( -\frac{1}{16\pi^2} K_0(m\sqrt{t})^2 + \frac{2}{\pi^3} K_0'(m\sqrt{t})^2 - \frac{1}{4\pi^2} K_0(m\sqrt{t}) K_0''(m\sqrt{t}) \right. \\ \left. - \left(\frac{16}{\pi^4} + \frac{1}{4\pi^2}\right) K_0''(m\sqrt{t})^2 + \frac{4}{\pi^3} K_0'(m\sqrt{t}) K_0'''(m\sqrt{t}) \right) + \dots \quad (61)$$

Again they are in full agreement with the fields found in Section 4.

We have also computed a few 3-site probabilities, when one of the insertion is a height 1 or a supercritical height. Then the same system as above can be used, the only difference is that the Laplacian has to be decorated by a local defect matrix, and only affects the calculation of the local diagrams. We have found for instance the connected probability  $P_{212}$  to have a height 1 and two heights 2 on a closed boundary, all separated by large distances, at the critical point (the expressions for off-critical 3-site probabilities are too long),

$$P_{212,\text{conn}} = \frac{2}{\pi^3} \left( \frac{3}{4} - \frac{2}{\pi} \right) \left( \frac{1}{m_{12} m_{13}^2 m_{23}^3} + \frac{1}{m_{12}^3 m_{13}^2 m_{23}} + \frac{1}{\pi^2} \frac{(6\pi - 24)^2}{m_{12}^2 m_{13}^2 m_{23}^2} \right. \\ \left. + \frac{6\pi - 24}{\pi} \left[ \frac{1}{m_{12} m_{13}^3 m_{23}^2} + \frac{1}{m_{12}^2 m_{13}^3 m_{23}} + \frac{1}{m_{12}^3 m_{13} m_{23}^2} + \frac{1}{m_{12}^2 m_{13} m_{23}^3} \right] \right) + \dots \quad (62)$$

where  $m_{ij}$  is the distance between the  $i^{\text{th}}$  and  $j^{\text{th}}$  site. The connected probability is equal to  $P_{212,\text{conn}} =$

$P_{212} - P_2 P_{12} - P_{22} P_1 - P_{21} P_2 + 2P_2^2 P_1$ . The previous formula for  $P_{212}$  is equivalent to that found by Jeng [7], but allows for a more direct comparison with the field theoretic result, as the various terms correspond to specific Wick contractions.

## Acknowledgments

We would like to thank Monwhea Jeng for instructive discussions and comparisons of our respective methods. They led us to realize that our early calculations had overlooked some diagrams, which however had no effect on the end results.

## References

- [1] J.L. Cardy, *Conformal Invariance*, in Phase Transitions and Critical Phenomena, Vol. 11, edited by C. Domb and J.L. Lebowitz.
- [2] P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. Lett. 59 (1987) 381.
- [3] S.N. Majumdar and D. Dhar, J. Phys. A: Math. Gen. 24 (1991) L357.
- [4] V.B. Priezzhev, J. Stat. Phys. 74 (1994) 955.
- [5] E.V. Ivashkevich, J. Phys. A: Math. Gen. 27 (1994) 3643.
- [6] S. Mahieu and P. Ruelle, Phys. Rev. E 64 (2001) 066130.
- [7] M. Jeng, *The four height variables of the Abelian sandpile model*, cond-mat/0312656.  
M. Jeng, *The four height variables, boundary correlations, and dissipative defects in the Abelian sandpile model*, cond-mat/0405594.
- [8] D. Dhar, Phys. Rev. Lett. 64 (1990) 1613.
- [9] S.N. Majumdar and D. Dhar, Physica A 185 (1992) 129.
- [10] T. Tsuchiya and M. Katori, Phys. Rev. E 61 (2000) 1183.
- [11] V. Gurarie, Nucl. Phys. B 410 (1993) 535.
- [12] H. Kausch, Nucl. Phys. B 583 (2000) 513.
- [13] M.R. Gaberdiel and H. Kausch, Nucl. Phys. B 538 (1999) 631.
- [14] J.G. Brankov, E.V. Ivashkevich and V.B. Priezzhev, J. Phys. I France 3 (1993) 1729.
- [15] M. Jeng, *Conformal field theory correlations in the Abelian sandpile model*, cond-mat/0407115.
- [16] P. Ruelle, Phys. Lett. B 539 (2002) 172.
- [17] G. Piroux and P. Ruelle, *Pre-logarithmic and logarithmic fields in sandpile model*, hep-th/0407143.