STUDIES IN SELF-ORGANISED CRITICALITY

by
Gunnar Pruessner

A thesis presented for the degree of
Doctor of Philosophy of the University of London
and the
Diploma of Imperial College

Imperial College London
Department of Mathematics
Huxley Building
180 Queen's Gate
London SW7 2BZ
United Kingdom

February 2004

Für den Tiger.

Abstract

Self-Organised Criticality (SOC) *has been suggested by* PER BAK, CHAO TANG *and* KURT WIESENFELD *to explain the abundance of scale-invariant spatio-temporal structures in nature. It can be observed in many slowly driven, highly interacting non-equilibrium systems which develop into a scale-invariant state without explicit tuning of any parameters. Focusing on particular features of these systems, toy models have been developed to study the phenomena in detail both numerically and analytically. The presence of scale-invariance indicates the presence of universality, so that apparently unrelated systems are intimately linked by their fundamental interactions and symmetries.*

Even after more than 16 years of research, the necessary and sufficient conditions for the appearance of SOC remain to a large extent unclear.

Within the enormous zoo of SOC-models, there is hardly any system that is well-behaved and non-trivial while nevertheless showing all the expected features of a scale-invariant system. However, reliable models are required to pursue the analytical understanding of SOC.

In this thesis, characteristic features of scale invariant systems are identified against the background of classical critical phenomena. They form the basis for a numerical simulation of the Forest Fire model on large scales. The results refute former findings: The Forest Fire model is shown to lack important characteristics of scale-invariance.

Nevertheless, a stochastic process, known as the Oslo model, is found to show consistent scale-invariant behaviour. In the second part of the thesis it is discussed in detail. It transpires that the Oslo model is in fact a discrete realisation of the quenched EDWARDS-WILKINSON *equation. This finding opens the door for further analytical investigations. Moreover, an anisotropic variant of it, representing a very large universality class, is solved exactly. This raises the issue of the rôle of a drift term in the quenched* EDWARDS-WILKINSON *equation, which is examined in some depth.*

Table of Contents

List of Tables

List of Figures

Acknowledgements

it's not that it didn't, it's that it now does
it's not just ironic, it's to do with a rush
transitions need making
and averages taking
so corporate searching must hush.

— Matthew G. Frederic

The acknowledgement section is probably the most read part of a PhD thesis. In fact, there is no reason to hesitate to praise other people's work. I am deeply indebted to all the people around me, from whom I have benefited so greatly. My PhD was and still is a fantastic time; Imperial and London are really great places.

I have had the great privilege to learn from Prof. Dr. Henrik Jensen. I am very grateful for what he taught me about science and good scientific practice, as well as for his patience, sincereness, honesty and the good time we have spend together with his family.

I am very grateful to Dr. Kim Christensen, especially for his seemingly infinite patience.

I am very happy to have met Dr. Matt Hall and thankful to him for being such a very good friend. I am also grateful to have become friends with Dr. Simone Avogadro di Collobiano, with his passion, discipline and admiration of good PSTricks.

I would like to thank Paul Anderson for lots and lots of good advice, lots and lots of time and plenty of exercises in discipline, Nicholas Moloney for the most elaborate and careful help and for having bizarre opinions about rather mundane issues, as well as Andy Thomas for his ingenuity and the most fantastic computing support.

Also, I want to thank Holger Bruhn and Inge Ruf for good times, good booze and interesting discussions, Prof. Dr. Yang Chen for good science, helpful advice and cheerful ranting, Dr. Martin Greenall for being Martin Greenall, Dr.

MATTHIAS KAULKE, OLIVER KILIAN and Dr. DAN MOORE for their continuous support, Dr. NIGEL LAWRENCE and the APS for some of their style files, OLE PETERS and MATTHEW STAPLETON for interesting and highly instructive discussions, and Dr. h. c. QUINCY THOEREN for plenty of useful advice on the important things in life.

I would like to thank all members of the Mathematical Physics section as well as the Condensed Matter Theory Group in the Physics Department and especially the research group of Prof. JENSEN as well as Dr. CHRISTENSEN's group, for the friendly, relaxed, interesting and very stimulating environment.

Finally, I would like to thank my family for their continuous support and especially for their enormous patience.

This thesis is dedicated to SONJA who I owe so much for the wonderful life we spend together.

Introduction

A phenomenological approach to Self-Organised Criticality

It is not easy to fully understand the history of Self-Organised Criticality (SOC). It seems it caused major uproar when it was first proposed by PER BAK, CHAO TANG and KURT WIESENFELD (Bak *et al.*, 1987) in 1987. Probably the most exciting message is: There are, apparently, scale-invariant systems, which do not have to be tuned to a critical value. After almost 60 years of research into critical phenomena[1] the crucial notion of a "critical point" seems to have disappeared.

The 1987 article has, to use an overstrained pun, truly triggered an avalanche. Literally thousands of articles have been published on SOC — BAK, TANG and WIESENFELD's paper alone has accumulated 1947 citations.[2] Looking at this sheer volume, but also at the phenomena reported, there can hardly be any doubt that SOC exists, and that it is, in fact, ubiquitous.

Indeed, sightings of SOC have been reported in every conceivable and inconceivable area of science, encompassing sociology (Roberts and Turcotte, 1998; Bentley and Maschner, 1999/2000), computer science (Gorshenev and Pis'mak, 2003), engineering (Carreras *et al.*, 2002) and biology (Sepkoski Jr., 1993). Further examples, most of them more profound, are given in Chapter 1. The key problem is that in many of these studies, SOC remains on a purely descriptive level. Rather than providing any insight into the underlying mechanisms and the nature of the problem, it is used as a name tag to label power law observations. Attempts to go further and to draw conclusions at least on the level of analogy sometimes look rather grotesque, or simply outrageous [(Turcotte, 1999, p. 1419), which contains further interesting references]:

[1] One might set the beginning of research on critical phenomena to the publication of LARS ONSAGER's paper in the *The Physical Review* in 1944 (Onsager, 1944) (which did not arrive in most of Europe before the end of the war). It is worth noting that ERNST ISING refuted the possibility of spontaneous order in his model, (Ising, 1925).

[2] Number of citations according to the "ISI Web of Knowledge℠", in particular the "ISI Web of Science®", namely the "Science Citation Index Expanded®", "Social Sciences Citation Index®" and "Arts & Humanities Citation Index®", `http://wos.mimas.ac.uk/`, 5th February 2004. According to the American Physical Society, `http://prola.aps.org/forward/PRL/v59/i4/p381_1`, there are 713 citations as of 20th January 2004.

> *Similarly, a war must began* [sic!] *in a manner similar to the ignition of a forest. One country may invade another country, or a prominent politician may be assassinated. The war will then spread over the contiguous region of metastable countries. Such regions of metastability could be the countries of the Middle East (Iran, Iraq, Syria, Israel, Egypt, etc) or of the former Yugoslavia (Serbia, Bosnia, Croatia, etc). These are then the metastable clusters. In some cases the metastable clusters could combine. Albania and Greece bridge the gap between the metastable clusters of the Middle East and the former Yugoslavia.*

One might speculate that one reason for the popularity of SOC and power laws is the fact that a few numbers, the exponents, seem to characterise an observation and make them comparable to other, possibly completely unrelated phenomena. So rather than getting lost in details, exponents seem to capture all the relevant information at once.

This, in fact, points to a scientifically more profound motivation for SOC: Universality. Critical, or rather scale-invariant behaviour is interesting because it is universal, *i.e.* seemingly different systems share the same critical properties such as critical exponents and amplitude ratios. This is usually caused by a common underlying interaction or symmetry.

Towards an analytical approach to SOC

Thus, even though questionable as the sole scientific aim, identification of common critical properties can teach us something about the underlying mechanisms.

However, before one can really appreciate the "ubiquity of power laws in nature", one should develop an insight into what causes these power laws and what they actually signify. That is, SOC must gain some explanatory power. The central question is thus:

What are the necessary and sufficient conditions for SOC?

The sufficient conditions point to the causes of SOC, the necessary conditions point to phenomena caused by SOC. The most realistic approach to that question is to make the phenomena under the label "SOC" accessible to established physical and mathematical techniques. Those techniques which have already proven their power by tackling classical critical phenomena seem most promising.

As a first step, one has to define SOC carefully and identify toy models which undoubtedly show at least some of the expected features. In a second step, these models can be investigated in detail. This also sets out the philosophy of the thesis.

Outline of the thesis

First, SOC as a phenomenon is reviewed. Without redefining SOC, generic scale-invariance is identified as its most crucial feature. Some toy models are presented, the Forest Fire model in great detail. The latter can be thought of as a good example of the problems for which SOC meanwhile is notorious for: Sufficiently close inspection typically reveals that the phenomenon or model under consideration is just not scale-invariant.

However, most remarkably, there are proper, "sandpile-like" models which show generic scale invariance. One such model is the Oslo model, to which the entire second part of this thesis is dedicated.

Overview chapter by chapter

Part I

Chapter 1. Overview: This chapter contains an overview on SOC mainly in the form of a literature review: The meaning of the term, observations, established models. Moreover, it discusses a very simple mechanism supposedly explaining SOC, which continues to gain support in the community, but has some serious shortcomings.

This chapter gives an introductory overview.

Chapter 2. Scaling: Scaling and power laws are central theoretical, experimental and numerical themes in SOC. This chapters tries to demonstrate some intimate links between "classical critical phenomena" and SOC. An understanding of the key concept of simple scaling allows us to give a clear prescription for good numerical analyses. In particular, moment analysis and universal amplitude ratios are introduced and discussed in detail.

This chapter develops some key tools for the analysis of SOC.

Chapter 3. The Rôle of Conservation: Very early in the history of SOC, conservation and dissipation was discovered to be a central issue. This chapter presents some results of a LANGEVIN approach to these phenomena. Additionally, some details of the OFC model are briefly discussed. Today, the OFC model enjoys paradigmatic status among the models which supposedly display SOC in the presence of bulk dissipation. The bulk of this chapter concerns a solvable, random neighbour model, which shows a power law event size distribution in the presence of dissipation.

This chapter is mainly analytical. Results have been published in (Pruessner and Jensen, 2002b).

Chapter 4. The Forest Fire Model: The Forest Fire model is another established, dissipative model of SOC. A large scale numerical analysis, however, reveals that there is actually no reason to suspect that this model is scale-invariant in any limit. The algorithm which made the simulation possible is presented in detail.

This chapter is mainly numerical and contains many technical details. Results have been published in (Pruessner and Jensen, 2002a, 2003b). The techniques developed have been further exploited in (Moloney and Pruessner, 2003; Pruessner and Moloney, 2003, 2004).

Part II

Chapter 5. The Oslo Model and Its Variants: The Oslo model has turned out to be one of the few models displaying SOC and all features expected from classical critical phenomena. The model is introduced together with a number of its variants. An overview of numerical results is presented at the end of the second section. Furthermore, an operator approach is introduced, which will be explored in a similar manner in Chapter 8.

This chapter gives an introductory overview on the Oslo model.

Chapter 6. The Oslo Model and the qEW: Most remarkably, the equation of motion for the Oslo model is a discrete realisation of the quenched EDWARDS-WILKINSON (qEW) equation. This discovery allows us to investigate the Oslo model on a well-founded analytical level. Some concrete calculations show full correspondence to analytical results for the qEW found in the literature. The same method can be applied to the periodic Oslo model, but this leads to some serious difficulties.

This chapter is mainly analytical. Results have been published in (Pruessner, 2003c).

Chapter 7. Universality, Anisotropy and Crossover: Exploring the universal features of the Oslo model further, one can introduce anisotropy which turns out to be relevant. Thus, any degree of anisotropy drives the Oslo model eventually (in the thermodynamic limit) to the "fixed point" of the Totally Asymmetric Oslo Model (TAOM). For very small degrees of anisotropy a crossover behaviour is observed and analysed. The TAOM can be solved exactly, see Chapter 8. Because of its link to the qEW, this amounts to an exact solution of the qEW equation with an extra drift term. Based on the qEW, some exact calculations regarding toppling frequency and avalanche size can be performed quite easily. The latter can also be tackled using a random walker approach. Implications for experimental systems are discussed.

This chapter is partly numerical and partly analytical. Results have been published in (Pruessner and Jensen, 2003a).

Chapter 8. Exact Solution of the TAOM: This chapter contains details of the calculations leading to the exact solution of the TAOM as mentioned above. A MARKOV matrix ansatz leads to a recursive expression for the eigenvectors and finally gives a moment generating function of the avalanche size distribution. After mapping the model on a reaction-diffusion process, a continuum theory is developed which fully agrees with the results on the lattice.

This chapter is mainly analytical. Results are to be published in (Pruessner, 2003b).

Chapter 9. The Thermal EW Equation with Drift: It is remarkable that the qEW equation becomes solvable, apparently solely due to the presence of a drift term — thereby, of course, leaving the qEW universality class. What can one infer from the critical properties of the model with drift to the model without drift? The (asymptotic) solution of the *thermal* EDWARDS-WILKINSON (EW) equation with drift addresses this problem in detail. In this context dimensional analysis is discussed, as it seems to suggest incorrect results for the EW with drift.

This chapter is mainly analytical. Results are to be published in (Pruessner, 2003a).

The key chapters are probably Chapter 6 (The Oslo Model and the qEW) and Chapter 7 (Universality, Anisotropy and Crossover).

Style

This thesis is comparatively long, which certainly needs justification. The guiding principle was to provide some background information and extra details to the published work. That way, the thesis might actually be of some use. Here are some precautions and remarks regarding the style:

- Whenever results are already published, detailed calculations and discussions are presented, whenever I felt that they might help to develop the problem further.
- Material presented in the thesis that has not been published contains much more detail.
- I did not hesitate to use references wherever possible. It should become clear from the context which references have been used in detail and which references only serve as a “backup”.
- Even though the chapters relate to each other, they should be completely self-contained — which is one more reason for the length of the thesis. Nevertheless

they are ordered hierarchically, either in the sense that chapters pick up some of the themes developed in preceeding chapters or simply in a temporal sense (*i.e.* the material in later chapters has typically been developed later).

- Every chapter starts with an abstract and ends with a summary.
- Many results presented in this thesis have naturally raised new questions. Furthermore, there are cases, where the results must be regarded as preliminary or incomplete. The thesis contains marginal notes in the form “Open issue” if a detailed investigation is still missing or a method still needs clearing up. “Open problem” stand for a concrete problem, such as a contradiction or a controversy, which requires further investigation. Finally “Open question” or “Open task” indicates a concrete question to be answered or concrete task to be performed. They are listed on page 29 in short form.
- Many discussions might appear rather mundane and maybe not even suitable for a thesis. I have tried to use the thesis as a chance to write down the status of many of the issues I have touched during my PhD — that actually *is* a collection of rather mundane problems!
- Finally, in contrast to published work, I did not hesitate to present arguments and results which turn out to be dead ends.

List of Acronyms

AS	**A**bsorbing **S**tate *or sometimes* **A**bsorbing **S**tate phase transition
BC	**B**oundary **C**onditions
BTW	**B**AK-**T**ANG-**W**IESENFELD
DS-FFM	**D**ROSSEL-**S**CHWABL **F**orest **F**ire **M**odel
DP	**D**irected **P**ercolation
EW	**E**DWARDS-**W**ILKINSON (equation)
EWd	**E**DWARDS-**W**ILKINSON equation with **d**rift
FBC	**F**ixed **B**oundary **C**onditions
FFM	**F**orest **F**ire **M**odel *see also* DS-FFM
FSS	**F**inite **S**ize **S**caling
KPZ	**K**ARDAR-**P**ARISE-**Z**HANG
LE	**L**ANGEVIN **E**quation
LHS	**L**eft **H**and **S**ide *see also* RHS
MCS	**M**onte **C**arlo **S**tep
MF	**M**ean **F**ield
OFC	**O**LAMI-**F**EDER-**C**HRISTENSEN
PBC	**P**eriodic **B**oundary **C**onditions
PDE	**P**artial **D**ifferential **E**quation
PDF	**P**robability **D**ensity **F**unction
qEW	**q**uenched **E**DWARDS-**W**ILKINSON (equation)
qEWd	**q**uenched **E**DWARDS-**W**ILKINSON equation with **d**rift
RG	**R**enormalisation **G**roup
RHS	**R**ight **H**and **S**ide *see also* LHS
RNG	**R**andom **N**umber **G**enerator
SOC	**S**elf **O**rganised **C**riticality

List of Symbols

This is an incomplete list of some frequently used symbols.

Symbol	Explanation
a	metric factor
b	metric factor
d	spatial dimension
D	cutoff exponent
D	diffusion constant
E	expectation value
$\dot{E}$	external driving frequency
$E(t)$	external drive (boundary condition, height)
f	probability to ignite a tree in the Forest Fire model
h_i	height of a ricepile at position i
$h(x)$	number of charges of a site, or height of an interface over a substrate at position x
g_n	nth universal moment ratio
H	number of topplings of a site
$\mathbf{H}$	unit of interface height
j_n	flux of the $n$the moment in the exact solution of the TAOM
k_n	nth wavenumber
L	linear system size
L_{X}	crossover length
$\mathbf{L}$	unit of length
M	magnetisation or, in general, an order parameter
M_0	cutoff in the order parameter distribution
$\bar{n}(s)$	cluster size distribution
$\mathbf{O}$	operator in a MARKOV process
$\mathcal{P}(s)$	avalanche size distribution, or, in general, a distribution function

continued on next page

continued from previous page

Symbol	Explanation
p	probability to grow a tree in the Forest Fire model, or probability for various processes otherwise
S	local update in a Markovian sandpile without toppling
s	avalanche size
t	time
T_{M}	burning time in the Forest Fire model
T	unit of time
T	local update in a Markovian sandpile with toppling
U	local update in a Markovian sandpile with double toppling
v	driving velocity
v	anisotropy parameter, drift
v_{X}	crossover drift
w^2	width of an interface
x	spatial coordinate or exponent of the crossover drift
z_i	slope of a ricepile at position i
z_i^c	critical slope of a ricepile at position i
$z(x)$	interface position over a substrate at position x with trivial contribution removed
α	Driving position as fraction of system size
η	noise
θ	crossover exponent in the Oslo model or driving parameter in the Forest Fire model
$\theta(x)$	HEAVISIDE theta function
λ	cutoff exponent in the Forest Fire model
μ	auxiliary exponent
ν	exponent of the correlation length
ξ	correlation length
ρ	density of particles, trees etc.
τ	avalanche exponent
$\Upsilon(n)$	indicator function; is 0 for even n, -1 otherwise
φ	dimensionless form of h
ψ_n	nth conditional moment in the exact solution of the TAOM
ω	frequency

Open Problems

The following is a list of open problems etc. as described in the introduction, page 24:

Publications

The following publications are based on or related to material presented in this thesis.

K. Christensen, N. R. Moloney, O. Peters and G. P., 2004, *Avalanche Behavior in an Absorbing State Oslo Model*
submitted to *Phys. Rev. E*, preprint `cond-mat/0405454`, (Christensen *et al.*, 2004)

G. P., 2003, *Exact solution of the totally asymmetric Oslo model*
accepted for publication in *J. Phys. A*, preprint `cond-mat/0402564`, (Pruessner, 2003b)

G. P., 2003, *Drift causes anomalous exponents in growth processes*
accepted for publication in *Phys. Rev. Lett.*, preprint `cond-mat/0404007v2`, (Pruessner, 2003a)

G. P. and N. R. Moloney, 2003, *Winding clusters in percolation on the Torus and the Möbius strip*
accepted for publication in *J. Stat. Phys.*, preprint `cond-mat/0310361`, (Pruessner and Moloney, 2004)

G. P. and H. J. Jensen, 2003, *Alternative efficient algorithm for the Forest Fire Model*
accepted for publication in *Phys. Rev. E*, preprint `cond-mat/0309173`, (Pruessner and Jensen, 2003b)

G. P. and H. J. Jensen, 2003, *Anisotropy and universality: the Oslo model, the rice pile experiment and the quenched Edwards-Wilkinson equation*
Phys. Rev. Lett. **91**, 244303-1–4, preprint `cond-mat/0307443`, (Pruessner and Jensen, 2003a)

G. P. and N. R. Moloney, 2003, *Numerical results for crossing, spanning and wrapping in two-dimensional percolation*
J. Phys. A **36**, 11213–11228, preprint `cond-mat/0309126`, (Pruessner and Moloney, 2003)

N. R. MOLONEY and G. P., 2003, *Asynchronously parallelized percolation on distributed machines*
Phys. Rev. E **67**, 037701-1–4, preprint `cond-mat/0211240`, (Moloney and Pruessner, 2003)

G. P., 2003, *Oslo rice pile model is a quenched Edwards-Wilkinson equation*
Phys. Rev. E **67**, 030301(R)-1–4, preprint `cond-mat/0209531`, (Pruessner, 2003c)

G. P. and H. J. JENSEN, 2002, *Broken scaling in the forest-fire model*
Phys. Rev. E **65**, 056707-1–8, preprint `cond-mat/0201306`, (Pruessner and Jensen, 2002a)

G. P. and H. J. JENSEN, 2002, *A solvable non-conservative model of self-organised criticality*
Europhys. Lett. **58**(2), 250–256, preprint `cond-mat/0104567`, (Pruessner and Jensen, 2002b)

Part I

SOC in general

Chapter 1

Overview

*The acronym "SOC" stands for "**S**elf-**O**rganised **C**riticality" and is a subsection of statistical mechanics. It is concerned with driven systems which contain many coupled degrees of freedom and develop to a scale-invariant state without explicit tuning of parameters. SOC asks not only for the properties of these systems and for their common features, but also whether criticality due to self-organisation is possible at all.*

In the following sections the meaning of SOC is explained in some detail. Its significance is motivated by referring to natural phenomena as well as to theoretical considerations. Sec. 1.3 contains an overview of the most prominent model systems used in the literature. The last section contains a discussion of a recent suggestion about "how SOC works".

1.1 Meaning of Criticality and Self-Organisation

Criticality by its common definition usually means a high susceptibility to external perturbations. In traditional models displaying critical phenomena[1], such as ferromagnetic systems like the ISING model, the liquid-vapour transition at the critical point (Stanley, 1971) or percolation (Stauffer and Aharony, 1994), this naive interpretation is still valid in the context of statistical mechanics. However, in the following, criticality refers to the absence of a typical length scale; at the critical point all length scales are relevant. On a handwaving level, this entails divergent fluctuations due to external perturbations, since in the absence of any length scale the response to a perturbation cannot give rise to a length scale either.

Most non-critical systems are governed by a hierarchy of scales; certain effects and certain interactions dominate the behaviour of the systems at certain scales. Close to the critical point these scales get separated and usually [there are exceptions,

[1] In the following called "classical" or "traditional critical phenomena'.

for example, in cases of dimensional crossover (Novotny, 1996; Anisimov *et al.*, 1998)] one expects then only one length scale to characterise or dominate the behaviour of the system, like the average cluster size in percolation or the correlation length in the ISING model. It is a well-established misconception that systems at criticality "forget all about their microscopic details". In fact, at least in classical critical phenomena, it is the tuning of *microscopic* couplings, which give rise to the critical behaviour, and, most importantly, only due to the presence of microscopic scales, anomalous exponents can be observed at all (Goldenfeld, 1985) (see also Chapter 9).

At the critical point, the remaining characteristic length disappears, usually by diverging. The alternative case of a vanishing characteristic length is usually trivial and therefore ignored. In the following, "criticality" in "SOC" will refer to the absence of a characteristic scale.

The term "self-organised" is used in a much less technical sense as the term "criticality" above. What "self-organisation" indicates here is that the system is not tuned to its critical state, it develops into this stage from (almost) any initial configuration. Its evolution — either stochastic or deterministic — is determined by *local* rules, *i.e.* if there is any tuning mechanism, it must be intrinsic to the dynamical rules, rather than explicitly given.

Such an explicit rule might be: "Tune the temperature of the ISING model up if there is a non-vanishing magnetisation, and tune it down if the net-magnetisation vanishes." Such a model might trivially develop to its critical state, as will be discussed in sec. 1.4, but would make use of global measures like the net magnetisation (see page 63).

The requirement of local rules does not mean that the resulting processes have to be local, in fact system spanning events are typically the most interesting.

The term "organisation" hints to some kind of organisation or evolution and this is usually understood as a non-equilibrium process. Indeed, SOC is concerned with system far from equilibrium, *i.e.* the systems are driven externally, do not obey detailed ballance and do not relax towards a state, where detailed balance is obeyed. Nevertheless, the systems are (usually) studied in the stationary state. For a stationary state to exists under the presence of an external drive, the systems need a dissipation mechanism. The rate, as well as spatial and temporal correlation of the dissipation, are typical observables of an SOC model, apart from the statistical properties of the configurations themselves. In the histograms of the observables and correlations in the events power laws are observed, which will be discussed in more detail in chapter 2.

As will be shown explicitly in sec. 1.3 for a number of popular models, the typical features of almost all SOC models are:

- (Lattice models) Most models are realised as sites on a lattice, each of which having a degree of freedom, either continuous or discrete.

- (Relaxation) Interaction between sites are usually nearest neighbour like and governed by local rules.
- (Drive) There is a slow external drive, which triggers a finite relaxation event. Usually the models are driven only during their quiescent phase, *i.e.* after they are completely relaxed.
- (Threshold) Most systems contain some threshold, *i.e.* activity takes place only, if certain thresholds are reached.
- (Dissipation) To compensate the external drive, the relaxation somehow includes a dissipation mechanism (boundary or bulk dissipation).
- (Observables) Typical observables are: histogram of the magnitude of dissipative or relaxation events, correlation length, occupation probabilities etc.
- (Power laws) Power laws are observed in the observables, very similar to the cluster size distribution of standard percolation.

Or to say it shortly: "We will expect SOC behaviour in slowly-driven interaction-dominated threshold systems" (Jensen, 1998, p. 126).

To summarise this section, SOC deals with systems, which have many interacting degrees of freedom. These systems evolve under local updating rules, which are stochastic or deterministic. Apparently, they evolve into a scale-invariant state without explicit tuning of parameters.

1.1.1 SOC and generic scale invariance

In the early days of SOC, there was still some hope to establish a sharper definition of the term SOC. One of the most popular definition has been introduced by GRINSTEIN, LEE and SACHDEV in 1990 (Grinstein *et al.*, 1990) (see also chapter 3) in an effort to remove trivial cases of generic scale invariance from SOC. They

> *. . . reserve the term SOC for the situations where the correlation length is infinite, so that not only temporal but spatial correlations are long ranged, since in the presence of conservation laws even equilibrium system under generic conditions exhibit 'long-time tails', i.e., correlations that decay algebraically in time at a given point in space, even though spatial correlations decay exponentially.*

This definition has not survived — not least because it would exclude all models which are not spatially extended, such as any random neighbour model.

The random walker is widely regarded as the simplest case of SOC, even though its scale-invariant properties exclusively derive from the underlying noise. But to exclude the random walker from the class of SOC models because of that would be very dangerous, because many other models are expected to depend strongly on an underlying noise term; in fact, *all* models discussed in sec. 1.3 contain a noise-source and therefore can be regarded as mechanisms which just process the noise.

1.1.1.1 Interface literature

The arguments by GRINSTEIN *et al.* presented in (Grinstein *et al.*, 1990) are based on a LANGEVIN or interface growth background. In recent years, there has been a remarkable tendency in the interface-community to incorporate SOC into their themes as another name for "generic scale invariance", very much in contrast to GRINSTEIN *et al.*.

In his review (Lässig, 1998) MICHAEL LÄSSIG explicitly mentions SOC as a case of generic scale invariance in slowly driven systems, which is very similar to the generic scale invariance observed in interface problems. Similarly, JOACHIM KRUG (Krug, 1997), mentions SOC as a concept to study mechanisms possibly responsible for generic scale invariance. He stresses that the association of scale invariance and critical point behaviour, so strongly pronounced in the term SOC, is lead by equilibrium critical phenomena. However, many non-equilibrium phenomena are known to exhibit scale invariance without a specific critical point. Consequently, he advocates to reserve the term SOC to models which display scale invariance through a *separation of time scales* between driving and relaxation.

Finally BARABÁSI and STANLEY discuss SOC in their review (Barabási and Stanley, 1995) on surface growth. This is motivated by an article by HWA and KARDAR (Hwa and Kardar, 1989) on a LANGEVIN approach to the sandpile model, which will be, together with the article by GRINSTEIN *et al.* mentioned above, discussed in chapter 3.

1.2 Power Laws in Nature

As will be explained in chapter 2, power laws are the fingerprint of criticality. The research in the field of SOC is motivated by the apparent ubiquity of power laws in nature. They are found virtually everywhere, although it sometimes seems to require a significant bias to claim them. The most striking problem in order to identify power laws definitely is the number of decades available to analyse the behaviour. There seems to be a general inclination towards finding power laws in the scientific community. A similar problem was reported three years ago for the case of fractals in nature (Malcai *et al.*, 1997; Avnir *et al.*, 1998). It turned out that the majority of power law claims were based on only 1.3 decades. Therefore, it seems reasonable to treat some claims regarding the discovery of power laws with great care, especially if they lead to conclusions of enormous implications [for example (Smethurst and Williams, 2001) and comments, for example (Ball, 2001; The British Library Science Technology and Business (STB), 2001; Sornette, 2001)].

Nevertheless, power laws are observed reproducibly and reliably over many decades, and occur very often in natural processes, such as

- Earthquakes (Johnston and Narva, 1985) [after (Bak, 1996)]: Earthquakes are the standard example of power laws in nature, not only because they have been studied and recorded literally for ages, but also because it has been noted very early that their distribution obeys a power law. The Gutenberg-Richter law says that the cumulated probability density $\mathcal{P}_c(E)$ of obtaining an earthquake of energy E is a power law, $\mathcal{P}_c(E) \propto E^{-B}$ [(Gutenberg and Richter, 1956) according to (Olami *et al.*, 1992)]. The exponent B varies in the range $[0.80, 1.05]$, depending on the fault under consideration, but it is truly an exponent, *i.e.* it does not vary with the size of the earthquake. As all earthquakes follow the same statistics, this justifies the remarkable assumption that the underlying mechanisms of small and large earthquakes is the same! In addition, the intensity of aftershocks decays scalefree, known as the OMORI law. Recently it has been shown by BAK *et al.* (Bak *et al.*, 2002) that the OMORI law and the Gutenberg-Richter law can be unified.

- Rainfall (Peters *et al.*, 2002): The time signal of rain, *i.e.* the rain rate (a volume per area and time) can be recorded with very high precision. The time signal allows one to define events (rate nonzero), event sizes (total volume) and interoccurance times, which is the time between two events. The interoccurance time, as well as the event size distribution, show a power law distribution. So far this is checked for only one spot on the earth, but the data look very promising.

- Coastlines (Feder, 1988) [after (Bak, 1996)]: This is the classical example of fractals in nature and became very popular in the context of chaos and fractals [for example (Mandelbrot, 1983)]. The length of a coastline does not change linearly with the length scale of the embedding space, *i.e.* taking a linearly increasing patch of a map and holding the absolute resolution (lower cutoff) fixed, the length of the coastline measured increases with an exponent different from one.

- Elementary particles (Meng Ta-chung *et al.*, 1999; Rittel, 2000): Instead of using the framework of Quantum Chromodynamics, it is possible to approach inelastic diffractive high-energy scattering processes on a phenomenological level in terms of SOC. The resulting predictions agree very well with experimental data. The idea is that certain colourless objects obtained in so-called LRG events

(Large Rapidity Gap) have a scalefree size and lifetime distribution. These events can be seen in highly inelastic electron-proton scattering experiments. The conclusion is that the proton develops to a scalefree state on the gluon level. It is remarkable how consistent these assumptions are with experiments. Moreover it is possible to derive the scattering cross section for (anti-)proton-proton scattering, which is also in impressive agreement with experiments.

- Evolution (Sepkoski Jr., 1993) [after (Bak, 1996)]: Based on the marine fossil records there are some indicators which allow (again on a phenomenological level) the conclusion that extinction rates are scalefree. Introducing a temporal grid of the geological history (time intervals of roughly $25,000$ years) one counts into a histogram the number of time intervals where the relative extinction (number of species getting extinct between two time slots over their total number) was within a certain range. The results seem to indicate that there is no typical event size, in the sense of a scalefree distribution. Recently this result has been criticised as statistical artifact (Kirchner and Weil, 1998).

1.3 Overview of Established Models

As explained above, SOC is the theory of critical phenomena which seem to occur in natural systems without tuning. SOC strives to identify the necessary and sufficient conditions for this behaviour, usually by means of model systems, which are much easier controllable than real-world experiments. They are designed to mimic certain features, which are assumed to be important: Spatial order, temporal order, conservation, certain types of external drive (random, homogeneous, deterministic) etc. These models are mainly numerical, *i.e.* they are designed to be studied in computer simulations and the vast majority of publications in the field of SOC consider these models, mostly in a numerical way.

Although Molecular Dynamics is a standard technique for simulating spatially continuous models, virtually all SOC models live on a lattice and are therefore investigated within the framework of Monte Carlo. In the following sections the most popular models are defined.

1.3.1 The BTW model

The BAK-TANG-WIESENFELD sandpile model (Bak *et al.*, 1987) (BTW model for short) is the paradigm of SOC. Originally invented to explain $1/f$ noise (this is a form of noise with logarithmic correlations), it quickly became a research topic on its own (Bak *et al.*, 1988; Christensen *et al.*, 1991). To see to what extend it actually resembles a sandpile, we first give the original definition in terms of *heights*. This, however, is meaningful only in $d = 1$, so the first definition is applicable only to the one-dimensional model.

On a one-dimensional lattice (*i.e.* in $\mathbb{Z}$) each node has a height h_i and a slope derived from this quantity, defined as $z_i = h_i - h_{i+1}$. The model evolves as follows (Jensen, 1998):

- Perturbation: A grain is added to a randomly chosen site k, *i.e.*

$$\begin{aligned} h_k &\rightarrow h_k = h_k + 1 \\ &\quad \text{and therefore} \\ z_k &\rightarrow z_k = z_k + 1 \\ z_{k-1} &\rightarrow z_{k-1} = z_{k-1} - 1 \end{aligned}$$

- Relaxation (Toppling): If the slope at site k is larger than a certain threshold z_c, that site relaxes by moving one grain from k to $k+1$, *i.e.* it slides down and decreases the slope:

$$\begin{aligned} &\quad \text{If } z_k > z_c \qquad (1.1) \\ h_k &\rightarrow h_k = h_k - 1 \\ h_{k+1} &\rightarrow h_{k+1} = h_{k+1} + 1 \\ &\quad \text{and therefore} \\ z_k &\rightarrow z_k = z_k - 2 \\ z_{k\pm1} &\rightarrow z_{k\pm1} = z_k + 1 \end{aligned}$$

Starting from an empty lattice, the system is perturbed until a slope reaches the threshold. Then the relaxation rule is applied until no slope is above the threshold anymore. The toppling of a site is also depicted in Fig. 1.1.

In the definition given above, it seems that the total slope is invariant under evolution, as all rules keep it constant. However, for a finite system of size N it is

$$Z = \sum_{i=1}^{N} z_i = h_1 - h_{N+1} \tag{1.2}$$

reflecting a problem of the rules at the boundaries (h_{N+1} is not defined). The total height,

$$H = \sum_{i=1}^{N} h_i \tag{1.3}$$

can only increase. Moreover, the rules are anisotropic, as the slope is defined only "towards the right". To justify that and in order to make the model consistent, bound-

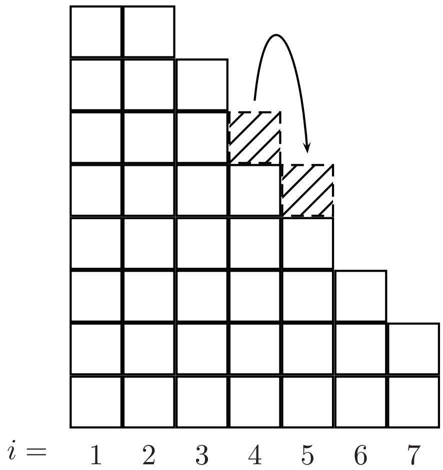


Figure 1.1: A relaxation event in the BTW model. The column of grains at site $i-1=4$ topples, because its slope ($z_{i-1} = h_{i-1} - h_i$) is above the threshold, here $z_4^c = 1$. The toppling grain is shown as a dashed and hatched box. Note that in the next step column i will topple, since it has become unstable, and an avalanche of total size 2 will have formed. Figure drawn after (Jensen, 1998).

aries are introduced: h_0 never topples and h_{N+1} is always 0 (see Fig. 1.1), so that there is a sink (h_{N+1}) and a source (random sprinkling) for height and the total slope becomes simply h_1.

One can streamline the definition a bit in order to generalise it to higher dimensions. The height variables h_i get abandoned and only the slopes z_i are considered. Moreover the perturbation becomes non-conservative with respect to z_i:

- Perturbation: $z_j \to z_j + 1$
- Relaxation: For $z_j > z_c$ (z_j is said to be "unstable") do $z_j \to z_j - q$ and $z_{k.\text{nn}.j} \to z_{k.\text{nn}.j} + 1$ for all q nearest neighbours k of site j.

Here q denotes the coordination number, so that the model is now defined on arbitrary lattices in arbitrary dimensions. Since the perturbation is non-conservative, units of slope need to dissipate somewhere (note that the relaxation is conservative with respect to this quantity). Different scenarios are possible at the boundaries: Either a unit of z falls over the edge, *i.e.* dissipates, or q is reduced to the true number of recipients for events at the boundaries, *i.e.* relaxation is conservative again. In any case there must be dissipative processes somewhere, otherwise the model will finally reach a state where it runs forever.

The most important observable in these models is the avalanche size and its distribution. An avalanche is the entire relaxation process after a perturbation until the system reaches a new stable configuration, its size is the number of times sites have discharged.

As the BTW model is somehow the foundation of SOC, it has been studied in great detail. In one dimension (height picture, right hand end open), one can easily derive the probability density function for different avalanche sizes, or "avalanche size distribution" for short. The system develops into a staircase-like configuration and arrives there back after each relaxation. The avalanche size is simply determined by the distance between the location of the perturbation and the right hand end, where the sand leaves the system. The avalanche size distribution is therefore uniform with a sharp cutoff, $\mathcal{P}(s) = (1/L)\theta(L-s)$.

The standard form of a scalefree distribution is "simple scaling"[2], as known from classical critical phenomena such as percolation:

$$\mathcal{P}(s) = as^{-\tau}\mathcal{G}(s/s_0) \quad \text{for} \quad s > s_l \tag{1.4}$$

with a system specific constant a, the lower cutoff c_l, and critical exponent τ. Moreover $s_0 = bL^D$ with D another independent exponent and b another system dependent amplitude. The function $\mathcal{G}$ is the scaling function and expected to be universal, *i.e.* it is the same among all members of the universality class. To bring $\mathcal{P}(s)$ of the one-dimensional BTW model in this form, one writes

$$\mathcal{P}(s) = \frac{1}{s}\left(\frac{s}{L}\theta\left(1-\frac{s}{L}\right)\right) \tag{1.5}$$

so that $a = 1$, $\tau = 1$, $D = 1$, $b = 1$ and $\mathcal{G}(x) = x\theta(1-x)$. This, however, is a somewhat pathological, since $\mathcal{G}(0) = 0$. Usually it is expected that the scaling function acquires a finite value at $x = 0$, so that for sufficiently large system sizes [see (1.4)] tends to $\mathcal{P}(s) = as^{-\tau}$. This, however, is terribly misleading, as it is *a priori* unknown, whether this asymptotic regime is reached yet or not. The standard technique to identify the exponent is therefore a data collapse (Bhattacharjee and Seno, 2001).

In two dimensions the exponent has been found numerically to be $\tau \approx 1.22$ (Dorn *et al.*, 2001).

One of the most important properties, which has lead to an impressive amount of exact results [for example certain height probabilities, see (Priezzhev, 1994), but also (Priezzhev *et al.*, 1996; Ktitarev *et al.*, 2000)], is the fact that the BTW model in the definition above is Abelian: Not only do so-called *allowed states* form a group, but the outcome of an operation of the relaxation rule on a configuration is independent of the order in which it is performed, if more than one site is unstable (Dhar, 1990, 1999c)[3]. Knowing in advance the number of times all sites will topple within a single

[2] Actually simple scaling describes distribution with a *single relevant scale*, that is divergent in some limit (here the thermodynamic limit) and therefore *leads asymptotically* to a scale*free* distribution.

[3] If an over-critical site in the BTW model would redistribute *all* slope units to its neighbours, not only "bunches" of q, the model would *not* be Abelian.

avalanche even allows one to perform the topplings in arbitrary order. This is possible since the change in z for a given site depends only on the number of times it has received a unit from its neighbours and the number of times it has toppled.

1.3.2 The OFC model

The OFC-model, named after **O**LAMI, **F**EDER and **C**HRISTENSEN, is derived from the BURRIDGE-KNOPOFF spring-block model. It was supposed to model earthquakes, and is therefore also known as the earthquake model in literature (Olami *et al.*, 1992). Like the BTW model, it lives on a d-dimensional hyper-cubic lattice, but unlike the BTW, a *continuous* variable $0 \leq z_i < z_c$ called "force", where $z_c = 1$, is assigned to each node i. What distinguishes this model from most other models, is the fact that it is completely deterministic, as we shall see in the updating rules:

- Perturbation: Find the largest value $z_m = \max\{z_i | 0 < i \leq N\}$ among all N sites. Increase the variable on all sites by $z_c - z_m$.
- Relaxation: Distribute the force of all unstable sites j with $z_j \geq z_c$ to their q nearest neighbours, by assigning $z_{k.nn.j} \to z_{k.nn.j} + \alpha z_j$. Now set $z_j \to z_j = 0$. Repeat this step until $z_i < z_c$ for all sites. An uninterrupted sequence of relaxation events constitutes an avalanche. The parameter $\alpha \in [0, 1/q]$ is called degree of conservation, because it measures to what extend force is dissipated.

These rules represent a simplified mechanism of the way stress spreads in faults caused by and causing tectonic movement: If a volume element relaxes, its change in strain gives rise to a change in stress for its neighbours. In leading order this effect is linear. Because of the anisotropy of tectonic plates, the *two* dimensional OFC model is therefore thought of as a simplified model of earthquakes. If it contains all the important ingredients, it should show the right exponents.

After the perturbation step, which can also be seen as a continuous drive, *at least one* site (in simulations virtually always *exactly one*) is critical, *i.e.* has a value $z_m = z_c$. However, the fact that more than one site can topple initially, seems to represent a breakdown of its deterministic behaviour: The problem is that the relaxation is supposed to take place in parallel and if two neighbouring sites topple, they could topple on top of each other and therefore the outcome of such a relaxation is dependent on the order. The relaxation does not commute, *i.e.* the model cannot be Abelian! One usually solves the problem by performing the relaxation in parallel: All unstable sites topple at once and all resulting contributions of force to other sites are added afterwards, *i.e.* if two neighbouring sites topple at once, they first both topple and receive the contribution from each other only afterwards. This cures the model from being undetermined within a single update during an avalanche, but does not make all updates commuting.

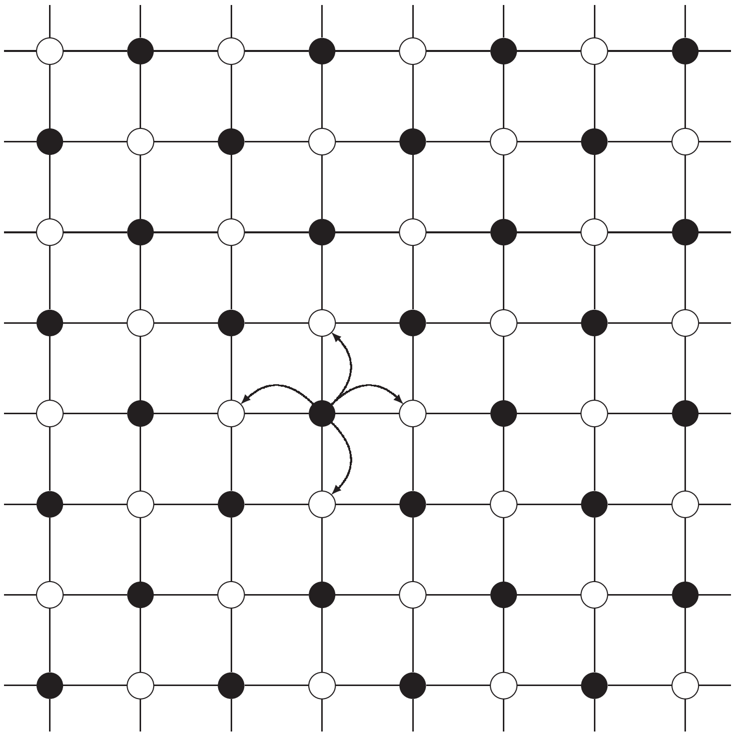


Figure 1.2: The two sub-lattices of a square-lattice, marked by filled and white circles. The set of all nearest neighbours of one sub-lattice is the other one. If only sites on one sub-lattice topple at one time step, only sites on the other will topple in the next one.

As shown now, the problem occurs on a hyper-cubic lattice only if more than one site *initially* topples at the same time. If there is only one so-called initial seed, the problem cannot occur: If all relaxing sites are on one sub-lattice of the square lattice (see Fig. 1.2) they can trigger events only on the other sub-lattice. So in the next time step, all events will take place on the other sub-lattice. The problem mentioned above occurs therefore only if the initial seeds are on different sub-lattices. And this is the only case where one needs to keep track of contributions and order of toppling as described above — in the case of only one initial seed the relaxations within each time step commute (but still not within an entire avalanche).

It is clear from the beginning that the statistical properties of the model depend on the degree of non-conservation α. For $\alpha = 1/q$ it is completely conservative. As in the BTW model, full conservation of the dynamical variable drives the system inevitably into a state where it runs forever. Periodic boundary conditions are therefore not reasonable in the conservative limit, and the standard boundary conditions are open like in the BTW model. For the non-conservative regime, it is well established that periodic boundaries lead — after a transient — to a periodic state (Jensen, 1998).

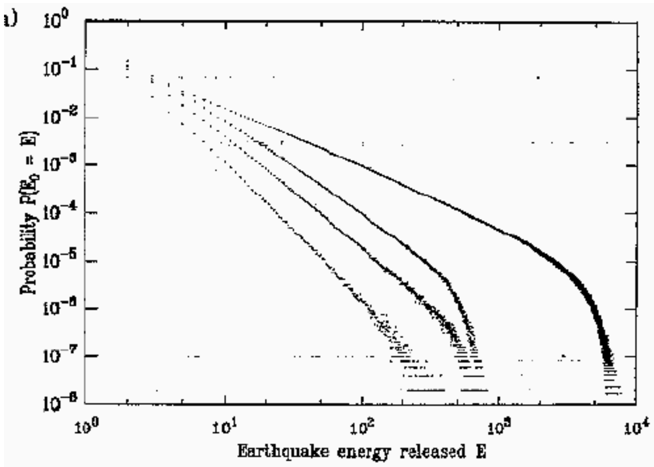


Figure 1.3: Cited after (Olami *et al.*, 1992, Fig. 2a): "Simulation results for the probability density of having an earthquake of energy E as a function of E for a 35×35 system. The curves correspond to $\alpha = 0.25, 0.20, 0.15$ and 0.10. The slope of the curves becomes steeper as the α value is decreased."

So, some open boundaries are necessary in any case. It has been argued that this inhomogeneity is the cause for statistics which resemble scalefree distribution (Middleton and Tang, 1995). The role of anisotropy and non conservation is still unclear: Starting out from some LANGEVIN equations, some authors argue that either of them are sufficient to observe SOC (Grinstein *et al.*, 1990).

The standard observables of the OFC model are the total amount of energy redistributed, the number of sites which have toppled, and the total dissipation, where each relaxation event contributes $z_j(1 - q\alpha)$. Other observables are spatial, like the area covered by the avalanche. The distribution of the avalanche sizes turns out to be scalefree and the exponent depends on α. For $\alpha \approx 0.2$ and $d = 2$ the exponent found relates very well to the exponent B of Gutenberg-Richter law found in real earthquakes mentioned in sec. 1.2.

That the exponents are continuous in α is highly remarkable, since continuous exponents to a certain extent undermine the concept of universality which relies on the idea that exponents depend only on dimensionality, symmetries and other discrete parameters, rather than a tunable quantity. An example for the histograms and different exponents is shown as a double logarithmic plot in Fig. 1.3 [the figure is copied from the original *Phys. Rev. Lett.* (Olami *et al.*, 1992)].

It is still debated whether and if so for which value of α the OFC model is truly critical. It is widely accepted that it is scalefree in the conservative limit, *i.e.* $\alpha = 1/q$. For the other extreme, $\alpha = 0$ it is easy to see that the sites completely decouple — but there is no proof that the behaviour for $\alpha \to 0$ corresponds to $\alpha = 0$.

It is — actually — easy to see that the random neighbour version of the OFC model cannot have a divergent average dissipation (Lise and Jensen, 1996; Chabanol and Hakim, 1997; Bröker and Grassberger, 1997) and therefore to exclude exponents ≤ 2 for the dissipation distribution. The argument is that in the random neighbour model the energy inflow between two topplings is constant, because the density of sites at z_c scales as N, N being the total number of sites in the system, and therefore the average distance between z_m and z_c scales as $1/N$. The global inflow of energy is then $N(z_c - \langle z_m \rangle) \in \mathcal{O}(1)$. But this means that the average outflow is also constant (see also (1.7)). Mapping the model onto a random walker along an absorbing wall (Bröker and Grassberger, 1997) to derive the exponent for the dissipation rate and onto a branching process for the distribution of durations, produces the standard (Mean Field) exponents $3/2$ and 1. This rules out criticality in the non-conservative regime at the same time.

1.3.3 The Forest Fire model

This subsection only sets the Forest Fire model in the context of other models. There is an entire chapter dedicated to its detailed analysis (see Chapter 4).

The first model, the BTW-model, is conservative with respect to the dynamical variable. The OFC-model is either conservative or non-conservative, depending on the parameter α, but might be considered as a model of extremal dynamics, as described below. The Forest-Fire Model (FFM) is the standard model of SOC for complete non-conservation: The dissipative events are complete and happen only within the bulk. There is no spreading of force, just like $\alpha = 0$ in the OFC model. The usual boundary conditions for the FFM are periodic.

The strength of the external drive in all lattice models described so far is either unity or given by the system size. For the response of the systems to an external perturbation to diverge, it is therefore necessary that either energy can be accumulated, or the external drive must diverge.

In the case of the Forest Fire model the external drive is a tunable parameter and the model becomes supposedly scalefree in the thermodynamic limit for divergent external drive. One might argue that this is a form of external tuning, but the divergence of the external drive is a natural requirement and trivial in the sense that not an external drive of particular strength is necessary.

A first version of the FFM was proposed by BAK, CHEN and TANG (Bak *et al.*, 1990) but later turned out (Grassberger and Kantz, 1991; Moßner *et al.*, 1992) not to be critical. The following definition is known as the *critical* DROSSEL-SCHWABL Forest Fire model (Drossel and Schwabl, 1992): As in standard site percolation, on a d-dimensional hyper-cubic lattice each node is either occupied or empty. A configuration of such a lattice is updated as follows:

- Choose randomly θ^{-1} sites, one after the other. If its state is "empty" turn it into "occupied".
- Choose one site randomly. If it is empty continue with the first step. Otherwise turn the state of all sites in the entire cluster with which the chosen site is connected to "empty" (a cluster is defined in the standard fashion via nearest neighbour interactions). In this case one says "the cluster was struck by lightning" and the update is considered to be successful. Continue with the first step.

This model becomes critical for $\theta^{-1} \to \infty$, in the sense that the cluster size distribution becomes scalefree, as known for example in percolation. However, the definition above is actually already a simplification. It is instructive to see how it is derived from the original definition of the critical model, which contained a third state of sites, namely "burning". All lattice sites are updated simultaneously according to one of the following rules (Jensen, 1998):

- An empty site becomes occupied with probability p.
- A burning site becomes empty.
- An occupied site becomes burning, if at least one of its nearest neighbours is burning or ...
- ... with probability f, independent of the state of the neighbours.

It is immediately clear that this definition lives on a microscopic timescale: In the definition before, thousands (according to the parameter θ^{-1}) of trees are "planted" at once and fires are instantaneous (second step). In the second description, the model obviously contains at least two timescales, p^{-1} and f^{-1}. The latter is clearly the frequency of fires measured in microscopic time-units, *i.e.* in update steps as used in the second set of rules. It is reasonable to impose

$$p, f \to 0 \tag{1.6}$$

for criticality, as the corresponding timescales must be large compared to the microscopic one. This means, that fire spreading (the process which defines the microscopic timescale) must become instantaneous. But more than this is required for this model to develop into the critical state, as it is shown now:

In the stationary state, the inflow of trees by growing must compensate the outflow of trees by burning, so that

$$p\rho_e = f\rho_o\langle s\rangle \tag{1.7}$$

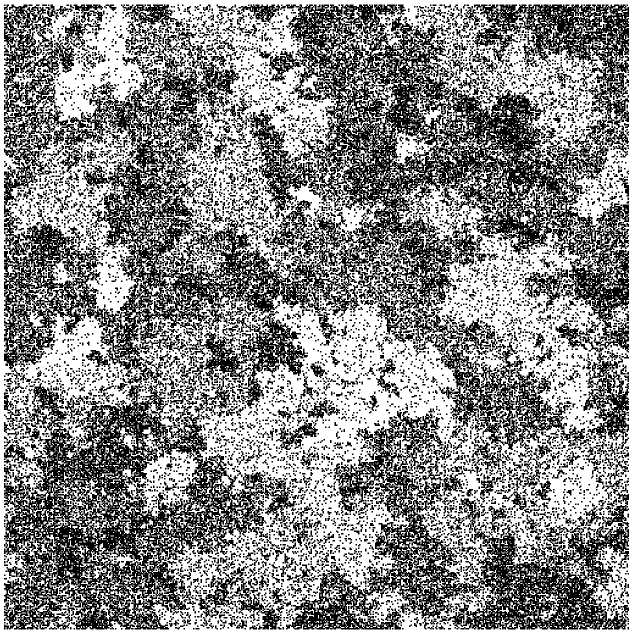


Figure 1.4: The two-dimensional Forest Fire model. Occupied sites are black. The model develops a patchy structure because regions which become completely empty due to the removal of a cluster become homogeneously reforested.

where ρ_e (ρ_o) denotes the density of empty (occupied) sites and $\langle s \rangle$ the expected size of fires (Clar *et al.*, 1996). In the thermodynamic limit, this is *not* a mean field or single side approximation, as one might suspect — to see that one simply multiplies both sides of (1.7) by a number of time steps T, so that the LHS is a total number of trees grown in this time and the RHS is the product of $Tf\rho_o$ and $\langle s \rangle$. The first factor is just the expected number of times a fire was triggered, the second factor its average size, so that the product becomes the total number of burnt trees.

Provided that neither ρ_e nor ρ_o behave singularly, the average fire size scales as

$$\langle s \rangle \propto \frac{p}{f} \quad . \tag{1.8}$$

This makes sense, as p/f is related to the number of trees grown between two burning events. For a large scale structure to form it is therefore necessary to have $p \gg f$. Together with the requirement (1.6) from above this gives

$$1 \gg p \gg f \tag{1.9}$$

This however is not enough, because it just fixes the relation between the microscopic timescales (spreading, growing, triggering). In addition the macroscopic timescale needs to be related. This is the relation of the entire burning process to growing and burning, measured in microscopic units. The entire burning (not only the spreading!) should be instantaneous compared to both of them, so especially

$$\frac{p}{f} \ll p^{-1} \tag{1.10}$$

if one assumes the timescale of burning to be of the same order as the average burnt cluster size [the relation between burning time and size is certainly more complicated, but this does not matter for the argument, for details see (Clar *et al.*, 1996)]. Both sides of (1.10) are measured in units of the microscopic timescale, which is assumed to be dimensionless. Both relations (1.9) and (1.10) together can be written in the form of timescales as

$$\frac{p}{f} \ll p^{-1} \ll f^{-1} \tag{1.11}$$

This relation is known as the double separation of timescales in the FFM: Burning time is much shorter than growing time which in turn is much shorter than lightning.

By introducing burning, which is perfectly instantaneous compared to growing and triggering, one realizes the first inequality in (1.11) perfectly and all that is left is

$$p^{-1} \ll f^{-1} \quad . \tag{1.12}$$

In such a system the shortest timescale is p^{-1}, so one can simply write everything in terms of this scale. Trees are then grown with frequency 1 and fires are triggered with frequency $f/p \ll 1$ (as all frequencies are then ≤ 1, they can be considered as probabilities). Such a model is then exactly represented by

- Choose a site at random. If it is occupied go to the next step. Otherwise turn it into occupied. Continue with first step.
- With probability f/p trigger a fire, *i.e.* remove the entire cluster connected to the site chosen. Continue with the first step.

On average p/f growing trials happen before a fire is triggered (Poisson process). A possibly more computer friendly version of the rules is therefore the first one given in this section, where $\theta^{-1} = p/f$.

The observable in the FFM is the cluster size distribution of the burnt cluster. As this cluster is chosen randomly, the distribution of burnt clusters is simply the size-biased distribution of the global cluster size distribution $n(s)$, which is the site normalised density of clusters of size s in the system, so that the distribution of the burnt cluster then reads $n(s)s$. This distribution again is assumed to behave like

$$n(s) = s^{-\tau}\mathcal{G}(s/s_0) \tag{1.13}$$

where $\mathcal{G}$ is the scaling function. Further observables consider mainly the geomet-

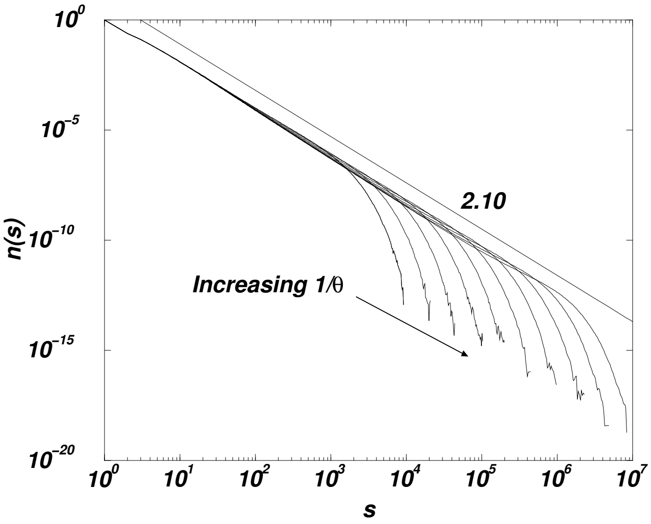


Figure 1.5: The cluster size distribution $n(s)$ of the FFM. The arrow points in the direction of increasing θ^{-1}, which are from left to right 125, 250, 500, 1000, 2000, 4000, 8000, 16000, 32000 and 64000. The exponent is roughly 2.10, but apparently fits quite well to 2.14 for smaller values of θ^{-1}.

ric properties of the clusters, for example the maximum Manhattan distance from a randomly chosen starting point to all other points within the same cluster.

The FFM is obviously closely related to percolation. What makes it different from standard percolation is the presence of additional *global* occupation correlations; these correlations are different from those usually investigated in percolation, which are quantified by the correlation function for sites being occupied and belonging to the same cluster. The global two-point correlation function for occupation is just a δ-peak in the case of percolation, as occupation is uncorrelated. This is not true for the FFM, where the long-ranged removal procedure introduces correlations, by setting the average density of sites within a region which was struck by lightning to the same value. Those "patches" which are uniformally reforested afterwards then show a correlation for their occupation: If two sites within the same (former) patch are considered and one of them is occupied, the other one is occupied with a certain probability depending on the time since the lightning, even if the two sites are now not in the same cluster anymore.

In spite of its differences to standard percolation, the FFM has inherited its notation and some simulation techniques from percolation as well as the value of its upper critical dimension, which is expected to be 6 (Christensen *et al.*, 1993; Clar *et al.*, 1994). An analytical approach to the FFM is still lacking, even for $d = 1$ the results are disappointingly rough and vague (Drossel *et al.*, 1993). Therefore virtually all results for the FFM are numerical; they consistently suggest that the FFM is not in the same universality class as percolation, *i.e.* the exponents consistently differ and converge only above the upper critical dimension, where both models become trivial (*i.e.* correlations of whatever nature become irrelevant) and they simply share the same theory.

SOC is notorious for the problem that careful investigations of the models reveal an unexplainable variety of effects and structures. The FFM is not an exception (Pruessner and Jensen, 2002a; Grassberger, 2002) — recent speculations (Schenk *et al.*, 2002) try to map it again on percolation (at least for $d = 2$) to account for defects in the scaling behaviour of $n(s)$. The idea is that the FFM is effectively a superposition of percolation histograms due to the patchy structure as shown in Fig. 1.4. If a local density p has frequency $w(p)$, the resulting cluster size histogram is

$$\int_{0,1} dp w(p) n_{\text{perc}}(s,p) \propto s^{-(\tau_{\text{perc}}+\sigma)} \quad . \tag{1.14}$$

where $n_{\text{perc}}(s,p) \propto s^{-\tau_{\text{perc}}} \mathcal{G}(-s/s_0)$ is the histogram of standard percolation. As usual $\mathcal{G}$ denotes the cutoff function with $s_0 \propto |p - p_c|^{-1/\sigma}$.

1.3.4 Extremal dynamics: the BAK-SNEPPEN model

The OFC model is in a certain sense a model of extremal dynamics, which in general means that the dynamics is defined with respect to the extreme value among local variables. The BAK-SNEPPEN (BS) model (Bak and Sneppen, 1993) is probably the most carefully studied model of this class. Defined in general again on a d-dimensional lattice with N nodes, each of them, identified by i, gets assigned a fitness $z_i \in [0, 1]$ and is updated as follows:

- Find the smallest value $z_s = \min\{z_i | 0 < i \leq N\}$ among all sites.
- Assign this site and its nearest neighbours new fitnesses, which are randomly chosen from the interval $[0, 1]$.

Since the fitnesses are chosen randomly, the probability of a degeneracy of the minimum vanishes even for a finite system. This model was invented as an oversimplified version of biological evolution, where the extinction of the least fittest species leads to the extinction of the species depending on it. The ecological niche appearing is immediately occupied by mutants with random fitnesses.

The most important observable is the minimal fitness in the lattice,

$$f_{\text{min}}(t) = \min\{z_i(t) | 1 < i \leq N\} \quad . \tag{1.15}$$

where its time dependence is measured in units of updates. Starting from a completely random configuration, it is clear that $f_{\text{min}}(t)$ will increase often, but it cannot reach 1 and there is always a probability that a newly chosen fitness is below the

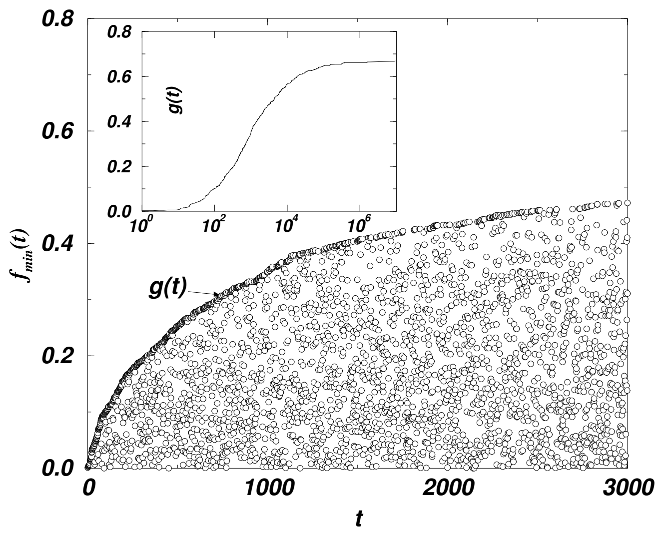


Figure 1.6: The transient behaviour of $f_{\min}(t)$ and $g(t)$ (see text). The thick line is $g(t)$ and consists of dense values of $f_{\min}(t)$, which changes often but in small steps, because the values of z_i are very dense and if the system gets rid of one small z_i, there is another, only slightly larger value somewhere else in the system. Moreover, $f_{\min}(t)$ decreases for small t only rarely because the probability for an update to arrive below the current $g(t)$ is only $g(t)$. Dimension is $d = 1$ and $N = 1000$ as in Fig. 1.8. The inset shows the long time behaviour of $g(t)$ in a log-linear plot obtained in the same simulation.

current value of $f_{\min}(t)$. The maximum value of the function within a time window defines the so-called gap function [(Paczuski *et al.*, 1996) after (Rittel, 2000)] $g(t) = \max\{f_{\min}(t')|0 \leq t' \leq t\}$. It is the largest free region between 0 and the fitnesses within all time steps up to t. The evolution of this quantity is shown in Fig. 1.6 for short times and logarithmically for long times in the inset of the same figure. Since the actual value of $f_{\min}(t)$ is caused by randomly chosen values, this quantity seems to be distributed completely randomly under the envelope $g(t)$. Looking a bit more carefully reveals that there is a gradient towards smaller values in the density of $f_{\min}(t)$ values below $g(t)$. This makes sense, because if there is more than one $z_i < g(t)$ then $f_{\min}(t)$ will be the smallest one, *i.e.* for $f_{\min}(t) \neq g(t)$ this quantity is more likely to be close to 0 than to $g(t)$. Since the updated value of z_i is chosen randomly, it is also not surprising that the PDF (probability density function) above $g(t)$ is uniform. Fig. 1.7 shows the distribution for different system sizes in the stationary state.

Fig. 1.8 shows three different configurations of the model with time increasing from left to right: The homogeneous initial configuration, a transient configuration, which has typically almost no z_i below $g(t)$ and a late configuration, when $g(t)$ changes only very slowly and the typical configuration has a number of sites in the gap.

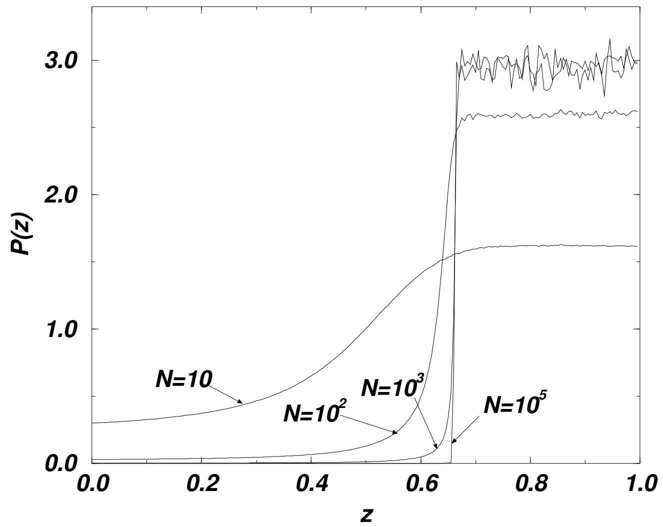


Figure 1.7: The distribution of fitnesses for different system sizes. For $N = 10^5$ the jump at f_c becomes very sharp. Note the strong fluctuation in the large z region. All data from 10^8 updates for equilibration and the same number for statistics.

From its definition, it is clear that $g(t)$ is monotonic and bound, *i.e.* it must converge. However, it is not necessarily increasing, so its limit is *a priori* unknown. For a finite system, there is a finite probability density for all fitnesses being 1, so for any finite system $\lim_{t\to\infty} g(t)$ must be 1. However, this is not the value suggested in Fig. 1.6. But applying first the thermodynamic limit and starting from a completely random configuration, means that $\lim_{t\to\infty} g(t) = 0$, as there are infinitely many sites at (around) 0 (for the mathematician it might be more appropriate to say that the entire real interval $[0, 1]$ consists of cluster points, *i.e.* $\mathbb{Q}$ and $\mathbb{R}$ are dense). Fig. 1.7 helps to understand the asymptotic behaviour of $g(t)$: If the fitnesses are distributed according to the PDF shown, a further increase of $g(t)$ in time will be stopped by the jump in density. However, for this mechanism to work perfectly one needs the thermodynamic limit, otherwise $g(t)$ just starts to creep at the edge of the distribution. Therefore, it does not seem to be reasonable to define an asymptotic value of $g(t)$ in any sense. The magic number around $2/3$, as seen in Fig. 1.6 and Fig. 1.7 must be quantified using a different approach.

An f_0-avalanche is defined as the event when all fitnesses initially above a threshold f_0 are perturbed such that for a certain time there are some below f_0. The event ends as soon as all fitnesses are again above f_0. For a certain value of f_0, namely $f_0 = f_c$ one obtains scalefree avalanches, *i.e.* their distribution in size and duration follows a power law. The exponent of this power law is not easy to measure and still debated (Datta *et al.*, 2000; Grassberger, 1995). As mentioned above, the distribution above f_c seems to be uniform, and asymptotically one expects a step function for the

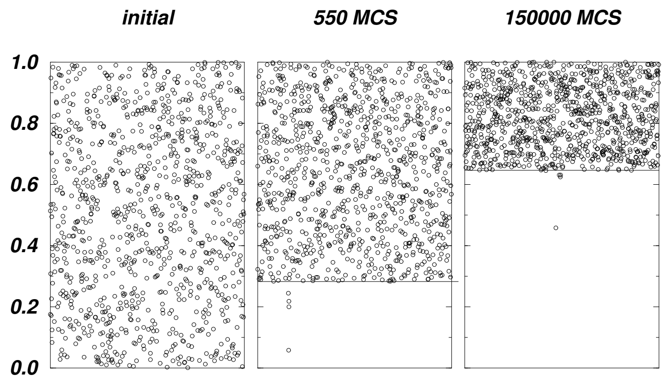


Figure 1.8: The configuration for a system of size $N = 1000$ after three different t. Typically only a few number of sites are active (below $g(t)$) and even less for early t. The value of the gap function is indicated by a straight line.

distribution of fitnesses. The idea is that if the system is set up in this way, small perturbation will produce scalefree $f_0 = f_c$-avalanches, but these avalanches are *just* not over-critical, *i.e.* the sum over the probabilities of finite avalanches is exactly 1. And choosing a slightly larger f_0 will, in the thermodynamic limit, result in infinite avalanches. It is therefore natural to define f_c as the f_0-value of avalanches to become scalefree. This would be a dynamical definition of f_c.

It would be mathematically much more convenient to define f_c, which somehow holds the key for almost all effects obtained in the model, using only static or stationary properties of the system. This could be done as follows: Let $\mathcal{P}(f,t;N)$ be the PDF of the fitnesses f in a system of size N at time t, then f_c is the point, where $\mathcal{P}(f,t;N)$ makes a jump. The thermodynamic limit must be applied after $t \to \infty$, otherwise P does not change in time, so we define

$$\widetilde{P}(f) = \lim_{N\to\infty} \lim_{t\to\infty} \langle \mathcal{P}(f,t;N) \rangle \tag{1.16}$$

where the average $\langle\rangle$ runs over an ensemble. We assume that both limits exist. Then $f_c \in]0,1[$ is the point, where $\widetilde{P}(f)$ is discontinuous, *i.e.* where

$$\liminf_{f\to f_c} \mathcal{P}(f) \neq \limsup_{f\to f_c} \mathcal{P}(f) \tag{1.17}$$

A simple but extremely slow version of the model is easy to implement, see Fig. 1.9. The major flaw is that the entire lattice is scanned again and again for finding the minimal site. It is much more efficient to simply keep track of it; if all sites are somehow organised according to their value, one only needs to update this structure, whenever the value of a site changes. These data are probably best stored in form of a binary search tree (in (Grassberger, 1994) a hashed, degenerate tree is proposed for

```
#include <stdio.h>
#include <stdlib.h>

/* Some parameters */
#define N (1000)                 /* System size, i.e. length */
#define INIT_THRESHOLD (0.0)     /* Init sites ran-
domly above this thershold */
#define MAX_LOOPS (1000000)      /* Number of itterations */

/* Use the poor(?) random number generator provided by libc.
 * They are choosen from [0,1] rather than [0,1[ */
#define RND() (((double)rand())/((double)RAND_MAX))

double site[N];

int main(int argc, char *argv[])
{
double g;
int min_i, i, loop;

/* Init */
for (i=0; i<N; i++)
  site[i]=RND()*(1.0-INIT_THRESHOLD)+INIT_THRESHOLD;
g=-1.0;

/* Main loop */
for (loop=0; loop<MAX_LOOPS; loop++)
  {
  /* Find minimum */
  for (min_i=0, i=1; i<N; i++)
    if (site[i]<site[min_i]) min_i=i;
  if (site[min_i]>g) g=site[min_i];
  printf("%i %g %g\n, loop, site[min_i], g);

  /* Update the minimum and its two neighbours */
  site[min_i]=RND();
  site[(min_i+1)%N]=RND();
  site[(min_i-1+N)%N]=RND();
  }
return(0);
}
```

Figure 1.9: A very simple implementation of the BAK-SNEPPEN model as a C-program.

a similar purpose in the OFC model).

This remarkably simple model has proven to be rock stable against any attempt to solve it. How frustratingly slow the progress in this field is can be seen in a recent impressive work of RONALD MEESTER and DIMITRI ZNAMENSKI (Meester and Znamenski, 2001), who prove for a simplified model the existence of a mean fitness, which is bounded away from 1.

It should be mentioned that the value of f_c is given in (Paczuski *et al.*, 1996) as $f_c = 0.66702(3)$ in (d=1) and $f_c = 0.328855(4)$ in $d = 2$. GRASSBERGER finds $f_c = 0.66702(8)$ for the one-dimensional case (Grassberger, 1995). Although one-dimensional models are mostly designed to be solvable, the BS model is notorious for its unusual behaviour. For example $f_c \neq 2/3$ is generally accepted.

The random neighbour version of this model is fairly straight-forward to solve (Rittel, 2000). The exponents are then those of the critical branching process (time-exponent 1, where time needs then to be redefined) and the random walker (size exponent 3/2) as already mentioned above. That the nearest neighbour property is of fundamental importance becomes clear when calculating f_c in the the random neighbour model: The average number of newly activated sites must be 1 for each update, *i.e.* $f_c = 1/3$ in case of overall 3 updates.

1.4 The VESPIGNANI mechanism[4]

For obvious reasons, it is very interesting to try to reduce SOC behaviour to established critical phenomena. The hope is that if a mapping of SOC models to existing, well-understood models can be done, it would be possible to identify "where criticality comes from" in SOC.

One such ansatz, under the rather ambitious title "How self-organised criticality works"[5] was published by VESPIGNANI and ZAPPERI (Vespignani and Zapperi, 1998)[6] and re-adopts ideas first introduced by TANG and BAK (Tang and Bak, 1988)[7]. It was later explained in detail in (Dickman *et al.*, 2000). The aim is to identify a link between SOC and absorbing state phase transitions (AS). To describe the latter in detail is beyond the scope of this thesis; we refer especially to the review by HINRICHSEN (Hinrichsen, 2000).

In the following, AS is briefly introduced to set the stage for the "VESPIGNANI mechanism", that is the mechanism VESPIGNANI proposes as an explanation for SOC. In the following subsections the fundamental shortcomings of this mechanism are explained. One should stress that the VESPIGNANI mechanism enjoys increasing acceptance in the SOC community.

1.4.1 A brief overview of AS

The paradigmatic example of AS is directed percolation (DP), which is described in the following for a $1+1$-dimensional system, as a dynamical model. The dynamical interpretation is more obviously understandable as an absorbing state phase transition; the notation of the dimensionality $1+d$, refers to an anisotropy, where one dimension acquires a special rôle. In the dynamical representation, this dimension is time.

[4] I would like to thank OLE PETERS for the many discussions we have had about this subject.

[5] In obvious relation to the even more ambitious title of the popular science book by PER BAK (Bak, 1996).

[6] There is a second issue raised in this paper, which is whether the critical behaviour of SOC can in principle be recovered from an AS-type study. There are two contradicting papers with an emphasis on this problem, both by VESPIGNANI, DICKMAN, MUÑOZ and ZAPPERI, (Vespignani *et al.*, 1998) and (Dickman *et al.*, 2000).

[7] Thanks to OLE PETERS for pointing that out to me.

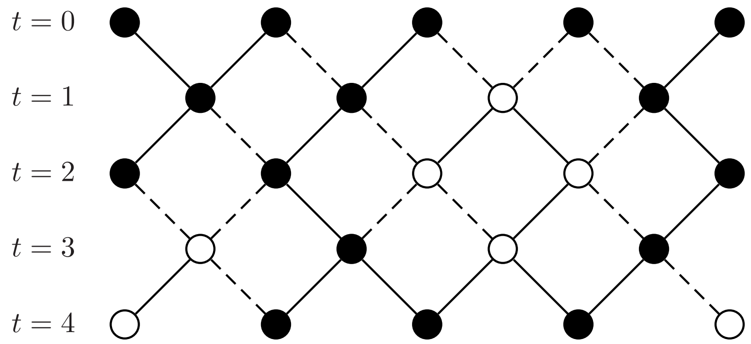


Figure 1.10: An example of directed percolation [similar to figure 11 in the preprint of (Hinrichsen, 2000)]. The time evolves from the top row to the bottom row. Bonds are active (solid lines) with probability p, otherwise they are inactive (dashed lines). Sites in generation t are occupied (black circles), if either their left or right neighbour in the preceeding generation ($t-1$) is occupied and connected via an active bond. Otherwise they are unoccupied (white).

Fig. 1.10 shows an example of DP in the dynamical interpretation: starting from a completely occupied lattice at time $t = 1$, sites in the next generation $t + 1$ are occupied, if they are connected to an occupied site in the preceeding generation t via an active bond. Bonds are active with probability p. In this interpretation, a new one dimensional lattice is generated from the preceeding one in every time step, independent from any history, *i.e.* the probability distribution of the configuration of the lattice is a Markovian process (van Kampen, 1992), parametrised by the probability p.

This model possesses exactly one absorbing state, namely the empty lattice: Once the lattice is empty, there is no way to escape from this configuration. Thus, the configuration is, what is known as "absorbing" in stochastic processes (van Kampen, 1992). However, even though the probability to reach the absorbing state is finite for $p < 1$ and any finite system, this probability decays extremely fast in the system size at sufficiently large p. In this large p region, the system virtually never reaches the absorbing state and remains in the active transient for a very long time. Therefore, this regime is called the "active phase".

The most commonly investigated observable is the stationary occupation density, which is the ensemble averaged fraction of occupied or active sites in the stationary state. If $\left\langle \rho_a^{AS}(t) \right\rangle$ denotes the ensemble averaged fraction of active sites after time t,

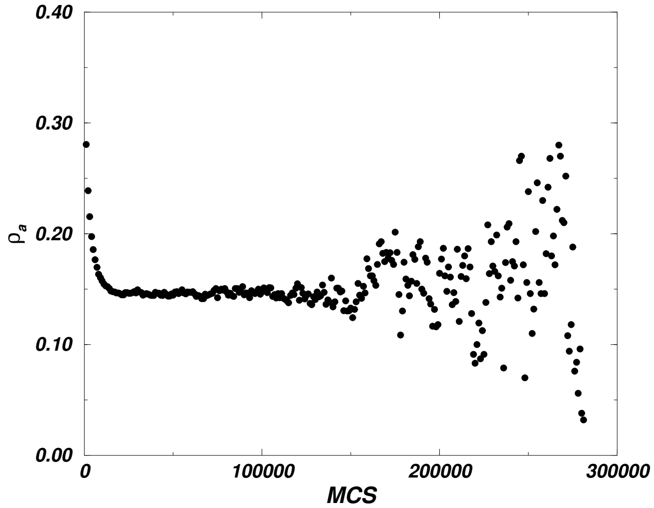


Figure 1.11: An example of the temporal evolution of the density of active sites in directed percolation at $p = 0.6435$ for a one-dimensional system of size $L = 500$. The time is given in Monte Carlo Steps (MCS).

then for any finite system and any $p < 1$

$$\lim_{t\to\infty} \left\langle \rho_a^{AS}(t) \right\rangle = 0 \; . \tag{1.18}$$

However, this is not what is numerically found. Numerically one has to stop the process after a finite number of generations and in the high-p region it becomes increasingly unlikely that the absorbing state is found within accessible times. What one is interested in, is the *stationary activity* in the thermodynamic limit,

$$\rho_a^{\text{AS, stat}} = \lim_{t\to\infty} \lim_{L\to\infty} \left\langle \rho_a^{AS}(t) \right\rangle \tag{1.19}$$

One established, numerical procedure (Dickman, 2003) to study this quantity which is theoretically inaccessible by numerics, is to take the average activity *conditional* to activity; for every given p the average activity of an ensemble of *active* realisations is taken for a series of times. By visual inspection, the activity is derived as the plateau before the onset of strong fluctuations. An example for the time-dependent conditional activity is shown in Fig. 1.11.

The stationary activity $\rho_a^{\text{AS, stat}}$ is generally accepted as the order parameter describing a continuous phase transition between the absorbing and the active phase. The transition is driven by the probability p and takes place at some non-trivial value p_c. Correspondingly, the critical exponent β is defined as

$$\rho_a^{\text{AS, stat}} = A_{\rho_a^{\text{AS, stat}}} (p - p_c)^\beta \text{ for } p > p_c \; , \tag{1.20}$$

with an amplitude $A_{\rho_a^{\text{AS, stat}}}$, just like for classical critical phenomena. Naturally, other exponents, characterising the correlation lengths in the two directions, the time-integrated activity etc., can be defined in a straight forward way. For the sake of the following arguments, the only other important exponent is the one describing the divergence of the spatial correlation length (as opposed to the temporal correlation length)

$$\xi_\perp = A_{\xi_\perp} |p - p_c|^{-\nu_\perp} , \tag{1.21}$$

again with an amplitude $A_{\xi_\perp}$. As expected from classical critical phenomena, $\rho_a^{\text{AS, stat}}$ shows *finite size scaling*,

$$\rho_a^{\text{AS, stat}}(L, p) = L^{-\beta/\nu_\perp} \mathcal{G}\left(\frac{p - p_c}{L^{-1/\nu_\perp}}\right) \tag{1.22}$$

with system size L and a universal finite size scaling function $\mathcal{G}(x)$. One expects the size Δp of the finite size scaling region to scale like $L^{-1/\nu_\perp}$.

It is most remarkable, that the understanding of $1+1$ DP is still far from an exact solution, as opposed to the extremely well understood equilibrium critical phenomena in two dimensions, such as percolation (Cardy, 1992; Langlands *et al.*, 1992; Aizenman, 1997). There are, however, series expansions (most prominently by IWAN JENSEN) of remarkable accuracy (Jensen, 1999). Although it is very hard to estimate the error of a series expansion, it is generally accepted that the exponents of DP are not rational numbers, which is very different to the general results obtained in equilibrium critical phenomena (Henkel, 1999). In this context, it is interesting to consider the recent work by CLEMENT SIRE (Sire, 2002), who finds, using an improved mean field theory,

$$\beta = \frac{1}{2}\left(1 - \frac{1}{\sqrt{5}}\right) = 0.276393202250... \tag{1.23}$$

very close to the best known estimate by a series expansion, $\beta = 0.276486(8)$ (Jensen, 1999).

1.4.2 AS picture of SOC

In terms of AS, SOC models can, quite generally, be described as follows: An SOC model has an active phase, where the updating rules operate, and an inactive phase, where the system is quiescent. Without the external drive, the SOC would stay there forever, *i.e.* this is an absorbing state. Above it has been pointed out that the driving takes place only if the system is in this state. This suggests a global mechanism, which makes sure that the system is in the quiescent state whenever a grain is introduced. In order to avoid that, one generally uses the notion of an infinitely slow drive, so that there is a constant external driving rate h, which is, however, so small that the model is effectively driven only if it is completely relaxed. From the point of view of AS, it is this external drive, which brings the SOC model from the absorbing phase back into the active phase. It would remain there forever, if dissipation would not restrict the event sizes. Denoting the fraction of sites active in the system at time t as $\rho_a^{SOC}(t)$, one might introduce a bulk dissipation rate ϵ such that the dissipation at time t is given by $\epsilon\rho_a^{SOC}(t)$; thus, ϵ is the dimensionless probability to dissipate a grain during toppling. In most of the models described in sec. 1.3, this dissipation takes place at the boundaries, but it seems that in some cases, it can be replaced by a bulk dissipation or at least there are models apparently displaying SOC with bulk dissipation (Olami *et al.*, 1992; Manna *et al.*, 1990; Vespignani and Zapperi, 1997, 1998; Dickman *et al.*, 1998; Barrat *et al.*, 1999).

This dissipation will eventually bring the system back into an absorbing state. In the stationary state of this procedure, dissipation and drive (measured as inflow per site and unit time) must compensate, so that

$$h = \epsilon \left\langle \rho_a^{SOC} \right\rangle \tag{1.24}$$

where $\langle\rangle$ denotes the ensemble average (in the stationary state). It is important to note that for sufficiently small h, the activity $\rho_a^{SOC}(t)$ at time t is most of the time 0. It is therefore not the same activity as the one described above which is a conditional average. One can see this immediately by considering very small h: by reducing the external drive further, one can make $\rho_a^{SOC}(t)$ arbitrarily small.

If now the external rate is tuned down to 0 in order to prevent the relaxation events to overlap (Corral and Paczuski, 1999) and thereby introducing a timescale, the average activity is bound to converge to 0 as well. The average size of the dissipation event $\langle s \rangle$ is the integrated activity between two driving steps; the time between two of those steps is (in appropriate units) $1/h$, so that

$$\langle s \rangle = \frac{\left\langle \rho_a^{SOC} \right\rangle}{h} = \frac{1}{\epsilon} \ . \tag{1.25}$$

This equation is indeed exact and does not neglect any correlation: If the activity is observed over a time span T, then the total number of events in this time span is $\int_0^T dt \rho_a^{SOC}(t)$. In this time $N(T)$ avalanches take place, so that

$$\langle s \rangle = \lim_{T\to\infty} \frac{\int_0^T dt \rho_a^{SOC}(t)}{N(T)} = \left\langle \rho_a^{SOC} \right\rangle \lim_{T\to\infty} \frac{T}{N(T)} = \left\langle \rho_a^{SOC} \right\rangle \frac{1}{h} \ . \tag{1.26}$$

From the two equations, (1.24) and (1.25), it is clear that two limits are needed in order to see scale-invariant behaviour: $h/\epsilon \to 0$ in order to arrive at $\rho_a^{SOC} = 0$ and $\epsilon \to 0$ in order to arrive at a divergent avalanche size. Even if one does not impose

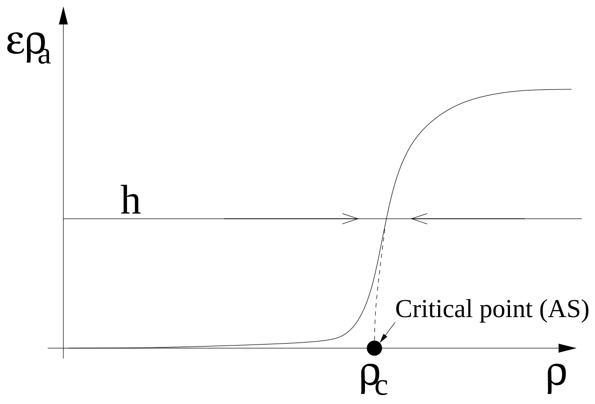


Figure 1.12: An illustration of VESPIGNANI's mechanism to explain SOC by absorbing state phase transition. The dissipation is proportional to ρ_a^{SOC} and occurs with rate ϵ, the external drive runs at constant rate h. The intersection of h and $\epsilon\rho_a^{SOC}$ determines the particle density ρ in the stationary state.

a divergent avalanche size, any finite ϵ seems, at least on a handwaving level, to suppress system spanning events, *i.e.* it establishes a characteristic size of dissipative events. In total, one arrives at $h, \epsilon \to 0$ and $h/\epsilon \to 0$.

The limit $h \to 0$ is actually meaningless in this form, because it could be implemented by a redefinition of the units of time. What is actually meant by this limit is that the time an avalanche runs should be much shorter than the time between two "kicks" of the system. In the Forest Fire model, where similar limits are needed, this has been properly identified [for example (Clar *et al.*, 1996)] very early in the literature, see also the discussion on timescales in the Oslo model Sec. 6.1.3.3, page 253.

In SOC all these limits are perfectly obeyed: h is always chosen to be a slow rate compared to any other mechanism, ϵ vanishes implicitly with the system size and ϵ/h diverges, even though not by explicit tuning.

Fig. 1.12 shows a cartoon[8] of the dissipation $\epsilon\rho_a^{SOC}$ as a function of the overall density of particles, ρ, which, in turn, depends on the external driving rate h and the dissipation. On the microscopic timescale[9] the resulting equation of motion for ρ can be written as

$$\dot{\rho} = h - \epsilon\rho_a^{SOC}(t) \ , \tag{1.27}$$

[8] It would certainly be very useful to produce real numerical data for such a diagram. However, it requires a meaningful definition of activity in the model under consideration. Open task

[9] Just like in the Forest Fire model, the microscopic timescale is where local updates take place, while the macroscopic timescale evolves from avalanche to avalanche. On the macroscopic timescale avalanches are instantaneous.

which, remarkably, contains global measures.[10] Clearly, in the stationary state $\langle \dot{\rho} \rangle = 0$ and therefore Eq. (1.24). As illustrated in Fig. 1.12, the density ρ increases whenever $h > \epsilon \rho_a^{SOC}(t)$ and decreases when $h < \epsilon \rho_a^{SOC}(t)$. The full line shows $\langle \rho_a^{SOC} \rangle(\rho)$ as obtained in a finite system, while the dotted line shows the behaviour in the thermodynamic limit. The latter allows the definition of a critical density ρ_c, where an AS phase transition takes place.

1.4.3 Criticism I: comparing AS and SOC

The preceeding section contained merely a *description* of SOC in terms of AS. This description itself does not posses any explanatory power. It has a potential for an explanation by pointing out that the critical point of SOC can be interpreted as an AS critical point. This statement can be made much stronger by claiming that the same critical behaviour of the SOC model (by tuning h and ϵ and therefore implicitly ρ) can be observed in the corresponding AS model (by tuning only the density ρ, without ϵ and h). This is not obvious — for example, the behaviour of an ensemble of an AS model at particle density ρ is not necessarily the same as the behaviour of the SOC model with the same ensemble-*averaged* particle density.

It seems easy to compare AS and SOC numerically: First one investigates the AS behaviour of a model system by fixing the density ρ, then one repeats the experiment for SOC, either by explicitly tuning h and ϵ, or by introducing an external drive which acts only at quiescent time and dissipation on the boundaries. The main problem occurs when looking for an observable to compare between the two models: the focus of AS is on the microscopic scale, when activity is spreading through the system from one time-step to the other, while the main focus of SOC rests on macroscopic properties, such as avalanches, which are instantaneous and do not have a microscopic structure. Abelian models (see Sec. 1.3.1, page 40, especially page 52) contain the extra difficulty that their statistical properties do not depend on the specific choice of the order of updates and therefore are very flexible regarding the definition of the timescale. This is certainly not expected for AS models.

Focusing on the SOC point of view, one can measure the avalanche size distribution $\mathcal{P}(s;\rho)$ in AS models while driving ρ from 0 to a maximum value. There is no dissipation which would make possible the arbitrary tuning of ρ in AS. Just like in SOC, the density is increased only in the quiescent state. Assuming simple scaling

[10] Actually Eq. (1.27) is correct only asymptotically (thermodynamic limit), because in (1.27) dissipation and drive are expressed as global, *de facto* spatially averaged properties, while in the model they are realised as purely local processes. Therefore in a finite system there are additional fluctuations on top of the *average* rates ϵ, h.

(1.4) with the cutoff $s_0 = bL^D$ one has

$$\langle s^n \rangle_{\mathcal{P}(\rho)} = \int_0^\infty s^n \mathcal{P}(s;\rho) = A_{s^n} L^{D(1+n-\tau)} + \ldots \tag{1.28}$$

where A_{s^n} denotes an unknown amplitude and $\ldots$ stands for finite size corrections. The subscript on $\langle s^n \rangle_{\mathcal{P}(\rho)}$ stresses that this an average over the distribution $\mathcal{P}(s;\rho)$. If one plots the nth moment of the avalanche size distribution as the ratio $\langle s^n \rangle_{\mathcal{P}(\rho)}/L^{\gamma_n}$ versus ρ, then curves for different L will intersect at one specific ρ, if one chooses the correct γ_n. Plotting these estimates of γ_n as a function of n, one can estimate τ and D by fitting against $D(1+n-\tau)$.

This method was developed by the author in a collaboration with KIM CHRISTENSEN, NICHOLAS MOLONEY and OLE PETERS. Within this collaboration the Oslo model (see Chapter 5 and also Sec. 6.3, page 273) was investigated. The main results (Christensen *et al.*, 2004) are

- The exponent D is the same in the original Oslo model and the AS variant.
- The exponent τ is *not* the same, however, it gets close to a value known from the bulk-driven variant of the Oslo model (see Sec. 5.2.2.1, page 220). There are good arguments, why this happens.
- The universal moment ratios do not correspond to those of any known (SOC) variant of the Oslo model, which can easily be understood from the difference in boundary conditions, see Sec. 2.2.1.1, page 77.

1.4.4 Criticism II: converting AS to SOC

The crucial claim regarding the relation between AS and SOC, particularly explictly expressed in (Dickman *et al.*, 2000), is that SOC can be obtained in an AS model, simply by tuning h, ϵ and h/ϵ to 0, which brings the model automatically to the AS critical point. One might argue that neither a bulk dissipation ϵ nor a fixed external driving rate h can be found in most models of SOC. Nevertheless, the statement that under very general circumstances AS models can be driven in a way that make them behave like SOC models, is a very strong one.

In the following, it will be shown that the exponents obtained in an AS model in the limits of vanishing h, ϵ and h/ϵ, depend crucially on the way this tuning is performed.

The order parameter of the AS model is actually the conditional activity (*i.e.* the zero activity of quiescent time stretches is ignored), which was introduced above as $\rho_a^{\text{AS, stat}}$. The relation between the conditional activity $\rho_a^{\text{AS, stat}}$ and the "instantaneous activity" ρ_a^{SOC} is unknown, and might be a very complicated function. If h is very small, ρ_a^{SOC} will roughly be linear in h, however, for large h this is probably different. For the sake of the following argument, it is only important to know that there is a relation. If h and ϵ are given, then $\rho_a^{\text{AS, stat}} = \rho_a^{\text{AS, stat}}(h, \epsilon)$.

VESPIGNANI *et al.* are not concerned with these subtleties; their explanation is rather fuzzy in this respect. This poses a major obstacle to any solid criticism, as the counter argument can hardly be sharper than the claim.

The only scaling behaviour accessible to most SOC systems is finite size scaling — SOC models ideally have only one tuning parameter, the system size. This is even true for the behaviour of distributions like the avalanche size distribution $\mathcal{P}(s)$, the cutoff of which is only a function of the system size. In fact, as discussed in Chapter 2, the only reliable way to determine the exponents directly from such a distribution is to make a data collapse among results for different system sizes. Remarkably, mean field arguments have been used to substantiate the claims by VESPIGNANI *et al.* (Vespignani and Zapperi, 1998; Lübeck, 2003a), which, however, are bound to show improper finite size scaling (see for example the problems in Sec. 3.3.5.2, page 126.).

The idea is now to ask the following question: Given a certain sequence of choices of $h(L)$ and $\epsilon(L)$ with system size L, what values are assumed by observables? Subsequently, which exponents are derived from their scaling behaviour?

To this end, one assumes that in leading order the choice of $h(L)$ and $\epsilon(L)$ leads to[11] $\rho^* - \rho_c = qL^{-1/\mu}$, where $\mu > 0$ and ρ^* now means the *average* particle density in the system, which plays the rôle of the effective particle density of the underlying AS transition. It is clear that the actual density fluctuates around the average ρ^*, but by decreasing ϵ, one can reduce these fluctuations (see sec. 1.4.4.1) almost arbitrarily.

If one accepts the correspondence between AS and SOC, then the observables obtained in the SOC model should correspond to those in the AS model at the density ρ^*. Especially, the stationary, conditional activity $\rho_a^{\text{AS, stat}}$ is expected to behave like[12]

$$\rho_a^{SOC}(h, \epsilon) = \rho_a^{\text{AS, stat}}(\rho) = L^{-\frac{\beta}{\nu_\perp}} \mathcal{G}\left(\frac{\rho^* - \rho_c}{L^{-1/\nu_\perp}}\right) \tag{1.29}$$

where — at least within a certain range of ρ and sufficiently large L — the scaling function $\mathcal{G}$ is expected to behave like $\mathcal{G}(x) \propto x^\beta$ for large positive arguments (thermodynamic limit in the active, *i.e.* "low temperature" phase), to converge to a constant for small arguments (finite size scaling) and to vanish[13] for divergent negative argu-

[11] Every choice of h, ϵ has a particular ρ associated with it, either via $h/\epsilon = \langle \rho_a^{SOC} \rangle(\rho)$, see Eq. (1.24) and Fig. 1.12, or, alternatively, via $\rho_a^{SOC}(h, \epsilon) = \rho_a^{\text{AS, stat}}(\rho)$.

[12] Using the notation introduced in Chapter 2 this corresponds to $\partial_H \mid_{H=0} f_s(t, H; L)$, Eq. (2.13), page 77, with $H = 0$ and t replaced by $\xi^{1/\nu}$ and $-d + \Delta/\nu = -\beta/\nu$, Eq. (2.9), page 76.

[13] In the Ising model, one can show that the scaling function must decay like $\mathcal{G}(x) \propto (-x)^{-\gamma/2}$, based on the idea that in the high temperature phase, patches of size ξ become independent so that the magnetisation density scales like $L^{-d/2}$.

ments (thermodynamic limit in the absorbing, *i.e.* "high temperature" phase). This is a standard result in classical critical phenomena (Stauffer and Aharony, 1994; Cardy, 1997). Plugging in $\rho^* - \rho_c = qL^{-1/\mu}$ then gives immediately the large L behaviour as

$$\rho_a^{\text{AS, stat}}(\rho(L)) \propto \begin{cases} L^{-\beta/\mu} & \text{for } \mu > \nu_\perp \text{ and } q > 0 & (1.30a) \\ 0 & \text{for } \mu > \nu_\perp \text{ and } q < 0 & (1.30b) \\ L^{-\beta/\nu_\perp} & \text{for } \mu \leq \nu_\perp & (1.30c) \end{cases}$$

Apparently only (1.30c) leads to appropriate finite size scaling behaviour of the order parameter, independent of the specific choice μ. Apart from the vanishing amplitude in (1.30b), this result is of course not restricted to the observable $\rho_a^{\text{AS, stat}}$; any observable obeying a scaling form like (1.29) will show similar behaviour.

From (1.30) it is clear that the scaling of ρ^* must be chosen such that $\nu_\perp \geq \mu$. Otherwise, the AS version produces exponents which depend on μ, and in general cannot be the same as those observed in SOC, because μ is (supposedly) completely arbitrary. However, the identity of the exponents in AS and SOC is a crucial part of the claim. So, how does one have to scale ρ^* via h and ϵ, so that $\rho^* - \rho_c$ scales with an exponent $\mu \leq \nu_\perp$? There is no simple way without knowing the correct exponents, as is shown in the following.

The value of ρ^* is chosen only implicitly via $\rho_a^{\text{AS, stat}}(h, \epsilon)$. Assuming that $\rho_a^{\text{AS, stat}}$ is tuned such that $\rho_a^{\text{AS, stat}} \propto L^{-\omega}$, with a certain, unknown exponent ω, how does $\rho^* - \rho_c$ behave? According to (1.29) it is

$$\rho^* - \rho_c \propto \begin{cases} L^{-\omega/\beta} & \text{for } \frac{\beta}{\nu_\perp} - \omega > 0 & (1.31a) \\ L^{-1/\mu} & \text{with } \mu > \nu_\perp \text{ for } \frac{\beta}{\nu_\perp} - \omega < 0 & (1.31b) \\ L^{-1/\mu} & \text{with } \mu \leq \nu_\perp \text{ for } \frac{\beta}{\nu_\perp} - \omega = 0 & (1.31c) \end{cases}$$

It is worth commenting the three cases. The first one, (1.31a), comes from a divergent product $\rho_a^{\text{AS, stat}}(\rho^*(L))L^{\beta/\nu_\perp}$ and therefore divergent $\mathcal{G}(x)$ in (1.29), so that apparently $\mathcal{G}(x) \propto x^\beta$, which is the only way $\mathcal{G}$ can diverge. Since $\beta/\nu_\perp - \omega > 0$ the resulting asymptote $L^{-\omega/\beta}$ has an exponent $1/\mu = \omega/\beta$ obeying $\mu > \nu_\perp$; the system leaves the finite size scaling region, the width of which scales like $L^{1/\nu_\perp}$, towards the active, or "low temperature phase".

For (1.31b), $\mathcal{G}$ converges to 0, so that its argument apparently diverges negatively, *i.e.* $\rho^* - \rho_c$ vanishes slower than its denominator[14] $L^{-1/\nu_\perp}$ in (1.29); the system leaves the finite size scaling region towards the absorbing, or "high temperature phase".

The last case, (1.31c), is obtained if $\rho_a^{\text{AS, stat}} L^{\beta/\nu_\perp}$ remains constant, which means

[14]So, actually, (1.31b) should be $\rho^* - \rho_c \in \omega\left(L^{-1/\nu_\perp}\right)$, meaning that it decays strictly slower than $L^{-1/\nu_\perp}$, see (Knuth, 1997, page 110).

that the argument of $\mathcal{G}$ is either constant or vanishes, so that $\rho^* - \rho_c \propto L^{-1/\nu_\perp}$, or faster; then the system stays in the finite size scaling region.

The crucial outcome of (1.31) is that $\mu > \nu_\perp$ is obtained for any $\omega \neq \beta/\nu_\perp$. While the specific choice $\omega = \beta/\nu_\perp$ leads apparently to the correct finite size scaling behaviour of $\rho_a^{\text{AS, stat}}$, (1.30c), any other choice leads to a an effective ρ^* which gives ω dependent exponents; there seems to be no direct way to derive any "correct exponent" from the scaling behaviour of the observed activity, moreover, the model leaves the finite size scaling region, since $\mu > \nu$, so that $\xi/L \to 0$, while still $\xi \to \infty$. Eq. (1.31a) predicts $1/\mu = \omega/\beta$, so $\rho_a^{\text{AS, stat}} \propto L^{-\omega}$, as tuned anyway. So, if $\rho_a^{\text{AS, stat}}$ is tuned like $L^{-\omega}$, then ω is the exponent one will see for β/μ in finite size scaling, see (1.30a). Eq. (1.31b) is somewhat peculiar in this respect: It applies to $\omega > \beta/\nu_\perp$ and predicts $1/\mu < 1/\nu_\perp$, so that $\omega > \beta/\mu$, while $\omega = \beta/\mu$ might be expected. The answer to this spurious inconsistency is that in the "high temperature" phase the order parameter does *not* scale like $|\rho^* - \rho_c|^\beta$, and therefore no scaling $L^-(\beta/\mu)$ of $\rho_a^{\text{AS, stat}}$ is expected — it simply decays in a way such that $\rho_a^{\text{AS, stat}} \propto L^{-\omega}$, just as tuned. In fact, (1.30b), applicable because of the negative divergence of the argument in (1.31b), does not necessarily predict algebraic behaviour of $\rho_a^{\text{AS, stat}}$ for an algebraic decay of $\rho^* - \rho_c$ and vice versa. To summarise, the difference $\rho^* - \rho_c$ just decays somehow slower than $L^{-1/\nu_\perp}$ and in a way to reproduce $\rho_a^{\text{AS, stat}} \propto L^{-\omega}$ as setup by tuning in the first place.

Thus, AS critical behaviour cannot be obtained by an arbitrary choice for the limits $h, \epsilon, h/\epsilon \to 0$. Only a choice such that $\rho_a^{\text{AS, stat}} \propto L^{-\beta/\nu_\perp}$, can produce *universal* critical behaviour, *i.e.* reproduce the AS results. If this behaviour of $\rho_a^{\text{AS, stat}}$ is obtained for a wide variety of choices of $h(L)$ and $\epsilon(L)$, then this is a remarkable, highly non-trivial property of $\rho_a^{\text{AS, stat}}$. This property is then what makes SOC work.

An alternative formulation of the criticism is to say that the mechanism proposed can reach ρ_c within an accuracy of $L^{-1/\nu_\perp}$ only if h and ϵ are tuned such that $\rho_a^{\text{AS, stat}} \propto L^{-\beta/\nu_\perp}$, *i.e.* in order to see proper finite size scaling, one needs to know the exponents beforehand. This is the reason why the argument above sounds circular. It actually *is* circular, simply because the mechanism is circular; we ask "How do you have to tune $\rho_a^{\text{AS, stat}}$ in order to see proper finite size scaling?" And the answer is "You need to tune *exactly* according to finite size scaling." What has been proven above is that there is no other way than that. All that was needed was that a function $\rho_a^{\text{AS, stat}}(h, \epsilon)$ exists, that this functions follows standard AS behaviour (1.29) and that the whole model can, in the spirit of AS, be parametrised by ρ^*.

Finally, one might consider the whole AS picture as misleading. Firstly, as discussed above, a fixed inflow h and a dissipation rate ϵ may not be appropriate to describe SOC models in general. Secondly, even accepting the parametrisation, the critical behaviour obtained in SOC is not necessarily the same as the one in AS. As

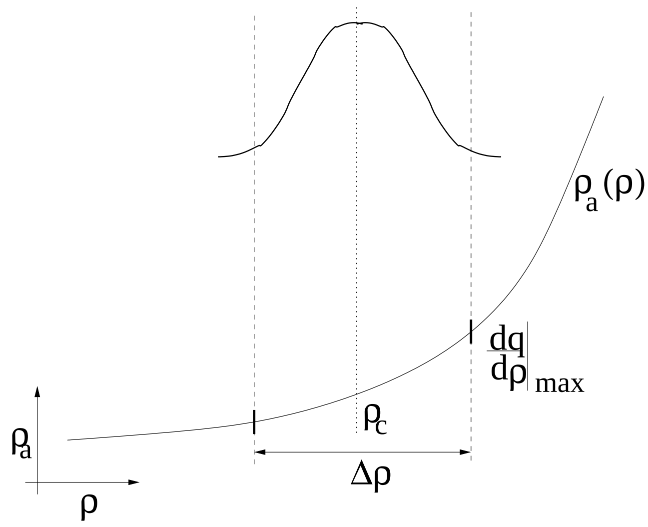


Figure 1.13: A cartoon of the fluctuation of ρ based on Eq. (1.27). The width of the region of ρ values the system is typically found in, is marked by $\Delta\rho$, the average by ρ_c. The maximum slope of the average activity is marked as such. In the upper part the cartoon of a probability density function shows the position probability as a function of ρ.

mentioned above, there is no obvious correspondence between ρ_a^{SOC} and $\rho_a^{\text{AS, stat}}$. Moreover, the observables obtained in SOC by fluctuating around a density ρ are not necessarily identical to those obtained in AS at the mean value of this density, ρ^*.

1.4.4.1 Supplement: fluctuations

By examining fluctuations, it is possible to find an upper bound for the convergence of $\epsilon \to 0$. For the following derivation we will assume $\left\langle \rho_a^{SOC} \right\rangle \propto \rho_a^{\text{AS, stat}}$ and $\sigma^2\langle(\rho_a^{SOC}\rangle) \propto \sigma^2\langle(\rho_a^{\text{AS, stat}}\rangle)$.

From (1.27) one can estimate the fluctuations in ρ from the fluctuations in $\rho_a^{SOC}(\rho)$. By dimensional analysis one finds a variance of ρ

$$\sigma^2\langle(\rho\rangle) = \epsilon \left(\left. \frac{dq}{d\rho} \right|_{\max} \right)^{-1} \sigma^2\langle(\rho_a^{SOC}\rangle) \times \text{const.} \tag{1.32}$$

where $q(\rho)$ is the average value of ρ_a^{SOC} obtained at a given ρ. The derivative is to be evaluated, where it has a maximum, in order to determine the minimum fluctuations in ρ. If the width of the fluctuations in ρ becomes larger than the width of the finite size scaling region, finite size scaling should break down, because the system would spend more and more time outside the finite size scaling region.

The fluctuations $\sigma^2\langle(\rho_a^{SOC}\rangle)$ in the finite size scaling region scale at least like $L^{-2\beta/\nu_\perp}$, which is the scaling of the first moment squared; "at least" here means that the exponent γ in

$$\sigma^2\langle(\rho_a^{SOC}\rangle) \propto L^{\frac{\gamma}{\nu_\perp}-d} \tag{1.33}$$

obeys $\gamma - \nu_\perp \geq -2\beta$, where the equal sign applies if the RUSHBROOKE and the JOSEPHSON scaling law apply (Pfeuty and Toulouse, 1977). The function $q(\rho)$ is supposedly simply $\rho_a^{\text{AS, stat}}(\rho)$, Eq. (1.29). The maximum slope of this function is obtained outside the finite size scaling region, because ρ_a^{SOC} and therefore $\rho_a^{\text{AS, stat}}$ is flattened within the finite size scaling region due to the scaling function. Thus, one estimates $q(\rho) = \rho^\beta$ and therefore

$$\left.\frac{dq}{d\rho}\right|_{\max} \propto L^{-\frac{\beta-1}{\nu_\perp}} \tag{1.34}$$

where the derivative has been evaluated at the boundary of the finite size scaling region, see Fig. 1.13, which is the region *at least* accessible by the equation of motion, because of non-vanishing fluctuations within there. The slope is overestimated, *i.e.* the exponent on the RHS of (1.34) is in fact rather more negative. Thus there is a non-vanishing constant such that

$$\sigma^2\langle(\rho\rangle) > \epsilon L^{\frac{\beta-1}{\nu_\perp}} L^{-\frac{2\beta}{\nu_\perp}} \times \text{const} , \tag{1.35}$$

while the width of the finite size scaling region scales like $L^{-2/\nu_\perp}$. If ϵ remains constant, one needs

$$\frac{\beta+1}{\nu_\perp} > \frac{2}{\nu_\perp} \tag{1.36}$$

or $\beta > 1$ for ρ not to overshoot due to fluctuations $\sigma^2\langle(\rho\rangle)$. If ϵ changes in L like $L^{-\sigma}$, one arrives at the relation

$$\beta + \sigma > 1 . \tag{1.37}$$

If this is not obeyed, the fluctuations in the activity will eventually drive the system outside the finite size scaling region.

1.4.4.2 Supplement: the ISING argument

The criticism on the mechanism proposed by VESPIGNANI *et al.* becomes especially vivid when realising that it ignores fifty years of computational physics based on Monte Carlo: Why should one run massive Monte-Carlo simulations to pin down the critical values of the parameters, such as temperature or external field, if there is a very simple, self-organised mechanism, which drives the model automatically to the critical point? To illustrate this point further, in the following the mechanism is translated to the ISING model.

The VESPIGNANI mechanism needs only two ingredients: external drive and dissipation. While it makes the argument certainly more appealing that there is a natural relation between activity and particle density, this is formally irrelevant. Thus

$$-\dot{T} = h - \epsilon M \tag{1.38}$$

is a perfect realisation of this mechanism for the ISING model. Here, T denotes the temperature (T_c being the critical temperature) and M the instantaneous magnetisation. To claim that this mechanism would automatically lead to proper finite size scaling makes it very suspicious: Why would one invest considerable amounts of CPU-time to find T_c and do standard finite size scaling if there is such a simple mechanism, which automatically produces these values? One would usually expect that an "economic principle of computational physics" applies: The quality of the estimates for numerical values, such as exponents and amplitude ratios, is roughly constant among different approaches to a problem in a mature field for a given amount of CPU-time, provided that the best available algorithms are used. Clearly, this ignores effects of, for example, self-averaging, dynamical exponents or any progress in general, but it is a fair representation of the minute achievements possible in a *mature* field.

Of course all arguments presented in sec. 1.4.4 apply here as well and explain, why one cannot use (1.38) to do finite size scaling. For the ISING model it translates to tuning h/ϵ exactly like $L^{-\beta/\nu}$. Only in that case one can observe correct finite size scaling exponents and only in that case T_c is reached within an accuracy of $L^{-1/\nu}$ (Hasenbusch, 1999). While many other choices lead to an approach of T_c in the thermodynamic limit, they lead to finite size scaling behaviour dependent on ω in $h/\epsilon \propto L^{-\omega}$. Can one derive any useful information from these exponents?

Open task

1.4.5 Discussion

During all this critique one must bear in mind that there is a choice of an equation of motion like (1.27) or (1.38) such that proper finite size scaling is observed. This is because apparently numerical physicist *are* able to observe proper finite size scaling, while changing the density or the temperature in time, *i.e.* with an equation like (1.27) or (1.38) but with a more complicated right hand side.

In fact, the argument above cannot and should better not rule out the existence of such a "proper equation of motion", which makes proper finite size scaling possible. Indeed, this is not what it claims; it merely claims that the choice of h and ϵ is not arbitrary. One standard choice is $h = 0$, $\epsilon = 0$, namely determining T_c with very high accuracy using (for example) the behaviour of the BINDER cumulant (Landau and Binder, 2000) and then using this estimate for finite size scaling *without changing the temperature* any further. Most remarkably, using this technique makes results obtained from larger L (when using the BINDER cumulant to determine T_c) enter the simulation for smaller L (when doing the finite size scaling). Moreover, the accuracy with which T_c can be determined from finite size scaling scales with the size of the region where the correlation length is of the order of the system size, *i.e.* the region scales like $L^{-1/\nu}$ which is also the leading order correction (Hasenbusch, 1999).

It is worth pointing out, that in the Forest Fire model the L-dependence of the choice of f/p, which is closely related to h/ϵ, has not received much attention. The finite size behaviour of the model discussed in Sec. 3.3 (see page 126) suffers from the same problem.

One might argue that the VESPIGNANI mechanism is a perfect explanation for the wide variety of non-trivial exponents observed in SOC. This mechanism predicts non-universal exponents, because of the dependence of the exponents on the scaling of ϵ and h in L. But apart from that, the mechanism also predicts that the models sooner or later leave the finite size scaling region, *i.e.* in the thermodynamic limit the ratio ξ/L vanishes. This is not what is observed. On the contrary, in SOC, accepting the description in terms of ϵ and h, these parameters are tuned implicitly by the model rules and produce *universal*, *scale-invariant* behaviour. These crucial properties are rejected by the AS approach.

But after all this critique — how then is SOC possible?

Because of the presence of the AS critical point for the SOC models, at least for those with bulk dissipation, the *only* way to obtain the same critical behaviour in SOC as in AS is by proper tuning. Otherwise the AS behaviour is either not seen, or SOC is non-universal and depends on the tuning, which is, however, not what is observed. The only way to get around the tuning necessity of h and ϵ in order to keep $\rho_a^{\text{AS, stat}}(h,\epsilon)$ at the right value, is that $\rho_a^{\text{AS, stat}}(h,\epsilon)$ somehow does not depend on them. Whatever the way to avoid tuning, *this* is what explains SOC. The status of VESPIGNANI's mechanism as well as this issue needs further investigation. Open issue

1.5 Summary

This chapter provides an overview of some general features of SOC.

- In Sec. 1.1 the actual meaning of the term "SOC" is discussed.
- The next section, Sec. 1.2, discusses the "apparent ubiquity of power laws in nature".
- The models of paradigmatic status, such as BTW, OFC and DS-FFM, are discussed in Sec. 1.3.
- VESPIGNANI *et al.* have proposed a trivial mechanism of how SOC works. It is discussed and criticised in Sec. 1.4. It transpires that there are many good arguments why this mechanism is bound to fail.

Chapter 2

Scaling

This chapter is a collection of basic analytical results and the subsequent numerical techniques. Most of the discussion is very detailed and contains plenty of technical remarks, which hopefully increases its usability; this chapter is meant as a manual. Most examples are taken from numerical simulations of the Oslo model or percolation (which is not discussed in detail in this thesis).

2.1 Introduction

Finite Size Scaling is obtained in many different areas of critical phenomena. Most traditional critical phenomena are studied in terms of moments of an underlying distribution, while the distributions themselves are subject to the investigation typically only in percolation and Self-Organised Criticality. In the latter field there is alarming confusion about scaling and how to analyse it [see for example (Pastor-Satorras and Vespignani, 2000a,b)]. This chapter introduces these concepts, tries to clarify some of the confused issues[1] and to identify some important relations between different scaling assumptions.

Scaling behaviour is the hallmark of criticality (Fisher, 1967), as it implies scale invariance: The statistical properties of a system on one scale are, apart from some factors, identical to the statistical properties of the system on another scale. The factors depend only on the ratio of the two scales and not on the scales themselves. The scales might be the system sizes, in which case the scaling is obtained right on the critical point and is then known as "Finite Size Scaling" (Cardy, 1988). Alternatively the scales are the correlation lengths, in which case the scaling is obtained away from the critical point.

In the following, various features of scaling are discussed with special attention to their application to SOC. Nevertheless, many of them are presented with their

[1]While probably introducing new misunderstandings.

relation to equilibrium critical phenomena. The aim of this chapter is to provide a general theoretical basis for the numerical and analytical analysis of SOC models.

Most of the following derivations are based on a presumed probability density function $P(M)$ of a quantity M. This quantity will be the total magnetisation in case of equilibrium statistical mechanics and the avalanche size in case of SOC. However, this identification is only for illustration purposes and most of the results are valid under more general circumstances.

2.2 Moments and Distributions

The traditional observables in magnetic systems, such as the ISING model, are internal energy, specific heat, magnetisation and susceptibility. Also the higher moments are of interest, such as those used in the BINDER cumulant [see, for example the review (Landau and Binder, 2000)], which effectively measures the deviation of the underlying distribution from a Gaussian. This section shows various equivalent representation of such an underlying distribution.

In the following the term "distribution" is used synonymously with probability density function (PDF). It will *not* refer to the cumulative distribution usually preferred by mathematicians. The latter is related to a PDF $\mathcal{P}(M)$ by $\int_{-\infty}^{M+0^+} dM'\mathcal{P}(M')$. The moments of the distribution $\mathcal{P}(M)$ are defined by

$$\langle M^n \rangle \equiv \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} dM M^n \mathcal{P}(M) \ , \tag{2.1}$$

where especially $\langle 1 \rangle = 1$ is the normalisation of the PDF. The moment generating function (MGF) is $\langle \exp(HM) \rangle$, because

$$\langle M^n \rangle = \left.\frac{d^n}{dH^n}\right|_{H=0} \sum\!\!\!\!\!\!\int_{-\infty}^{\infty} dM e^{HM} \mathcal{P}(M) = \left.\frac{d^n}{dH^n}\right|_{H=0} \langle \exp(HM) \rangle \ . \tag{2.2}$$

To escape all problems of existence of the integrals, convergence and analyticity, all the derivations are restricted to PDF's which are strictly 0 beyond a finite cutoff, as obtained, for example, for the distribution of the magnetisation density in a finite ferromagnetic system.

Because all moments exist, one can try to Taylor expand $\langle \exp(HM) \rangle$:

$$\left\langle e^{HM} \right\rangle = \sum_n \frac{H^n}{n!} \langle M^n \rangle \tag{2.3}$$

According to (2.1), $\langle \exp(HM) \rangle$ is the LAPLACE transform of $\mathcal{P}(M)$, so that (2.3) allows to reconstruct $\mathcal{P}(M)$ from the set of all moments using an inverse LAPLACE transform. This suggest a unique $\mathcal{P}(M)$ for a given set of moments. However, this holds only if the RHS of (2.3) actually exists and converges. This is assumed everywhere below.[2]

The cumulant generating function (CGF) is $\ln \langle \exp(HM) \rangle$, where the nth cumulant $\langle M^n \rangle_c$ is defined as

$$\langle M^n \rangle_c \equiv \frac{d^n}{dH^n} \ln \langle \exp(HM) \rangle \ . \tag{2.4}$$

Of course, there is a close relation between the free energy and the cumulant generating function, which will be discussed in sec. 2.2.1. For $n = 0$ (2.4) is well behaved and gives $\langle 1 \rangle_c = 0$.

Provided that all moments are finite, it is obvious how to get from the PDF to the generating functions and from there to the moments and the cumulants. The only technical difficulty is to find a closed form of the RHS in (2.3) and to perform the inverse LAPLACE transform to derive the PDF from it. In sec. 2.3 it is shown that under very general conditions, the task boils down to determine a universal scaling function.

2.2.1 Free energy

In a ferromagnetic system the cumulants of the magnetisation at zero-field are given as derivatives of the free energy by the conjugated field:

$$L^{-d} \langle M^n \rangle_c = -(kT)^{n-1} \left. \frac{d^n}{dH^n} \right|_{H=0} f(t, H) \tag{2.5}$$

where $\langle M^n \rangle_c$ is again the nth cumulant of the total magnetisation M at vanishing external field H, f is the free energy density, T is the temperature and $t = (T - T_c)/T_c$ the reduced temperature around the critical point at $H = 0$ and $T = T_c$. The reason for having the total magnetisation on the LHS is that the Hamiltonian in the Boltzmann factor in the partition sum contains the *total* magnetisation. Thus, deriving the free energy *density* gives only one factor L^{-d}. The powers of kT appear for a similar reason. For the following arguments, these powers are completely irrelevant.

In the thermodynamic limit, close to the critical point, the *singular* part of the free energy density f_s obeys (Cardy, 1997)

$$b^{-d} f_s(b^y t, b^{y\Delta} H) = f_s(t, H) \tag{2.6}$$

[2]This it is actually a very important detail: The moments could all exist but the sum might not, for example if $\langle M^n \rangle = (2n)!$. It seems that something similar happens for $\mathcal{P}(M) = \exp -t^\beta$ for $\beta > 1/2$ (Akhiezer, 1965; Chen and Lawrence, 1999). Thanks to YANG CHEN for advice on this problem.

where the exponents y and Δ are *a priori* unknown. From (2.5) and (2.6) one has immediately

$$L^{-d}\langle M^n\rangle_c(t) = -b^{-d+ny\Delta} f_s^{(n)}(b^y t, 0) \; , \tag{2.7}$$

where $f_s^{(n)}(b^y t, 0)$ is the nth derivative by external field H of the singular part of the free energy density evaluated at temperature $b^y t$ and vanishing external field. It is important to note that (2.7) applies only to the *singular* behaviour of $\langle M^n\rangle_c$ at the critical point.

Choosing $b^y t = 1$ one arrives at the final expression

$$\langle M^n\rangle_c(t) = L^d t^{d/y-n\Delta} a_n^{\pm} \; , \tag{2.8}$$

where $a_n^{\pm}$ denote unknown amplitudes, dependent on the sign of the reduced temperature, but otherwise independent from temperature and external field. Using RUSHBROOKE, WIDOM and JOSEPHSON scaling (hyperscaling) (Pfeuty and Toulouse, 1977) one derives for the so-called "gap exponent" Δ

$$\Delta = \beta + \gamma = \beta\delta = \nu d - \beta \tag{2.9}$$

and for y

$$y = \frac{1}{\nu} \; . \tag{2.10}$$

In the following, a theory obeying these standard scaling laws (RUSHBROOKE and JOSEPHSON, but also WIDOM and FISHER) will be called "equilibrium critical phenomena".

For finite systems, the free energy density can be re-parametrised[3] in terms of $\xi(t)/L$ rather than t, where the correlation length ξ is supposedly measured in an infinity system. At the critical point this yields

$$\langle M^n\rangle_c(L) = L^{n\Delta'} a'_n \tag{2.11}$$

with $\Delta' = \Delta/\nu = d - \beta/\nu$ the gap exponent for finite size scaling and a'_n new, appropriately defined, L-independent amplitudes related to the amplitudes in (2.8). One easily confirms $\Delta' = d - \beta/\nu$ and $2\Delta' = d + \gamma/\nu$ for the scaling behaviour of the total magnetisation and the total susceptibility.

The crucial difference between Eq. (2.8) and Eq. (2.11) is that the latter equation has only a power of L linear in n on the RHS, while the former has additionally a constant power of L and t as well. The structure of $\langle M^n\rangle_c(L)$ makes it possible to derive the behaviour of the moments as a function of L by induction: Starting with

[3] Finite size scaling is introduced a bit less *ad hoc* below.

$\langle M \rangle_c(L) = \langle M \rangle(L)$ one notes that all cumulants are sums of products of moments, where each product has the same dimension, *i.e.* the same total power of M. Thus, if $\langle M^m \rangle(L) \propto L^{m\Delta'}$ holds for all $m < n$, this is also true for $m = n$ because of (2.11). Thus, there are amplitudes a''_n so that the *moments* obey

$$\langle M^n \rangle(L) = L^{n\Delta'} a''_n . \tag{2.12}$$

The same argument cannot be repeated for the t-dependence of M in (2.8), because of the presence of additional powers of L and t. Below we will see that this is what makes it impossible to conclude that $\mathcal{P}(M)$ obeys so called "simple scaling" outside the finite size scaling regime.

2.2.1.1 Universal scaling function: remarks

The scaling behaviour of the free energy can be expressed in a more formal way. The singular part of the free energy of a finite system behaves around the critical point like (Privman *et al.*, 1991)

$$f_s(t, H; L) = L^{-d} Y(K_t t L^{1/\nu}, K_h H L^{\Delta/\nu}) \tag{2.13}$$

with two non-universal amplitudes K_t and K_h, which are often called "metric factors" (Privman and Fisher, 1984). The scaling function $Y(x, y)$ is universal among different models, but it depends on the boundary conditions and the geometric shape of the system. This applies especially to finite size scaling at the critical point, where

$$\lim_{x \to 0} Y(x, y) = Y(0, y) \tag{2.14}$$

removes one of the metric factors.

In the thermodynamic limit the free energy density converges to a universal function (Privman *et al.*, 1991; Cardy, 1997)

$$Y(x, y) \to x^{d\nu} Q_\pm(y x^{-\Delta}) \ \text{ for } x \to \pm\infty , \tag{2.15}$$

so that

$$\lim_{L \to \infty} f_s(\xi, H; L) = K_\xi \xi^{-d} \widetilde{Q}_\pm(\widetilde{K} H \xi^{\Delta/\nu}) \tag{2.16}$$

now parametrised in ξ rather than t, to be compared to[4] (2.15),

$$\lim_{t\to 0} f_s(t, H; L) = \lim_{\xi\to\infty} f_s(\xi, H; L) = L^{-d} Y(0, K_h H L^{\Delta/\nu}) \tag{2.17}$$

Thus, $\widetilde{Q}_\pm$ and $Q_\pm$ respectively (the two signs are for high temperature phase $(+)$ and low temperature phase $(-)$ respectively) is observed in the thermodynamic limit only, provided that the reduced temperature does not vanish before the limit is taken. Most remarkably, $Q_\pm$ does not depend on any details of the finite system, like shape, aspect ration, boundary conditions etc. One is therefore faced with two different notions of universality: Universal finite size scaling, observed in the limit $\xi/L \to \infty$ and parametrised by L, (2.17), which is blind to microscopic details (next nearest neighbour interaction etc.) and a more general concept of universality, observed only in the limit $\xi/L \to 0$ and parametrised by ξ, (2.16), which even ignores shapes etc. Even the weak, first form of universality in finite size scaling is a meaningful, useful concept, as it helps to identify universality classes and what supposedly determines them, namely the fundamental properties, interactions and symmetries of the microscopic dynamics or rules. *Finite size scaling is the only scaling accessible to SOC.*

2.3 Simple Scaling

As known from percolation (Stauffer and Aharony, 1994), where the cluster size distribution has a similar form, one says that a PDF $\mathcal{P}(M)$ obeys simple scaling if

$$\mathcal{P}(M; L, \xi) = a M^{-\tau} \mathcal{G}\left(\frac{M}{M_0(L)}, \frac{L}{\xi}\right) \quad \text{for } M > M_l \tag{2.18}$$

where a is a non-universal constant (metric factor), M_l is the so-called "lower cutoff", ξ is the correlation length and L is the system size. The universal function[5] $\mathcal{G}$ is called the "scaling function" or "cutoff function", as it effectively cuts off an integral over M to an integral over a finite range of M-values. Usually, in a finite system all positive moments exist[6]

$$\forall_{n\geq 0} \int_0^\infty dx x^{n-\tau} \mathcal{G}(x, y) = \text{finite}\ , \tag{2.19}$$

[4] The re-parametrisation of t by ξ is, of course, incorrect; ξ diverges like $t^{-\nu}$ only for $H = 0$, so that formally, we have to do all derivations first, then set $H = 0$ and then re-parametrise t. Moreover, formally t does not prevail over H as a tuning parameter, so that H could be re-parametrised by ξ at $t = 0$ equally well. It is just far more established to re-parametrise t rather than H by ξ.

[5] $\mathcal{G}$ is a universal scaling function, dependent on the boundary conditions, shape, aspect ratio etc., which has two arguments; the H dependence of f_s, (2.13) has been replaced by an M dependence.

[6] This is because a finite system can only produce finite "events", so that all (positive) moments must be finite as well. For the origin of this integral see (2.29) and (2.41).

so that the scaling function must decay faster than any power of its first argument,

$$\forall_{n\geq 0} \lim_{x\to\infty} x^{n-\tau}\mathcal{G}(x,y) = 0 \, . \tag{2.20}$$

For example, this is obeyed by $\mathcal{G}(x) = \exp(-x)$. The fact that there are usually two scaling functions, one for the high temperature phase and one for the low temperature phase has been incorporated into $\mathcal{G}$ by displaying two shapes depending on the sign of the correlation length, which is chosen according to the phase.

It is important to keep in mind that all following derivations are based on a lower cutoff M_l being independent of L. Especially a lower cutoff which diverges in L and therefore renders — in some sense — the entire distribution $\mathcal{P}(M)$ asymptotically non-universal is not compatible with simple scaling.

The function $M_0(L)$ in Eq. (2.18) is called the cutoff or characteristic size. It includes another metric factor b, because one expects a scaling behaviour of the cutoff[7]

$$M_0(L) = bL^D \tag{2.23}$$

At the critical point, (2.18) simplifies to

$$\lim_{\xi\to\pm\infty} \mathcal{P}(M;L,\xi) = aM^{-\tau}\mathcal{G}\left(\frac{M}{M_0(L)},0\right) \text{ for } M > M_l \tag{2.24}$$

with $\mathcal{G}(x,0)$ being the finite size scaling function depending on shape, aspect ratio, boundary conditions etc. In the context of SOC, often (2.24) alone is called "simple scaling" rather than the full form (2.18). If the thermodynamic limit is taken first, one expects

$$\lim_{L\to\infty} \mathcal{P}(M;L,\xi) = \widetilde{a}M^{-\tau}\widetilde{\mathcal{G}}\left(\frac{M}{\widetilde{M_0}(\xi)}\right) \text{ for } M > \widetilde{M_l} \tag{2.25}$$

where $\widetilde{\mathcal{G}}$ is universal, even independent of shapes, aspect ratios and boundary conditions. The cutoff $\widetilde{M_0}(\xi)$ of this scaling function is a function of ξ rather than L and

[7] By restricting M to the domain $[0,\infty[$ and ignoring M_l one might note that the CGF of $\mathcal{P}(M;L,\xi)$ is

$$\ln \int_0^\infty dM\mathcal{P}(M;L,\xi)e^{-HM} \tag{2.21}$$

which is, up to some irrelevant pre-factors, apparently, (2.7), identical to $L^d f_s(\xi,H;L)$, now with ξ instead of t, as in (2.16). Eqs. (2.18) and (2.18) then give together with (2.13)

$$\ln\left((bL^D)^{1-\tau}a\right) + \ln\int_0^\infty dx\mathcal{G}\left(x,\frac{L}{\xi}\right)e^{-xHbL^D} = Y\left(K'_\xi\frac{L}{\xi}, K_h H L^{\Delta/\nu}\right) \tag{2.22}$$

so that $1-\tau = 0$ to remove the L dependence in the first term and $D = \Delta/\nu$ to get the correct exponents in the second argument of Y. This is confirmed by (2.30). However, note that $M_l = 0$ implies $\tau = 1$, see paragraphs after (2.48) and (2.54).

has a new metric factor $\widetilde{b}$ in

$$\widetilde{M_0}(\xi) = \widetilde{b}\xi^D \tag{2.26}$$

with the same exponent D as in (2.23).

Even though most results are unaffected by the fact that $\widetilde{\mathcal{G}}(x)$ and $\mathcal{G}(x,0)$ cannot be expected to be the same, it is important to keep that in mind; finite size scaling produces a single function $\mathcal{G}(x,0)$, while critical scaling[8] produces two scaling functions depending on the sign of ξ. Especially, one *cannot* incorporate everything into a single function $\mathcal{G}'$, with

$$\mathcal{P}(M; L, \xi) = aM^{-\tau}\mathcal{G}'(M/M_0'(L, L/\xi)) \tag{2.27}$$

even though one might generalise the cutoff itself to [(2.23) and (2.26)]

$$M_0(L, L/\xi) = \begin{cases} bL^D & \text{for } L \ll \xi \text{ i.e. finite size scaling} & (2.28\text{a}) \\ \widetilde{b}\xi^D & \text{for } L \gg \xi \text{ i.e. critical scaling }, & (2.28\text{b}) \end{cases}$$

see the analogous forms (2.16) and (2.17) respectively.

It is very instructive to derive the behaviour of the moments from (2.18). All of the following calculations are done for a statistical quantity M on the continuous interval $M \in [0, \infty]$. Generalisations to discrete quantities or quantities on another (infinite) set are straight forward.

For the time being, the lower cutoff is assumed to vanish, so that one finds in the finite size scaling regime

$$\langle M^n \rangle(L) = a\left(bL^D\right)^{(1+n-\tau)} \int_0^\infty dx x^{n-\tau}\mathcal{G}(x) \tag{2.29}$$

which can be compared to Eq. (2.12) to yield for ferromagnetic phase transitions

$$\tau = 1 \qquad D = \Delta' = \frac{\Delta}{\nu} \ . \tag{2.30}$$

For this choice of exponents and with the other parameters obeying

$$ab^n \int_0^\infty dx x^{n-1}\mathcal{G}(x) = a_n'' \tag{2.31}$$

simple scaling, (2.18), describes the behaviour of the moments of the total magnetisation in the finite size scaling regime. Apart from the two parameters a and b, the scaling function $\mathcal{G}(x)$ is fixed by the a_n''. Since the distribution $\mathcal{P}(M)$ is uniquely given by the set of moments, simple scaling *must* be obeyed by the PDF, if the moments are

[8]"Critical scaling" refers to the behaviour away from the critical point, when the thermodynamic limit is taken before $T \to T_c$.

given by (2.29) and (2.31),

$$\mathcal{P}(M;L) = aM^{-1}\mathcal{G}(M/(bL^{\Delta'})) \tag{2.32}$$

There is a more general statement regarding Eq. (2.29). If there are $n \neq k$ so that

$$\left\langle M^n \right\rangle^k (L) \propto \left\langle M^k \right\rangle^n (L) \tag{2.33}$$

and simple scaling in the form (2.29) applies, then $(1-\tau)n = (1-\tau)k$, *i.e.* $\tau = 1$ and therefore $\langle M^n \rangle(L) \propto L^{Dn}$ which means that (2.33) is true for *all* k, n for which the moments exist. However, a relation like (2.33) is indeed found for all equilibrium phase transitions with a RUSHBROOKE and a JOSEPHSON scaling law, namely

$$\begin{aligned} \langle M \rangle^2 &\propto L^{-2\beta/\nu+2d} \\ \left\langle M^2 \right\rangle_c &\propto L^{\gamma/\nu+d} \ . \end{aligned}$$

RUSHBROOKE implies $-2\beta/\nu + 2d = \gamma/\nu + (\alpha-2)/\nu + 2d$ and JOSEPHSON implies $\gamma/\nu + (\alpha-2)/\nu + 2d = \gamma/\nu + d$, so that $\langle M \rangle^2 \propto \left\langle M^2 \right\rangle_c$ and therefore $\langle M \rangle^2 \propto \left\langle M^2 \right\rangle$, leading to $\tau = 1$ via (2.33). The fact that simple scaling with $\tau = 1$ applies to the order parameter in a ferromagnetic phase transition has already been pointed out by AJI and GOLDENFELD (Aji and Goldenfeld, 2001) for the magnetisation density m. In fact, its probability density $\mathcal{P}(m)$ is related to $\mathcal{P}(M)$ via $\mathcal{P}(m)dm = \mathcal{P}(M(m))dM$ where $M(m) = mL^d$, leading to[9]

$$\mathcal{P}(m;L) = am^{-1}\mathcal{G}(m/(bL^{-\beta/\nu})) \ . \tag{2.34}$$

It is worth noting that a scaling with $\tau = 1$ implies that the distribution $\mathcal{P}(M)$ can be made to collapse by rescaling by the average. Even more general, since for $\tau = 1$ the rescaled PDF $\mathcal{P}(M;L)L^D$ collapses for different L if plotted against M/L^D, and $\langle M^n \rangle^{1/n} \propto L^D$, any $\mathcal{P}(M;L)\langle M^n \rangle^{1/n}$ plotted against $M/\left\langle M^k \right\rangle^{1/k}$ collapses for any n, k (Koba *et al.*, 1972; Nicolaides and Bruce, 1988; Botet and Płoszajczak, 2000)[10]. Vice versa, such a collapse uniquely indicates $\tau = 1$. It is interesting to note, that a finite BINDER-cumulant is only possible for $\tau = 1$, because $\left\langle M^4 \right\rangle / \left\langle M^2 \right\rangle^2$ diverges for all $\tau > 1$. This is consistent with the fact that a non-convergent or vanishing BINDER-cumulant usually indicates the breakdown of hyperscaling (in case of a vanishing BINDER-cumulant it is the onset of Gaussian fluctuations). But exactly hyperscaling has been used to prove that the exponent τ is equal to unity[11].

---

[9] Eq. (2.9) and Eq. (2.30) give $d - D = \beta/\nu$.

[10] Thanks to NICHOLAS MOLONEY for pointing out these references to me.

[11] Thanks to ANDREA GAMBASSI for interesting discussions about that point.

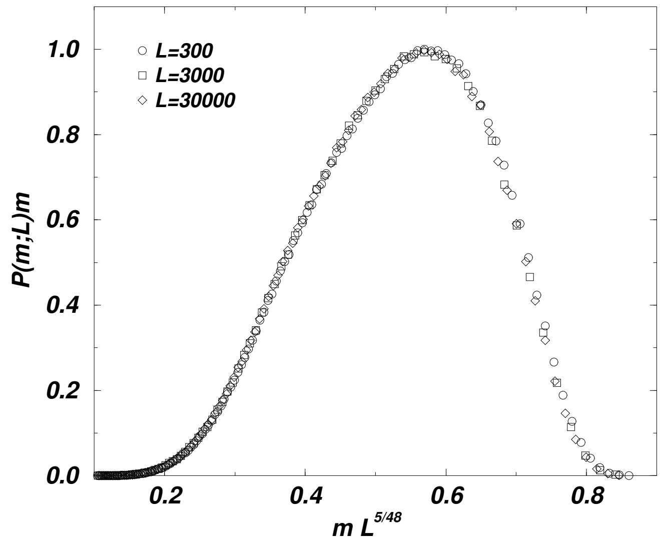


Figure 2.1: The probability $\mathcal{P}(m;L)$ that the largest cluster in a two-dimensional (site) percolation system at the critical point covers a fraction m of all sites in a system of size L, plotted in the form $\mathcal{P}(m;L)m$ versus $mL^{\beta/\nu}$ for $L = 300, 3000, 30000$. Since m is the order parameter, $\mathcal{P}(m;L)$ is expected to obey Eq. (2.34). That this is the case is manifested in the fact that data from different system sizes collapse on a single curve.

2.3.1 Percolation

One might ask, why $\tau \neq 1$ is found in *the* classical critical distribution function, namely in percolation, which is a standard equilibrium critical phenomenon by the mapping to the Potts model with $q \to 1$ (Kunz and Wu, 1978).

First of all, in classical percolation, what is usually measured in a distribution and characterised by an exponent τ is the *overall* cluster-size distribution, $\mathcal{P}_a(s)$, *i.e.* the density of clusters of size s per unit volume. However, the order parameter in percolation, *i.e.* the observable with respect to which RUSHBROOKE and JOSEPHSON scaling apply, is the fraction covered by the largest cluster, m. As shown in Fig. 2.1 for two-dimensional percolation, the order parameter indeed obeys (2.34), because $\mathcal{P}(m;L)m$ collapses (in the finite size scaling region) for different system sizes L to a single function if plotted versus $mL^{\beta/\nu}$, where $\beta/\nu = 5/48$.

However, the distribution of the fraction covered by the largest cluster is of course contained in the overall cluster size distribution. In fact, one would expect that the largest clusters are those contained in the bump at the tail of the distribution. Therefore, one expects that $\tau = 1$ for $\mathcal{P}_a(s;L)$ as well.

This transpires to be true, provided that the correct representation $\mathcal{P}_a(s;L)$ is used, namely one where $\mathcal{P}(m;L)$ becomes just a subset of it. Defining $\tilde{\mathcal{P}}_a(\tilde{s};L)$ to be the average number of clusters covering a fraction $\tilde{s} = s/L^d$ of the entire system per realisation, one has

$$\tilde{\mathcal{P}}_a(\tilde{s}; L) d\tilde{s} = L^d \mathcal{P}_a(s; L) ds \ , \tag{2.35}$$

where the factor L^d on the RHS is to bring up $\mathcal{P}(s; L)$ to an average number, rather than a density. The distribution of the order parameter is contained in $\tilde{\mathcal{P}}_a(\tilde{s}; L)$ as the average number of the largest cluster to cover a particular fraction of the entire system. From the behaviour of $\mathcal{P}_a(s; L)$ in the finite size scaling region

$$\mathcal{P}_a(s; L) = a s^{-\tau} \mathcal{G}(s/(bL^{1/(\nu\sigma)})) \tag{2.36}$$

with parameters a and b, exponents $\nu\sigma = 48/91$ obeying $d - \beta/\nu = 1/(\nu\sigma)$ from hyperscaling and $\tau = 187/91$ obeying $\tau - 1 = d\nu\sigma$ from other scaling laws (Stauffer and Aharony, 1994) and a scaling function $\mathcal{G}$, one has

$$\begin{aligned}\tilde{\mathcal{P}}_a(\tilde{s}; L) = L^{2d} \mathcal{P}_a(\tilde{s} L^d; L) = a L^{2d} \left(\tilde{s} L^d\right)^{-\tau} \mathcal{G}(\tilde{s}/(bL^{-\beta/\nu})) = \\ a L^{2d - \tau d} \tilde{s}^{-\tau} \left(b^{-1} \tilde{s} L^{\beta/\nu}\right)^{(\nu/\beta)(d\tau - 2d)} \tilde{\mathcal{G}}(\tilde{s}/(bL^{-\beta/\nu})) = \\ a b^{1-\tau} \tilde{s}^{-1} \tilde{\mathcal{G}}(\tilde{s}/(bL^{-\beta/\nu}))\end{aligned} \tag{2.37}$$

using a new scaling function $\tilde{\mathcal{G}}(x) x^{\tau - 1} = \mathcal{G}(x)$ and the scaling laws cited above, leading to $(\nu/\beta) d(\tau - 2) = \tau - 1$. Thus, expressed in these units, the overall cluster size distribution has $\tau = 1$.

While this transformation looks a bit artificial, it becomes less so, if one writes down the *average number* distribution of clusters of size s (rather than their density per unit volume):

$$L^d \mathcal{P}_a(s; L) = a L^d s^{-\tau} \mathcal{G}(s/(bL^{1/(\nu\sigma)})) = a b^{-\nu\sigma d} s^{-1} \tilde{\mathcal{G}}(s/(bL^{1/(\nu\sigma)})) \ , \tag{2.38}$$

where the power of L has been absorbed into a new scaling function $\tilde{\mathcal{G}}(x) x^{\nu\sigma d} = \mathcal{G}(x)$ and $\nu\sigma d = \tau - 1$ has been used. Fig. 2.2 shows such a number distribution for a site percolation system in the form $L^d \mathcal{P}_a(s; L) s$ versus $s/L^{1/(\nu\sigma)}$. Of course, the average number distribution is closely related to $\tilde{\mathcal{P}}_a(\tilde{s}; L)$, as can be seen in (2.36).

2.3.2 Cumulants and lower cutoff

The reason why no conclusions can be drawn for τ of the distribution of the order parameter away from the critical point[12] is because of the inconvenient form of the cumulants (2.8), which does not allow us to infer a behaviour like (2.12) for the *moments*. It is therefore very interesting to see how the cumulants of a distribution obey-

[12] Apart from has been said in footnote 7 on page 79.

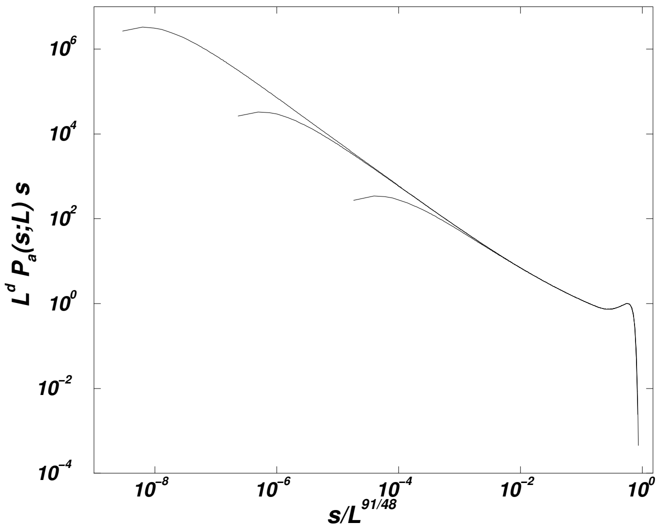


Figure 2.2: The number distribution, *i.e.* average number per realisation, $L^d\mathcal{P}_a(m; L)$ of clusters of size s for various system sizes ($L = 300, 3000, 30000$), plotted in the collapsing form $L^d\mathcal{P}_a(m; L)s$ versus $s/L^{91/48}$. The data are binned and shown in a double-logarithmic plot. The collapse confirms Eq. (2.38).

ing simple scaling look like.

Before the cumulant generating function is calculated, a non-vanishing lower cutoff is introduced. Later in this section it will become clear why this is necessary. For the time being, it is only noted that the general form of a PDF obeying *simple scaling*, (2.18), is

$$\mathcal{P}(M; L, \xi) = \begin{cases} aM^{-\tau}\mathcal{G}\left(\dfrac{M}{M_0(L)}, \dfrac{L}{\xi}\right) & \text{for } M > M_l \qquad (2.39\text{a}) \\ f(M; L, \xi) & \text{otherwise} \qquad (2.39\text{b}) \end{cases}$$

There is some crossover behaviour in the region of $M \approx M_l$, which could in principle be captured by the function $f(M; L, \xi)$. However, for simplicity, in the following the (non-universal) function $f(M; L, \xi)$, which describes the behaviour of the distribution for small arguments, will be assumed to dependent only on M. In the two limits as discussed above in (2.24) and (2.25), the same behaviour is of course expected again in (2.39) for $M > M_l$. In the following, $M_0(L)$ and $\widetilde{M_0}(\xi)$ are both referred to by M_0 alone. The same applies to $\mathcal{G}(x)$, which either refers to $\mathcal{G}(x, 0)$ as in (2.24) or to $\widetilde{\mathcal{G}}(x)$ as in (2.25). See, however, the remarks around (2.27).

Apparently one can split the integration for $\langle M^n\rangle$ into two parts, one for small $M < M_l$, one for large $M > M_l$. Defining

$$f_n = \int_0^{M_l} dM\ M^n f(M) \qquad (2.40)$$

one has

$$\langle M^n \rangle = f_n + aM_0^{1+n-\tau} \int_{M_l/M_0}^{\infty} dx\, x^{n-\tau} \mathcal{G}(x) \ . \tag{2.41}$$

To determine the asymptotic behaviour of the moment, one splits the remaining integral in two parts, defining a threshold y, below which $\mathcal{G}(x)$ can be expanded up to finite order to satisfying accuracy. This is the crucial assumption for the following calculations: $\mathcal{G}(x)$ is analytic around $x = 0$. Defining

$$q_n^{\infty}(y) = \int_{y}^{\infty} dx\, x^{n-\tau} \mathcal{G}(x) \tag{2.42}$$

$$q_n^{1}(M_l/M_0, y) = \int_{M_l/M_0}^{y} dx\, x^{n-\tau} \mathcal{G}(x) \tag{2.43}$$

for fixed y one has

$$\langle M^n \rangle = f_n + aM_0^{1+n-\tau} \left(q_n^{\infty}(y) + q_n^{1}(M_l/M_0, y) \right) \tag{2.44}$$

where q_n^1 can be expanded like

$$\begin{aligned} q_n^{1}(M_l/M_0, y) &= \frac{1}{1+n-\tau} \left(y^{1+n-\tau} - \left(\frac{M_l}{M_0} \right)^{1+n-\tau} \right) \mathcal{G}(0) \\ &+ \frac{1}{2+n-\tau} \left(y^{2+n-\tau} - \left(\frac{M_l}{M_0} \right)^{2+n-\tau} \right) \mathcal{G}'(0) \\ &+ \ \dots \end{aligned} \tag{2.45}$$

for any $1 + n - \tau \neq 0$. The special case $n = \tau - 1$ is simply

$$\left\langle M^{\tau-1} \right\rangle = f_{\tau-1} + aq_{\tau-1}^{\infty}(y) + a \int_{M_l/M_0}^{y} dx\, x^{-1} \mathcal{G}(x) \tag{2.46}$$

where the integral $q_{\tau-1}^{\infty}(y)$ exists and is finite for all $y > 0$, as $\mathcal{G}(x)$ is finite everywhere, it even decays fast enough such that all its moments exist, see (2.19).[13] Expanding $\mathcal{G}(x)$ again around $x = 0$, the rightmost integral in (2.46) is

$$\int_{M_l/M_0}^{y} dx\, x^{-1} \mathcal{G}(x) = \left(\ln(y) - \ln\left(\frac{M_l}{M_0} \right) \right) \mathcal{G}(0) + \left(y - \frac{M_l}{M_0} \right) \mathcal{G}'(0) + \dots \tag{2.47}$$

which is asymptotically (in divergent M_0) logarithmically divergent for $\mathcal{G}(0) \neq 0$ and convergent for $\mathcal{G}(0) = 0$; there is no other "counter-divergence" for example in $q_{\tau-1}^{\infty}(y)$, the value of which depends only on y, which can be hold fixed. The case

[13]For $y > 0$ the integral (2.42) could diverge only if $\mathcal{G}(x)$ would diverge at some finite value x or if it would not converge because of the upper bound. The latter, however, cannot be the case if all moments exist, (2.19). Problems at the lower bound are not important provided that $y > 0$.

$n = \tau - 1$ is particularly important for $\tau = 1$, because it represents the normalisation. This is the first fundamental result of this section: If simple scaling in the form (2.39) applies and $\tau = 1$, then $\mathcal{G}(0) = 0$, otherwise the distribution cannot be normalised; of course that does *not* entail $\mathcal{P}(0)$ to vanish in case of $\tau = 1$, in fact it does *not* if $\mathcal{G}(x)$ is linear in x for small arguments.

The leading order for divergent M_0 in the case of $1 - \tau + n < 0$ is according to (2.44) and (2.45)

$$\langle M^n \rangle = f_n - a \frac{M_l^{1+n-\tau}}{1+n-\tau} \mathcal{G}(0) + \dots \tag{2.48}$$

where all terms hidden in $\dots$ vanish as M_0 diverges. For $n = 0$ this result implies $M_l > 0$ if $\tau > 1$ and $\mathcal{G}(0) \neq 0$, otherwise the distribution cannot be normalised.[14] Of course the opposite does not hold: $\tau = 1$ does *not* imply $M_l = 0$.

Similarly for $1 - \tau + n > 0$; the leading term of the result reads

$$\langle M^n \rangle = a M_0^{1+n-\tau} \left(q_n^\infty + \frac{y^{1+n-\tau}\mathcal{G}(0)}{1+n-\tau} + \frac{y^{2+n-\tau}\mathcal{G}'(0)}{2+n-\tau} + \dots \right) \tag{2.49}$$

where the dots now include further terms which are not asymptotically vanishing in M_0, but can be prepared so by choosing smaller and smaller y as M_0 increases. Moreover all powers of y are contained in the lower boundary of the integral $q_n^\infty(y)$, see (2.42). In fact, since f_n and M_l become asymptotically irrelevant, one has

$$\langle M^n \rangle = a M_0^{1+n-\tau} g_n \tag{2.50}$$

where

$$g_n \equiv \int_0^\infty dx\, x^{n-\tau} \mathcal{G}(x) \ . \tag{2.51}$$

Thus, all moments $n > \tau - 1$ diverge with divergent M_0. For $n = 0$ Eq. (2.49) would imply a divergent normalisation if $\tau < 1$, which is therefore impossible.[15]

Thus, there are only two relevant cases: The marginal case $\tau = 1$ as found for the order parameter in ferromagnetic phase transitions and which entails $\mathcal{G}(0) = 0$. The general case is $\tau > 1$ and it is abundantly found in SOC; as seen above, such a value of τ necessarily requires $M_l > 0$, if $\mathcal{G}(0) \neq 0$.

All results have been derived as asymptotes for divergent M_0. This is of course especially applicable for divergent L in the finite size scaling region or divergent ξ in the critical region away from the critical point, if the thermodynamic limit has been taken first. In these cases one can simply replace M_0 by bL^D or $\widetilde{b}\xi^D$, see (2.23) and

---

[14] Clearly, a negative lower cutoff makes sense only if the distribution $\mathcal{P}(M)$ is defined for negative arguments.

[15] The crucial point is that $\tau < 1$ would make the normalisation divergent in the large M region. Moreover, to avoid asymptotic violation of $\langle M^2 \rangle \geq \langle M \rangle^2$ one needs $D(3-\tau) \geq 2D(2-\tau)$ so that $\tau \geq 1$ for any $D > 0$.

(2.26), respectively.

The case $\tau = 1$ implies $\langle M \rangle \propto M_0$. Therefore, the confusion of the average with the characteristic size does not cause any trouble in equilibrium critical phenomena. However, when $\tau > 1$, it is highly important to distinguish average and cutoff (or characteristic size).

2.3.2.1 Cumulant generating function

Along the lines of the calculations above, one can now compute the cumulant generating function of $\mathcal{P}(M; L, \xi)$. To this end the characteristic function (van Kampen, 1992) has to be calculated, which boils down to

$$
\begin{aligned}
\hat{\mathcal{P}}(z; L, \xi) &\equiv \int_0^\infty dM\ e^{zM} \mathcal{P}(M; L, \xi) &\text{(2.52a)}\\
&= \int_0^{M_l} dM\ e^{zM} f(M) &\text{(2.52b)}\\
&\quad + a M_0^{1-\tau} \left(\int_y^\infty dx\ e^{zM_0 x} x^{-\tau} \mathcal{G}(x) + \int_{M_l/M_0}^y dx\ e^{zM_0 x} x^{-\tau} \mathcal{G}(x) \right)
\end{aligned}
$$

In this form it becomes particularly clear why $M_l > 0$ for $\tau > 1$: if $M_l = 0$, then the first integral vanishes and the two integrals in the brackets can be merged to a single integral, producing a result, which only depends on the product xM_0, say

$$
\hat{\mathcal{P}}(z; L, \xi) = a M_0^{1-\tau} \int_0^\infty dx\ e^{zM_0 x} x^{-\tau} \mathcal{G}(x) = a M_0^{1-\tau} \hat{\mathcal{G}}(zM_0)\ . \tag{2.53}
$$

Taking the logarithm of this equation gives an expression which depends on τ only implicitly via $\hat{\mathcal{G}}(zM_0)$; in fact one gets

$$
\langle M^n \rangle_c = M_0^n c_n \tag{2.54}
$$

with some constants c_n, implying a scaling like in (2.33), as discussed in sec. 2.3 and therefore $\tau = 1$. Hence, again [see after (2.48)], $M_l = 0$ implies $\tau = 1$ or no scaling at all. See also footnote 7 on page 79.

With $M_l > 0$ the resulting CGF is more complicated; defining

$$
\begin{aligned}
\hat{f}(z) &= \int_0^{M_l} dM\ e^{zM} f(M) &\text{(2.55a)}\\
\hat{\mathcal{G}}(z) &= \int_y^\infty dx\ e^{zM_0 x} x^{-\tau} \mathcal{G}(x) &\text{(2.55b)}\\
\hat{\mathcal{H}}(z, M_l/M_0) &= \int_{M_l/M_0}^y dx\ e^{zM_0 x} x^{-\tau} \mathcal{G}(x) &\text{(2.55c)}
\end{aligned}
$$

one can write $\ln(\hat{\mathcal{P}}(z))$ in the form

$$\ln(\hat{\mathcal{P}}(z)) = \ln(\hat{f}(z)) + \ln\left[1 + aM_0^{1-\tau}\left(\frac{\hat{\mathcal{G}}(z)}{\hat{f}(z)} + \frac{\hat{\mathcal{H}}(z, M_l/M_0)}{\hat{f}(z)}\right)\right] \tag{2.56}$$

which, in general, does not give a simple form of scaling for the cumulants $\langle M^n \rangle_c$. That finally settles the discussions about the difference between (2.8) and (2.11) in sec. 2.2.1: either $\tau = 1$, but then (2.54) applies which is different from (2.8) (critical scaling) but fits (2.11) (finite size scaling) perfectly. Or $\tau > 1$ but then simple scaling does not imply any simple form of the cumulants, especially not (2.8). It remains unclear how to characterise the PDF for a cumulant scaling like (2.8); apparently it is not compatible to simple scaling of the form (2.39) with $M_l = 0$, because then (2.8) would be of the form (2.54).

Open question

2.3.2.2 Corrections to scaling

From equilibrium critical phenomena, one is used to the idea that scaling forms like Eq. (2.39) usually contain sub-leading corrections, as famously shown by WEGNER (Wegner, 1972). As will be shown below, these corrections are expected to be less important as the system size is increased. Maybe for that reason, these "finite size corrections" are sometimes confused with "finite size scaling".

The numerical techniques discussed in the next section will mainly focus on single moments. Thus, the main interest in corrections to scaling is for their effect to the scaling of moments, not so much for the way they change the scaling form (2.39) itself.

From Eq. (2.45) it is already clear that sub-leading terms will always enter at least in the form

$$\langle M^n \rangle = \begin{cases} a_0 + a_0' M_0^{1+n-\tau} + a_1 M_0^{-1} + a_2 M_0^{-2} \dots & \text{for } 1+n-\tau < 0 \quad \text{(2.57a)} \\ b_0 M_0^{1+n-\tau} + b_0 + +b_1 M_0^{-1} + b_2 M_0^{-2} \dots & \text{for } 1+n-\tau > 0 \quad \text{(2.57b)} \end{cases}$$

with constants $a_n^{(\prime)}$ and b_n. The seemingly paradox situation, that the 0th moment cannot have any corrections while it acquires some in the general form (2.39) is resolved by noting that there is a crossover behaviour around M_l and there are indeed WEGNER's corrections to scaling. Thus, in general, one expects the scaling for $1+n-\tau > 0$ to be of the form (Landau and Binder, 2000)

$$\langle M^n \rangle = M_0^{1+n-\tau}(c_0 + c_1 M_0^{-\omega_1} + c_2 M_0^{-\omega_2} + \dots) \tag{2.58}$$

with constants c_n and $0 < \omega_1 < \omega_2 < \dots$; these powers are sometimes called "confluent singularities" (Privman *et al.*, 1991).

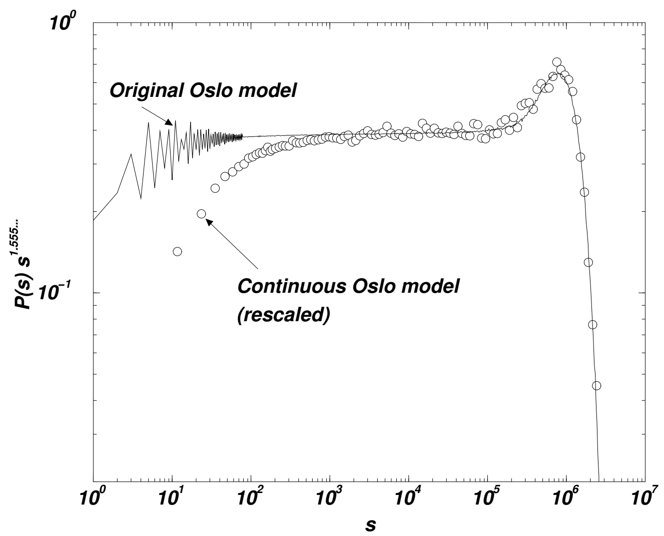


Figure 2.3: Comparison of the scaling function of the original and the continuous Oslo model (see chapter 5). The scaling function is obtained by multiplying the PDF (distribution of the avalanche sizes s) by an appropriate power of s, in this case $\tau = 1.555\ldots$. One can clearly see the onset of the actual scaling behaviour and the presumably non-universal behaviour of the distribution below the lower cutoff.

Most remarkably, the insight that corrections to scaling of the WEGNER type are *necessary* in order to explain the scaling behaviour of the moments of the avalanche size distribution in the DS-FFM, has been celebrated in an article by PASTOR-SATORRAS and VESPIGNANI (Pastor-Satorras and Vespignani, 2000a), ignoring many of the details discussed above. It is worth stressing that a scaling form (2.58) is what is usually expected in equilibrium critical phenomena on textbook level (Landau and Binder, 2000).

2.3.3 Scaling function

One of the most appealing features of critical phenomena is universality. Universality was almost completely empty, if that just meant that exponents are the same; the way to decide that two seemingly different models belong to the same universality class is by finding the same (independent) exponents. But universality means that even more critical properties are identical, most prominently the scaling function ($\mathcal{G}(x)$ in (2.18)), which encodes the universal moment ratios.

Up to constants, the scaling function can be obtained by multiplying the PDF $\mathcal{P}(M)$ with the appropriate power of its argument. Fig. 2.3 shows the so-produced scaling function of the original Oslo model (see chapter 5) together with a continuous variant (see Sec. 5.2.1.2). The observable is the avalanche size s, the exponent τ is $\approx 1.555\ldots$. The data of the continuous model is for technical reasons much noisier than the data for the original model. For the continuous model, ordinate and abscissa have been rescaled to accommodate for different non-universal constants a and b as introduced in (2.18).

These two constants, or metric factors, a and b, are two non-universal degrees of freedom, which could in principle be absorbed into the scaling function $\mathcal{G}(x)$, even though one certainly want the scaling function to be dimensionless. However, absorbing the constants would render the scaling function non-universal. The idea is to impose two *size and system independent* properties on the scaling function to fix the two metric factors. These two properties must be chosen such that scaling functions obtained in different systems or at different system sizes collapse if there is *any* choice of a and b which would do the job. One obvious choice is to fix the absolute maximum of the scaling function to sit exactly at $x = 1$ and $y = 1$. However, some scaling function obtain their maximum at $x \to 0$, for example the scaling function of one-dimensional percolation $\exp(-x)$.

As will be shown below, a very useful choice is to impose

$$g_1 = \int_0^\infty x^{1-\tau}\mathcal{G}(x) \quad = \quad 1 \tag{2.59a}$$

$$g_2 = \int_0^\infty x^{2-\tau}\mathcal{G}(x) \quad = \quad 1\,, \tag{2.59b}$$

for $2 > \tau \geq 1$, where g_n have been defined in (2.51). It is important to realise that these conditions can *always* be met, *i.e.* if there is a scaling function $\mathcal{G}(x)$ given by imposing whatever conditions to fix a and b, there is a choice of a' and b' such that $\mathcal{G}'(x) = a'\mathcal{G}(b'x)$ and $\mathcal{G}'(x)$ obeys (2.59). In fact, this is almost a trivial statement, assuming that the two integrals $\int_0^\infty x^{1-\tau}\mathcal{G}(x) = g_1$ and $\int_0^\infty x^{2-\tau}\mathcal{G}(x) = g_2$ actually exist, as they have to if the corresponding moments exist. In that case, one easily shows that the choice $b' = g_2/g_1$ and $a' = b'^{2-\tau}/g_1$ leads to $\mathcal{G}'(x)$ obeying (2.59).

2.3.3.1 Universal moment ratios

Based on (2.50), which is valid for $n > \tau - 1$, and (2.59) one can easily show that[16]

$$\frac{\langle M^n\rangle\langle M\rangle^{n-2}}{\langle M^2\rangle^{n-1}} = g_n \tag{2.60}$$

[16]In the presence of a free energy f obeying (2.5) and (2.13), already ratios like $\langle M^{2n}\rangle/\langle M^2\rangle^n$ are universal in the finite size scaling regime, since the moments are polynomials of derivatives of f, $\partial^m f/\partial H^m$, each of which having an appropriate power of the metric factor K_h in front of it, Eq. (2.13). These metric factor are the only non-universal quantities and cancel in an expression like $\langle M^{2n}\rangle/\langle M^2\rangle^n$. Of course, similar conclusions cannot be drawn if the thermodynamic limit is taken before $t \to 0$, where the cumulants obey (2.8), because in that case the moments are *not* just simple polynomials of the cumulants, see the comment at the end of sec. 2.2.1. Also, this conclusion cannot be reached by just putting $\tau = 1$ in (2.61) according to (2.30), because of the remaining power of a.

with g_n defined in (2.51), which is universal as $\mathcal{G}(x)$ is so. It is worth noting that (2.59) has only been used to identify the RHS of Eq. (2.60) and therefore its universality. However, $g_1 = g_2 = 1$ is not needed to actually compute the LHS of Eq. (2.60) from numerics. As one easily checks by evaluating at $n = 1$ and $n = 2$, (2.60) is compatible with $g_1 = g_2 = 1$.

Eq. (2.60) is just one choice of a general way to construct universal moment ratios. Defining

$$\mu_{n,m} \equiv \frac{\langle M^n \rangle}{\langle M^m \rangle^{n/m}} = \left(a(bM_0)^{1-\tau}\right)^{1-n/m} \frac{g_n}{g_m^{n/m}} \tag{2.61}$$

any ratio $\mu_{n,m}/\mu_{k,l}^q$ is universal provided that $(1 - k/l)q = 1 - n/m$. Eq. (2.60) corresponds to $m = 2$, $k = 1$, $l = 2$ and $q = 2 - n$. It is worth noting that n does not need to be integer valued, so that $g_{m+\tau}$ is proportional the mth moment (instead of $m - \tau$) of $\mathcal{G}(x)$ (2.51), $m \in \mathbb{N}$.

For $\tau \geq 2$, (2.60) breaks down, because (2.50) does not apply for $n = 1$. Nevertheless, similar expressions can be found via (2.61).

The case $\tau = 1$ deserves special attention, not least because universal moment ratios[17] are well-established [for example (Salas and Sokal, 2000)] in equilibrium critical phenomena (at least in finite size scaling). Again, based on (2.50) one has

$$\frac{\langle M^n \rangle}{\langle M^2 \rangle^{n/2}} = a^{1-n/2} g_n \tag{2.62}$$

where $g_2 = 1$ has been used again and a is in fact dimensionless. The reason why these amplitude ratios are universal (Privman and Fisher, 1984), which implies a universal a, provided that (2.59) are obeyed, is because they are based on a universal singular part of the free energy as discussed in sec. 2.2.1. As suggested in (2.13) and supported by dimensional analysis, all powers of the order parameter contained in the cumulants are multiplied by the same non-universal pre-factor, a power of the metric factor K_h. Their *ratio* however, is determined by the universal, singular part of the free energy. This can be seen by summing appropriate powers (van Kampen, 1992) of cumulants to produce individual moments. According to (2.13) they all contain the same leading pre-factor K_h^n which contains all non-universality and cancels in a ratio like (2.62). Hence, the ratio (2.62) is universal for equilibrium critical phenomena — if b is fixed, a is so too.

Numerically, one finds that this result extends to the cluster size distribution of percolation (Moloney and Pruessner, 2003); handwavingly one can understand this by the fact that the order parameter distribution is contained in the overall cluster

[17] Again, either by taking $L \to \infty$ first and then obtaining the behaviour in ξ or by taking $\xi \to \infty$ first and then obtaining the scaling in L, which, however, produces geometry dependent ratios, see Sec. 2.2.1.1, page 77.

Open question

size distribution as discussed in sec. 2.3.1. However, it remains an open question, whether it is possible to derive the universality of a from properties of $\mathcal{P}(s)$ (for percolation) directly.

There is another way to observe the universality of a: Considering the normalisation of $\mathcal{P}(M)$, the importance of which has already been pointed out around (2.46), one finds

$$\langle M^0 \rangle = f_0 + a(y\mathcal{G}'(0) + q_0^\infty(y)) - a\frac{M_l}{M_0}\mathcal{G}'(0) + \dots \tag{2.63}$$

using $\mathcal{G}(0) = 0$ found for $\tau = 1$ in sec. 2.3.2. The rightmost term and all terms included in $\dots$ vanish asymptotically. Since (2.63) is independent from y for sufficiently large M_0, and f_0 vanishes if there is no lower cutoff, as for example in magnetisation density distributions, the RHS of (2.63) contains only universal properties of $\mathcal{G}(x)$, multiplied by a. As this must give identity, $\langle M^0 \rangle = 1$, the universality of a follows.

2.4 Numerical Techniques in SOC

Skimming through the SOC literature it seems that only a few people are actually concerned about using or even establishing standard numerical techniques to produce accurate results with reliable error bars. It is important to stress that due to the absence of temperature-like parameters, finite size scaling is (supposedly) the *only* technique available in SOC. Thus, again M_0 is identical to bL^D.

2.4.1 Data collapse

The most common method of data analysis is to produce a data collapse, *i.e.* to collapse the data for different system sizes on to a single function. This technique is based on simple scaling (2.39); however, there is not a unique way of producing a data collapse. One way is to try to find an exponent τ and for each L consider a value of $M_0(L)$, such that a plot $\mathcal{P}(M)M^\tau$ versus $x = M/M_0(L)$ collapses on to a single function $a\mathcal{G}(x)$. This has been used to produce the plots shown in Fig. 2.1, Fig. 2.2 as well as Fig. 2.3. Alternatively one can plot $\mathcal{P}(M)M_0^\tau$ versus $x = M/M_0$ which collapses to $ax^{-\tau}\mathcal{G}(x)$. To illustrate the difference, Fig. 2.4 shows the results of these techniques. It is indeed a rather mundane reason, why the method to collapse to $a\mathcal{G}(x)$ (left plot) is superior than to collapse to $ax^{-\tau}\mathcal{G}(x)$ (right plot): since the latter drops off continuously over the entire range, one needs to allow for a much wider range in the ordinate. Thus, details, especially those close to the cutoff, are not as clearly visible as in the collapse to the scaling function $a\mathcal{G}(x)$ alone. In principle there are infinitely many possibilities to collapse $\mathcal{P}(M)$, since for any μ the product

$$\mathcal{P}(M)M^\tau \left(\frac{M}{M_0}\right)^\mu \tag{2.64}$$

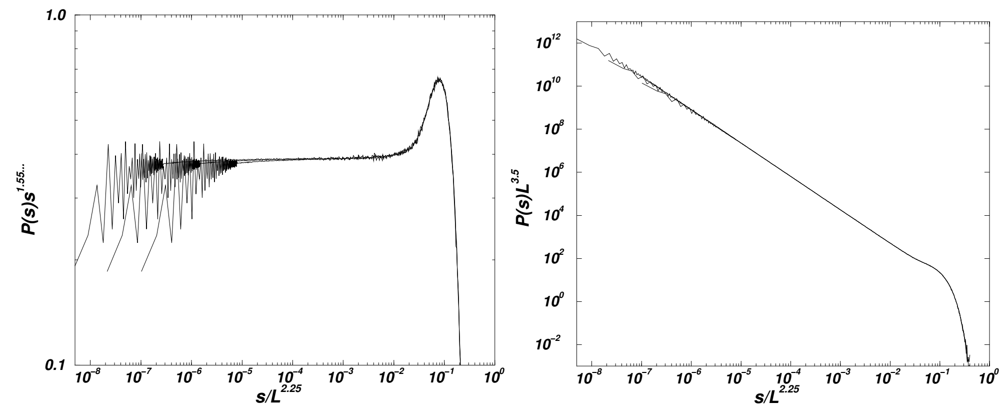


Figure 2.4: Comparison of different ways of collapsing data which follows simple scaling in a double logarithmic plot. The data is taken from the original Oslo model, where s is the avalanche size. The left plot shows $\mathcal{P}(s)s^{\tau}$ versus s/L^D, the right one $\mathcal{P}(s)L^{D\tau}$ versus the same s/L^D. Due to the continuous decrease, the right plot cannot show as much details as the left. Data are taken from the original Oslo model with $L = 1280, 2560, 5120$, and exponents $\tau = 1.555\ldots$ and $D = 2.25$, *i.e.* the PDF scales like $\mathcal{P}(s) = as^{-\tau}\mathcal{G}(s/(bL^D))$.

is a function only of M/M_0 and therefore collapses when plotted against this ratio.

In any case, these plots usually span many orders of magnitude and therefore it is standard to plot them double-logarithmically. Fig. 2.5 shows the same data on a linear scale, stressing the advantages of the plot of $\mathcal{P}(s)s^{\tau}$.

To determine an error bar or any measure of quality of a data collapse is a serious technical difficulty (Bhattacharjee and Seno, 2001). The reason is that two histograms based on two different system sizes have most likely not a single value of $x = s/L^D$ in common. In fact, the data consists of sets of point for each system size, for which one is supposed to test, whether they fall on the same continuous function. One qualitative way is to plot the binned (see Sec. 2.4.2) histograms $\mathcal{P}(M; L)$ with error bars on each point and connecting these points. If the resulting line for one system size lies within the error bar of the data points of the other system size, then the collapse is trustworthy. A technique like this has been employed in (Pruessner and Jensen, 2002a).

Such a technique can only give a qualitative answer and can much better be used convincingly to prove that a proper data collapse is not possible, rather than to prove the opposite. As will be explained in detail in chapter 4, the error bars needed for each data point must be derived from the variance of the time series of histograms, which can be computationally very intense.

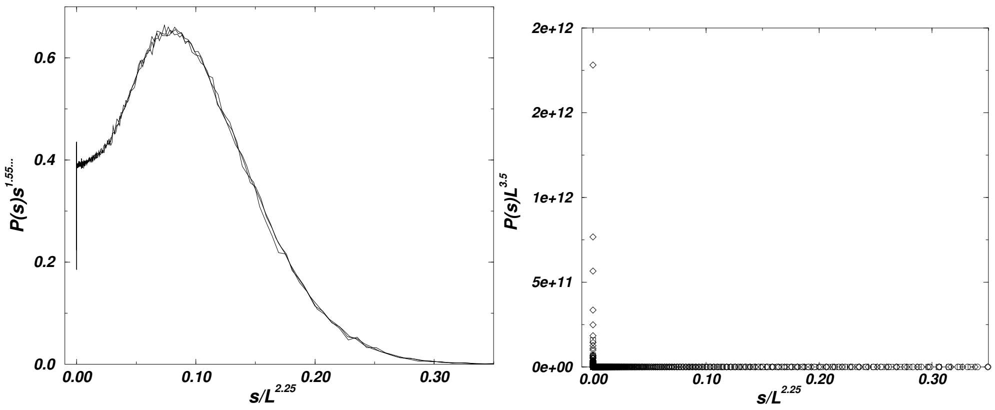


Figure 2.5: Similar to Fig. 2.4 and using the same data, comparison of different ways of collapsing data, but now in a linear plot. The right plot shows $\mathcal{P}(s)L^{D\tau}$ versus s/L^D using symbols to make the data visible, the left plot shows $\mathcal{P}(s)s^\tau$.

2.4.1.1 The lower cutoff and how to determine τ

When doing a data collapse, one only assumes that there is a lower cutoff and tries to perform the collapse. One focuses on the behaviour of the distribution for large arguments, not only in order to avoid problems with the lower cutoff, but also to keep corrections as low as possible. The easiest way to perform a collapse, is to concentrate on a "landmark" in the distribution, such as a pronounced bump before the PDF drops down, and to change $M_0(L)$ and exponent τ, until $\mathcal{P}(M)M^\tau$ collapses for various L onto the same function for large M when plotted versus M/M_0. There is no need to make assumption about the lower cutoff and the exact scaling of M_0; its dependence on L can be extracted in a second step. The only assumption is that the landmark is not affected by the lower cutoff or further corrections.

The scaling of the landmark itself in a double logarithmic plot of the data gives the exponent τ; if for example the maximum of the scaling function is at x^*, then the observable M at this maximum is $M^* = x^* M_0(L)$, so that $\mathcal{P}(M = M^*; L) = a(M^*)^{-\tau}\mathcal{G}(x^*)$. There are only constants on the RHS, apart from M^*, which is an implicit function of L. Thus, the set of points $(\mathcal{P}(M = M^*; L), M^*)$ describes are line with slope $-\tau$ (see Fig. 2.7). Of course, this "trick" is based on the assumption that one actually identifies in $\mathcal{P}(M)$ a feature of the scaling function. However, due to the presence of the pre-factor $M^{-\tau}$ one can never be sure whether one sees a feature of $\mathcal{G}(x)$ alone and not of the product $M^{-\tau}\mathcal{G}(M/M_0)$.

One might be tempted to measure the exponent τ simply from the (intermediate) straight-line behaviour of $\mathcal{P}(M)$, *i.e.* above the lower cutoff and sufficiently far away from the upper cutoff. For large cutoffs one might assume that $\mathcal{G}(M/M_0)$ hardly changes, and therefore $\mathcal{P}(M)$ to be effectively proportional to $M^{-\tau}$, so that the latter exponent reveals itself as the slope of $\mathcal{P}(M)$ in a double logarithmic plot. This,

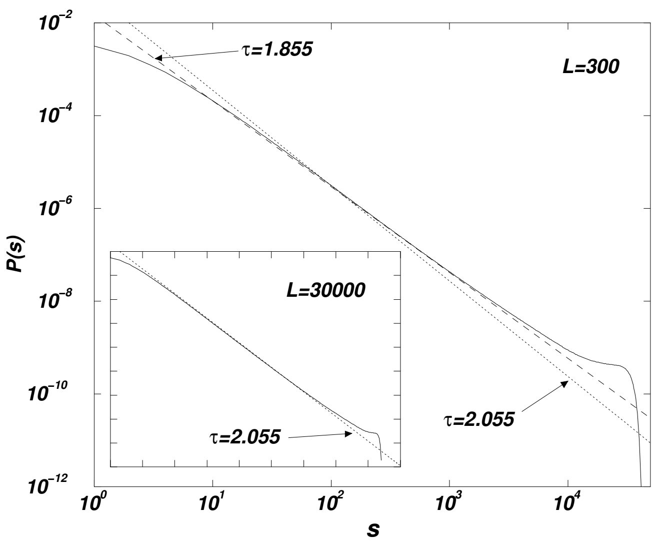


Figure 2.6: Cluster size distribution $\mathcal{P}(s)$ for site percolation with $L = 300$ (main panel) and $L = 30000$ (inset). Neither of these systems should be considered as particularly small. However, the best estimate for the intermediate slope of $\mathcal{P}(s)$ for $L = 300$ in a double logarithmic plot, is, as shown as a dashed line, $\tau = 1.855$ which is 10% off the correct value $\tau \approx 2.055$ shown as a dotted line. In the inset, the correct exponent is shown to fit fairly good the intermediate slope of the data for $L = 30000$.

however, does not work: From a single set of data, it is impossible to judge where the lower cutoff is, therefore it is impossible to decide, what "intermediate" means. Moreover, it is *a priori* unclear where the scaling function can be ignored, *i.e.* in which region of M values the distribution is dominated by the $M^{-\tau}$.

The danger of ignoring the scaling function becomes particularly clear when expanding the scaling function for small arguments:

$$\mathcal{P}(M) = aM^{-\tau}\left(\mathcal{G}(0) + \frac{M}{M_0}\mathcal{G}'(0) + \dots\right) \tag{2.65}$$

If $\mathcal{G}(0)$ vanishes [see discussion after (2.47)], the apparent exponent, which would be measured in a double logarithmic plot, is $1 - \tau$. This, however, would ignore that the cutoff is allowed to enter the scaling only in the form M/M_0. The result is even more misleading if $\mathcal{G}(x)$ is not analytic and can be expanded like $x^{\mu_0}a_0 + x^{\mu_1}a_1 + \dots$ with non-integer exponents μ_i.

Even if $\mathcal{G}(0) \neq 0$, a direct "straight line fit" can be very misleading. Fig. 2.6 shows the high-quality PDF of the cluster size distribution for system size $L = 300$ and open boundaries.

The idea mentioned above to concentrate on a landmark is further illustrated by

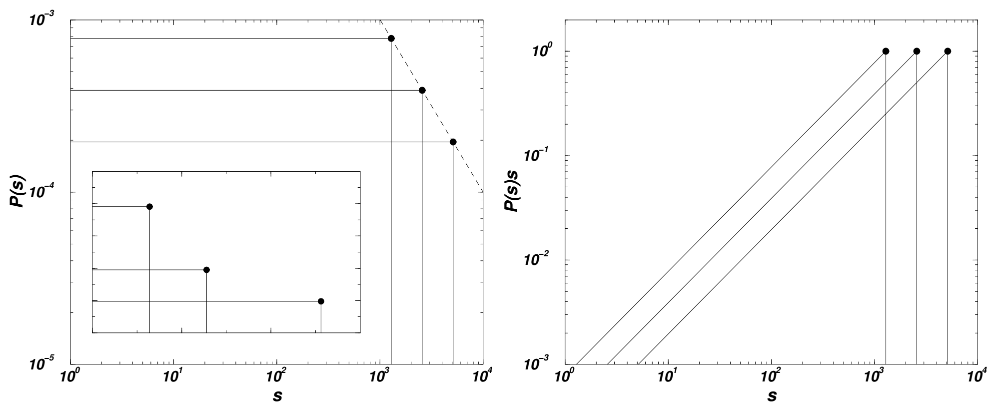


Figure 2.7: Left panel: Avalanche size distribution $\mathcal{P}(s)$ for the randomly bulk driven BTW model. The inset shows the data on a linear scale, the main panel shows it on a logarithmic scale. The "obvious" exponent is $\tau = 0$. The landmark used to facilitate the collapse is marked by circles and connected by a dashed line, which has slope $\tau = -1$. Right panel: The only way to bring all landmarks up to the same height is by rescaling $\mathcal{P}(s)$ by s, indicating an exponent $\tau = 1$.

studying a PDF with a scaling function that vanishes at $x = 0$. A good example[18] is the randomly (bulk) driven one-dimensional BTW model (see Sec. 1.3.1, page 40). The PDF of the avalanche sizes s found in this model is a uniform distribution in the interval $[1, L]$, *i.e.*

$$\mathcal{P}(s) = L^{-1}\theta(L - s) = s^{-1}\frac{s}{L}\theta(1 - \frac{s}{L}) \tag{2.66}$$

where $\theta(x)$ is the HEAVISIDE θ-function. The exponent of this PDF is $\tau = 1$ and the scaling function is $(s/L)\,\theta(1 - s/L)$, which vanishes for $s = 0$. When examining the non-rescaled PDF, shown in Fig. 2.7, it is tempting to claim $\tau = 0$. However, when trying to collapse the data, especially the "landmarks", it becomes clear that one needs a non-vanishing exponent to get the landmarks all up to the same height via $s^\tau \mathcal{P}(s)$ which leads directly to the collapse when shifting horizontally using a cutoff $s_0(L)$.

The paradox outcome of a collapse like the one shown in Fig. 2.7 is the reason to question the physical relevance[19] of cases with $\mathcal{G}(0) = 0$: In most natural system one cannot derive the data for various system sizes and perform a collapse. The only accessible property of the distribution is in fact the slope in a double logarithmic plot. If that gives spurious results for the exponent τ as defined by simple scaling, can that definition be reasonable?

[18]Thanks to KIM CHRISTENSEN for explaining that to me.
[19]Note again $\tau = 1$ entails $\mathcal{G}(0) = 0$, as discussed after (2.47).

2.4.2 Binning

Numerical simulations produce histograms which need to be post-processed in order to extract relevant information from them. The main problem is the sparse coverage of rare events. Since most of the analysis is based most naturally on double-logarithmic data, a histogram for integer valued events (such as avalanche sizes in most sandpile-like systems) or a histogram with equally sized bins has a large density of points on the abscissa for large arguments. At the same time very few events actually take place there. One way of curing this problem is to bin the data.

The idea is to define intervals $[b_i, b_{i+1}]$ and to average the numerically calculated histogram $h(s)$ over these intervals:

$$h'(i) = \frac{1}{b_{i+1} - b_i} \sum\!\!\!\!\!\!\!\int_{[b_i, b_{i+1}[} dM\ h(M) \tag{2.67}$$

The key question is where to place the bins b_i. There are two standard methods.

2.4.2.1 Exponential or logarithmic binning

Apart from the confusion about the name, exponential binning as it will be called throughout this thesis, means that the b_i are chosen according to $b_{i+1} = bb_i$ with a basis b, the value of which mainly depends on the number of data point one is aiming at. Clearly, if the goal is to have about N data points, and the histogram is spread between b_0 and b_N, then $b = (b_N/b_0)^{1/N}$ and $b_i = b_0 b^i$

When plotting the resulting binned data $h'(i)$, one has to decide about the position on the abscissa of data point, or plot a staircase-like function. Most naturally one takes the geometric mean of the two boundaries of the bin, but whatever method is applied it should never play a rôle for any numerical conclusion. If it does, the numerical result is rendered useless, because it depends on something which is mainly a question of taste.

The main advantage of exponential binning is that the data points are equally spaced on the logarithmic scale, which gives nice, smooth curves. If the points are dense enough, a change in the method of determining the position of each bin on the abscissa leads only to a small shift and maybe an invisible distortion of the plot. The main disadvantage is that for $\tau > 1$ the number of "events" per bin decreases with increasing b_i, assuming power law distributed raw data;

$$\sum\!\!\!\!\!\!\!\int_{[b_i, b_{i+1}[} dM\ M^{-\tau} = \frac{(b_0 b^i)^{1-\tau}}{\tau - 1}(1 - b^{1-\tau}) \ . \tag{2.68}$$

The decreasing number of events means that for a finite time simulation the error

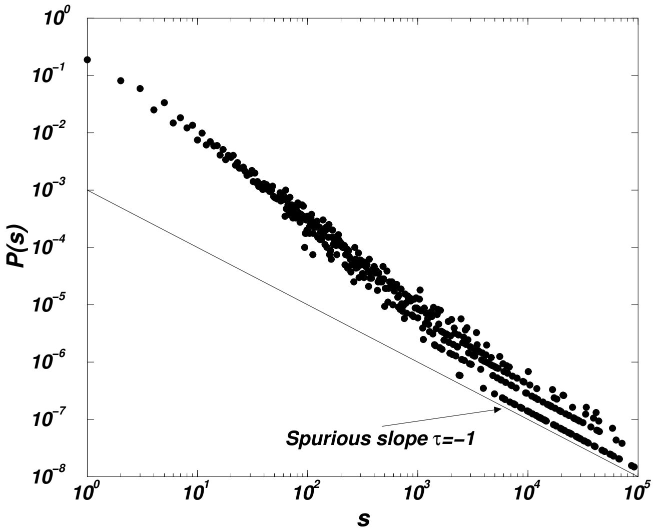


Figure 2.8: If the sample is too sparse in the tail of the distribution, exponential binning produces a spurious slope $\tau = 1$.

bar increases with increasing event size. Below, an alternative binning method is proposed which resolves this problem.

Another, more dangerous problem is that exponential binning produces an spurious exponent $\tau = 1$ if the sample size is small, as shown in Fig. 2.8. Depending on the system under consideration, this can be very deceptive. It is caused by having too many bins and sparse statistics in the tail of the histogram. Most of the bins are then empty, *i.e.* they have 0 density and therefore drop out of a double logarithmic plot. Noting that the width of the bin is $(b-1)b_0b^i = (b-1)b_i$ the density for those with a single event is then $1/((b-1)b_i) \propto b_i^{-1}$ leading to a sequence of points with slope -1. Those bins with 2 events in them lead to a parallel line nearby.

2.4.2.2 Power law binning

To tackle the problem of decreasing subsample sizes per bin, one might impose that b_i is chosen such that

$$\int_{b_i}^{b_{i+1}} dss^{-\tau} = c \tag{2.69}$$

with a constant c to be determined according to the number of bins requested, which, for $\tau > 1$, simply gives

$$b_i = b_0 \left(1 - ic(\tau-1)b_0^{\tau-1}\right)^{1/(1-\tau)} , \tag{2.70}$$

and for $\tau = 1$

$$b_i = b_0 e^{ic} , \tag{2.71}$$

which is exponential binning with basis $b = \exp(c)$.

This method of binning has two major disadvantages:

- The quantity τ to be estimated from the data analysis enters the data analysis itself. That might skew the results.

- For $\tau > 1$ the bin sizes increase even on a logarithmic scale to catch more and more data. For the large event region, that means that the gaps between two data points become so large, that their exact position starts to matter for the interpretation of the data.

2.4.3 Determining the cutoff

Rather than tuning the cutoff $M_0(L)$ until the collapse looks reasonable, it would be nice to have a technique to determine it (or any multiple) directly from the data. As indicated above , this could be the position of the maximum of the "bump" in the PDF towards the end. An easier method sometimes employed [for example (Nordhagen, 2003)] is to define a very small threshold P^* and to define the cutoff M_0^* by

$$\mathcal{P}(M_0^*) = P^* \; . \tag{2.72}$$

The idea is that for large arguments the scaling function drops down extremely fast, so $\mathcal{P}(M)$ gets extremely small very quickly, as illustrated in Fig. 2.9. The problem with this technique is the uncontrolled value P^*; for sufficiently large M_0 the Eq. (2.72) gives a constant M_0^*.

As all data in the plot Fig. 2.9 are normalised, it also illustrates that the small bump caused by the scaling function at the end of the distribution must be large enough to make $\mathcal{P}(s; L)$ larger than $as^{-\tau}\mathcal{G}(0)$ (if $\mathcal{G}(0)$ exists). If the distribution $\mathcal{P}(s; L_1)$ would fully cover the distribution $\mathcal{P}(s; L_2)$, they cannot be both normalised. In fact, all events in the bump "lurking over the power law" are redistributed in the tail of the distribution of a larger system.

2.4.4 Moment analysis

While a data collapse contains a lot of information and is very useful to determine the overall behaviour of a system and even to get a fairly good estimate for its exponents, it is very hard to quantify the quality of a data collapse. On the other hand, moments of the distribution contain much less information, however, they allow not only to determine the value of the exponents, but also to quantify their estimation error.

Also, it is numerically much easier and has much less computational impact to

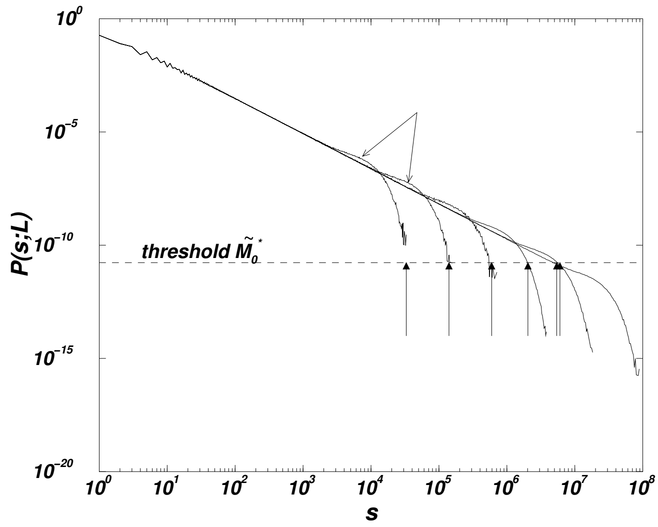


Figure 2.9: Avalanche size distribution $\mathcal{P}(s; L)$ for the Oslo model. System sizes are $L = 160, 320, \ldots, 5120$. The position of the cutoff is marked by the up-pointing arrows. For small system sizes that technique gives reasonable results, but it fails as soon as the true cutoff becomes too large. Also indicated (two connected arrows) is the small area of an PDF for a smaller system size not covered by the larger one. This area contains all the probability which can be redistributed in the PDF of the larger system sizes.

calculate moments [20] rather than to store entire distributions, which could, in fact, be arbitrarily fine-grained.

In principle one can calculate any moment, but in order to calculate τ and D, it is most straight-forward to calculate those integer moments $n > \tau - 1$, for which Eq. (2.50) applies, such that a double logarithmic plot of $\langle M^n \rangle$ versus L at fixed n gives the exponents

$$\gamma_n = D(1 - \tau + n) , \tag{2.73}$$

since

$$\langle M^n \rangle = a(bL^D)^{1+n-\tau} g_n , \tag{2.74}$$

using $M_0 = bL^D$. In order to account for corrections, one can also fit the moments against

$$\langle M^n \rangle = (bL^D)^{1+n-\tau}(g_n + c_1(bL^D)^{-\omega_1}) . \tag{2.75}$$

Fig. 2.10 shows the scaling in L of the second moment of the avalanche size of the Oslo model. To estimate the accuracy, one can plot $\langle s^n \rangle / L^{\gamma_n}$ with the estimated γ_n versus L. The resulting set of γ_n ($n = 1, 2, \ldots, 6$ is typically not too noisy) can then be fitted against $D(1 - \tau + n)$ to give D and τ (in the range $n > 1 - \tau$). Fig. 2.11 shows

[20] This rather basic "technique", actually the most basic technique in traditional numerical estimation of exponents in equilibrium critical phenomena, has apparently been introduced to SOC by DE MENECH, STELLA and TEBALDI (De Menech *et al.*, 1998; Tebaldi *et al.*, 1999).

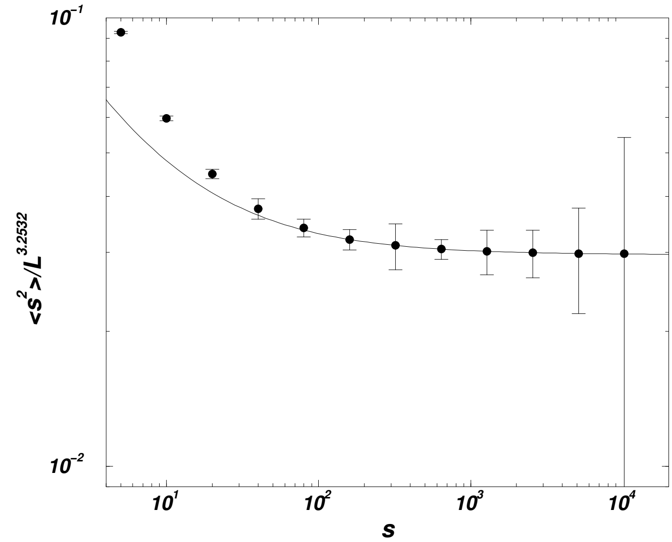


Figure 2.10: The rescaled second moment $\langle s^2 \rangle / L^{\gamma_2}$ of the avalanche size distribution of the original Oslo model as a function of the system size L. The full line is based on a (four-parameter) fit of the data in the range $L \in [40, \ldots, 10240]$ against (2.58) with one correction; $\langle s^2 \rangle = c_0 L_2^{\gamma} + c_1 M_0^{\gamma_2 - \omega_1}$.

γ_n as a function of n. One could calculate γ_n even for non-integer n two get more data points within the accessible range of moments; as shown in (Pastor-Satorras and Vespignani, 2000b) the γ_n pick up "suddenly" at $n = 1 - \tau$. Around this value of n one expects strong corrections.

It is worth noting that plotting $\langle M^n \rangle$ for fixed L versus n cannot produce estimates for the exponents, since this data is affected by g_n to unknown extend; nevertheless, it contains interesting information about the behaviour of the moments ratios g_n.

The fits to calculate the exponents from these plots can be done using standard tools (Press *et al.*, 1992), either directly on the moments or on their logarithms. The estimates for their errors, used in the fits to calculate an error of the fitted parameters, come from the standard estimator[21] for the variance of the estimator for the mean (Brandt, 1998),

$$\sigma^2(\langle M^n \rangle) = \frac{1}{(2\tau_c + 1)N} \frac{N}{N-1} \left(\langle M^{2n} \rangle - \langle M^n \rangle^2 \right) , \tag{2.76}$$

where N is the size of the sample and τ_c the correlation time (Müller-Krumbhaar and Binder, 1973; Madras and Sokal, 1988), which can be calculated in various ways (Anderson, 1964) and might have to account for anti-correlations (see for example the

[21]For this small section on numerics, it does not make much sense to introduce extra symbols to distinguish estimators based on numerical data and real averages calculated from the distribution of the infinite ensemble.

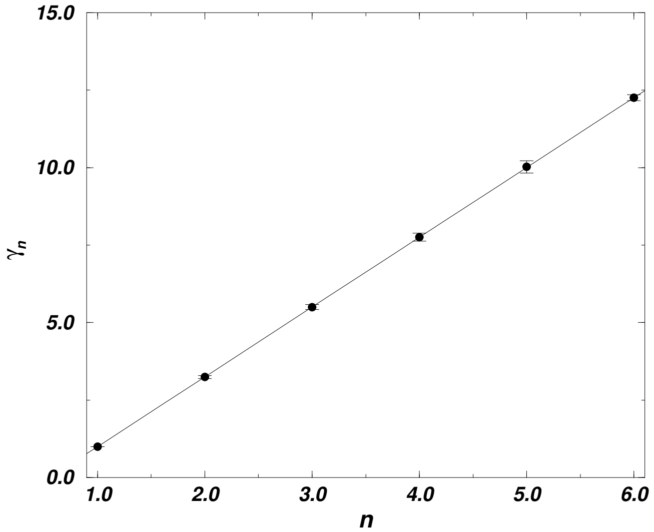


Figure 2.11: The exponent γ_n as a function of n from fits of the moments of the avalanche size distribution of the Oslo model. The straight line is a fit of the γ_n to $D(1-\tau+n)$.

Oslo model, Sec. 5.2.3, page 224). The RHS, $\left\langle(\langle M^n\rangle - M^n)^2\right\rangle$ is always positive which means that the exponent of $\left\langle M^{2n}\right\rangle$, $(1-\tau+2n)$, must be larger than or equal to the exponent of $\langle M^n\rangle^2$, which is $2(1-\tau+n)$ for all $n \geq 0$, if simple scaling and especially (2.50) holds. Thus $\tau \geq 1$, where $\tau = 1$ implies that the two terms on the RHS of (2.76) scale equally. For $\tau > 1$, the second, subtracted term becomes asymptotically irrelevant. Thus, the relative variance scales like

$$\frac{\left\langle(\langle M^n\rangle - M^n)^2\right\rangle}{\langle M^n\rangle^2} \propto M_0^{\tau-1} , \tag{2.77}$$

indicating an massively increasing demand of computing time with system size, because not only does the variance increase, it also become (typically) computationally more and more complex to simulate larger and larger systems. In that sense, the concept of self-averaging (Ferrenberg *et al.*, 1991) is completely changed. It is clear that one needs to simulate larger and larger system sizes in order to avoid problems with corrections, due to the presence of a lower cutoff and the higher order terms discussed in sec. 2.3.2.2. There is no systematic way to discriminate events below the lower cutoff when calculating the moments, which makes the moment analysis different from data collapse, which does this explicitly.

2.4.5 Universal moment ratios

While the scaling of the moments are the handle to identify the exponents, their amplitudes ratios characterise the scaling function. Universal amplitude ratios allow an easily accessible quantitative comparison of scaling functions found in different systems. According to (2.60) it is straight-forward to calculate these ratios. However, contrary to common believe, the best estimator[22] of a ratio of averages is not necessarily the ratio of the best estimator of the averages. While one might ignore these subtleties and simply apply Eq. (2.60) to the estimators of the averages, one often cannot ignore them for the error.

To illustrate the problem, we consider the standard error propagation of the ratio $\langle X\rangle/\langle Y\rangle$ for two observables X and Y. The variance for this ratio is usually estimated as $\sigma^2\langle(X\rangle)/\langle Y\rangle^2 + \sigma^2\langle(Y\rangle)\langle X\rangle^2/\langle Y\rangle^4$. For non-vanishing variances of X and Y this gives a non-vanishing result. However, this is certainly wrong for a strong correlation like $X = Y$. In fact, standard error propagation completely ignores correlations. This is unacceptable, when calculating ratios between averages of different powers of the *same* observable.

The Jackknife (Efron, 1982; Berg, 1992) provides a way out. For $\langle X\rangle/\langle Y\rangle$ is suggests (in leading order of N)

$$\frac{\sigma^2\langle(X\rangle)}{\langle Y\rangle^2} + \frac{\sigma^2\langle(Y\rangle)\langle X\rangle^2}{\langle Y\rangle^4} - 2\frac{\langle X\rangle\langle Y\rangle}{\langle Y\rangle^4}\Big(\langle XY\rangle - \langle X\rangle\langle Y\rangle\Big)\,, \tag{2.78}$$

with an additional, third term accounting for correlations. Indeed, this term would cancel the other terms if $X \propto Y$.

Using the same technique for the universal amplitude ratios, one finds for the variance of the estimator of g_n (*i.e.* the square root of it is the error):

$$\begin{aligned}\sigma^2(g_n) = &\frac{1}{N}\frac{\langle M^n\rangle^2\langle M\rangle^{(2n-4)}}{\langle M^2\rangle^{(2n-2)}}\\ &\times\left\{\frac{\left\langle M^{2n}\right\rangle - \langle M^n\rangle^2}{\langle M^n\rangle^2} + (2n-4)\frac{\left\langle M^{(n+1)}\right\rangle - \langle M^n\rangle\langle M\rangle}{\langle M^n\rangle\langle M\rangle}\right.\\ &+ (n-2)^2\frac{\left\langle M^2\right\rangle - \langle M\rangle^2}{\langle M\rangle^2} + (n-1)^2\frac{\left\langle M^4\right\rangle - \left\langle M^2\right\rangle^2}{\langle M^2\rangle^2}\\ &- 2(n-1)\frac{\left\langle M^{n+2}\right\rangle - \langle M^n\rangle\left\langle M^2\right\rangle}{\langle M^n\rangle\langle M^2\rangle}\\ &\left.-2(n-1)(n-2)\frac{\left\langle M^3\right\rangle - \left\langle M^2\right\rangle\langle M\rangle}{\langle M^2\rangle\langle M\rangle}\right\} + \mathcal{O}(N^{-2})\end{aligned} \tag{2.79}$$

[22] A "good estimator" is not biased, *i.e.* the expectation value of the estimator is identical to the expectation value to be estimated, and it is consistent, *i.e.* it converges to the expectation value for divergent sample size. (Brandt, 1998)

2.5 Summary

Power laws and scaling are the fingerprints of criticality. The SOC literature displays widespread inconsistencies regarding appropriate methods of analysis and the relation of SOC to standard results in statistical physics. The individual sections in this chapter can be summarised as follows:

- After an introduction, some general relations between moments, distributions and generating functions are established in Sec. 2.2. It is found that the observables used in classical critical phenomena and SOC are closely related.
- In Sec. 2.3 simple scaling is discussed in detail. Most importantly, the exponents τ and D are introduced and related to the behaviour of the singular part of the free energy in the case of classical critical phenomena. Moreover, corrections are identified as well as the lower cutoff, which is often neglected in the literature.
- Simple scaling is the central theme of SOC, its notion lays the foundation for an investigation of numerical techniques in SOC, see Sec. 2.4. They boil down to two main tasks when characterising a critical model: Determination of the moments τ and D via a moment analysis, and determination of the universal moment ratios to characterise the universal scaling function. For all these properties an error bar can be determined as well.

Chapter 3

The Rôle of Conservation

In the following, some key ideas of HWA *and* KARDAR *as well as* GRINSTEIN *et al.* *are discussed briefly. Then, some objections against the criticality of the OFC model are presented; they represent an unfinished, yet quite interesting project. In the following section, a solvable mean field model is presented, the results of which have been published in (Pruessner and Jensen, 2002b). The last section of this chapter contains a random walker approach to the finite size scaling problems found in the mean field model. Again, this is unfinished work, however, it shows an elegant way of tackling random walker problems, based on orthogonal polynomials.*

Very early on, conservation has been recognised as an important feature in systems (supposedly) displaying SOC behaviour, most famously in the articles by HWA *and* KARDAR *(Hwa and Kardar, 1989) and* GRINSTEIN, LEE *and* SACHDEV *(Grinstein* et al.*, 1990). Naively, one might think that a lack of conservation necessarily leads to a characteristic scale in the distribution of event sizes. Therefore, the development of the* OLAMI-FEDER-CHRISTENSEN *(OFC) model (Olami* et al.*, 1992) and the* DROSSEL*-Schwabl Forest-Fire-Model (DS-FFM) (Drossel and Schwabl, 1992) apparently displaying SOC in spite of dissipation, was and still is very surprising.*

3.1 Conservation: A LANGEVIN Approach

It is fair to say that HWA and KARDAR made the use of LANGEVIN equations (LE's) popular in SOC (Hwa and Kardar, 1989). They consider the spatio-temporal evolution of the height profile $h(\mathbf{x}, t)$ of a sandpile over a substrate $\mathbf{x}$ in time. Due to the presence of a particular "transport direction" (grains toppling downhill), the resulting LE is anisotropic. The most general form of it is

$$\frac{\partial}{\partial t} h(\mathbf{x}, t) = \mathcal{F}(h(\mathbf{x}, t)) + \eta(\mathbf{x}, t) \ , \tag{3.1}$$

where $\mathcal{F}$ is some functional describing the deterministic dynamics and $\eta(\mathbf{x}, t)$ is a noise, obeying

$$\begin{aligned} \langle \eta(\mathbf{x}, t) \rangle &= 0 & \text{(3.2a)} \\ \big\langle \eta(\mathbf{x}, t)\eta(\mathbf{x}', t') \big\rangle &= \Gamma^2 \delta^d(\mathbf{x} - \mathbf{x}')\delta(t - t') & \text{(3.2b)} \end{aligned}$$

in d spatial dimensions. The simplest equation of motion including the leading non-linearity is

$$\mathcal{F}(h(\mathbf{x}, t)) = \nu_{\parallel}\partial_{\parallel}^2 h(\mathbf{x}, t) + \nu_{\perp}\nabla_{\perp}^2 h(\mathbf{x}, t) - \frac{\lambda}{2}\partial_{\parallel} h(\mathbf{x}, t) + \eta(\mathbf{x}, t) \tag{3.3}$$

with the first term being the Laplacian in the flow-direction, the second covering the other directions and the third the leading non-linearity. The two constants $\nu_{\parallel}$ and $\nu_{\perp}$ are the diffusion constants parallel to the transport direction and perpendicular to it. The absence of certain other terms in (3.3) is a crucial point and is justified below. It is important to note that the splitting of the Laplacian cannot be undone by a spatial transformation, without affecting the noise; the form of the noise correlator (3.2b) would change if the Laplacian was changed into an isotropic ∇^2.

HWA and KARDAR refer to conservation on average, when they say that

> [t]*he conservative nature of the* deterministic *part* [,$\mathcal{F}(h(\mathbf{x}, t))$,] *of the dynamics rules out terms such as* $-h/\tau$ *in the equation of motion. (Such a term would introduce a characteristic time* τ *and characteristic length scales in the problem that would remove scale invariance.)*

The latter is in line with the general confusion that a characteristic length scale other than the system size automatically destroys scale invariance. This is not true; on the contrary, *only* the presence of an additional scale actually makes possible anomalous scaling, *i.e.* scaling behaviour deviating from dimensional analysis. This point is further illustrated in Chapter 9.

Nevertheless, the crucial point is that conservation rules out certain terms and apparently their absence allows the solution of (3.1) to be generically scale-invariant. This is further substantiated by a dynamic RG analysis in (Hwa and Kardar, 1989). The exponents from these calculations agreed fairly well with the numerical results by BAK, TANG and WIESENFELD. Almost 15 years later, it is highly questionable whether the BTW model develops into a scale-invariant state at all (Dorn *et al.*, 2001).

GRINSTEIN *et al.* use the approach introduced by HWA and KARDAR, to generalise their results. They analyse conservation in much finer detail and distinguish between a) conservation on average, where $\mathcal{F}$ is conservative, and the noise fluctuates according to (3.2), and b) strict conservation, where the noise can be written as a gradient of another noise, so that

$$\left\langle \eta(\mathbf{x},t)\eta(\mathbf{x}',t')\right\rangle = \sum_i^d (\Gamma_i^2 \partial_i^2)\delta^d(\mathbf{x}-\mathbf{x}')\delta(t-t') \tag{3.4}$$

where the sum, running over all spatial dimensions, allows the noise to have different amplitudes in all directions. At the same time, the deterministic part of the LE is rewritten as

$$\mathcal{F}(h(\mathbf{x},t)) = \nu\nabla^2 h(\mathbf{x},t) + \eta(\mathbf{x},t) \ , \tag{3.5}$$

now prohibiting a spatial transformation to make the noise isotropic. In case of a noise correlator (3.4) the LE (3.1) can now be written as a gradient, so the equation is strictly conservative provided that the integral over the boundaries vanishes.

Essentially based on power-counting arguments, it is then possible to show that correlations decay exponentially in space *only* if $\Gamma_1^2 = \Gamma_2^2 = \cdots = \Gamma_d^2$. In that case, a Hamiltonian can be found for the problem and detailed balance holds. In turn that means that a very small anisotropy in the noise is enough to make produce generic scale invariance. In $d = 1$, there is in fact "no SOC" in the strictly conservative case.

A noise being conservative only on average,

$$\left\langle \eta(\mathbf{x},t)\eta(\mathbf{x}',t')\right\rangle = \Gamma^2\delta^d(\mathbf{x}-\mathbf{x}')\delta(t-t') \tag{3.6}$$

which breaks detailed balance and therefore renders the problem non-equilibrium again. In that case, SOC is to be expected in all dimensions.

In some models, it is not obvious how to distinguish strict conservation and conservation on average. For example, in the Oslo model, slope-grains can get "buried". They do not disappear from the system and are bound to reappear sooner or later. Thus, they are conserved; however, they get trapped randomly, so that one could also argue that the trapping represents a noise term, which renders the model only conservative on average.

Conservation can lead to scaling relations[1]; in the Oslo model, the average conservation of grains leads immediately to the scaling behaviour of the average avalanche size and therefore to a scaling law, relating τ and D. The argument can be extended to an exact result for the average avalanche size, using a random walker approach to the behaviour of individual grains. This, however, does not contain any information about the *collective* behaviour of these grains.

[1] In general, arguments about particle creation and annihilation, especially non-conservation, can lead to rigorous results for the first moment of the avalanche size distribution.

3.1.1 Discussion

The authors cited above have introduced a LANGEVIN approach to SOC [2]. Probably the most remarkable outcome of these papers is that scale invariance is the rule and exponential correlations the exception, mostly found in equilibrium models. But, do the two papers presented above explain SOC?

They explain generic scale invariance found in an enormous class of problems, which can be described by the LE's introduced above. There are, however, a number of problems:

- Anisotropy: *Prima facie*, it seems that the anisotropy is a natural ingredient in these sandpile models (for example the Oslo model chapter 5), since grains topple downhill. However, analysing the toppling rules, one finds that the anisotropy comes purely from the definition of slope, see Sec. 5.1.3.1, page 211. Expressed in different variables, the anisotropy completely disappears. All what is left are the boundary conditions, which might be different at different boundaries. However, the rôle of boundary conditions is not discussed in the papers or even explicitly neglected.[3]

- Thermal noise: One can regard a LANGEVIN equation as a transformation prescription for the noise term; the solution of the equation is a functional of the noise and for every sampling of the noise, there is a unique solution corresponding to it. Thus, one might not be too surprised, if a scale-invariant noise, $\eta(b\mathbf{x}, b^{z'}t) = b^{\alpha'}\eta(\mathbf{x}, t)$ with rescaling factor b and exponents z' and α' transforms into something which is scale-invariant again.

- Models: Even though viewed in this light the result is not very surprising, it is still valid. The remaining question is then, whether there are any relevant models, obeying the LE's discussed. From a comparison with the numerical values of the exponents, this does not seem to be the case; at least the LE's are not the generic class for sandpile models. It seems that one of the crucial ingredients of SOC, thresholds (see Sec. 1.1, page page 36), is not properly represented. Since it is nevertheless possible to write down a (noisy) equation of motion for these models, one might argue that such thresholds need to be expressed as multiplicative and/or quenched noise (see Chapter 6).

What is not discussed in these articles, is the case of a non-conservation deterministic part of the equation, which also includes a noise with drift that can be absorbed in $\mathcal{F}$, violating $\langle\eta\rangle = 0$. Most remarkably, there are models apparently ex-

[2]As was done in parallel by GARRIDO, LEBOWITZ, MAES and SPOHN (Garrido *et al.*, 1990) and others
[3]Most remarkably, the possible relevance of a drift term (Chapter 9) is ignored in (Grinstein *et al.*, 1990).

hibiting SOC, while belonging to this class, for example the OFC model (see next section) or the forest fire model (see Chapter 4). It was conjectured that any degree of non-conservation can always by compensated by an external drive (Dickman *et al.*, 2000), which will be discussed in sec. 3.3. However, it is not always obvious whether non-conservation apparent in one description of a model, is visible in other descriptions as well — it might well be possible that a model looks completely non-conservative in one quantity, but is conservative in the quantity governed by the LE. The anisotropic Oslo model, discussed in Sec. 7.1, represents such a case: Expressed in terms of height, it creates grains, while being completely conservative (apart from boundaries) when expressed in terms of slopes. Again, height and slope are related by an arbitrary choice of the definition of "downhill".

3.2 The OFC Model

The OFC model has already been introduced in Sec. 1.3.2, page page 44. In this section, an argument about why one cannot expect scale invariance in the random neighbour version of it was already outlined. The following argument for the nearest neighbour version is neither rigorous nor finished, but it provides a very strong necessary condition for a divergent first moment of the avalanche size distribution.[4] It therefore offers an alternative way of settling the issue whether the OFC model is critical in the non-conservative regime or not. Open issue

We recapitulate the most important features of the OFC model on a square lattice in this context. Every site i has a height $z_i \in [0, z_c]$. All sites are driven homogeneously by a small amount ϵ such that $z_i \to z_i + \epsilon$ until a site gets above a threshold z_c (Grassberger, 1994). Every site i exceeding the threshold z_c topples by redistributing z_i according to the rule $z_j \to z_j + \alpha z_i$ where j are all nearest neighbours, and $z_i \to 0$. The fact that a site redistributes everything to the neighbours makes the model non-Abelian. The model is closely related to the MANNA model (Manna, 1991), which is discrete and has randomised topplings, and even closer to the ZHANG model (Zhang, 1989). The ZHANG model is driven randomly by an amount δ; the OFC model is recovered in the ZHANG model in the limit $\delta \to 0$, which mimics homogenous drive.

In the stationary state, one expects a height distribution $\mathcal{P}(z)$ as depicted in Fig. 3.1. It shows the average fraction of sites in the system, which have a height in the interval $[z, z + dz]$. The average is taken *after* each avalanche. The total inflow

[4]This theme repeats again and again: The behaviour of the first moment can be related to the behaviour of the individual degrees of freedom. Higher moments, however, relate to their *collective* behaviour, which requires the notion of an avalanche, as a single, coherent event, which connects them.

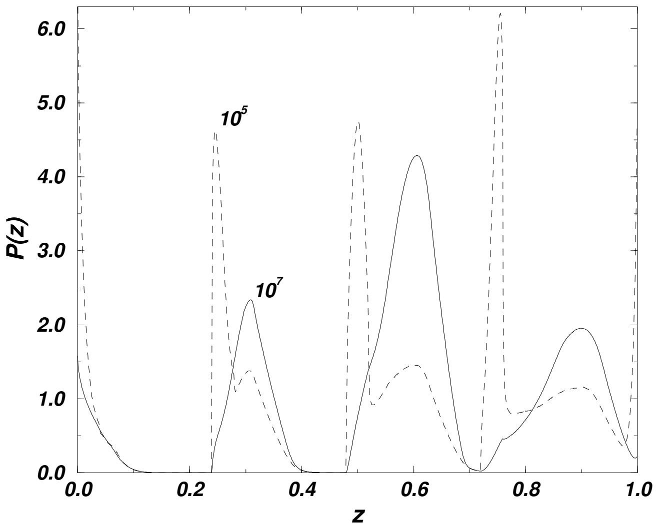


Figure 3.1: Height distribution in the two dimensional OFC model on a square lattice for $L = 256$, open boundaries and $\alpha = 0.24$ after 10×10^6 avalanches, averaged over 10^6 avalanches. The four bumps correspond to four charges of a site. There is a general problem of slow convergence in the OFC model (not discussed here), which is illustrated by the distribution after 10^5 avalanches (averaged over the same number) shown as a dashed line.

due to the external drive, $\Delta E(t)$, is given by

$$\Delta E(t) = N(z_c - z_{\text{max}}(t)) \tag{3.7}$$

with N being the total number of sites and z_{max} the (currently) largest height in the system. Of course, this value fluctuates over time, which is why the quantities above are shown as a function of t. The time in this context is best taken to be the macroscopic time, *i.e.* an integer numbering sequentially the avalanches.

In the non-conservative regime, the average dissipation $\Delta E'$, *i.e.* the average amount dissipated by the system per avalanche, is bounded by

$$\Delta E' \geq (1 - q\alpha)\langle s \rangle \tag{3.8}$$

where $\langle s \rangle$ is the average avalanche size measured as the average number topplings, q is the coordination number and $\alpha \leq 1/q$ is the "degree of conservation". The relation sign in Eq. (3.8) becomes an equal sign, if the boundaries are non-dissipative. In the stationary state $\Delta E' = \langle \Delta E \rangle$, so that

$$\langle s \rangle \leq N \frac{z_c - \langle z_{\text{max}} \rangle}{1 - q\alpha} \tag{3.9}$$

It seems straight forward to estimate $\langle z_{\text{max}} \rangle$ from $\mathcal{P}(z)$: If it was only one site distributed according to it, then the probability that this site had a value lower than z' is $\int_0^{z'} dz \mathcal{P}(z)$. For N *independent* sites, the probability that all z_i are below z' is $\left(\int_0^{z'} dz \mathcal{P}(z)\right)^N$, leading to

$$\mathcal{P}_{\max}(z_{\max}) = \left.\frac{d}{dz'}\right|_{z'=z_{\max}} \left(\int_0^{z'} dz \mathcal{P}(z)\right)^N = N\mathcal{P}(z_{\max}) \left(\int_0^{z_{\max}} dz \mathcal{P}(z)\right)^{N-1} \quad (3.10)$$

where $\mathcal{P}_{\max}(z_{\max})$ denotes the PDF of the maximum z_i. Thus, the average maximum is in the limit of large N

$$\langle z_{\max} \rangle \to z_c - \frac{1}{\mathcal{P}(z_c)(N+1)} \quad (3.11)$$

where $\mathcal{P}(z)$ has been expanded around $z = z_c$. Even though Fig. 3.1 does not seem to indicate a vanishing $\mathcal{P}(z_{\max})$, one can find similar results even in that case. The crucial point is that the population density around z_c increases linearly in N, so that the distance to z_c of the largest value in this population vanishes like $1/N$. Thus, in case of uncorrelated heights, there is a constant c such that

$$\langle s \rangle \leq c \frac{N}{(N+1)(1-q\alpha)} \quad (3.12)$$

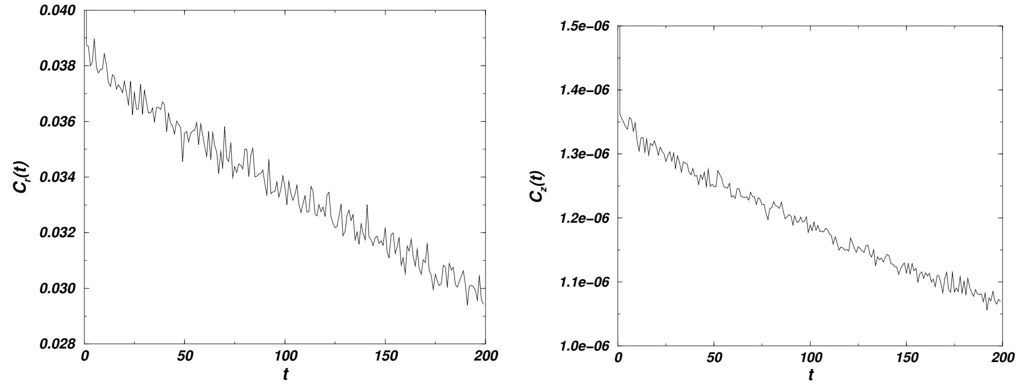
with $c = 1/\mathcal{P}(z_c)$ if the density around z_c is finite. For $\alpha < 1/q$ the RHS of Eq. (3.12) remains finite for all N, so that $\langle s \rangle$ does not diverge. However, that contradicts numerical findings (namely $\tau < 2$) (Lise and Paczuski, 2001b,a; Boulter and Miller, 2003), even though the latter remain inconclusive.

The most serious assumption made above was that the heights are uncorrelated. One can see that immediately by considering the extreme case, namely that all sites have exactly the same value after an avalanche. This value would be randomly chosen from the distribution $\mathcal{P}(z)$, so that this ensemble of δ-peaked distribution gives on average $\mathcal{P}(z)$. In that case, the average maximum does not get closer to z_c as the system size is increased.

A numerical study of the average position of the maximum is exactly identical to a numerical study of the average avalanche size. Plenty of these studies can be found in the literature and they all remain inconclusive. So, by focusing on another approach, one might find stronger evidence in favour or against a divergent first moment.

The most obvious question to ask is whether there are correlations between the consecutive $z_{\max}$, in space or in amount. Fig. 3.2(a) shows the correlation function

$$C_r(t) = \frac{\langle \mathbf{r}_{\text{in}}(t')\mathbf{r}_{\text{in}}(t'+t)\rangle - \langle \mathbf{r}_{\text{in}}(t')\rangle\langle \mathbf{r}_{\text{in}}(t'+t)\rangle}{\langle \mathbf{r}_{\text{in}}(t')^2\rangle - \langle \mathbf{r}_{\text{in}}(t')\rangle^2}, \quad (3.13)$$



(a) Correlation function $C_r(t)$. **(b)** Correlation function $C_z(t)$.

Figure 3.2: Correlation function $C_r(t)$ of the position of the seed of the avalanches and correlation function $C_z(t)$ of the value of the maximum site at time t in the two-dimensional OFC model on a square lattice for $L = 256$, open boundaries and $\alpha = 0.24$, after 10^6 avalanches, averaged over 3×10^6 avalanches.

where $\mathbf{r}_{\text{in}}(t)$ is the location of the the site with the largest height after the tth avalanche. The average $\langle\rangle$ runs over t' and is — supposedly — identical to the ensemble average. The plot indicates that there is almost no correlation between the location of consecutive "seeds" of avalanches. However, the correlations are markedly higher or at least more structured for $\alpha = 0.24$ than for $\alpha = 0.25$ (not shown). Similarly for the correlation function of $z_{\max}(t)$, $C_z(t)$, shown in Fig. 3.2(b).

In order to identify which sites actually trigger the avalanches, Fig. 3.3 shows the probability for each site to be the seed of an avalanche[5]. As already suggested by MIDDLETON and TANG (Middleton and Tang, 1995), the boundaries are typically triggering an avalanche. That seems to be surprising, because the boundaries are most dissipative, so one should expect boundary sites typically to be at very low height. However, most of the height driven into the bulk leaves the system *via* the boundaries. Moreover, having less neighbours means also less charges; Fig. 3.4 supports that view. This figure shows the average time between two hits.

One major problem in the OFC model is its enormous, practically unknown equilibration time. This problem has rendered a comparison of (3.10) to the numerical result very problematical. It is illustrated in terms of the PDF $\mathcal{P}(z_{\max})$ in Fig. 3.5: The PDF changes drastically with the equilibration time. The numerical problems with the OFC model have been discussed by DROSSEL (Drossel, 2002), claiming that one essential mechanism in the OFC model is the limited numerical accuracy when simulated. This is consistent with the findings in the next section, where a model is

[5]Fig. 3.3 remains essentially unchanged if avalanches of size 1, *i.e.* those which do not actually spread, are omitted.

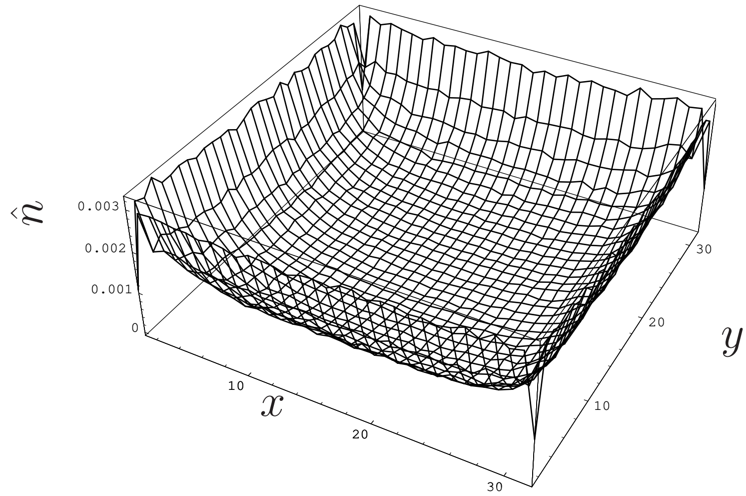


Figure 3.3: Average number $\hat{n}$ of times a site was the seed of an avalanche in the OFC model, *i.e.* probability to be seed per site. Simulation parameters: $L = 32$, open boundaries, $\alpha = 0.24$, 2×10^5 avalanches for transient and 5×10^5 for statistics.

presented, which becomes particularly difficult to handle numerically, whenever α is of the form $1/n$, $n \in \mathbb{N}$, see Sec. 3.3.4.

3.3 A Solvable Random Neighbour Model

As an example of non-conservative SOC, we present the first solvable non-conservative sandpile-like *critical* model of Self-Organised Criticality (SOC). This will help to substantiate on mean field level the suggestion by VESPIGNANI and ZAPPERI (Vespignani and Zapperi, 1998) that a lack of conservation in the microscopic dynamics of an SOC-model can be compensated by introducing an external drive and thereby re-establishing criticality. The model shown is critical for all values of the conservation parameter. The analytical derivation follows the lines of BRÖKER and GRASSBERGER (Bröker and Grassberger, 1997) and is supported by numerical simulation. In the limit of vanishing conservation the Random Neighbour DROSSEL SCHWABL Forest Fire model (R-DS-FFM) is recovered.

3.3.1 Introduction

The rôle of conservation in SOC-models is an old issue (Grinstein *et al.*, 1990; Middleton and Tang, 1995; Jensen, 1998) and is still unsettled. The number of non-conservative models which are definitely critical is, however, strikingly small. The Random Neighbour DROSSEL SCHWABL Forest Fire model (R-DS-FFM) (Christensen *et al.*, 1993) is one of them, while the Random Neighbour OLAMI-Feder-Christensen

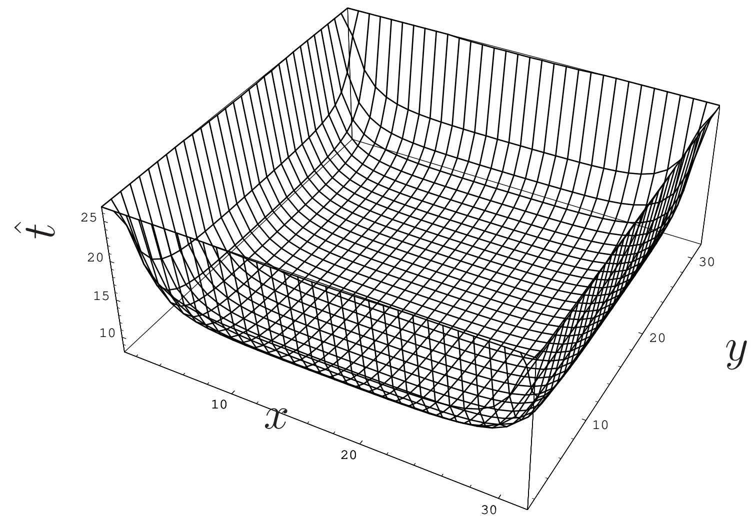


Figure 3.4: Average time $\hat{t}$ (measured in number of avalanches) between two charges of a site. Simulation parameters: $L = 32$, open boundaries, $\alpha = 0.24$, 2×10^5 avalanches for transient and 5×10^5 for statistics.

model (R-OFC) has been shown not to be critical in the non-conservative regime (Bröker and Grassberger, 1997; Chabanol and Hakim, 1997). The nearest neighbour OFC model is widely accepted to be critical in the conservative limit, but whether this model is critical in the non-conservative regime is still debated (Middleton and Tang, 1995; Lise and Paczuski, 2001b) — a way to tackle the problem has been proposed above.

In (Vespignani and Zapperi, 1998; Dickman *et al.*, 2000) it has been suggested that non-conservation in the microscopic dynamics can be compensated by an external drive in order to re-establish criticality. Applying this concept directly to a model known to be non-critical in its original definition provides the ideal basis to identify the effect of such an external drive. In this letter such a model is defined and solved semi-analytically. The results are compared to simulations and the (trivial) critical exponents are extracted. Several limits are discussed.

3.3.2 The model

The model, which is derived from the DS-FFM (Drossel and Schwabl, 1992) and the ZHANG model (Zhang, 1989), has three main parameters: N is the total number of sites, which diverges in the thermodynamic limit. The number of randomly chosen "neighbours" is given by n, where $n = 4$ in all examples, corresponding to a two-dimensional square lattice. The conservation parameter is α. The degree of non-conservation is then $1 - n\alpha$, as it is shown below. Each site $i \in \{1, 2, \cdots, N\}$ has associated a value z_i for its "energy". Sites with $0 \leq z_i < 1-\alpha$ are said to be "stable",

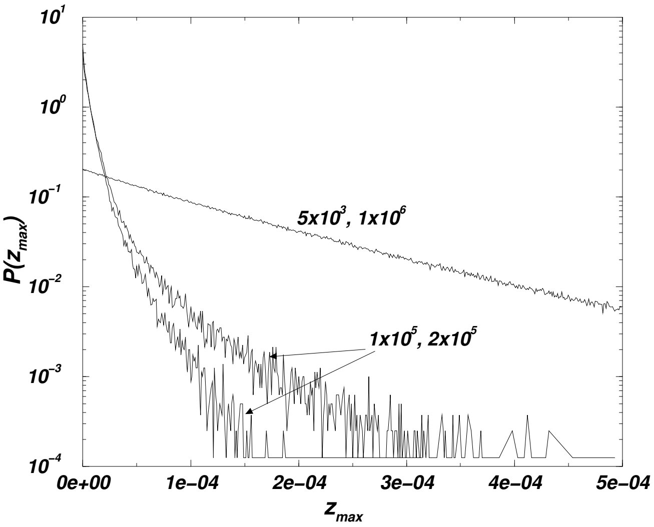


Figure 3.5: The distribution of the maximum $\mathcal{P}(z_{\max})$ for two different equilibration times and two different initial seeds of the random number generator. System size $L = 200$, open boundaries, $\alpha = 0.24$. The number of avalanches rejected as transient and accepted as statistics are given in the form "transient, statistics".

sites with $1-\alpha \leq z_i < 1$ are called "susceptible" and sites with $1 \leq z_i$ are "active". Negative energies are not allowed. The probability density function (PDF) for the variable z_i is $P(z)$ with $z \in [0, 1[$ and is defined only when no sites are active. $P(z)$ contains most of the stationary properties of the model.

The dynamics of the model are defined as follows: After an initial choice of z_i $(i = 1, 2, ..., N)$ from a uniform distribution in the interval $[0, 1[$, the model is updated by repeatedly (i) "driving", (ii) "triggering" and (iii) "relaxing" the system. During the drive (i), $i = 1, ..., \theta^{-1}$ sites are chosen randomly $[\theta^{-1} = p/f$ in the notation of (Drossel and Schwabl, 1992)], one after the other, and their energies z_i are set to $1-\alpha$, if the site is stable, otherwise z_i remains unchanged. Subsequently one random site j is chosen and if it is susceptible, the system is triggered (ii) by setting the energy of the chosen site to 1, *i.e.* making it active (initial seed). Otherwise z_j remains unchanged and the model is driven again by repeating (i). As long as N and θ^{-1} are finite, the system will escape from the driving loop sooner or later. In the thermodynamic limit this is ensured by a non vanishing density of susceptible sites.

During the relaxation (iii) the energy of each active site i is redistributed according to the conservation parameter α to n randomly chosen sites j and the energy of z_i is then set to 0:

$$z_j \to z_j + \alpha z_i \qquad z_i \to 0 \tag{3.14}$$

Each visit or "toppling" (3.14) defines a microscopic time step and dissipates exactly $(1-n\alpha)z_i$ energy units. The sites j are chosen randomly one after the other and are not necessarily different. In the thermodynamic limit the probability of choosing a target site which is already active or was already charged during the same avalanche, vanishes and therefore the order of these visits is irrelevant. In this very restricted sense the model might be considered as "Abelian". In contrast, sites in finite systems have always a finite probability to get charged more than once. Nevertheless, this probability decreases rapidly with increasing system size.

3.3.3 Calculations

The number of active sites relaxed by (3.14) defines the avalanche size s, which is always positive due to the initial seed. In the stationary state the avalanches must dissipate, on average, the same amount of energy as is supplied by the external drive and the initial seed. The average dissipation depends on the avalanche size weighted average energy of active sites $\overline{z_{\text{act}}}$, which is equivalent to the average energy of active sites per toppling. Therefore

$$(1 - n\alpha)\overline{z_{\text{act}}}\langle s\rangle = \theta^{-1}\frac{\overline{p_{\text{st}}}}{\overline{p_{\text{c}}}}(1 - \alpha - \overline{z_{\text{st}}}) + (1 - \overline{z_{\text{c}}}) \tag{3.15}$$

must hold exactly in the stationary state even for finite systems and does not introduce any approximation. Here $\langle s\rangle$ is the average avalanche size, $\overline{p_{\text{st}}}$ ($\overline{p_{\text{c}}}$) is the density of stable (susceptible) sites (the drive stage is, on average, repeated $1/\overline{p_{\text{c}}}$ times), $\overline{z_{\text{st}}}$ and $\overline{z_{\text{c}}}$ are the average energy of stable and susceptible sites respectively. As in (Bröker and Grassberger, 1997) the only crucial assumption is that $\langle s\rangle/N$ as well as θ^{-1}/N vanishes in the thermodynamic limit, which turns out to be entirely consistent with the results. This assumption allows us, for example, to assume the distribution $P(z)$ to be essentially unaffected by external drive or relaxation for sufficiently large systems.

From (3.15) it is clear that in general $\langle s\rangle$ diverges for diverging θ^{-1} or vanishing dissipation rate $1-n\alpha$. From the microscopic dynamics it is clear that there is always a non-vanishing fraction of sites with $z = 0$, therefore $(1 - \alpha) - \overline{z_{\text{st}}}$ is finite and a divergence of θ^{-1} entails a divergence of $\langle s\rangle$, *which is a signature of criticality*.

In the following outline of the actual calculation, which is adapted from (Bröker and Grassberger, 1997), the PDF's of the model are derived.

After an avalanche, each site belongs to one of $m + 2$ classes, where $m = \lfloor 1/\alpha \rfloor$. The index $k = 0, 1, \cdots, m$ of the class indicates the number of charges received from other toppling sites since their last toppling, while $k = m + 1$ indicates the class of sites, whose energy has been set by external drive. A site charged more than m times must be active. For each of these classes a conditional distribution function $Q_k(z)$ is introduced, describing the distribution of energy among non-active sites, which have been charged $k = 0, ..., m$ times or externally driven, $k = m + 1$. The distribution of sites which have not been charged since their last toppling, $Q_0(z)$, is a delta peak at $z = 0$. For convenience the normalisation of $Q_0(z)$ is chosen to be unity and all other distribution functions are normalised relative to class 0. The distribution of sites which have not changed after once being driven externally, $Q_{m+1}(z)$, is obviously a weighted delta peak at $1 - \alpha$. If the fraction of these sites is g ("accumulated susceptible sites"), $P(z)$ can be written as

$$P(z) = \mathcal{N} \sum_{k=0}^{m+1} (1-g) Q_k(z) \tag{3.16}$$

where $\mathcal{N}(1-g)Q_{m+1}(z) = g\delta((1-\alpha)-z)$ and $\mathcal{N} < 1$ is an appropriately chosen normalisation. The upper bound for the energy of an active site is the geometric sum $1 + \alpha(1 + \alpha(\cdots)) = 1/(1-\alpha)$, neglecting double charges. Therefore, the support of the distribution function of active sites $C(z)$ is $[1, 1/(1-\alpha)[$. If this distribution is normalised, the expected increase per avalanche in the class $k > 0$ is given, in the thermodynamic limit (where multiple toppling can be neglected), by the convolution

$$n\langle s\rangle \int_1^{1/(1-\alpha)} dz' C(z') Q_{k-1}(z - \alpha z') \ , \tag{3.17}$$

where the factor $n\langle s\rangle$ takes into account the expected total number of charges. There are three different ways in which the classes $k < m+1$ may be decreased:
1) By charges, $Q_k(z) n\langle s\rangle$
2) By external drive, $Q_k(z)\theta^<((1-\alpha)-z)\theta^{-1}\overline{p_c}^{-1}$, where $\theta^<$ is the HEAVISIDE step function with $\theta^<(0) = 0$.
3) By initial seed, $Q_k(z)\theta^>(z-(1-\alpha))/\overline{p_c}$, where $\theta^>(0) = 1$ correspondingly.
Adding these contributions together and assuming stationarity leads to m equations for Q_k, $k = 1, \cdots, m$:

$$Q_k(z) l(z) = \int_1^{1/(1-\alpha)} dz' C(z') Q_{k-1}(z - \alpha z') \ , \tag{3.18}$$

where

$$l(z) = 1 + \frac{\theta^{-1}}{\overline{p_c} n\langle s\rangle} \theta^<((1-\alpha)-z) + \frac{1}{\overline{p_c} n\langle s\rangle} \theta^>(z-(1-\alpha)) \tag{3.19}$$

has been used. For diverging θ^{-1}, the last term in (3.19) becomes irrelevant and the RHS of (3.15) is dominated by the first term, meaning that the initial seed becomes irrelevant compared to the external drive. It is reasonable to restrict the range of α so that single charges cannot activate a site, $\alpha/(1-\alpha) < 1 \Leftrightarrow \alpha < 1/2$ (due to $n\alpha < 1$, this is a restriction only for $n = 1$). For $k = 1$ (3.18) can be written as

$$C\left(\frac{z}{\alpha}\right) = \alpha Q_1(z) l(z) \ , \tag{3.20}$$

due to the particularly simple form of $Q_0(z)$.

Since $Q_0(z) = \delta(z)$, by definition, and $\mathcal{N}(1-g)Q_{m+1}(z) = g\delta((1-\alpha)-z)$ as mentioned above [see (3.16)], one further equation is necessary in order to find $m+3$ distributions Q_k, $k = 0, \cdots, m+1$ and $C(z)$. This equation concerns the construction of the distribution of active sites $C(z)$. Since active sites are created due to charging or as the initial seed the average distribution of the number of those sites per avalanche is given by

$$\langle s \rangle C(z) = n \langle s \rangle \int_1^{1/(1-\alpha)} dz' C(z') P(z - \alpha z') + \delta(z-1) \tag{3.21}$$

where the δ-function represents the initial seed.

Although it is *a priori* unknown whether there exists a stable solution, or whether it is unique, the set of equations given above is enough to start an iteration procedure in order to find a solution. The scalar parameters required are $\langle s \rangle$ from (3.15), $\overline{p_{\mathrm{st}}}$, $\overline{p_{\mathrm{c}}}$, $\overline{z_{\mathrm{st}}}$, $\overline{z_{\mathrm{c}}}$, which are easily derived from (3.16) and $\overline{z_{\mathrm{act}}}$, the first moment of $C(z)$. While n and θ^{-1} parameterise the problem, g remains the only unknown quantity, which is found to be

$$g = \frac{\overline{p_{\mathrm{st}}}\theta^{-1}}{n\langle s \rangle \overline{p_{\mathrm{c}}} + 1} \tag{3.22}$$

by comparing the in- and outflow of class $m+1$, the externally driven sites, per avalanche.

3.3.3.1 Invariant I

As a test for the derivations above, one can derive invariants, *i.e.* quantities whose value is independent from the specific form of the solution. This is of particular interest if these invariants are easy to derive from the original model but do not enter directly the solution.

The first such object is the normalisation of $C(z)$. Since $C(z)$ is normalised, it suffices to impose

$$\langle s \rangle C(z) = \mu \left(n \langle s \rangle \int_1^{1/(1-\alpha)} dz' C(z') P(z - \alpha z') + \delta(z-1) \right) \tag{3.23}$$

with an arbitrary $\mu > 0$, rather than the stronger Eq. (3.21) which implies $\mu = 1$. This quantity has a direct physical meaning; the average number of topplings triggered by the toppling of a site, *i.e.* the average number sites which become active by the relaxation of a site is

$$\sigma = n \int_1^{1/(1-\alpha)} dz \int_1^{1/(1-\alpha)} dz' C(z') P(z - \alpha z') \tag{3.24}$$

which is related to (3.23) by $\sigma = 1/\mu - 1/\langle s \rangle$. Of course, it is well known that the branching ratio in these processes is simply $1 - 1/\langle s \rangle$ (Christensen *et al.*, 2001; Harris, 1963), so that $\mu = 1$. The consistency check means now to test whether $\mu = 1$ follows directly from (3.23).

To this end, the distribution of sites activated after k charges is introduced,

$$Q_k^>(z) = (1 - \theta^<(1-z)) \int_1^{1/(1-\alpha)} dz' C(z') Q_{k-1}(z - \alpha z') \; . \tag{3.25}$$

Moreover, the θ-function implicit in Eq. (3.18) and Eq. (3.19) is made explicit:

$$Q_k^< l(z) = \theta^<(1-z) \int_1^{1/(1-\alpha)} dz' C(z') Q_{k-1}^<(z - \alpha z') \; , \tag{3.26}$$

so that by defining

$$\begin{aligned} l^\circ(z) &= 1 + \theta^<(1-z) \frac{\theta^{-1}}{\overline{p_c} n \langle s \rangle} \theta^<((1-\alpha) - z) \qquad (3.27) \\ &\quad + \theta^<(1-z) \frac{1}{\overline{p_c} n \langle s \rangle} \theta^>(z - (1-\alpha)) \\ Q^\circ(z) l^\circ(z) &= \int_1^{1/(1-\alpha)} dz' C(z') Q_{k-1}^<(z - \alpha z') \; , \qquad (3.28) \end{aligned}$$

one can express the distribution of sites charged k times in two parts. One part, $Q_k^<$, is not activated by the charges, while the other part, $Q_k^>$ is:

$$\begin{aligned} Q_k^< &= \theta^<(1-z) Q^\circ(z) \qquad (3.29a) \\ Q_k^> &= (1 - \theta^<(1-z)) Q^\circ(z) \qquad (3.29b) \end{aligned}$$

It is important to note that $Q_{m+1}^>(z)$ has nothing to do with $Q_{m+1}(z)$, which has only been introduced as a shorthand expression for the sites being moved to $z = 1-\alpha$ by the external drive. Using (3.16) one finds

$$\int_1^{1/(1-\alpha)} dz' \int_1^{1/(1-\alpha)} dz C(z') P(z - \alpha z') = g + (1-g) \mathcal{N} \int_1^{1/(1-\alpha)} dz \sum_{k=1}^{m+1} Q_k^>(z) \; , \tag{3.30}$$

where the integration over the δ-function requires

$$\underset{z'\in[1,1/(1-\alpha)]}{\forall}\quad \underset{z\in[1,1/(1-\alpha)]}{\exists}\quad (1-\alpha) = (z - \alpha z') \tag{3.31}$$

which can easily be proven, provided that $\alpha^2 \leq \alpha$.

To find the RHS integral of (3.30) one integrates (3.28),

$$\begin{aligned}
\int_0^{1/(1-\alpha)} dz\, Q_k^\circ(z) l^\circ(z) &= \int_1^{1/(1-\alpha)} dz' \int_0^1 dz\, Q_{k-1}^<(z) C(z') && \text{(3.32a)}\\
&= \int_0^{1/(1-\alpha)} dz \int_1^{1/(1-\alpha)} dz'\, Q_{k-1}^<(z-\alpha z') C(z') && \text{(3.32b)}\\
&= \int_0^1 dz\, Q_{k-1}^<(z)\ , && \text{(3.32c)}
\end{aligned}$$

where $Q_k^<(z) \equiv 0$ for $z \notin [0,1]$ has been used as well as

$$\underset{z'\in[1,1/(1-\alpha)]}{\forall}\quad \Big[-\alpha z', \frac{1}{1-\alpha} - \alpha z'\Big] \supseteq [0,1] \tag{3.33}$$

It is useful to rewrite the result (3.32c) as

$$\int_0^{1/(1-\alpha)} dz \Big(Q_k^<(z) - Q_{k+1}^\circ(z)\Big) = \int_0^1 dz\, Q_{k+1}^<(z)\Big(l^\circ(z) - 1\Big) \tag{3.34}$$

using the HEAVISIDE function in the definition of $l^\circ(z)$. Since Q_{k+1}° includes $Q_{k+1}^<$, summing the LHS over k cancels these terms and one arrives at

$$\begin{aligned}
\int_0^{1/(1-\alpha)} dz \sum_{k=0}^m \Big(Q_k^<(z) - Q_{k+1}^\circ(z)\Big) &= \int_0^{1/(1-\alpha)} dz \Big(Q_0^<(z) - \sum_{k=0}^m Q_{k+1}^>(z)\Big) && \text{(3.35)}\\
&= 1 - \int_1^{1/(1-\alpha)} dz \sum_{k=1}^{m+1} Q_k^> && \text{(3.36)}
\end{aligned}$$

where in the first line $Q_{m+1}^> \equiv 0$ has been used. This gives together with the RHS of (3.34) the exact result

$$\int_1^{1/(1-\alpha)} dz \sum_{k=1}^{m+1} Q_k^> = 1 - \int_0^1 dz \sum_{k=1}^m Q_k^<(z)\Big(l^\circ(z) - 1\Big) \tag{3.37}$$

However, in order to give a final result for the value of μ in (3.23), it still remains to calculate the RHS of Eq. (3.37). Therefore one notes that

$$\begin{aligned}\int_0^1 dz \sum_{k=1}^{m} Q_k^<(z)\theta^<((1-\alpha)-z) &= \int_0^1 dz \sum_{k=0}^{m} Q_k^<(z)\theta^<((1-\alpha)-z) - 1 \\ = \frac{\mathcal{N}_{\hat{P}}}{(1-g)} \int_0^1 dz\, \widetilde{P}(z)\theta^<((1-\alpha)-z) - 1 &= \frac{\mathcal{N}_{\hat{P}}}{(1-g)}\overline{p_{\mathrm{st}}} - 1 \end{aligned} \tag{3.38}$$

and

$$\begin{aligned}&\int_0^1 dz \sum_{k=1}^{m} Q_k^<(z)(1-\theta^<((1-\alpha)-z)) = \int_0^1 dz \sum_{k=0}^{m} Q_k^<(z)(1-\theta^<((1-\alpha)-z)) \\ &= \frac{\mathcal{N}_{\hat{P}}}{(1-g)} \int_0^1 dz \left(\widetilde{P}(z)(1-\theta^<((1-\alpha)-z)) - g\delta((1-\alpha)-z)\right) = \frac{\mathcal{N}_{\hat{P}}}{(1-g)}(\overline{p_{\mathrm{c}}} - g)\ . \end{aligned} \tag{3.39}$$

For (3.37) this gives

$$\int_1^{1/(1-\alpha)} dz \sum_{k=1}^{m+1} Q_k^> = 1 - \frac{\theta^{-1}}{n\langle s\rangle}\Big(\frac{\mathcal{N}_{\hat{P}}}{(1-g)}\overline{p_{\mathrm{st}}} - 1\Big) - \frac{1}{n\langle s\rangle\overline{p_{\mathrm{c}}}}\Big(\frac{\mathcal{N}_{\hat{P}}}{(1-g)}(\overline{p_{\mathrm{c}}} - g)\Big) \quad , \tag{3.40}$$

which can now be inserted into (3.30). Integrating over (3.23) this gives

$$1 = \mu\left\{ n\left(g + (1-g)\mathcal{N}_{\hat{P}}^{-1}\left[1 - g_0\Big(\frac{\mathcal{N}_{\hat{P}}}{(1-g)}\overline{p_{\mathrm{st}}} - 1\Big) - \frac{g_0}{\theta^{-1}\overline{p_{\mathrm{c}}}}\Big(\frac{\mathcal{N}_{\hat{P}}}{(1-g)}(\overline{p_{\mathrm{c}}} - g)\Big)\right]\right) + \frac{1}{\langle s\rangle}\right\}, \tag{3.41}$$

using

$$g_0 = \left(\frac{1/\theta}{n\langle s\rangle}\right) \tag{3.42}$$

and finally

$$\mu = \mathcal{N}_{\hat{P}}\big(n(1-g)(1+g_0)\big)^{-1}\ . \tag{3.43}$$

In order to show that this is in fact 1, one imposes stationarity on class $k = 0$. The fraction of sites in this class is given by the integral over $P(z)$ from 0 to "infinitesimal above 0", which gives a pre-factor $\mathcal{N}_{\hat{P}}^{-1}(1-g)$, since $Q_0^<$ is normalised. The inflow to this class by imposing conservation of sites is exactly $\langle s\rangle$, which includes the initial seed. The outflow is due to external drive and hits of toppling sites, therefore

$$\langle s\rangle = \mathcal{N}_{\hat{P}}^{-1}(1-g)\theta^{-1} + \mathcal{N}_{\hat{P}}^{-1}(1-g)n\langle s\rangle\ . \tag{3.44}$$

Rewriting this equation one arrives at

$$\mathcal{N}_{\hat{P}}^{-1}(1-g) = \frac{\langle s\rangle}{n\langle s\rangle + \theta^{-1}} = \frac{1/n}{1+g_0} \quad , \tag{3.45}$$

which gives together with (3.43) as expected $\mu = 1$.

3.3.3.2 Invariant II

The second check for consistency is to derive (3.15) directly from $Q_k^\circ(z)$ and (3.23):

$$\begin{aligned}
\int_1^{1/(1-\alpha)} dz\, zC(z) &= \overline{z_{\text{act}}} \\
&= \int_1^{1/(1-\alpha)} dz\, z\mu\Big(n\int_1^{1/(1-\alpha)} dz'\,\widetilde{P}(z-\alpha z')C(z') + \frac{1}{\langle s\rangle}\delta(z-1)\Big) \\
&= \mu\Big(\frac{1}{\langle s\rangle} + n\int_1^{1/(1-\alpha)} dz \int_1^{1/(1-\alpha)} dz'\, zg\delta((1-\alpha)-(z-\alpha z'))C(z') \\
&\quad + n\int_1^{1/(1-\alpha)} dz \int_1^{1/(1-\alpha)} dz'\, z(1-g)\mathcal{N}_{\hat{P}}^{-1}\sum_{k=0}^{m} Q_k^<(z-\alpha z')C(z')\Big) \\
&= \mu\Big(\frac{1}{\langle s\rangle} + ng(1-\alpha) + ng\alpha\overline{z_{\text{act}}} \\
&\quad + n(1-g)\mathcal{N}_{\hat{P}}^{-1}\int_1^{1/(1-\alpha)} dz\, z\sum_{k=1}^{m+1} Q_k^>(z)\Big)
\end{aligned}$$

by inserting the definition of $P(z)$ (3.16) and of $\overline{z_{\text{act}}}$. Just like in (3.30) the last integral is to be evaluated and the same technique as above can be used. Integrating the first moment of (3.28) using the definition of $l^\circ(z)$ (3.27) gives

$$\int_0^{1/(1-\alpha)} zQ_k^\circ(z)l^\circ(z) = \int_0^{1/(1-\alpha)} dz\int_1^{1/(1-\alpha)} dz'\, zQ_{k-1}^<(z-\alpha z')\widetilde{C}(z') \tag{3.47}$$

$$= \alpha\overline{z_{\text{act}}}\int_0^1 dz\, Q_{k-1}^<(z) + \int_0^1 dz\, zQ_{k-1}^<(z) \tag{3.48}$$

where (3.33) has been used. Rewriting this equation similar to (3.34) yields

$$\int_0^{1/(1-\alpha)} z\Big(Q_k^<(z) - Q_{k+1}^\circ(z)\Big) = \int_0^1 dz\, zQ_{k+1}^<(z)(l^\circ(z)-1) - \alpha\overline{z_{\text{act}}}\int_0^1 dz\, Q_k^<(z) \tag{3.49}$$

The sum over the LHS is

$$\sum_{k=0}^{m}\int_0^{1/(1-\alpha)} dz\, z\Big(Q_k^<(z) - Q_{k+1}(z)\Big) = -\int_1^{1/(1-\alpha)} dz\, z\sum_{k=0}^{m} Q_{k+1}^>(z) \quad .$$

while the sum over the RHS of (3.49) gives

$$\int_0^1 dz\, z\sum_{k=0}^{m} Q_{k+1}^<(z)(l(z)-1) = \frac{\mathcal{N}_{\hat{P}}}{(1-g)}\Big(g_0\overline{z_{\text{st}}p_{\text{st}}} + \frac{g_0}{\theta^{-1}\overline{p_{\text{c}}}}\big(\overline{z_{\text{c}}p_{\text{c}}} - g(1-\alpha)\big)\Big) \; . \tag{3.50}$$

The sum over the second term is simply

$$\int_0^1 dz \sum_{k=0}^{m} Q_k^<(z) = \mathcal{N}_{\hat{P}} \tag{3.51}$$

The sum of all contributions gives for (3.49)

$$\begin{aligned}\int_1^{1/(1-\alpha)} dz\, z \sum_{k=0}^{m} Q_{k+1}^>(z) &= \frac{\mathcal{N}_{\hat{P}}}{(1-g)} \\ &\times \left(\alpha\overline{z_{\mathrm{act}}}(1-g) - g_0\overline{z_{\mathrm{st}}p_{\mathrm{st}}} - \frac{g_0}{\theta^{-1}\overline{p_{\mathrm{c}}}}\big(\overline{z_{\mathrm{c}}p_{\mathrm{c}}} - g(1-\alpha)\big)\right) \quad ,\end{aligned} \tag{3.52}$$

which now can be inserted in (3.46a) to give

$$\begin{aligned}\frac{\overline{z_{\mathrm{act}}}}{n\mu} &= \frac{1}{n\langle s\rangle} + g(1-\alpha) + g\alpha\overline{z_{\mathrm{act}}} \\ &+ \alpha\overline{z_{\mathrm{act}}}(1-g) - g_0\overline{z_{\mathrm{st}}p_{\mathrm{st}}} - \frac{g_0}{\theta^{-1}\overline{p_{\mathrm{c}}}}\big(\overline{z_{\mathrm{c}}p_{\mathrm{c}}} - g(1-\alpha)\big) \end{aligned} \tag{3.53}$$

$$= g_0\overline{p_{\mathrm{st}}}\big((1-\alpha) - \overline{z_{\mathrm{st}}}\big) + \alpha\overline{z_{\mathrm{act}}} + \frac{1}{n\langle s\rangle}(1-\overline{z_{\mathrm{c}}}) \tag{3.54}$$

where $g_0/\theta^{-1} = (n\langle s\rangle)^{-1}$ has been used. This can be rewritten as the final result

$$\langle s\rangle = \frac{((1-\alpha) - \overline{z_{\mathrm{st}}})\theta^{-1}\overline{p_{\mathrm{st}}} + (1-\overline{z_{\mathrm{c}}})}{(\mu^{-1} - n\alpha)\overline{z_{\mathrm{act}}}} \tag{3.55}$$

which differs from (3.15) only by the a factor μ^{-1} instead of 1 in the denominator. Again one finds $\mu = 1$ as soon as Eq. (3.15) is imposed as a "physical" expression since it relies on conservation, which is not inherent to the recursion relation Eq. (3.18).

To summarise these two sections, whenever conservation is imposed on the process, one arrives at an equation for $\langle s\rangle$ as seen above and/or $\mu = 1$. These two quantities are the physical expression of conservation.

3.3.4 Comparison to numerics

All the equations above can alternatively be derived from the micro-dynamics of the system. This ensures that the solution is exact in the thermodynamic limit given the stationarity assumption.

The implementation of the iteration procedure is straight-forward. As a criterion for termination, one could check whether $C(z)$, as defined by (3.21), is properly normalised (Bröker and Grassberger, 1997), as it can be proven that it must be correctly normalised if it is a solution. However, it would be sufficient to assume $C(z)$ proportional to the RHS in (3.21). Moreover, in the numerical procedure the quality of the normalisation of $C(z)$ depends strongly on the resolution of the grid chosen, when-

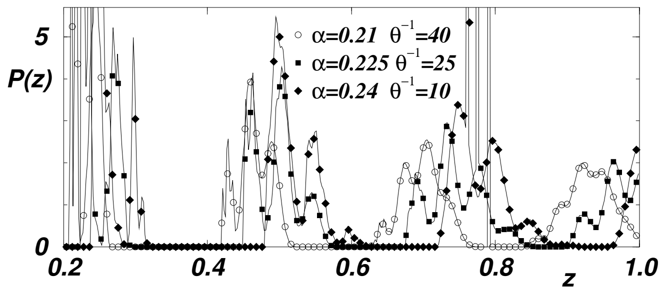


Figure 3.6: Distribution of energy, $P(z)$, for $n = 4$ and different values of α. Since there is only a delta peak in $0 \leq z \leq 0.2$, results for $z < 0.2$ are cut off. Continuous lines indicate results from theory (grid size 32000, integrated in 125 bins), points represent results from numerical simulations with $N = 10^6$, 10^6 avalanches for equilibration and $5\ 10^6$ avalanches for statistics (125 bins for the histogram).

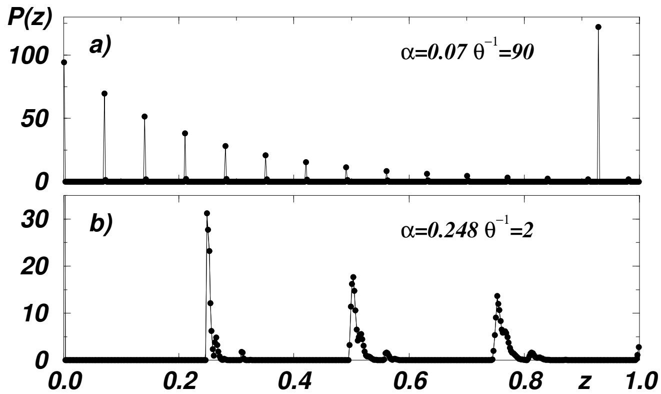


Figure 3.7: $P(z)$ as in Fig. 3.6, but for $\alpha = 0.07$ and $\alpha = 0.248$ and 500 bins.

ever $C(z)$ changes rapidly as function of z. Therefore convergence of the iteration procedure is better verified by checking whether $C(z)$ approaches a fixed point, *i.e.* is invariant under (3.21). Since the distribution is expected to be highly non-analytic - there are at least two δ-functions in $P(z)$ - sophisticated integration routines are inappropriate. For $n = 4$ the procedure quickly converges for $0.07 < \alpha < 0.24$, all non-pathological initial values tested lead to the same stable solutions. Only for small values of $\langle s \rangle$, when the delta peak of the initial seed starts to propagate through the distribution, a large grid is required for sufficient resolution. The same problem appears close to the commensurable limits mentioned below. In Fig. 3.6 numerical simulations of the model are compared to the numerical solution of the analytical approach. Although the PDF is very structured, discrepancies are small and can be reduced by increasing the resolution of the underlying grid.

3.3.5 Discussion

The distribution $C(z)$ collapses to a δ-function in at least two limits. Firstly, when $\alpha \to 0$, the R-DS-FFM-limit, the probability for a site to become susceptible due to a number of charges vanishes as $q^{(1-\alpha)/\alpha}$, where $q < 1$ is the product of the probability that a site receives a charge from a relaxing site and the probability that a site is not driven externally between two hits. Hence, for $\alpha \to 0$ the mechanism of "growing by charges" becomes negligible and the external drive becomes the dominating source for susceptible sites. The dynamics now become equivalent to the R-DS-FFM: stable sites = empty sites, susceptible sites = trees, and active sites = fires. Furthermore, as $\alpha \to 0$, the support of $C(z)$ becomes smaller and smaller and the distribution of active sites is strongly peaked at $z = 1$, collapsing to a δ-function. Therefore, the distributions Q_k, $k > 0$ are less smeared out, as shown in Fig. 3.7(a) for a small value of α. Assuming $C(z)$ to be a δ-function, one can easily reconcile the results in (Christensen *et al.*, 1993, Eqs. (3) and (7)). The assumption that $P(z)$ is unaffected by single avalanches corresponds to $p, f \to 0$ in the SOC-limit of the DS-FFM (Clar *et al.*, 1996).

In the second limit, $\alpha \to 1/n$, the model becomes conservative, but more important, α becomes more and more commensurable in the sense that a site charged $m = n$ times is almost always active and therefore the support of Q_m vanishes, as it is squeezed between $n\alpha$ and 1. When α is very close to $1/n$, most of the active sites are provided by Q_{m-1} and their average energy is just above 1, *i.e.* $C(z)$ becomes more and more δ shaped and so do the Q_i, as shown in Fig. 3.7(b). The same behaviour is obtained whenever $k\alpha = 1$ for $k \in \mathbb{N}$.

3.3.5.1 Exponents

The critical exponents of the model, for all $\alpha \in]0, 1/n[$, can be obtained by mapping it on to a branching process (Harris, 1963) in order to identify the critical exponent $b = 2$, where $\mathcal{P}(t) \propto t^{-b}$ is the exponent of avalanche duration. The exponent $\tau = 3/2$, found by mapping the model on to a random walker along an absorbing barrier, is the exponent of avalanche sizes, $\mathcal{P}(s) \propto s^{-\tau}$. Formally these exponents arise only for diverging cutoffs in the distributions, which are controlled by the average number of active sites produced per single toppling, the branching ratio σ. The cutoffs diverge for $\sigma \to 1$.

However, the mapping is non-trivial, except when $C(z)$ is a δ-function. This is because a distribution of active sites entails a distribution in the branching ratio, *i.e.* the branching probability itself becomes a random variable. However, it is not necessary to consider the explicit time dependence of the branching probability. In order to justify the mapping it is sufficient, though not less accurate, to consider the ensemble average of the process with an annealed disorder in the branching probability, *i.e.* writing the probabilities for one given node (active site) to branch into k new nodes (active sites) as

$$\mathcal{P}(1 \to k) = \left\langle \binom{n}{k} p^k (1-p)^{n-k} \right\rangle_p \tag{3.56}$$

where p denotes the branching probability (which is a function of the energy of the site) and $\langle\rangle_p$ denotes the weighted average over the probabilities. Therefore $\sigma = \sum_k k\mathcal{P}(1 \to k)$. This branching process is characterised by the same generating functions as the standard branching process (Harris, 1963), which becomes critical for $\sigma = 1$. Hence the condition for criticality is

$$\sum_{k=0}^{\infty} k \left\langle \binom{n}{k} p^k (1-p)^{n-k} \right\rangle_p = n\langle p \rangle_p = 1 \tag{3.57}$$

which is the (average) branching ratio, according to (3.21) given by

$$n\langle p \rangle_p = n \int_1^{1/(1-\alpha)} dz \int_1^{1/(1-\alpha)} dz' \, C(z') P(z - \alpha z') = 1 - \frac{1}{\langle s \rangle} \, . \tag{3.58}$$

Defining $C(z,t)$ to be the averaged distribution of active sites after t updates, $C(z,t)$ will only gradually change from an initial delta peak at $z = 1$ towards $C(z)$, the latter being then the active-site weighted average of $C(z,t)$. However, any deviation from $C(z)$ decays exponentially fast, which can be seen by investigating the MARKOV chain of the repetitive convolution of $C(z,t)$ with $P(z)$ as in (3.21). Therefore the cutoff, introduced by the deviation of σ from 1, is dominated by the asymptotic iteratively stable limit of $C(z)$ only. Since the asymptote is approached exponentially fast the transient cannot influence the value of the exponents. The same arguments apply for the random walker approach, therefore $b = 2$ and $\tau = 3/2$ is true for all $\alpha \in]0, 1/n[$.

3.3.5.2 Finite systems

The calculations above are *a priori* valid only in the thermodynamic limit. However, a simulation of the model must consider a finite system. Moreover the model relies on several assumptions, which entail certain finite size scaling: θ_N^{-1}/N (the index indicates the value to be measured in a system of size N) as well as $\langle s \rangle_N / N$ must vanish for diverging N, while $\theta_N^{-1}/\langle s \rangle_N$ must remain constant. It is a well known problem in the DS-FFM (Clar *et al.*, 1994) that the number of trees grown between two ignitions is a parameter, θ_N^{-1}, which needs to be *tuned* according to the system size; it is supposed to diverge, but its value is restricted by system size. An inappropriately chosen parameter produces a small value of the cutoff or a bump in the distribution

Open problem

function of avalanche sizes. Nevertheless, $P(z)$ depends only weakly on θ_N^{-1}. As a more quantitative measure for the "right choice of θ_N^{-1}" we compared g_N to g [see Eq. (3.22)] in the thermodynamic limit. Assuming a cutoff of order $\mathcal{O}(N)$ in $\mathcal{P}(s)$ of a finite system, the scaling is $\langle s\rangle_N = \int ds\mathcal{P}(s)s \in \mathcal{O}(N^{1/2})$ and thus $\theta_N^{-1} \in \mathcal{O}(N^{1/2})$. For a more quantitative picture one can map the avalanche on to a random walker along an absorbing barrier with time dependent walking probability $p(t)$ (in the sense of (Fisher, 1984) a drinking rather than a drunken random walker). An idea how to solve this technically very interesting problem is presented in the next subsection.

In summary, a solvable SOC model, critical in the entire regime of the conservation parameter, has been defined and the main properties have been derived. The critical exponents are as expected the trivial exponents of a critical branching process and a random walker. The model clarifies the rôle of the external drive and represents an explicit example of the recovery of criticality by introducing an external drive.

3.4 Supplement: Random Walker Approach

In order to be able to estimate the scaling of the "proper choice" of θ_N^{-1}, one could map the process described above on a random walker with time-dependent drift and diffusion constant. The reason for this time-dependence is the gradual decrease of susceptible sites in the system, so that finally the dissipation kills off any avalanche. If $\phi(t,x;x_0)$ is the probability to have x active sites in the system after t updates, starting from an initial activity x_0, then the PDE to solve is

$$\partial_t\phi(t,x) = d(t)\partial_x^2\phi(t,x) - v(t)\phi(t,x) \tag{3.59}$$

with boundary conditions

$$\begin{aligned} \phi(t,x=0;x_0) &= 0 && (3.60a)\\ \lim_{t\to 0}\phi(t,x;x_0) &= \delta(x-x_0)\ ,\ x_0>0 && (3.60b) \end{aligned}$$

where the last line imposes the solution to be the propagator of the problem, so that general initial conditions, given by a source function $s(x)$ would be solved by $\int dx'\ s(x')\phi(t,x;x')$. Here, $v(t)$ is the time-dependent drift and $d(t)$ is the time-dependent diffusion constant.

Without boundary condition (3.60a) the problem is trivial — the normalised solution for $x \in [-\infty,\infty]$ is (as one might expect) given by

$$\phi_0(t,x;x_0) = \frac{1}{\sqrt{4\pi D(t)}}e^{-\frac{(x-x_0-V(t))^2}{4D(t)}} \tag{3.61}$$

where $D(t)$ and $V(t)$ are simply the integrals of $d(t)$ and $v(t)$:

$$D(t) \equiv \int_0^t dt' \, d(t') \tag{3.62a}$$

$$V(t) \equiv \int_0^t dt' \, v(t') \tag{3.62b}$$

assuming integrability of $d(t)$ and $v(t)$ respectively.

Since $v(t)$ and $d(t)$ are independent, they represent two independent (time dependent) scales, and it is impossible to gauge $v(t)$ away, as can be done for the time-independent problem,

$$\partial_t \phi(t, x; x_0) = d\partial_x^2 \phi(t, x; x_0) - v\partial_x \phi(t, x; x_0) \tag{3.63}$$

by assuming $\phi(t, x; x_0) = g(x, t)\widetilde{\phi}(t, x; x_0)$ where $\widetilde{\phi}$ obeys (3.63) with $v = 0$, which gives $g(x, t) = c \exp(vx/(2d) - v^2 t/(4d))$.

It is more convenient to rewrite the original problem (3.59) as

$$\partial_t \phi(t, x; x_0) = \partial_x^2 \phi(t, x; x_0) - w(t)\phi(t, x; x_0) \tag{3.64}$$

so that a $\widetilde{\phi}(t, x) \equiv \phi(\alpha(t), x)$ solves the original problem, for $\dot{\alpha}(t) = d(t)$ and $w(t) = v(t)/\dot{\alpha}(t)$ as

$$\partial_t \widetilde{\phi} = \dot{\alpha}\partial_t \phi = \dot{\alpha}\left(\partial_x^2 \phi - w\phi\right) \tag{3.65}$$

using $d(t) \neq 0$ for $t \geq 0$. Therefore we can use $D(t) = t$ from now on, so that the propagator becomes

$$\phi_0(t, x; x_0) \equiv \frac{1}{\sqrt{4\pi t}} e^{-\frac{(x - x_0 - W(t))^2}{4t}} \quad . \tag{3.66}$$

with $W(t) \equiv \int_0^t dt' \; w(t')$. Since (3.66) solves the equation and the boundary conditions up to Eq. (3.60a), the goal of the remaining section is to construct a corresponding "source field", the convolution with which makes the result obey Eq. (3.60). In a sense the status of the PDE has changed: It is now to be solved for $x \in [-\infty, \infty]$, while the source field is to be chosen such that (3.60a) is obeyed everywhere and (3.60b) is obeyed in the upper half plane, *i.e.*

$$\lim_{t \to 0} \theta(x)\phi(t, x; x_0) = \theta(x)\delta(x - x_0) \; , \; x_0 > 0 \tag{3.67}$$

The idea behind this approach is that *any* field can be expressed in terms of this propagator and therefore *any* solution, say $\varphi(t, x; x_0)$ could serve as a source field for the solution, in the sense that

$$\int_{-\infty}^{+\infty} dx' \; \varphi(t_0, x'; x_0)\phi_0(t - t_0, x; x') \tag{3.68}$$

solves the problem for $t \geq t_0$ — in other words Eq. (3.68) is a representation of $\varphi(t, x; x_0)$.

A remark on this are worth to be mentioned. Since ϕ_0 as defined in (3.66) obeys the PDE (3.64), it is clear that (3.68) does also obey the PDE for $t \geq t_0$. In other words, there are *two* solutions for the problem

$$\lim_{t \to t_0} \Phi(t, x; x_0) = \varphi(t_0, x; x_0) \tag{3.69}$$

$$\partial_t \Phi(t, x; x_0) = \partial_x^2 \Phi(t, x; x_0) - w(t)\Phi(t, x; x_0) \tag{3.70}$$

namely

$$\Phi_1(t, x; x_0) = \varphi(t, x; x_0) \tag{3.71}$$

and

$$\Phi_2(t, x; x_0) = \int_{-\infty}^{+\infty} dx' \; \varphi(t_0, x'; x_0)\phi_0(t - t_0, x; x') \tag{3.72}$$

From the uniqueness of the forward diffusion with given initial condition it follows that $\Phi_1(t, x; x_0) = \Phi_2(t, x; x_0)$ for $t \geq t_0$. In the limit $t_0 \to 0$ the uniqueness for $t \geq 0$ follows.

3.4.1 Constructing the source field

The aim is now to find a source function $\bar{s}(x; x_0)$ such that

$$\phi(t, x; x_0) = \int_{-\infty}^{+\infty} dx' \; \bar{s}(x'; x_0)\phi_0(t, x; x') \tag{3.73}$$

solves the PDE (3.64) with BC (3.60). The source field has the property

$$\mathcal{N}^{-1}\bar{s}(x'; x_0) = \delta(x - x_0) + \hat{s}(x'; x_0) \tag{3.74}$$

where $\mathcal{N}$ is a normalisation constant and $\hat{s}(x; x_0) = 0$ for $x \geq 0$. In order to obey (3.60a) one therefore needs

$$\int_{-\infty}^{0} dx' \; \hat{s}(x'; x_0)\phi_0(t, 0; x') = -\phi_0(t, 0; x_0) \; . \tag{3.75}$$

Using (3.66) and $s(x; x_0) = \hat{s}(-x; x_0)$ it is

$$\int_{0}^{\infty} dx' \; s(x'; x_0) e^{-\frac{(x' - W(t))^2}{4t}} = -e^{-\frac{(x_0 + W(t))^2}{4t}} \tag{3.76}$$

The difference in the sign in the exponential in this equation makes the mirror charge solution possible only for $W(t) = 0$ or $W(t) = bt$, b being a constant. That can be seen by expanding the square in the exponent on both sides and obtaining then simply $\exp(-x'^2/(4t) + x'b/2)$ in the integral, so that the solution becomes via

$$\int_0^\infty dx'\ s(x';x_0) \exp((x'+x_0)b/2) \exp(-x'^2/(4t)) = -\exp(-x_0^2/(4t)) \tag{3.77}$$

simply

$$s(x';x_0) \exp((x'+x_0)b/2) = \delta(x'-x_0) \quad , \tag{3.78}$$

which is the mirror charge solution $s(x';x_0) = -\delta(x'-x_0)e^{(x_0 b)}$. However, as $W(t) = bt$ has been used here, $w(t)$ is just a constant, so that this solution could easily be obtained by gauging the drift away.

In general the problem consists of finding a solution $s(x;x_0)$ of the integral equation (3.76). Introducing $\widetilde{x} = \frac{(x'-W(t))}{2\sqrt{t}}$ makes it possible to rewrite the integral as

$$\int_{\mu(t)}^\infty d\tilde{x}\ \sigma(t,\tilde{x};x_0)e^{-\tilde{x}^2} = g(t;x_0) \tag{3.79}$$

using

$$\mu(t) \equiv -W(t)/(2\sqrt{t}) \tag{3.80}$$

$$g(t;x_0) \equiv -\frac{1}{2\sqrt{t}} \exp\Big(-\frac{(x_0+W(t))^2}{4t}\Big) \tag{3.81}$$

$$\sigma(t,\tilde{x};x_0) \equiv s(2\sqrt{t}(\tilde{x}-\mu(t));x_0) \tag{3.82}$$

This is nothing else than an integral transformation, an ansatz for a solution of which is discussed below. *This ansatz will fail*, yet it provides a very promising route for similar problems. One of the key questions yet to be answered is whether the integral (3.79) actually exists, *i.e.* whether it converges.

First σ is expressed in terms of polynomials in x with an additional parameter for the weight which are defined by a orthogonality relation:

$$\int_u^\infty dx\ p_i(x,u)p_j(x,u)e^{-x^2} = h_i(u)\delta_{ij} \tag{3.83}$$

For uniqueness one imposes $\kappa_{n,n}(u) = 1$ (monic polynomials) in

$$p_n(x,u) = \sum_{i=0}^n \kappa_{n,i}(u)x^i\ . \tag{3.84}$$

The function σ can now be rewritten as

$$\sigma(t,\tilde{x};x_0) = \sum_{i=0}^{\infty} a_i(t,u;x_0)p_i(\tilde{x},u) \tag{3.85}$$

which has the interesting property

$$\partial_u \sigma(t,\tilde{x};x_0) = 0 \tag{3.86}$$

In this notation the problem reduces to determine a_i from

$$\langle \sigma(t,\tilde{x};x_0)|p_i(\tilde{x},u)\rangle_u \equiv \int_u^{\infty} d\tilde{x}\ \sigma(t,\tilde{x};x_0)p_i(\tilde{x},u)e^{-x^2} = h_i(u)a_i(t,u;x_0) \tag{3.87}$$

which defines the projection operator $\langle\cdot|\cdot\rangle_u$ used henceforth. As a shorthand notation $\langle\cdot\rangle_u \equiv \langle\cdot|1\rangle_u$ is introduced. One non-trivial information, (3.79), about $\sigma(t,\tilde{x};x_0)$ can then be written as

$$\langle \sigma(t,\tilde{x};x_0)|1\rangle_{\mu(t)} = a_0(t,\mu(t);x_0)\,h_0(\mu(t)) = g(t;x_0) \tag{3.88}$$

using $p_0(\tilde{x},u) = 1$ by definition. The only other information about about $\sigma(t,\tilde{x};x_0)$ is (3.82)), which implies

$$2t\partial_t\sigma(t,\tilde{x};x_0) = \big(\tilde{x} - \mu(t) - 2t\dot{\mu}(t)\big)\,\partial_{\tilde{x}}\sigma(t,\tilde{x};x_0) \tag{3.89}$$

with $\dot{\mu}(t) \equiv \frac{d}{dt}\mu(t)$.

3.4.2 General properties of the polynomials

Since p_n is orthogonal to all polynomials of degree less than n, one has the standard recurrence relation

$$xp_i(x,u) = p_{i+1}(x,u) + \alpha_i(u)p_i(x,u) + \beta_i(u)p_{i-1}(x,u) \tag{3.90}$$

A priori the quantities $\alpha_i(u)$ and $\beta_i(u)$ are unknown, but they can be related to $h_i(u)$. From (3.90) and (3.83) one finds by straight-forward integration

$$\begin{aligned}
\beta_i(u) &= \frac{h_i(u)}{h_{i-1}(u)} && \text{(3.91a)}\\
\alpha_i(u)h_i(u) &= \frac{1}{2}p_i^2(u,u)e^{-u^2} && \text{(3.91b)}\\
\beta_{i+1} + \alpha_i^2 + \beta_i &= u\alpha_i + i + \frac{1}{2} && \text{(3.91c)}\\
ih_{i-1}(u) &= -p_i(u,u)p_{i-1}(u,u)e^{-u^2} + 2h_i(u) && \text{(3.91d)}
\end{aligned}$$

In general one finds for $p_i'(x,u) \equiv \frac{\partial}{\partial x} p_i(x,u)$

$$\langle p_i' | p_j \rangle_u = \begin{cases} 0 & \text{for } j \geq i \\ -p_i(u,u) p_j(u,u) e^{-u^2} + 2h_i(u) & \text{for } j = i-1 \\ -p_i(u,u) p_j(u,u) e^{-u^2} & \text{for } j < i-1 \end{cases} \tag{3.92}$$

3.4.2.1 The rôle of u

The parameter u looks like a simple generalisation of the set of polynomials and for $u \to -\infty$ one recovers Hermité's polynomials. But in the light of (3.88) it looks like it is the t-dependence and hence only the dependence on u, which makes the problem well defined, as (3.76) must be obeyed for all t.

To find out more about the u dependence, one can derive $\langle p_i | p_j \rangle_u = \delta_{ij} h_i(u)$ by u. By definition (3.84) $\partial_u p_i$ is a polynomial of order $i-1$, so that $\langle \partial_u p_i | p_j \rangle_u = 0$ for $j \geq i$. In conjunction with (3.91b) the case $i = j$ gives rise to

$$h_i'(u) = -2\alpha_i(u) h_i(u) \tag{3.93}$$

while for $j < i$ one finds

$$\langle \partial_u p_i | p_j \rangle_u = \begin{cases} 0 & \text{for } j \geq i \\ p_i(u,u) p_j(u,u) e^{-u^2} & \text{for } j < i \end{cases} \tag{3.94}$$

Together with (3.92) one finds $\langle p_i' | p_j \rangle_u = -\langle \partial_u p_i | p_j \rangle_u$ for $j \geq i$ or $j < i-1$. The case $j = i-1$ requires an extra term, namely $\langle p_i' | p_{i-1} \rangle_u = -\langle \partial_u p_i | p_{i-1} \rangle_u + 2h_i$. Together with (3.91a) this leads to the conclusion

$$p_i' + \partial_u p_i = 2\beta_i(u) p_{i-1}(x,u) \tag{3.95}$$

which is true even for $i = 0$, by defining $p_{-1} \equiv 0$.

There is another property of this system of polynomials which will be used below. Based on (Chen and Ismail, 1997) one easily finds

$$(x-u) p_i' = i p_i + 2\beta_i (\alpha_i + \alpha_{i-1} - u) p_{i-1} + 2 \frac{h_i}{h_{i-2}} p_{i-2} \tag{3.96}$$

using (3.91d), (3.91b) and (3.91a).

3.4.2.2 Explicit form

Out of curiosity, the first few p_i can be calculated explicitly. Defining the error function $\mathcal{E}(x)$ as

$$\mathcal{E}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x dx' e^{-x'^2} \tag{3.97}$$

one finds after some algebra

$$\begin{aligned}
\langle 1 \rangle_u &= \frac{\sqrt{\pi}}{2}\big(1-\mathcal{E}(u)\big) && (3.98a)\\
\langle x \rangle_u &= \frac{1}{2}e^{-u^2} && (3.98b)\\
\langle x^2 \rangle_u &= \frac{\sqrt{\pi}}{4}\big(1-\mathcal{E}(u)\big) + \frac{1}{2}ue^{-u^2} && (3.98c)\\
\langle x^3 \rangle_u &= \frac{1}{2}(1+u^2)e^{-u^2} && (3.98d)\\
\langle x^4 \rangle_u &= \frac{3\sqrt{\pi}}{8}\big(1-\mathcal{E}(u)\big) + \frac{1}{4}(3u+2u^3)e^{-u^2} && (3.98e)\\
\langle x^5 \rangle_u &= \frac{1}{2}(2+2u^2+u^4)e^{-u^2} && (3.98f)\\
\langle x^6 \rangle_u &= \frac{15\sqrt{\pi}}{16}\big(1-\mathcal{E}(u)\big) + \frac{1}{8}(15u+10u^3+4u^5)e^{-u^2} && (3.98g)\\
\langle x^7 \rangle_u &= \frac{1}{2}(6+6u^2+3u^4+u^6)e^{-u^2} && (3.98h)\\
\langle x^8 \rangle_u &= \frac{105\sqrt{\pi}}{32}\big(1-\mathcal{E}(u)\big) + \frac{1}{16}(105u+70u^3+28u^5+8u^7)e^{-u^2} && (3.98i)
\end{aligned}$$

and their interrelation is found to be

$$\begin{aligned}
\langle 1 \rangle_u &= \frac{\sqrt{\pi}}{2}\big(1-\mathcal{E}(u)\big) && (3.99a)\\
\langle x \rangle_u &= \frac{1}{2}e^{-u^2} && (3.99b)\\
\langle x^2 \rangle_u &= \frac{1}{2}\langle 1 \rangle_u + u\langle x \rangle_u && (3.99c)\\
\langle x^3 \rangle_u &= \langle x \rangle_u + u^2\langle x \rangle_u && (3.99d)\\
\langle x^4 \rangle_u &= \frac{3}{2}\langle x^2 \rangle_u + u^3\langle x \rangle_u && (3.99e)\\
\langle x^5 \rangle_u &= 2\langle x^3 \rangle_u + u^4\langle x \rangle_u && (3.99f)\\
\langle x^6 \rangle_u &= \frac{5}{2}\langle x^4 \rangle_u + u^5\langle x \rangle_u && (3.99g)\\
\langle x^7 \rangle_u &= 3\langle x^5 \rangle_u + u^6\langle x \rangle_u && (3.99h)\\
\langle x^8 \rangle_u &= \frac{7}{2}\langle x^6 \rangle_u + u^7\langle x \rangle_u && (3.99i)
\end{aligned}$$

which can be written as

$$\langle x^{n+1} \rangle_u = \frac{n}{2}\langle x^{n-1} \rangle_u + u^n\langle x \rangle_u \quad . \tag{3.100}$$

The proof of (3.100) is easily obtained by partial integration,

$$\int_u^\infty dx\; nx^{n-1}e^{-x^2} = u^n e^{-u^2} + 2\int_u^\infty dx\; x^{n+1}e^{-x^2}\;. \tag{3.101}$$

It reduces all efforts to calculate $\langle 1\rangle_u$ and $\langle x\rangle_u$. One notes that $\langle x^n\rangle_u$ for odd n could be expressed in terms of $\langle x\rangle_u$ only, while even n depend on $\langle 1\rangle_u$ and $\langle x\rangle_u$. Only for $u=0$ they become independent from $\langle x\rangle_u$.

The polynomials p_i and the function h_i can now be written down explicitly for some i. After some tedious algebra, one finds

$$\begin{aligned} p_0(x,u) &= 1; & h_0(u) &= \langle 1\rangle_u \\ p_1(x,u) &= x - \tfrac{\langle x\rangle_u}{\langle 1\rangle_u}; & h_1(u) &= \langle x^2\rangle_u - \tfrac{\langle x\rangle_u^2}{\langle 1\rangle_u} \\ p_2(x,u) &= x^2 + \tfrac{\langle x^3\rangle_u\langle 1\rangle_u - \langle x^2\rangle_u\langle x\rangle_u}{\langle x\rangle_u^2 - \langle x^2\rangle_u\langle 1\rangle_u}x + \tfrac{\langle x^2\rangle_u^2 - \langle x^3\rangle_u\langle x\rangle_u}{\langle x\rangle_u^2 - \langle x^2\rangle_u\langle 1\rangle_u}; \end{aligned} \tag{3.102}$$

and

$$h_2(u) = \frac{\langle x^2\rangle_u^3 + \langle 1\rangle_u\langle x^3\rangle_u^2 + \langle x\rangle_u^2\langle x^4\rangle_u - 2\langle x\rangle_u\langle x^2\rangle_u\langle x^3\rangle_u - \langle 1\rangle_u\langle x^2\rangle_u\langle x^4\rangle_u}{\langle x\rangle_u^2 - \langle 1\rangle_u\langle x^2\rangle_u} \tag{3.103}$$

3.4.3 Determining $\sigma(t,\tilde{x};x_0)$

In order to find a recursion relation for the a_i in (3.85) $u=\mu(t)$ is chosen and the total differential with respect to time is calculated:

$$\begin{aligned} \frac{d}{dt}\sigma(t,\tilde{x};x_0) &= \sum_{i=0}^\infty \dot{a}_i(t,\mu(t);x_0)p_i(\tilde{x},\mu(t)) + \dot{\mu}(t)\partial_u\big|_{u=\mu(t)}\Big(\sum_{i=0}^\infty \dot{a}_i(t,u;x_0)p_i(\tilde{x},u)\Big) \\ &= \sum_{i=0}^\infty \dot{a}_i p_i + \dot{\mu}(t)\sum_{i=0}^\infty \big(\partial_u\big|_{u=\mu(t)}a_i(t,u;x_0)\big)p_i(\tilde{x},u) \qquad (3.104a) \\ &\quad + \sum_{i=0}^\infty a_i(t,\mu(t);x_0)\big(\partial_u\big|_{u=\mu(t)}p_i(\tilde{x},u)\big) \\ &= \sum_{i=0}^\infty \big(\frac{d}{dt}a_i(t,\mu(t);x_0)\big)p_i(\tilde{x},u) + \dot{\mu}\sum_i a_i\partial_u p_i \qquad (3.104b) \end{aligned}$$

where $\dot{a}_i$ denotes the partial derivative of a_i by t and $\dot{\mu}(t) \equiv \frac{d}{dt}\mu(t)$ as already used above.

Using (3.95) and adding $\dot{\mu}\partial_{\tilde{x}}\sigma$ to (3.104b) gives $\sum_i a_i(\partial_u + \partial_{\tilde{x}})p_i$ on the RHS, so

$$\partial_t\sigma + \dot{\mu}\partial_{\tilde{x}}\sigma = \sum_i \big(\frac{d}{dt}a_i\big)p_i + 2\dot{\mu}\sum_i \beta_i a_i p_{i-1} \tag{3.105}$$

Note that although $\partial_t\sigma = \sum \dot{a}_i p_i$ these terms do not cancel out, as $\frac{d}{dt}a_i$ contains also a derivative by u. One should stress that this equation is of course only true for $u = \mu(t)$.

The second expression needed is $(\tilde{x} - \mu)\partial_{\tilde{x}}\sigma$, which can be determined by means of (3.96):

$$(\tilde{x} - \mu)\partial_{\tilde{x}}\sigma \quad = \quad \sum_i a_i \big(ip_i + 2\beta_i\,(\alpha_i + \alpha_{i-1} - \mu)p_{i-1} + 2\frac{h_i}{h_{i-2}}p_{i-2}\big) \tag{3.106}$$

Using (3.89), one arrives at

$$2t\partial_t\sigma \quad = \quad (\tilde{x} - \mu - 2t\dot{\mu})\partial_{\tilde{x}}\sigma \tag{3.107a}$$

$$\Leftrightarrow$$

$$2t\Big(\partial_t\sigma + \dot{\mu}\partial_{\tilde{x}}\sigma\Big) \quad = \quad (\tilde{x} - \mu)\partial_{\tilde{x}}\sigma \tag{3.107b}$$

$$\Leftrightarrow$$

$$2t\Big(\sum_i (\frac{d}{dt}a_i)p_i + 2\dot{\mu}\sum_i \beta_i a_i p_{i-1}\Big) \quad = \quad \sum_i a_i\big(ip_i + 2\beta_i\,(\alpha_i + \alpha_{i-1} - \mu)p_{i-1} + 2\frac{h_i}{h_{i-2}}p_{i-2}\big) \tag{3.107c}$$

Calculating the projection $\langle\cdot|i\rangle_u$ for the last two lines gives

$$2t\left(\frac{d}{dt}a_i\right) + 4t\dot{\mu}\beta_{i+1}a_{i+1} = ia_i + 2\beta_{i+1}(\alpha_{i+1} + \alpha_i - \mu)a_{i+1} + 2\frac{h_{i+2}}{h_i}a_{i+2} \tag{3.108}$$

However, it is more comfortable to work with the variable $\hat{a}_i(t,\mu(t)) \equiv a_i(t,\mu(t))h_i(\mu(t))$ so that (3.88) becomes

$$\hat{a}_0(t,\mu(t);x_0) = g(t;x_0) \tag{3.109}$$

and relation (3.108) finally becomes

$$2t(\frac{d}{dt}\hat{a}_i) + 4t\dot{\mu}(\hat{a}_{i+1} + \alpha_i\hat{a}_i) = i\hat{a}_i + 2(\alpha_{i+1} + \alpha_i - \mu)\hat{a}_{i+1} + 2\hat{a}_{i+2} \tag{3.110}$$

and specifically for $i = 0$:

$$2t(\frac{d}{dt}\hat{a}_0) + 4t\dot{\mu}(\hat{a}_1 + \alpha_0\hat{a}_0) = 2(\alpha_1 + \alpha_0 - \mu)\hat{a}_1 + 2\hat{a}_2 \tag{3.111}$$

It is important to stress that all a_i are of the form $a_i(t,\mu(t))$. Especially a_0 on the LHS, which is the only known one (3.88).

Eq. (3.110) and Eq. (3.111) seem to solve the problem in form of a recurrence relation for the a_i; whatever σ is, it must obey the recurrence relation. However, surprisingly it involves two unknown functions, a_1 and a_2.

3.4.4 Discussion

The final result, especially in the form (3.111) seems to suggest that one only needs to choose a_1 and a_2 to find a solution to the original problem. However, that turns out to be wrong. In fact, one can show that

$$\partial_u a_i(t,u) = -2\beta_{i+1} a_{i+1}(t,u) \tag{3.112}$$

so that the "solution" (3.111) can be brought in the form

$$\frac{d}{dt} a_0 = \left(\dot{\mu} - \frac{\alpha_0 + \alpha_1 - \mu}{2t}\right) \partial_u a_0 + \frac{1}{4t} \partial_u^2 a_0 \tag{3.113}$$

which is to be solved with boundary condition (3.109) in the domain $u > \mu(t)$. At first glance it is surprising that this fairly involved approach finally yields essentially the original problem. However, such a relation actually follows trivially from the observation that the contribution to the full solution (3.73) is essentially

$$\int_0^\infty dx' s(x'; x_0) \frac{1}{\sqrt{4\pi t}} e^{-\frac{(x+x'-W(t))^2}{4t}}$$

which can be rewritten as

$$\int_{\frac{x-W}{2\sqrt{t}}}^\infty dx' s(2\sqrt{t}(\tilde{x} - \frac{x-W}{2\sqrt{t}}); x_0) e^{-\tilde{x}^2}$$

In fact, this is *identical* to $\hat{a}_0(t, \frac{x-W}{2\sqrt{t}}; x_0)$, *i.e.* there is a very simple relation between the full solution ϕ of the initial PDE (3.59) and the first coefficient a_0 in the series (3.85) of the source term used in (3.73).

Nevertheless, the technique introduced above seems to provide an interesting path to tackle problems of the form (3.59). What makes it so promising is the fact that the problem strongly benefits from the many generally known properties of orthogonal polynomials and parametrised weighting functions (Szegö, 1975). It might be worthwhile to pursue the approach presented above a little bit further and see, which non-trivial information can be derived from it.

Open issue

3.5 Summary

This chapter is more diverse than the others. Conservation has been identified early in the research literature as an important feature in SOC. The next chapter is dedicated solely to a single model, a characteristic of which is non-conservation.

The chapter above contains the following sections;

- The first section, Sec. 3.1, contains a short account of the first LANGEVIN-based approaches to SOC. The key message is that a wide class of toy models generically develop into a state which can justifiably be called SOC. These toy models are expressed as LANGEVIN-equations involving a scale-free, thermal noise.

- The section on the OFC model, Sec. 3.2, contains some arguments about the possible scale invariance of the OFC model. It represents unfinished research, which requires further numerical as well as analytical insight.
- In Sec. 3.3 a solvable random-neighbour model is presented and subsequently solved. In this section, one of the arguments brought forward by VESPIGNANI *et al.*, is substantiated.
- In the supplemental section Sec. 3.4 an (disputably) elegant mathematical method for solving random walker problems is applied to the open question of the finite size behaviour of the solvable model.

Chapter 4

The Forest Fire Model

The DROSSEL-SCHWABL FOREST FIRE MODEL *is one of the best studied models of non-conservative self-organised criticality. However, using a new algorithm which allows us to study the model on large statistical and spatial scales, it has been shown to lack simple scaling. We thereby show that the considered model is not critical. This chapter presents the algorithm and its parallel implementation in detail, together with large scale numerical results for several observables. The algorithm can easily be adapted to related problems such as percolation.*

4.1 Introduction

The assumption that SOC (Jensen, 1998) is the correct framework to describe and explain the ubiquity of power laws in nature, has been greatly supported by the development of non-conservative models, because natural processes are typically dissipative. Contrary to these models, analytical work has suggested, that the deterministic part of the dynamics must be conservative in order to obtain scale invariance (Hwa and Kardar, 1989; Grinstein *et al.*, 1990). However, on a mean field level, this is not necessarily true (Vespignani and Zapperi, 1998), which has been exemplified in an exact solution of a model, that has a forest fire-like driving (Pruessner and Jensen, 2002b). However, as a random neighbour model, the latter lacks spatial extension.

The DROSSEL-SCHWABL Forest Fire model (DS-FFM) (Drossel and Schwabl, 1992) is one of the few spatially extended, dissipative models, which supposedly exhibit SOC. Contrary to the OLAMI-FEDER-CHRISTENSEN stick-slip model (Olami *et al.*, 1992), where criticality is still disputed [for recent results see for example (Lise and Paczuski, 2001b,a; Boulter and Miller, 2003)], for the DS-FFM the asymptotic divergence of several moments of its statistics, and therefore the divergence of an upper cutoff can be shown rigorously. Although this might be considered as a sign of criticality, it is far from being a sufficient proof. In equilibrium thermodynamics "criticality” usually refers to a divergent correlation length (Binney *et al.*, 1998; Stanley, 1971) in the two-point correlation function, which is associated with a scale-invariant or power law like behaviour. This is how the term “criticality” is to be interpreted in SOC: Observables need to be scale-invariant[1], i.e. power laws in the statistics. There are many examples of divergent moments without scale invariance, such as the over critical branching process (Harris, 1963) or over critical percolation (Stauffer and Aharony, 1994).

Thus, there is *a priori* no reason to assume that the DS-FFM is scale free. However, there are many numerical studies, which suggest so (Drossel and Schwabl, 1992; Christensen *et al.*, 1993; Clar *et al.*, 1994), one of them, however, suggests the breakdown of simple scaling (Grassberger, 1993). Since an analytical approach is still lacking, numerical methods are required to investigate this problem. In this chapter, we propose a new, very fast algorithm to simulate the DS-FFM with large statistics and on large scales. The implementation of the algorithm, has produced data of very high statistical quality. Some of the results have been already published elsewhere (Pruessner and Jensen, 2002a).

The structure of the chapter is as follows: The next section contains the definition of the model together with its standard observables and their relations. Then the algorithm is explained in detail. The section finishes with a detailed discussion on the changes necessary to run the algorithm on parallel or distributed machines. In the third section results for the two-dimensional FFM are presented and analysed. The chapter concludes with a summary in the fourth section.

4.2 Method and Model

This section is mainly technical: After defining the model, all relevant details of the implementation are discussed. Apart from concepts such as the change from a tree oriented algorithm to a cluster oriented algorithm, concrete technical details are given, for example memory requirements and methods for handling histograms. The section also contains a description of the performance analysis of the implementation. A parallelised version of the algorithm is introduced an discussed in the last section.

4.2.1 The model

A Forest Fire model was first proposed by BAK, CHEN and TANG (Bak *et al.*, 1990) and changed later by DROSSEL and SCHWABL (Drossel and Schwabl, 1992) to what is now known as *the* Forest Fire Model (or DS-FFM as we call it): On a d dimensional

[1] In a finite system the distributions are not expected to be free of any scale, but to be dominated asymptotically by one scale only, which diverges in the thermodynamic limit.

lattice of linear length L, each site has a variable associated with it, which indicates the state of the site. This can either be "occupied" (by a tree), "burning" (occupied by a fire) or "empty" (ash). In each time step, all sites are updated in parallel according to the following rules: If a site is occupied and at least one of its neighbours is burning, it becomes burning in the next time step. If a site is occupied and none of its neighbours is burning, it becomes burning with probability f. If a site is empty, it becomes occupied with probability p. If a site is burning it becomes empty in the next time-step with probability one. As these probabilities become very small, they are better described as rates in a Poisson like process. From a simple analysis it is immediately clear (Clar *et al.*, 1994), that the model can become critical only in the limit $p \to 0$ and $f \to 0$. In this limit, the burning process becomes instantaneous compared to all other processes (see also Sec. 4.2.2.2) and can be represented by the algorithm shown in Fig. 4.1.

```
FOREVER {
   /* Choose a site randomly */
   rn = random site;
   /* If empty occupy with probability p */
   IF (rn empty) THEN {
      with probability p: rn=occupied;
   } ELSE {
   /* If occupied start a fire with probability f */
      with probability f:
         burn entire cluster connected to rn;
   }
}
```

Figure 4.1: The naïve, basic algorithm of the DS-FFM

Compared to the instantaneous burning, both of the remaining processes are slow. In Sec. 4.2.2.2 it is shown that $p \gg f$ is required (Clar *et al.*, 1994) for criticality, so that $f/p < 1$ and the algorithm in Fig. 4.1 can be written as Fig. 4.2, which is faster than the former, because the number of random choices of a site is reduced, but equivalent otherwise.

The line `with probability p` makes sure that the occupation attempt still happens with probability p and the burning attempt still occurs with $pf/p = f$. Of course, the line is completely meaningless, because the alternative, which occurs with probability $1 - p$ is no action at all. It therefore can be omitted. Then every randomly picked empty site will become occupied, while burning happens with the reduced probability f/p.

This rescaling of probabilities is only possible in this form if the two processes are independent, which is the case because a new occupation can only occur for empty sites, while a burning attempt operates only on occupied sites. If both processes were to operate on the same type of site, a reduced probability $(1 + f/p)^{-1}$ would decide

```
FOREVER {
   /* The following line is without effect */
   with probability p: {
      rn = randomly chosen site;
      IF (rn empty) THEN {
         rn=occupied;
      } ELSE {
         with probability f/p:
            burn entire cluster connected to rn;
      }
   }
}
```

Figure 4.2: A faster algorithm, doing essentially the same as the one shown in Fig. 4.1.

between the two alternatives.

The implementation shown in Fig. 4.2 (without the meaningless line) has been used for example in (Honecker and Peschel, 1997; Henley, 1993). However, probably for historical reasons, the model is usually (Grassberger, 1993; Clar *et al.*, 1994; Schenk *et al.*, 2000) implemented as shown in Fig. 4.3, where trees are grown in chunks of p/f between two lightning attempts. Although this means that sites become re-occupied only in chunks of p/f, it turns out that apart from peaks in the histogram of the time series of global densities of occupied sites (Schenk *et al.*, 2000), the statistics do not depend on these details. Only in order to avoid any confusion, all data for this chapter have been produced by means of the algorithm in Fig. 4.3. Moreover this algorithm is much more suitable for parallelisation (see Sec. 4.2.5).

```
FOREVER {
   /* This is just a loop to occupy the
    * right number of sites */
   REPEAT p/f TIMES {
      rn = randomly chosen site;
      IF (rn empty) THEN {rn=occupied;}
   }
   rn = randomly chosen site;
   IF (rn occupied) THEN {
      burn entire cluster connected to rn;
   }
}
```

Figure 4.3: The traditional implementation.

4.2.2 Statistical quantities

The objects of interest in the DS-FFM are clusters formed by occupied sites: Two trees belong to the same cluster, if there exists a path between them along nearest neighbouring, occupied sites. The cluster in the DS-FFM correspond to avalanches in sandpile-like models (Jensen, 1998). The cluster, which is burnt at each burning step can be examined more closely, so that various geometrical properties can be determined either as averages (and higher moments) or as entire distribution: Mass (in the following this term is used synonymously to size), diameter, time to burn it etc. The last property is better expressed as the maximum length for all paths parallel to the axes and fully within the given cluster, connecting the initially burnt tree and each tree within the same cluster. It is the maximum number of nearest neighbour moves one has to make to reach all sites in the same cluster, in this sense a "Manhattan distance" (Cormen *et al.*, 1990, pp. 194, 912). As trees catch fire due to nearest neighbours only, this maximum distance is the total burning time of the entire cluster. In the definition above, the "time to burn" T_{M} becomes a purely geometrical property of the cluster and therefore independent from the actual implementation (see sec. 4.2.3.4) of the burning procedure.

4.2.2.1 Cluster size distribution

The most prominent property of the model, however, is the size distribution of the clusters, $\bar{n}(s)$, which is the single-site normalised number density of clusters of mass s, i.e. the number of clusters of size s per unit volume. The average cluster size, i.e. the average size of a cluster a randomly chosen occupied site belongs to, is correspondingly defined as

$$\widetilde{s} = \frac{\sum_s s^2 \bar{n}(s)}{\sum_s s\bar{n}(s)} \quad . \tag{4.1}$$

As indicated by the bar, $\bar{n}(s)$ denotes the *expected* distribution, i.e. something to be *estimated* from the observables. On average, the probability that a randomly chosen site belongs to a cluster of size s is then $s\bar{n}(s)$. If $n_t(s)$ denotes the cluster size distribution of the configuration at time t (see below), then one expects

$$\langle n_t(s) \rangle = \bar{n}(s) \quad . \tag{4.2}$$

where $\langle\rangle$ denotes the ensemble average (as opposed to $\widetilde{\ }$, which denotes the average over $\bar{n}(s;\theta)$). Assuming ergodicity, one has

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} A_t \to \langle A \rangle \tag{4.3}$$

for an arbitrary quantity A_t measured at each step t of the simulation. The limit exists for all bound observables A_t.

Regarding the time t, it is worth noting that a step in the simulation is considered completed, i.e. $t \to t+1$ if the randomly chosen site for the lightning attempt was occupied, i.e. the attempt was successful, so that T is the number of burnt clusters.

For sufficiently large systems, the changes of the system due to growing or lightning are almost negligible, and so are the differences between averages taken over all lightning attempts or all *successful* lightning attempts. Also, the distributions found directly before and directly after burning tend to the same expectation value for sufficiently large systems, see sec. 4.3.3.1. It is noted only for completeness, that in this chapter the cluster size distribution $n_t(s)$ has been measured directly *after* the burning procedure. Therefore $n_t(s)$ does not include the cluster burnt at time step t, just like $n_{t+1}(s)$ does not in an implementation, where the distribution is measured *before* burning.

Introducing
$$\bar{\rho} = \sum_{s=1} s\bar{n}(s) \tag{4.4}$$
as average density of occupied sites, the expected distribution of burnt clusters is $s\bar{n}(s)/\bar{\rho}$. To see this, $\mathcal{P}_t^{\mathrm{b}}(s)$ is introduced, denoting the distribution of clusters burnt in the tth step of the simulation. This distribution contains only one non-zero value for each t, namely $\mathcal{P}_t^{\mathrm{b}}(s) = 1$ for the size s of the cluster burnt at time t, and $\mathcal{P}_t^{\mathrm{b}}(s) = 0$ for all other s. Therefore
$$\sum_{s=1}^{N} \mathcal{P}_t^{\mathrm{b}}(s) = 1 \tag{4.5}$$
where N is the number of sites in the system, $N = L^d$, which is also the maximum mass of a cluster. Since the site where the fire starts is picked randomly, the cluster burnt in time step $t+1$ is drawn randomly from the distribution $n_t(s)$ with a probability proportional to the mass of the cluster. The normalisation of the distribution $s\bar{n}(s)$ is given by (4.4), so that for t large enough, the effect of the initial condition can be neglected,
$$\left\langle \mathcal{P}_t^{\mathrm{b}}(s) \right\rangle = s\bar{n}(s)/\bar{\rho} \, . \tag{4.6}$$

In the stationary state the average number of trees, $\bar{\rho}$ is related to $\widetilde{s}$ by (Clar *et al.*, 1994)
$$\widetilde{s} = \frac{1-\bar{\rho}}{\theta\bar{\rho}} \, . \tag{4.7}$$
This equation, as well as (4.6), is strictly only exact if the density of occupied sites is constant over the course of the growing phase. For very large system sizes (4.7) holds almost perfectly, as shown in Tab. 4.3 (see page 185); however, note the remarks in Sec. 4.3.3.1.

For a coherent picture $\mathcal{P}_t^{\mathrm{a}}(s)$ is introduced, which is the histogram of *all* clusters, i.e. $\sum_s \mathcal{P}_t^{\mathrm{a}}(s)$ is the number of clusters in the system at time t. According to the definition of $\bar{n}(s)$ it is
$$\langle \mathcal{P}_t^{\mathrm{a}}(s) \rangle = N\bar{n}(s) \quad , \tag{4.8}$$

and correspondingly

$$\rho_t = \frac{1}{N}\sum_s s\mathcal{P}_t^{\mathrm{a}}(s) \tag{4.9}$$

with $\langle \rho_t \rangle = \bar{\rho}$. Since (4.6) and (4.8) differ on the RHS only by constants rather than by random variables, both distribution, $\mathcal{P}_t^{\mathrm{b}}(s)$ and $\mathcal{P}_t^{\mathrm{a}}(s)$, are estimators of the expected distribution $\bar{n}(s)$. Clearly, the burnt cluster distribution $\mathcal{P}_t^{\mathrm{b}}(s)$ is much sparser than than $\mathcal{P}_t^{\mathrm{a}}(s)$ and the estimator for $\bar{n}(s)$ derived from this quantity, is therefore expected to have a significant larger standard deviation. On the other hand, its autocorrelation time is expected to be considerably smaller than that of $\mathcal{P}_t^{\mathrm{a}}(s)$, because on average only $p/f+1$ entries ($\bar{\rho}p/f$ sites are occupied in each "growing loop", which is repeated on average $1/\bar{\rho}$ times) of the latter are changed between two subsequent measurements, corresponding to the number of newly occupied sites plus the cluster which is burnt down. So, $\mathcal{P}_t^{\mathrm{a}}(s)$ provides a much larger sample size, but is also expected to be much more correlated. In order to judge, whether it is wise to spend CPU time on calculating the full $\mathcal{P}_t^{\mathrm{a}}(s)$ rather than only $\mathcal{P}_t^{\mathrm{b}}(s)$, as it was done in the past (Clar *et al.*, 1994), these competing effects need to be considered, by calculating the estimate for the standard deviation of the estimator of $\bar{n}(s)$ from both observables, which is discussed in detail in Sec. 4.2.4.

4.2.2.2 Timescales

In order to obtain critical behaviour in the FFM, a double separation of timescales is required (Clar *et al.*, 1996)

$$f \ll p \ll \left(\frac{f}{p}\right)^{\nu'} \quad , \tag{4.10}$$

with some positive exponent ν'. The left relation, $f \ll p$, entails $f/p \to 0$ and therefore (4.10) entails $p \to 0$ and $f \to 0$. This is also the case for

$$f \ll p \ll 1 \quad , \tag{4.11}$$

and therefore leads to the same prescription to drive the system, however (4.10) entails (4.11) but not vice versa. This can be seen by noting that (4.10) entails the non-trivial relation $p^{1+1/\nu'} \ll f \ll p$. Some authors, however, just state (4.11) (Grassberger, 1993; Vespignani and Zapperi, 1998). The three scales involved are due to three different processes and their corresponding rates:

1. The timescale on which the burning happens, the typical time of which is handwavingly estimated as the average number of sites in a burnt cluster, $\widetilde{s} \propto p/f$. A more appropriate assumption is that the typical burning timescales like a power of the average cluster size (Clar *et al.*, 1996). This should be

distinguished from the scaling of the *average* time it takes to burn a cluster, because the *typical* time represents the characteristic scale of the burning time distribution, which might be very different from its average.

2. The timescale of the growing, which is $1/p$.

3. The timescale of the lightning, $1/f$.

Burning must be fast compared to growing, so that clusters are burnt down, before new trees grow on it edges (Clar *et al.*, 1996), i.e. $(p/f)^{\nu'} \ll 1/p$ or $(f/p)^{\nu'} \gg p$. In order to obtain divergent cluster sizes, growing must be much faster than lightning, i.e. $p \gg f$. Thus, the double separation reads as stated in (4.10). By making the burning instantaneous compared to all other processes, the dynamics effectively loses one timescale. In this case, the rates f and p, measured on this microscopic timescale, vanish, i.e. $f = 0$ and $p = 0$, so that the right relation of (4.10) is perfectly met, provided that p/f does not vanish. However, the ratio f/p remains finite, and $f \ll p$ is still to be fulfilled. A finite f/p means that one rate provides a scale for the other. Measuring the rates on the macroscopic timescale, defined by the sequence of burning attempts, f becomes 1 in these new unities, and p becomes $p/f \equiv \theta^{-1}$. The notation $\theta = f/p$ corresponds to (Vespignani and Zapperi, 1998), which is, unfortunately, the inverse of θ used in (Grassberger, 1993). Eq. (4.10) then means $\theta \to 0$. At first sight, this result seems paradoxical, since $\theta = 0$ is incompatible with instantaneous burning's compliance with $p \ll \theta^{\nu'}$. However, this problem does not appear in the *limit* $\theta \to 0$. In a finite system, one cannot make θ arbitrarily small, as the system will asymptotically oscillate between the two states of being completely filled and completely empty. On the other hand, for fixed θ and sufficiently large system sizes, a further increase in system size will leave the main observables, such as ρ_t and $\mathcal{P}^{\mathrm{a}}$ (see Sec. 4.2.2.1), essentially unchanged. These asymptotic values, namely the observables at a given θ in the thermodynamic limit, are to be measured.

4.2.2.3 Scaling of the cluster size distribution

Assuming that finite size effects do not play any rôle, i.e. for θ not too small, the ansatz

$$\bar{n}(s;\theta) = s^{-\tau}\mathcal{G}(s/s_0(\theta)) \tag{4.12}$$

as obtained in percolation (Stauffer and Aharony, 1994) is reasonable for s larger than a fixed lower cutoff. In the following, the additional parameter θ in $\bar{n}(s;\theta)$ is omitted, whenever possible. The quantity $s_0(\theta)$ is the upper cutoff and supposed to incorporate all θ dependence of the distribution. It can be shown easily (Clar *et al.*, 1994) that the second moment of $\bar{n}(s;\theta)$ [see (4.1)] diverges in the limit $\theta \to 0$ and $L \to \infty$, so that s_0 must diverge with $\theta \to 0$. Here, $\mathcal{G}(x)$ plays the rôle of a cutoff function, so that $\lim_{x\to\infty} \mathcal{G}(x) = 0$ and falls off faster than any power for large x, because all moments of $\bar{n}(s;\theta)$ are finite in a finite system. For finite x, $\mathcal{G}(x)$ can show any structure and does not have to be constant. However, assuming $\lim_{s_0\to\infty} \bar{n}(s;\theta)$ finite, $\mathcal{G}(s/s_0)$ can be regarded as constant in s for sufficiently large s_0, so that $\bar{n}(s;\theta)$ behaves like a power law, $s^{-\tau}$, for certain s. However, *a priori* it is completely unknown, whether s_0 is large enough in that sense and the *only* way to determine τ directly from $\bar{n}(s;\theta)$ is via a data collapse. It is already known that "simple scaling" (4.12) does not apply in the presence of finite size effects (Schenk *et al.*, 2000).

The assumption (4.12) states that the FFM is scale-free in the limit $s_0(\theta) \to \infty$ and *defines* the exponent τ which characterises the scale invariance. One cannot stress enough, that with the breakdown of (4.12), the proposed exponent is undefined, unless a new scaling behaviour is proposed. It has been pointed out that (4.12) certainly contains corrections (Pastor-Satorras and Vespignani, 2000a). This asymptotic character of the universal scaling function is well known (Wegner, 1972) from equilibrium critical phenomena.

While GRASSBERGER concludes that the ansatz (4.12) "cannot be correct" (Grassberger, 1993), this is rejected in (Schenk *et al.*, 2000). However, the latter authors do not actually investigate $\mathcal{G}(x)$ and simply plot their estimate of $s\bar{n}(s;\theta)$ vs. $s/s_0(\theta)$. In the result section it is shown that there is no reason to believe that (4.12) could hold in any finite system.

4.2.2.4 Other distributions

The exponent τ as defined in (4.12) can be related to exponents of other assumed power laws. To this end, the distribution $\mathsf{P}(s, T_\mathrm{M};\theta)$ is introduced, which is the joint probability density function (PDF), for a cluster burnt to be of mass s and burning time (see sec. 4.2.2) T_M at given θ. Then it is possible to define conditional expectation values as (Christensen *et al.*, 1991).

$$\mathsf{E}(s|T_\mathrm{M};\theta) = \sum_{s'} s'\mathsf{P}(s', T_\mathrm{M};\theta) \tag{4.13}$$

$$\mathsf{E}(T_\mathrm{M}|s;\theta) = \sum_{T_\mathrm{M}'} T_\mathrm{M}'\mathsf{P}(s, T_\mathrm{M}';\theta) \quad . \tag{4.14}$$

Moreover it is clear that $\bar{n}(s;\theta)$ is just a marginal distribution, i.e.

$$s\bar{n}(s;\theta) = \sum_{T_{\mathrm{M}}'} \mathsf{P}(s, T_{\mathrm{M}}';\theta) \equiv \mathsf{P}_s(s;\theta) \quad . \tag{4.15}$$

In the assumed absence of any scale, it is reasonable to define for the distribution of T_{M} similar to (4.12)

$$\mathsf{P}_{T_{\mathrm{M}}}(T_{\mathrm{M}};\theta) = T_{\mathrm{M}}{}^{-b}\mathcal{G}_{T_{\mathrm{M}}}(T_{\mathrm{M}}/T_{\mathrm{M}0}(\theta)) \tag{4.16}$$

and for the relation between $\mathsf{E}(s|T_{\mathrm{M}})$ and T_{M}:

$$\mathsf{E}(s|T_{\mathrm{M}}) \propto T_{\mathrm{M}}{}^{\mu'} \tag{4.17}$$

To avoid confusion, it is important to keep in mind that the absence of scales is not a physical or mathematical necessity: The system could as well "self-organise" to any other, sufficiently broad distribution, which could have an intrinsic, finite scale, i.e. a natural constant characterising the features of the distribution. This looks much less surprising considering the fact that standard models of critical phenomena (Stanley, 1971) like the ISING model, possess such a scale everywhere apart from the critical point.

An additional assumption is necessary in order to produce a scaling relation:

$$\mathsf{P}_{T_{\mathrm{M}}}(T_{\mathrm{M}};\theta)dT_{\mathrm{M}} = \mathsf{P}_s(\mathsf{E}(s|T_{\mathrm{M}});\theta)d(\mathsf{E}(s|T_{\mathrm{M}};\theta)) \tag{4.18}$$

where $\mathsf{P}_{T_{\mathrm{M}}}$ and P_s denote the marginal distributions of $\mathsf{P}(s, T_{\mathrm{M}};\theta)$, which leads — assuming sufficiently large s_0 and $T_{\mathrm{M}0}$ — to

$$b = 1 + \mu'(\tau - 2) \tag{4.19}$$

using $\mathsf{P}_s = s\bar{n}(s;\theta)$ and (4.12). Eq. (4.18) is based on the idea that a cluster requiring burning time T_{M} is as likely to occur as a cluster of the size corresponding to the average taken conditional to the burning time T_{M}. If the distribution $\mathsf{P}(s, T_{\mathrm{M}};\theta)$ is very narrow, such that $\mathsf{E}(s|T_{\mathrm{M}})$ is virtually the only value of s with non-vanishing probability[2], this condition is met. However, the distribution can have any shape and still obey the assumption, as illustrated in Fig. 4.4.

Scaling relation (4.19) can only be derived via (4.18), which cannot be mathematically correct, as P_s is actually only defined for integer arguments, while in general $\mathsf{E}(s|T_{\mathrm{M}})$ is not integer valued. However, the scaling relation might hold in some limit.

[2]The extreme case would be $\mathsf{P}(s, T_{\mathrm{M}};\theta) = \delta(s - f(T_{\mathrm{M}}))g(T_{\mathrm{M}})$ with a monotonic function $f(T_{\mathrm{M}})$ representing the conditional average.

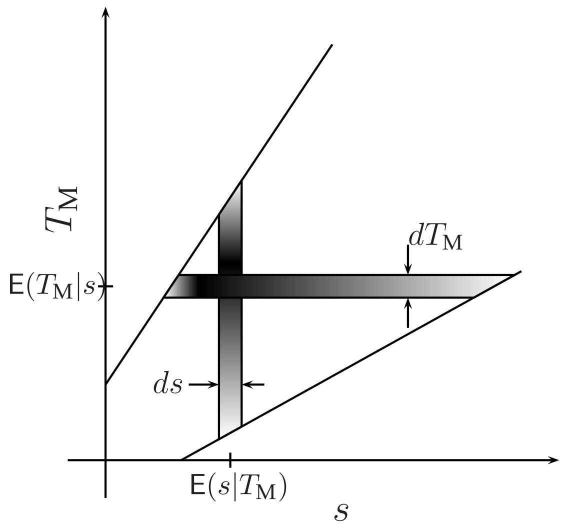


Figure 4.4: A schematic joint PDF $P(s, T_M'; \theta)$. The gray shading is used to indicate the density and the straight lines indicate roughly the limits of the distribution. While a narrower distribution would most easily obey (4.18), it does not necessarily have to be sharply peaked. In this example the weighted areas of the horizontal and the vertical stripes might be the same. They cross at the conditional averages.

The exponent defining the divergence of s_0 in (4.12) is defined as

$$s_0(\theta) = \theta^{-\lambda} \tag{4.20}$$

leading together with (4.1) and (4.7) to the scaling relation (Clar *et al.*, 1996)

$$\lambda(3 - \tau) = 1 \quad . \tag{4.21}$$

The corresponding exponent for T_{M0} in (4.16) as

$$T_{M0}(\theta) = \theta^{-\nu'} \tag{4.22}$$

The assumption $T_{M0} = E(T_M|s_0) \propto s_0^{1/\mu'}$ then gives the scaling relation

$$\nu' = \frac{\lambda}{\mu'} \tag{4.23}$$

It is interesting to note that this assumption is consistent with the assumption that clusters, which have a size of the order $s_0(\theta)$ need of the order T_{M0} time to burn. In that case one has $P_{T_M}(T_{M0}; \theta)dT_M = P_s(s_0; \theta)ds$ and as $T_{M0} \propto s_0^{\nu'/\lambda}$, one has using (4.16) and (4.12):

$$(1 - b)\frac{\nu'}{\lambda} = 2 - \tau \tag{4.24}$$

corresponding to (4.19) with (4.23).

4.2.3 The implementation

In this section a new implementation of the DS-FFM is discussed. An implementation especially capable to handle large scales has been proposed by HONECKER (Honecker, 1997) earlier. The most prominent feature of it is the bitwise encoding of the model, which significantly reduces memory requirements. Some of the properties investigated, profit from this scheme of bitwise encoding, because bitwise logical operators can be used to determine for example correlations, and operate on entire words "in parallel". However, in this implementation it would have been inefficient to count all clusters, i.e. $\bar{n}(s)$ is determined via $\mathcal{P}^{\mathrm{b}}(s)$ rather than $\mathcal{P}^{\mathrm{a}}(s)$.

In contrast to standard implementations (Clar *et al.*, 1994; Schenk *et al.*, 2000; Honecker and Peschel, 1997), where $\bar{n}(s)$ is derived from $\mathcal{P}^{\mathrm{b}}(s)$, the philosophy of the implementation presented in this chapter is to count *all* clusters efficiently by keeping track of their growing and disappearance, so that $\bar{n}(s)$ is derived from $\mathcal{P}^{\mathrm{a}}(s)$. By comparing the standard deviation of the estimates, and the costs (CPU time), the efficiency is found to be at least one order of magnitude higher. At the same time, the complexity of the algorithm is essentially unchanged, namely $\mathcal{O}(\theta^{-1}\log(N))$ instead of $\mathcal{O}(\theta^{-1})$, while a naïve implementation of the counting of all clusters is typically of order $\mathcal{O}(N)$. In the following the algorithm is described in detail. Because of its close relation to standard percolation, the algorithm presented below is also applicable for this classical problem of statistical mechanics. In fact, the percolation algorithm recently proposed by NEWMAN and ZIFF (Newman and Ziff, 2000, 2001) is very similar. Based on many principles presented in this chapter, an asynchronously parallelised version for percolation has been developed recently (Moloney and Pruessner, 2003; Pruessner and Moloney, 2003, 2004).

4.2.3.1 Tracking clusters

Usually each site is represented by a two-state variable, which indicates whether the site is occupied or empty. The variable does not need to indicate the state "burning", because the burning procedure is instantaneous compared to all other processes and can be implemented without introducing a third state (see sec. 4.2.3.4). In order to keep track of the cluster distribution, each site gets associated two further variables (in an actual implementation the number of variables can be reduced to one, see sec. 4.2.3.2), one which points (depending on the programming language either directly as an address or as an index) to its "representative" and one which contains the mass of the cluster the given site is connected to. The representative of a site is another site of the same cluster, but not necessarily and in fact typically not a nearest neighbour. This is shown in Fig. 4.5. If a site is empty, the pointer to a representative is meaningless. The pointer of representatives form a tree-like structure, because representatives might point to another representative, as shown in Fig. 4.6. A site which points to itself and is therefore its own representative, is called a “root” site, since it forms the root of the tree like structure. Only at a root site, the second variable, denoting the mass of the cluster, is actually meaningful and indicates the mass of the entire cluster. Each cluster is therefore uniquely identified by its root site: Any two sites, which belong to the same cluster have the same root and vice versa. By construction of the clusters (shown below), it takes less than $\mathcal{O}(\log N)$ to find the root of any site in the system.

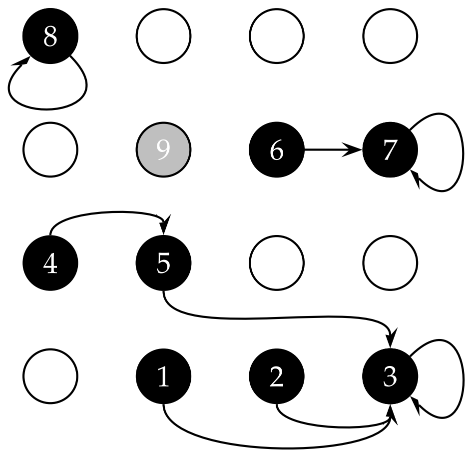


Figure 4.5: All occupied sites (black) on the lattice point to a representative. The site pointing to itself is the root of the cluster. The site shown in light gray is the one which is about to become occupied, as shown in Fig. 4.7. The labels on the sites are just to uniquely identify them in other figures.

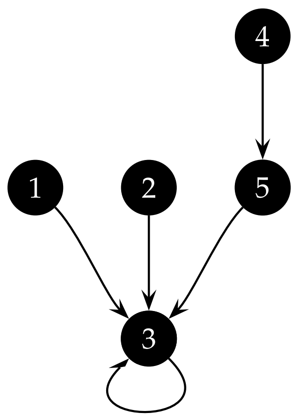


Figure 4.6: The tree-like structure of the largest cluster in Fig. 4.5.

The algorithm is a dynamically updated form of the HOSHEN-KOPELMAN algorithm (Hoshen and Kopelman, 1976). The same technique has recently been used to simulate percolation efficiently for many different occupations densities (Newman and Ziff, 2000). The method described in the following differs from (Newman and Ziff, 2000), by not only growing clusters, but also removing them. While one of the main advantages of the original HOSHEN-KOPELMAN algorithm is its strong reduction of memory requirements to $\mathcal{O}(L^{d-1})$, the algorithm described here only makes use of the data representation proposed by HOSHEN and KOPELMAN, so that the memory requirements are still $\mathcal{O}(L^d)$.

In the following the technique, how to create and to update the clusters, is described in detail.

Starting from an empty lattice, the first site becomes occupied by setting the state variable. Since this site cannot be member of a larger cluster, its representative is the site itself. Therefore the mass variable must be set to one. The same pattern applies to all other sites which get occupied, as long as they are isolated. The procedure becomes more involved, when a site induces a merging of clusters. This is the case whenever one or more neighbours of the newly occupied site are already occupied. In general the procedure is then as follows:

- Find the root of all neighbouring clusters.
- Reject all roots, which appear more than once in order to avoid double counting.
- Identify the largest neighbouring cluster.
- Increase the mass variable of the root of this cluster by the mass of all remaining clusters (ignoring those which have been rejected above) plus one (for the newly occupied site).
- Bend the representative pointers of the roots of all remaining clusters to point to the root of the largest cluster (keeps the tree height small, see below).
- Bend the representative pointers of the newly occupied site to point to the root of the largest cluster.

This procedure is depicted in Fig. 4.7, illustrating the join of the clusters shown in Fig. 4.5. As an optimisation, one could also bend the pointer of site 6 to point to site 3, which would effectively be a form of path compression. However, as shown below, the trees generated have only logarithmic height, so that the path compression possibly costs more CPU time than it saves for system sizes reachable with current computers[3]. It is important to note that only the root of the largest cluster is not redirected.

[3] Similarly for other forms of path compression, for example bending the pointer of the preceeding to the adjacent site in `find_root` (Fig. 4.8).

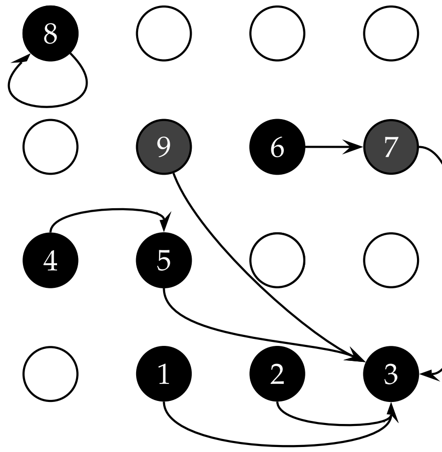


Figure 4.7: The configuration in Fig. 4.5 after occupying the highlighted site. Sites, the pointer of which have been changed, are shown in dark gray (site 6, 7 and 9).

```
/* Find the root of the cluster identified by start_index.
 * All sites are expected to have a pointer to their
 * representative in the array pointer_of. The result
 * is stored in index. */
index = start_index
WHILE ( index != pointer_of[index] ) {
  index=pointer_of[index]; }
```

Figure 4.8: The `find_root` algorithm. All sites are expected to have a pointer to their representative in the array `pointer_of`. The result of this procedure is `index`.

To find the root of a given site, which is necessary, whenever clusters are considered for merging, an algorithm like the one shown in Fig. 4.8 needs $\mathcal{O}(h_m(M(\mathcal{C})))$ time (worst case), where $h_m(M(\mathcal{C}))$ is the maximum height of a cluster containing $M(\mathcal{C})$ sites, $\mathcal{C}$ being the cluster under consideration.

All clusters are constructed by merging clusters, which might often involve single sites. These clusters are represented as trees, like the one shown in Fig. 4.6. In the following this representation is used. By construction, if at least two trees join, the resulting tree has either the height of the tree representing the largest cluster or the height of any of the smaller trees plus one — whatever is larger. Thus, by construction,

$$h_m(M) \geq h_m(M') \text{ for any } M \geq M' \quad , \tag{4.25}$$

so in order to find the maximum height of a tree of mass M, one has to consider the worst case when the smaller trees have maximum height. For a given, fixed M, this is the case when only two cluster merge, so

$$h_m(M) \leq \max\Big(\max_{M' \leq \lfloor \frac{M}{2} \rfloor} (h_m(M - M')), \max_{M' \leq \lfloor \frac{M}{2} \rfloor} (1 + h_m(M')) \Big) \quad , \tag{4.26}$$

where $\lfloor \frac{M}{2} \rfloor$ denotes the integer part of $M/2 \geq 0$, which is is the maximum size of the smaller cluster. The outer $\max$ picks the maximum of the two $\max$ running over all allowed sizes of the smaller cluster. Using (4.25),

$$h_m(M) \leq \max(h_m(M-1), 1 + h_m(\lfloor \frac{M}{2} \rfloor)) \tag{4.27}$$

so that

$$h_m(M) \leq \begin{cases} 1 + h_m(\lfloor \frac{M}{2} \rfloor) & \text{for } 1 + h_m(\lfloor \frac{M}{2} \rfloor) \geq h_m(M-1) \\ \\ h_m(M-1) & \text{otherwise} \end{cases} \tag{4.28}$$

With $h_m(1) = 1$ one can see immediately that

$$h_m(M) \leq \lceil \log_2(M) \rceil \tag{4.29}$$

by induction, nothing that $\lceil \log_2(M/2) \rceil = \lceil \log_2(M) \rceil - 1$, where $\lceil a \rceil \equiv \lfloor a \rfloor + 1$ for any $a \geq 0$. Hence

$$h_m(M) \in \mathcal{O}(\log(M)) \quad , \tag{4.30}$$

which is therefore the (worst case) complexity of the algorithm. It is worthwhile noting that all the algorithms considered are just one solution of the more general union-find (and also insert) problem (Cormen *et al.*, 1990).

As the tree constructed is directed, there is no simple way to find all sites which are pointing to a given site. This means that splitting trees is extremely expensive in terms of complexity. However, in the DS-FFM trees do not get removed individually, but always as complete clusters. Thus, no part of the tree structure needs to be updated during the burning (see Sec. 4.2.3.4).

4.2.3.2 Reducing memory requirements

The three variables (state, pointer, size) mentioned above would require a huge amount of memory: At least a bit for the state (but for convenience a byte), a word for the address and a word for the mass (actually depending on the maximum size of the clusters). However, as the pointers are only meaningful if the site is occupied, the representative pointer can also be used to indicate the state of a site: If it is 0 (or `NULL` if it is an address), the site is empty and occupied otherwise.

As mentioned above (Sec. 4.2.3.1), the mass variable is meaningful only at a root site. Since only a certain range of pointers is meaningful, the remaining range can be

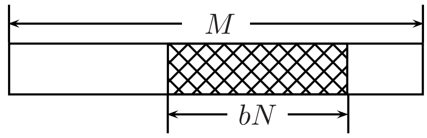


Figure 4.9: The memory layout when using addresses as pointers to representative. The hatched area is used for valid addresses, what remains left can be used to represent cluster masses, i.e. if the value of an address points into the white area, the value is interpreted as a mass.

used to indicate the mass of a cluster. Assuming that indeces can only be positive, negative numbers as the value of the pointer can be interpreted as self references and their modulus as total mass of the cluster. The concept is restricted to system sizes which are small enough that the space not occupied by meaningful pointers is large enough to store the mass information. How large is the maximum representable system size (not to be confused with memory requirements, which is N times word size)? For a word size of $b = 4$ byte, i.e. $M = 2^{8b}$ representable values in a word, the maximum system size is $N = 2^{31} - 1$, namely $-1 \cdots - N$ values for indicating masses, $1 \dots N$ for indeces and 0 for the empty site, summing up to $2N + 1 \leq M$, which is overruled by the memory required $bN \leq M$, as M is (usually) the maximal addressable memory for a single process.

When using addresses as pointers, it is less obvious how to identify the range of meaningless pointers which could be used to store the mass information. In order to distinguish quickly whether a given value is an address or a mass, the most obvious way is to use higher bits in the pointers. What is the range of meaningless addresses? The addresses are words, occupying bN bytes. If each byte is individually addressable (as usual), their value differs by at b, i.e. they span a range of bN different values. As shown in Fig. 4.9, the largest remaining continuous chunk of values, not used for references to representatives, has therefore at least size $\lceil (M - bN)/2 \rceil = (M - bN)/2$, assuming that the pointer values used, which is also the range of addresses where they are stored, spans a continuous range. If the $N + 1$ different cluster masses are to be represented as pointer values pointing into the meaningless region, one has $1 + N \leq (M - bN)/2$, i.e. $(b + 2)N + 2 \leq M$. If they do not have to be continuous, the condition is relaxed: $1 + N \leq M - bN$. Alternatively one can make use of the lower bits: If the pointers point to words in a continuous chunk of memory or at least are all aligned in the same way, then all pointers are identical $\pmod b$, i.e. all pointers p obey $p = c \pmod b$ where $0 \leq c < b$ is a constant. Since $b > 1$ one can use $p \neq c \pmod b$ to indicate that a given pointer value is to be interpreted as mass, which can easily be calculated via a bit-shift.

```
void *start_pointer, *root, *content;
/* start_pointer is the address of the site, the root
 * of which is to be found. root will always point to
 * the site currently under consideration, while content
 * is always the address root is pointing to.
 * The macro IS_SIZE verifies, whether the value given
 * is a size. */


/* Initialise: Assume that start_pointer is the root and
 * read its content. */
for (content=*((void **)(root=start_pointer));

/* Test whether root's content is actually a size */
    (!IS_SIZE(content));

/* Iterate: content is not a size, so the next candidate
 * is what root is currently pointing to.
 * content is updated accordingly*/
    content=*((void **)(root=content)));
```

Figure 4.10: An implementation of `find_root` in C using pointers to `void`.

In C it is reasonable to represent the sites as `void *` and interprete these as pointers to other sites, i.e. `void **`, so that the loop to search for a root just becomes the code shown in Fig. 4.10.

Representing each site by a word instead of a byte or even a bit (Honecker, 1997), still leads to reasonably small memory requirements for typical system sizes (for instance a system of size $N = 4096 \times 4096$ would require $64\,MB$). Since the algorithm has an almost random memory access pattern, it is not reasonable to implement it out of core (Dowd and Severance, 1998). In order to simulate even larger sizes, the following representation has been implemented: At the beginning of the simulation the entire lattice is splitted in cells so that whatever site in such a cell is occupied, it must belong to the same cluster as any other occupied site in the same cell, i.e. each site in the cell is nearest neighbour of all other sites. On an hyper-cubic lattice these cells have size 2, as depicted in Fig. 4.11: Each site within such a cell must belong to the same cluster if it is occupied. Therefore only one pointer is necessary to refer to its representative. On a triangular lattice these cells would have size 3. Since a pointer can be non-null, although not all sites in the cell are occupied, a new variable must represent the state of the sites in each cell, if not lower or higher order bits of the pointers can be used (see above). On the hyper-cubic lattice the memory requirement is therefore for each pair of sites 2 bit for the state and 1 word for the address or index of the representative. Storing the 2 bits in a byte (and keeping the remaining 6 bits unused), the memory requirements are therefore reduced to $(b+1)N/2$ bytes.

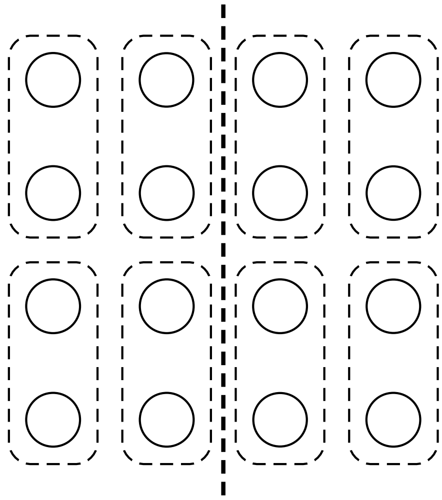


Figure 4.11: If occupied, each site within a dashed box belongs to the same cluster. On a triangular lattice the dashed patches would be triangular, each one containing three sites. The thick dashed line shows the orientation of the boundary between two consecutive slices in the parallelised code, see Sec. 4.2.5.

Using indices the maximum representable system size is given by $3/2N + 1 \leq M$ and using pointers with a size identification as shown in Fig. 4.9 the constraint is $1 + N \leq (M - \frac{bN}{2})/2$ in worst case.

4.2.3.3 Efficient histogram superposition

So far, only the maintenance of the cluster structure has been described. Since the masses of all clusters involved are known, it is simple to maintain a histogram of the cluster mass distribution: If a cluster of size s is burnt, the corresponding entry in $\mathcal{P}_t^{\mathrm{a}}(s)$ is decreased by one. If a cluster changes size, $\mathcal{P}_t^{\mathrm{a}}(s)$ is updated accordingly. For example, when two clusters of size s_1 and s_2 merge as a particular site is newly occupied during the growing procedure, $\mathcal{P}_t^{\mathrm{a}}(s_1)$ and $\mathcal{P}_t^{\mathrm{a}}(s_2)$ are decreased by one and $\mathcal{P}_t^{\mathrm{a}}(s_1 + s_2 + 1)$ is increased by one.

Naïvely, the average cluster size distribution is the average of $\mathcal{P}_t^{\mathrm{a}}(s)$, i.e.

$$\frac{1}{T} \sum_{t'=1}^{t} \mathcal{P}_{t'}^{\mathrm{a}}(s) \tag{4.31}$$

with T as the number of iterations. Depending on the resolution of the histogram, it would be very time consuming to calculate this sum for each s. Using exponential binning (which is in fact a form of hashing) in order to reduce the size of the histogram solves the problem only partly.

Ignoring any hashing, a naïve superposition, where each slot in the histogram needs to be read, has complexity $\mathcal{O}(TH)$, where H is the largest cluster mass in the histogram.

This problem is solved by noting that early changes in the histogram propagate though the entire sequence of histograms. Denoting the initial histogram as $\mathcal{P}_0^{\mathrm{a}}(s)$ and $\Delta\mathcal{P}_t^{\mathrm{a}}(s) = \mathcal{P}_{t-1}^{\mathrm{a}}(s) - \mathcal{P}_t^{\mathrm{a}}(s)$ then

$$\mathcal{P}_t^{\mathrm{a}}(s) = \mathcal{P}_0^{\mathrm{a}}(s) + \sum_{t'=1}^{t} \Delta\mathcal{P}_{t'}^{\mathrm{a}}(s) \tag{4.32}$$

and therefore

$$\sum_{t=1}^{T} \mathcal{P}_t^{\mathrm{a}}(s) = T\mathcal{P}_0^{\mathrm{a}}(s) + \sum_{t'=1}^{T} (T - t' + 1)\Delta\mathcal{P}_{t'}^{\mathrm{a}}(s) \quad . \tag{4.33}$$

By using this identity only the right hand side of (4.33) is maintained by increasing it by $T - t + 1$ when a new cluster is created at time t and by decreasing it by the same amount when it is destroyed. In this way, the complexity is only of order $\mathcal{O}(T(\theta^{-1} + 1))$, according to the number of clusters created and destroyed, i.e. the number of changes in the distribution. This concept becomes only problematic, if floating point numbers are used to store the histogram and the accuracy is so small that changes in the sum by 1 do not change the result anymore.[4] The maximum value in $\mathcal{P}_t^{\mathrm{a}}(s)$, where this does not happen, is given by the largest m with $m + 1 \neq m$ where m is a variable of the same type as $\mathcal{P}_t^{\mathrm{a}}(s)$. For floating point number, the value of m is related to the constant `DBL_EPSILON` (or `FLT_EPSILON` for single precision), which essentially characterises the length of the mantissa. The concrete value of m is actually platform, precision and type dependent. For an unsigned integer of size 4, this value would be $(2^{32} - 1) - 1$, corresponding to `ULONG_MAX` $- 1$, for double precision IEEE75 floating point numbers this value is `FLT_RADIX**DBL_MANT_DIG` $- 1$, i.e. $2^{53} - 1$.

Provided that the right hand side of (4.33) is below the threshold m discussed above for all s, this means that only a single histogram needs to be maintained. It is initialised with $T\mathcal{P}_0^{\mathrm{a}}(s)$ and updated with $\pm(T - t + 1)$ at time step t, when a cluster of size s appears or disappears. It is worth mentioning that this concept obviously even works in conjunction with binning (or any other hashing).

4.2.3.4 Implementation of the burning procedure

The burning procedure was implemented in the obvious way, without making use of the tree structure, as shown in Fig. 4.12. Although the burning procedure could also be implemented explicitly recursively, it is of course significantly faster when implemented iteratively. The usage of a stack in the procedure might be thought of as reminiscent of the underlying recursive structure.

[4] For integers the precision is not a problem, but the maximum representable number easily becomes a problem.

```
/* Initialise current_stack. */
CLEAR current_stack;
/* Put initial site on current_stack. */
PUT rn ON current_stack;
/* Sites are cleared right after they have entered the cur-
rent_stack. */
rn = empty;
/* The first loop runs until there is nothing left to
 * burn, i.e. next_stack was not filled during the inner loop. */
DO {
   /*  Clear next_stack so that it can get filled in the next loop. */
   CLEAR next_stack;
   /* The next loop runs as long as there are sites left to burn
    * in the current generation of the fire. */
   WHILE current_stack not empty {
      /* GET: remove the upmost element from current_stack and
       * put it in x */
      GET x FROM current_stack;
      /* Visit all neighbours */
      FOR all neighbours n of x {
         if (n occupied) {
            /* Put occupied sites on the current_stack of the
             * next generation of the fire */
            PUT n ON next_stack;
            n = empty
         }
      }
   }
/*  The next current_stack to be considered is next_stack. */
   current_stack = next_stack;
} WHILE current_stack is not empty
```

Figure 4.12: The burning procedure starting at `rn`. In an actual implementation the copying of `next_stack` to `current_stack` can easily be omitted by repeating the code above with `current_stack` and `next_stack` interchanged, similar to a red-black approach (Dowd and Severance, 1998).

The number of times the outer loop in Fig. 4.12 runs, defines the generation of the fire front and gives T_M; other properties of the burnt cluster can be extracted accordingly. The most important resource required by this procedure are the stacks: One for the currently burning sites and one for the sites to be burnt in the next step. There is no upper limit known for the number of simultaneously burning sites, apart from the naïve $N/2$ on a hyper-cubic lattice, which comes from the observation that sites, which belong to the same generation of the fire, must reside on the same sub-lattice (even or odd).

On the other hand it is also trivial to find the maximal number of sites which burn at the same time, if the fire starts in a completely dense forest, i.e. in a lattice with $\rho = 1$. Obviously the size of the tth generation is then given by $4(t-1)$ for

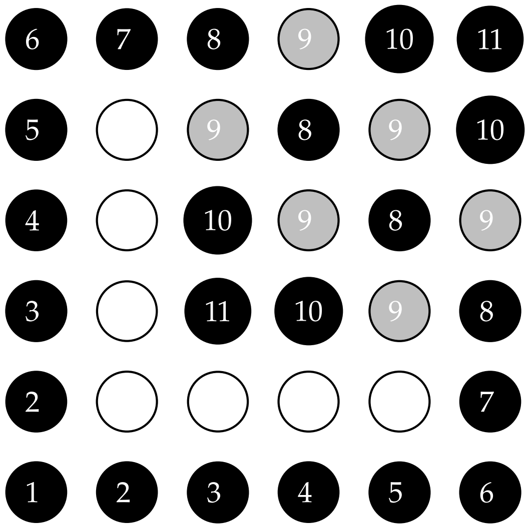


Figure 4.13: The burning order for a 6×6 patch of sites, where seven sites are not occupied and form a barrier, such that some sites behind it burn later, together with the fire front propagating away from the starting point of the fire at the lower left hand corner. The sites belonging to the largest set of trees burning at the same time are shown in light gray, unoccupied sites are shown in white, occupied sites in black. The numbers indicate the generation of the fire, which is one plus the Manhattan distance from the starting point of the fire along occupied sites.

$t > 1$ and 1 at the beginning, $t = 1$. Since the sum of these numbers is the number of sites which is reachable within a certain time t, *the sum* is also an upper limit for the number of simultaneously burning sites. Indeed, the actual number can easily be larger than $4(t-1)$, caused by arrangements of wholes in the lattice, which delay the fire spreading at certain sites, so that they burn together with a larger fire front. Such a construction is shown in Fig. 4.13.

Of course it is neither reasonable, nor practically possible to provide enough memory for the theoretical worst case, i.e. two stacks each of size $N/2$. Indeed the typical memory requirements seem to be of order $\mathcal{O}(\sqrt{\theta^{-1}})$, as shown in Tab. 4.1, where $f_{\max}$ denotes the largest fire front observed during the simulation. Providing stacks only of size $4L$ turned out to be a failsafe, yet pragmatic solution. Formally one could implement a slow out-of-core algorithm in the rare yet possible case the memory for the stack is insufficient, i.e. use hard-disk space to maintain it. In fact, this is what *de facto* happens if one uses a stack of size $N/2$ on a virtual memory system.

4.2.3.5 Complexity of the algorithm

The overall complexity of the algorithm has two contributions: The "growing" part, where new clusters are generated from existing ones and the "burning" part. The

System	L	θ^{-1}	ζ	$f_{\max}$
ap3000,2	8000	4000	1.51	
ap3000,2	8000	8000	1.52	
ap3000,4	16000	4000	1.34	
ap3000,4	16000	8000	1.48	
ap3000,4	16000	16000	1.37	
ap3000,4	16000	32000	1.41	
cluster,10	32000	4000	2.71	
cluster,10	32000	64000	3.81	
cluster,100	32000	32000	1.76	
single1	1000	500	1.41	216
single1	2000	1000	1.41	326
single1	4000	125	1.42	106
single1	4000	250	1.47	172
single1	4000	500	1.48	255
single1	4000	1000	1.53	317
single1	4000	2000	1.50	518
single1	4000	4000	1.57	646
single1	4000	8000	1.48	907
single1	4000	16000	1.45	1327
single2	8000	4000	2.11	687
single2	8000	8000	2.11	912
single2	8000	16000	2.09	1415

Table 4.1: Performance data for different parameters and setups. "ap3000,2" denotes a parallel run on two nodes on an AP3000, accordingly "ap3000,4". "cluster,10" denotes a cluster of 25 Intel machines, connected via an old 10 MBit network, "cluster,100" denotes the same cluster on a 100 MBit network. "single1" and "single2" denote two different types of single nodes. The largest fire front, $f_{\max}$, was only measured on these systems. The quantity ζ is the ratio of the average time (real time on the parallel systems in order to include communication overhead, user time on single nodes) for one successful update during statistics, i.e. when all data structures need to be maintained, and equilibration (transient) i.e. when the standard representation is used.

time needed for the burning part is proportional to the number of sites burnt and therefore expected as $\mathcal{O}(\langle s \rangle)$ [see (4.1) and (4.7)] and $\mathcal{O}(N)$ in the worst case. Since $\widetilde{\rho}$ in (4.7) is bound, the complexity of "burning" is $\mathcal{O}(\theta^{-1})$ (expected). The complexity of "growing" is estimated by the average number of sites newly occupied, θ^{-1}, times the worst-case complexity (4.30) to find the root of any given site, because up to four roots need to be found at each tree growing. According to (4.30) the worst case complexity to find the root of any given site is $\mathcal{O}(\log(N))$, leading to an overall complexity for "growing" of $\mathcal{O}(\log(N)\theta^{-1}) \supset \mathcal{O}(\theta^{-1})$. In practice the logarithmic correction is negligible, especially since $\log(N)$ is an extreme overestimate of the average case and therefore essentially the same runtime-behaviour is expected for both procedures (Newman and Ziff, 2001).

Implementations like the one in (Honecker, 1997) avoid this logarithmic factor by counting only the burnt cluster and therefore arrive at an overall complexity of $\mathcal{O}(\theta^{-1})$.

The algorithm presented has therefore only a negligibly higher computational complexity compared to implementations which measure only $\mathcal{P}^{\mathrm{b}}$. This is corroborated by the comparison of the CPU time per burnt cluster during equilibration, i.e. the transient, when the cluster structure do not need to be maintained and the algorithm used is the standard implementation, to the CPU time per burnt cluster during statistics, i.e. when observables are actually measured and especially $\mathcal{P}^{\mathrm{a}}$ is produced. This ratio is shown as ζ in Tab. 4.1 and Tab. 4.2. It varies only slightly with L or θ^{-1}.

Apparently the algorithm presented offers more statistics, however it suffers from one limitation: It requires about $(b+1)N/2$ bytes memory (see Sec. 4.2.3.2), compared to $N/8$ bytes in bitwise implementations like (Honecker, 1997), i.e. typically a factor 20 more. In order to ascertain whether this disadvantage is acceptable with respect to the statistical gain, one has to determine the standard deviations of the calculated quantities for both implementations.

4.2.4 Calculating the standard deviation

In order to compare the two algorithm rigorously, it is necessary to estimate the standard deviation of the estimators for $\bar{n}(s)$ produced by them (Müller-Krumbhaar and Binder, 1973; Landau and Binder, 2000):

$$\begin{aligned}
\sigma^2_{\mathcal{P}^{\mathrm{b}}}(s) &= \tfrac{2\tau_{\mathcal{P}^{\mathrm{b}}}+1}{T-1}\left(\left\langle \mathcal{P}^{\mathrm{b}}_t(s)^2\right\rangle - \left\langle \mathcal{P}^{\mathrm{b}}_t(s)\right\rangle^2\right) \\
\sigma^2_{\mathcal{P}^{\mathrm{a}}}(s) &= \tfrac{2\tau_{\mathcal{P}^{\mathrm{a}}}+1}{T-1}\left(\left\langle \mathcal{P}^{\mathrm{a}}_t(s)^2\right\rangle - \left\langle \mathcal{P}^{\mathrm{a}}_t(s)\right\rangle^2\right)
\end{aligned} \tag{4.34}$$

Here $\tau_{\mathcal{P}^{\mathrm{b}}}$ and $\tau_{\mathcal{P}^{\mathrm{a}}}$ are the correlation times of the two quantities. Calculating the correlation time in the standard fashion by recording the history $\mathcal{P}^{\mathrm{a}}_t(s)$ and $\mathcal{P}^{\mathrm{b}}_t(s)$ for each s would mean to store millions of floating point numbers. Therefore it was decided to restrict these calculations to just a small yet representative set of s values. The result shows that the standard deviation does not fluctuate strongly in s.

Because of the special form of $\mathcal{P}^{\mathrm{b}}_t(s) \in 0, 1$, its variance is particularly simple,

$$\left\langle \mathcal{P}^{\mathrm{b}}_t(s)^2\right\rangle = \left\langle \mathcal{P}^{\mathrm{b}}_t(s)\right\rangle \tag{4.35}$$

so that

$$\sigma^2_{\mathcal{P}^{\mathrm{b}}}(s) = \frac{2\tau_{\mathcal{P}^{\mathrm{b}}}+1}{T-1}\left\langle \mathcal{P}^{\mathrm{b}}_t(s)\right\rangle\Big(1-\left\langle \mathcal{P}^{\mathrm{b}}_t(s)\right\rangle\Big) \quad . \tag{4.36}$$

The correlation time of $\mathcal{P}^{\mathrm{b}}_t(s)$ is expected to be extremely small, not only on physical grounds — an cluster can only burn down once — but also because of the extreme dilution of $\mathcal{P}^{\mathrm{b}}_t(s)$, as it was described in Sec. 4.2.2.1. For fixed s, most of the $\mathcal{P}^{\mathrm{b}}_t(s)$ are

0. In contrast, the $\mathcal{P}^{\mathrm{a}}_t(s)$ are expected to have a large correlation time, because "only" $\theta^{-1}+1$ entries are changed between two subsequent histograms.

The correlation function is calculated in the symmetric way as proposed in (Anderson, 1964), here for an arbitrary quantity A_t:

$$\phi^{AA}_{t'} = \frac{\langle A_t A_{t+t'}\rangle_{T-t'} - \langle A_t\rangle_{T-t'}\langle A_{t+t'}\rangle_{T-t'}}{\langle A_t^2\rangle_T - \langle A_t\rangle_T^2} \tag{4.37}$$

where $\langle\rangle_{T-t'}$ denotes the average taken over time t from $t = 1$ to $t = T - t'$. The quantity $\phi^{AA}_{t'}$ was fitted to $\exp(-t/\tau_A)$ in order to find the correlation time τ_A. The results are given in Tab. 4.2.

As described in (4.6) and (4.8), the two estimators for $\bar{n}(s)$ differ slightly. However, except for $\bar{n}(s)$ only constant values appear on the RHS of (4.6) and (4.8), so that the relative errors of $\langle\mathcal{P}^{\mathrm{b}}_t(s)\rangle_T$ and $\langle\mathcal{P}^{\mathrm{a}}_t(s)\rangle_T$ are also the relative errors of the estimators for $\bar{n}(s)$ derived from them. These relative errors are shown in Tab. 4.2 as well. Their ratio is given as α and is an indicator for the advantage of the algorithm proposed. If the relative error is to be improved by a factor q, one needs to invest q^2 CPU-time, i.e. if the algorithm proposed in this chapter costs a factor ζ more CPU-time, and the gain in the relative error α, the total gain is α^2/ζ. The values for this quantity are also given in Tab. 4.2.

According to the table, for fixed θ relative errors and the correlation times are only weakly affected by an increase in system size. At first sight, this is counter-intuitive, as the number of passes (Henley, 1993; Honecker and Peschel, 1997), the mean number of times a site has been visited between two lightnings, decreases inversely proportional to the total number of sites in the system: $1/(\theta\bar{\rho}L^2)$, see Sec. 4.3.4. Assuming that this number is essentially responsible for the error, suggests to keep the number of passes constant among different L. However, this is apparently not the case, possibly because of self-averaging (Ferrenberg *et al.*, 1991) effects.

The table also shows various tendencies, which are worth mentioning. First of all, the total gain becomes smaller for larger avalanche size s. The B in front of some of the values indicates that a bin around the s value was investigated, i.e. the time series of

$$\sum_{s'\in\mathcal{B}} \mathcal{P}^{\mathrm{a,b}}(s') \tag{4.38}$$

was considered, where $\mathcal{B}$ is a set of (consecutive) s values, representing the bin. For larger values of s, these sets get exponentially larger, which is necessary for a reasonably large number of events as basis for the estimators. The general tendency that the proposed algorithm is even more efficient at small s is not surprising: $\mathcal{P}^{\mathrm{b}}$ samples from $s\bar{n}(s)$, while $\mathcal{P}^{\mathrm{a}}$ samples only from $\bar{n}(s)$, i.e. $\mathcal{P}^{\mathrm{b}}$ "sees" larger cluster more often. *Nevertheless $\mathcal{P}^a$ still is advantageous by roughly a factor* 5. The empty entries in

L	θ^{-1}	ζ	s	$\tau_b(s)$	$\tau_a(s)$	$\frac{\sqrt{\sigma^2_{\mathcal{P}^b}(s)}}{\langle \mathcal{P}^b_t(s)\rangle}$	$\frac{\sqrt{\sigma^2_{\mathcal{P}^a}(s)}}{\langle \mathcal{P}^a_t(s)\rangle}$	α	α^2/ζ
4000	4000	1.57	10	–	–	0.0138∗	–	–	–
			100	0.170	23.6	0.0637	0.00099	64.3	2633.4
			B 10^3	0.028	14.2	0.0450	0.00191	23.6	354.8
			B 10^4	0.006	10.0	0.0412	0.00470	8.8	49.3
			B 10^5	–	7.2	0.0662∗	0.02104	3.1	6.1
4000	16000	1.45	10	0.013	39.9	0.0141	0.00056	25.4	444.9
			100	0.126	28.8	0.0608	0.00127	48.0	1589.0
			B 10^3	0.006	4.7	0.0457	0.00175	26.1	469.0
			B 10^4	0.013	2.9	0.0512	0.00332	15.4	163.6
			B 10^5	–	2.2	0.0433	0.00795	5.4	20.1
8000	1000	–	10	0.131	–	0.0154	–	–	–
			100	0.122	284.6	0.0602	0.00158	38.1	–
			B 10^3	0.028	236.5	0.0399	0.00337	11.8	–
			B 10^4	0.016	163.5	0.0397	0.00878	4.5	–
8000	4000	2.11	10	0.122	78.2	0.0154	0.00052	29.8	420.9
			100	0.132	16.4	0.0634	0.00087	72.9	2518.7
			B 10^3	0.022	8.2	0.0438	0.00147	29.7	418.1
			B 10^4	0.005	5.5	0.0442	0.00241	18.3	158.7
			B 10^5	–	4.2	0.0409∗	0.01006	4.1	8.0
			B $2\cdot10^5$	–	3.8†	0.0635∗	0.02055	3.1	4.6
8000	16000	2.09	10	–	262.5	0.0139∗	0.00068	20.5	201.1
			100	0.131	56.1	0.0629	0.00087	72.0	2480.4
			B 10^3	0.014	19.0	0.0467	0.00115	40.6	788.7
			B 10^4	0.009	11.1	0.0503	0.00296	17.0	138.3
			B 10^5	0.006	8.3	0.0411	0.00689	6.0	17.2
			B $2\cdot10^5$	–	7.5	0.0423∗	0.00947	4.5	9.7
			B $5\cdot10^5$	–	7.0†	1.1106∗	0.33331	3.3	5.2

Table 4.2: Correlation times τ_b and τ_a of the corresponding observables $\mathcal{P}^b$ and $\mathcal{P}^a$ as a function of s and for different parameters L, θ^{-1}. Values of s marked by "B" are results for bins around the s value indicated. For each set of parameters, the quantity ζ is given. It denotes the ratio between the average CPU-time for one successful update during equilibration (transient) and during statistics, see also Tab. 4.1. The two fractions $\frac{\sqrt{\sigma^2_{\mathcal{P}^b}(s)}}{\langle \mathcal{P}^b_t(s)\rangle}$, $\frac{\sqrt{\sigma^2_{\mathcal{P}^a}(s)}}{\langle \mathcal{P}^a_t(s)\rangle}$, their ratio α and α^2/ζ are derived. $*$ marks cases, wheres $\tau_b(s) = 0$ has been assumed. † marks values of $\tau_a(s)$, which have been extrapolated from $\tau_a(s)$ for smaller s.

Tab. 4.2 are due to numerical inaccuracies or simply missing simulations for certain parameters. Some entries are estimated and marked as such.

There is an additional correlation not mentioned so far: The individual points in the estimator of the distribution $\mathcal{P}^a$ are not independent. There are "horizontal correlations", i.e. $\mathcal{P}^a(s)$ is correlated for different values of s. These are additional correlations due to clusters of small sizes, which are likely to grow and propagate through s in $\mathcal{P}_t^{\mathrm{a}}(s)$ for consecutive time steps, i.e.

$$\left\langle \mathcal{P}_t^{\mathrm{a}}(s)\mathcal{P}_{t'}^{\mathrm{a}}(s')\right\rangle - \left\langle \mathcal{P}_t^{\mathrm{a}}(s)\right\rangle\left\langle \mathcal{P}_{t'}^{\mathrm{a}}(s')\right\rangle \quad . \tag{4.39}$$

This correlation is at least partly captured by the correlations measured for the binned data. It is to be distinguished from the correlations of *independent* realisations, where correlations are expected in the cluster size distribution also, i.e.

$$\left\langle \mathcal{P}_t^{\mathrm{a}}(s)\mathcal{P}_t^{\mathrm{a}}(s')\right\rangle - \left\langle \mathcal{P}_t^{\mathrm{a}}(s)\right\rangle\left\langle \mathcal{P}_t^{\mathrm{a}}(s')\right\rangle \quad . \tag{4.40}$$

This must be taken into account as soon as estimates of $\bar{n}(s)$ for different s are compared, as it is done when an exponent is calculated by fitting. This effect is also present for $\mathcal{P}^{\mathrm{b}}$, which is, however, diluted so enormously that it influences the outcome only in an insignificant way.

The horizontal correlations could be estimated using a Jackknife scheme (Efron, 1982), similar to that used to calculate the error bar of the exponent from the time evolution of a quenched ISING model (Pruessner *et al.*, 2001). While it is certainly essential for the careful estimation of the error bar of an exponent, it is irrelevant for the discussion in this chapter, as it is quantitatively based only on *local* comparisons of error bars (overlaps), while its global properties, i.e. shape and collapse with other histograms estimated, is not concerned with errors bars. Some authors even seem to dismiss the relevance of these correlations completely (Newman and Ziff, 2001).

4.2.5 Parallelising the code

Constructing clusters and keeping track of clusters rather than of single sites seems to be in contradiction to any attempt to run the algorithm distributed, that is splitting the lattice into S *slices* (one-dimensional decomposition — as periodic boundaries apply, the slices may better be called cylinders). Moreover, there is a general problem of parallelisation which becomes apparent in this context: The usual bottleneck of parallel systems is the communication layer. In order to keep the communication between sub-lattices as low as possible, fast parallel code on a lattice requires as few interaction between slices as possible, while the whole point of doing physics on large lattices is the assumption of significant interaction between their parts. It is this fundamental competition of requirement and basic assumption which makes successful parallel code so rare and which seems to indicate that problems must have very specific characteristics in order to be parallelisable in a reasonable way.

However, it is indeed possible to run the algorithm described above on parallel machines successfully in the sense that it does not only make use of the larger amount of (distributed) memory available, but also of the larger amount of computing capabilities. In fact, the code was successfully rewritten using MPI (Gropp *et al.*, 1999) and has been run on two systems with distributed memory: The massively parallel machine AP3000 at the Department of Computing at Imperial College and on a cluster of workstations (25 nodes).

In the following the most important design characteristics are described which proved important in order to make the code running reasonably fast. This concerns mainly the statistics part, but the equilibration also needs some tricks.

MPI assures that packets sent from one node to another in a certain order are received in exactly the same order — in the language of MPI this means that the message ordering is preserved in each particular communicator. But, how different communicators relate to each other, i.e. how one stream of packets relates to another one is not specified. If, for instance, node A sends a packet to node B, and then to node C, which then sends a packet to node B, this packet might arrive earlier at B then the packet first send by A, see Fig. 4.14.

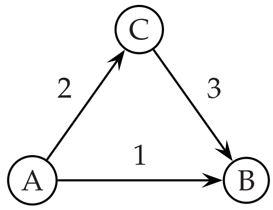


Figure 4.14: Nodes A, B and C send messages in the order indicated. However, it might well happen that the message sent last by node C to node B, namely message 3, arrives at that node before message 1, sent *before* message 2 was sent, which arrived *before* message 3 was sent.

However, it is one of the main goals of parallelisation, to avoid any kind of synchronisation, which is extremely expensive. Even in a master-slave design, as it was chosen here, one encourages communication between the slaves, whenever they can anticipate what to do next or can indicate each other what to do next.

As explained above (sec. 4.2.1) an update consists essentially of two steps: Growing and burning. Both processes now are distributed among the slices. The growing procedure is realized by trying to grow θ^{-1}/S trees in each slice. This is not an exact representation of a growing procedure taking place on the entire lattice at once, because the latter has a non-vanishing probability to grow all trees at one particular spot, while the parallelised version distributes them evenly among the different slices. Provided that θ^{-1} is large compared to S, this effect can certainly be neglected. The advantage of the procedure is that the growing procedure at each slice does not need to be conducted by the master. The burning procedure is more complex, as the fire starts at one particular site of the entire lattice, so that it must be selected by the master. The exact procedure of the possibly following burning process depends on the stage of the algorithm.

In the following the procedures are explained in terms of "sites" rather than "cells", as introduced in Sec. 4.2.3.2. Using cells instead of sites makes the code slightly more complicated, but the changes are obvious. If the cells are oriented parallel to the borders of slices (see Fig. 4.11), so that its width is a multiple of 2 in case of a hyper-cubic lattice, the algorithm runs considerably faster, as the communication between the nodes is reduced by the same factor.

4.2.5.1 Equilibration

During the equilibration phase it is not necessary to keep track of all clusters. Nevertheless there is some statistics, which is very cheap to gather: The distribution of burnt clusters and the density of trees. The latter is very simple, as this number changes in time only by the number of grown trees minus the number of burnt trees. This is also a crosscheck for the overall statistics, as the tree density is equivalent to the probability of a site to belong to *any* cluster (4.4).

The burning is implemented as follows: The master chooses a site from the entire lattice and sends the corresponding slice (slave) the coordinate and (implicitly) an identifier which uniquely identifies this request within this update step. The slice's response consists of the number of sites burnt (possibly 0), the identifier referring to the initial request and possibly up to two further, new, unique identifiers. These identifiers refer to the two possible sub-requests to the right and left neighbouring slice due to a spreading of the fire. If a slice contacts another slice, it does so by sending the coordinates of sites, which are on fire in the sending patch, together with a unique identifier. The slice contacted sends it result to the master, again together with the identifier and possibly two new ones, corresponding to the possibly two contacted neighbouring slices. In this way the master keeps track of "open (sub)requests", i.e. requests the master has been told about by receiving an answer containing information about sub-requests, which have not been matched by receiving a corresponding answer. The structure of requests forms a tree-like structure and if there are no open requests, the master must have received all answers of the currently burning fire. It is very important to make it impossible that by a delay of messages some answers are not counted, as it would be, if the master would just count open requests, without identifying them individually. It can easily happen that the master receives an answers for a request, without having received the information about the very existence of the request. It is worth to mention that in this scheme the order of burnings is irrelevant, if the burn-time is not measured, as it was done here.

Adding up the number of burnt sites gives the total size of the burnt cluster. This number is finally sent to all slices. If it is nonzero, the step is considered as successful.

After equilibration the cluster structure of pointers and roots as described above (see Sec. 4.2.3.1), needs to be constructed. This is done in a naïve manner: Keeping track of sites, which have already been visited, every site is visited once. The first site visited in each cluster becomes root of all sites connected to it, which become marked as visited. The procedure corresponds to the burning procedure described above (see sec. 4.2.3.4).

Each slice maintains a local histogram $\mathcal{P}^{\mathrm{a}}$, which contain all clusters, which do not have a site on the border to another slice. Otherwise, they are maintained at the master's histogram, as discussed below. In this case the (local) root site of these clusters are moved to the border. As periodic boundary conditions apply, the only boundaries are those with other slices.

4.2.5.2 Collecting statistics

After finishing the equilibration phase another concept needs to be applied in order to count the total cluster size distribution $\mathcal{P}_t^{\mathrm{a}}(s)$. At every update of the lattice each slice must keep track of the clusters in the same way as it was described in section sec. 4.2.3.1. Clusters, which do not contain a site at a border to another slice are maintained locally, i.e. at each node as a *local histogram*. However, if a cluster contains a site at a border, it might span several slices. As soon as a cluster acquires a site at the border, it is removed from the local histogram and the site under consideration becomes the root of the cluster. The algorithm ensures that a cluster with at least one site on the border has its root at the border.

During all processes (growing or burning), the size of all clusters is updated as usual, independent from the location of the root. If the status of a border site changes, its new value or its change is put on a stack together with its coordinate. During the growing procedure the following changes of the status are possible:

- *New occupation:* Change in occupation information for a site (cell); If this is the only change, then it must have been already occupied (this is only possible, in an implementation using cells). If this is not the case, the reference information pointing to the root site of the given cluster, must be updated also, see next point.

- *Merging border clusters:* Change of the reference information for a site (cell); This can only happen if the site (cell) was (completely) unoccupied at the time of the change or did contain a size information, i.e. it was itself a root.

- *General merging of clusters:* Change in size information for a site (cell); Only an increase is possible, so that any change can be represented by a single number indicating the size difference.

For each border site changing at each slice, the corresponding information are sent to the master. Typically the number of messages is not very large, because the total number of sites updated during a single growing phase is limited by θ^{-1}/S. The expected number of these message is not given by the fraction of border sites in each slice, because changes in all border *clusters* (i.e. clusters with at least one site in the border) affect the border *sites*, as the root of each border cluster is a border site.

However, the data regarding the updates in the border do not need to be send from the slaves to the master, if the burning attempt following the growing fails, i.e. if an empty site has been selected for lightning. Of course it is much more efficient not to send any data if not necessary. As there is only a finite number of sites in each slice, the theoretical limit of updates of border sites is bound by this number. However, it is sufficient to allocate a reasonable amount of memory ($2L$ turned out to be enough) for the stack of messages to be sent and check its limits, similar to the stack used in the burning procedure described in Sec. 4.2.3.4. Henceforth the sending of the update information of the border is called "sending the border".

The master maintains a copy of the state of the border sites and updates a *global histogram* of border clusters. By sending the changes on the border to the master as described above, the master can update its copy of the configuration of the borders as well as the global histogram. At the end of the simulation all histograms (S slaves histograms plus the global histogram maintained by the master node) are summed to produce the total $\mathcal{P}^{\mathrm{a}}$.

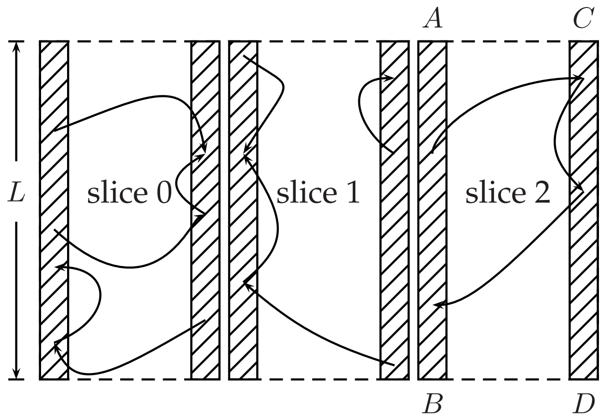


Figure 4.15: The slices, three of which are shown here, maintain the references for all clusters within each slice (illustrated by arrows), even for border clusters. The references *between* slices, however, are maintained by the master. The variables $A = 0$, B=L-1, $C = I$ and $D = I + L - 1$ are the indices used for references within each slice.

As suggested in Fig. 4.15, the slices maintain the pointers within each slice and these references are not changed by the master, which only connects *between* slices. If a reference at the border changes at a slice, the master receives a message to apply the corresponding changes (joining two clusters), if the size of a cluster changes, the master updates the corresponding unique root etc. These changes are indicated by the slaves and the master only realises them in the copy of the border sites. Only if a change in occupation occurs, the master must actually perform some non-trivial operations, because a newly occupied site might introduce a new connection *between* borders of different slices. From the point of view of the master, only borders belonging to two different, neighbouring slices are directly connected and therefore to be maintained by the master, while the connectivity of the borders *within* each slice is indicated and maintained by the corresponding slave. Apart from that, the master maintains the slice spanning structures in exactly the same way as the slaves, e.g. a cluster having multiple roots among the various slices has a unique root at the master etc.

The question arises how the master best keeps track of the changes of the borders. Ideally, a change of reference of a site at the boundary is communicated to the master simply by sending the new pointer value (index). By choosing a reasonable indexing scheme, this is indeed possible. If the value of the reference is within 0 and $L - 1$, where L is the width in terms of number of sites (or cells) (see Fig. 4.15), the reference denotes a site in the left border within the same slice. Similarly, if the value of a reference is within I and $I + L - 1$, where I denotes the first index in the last column, a reference with such a value is bound to point to the right border of the same slice. If the master uses indexes of the range $[L, I - 1]$ for denoting cross-references between slices the references are therefore unambiguous and no translation is necessary between indeces used by the slices and indeces used by the master.

During the burning procedure the master can make use of its knowledge about the borders. The site selected for starting the fire is most likely a bulk size, so that the corresponding slave needs to be contacted for the occupation information. Three outcomes are possible:

- The site is unoccupied. Nothing happens, all slices get signalled to continue with growing.
- The site is occupied, but does not contain a border site. In this case the slice contacted can send back the size of the burnt cluster (an information it knows even without actually doing the burning as the size is stored in the root, which needs to be found anyway in order to find out whether the cluster is a border cluster) and the master can signal all other slices to send the border and to continue. After receiving the borders it can update the histogram.[5]

[5]One might be inclined to postpone the sending of the borders to a time, when it is really needed. However, after a successful burning the time t is increased and this enters the histogram (see Sec. 4.2.3.3). Ignoring this change for a large number of steps would introduce uncontrollable deviations

- The site is occupied and contains a border site. In this case the slice sends the reference of the border site back to the master, which then contacts all slices to send the most recent border update. It updates the border and the histogram, deletes the cluster which is going to burn and sends the "burning borders", i.e. a list of all border sites which will be affected by the burning procedure to the slices in form of a stack as described in Sec. 4.2.3.4. The slaves use this stack as the initial stack of the burning procedure and delete the corresponding sites. No communication between the slices is necessary.

The global histogram contains much larger clusters than the local histograms. In order to keep memory requirements low, even for histograms of resolution unity, it is reasonable to introduce a threshold, above which slaves use the global histogram to maintain $\mathcal{P}^{\mathrm{a}}$ even for local clusters (i.e. non border cluster). For that purpose a histogram "appendix" has been introduced. This is a finite stack, which stores the size of the cluster s together with the value of $t' = \pm(T - t + 1)$ as described in Sec. 4.2.3.3. During the growing phase when such large clusters grow quickly, one would obtain a sequence of stack entries of the form $(s, t'), (s, -t'), (s + 1, t'), (s + 1, -t'), (s + 2, t'), \dots$, corresponding to entering the appendix, (s, t'), increasing in size by 1, which gives $(s, -t'), (s + 1, t')$ etc. As soon as a cluster is larger then the upper cutoff each update causes two entries, of the form $(s, -t'), (s+1, t')$, the first for the deletion from the histogram, the second from the increase in the next slot. These entries possibly cancel, for example the sequence above is equivalent to the single entry $(s + 2, t')$. It turned out to be highly efficient to perform this cancellation, i.e. to check the last entry in the appendix for being the negative entry of the one to be done.

As the maximum size of the appendix is finite, it must be emptied from time to time. The information about the size of the appendix of each slave is sent to the master together with the information about the borders. If a possible overflow is detected ($2/3$ of the maximum size in the implementation presented) the master requests all slices to send the content of the appendices and applies it to the global histogram. The slices then empty their appendices.

4.2.5.3 The random number generator

The random number generator (RNG) acquires a crucial role when used in a parallel environment. With M the number of iterations, the expected number of calls of the RNG is $M\theta^{-1}/\rho$ (for $M \approx 10^7, \theta^{-1} \approx 5\times10^4$ this is more than 5×10^{11}), so that an RNG as `ran1` in (Press *et al.*, 1992) with a period of only $\approx 2 \times 10^9$ is insufficient. Therefore `ran2` in (Press *et al.*, 1992) was used for all simulations, both parallel and non-parallel,

of the estimator of the histogram from its true value.

which has a period of $> 2 \times 10^{18}$. If the number of RNG calls is small enough, one can compare results obtained by means of `ran1` and `ran2`. No significant deviation was found.

In the parallel implementation, each slave requires an independent sequence of random numbers. This is a classical problem in parallel computing (Aluru *et al.*, 1992; Coddington, 1996). The simplest solution is to divide a single sequence $r_1, r_2, \ldots$ into distinct subsequences. This can be done either by a leapfrog scheme (Coddington, 1996; Entacher, 1999), where each subsequence consists of random numbers which are S calls away, i.e. S subsequences of the form $r_u, r_{S+u}, r_{2S+u}, \ldots$ with $u = 1, 2, \ldots, S$ unique at each slave, or by splitting the sequence (Coddington, 1996), so that each subsequence consists of consecutive RNG calls, i.e. $r_{1+uX}, r_{2+uX}, r_{3+uX}$ again with $u = 1, 2, \ldots, S$ and offset X large enough to avoid any overlap. The latter scheme has the advantage that the sequence consists of consecutive RNG calls and therefore has been used in the following. The implementation of the offset X at each slave is easily realised by restoring all state variables of the RNG, which have been produced once and for all in a single run producing all XS random numbers and saving the state variables on a regular basis. However, such a technique is advisable only if the RNG calls do not dominate the overall CPU time, in which case it would take almost as long as the simulation itself to produce the random numbers required for it.

4.3 Results

The sections above were only concerned with the technical issues of the model and its implementation. Some of the actual results from the simulation carried out using the new algorithm have been published already (Pruessner and Jensen, 2002a). This article was focused on $\bar{n}(s)$. The main outcome was that the standard scaling assumption (4.12) is not supported by numerics, so the main conclusion was that the model *is not scale-invariant*.

In the following these results are shortly restated and discussed. Other observables are connected with this observation to see, whether it is only $\bar{n}(s)$ which lacks scale-invariance. All results presented are based on the same simulations, the parameters of which are given in Tab. 4.3 (page 185).

4.3.1 Cluster size distribution

Before the actual findings are discussed, it is important to consider how to avoid finite size effects, which otherwise might damage the results. Usually, finite size effects are avoided by keeping the correlation length ξ small compared to the system size. However, it requires a significant amount of CPU-time to actually determine the

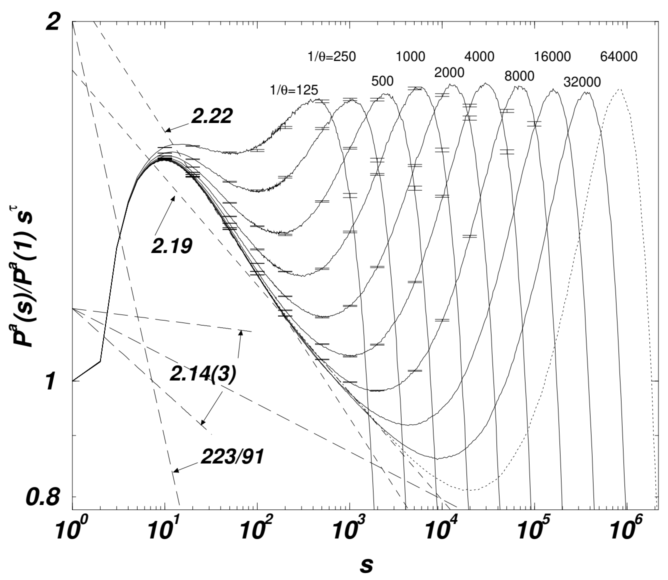


Figure 4.16: The rescaled and binned histogram $\frac{\mathcal{P}^{\mathrm{a}}(s)}{\mathcal{P}^{\mathrm{a}}(1)} s^{\tau^*}$, where $\tau^* = 2.10$ for $\theta^{-1} = 125, 250, 500, \ldots, 32000, 64000$ (as indicated) in a double logarithmic plot. The linear size L is chosen according to the bold printed entries in Tab. 4.3 (page 185) and large enough to ensure absence of finite size effects. The error bars are estimated from shorter runs. The rightmost histogram (dotted, $\theta^{-1} = 64000$) could not be cross-checked by another run, see text. The dashed lines belong to different exponents, whose value is specified as the sum of the slope in the diagram and τ^*, i.e. a horizontal line would correspond to an exponent 2.1. The shortly dashed lines represent estimated exponents for different regions of the histogram (2.22 for s within approx. $[20, 200]$ and 2.19 for s within $[200, 2000]$), the other exponents are from literature, namely 2.14(3) in (Clar *et al.*, 1994, 1996) and $223/91 \approx 2.45$ in (Schenk *et al.*, 2002). Since it was impossible to relate these exponents to any property of the data, the exact position of the lines associated with them was chosen arbitrarily.

correlation length. Moreover, *a priori* it would not be clear, which ratio ξ/L to choose in order to avoid finite size effects.

4.3.2 Avoiding finite size effects

Throughout this chapter we initially performed $5 \cdot 10^6$ successful updates (as defined in Sec. 4.2.2.1) as transient (and therefore rejected them) and the same number for producing statistics, apart from runs for calculating error bars, where only 10^6 updates has been used for statistics, see below. It is known that the transient can be very long (Honecker and Peschel, 1997) (note that the time unit in (Honecker and Peschel, 1997) is expressed in our units by multiplying it with $\theta^{-1}/(\rho L^2)$), but in all cases presented the number of initial steps seemed to be more than sufficient. Numerical checks indicate that the cluster size distribution is very stable against the size of the transient, i.e. even a transient, which is presumably too short, still produces reasonable results for $n(s)$.

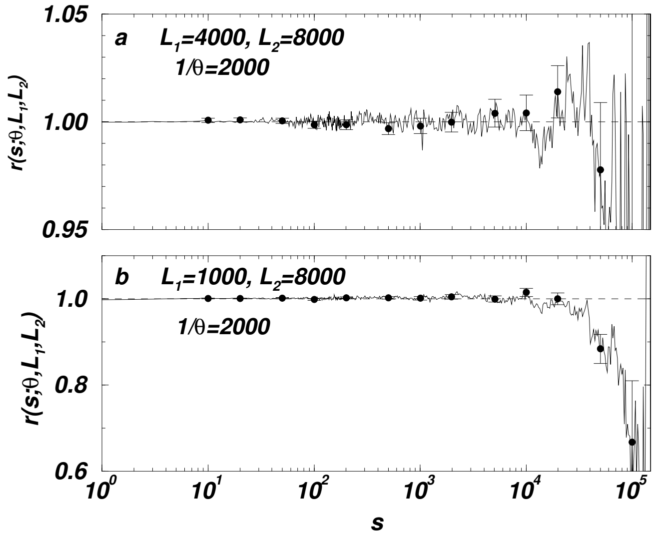


Figure 4.17: Ratio $r(s;\theta,L_1,L_2) = \bar{n}(s;\theta,L_1)/\bar{n}(s;\theta,L_2)$ with $\theta^{-1} = 2000$ for two pairs L_1, L_2 with error bars (one standard deviation; the error bars as well as the data shown are exponentially binned). The data are from short runs (10^6 updates for statistics). Finite size effects have been considered negligible under the condition that (almost all) error bars for this ratio have covered 1 (marked by a dashed line) in the relevant range. a) $L_1 = 4000$ and $L_2 = 8000$: Almost no finite size effects, the deviation from 1 is probably due to noise. Note the fine scale of the ordinate. b) $L_1 = 1000$ and $L_2 = 8000$: Systematic, strong finite size effects for $s \gtrapprox 10^4$. The scale of the ordinate is five times larger than in a). Data of this quality have been dismissed.

All systems have been initialised by a random independent distribution of trees with density 0.41.

The standard deviation of the binned histogram is not completely trivial to calculate. In particular, its computation requires a significant amount of CPU time, and was therefore only calculated for the smaller system sizes (up to $L = 8000$) and in shorter runs (only 10^6 updates for statistics, but $5 \cdot 10^6$ for transient). We resorted to visual examination for the larger systems when comparing $\bar{n}(s;\theta,L)$ for different system sizes. Fig. 4.17a and b show the ratio of $\bar{n}(s;\theta,L)$ for two different system sizes. A deviation of this ratio from 1 indicates a difference in the statistics and therefore the presence of finite size effects. Fig. 4.17a shows a typical case we accepted as reasonable agreement. Here $L_1 = 4000$ and $L_2 = 8000$ do not seem to differ for $\theta^{-1} = 2000$. Fig. 4.17b shows a case of finite size corrections we have dismissed (note the different scales in the two graphs). It differs from Fig. 4.17a only by $L_1 = 1000$.

Fig. 4.18 illustrates the strong agreement of $\bar{n}(s;\theta)$ at the same value of θ for the same two different sizes L as in Fig. 4.17a. The two sets of data are virtually indistinguishable, but in this kind of plot it is also almost impossible to see a difference between the data of $L_1 = 1000$ and $L_2 = 8000$, as shown in the inset of Fig. 4.18. This is also the case with the rescaled data below, and the use of very large systems

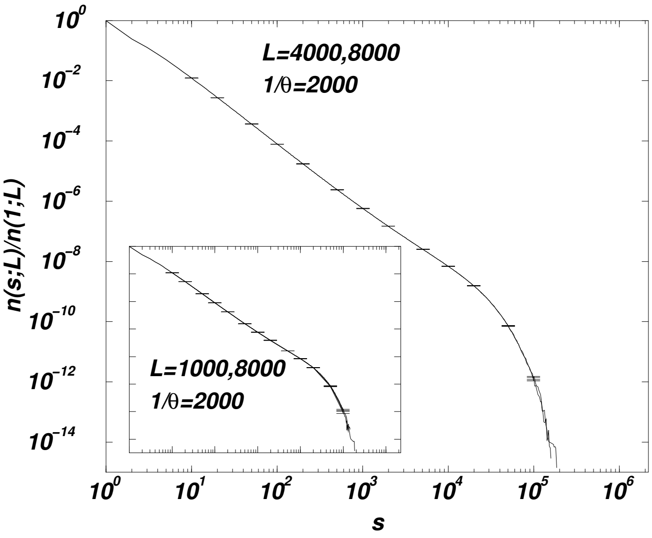


Figure 4.18: The binned histogram $\bar{n}(s;\theta,L)$ for two different values of L and fixed θ as in Fig. 4.17a. In this plot the two histograms are virtually indistinguishable. However, note that the deviations shown on Fig. 4.17b would also hardly be visible in this type of plot, as shown in the inset.

throughout this chapter might therefore be "overcautious" in avoiding finite size effects, although such large sizes are obviously required for an accurate *quantitative* analysis of this model. However, when it comes only to qualitative analysis, such a judgement seems to be justified. On the other hand, an increase in system size hardly increases the computing time and affects "only" the memory requirements, which forced us to implement the algorithm for parallel machines. The side effect of using multiple CPUs at the same time is a significant reduction of the simulation time especially for large values of θ^{-1}, a fact which compensates the complications of parallel coding.

Another indicator for the absence of finite size effects is the scaling of the standard deviation of ρ: If the lattice can be split into independent parts, i.e. if subsets of the lattice can be considered as independent, the standard deviation of ρ should scale like $1/L$ for different values of L at given θ^{-1}. Such a behaviour can be seen in Tab. 4.3, although the standard deviation of ρ could be calculated only roughly. This might explain the slight mismatch for $\theta^{-1} = 32000$, $L = 16000, 32000$.

For the highest values of θ^{-1} we could not yet do the comparison to another system, so the curve for the largest value of θ^{-1} in Fig. 4.16 is dotted, as their quality is not known. However, it is reasonable to assume that it is not affected by finite size scaling.

Comparing the different histograms $\bar{n}(s;\theta)$ for different values of θ^{-1} in a plot enables us not only to find the exponent τ, but also to find the universal function

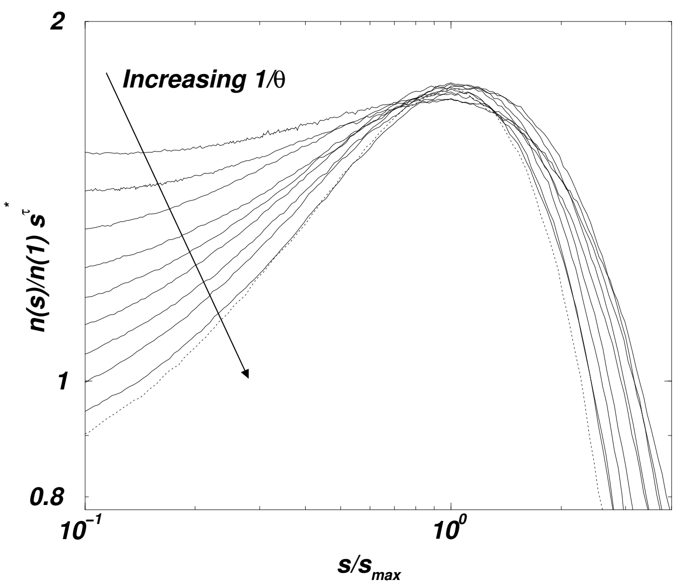


Figure 4.19: The rescaled and binned histogram $\bar{n}(s;\theta)s^{\tau^*}$, versus $s/s_{\max}(\theta)$, where $\tau^* = 2.10$ for $\theta^{-1} = 125, 250, 500, \cdots, 32000, 64000$ in a double logarithmic plot. The scales $s_{\max}(\theta)$ by which the histograms have been shifted are the maxima marked in Fig. 4.16, so that a data collapse would be possible. The arrow indicates the order of the data in increasing θ^{-1}.

$\mathcal{G}$ as defined in Eq. (4.12). A rough, naïve estimate of τ is given by $\bar{n}(s;\theta)$ fitted against $s^{-\tau}$, which gives a value of $\tau^* \approx 2.1$ in our case. Plotting now $\bar{n}(s;\theta)s^{\tau^*}$ double logarithmically should allow us to find the "true" value of τ by performing a data collapse, i.e. choosing τ^* in such a way that horizontal shifts (corresponding to the choice of the scale $s_0(\theta)$ in the scaling function) make all curves collapse. This is shown in Fig. 4.16, where $\tau^* = 2.1$ was chosen so that the maxima for the second bumps are almost equally high: denoting their position on the abscissa for each value of θ by $s_{\max}(\theta)$, we have chosen τ^* such that

$$\bar{n}(s_{\max}(\theta);\theta)\; s_{\max}^{\tau^*}(\theta) \approx \text{const.} \quad . \tag{4.41}$$

According to (4.12) the constant is simply the maximum value of $\mathcal{G}$, namely $\mathcal{G}(s_{\max}(\theta)/s_0(\theta))$, where the value of the argument is therefore the same for all θ.

The value of τ^* is close to (but not within the error of) the exponent found in the literature, $\tau = 2.14(3)$ (Clar *et al.*, 1994, 1996) [($\tau = 2.15(2)$ in (Grassberger, 1993), $\tau = 2.159(6)$ in (Honecker and Peschel, 1997)], which is shown in the same figure for comparison. However, it is impossible to force the minima (see the down pointing marks in Fig. 4.16) to the same height while maintaining the constraint that the maxima remain aligned, i.e. these minima cannot be a feature of the same universal scaling function. Otherwise (4.12) would hold and the quantity

$$\bar{n}(s_{\text{min}}(\theta);\theta)\ s_{\text{min}}^{\tau^*}(\theta) \quad , \tag{4.42}$$

where $s_{\text{min}}(\theta)$ denote the position of the minima, would assume the same value for all θ, because they are local minima of $\mathcal{G}$, which is supposed to be the same for all θ.

Since these minima cannot be included in the simple scaling defined in (4.12), they must be explicitly excluded by introducing a lower cutoff, so that simple scaling supposedly sets in only above these cutoffs, excluding especially the minima. However, such a lower cutoff would apparently have to diverge for $\theta^{-1} \to \infty$ – something that is certainly beyond any established concept of scaling. Even when accepting this peculiar scaling behaviour, a data collapse for the second bump still seems to be unsatisfactory, as shown in Fig. 4.19.

If one accepts a divergent lower cutoff of the scaling function, one has to face the fact that this would describe the behaviour of $\bar{n}$ in a region, which becomes physically less and less interesting in the limit $\theta^{-1} \to \infty$, because the vast majority of events are situated at small s and as the second bump moves out to infinity, the scaling function hence covers a smaller and smaller part of $\bar{n}$. However, only a region of $\bar{n}$ which covers a non-vanishing fraction of events can be *physically relevant*.

Concentrating now on the behaviour of $\bar{n}$ up to the minimum (see arrows pointing downwards in Fig. 4.16), one finds that this region is also badly described by a function like (4.12). First of all, the question of which region is supposedly described by the function needs to be answered. A unique lower cutoff and a θ dependent upper cutoff needs to be found. At first view it looks appealing to choose these two marks such that they cover the set of data, where the curves fall on top of each other. In this case the lower cutoff would be 1 and the upper cutoff, $s_{\text{naïve}}$, would have a value smaller than the minima marked by downwards pointing arrows in Fig. 4.16. However, this would be described by a function like

$$\bar{n}(s;\theta) = f(s)\mathcal{G}(s/s_{\text{naïve}}) \tag{4.43}$$

rather than (4.12). Note the *parameter independent* function $f(s)$ describing the shape of the curve, while $\mathcal{G}(s/s_{\text{naïve}})$ is a sharp cutoff function. Eq. (4.43) does not allow for an exponent, $f(s)$ is an arbitrary function. Writing it as

$$f(s) = s^{-\tau}(a_0 + \text{higher order corrections}) \tag{4.44}$$

defines τ to be the steepest descent of this part of the curve and gives a value between $\tau_{\text{stp.}} = 2.22$ and $\tau_{\text{stp.}} = 2.19$ (see Fig. 4.16).

This concept appears to be rather naïve – on the other hand, it is hard to assume

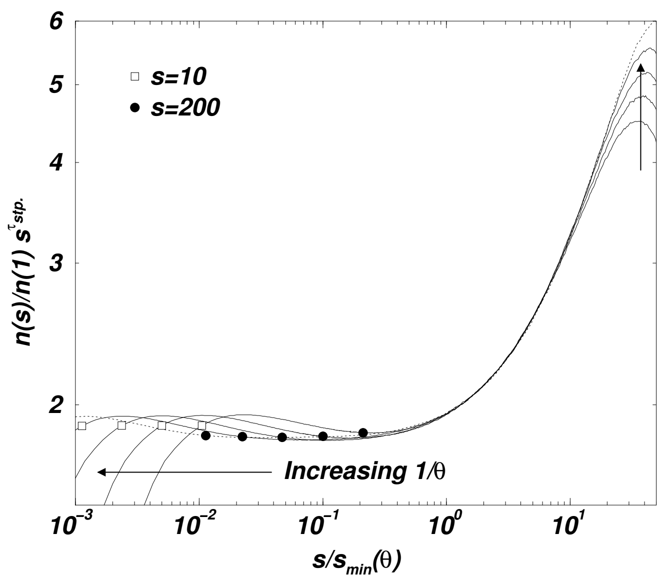


Figure 4.20: The rescaled and binned histogram $\bar{n}(s;\theta)s^{\tau_{\text{stp.}}}$, versus $s/s_{\text{min}}(\theta)$, where $\tau_{\text{stp.}} = 2.19$ for $\theta^{-1} = 4000, 8000, 16000, 32000, 64000$ in a double logarithmic plot. The scales $s_{\text{min}}(\theta)$ by which the histograms have been shifted are slightly different from the minima marked in Fig. 4.16, to make the collapse as good as possible. The squares and the filled circles mark $s = 10$ and $s = 200$, respectively, for orientation and relation to other figures. The arrows indicate the order of the data in increasing θ^{-1}.

that (4.12) can still hold: it would correspond to (4.43) with $f(s)$ replaced by s^τ, which is a straight line in a double logarithmic plot. Therefore (4.12) can apply only to a region in Fig. 4.16 where the data that fall on top of each other form a straight line. Those features not already collapsing would then collapse when properly tilted (choosing the right τ) and shifted (choosing the right s_0). Introducing a lower cutoff at $s = 10$ and discarding the data for $\theta^{-1} \leq 2000$ then leads to a data collapse in a narrow range as shown in Fig. 4.20. It is worthwhile mentioning that even for some $10 < s < 200$, namely for values of s between the squares and the filled circles, none of the data collapse. The exponent used in this "collapse" is $\tau_{\text{stp.}} = 2.19$, as mentioned above.

By considering the function $f(s)$ it becomes apparent that $\bar{n}$, and therefore the model, cannot be scale free: it depends on the fixed, microscopic scale $s = 1$. This entails that it is always possible to tell θ^{-1} by looking only at the *shape* of $\bar{n}$; a diagram showing only this shape, without any scales on the axes, reveals θ^{-1}, since a scale is intrinsically given by the features of $f(s)$. One would only need to rescale and tilt it until it fits the plot Fig. 4.16 and one could identify θ^{-1}. Only if $f(s)$ were scale free, i.e. a straight line in a double logarithmic plot, would this not be possible.

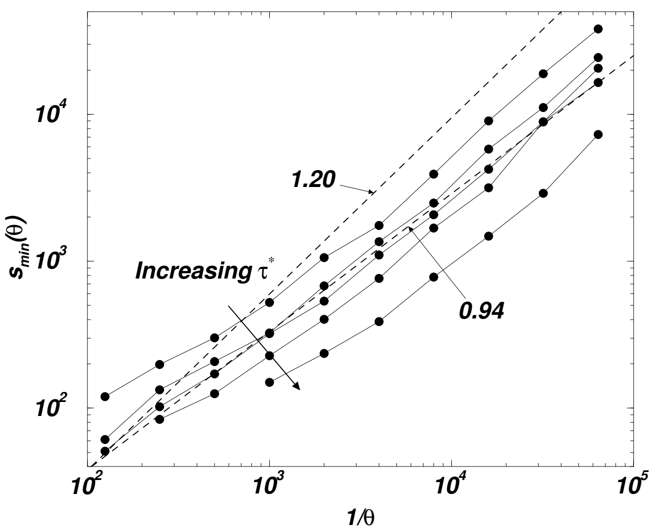


Figure 4.21: The position of the minimum in the binned and rescaled histogram for different values of $\tau^* = 2.04, 2.08, 2.10, 2.12, 2.16$. The exponents shown in the plot are for comparison only.

4.3.3 Two length scales

That $\bar{n}$ contains features to define at least two scales, which apparently diverge in θ^{-1} with different exponents, becomes clear when analysing the scaling of the minima and maxima as marked in Fig. 4.16, using the definitions

$$s_{\min}(\theta) \propto \theta^{-x_{\min}} \tag{4.45}$$

$$s_{\max}(\theta) \propto \theta^{-x_{\max}} \quad . \tag{4.46}$$

Of course, the concrete position of the extrema of $\bar{n}(s;\theta)s^{\tau^*}$ depends on its tilt, i.e. on the choice of τ^*. However, their *scaling* in θ^{-1} does not depend strongly on this choice. In particular $x_{\min}$ and $x_{\max}$ are different for all choices of τ^*. A plot of $s_{\min}(\theta)$ versus θ^{-1} for different values of τ^* is shown in Fig. 4.21. For small values of θ^{-1} the minimum is not pronounced enough to survive for large values of τ^*, so these curves do not give a data point. Using a linear fit of $\log s_{\min}(\theta)$ versus $\log \theta^{-1}$ of the minimum as found in the rescaled (τ^*) and binned histogram, gives an "exponent" between $x_{\min} = 0.93$ and $x_{\min} = 0.98$. The same procedure applied to the maxima gives an "exponent" in the range $x_{\max} = 1.18$ and $x_{\max} = 1.22$, shown in Fig. 4.22. One may expect that $x_{\min}$ tends towards $x_{\max}$ for decreasing τ^*, as $s_{\min}$ increases and might enter the scaling region of $s_{\max}$, but neither "exponent" exhibits a systematic variation, and the quality of the fit certainly suffers from the rough procedure that searches for the extrema in the *binned* histogram. This is unfortunately necessary because of statistical fluctuations, in conjunction with the absence of error bars for all

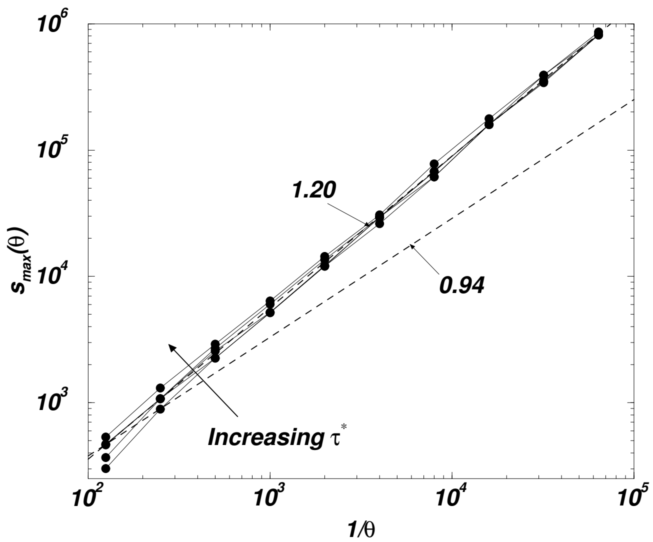


Figure 4.22: The position of the maximum in the binned and rescaled histogram for different values of $\tau^* = 2.04, 2.08, 2.10, 2.12, 2.16$. The exponents shown in the plot are for orientation only.

data points.

The scale of the clusters, $s_{\text{min/max}}$ is related to the correlation length ξ by the fractal dimension μ, i.e. [see (Clar *et al.*, 1994)]

$$s_{\text{min/max}} \propto \xi^{\mu_{\text{min/max}}} \quad . \tag{4.47}$$

Since $\xi \propto \theta^{-\nu}$, one should expect $\nu = x_{\text{min/max}}/\mu_{\text{min/max}}$. The minima are supposed to be dominated by smaller, fractal events [see (Schenk *et al.*, 2002)], so $\mu_{\text{min}} = 1.96(1)$ (Clar *et al.*, 1994) and therefore $\nu_{\text{min}} \in [0.47, 0.50]$. The maxima are more likely to be dominated by compact fires, so $\nu_{\text{max}} \in [0.59, 0.61]$. It is unclear how the two exponents $\nu_{\text{min/max}}$ are related exactly to the exponents of the two correlation lengths found by HONECKER and PESCHEL (Honecker and Peschel, 1997) for the connected correlation function $\nu = 0.576(3)$ and for the tree-tree correlation function $\nu = 0.541(4)$.

4.3.3.1 Finite size scaling

The failure of the DS-FFM to obey proper finite size scaling has been observed in (Schenk *et al.*, 2000) already. In the following some finite size scaling principles have been applied in a straight forward manner and subsequently ruled out.

Ignoring changes in $\mathcal{G}$ (see Sec. 2.3, page 78, especially page 80), one can incorporate finite size scaling and scaling in θ into a generalised form of the scaling behaviour of s_0

$$s_0(\theta, L) = \theta^{-\lambda} m\left(\theta L^{\sigma}\right) \tag{4.48}$$

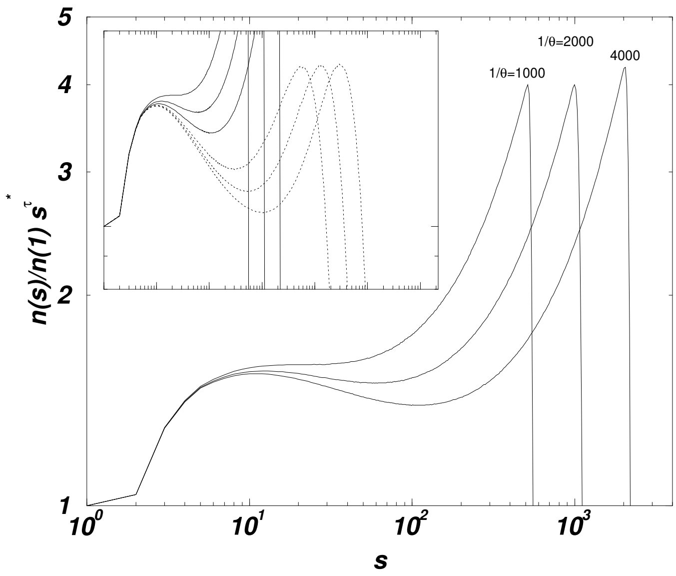


Figure 4.23: The rescaled and binned histogram $\bar{n}(s;\theta)s^{\tau^*}$ (again $\tau^* = 2.10$), for a modified model, where the largest cluster in the system is removed after each driving step, for $\theta^{-1} = 1000, 2000, 4000$ (as indicated) with linear sizes $L = 2000, 2000, 4000$. The inset shows the same data on the scale of Fig. 4.16 for comparison. The data for $\theta^{-1} = 1000, 2000, 4000$ of the original model as shown in Fig. 4.16 are dotted. The peculiar behaviour of the different height-scaling of the minimum and the maximum is again visible.

where $m(x)$ is a crossover function describing the dependence of s_0 on the two parameters θ and L. For sufficiently large argument x, the crossover function is expected to approach a constant, such that (4.20) is recovered. For small arguments, however, the dependence of the cutoff is expected to be strongly dominated by L, just like in equilibrium critical phenomena, where L takes over the rôle of ξ for sufficiently small systems. Thus, for small arguments $m(x) \propto x^\lambda$, so that for sufficiently small L, s_0 becomes independent of θ.

Generic models of SOC do not have any tuning parameters other than the system size, so that the cutoff s_0 is only a function of L. In this sense, finite size scaling is the only scaling behaviour in SOC, and a failure of the model to comply to finite size scaling is identical to the failure to comply to simple scaling altogether. Therefore, one might be surprised to see a simple scaling analysis *and* a finite size scaling analysis of an SOC model. However, the Forest Fire Model is different in this respect, as it has the additional parameter θ, which is, supposedly, finite only because of the finiteness of the system size. In the thermodynamic limit it supposedly disappears as a free parameter.

As seen above (see Fig. 4.16), the θ-dependence of $\bar{n}(s;\theta)$ can not be captured by s_0 in the scaling function alone. However, the scaling form (4.12) would remain valid

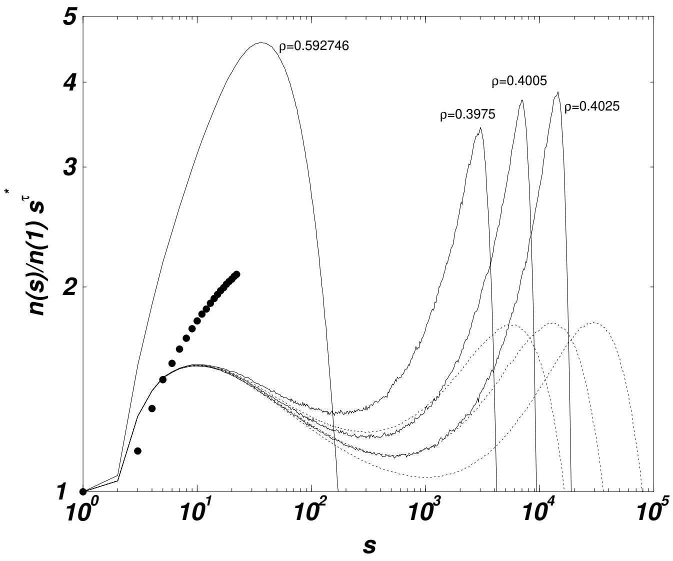


Figure 4.24: The rescaled and binned histogram $\bar{n}(s;\rho)s^{\tau^*}$ (again $\tau^* = 2.10$), for a modified model, where the largest cluster in the system is removed in each relaxation step and the corresponding number of trees is filled back into the system afterwards. The three small values of ρ chosen, $\rho = 0.3975, 0.4005, 0.4025$ correspond (up to the last digit) to the values of the tree density for $\theta^{-1} = 1000, 2000, 4000$ respectively, see Tab. 4.3. The linear size was $L = 1000, 2000, 4000$. The corresponding data of the original model are shown dotted. The peculiar behaviour of the different height-scaling of the minimum and the maximum is again visible (a correct tilt τ^* would make it even more pronounced), but disappears obviously for $\rho = \rho_{\text{perc}}$ – for these data it is relevant to mention that $\bar{n}(s)$ was measured *after* the relaxation. The filled circles show the exact results for the lattice animals (Stauffer and Aharony, 1994; Sykes and Glen, 1976; Mertens, 1990) at $\rho = \rho_{\text{perc}}$.

in some sense, if in the finite size scaling regime the L dependence of $\bar{n}(s;\theta)$ enters s_0 only. Therefore the original form (4.12) is generalised to

$$\bar{n}(s;\theta,L) = s^{-\tau}\mathcal{G}(s/s_0(\theta,L)) \tag{4.49}$$

ignoring that it has been shown above already that it does not hold in the limit where $\bar{n}(s;\theta,L)$ becomes independent of L. In this section the dependence of $\bar{n}(s;\theta,L)$ on L is investigated, in the limit of large θ^{-1} and small L. A similar study has been performed by SCHENK *et al.* (Schenk *et al.*, 2000), however on much smaller scales and using $\mathcal{P}^{\text{b}}$.

If the form (4.49) holds, it should be possible to collapse $\bar{n}(s;\theta,L)$ for different L by choosing the correct τ and s_0, just like for the cluster size distribution of standard percolation. This turns out not to be the case, as can be seen in Fig. 4.25: As *smaller* L, as *stronger* the changes of shape of $\bar{n}(s)$ for any θ tested. Consequently, (4.49) does not hold, and as s_0 is only *defined* via its rôle as cutoff in (4.49), s_0 is undefined and (4.48) remains meaningless.

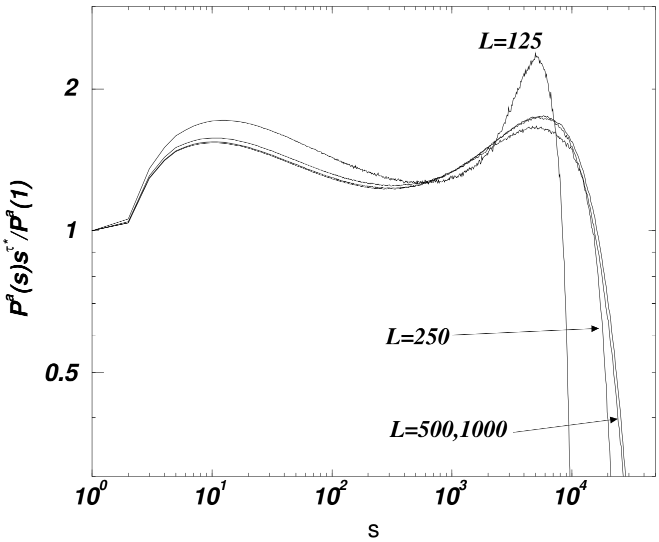


Figure 4.25: Plot of the rescaled PDF $\mathcal{P}^{\mathrm{a}}(s;\theta,L)s^{\tau^*}/\mathcal{P}^{\mathrm{a}}(1;\theta,L)$ for fixed $\theta^{-1}=1000$ and different system sizes, $L=125,250,500,1000$. The different shapes make it impossible to collapse the data, as would be expected from a finite size scaling ansatz (4.49) and (4.48).

One might argue that the average density of trees, $\bar{\rho}$ (see (4.4)), is the relevant parameter of $\bar{n}(s)$, so that $\bar{n}(s)$ has the same shape for different, sufficiently small L and constant $\bar{\rho}$. However, as shown in Fig. 4.26, for any value of θ, there is a value of L, such that $\bar{\rho}$ varies considerably with decreasing L. Especially, there seems to be a maximum tree density for every system size, so that for large values of $\bar{\rho}$, there is a smallest system size L, below which this density cannot be reached. This maximum increases monotonically with system size, so that the maximum for every finite system size is smaller than the expected average tree density in the thermodynamic limit, which is according to Tab. 4.3 (page 185) larger than 0.40777 and was recently conjectured to be as large as $0.5927\dots$ (Grassberger, 2002), namely the critical density of site percolation (Newman and Ziff, 2000). Accepting this limitation, Fig. 4.27 shows an example for three $\bar{n}(s)$ with roughly the same $\bar{\rho}$ and different L and θ. Most surprisingly two of the histograms collapse already without rescaling, while the third ($L=500$) reveals the same problems as visible in Fig. 4.16. Hence, finite size scaling does also not work for fixed $\bar{\rho}$.

That large densities of trees cannot be reached by small system sizes is related to the specific way the histograms are generated and the density measured: Is it before or after each (successful) burning? For sufficiently large systems, it becomes irrelevant when to do it, because two histograms, one measured before, the other one right after the burning, differ only by one cluster. Also the question, whether to average only over successful burnings is irrelevant, because the difference between a

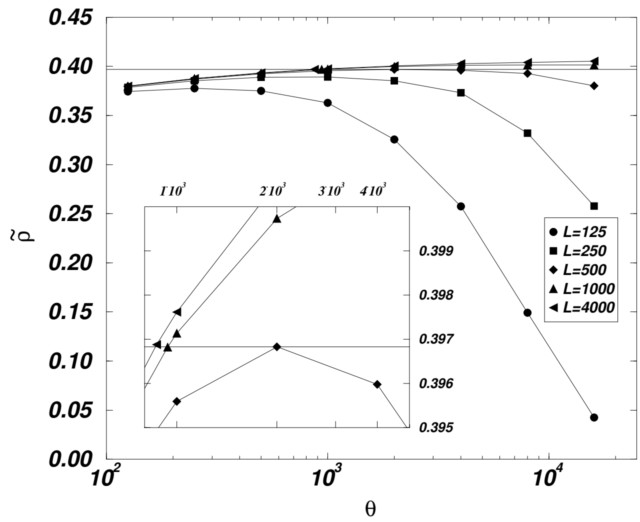


Figure 4.26: The average density of trees, $\bar{\rho}$, as a function of θ and for various L. For sufficiently small systems, the maximum in $\bar{\rho}$ is much smaller than the expected density at the "critical point", which is larger than 0.40777 found as in Tab. 4.3 (page 185). The straight line marks $\rho = 0.396827$, the density chosen in Fig. 4.27. The inset is a magnification of the crossing of the straight line with the simulation data, and shows all three values of θ, L used in Fig. 4.27.

histogram before and after the burning is only one cluster.

Clearly, for small systems, the difference between the histogram before and after the burning, is just the one enormous cluster of size $\mathcal{O}(\theta^{-1})$. Fig. 4.28 shows the difference. Even though in principle every density is reachable for every system size if the histogram is measured before burning, the newly defined histograms do not have a considerably different shape, so that a collapse remains impossible. For example, the problems shown in Fig. 4.25 become even more pronounced, if the histogram is taken before burning.

Surprisingly and actually in contradiction to what has been said in Eq. (4.6), there is a discrepancy between the cluster size distribution of burnt clusters, $\mathcal{P}^{\mathrm{b}}$, and the overall cluster size distribution $\mathcal{P}^{\mathrm{a}}$, even if the latter is measured *before* the burning takes place. This sounds paradoxical, because the random picking of a cluster to be burnt is just a sampling of $\mathcal{P}^{\mathrm{a}}$. This cannot be caused by the correlation between those samples, due to the fact that $n_{t+1}(s)$ is actually a function of the cluster chosen at t — a correlation like this would be equally picked up by $\mathcal{P}^{\mathrm{a}}$. The reason for this discrepancy is the fact that a site picked randomly as the starting point of the next fire is necessarily occupied. Therefore $n_t(s)$ with a low occupation density enter $\mathcal{P}^{\mathrm{b}}$ overweightedly. As low density states contain much more small clusters then large ones, $\mathcal{P}^{\mathrm{b}}$ overestimates the probability of small clusters. A sample of $\mathcal{P}^{\mathrm{b}}$ at a low density is indistinguishable from a sample at high density, while a sample for $\mathcal{P}^{\mathrm{a}}$ trivially

θ^{-1}	L	$n(1)$	$\widetilde{s}$	$\bar{\rho}$	$\frac{1-\bar{\rho}}{\theta\bar{\rho}}$
125	*1000*	0.04553	204.07	0.37973	204.18
125	1000	0.04552	203.81	0.37977	204.15
125	*4000*	0.04553	203.88	0.37983	204.10
125	**4000**	0.04552	203.77	0.37983	204.10
250	*1000*	0.04451	395.03	0.38756	395.06
250	1000	0.04452	394.08	0.38750	395.15
250	*4000*	0.04454	394.97	0.38766	394.89
250	**4000**	0.04454	395.29	0.38765	394.91
500	*1000*	0.04380	764.73	0.39316	771.75
500	1000	0.04380	764.81	0.39315	771.77
500	*4000*	0.04382	771.12	0.39343	770.88
500	**4000**	0.04382	771.90	0.39343	770.87
1000	*1000*	0.04328	1495.36	0.39716	1517.91
1000	1000	0.04328	1490.05	0.39714	1518.00
1000	*4000*	0.04331	1510.85	0.39761	1515.00
1000	**4000**	0.04331	1513.13	0.39764	1514.81
1000	*8000*	0.04332	1510.10	0.39763	1514.91
2000	*4000*	0.04296	2976.34	0.40053	2993.35
2000	**4000**	0.04297	2990.50	0.40054	2993.15
2000	*8000*	0.04297	2995.67	0.40060	2992.56
4000	*4000*	0.04273	5929.24	0.40258	5935.91
4000	4000	0.04273	5930.97	0.40249	5938.03
4000	*8000*	0.04274	5931.32	0.40261	5935.15
4000	**8000**	0.04273	5935.36	0.40256	5936.47
8000	*4000*	0.04255	11786.97	0.40405	11799.72
8000	4000	0.04255	11788.90	0.40406	11799.07
8000	*8000*	0.04257	11801.31	0.40412	11795.98
8000	**8000**	0.04257	11792.82	0.40413	11795.38
16000	*4000*	0.04244	23430.01	0.40525	23481.82
16000	*8000*	0.04243	23466.93	0.40540	23467.22
16000	8000	0.04243	23446.10	0.40542	23465.64
16000	**16000**	0.04245	23449.31	0.40541	23466.57
32000	16000	0.04232	46443.83	0.40660	46701.82
32000	**32000**	0.04233	46731.44	0.40662	46698.51
64000	**32000**	0.04220	91148.64	0.40777	92952.40

Table 4.3: Parameters and results for different choices of L and θ^{-1}. The average cluster size is denoted by $\widetilde{s}$, for definition see (4.1), but due to a truncation in the histogram for some of the simulations in the range $2000 \leq \theta^{-1} \leq 16000$, the number presented is actually the average size of the burnt cluster. In the stationary state it is — apart from small corrections — also given by $(1-\bar{\rho})/(\theta\bar{\rho})$, see (4.7). Values of θ^{-1} and L printed in bold indicate results shown in Fig. 4.16, the other results are only for comparison. All data are based on $5 \cdot 10^6$ (successful) updates (see Sec. 4.2.2.1) for the transient and statistics, apart from those printed in italics which are based on short runs ($5 \cdot 10^6$ updates for the transient and $1 \cdot 10^6$ updates for statistics).

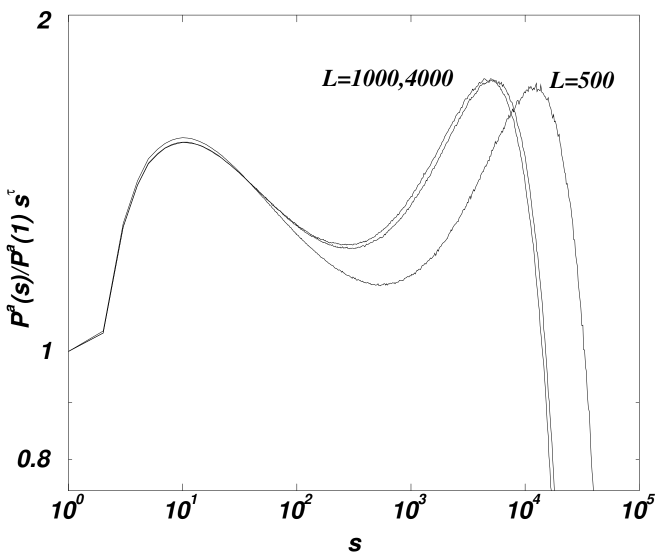


Figure 4.27: Plot of the rescaled PDF $\mathcal{P}^{\mathrm{a}}(s;\theta,L)s^{\tau^*}/\mathcal{P}^{\mathrm{a}}(1;\theta,L)$ for fixed $\bar{\rho}\approx 0.397$: $L=500$ with $1/\theta=2000$ ($\bar{\rho}=0.396827$), $L=1000$ with $1/\theta=940$ ($\bar{\rho}=0.396825$) and $L=4000$ with $1/\theta=870$ ($\bar{\rho}=0.396883$). Again, a data collapse is impossible.

contains the information about the density. To illustrate that, one might imagine a sequence of configurations, which consists of one state, with exactly one cluster of size 1, and a second state, with exactly one cluster of size L^2. The two configurations appear with a frequency such that a cluster of size 1 is burnt down as often as a cluster of size L^2. The resulting $\mathcal{P}^{\mathrm{a}}$ reports that a randomly chosen site belongs to a cluster of size L^2 with probability $\frac{1}{2}$ and to a cluster of size 1 with probability $1/(2L^2)$, while $\mathcal{P}^{\mathrm{b}}$ incorrectly reports the same probability for both cluster sizes. The problem can actually already be spotted in (4.6), which contains a ρ on the RHS, which should rather be $\rho(t)$. The problem disappears in the limit where $\rho(t)$ hardly changes in time, i.e. in the limit of $\theta^{-1}\ll L^2$.

It is also clear, why (4.7) breaks down for small systems and large θ^{-1}: The average size of the burnt cluster tends to L^2, while the density tends to 0. Apparently (4.7) must be incorrect for $\rho<(L^2\theta+1)^{-1}$.

4.3.3.2 Scaling of the moments of $\mathcal{P}^{\mathrm{a}}$

According to (4.12), (4.20) and (4.8) the nth moment of $\mathcal{P}^{\mathrm{a}}$ should scale like (this analysis has apparently been introduced to SOC by DE MENECH *et al.* (Pastor-Satorras and Vespignani, 2000a; De Menech *et al.*, 1998; Tebaldi *et al.*, 1999), see also Sec. 2.4, page 92)

$$\widetilde{s^n}=\frac{\sum_s s^n s\bar{n}(s;\theta)}{\sum_s s\bar{n}(s;\theta)}=q_n\theta^{-\lambda(2+n-\tau)}+\text{ corrections}\quad, \tag{4.50}$$

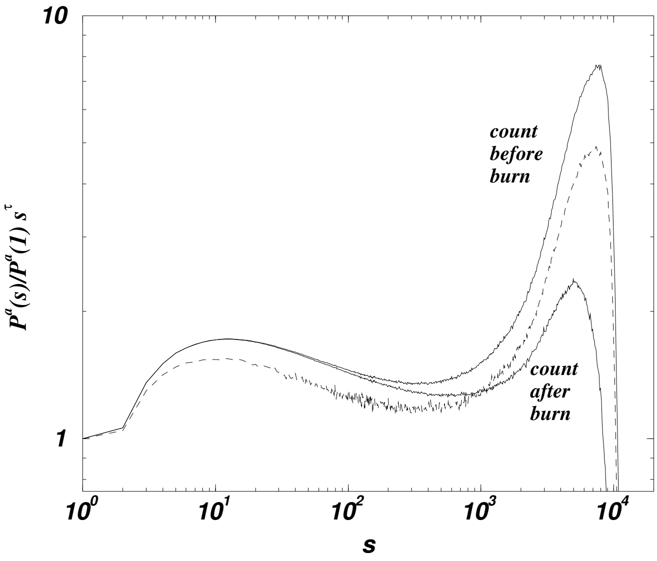


Figure 4.28: Comparison between the rescaled and binned histograms measured before and after the burning for small $L = 125$ and large $\theta^{-1} = 1000$. As expected, only the statistics for large s is affected. The dashed line shows the data for $\mathcal{P}^{\mathrm{b}}(s)$.

where q_n is a non-universal amplitude (see Sec. 4.3.3.3) and λ is also known as a gap exponent (Pfeuty and Toulouse, 1977). The corrections are due to the lower cutoff and the asymptotic character of the scaling, which is expected only for "sufficiently small θ" (Wegner, 1972). In turn, one can infer a scaling form like (4.12) if the moments scale in the form of (4.50).

Contrary to what is observed in an attempt of a data collapse, it turns out that the moments follow beautifully this scaling behaviour. Fig. 4.29 shows the scaling of the moments for $n = 2, 3, 5, 10$. By simply fitting the double logarithmic data to a straight line, i.e.

$$\log(\widetilde{s^n}) = a'_n - \sigma_n \log(\theta) \tag{4.51}$$

one can derive an estimate of the exponents σ_n and in turn compare them to the expected linear behaviour:

$$\sigma_n = \lambda(2 + n - \tau) \quad . \tag{4.52}$$

The resulting estimates, using $n = 2, \ldots, 8$ and $\sigma_1 = 1$ from (4.1), are $\lambda = 1.0808\ldots$ and $\tau = 2.0506\ldots$, where no statistical error is given because the systematic error, due to neglecting of the lower cutoff as well as the corrections (4.50), is expected to be much more important. By using the assumption $\sigma_1 = 1$, this result is consistent with (4.21). The results are shown in Fig. 4.30.

The exponent found for τ is remarkably close to the accepted value of standard 2D percolation, $187/91 = 2.054945\ldots$. However, if one leaves out the results for $\theta^{-1} = 64000$, which seem to be a bit off the lines shown in Fig. 4.29, one finds a

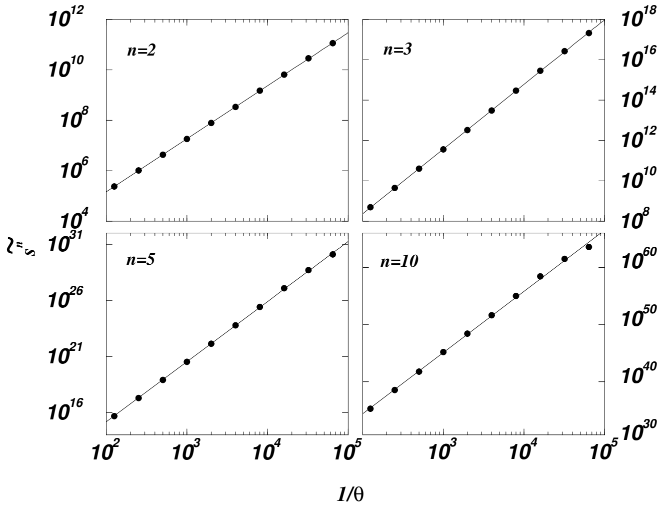


Figure 4.29: Scaling of the nth moments of $\mathcal{P}^{\mathrm{a}}$ in double logarithmic plots. The straight lines show the results of a fit as $\exp(a'_n)\theta^{-\sigma_n}$, see (4.50).

slightly larger value for the exponent, namely $\tau = 2.0864$ and $\lambda = 1.0998\ldots$. This is much closer to the $\tau^* = 2.10$ used above. For comparison to values found in the literature, see Tab. 4.4.

It is very remarkable that the resulting estimates for the exponents are so impressively consistent, even though in Sec. 4.3.1 it turned out, that the scaling assumption (4.12) does not actually hold; one would much rather expect a failure of the moments to comply with (4.50), or a failure of the exponents to comply with (4.52). Apparently the moments are hiding the breakdown of simple scaling. Therefore it is interesting to analyse the behaviour of the presumably universal amplitude ratios, which are solely a property of the (presumed) scaling function.

Another explanation for the moments being well behaved is the following: According to (Pruessner and Jensen, 2002a) one might expect the moments to behave like

$$\int_1^{\theta^{-x_{\min}}} ds f(s) s^n + \int_{\theta^{-x_{\min}}}^{\infty} ds s^{n-\tau} \mathcal{G}(s/\theta^{-x_{\max}}) \tag{4.53}$$

where the first integral describes the behaviour up to the minimum, which scales like θ^{-xmin} ($x_{\min} \approx 0.95$) and the second integral the behaviour from the minimum on. Because Fig. 4.16 indicates already that the scaling function $\mathcal{G}$ does not collapse using a scale $\theta^{-x_{\max}}$ this scaling does not work and can therefore be only an approximation. While the first integral is bound by $\mathcal{O}(\theta^{-x_{\min}(n+1)})$ the second integral gives $\mathcal{O}(\theta^{-x_{\max}(1+n-\tau)})$ asymptotically, which dominates the moments for $x_{\min}(n+1) < x_{\max}(1+n-\tau)$, which leads to $n > 9.08$ using $x_{\max} \approx 1.2$ and $\tau \approx 2.1$. Fig. 4.29 shows clearly a deviation from the straight line behaviour for $\theta^{-1} = 64000$ and $n = 10$ and even for $n = 5$. It remains unclear whether this is due to the effect

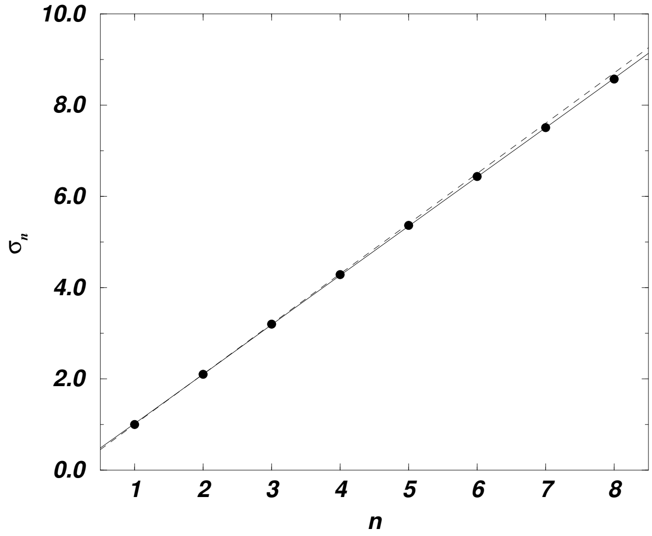


Figure 4.30: Exponents σ_n of the scaling of $\widetilde{s^n}$ in θ vs. n. The slope of this curve gives λ and τ can be derived from the offset. The straight, full line shows the results $\lambda = 1.0808\ldots$ and $\tau = 2.0506\ldots$, the dashed line shows $\lambda = 1.0998\ldots$ and $\tau = 2.0864\ldots$ from a fit excluding $\theta^{-1} = 64000$.

discussed or simply a finite size problem. According to the findings presented in Sec. 4.3.3.3 the latter might well be the case.

It is worthwhile pointing out, that the analysis in this section arrives at estimates for the critical exponents very close to those obtained by PASTOR-SATORRAS and VESPIGNANI (Pastor-Satorras and Vespignani, 2000a), who, however, allow for the corrections in (4.50) which were omitted above.

reference	method	τ	λ
(Christensen *et al.*, 1993)	$P(s)$	$2.16(5)$	–
(Henley, 1993)	$P(s)$	$2.150(5)$	$1.167(15)$
(Grassberger, 1993)	$P(s)$	$2.15(2)$	$1.08(2)$
(Clar *et al.*, 1994)	$P(s)$	$2.14(3)$	$1.15(3)$
(Honecker and Peschel, 1997)	$P(s)$	$2.159(6)$	$1.17(2)$
(Pastor-Satorras and Vespignani, 2000a)	moments	$2.08(1)$	$1.09(1)$
(Schenk *et al.*, 2002)	theoretical and $P(s)$	$2.45\ldots$	1.1
(Grassberger, 2002)	$P(s)$	2.11	1.08

Table 4.4: Exponents of the Forest Fire model found in the literature. The first column indicates the source, the second column the method. $P(s)$ denotes a direct analysis of $\bar{n}(s;\theta)$, which sometimes may have been just an estimate of the slope of $\bar{n}(s;\theta)$ rather than a data collapse. For details the original sources should be consulted. The entry "moments" refers to an analysis of the moments of $P(s)$, the entry "theoretical" to theoretical considerations regarding the relation of the Forest Fire model to percolation.

4.3.3.3 Universal amplitude ratios

In general simple scaling involves two additional non-universal parameters a and b,

$$\bar{n}(s;\theta) = as^{-\tau}\mathcal{G}\left(\frac{s}{bs_0}\right) \quad . \tag{4.54}$$

For $1 < \tau < 2$ the lower cutoff becomes asymptotically irrelevant compared to the upper cutoff for all moments $n \geq 1$ — indeed the effective τ of $s\bar{n}(s;\theta)$ fulfils this condition as $2 < \tau < 3$ (Clar *et al.*, 1996). Neglecting the lower cutoff then gives for the nth moment of $s\bar{n}(s;\theta)$

$$\widetilde{s^n} = a(bs_0)^{1+n-\tau}g_n \tag{4.55}$$

with

$$g_n \equiv \int_0^\infty dx x^{1+n-\tau}\mathcal{G}(x) \tag{4.56}$$

In order to construct universal amplitude ratios, one needs to get rid of all exponents and parameters. This can be achieved by considering

$$\frac{\widetilde{s^n}}{\widetilde{s^2}^{n/2}} = \left(a\left(b\theta^{-\lambda}\right)^{(1-\tau)}\right)^{1-n/2}\frac{g_n}{g_2^{n/2}} \tag{4.57}$$

and (4.57) with $n = 1$)

$$\frac{\widetilde{s}}{\widetilde{s^2}^{1/2}} = \left(a\left(b\theta^{-\lambda}\right)^{(1-\tau)}\right)^{1/2}\frac{g_1}{g_2^{1/2}} \quad . \tag{4.58}$$

If one now multiples (4.57) with the $n-2$th power of (4.58), everything cancels apart from the g_n:

$$\frac{\widetilde{s^n}}{\widetilde{s^2}^{n/2}}\frac{\widetilde{s}^{n-2}}{\widetilde{s^2}^{(n-2)/2}} = \frac{g_n g_1^{n-2}}{g_2^{n-1}} \tag{4.59}$$

It is worth noting that for a trivial case, where $\widetilde{s^n} \propto \widetilde{s}^n$, the effective exponent τ is necessarily unity and (4.57) as well as (4.58) are already independent of θ.

A further simplification is to impose $g_1 = 1$ and $g_2 = 1$, which fixes the two free parameters a and b in (4.54), so that

$$g_n = \frac{\widetilde{s^n}\widetilde{s}^{n-2}}{\left(\widetilde{s^2}\right)^{(n-1)}} \tag{4.60}$$

for $n \geq 1$. In Fig. 4.31 this quantity is shown for $n = 3, 4, 5, 6$. Now, for $\theta^{-1} = 64000$ a deviation is clearly visible — in turn that means that $\theta^{-1} = 64000$ requires at least systems of the size $L = 64000$, which might explain the large value of $\bar{\rho}$ obtained in (Grassberger, 2002). Apart from that, this analysis agrees with the result found in

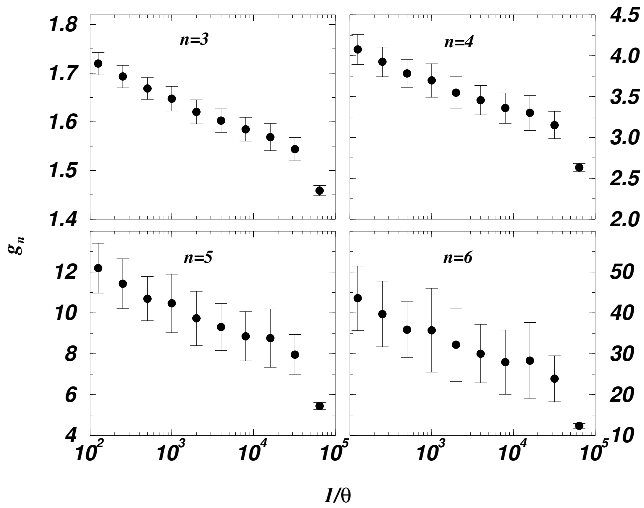


Figure 4.31: The supposedly universal amplitude ratio g_n (4.60) for $n = 3, 4, 5, 6$. The error bars are based on a Jackknife scheme (Efron, 1982; Pruessner *et al.*, 2001) using a roughly estimated correlation time of 50, see Tab. 4.2.

Sec. 4.3.1: The supposedly universal amplitude ratios keep changing with θ and an asymptote cannot be estimated, i.e. the scaling (4.12) is broken.

4.3.3.4 Burning time distribution

Another distribution of interest is the distribution of burning times, $\mathsf{P}_{T_\mathrm{M}}(T_\mathrm{M};\theta)$. The statistics are comparatively small for this quantity, as the burning time is defined only for the cluster removed. However, they still seem to be good enough to allow us to make a statement about their scaling behaviour. The rescaled data, $\mathsf{P}_{T_\mathrm{M}}(T_\mathrm{M};\theta){T_\mathrm{M}}^{b^*}$ with a trial exponent $b^* = 1.24$ can be seen in Fig. 4.32. The intermediate part of the distribution between $T_\mathrm{M} = 4$ and the maximum seems to bend down as θ^{-1} increases, but the developing dip is much less pronounced than in Fig. 4.16. Nevertheless, the region where a data collapse seems possible moves out towards larger values of T_M, which again prohibits simple scaling. Assuming that the bending might become weaker for sufficiently large T_M leads to a data collapse shown in Fig. 4.33, using an exponent $\nu' = 0.6$ as defined in Eq. (4.22). However, only for values of $T_\mathrm{M} \approx T_{\mathrm{M}0}$ the data possibly collapse. Again, this violates the assumption of simple scaling, namely that there is a *constant* lower cutoff above which the behaviour is universal.

The only remaining exponent of those defined in sec. 4.2.2.4, μ', relates the statistics of s and T_M. It requires the bivariate distribution $\mathsf{P}(s, T_\mathrm{M};\theta)$, as the exponent is derived from $\mathsf{E}(T_\mathrm{M}|s) \propto s^{1/\mu'}$, which is essentially equivalent to Eq. (4.17). The distribution $\mathsf{P}(s, T_\mathrm{M};\theta)$ is shown in Fig. 4.34. At first glance the assumption of a power law dependence of s and T_M seems to be confirmed. Also the width of the distribution

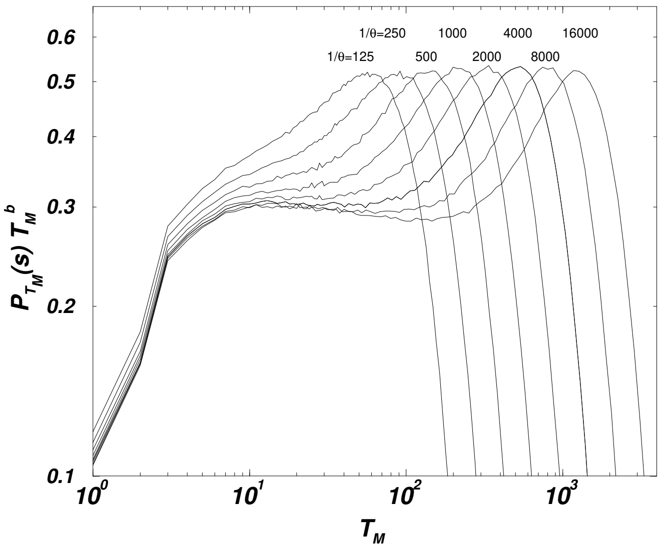


Figure 4.32: The rescaled probability distribution of the burning time, $\mathsf{P}_{T_\mathrm{M}}(T_\mathrm{M};\theta)$. Similar to Fig. 4.16 a dip seems to form between the low T_M region and the maximum, which again renders a data collapse impossible.

seems to be very small, with almost no change over 5 orders of magnitude in s. However, the plot is double logarithmic, so that the width roughly scales like the slope, which is about 0.6, as shown by straight lines. This matches perfectly the exponent chosen to rescale P (see caption of Fig. 4.34).

By inspecting $\mathsf{E}(T_\mathrm{M}|s;\theta)$ and $\mathsf{E}(s|T_\mathrm{M};\theta)$ for various θ, one can determine μ' as slope in a double logarithmic plot. Fig. 4.35 shows that μ' remains ambiguous and deviations from the expected behaviour do not vanish as θ^{-1} is increased. Asymptotically one might expect $1/\mu' \approx 0.62$, while $(\tau^* - 2)/(b^* - 1)$ suggests $1/\mu' \approx 0.417$. The value of 0.62 is consistent with the rough estimate 0.6 made in Fig. 4.34. Fig. 4.35 also shows two other exponents, 0.53 and 0.7, the former being in line with the value found in literature of $0.529(8)$ (Clar *et al.*, 1994).

Conclusively it is noted that the other observable available in this study, T_M, does not seem to provide an alternative way to ascribe the DS-FFM critical behaviour in the sense of the scaling behaviour as proposed in the literature.

4.3.4 Tree density as a function of time

As mentioned above (see Sec. 4.3.3.1), the density of trees, $\bar{\rho}$, is actually a function of time. Initially, it is periodic around the average value, with an amplitude that depends mainly θ. This amplitude decays in time and after sufficiently long times $\rho(t)$ looks like a random walk around $\bar{\rho}$.

Fig. 4.36 illustrates how the period and the amplitude depends on θ and L: The

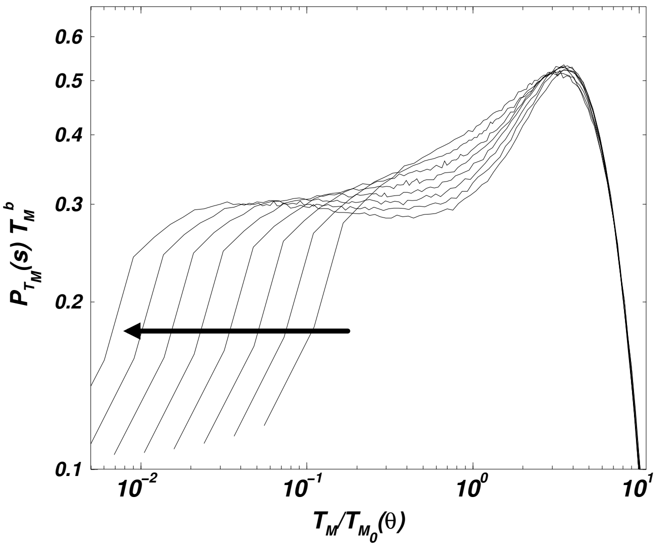


Figure 4.33: Attempt of a data collapse for $\mathsf{P}_{T_M}(T_M;\theta)$. Only at the far end of the scaling function at the descent from the maximum, the data seem actually to collapse. This, however, is not sufficient for a data collapse. The big arrow points in the direction of increasing θ^{-1}.

period is proportional to θL^2, while the amplitude mainly depends on θ, i.e. the strength of the influx $\propto \theta^{-1}$. The reason for the former is easy to understand: θ^{-1}/L^2 is proportional to the fraction of newly grown trees (Honecker and Peschel, 1997); the change of the tree density is roughly

$$\frac{d}{dt}\rho = \frac{1-\rho}{\rho}\frac{1}{\theta L^2} - \eta(\rho(t), t) \tag{4.61}$$

assuming that it hardly changes during the growing. Otherwise, one would have to introduce a microscopic timescales, which make it possible to measure of tree density on the timescale on which the trees are grown. The pre-factor $(1-\rho)/\rho$ takes into account that only empty sites can be re-occupied and that an occupied site is required for the burning to start. The second term on the right hand side, $\eta(\rho(t), t)$, is a noise, which represents the burning of the trees. From this equation one can already expect that the period is roughly linear in $\theta L^2 \bar{\rho}/(1-\bar{\rho})$. This has already been measured in detail by HONECKER and PESCHEL (Honecker and Peschel, 1997); the numerical results presented here (Fig. 4.36) are fully consistent with their results.

Apart from the relevance of the periodic behaviour for the equilibration time, the periodic behaviour of $\rho(t)$ is physically of great significance: What distinguishes the state of the system for a given ρ at the ascending and the descending branches? Trivially, the sequence of configurations of the system is Markovian, while the tree density alone as a time series is certainly not. The configuration somehow manages to “remember” whether the tree density was increasing or decreasing during the last

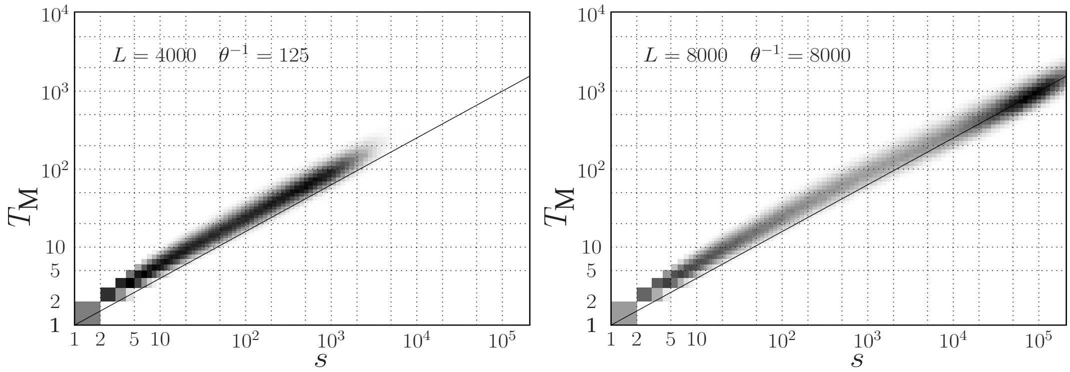


Figure 4.34: Binned density plots of $\mathsf{P}(s, T_\mathrm{M}; \theta)$ for different values of θ on a double logarithmic scale. High densities are presented as dark areas. For better presentation, $\mathsf{P}(s, T_\mathrm{M}; \theta)$ has been multiplied by a factor $s^{1.7}$, tilting the distribution similar to those shown in Fig. 4.16, so that the second maxima in the distribution, those at large s and T_M, are roughly as high as the first maxima, i.e. they show in the plot as dark as around $s = 5$. Since $\mathsf{P}(s, T_\mathrm{M}; \theta)$ is a histogram only of burnt clusters, it contains a factor s compared to $\bar{n}(s)$ (see discussion around (4.4)). Therefore, the exponent 2.7 needs to be compared to $\tau^* = 2.10$, indicating that the width of $\mathsf{P}(s, T_\mathrm{M}; \theta)$ roughly scales like $s^{0.6}$, so that the reduced height of $\mathsf{P}(s, T_\mathrm{M}; \theta)$ is caused by an increase in width. This coincides well with the slope of the distribution, as shown by a straight line. Thus, the relative width remains roughly constant.

update, in order to keep $\rho(t)$ periodic.

One explanation for this behaviour might be a "growing-and-harvesting" concept: From the initially completely random tree distribution larger and larger patches are formed, so that larger and larger patches are harvested by lightning. When the density reaches the maximum, for a while the patches harvested remain large compared to the amount grown. This is because the growing process does not actually produce those large patches itself, but makes them available to the harvesting by continuously connecting smaller patches in areas, where the lightning has not yet struck. This process goes on, until almost all the trees are newly grown, i.e. the trees are distributed almost randomly, apart from the spatial correlation in density. The period of this process would be proportional to the time it takes to renew the entire system, which is $L^2\theta\bar{\rho}/(1-\bar{\rho})$, namely L^2 divided by $\widetilde{s}$, see (4.7).

The time-dependent tree density gives only a hint of what actually happens in the system. It would be very instructive to study the two-point correlation function as a function of time to answer the question, whether the explanation above is actually valid.

4.3.5 Discussion

Prima facie the DS-FFM looks like a percolation process, and one might naïvely think that it is indeed a percolation process which organises itself to the critical den-

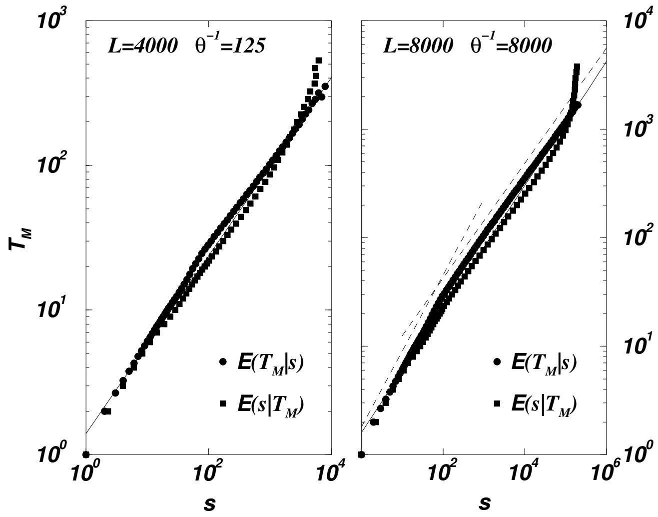


Figure 4.35: $\mathsf{E}(T_{\mathrm{M}}|s;\theta)$ and $\mathsf{E}(T_{\mathrm{M}}|s;\theta)$, based on the binned histogram $\mathsf{P}(s, T_{\mathrm{M}};\theta)$ for different values of θ^{-1}. The straight lines in the plots are $1.4s^{0.615}$ for $\theta^{-1} = 125$ (left hand plot) and $1.6s^{0.57}$ for $\theta^{-1} = 8000$. The two dashed lines in the right hand plot show alternative exponents $1/\mu' = 0.7$ and $1/\mu' = 0.53$, which are consistent with data either for small values of s or for large values.

sity: sites are occupied randomly and independently and (at least in the thermodynamic limit) there is only one cluster which is removed with non-vanishing probability, namely the largest. In this way the density of occupied sites is automatically reduced below the percolation threshold whenever the threshold is reached. It is puzzling how remarkably close the tree density in the DS-FFM is to the density of *empty* sites in critical site percolation on a square lattice ($\rho_{\mathrm{FFM}} \approx 0.4078$ and $1 - \rho_{\mathrm{perc}} = 0.40725379(13)$ (Newman and Ziff, 2000) respectively). However, the removal process involved in the DS-FFM introduces spatial correlations which are not present in standard percolation. These correlations are expressed, for example, in the form of a patchy tree density distribution (Schenk *et al.*, 2002).

The purpose of this section is *not* to add yet another model to the enormous zoo of SOC models. However, in order to investigate certain features of the given model and identify underlying mechanisms, it makes sense to modify it slightly. The outcome for the histogram of the DS-FFM modified such that the *largest* cluster is removed after each driving step is shown for a few values of θ^{-1} in Fig. 4.23. The distinctive feature of a minimum which scales differently from the maximum is again present, as the peaks of the maxima have approximately the same height, while the height of the local minima varies among different values of θ. The inset of this figure shows the histogram on the same scale as Fig. 4.16 together with the data of the original model (dotted) with the corresponding values of θ. One can understand that they do not fall on top of each other, because the relaxation rule in the modified model erases

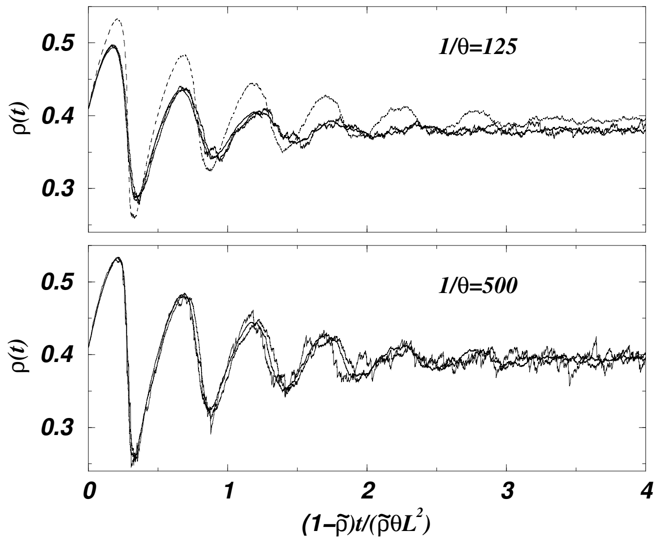


Figure 4.36: The density of trees as a function of time, plotted versus the rescaled time $(1 - \bar{\rho})t/(\theta\bar{\rho}L^2)$. Upper panel: Plot for $\theta^{-1} = 125$ and $L = 1000, 2000, 4000$ with an additional plot for $\theta^{-1} = 500$ and $L = 4000$ shown as dashed line, for comparison of period and amplitude. Lower panel: Same plot for $\theta^{-1} = 500$ and $L = 1000, 2000, 4000$.

much larger clusters than in the original model.

Fig. 4.24 shows a second modification of the model, where again the *largest* cluster is removed during relaxation and in addition the driving is changed such that the density of trees, ρ, is the same before each relaxation; the trees removed during the relaxation are just filled in randomly afterwards. This model differs from standard percolation only by its updating scheme [6]. In order to compare the outcome with the original model, the values of ρ have been chosen close to the values given in Tab. 4.3. Indeed, the feature of different scaling of the extrema is still present, but it disappears completely if the density is increased to $\rho_{\text{perc}} = 0.592746$ (Newman and Ziff, 2000), which is shown in the same figure as the large bump. This curve does not vary much if a much smaller system size is simulated at this density, so we expect it essentially to be free of finite size corrections. Since it represents a *correlated* percolation process, it is just consistent that this bump does not cover the exact results for the lattice animals of standard percolation (Stauffer and Aharony, 1994; Sykes and Glen, 1976; Mertens, 1990) at $\rho = \rho_{\text{perc}}$ shown as filled circles in Fig. 4.24. The dotted graphs in the figure show the corresponding data of the original model. Again they do not match apart from the region of very small s. Unfortunately the simulations of the so-modified model are very expensive in CPU time, because the mass of the

[6] Actually, it also differs from standard percolation because it fixes the number of occupied sites rather than simply the probability of being occupied. However, this difference becomes irrelevant for sufficiently large systems.

largest cluster needs to be refilled after each relaxation, so that only 50.000 updates for transient and statistics could be done.

Since the feature of different scaling survives the modifications described above, it seems reasonable to assume that any relaxation rule that favours the largest cluster leads to the peculiar behaviour. Its disappearance at high densities can be explained by the extremely small cutoff in the distribution, which leads to a domination of the statistics by very small clusters, while a single, enormously large one dominates the burning (the average size of the burnt cluster for $\rho = \rho_{\text{perc}}$ was 355811). However, much more careful and detailed investigations of models like the modification described above are required to gain a full understanding of the underlying mechanisms. In particular, this should include a modification of the rules such that the feature disappears.

HONECKER and PESCHEL (Honecker and Peschel, 1997) have calculated the correlation length not only for the probability that two sites belong to the same cluster, but also for the probability that two sites are occupied at all. The correlation function for the latter is of course a δ peak in ordinary percolation, as there are no spatial correlations for the distribution of occupied sites by construction. However, in the DS-FFM the correlation length for this quantity, ξ, is finite and seems to diverges when approaching the critical point. It is highly remarkable that HONECKER and PESCHEL conclude from their simulations that this correlation length diverges slightly *slower* than the correlation length of the probability for two sites to belong to the same cluster, ξ_s. This seems to indicate that for sufficiently large scales the spatial correlation of the occupation probability becomes arbitrarily small, so that on sufficiently large scales the DS-FFM occupation is uncorrelated and therefore tends to standard percolation. In other words, it seems to be possible to rescale or "renormalise" the DS-FFM to make the occupation correlation arbitrarily small. This would introduce higher order interactions, as known from standard real space renormalisation group and would explain the difference in critical density between the rescaled DS-FFM and standard percolation. However, if this "mapping" is valid, one should find the exponent for the divergence of ξ_s/ξ to be the same as in standard percolation, but this is precluded by numerics.

It has been suggested at least twice (Honecker and Peschel, 1997; Schenk *et al.*, 2002), that the DS-FFM is a superposition of cluster distributions $n_{\text{perc}(s,p)}$ of standard percolation for a whole range of concentrations p, weighted by a certain distribution function $w(p)$, i.e. $\int_0^1 dp\, w(p) n(s,p)$. Obviously such an assumption neglects spatial correlations. We recall the following result from standard percolation theory (Stauffer and Aharony, 1994),

$$n(s,p) \propto s^{-\tau} \mathcal{C}(-s/(p-p_c)^{-1/\sigma}) \quad , \tag{4.62}$$

where $\mathcal{C}$ denotes the cutoff function and the exponents σ and τ_{perc} have their standard definitions. Assuming that the weighting function $w(p)$ is analytic around the critical concentration in standard percolation, p_c, (4.62) leads to

$$\int_0^1 dpw(p)n(s,p) \propto s^{-(\tau_{\text{perc}}+\sigma)} \quad . \tag{4.63}$$

This gives rise to an exponent $\tau = 223/91 \approx 2.45$, however, this is definitely not supported by numerics (see Fig. 4.16).

It remains completely unclear how to characterise the scaling of the DS-FFM in two dimensions. Apparently it is not a mere superposition of two simple scalings, as recently speculated (Schenk *et al.*, 2002). Moreover the model does not seem to be scale free as described above and it does not seem to be possible to identify a unique power law behaviour of the cluster size distribution. Nevertheless *effective* power law behaviour over restricted regions has clearly been produced by the model, making it potentially relevant to observation.

Open problem

All we can conclude is that *the DS-FFM is not critical* in the sense of simple scaling. It reminds us that a divergent moment (here $\langle s \rangle$, the second moment) can be regarded as a unique sign of emergent scale invariance only if we are certain that one single scale is sufficient to characterise the system. If there is more than one relevant scale, different properties of the system might depend on different scales which may or may not diverge.

4.3.5.1 Alternative scaling approaches

From the results presented above it becomes clear that the Forest Fire model does not show the scaling behaviour expected for a system, which becomes critical in the appropriate limit (namely $L \to \infty$ and $\theta^{-1} \to \infty$). One might argue that another scaling ansatz could lead to a distribution which is asymptotically scalefree in this limit, for example a multi-fractal ansatz (Tebaldi *et al.*, 1999) or the one proposed in (Schenk *et al.*, 2002), where more than one scale is assumed to govern the model. For an asymptotically scalefree distribution, the scales have to diverge or to vanish in the appropriate limit. It has been suggested already very early (Honecker and Peschel, 1997) that more than one characteristic length scale can be found in the Forest Fire model.

However, changing the scaling assumption would entail a new *definition* of the exponents τ, D etc., which would therefore prohibit comparison with other results which are based on the assumption of simple scaling (4.12). Moreover, introducing multiple scales would stretch the notion of universality, especially the universality of the scaling function, to its limits. As can be seen in Fig. 4.16, the shape of the distribution function *is not universal*, i.e. the shape of this function is unique for every single

θ^{-1}, even for $L \to \infty$. This is in direct contradiction to the concept of universality, scaling and scale invariance.

However, it might be possible to reestablish simple scaling by introducing another mechanism in the model, as was done for example in the "autoignition Forest Fire model" (Sinha-Ray and Jensen, 2000). If there were, for example a mechanism parameterised by u, such that

$$\bar{n}(s;\theta,u) = s^{\tau}\mathcal{G}(s/s_0(\theta,u)) \tag{4.64}$$

then simple scaling might be reestablished possibly by choosing an appropriate $u = u(\theta)$; even the cutoff, s_0, which were assumed to diverge with θ^{-1}, would then effectively depend only on θ. Currently, there is no hint, what this new parameter u could be.

STEFANO LISE and MAYA PACZUSKI (Lise and Paczuski, 2001b) suggested for a similar problem in the OFC model (Olami *et al.*, 1992) to define an exponent τ by the slope of the distribution $\mathcal{P}^{\mathrm{a}}(s)$, imposing the remaining background, $\mathcal{F}(s, L, \theta^{-1})$, to be as straight as possible:

$$\ln\left(\mathcal{P}^{\mathrm{a}}(s)\right) = -\tau\ln(s) + \mathcal{F}(s, L, \theta^{-1}) \tag{4.65}$$

This ansatz, in fact based on a multiscaling ansatz, would indeed allow the measurement of an exponent, however, with some degree of ambiguity. The crucial problem with this approach is that, firstly, it again does not allow any direct comparison to other models, where the exponents are defined via (4.12) and that, secondly, the notion of a presumably universal exponent hides the fact of broken scaling.

From Sec. 4.3.1 one might conclude that there does not even exists a limiting distribution for $\bar{n}(s;\theta)$. However, even if it exists, that does not mean that simple scaling is obeyed and if it does, it is still open whether the exponents are non-trivial or not and whether the model posses any spatio-temporal correlation which do not vanish on sufficiently large scales.

4.3.6 Summary

Using a new method for simulating the Forest Fire model on large scales, it is possible to make clear statements about the validity of the scaling assumption of this model. The two observables investigated in this chapter suggest the model does not develop into a scale-invariant state.

The method is based on the HOSHEN-KOPELMAN algorithm (Hoshen and Kopelman, 1976) and uses a master/slave parallelisation scheme to simulate the model on very large scales and very large sample sizes. The key to the parallelisation is to decompose the lattice in strips and to encode the connectivity of these strips in the border sites. Clusters crossing these strips are then maintained by the master node, while clusters within a strip are maintained on the local nodes. There is almost no data exchange apart from the border configuration, which lowers the impact on the network linking the nodes.

The resulting distribution $\mathcal{P}^{\mathrm{a}}(s)$ is, different from other simulations found in the literature, the distribution of *all* clusters in the system, rather than just the burnt clusters. The resulting statistics then allows to draw clear conclusions as to what extend the model does actually obey the scaling assumption. This turns out not to be case. The violation of scaling is also observed in the distribution of the burning time. Conclusively we find that there is no reason to assume that the DROSSEL-SCHWABL Forest Fire model develops into a critical state. This is in line with the conclusion by GRASSBERGER (Grassberger, 2002), who however, still finds some signs that the Forest Fire model will finally show some characteristics of standard percolation.

4.4 Supplement: Relation to Real Forest Fires

The name of the Forest Fire model is unfortunate, as it might suggest relevance to real forest fires, as they are observed on a regular basis for example in the US. Clearly the FFM in the DROSSEL and SCHWABL version and to some extend even in the BAK, CHEN and TANG version resemble some features of a real forest fire. However, these dynamical models are by no means meant to model the processes of real forest fires, because the latter contain many more parameters than those captured by the models: Winds, landscape, political borders, type of wood and so on.

Assuming that the FFM captures all *relevant* features of real forest fires, one can compare universal properties of the model and real fires. These universal features are asymptotics, so that the FFM must be studied in the limit of $L \to \infty$. However, it remains unclear, whether this asymptotic region can be reached by real forest fire, because there is no obvious way to relate the system size of the model to the system size in reality. This is not only a technical problem, but a matter of principle: While the scaling behaviour of two models in the same universality class is not affected by presumably *irrelevant* features, such as next nearest neighbour interaction, lattice topology and so on, this is certainly not the case for non-universal properties. This is the reason why T_c is different for different (real) ferromagnets. So, while 10^7 trees is fairly large in the FFM, it might not be so in real forests.

So, only assuming that a real forest fire has reached its asymptotic behaviour, one can possibly (see notes in Sec. 2.4.1.1) compare universal features. Moreover, while universal features help to identify interactions which are relevant on the large scale, non-universal features might be much more important. This includes the critical tree density, fire velocity etc.

It has been suggested [for example (Malamud *et al.*, 1998)] that by allowing small fires to burn, one might prevent large fires. This trivial sounding argument is highly misleading: It might well be that many desirable properties of a forest are — contrary to the suggestion — optimised by extinguishing as many fires as possible. For example, the average tree density might be higher, even though large fires destroy sometimes large patches. Also, the number of fatalities might possibly be lower.

Transferring the principle of "sacrificing small patches for the sake of large ones" to epidemics leads to the tasteless result that pruning populations is an efficient way to prevent large epidemics. But that apparently cannot be the aim of disease control.

The real power of the FFM is as a model of non-conservative SOC; its sole purpose is to help to understand the fundamental principles which allow certain systems to develop into a scale-invariant state without explicit tuning of parameters, even if the underlying dynamics is highly dissipative.

4.5 Summary

In this chapter the DROSSEL-Schwabl Forest Fire model has been studied in detail. The chapter is split up in the following sections

- After a short introduction, in Sec. 4.2 the model is defined and the method (which allows a study of the model on very large statistical and spatial scales) discussed in detail. In particular, this discussion comprises a elaborate discussion of the problems encountered by a parallelised implementation of the algorithm.
- The results are presented in Sec. 4.3. Former claims about the scale invariance of the model can be rejected. The key message is: There is no reason to assume that the DS-FFM is scale invariant.
- In a short supplemental section, Sec. 4.4, the relation to real forest fires is critically assessed.

Part II

The Oslo Model

Chapter 5

The Oslo Model and Its Variants

After presenting so many problems with SOC models, it is time to show that there are models which indeed display "proper" SOC behaviour. In fact, there is hardly any other model in SOC, which is so robust and shows all features expected from equilibrium critical phenomena. In the following sections the original Oslo model is defined then some variants are presented which are apparently in the same universality class. An operator approach is introduced which casts the Oslo model into an operator algebra, which might open the door for an exact solution.

5.1 The Model

The Oslo model is motivated by the famous ricepile experiments by FRETTE *et al.* (Frette *et al.*, 1996)[1], as a somewhat more formal way to analyse the dynamics of a relaxing ricepile (Christensen *et al.*, 1996). It is originally defined in terms of integer-valued heights ("height model") of rice columns over a — usually one-dimensional — grid. Generalisations to higher dimension are plagued by ambiguities, so that the remainder of this section focuses on the one-dimensional case. The microscopic dynamics of the model is analysed in great detail, in order to determine the physical meaning of different boundary conditions.

5.1.1 Definition in the height picture

The Oslo model is defined in one dimension ($d = 1$ is therefore the dimension of the interface describe in later chapters) as follows: On each vertex of a one-dimensional lattice a height h_i is defined, $i = 1, \cdots, L$. Defining $h_{L+1} = 0$, the slope z_i of site i is defined as

$$z_i = h_{i+1} - h_i \ . \tag{5.1}$$

[1] Recently a similar experiment in two dimensions was performed (Aegerter *et al.*, 2000).

For each site there is in addition a critical slope $z_i^c \in 1, 2$. The heights evolve as follows (Christensen *et al.*, 1996):

- Driving: $h_1 \to h_1' = h_1 + 1$, thus $z_1' = z_1 + 1$
- Relaxation (or toppling): As long as there is a any i with $z_i > z_i^c$ remove a grain from h_i ($h_i \to h_i' = h_i - 1$) and, provided that $i < L$, increase or "charge" h_{i+1} by 1 grain ($h_{i+1} \to h_{i+1}' = h_{i+1} - 1$). Thus $z_i \to z_i' = z_i - 2$ and $z_{i\pm 1} \to z_{i\pm 1}' = z_{i\pm 1} + 1$ for $i < L$; for $i = L$ the rule is $z_L \to z_L' = z_L - 1$ and $z_{L-1} \to z_{L-1}' = z_{L-1} + 1$. After each relaxation, choose a new z_i^c randomly from $1, 2$ with probabilities p and $q = 1 - p$. Repeat the relaxation until all sites obey $z_i \leq z_i^c$.

The Oslo model is very similar to the BTW model, with two important differences: i) The Oslo model is driven only on site 1, at least in its original definition. ii) The BTW has fixed z_c^i — in fact they could vary in space, but the BTW produces the same statistics provided that the z_i^c remain constant in time. A toppling process is shown in Fig. 5.1.

The central observable in the Oslo model is the avalanche size s, measured as the number of topplings, *i.e.* the number of times the second rule have been applied until the pile is fully relaxed again. Alternatively, one can investigate the number of charges, see sec. 5.1.5. The main claim, which is strongly supported by numerics and analytical approaches, is that the PDF of the avalanche sizes follows simple scaling (see Chapter 2), *i.e.*

$$\mathcal{P}(s) = a s^{-\tau} \mathcal{G}(s/(bL^D)) \tag{5.2}$$

with exponents τ (sometimes called "avalanche exponent") and D (cutoff exponent) and two unknown system dependent, non-universal parameters a and b.

To simplify the following discussion, we introduce the terms "stable" for sites with $z_i \leq z_i^c$, "susceptible" for sites with $z_i = z_i^c$ and "active" for sites with $z_i > z_i^c$. To avoid confusion, the term "grain" refers to height h_i and "slope units" refers to slope z_i.

5.1.2 Abelian nature I

It is important to realise that the order of relaxations within the set of active sites does not matter. This can be shown in a straight-forward manner based on a operator approach (Sec. 5.3.3.1), or handwavingly as follows [see also (Dhar, 1999c)[2]]:

It will be shown in the following that if the sequence of z_i^c assigned after each toppling is fixed at each site, then the outcome of a "kick" (application of a driving

[2]Below, Sec. 5.3, page 225, an operator approach is introduced, which is, however, unrelated to DHAR's operators in (Dhar, 1999c). The latter operate on individual states, while the operators below act on the *distribution* of states.

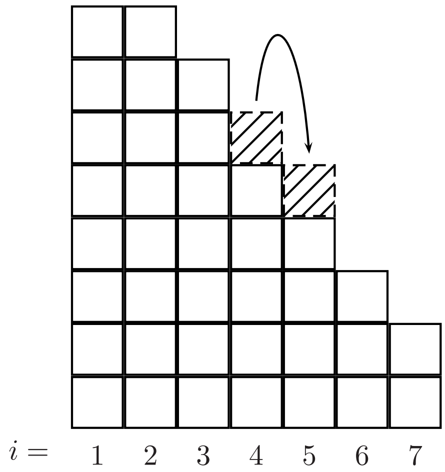


Figure 5.1: A toppling (marked by the arrow) in the Oslo model at site $i = 4$ transports one grain to the right and changes the slopes according to $z_4 \to z_4 - 2$, $z_3 \to z_3 + 1$ and $z_5 \to z_5 + 1$. The toppling grains are shown as dashed and hatched boxes. The process is the same as in the BTW model, see Fig. 1.1, page 42.

step) is also fixed. To see that, one assumes that at a given instance more than one site is active. As long as this is not the case, the order of updates would be fixed by causality.[3] Obviously the order of updates cannot matter, if none of the active sites share a site, whose value will change during the update. So, one assumes that this is the case, *i.e.* there are two sites separated by a common neighbour or right next to each other.

The question which needs to be answered is whether it is possible that the final state changes if the order of topplings is changed, *i.e.* whether it is possible that any site has a different value for h_i (or z_i) or z_i^c after all relaxations have been applied depending on the order of updates. The height h_i depends only on the number of grains a site has received during the updating procedure and the number of times it has toppled. *The order of charges and topplings is irrelevant for the height. Moreover, the number of toppling depends only on the* total *number of grains received, if the sequence of z_i^c is fixed.* In turn, the value of z_i^c depends only on the number of times a site has toppled. Thus, the final state of a site cannot change, if the two neighbouring or consecutive sites topple in a different order.

Thus, swapping the order of updates of two active sites does not change the resulting state. Therefore any permutation of updates within a given set of active sites

[3]It is important to bear in mind that the assertion that a model is Abelian or non-Abelian makes sense only if there are alternative updating schemes allowed by the rules of the model. If the rules fix, for example by causality, the order of *all* updates and the external drive is localised, then one simply cannot say anything about an equivalence of an *alternative* updating order in *this* model, see (Hughes and Paczuski, 2002).

leads to the same final state. Additionally, by stopping the updating procedure and considering the newly charged sites in the current configuration, one arrives at the statement, that any permutation of updates within any given set of charged sites, leads to the same result.

The irrelevance of the order of updates goes even further: As long as the final value of z_i^c is known and provided that it is known how often each site is supposed to topple, the final state is fixed too. Of course, this fact is not of such a big practical value, since the topplings are genuinely *caused* by other topplings and this process of triggering depends on the current value of z_i^c; there is no way of finding the number of topplings without actually performing it.

The independence of the order allows one to define the model without fixing the order. It is worth noting that it is possible to maintain an updating scheme, where the set of sites which are currently active, are either from the odd or from the even sub-lattice. This is because topplings on the odd sub-lattice charge only sites on the even sub-lattice and vice versa.[4]

Noting that the order of updates is irrelevant resembles a feature of the BTW model (Bak *et al.*, 1987), which allowed DHAR (Dhar, 1990) to develop the paradigm of the Abelian Sandpile. However, the situation here is different from the BTW model: In the BTW model, it is sufficient to know which site has toppled in order to reconstruct the initial state, *i.e.* each time evolution operator has an inverse. This is not the case for the Oslo model, because by knowing which site has been hit last by a grain plus its current state (where the state might include z_i^c), it is not possible to reconstruct the former state: More than one state can cause the current state by increasing the height at one site; for example $z_i = 1$ can be caused by $z_i = 0 \to z_i = 1$ or by $z_i = 2 \to z_i = 1$ with former $z_i^c = 2$. This ambiguity means that there is no way to construct an inverse operator and therefore the Oslo model does not posses an inverse operator.

It is important to stress that "Abelian" in the literature refers to the fact that a model's observables do not change if the order of updates is changed. It does *not*

[4]There is a surprising, yet very instructive misconception of the proof above (Paczuski and Bassler, 2000), which is due to a confusion of the configuration of the system with its entire state including the (auxiliary) state of the history of the z_i^c: Fixing the sequence of z_i^c it seems that one arrives at a deterministic model. In fact, having the entire sequence of z_i^c and the current configuration of the model given as the number of charges on each site, plus the initial condition, it is possible to reconstruct its entire history, simply by redoing all updates starting with the initial state, until one arrives at the current state. Thus, conditional to the complete knowledge of those z_i^c, there is an inversion operation for updates. This seems to indicate (Dhar, 1999c) that all states are equally likely, because there is exactly one preceeding state and one succeeding state, so that each one is visited with the same frequency. This however is wrong: The description above gives a deterministic path through phase space only using a particular set of z_i^c. But some configurations might be visited much more often leading to one succeeding configuration at one time and to another at another time, depending on the current values of z_i^c. Only if one would include the future and the past z_i^c into the configuration, all these configurations would be visited equally, namely each one exactly once or never.

mean that there is a way to construct an inverse: It is *always* possible to construct an inverse, provided that enough information is supplied. A stronger definition of "Abelian" in this context would be: The expectation value of all observables are independent under a change of the specific microscopic dynamics.

If a model is Abelian, it does not make much sense to discuss properties of its microscopic dynamics, because the latter is (within a wide range) arbitrary. So, assigning a dynamical critical exponent to the Oslo model (Hughes and Paczuski, 2002) is similar to assigning a dynamical exponent to the ISING model; it makes sense only if the dynamics is actually specified.

5.1.3 Simplification of the boundary and the slope picture

The model described above in terms of height h_i can be simplified by removing the notion of heights, so that the configuration is entirely given by the values of the slopes, from which one can easily derive the heights by integrating, if needed. The "slope picture" of the model describes the model in terms of slope units toppling from one site to another.

While the behaviour at the boundary was completely naturally in the light of the height description, it seems a bit unfortunate that in the slope description the right hand boundary only evolves like $z_L \to z'_L = z_L - 1$ at toppling, while all other sites do $z_i \to z'_i = z_i - 2$. What does it physically mean to replace the evolution rule on the right hand side by $z_L \to z'_L = z_L - 2$? It behaves as if there would be a site at $L+1$ which receives the grain falling off the site at $i = L$ but never topples back, *i.e.* having a neighbouring site which obeys $h_{L+1} - h_{L+2} \leq 1$ always.

This little change in the dynamics smoothens the definition of the model in the slope picture. It evolves now according to

- Driving: $z_1 \to z'_1 = z_1 + 1$
- Relaxation: For all $z_i > z_i^c$, $i = 1, \cdots, L$, do $z_i \to z'_i = z_i - 2$ and $z_{i\pm1} \to z'_{i\pm1} = z_{i\pm1} + 1$ and choose a new random z_i^c, ignoring updates outside the domain $i = 1, \cdots, L$.

In the following this version of the model will be called "Oslo model with simplified boundary conditions", while the original version will be called "Oslo model with original boundary conditions".

It is worth noting that in the height picture the relaxation is anisotropic (grains are falling only towards larger i) and the dynamics of the model is dissipative only on the right hand boundary: *All* grains must propagate through the entire system, causing a toppling at each site the pass, until they leave the system. Defining the size of an avalanche s as the number of applications of the relaxation rule until the system obeys $z_i \leq z_i^c$ for $i = 1, \cdots, L$ again, it follows

$$\langle s \rangle = L \quad \text{(original boundaries)} \tag{5.3}$$

in the stationary state, where $\langle\rangle$ is an ensemble average over realisations of the random z_i^c. For many variations the exact $\langle s \rangle$ can be calculated quite easily by assuming random walker like behaviour for individual *slope* units (see sec. 7.4).[5]

In the slope picture the anisotropy mentioned above is different. During toppling he slope is distributed isotropically among the neighbours with an open boundary only on the *left* hand side in the original definition and two open boundaries in the simplified version. So, it seems that dissipation and conservation at the boundary change meaning when moving from a height picture to a slope picture. However, this correspondence is imperfect, since a nonconservative right end in the slope formulation does not correspond to a conservative right end in the height formulation. The conservation in the bulk (of slope units or height units) is unaffected by these changes.

In the simplified version the simple equation above (5.3) cannot hold any longer. At first glance it might seem that the two dissipative boundaries in the slope picture must translate back into two conservative boundaries in the height picture. This is not true because of the anisotropy of the dynamics in terms of height units; a right hand boundary which is dissipative in terms of slope is even more so in terms of height, see Fig. 5.2. Again, grains (height units) can leave the system only via the right hand boundary, but since a toppling of the rightmost site leads to a loss of one more slope unit ($z_L \to z_L - 2$ instead of $z_L \to z_L - 1$), each such toppling removes not only one grain, but in addition L other grains, as the rightmost slope, where we would start integrating, is reduced by one.

To illustrate that, one considers an example of a lattice of size $L = 2$. The evolution with original boundaries in height units (superscripts denote the value of z_i^c, subscripts the value of z_i) goes like

[5]From Eq. (5.2), one has

$$\langle s^n \rangle = a(bL^D)^{1+n-\tau} \tag{5.4}$$

as shown in Sec. 2.3, page 78. Thus, for $\langle s \rangle \propto L$, as typically found for the kind of driving introduced in sec. 5.1.1, one has (Paczuski and Boettcher, 1996)

$$D(2 - \tau) = 1 \tag{5.5}$$

or in general

$$D(2 - \tau) = \mu \text{ for } \langle s \rangle \propto L^\mu \; . \tag{5.6}$$

h_1	h_2	total content
3_2^1	1_1^1	4
2_0^1	2_2^1	4
2_1^1	1_1^1	3

The same initial configuration now updated with a simplified left boundary leads to

h_1	h_2	total content
3_2^1	1_1^1	4
2_0^1	2_2^1	4
1_1^1	0_0^1	**1**

The average avalanche size for these boundary conditions can be derived easily in the height picture: Depending on how often the rightmost site topples per grain fed into the system, in the following denoted by $\hat{s}_L$, its dissipation rate is $\hat{s}_L(1+L)$, removing 2 grains which have toppled until $i = L$, 1 grain which has toppled until $i = L-1$, and 1 grain everywhere else, see Fig. 5.2. One of the two grains dissipated from $i = L$ does so by leaving the system properly, all other grains "magically disappear" due to the relaxation rule at the boundary; a grain at $i = L$ just removed by the toppling of the rightmost site has toppled already $L-1$ times and so on up to $i = 1$ where the grain had toppled 0 times, in total $L(L+1)/2$. This is the only mechanism for height units to dissipate, so for $1+L$ units coming in, this must lead on average to a single dissipation event at $i = L$. Thus, the average number a site has toppled before leaving the system is $[L(L+1)/2]/(L+1)$, *i.e.*

$$\langle s \rangle = L/2 \quad \text{(simplified boundaries)} \tag{5.7}$$

5.1.3.1 Directedness

What makes the slope picture so much more appealing than the height picture of the Oslo model is the fact that the height picture contains an anisotropy, which is not only mathematically completely spurious, but also physically unjustified.

Above, it has been discussed that the slope units relax completely isotropic. With or without simplified boundaries, the model is perfectly well-defined in the slope picture, especially the height picture can be derived from it. So how is it possible that the model is isotropic in one "language" and anisotropic in another, completely equivalent "language"? The reason is the definition of slope and height, see Eq. (5.1). The anisotropy enters simply when the slope is calculated only towards the right

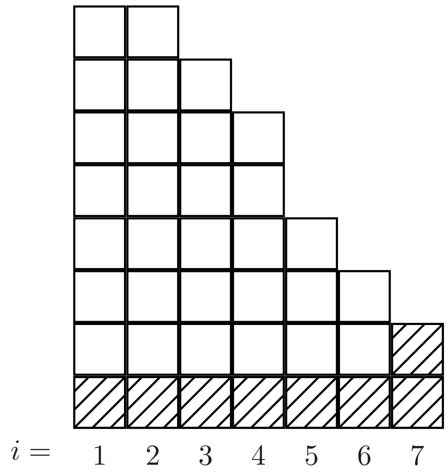


Figure 5.2: If the rightmost site topples once, $L + 1$ height grains dissappear from the system (hatched).

neighbour. Thus, it appears natural to integrate the slope from right to left in order to restore the height picture,

$$h_i = \sum_{j=L}^{i} z_j \tag{5.8}$$

where $h_{L+1} = 0$ has been used.

That this anisotropy is completely unphysical can be seen by the fact that grains can *only* topple to the right. There is no mechanism for toppling to the left. If the model was driven in the centre at $i = L/2$ from scratch (*i.e.* initially empty substrate), there would not develop a heap, but an asymmetric pile as shown in Fig. 5.3.

Conclusively, the Oslo model is a critical height model, *i.e.* a model with a certain height threshold on the sites, if formulated in terms of slope. It is a critical slope model (*i.e.* there is a threshold slope), if formulated in terms of heights. The threshold z_i^c in slope z_i of height h_i is identical to the threshold in amount or "height" of slope.[6]

5.1.4 Toppling frequency

The toppling frequency, *i.e.* the number of times each site receives a height grain per grain introduced into the system is exactly 1 in the original model, since the *only* way for a grain to leave the system is the right boundary and it takes exactly L topplings for a grain to leave the system starting from the leftmost site.

For the system with simplified boundaries, it is a bit more difficult to determine

[6]In one dimension slope models directly translate to height models and vice versa. In higher dimensions this mapping breaks down because of the ambiguity of a scalar slope to be mapped onto a vector-valued gradient.

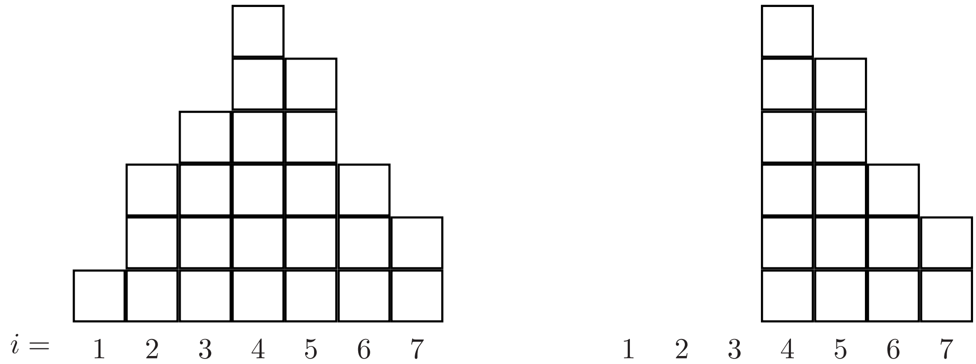


Figure 5.3: Left: A rice pile as one would expect it if it was driven at site $i = 4$, *i.e.* new grains enter only at this site. Right: What really happens in the Oslo model in terms of height is unphysical. For similar considerations in terms of slope, see sec. 5.2.2.

the toppling frequency. The argument above suggests that the expected toppling frequency of the rightmost site is given by

$$\langle \hat{s}_L \rangle = \frac{1}{L+1} \tag{5.9}$$

in the stationary state. What about the other toppling frequencies?

First one notes that the toppling frequency at a site is the same as the charging frequency, *i.e.* the number of times a site is charged (in height units) per initial "kick" of the system. However, the following derivation is better done in terms of slope units, where a single toppling removes two units of slope. Denoting the charging frequency in terms of slope of site i as s_i one has $\langle s_i \rangle = 2\langle \hat{s}_i \rangle$ in the stationary state. Thus, one arrives at a set of equations by imposing that the hits a site received are either caused by external drive or neighbours:

$$\begin{aligned}
\langle s_L \rangle &= \langle \hat{s}_{L-1} \rangle = \langle s_{L-1}/2 \rangle \\
\langle s_{L-1} \rangle &= \langle s_L/2 \rangle + \langle s_{L-2}/2 \rangle \\
\langle s_{L-2} \rangle &= \langle s_{L-1}/2 \rangle + \langle s_{L-3}/2 \rangle \\
\vdots \quad &\quad \vdots \\
\langle s_1 \rangle &= \langle s_2/2 \rangle + \dot{E}
\end{aligned} \tag{5.10}$$

The last line contains the driving frequency $\dot{E}$ explicitly, which was set to $\dot{E} = 1$ previously.

Eq. (5.10) can be solved easily by noting that each line (apart from boundaries) represents just a discretised LAPLACE equation, $\partial_i^2 s = 0$. Thus, the solution is linear in i, *i.e.* $s_i = ai + b$ with a and b to be determined. In order to fulfil the boundary condition which read $s_0 = 2E$ and $s_{L+1} = 0$, one sets $b = 2E$ and $a = -b/(L+1)$, so that

$$\langle s_i \rangle = 2\dot{E}\frac{L+1-i}{L+1} = 2\langle \hat{s}_i \rangle \tag{5.11}$$

or

$$\langle s_i \rangle = \frac{L+1-i}{L}\langle s_1 \rangle \tag{5.12}$$

using $2E = \frac{L+1}{L}\langle s_1 \rangle$. The case $i = L$ is fully consistent with (5.9), $\langle s_L \rangle = 2\dot{E}/(L+1)$ and therefore $\langle \hat{s}_L \rangle = \dot{E}/(L+1)$.

For the integrated charging frequency one finds

$$\sum_{i=1}^{L} \langle s_i \rangle = \dot{E}L = \frac{1}{2}(L+1)\langle s_1 \rangle \quad . \tag{5.13}$$

All charges, apart from the external drive, are caused by topplings, *i.e.* the integrated toppling frequency is $(\dot{E}L - \dot{E} + \langle s_L \rangle/2 + \langle s_1 \rangle/2)/2$, where extra contributions for the topplings at $i = 1$ and $i = L$ have been added, which lead only to charges of the fixed boundary which is not received by any site in the system. Therefore the expected total toppling frequency is

$$\frac{E}{2}\frac{(L-1)(L+1)+1+L}{L+1} = \dot{E}\frac{L}{2} \tag{5.14}$$

which corresponds of course to the average avalanche size (5.7).

An extremely interesting constant is the charging frequency of the first site for non-zero avalanches, *i.e.* the number of times the first site is hit between two avalanches of finite size. It would contain some non-trivial information about the state of the first site, as it involves the probability of the first site to be susceptible, about, however, nothing can be said, because the formalism above has no notion of a collective event such as an avalanche.

It is important to note that all the averages taken above where based on the external driving rate, *i.e.* each "kick" was an external perturbation. It was not imposed that the system actually relaxes at all, therefore avalanches of size 0 have been included.

5.1.4.1 Remark: density of susceptible sites

It might be interesting to measure then fraction of susceptible sites with $z_i = z_i^c$ on the macroscopic timescale, *i.e.* from avalanche to avalanche. However, this quantity, say $\rho_c(t)$, behaves trivial: the total number of height grains needed to fill a system of size

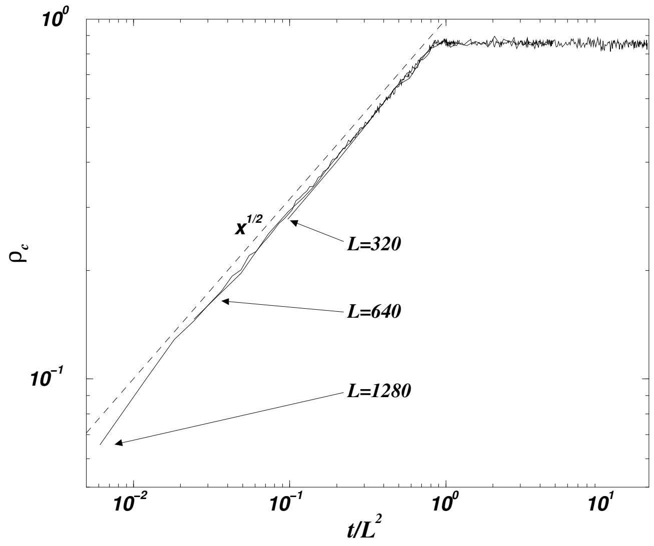


Figure 5.4: A data collapse according to (5.15) for three different system sizes as indicated of the temporal behaviour of the density of susceptible sites ($z_i = z_i^c$) $\rho_c(t)$ in the Oslo model. The time is measured on the macroscopic scale, *i.e.* it is the number of initial kicks. The dashed line shows $\sqrt{x}$.

L is proportional to L^2, see Fig. 5.1. The time it takes to fill in n grains in an empty system is exactly n. As long as the right boundary is not reached by the heap, the density of susceptible sites is just proportional to the distance covered so far, which is, in turn, proportional to $\sqrt{n}$, so that the density is simply $\sqrt{n}/L$. Thus, one expects

$$\rho_c(t) = \frac{\sqrt{t}}{L} f\left(\frac{t}{L^2}\right) , \tag{5.15}$$

where $f(x)$ is constant for $x \leq x_0$ and proportional to $1/\sqrt{x}$ for $x > x_0$ in order to keep $\rho_c(t)$ on a constant level in the stationary state. The constant x_0 is apparently related to the average slope in the stationary state $\overline{z}$ by $\overline{z}/2 = x_0$. An example of this behaviour is shown in Fig. 5.4.

5.1.5 Definition of avalanche and boundary conditions

Before the numerical differences between the Oslo model with original or with simplified boundary conditions are studied, one should first agree on the observable. The literature is not consistent on how to define an avalanche: Is it the total number of topplings in the system between to initial kicks or is it the total number of charges? And are these quantities measured for slope units or for height units? In case of the latter, in the original model with original boundary conditions the only difference between toppling and charging occurs at the right boundary, because if the rightmost site topples, it does not charge any other site. In the following, all events will be

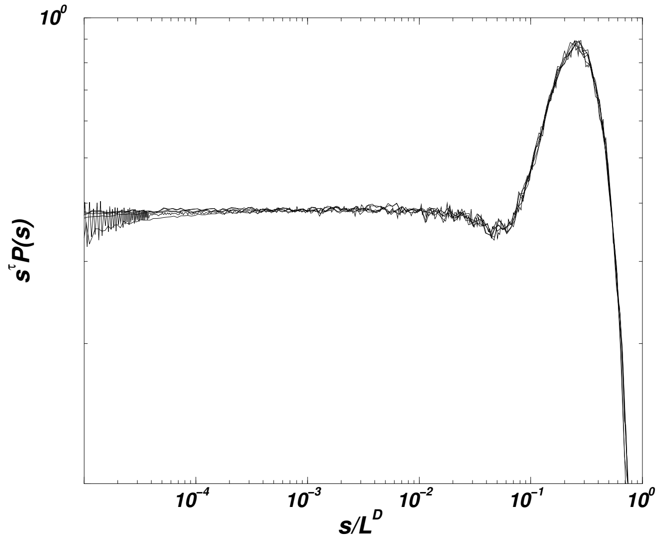


Figure 5.5: Data collapses of binned and rescaled data, according to (5.2), for the Oslo model with original boundary conditions. System sizes are $L = 640, 1280, 2560$, exponents are $\tau = 1.555\ldots$ and $D = 2.25$. The data for the avalanches meassured in number of charges have been rescaled to collapse with the data for the avalanches meassured in number of toppling. The former has been plotted as $(1/2)^{\tau-1} s^\tau \mathcal{P}(s)$ versus $(1/2)s/L^D$.

measured in slope units, where each toppling distributes two slope charges.

In case of the original Oslo model, the avalanche size measured in terms of slope charges and the avalanche size measured in terms of slope topplings are directly related by a factor 2, apart from boundary effects, *i.e.* charges "falling off" the system. Therefore, there is no reason to expect any significant difference between these two observables[7]. This is substantiated by the example shown in Fig. 5.5. *Henceforth, avalanches sizes will be measured as number of topplings if not stated otherwise.*

Remarkably, and in fact in contrast to most other SOC models, the Oslo model is very insensitive to changes in the updating rules (see sec. 5.2). *Prima facie*, it seems indeed that numerically there is no difference between the two different *boundary conditions* introduced above. Only for longer runs and especially when investigating the amplitude ratios, it becomes clear that only the exponents are apparently the same, while the amplitudes ratios are not. Fig. 5.6 compares the avalanche size distribution for three different linear sizes of the model, $L = 640, 1280, 2560$. The resulting exponents and amplitude ratios (see Sec. 2.3.3.1) are shown in Tab. 5.1 (Sec. 5.2.3 contains more details on the parameters of the simulations).

[7]Maybe the only surprise when investigating the avalanche size distribution when measured in terms of number of charges is the entry at size $s = 3$ — it is not possible to obtain this number of charges within a single avalanche: The first site is hit, so $s = 1$. In order to make $s > 1$, this site must relax, so $s = 2$. Again, this site must relax, because there is no other site possibly active. If now $L > 2$, two charges are received by the first and third site, so $s \geq 4$.

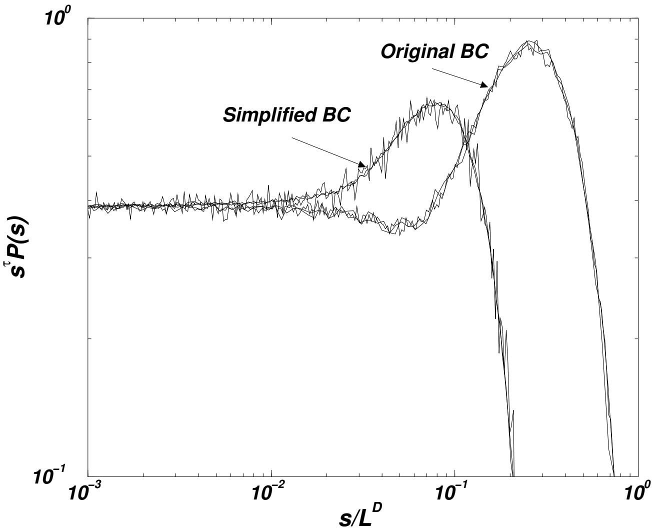


Figure 5.6: Data collapse of binned and rescaled data for the Oslo model with original boundary conditions and with simplified boundaries, focussing on the "bump" of the scaling function. System sizes are $L = 640, 1280, 2560$, exponents are $\tau = 1.555\ldots$ and $D = 2.25$. There are indeed differences between the two boundary conditions: the most significant is an additional dip in front of the bump in case of the original BC.[8]

5.2 Variants

Even though the effects of the boundary conditions detailed above might look surprising in the light of universality, there is strong support for universality in the Oslo model. In fact, all of the following models seem to have the same scaling function and exponents, see Tab. 5.1 (also for more details). They have all been simulated with simplified boundary conditions and with avalanches defined as the number of topplings. From now on "original Oslo model" will refer to the definition in sec. 5.1.1; in addition the boundary conditions will be specified either as "simplified boundaries" (sec. 5.1.3) or as "original boundaries". In later chapters the term "original Oslo model" (OOM) will in fact refer to the original definition with simplified boundaries.

Some variants of the Oslo have already been discussed in the literature. SHU-DONG ZHANG (Zhang, 1997) found that a variation of the Oslo model where the critical slope z_i^c is randomly (and uniformly) chosen from $\{1, 2, 3, 4\}$ has the same critical behaviour than the original model. Bengrine *et al.* (Bengrine *et al.*, 1999a,b) studied the change in behaviour if the probability p is changes, which fixes the rate with which $z_c^i = 1$ is chosen. Again, no changes in the exponents are found.

[8]Thanks to MATTHEW STAPLETON for pointing that out to me.

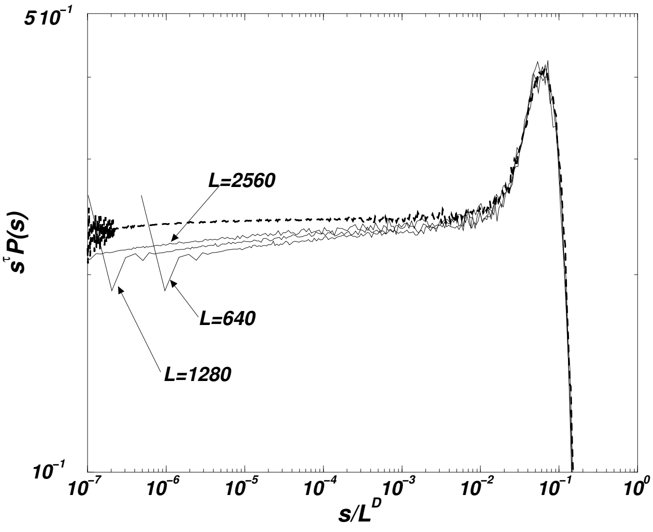


Figure 5.7: Data collapse of binned and rescaled data for the Oslo model with simplified boundaries and next nearest neighbour interaction. System sizes are $L = 640, 1280, 2560$, exponents used (for the collapse) are $\tau = 1.555\ldots$ and $D = 2.25$. The thick, dashed line shows the original Oslo model (simplified boundary conditions), plotted in the rescaled form $(0.8)s^\tau\mathcal{P}(s)$ versus $(0.625)s/L^D$. While the next nearest neighbour version collapses nicely, the collapse with the original model is good, but not perfect.

5.2.1 Changes of the microdynamics

In the following some variants of the Oslo model (slope picture, avalanches are measured as number of topplings, boundaries simplified) are discussed. Firstly, some models with new updating rules are presented, in the next section models with new driving. In the field-theoretical description the changed driving is a change of the boundary condition.

5.2.1.1 Next nearest neighbour interaction

In classical critical phenomena, such as the ISING model, it is well known that the introduction of next nearest neighbour interaction, or any kind of short-ranged interaction does not change the universality class (Fisher *et al.*, 1972) (but see (Barber, 1978), for example). A similar behaviour is expected for the sandpile models defined above.

The idea is that if a site i topples, it randomly distributes the two slope units among, say, four surrounding sites $i-2, i-1, i+1$ and $i+2$ with equal probability. If slope units hit a target site outside the range $\{1, \ldots, L\}$ they disappear from the system. Fig. 5.7 shows a data collapse with the original Oslo model (again, both models use simplified boundary conditions). While such a data collapse gives only a qualitative impression, Tab. 5.1 shows the resulting amplitude ratios and estimates for the exponent D. In fact, based on this data alone, it is very reasonable to assume that these two models belong to the same universality class.[9].

Generalisations of this model are obvious: One could widen the range of reachable neighbours, study the transition to long range interaction, weighten the probability to reach a site by the distance etc. Especially one could make the process anisotropic, by toppling in a favorite direction. Such a model is investigated in detail in Sec. 7.1.

5.2.1.2 Continuous Oslo model

Another, probably more important variant of the Oslo model is a continuous version. What makes it so important is the fact that a field theory will be formulated in terms of continuous variables, *i.e.* apart from being continuous in space and time, continuous in the state.

The most natural way to change the OOM into a model continuous in the state variable is to pick the slope threshold randomly from a uniform distribution, $z_c \in [0, 2]$, as well as the amount distributed among the nearest neighbours, $\Delta z_r \in [0, t_r]$ and $\Delta z_l \in [0, t_l]$. Again, to maintain conservation, the toppling site is reduced by the sum of these two amounts, *i.e.* $z_i \to z_i - (\Delta z_r + \Delta z_l)$. Anisotropy could be introduced by a non-vanishing difference $v = t_r - t_l$, or by changing the probability disrtibutions of Δz_r and Δz_l.

Fig. 5.8 shows the scaling function for this model. Again the data collapse looks very convincing, as well as the numerical results listed in Tab. 5.1.

Of course, the particular choice of threshold range, interaction range and concrete mechanism is pretty arbitrary. But exactly this fact strengthens the claim of strong universality in the Oslo model: It is extremely robust against changes of the microscopic dynamics.

5.2.2 Changes in the drive

This section focusses on changes in the way of driving the model, rather than changing the microscopic dynamics. Originally the model is "point-driven" at the first site, *i.e.* $h_1 \to h_1 + 1$ and therefore $z_1 \to z_1 + 1$. If the driving mechanism (defined in terms of height) is moved away from $i_0 = 1$, *i.e.* $h_{i_0} \to h_{i_0} + 1$ it has been shown in Fig. 5.3 that the relaxation becomes unphysical. However, switching to the slope picture and driving with slope units resolves the problem of the spurious anisotropy. Instead, it introduces another unphysical process, as increasing the slope somewhere in the bulk corresponds to introducing a layer of (height) grains left from the driven

[9] In this context it is interesting to remember that next nearest and higher order neighbour interactions can in princple be used to tune leading corrections to scaling to 0 (Hasenbusch, 2001).

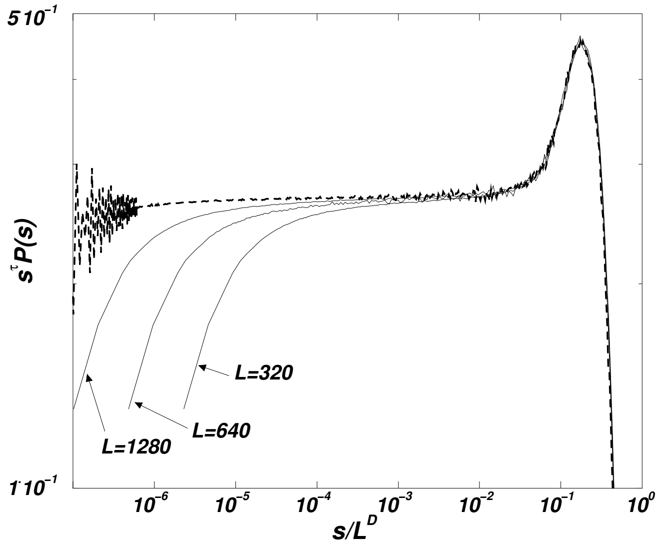


Figure 5.8: Data collapse of binned and rescaled data for the continuous variant of the Oslo model with simplified boundaries. System sizes are $L = 320, 640, 1280$, exponents used are $\tau = 1.555\ldots$ and $D = 2.25$. The thick, dashed line shows the original Oslo model (simplified boundary conditions), plotted in the rescaled form $(0.69)s^\tau\mathcal{P}(s)$ versus $(2.25)s/L^D$. The continuous data collapse nicely with the original model.

site, see Fig. 5.9, similar to Fig. 5.2. One must therefore abandom any physical correspondence when considering slope driving in the bulk.

5.2.2.1 Random drive

The first example is the Oslo model with simplified BC and random drive, *i.e.* an avalanche is initiated by chosing a site i_0 randomly from $i_0 \in \{1, \ldots, L\}$ and driving according to $z_{i_0} \to z_{i_0} + 1$. The resulting average avalanche size is easily estimated:

$$\begin{aligned} \langle s \rangle &= \frac{1}{L}\sum_{i_0=1}^{L}\sum_{k=1}^{i_0}(L-k+1) \\ &= \frac{1}{3}L^2 + \frac{1}{2}L + \frac{1}{6} \end{aligned} \tag{5.16}$$

where the first line sums firstly over all initial positions i_0 (each appearing with probability $1/L$) and then over all "spurious grains" introduced (see Fig. 5.9) at position $k = 1 \ldots i_0$. Such a spurious grain needs $L - k + 1$ topplings to leave the system via the right hand open boundary. The leading term is of course identical to

$$\frac{1}{L}\int_0^L dx_0 \int_0^{x_0} dy(L-y) = \frac{1}{3}L^2 \tag{5.17}$$

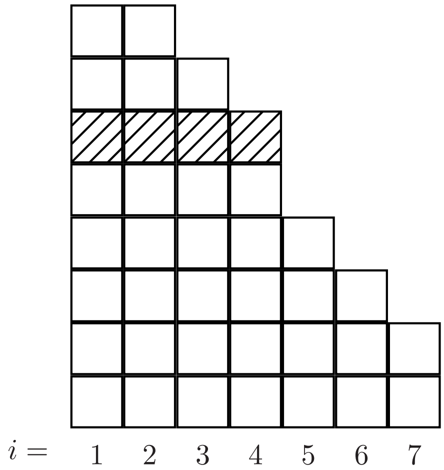


Figure 5.9: A slope grain is added on site $i = 4$, adding effectively an entire layer of of height grains left from it.

Therefore, the scaling law (5.5) changes to

$$D(2 - \tau) = 2 \tag{5.18}$$

A not completely satisfying data collapse for this model is shown in Fig. 5.10. However, it is immediately clear that it does not collapse at all with the OOM.

Most remarkably, the scaling function seems to drop to 0 for small arguments. The resulting exponents are $\tau = 1.1043(8)$ and $D = 2.233(2)$ [see Tab. 5.1, but also (Paczuski *et al.*, 1996; Malthe-Sørenssen, 1999; Dickman *et al.*, 2000)] . The crucial outcome is that the exponent D might well be the same as in the original Oslo model. This is consistent with the field theoretic approach (see Chapter 6), which is directly related only to the exponent D, not to τ. It would be very interesting to compare the result to the bulk-driven MANNA model, which should suffer from similar problems. Open question

Furthermore, ZHANG (Zhang, 1997) has also studied the case of randomly added *height* grains and finds $D = 1.25(10)$ and $\tau = 1.20(6)$, consistent with $D(2 - \tau) = 1$ as expected from $\langle s \rangle = L/2$ in this model. One should stress however, that a random addition of height grains is not easily dealt with by a field theory. A comprehensive field theoretic study of random drive is still to be done. Open problem

The current interpretation is that D is the "characteristic exponent" of the Oslo model, while τ can only be derived with respect to model-specific features. Those are imposed, for example, by conservation, which allows the derivation of $\langle s \rangle \propto L^{\mu}$, see (5.5). Moreover, the scaling function depends additionally on boundary conditions and possibly other properties.

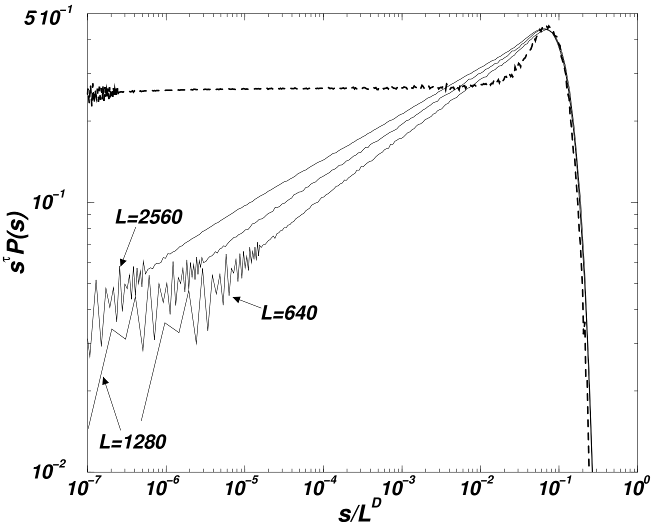


Figure 5.10: Attempt of an data collapse of the binned and rescaled data for the bulk driven variant of the Oslo model with simplified boundaries. System sizes are $L = 640, 1280, 2560$, exponents used are $\tau = 1.111\ldots$ and $D = 2.25$. The thick, dashed line shows the original Oslo model (simplified boundary conditions), plotted in the rescaled form $(0.68)s^{1.555\ldots}\mathcal{P}(s)$ versus $(0.9)s/L^D$. The collapse of the bulk driven data is not satisfying; they certainly do not collapse with the scaling function of the original model.

5.2.2.2 The tilted sandpile

There is another model known in the literature (Malthe-Sørenssen, 1999), which is designed along the lines of the Oslo model, but has a different driving mechanism and slightly different dynamics. The slopes z_i as well as the critical slopes z_i^c are continuous variables. The former are updated as in the original model, the latter are chosen randomly from a uniform distribution. During the driving phase all z_i are increased gradually by the same amount until a toppling takes place. Just like in the OFC model (see Sec. 1.3.2 and Sec. 1.3.2), the driving stops while an avalanche is running, so that the avalanches are instantaneous compared to all other processes. It is worth noting that a homogeneous drive can in fact be replaced by a random drive by an amount ϵ, so that the homogeneous drive is recovered in the limit $\epsilon \to 0$; the local fluctuations in the drive are only of order $\sqrt{\epsilon}$ and vanish therefore as well. Moreover, in this particular model, these fluctuations can be "absorbed" into the fluctuations of z_i^c; there is no way to determine whether a delayed or premature occurrence of a toppling is due to a fluctuation of the inflow or a fluctuation in the local threshold.

5.2.2.3 MANNA model

Remarkably, even the Abelian version of the MANNA model (Manna, 1991; Dhar, 1999a,b) in one dimension seems to belong to the Oslo universality class[10]. Well supported by numerics, this applies to the randomly driven model as well as to the point-driven version. In fact, the exponent $\tau = 1.10(2)$ presented above is found in the literature (Paczuski *et al.*, 1996; Dickman *et al.*, 2000) for the randomly driven MANNA model.

In the MANNA model, all z_i^c are set to $z_i^c = 1$. This would lead to the BTW model, if there was not an additional randomness in the update: When toppling, each slope unit moves independently and with equal probability to the left or to the right. It is important that contrary to the original MANNA model (Manna, 1991), slope units topple only pairwise; originally, if a site relaxes, it would redistribute *all* particles, which makes it very important whether a particle arrives before or after a relaxation, and therefore renders the model non-Abelian.

Other Models similar to the Abelian MANNA model will be discussed in Sec. 7.1.

If the MANNA model is in the same universality class as the Oslo model, this is very interesting for two main reasons:

i) The MANNA model is sometimes seen as the standard representative of an entire universality class (Alava, 2003; Vespignani *et al.*, 2000). The models in this universality class are supposedly described by a REGGEON Field Theory, which is the usual field theoretic approach to directed percolation (Sec. 1.4.1) and other absorbing state phenomena. The Oslo model, on the other hand, is described by a quenched EDWARDS- WILKINSON equation. Thus, the MANNA model might provide a path to join these two field theoretic approaches.

ii) The MANNA model does not contain randomly chosen critical slope. The only randomness lies in the random choice of the direction the particles are spilled. This makes it numerically as well as analytically very interesting, see Sec. 7.3.3.1, page 294.

Especially the continuous Oslo model discussed above (sec. 5.2.1.2) should also be translatable into a MANNA model, *i.e.* it should be possible to fix the random threshold z_i^c to a specific value $z_c^i = z^c$ without changing the critical behaviour, at least not the exponents. If this is numerically confirmed, it is certainly one of the most attractive directions to pursue the field theoretical approach discussed in chapter 6.

Model	g_3	g_4	g_5	D	γ_1
original BC [a]	$1.685(3)$	$3.55(2)$	$8.74(9)$	$2.2509(6)$	1
simplified BC [b]	$1.813(2)$	$4.142(12)$	$11.08(7)$	$2.2498(3)$	1
next nearest NB [c]	$1.812(4)$	$4.16(3)$	$11.26(17)$	$2.21(2)$	1
continuous [d]	$1.8129(11)$	$4.148(7)$	$11.11(4)$	$2.2476(2)$	1
random drive [e]	$1.49(2)$	$2.71(8)$	$5.7(3)$	$2.233(2)$	2

[a]original Oslo model with original boundaries, Sec. 5.1.1 on page 205.
[b]original Oslo model with simplified boundaries, Sec. 5.1.3 on page 209.
[c]Oslo model with next nearest neighbour interactions, Sec. 5.2.1.1 on page 218.
[d]continuous version of the Oslo model, Sec. 5.2.1.2 on page 219.
[e]randomly, bulk driven Oslo model, Sec. 5.2.2.1 on page 220.

Table 5.1: All numerical results are based on system sizes in a range of typically $40, 80, \ldots, 2560$ and 10^9 avalanches. The numerical errors are highly unreliable, see text. The three models expected to be in the same universality class are printed in bold. For a discussion of the amplitude ratios see Sec. 2.3.3.1.

5.2.3 Overview of numerical results

The correlation time is not completely straightforward to define, because the sequence of avalanche sizes is not a MARKOV process, so that actually there is no reason to assume[11] that the autocorrelation function behaves like

$$\frac{\langle s_t s_{t+t'}\rangle - \langle s\rangle^2}{\langle s^2\rangle - \langle s\rangle^2} = \exp(-t'/T)(1 + \text{corrections}) \ . \tag{5.19}$$

Indeed, as one might expect, the avalanche sizes are *anti*correlated, captured by a correlation function of the form

$$\begin{cases} 1 & \text{for } t = 0 \\ c_0 \exp(-t/T_0) + c_1 \exp(-t/T_1) & \text{otherwise} \end{cases} \tag{5.20}$$

It turns out that both constants are very small, for example, for $L = 640$ one finds $c_0 = -0.00105$, so that one might regard the process as uncorrelated. On the other hand, the correlation times found by this procedure behave consistently and systematically: for the OOM with simplified boundary conditions T_0 is almost perfectly linear (prefactor 0.65) in L and for the AOM in the totally asymmetric limit (see Sec. 7.1) T_0 is proportional to $\sqrt{L}$. For example, at $L = 640$ the correlation time in the original Oslo model was found to be $T_0 = 395$ and $T_1 = 38$.

Because of these complications, it was decided not to take the correlation into account at all. The errors of the amplitude ratios have therefore been calculated based

[10]In fact, MATTHEW STAPLETON recovered the MANNA model while trying to simplify the Oslo model.

[11]For observables which can be written as a matrix acting on the state distribution of the MARKOV process, this can be shown to hold in general.

on a Jackknife scheme (Sec. 2.4.5, page 102), but with correlation time T_0 set to 0. The resulting set of data has been fitted against

$$g_n + g_n'/L^{\gamma'} \ . \tag{5.21}$$

The omission of the correlation time leads to a reduced goodness-of-fit (Press *et al.*, 1992), yet the resulting values are perfectly reasonable.

Similarly, the exponents have been determined by a moment analysis (Sec. 2.4.4), by fitting the moments against

$$\langle s^n \rangle = a_0 L^{\gamma_n} + a_1 L^{\gamma_n - 1} \ . \tag{5.22}$$

Since γ_1 is usually known[12], the resulting set of exponents was fitted against

$$\gamma_1 + D(n-1) \ .$$

Again, the goodness-of-fit is significantly reduced and one finds results with almost ridiculously small error bars: $D_{\text{OOM}} = 2.24984(31)$ and therefore $\tau_{\text{OOM}} = 1.555525(61)$. Thus, all error bars given in Tab. 5.1 are very unreliable and should not be taken too serious. However, an alternative numerical approach is yet to be found. Open question

For some of the exponents and amplitude ratios, the numerical results are much poorer than one could have hoped. For example, the cutoff exponent for the next nearest neighbour model seem to match the exponent of the OOM only qualitatively, see Tab. 5.1. In case of the nearest neighbour version, only system sizes from $L = 640$ on could be included in the fits for D. The central question is how to determine the error properly — what is a "good fitting range", a proper set of confluent singularities to include, a proper error of the raw data etc.

5.3 Operator Approach

One of the most promising approaches to *solve* the Oslo model in its original form directly on the lattice, *i.e.* without continuum approximation, is using an operator approach. The idea is to write the Oslo model as a MARKOV process, with a MARKOV matrix representing an "initial kick" and the subsequent relaxation of the entire lattice. In the following, the formalism [see for example (Dhar, 1999c)] is shortly outlined and used to convince ourselves about the Abelian property.

[12] For $D(2-\tau) = \gamma_1$ one has $\mathcal{P}(s) = aL^{\gamma_1} s^{-2} \mathcal{G}'(s/(bL^D))$ from (5.2).

5.3.1 Notation and remarks

For a system of size L, the Oslo model has 3^L stable states, corresponding to $z_i \in \{0, 1, 2\}$. The probability distribution of these states can be represented in form of a vector with 3^L entries, each one representing the probability that the system is found in a particular state. A single site system would be represented by a vector

$$\begin{pmatrix} p_0 \\ p_1 \\ p_2 \end{pmatrix} \tag{5.23}$$

with $p_0 + p_1 + p_2 = 1$. In order to use a tensor product formalism (see chapter 8 for more details), the distribution vector $|P_t\rangle_L$ for a system[13] of size L is arranged such that the probability for the system to be in state $(z_1, z_2, \cdots, z_L)$ is found in row

$$r = \sum_{i=1}^{L} 3^{L-i} z_i \ . \tag{5.24}$$

It is very important to keep in mind that the vector $|P_t\rangle$ contains all possible states of the system to avoid its confusion with a vector like $(z_1, z_2, \cdots, z_L)$.

The MARKOV matrix corresponding to adding a single slope unit to site i and relaxing the entire system of size L is denoted $\mathbf{O}_{L,i}$. The Abelian property of the Oslo model then reads

$$\left[\mathbf{O}_{L,i}, \mathbf{O}_{L,j}\right]_- = 0 \tag{5.25}$$

This property is investigated in detail in sec. 5.3.3.1.

Using the new notation, a stationary state is then any state $|P\rangle_L$ with

$$\mathbf{O}_{L,i} |P\rangle_L = |P\rangle_L \ . \tag{5.26}$$

Such a state exists, because $\langle 0|_L$ with

$$\langle 0|_L = (\underbrace{1, 1, \ldots, 1}_{3^L \text{ times}}) \tag{5.27}$$

is a left-hand eigenvector of $\mathbf{O}_{L,i}$ with eigenvalue 1, since $\mathbf{O}_{L,i}$ is Markovian. This means that the normalisation $\langle 0|P\rangle_L$ is invariant under the operation of the Markovian operator. It follows that every MARKOV matrix has *at least* one right hand eigenvector with eigenvalue 1. For details on decomposable, reducible and splitting MARKOV matrices see (van Kampen, 1992).

[13] The fancy bra-ket notation is very useful and therefore very popular in the field of stochastic processes in one dimension (Hinrichsen, 2000; Derrida and Evans, 1997).

Of course, in the context of the Oslo model, one is particularly interested in a stationary distribution $|P^*\rangle_L$ of the form

$$\lim_{n\to\infty} \mathbf{O}_{L,1}^n |P_0\rangle_L = |P^*\rangle_L \ , \tag{5.28}$$

i.e. a final distribution $|P^*\rangle_L$ to which every initial state (distribution) $|P_0\rangle_L$ converges, charging the very first site $i = 1$ again and again.[14] However, such a state does not necessarily exist, because $\mathbf{O}_{L,1}^n$ might have eigenvalues λ which are any root of 1. For example, the MARKOV matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{5.29}$$

has eigenvectors

$$\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \quad \lambda = 1 \tag{5.30a}$$

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \lambda = -1 \tag{5.30b}$$

and therefore any pure initial state, *i.e.* any initial $|P_0\rangle_L$ with a single 1, oscillates forever. It is worth noting that the normalisation of every eigenvector $|P\rangle_L$ of a MARKOV matrix, which has not eigenvalue $\lambda = 1$ is necessarily 0, since

$$\langle 0|_L \, |P\rangle_L = \langle 0|_L \, \mathbf{O}_i \, |P\rangle_L = \langle 0|_L \, \lambda \, |P\rangle_L \ . \tag{5.31}$$

Moreover, there is no guarantee that a Markovian matrix has enough eigenvectors to span the entire space. For example, the MARKOV matrix

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \tag{5.32}$$

has only eigenvectors

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \text{with} \qquad \lambda = 1 \tag{5.33a}$$

[14] The repeated charging is denoted by the power n in (5.28).

$$\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad \text{with} \qquad \lambda = 0 \tag{5.33b}$$

which of course do not span the entire $\mathbb{R}^3$.

This spoils the naïve ansatz to prove convergence of the form (5.28), even if there is only one right hand eigenvector with eigenvalue that is a root of 1; in that case on might want to write

$$|P_0\rangle_L = \sum_{j=1}^{3^L} a_j \, |e_j\rangle_L \tag{5.34}$$

where $|e_i\rangle_L$ are the eigenvectors of $\mathbf{O}_i$. From (5.31), operating with $\langle 0|_L$ on the LHS immediately gives $a_{j_0} = 1$ if $|e_{j_0}\rangle_L$ is the eigenvector with eigenvalue 1. If (5.34) holds, one has

$$(\mathbf{O}_i)^n \, |P_0\rangle_L = \sum_{j=1}^{3^L} \lambda_{L,j}^n a_j \, |e_j\rangle_L \tag{5.35}$$

where $\lambda_{L,j}$ are the eigenvalues of eigenvector j. For $n \to \infty$ one would recover (5.28).

The vectors obtained in Eq. (5.33) seem to suggest that $(0,0,1)^\mathsf{T}$ might not converge, *i.e.* that

$$\lim_{n\to\infty} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}^n \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{5.36}$$

does not exist. However, one can easily prove explicitly that it does. The same can be shown for the entire Euclidean basis. Clearly, if an entire set of basis vectors converges, so does every vector of the space spanned by them. Thus, even though a MARKOV matrix might not have enough eigenvectors to span the entire vector space, every vector in this space taken as initial distribution might still converge to a unique stationary state.

These are just some caveats of discrete time MARKOV chain. The oscillations mentioned above dissappear immediately for continuous time chains (van Kampen, 1992). Nevertheless, one should keep these remarks in mind when dealing with lattice models like the Oslo model.

5.3.2 The Oslo model

The first step to analyse the Oslo model in terms of operators, is to construct them explicitly. The fundamental observation is that *an individual site charged n times instantaneously can be updated sequentially.* This follows directly from the definition of the model; it simply does not make any difference, whether z_i is first increased by n and then relaxed according to the rules, or to update it sequentially. Provided one does not mix charges produced at different (microscopic[15]) time steps, this does *not* make use of the Abelian property. It follows solely from the fact that the results of a sequential and an instantaneous update are indistinguishable, if the same sequence of z_i^c is drawn.

The fact that updates can be locally serialised is the crucial property that allows to write an operator putting n charges on a site as the nth power of the operator which puts a single charge on this site. For the time being, this does *not* mean that the resulting topplings can be performed in any order, for example, if n charges arrive at a site to perform all updates resulting from the first charge first, then do all updates from the second and so on. It only means that the *local* update does not require a special operator for every number of charges arriving.

As will be used in Chapter 6 (see Fig. 6.1), if $z_i = 0$ is charged once, it must change to $z_i = 1$. Depending on the value of z_i^c it can go over into $z_i = 2$ if charged once at $z_i = 1$ or go down to $z_i = 0$ by toppling. The latter happens with the same probability as $z_i^c = 1$ has been drawn at the last toppling, *i.e.* with probability p. Finally, if charged at $z_i = 2$ it necessarily topples and goes down to $z_i = 1$. Thus, for a system of size $L = 1$ there are two matrices

$$\mathsf{S} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & q & 0 \end{pmatrix} \qquad \mathsf{T} = \begin{pmatrix} 0 & p & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.37}$$

where the first one, S, is the partition of the complete MARKOV matrix

$$\mathbf{O}_{L=1,1} = \mathsf{S} + \mathsf{T} \tag{5.38}$$

which describes transitions in the single site system without toppling, and T describes transitions with toppling. The correct order of rows and columns can easily be checked by explictly calculating the resulting distribution of states for $\mathbf{O}_{L=1,1}(0,0,1)^{\mathsf{T}}$ etc.

Starting from a single site, one can easily build up the operators for larger systems. Similar to the methods used in Chapter 8, one defines the single site operators

$$\mathsf{S}_{L,i} = \mathbf{1}^{\otimes i-1} \otimes \mathsf{S} \otimes \mathbf{1}^{\otimes L-i-1} \tag{5.39a}$$
$$\mathsf{T}_{L,i} = \mathbf{1}^{\otimes i-1} \otimes \mathsf{T} \otimes \mathbf{1}^{\otimes L-i-1} \tag{5.39b}$$

which charge site i in a system of size L. The identity $\mathbf{1}$ has rank 3, so that the resulting matrices have the correct rank 3^L. One can immediately prove explicitly

[15]Of course, the notion of a microscopic time makes sense only if it is fixed somehow. Thus, for the time being, one can imagine it is given by parallel updates.

the commutators

$$\begin{aligned} \left[\mathsf{S}_{L,i}, \mathsf{S}_{L,j}\right]_- &= 0 && (5.40a)\\ \left[\mathsf{T}_{L,i}, \mathsf{T}_{L,j}\right]_- &= 0 && (5.40b)\\ \left[\mathsf{T}_{L,i}, \mathsf{S}_{L,j}\right]_- &= \delta_{i,j} Q_{L,i} && (5.40c) \end{aligned}$$

with an additional matrix $Q_{L,i}$, which can easily be calculated explicitly, to be $Q_{L,i} = \mathbf{1}^{\otimes i-1} \otimes Q \otimes \mathbf{1}^{\otimes L-i-1}$, where

$$Q = \begin{pmatrix} p & 0 & 0 \\ 0 & q-p & 0 \\ 0 & 0 & -q \end{pmatrix} \tag{5.41}$$

If a site topples, it charges its neighbours, so that

$$\mathbf{O}_{L,i} = \mathsf{S}_{L,i} + \mathbf{O}_{L,i-1} \mathbf{O}_{L,i+1} \mathsf{T}_{L,i} \tag{5.42}$$

where $\mathbf{O}_{L,0} = \mathbf{O}_{L,L+1} = \mathbf{1}^{\otimes L}$ implements the simplified boundary conditions. Recently, DHAR has shown (Dhar, 2003) some very interesting properties of these matrices using original boundary conditions, but in the following these differences are insignificant.

Eq. (5.42) is the central *definition* of the operator. However, there is a fundamental problem associated with this definition, which is often overlooked: The claim is that it leads to a correct representation of the Oslo model. The problem is that Eq. (5.42) prescribes a specific order of updates, namely that the charges resulting from a toppling are to be evaluated sequentially but non-local; if site i topples, the resulting charge on site $i+1$, which might lead to an entire avalanche, is first completed, before the charge on site $i-1$ is even considered. In fact, the operation $\mathbf{O}_{L,i+1}\mathsf{T}_{L,i}$ removes even a particle from the system, because the charge arriving at i leads to a reduction of slope by 1 unit ($\mathsf{T}_{L,i}$) and an additional charge at $i+1$, *i.e.* no net change instead of a net change by $+1$. Thus, contrary to the tacit assumption that (5.42) is a proper starting point for subsequent calculations, especially to show the Abelian property, it actually makes already use of it. For example, by symmetry it would follow immediately that

$$\mathbf{O}_{L,i} = \mathsf{S}_{L,i} + \mathbf{O}_{L,i+1} \mathbf{O}_{L,i-1} \mathsf{T}_{L,i} \tag{5.43}$$

is an equally proper representation of the Oslo model as well, which means that

$$\mathbf{O}_{L,i+1} \mathbf{O}_{L,i-1} \mathsf{T}_{L,i} = \mathbf{O}_{L,i-1} \mathbf{O}_{L,i+1} \mathsf{T}_{L,i} \ . \tag{5.44}$$

This, however, needs to be proven first.

5.3.2.1 Operators as trees

It is clear that the matrix $\mathbf{O}_{L,i}$ must somehow be representable as the sum of products of the "atomic" matrices $\mathsf{S}_{L,j}$ and $\mathsf{T}_{L,j}$. One only needs to follow the original definition in sec. 5.1.1 to construct it. First, $\mathbf{O}_{L,i}$ can either lead to no toppling, $\mathsf{S}_{L,i}$, or to a toppling, charging the two neighbouring sites:

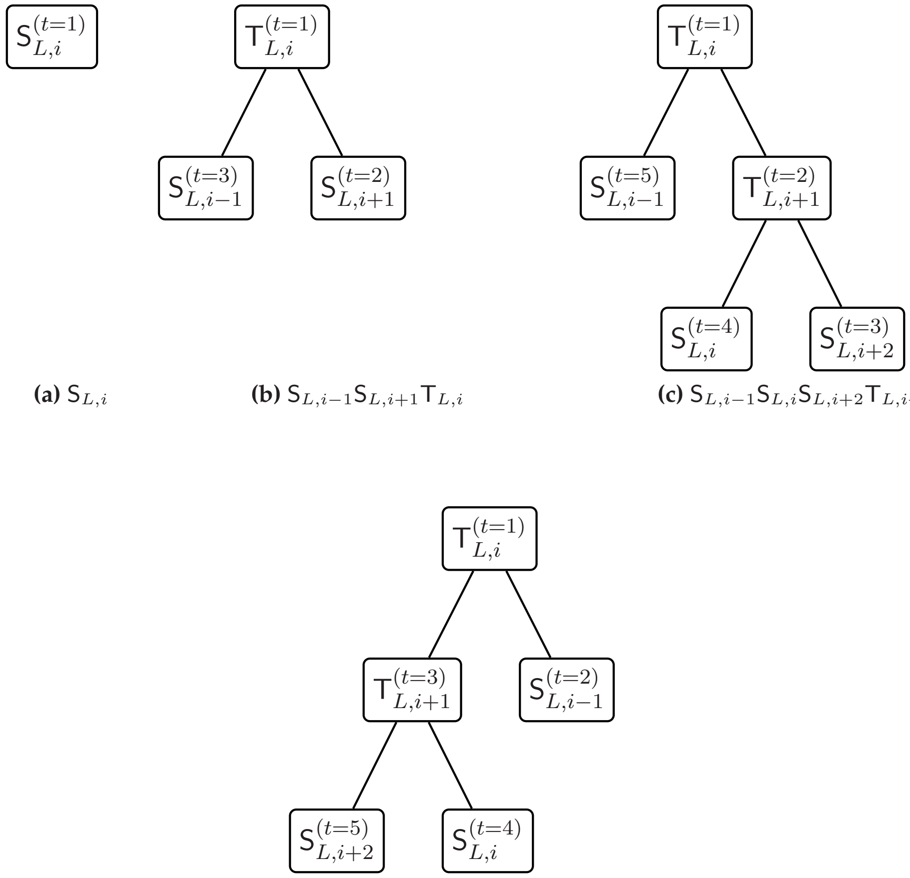
$$\mathbf{O}_{L,i} = \mathsf{S}_{L,i} + (\dots)\mathsf{T}_{L,i} \tag{5.45}$$

where the brackets, $(\dots)$, include all terms of the subsequent charging of the neighbouring site. It is important that $(\dots)$ is applied *after* $\mathsf{T}_{L,i}$ has been applied (to the state distribution), as $(\dots)$ happens only if $\mathsf{T}_{L,i}$ gives a non-zero result. The operation of $\mathsf{T}_{L,i}$ on some pure states leads to $(0, 0, \dots, 0)^\mathsf{T}$, if site i is in state 0, so that a single charge cannot lead to a toppling.

The next step is to perform the "expansion" of $(\dots)$. This requires a decision about the order of updates. The original definition leaves this choice open and one aim of this section is to show, that the ordering scheme does not change the resulting operator. It is important to distinguish that from the proof that $\mathbf{O}_{L,i}$ is Abelian (5.25): For the time being the aim is to show that the representation of $\mathbf{O}_{L,i}$ does not change if the ordering scheme of the updates is changed.

The problem is more than just the question whether to write the operator $\mathbf{O}_{L,i}$ in the form (5.42) or as (5.43). The central question is whether a specific choice for the ordering of the operations changes the resulting operator. If that is true, then a choice like (5.42) fixes the ordering in an important way. The Oslo model in its original form would then be under-defined.

For further investigation, a tree representation of the expansion is introduced, see Fig. 5.11. There are two types of nodes, internal (branching) nodes, which correspond to $\mathsf{T}_{L,i}$ and leaves, which correspond to $\mathsf{S}_{L,i}$. Each node has a label indicating the operator and the time order in the form $(t = 1, 2, \dots)$. This time label indicates the order of updates in the system and therefore the order in the expansion of the operator. Fig. 5.11 shows the first four terms in the expansion of $\mathbf{O}_{L,i}$ in the form



(a) $\mathsf{S}_{L,i}$ **(b)** $\mathsf{S}_{L,i-1}\mathsf{S}_{L,i+1}\mathsf{T}_{L,i}$ **(c)** $\mathsf{S}_{L,i-1}\mathsf{S}_{L,i}\mathsf{S}_{L,i+2}\mathsf{T}_{L,i+1}\mathsf{T}_{L,i}$

(d) $\mathsf{S}_{L,i+2}\mathsf{S}_{L,i}\mathsf{T}_{L,i+1}\mathsf{S}_{L,i-1}\mathsf{T}_{L,i}$

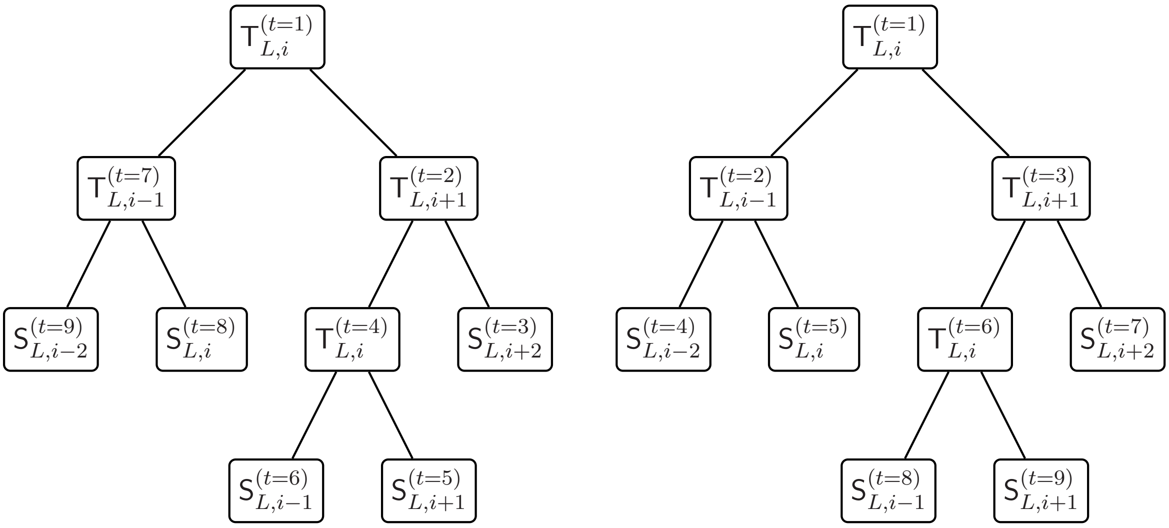
Figure 5.11: The first four terms in the expansion of $\mathbf{O}_{L,i}$ following the prescription (5.42). The equation below each figure shows the term corresponding to the figure.

(5.42) leading to

$$\begin{aligned}
\mathbf{O}_{L,i} &= \mathsf{S}^{(t=1)}_{L,i} \\
&+ \mathsf{S}^{(t=3)}_{L,i-1}\mathsf{S}^{(t=2)}_{L,i+1}\mathsf{T}^{(t=1)}_{L,i} \\
&+ \mathsf{S}^{(t=5)}_{L,i-1}\mathsf{S}^{(t=4)}_{L,i}\mathsf{S}^{(t=3)}_{L,i+2}\mathsf{T}^{(t=2)}_{L,i+1}\mathsf{T}^{(t=1)}_{L,i} \\
&+ \mathsf{S}^{(t=5)}_{L,i+2}\mathsf{S}^{(t=4)}_{L,i}\mathsf{T}^{(t=3)}_{L,i+1}\mathsf{S}^{(t=2)}_{L,i-1}\mathsf{T}^{(t=1)}_{L,i} \\
&+ \dots
\end{aligned} \tag{5.46}$$

For simplicity one defines $\mathsf{S}_{L,0} = \mathsf{S}_{L,L+1} = \mathbf{1}^{\otimes L}$. Clearly, the time labels must follow causality: The sequence of time labels obtained while passing from the root to any node is strictly monotonically increasing.

To illustrate the problem a bit further, Fig. 5.12 shows two examples of possible



(a) $\mathsf{S}_{L,i-2}\mathsf{S}_{L,i}\mathsf{T}_{L,i-1}\mathsf{S}_{L,i-1}\mathsf{S}_{L,i+1}\mathsf{T}_{L,i}\mathsf{S}_{L,i+2}\mathsf{T}_{L,i+1}\mathsf{T}_{L,i}$ **(b)** $\mathsf{S}_{L,i+1}\mathsf{S}_{L,i-1}\mathsf{S}_{L,i+2}\mathsf{T}_{L,i}\mathsf{S}_{L,i}\mathsf{S}_{L,i-2}\mathsf{T}_{L,i+1}\mathsf{T}_{L,i-1}\mathsf{T}_{L,i}$

Figure 5.12: Two ways of labelling the same tree, leading to different terms. Fig. 5.12(a) shows the labelling corresponding to (5.42), which is in fact a depth-first-search (Cormen *et al.*, 1990). Fig. 5.12(b) shows a parallel update.

ordering schemes for a very simple sequence of topplings, leading to the same operations performed at each site, but in different orders.

5.3.3 Rearranging trees

The crucial observation is that (5.40) indicates that a resulting term from every time-ordered tree can be brought into the form

$$\Pi_{L,0}\Pi_{L,1}\dots\Pi_{L,L}\Pi_{L,L+1} \tag{5.47}$$

where each $\Pi_{L,i}$ is a product of matrices $\mathsf{S}_{L,i}$ and $\mathsf{T}_{L,i}$. For example Fig. 5.12(a) gives

$$\underbrace{\mathsf{S}_{L,i-2}}_{\Pi_{L,i-2}}\underbrace{\mathsf{T}_{L,i-1}\mathsf{S}_{L,i-1}}_{\Pi_{L,i-1}}\underbrace{\mathsf{S}_{L,i}\mathsf{T}_{L,i}\mathsf{T}_{L,i}}_{\Pi_{L,i}}\underbrace{\mathsf{S}_{L,i+1}\mathsf{T}_{L,i+1}}_{\Pi_{L,i+1}}\underbrace{\mathsf{S}_{L,i+2}}_{\Pi_{L,i+2}}$$

while Fig. 5.12(b) gives

$$\underbrace{\mathsf{S}_{L,i-2}}_{\Pi_{L,i-2}}\underbrace{\mathsf{S}_{L,i-1}\mathsf{T}_{L,i-1}}_{\Pi_{L,i-1}}\underbrace{\mathsf{T}_{L,i}\mathsf{S}_{L,i}\mathsf{T}_{L,i}}_{\Pi_{L,i}}\underbrace{\mathsf{S}_{L,i+1}\mathsf{T}_{L,i+1}}_{\Pi_{L,i-1}}\underbrace{\mathsf{S}_{L,i+2}}_{\Pi_{L,i+2}}$$

which is incompatible in the term $\Pi_{L,i}$. The latter is $\mathsf{S}_{L,i}\mathsf{T}_{L,i}\mathsf{T}_{L,i}$ in Fig. 5.12(a) and $\mathsf{T}_{L,i}\mathsf{S}_{L,i}\mathsf{T}_{L,i}$ in Fig. 5.12(b).

To restate the problem again: If a particular order of updates leads to an expansion of the operator which includes, for example, a term like Fig. 5.12(a), and another order leads to an expansion including a term like Fig. 5.12(b), then these different time ordering might lead to a different operator altogether. However, the claim is: There is a one-to-one mapping between each term contained in an expansion of $\mathbf{O}_{L,i}$ obtained with different orderings.

The idea to prove that is to take the time-ordered tree produced by one ordering scheme and rearrange it to obtain the time-ordered tree of another ordering, corresponding to the same operation. The expansion of $\mathbf{O}_{L,i}$ produces all possible binary trees, independent of the ordering; *i.e.*, since *every* tree is generated in every ordering[16], this proves that every ordering contains the same set of terms.

So, now we show that every time-ordering produces the same set of terms as every other time ordering, by taking the tree representing a certain term in one time-ordering and rearrange it such that produces the same term in another time-ordering. By keeping the tree complete (*i.e.* not splitting it) and by showing that all nodes are visited, one then establishes that the new time-ordering can produce the same term by a valid tree[17].

The procedure to rearrange the tree works as follows: Starting from the root, which is certainly the first occurrence of $\mathsf{T}_{L,i}$ (or $\mathsf{S}_{L,i}$ in a trivial case), nodes of a given tree in a given, say "original", time-order are visited and labelled in the new time-order, one after the other. The procedure stops, as soon as the tree is completely labelled by new time-labels. During the entire procedure the tree remains valid but gets rearranged. This rearrangements takes place in regions only, where no new time labels have been assigned yet.

To facilitate the procedure, the term resulting from the original time ordering is represented in from of those $\Pi_{L,i}$ mentioned above, with the original time labels still attached. *The idea is that all nodes from position i are visited by the new time-ordering in the order prescribed by $\Pi_{L,i}$.*

If a node $\mathsf{X}_{L,i}$ gets visited in the new time ordering, which does not correspond to the prescription, in original time-label or value, then the subtree currently dangling on that node is swapped with the entire subtree dangling on the node which is actually supposed to be visited. Since it has not been visited before, neither of these subtrees can wear new labels. Repeating this procedure, one rearranges the tree until it is entirely covered by new labels. It cannot happen that some nodes do not get visited , because the tree remains valid and for every valid tree a complete new la-

[16]But of course, the same tree has different time-labels for different time-ordering schemes, as seen in Fig. 5.12.

[17]In the following a tree is called "valid" if all leaves are of type $\mathsf{S}_{L,i}$ and all internal nodes are of type $\mathsf{T}_{L,i}$, each having two children (direct descendants) $\mathsf{X}_{L,i-1}$ and $\mathsf{X}_{L,i+1}$ where X can be either S or T.

belling exists. Also, the tree changes only where not labelled yet by new labels. The procedure terminates because there is only a finite number of nodes to be labelled.

Fig. 5.13 shows the procedure for rearranging Fig. 5.12(a) such that it produces the same term as Fig. 5.12(b). In this example the $\Pi_{L,i}$ are

$$\begin{aligned}
\Pi_{L,i-2} &= \mathsf{S}^{(t=9)}_{L,i-2} \\
\Pi_{L,i-1} &= \mathsf{T}^{(t=7)}_{L,i-1}\mathsf{S}^{(t=6)}_{L,i-1} \\
\Pi_{L,i} &= \mathsf{S}^{(t=8)}_{L,i}\mathsf{T}^{(t=4)}_{L,i}\mathsf{T}^{(t=1)}_{L,i} \\
\Pi_{L,i+1} &= \mathsf{S}^{(t=5)}_{L,i+1}\mathsf{T}^{(t=2)}_{L,i+1} \\
\Pi_{L,i+2} &= \mathsf{S}^{(t=3)}_{L,i+2}
\end{aligned}$$

The time labels are now represented in the form $[t_{\mathrm{old}}, t_{\mathrm{new}}]$.

Looking at Fig. 5.13(c), the resulting terms are correctly

$$\begin{aligned}
\Pi_{L,i-2} &= \mathsf{S}^{[9,8]}_{L,i-2} \\
\Pi_{L,i-1} &= \mathsf{T}^{[7,6]}_{L,i-1}\mathsf{S}^{[6,2]}_{L,i-1} \\
\Pi_{L,i} &= \mathsf{S}^{[8,9]}_{L,i}\mathsf{T}^{[4,4]}_{L,i}\mathsf{T}^{[1,1]}_{L,i} \\
\Pi_{L,i+1} &= \mathsf{S}^{[5,7]}_{L,i+1}\mathsf{T}^{[2,3]}_{L,i+1} \\
\Pi_{L,i+2} &= \mathsf{S}^{[3,5]}_{L,i+2}
\end{aligned}$$

now ordered using the new ordering.

It is worth noting that the rearrangement of trees maintains causality, provided that the new time labelling does so. This can be ensured by labelling the nodes sequentially, each at the time of its expansion.

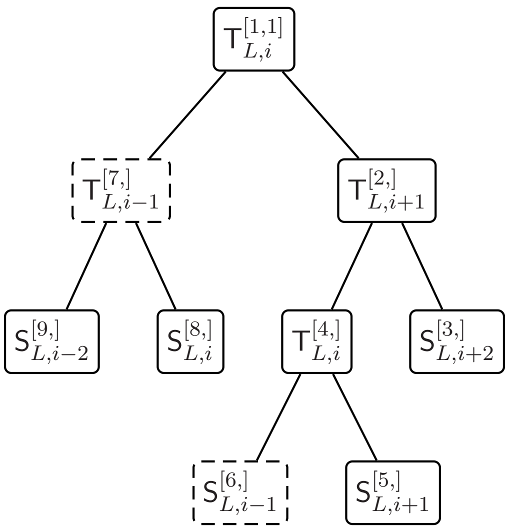
Conclusively, each term obtained in one time-ordering can be obtained in any other time ordering by explicitly rearranging the corresponding trees.

5.3.3.1 Abelian nature II

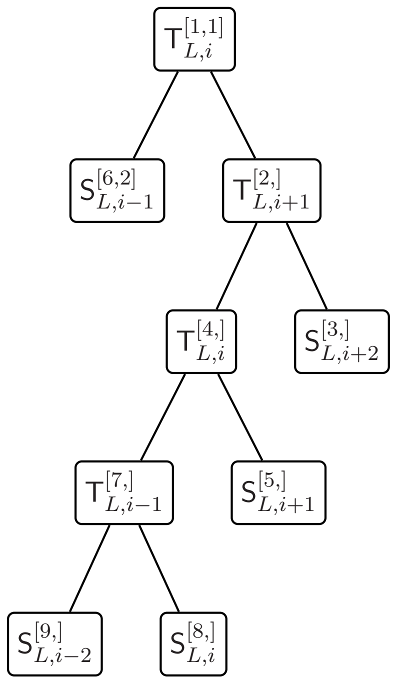
The procedure described above can also be used to produce the correct sequence of operations from any *two* trees considered at the same time. The $\Pi_{L,i}$ are constructed from multiplying the resulting terms of both trees and contain terms from both in the ordering prescribed by the time order.

The aim of the following is to prove the commutator $\mathbf{O}_{L,i}\mathbf{O}_{L,j} = \mathbf{O}_{L,j}\mathbf{O}_{L,i}$ within the *same* time ordering. We assume that the product $\mathbf{O}_{L,i}\mathbf{O}_{L,j}$ is given and want to show that every term in $\mathbf{O}_{L,i}\mathbf{O}_{L,j}$ is also contained in $\mathbf{O}_{L,j}\mathbf{O}_{L,i}$.

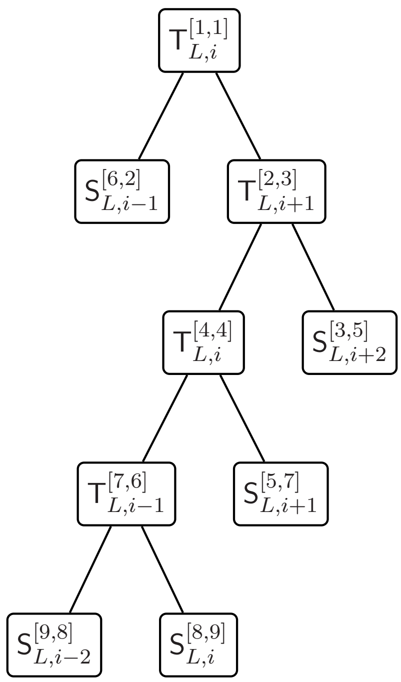
The idea is that two trees are given, one representing a term from $\mathbf{O}_{L,i}$, the other one representing a term from $\mathbf{O}_{L,j}$. When labelling the trees in the product $\mathbf{O}_{L,i}\mathbf{O}_{L,j}$,



(a) The dashed nodes get exchanged in the next step



(b) The nodes are exchanged, rest of the labeling starts.



(c) Relabeling finished.

Figure 5.13: Starting from the original tree wearing original labels (Fig. 5.12(a)), a new labelling is done by the procedure described in the text. From Fig. 5.13(a) to Fig. 5.13(b) the tree actually changes. In the last step, labels are assigned to all remaining nodes without making necessary any further rearrangement.

the trees in $\mathbf{O}_{L,i}$ are labelled *after* $\mathbf{O}_{L,j}$ has been labelled. Picking now the root in the tree of $\mathbf{O}_{L,i}$ (formerly the *second* operation) and use it as a starting point for rearrangements based on the time order prescribed by the $\Pi_{L,k}$ from $\mathbf{O}_{L,i}\mathbf{O}_{L,j}$, one can reshape the two trees until the first one terminates. After that, one continues with the root of the first tree (from $\mathbf{O}_{L,j}$).

The procedure applied to two trees within the same time ordering is very similar to the procedure discussed above to identify corresponding terms in different time orderings. Again, the procedure terminates after all nodes have been visited exactly once and in the right order. Since both trees remain valid and both trees will keep having root nodes at position i and j, even though the root nodes will possibly be exchanged with subtrees during the course of the rearrangement, one arrives at two valid trees, which *must* necessarily be contained in the set of terms produced by $\mathbf{O}_{L,j}\mathbf{O}_{L,i}$. Thus

$$\mathbf{O}_{L,i}\mathbf{O}_{L,j} = \mathbf{O}_{L,j}\mathbf{O}_{L,i} \tag{5.48}$$

which proves the Abelian property.

The reason why the procedure described in Sec. 5.3.3 works on any number of trees is because it does not make use of the fact that the trees under consideration are actually connected. The procedure rearranges the trees according to a prescribed time ordering, while the roots remain operations on the same position i (swapping subtrees cannot change the position label).

In fact, the above procedure can be used for many other models as well. The only properties used in showing the Abelian property of the $\mathbf{O}_{L,i}$ and the irrelevance of the ordering scheme are

Operator representation: The updates of the model ($\mathbf{O}_{L,i}$) must be expressible as products of operators ($\mathsf{S}_{L,i}, \mathsf{T}_{L,i}, \mathbf{O}_{L,i}$), so that one can actually represent them as trees.

Commutator: Non-local operators commute ($\mathsf{S}_{L,i}, \mathsf{T}_{L,j}$ for $i \neq j$), while local operators acting on each other might not, see Eq. (5.40), so that the $\Pi_{i,L}$ can be constructed.[18]

Sequential *local* updates: Multiple *local* charges must be equivalent to powers of those operators ($\mathbf{O}_{L,i}$). This is needed to make sure that the charging of a site can actually be written as a product of *independent* single charges, as has been done above. No special operators are needed for multiple charges. Otherwise, within a single term in the expansion of an operator, the subtrees might depend on each other in order to act with the correct operator depending on the *total*

[18]Clearly, the commutator of $\mathbf{O}_{L,i}$ is exactly to be found, so this operator is not relevant for this rule.

number of charges at a given instant or so. Such an interdependence might not be preserved under rearrangements of the trees. Note however, that this condition is implicit in the first one.[19]

The last point is the reason, why the MANNA model is not Abelian in its original definition; since an active ($z_i > z^c$) site i discharges down to $z_i = 0$ whenever it is updated, it is crucially important whether it is charged by another slope unit before or after the relaxation.

5.3.4 Oslo algebra

After clearing up the issue that the operations in the Oslo model can be written in any order, one can accept the expansion Eq. (5.42), which is repeated here together with the commutators:

$$\begin{aligned}
\mathbf{O}_{L,i} &= \mathsf{S}_{L,i} + \mathbf{O}_{L,i-1}\mathbf{O}_{L,i+1}\mathsf{T}_{L,i} \text{ for } 1 \le i \le L && \text{(5.49a)} \\
& \quad \text{where } \mathbf{O}_{L,0} = \mathbf{O}_{L,L+1} = \mathbf{1}^{\otimes L} \\
\left[\mathsf{S}_{L,i}, \mathsf{S}_{L,j}\right]_- &= 0 && \text{(5.49b)} \\
\left[\mathsf{T}_{L,i}, \mathsf{T}_{L,j}\right]_- &= 0 && \text{(5.49c)} \\
\left[\mathsf{T}_{L,i}, \mathsf{S}_{L,j}\right]_- &= \delta_{i,j} Q_{L,i} && \text{(5.49d)}
\end{aligned}$$

From what has been shown in the preceeding section (sec. 5.3.3), this alone is enough to show the irrelevance of ordering in the expansion of the operator and the Abelian property.

However, the Oslo matrices have some other features as well. First of all

$$\begin{aligned}
\mathsf{S}^n_{L,i} &= 0 \quad \text{for } n > 2 && \text{(5.50a)} \\
\mathsf{T}^n_{L,i} &= 0 \quad \text{for } n > 2 && \text{(5.50b)}
\end{aligned}$$

which means that a site cannot absorb more than two slope units and cannot topple more then twice in a row.

For sufficiently small systems ($L \le 7$, possibly $L \le 8$), it is fairly straight forward to derive $\mathbf{O}_{L,i}$ directly and exactly, using a simple algorithm to expand it in the form of Eq. (5.49). The idea is to start the procedure of kicking and toppling once for each initial configuration and to calculate the (cumulated) probability to arrive at a certain final configuration. Whenever a random number would be drawn in a Monte-Carlo implementation, the current configuration goes on a stack and both

[19] It seems that all conditions can be summarised as: If a set of Markovian operators $\mathbf{O}_{L,i}$ can be written as a sum of products of non-locally commuting operators $\mathsf{S}_{L,i}, \mathsf{T}_{L,i}, \ldots$, then the $\mathbf{O}_{L,i}$ commute themselves and the ordering scheme in the expansion is irrelevant.

possible outcomes are taken into account with appropriate probabilities. The resulting matrix can be diagonalised. DHAR has recently proven (Dhar, 2003) that, using the original boundary conditions introduced in Sec. 5.1.1, page 205, a single kick of the initial condition $z_i = 2\forall_i$ produces the stationary state. Knowing the stationary state exactly would imply the knowledge of the roughness exponent in the interface picture (Sec. 6.1.5, page 255).

It is therefore indeed more than an exercise in algebra to pursue this approach a bit further. However, so far results have been very limited. For example, the attempt to generalise the observation that Open issue

$$\mathbf{O}_{2,1} = \mathsf{S}_{2,1} + \mathsf{S}_{2,2}\mathsf{T}_{2,1} + \mathbf{O}_{2,1}\mathsf{T}_{2,2}\mathsf{T}_{2,1} \tag{5.51}$$

entails

$$\mathbf{O}_{2,1} = (\mathsf{S}_{2,1} + \mathsf{S}_{2,2}\mathsf{T}_{2,1})\,(1 - \mathsf{T}_{2,2}\mathsf{T}_{2,1})^{-1} \tag{5.52}$$

to larger L failed. Moreover, it is numerically extremely convincing, yet remains unproven that the stationary state of $\mathbf{O}_{L,i}$ is the same as the stationary state of $\mathbf{O}_{L,L+1-i}$ — it is not even clear yet which property of $\mathbf{O}_{L,i}$ is exactly causing this. However, it seems that it cannot be observed using the original boundary conditions. Open problem

One of the biggest problems lies in the fact that certain configurations are obviously not accessible by the (quasi-) dynamics (Chua and Christensen, 2002); even though this is numerically a very convincing fact, which can be proven for L small enough such that the operator $\mathbf{O}_{L,i}$ can be calculated exactly, it could not yet be proven for general L.

Probably the most promising paths are those using Density Matrix Renormalisation Group-style techniques (Peschel *et al.*, 1999), which have been used already successfully for one-dimensional stochastic models (Kaulke, 1999). The key idea is to iteratively extend the operator $\mathbf{O}_{L,1}$ to larger system sizes, by devising a method to extend $\mathbf{O}_{L,1}$ to $\mathbf{O}_{L+1,1}$. One way would be to write $\mathbf{O}_{L,1}$ in an expansion over the number of grains spilled by the leftmost site,

$$\mathbf{O}_{L,1} = \sum_{j=0}^{\infty} \Omega_{L,j} \tag{5.53}$$

which starts an iteration, if the dissipative identity at $i = 0$, $\mathbf{O}_{L,0} = \mathbf{1}^{\otimes L}$ is replaced by a site. However, it is not even clear yet how to formally expand $\mathbf{O}_{L,1}$, let alone, how to derive the iteration prescription for $\{\Omega_{L,j}\} \rightarrow \{\Omega_{L+1,j}\}$.

One way out might be to actually write down a time-evolution operator, which operates on the *microscopic* timescale, rather than the operator $\mathbf{O}_{L,i}$ which triggers and relaxes the system completely and therefore operates on the macroscopic timescale.

5.3.4.1 MANNA algebra

The Abelian MANNA model[20] provides an alternative path to understand sandpile-like models algebraically. However, the more attractive algebraic properties come together with infinitely many terms, as the MANNA model has a finite probability to continue to topple at each update. The model has already been introduced in Sec. 5.2.2.3, page 223, and its expression in terms of operators is straight forward; most of the results derived for the Oslo model above extend quite easily to the MANNA model. Since there are only two possible stable states per site the matrices can be understood as flipping a spin. Denoting creation and annihilation operators[21] by

$$C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{5.54a}$$

$$C^\dagger = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{5.54b}$$

one can express the MANNA model operator by

$$\mathbf{O}_{L,i} = C^\dagger_{L,i} + \frac{1}{4}(\mathbf{O}_{L,i-1} + \mathbf{O}_{L,i+1})^2 C_{L,i} \tag{5.55}$$

with appropriate boundary conditions

$$\mathbf{O}_{L,0} = \mathbf{O}_{L,L+1} = \mathbf{1}^{\otimes L} \tag{5.56}$$

and $C^{(\dagger)}_{L,i} = \mathbf{1}^{\otimes i-1} C^{(\dagger)} \mathbf{1}^{\otimes L-i}$. Of course, $\mathbf{1}$ is now a 2×2 matrix. The Abelian nature of the MANNA model follows from the same proof as in sec. 5.3.3.1, just that there are *four* different nodes now: One leaf node $C^\dagger_{L,i}$ and three different internal, branching nodes, depending on which two operators branch off, see Eq. (5.57).

In Eq. (5.55) it has been assumed that there is equal probability for each toppling particle to get redistributed to either side; while this simplifies the expression compared to

$$\mathbf{O}_{L,i} = C^\dagger_{L,i} + (p\mathbf{O}^2_{L,i-1} + (1-p-q)\mathbf{O}_{L,i-1}\mathbf{O}_{L,i+1} + q\mathbf{O}^2_{L,i+1})C_{L,i} \quad \text{where } 0 \leq p+q \leq 1 \tag{5.57}$$

the latter is much more flexible and could also incorporate the one-dimensional BTW model by choosing $p = q = 0$ and appropriate driving. In fact, with the simplified boundary conditions (5.56) and accepting the conjecture that the MANNA model (for

[20] Thanks to MATTHEW STAPLETON for pointing out that his model, which turned out to be a "generalised MANNA model", has critical properties very similar to those found in the Oslo model.

[21] Thanks to ALVIN CHUA for pointing out this elegant analogy.

$p = q$) is in the same universality class as the Oslo model, the expected phase diagram is shown in Fig. 5.14. Obviously, for $p = q = 1/4$ Eq. (5.55) is recovered.

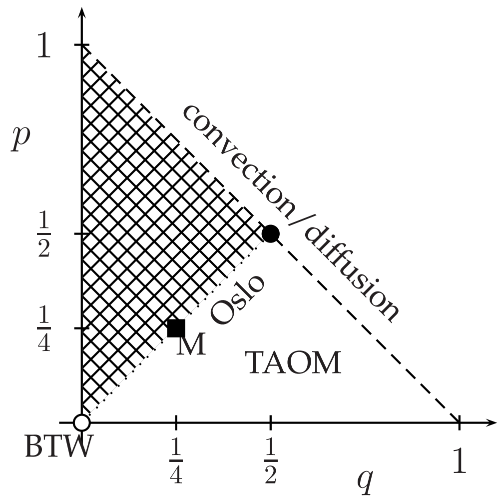


Figure 5.14: Conjectured phase diagram of the MANNA model with simplified boundary conditions — what is shown needs additional numerical and analytical backup, but is very reasonable. The crosshatched area is not expected to be scale-invariant. The BTW model (open circle) is recovered for $p = q = 0$, the Abelian form of the original MANNA model if recovered at $p = q = 1/4$ (black square). The whole (dotted) line $p = q$ should belong to the Oslo universality class, unless for $p = q = 1/2$ (black circle) which is purely diffusive. For $p + q = 1$ (dashed line) but $p \neq q$ the slope units propagate in pairs with a net drift (convection), so that for $q > p$ trivial exponents should be obtained (see text). For $p + q = 1$ and $q < p$, *i.e.* where the dashed line touches the crosshatched area, no scaling behaviour is expected. The TAOM (see Chapter 7, page 281)is recovered for all $q > p$ and $p + q < 1$.

Any even number of charges and especially a double charge leaves the state of the site unchanged. In fact, if $p + q = 1$ in (5.57), then there is only one double-charge propagating through the system (see below) and the configuration does not change between two non-vanishing avalanches. The distribution of avalanche sizes then corresponds to the distribution of live-times of a random walker between two absorbing walls. For $p \neq q$ there is a net drift, which represents a cutoff, that is either a constant (drift pointing towards left boundary) or proportional to L (drift pointing to right boundary). In the former case, the scaling is characterised by the position of the initial kick, which is a constant. The latter case produces avalanche moments $\langle s^n \rangle \propto L^n$ and therefore $D = 1$ and $\tau = 1$. Only for $p = q = 1/2$ there is no "simple cutoff" and one recovers the return-time distribution of a random walker (Redner, 2001), *i.e.* $\tau = 3/2$ and $D = 2$, the latter from the time it takes a random walker to span the system. The universality class of the TAOM (see Chapter 7) with $\tau = 4/3$ and $D = 3/2$ is expected for the whole region $q > p$, unless $p + q = 1$. If the right

Open issues

boundary condition is the original one of the Oslo model, "jamming"[22] is expected on the right hand side and avalanches run "forever". [23]

The effect of double charging can be directly derived from (5.57) using the commutators

$$\begin{aligned}
\left[\mathbf{O}_{L,i}, \mathbf{O}_{L,j}\right]_- &= 0 && (5.58a)\\
\left[C^\dagger_{L,i}, C_{L,j}\right]_- &= \delta_{i,j}\mathbf{1}^{\otimes i-1} \otimes Q \otimes \mathbf{1}^{\otimes L-i} && (5.58b)\\
\left[C^\dagger_{L,i}, C_{L,i}\right]_+ &= \mathbf{1}^{\otimes L} \ . && (5.58c)
\end{aligned}$$

with

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ . \tag{5.59}$$

Moreover, similar to Eq. (5.50)

$$\begin{aligned}
\left(C^\dagger_{L,i}\right)^2 &= 0 && (5.60a)\\
\left(C_{L,i}\right)^2 &= 0 && (5.60b)
\end{aligned}$$

and

$$C_{L,i}C^\dagger_{L,i}C_{L,i} = C_{L,i} \ . \tag{5.61}$$

Then one has

$$\mathbf{O}^2_{L,i} = C^\dagger_{L,i}WC_{L,i} + WC_{L,i}C^\dagger_{L,i} + WC_{L,i}WC_{L,i} \tag{5.62}$$

where

$$W = p\mathbf{O}^2_{L,i-1} + (1-p-q)\mathbf{O}_{L,i-1}\mathbf{O}_{L,i+1} + q\mathbf{O}^2_{L,i+1} \ , \tag{5.63}$$

so that $\mathbf{O}^2_{L,i}C_{L,i} = WC_{L,i}$ using (5.61). Moreover $\mathbf{O}_{L,i}C^\dagger_{L,i} = WC_{L,i}C^\dagger_{L,i}$ and therefore

$$\mathbf{O}^2_{L,i}C^\dagger_{L,i} = W\mathbf{O}_{L,i}C_{L,i}C^\dagger_{L,i} = WC^\dagger_{L,i}C_{L,i}C^\dagger_{L,i} \tag{5.64}$$

noting that $\left[\mathbf{O}_{L,i}, W\right]_- = 0$ because of (5.58a). Thus

$$\mathbf{O}^2_{L,i}(C^\dagger_{L,i}C_{L,i} + C_{L,i}C^\dagger_{L,i}) = WC^\dagger_{L,i}C_{L,i} + WC_{L,i}C^\dagger_{L,i} \ , \tag{5.65}$$

which gives with (5.58c) and (5.63) the desired result, namely that charging site i twice is like charging the neighbours appropriately but leaving site i *externally* un-

[22] Thanks to KIM CHRISTENSEN and MATTHEW STAPLETON for pointing that out. One might ask, in what sense this is still SOC, if any net-drift leads to non-critical behaviour.

[23] All exponents cited in this paragraph are preliminary and require additional analytical and numerical backup. Especially the calculation of higher moments of the lifetime of a random walker between two absorbing walls should be accessible by rigorous methods, at least in the continuum.

charged[24],

$$\mathbf{O}_{L,i}^2 = W = p\mathbf{O}_{L,i-1}^2 + (1-p-q)\mathbf{O}_{L,i-1}\mathbf{O}_{L,i+1} + q\mathbf{O}_{L,i+1}^2 \ . \tag{5.66}$$

For $p+q=1$ this simplifies to a process, where any site, if charged, is charged twice. If (5.66) is expanded[25] for $p+q=1$, it will produce infinitely many terms finally terminating in either $\mathbf{O}_{L,0}$ or $\mathbf{O}_{L,L+1}$ and each one carrying a certain probability. All these probabilities will necessarily add up to 1, *i.e.*

$$\mathbf{O}_{L,i}^2 = \mathbf{1}^{\otimes L} \text{ for } p+q=1 \ . \tag{5.67}$$

Getting back to the general case of $p+q \neq 1$ or just (5.55), what makes the analytical treatment of the MANNA model a bit inconvenient is the fact that, in contrast to the Oslo model, its recursive definition in general produces infinitely many terms. This is because there are ballistic terms ($\mathbf{O}_{L,i}^2$) in (5.57) which do not necessarily transport anything closer to the boundary and therefore make the expansion necessarily terminate after a finite number of terms. To illustrate that, one can "follow the double charges" discussed above, *i.e.* when doing the expansion of (5.57) one follows only the "branch" which produces double charges in an appropriate direction. For example, expanding initially $\mathbf{O}_{L,1} = q\mathbf{O}_{L,2}^2 C_{L,1}$ and further charges as

$$\mathbf{O}_{L,2}^2 \rightarrow p\mathbf{O}_{L,1}^2 \tag{5.68a}$$
$$\mathbf{O}_{L,1}^2 \rightarrow q\mathbf{O}_{L,2}^2 \tag{5.68b}$$

does not lead to a finite expansion. Instead one obtains

$$\mathbf{O}_{L,1} \rightarrow qpqpq\ldots\mathbf{O}_{L,2}^2 \ . \tag{5.69}$$

Therefore, even the exact stationary state of a system of size $L=2$ is not as simple to calculate as for the Oslo model.

One interesting way to explore the MANNA model is to make the spin-operators $C_{L,i}^\dagger$ and $C_{L,i}$ fermionic. In the current form, (5.58c) is the correct fermionic anticommutation relation for a single site. However, on different sites the operators commute, see Eq. (5.58b). Moreover, Open issue

$$\left[C_{L,i}, C_{L,j}\right]_- = 0 \tag{5.70a}$$
$$\left[C_{L,i}^\dagger, C_{L,j}^\dagger\right]_- = 0 \ , \tag{5.70b}$$

[24] Of course, if the neighbours topple, they might change the state of the site initially charged twice, but the change is not directly due to the initial charging.

[25] "Expansion" here refers to the same procedure as discussed in sec. 5.3.2.1.

while fermionic operators should anti-commute. However, fermionisation of spin operators is a standard problem (Mahan, 1990, pp. 46–53); most prominently, in the ISING model the JORDAN-WIGNER transform has been used for this purpose (Wannier, 1987, pp. 356–365).

5.4 Outlook and Summary

Having discussed the Oslo model and its various variants, this model seems to be the most promising route towards an analytical understanding of SOC. The Oslo model shows all universal features expected from equilibrium statistical mechanics, especially a universal finite-size scaling function, which only depends on boundary conditions, but not on local properties such as next nearest neighbour interaction.

5.4.1 Summary

Here is a summary of the sections in this chapter;

- The model is defined in Sec. 5.1, first in its original form and then with simplified boundaries. Various properties can be derived exactly in the stationary state.
- It is most encouraging that the Oslo model can be changed quite dramatically in its microscopic rules without changing the critical, presumably universal features. The variants are described in Sec. 5.2. Numerical results are listed and briefly discussed in Sec. 5.2.3.
- As shown in Sec. 5.3, an operator approach can be developed for the Oslo model. This seems to be a promising, but not yet complete path to an exact solution. As shown in Chapter 6, an exact solution of the Oslo model amounts to an exact solution of the quenched EDWARDS-WILKINSON equation.
- In particular, standard techniques from one-dimensional stochastic processes, such as a matrix product ansatz (Derrida and Evans, 1997) might be suitable for the Oslo model. The MANNA model can be treated similarly, see Sec. 5.3.4.1. An operator algebra similar to the one discussed above will be used in Chapter 8 to solve the "totally asymmetric Oslo model".

In the next chapter, a field theoretic approach is introduced (essentially by identifying the correct dynamical variable) so that an exact equation of motion can be written down. This equation can be explored in various ways, especially by studying the periodic Oslo model (which is an AS model), Sec. 6.3, page 273, and models with anisotropy, which will be discussed in Chapter 7.

Chapter 6

The Oslo Model and the Quenched EDWARDS-WILKINSON Equation

Having identified the Oslo model as a "true model of SOC" in the previous chapter, it is highly desirable to find an analytical framework which puts the model on a firm theoretical footing. In this chapter the Oslo model is identified as a discrete realisation of the quenched EDWARDS-WILKINSON *equation (Nattermann* et al.*, 1992). This is made possible by choosing the correct dynamical variable and identifying its equation of motion. It establishes for the first time an exact link between SOC models and the field of interface growth with quenched disorder. This connection is obviously very encouraging as it suggests that established theoretical techniques can be brought to bear with full strength on some of the hitherto elusive problems of SOC.*

6.1 Towards a Field Theory

In this section it will be shown which dynamical variable in the Oslo model obeys the quenched EDWARDS-WILKINSON (qEW) equation and how to derive it. Remarkably, the equation of motion will operate on the microscopic timescale, *i.e.* the qEW equation is just *one possible* realisation of the Oslo model. Other dynamics are permitted as well. This stresses once more the ambiguity of the dynamical exponent, see Sec. 6.1.3.3.

6.1.1 Introduction

The Oslo rice pile model (Oslo model hereafter) as discussed in Chapter 5 was originally intended to model the relaxation processes in real rice piles (Frette *et al.*, 1996). Meanwhile, it has been subject to many investigations and publications in its own right. The model as defined below supposedly develops into a scale free state without the explicit tuning of external parameters, and is therefore regarded as an example of Self-Organised Criticality (SOC) (Jensen, 1998). In fact, contrary to many other "standard" models of SOC (Datta *et al.*, 2000; Dorn *et al.*, 2001; Lise and Paczuski, 2001a; Pruessner and Jensen, 2002a), it shows a reliable and consistent (simple) scaling behaviour and is robust against certain changes in the details of the dynamics (Zhang, 1997; Bengrine *et al.*, 1999a,b). The most prominent observable in the model, the avalanche size s, is governed by a probability distribution $\mathcal{P}(s)$ which obeys simple scaling (Sec. 2.3, page 78),

$$\mathcal{P}(s) = s^{-\tau}\mathcal{G}(s/s_0) \text{ and } s_0 = L^D, \tag{6.1}$$

where L denotes the system size and τ and D are critical exponents, consistently reported to be $\tau = 1.55(10)$ and $D = 2.25(10)$ (Christensen *et al.*, 1996; Paczuski and Boettcher, 1996; Zhang, 1997; Bengrine *et al.*, 1999a; Corral and Paczuski, 1999; Bengrine *et al.*, 1999b). These two exponents are related by (Christensen *et al.*, 1996; Paczuski and Boettcher, 1996)

$$D(2-\tau) = 1\ , \tag{6.2}$$

which can be proven easily given that the first moment of $\mathcal{P}(s)$, $\langle s \rangle$, scales like L.

In the following the model is defined again explicitly, the relevant dynamical variable extracted and its equation of motion derived, which turns out to be a discretised quenched EDWARDS-WILKINSON (qEW)[1] equation. By analysing the essential characteristics of the model on the lattice, such as uniqueness of the solution and symmetries, it is then possible to construct the continuum theory, which can subsequently be examined using standard methods.

The model (Christensen *et al.*, 1996) is defined on a one-dimensional grid of size L, where each site $i = 1 \cdots L$ has slope z_i and critical slope $z_i^c \in \{1, 2\}$. Starting from an initial configuration with $z_i = 0$ and z_i^c random everywhere, the model evolves according to the following update rules: 1) (Driving) Increase z_1 by one. 2) (Toppling) If there is an i with $z_i > z_i^c$ decrease z_i by 2 and increase its nearest neighbours by one, $z_{i\pm1} \to z_{i\pm1} + 1$, provided that $1 \leq i \pm 1 \leq L$. A new z_i^c is chosen at random, 1 with probability p and 2 with probability $q \equiv 1 - p$. 3) Repeat the second step until $z_i \leq z_i^c$ everywhere. Then proceed with the first step. The order of updates is irrelevant in this model and the original definition does not fix it explicitly. Therefore the microscopic (fast) timescale is *a priori* undefined.

The avalanche size s is defined as the number of charges, *i.e.* apart from boundary

[1]Sometimes the quenched EDWARDS-WILKINSON equation is called the "BRUINSMA-AEPPLI-KOPLIK-LEVINE" equation, after the four authors who first considered it (Bruinsma and Aeppli, 1984; Koplik and Levine, 1985).

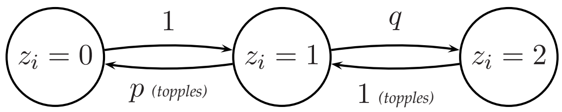


Figure 6.1: Each site can be in one of three states and changes stepwise between them, whenever it receives a charge. The labels indicate the probability of the move and whether it entails a toppling.

effects, twice the number of times the second rule is applied between two consecutive application of the first rule. For convenience the model is dissipative on both boundaries, where one of the two "units" lost by the boundary site during toppling leaves the system.

6.1.2 Equation of motion for the proper dynamical variable

A few years ago PACZUSKI and BOETTCHER translated the Oslo model into the language of interfaces in random media (Paczuski and Boettcher, 1996). However, the evolution of the dynamical variable $H(x,t)$, which is the total number of topplings of site x, was given by $\partial_t H = \theta(\partial_x^2 H - \eta(x,H))$, where ∂_t is defined in discrete time, *i.e.* $\partial_t H \equiv H(x,t+1) - H(x,t)$ and ∂_x^2 is the lattice Laplacian, so that x is actually an index. The last term, $\eta(x,H)$, represents a quenched noise. The HEAVISIDE θ-function makes this equation of motion highly nonlinear and analytically almost intractable (Díaz-Guilera, 1992). PACZUSKI and BOETTCHER have already conjectured that the Oslo model is in the same universality class as qEW (Nattermann *et al.*, 1992). More recently, MIKKO ALAVA has suggested that certain other sandpile models are described by the qEW (Alava, 2002). It is, however, important to realize that no rigorous and exact link has so far been established between SOC models and the qEW equation.

The crucial step to make this correspondence exact is to identify the proper dynamical variable. It is found in form of the number of times a site has been charged (*i.e.* received a unit from a neighbour during a toppling or by external drive, see below) $h(x,t)$, where x and t are discrete for the time being. There is a simple functional relation between $h(x,t)$ and $H(x,t)$, which can be obtained as follows: Each site can be in one of three stable configurations, $z_i \in 0,1,2$. When a site receives a unit from a neighbour, it changes state as shown in Fig. 6.1. Charging a site in state 0 necessarily leads to state 1 without toppling and the specific value of z_i^c is completely irrelevant at this stage. Similar for state 2: If a site receives a charge in this state, its z_i^c must be 2 and it must topple. The only point where the value of z_i^c actually matters, is in state 1, therefore it can be effectively chosen at random when necessary, so that the site topples with probability p (according to the probability of having $z_i^c = 1$) or increases to 2 with probability q (see Fig. 6.1). It is immediately clear that any even number of charges, say $m = 2n$, starting from $z_i = 1$ leads to state 1 again with n topplings. An odd number of charges, say $m = 2n + 1$, leads either to n topplings and state 2 or $n + 1$ topplings and state 0. This is illustrated in Fig. 6.1: The m charges lead to m steps along the arrows. Whenever one moves left, the site topples.

In order to write a functional relation between $h(x, t)$ and $H(x, t)$, the randomness in the decision of moving to the left or to the right from state 1 must be quenched in $h(x, t)$, *i.e.* it is not allowed to change unless $h(x, t)$ changes. This can be summarised as

$$H(x, t + 1) = \frac{1}{2} \left(h(x, t) + \eta(x, h(x, t))\right) \ , \tag{6.3}$$

where η is 0 whenever $h(x, t)$ is even, corresponding to state 1. If $h(x, t)$ is odd, η is either 1 (with probability p, state $z_i = 0$) or -1 ($z_i = 2$). Every sequence of $\eta(x, h)$ values maps uniquely to a sequence of z_i^c and vice versa, see Sec. 6.1.5. The equation above can easily be transformed to comply to any initial configuration, especially to $z_i(t = 0) \equiv 0$. Essentially, it is (6.3), which makes the exact identification of the Oslo model and qEW possible.

The final equation is derived by noting that obviously $h(x, t) = H(x-1, t)+H(x+1, t)$ with appropriately chosen boundary conditions (BC's) (see below), so that using the short hand notation $h^{\pm} = h(x \pm 1, t)$ and $\eta^{\pm} = \eta(x \pm 1, h^{\pm})$ the equation of motion is

$$h(x, t + 1) - h(x, t) = \frac{1}{2} \left(h^{-} - 2h(x, t) + h^{+} + \eta^{+} + \eta^{-}\right) \ , \tag{6.4}$$

which is the *exact* representation of the Oslo model as defined above, captured in a single equation. Its differential form is accordingly

$$\partial_t h(x, t) = \frac{1}{2}\partial_x^2 h(x, t) + \left(1 + \frac{1}{2}\frac{d^2}{dx^2}\right) \eta(x, h(x, t)) \ . \tag{6.5}$$

The right hand BC is

$$h(x = L + 1, t) \equiv 0 \tag{6.6}$$

(and $h(x = L, t) \equiv 0$ in the continuum), while the left hand BC provides the driving via

$$h(x = 0, t) = 2E(t) \ , \tag{6.7}$$

$E(t)$ being the total number of initial seeds (step 1 above) at time t. These seeds arrive at site $x = 1$ via the Laplacian (see Sec. 6.1.5). In the continuum, the simplest drive is

$$E(t) = vt \tag{6.8}$$

with v a driving velocity and t the microscopic time. Together with the BC's, Eq. (6.5) or the generalised form

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) + g\left(1+\lambda\frac{d^2}{dx^2}\right)\eta(x,h(x,t))\ , \tag{6.9}$$

where the correlator of η is now normalised, *i.e.* $\int dx \int dh \langle\eta\eta\rangle = 1$, describes the movement of an elastic band over a rough surface (Dong *et al.*, 1993) pulled by a transverse force acting at one end point only. Below it is shown that the λ-term disappears in the continuum, establishing the *first rigorous identification* of the Oslo model and the qEW equation. The same equation with different properties of the noise term and/or different BC's applies to other models, such as the BTW model [Sec. 1.3.1, page 40, (Bak *et al.*, 1987)], Fixed Energy Sandpiles [for example (Vespignani *et al.*, 2000)] or the tilted sandpile [Sec. 5.2.2.2, page 222, (Malthe-Sørenssen, 1999)]. Having identified the relevant dynamical variable h, the effect of modifications of the dynamical rules of the Oslo model, such as (Zhang, 1997; Bengrine *et al.*, 1999a,b), can be understood.

The equation above exemplifies a general "trick"[2] to get rid of θ-functions in equations of motion — they often appear in descriptions of sandpile-like systems [for example (Díaz-Guilera, 1992)]: One simply replaces $\theta(h-h_c)$ by $h+\eta(h)$ with an appropriately chosen sawtooth-like η. This does not necessarily simplify the problem, unless there is already a quenched noise present in the system. In this case the θ turns into a correlation in η. This is highly remarkable from the point of view of SOC, because the presence of "thresholds" is usually expected to be a crucial ingredient of SOC (Bak *et al.*, 1987; Cafiero *et al.*, 1995; Jensen, 1998). Moreover, the correlations in η, which are of fundamental significance in interface models (Nattermann *et al.*, 1992; Leschhorn *et al.*, 1997) and have been neglected in former mappings, now arise naturally from the dynamical description of the model.

6.1.3 Continuum theory

In order to construct the proper continuum theory, it is worthwhile to consider the formal solution of Eq. (6.9). It will turn out later that $E(t) = vt$ is sufficiently general, so that it makes sense to define

$$v(x) \equiv v\frac{L-x}{L} \tag{6.10}$$

and

$$h(x,t) = 2v(x)t + P_3(x) + z(x,t) \tag{6.11}$$

[2] The substitution was already used earlier (see (Alava, 2002) and references therein).

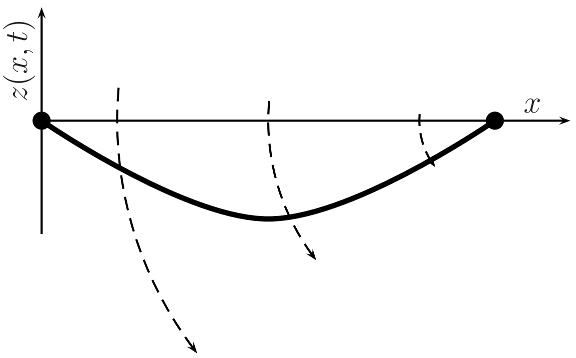


Figure 6.2: Cartoon of an elastic band (thick line) fixed at its two ends (black circles), while the rough background is moving with a velocity linearly increasing from right to left like a cutting disk rotating around $x = L$, as indicated by the arrows.

in order to homogenise the BC's. After the data shift $z(x,t)$ describes an interface fixed on its two boundaries, sliding over a surface which moves with velocity $2v(x)t$. Because $v(x)$ is linear in x and $v(x = L) = 0$, the movement corresponds to that of a rough disk rotating around $x = L$, see Fig. 6.2.

$P_3(x)$ in (6.11) is a third order polynomial only present to cancel the first term in the differential equation, *i.e.* $\mathsf{D}\partial_x^2 P_3 = 2v(x)$, with roots at $x = 0$ and $x = L$. Therefore

$$\partial_t z = \mathsf{D}\partial_x^2 z + g\eta_\lambda(x, h(x,t)) \tag{6.12}$$

with homogenous BC's. The term $\eta_\lambda(x, h(x,t))) \equiv (1 + \lambda \frac{d^2}{dx^2})\eta(x, h(x,t))$ is actually a functional of h. The initial condition of $z(x,t)$ is not $z(x, t = 0) \equiv 0$ as for h, because of the data shift above. But due to the homogenous BC's any initial condition decays, so that the initial sources, accounting for $z(x, t = 0) = -P_3(x)$, can be ignored. Then the formal solution is

$$z(x,t) = \sum_{n=1}^{\infty} z_n(t) \sin(k_n x) \tag{6.13}$$

with

$$z_n(t) = \frac{2g}{L} \int_0^t dt' \int_0^L dx' \eta_\lambda(x', 2v(x')t' + z(x',t)) \sin(k_n x') \; \exp(-k_n^2 \mathsf{D}(t - t')) \tag{6.14}$$

and $k_n = \frac{\pi n}{L}$ (see also Chapter 9).

In Eq. (6.14) η should, according to (6.11), be

$$\eta_\lambda(x', 2v(x')t + P_3(x') + z(x',t)) \;, \tag{6.15}$$

but P_3 can be ignored. The reason is that we assume the statistics of the interface to be invariant under a tilt of the substrate. Even more general, the statistics of $\eta(x,h)$

is the same as the statistics of $\eta(x, h + f(x))$ with arbitrary function $f(x)$,

$$\langle \eta(x,h)\eta(x',h') \rangle = \langle \eta(x,h+f(x))\eta(x',h'+f(x')) \rangle \tag{6.16}$$

which for $f(x) = \alpha x$ corresponds to tilting the interface, which can of course neither be seen by the Laplacian nor by ∂_t in Eq. (6.9).

6.1.3.1 The noise correlator

One can make use of this invariance also in a slightly different sense: According to Eq. (6.11), the tilt of $h(x,t)$ in x increases in time. Assuming stationarity of the relevant statistical properties (especially avalanches as defined below), this requires the solution to be invariant under tilt, which is also known as Galilean invariance (Meakin, 1998): $h' = h + \alpha x$ must produce the same statistics as h, which entails $\eta(x, a + \alpha x)$ to be equally likely as $\eta(x, a)$, so that

$$\langle \eta(x, a+\alpha x)\eta(x', a'+\alpha x') \rangle = \langle \eta(x,a)\eta(x',a') \rangle \,. \tag{6.17}$$

But assuming the standard form (Nattermann *et al.*, 1992) $\langle \eta(x,a)\eta(x',a') \rangle = \Delta_\|(x - x')\Delta_\perp(a-a')$, the correlator obeys for any $x-x'$ where $\Delta_\|(x-x')$ is finite, $\Delta_\perp(a-a') = \Delta_\perp(a-a'+\alpha(x-x'))$. This holds for any α, so if $\Delta_\|(x-x')$ was finite for any $x-x' \neq 0$, $\Delta_\perp$ would be bound to be a constant. This is impossible, because $\Delta_\perp$ must be non-vanishing somewhere and normalisable, so that $\Delta_\|(x-x')$ must vanish for any finite $x - x'$, *i.e.* it *must* be a δ-function.

6.1.3.2 No-passing

Next it can be shown that the Oslo model obeys MIDDLETON's no-passing (Middleton, 1992). For $\lambda \neq 0$ this will lead to a constraint on the noise which is incompatible with the δ correlation of $\Delta_\|$ in the continuum, so that λ must vanish in the continuum. Defining a partial ordering $\succeq$ for two configurations $h_1(t_1, x)$ and $h_2(t_2, x)$ of the interfaces as $h_1(t_1,x) \succeq h_2(t_2,x) \Leftrightarrow \forall_{x\in[0,L]} h_1(t_1,x) \geq h_2(t_2,x)$, one has to show that this order is preserved under the dynamics (Sethna *et al.*, 1993). With the "external field" being the BC's $E_1(t)$ and $E_2(t)$, one shows that if $h_1(t_0,x) \succeq h_2(t_0,x)$ for a given t_0 (which entails $E_1(t_0) \geq E_2(t_0)$) the interfaces can never "overtake" each other at $t \geq t_0$. By assuming the opposite, one only needs to prove that where the two interfaces "touch" for the first time, x_0, the velocity of h_1 is higher or equal to the velocity of h_2. For the model on the lattice (6.4), this is equivalent to

$$h_1^+ + \eta_1^+ + h_1^- + \eta_1^- \geq h_2^+ + \eta_2^+ + h_2^- + \eta_2^- \tag{6.18}$$

using the same notation as in (6.4). In the original discrete model, condition (6.18) follows immediately from $\eta(x, h) + h$ being a monotonically increasing function in h for any x. For the continuum equation (6.9) the corresponding calculation gives

$$\lambda g \partial_h \eta(x, h) \geq -\mathsf{D} \tag{6.19}$$

assuming that $\frac{d^2}{dx^2}\eta = \partial_x^2\eta + \partial_x h \partial_x \partial_h \eta + \partial_x h \partial_h \partial_x \eta + \partial_x^2 h \partial_h \eta + (\partial_x h)^2 \partial_h^2 \eta$ and that the interface is smooth in x_0 such that $\partial_x h_1(x_0, t) = \partial_x h_2(x_0, t)$ and $\partial_x^2 h_1(x_0, t) > \partial_x^2 h_2(x_0, t)$. For a noise with divergent width, $\Delta_{\|}(x) = \delta(x)$, Eq. (6.19) cannot hold for any $\lambda \neq 0$, *i.e.* a non-vanishing λ destroys no-passing. However, no-passing must be regarded as a crucial feature, as it ensures the asymptotic uniqueness of the configuration and is reminiscent of the irrelevance of the order of updates in the original model (*i.e.* to be Abelian, see Sec. 5.3.3.1, page 235), so that $\lambda = 0$ is a necessary condition for the equivalence of the continuum and discrete model.

This is physically justified: Assuming a smooth η, in the continuum approximation of Eq. (6.4) λ becomes proportional to the square of the lattice spacing and therefore vanishes in the continuum limit.

Keeping the λ term nevertheless, a naïve scaling analysis shows that it is irrelevant. Moreover, its FOURIER transform in Eq. (6.14) produces only a term $-g\lambda k_n^2$, because of the total derivative in η_λ. This can be absorbed into the bare propagator of a perturbative expansion in the style of (Nattermann *et al.*, 1992; Leschhorn *et al.*, 1997) in the form

$$\frac{2g(1 - \lambda k_n^2)}{L(\mathsf{D}k_n^2 + i\omega)}, \tag{6.20}$$

leading possibly to an ultraviolet divergence. Apart from that, the terms obtained for an renormalisation group treatment are structurally the same as in (Leschhorn *et al.*, 1997) as calculations show [see sec. 6.2, but also (Narayan, 2000)]. The only differences are due to the peculiar way of driving the interface (*i.e.* the term $2v(x)$, which is a mean velocity in (6.11), but also drives the model by moving the quenched noise in (6.14)) and the non-conservative nature of the interface (which makes sense only for a finite system) leading to the homogenous BC's and therefore to the $\sin(k_n x)$ rather than $\exp(2ik_n x)$ terms. In turn, the standard qEW problem (Nattermann *et al.*, 1992) corresponds to an Oslo model with periodic BC's and continuous, uniform drive, see Sec. 6.3, page 273.

The definition of the avalanche size s in the continuum is the area between the interface configurations at two times t_1 and t_2,

$$s = \int_0^L dx(h(x, t_2) - h(x, t_1)) \tag{6.21}$$

so that $\langle s \rangle = v\Delta t L$ with $\Delta t \equiv t_2 - t_1$, because $\langle z(x, t) \rangle$ is expected to be asymptotically independent of t, as a non-vanishing $\lim_{t\to\infty} \partial_t \langle z(x,t)\rangle$ with homogenous BC's would require support for a divergent curvature of the interface, see Fig. 6.2. Choosing $\Delta h \equiv \Delta t v$ constant for different system sizes L then preserves the property $\langle s\rangle \propto L$.

Due to the asymptotic uniqueness of the solution the system can either be driven in jumps of Δh separated by sufficiently long times, or driven very slowly taking "snapshots" of the configuration in order to calculate s.

6.1.3.3 Timescales in the qEW

The model possesses two characteristic microscopic timescales: One is the diffusive timescale $t_0 \equiv L^2/\mathsf{D}$, the other one is the non-trivial scale due to noise and drive, $t_g \equiv g^2/(v^3 L)$. One has to maintain a sufficiently large Δt to prevent distinct avalanches from merging, otherwise the central limit theorem would probably turn $\mathcal{P}(s)$ into a Gaussian. The SOC limit is usually identified with $v \to 0$, which makes sense only in the presence of an intrinsic scale for v.[3] The only combination of parameters (D, g and L, but $\lambda = 0$) which provides a "natural velocity" is $v_g \equiv (g^2\mathsf{D})^{1/3}/L$. The SOC condition $v \to 0$ is therefore already met by $v \ll v_g \propto L^{-1}$, which is however, not sufficient. According to Ref. (Paczuski and Boettcher, 1996) $\Delta t \gg L^z$ with $z \approx 1.42$, so that $\Delta h =$ const. entails $v \ll L^{-z}$, which therefore seems to be the correct condition for SOC, even though the microscopic time step in (Paczuski and Boettcher, 1996) is defined as a parallel update, which is not *exactly* (6.4).

6.1.4 Relation between cutoff and roughness exponent

The relation between the cutoff exponent D and roughness exponent χ of the qEW equation advocated as Open problem

$$D = 1 + \chi \tag{6.22}$$

in (Paczuski and Boettcher, 1996) has actually not been derived exactly. However, at least on a handwaving level it is extremely convincing: Fig. 6.3 shows two consecutive configurations of $h(x,t)$, denoted as $h(x,t_1)$ and $h(x,t_2)$. The hatched area between these configurations is the avalanche size (see also (Paczuski *et al.*, 1996, Fig. 3), (Bak, 1996, Fig. 29, p. 133) and Fig. 8.4, page 334); the cutoff size of these areas is determined by the cutoff length of the vertical fluctuations times the length. The former is expected to scale like L^χ, leading to a cutoff scaling like $L^{1+\chi}$. Clearly this applies only if the two configurations are correlated. Without correlations between the interfaces, their vertical distance is fully characterised by Δh only.

[3] The reason is that $v \to 0$ means that v, which is dimensionful, needs to be sufficiently small — compared to what? If no scale is given, then a trivial redefinition of the unit of velocity will make it arbitrarily small. In fact, an expression like $v \to 0$ is meaningless; what is really needed is $v/v_0 \to 0$.

Even though this is not an unusual line of arguments [see for example the "Phenomenological scaling theory", (Hinrichsen, 2000, chapter 3.3)], it is not a derivation. What is actually needed is a relation between the scaling of the width of the interface and the scaling of the avalanche sizes, which is in principle given by (6.21); for the second moment that means

$$\langle s^2 \rangle = \int_0^L dx \int_0^L dx' \langle (h(x,t_2) - h(x,t_1))(h(x',t_2) - h(x',t_1)) \rangle . \tag{6.23}$$

The RHS should be given by the two-time correlator $\langle (h(x,t_1) - h(x',t_2))^2 \rangle$. However, contrary to the suggestive form, this correlator is not only a function of $t_1 - t_2$ and $x_1 - x_2$. This is not so important for the temporal dependence, which is fully captured by $t_1 - t_2$ in the stationary state. The spatial dependence, however, can never be fully captured by $x_1 - x_2$ in a system, which is not translational invariant. Moreover, since this "full correlator" is evaluated everywhere on $[0, L]$, there is no way to avoid finite size effects.

While one expects (Kardar, 1998)

$$\lim_{L\to\infty, t\to\infty} \langle (h(x,t_1) - h(x',t_2))^2 \rangle = (x - x')^{2\chi} \mathcal{G}_\chi \left(\frac{|t_1 - t_2|}{|x - x'|^z} \right) \tag{6.24}$$

with dynamical exponent z and scaling function $\mathcal{G}_\chi$, this is *not* necessarily the case, if the roughness exponent χ is defined, for example, by the scaling behaviour of the width. The roughness exponent in (6.24) might be different from the roughness exponent from the scaling of the width (Galluccio and Zhang, 1995), in fact, (6.24) is not even necessarily the case for any χ, even if the width scales properly (see, for example, sec. 9.2.1.1). Clearly, if finite size effects matter, the scaling form (6.24) does not hold.

A naïve calculation of (6.23) gives

$$\langle s^2 \rangle = \int_0^L dx \int_0^L dx' \left(|x - x'|^{2\chi} (\mathcal{G}_\chi \left(\frac{|t_1 - t_2|}{(x - x')^z} \right) - \mathcal{G}_\chi(0)) \right) , \tag{6.25}$$

which certainly does not suggest the scaling expected from the handwaving explanation presented above, namely

$$\langle s^2 \rangle \propto L^{D(3-\tau)} = L^{1+D} = L^{2+\chi}, \tag{6.26}$$

according to Eq. (6.1), Eq. (6.2) and Eq. (6.22). But it is also not simply $L^{2+2\chi}$ and therefore $D = 1 + 2\chi$, as suggested by

$$\int_0^L dx \int_0^L dx' |x - x'|^{2\chi} , \tag{6.27}$$

because of the scaling function in (6.25). It remains an open question, how to relate D and χ in an exact manner. Fortunately, the exactly solvable model discussed in Chapter 8 (see especially Sec. 8.4, page 338) gives correctly $D = 1 + \chi$. One direct test for the validity of this equation in the Oslo model, which still needs to be performed, is to calculate the roughness exponent of the number of charges $h(x, t)$ *directly* in the original model on the lattice. Open task

6.1.5 Interface picture

Because of the equivalence of the Oslo model and the qEW, one can depict the Oslo model directly as an interface advancing through a random background. Fig. 6.3 shows such an interface configuration; site $x = 0$ has the rôle of an external drive. It is moved in steps of 2, so that it necessarily charges the next site, $x = 1$, once each move. The advantage of this procedure over a direct charging of site $x = 1$ is that one has to charge $x = 1$, but leave it otherwise free to receive further charges from its right neighbour. Moreover, it never receives any further charges from its left neighbour, which is the fixed, left boundary. For that reason, it is very convenient to simply set site $x = 0$ to a specific value and let the Laplacian transport the charge to the right neighbour,

$$\partial_x^2 h(1, t) = h(0, t) - 2h(1, t) + h(2, t) \to \partial_x^2 h(1, t) + 2 \ . \tag{6.28}$$

Of course, the charge of the site $x = 0$ is not included in the avalanche size. An example for an avalanche size is shown as the hatched area in Fig. 6.3 (see also Fig. 8.4, page 334).

The right hand boundary condition is simply fixed at $h(L + 1, t) = 0$ so it never evolves.

In Fig. 6.3 the values of $\eta(x, h)$ are shown in the boxes. They are only shown where they have actually been generated by the simulation; for example $\eta(x = 0, h)$ is never evaluated, but also $\eta(x = 3, h = 5)$ shows just a blank. This is because the site had received a double charge from the neighbours, which transported it directly from $h = 4$ to $h = 6$.

It is not entirely straight-forward to map between h and the interface configuration.[4] Clearly, there is no unique mapping from z_i to $h(x, t)$, because of an (almost) arbitrary[5] offset in $h(x, t = 0)$. But there is a unique mapping from $h(x, t)$ and $\eta(x, h)$ to z_i and z_i^c, which is detailed below.

Prima facie it looks irritating that η changes at each charge (*i.e.* whenever a site

[4] We stress again that h here is the number of charges received, while it denotes the height in Chapter 5.

[5] It is not completely arbitrary, because one probably wants to preserve the property that even z_i corresponds to odd h and vice versa.

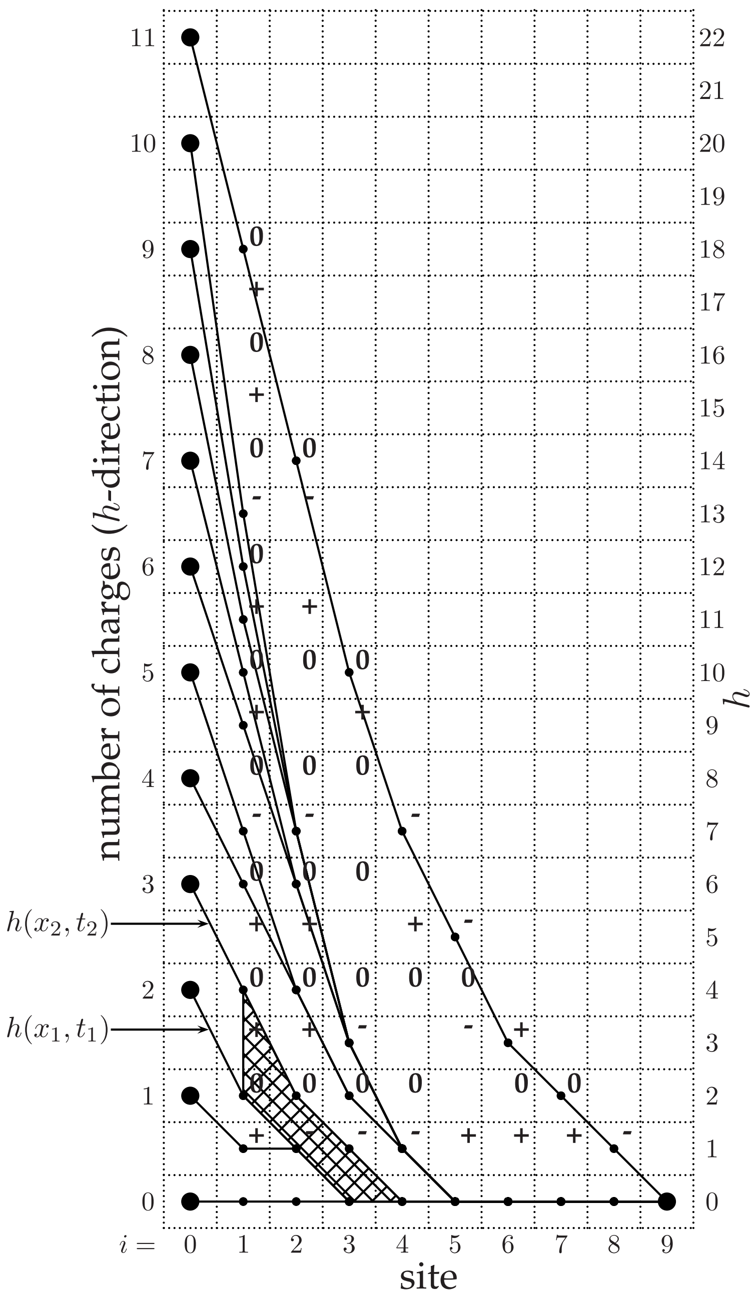


Figure 6.3: Eleven consecutive interface configurations. The avalanche size (measured as number of charges) is shown as a hatched area, which does not include the charge of site $i = 0$. The leftmost and the rightmost sites, which are fixed by boundary conditions, are shown a big black circles. The numbers on the ordinate on the left hand side indicate the number of charges and on the right hand side they indicate the interface height $h(x, t)$. The symbols in the boxes indicate the value of $\eta(x, h)$, where $+$ stands for $+1$ and $-$ for -1.

$h(x,t)$	$\eta(x,h(x,t))$	$\eta(x,h(x,t)+1)$	z_i	z_i^c	Remark
even	0	1	1	1	(topples at charge)
even	0	-1	1	2	
odd	1	0	0	$\eta(x,h(x,t)+2)$	(irrelevant, as z_i cannot topple)
odd	-1	0	2	2	(necessary, as z_i must topple)

Table 6.1: Mapping of h and η to z_i and z_i^c.

receives a unit from one of its neighbours), while z_i^c changes only at each toppling. But every single toppling is triggered by a charge, so that η changes at least as often as z_i^c. There is no trivial mapping between the value of z_i^c and $\eta(x,h(x,t))$, because the latter is a noise with respect to the field $h(x,t)$ mimicking the effect of the z_i^c.

It is important to keep in mind that z_i is supposed to represent *stable* configurations only. For the response to a single charge (any multiple charge can be decomposed into a sequence of single charges) it is then clear that the actual value of z_i^c plays a rôle only if $z_i = 1$ — in case that $z_i = 0$ a single charge cannot activate site i, if $z_i = 2$, then the site must topple for any value of z_i^c.

To derive the value of z_i^c for a given site at the time it receives a charge from its η value, one must consider $\eta(x,h(x,t)+1) - \eta(x,h(x,t))$. This quantity is either 1 or -1. Whenever it is positive, a toppling will occur if the site is charged, whenever it is negative z_i will increase and no toppling will take place. Together with the actual value of z_i [as derived from $h(x,t)$ and $\eta(x,h(x,t))$, see explanation after Eq. (6.3)] at the time the charge occurs, one arrives at the mapping shown in Tab. 6.1.

6.1.6 Boundary conditions

It is very instructive to reconsider the boundary conditions originally used in the Oslo model, see Sec. 5.1.1, page 205. There, the number of topplings of the rightmost site is, apart from the effect of z_L^c, just the number of charges, *i.e.*

$$H(x=L,t+1) = h(x=L,t) + \widetilde{\eta}(x,h(x=L,t)) \tag{6.29a}$$

$$h(x=L,t) = H(x=L-1,t) + H(x=L+1,t) \tag{6.29b}$$

where $H(x=L+1,t) \equiv 0$ is the "true" boundary condition and $\widetilde{\eta}(x,h(x=L,t)) \in \{0,-1\}$ where 0 corresponds to $z_L^c = 1$ and -1 to $z_L^c = 2$. The problem with this boundary condition is that it actually introduces a special equation of motion for the rightmost site in the bulk — it is not just a boundary condition.

One very interesting way to cast (6.29a) into the form (6.3) is to introduce $H(x = L+1,t) \equiv H(x=L,t)$, as if the rightmost boundary would "topple back" whatever it has received. From this point of view $x = L$ becomes a regular site in the bulk with a regular noise $\eta(x, h(x = L, t))$

$$\begin{aligned} H(x = L, t + 1) &= \frac{1}{2}\left[h(x = L, t) + \eta(x, h(x = L, t))\right] && \text{(6.30a)} \\ h(x = L, t) &= H(x = L - 1, t) + H(x = L + 1, t) \,. && \text{(6.30b)} \end{aligned}$$

It is worth noting that — correctly — this site can never be in the stable state $z_L = 0$, because it can reach this state only temporarily after toppling, but is automatically charged afterwards.

The new boundary condition $H(x = L + 1, t) \equiv H(x = L, t)$ in the continuum becomes a von-Neumann boundary condition,

$$\partial_x|_{x=L+1} H(x, t) = 0 \,. \tag{6.31}$$

To transform that into a boundary condition for h, one could in principle set $h(x = L + 1, t) = h(x = L, t)$ and $\eta(x = L + 1, h) = \eta(x = L, h)$, so that (6.30a) applies to $H(x = L + 1, t)$ as well. Thus

$$\begin{aligned} \partial_x|_{x=L+1} h(x, t) &= 0 && \text{(6.32a)} \\ \partial_x|_{x=L+1} \eta(x, t) &= 0 \,. && \text{(6.32b)} \end{aligned}$$

The big advantage of this boundary condition over the Dirichlet boundary condition (6.6) is that in the original boundary conditions all sites topple equally often on average (see Sec. 5.1.4, page 212). Therefore, $v(x)$, defined in (6.10), would become independent of x and would not enter η in (6.14) in such a complicated form. In fact, as it is demonstrated below, the current form of $v(x)$ causes serious problems. The usage of a simplified $v(x)$ is still to be explored.

Open issue

6.1.7 Direct translation into the height picture

I am indebted to OLE PETERS for pointing out that the height $\widetilde{h}(x, t)$ of the Oslo model in the height picture is given by

$$\widetilde{h}(x, t) = \widetilde{h}(x, t = 0) + h^*(x, t) - H(x, t) \tag{6.33}$$

where $h^*(x, t)$ is the number of charges received by a site in terms of height. This is a further subtlety: while a site can be charged in terms of slope by its right neighbour, this cannot happen in terms of height, because "height grains" only move to the right. Thus, $h^*(x, t) = H(x-1, t)$, and ignoring the standard problem of forward/backward derivatives, one has

$$\widetilde{h}(x, t) = \widetilde{h}(x, t = 0) - \partial_x H(x, t) \tag{6.34}$$

Moreover, the slope z_i (not to be confused with the FOURIER components in (6.14)) is just $\widetilde{h}(x,t) - \widetilde{h}(x+1,t)$, so that

$$z_i(t) = z_i(t=0) + \partial_x^2 H(i,t) \ . \tag{6.35}$$

Thus, height as well as slope have an immediate meaning in the field-theoretical description, even though at first glance they seem to be just the most natural quantities to describe the model as a sandpile. From this observation it was possible to derive the boundary conditions discussed in Sec. 6.1.6.

Unfortunately, it does not seem to be possible to write an equation of motion for the slope or the height in the same form as (6.9), because the noise cannot be quenched in a variable, which approaches a stationary state. Otherwise the evolution of the model would eventually become periodic.

6.2 Analytical Results

While initially a different outcome was expected, it turned out that the analysis of (6.9) follows exactly the results in (Leschhorn *et al.*, 1997). This has already been conjectured by ONUTTOM NARAYAN (Narayan, 2000). The main problem is the actual evaluation of the integrals, which seriously suffer from the fixed boundary conditions and the tilt. As is pointed out in sec. 6.1.6, this problem is possibly lifted by using the original boundary conditions (see Sec. 5.1.1, page 205).

The first step is to write down the propagator explicitly. Replacing η_λ back into (6.14) and using[6] $\eta(x=0,h) = \eta(x=L,h) \equiv 0$ one has

$$z_n(t) = \frac{2g}{L}\int_0^t dt' \int_0^L dx' (1-\lambda k_n^2)\eta(x', 2v(x')t' + z(x',t)) \sin(k_n x') \ \exp(-k_n^2 \mathsf{D}(t-t')) \tag{6.36}$$

One can hide the fact of an initial condition $z(x,t=0) \equiv 0$ by defining (Nattermann and Tang, 1992)

$$\eta(x,h) \equiv 0 \text{ for } h < 0 \tag{6.37}$$

so that the FOURIER transform[7] in t simply becomes

$$z_n(\omega) = \frac{2g(1-\lambda k_n^2)}{\sqrt{2\pi}(i\omega + \mathsf{D}k_n^2)} \int_{-\infty}^{\infty} dt' \int_0^L dx' \eta(x', 2v(x')t' + z(x',t)) \sin(k_n x') \, e^{-i\omega t'} \ . \tag{6.38}$$

[6]This is to translate $\frac{d^2}{dx^2}\eta$ to $k_n^2\eta$ by partial integration; this is not needed for the first derivative, because the sine vanishes on both ends.

[7]For reasons of readability we do not distinguish symbols for the actual function and its FOURIER transform, wherever possible.

It is worth stressing that it does not make much sense to take the thermodynamic limit at this point, since one expects the problem to be self-organised critical, *i.e.* one should see pure finite size scaling.

One can now expand η in powers of z_n, by writing a Taylor series for $\eta(x,h)$ in h and FOURIER transforming:

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dt'\eta(x',2vt'+z(x',t'))e^{-i\omega t'} =& \frac{1}{2v}\eta(x',\omega/(2v)) \\ &+\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dq\eta(x',q)qz(x',\omega-2vq)+\dots\end{aligned} \tag{6.39}$$

which can also be obtained using the expansion (Leschhorn *et al.*, 1997)

$$\eta(x,2vt+z(x,t)) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dq\eta(x,q)e^{iq(2vt+z(x,t))} \tag{6.40a}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dq\eta(x,q) \times\left(1+iqz(x,t)+\frac{1}{2}(iq)^2z(x,t)^2+\dots\right) . \tag{6.40b}$$

Up to second order one therefore arrives at

$$\begin{aligned}z_n(\omega) =& \frac{2g(1-\lambda k_n^2)}{L(\mathsf{D}k_n^2+i\omega)}\Bigg(\int_0^L dx'\eta\left(x',\frac{\omega}{2v(x')}\right)\frac{\sin(k_nx')}{2v(x')} \\ &+\int_0^L dx'\int_{-\infty}^{\infty} dq\sum_{m=1}^{\infty}\eta(x',q)\frac{iq\sin(k_nx')}{\sqrt{2\pi}}z_m(\omega-2v(x')q)\sin(k_mx')\Bigg) \\ &+\dots\end{aligned} \tag{6.41}$$

It is worth pointing out a couple of important differences between (6.41) and the calculations presented in (Nattermann *et al.*, 1992) and (Leschhorn *et al.*, 1997):

- Due to the boundary conditions the set of orthogonal functions used here is $\{\sin(k_nx)\}$ rather than $\{\exp(-ikx)\}$. This causes many technical problems when it comes to convolutions.
- As mentioned above, all calculations take place in a finite system and only make sense in a finite system. Thus, $\sum_m$ cannot be written as $\int_k$.
- The x-dependence of $v(x)$ causes massive problems, because $\int dx\sin(mx)\sin(nx)/x$ is not orthogonal. As suggested in Sec. 6.1.6, one could probably resolve this problem by introducing different boundary conditions.

- The propagator acquires an extra factor $(1-\lambda k_n^2)$, so that for large n and fixed ω the propagator converges to λ/D, which might cause problems when summing $\sum_k^\infty$. However, according to sec. 6.1.3.2, it seems reasonable to assume that the proper continuum theory is $\lambda = 0$.

6.2.1 Diagrammatic expansion

Eq. (6.41) can be extended to higher orders, but is much more conveniently expressed in a diagrammatic expansion. The rules are slightly more involved than in (Leschhorn *et al.*, 1997), because the integration over x in $z_n(\omega)$, for example Eq. (6.41), cannot be performed systematically due to the x-dependence of $v(x)$. This becomes more vivid, when rewriting the second integral in (6.41) as

$$\begin{aligned}
&\int_0^L dx' \int_{-\infty}^{\infty} dq \sum_{m=1}^{\infty} \eta(x',q) \frac{iq\sin(k_n x')}{\sqrt{2\pi}} z_m(\omega - 2v(x')q)\sin(k_m x') \\
= &\int_0^L dx' \int_{-\infty}^{\infty} \frac{dp}{2v(x')} \sum_{m=1}^{\infty} \eta\left(x', \frac{p}{2v(x')}\right) \frac{ip\sin(k_n x')}{2v(x')\sqrt{2\pi}} z_m(\omega - p)\sin(k_m x') \ .
\end{aligned} \tag{6.42}$$

In principle the integral

$$\int_0^L dx' \eta\left(x', \frac{p}{2v(x')}\right) \frac{\sin(k_n x')\sin(k_m x')}{(2v(x'))^2} \tag{6.43}$$

could now be performed. However, it is certainly not just $\delta_{m,n}\hat{\eta}_n(p)$, nor could the x-dependence of v be absorbed into a redefinition of the FOURIER transform, because in each order, higher (negative) powers of $2v(x')$ appear. This will become clearer in the diagrammatic expansion below. If $v(x)$ would not depend on x, the KRONECKER resulting from (6.43) would even remove the sum over m in the second integral in (6.41). This is exactly what happens in (Leschhorn *et al.*, 1997); it is therefore highly desirable to implement the new boundary conditions discussed in Sec. 6.1.6.

The rules for the diagrams are as follows:

1 A double arrow ($\Longrightarrow$) represents a term $z_n(\omega)$ with n and ω according to the label. All arrows can be (dummy-) labelled using the rules explained below, apart from one incoming arrow, which fixes the temporal and spatial variable n and ω respectively.

2 A single, solid arrow ($\longrightarrow$) represents the propagator, $\frac{2g(1-\lambda k_n^2)}{L(\mathsf{D}k_n^2+i\omega)}$, with n and ω according to the label.

3 A dashed line (- - - -) represents the noise in the form $\sin(k_n x)\eta(x, q/(2v(x)))/(2v(x))$, with q, n and x according to the label.

4 For each node ($\bullet$), integrate over all outgoing frequencies and all space variables. Multiply by $(iq/(2v(x)\sqrt{2\pi}))^s$, where s is the number of outgoing solid lines (double or single) and x and q are chosen according to the *single* noise connected to the node.

5 All outgoing arrows receive a sum $\sum_n \sin(k_n x)$ and an integral $\int dx$ with n and x according to their label.

Apart from the single ingoing arrow, which needs to be labelled, the diagrams are meaningful even without any other labels, because they all represent dummy variables. The labelling is done as follows:

a For each node provide dummy frequency variables, starting with the single node, whose incoming arrow is already labelled (usually ω). Leave one outgoing arrow unlabelled, such that the total incoming frequency equals the total outgoing frequency.

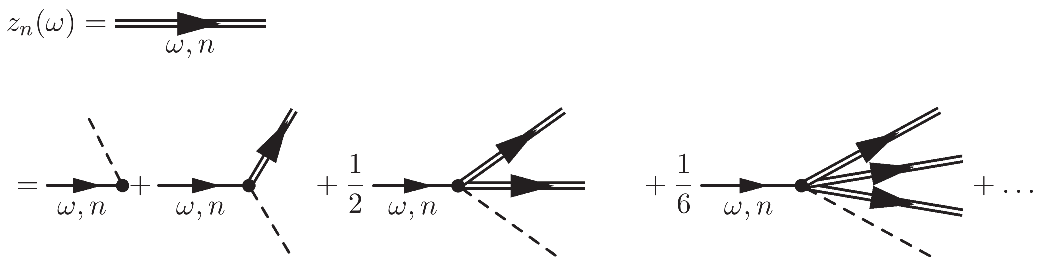
b The single noise term on each node obtains the same momentum as the incoming arrow (initially n).

c All outgoing legs but the noise obtain a new, individual (dummy) momentum variable $k, l, \ldots$.

d All outgoing legs obtain a new spatial variable (like x') to be integrated over at the node. It is the same for all legs on one node.

The labelling with two different spatial variables, n and x seems to be a bit clumsy, but is unavoidable, as pointed out around Eq. (6.42). In fact, this kind of labelling and the associated integrations will also work if it turns out that the boundary condition discussed in Sec. 6.1.6, which seem to make $v(x)$ a constant, are for some reason not applicable.

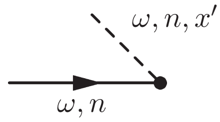
The expansion based on (6.40) is then



(6.44)

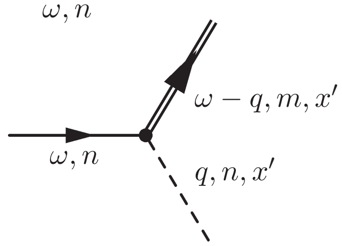
These diagrams are not dummy-labelled for better readability. The integrals associated with the first two terms have already been obtained in (6.41):

$$\overbrace{\underbrace{\frac{2g(1-\lambda k_n^2)}{L(\mathsf{D}k_n^2+i\omega)}}_{\omega,n}}^{\text{Rule 2}}\overbrace{\underbrace{\int_0^L dx'\eta\left(x',\frac{\omega}{2v(x')}\right)\frac{\sin(k_n x')}{2v(x')}}_{\omega,n,x'}}^{\text{Rule 3}} = \tag{6.45}$$



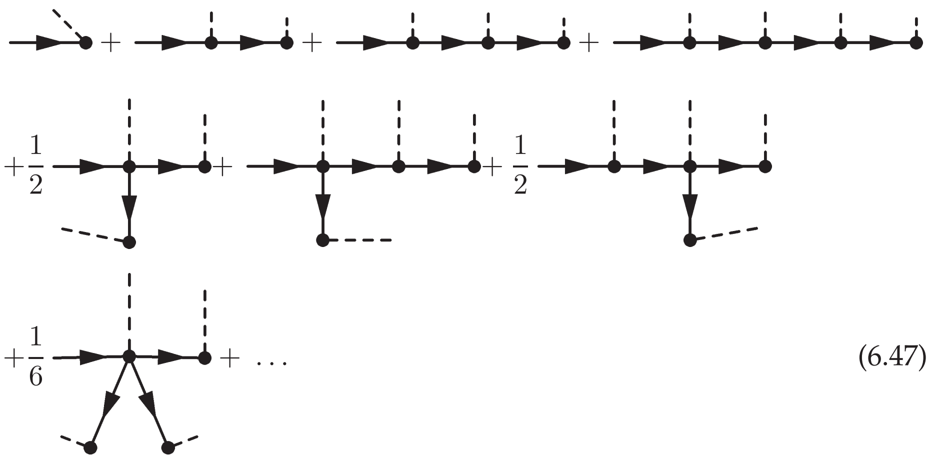
and

$$\overbrace{\underbrace{\frac{2g(1-\lambda k_n^2)}{L(\mathsf{D}k_n^2+i\omega)}}_{\omega,n}}^{\text{Rule 2}}\overbrace{\int_0^L dx'\sum_{m=1}^{\infty}\sin(k_m x')}^{\text{Rule 5}}\overbrace{\underbrace{\int_{-\infty}^{\infty}\frac{iq}{\sqrt{2\pi}2v(x')}dq}_{\bullet}}^{\text{Rule 4}}\overbrace{\underbrace{\eta\left(x',\frac{q}{2v(x')}\right)\frac{\sin(k_n x')}{2v(x')}}_{q,n,x'}}^{\text{Rule 3}}\overbrace{\underbrace{z_m(\omega-q)}_{\omega-q,m,x'}}^{\text{Rule 1}}$$

$=$ 

(6.46)

Based on the rules give above, one can now start the iteration procedure based on (6.44), by iteratively plugging in the first few orders for the double arrow. The resulting first few terms are

$z_n(\omega) \quad =$ 

(6.47)

When taking the average over the noise, the higher order correlators of the presumably Gaussian noise are obtained from WICK's theorem (Le Bellac, 1991; van Kampen, 1992); one can leave the labelling as discussed above and replace pairs of the form $\langle \eta\eta \rangle$ by their correlator, which may or may not be a δ-function.[8] Thus, taking the average of $z_n(\omega)$ simply means to connect all dashed legs in pairs, in all possible ways. If there is a leg left, this term does not contribute, because the noise enters in the form $\eta(x,\omega)$ with x and ω just being fixed arguments or even been integrated over, so that $\langle \eta(x,\omega) \rangle$ vanishes. This is of course different from $\langle \eta(x, 2v(x) + z(x,t)) \rangle$, which is the average noise experienced by the interface. In that expression $z(x,t)$ is supposedly a solution of (6.12) and is therefore correlated to η; one can see that immediately by writing $z(x,t)$ as a *functional* of η, $z(x,t;[\eta])$, as it is the solution of (6.9) for a *particular realisation* of η.

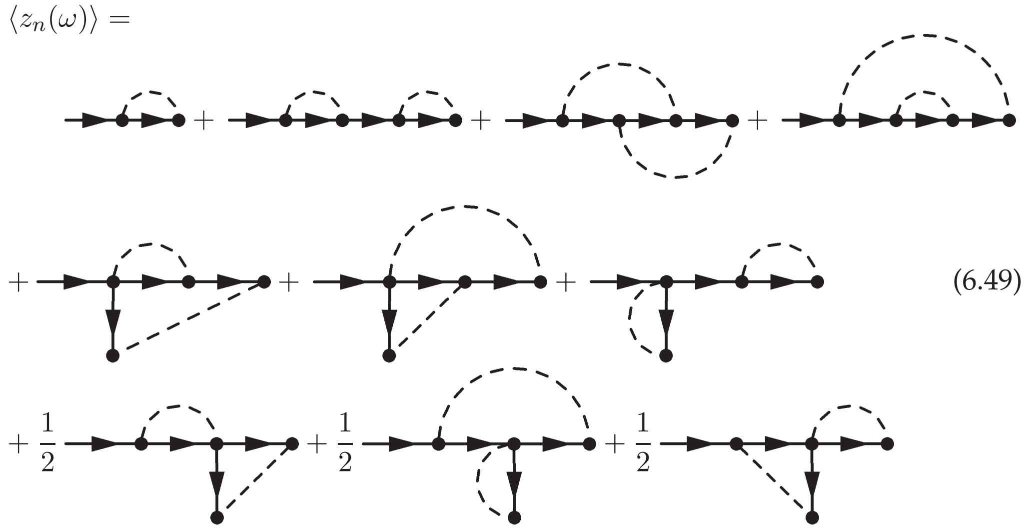
In order to test the continuum theory against the original model, it would be interesting to calculate the average number of charges as a function of the position. The average number of charges is simply $\langle h(x,t) \rangle$. Using (6.11) gives only a part of the answer

$$\langle h(x,t) \rangle = 2v(x)t + P_3(x) + \langle z(x,t) \rangle \ , \tag{6.48}$$

but the last term is non-trivial. However, it should converge for large t to a function only in x, otherwise the interface $z(x,t)$, fixed on both ends, would move and therefore bend forever. However, there is nothing to balance an arbitrary curvature, as mentioned already around (6.21). On the other hand, it is also clear that the interface is "dragged" through the random medium, so one expects a non-trivial form of $\langle z(x,t) \rangle$.

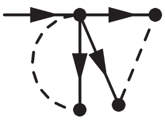
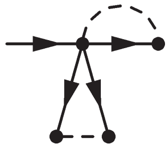
Thus, taking the average of (6.47) gives

$\langle z_n(\omega) \rangle =$



(6.49)

[8]Usually it is a δ-function in x but not in h, (Leschhorn *et al.*, 1997).

$+\frac{1}{3}$  $+\frac{1}{6}$  $+\ldots$

which are essentially the diagrams shown in (Leschhorn *et al.*, 1997, Fig. 14). Each diagram in (6.47) with s noise legs produces $s!/(2^{s/2}(s/2)!)$ diagrams with $s/2$ loops, provided that s is even. Each loop contains a term of the form $\langle\eta\eta\rangle$ and is therefore according to (6.12) preceeded by a factor g^2. This can also be seen from the fact that the propagator contains a factor g, and there is one node per propagator and each node contains exactly one noise term.[9] In other words, an expansion in powers of g^2 is exactly an expansion loop-by-loop.

6.2.1.1 An exact result

Unfortunately the diagrams are extremely hard to calculate, mainly because of the presence of $v(x)$. The only accessible term is

.

Assuming a noise correlator of the form

$$\big\langle\eta(x,h)\eta(x',h')\big\rangle = \delta(x-x')\Delta(h-h') \tag{6.50}$$

with unknown function $\Delta(h-h')$ means to waive the initial condition $z(x,t=0)\equiv 0$, see Eq. (6.37). Then, in the stationary state, one finds

$$\begin{aligned} = \quad & \delta(\omega)\frac{1}{\sqrt{2\pi}}\frac{2g(1-\lambda k_n^2)}{L(\mathsf{D}k_n^2+i\omega)} \\ & \times\int_0^L dx'\int_{-\infty}^{\infty}dq\sum_{m=1}^{\infty}\frac{2g(1-\lambda k_m^2)}{L(\mathsf{D}k_m^2+2iqv(x'))}iq\widetilde{\Delta}(q)\sin^2(k_m x')\sin(k_n x') \end{aligned} \tag{6.51}$$

where $\widetilde{\Delta}(q)$ is the FOURIER transform of $\Delta(h)$. The leading $\delta(\omega)$ is not surprising, because one expects a time-independent behaviour.

In order to continue in closed form, one has to take $\lambda=0$. Noting that

$$\sum_n k_n^2\langle z_n(\omega)\rangle\sin(k_n x) = -\big\langle z''(x,\omega)\big\rangle \tag{6.52}$$

[9] It certainly would make sense to absorb the g into η and so in its correlators.

where $z''(x,\omega) = \frac{d^2}{dx^2} z(x,\omega)$ and taking into account the completeness of $\sin(k_n x)$ then gives

$$\left\langle z''(x,t) \right\rangle = -\frac{g^2}{\mathsf{D}L\pi} \int_{-\infty}^{\infty} dq \sum_{m=1}^{\infty} \frac{iq\widetilde{\Delta}(q)}{\mathsf{D}k_m^2 + 2iqv(x)} \sin^2(k_m x) \tag{6.53}$$

and after integration

$$\left\langle z''(x,t) \right\rangle = -\frac{g^2}{\mathsf{D}L\pi} \sum_{m=1}^{\infty} \frac{\sqrt{2\pi}}{\mathsf{D}k_m^2} \sin^2(k_m x) \int_0^{\infty} du e^{-u/\xi_m(x)} \Delta''(u) \;, \tag{6.54}$$

where $\xi_m = 2v(x)/(\mathsf{D}k_m^2)$. It is very interesting to see that the second derivative of the correlator appears here, since this supposedly plays a crucial rôle in the quenched EDWARDS-WILKINSON equation (Leschhorn *et al.*, 1997; Wiese, 2002). The result is hardly more than a reassurance that one has not lost track completely; approximating

$$\int_0^{\infty} du e^{-u/\xi_m(x)} \Delta''(u) \approx \int_0^{\xi_m(x)} du \Delta''(u) = \Delta'(\xi_m(x)) - \Delta(0) < 0 \tag{6.55}$$

where the latter relation relies on the idea that the correlation function gets "flatter" for larger arguments, one can at least convince oneself that $\langle z''(x,t)\rangle > 0$, *i.e.* the interface is really "dragged behind".

6.2.1.2 Meaning of $\chi > 1$

Usually, $\chi > 1$ is associated with overhangs in $h(x,t)$ (Fisher, 1986; Krug and Spohn, 1991; Jensen, 1995; Barabási and Stanley, 1995). In the current context, however, $h(x,t)$ *must* be a single-valued function. The explanation for this inconsistency is intriguingly simple: the interface bends between the two fixed boundary, possibly with a finite average curvature as it is dragged over the surface. If χ refers to the width w^2, where

$$w^2 = \left\langle \overline{h(x,t)^2} \right\rangle - \left\langle \overline{h(x,t)}^2 \right\rangle \tag{6.56}$$

with $\overline{h}$ denoting the spatial average, then even $h(x) = x^2$ would have a width of $w^2 = (4/45)L^4$. One can now define the roughness via FAMILY-VICSEK scaling (Family and Vicsek, 1985),

$$w^2 = aL^{2\chi}\mathcal{G}\left(\frac{t}{bL^z}; \ldots\right) \tag{6.57}$$

where z is the dynamical exponent, $\mathcal{G}$ is the scaling function, a and b are metric factors and ... denotes further parameters.[10] Therefore the parabolic $h(x)$ simply gives $\chi = 2$, even without any fluctuations. So, $\chi < 2$ even indicates that on spatial and

[10]These further parameter are crucial to avoid trivialisation of the problem, Sec. 9.4.4, page 371.

ensemble average, the curvature is *not* finite. Of course this explanation depends fundamentally on the definition of χ, which might also be defined via the structure factor, a self-similarity assumption etc. This discussion is closely related to the problem of the relation between D and χ in Sec. 6.1.4, page 253.

While the natural explanation of $\chi > 1$ sounds quite reassuring, it actually is a reason for concern. The problem is that the qEW equation does not have such a natural explanation for $\chi > 1$, because it is dragged homogeneously, by a constant force, over the surface. So, if the Oslo model is a qEW, how is it possible that these two models have different mechanism responsible for $\chi > 1$? Open problem

6.2.2 Remarks on numerics

One of the most obvious tests of the theory above is to integrate (6.9) numerically and compare the results to the discrete model. That way, problems regarding the transition to a continuum theory become immediately obvious. The numerical study of (6.9) with boundary conditions and driving as discussed stands at its beginning. It would be very reassuring to obtain the same results for (6.9) as for the original Oslo model. However, it is already clear that the numerical study of the qEW equation is not straight forward; usually alternative models, which supposedly belong to the qEW universality class are simulated [see, for example, (Leschhorn, 1993; Amaral *et al.*, 1995)]. Thus, in the following only preliminary numerical studies are presented, all based on a direct integration of (6.9).

In fact, these studies indeed suggest that (6.9) with $\lambda = 0$ is a valid continuous description of the Oslo model: Fig. 6.4 compares a data collapse for different system sizes of the avalanche size distribution obtained from (6.9) in conjunction with (6.21) and from the original, discrete Oslo model. The best collapse is obtained with $\tau = 1.55$ for both models. As mentioned in sec. 6.1.4, the scaling law $D = 1 + \chi$ (6.22) (Paczuski and Boettcher, 1996) remains applicable as long as the two configurations at t_1 and t_2 are correlated. It is in perfect agreement with numerical results (Leschhorn, 1993; Le Doussal *et al.*, 2002) for the qEW model [11].

The data for the qEW equation have been obtained by direct integration of (6.9) on a lattice with discrete time steps, *i.e.* in a Euler-discretisation (Press *et al.*, 1992). The only continuous (up to numerical resolution) quantity in these simulations is $h(x,t)$. The discretisation of Eq. (6.9) introduces a number of new parameters in the simulations, which are not always easy to control.

The first step is to set $g = 0$ and to test, whether $h(x,t) = 2v(x)t + P_3(x)$ is obtained in the numerics. As mentioned above (Sec. 6.1.3, on page 250), corrections are expected for the initial behaviour, because the data shift violates $h(x,t=0) = 0$.

[11] Note for $\chi = 5/4$ one has $D = 9/4$ and $\tau = 14/9$.

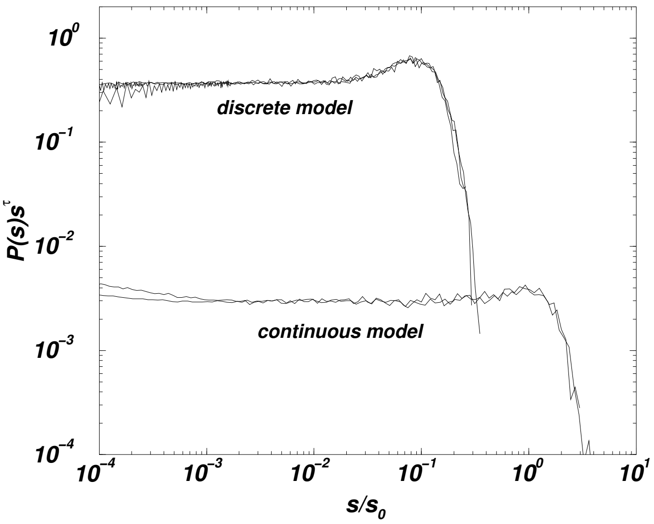


Figure 6.4: Comparison of a data collapse according to (6.1) for system sizes between $L = 128$ and $L = 512$ for the qEW equation (marked as "continuous") and the discrete Oslo model (marked as "discrete"). The same value of $\tau = 1.55$ collapses all curves within each model onto its scaling function. Due to the omission of the non-universal constants in Eq. (6.1) the two resulting curves are shifted relative to each other. See also Fig. 5.6, page 217.

This could of course be corrected by adding

$$\sum_{n=1}^{\infty} \frac{2}{L} \int_0^L dx' P_3(x') \sin(k_n x') e^{-\mathsf{D} k_n^2 t} \sin(k_n x) \tag{6.58}$$

which, however, vanishes in the stationary state, $\lim_{t\to\infty}$.

One problem is, for example, that the interface cannot be pulled arbitrarily fast. Between to driving steps, it takes, depending on the specific implementation of the lattice Laplacian, a finite number of time steps to "communicate" the new position of the boundary through the system. If pulled too fast, deviations from the expected analytical behaviour are observed, which get stronger over time.

In order to capture the effects of discretisation, one could in principle write

$$h(x, t+1) - h(x, t) = \sum_{i=1} \frac{1}{i!} (\partial_t)^i h(x, t) = (e^{\partial_t} - 1) h(x, t) \tag{6.59}$$

and

$$\partial_x^2 h(t) = \sum_{i=1} \frac{2}{(2i)!} (\partial_x)^{2i} = 2(1 - \cos(\partial_x)) \tag{6.60}$$

so that the PDE without noise the reads

$$(e^{\partial_t} - 1) h(x, t) = 2\mathsf{D}(1 - \cos(\partial_x)) h(x, t) \ . \tag{6.61}$$

As it is still linear, the solution does not change fundamentally; with the ansatz $h(x,t) = \sin(k_n x)\exp(\mu_n t)$ one finds

$$e^{\mu_n} - 1 = 2\mathsf{D}(\cos(k_n) - 1) \tag{6.62}$$

so that the set of orthogonal functions is simply

$$h(x,t) = \sin(k_n x)(2\mathsf{D}\cos(\lambda_n) + (1 - 2\mathsf{D}))^t \ . \tag{6.63}$$

Luckily it turns out that this more complicated, "proper continuum limit" has no numerical advantage over $\sin(k_n x)\exp(-\mathsf{D}k_n^2)$.

Conclusively, one should consider the direct numerical integration of the boundary driven qEW equation with homogenous boundary conditions as an open issue. Open issue

6.2.3 Remarks on the mean field theory

The random neighbour Oslo model should be consistent with the mean field (MF) theory of the qEW equation. Clearly, the random neighbour version of the Oslo model gives $\tau = 3/2$, namely exactly the return-time exponent of a random walker (see Sec. 3.3, page 113). On the other hand, the MF exponent of the qEW equation is $\chi = 0$ with upper critical dimension[12] $d_c = 4$ (see below), which seems to suggest $D = 1$ by Eq. (6.22), contrary to $\tau = 3/2$ which seems to suggest $D = 2$ by the scaling law $D(2-\tau) = 1$ (6.2). Why are these two exponents inconsistent?

The answer is that neither $D = 1 + \chi$ nor the scaling law $D(2-\tau) = 1$ necessarily hold in higher dimensions, where the mean field theory becomes actually exact. Both scaling laws depend crucially on the boundary conditions and the exact implementation of the higher dimensional models. All this is much more comprehensible in the randomly, bulk-driven MANNA model, which is well-defined in any spatial dimension d. In that case, it is probably reasonable to assume that the average avalanche size scales like L^2 in any dimension, because the particles can be thought of as independent random walkers (see Sec. 7.4, page 299). For $\tau = 3/2$ that entails $D = 4$ from $D(2-\tau) = 2$. There is no reason to expect that scaling law to break down in any dimension and there is also no reason to expect that τ changes to a different value at higher dimensions, once it has reached it random neighbour value. Thus $D = 4$ is expected to remain valid for spatial dimensions $d \geq 4$.

Moreover, for the isotropic MANNA model, one should expect $D = d + \chi$, which gives $D = 4$ *at* the upper critical dimension. So, *at* $d = 4$ the MF theory of the interface

[12] Here and in the following (spatial) dimension will refer to the dimension of the substrate above which the interface grows. This differs from the notation of, for example, NARAYAN and FISHER (Narayan and Fisher, 1993), but underlines the fact that the interface field $h(x,t)$ does not necessarily have to be a spatial distance.

picture is fully consistent with the random neighbour approach from the sandpile picture. However, in order to remain $D = 4$ for $d = 4$ and larger, the scaling law $D = d + \chi$ has to break down. What mechanism is responsible for this breakdown? Does the hyperplane where the interface is lifted get fractal?

Open problem

6.2.4 Dimensional analysis of the qEW

In this section, dimensional analysis (for more details about this method, see Sec. 9.4, page 363) is carried out on the qEW. Surprisingly[13], dimensional analysis produces the roughness exponent correctly in leading order of $\epsilon = d_c - d$ where $d_c = 4$. This leading order correction has once been claimed to be the only correction in ϵ, *i.e.* there are no higher order corrections (Narayan and Fisher, 1993), which has been disproved meanwhile (Le Doussal *et al.*, 2002, 2003).

The solution of the LANGEVIN equation,

$$\partial_t z(\mathbf{x}, t) = \mathsf{D}\nabla^2 z(\mathbf{x}, t) + g\eta(\mathbf{x}, z(\mathbf{x}, t)) \tag{6.64}$$

with a noise term obeying

$$\langle \eta(\mathbf{x}, h)\eta(\mathbf{x}', h') \rangle = \delta^d(\mathbf{x} - \mathbf{x}')\delta(h - h') \tag{6.65}$$

where $\delta^d(\mathbf{x} - \mathbf{x}')$ denotes the d-dimensional DIRAC δ-function, can be expressed as

$$z(\mathbf{x}, t; \mathsf{D}, L, g) = L^{\frac{4-d}{3}} \left(\frac{g}{\mathsf{D}}\right)^{\frac{2}{3}} \varphi\left(\frac{\mathbf{x}}{L}, \frac{t}{L^2/\mathsf{D}}\right) , \tag{6.66}$$

where φ is dimensionless. Eq. (6.64) is a generalisation and simplification of (6.12) in higher spatial dimensions d. The correlator (6.65) has actually already been ruled out as a possible choice in the qEW [see Sec. 6.1.3.1, page 251, as well as (Koplik and Levine, 1985; Nattermann *et al.*, 1992)]. But being the simplest possible choice, it is certainly a good starting point for a dimensional analysis.

It is crucial for the further interpretation of (6.66) that *all* free parameters are listed on the LHS (see Sec. 9.4.6, page 375). Otherwise, there are also additional parameters on the RHS. Integrating out $\mathbf{x}$ in order to determine the width according to the definition (6.56), and evaluating in the stationary limit gives therefore a function

$$w^2(\mathsf{D}, L, g) = L^{2\frac{4-d}{3}} \left(\frac{g}{\mathsf{D}}\right)^{\frac{4}{3}} \mathcal{C} \tag{6.67}$$

with a constant $\mathcal{C}$. This constant might under special circumstances diverge or vanish. It is worth noting that with such a small set of parameters there *are no corrections to*

[13] Remarkably, this is not discussed in the literature.

(6.67) possible.[14] Therefore the roughness exponent is

$$\chi = \frac{4-d}{3} \tag{6.68}$$

and the upper critical dimension is identified as $d_c = 4$, above which the roughness exponent becomes negative and an ultraviolet divergence appears (Krug, 1997).

For the original qEW, Eq. (6.69) (see below), one can also argue (Leschhorn *et al.*, 1997) that above $d = 4$ the curvature contribution in (6.64) is always stronger than the random force contribution, averaged over a given area. Thus, the interface is flat and the force exerted by the random background on the interface scales only like $L^{d/2}$ compared to L^d of the force F. Thus, above $d = 4$, in the thermodynamic limit even the weakest external force always overcomes the pinning force. However, below $d = 4$, the interface is flat only on scales below the Larkin-length (Bruinsma and Aeppli, 1984) and adapts to the random background force on larger scales, *i.e.* it gets pinned.

Remarkably, the simple "calculation" (6.67) gives exactly the same result as the single loop, very involved calculation by NATTERMANN *et al.* (Nattermann *et al.*, 1992) and NARAYAN and FISHER (Narayan and Fisher, 1993). As mentioned above, the latter authors even claimed that there are no further corrections to this result in higher loops.

However, there are at least two reasons, why (6.67) and therefore (6.68) is actually wrong.

6.2.4.1 Noise correlator

The noise correlator chosen in (6.65) is a valid choice within the substrate (*i.e.* the $\delta^d(\mathbf{x} - \mathbf{x}')$ term, see Sec. 6.1.3.1, page 251, and Eq. (6.50)), but not in the h-direction. Above (page 249), it has already been stated that η is correlated in the h direction because of the sawtooth-like noise required in the discrete model. In fact, the entire h-correlator remains visible even in an RG approach (Nattermann *et al.*, 1992; Leschhorn *et al.*, 1997; Le Doussal *et al.*, 2002) and it never simplifies to a δ-function. Moreover, the pinning force on the interface scales like $\Delta(0)^{2/(4-d)}$ (Nattermann *et al.*, 1992), so that for the "non-SOC" approach to the qEW, a δ-function for the noise correlator in the h-direction is not a valid choice.

Such a divergent force does not seem to pose a direct physical problem to the SOC-approach[15], because in SOC, the driving velocity v is fixed. Yet, the driving site must somehow transmit the force to the other sites using the curvature of the inter-

[14] For a similar problem see Sec. 9.4.4, page 371.

[15] "SOC-approach" to the qEW equation here means the boundary driven Eq. (6.12) with $\lambda = 0$ as opposed to Eq. (6.69) with explicit external driving force F.

face for pulling. Thus, the curvature would have to diverge as well, and therefore it seems that a δ-correlated noise is unphysical even in the SOC approach.

Moreover, there is an additional twist in having a δ-correlated noise: Dimensional reduction (Efetov and Larkin, 1977) predicts that a disordered system in d dimensions at zero temperature can be mapped to a system without disorder in $d-2$ dimensions at finite temperature. Thus, dimensional reduction immediately predicts $\chi = (4-d)/2$, namely the roughness exponent of the thermal EDWARDS-WILKINSON equation at $d' = d-2$. So, does dimensional analysis contradict dimensional reduction? It does not seem so: The proof of dimensional reduction relies on the uniqueness of the ground state and the analyticity of the noise correlator (Wiese, 2002). This is the reason why it does not apply to the standard (original) qEW problem, (6.69), where the correlator apparently renormalises to a non-analytical function. The same can be assumed for the SOC-approach to the qEW equation. Similarly it should not apply to a δ-correlated noise, but whether this is the only reason needs to be cleared up.

Open question

6.2.4.2 Additional scales

One of the key assumptions for the validity of Eq. (6.66), (6.67) and therefore (6.68) is that all variables parametrising the problem are listed on the LHS (see Sec. 9.4, page 363). However, in the original qEW equation has an extra term, namely the uniform force F over the entire interface,

$$\partial_t z(\mathbf{x}, t) = \mathrm{D}\nabla^2 z(\mathbf{x}, t) + g\eta(\mathbf{x}, z(\mathbf{x}, t)) + F \ . \tag{6.69}$$

Without any driving, the interface would remain stuck and (6.67) would apply, but $\mathcal{C}$ would vanish. The presence of F, however, renders (6.67) invalid.

There is a related parameter in the SOC version of the qEW equation, namely the driving velocity v, see Eq. (6.8). *Prima facie*, it seems this parameter resolves the riddle in case of the SOC model as well. Clearly, it also must not vanish, in order to make the interface move and make it develop a non-vanishing roughness. However, while F needs to be larger than a particular critical force F_c, there is no such threshold in the SOC version. A rough interface is obtained[16] for any $v > 0$ and in the stationary regime, *i.e.* after taking $\lim_{t\to\infty}$, one can even take the limit $v \to 0$ without changing the roughness. Thus, in the SOC version, there is *no additional scale from the driving*.

However, there is a source for additional scales even in the SOC version: As soon as the noise-correlator is non-trivial, it contains scales as well. For example, $\Delta(0)$, which is certainly finite, could be such an additional scale. So, would the result (6.68) be really be obtained for the correlator (6.65)? One way to get closer to an

Open problem

Open task

[16] Of course, the same arguments regarding flatness above $d_c = 4$ apply.

answer might be to apply dimensional analysis to the *discrete, exact* equation of motion, Eq. (6.4), page 248, in a sensible way.

6.3 Supplement: The Periodic Oslo Model

The concepts above can be applied to the periodic Oslo model[17] as well. In this model the boundaries are periodically closed, which removes all dissipation mechanisms from the model. That way, the Oslo model becomes an absorbing state model. Apart from that, the rules are exactly the same as for the original Oslo model with simplified boundary conditions, except for h_{L+1} being identical to h_1 and h_0 being identical to h_L. Just like the original model, the periodic model is driven on site $i = 1$, even though a random drive is maybe more natural. In the continuum limit, the LANGEVIN equation for the point-driven model is then

$$\partial_t h = \mathsf{D}\partial_x^2 h + (1 + \lambda \frac{d^2}{dx^2})\eta + \delta(x)vt \tag{6.70}$$

with the external drive

$$E(t) = vt \; . \tag{6.71}$$

This equation is much closer to the original quenched EDWARDS-WILKINSON equation, where $\delta(x)vt$ is replaced by a homogeneous force F. Just like above, it is certainly helpful to remove the background, *i.e.* to subtract the solution of the PDE without the quenched noise

$$\partial_t h_0 = \mathsf{D}\partial_x^2 h_0 + \delta(x)vt \; . \tag{6.72}$$

Using the eigenfunction $\exp(-ik_n x - \mathsf{D}k_n^2 t)$ with $k_n = 2\pi n/L$ one finds for $x \in [-L, L]$, which covers *two periods*,

$$h_0(x,t) = \frac{vt}{2L\mathsf{D}}\Pi_2(x) + \frac{1}{4!}\frac{v}{L\mathsf{D}^2}\Pi_4(x) + \frac{1}{2}\frac{v}{L}t^2 \tag{6.73}$$

with

$$\Pi_2(x) = x^2 - L|x| + \frac{1}{3!}L^2 \tag{6.74a}$$

$$\Pi_4(x) = x^4 - 2Lx^2|x| + L^2x^2 - \frac{1}{30}L^4 \; . \tag{6.74b}$$

Just as above (Sec. 6.1.3, on page 250), one can ignore the initial condition $h_0(x,t) = 0$ in the long time limit, since the modes decay as fast as $\exp(-\mathsf{D}k_n^2 t)$. Only $n = 0$

[17] This model has been studied together with KIM CHRISTENSEN, NICHOLAS MOLONEY and OLE PETERS.

requires special attention and produces the term $vt^2/(2L)$ in (6.73). After a data shift

$$h(x,t) = h_0(x,t) + z(x,t) \tag{6.75}$$

the remaining problem is

$$\partial_t z = \mathsf{D}\partial_x^2 z + \eta(x, h(x,t)) \; . \tag{6.76}$$

The noise term, however, is in contrast to (6.15) very difficult to handle, as it actually reads

$$\eta(x, h(x,t)) = \eta(x, \frac{vt}{L\kappa}\Pi_2(x) + \frac{v}{L\kappa^2}\Pi_4(x) + \frac{1}{2}\frac{v}{L}t^2 + z(x,t)) \tag{6.77}$$

which is equivalent to

$$\eta(x, \frac{vt}{L\kappa}\Pi_2(x) + \frac{1}{2}\frac{v}{L}t^2 + z(x,t)) \quad . \tag{6.78}$$

Open issue

The t^2-term makes it very hard to find a closed expressions for the FOURIER transform of the noise. It remains an open issue how to proceed.

6.3.1 Dropping sources, dropping particles

The t^2-term points to a confusing inconsistency between the sandpile model and the continuum PDE of it. In the sandpile model the driving is performed by dropping slope units on the pile, while in the PDE representation this process becomes an ever increasing source term, $\delta(x)vt$; it seems to be much more natural to write $\delta(x)v$ representing a localised source at $x = 0$ driving the system with velocity v.

However, what *is* governed by the PDE are not slop-units but their charges. The number of hits received a the driven site $x = 0$ from the external drive slowly increases. In the discrete, exact representation, one writes (see (6.4))

$$h(x,t+1) = \frac{1}{2}\left(h^- + h^+ + \eta^+ + \eta^-\right) + \delta_{x,1}E(t) \; , \tag{6.79}$$

with periodic boundary conditions $h(x=0,t) \equiv h(x=L,t)$ and $h(x=L+1,t) \equiv h(x=1,t)$. The external drive is represented by $E(t)$, the number of charges additionally having arrived at site $x=1$. Contrary to (6.4), the equation above has is not a rate equation, *i.e.* it has no time-derivative on the LHS. Correspondingly, $E(t)$ is not a rate either, it is a total amount. Only by subtracting $h(x,t)$ the LHS becomes a (discrete) time-derivative and the *conservative* Laplacian forms on the RHS, leaving $\delta_{x,1}E(t)$ and the noise terms unchanged, which now describe an increase per unit time, i.e they are source terms.[18] This apparent change in character becomes less con-

[18] In other words: $E(t)$ can take on its rôle as a source term only because $h(x,t)$ can be absorbed into

fusing if the LHS and therefore also the source terms on the RHS are interpreted as velocity fields.

As $E(t) = vt$ in the diffusion equation it has to be interpreted as a slowly increasing external source. A source, the strength of which increases linear in time, gives rise to a total volume quadratic in time — and this is where the t^2 in Eq. (6.73) is from.

In fact, if the driving $E(t)$ was just a constant, that would again lead to standard depinning. But that would *not* capture the original idea of the Oslo model as an AS model: The system is kicked at a slow, constant rate and it relaxes between kicks into a unique state. In order to study the approach to the depinned phase a constant E is useless. The increase in E and therefore the t^2-term are crucial features of the model.

6.3.2 Numerical results

The periodic Oslo model has been studied numerically fairly extensively as an AS model by Kim Christensen, Nicholas Moloney, Ole Peters and the author. The idea was to measure the avalanche size distribution as a function of the density of slope-units ρ in the system. If the AS to SOC correspondence as discussed in Sec. 1.4, page 57, indeed applies, then the distribution should show proper finite size scaling at a particular value of ρ. If it obeys simple scaling, then one expects

$$\langle s^n \rangle = a(bL^D)^{1+n-\tau} g_n , \tag{6.80}$$

as discussed around Eq. (2.74), page 100. Thus, for a particular choice of D and τ, the value

$$\frac{\langle s^n \rangle}{L^{D(1+n-\tau)}} \tag{6.81}$$

is the same for different system sizes. One can therefore simply plot this ratio for different sizes L versus the density ρ and search for a particular value of ρ, where all curves cross. Preliminary results suggest that such a point exist and that the value of D is surprisingly close to $D = 2.25$.

6.4 Supplement: Alternative Interface Approach

An alternative, very pragmatic path[19] to understand the relation between sandpiles and interfaces has been proposed by Mikko Alava and Bækgaard Lauritsen (Alava and Lauritsen, 2001). Effectively they write down an equation of motion for an appropriate observable of a sandpile-like model, for instance [see Eqn. (5) in (Alava

a conservative Laplacian.

[19] One might call it "reverse engineering".

and Lauritsen, 2001)]

$$\frac{\Delta H}{\Delta t} = \mathsf{D}\nabla^2 H + \widetilde{\eta}(x, H) + F(x, t) \tag{6.82}$$

where $H(x, t)$ counts the number of topplings at site x and time t. The term $F(x, t)$ represents the external drive[20] while $\widetilde{\eta}$ is a specially prepared quenched noise; the idea is to run a model according to the rules it is defined by, obtain the spatio-temporal evolution of the observable $H(x, t)$, and calculate from that which values the noise $\widetilde{\eta}$ had to assume so that Eq. (6.82) is obeyed.

There is no proof that such a $\widetilde{\eta}(x, H)$ exists for every model, let alone whether it is unique, but it has been obtained and studied numerically, for example for the BTW model and the MANNA model. That way, the study of different sandpile-like models becomes merely a study of (6.82) with different ensembles of noise, $\{\widetilde{\eta}\}$. The question is then: What can we know and learn from the properties of these ensembles?

6.5 Summary and Conclusion

The established relationship between Oslo model and qEW equation is presently pursued in order to develop a direct approach to the critical exponent τ, clear up the rôle of the noise and clarify the relation between noise and drive. The framework used here is also promising for other models, such as the BTW model (Bak *et al.*, 1987), various other sandpile models (Vespignani *et al.*, 2000; Malthe-Sørenssen, 1999) and the ZHANG model (Zhang, 1989).

It was conjectured much earlier (Paczuski and Boettcher, 1996) and widely anticipated [for a review on the "good arguments" for it, see (Alava, 2003)] that the Oslo model is a quenched EDWARDS-WILKINSON equation, mainly based on numerics and some handwaving arguments. With the work presented above, the analogy of interfaces and sandpiles (Paczuski *et al.*, 1996) gains further support. However, there are some open questions which must be addressed in the future.

One of the most urgent tasks is to find an exact relation between τ and χ, or rather the necessary conditions for such a relation (see Sec. 6.1.4, page 253). This is closely related to the issue mentioned in Sec. 6.2.1.2, namely that there are two different mechanism operating in the Oslo model and the qEW which cause $\chi > 1$. Moreover, the relation between the result of dimensional analysis and dimensional reduction (Sec. 6.2.4) needs to be cleared up.

Even though the focus of the study of the Oslo model as a qEW is not the numerics, it would be very worthwhile to clear up the various possibilities to treat

[20] This is a so-called columnar noise, *i.e.* a noise which is constant along the h-direction

the equation numerically, this, however, might require more sophisticated numerical techniques.

Moreover, it would be interesting to extend the study to the Abelian version of the MANNA model (see Sec. 7.3.3.1, page 294). However, in this model the quenched noise comes only into play for the *direction* in which particles are redistributed. However, as shown in Chapter 7, it is actually reasonable to discount this randomness, which would amount to discounting all quenched noise in the MANNA model altogether. Since the MANNA model is widely accepted to be described by REGGEON Field Theory (Alava, 2003; Vespignani *et al.*, 2000) just like DP, an exact mapping to the qEW equation would provide an interesting link between sandpile-like models, the quenched EDWARDS-WILKINSON equation and the field theories of DP.

6.5.1 Summary

In conclusion, the Oslo model has been reduced to a quenched EDWARDS-WILKINSON equation. In the continuum limit the qEW becomes the *exact* equation of motion for the Oslo model. This not only makes possible an analytical approach to exponents of an SOC-model, but also gives insight into the nature of avalanche-like behaviour and the relation between SOC and other theories of critical phenomena. It provides the perfect test bed for analytical methods proposed for SOC.

It is worth pointing out that the qEW equation originally belongs to the realm of classical critical phenomena with a parameter F (force) to be tuned to a certain critical value F_c, see Eq. (6.69). By mapping the Oslo model to the qEW equation, this equation has thereby been placed in the realm of SOC. This *does* violate the economic principle of computational physics, see Sec. 1.4.4.2, page 69.

The sections in this chapter can be summarised as follows

- In Sec. 6.1, the relevant variables are identified, for which an equation of motion can be expressed in closed form without the use of a HEAVISIDE θ-function. This equation is exact on the lattice and has a well justified continuum form. The resulting equation of motion is in fact a quenched EDWARDS-WILKINSON equation, the roughness exponent of which can be directly related to the critical exponents of the Oslo model. The values of these exponents when obtained numerically are perfectly consistent.

- The original quenched EDWARDS-WILKINSON (qEW) equation describes a depinning transition of an interface at a certain critical force. The Oslo model does not posses such a particular critical value of any parameter, since it is a model of SOC. It is therefore interesting to compare details of the analytical treatment of the qEW for the Oslo model, which is purely driven via the boundary conditions, to the results for the qEW equation of an interface found in the literature.

Results for this are presented in Sec. 6.2.

- The same methods can be applied to the periodic Oslo model, which represents an absorbing state model (Sec. 6.3). Further analytical progress is hindered by quadratic terms.
- Finally, Sec. 6.4 gives a very pragmatic approach to mapping sandpile-like models to the qEW. The task boils down to identifying an ensemble of proper choices for the noise, which is, in a sense, an exercise in “reverse engineering”.

Chapter 7

Universality, Anisotropy and Crossover in the Oslo Model

This chapter focuses on universal features of the Oslo model. It is shown that any amount of anisotropy moves the Oslo model to another known *universality class, the exponents of which can be derived exactly. This amounts to an exact solution of the quenched* EDWARDS-WILKINSON *equation with a drift term. Based on this equation or, alternatively, a random walker approach, the toppling frequency and the average avalanche size are calculated exactly. Moreover, the crossover behaviour is briefly analysed. It is argued that anisotropy is likely to be experimentally relevant and may explain why consistent exponents have not been extracted in real rice pile experiments.*

7.1 Oslo Model with Anisotropy

The idea to introduce explicit anisotropy into the Oslo model originated from a discussion with DEEPAK DHAR about the question, whether the Oslo model is intrinsically anisotropic, as grains are transported through the system. There are good reasons to regard the anisotropy in the Oslo model as spurious (see Sec. 5.1.3.1, page 211), so it is reasonable to study the effect of explicit anisotropy.

The general interest and relevance of studies in SOC rely on the assumption, guided by equilibrium critical phenomena, that the critical behaviour of scale-invariant systems falls into universality classes determined solely by a few general characteristics of the system, such as symmetry and dimension. So-called "relevant" parameters can decide which of the symmetries the system is asymptotically dominated by.

The rôle of anisotropy in SOC has been highlighted very early by HWA and KARDAR (Hwa and Kardar, 1989) and GRINSTEIN *et al.* (Grinstein *et al.*, 1990), who used anisotropic LANGEVIN equations to describe sandpiles (see Sec. 3.1, page 105). On the cellular automata level, KADANOFF *et al.* (Kadanoff *et al.*, 1989) have conjectured that the net flux of particles is a relevant parameter. This chapter confirms this conjecture for the Oslo model, which shows a clear-cut and consistent relevant dependence on anisotropy. This is of great importance for the interpretation of experimental results (Frette *et al.*, 1996) and more generally for the much studied quenched EDWARDS-WILKINSON (EW) equation (Nattermann *et al.*, 1992; Pruessner, 2003c), see Chapter 6. Moreover, contrary to suggestions in former studies, the switch between different universality classes (crossover) is *not* triggered by the introduction of stochasticity (Dhar and Ramaswamy, 1989; Tadić and Dhar, 1997; Pastor-Satorras and Vespignani, 2000b; Kloster *et al.*, 2001) nor by multiple topplings (Vazquez, 2000; Pastor-Satorras and Vespignani, 2000b; Paczuski and Bassler, 2000; Priezzhev *et al.*, 2001).

Similar to TSUCHIYA AND KATORI (1999B,A), the system size at crossover, L_{X}, depends on the strength of the anisotropy[1] v. Two possible mechanisms causing anisotropy in experiments are exemplified, one of which vanishes with the system size L fast enough to keep $\mathsf{v}L$ constant. This represents a marginal case and consequently makes a unique identification of the critical exponents impossible.

7.1.1 Definition of the models

The original Oslo model with simplified boundary condition (see Sec. 5.1.3, page 209, this model is called "OOM" henceforth) consists, in one dimension, of a lattice of sites $i = 1, \dots, L$. Two coupled dynamical variables are associated with each lattice site: the primary variable $z_i \in \{0, 1, 2, \dots\}$ and the threshold variable $z_i^c \in \{1, 2\}$. The initial configuration consists of $z_i = 0\ \forall i$ and a random configuration of the z_i^c. The system is driven by increasing z_1 by one (a "slope unit") followed by a relaxation of all sites $1 \leq i \leq L$ for which $z_i > z_i^c$ ("over-critical" sites). In case a site i is over-critical, the following updates are performed ("toppling" or "relaxation"): $z_i \to z_i - 2$ and $z_{i\pm1} \to z_{i\pm1} + 1$ and, importantly, the existing value of the threshold z_i^c is afterwards replaced by 1 with probability p and by 2 with probability $1 - p$. The boundaries are updated the same way except that for $i = 1$ ($i = L$) addition on site 0 ($L + 1$) is omitted.

Now a tunable degree of anisotropy is introduced into the dynamics. An over-critical site i is relaxed in the following way. Only left movement: with probability $p_l(1-p_r)$ perform the updates: $z_i \to z_i - 1$ and $z_{i-1} \to z_{i-1} + 1$. Only right movement: with probability $p_r(1 - p_l)$ perform the updates $z_i \to z_i - 1$ and $z_{i+1} \to z_{i+1} + 1$. Both left and right movement: with probability $p_l p_r$ perform the updates $z_i \to z_i - 2$ and

[1] The strength of the anisotropy v must be distinguished from the driving velocity v. That the two symbols are so similar is unfortunate, but the context should lift any ambiguity.

$z_{i\pm1} \to z_{i\pm1} + 1$. A new z_i^c is chosen, at random, after every successful update, *i.e.* when at least one unit has been redistributed. We call this version of the model the anisotropic Oslo model (AOM). The strength of the anisotropy is described by the drift velocity

$$\mathsf{v} = (p_r - p_l)/(p_r + p_l) \tag{7.1}$$

which is the net flux of slope units through the system. Clearly it is only sensible to study $p_r + p_l > 0$. The case $p_r = p_l = 1$ corresponds exactly to the OOM, while $p_r = p_l \neq 1$ represents a stochastic variant of the OOM. The avalanche exponents for the extreme, totally asymmetric (TAOM, "Totally Asymmetric Oslo Model") case $p_l = 0$ and $p_r = 1$ can be obtained exactly [see chapter 8 and (Pruessner, 2003b)] and describe, as we shall see below, the scaling behaviour for all $\mathsf{v} > 0$. We are interested in the statistics of the sizes, s, of the avalanches of relaxation induced by the driving $z_1 \to z_1 + 1$. The size of an avalanche is in both versions of the model defined as the number of times the relaxation rule was successfully applied after the drive $z_1 \to z_1 + 1$ in order to make $z_i < z_i^c \ \forall_i$ yet again. Thus $s \geq 0$. The model remains essentially unchanged if the case of no slope redistribution is omitted, as was done in the numerical results presented below.

In the totally anisotropic or asymmetric limit (Priezzhev *et al.*, 2001) the model resembles some features of other exactly solved, *directed* models (Dhar and Ramaswamy, 1989; Maslov and Zhang, 1995; Paczuski and Bassler, 2000; Kloster *et al.*, 2001; Priezzhev *et al.*, 2001). In two dimensions a very similar model has been studied numerically (Vazquez, 2000). However, one should stress that contrary to some other "exact" solutions, the model is solvable directly on the lattice and without assuming any scaling behaviour (Chapter 8). Also the amplitudes of the moments can be calculated exactly.

It is a tedious, but straightforward task to show that the AOM is "Abelian" (see Sec. 5.3.3.1, page 235), *i.e.* the order of updates is irrelevant for its statistical properties. Since the micro-dynamics which prescribes the order of updates is irrelevant, there is no unique way to define a microscopic timescale. Presumably universal exponents of the duration of avalanches are therefore mainly a property of the arbitrary choice of the micro-dynamics. According to DAVID HUGHES and MAYA PACZUSKI (Hughes and Paczuski, 2002), a non-Abelian variant of an Abelian model may or may not remain in the same universality class.

Moreover, one notes that the OOM as well as the AOM contains multiple topplings, *i.e.* a single site can relax several times during a single avalanche.

7.1.1.1 Other models

Similarly, other models presented in Chapter 5 can be made anisotropic as well. For example, it is straight forward to implement a continuous, anisotropic Oslo model (see Sec. 5.2.1.2, page 219) or a anisotropic Oslo model with next nearest neighbour interaction (see Sec. 5.2.1.1). Preliminary numerical results suggest full agreement with the results for the AOM presented below.

It is more interesting to test, whether the observations depend on the way the anisotropy is implemented. The key question here is whether it is the isotropic or anisotropic redistribution of slope units which makes the difference, or maybe just the fact that sometimes only a single unit is redistributed. This is actually also addressed in the case of the AOM at $p_r = p_l \neq 1$, but it deserves full attention in the following way:

anisotropic OOM — This model is identical to the AOM up to the fact that the alternative to toppling to the right (left) with probability p_r (p_l) is toppling to the left (right), rather than leaving the slope units at the toppling site. That way, always two units get redistributed. For $p_r = p_l = 1$ or $p_r = p_l = 0$ the anisotropic OOM is identical to the OOM. The anisotropy is given by $\mathsf{v} = p_r - p_l$.

anisotropic TAOM — This model[2] is an anisotropic version of the TAOM. The rules are the same as for the TAOM (AOM with $p_r = 1$ and $p_l = 0$, see footnote 3 on page 284), but the single, redistributed slope unit is moved to the right with probability p_r only, otherwise it is moved to the left. The anisotropy is given by $\mathsf{v} = 2p_r - 1$.

Numerical data for both models are presented in the next section.

7.2 Numerical Results

In Fig. 7.1 is is demonstrated that the avalanche size distribution $\mathcal{P}(s; L, p_r, p_l)$ follows simple (finite size) scaling

$$\mathcal{P}(s; L, p_r, p_l) = a(p_r, p_l) s^{-\tau} \mathcal{G}\left(\frac{s}{b(p_r, p_l) L^D}\right) \text{ for } s > s_l \tag{7.2}$$

where $\mathcal{G}$ is the universal scaling function, $a(p_r, p_l)$ and $b(p_r, p_l)$ are two anisotropy and system dependent parameters, and s_l is the lower cutoff independent of L. The values for the scaling exponents of the anisotropic model are $\tau_a = 4/3$ and $D_a = 3/2$, which can be derived exactly in the asymmetric limit $\mathsf{v} = 1$ (Pruessner, 2003b) and represent the known universality class of the stochastic, directed sandpile in *two* dimensions (Pastor-Satorras and Vespignani, 2000b; Paczuski and Bassler, 2000;

[2]With a rather silly name, "anisotropic totally asymmetric Oslo model".

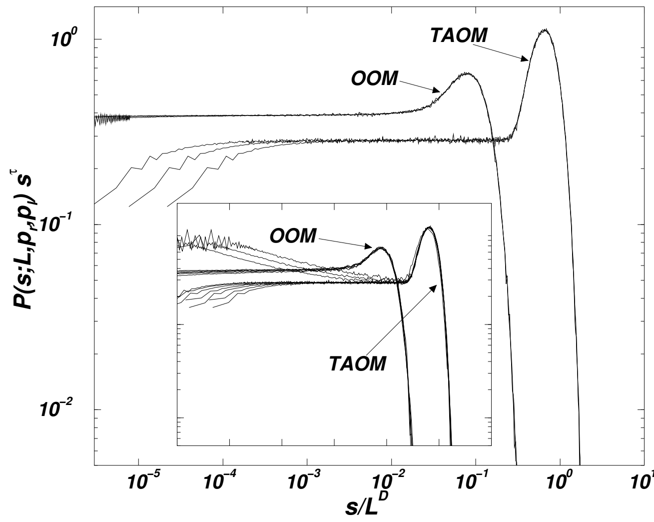


Figure 7.1: Main panel: Data collapse ($\mathcal{P}(s; L, p_r, p_l)s^\tau$ vs. s/L^D) of the normalised and binned distributions for the two extreme cases ($p_r = 1$ with $p_l = 0$ and $p_l = 1$) for $L = 1280, 2560, 5120$. The rescaling was done using $\tau = 1.333\ldots$ and $D = 1.5$ for the AOM and $\tau = 1.555\ldots$ and $D = 2.25$ for the OOM. Inset: Distributions for three choices of $p_r > p_l$, namely $(p_r, p_l) = (1.0, 0.95), (1.0, 0.25), (0.75, 0.25)$, and two choices of $p_r = p_l$ (0.75 and 0.25), for $L = 640, 1280, 2560$. The data collapse of each tuple (p_r, p_l) would form a single line, thereby fixing τ and D. By tuning a and b in Eq. (7.2), as done in the inset, the result collapses with one of the extreme cases shown in both panels.

Kloster *et al.*, 2001), which is in turn closely related to the directed sandpile (Dhar and Ramaswamy, 1989). Numerically, these exponents have been found for all $\mathsf{v} > 0$ studied at sufficiently large system sizes $L \gg L_\mathsf{X}$. The crossover that occurs around L_X is discussed in detail below. The two exponents of the AOM are to be compared with the exponents for the OOM, of $\tau_o = 1.556(4)$ and $D_o = 2.25(2)$. Since the average avalanche size scales linearly with the system size, the exponents are related by $D(2 - \tau) = 1$ (Christensen *et al.*, 1996; Paczuski and Boettcher, 1996).

The easiest way to derive the exponents from numerical data is by analysis of the moments (Tebaldi *et al.*, 1999), which scale according to (7.2) for $n > \tau - 1$ in leading order like [Eq. (2.57), page 88]

$$\langle s^n \rangle = \int_0^\infty ds s^n \mathcal{P}(s; L, p_r, p_l) = a(bL^D)^{1+n-\tau} g_n + \ldots \qquad (7.3)$$

where g_n is discussed below and $\ldots$ denotes sub-leading terms, especially WEGNER's corrections to scaling (Wegner, 1972). In the following, the crossover is studied by means of the rescaled second moment, $\langle s^2 \rangle / L^{5/2}$, which is shown in Fig. 7.2. For non-vanishing anisotropy, $\mathsf{v} > 0$, the ratio approaches a constant as $L \to \infty$. For very small but finite values of v and L the rescaled moment increases with L like $L^{0.75}$, corresponding to the OOM behaviour, but at $L \approx L_\mathsf{X}(\mathsf{v})$ it crosses over and

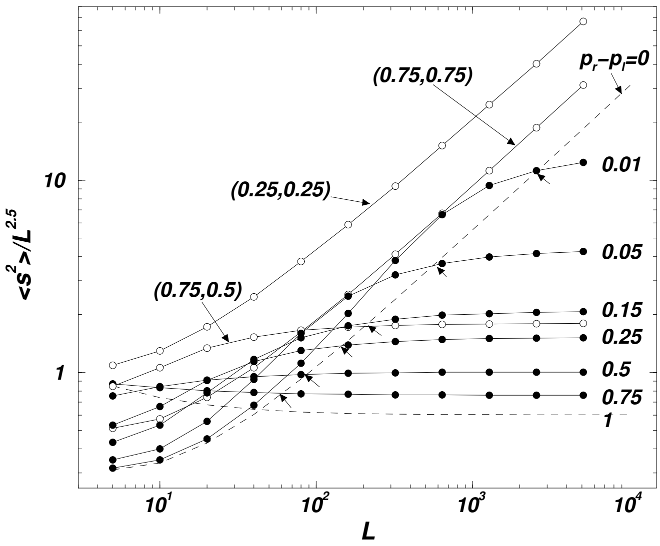


Figure 7.2: Scaling of $\langle s^2 \rangle / L^{2.5}$ for different anisotropies. The filled circles are results for $p_r - p_l$ as indicated and $p_r = 1$. Open circles show other parameters (p_r, p_l). The dashed lines are the two extreme cases OOM ($p_r = p_l = 1$) and the solvable model ($p_r = 1, p_l = 0$). The short arrows mark the approximate crossover.

eventually converges to a finite constant. Below, we shall relate the behaviour of $L_{\mathsf{X}}(\mathsf{v})$ to the effective anisotropy relevant to an experiment of a given size. Here it is emphasised that Fig. 7.2 clearly demonstrates that the universality class of the extremely anisotropic case, $p_r = 1$, $p_l = 0$, contains all systems with non-vanishing anisotropy $\mathsf{v} > 0$. That renders the OOM with $\mathsf{v} = 0$ a special case; remarkably, even for $p_r = p_l \neq 1$ the model still shows OOM behaviour.[3] Thus, it is not the stochasticity itself (Dhar and Ramaswamy, 1989; Tadić and Dhar, 1997; Pastor-Satorras and Vespignani, 2000b) which induces the change in critical behaviour.

7.2.1 Universal amplitude ratios[4]

Eq. (7.2) allows the definition of universal amplitude ratios

$$g_n = \frac{\langle s^n \rangle \langle s \rangle^{n-2}}{\langle s^2 \rangle^{(n-1)}} \tag{7.4}$$

[3]One might think that all models with $p_r = p_l =$ const. are identical, because unsuccessful updates are discounted. However, for small p_r, p_l the probability of the redistribution of only *one* slope unit gets larger. For the AOM, however, $\mathsf{v} = 1$ implies $p_l = 0$, so that the actual value of $p_r > 0$ is irrelevant, as the right movement is the only possible and unsuccessful updates are discounted. Thus, for $\mathsf{v} = 1$ the AOM is identical to the TAOM.

[4]More details on universal amplitude ratios can be found in Sec. 2.3.3.1, page 90, and Sec. 2.4.5, page 102.

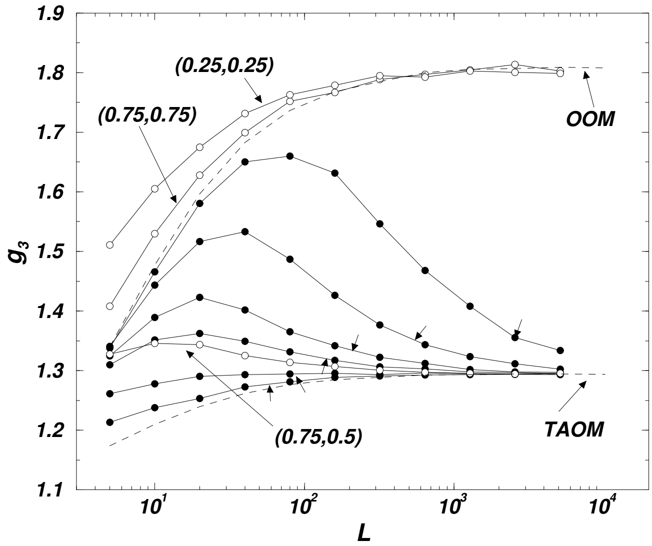


Figure 7.3: Scaling of g_3 [see Eq. (7.4)] for different anisotropies. The filled circles are results for $p_r - p_l$ in the same vertical order as shown in Fig. 7.2 and $p_r = 1$. Open circles show other parameters (p_r, p_l) as indicated. The dashed lines are the two extreme cases OOM ($p_r = p_l = 1$) and the solvable model (TAOM, $p_r = 1, p_l = 0$). The short arrows mark the crossover points in Fig. 7.2.

which can easily be proven to be asymptotically independent of a, b and L. The two constraints on $\mathcal{G}$, which fix the metric factors a and b in (7.2), can be chosen such that g_n are the moments of $x^{-\tau}\mathcal{G}(x)$ as used in (7.3), namely by imposing that $\int_0^\infty x^{1-\tau}\mathcal{G}(x) = \int_0^\infty x^{2-\tau}\mathcal{G}(x) = 1$. The universal amplitude ratio g_3 as shown in Fig. 7.3 and similarly for g_4 in Fig. 7.4 indicate not only the same crossover behaviour as observed in Fig. 7.2, but also the universality of $\mathcal{G}$.

7.2.2 Numerical results for alternative models

The alternative models described in Sec. 7.1.1.1 produce numerical results fully consistent with those presented above; whenever the net transport of slope units through the system is positive, the TAOM behaviour is observed in the thermodynamic limit. Only for $\mathsf{v} = 0$ the OOM is obtained.

Fig. 7.5(a) shows the data of the anisotropic OOM corresponding to Fig. 7.2. As expected, the smaller $\mathsf{v} = p_r - p_l$, the larger the system size at crossover to the TAOM behaviour. For $p_r = p_l = 1$ the “anisotropic OOM” is identical to the OOM. Similarly, Fig. 7.5(b) is the plot for the anisotropic OOM of g_3 corresponding to Fig. 7.3.

The fact that OOM behaviour is observed in the anisotropic OOM for any $p_r = p_l$ is documented in Fig. 7.6, which shows the scaling of the second and third moments in this model with varying the value of $p_r = p_l$.

Similar results are found for the anisotropic TAOM . Here the drift velocity is $\mathsf{v} =$

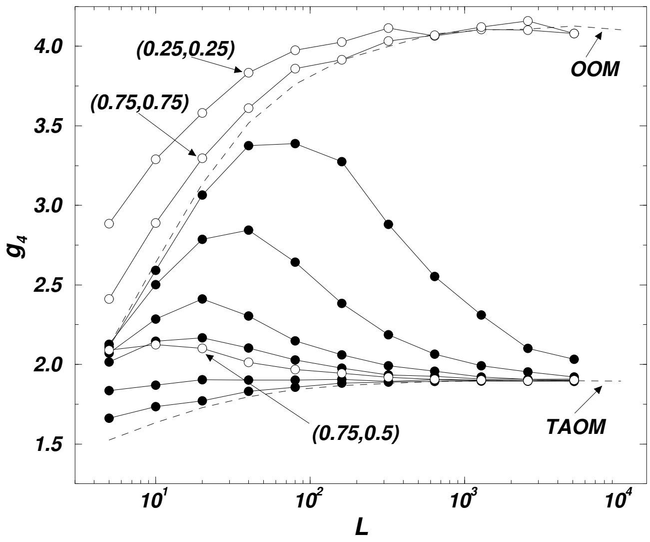


Figure 7.4: Similar to Fig. 7.3, scaling of g_4 [see Eq. (7.4)] for different anisotropies. The filled circles are results for $p_r - p_l$ in the same vertical order as shown in Fig. 7.2 and $p_r = 1$. Open circles show other parameters (p_r, p_l) as indicated. The dashed lines are the two extreme cases OOM ($p_r = p_l = 1$) and the TAOM ($p_r = 1, p_l = 0$).

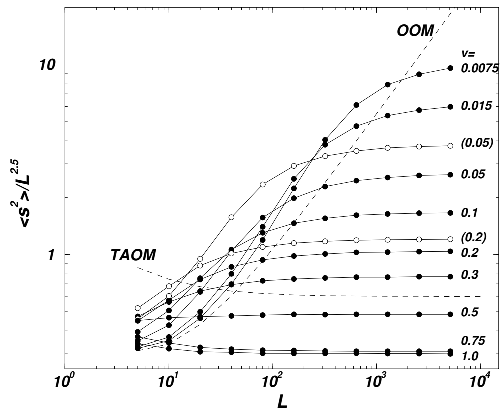
$2p_r - 1$. Fig. 7.7(a) shows again the behaviour of the second moment and Fig. 7.7(b) the behaviour for g_3. Finally, Fig. 7.8 shows g_3 as a function of v for different system sizes. The larger the system, the sharper the jump at small v towards the value of g_3 in the OOM.

7.3 Relation to the Quenched EDWARDS-WILKINSON Equation

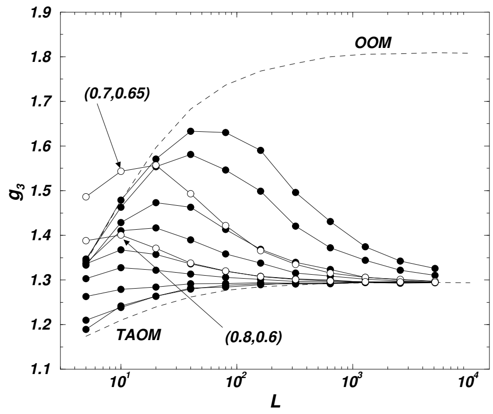
The importance of the above result is highlighted when we recall that the OOM in the continuum limit is described by the qEW equation, see Chapter 6. A similar derivation (details below) shows that the AOM is a qEW equation with an additional drift term

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) - \mathsf{v}\partial_x h(x,t) + \eta(x, h(x,t)), \tag{7.5}$$

where D is the diffusion constant and v the anisotropy or drift velocity as defined above. Just like in the OOM, the quenched noise, $\eta(x, h(x,t))$, represents the randomly chosen z_i^c and $h(x,t)$ is the number of charges received by site x at time t. The quenched nature of the noise makes it difficult to solve (7.5) directly. Together with the boundary conditions (see Chapter 6 and Chapter 9), the noise prevents the drift term $\mathsf{v}\partial_x h$ from being absorbed by a Galilean transformation. However, the above results determines the roughness exponent via $\mathsf{D} = 1 + \chi$ (Paczuski and Boettcher, 1996) to be $\chi = 1/2$ for $\mathsf{v} > 0$, as already suggested in another case of anisotropic de-



(a) Scaling of the reduced second moment in the anisotropic OOM.



(b) Scaling of the universal amplitude ratio g_3 in the anisotropic OOM.

Figure 7.5: Scaling in the anisotropic OOM. **Fig. 7.5(a)**: Rescaled second moment $\langle s^2 \rangle / L^{2.5}$ in the anisotropic OOM versus L for various values of v with $p_r = 1$ (filled circles). For larger v the TAOM behaviour sets in much earlier. The dashed lines are the corresponding data of the OOM and the TAOM. The data with opaque circles and v $= 0.2$ (in brackets) is for $p_r = 0.8$ and $p_l = 0.6$. Similar for v $= 0.05$ for $p_r = 0.7$ and $p_l = 0.65$. **Fig. 7.5(b)**: Moment ratio g_3 for different values values of v (from top to bottom the same order as in Fig. 7.5(a)) in the anisotropic OOM ($p_r = 1$ for all filled symbols) and different system sizes L. The upper and the lower dashed lines are the g_3 data of the OOM and the TAOM. For $p_r = p_l = 1$ the anisotropic OOM is identical to the OOM.

pinning (Tang *et al.*, 1995). For v $= 0$ numerics for the OOM suggest correspondingly that $\chi = 1.25(2)$.

7.3.1 Detailed derivation of the qEW for the AOM

As a first step towards the qEW for the AOM, a LANGEVIN equation for the TAOM is introduced, which is particularly simple, because in the TAOM toppling sites charge only their right neighbour, according to $p_l = 0$ and $p_r = 1$. Using exactly the same approach as for the OOM (see Sec. 6.1, page 245) one derives for the TAOM the following two equations:

$$H(x, t+1) = h(x,t) + \eta(x, h(x,t)) \tag{7.6}$$
$$h(x,t) = H(x-1, t) \tag{7.7}$$

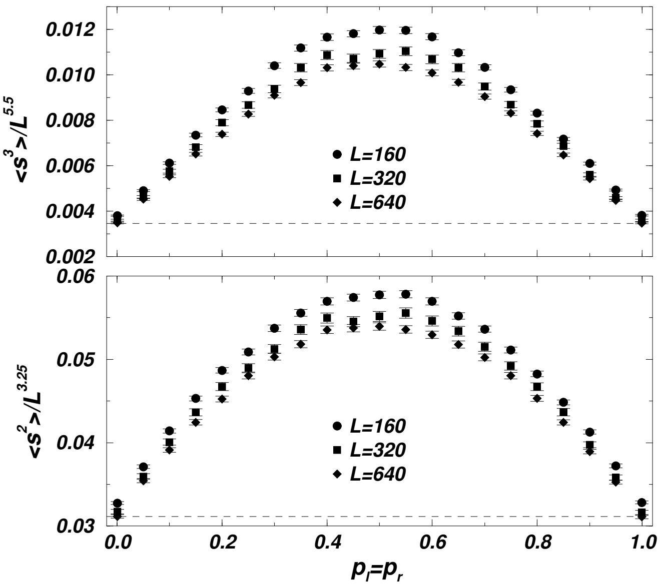
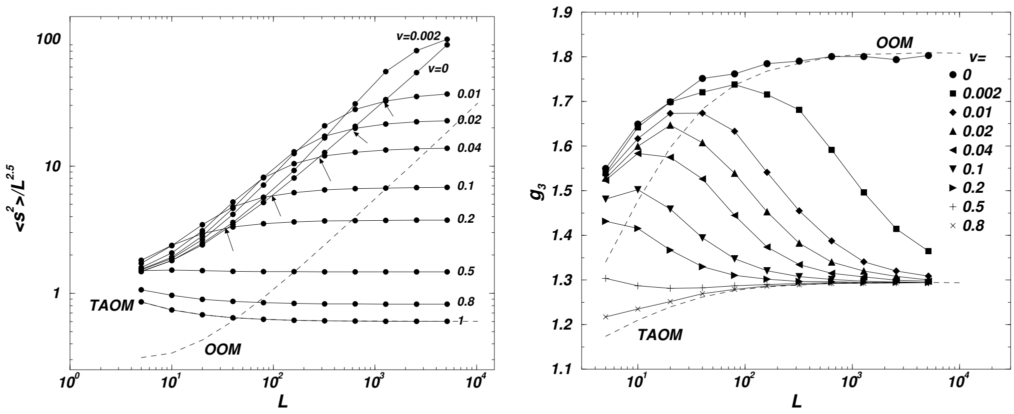


Figure 7.6: The rescaled second and third moments, $\langle s^2 \rangle / L^{3.25}$ and $\langle s^3 \rangle / L^{5.5}$, of the anisotropic OOM in the isotropic regime for various different values for $p_r = p_l$, and $L = 160, 320, 640$. The small variation of the ratio in L is probably only due to finite size effects. The straight dashed line indicates the amplitude found in the OOM.

where $H(x, t)$ is the number of topplings a site x has performed at time t and $h(x, t)$ is the number of charges it has received.[5] The quenched noise $\eta(x, h)$ describes the effect of the critical slope z_i^c: If $\eta(x, h+1) = \eta(x, h) - 1$, then a newly arriving charge does not trigger a toppling. The exact meaning of η can be derived as follows: The initial configuration is $z_i = 0$ everywhere as well as $h(x, t=0) = 0$ and $H(x, t=0) = 0$. If a site has received h charges and has now value z, then $H = h - z$, because of conservation of slope units. From (7.6) it follows that $\eta = -z$ in the stable state.

In the TAOM every site can only be in one of *two* stable states: $z = 1$ or $z = 2$ (see Chapter 8). Therefore η is a sequence of -1 and -2 in t. If $z_i^c = 1$ the only *stable* state is $z_i = 1$, for $z_i^c = 2$ both states are possible. Thus $\eta = -2$ entails $z_c = -2$. However, $\eta = -1$ just means $z = 1$ and *a priori* nothing is known for z^c. But if $\eta(x, h) = -1$ and $\eta(x, h+1) = -1$, a further charge would not increase z, so that z^c must be 1. Otherwise, *i.e.* $\eta(x, h) = -1$ and $\eta(x, h+1) = -2$, the critical slope is $z^c = 2$. The

[5]As in Chapter 6 the index i refers to a site in the discrete, sandpile-like description and x refers to the description in terms of $h(x, t)$ and $H(x, t)$.



(a) Scaling of the reduced second moment.

(b) Scaling of the reduced second moment.

Figure 7.7: Scaling in the in the anisotropic TAOM . **Fig. 7.7(a)**: Rescaled second moment $\langle s^2 \rangle / L^{2.5}$ in the anisotropic TAOM versus L for various values of v. For larger v the TAOM behaviour is observed much earlier. At $\mathsf{v} = 1$ the anisotropic TAOM is identical to the TAOM. The dashed lines are for the corresponding data of the OOM and the TAOM. The line for $\mathsf{v} = 0$ runs parallel to the OOM data. The arrows show the estimated crossover as shown in Fig. 7.11. **Fig. 7.7(b)**: Moment ratio g_3 in the anisotropic TAOM for different v as indicated, The smaller v the later the crossover to the TAOM sets in.

complete mapping is therefore (compare to Tab. 6.1, page 257):

$$
\begin{aligned}
\eta(x, h(x,t)) = -2 \quad &\Rightarrow \quad z^c(x,t) = 2 \\
\left.\begin{aligned} \eta(x, h(x,t)) &= -1 \\ \eta(x, h(x,t)+1) &= -1 \end{aligned}\right\} \quad &\Rightarrow \quad z^c(x,t) = 1 \\
\left.\begin{aligned} \eta(x, h(x,t)) &= -1 \\ \eta(x, h(x,t)+1) &= -2 \end{aligned}\right\} \quad &\Rightarrow \quad z^c(x,t) = 2
\end{aligned}
\tag{7.8}
$$

This is an example for a mapping the sequence of η values to z^c:

h	1	2	3	4	5	6	7	8	9	10	11
$\eta(x,h)$	-1	-1	-1	-2	-2	-2	-1	-1	-2	-1	-2
$z_i^c(h)$	1	1	2	2	2	2	1	2	2	2	2

(7.9)

The reason for this apparent asymmetry, namely that $z^c = 2$ is much more likely than $z^c = 1$, is that if a site relaxes and draws a new random z^c, then $z^c = 2$ will render it stable for $z = 2$, while $z^c = 1$ leads to an additional relaxation in that case. For example, if $z = 3$ relaxes to $z = 2$ and draws the new random threshold $z^c = 2$ then nothing more happens. However, if it draws $z_c = 1$ another toppling occurs, after which another z^c is to be drawn. This makes $z^c = 1$ less likely than $z^c = 2$.

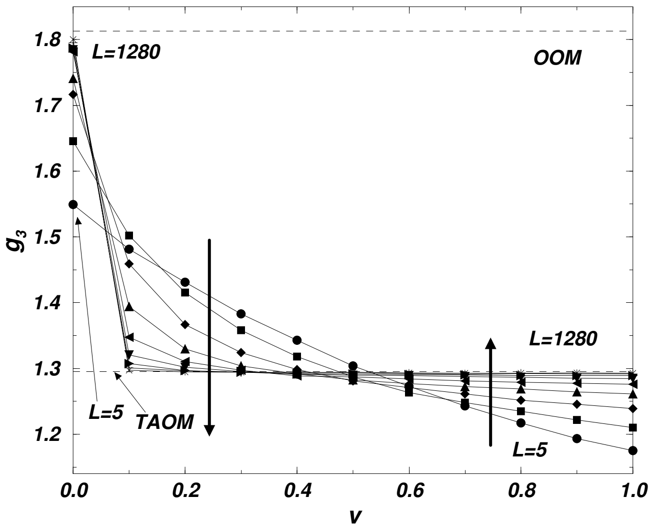


Figure 7.8: Moment ratio g_3 for different values of v in the anisotropic TAOM as indicated and different system sizes $L = 5, 10, 20, \ldots, 1280$. The arrows points in the direction of increasing L. The upper and the lower dashed lines are the estimated asymptotic g_3 data of the OOM and the TAOM. For $\mathsf{v} = 1$ the anisotropic TAOM is identical to the TAOM.

Putting (7.6) and (7.7) together and subtracting $h(x+1,t)$ on both sides leads to

$$h(x+1,t+1) - h(x+1,t) = h(x,t) - h(x+1,t) + \eta(x, h(x,t)) \tag{7.10}$$

or

$$\partial_t^+ h(x+1,t) = -\partial_x^+ h(x,t) + \eta(x, h(x,t)) \quad . \tag{7.11}$$

As indicated by the superscript, both differentials are "forward differences". One can rewrite this equation as

$$\partial_t^+ h(x,t) = -\partial_x^- h(x,t) + \eta(x, h(x-1,t)) \quad , \tag{7.12}$$

to make the derivatives local. Interpreting this equation on a lattice, one can think of an interface where the sites sitting *right of* a negative slope, try to pull up the interface to compensate the slope, as illustrated in Fig. 7.9. The noise $\eta(x, h(x-1,t))$ adds some extra random slope everywhere.

7.3.2 Boundary conditions for the TAOM

The left hand boundary condition of the LANGEVIN equation is where the system is driven, which can be realised by setting the site at $x = 0$ to a particular value, so that

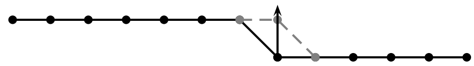


Figure 7.9: This figure illustrates how a kink propagates in the TAOM: On a given site the "backwards slope" decides about propagation, which then "relaxes" this slope (new configuration shown in gray), but produces a new "fault" at the right. This "fault" propagates through the system towards the right. For illustration purposes the effect of the noise is not shown.

the site at $x = 1$ receives the correct number of charges:

$$h(x = 0, t) = E(t) \text{ and } \eta(x = 0, h) \equiv -1 \tag{7.13}$$

where the noise just needs to be fixed to any value. The term $E(t)$ is the driving term, *i.e.* $E(t)$ is the number of times an avalanche has been started at time t, which is the *microscopic* time (just like in Chapter 5). For example

$$E(t) = \left\lceil \frac{t}{L+1} \right\rceil \tag{7.14}$$

would be appropriate, because an avalanche cannot last longer than it takes to sweep through the entire system (the $+1$ is there because the system size with boundary is $L + 1$). The avalanche size is then measured as

$$\int_1^L h(x, t+1) - h(x, t) dx \tag{7.15}$$

The right hand boundary turns out to be irrelevant. With nearest neighbour interactions only, the only site which could be in contact with it, is $x = L$. But it turns out that the right hand boundary gets never actually evaluated, as can be seen by setting $x = L$ in (7.12).

7.3.3 LANGEVIN approach to the full model

The results for the TAOM above and the OOM in Chapter 6 can be captured in a single equation as follows: $H(x, t)$ is the number of slope units which have toppled from site x. This can be split into two contributions H_r and H_l, denoting the amount distributed towards the right and towards the left, *i.e.*

$$H(x, t) = H_r(x, t) + H_l(x, t) \tag{7.16}$$

Just like in Chapter 6 (Sec. 6.1.2, page 247), the total number of topplings is controlled by a quenched noise η, so that

$$H(x,t+1) = h(x,t) + \eta(x,h(x,t)) \tag{7.17}$$

where $h(x,t)$ is the number of charges received by the site at x,

$$h(x,t) = H_r(x-1,t) + H_l(x+1,t) \quad . \tag{7.18}$$

Thus, η represents the number of units stored at a given site, $-\eta = z - z_0$, where z_0 is the initial value. That can be easily checked for $H_r(x,t) = H_l(x,t)$ as in the OOM. In that case, as shown in Chapter 6, starting from $z = 1$ and $h = 0$ everywhere, one finds $z = 1$ corresponds to $\eta = 0$, $z = 2$ corresponds to $\eta = -1$ and $z = 0$ to $\eta = 1$, just as it should (see also Tab. 6.1, page 257).[6] To quantify the anisotropy, we introduce

$$\xi(x,H(x,t)) = H_r(x,t) - H_l(x,t) \tag{7.19}$$

which is 0 in the OOM. From that we have $H - \xi = 2H_l$ and $H + \xi = 2H_r$. Now one derives

$$\begin{aligned}
& h(x,t+1) \\
= \;& H_r(x-1,t+1) + H_l(x+1,t+1) && (7.20)\\
= \;& \frac{1}{2}H(x-1,t+1) + \frac{1}{2}\xi(x-1,H(x-1,t+1)) \\
+ \;& \frac{1}{2}H(x+1,t+1) - \frac{1}{2}\xi(x+1,H(x+1,t+1)) && (7.21)\\
= \;& \frac{1}{2}h(x-1,t) + \frac{1}{2}\eta(x-1,h(x-1,t)) + \frac{1}{2}\xi^+ \\
+ \;& \frac{1}{2}h(x+1,t) + \frac{1}{2}\eta(x+1,h(x+1,t)) - \frac{1}{2}\xi^- && (7.22)\\
& && (7.23)
\end{aligned}$$

where we have used the obvious abbreviations $\xi^{\pm} = \xi(x \pm 1, H(x \pm 1, t+1))$. Introducing a lattice Laplacian and (forward) time derivate one has

$$\partial_t^+ h = \frac{1}{2}\partial_x^2 h + \frac{1}{2}(\eta^+ + \eta^-) + \frac{1}{2}(\xi^- - \xi^+) \tag{7.24}$$

[6] Note that the variable H used in Chapter 6 is half the H used here. In Chapter 6 it represents the number of topplings rather than the number of units redistributed.

In the OOM, this simplifies to the form given in Eq. (6.4), page 248, as $\xi \equiv 0$. However, in general ξ is a random variable. To understand its rôle, we introduce $\tilde{\xi}$

$$\xi(x, H) = H(\mathsf{v} + \tilde{\xi}(x, H)) \tag{7.25}$$

where the velocity v is now defined just like above, (7.1) $\mathsf{v} = (\langle H_r \rangle - \langle H_l \rangle)/\langle H \rangle$, using ensemble averages, so that $\langle H_{r,l} \rangle = p_{r,l}\langle H \rangle/(p_r + p_l)$. Thus for example

$$\begin{aligned}
\xi^+ &= \xi(x+1, H(x+1, t+1)) &(7.26)\\
&= (h(x+1,t) + \eta(x+1, h(t, x+1)))\\
&\times (\mathsf{v} + \tilde{\xi}(x, H(x+1, t+1))) &(7.27)
\end{aligned}$$

so that

$$\begin{aligned}
\partial_t^+ h &= \frac{1}{2}\partial_x^2 h + \frac{1}{2}(\eta^+ + \eta^-) &(7.28)\\
&+ \frac{1}{2}\mathsf{v}(h^- - h^+) + \frac{1}{2}(\tilde{\xi}^- h^- - \tilde{\xi}^+ h^+) &(7.29)\\
&+ \frac{1}{2}(\mathsf{v} + \tilde{\xi}^-)\eta^- - \frac{1}{2}(\mathsf{v} + \tilde{\xi}^+)\eta^+ &(7.30)
\end{aligned}$$

The crucial point is now that $\tilde{\xi}$ vanishes for sufficiently large H, because it quantifies only the deviation from the prescribed anisotropy. One expects it to decay like $1/\sqrt{H}$. That sounds dangerous,[7] because it might represent a relevant perturbation or not vanish fast enough. However, the latter objection does not hold; one of the key assumptions is that the interface develops into a stationary state and all statistics becomes time translation invariant. So, after "arriving" in this stationary state, nothing changes anymore, but $\tilde{\xi}$ keeps decaying. For $t \to \infty$ it vanishes, but this limit must be equivalent to the stationary state.

The term $h^- - h^+$ in (7.28) can be written as $-2\partial_x h$, now with a symmetric derivative. Thus

$$\partial_t^+ h = \frac{1}{2}\partial_x^2 h - \mathsf{v}\partial_x h + \frac{1}{2}(\eta^+ + \eta^-) + \frac{1}{2}\mathsf{v}(\eta^- - \eta^+) \tag{7.31}$$

For $\mathsf{v} = 0$, corresponding to the OOM, one recovers the result in Chapter 6, which is based on $\xi = 0$. For $\mathsf{v} = 1$, one also has $\xi = 0$ on the lattice[8], because that means that

[7] One reason why that actually appears to be wrong is that if v alone parametrises the model completely, why then is $p_r = p_l = 0.75$ different in the amplitude from $p_r = p_l = 1$, see Fig. 7.2? Note that $\mathsf{D} = 1/2$ in any model on the lattice.

[8] But there are some dangerous ambiguities on the lattice: For example, $\mathsf{D} = 1/2$ and $\mathsf{v} = 1$ in (7.34) is identical to $\mathsf{D} = 0$ and $\mathsf{v} = 1$ if the single derivative is replaced by its backwards form:

$$\frac{1}{2}(h^+ - 2h + h^-) - \frac{1}{2}(h^+ - h^-) = -h + h^- .$$

every single slope unit moves to the right, corresponding to the TAOM. Plugging in $\mathsf{v} = 1$ and resolving the spatial derivatives, one has

$$\begin{aligned}\partial_t^+ h &= \frac{1}{2}h^+ - h + \frac{1}{2}h^- - \frac{1}{2}(h^+ - h^-) \\ &+ (\eta^+ + \eta^-) + (\eta^- - \eta^+) \qquad (7.32) \\ &= h - h^- + \eta^- \qquad (7.33)\end{aligned}$$

and the last line is indeed (7.10).

In the OOM it has turned out that all noise terms can be collected into a single effective noise term, such that

$$\partial_t h = \mathsf{D}\partial_x^2 h - \mathsf{v}\partial_x h + \eta(x, h) \qquad (7.34)$$

is the final equation, the quenched EDWARDS-WILKINSON equation with drift (qEWd).

7.3.3.1 The MANNA model and the qEW

After the results discussed above had been published (Pruessner and Jensen, 2003a), a similar calculation for the MANNA model (Sec. 5.2.2.3, page 223) suggested that the argument for $\tilde{\xi} \to 0$ presented above leads to serious problems. In the MANNA model again, $H(x,t)$ is the total number of topplings and $h(x,t)$ the number of charges received. Then

$$h(x,t) = H(x-1,t) + \overline{\eta}(H(x-1,t), x-1) + H(x+1,t) - \overline{\eta}(H(x+1,t), x+1) + \qquad (7.35)$$

where $\overline{\eta}(H,x) = \sum_{H'=1}^{H} \eta(H',x)$ is the sum over the noise η, which decides about the distribution of the two slope units during toppling: it is $+1$ if both units have been redistributed to the right, -1 for both left and 0 if one has been redistributed to either side. Since

$$H(x,t+1) = \left\lfloor \frac{h(x,t)}{2} \right\rfloor = \frac{h(x,t) + \Upsilon(h(x,t))}{2} \qquad (7.36)$$

with $\Upsilon(h) = 0$ if h is even and -1 otherwise, one arrives at

$$\partial_t h = \frac{1}{2}\partial_x^2 h(x,t) + \Upsilon(h(x-1,t)) + \Upsilon(h(x+1,t)) + \overline{\eta}(H(x-1,t), x-1) - \overline{\eta}(H(x+1,t), x+1) \,. \qquad (7.37)$$

Of course, this equation can be written in terms of h only, by using (7.36).

Open issue

There are a number of problems with this equation. First of all, it does not seem to be possible to absorb E into the noise as periodicity like in Sec. 6.1.2, page 247, because of the different signs in front of the two noise terms. This could possibly be fixed, for example by changing the sign of $E(h(x+1,t))$ and changing the initial condition. More serious, however, is the problem that according to the arguments presented in the preceeding section, in the stationary state $\overline{\eta} \propto H$ with fluctuations, which eventually become irrelevant. If that is true, the equation of motion for the MANNA model would not contain a quenched noise. Maybe that is resolved by the derivative-like form of the last two terms.

7.3.4 Boundary conditions of the qEWd

In general the right hand boundary condition of (7.28) must be $h(x = L+1, t) = 0$, which just states that all slope units are lost there and nothing enters from there. The same applies to the driving term: One simply *fixes* $h(x = 0, t) = E'(t)$ to a particular value and chooses the noise terms such that their combined effect on site $x = 1$ is as if it was "kicked" an appropriate number of times. After fixing all noise terms at $x = 0$ to 0, the only terms involving the left boundary which enter into $\partial_t h(x = 1, t)$ are the h^- terms: $\frac{1}{2}h^- + \frac{1}{2}vh^-$, so that $h(x = 0) = E'(t)$ provides $\frac{1}{2}(1 + \mathsf{v})E'(t)$ kicks to the site at $x = 1$.

Therefore, the left hand boundary condition is

$$h(x = 0, t) = \frac{2E(t)}{1 + \mathsf{v}} , \tag{7.38}$$

where $E(t)$ is now the number of initial kicks delivered to the system in total until (microscopic) time t, very similar to Eq. (6.7), page 248. The right hand boundary is $h(x = L+1, t) \equiv 0$ (clearly the $+1$ disappears in the continuum).

7.3.4.1 Again the TAOM

All that is very straight forward, but the right hand boundary now seems to contradict the expected shape of h in the TAOM: In the TAOM the right hand side is the only point where slope units can leave the system. So, every unit must pass through every single site, so that on average every site topples with the same frequency;[9] the velocity of the *interface* is everywhere the same:

$$\langle \partial_t h(x,t) \rangle = c \tag{7.39}$$

[9]This is very similar to the OOM with original boundary condition, Sec. 5.1.4, page 212.

where $\langle\rangle$ denotes the ensemble average. Of course, this constant is time-independent only if the drive is constant. Thus, in general

$$c = \frac{2}{1+\mathsf{v}}\dot{E}(t) \tag{7.40}$$

Depending on the boundary conditions, it is then $\langle h(x,t)\rangle = tc$. The last site in the bulk, just left from the right boundary, has therefore an ever growing value of h, while the boundary site stays at $h = 0$: A jump develops. Does this make sense?

Yes, in fact, it does. First of all, one can solve the stationary velocity equation (see below) and arrives at an expression, which exactly develops such a kink. This, however, does only happen if D vanishes, in fact v/D must diverge. But as discussed above (see footnote 8 on page 293), the TAOM corresponds actually to $\mathsf{D} = 1/2$ and $\mathsf{v} = 1$. A vanishing diffusion constant is obtained only if the gradient term is a backward slope only. This ambiguity is a lattice versus continuum problem.

7.3.5 Toppling frequency

There is a very elegant way to calculate the *exact* toppling frequency for the full, generalised qEW equation with drift, Eq. (7.34). This leads to an expression for the average avalanche size, which can also be calculated by other means, see Sec. 7.4.

The average charging frequency is $\langle\partial_t h\rangle$ and expected to approach a stationary value, $\left\langle\dot{h}\right\rangle = f(x)$, so by taking the ensemble average in Eq. (7.34) and deriving once by time

$$0 = \mathsf{D}\frac{d^2}{dx^2}f(x) - \mathsf{v}\frac{d}{dx}f(x)\ , \tag{7.41}$$

where $\partial_t\langle\eta(x,h)\rangle = \left\langle\dot{h}\partial_h\right\rangle\eta(x,h) = 0$ has been used, which is part of the stationarity assumption. Actually, one has only

$$\partial_t^2\left\langle\dot{h}\right\rangle = 0 = \partial_t\left(\mathsf{D}\partial_x^2 f(x) - \mathsf{v}\partial_x f(x)\right) + \partial_t^2\langle\eta(x,h)\rangle\ , \tag{7.42}$$

so that $\partial_t^2\langle\eta(x,h)\rangle = 0$, *i.e.* $\langle\eta(x,h)\rangle = c_1 t + c_0$, with unknown constants c_1 and c_0. That $c_1 = 0$ follows naturally on the lattice, because there the noise is bound.

Eq. (7.41) is a linear, homogenous ODE, so it can easily be solved for any drive

$$f(x=0) = \frac{2}{1+\mathsf{v}}\frac{d}{dt}E(t)\ . \tag{7.43}$$

Here we have the simple drive $E(t) = \dot{E}t$, where $\dot{E}$ is constant.[10] Thus, taking $f(x =$

[10]In Chapter 6 $\dot{E}$ was called v [see Eq. (6.71), page 273], but this variable can easily be confused with the anisotropy parameter v.

$0) = \dot{E}$ and $f(x = L) = 0$ as right hand BC, one immediately has

$$f(x) = \frac{2\dot{E}}{1+\mathsf{v}} \frac{e^{-\frac{\mathsf{v}}{\mathsf{D}}(L-x)} - 1}{e^{-\frac{\mathsf{v}}{\mathsf{D}}L} - 1} . \tag{7.44}$$

For $\mathsf{v} \to 0$ one has

$$\lim_{\mathsf{v}\to 0} f(x) = 2\dot{E}\frac{L-x}{L} , \tag{7.45}$$

up to discretisation effects in full agreement with Eq. (5.11), page 214. For very small D (7.44) is roughly constant everywhere on the lattice apart from a very small "boundary layer" of size D/v where it drops to 0. This is exactly what is expected from sec. 7.3.4.1. These analytical results are still to be tested numerically. Open task

The spatial integral gives the total toppling frequency, *i.e.* the total number of charges if driving with speed $\dot{E}$, so dividing by $\dot{E}$ it is the average avalanche size, measured as the number of charges instead of topplings.

$$\langle s \rangle = \left(\frac{L}{1 - e^{-\frac{v}{D}L}} - \frac{D}{v} \right) \frac{2}{1+\mathsf{v}} \tag{7.46}$$

The limit $\mathsf{v} \to 0$ gives, as expected,

$$\langle s \rangle = L \tag{7.47}$$

see Eq. (5.7), page 211, the factor $1/2$ in which comes from the fact the it measures the avalanche size as the number of topplings. It is worth stressing that the form of $E(t)$ in this section comes from (7.38), which is a lattice result.

It is worthwhile to extend this calculation to the problem of a source in the bulk at position αL, *i.e.* Eq. (7.34) now becomes

$$\partial_t h = \mathsf{D}\partial_x^2 h - \mathsf{v}\partial_x h + \eta(x,h) + E(t)\delta(x - \alpha L) \tag{7.48}$$

with homogenous boundaries $h(x=0,t) = h(x=L,t) = 0$. With $\left\langle \dot{h} \right\rangle = f(x)$ the problem is now

$$0 = \mathsf{D}\frac{d^2}{dx^2} f(x) - \mathsf{v}\frac{d}{dx} f(x) + \dot{E}\delta(x - \alpha L) , \tag{7.49}$$

with BC $f(x=0) = f(x=L) = 0$. The solution of this inhomogeneous linear ODE is

$$f(x) = \frac{\dot{E}}{\mathsf{v}} \left\{ \frac{e^{\frac{\mathsf{v}}{\mathsf{D}}L(1-\alpha)} - 1}{e^{\frac{\mathsf{v}}{\mathsf{D}}L} - 1} \left(e^{\frac{\mathsf{v}}{\mathsf{D}}x} - 1 \right) + \left(1 - e^{\frac{\mathsf{v}}{\mathsf{D}}(x-\alpha L)} \right) \theta(x - \alpha L) \right\} , \tag{7.50}$$

where $\theta(x)$ is the HEAVISIDE θ-function. The resulting average avalanche size is

$$\langle s \rangle = \frac{L}{\mathsf{v}} \left\{ \frac{1 - e^{-\frac{\mathsf{v}}{\mathsf{D}}L\alpha}}{1 - e^{-\frac{\mathsf{v}}{\mathsf{D}}L}} - \alpha \right\} \tag{7.51}$$

which will be recovered in Sec. 7.4.[11]

Since the second spatial derivative ∂_x^2 of the toppling frequency [(7.44) or (7.50)] does not vanish in the case of $\mathsf{v} \neq 0$, the interface $h(x,t)$ develops an ever increasing slope and curvature. Nevertheless, (7.44) and (7.50) are *stationary states*, and adding

$$h'(x,t) = h(x,t) + f(x)t_0 \tag{7.54}$$

is — apart from the initial condition — a solution of (7.34) [or (7.48)] for $E(t)$ being replaced by $E(t)+\dot{E}t_0$. It might look unphysical to have an ever increasing curvature in the interface; in fact on page 253 and page 264 it has been argued that such a curvature is not supported by the interface. This remains true in the case of $\mathsf{v} = 0$, for which Eq. (7.45) holds or for (7.48)

$$\lim_{\mathsf{v}\to 0} f(x) = \frac{\dot{E}}{\mathsf{D}} \left(x(1-\alpha) = (x-\alpha L)\theta(x-\alpha L)\right) \tag{7.55}$$

which is only *linear* in x. For non-vanishing v, however, an ever increasing curvature can be balanced by the convection term.

For small α, one should expect that (7.51) becomes similar to the boundary driven case (7.46). Above, it has been estimated from consideration *on the lattice* that the boundary needs to be driven as $2E/(1+\mathsf{v})$, (7.38) in order to deliver height $E(t)$ to site $x=1$, so that (7.46) should correspond to (7.51) at least in the limit of large L, where the region $[0, \alpha L]$, which is inconsistent in these two picture, vanishes. At least for vanishing v, this is confirmed, see (7.47) and (7.66).

---

[11] There is a little, surprising twist in this section: In Eq. (7.44) $f(x)$ has the same dimensions as $\dot{E}$, while in Eq. (7.50) it has the dimension of $\dot{E}/\mathsf{v}$. Similarly the difference in the dimension of $\langle s \rangle$ in (7.46) and (7.51). The reason for this inconsistency has actually already been discussed in Sec. 6.3.1, page 274. In the boundary-driven case, the driving term $E(t)$ retains its meaning during the transition from the lattice to the continuum as a certain, given number of charges at $h(x=0,t)$. However, in the bulk-driven case, the difference equation governing the lattice model is to be interpreted as an equation of motion of local averages. The driving term in an expression like

$$h(x,t+1) - h(x,t) = h(x+1,t+1) - 2h(x,t) + h(x-1,t) + E(t)\delta_{x,\alpha L} \tag{7.52}$$

with a KRONECKER-δ fixing the position of the driving in the bulk must be interpreted as the total (integrated) influx during time $\Delta t = 1$; it should better be written as

$$h(x,t+\Delta t) - h(x,t) = h(x+\Delta x,t+1) - 2h(x,t) + h(x-\Delta x,t) + E(t)\delta_{x,\alpha L}\Delta t \tag{7.53}$$

So, $E(t)\delta_{x,\alpha L}$ has the dimension of a number of slope units per time and space.

The inconsistency in the dimensions is fixed by D, L and v, which are dimensionless on the lattice, but not so in the continuum.

The only reason why these inconsistencies come up again and again is because one starts out from a form like $h(x,t+1) = h(x+1,t+1) + h(x-1,t) + E(t)\delta_{x,\alpha L}$ where $E(t)$ is just a number of charges. Then one changes this into the difference equation (7.52). In principle, this is the same change of dimension as observed in the transition from $h(t+1) = E$, where E has the same dimension as h to $h(t+1) - h(t) = E$, where E is "naturally" a rate.

7.4 Exact Average Avalanche Size

The toppling frequency in the stationary state is one way to calculate the average avalanche size. The other way is to consider each slope unit fed into the model as a random walker. This is possible because the slope units are moved through the system at random *whenever they move*[12] — they might get stuck or buried but provided that they move, they do that diffusively. Thus, for a given slope unit, a time-step elapses only if it moved, and the average time they spend in the system is identical to the average number of moves, *i.e.* charges. Thus, the average avalanche size is given by the mean time a random walker spends in the system, until it leaves through one of the boundaries.

That way, the anisotropy v simply becomes the net drift of a random walker between two absorbing boundaries, with diffusion constant D. The latter follows from the probabilities on the lattice. In the following, all results presented are for the anisotropic TAOM , simply because it was the first model studied in this context and it is slightly simpler to follow, because it depends only on one parameter, p_r. For the anisotropic TAOM it is $\mathsf{D} = 2p_r(1 - p_r)$ (binomial-to-Gaussian mapping) and $\mathsf{v} = 2(p_r - 1/2)$, see Sec. 7.2.2, page 285. Thus $\mathsf{D} = \frac{1}{2}(1 - \mathsf{v}^2)$ in the following.

The problem of a biased diffusion between two absorbing walls in one dimension is straight forward and has been discussed in detail by Farkas and Fülöp (Farkas and Fülöp, 2001). Their calculations give for the anisotropic TAOM

$$\langle s \rangle_{\text{aniso TAOM}} = \frac{L}{\mathsf{v}} \left(\frac{1 - e^{-\frac{\alpha L \mathsf{v}}{D}}}{1 - e^{-\frac{L\mathsf{v}}{D}}} - \alpha \right) \tag{7.56}$$

where $\alpha = 1/L$ is the "continuum approximation" (with $\Delta x = 1$) of the point where the particles are fed in, as in Sec. 7.3.5.

Because of the presence of two length scales, D/v and L, it is actually not too surprising that the continuum result (7.56) deviates slightly from the simulations on the lattice, see Fig. 7.10. The result for a simulation of the anisotropic TAOM is also compared to a pure diffusion process on the lattice, which is shown in the same figure. The actual cause of the lattice/continuum problem is that the continuum transition is usually performed by rescaling time by a factor b^2 and space by a factor b. However, the presence of a net drift introduces a fixed ratio between space and time direction.

However, the lattice problem is actually solvable in closed form (Grimmett and

[12] In a concrete implementation one might have to formally introduce this randomness by claiming that a stack of (indistinguishable) slope units is randomly mixed whenever it is updated to avoid biases, for example that the top slope unit has a higher probability to move to the right than the unit below.

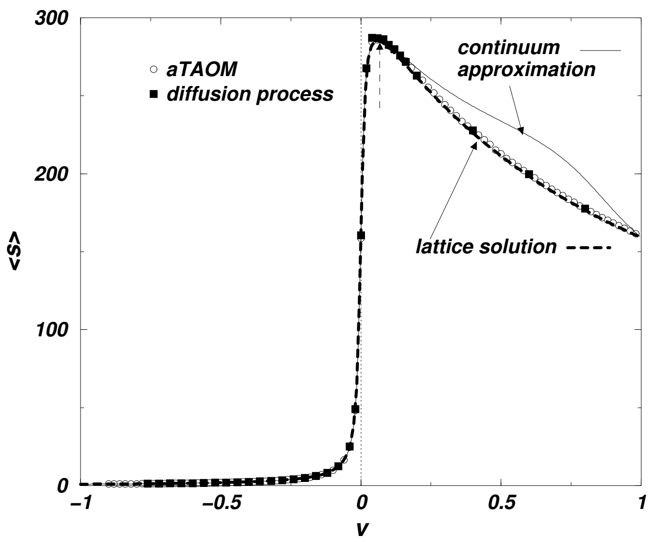


Figure 7.10: Comparison of the average avalanche size $\langle s \rangle$ in the anisotropic TAOM to a random walk in the continuum (7.56). The simulation results of the anisotropic TAOM are shown as open circles, the results from a pure diffusion process on a lattice as filled squares. While they correspond perfectly, the continuous approximation (straight line) seems to deviate in the intermediate regime especially around $\mathsf{v} = 0.5$. The dotted perpendicular line marks $\mathsf{v} = 0$, the dashed arrow marks the crossover as derived from Fig. 7.11. The proper solution on the lattice (7.58) is shown as a thick dashed line, which agrees perfectly with the numerical results.

Stirzaker, 1992, p. 74).[13] The key idea is not to start with the actual diffusion process as a function of space and time, but to write down the average number of moves U_k of a random walker as a function of its starting point k:

$$U_k = p_r(1 + U_k) + (1 - p_r)(1 + U_{k-1}) \tag{7.57}$$

with appropriate boundary conditions. The solutions is

$$U_k = \begin{cases} k(N-k) & \text{for } \mathsf{v} \neq 0 \quad (7.58a) \\ -\dfrac{1}{\mathsf{v}} \left(k - L \dfrac{1 - \left(\frac{1-\mathsf{v}}{1+\mathsf{v}}\right)^k}{1 - \left(\frac{1-\mathsf{v}}{1+\mathsf{v}}\right)^L} \right) & \text{otherwise} \quad (7.58b) \end{cases}$$

where the first line is of course the limit of the second, $\mathsf{v} \to 0$. This solution fully agrees with the numerical results shown in Fig. 7.10.

[13] I am indebted to MATTHEW STAPLETON for pointing out this reference to me.

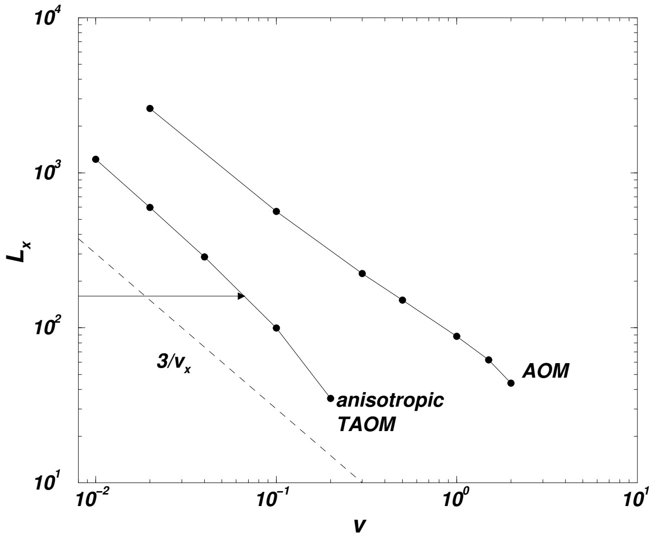


Figure 7.11: Scaling of the crossover system size versus the anisotropy $\mathsf{v_X}$ in the anisotropic TAOM (see Fig. 7.7(a), page 289) and the AOM (Fig. 7.2, page 284). The dashed line corresponds to (7.67). The horizontal arrow marks the crossing velocity for $L = 160$, which is shows as arrow in Fig. 7.10 and Fig. 7.12.

7.4.1 Details in the continuum

Focusing on the continuum problem, it is very instructive to study the solution (7.56) in some detail. For further analysis it is very convenient to define a function

$$f_b(x) = \frac{1 - e^{xb}}{(1 - e^{x/b})xb^2} + \frac{1}{b} - \frac{1}{x} \tag{7.59}$$

so that

$$\langle s \rangle_{\text{aniso TAOM}} = -\frac{L^2 \alpha^{3/2}}{\mathsf{D}} f_{\sqrt{\alpha}} \left(-\frac{\sqrt{\alpha} L \mathsf{v}}{\mathsf{D}} \right) + \frac{L^2 \alpha}{\mathsf{D}} \ . \tag{7.60}$$

The function $f_b(x)$ behaves for $x < 0$ asymptotically in b (more specifically $-x \gg b$ and $x^2 b \ll |x|$ and $x^2 b \ll 1/|x|$) like

$$\lim_{b \to 0} f_b(x) = -\frac{1}{2} x - \frac{1}{x} \ . \tag{7.61}$$

Moreover

$$\lim_{x \to 0} f_b(x) = \frac{1}{2} \frac{1 + b^2}{b} \tag{7.62a}$$

$$\lim_{x \to 0} \frac{d}{dx} f_b(x) = \frac{1}{12 b^2} \ . \tag{7.62b}$$

The function has a minimum at $x = -\sqrt{2}$ where $f_b(-\sqrt{2}) = \sqrt{2}$.

Now, the sharp crossover shown in Fig. 7.10, where $\langle s \rangle$ drops from about $2L$ to close to 0, can be quantified. The idea is to find a "GINZBURG-criterion" which would allow us to make a quantitative statement in which small region around $\mathsf{v} = 0$ to expect OOM behaviour. To do that explictly, one sets $\mathsf{D} = (1/2)(1 - \mathsf{v}^2)$ from above and $\alpha = 1/L$ (noting that these results from the lattice violate dimensional consistency). The maximum average avalanche size in the anisotropic TAOM is therefore at

$$\mathsf{v}_{\max} = \frac{\mathsf{D}}{L\sqrt{\alpha/2}} = \frac{1}{\sqrt{2L}}\left(1 - \frac{1}{2L} + \dots\right) \tag{7.63a}$$

where

$$\langle s \rangle_{\max} = \frac{L^2\alpha}{\mathsf{D}}\left(1 - \sqrt{2\alpha}\right) = 2L(1 - \sqrt{\frac{2}{L}} + \dots) \,. \tag{7.63b}$$

It is reasonable to assume that somewhere in the region $[0, \mathsf{v}_{\max}]$ the crossover occurs. Unfortunately, this regions contracts only as fast as $1/\sqrt{L}$, while one expects the crossover to occur, when the ballistic motion dominates over the diffusion, $L^2/\mathsf{D} > L/\mathsf{v}$, leading to an crossover length

$$L_{\mathsf{X}}(\mathsf{v}) \propto \frac{\mathsf{D}}{\mathsf{v}} \tag{7.64}$$

and a crossover anisotropy

$$\mathsf{v}_{\mathsf{X}} \propto \frac{\mathsf{D}}{L} \tag{7.65}$$

This has been confirmed numerically, based on heuristic estimation of L_{X}, as shown by the marks in Fig. 7.2, page 284. Plotting L_{X} versus v as derived from these marks gives a plot shown in Fig. 7.11. The exponent associated with the scaling of the crossover anisotropy v_{X} in L is x, *i.e.* $\mathsf{v}_{\mathsf{X}} \propto L^{-x}$.

For small v, the average avalanche size in the anisotropic TAOM behaves like

$$\langle s \rangle = L\alpha(L-1) + \frac{L^3\alpha}{12\mathsf{D}^2}\mathsf{v} + \mathcal{O}(\mathsf{v}^2) \,, \tag{7.66}$$

using (7.60) and (7.62). The limit $\mathsf{v} \to 0$ is fully consistent with Eq. (7.47). Using approximation (7.66) to calculate v_{X}, one finds

$$\mathsf{v}_{\mathsf{X}} \approx \frac{3}{L} \tag{7.67}$$

by equating $\langle s \rangle$ from (7.66) and $\langle s \rangle_{\max}$ from (7.63b). The various estimates for the crossover velocity are shown in Fig. 7.12. A comparison to the numerical results in Fig. 7.11 supports (7.67) very well.

One interesting direct implication of (7.66) for the scaling law $D(2 - \tau) = 1$ in the OOM and the TAOM is that for a scaling of α like $\alpha \propto L^{-\mu}$, one has in general $\langle s \rangle \propto L^{2-\mu}$. Especially, the random drive corresponds to a constant α, as does a fixed

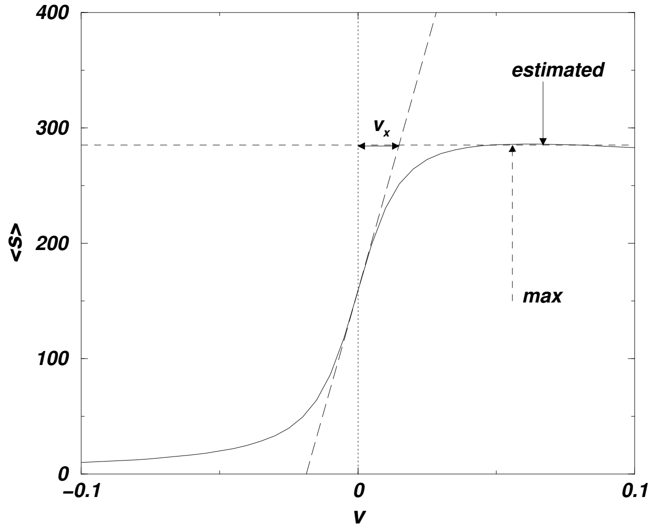


Figure 7.12: Equation (7.56) for $L = 160$, as already shown in Fig. 7.10. The asymptotic maximum average avalanche size is marked by a dashed horizontal line, the dashed arrow indicates the position of this maximum, see (7.63b). The long dashed line is (7.66) and the double arrow is v_X derived from this equation. The arrow from top (labelled "estimated") is the estimate for the crossover velocity as derived from Fig. 7.11.

relative position of the feeding, for example in the centre of the system at $x = L/2$. Thus, in general

$$\gamma_1 = 2 - \mu \leq 2 \ , \tag{7.68}$$

with γ_1 introduced in Sec. 5.2.3, page 224 [and Eq. (2.73), page 100] as $\langle s \rangle \propto L^{\gamma_1}$.

7.5 Towards a Universal Crossover Function

It would be very satisfying to cast the crossover behaviour observed in the anisotropic Oslo model into a single scaling function. Similarly to (Tsuchiya and Katori, 1999b), one might make an ansatz like

$$\langle s^n \rangle (L, \mathsf{v}) = L^{D_a(n-1)+1} \mathsf{v}^{-\theta_n} \mathcal{G}_n \left(\frac{\mathsf{v}}{L^{-x}} \right) \tag{7.69}$$

where θ_n is the crossover exponent for the $n$the moment and $\mathcal{G}_n$ the crossover function, which mediates between AOM behaviour for large arguments and OOM behaviour for small arguments. The latter entails

$$\mathcal{G}_n(z) \xrightarrow{z \to 0} \mathcal{O}(z^{\theta_n}) \tag{7.70}$$

in leading order with $\theta_n = \frac{D_o - D_a}{x}(n-1)$. Thus, plotting the nth moment of the avalanche size distribution as $\mathsf{v}^{\theta_n} \langle s^n \rangle / L^{D_a(n-1)+1}$ versus $\mathsf{v} L^x$ should lead to a data

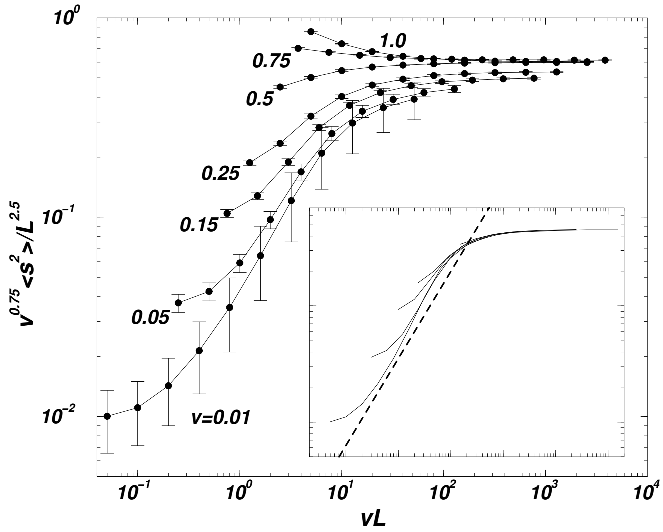


Figure 7.13: Attempt of an data collapse for the crossover according to Eq. (7.69). The data collapse fails, but a fairly good collapse is obtained by rescaling the data "by hand" to appropriate values. The deviation for small values of $\mathsf{v}L$ are acceptable as finite size corrections. The thick dashed line shows the expected slope $z^{0.75}$, see Eq. (7.70).

collapse, revealing $\mathcal{G}_n(z)$. This has been done for the second moment in Fig. 7.13, with $\theta_2 = 0.75$. Unfortunately, the data do not collapse. While one can understand some deviation from $\mathcal{G}_n(z)$ for small L due to finite size corrections, all curves should asymptotically converge.

Eq. (7.69) can be put on a more general basis (Lübeck, 2003b) by assuming generalised homogeneity for the avalanche size distribution,

$$\mathcal{P}(s; L, \mathsf{v}) = b^{\gamma}\mathcal{P}(bs; b^{\alpha}L, b^{\beta}\mathsf{v}) \tag{7.71}$$

with unknown exponents α, β and γ. For $\mathsf{v} = 0$ we know already the scaling behaviour (it is the one of the OOM), so that one finally arrives at an expected scaling form

$$\mathcal{P}(s; L, \mathsf{v}) = as^{-\tau_o}\mathcal{G}\left(\frac{s}{bL^{D_o}}, \mathsf{v}L^x\right) \tag{7.72}$$

which allows the derivation of (7.69).

Alternatively, one can force a collapse, as shown in the inset of Fig. 7.13 by plotting $a_n(\mathsf{v})\langle s^n\rangle/L^{D_a(n-1)+1}$ versus $b_n(\mathsf{v})L$. The resulting $a_n(\mathsf{v})$ and $b_n(\mathsf{v})$ can then be plotted versus v to determine their functional dependence on v. Fig. 7.14 shows the resulting behaviour, which is very close to the expected one: $a_2(\mathsf{v})$ should be proportional to $\mathsf{v}^{0.75}$ and $b_2(\mathsf{v})$ should be linear in v.

Open issue

It is not clear yet whether the deviation from the expected scaling is genuine or only due to numerical problems or finite size effects. In an alternative approach, one

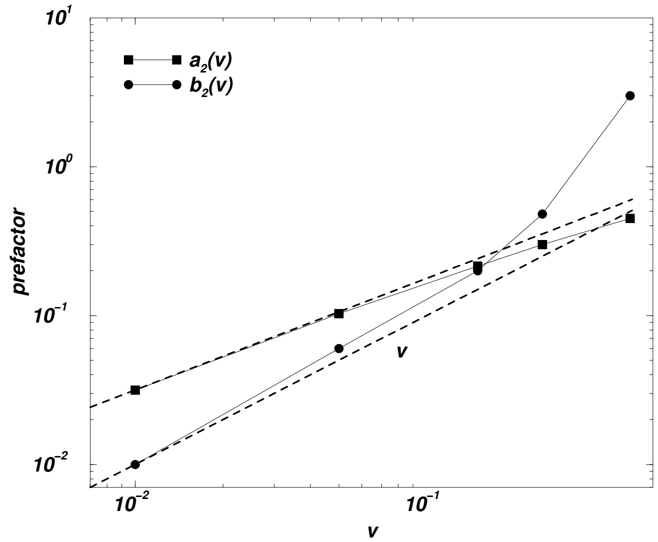


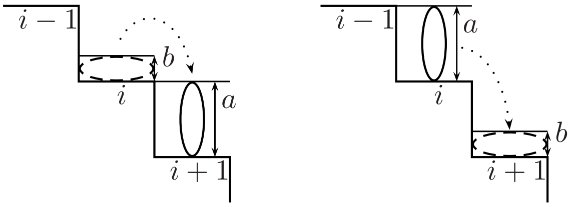
Figure 7.14: Scaling of the two sets of parameters used in the inset of Fig. 7.13 to force a data collapse. The parameters seem to scale nicely for sufficiently small v, as indicated by the two thick dashed lines.

could study $\mathcal{P}(s; L, \mathsf{v})$ directly.

7.6 Relation to Experiments

We stress that the anisotropy is in the amount of *slope* transported between sites involved in a relaxation event. Any net flux of the slope is eventually compensated by the toppling of the last site (see Sec. 7.6.1) and the slope is therefore asymptotically stationary.

One process leading to an anisotropic redistribution of slope arises when the toppling grain is elongated, see Fig. 7.15. Most remarkably, in the original experiment (Frette *et al.*, 1996) it was noted that only the elongated rice samples showed scale-invariant behaviour. A reorientation of a single grain as shown in Fig. 7.15(a), leads to a net flux of slope to the right (or to the left, as shown in Fig. 7.15(b)). It can happen only once, and in fact the occurrence of a net flux depends on how and whether the rice enters and/or leaves the system with a typical orientation. If that is the case, then an average reorientation is distributed among all topplings on its way through the system, *i.e.* $\mathsf{v} \propto 1/L$. Because $L_\mathsf{X} \propto 1/\mathsf{v}$ and $\mathsf{v} \propto 1/L$, (7.64), this represents a marginal case, because it is impossible to decide whether $L \gg L_\mathsf{X}(\mathsf{v})$ or not. The dependence of, say, the ratio $\langle s^2 \rangle / L^{2.5}$ on L would in this case be given by an (unknown) trajectory through the diagram in Fig. 7.2 since a change in L also leads to a change in v, which allows non-universal quantities to enter. Two possible scenarios are shown in Fig. 7.16.



(a) A stylised toppling: flow to the right

(b) A stylised toppling: flow to the left

Figure 7.15: A stylised toppling of a single grain. **Fig. 7.15(a)**: If an elongated grain of width b and height a topples from site i to site $i+1$, it reduces the height at i by b and increases the height at $i+1$ by a, thereby increasing the slope at site $i-1$ by b, decreasing the slope at i by $a+b$ and increasing the slope at $i+1$ by a. The net flux of slope is $a-b$ to the right. The opposite effect is shown in Fig. 7.15(b).

We do not know if the experiment by FRETTE *et al.* (Frette *et al.*, 1996) involves this complication, but the exponents extracted are not consistent.[14]

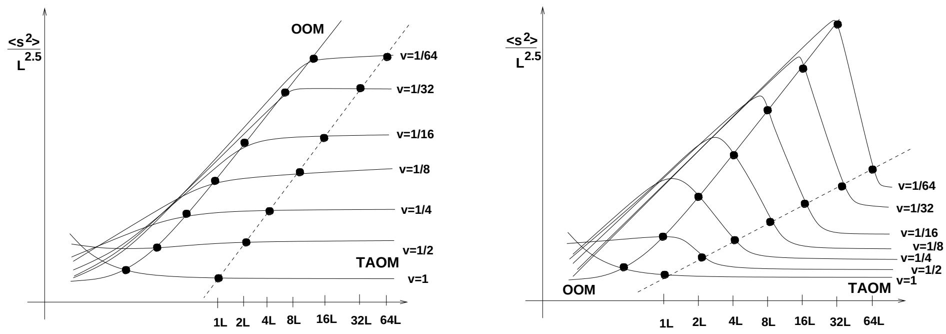
Local rearrangements, such as expansions (on the site losing a grain) and compressions (on the site receiving the grain) lead to an anisotropy not vanishing with L: Say a grain moves from i to $i+1$, then column i of height h_i expands by $\epsilon(h_i)$ and column $i+1$ is compressed by $\epsilon(h_{i+1})$. Then the changes of the slopes during toppling are $\Delta z_{i-1} = 1 - \epsilon(h_i)$, $\Delta z_i = -2 + \epsilon(h_i) + \epsilon(h_{i+1})$ and $\Delta z_{i+1} = 1 - \epsilon(h_{i+1})$. Assuming that the columns behave elastically, ϵ would be an increasing function of h, resulting in a net flow to the right, *i.e.* $\mathsf{v} > 0$. However, it remains unclear whether any of these effects can be seen in experimental systems.

Open problem

7.6.1 Boundary conditions

One possible problem why the considerations above are actually inapplicable are the boundary conditions. Opposite to the model introduced in Sec. 7.1.1, page 280, in the original model with original boundary conditions (Sec. 5.1.1, page 205), the right hand boundary was *closed* to slope units. This boundary condition was supposedly modelling the experiment — while one might actually argue that equally well the boundary condition could be to have vanishing slope at the right end, so that the pile smoothly runs out. However, accepting that a closed right boundary approximates the experiment better, one must ask: Is it possible to have an anisotropic redistribu-

[14]There are some problems in their derivation: The scaling ansatz is unphysical, as higher moments diverge even in finite systems. The result $\beta = \nu$ is due to a neglect of the lower cutoff. Moreover, the possible interpretation of $\alpha \approx 2.02$ as τ leads to a system size independent, finite first moment, which contradicts $\langle s \rangle \propto L$. In this light it would be very interesting to repeat the data analysis. It is remarkable, however, that a similar behaviour and exponent ($\alpha \approx 2.22$) has been observed in (Christensen *et al.*, 1996).



(a) Imaginative data close to the actual measurements, Fig. 7.2.

(b) Alternative imaginative data.

Figure 7.16: Cartoons of possible scenarios for measured values of $\langle s^2 \rangle$ as v changes with L; the measurement does not follows on solid line, which corresponds to a specific v, but keeps jumping tracks. The black blobs connected by the dashed line mark the exponent measured. The series of black circles towards the left marks the crossover.

tion of slope units in the bulk, with a net drift to the right, if the right most boundary is closed? Is it possible to have any net flux through the system?

In order to answer the question, it is reasonable to study the boundary driven BTW model with the original boundary conditions first. The difference to the Oslo model is only that $z_i^c \equiv 2$, so that the model develops into a unique (absorbing) state of the stationary distribution, which is $z_i \equiv z_i^c$. A four-site version of this model is shown in Fig. 7.17, where the movement of slope units is depicted by arrows. In fact, the rightmost and the leftmost sites do not relax isotropically, while the bulk does. The net-flux of units through the lattice vanishes, because each arrow pointing to the right has a corresponding arrow pointing to the left; what is really transported through the system is a configuration that allows the rightmost site to relax anisotropically.

Fig. 7.18 shows that it is possible to have anisotropic relaxations in the bulk even with a closed right hand boundary (again $z_i^c \equiv 2$). This is possible, provided that the rightmost site compensates the flux produced by its left neighbour. Again, all fluxes cancel and the net-flux vanishes. Moreover, the total height, shown as $\sum_i h_i$ increases due to the anisotropic relaxation between the third and the fourth row, for example, because a particle gets erected, see Fig. 7.15(a). If that could happen several times,[15] more and more height units would have to leave the system through the right boundary. That raises the question, whether the avalanche size should be measured at total height transported through the system, which increases with erected grains,

[15]This does not make much sense in case of reorientation like those shown in Fig. 7.15.

	z_1	z_2	z_3	z_4	h_1	h_2	h_3	h_4	$\sum_i h_i$
	2	2	2	2	8	6	4	2	20
kick!									
	3	2	2	2	9	6	4	2	21
	1	3	2	2	8	7	4	2	21
	2	1	3	2	8	6	5	2	21
	2	2	1	3	8	6	4	3	21
	2	2	2	2	8	6	4	2	20

Figure 7.17: Relaxation events in the BTW model (to be read from top to bottom) to illustrate anisotropic relaxations of the boundary sites. Solid arrows indicate the transport direction of slope units, dotted arrows indicate additions (external drive) and losses (relaxation of first site). The first part of the table lists the evolution of the local slopes z_i, the second part the heights (integrated slopes from right to left), and the third part the total number of grains in the system. The isotropic bulk relaxations only transport a configuration towards the right hand side, which allows the rightmost site to relax anisotropically. Every arrow pointing from site i to site $i \pm 1$ has a corresponding arrow pointing in the other direction, so that there is no net transport. Also, whenever a site in the bulk relaxes, it distributes equal amounts to both sides, so that this process is fully isotropic. Effectively every slope unit moved forward is moved backwards in the next time step.

or as number of topplings.

Open problem

The fact that the rightmost site produces a flux to the left large enough to compensate any incoming flow also implies that there is no net-flux through the lattice.[16] If there was any net flux, the system what simply jam, see page 242. Does that mean that there is no effective anisotropy possible with these boundary conditions? This question cannot be conclusively answered. However, it points once more to the task to repeat the study of the qEW with a Neumann condition. Moreover, having shown that anisotropy is a relevant field, it would not be surprising to find other relevant fields which lead to yet another universality class.

[16]This compensation comes about naturally if the grain is transported to the right boundary leaves the system without changing orientation.

	z_1	z_2	z_3	z_4	h_1	h_2	h_3	h_4	$\sum_i h_i$
	2	2	2	2	8	6	4	2	20
kick!									
	3	2	2	2	9	6	4	2	21
	1	3	2	2	8	7	4	2	21
	2	0	4	2	8	6	6	2	22
	2	2	0	4	8	6	4	4	22
	2	2	2	2	8	6	4	2	20

Figure 7.18: Similar to Fig. 7.17 (redistribution of two units shown as a double arrow) this figure shows the relaxation of a four site system, now with an anisotropic toppling at the second site, between the third and the fourth line. This corresponds to an event as shown in Fig. 7.15(a), where the grain gets erected when toppling from site 2 to site 3 and topples down the pile in this orientation. When it arrives at the rightmost site it leaves the system. Again, there is no net-flux through the system, even though one of the bulk relaxations is anisotropic.

7.7 Outlook and Summary

At this point, this thesis branches into two remaining chapters:

Chapter 8 demonstrates the exact solution of the TAOM, which is one of the very few exact results in SOC, which do not require any scaling assumptions.

Chapter 9 presents the calculation of the roughness exponent in the thermal EDWARDS-WILKINSON equation with drift.

To illuminate the motivation for these two chapters, we want to discuss the effect of the noise in greater detail.

The solution of the TAOM is shown in detail in Chapter 8 and since the TAOM represents the entire universality class of the quenched EDWARDS-WILKINSON equation with drift, the question naturally arises, whether it is possible to infer any exponents of the qEW equation from the exponents of the qEW equation with drift. To answer this question, the effect of a drift term in the one-dimensional EDWARDS-WILKINSON equation with *thermal* noise is studied in Chapter 9.

Moreover, one might ask, how the drift actually makes the problem solvable. It might be a general mechanism, which could make other quenched LANGEVIN equations solvable as well.

First of all, one has to answer the question, why it is not possible to gauge the drift away by a change of coordinates. If h solves

$$\partial_t h = \mathsf{D}\partial_x^2 h - \mathsf{v}\partial_x h + \eta \tag{7.73}$$

then, it seems

$$h'(x,t) = h(x + \mathsf{v}t, t) \tag{7.74}$$

solves

$$\partial_t h' = \mathsf{D}\partial_x^2 h' + \eta \ . \tag{7.75}$$

Thus, the qEW equation with drift is essentially the qEW. The point is, however, that η in (7.73) is $\eta(x,h)$, while in (7.75) it is $\eta(x + \mathsf{v}t, h')$. So, the time appears in the noise. Now, one might argue, the noise is essentially thermal and therefore (7.75) is the usual EW equation with thermal noise, the exponent of which is $\chi = 1/2$, exactly as calculated above, Sec. 7.3, page 286. So, is it all trivial then?

Not quite. Accepting the trick above, one might now go ahead and determine the roughness exponent of the original qEW without drift, by changing into a moving frame, $h'(x,t) = h(x + \mathsf{v}t, t)$, so that the problem

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) + \eta(x,h) \tag{7.76}$$

becomes

$$\partial_t h'(x,t) = \mathsf{D}\partial_x^2 h'(x,t) + \mathsf{v}\partial_x h'(x,t) + \eta(x + \mathsf{v}t, h') \tag{7.77}$$

If this noise term is now to be interpreted as a thermal noise, the roughness exponent of the qEW would be the roughness exponent of the EW equation with drift,

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) - \mathsf{v}\partial_x h(x,t) + \eta(x,t) \ . \tag{7.78}$$

So, what is the roughness exponent of the EW equation with drift? This drift *can* in fact be gauged away by changing the coordinate system (into the comoving frame), again $h'(x,t) = h(x + \mathsf{v}t, t)$, because the resulting noise

$$\partial_t h'(x,t) = \mathsf{D}\partial_x^2 h'(x,t) + \eta(x + \mathsf{v}t, t) \tag{7.79}$$

has a correlator

$$\begin{aligned}\left\langle \eta(x + \mathsf{v}t, t)\eta(x' + \mathsf{v}t', t')\right\rangle = \delta(x - x' + \mathsf{v}(t - t'))\delta(t - t') = \\ \delta(x - x')\delta(t - t') = \left\langle \eta(x,t)\eta(x',t')\right\rangle\end{aligned} \tag{7.80}$$

provided the original noise correlator is (as usual) $\langle\eta(x,t)\eta(x',t')\rangle = \delta(x-x')\delta(t-t')$. Therefore, the problem with drift, Eq. (7.78), cannot be distinguished from the problem without drift Eq. (7.79) and the roughness exponents are the same. Physically, that makes sense because the roughness exponent should not depend on the coordinate system of the observer.

Thus, since the roughness exponent of the EW equation with drift is the roughness exponent of the EW equation without drift, this should also be the roughness exponent of the qEW equation. That is certainly not supported by numerics. Thus, the trick introduced above [as used in Eq. (7.77)], which changes a quenched noise essentially into a thermal noise, does not work, at least not in general. *The drift does not make a quenched noise thermal.*

All the discussion above does not apply to the case of fixed boundary conditions (FBC), because a transformation like (7.74) changes $h(x=0,t)\equiv 0$ to $h(x=-\mathsf{v}t,t)\equiv 0$. Thus, in order for h' to obey to FBC, one needs moving boundaries in h. However, periodic boundary conditions (PBC) are not affected by the transformation (7.74), so that the results above, do apply in that case: The roughness exponent of the (thermal) EW equation with or without drift is the same, provided that periodic boundary conditions apply.

Finally, standard arguments of dimensional analysis (Sec. 9.4.2, page 366) suggest that the convection term $\mathsf{v}\partial_x h$ is relevant compared to the diffusion term $\mathsf{D}\partial_x^2 h$. In that case, it seems, that quenched and thermal EDWARDS-WILKINSON should be similarly affected by adding a drift term, but of course, not leading to the same exponents.

In order to study the effect of a drift term, it seems therefore reasonable to turn to the thermal problem and try to answer some of the following questions:

- Is it possible to say something about the qEW equation from the qEW with drift? Unfortunately, this transpires not to be possible.
- Does a drift term make the thermal EW equation trivially solvable? That will turn out not to be so.
- Does the effect of the drift term vary for different boundary conditions? Yes, indeed.
- Is the drift term relevant? Yes, but so is the diffusion term.

These questions are addressed in Chapter 9 and especially in Sec. 9.6.1, page 378.

7.7.1 Remark: SOC and relevant variables

One might ask in which sense the Oslo model without drift can be called "self-organised", since it has in fact a relevant tuning parameter, namely the drift v. Even though it must be tuned to a trivial value, namely to vanish[17], it still needs to be tuned. The most reasonable answer to this criticism is that the anisotropy is a *local* parameter, characterising *local dynamics*. This is different from a temperature-like parameter, which represents an energy scale of the *global* energy.[18]

7.7.2 Summary

The sections in this chapter can be summarised as follows:

- First, the anisotropic Oslo model, AOM, is defined, see Sec. 7.1.
- In the second section, Sec. 7.2, it is demonstrated that for any amount of anisotropy, the exponents of the original Oslo model change and are given by simple rational numbers which can all be obtained exactly, see Chapter 8. The crossover is studied numerically using a moment analysis, Eq. (7.3), and universal amplitude ratios, Eq. (7.4), and the crossover length is determined. The generalised model described above, continuously connects the established original Oslo model and an exactly solvable, directed variant. This variant has, compared to the original Oslo model, an enormous basin of attraction, so that the latter may be regarded a special case of the former.
- In Sec. 7.2.2 numerical results for alternative models are presented, which show the same crossover behaviour.
- Most remarkably, we find a change in critical behaviour of an SOC model, genuinely due to anisotropy, rather than stochasticity or the presence of multiple topplings.
- As illustrated in Sec. 7.3, the results are theoretically interesting because of their relation to the EW equation, the roughness exponent of which has been obtained in case of the presence of a drift term in one dimension to be $\chi = 1/2$.
- It is possible to derive from the qEW equation an exact expression for the toppling frequency in the model (Sec. 7.3.5) as well as for the average avalanche size. The latter can also be achieved by considering the slope units in the model as individually diffusing particles, Sec. 7.4.

[17] One might argue that an offset on v would make its "critical value" non-trivial. However, the important point is that v does not represent an independent scale in the analysis of the LANGEVIN equation. This remains true even with an offset, because then the anisotropy only enters together with this offset.

[18] This argument does not hold against percolation, which can be realised by local probabilities alone. The global observables are very similar to those in SOC. In fact, the most basic reason why percolation cannot be called SOC is because it is an equilibrium model.

- A presentation of some of the difficulties in identifying a universal crossover function (which characterises the change in scaling of the avalanche size distribution) has been attempted in Sec. 7.5.

- Moreover, according to our study, experiments are seriously complicated due to a coupling between system size and effective anisotropy. That might provide a clue as to the apparent difficulties in finding theoretically predicted exponents in the real world. However, there are complications in identifying the correct boundary conditions in experiments and different boundary conditions might lead to completely different behaviour, as discussed in Sec. 7.6.

Chapter 8

Exact Solution of the Totally Asymmetric Oslo Model

It has been found that a totally asymmetric variant of the Oslo model (Christensen et al.*, 1996) represents the entire universality class of the Oslo model with anisotropy [Sec. 7.1, page 279, and (Pruessner and Jensen, 2003a)]. The totally asymmetric model can be solved without scaling assumptions by finding the eigenvectors of the* MARKOV *matrix recursively, which can then be suitably modified to produce the moment generating function of the relevant observable. This method, the details of which are presented in this chapter, should be applicable to many other stochastic processes. The notation, as well as the general concept, is closely related to the operator approach introduced for the Oslo model in Sec. 5.3, page 225.*

8.1 Introduction

Self-organised criticality (SOC) was originally introduced (Bak *et al.*, 1987) as an approach to understand $1/f$-noise as well as the apparent abundance of power laws in nature, which is generally accepted as the sign of scale-invariance. The idea is that under very general circumstances driven stochastic processes develop into a scale-invariant state without the explicit tuning of parameters, contrary to what one would expect from equilibrium critical phenomena (Stanley, 1971).

A very large zoo of SOC models has been developed (Jensen, 1998), with each model having certain special features. However, based on large scale numerical simulations it has become increasingly clear that many of the models formerly thought of as representatives of entire universality classes or even paradigms for a specific type of model, are either not scale-invariant or at least do not follow simple scaling [Chapter 4 and (Boulter and Miller, 2003; Grassberger, 2002; Pruessner and Jensen, 2002a; Datta *et al.*, 2000)]. In fact, models of SOC are notorious for slow convergence and deviations from the expected behaviour.

Some models, however, show all features one would expect from a "self-organised critical" system: Consistent exponents and scaling, universality, crossover between different classes etc. One of these models is the so-called Oslo model (Christensen *et al.*, 1996), which was motivated by an experiment (Frette *et al.*, 1996). In the preceeding chapter, it has been shown that any (small) amount of anisotropy will drive this model eventually (in the thermodynamic limit) towards another "fixed point", which is represented by the "totally asymmetric Oslo model" (TAOM). As shown here, this model is solvable directly on the lattice without making any scaling assumptions. Consequently, it is not only possible to derive exponents, but also to calculate amplitudes of the moments of the relevant observable.

The TAOM is totally asymmetric in the sense that particles can move in one direction only, similar to the totally asymmetric exclusion process (TASEP) (Derrida *et al.*, 1992, 1993). The TASEP has been solved using a matrix product state ansatz (Derrida *et al.*, 1993; Derrida and Evans, 1997), so that it seems reasonable to apply similar techniques to the present model. However, there is a crucial difference between these two stochastic processes: The relevant observables in the TASEP exist on a microscopic timescale, *i.e.* there is an intrinsic timescale in the time-evolution of the microstate of the system. In contrast, in the TAOM the relevant observables are obtained by any dynamics which comply with a certain set of rules. In that sense, the specific (microscopic) dynamics of the TAOM are irrelevant. This is reflected in its theoretical treatment, in that the TASEP is updated homogeneously (all sites evolve equally) but the TAOM is perturbed once and is only observed after it is fully relaxed (separation of time-scales).

In the following, the model is defined in terms of rules on a lattice. Using a MARKOV matrix approach it is then solved and exponents and amplitudes derived. After mapping it to a reaction-diffusion process as well as various other processes, a more accessible continuum theory is described.

8.1.1 The model

The model is defined on a one-dimensional lattice of size L, where each site $i = 1, 2, \dots, L$ has assigned a slope $z_i \in \{1, 2\}$ and a critical slope $z_i^c \in \{1, 2\}$. From a flat initial configuration $z_i \equiv 1$ and z_i^c random, where $z_i^c = 1$ is chosen with probability p and $z_i^c = 2$ otherwise, the model evolves according to the following rules:

1. (Driving) Increase z_1 by one unit ("initial kick").

2. (Toppling) If there is an i where $z_i > z_i^c$, decrease z_i by 1 unit, $z_i \to z_i - 1$, and increase the right nearest neighbour $j = i + 1$ by 1, $z_j \to z_j + 1$ (charging). A

new z_i^c is chosen at random from $\{1, 2\}$, where $z_i^c = 1$ is chosen with probability p and $z_i^c = 2$ otherwise.

3. Repeat the second step until $z_i \leq z_i^c$ ("stable") everywhere. Then proceed with the first step.

During toppling, the right neighbour is charged of course only if it actually exists, *i.e.* $j \leq L$, otherwise the toppling site i relaxes without charging another site, so that a unit leaves the system. Apart from this boundary condition, the TAOM differs from the original Oslo model (Christensen *et al.*, 1996) in redistributing only a unit to the right, rather than one to each side.

It is important to note that the value of z_i^c is determined only *after* a site has discharged. Thus, if a *stable* site i is in state $z_i = 1$, its value of z_i^c could be randomly chosen in the moment when it is needed, *i.e.* when the site is charged again. If a *stable* site i is in state $z_i = 2$, then z_i^c has necessarily the value 2. When this site is charged, it will relax to $z_i = 2$ again and a new random z_i^c is drawn. If that is $z_i^c = 1$, then the site topples again and ends up in state $z_i = 1$, otherwise it remains in state $z_i = 2$.

If all sites are stable, *i.e.* $z_i \leq z_i^c$ for all $i \in [1, L]$, a configuration is fully described by the values of the z_i alone; if $z_i = 2$ then $z_i^c = 2$, otherwise z_i^c is random and has not yet been used in the dynamics.

The number of times the second rule is applied, that is the number of topplings, is the avalanche size s. The fundamental observable one is interested in is the probability density function of these sizes, $P(s)$, which is expected to obey simple scaling (see Sec. 2.3, page 78) above a fixed lower cutoff s_l,

$$P(s) = as^{-\tau}\mathcal{G}(s/s_0) , \tag{8.1}$$

with $s_0 = bL^D$, $\mathcal{G}$ the universal finite-size scaling function, and metric factors a and b (Privman and Fisher, 1984). Eq. (8.1) is the *definition* of the two exponents τ and D. It entails that the moments $\langle s^n \rangle$ of $P(s)$ behave like (Eq. (2.57), page 88)

$$\langle s^n \rangle = a(bL^D)^{1+n-\tau} g_n \text{ for } 1 + n - \tau > 0 \tag{8.2}$$

with universal amplitudes g_n (Sec. 2.3.3.1, page 90, and Sec. 2.4.5, page 102). Thus, assuming (8.1) one can derive τ and D from the behaviour of any two moments. Below, the exponents γ_n from $\langle s^n \rangle \propto L^{\gamma_n}$ will be used; Eq. (8.2) therefore means

$$\gamma_n = D(1 + n - \tau) . \tag{8.3}$$

The time series of avalanches, $s(t)$, itself is not Markovian,[1] while the sequence of stable configurations of the system, given by the vector $(z_1, z_2, \cdots, z_L)$, is. Since two consecutive stable configurations are not necessarily linked by a unique sequence of topplings (however, see Sec. 8.3, page 331, especially at the end, page 333), the sequence of avalanche sizes is not uniquely determined by the sequence of configurations. Nevertheless, in the form of a generating function this ambiguity can be built into the MARKOV matrix operating on the distribution vector of configurations, so that the avalanche size distribution can be determined by means of this specially prepared MARKOV matrix.

8.1.2 Abelian property

Put simply, if a model is Abelian (Dhar, 1999c), it means that the order of updates is irrelevant for its statistical properties. It is exceptionally simple to see this property here: Firstly, for an individual site there is no difference between a certain number of charges arriving at once or arriving sequentially. Secondly, if a site topples, it pours particles on its right neighbour, but it will never receive anything back from the neighbour. So, if a site at $z = 1$ has received 3 units, it topples at least twice, but for this site it does not make any difference whether it first moves one unit over to the right neighbour and waits until all sites to its right have relaxed, or whether it moves all units at once, 2 with probability $q \equiv 1 - p$ (namely the probability to have $z_i^c = 2$ after the second toppling) and 3 with probability p.

In this informal sense, the Abelian property allows the updating to run from left to right, completely relaxing each site during a sweep. If there is no toppling on site i, the avalanche has stopped and sites $j > i$ do not need to be checked for the toppling condition $z_j > z_j^c$ at all. This procedure makes very efficient Monte Carlo simulations possible. Moreover, it defines an activity a_i, which is the total number of charges received at site i during an avalanche. The activity will be used in Sec. 8.3.

8.2 MARKOV Matrix Approach

The tensor product $\otimes$ used here is explained in detail in (Hinrichsen, 2000). In particular it has the property (provided that A, B, A' and B' have appropriate ranks)

$$(A \otimes B) \odot (A' \otimes B') = (A \odot A') \otimes (B \odot B') \tag{8.4}$$

[1] Because the conditional probabilities to obtain avalanche sizes s_i at times t_i do not obey the MARKOV property

$$P(s_n, t_n | s_1, t_1; s_2, t_2; \ldots; s_{n-1}, t_{n-1}) = P(s_n, t_n | s_{n-1}, t_{n-1}) \, ,$$

where $t_1 < t_2 < \cdots < t_n$, see (van Kampen, 1992).

where $\odot$ stands for the appropriate operator: it is a matrix multiplication if A, B, A' and B' are matrices, it is a multiplication of a matrix and a vector if A and B are matrices and A' and B' are vectors or vice versa, or it is an inner product if they are all vectors. In particular, in the latter case it is

$$(a \otimes b)(a' \otimes b') = (aa')(bb') \ . \tag{8.5}$$

First, we consider a single site system, which can be in exactly two states, so that its distribution of states can be represented by a two-row vector. By convention, the upper row corresponds to $z = 1$ and the lower row to $z = 2$. Three matrices are introduced, corresponding to the three possible outcomes of a single initial charge.

The matrix S corresponds to a unit being absorbed, *i.e.* the site is in state $z_1 = 1$ and $z_1^c = 2$, which occurs with probability q. After the charge, the system is in state $z_i = 2$. Similarly, T corresponds to a single toppling due to the charge and U corresponds to a double toppling:

$$\mathsf{S} = \begin{pmatrix} 0 & 0 \\ q & 0 \end{pmatrix} \qquad \mathsf{T} = \begin{pmatrix} p & 0 \\ 0 & q \end{pmatrix} \qquad \mathsf{U} = \begin{pmatrix} 0 & p \\ 0 & 0 \end{pmatrix} \tag{8.6}$$

In the following, the aim is to find an expression for the moment generating function of the avalanche size distribution. To this end, each matrix is multiplied by an appropriate power of x, so that evaluating at $x = 1$ gives the usual MARKOV matrix of this process, and deriving by x before evaluating at $x = 1$ multiplies each process by the number of topplings occurring in it, and similarly for higher order moments (van Kampen, 1992).

It will be motivated only *a posteriori* that a dissipation process is required, say with probability $0 \leq \epsilon \leq 1$; this process corresponds to charging without changing the state, *i.e.* an identity operation $\mathbf{1}$, the latter being the 2×2 identity matrix. The resulting single site operator is therefore

$$\mathbf{O}_1(x) = \epsilon \mathbf{1} + \delta \left(\mathsf{S} + x\mathsf{T} + x^2 \mathsf{U} \right) = \begin{pmatrix} \epsilon + x\delta p & x^2 \delta p \\ \delta q & \epsilon + x \delta q \end{pmatrix} \tag{8.7}$$

with $\delta \equiv 1 - \epsilon$. The eigenvectors and eigenvalues of this matrix are found to be

$$\langle e_\lambda(x)| = \left(\frac{1}{x}, 1 \right) \qquad |e_\lambda(x)\rangle = \begin{pmatrix} xp \\ q \end{pmatrix} \qquad \lambda = \epsilon + x\delta \tag{8.8a}$$

$$\langle e_\mu(x)| = \left(-\frac{q}{x}, p \right) \qquad |e_\mu(x)\rangle = \begin{pmatrix} -x \\ 1 \end{pmatrix} \qquad \mu = \epsilon \tag{8.8b}$$

where $\mathbf{O}_1$ acts on bra-vectors $\langle|$ from the right and on ket-vectors $|\rangle$ from the left. The vectors are normalised such that

$$\langle e_a | e_b \rangle = \delta_{a,b} \tag{8.9}$$

with $\delta_{a,b}$ denoting the KRONECKER delta-function.[2] In order to distinguish vectors of different size, in the following they are often marked with an index L to indicate a size 2^L.

$\mathbf{O}_L(x)$ is the operator which adds a unit on site $i = 1$ and relaxes the entire lattice of size L. It is a matrix of size $2^L \times 2^L$ and defined as

$$\mathbf{O}_L(x) = \epsilon \mathbf{1}^{\otimes L} + \delta \left(\mathsf{S} \otimes \mathbf{1}^{\otimes (L-1)} + x\mathsf{T} \otimes \mathbf{O}_{L-1}(x) + x^2 \mathsf{U} \otimes \mathbf{O}^2_{L-1}(x) \right) \tag{8.10}$$

again with a dissipation rate ϵ, leaving the state unchanged. The bracket multiplied by δ consists of three terms: The first term charges the site without toppling and leaves the rest of the system unchanged by operating with the identity $\mathbf{1}^{\otimes (L-1)}$. The second term corresponds to a single toppling, which charges the remaining system of size $L-1$ once. This term is derived using the identity

$$\left(\mathsf{T} \otimes \mathbf{1}^{\otimes (L-1)} \right) (\mathbf{1} \otimes \mathbf{O}_{L-1}(x)) = \mathsf{T} \otimes \mathbf{O}_{L-1}(x) \ . \tag{8.11}$$

The third term is a double toppling of the site, giving rise to a double charge of the remaining system.

The Abelian property mentioned above (Sec. 8.1.2) can be expressed as the commutator for two charges on a system of size L, one at site $i = 1$, the other one at site $1 + L - L'$ with L' being the size of the subsystem starting from the site receiving the second charge,

$$\mathbf{O}_L(x) \left(\mathbf{1}^{\otimes (L-L')} \otimes \mathbf{O}_{L'}(x) \right) = \left(\mathbf{1}^{\otimes (L-L')} \otimes \mathbf{O}_{L'}(x) \right) \mathbf{O}_L(x) \tag{8.12}$$

where of course $L \geq L'$. The tensor multiplication used on $\mathbf{O}_{L'}$ and also in (8.11) ensures that both matrices have the same rank; they are "filled with identity" where they do not act. Eq. (8.12) simply states that it does not matter for the statistics whether the leftmost site of a right subsystem of size L' in a system of size L is charged first, followed by the leftmost site of the entire system, or vice versa. Due to the asymmetry in the dynamics, it is clear that a system of size L, initially charged at site i, has the same statistics as a system of size $L - i + 1$, charged at its leftmost site. It might be interesting, however, to formally prove Eq. (8.12), which should be easily feasible using established methods (Meester *et al.*, 2001; Dhar, 1999c).

[2] Of course, on the KRONECKER function the a and b are meant to be symbols rather than values, *i.e.* $\delta_{\mu,\lambda} = 0$, even if $x = 0$ so that $\mu = \lambda = \epsilon$.

The distribution of states at time t is the vector $|P_t\rangle_L$, which has rank 2^L, each row corresponding to the probability for the system to be in the state encoded by that row. The encoding follows from the row ordering convention introduced above and the use of the tensor product in (8.10). For example, for $L = 2$, the four rows encode the following four states:

$$\begin{pmatrix} 11 \\ 12 \\ 21 \\ 22 \end{pmatrix} . \tag{8.13}$$

For $x = 1$ the operator $\mathbf{O}_L(x)$ is simply the MARKOV matrix acting on $|P_t\rangle_L$, producing the distribution of states at time $t + 1$ (van Kampen, 1992)

$$|P_{t+1}\rangle_L = \mathbf{O}_L(1)\,|P_t\rangle_L \ . \tag{8.14}$$

There exists at least one eigenvector with eigenvalue 1, which is therefore a stationary distribution, see Sec. 5.3, page 225. If the eigenvectors represent a complete basis and the modulus of all other eigenvalues is less than unity, this stationary distribution is unique and reached by *any* initial distribution. The stationary distribution, denoted $|0\rangle_L$, is the focus of the following calculations. It is shown below that it is unique.

One very important bra-eigenvector with eigenvalue 1 of $\mathbf{O}_L(x = 1)$ is

$$\langle 0|_L \equiv (\underbrace{1, 1, \ldots, 1}_{2^L \text{ times}}) \tag{8.15}$$

by normalisation. As has been indicated above, for general x, the operator $\mathbf{O}_L(x)$ becomes a moment generating function of the avalanche size, if sandwiched between $\langle 0|_L$ and the stationary distribution:

$$Q_{L,n}(x;\epsilon) \equiv \langle 0|_L\, \mathbf{O}_L^n(x)\, |0\rangle_L \tag{8.16}$$

This can be seen from (8.10) containing an x for every toppling. When the operator acts on a distribution, for each transition from one state to another a power of x corresponding to the number of topplings enters and is multiplied by the probability to be in the initial state (given by the initial distribution) and the transition probability given by the transition matrix. The function $Q_{L,n}(x;\epsilon)$ for general n is then the generating function for avalanches caused by $n = 1, 2, \ldots$ initial kicks. In particular

$$\langle s^m\rangle_L = \left(x\frac{d}{dx}\right)^m \bigg|_{x=1} Q_{L,1}(x;\epsilon) \ . \tag{8.17}$$

Moreover, $Q_{L,1}(x;\epsilon)$ is also the probability generating function

$$P(k) = \frac{1}{k!} \left. \frac{d^k}{dx^k} \right|_{x=0} Q_{L,1}(x;\epsilon) \ , \tag{8.18}$$

but this property is not used in the following.

The aim of the following calculations is to find the generating function $Q_{L,1}(x;\epsilon)$ or at least the moments generated by it.

8.2.1 General eigenvectors and eigenvalues of $\mathbf{O}_L(x)$

It would be very helpful if $\mathbf{O}_L(x)$ could be written in the form

$$\mathbf{O}_L(x) = \sum_{i=0}^{2^L-1} |i(x)\rangle_L \, \lambda_{L,i} \, \langle i(x)|_L \ , \tag{8.19}$$

where $\langle i(x)|_L$ denote the left hand and $|i(x)\rangle_L$ the right hand eigenvectors of $\mathbf{O}_L(x)$ with eigenvalues $\lambda_{i,L}(x)$ and $i = 0 \dots 2^L - 1$. *A priori* it is not clear whether these vectors actually exist. In the following they are constructed and it is shown that setting $\epsilon = 0$ leads to fundamental problems.

Assuming that $|i\rangle_{L-1}$ is an eigenvector with eigenvalue $\lambda_{L-1,i}$ of $\mathbf{O}_{L-1}(x)$, the definition of $\mathbf{O}_L(x)$, Eq. (8.10), gives for an arbitrary vector $|e\rangle_1$

$$\mathbf{O}_L(x)\left(|e\rangle_1 \otimes |i\rangle_{L-1}\right) = \left\{ \left[\epsilon \mathbf{1} + \delta \left(\mathsf{S} + x\mathsf{T}\lambda_{L-1,i} + x^2 \mathsf{U}\lambda_{L-1,i}^2 \right) \right] |e\rangle_1 \right\} \otimes |i\rangle_{L-1} \tag{8.20}$$

where $|e\rangle_1$ contains two elements such that $|e\rangle_1 \otimes |i\rangle_{L-1}$ is a vector of 2^L elements. The matrix in the curly brackets is simply $\mathbf{O}_1(x\lambda_{L-1,i})$. So, if $|e\rangle_1$ is either $|e_\lambda(x\lambda_{L-1,i})\rangle$ or $|e_\mu(x\lambda_{L-1,i})\rangle$ from (8.8), then $|e\rangle_1 \otimes |i\rangle_{L-1}$ is an eigenvector of $\mathbf{O}_L(x)$ with eigenvalues $\epsilon + \delta(x\lambda_{L-1,i})$ or ϵ. Thus, based on (8.8), one can write the eigenvectors of $\mathbf{O}_L(x)$ as

$$|i\rangle_L = |e_\lambda(x\lambda_{L-1,i})\rangle \otimes |i\rangle_{L-1} \tag{8.21a}$$
$$\langle i|_L = \langle e_\lambda(x\lambda_{L-1,i})| \otimes \langle i|_{L-1} \tag{8.21b}$$
$$\left|i + 2^{L-1}\right\rangle_L = |e_\mu(x\lambda_{L-1,i})\rangle \otimes |i\rangle_{L-1} \tag{8.21c}$$
$$\left\langle i + 2^{L-1}\right|_L = \langle e_\mu(x\lambda_{L-1,i})| \otimes \langle i|_{L-1} \tag{8.21d}$$

and the eigenvalues as

$$\lambda_{L,i} = \epsilon + x\delta\lambda_{L-1,i} \tag{8.22a}$$
$$\lambda_{L,i+2^{L-1}} = \epsilon \tag{8.22b}$$

both with $i = 0, 1, \ldots, 2^{L-1} - 1$. To start the hierarchy, one defines

$$|0\rangle_1 = |e_\lambda(x)\rangle \tag{8.23a}$$
$$\langle 0|_1 = \langle e_\lambda(x)| \tag{8.23b}$$
$$|1\rangle_1 = |e_\mu(x)\rangle \tag{8.23c}$$
$$\langle 1|_1 = \langle e_\mu(x)| \tag{8.23d}$$

and the eigenvalues as

$$\lambda_{1,0} = \epsilon + x\delta \tag{8.24a}$$
$$\lambda_{1,1} = \epsilon \tag{8.24b}$$

Now it is clear why the quantity ϵ was necessary: For $\epsilon = 0$ all but one eigenvalues vanish, which can be seen from the hierarchy of eigenvalues obtained by iterating (8.22). Therefore, if the vanishing eigenvalue of $L-1$ is plugged into $\langle e_\lambda|$ or $\langle e_\mu|$ according to (8.21), the result is undefined, as can be seen from (8.8), so that the bra-eigenvectors cease to exist.

The fact that all but one eigenvalues vanish for $\epsilon = 0$ is very deceptive. Assuming that any initial condition $|P\rangle$ can be written in terms of the eigenvectors of $\mathbf{O}_L(1)$, say $\sum a_i |i\rangle$, this suggests $\mathbf{O}_L(1)|P\rangle = |0\rangle$. This, however, is wrong, because for vanishing ϵ the operator $\mathbf{O}_L(x)$ cannot be written in the form (8.19) for $L > 1$. And it must be wrong, because, for example, kicking an empty system once will not make it produce the stationary distribution.

If the eigenvectors of $\mathbf{O}_{L-1}$ are linearly independent, then one can show, using the construction (8.21), that the eigenvectors of $\mathbf{O}_L$ are linearly independent as well, provided that $|e_\lambda(x\lambda_{L-1,i})\rangle$ and $|e_\mu(x\lambda_{L-1,i})\rangle$ are linearly independent. This is not the case for $\epsilon = 0$ (see the ket vectors in (8.8) with $x = 0$) and this is the basic reason why $\epsilon \neq 0$ is needed for the time being. However, for any $\epsilon \neq 0$ one can apparently construct a diagonalising matrix for $\mathbf{O}_L$. Thus, it can be written in the form (8.19). Especially, the eigenvectors have the property (by induction)

$$\langle i(x)|j(x)\rangle_L = \delta_{i,j} \tag{8.25}$$

and as all 2^L eigenvectors are linearly independent, they must span the whole space so that

$$\sum_{i=0}^{2^L-1} |i\rangle_L \langle i|_L = \mathbf{1} \; . \tag{8.26}$$

In the form (8.19) the operator can now be applied to a stationary distribution to give

$$Q_{L,n}(x;\epsilon) = \sum_{i=0}^{2^L-1} \langle 0|i(x)\rangle_L \, \lambda_{L,i}^n \, \langle i(x)|0\rangle_L \tag{8.27}$$

8.2.1.1 The stationary distribution

From (8.21) the stationary distribution can be derived immediately. It is the eigenvector with eigenvalue 1 of $\mathbf{O}_L(1)$. Setting $x = 1$ in (8.22) it is clear that $\lambda_{L,i} = 1$ requires a $\lambda_{L-1,j} = 1$, which, together with (8.24), gives the unique $\lambda_{L,0} = 1$ provided that $\epsilon < 1$. If $\epsilon = 1$, then *all* eigenvalues are 1, but still all eigenvectors are linearly independent and therefore span the entire space, so that *all* initial distributions are stationary. This is not surprising because $\epsilon = 1$ simply means that any added particle immediately dissipates from the system, so that adding a particle is in fact just the identity operation.

For $0 \leq \epsilon < 1$ the stationary distribution is unique[3] and all other eigenvalues have modulus less than 1. The eigenvector corresponding to eigenvalue $\lambda_{L,0} = 1$ is, according to (8.21),

$$\langle 0|_L = \langle e_\lambda(1)|^{\otimes L} = (1,1)^{\otimes L} \tag{8.28a}$$

$$|0\rangle_L = |e_\lambda(1)\rangle^{\otimes L} = \begin{pmatrix} p \\ q \end{pmatrix}^{\otimes L} \tag{8.28b}$$

which is consistent with the notation for the stationary distribution and the normalisation eigenvector introduced in (8.16) and (8.15). The last line, Eq. (8.28b), indicates that the stationary state is a product measure, *i.e.* a state at one site does not depend on the state on any other site. In fact the spatial correlation function of sites $\{i_1, i_2, \dots\}$ can easily be calculated by "dressing" the states of the sites by appropriate powers of a variable x_i, in order to obtain the generating function of the correlators. The function

$$C(x_1, x_2, \dots, x_L) = \langle 0|_L \begin{pmatrix} px_1 \\ qx_1^{-1} \end{pmatrix} \begin{pmatrix} px_2 \\ qx_2^{-1} \end{pmatrix} \cdots \begin{pmatrix} px_L \\ qx_L^{-1} \end{pmatrix} = \prod_i^L (px_i + qx_i^{-1}) \tag{8.29}$$

is the generating function of the state-correlators, where state 1 stands for $z = 1$ and state -1 for $z = 2$. The states have the useful property that the joint contribution of two sites is 1 if both sites are in the same state and -1 otherwise. The average state is

[3] for $\epsilon = 0$ the eigenvectors do not span the entire space, but still the eigenvector with eigenvalue 1 can be constructed by (8.21) and is unique.

obtained from

$$x_i \left. \frac{d}{dx_i} \right|_{x_1,\ldots,x_L=1} C(x_1, x_2, \ldots, x_L) = p - q \tag{8.30}$$

Correspondingly, the connected two point correlation function of sites i and j is given by

$$x_i \frac{d}{dx_i} x_j \left. \frac{d}{dx_j} \right|_{x_1,\ldots,x_L=1} C(x_1, x_2, \ldots, x_L) - (p-q)^2 = \begin{cases} 4pq & \text{for } i = j \quad (8.31a) \\ 0 & \text{otherwise} \quad (8.31b) \end{cases}$$

This confirms the absence of correlations and is fully consistent with the expected variance of the state.

8.2.2 The hierarchy of generating functions

Using (8.27), one can now calculate the generating function $Q_{L,n}(x;\epsilon)$, by plugging the hierarchy of eigenvectors (8.21) and eigenvalues (8.22) into (8.27) and using the properties of $\otimes$, see Eq. (8.4). For $n = 1$ it is

$$\begin{aligned} Q_{L,1}(x;\epsilon) &= \sum_{i=0}^{2^{L-1}-1} x\delta\lambda_{L-1,i} \left\langle 0|i \right\rangle_{L-1} \left\langle 0|e_\lambda(x\lambda_{L-1,i}) \right\rangle_1 \times \left\langle i|0 \right\rangle_{L-1} \left\langle e_\lambda(x\lambda_{L-1,i})|0 \right\rangle_1 \\ &\quad + \sum_{i=0}^{2^L-1} \epsilon \left\langle 0|i(x) \right\rangle_L \left\langle i(x)|0 \right\rangle_L \end{aligned}$$

where the term proportional to ϵ comes from the ϵ in every $\lambda_{L,i}$ (see Eq. (8.22)). From (8.26) it is clear that the last sum gives 1. The two projections give

$$\begin{aligned} \left\langle 0|e_\lambda(x\lambda_{L-1,i}) \right\rangle_1 &= x\lambda_{L-1,i}p + q \\ \left\langle e_\lambda(x\lambda_{L-1,i})|0 \right\rangle_1 &= \frac{p}{x\lambda_{L-1,i}} + q \end{aligned}$$

so that

$$\begin{aligned} Q_{L,1}(x;\epsilon) &= \epsilon + \sum_{i=0}^{2^{L-1}-1} x\delta\lambda_{L-1,i} \left\langle 0|i \right\rangle_{L-1} \left\langle i|0 \right\rangle_{L-1} \\ &\quad \times \left(p^2 + q^2 + pq \left(x\lambda_{L-1,i} + \frac{1}{x\lambda_{L-1,i}} \right) \right) \\ &= \epsilon + \delta\big((p^2+q^2) x Q_{L-1,1}(x;\epsilon) \\ &\quad + pq(x^2 Q_{L-1,2}(x;\epsilon) + Q_{L-1,0}(x;\epsilon)) \big) \end{aligned} \tag{8.32}$$

where Eq. (8.27) has been used in the last line. Of course, the generating function $Q_{L,n}(x;\epsilon)$ is *defined*, (8.16), for all ϵ and therefore one can take the limit $\epsilon \to 0$. This limit should not cause any problems, as ϵ has only been used to construct the eigenvectors. In fact, the limit must be identical to setting $\epsilon = 0$ in (8.32), as can be shown from (8.32) by induction in L.[4] This finally gives

$$Q_{L,1}(x;0) = (p^2+q^2)xQ_{L-1,1}(x;0) + pq(x^2Q_{L-1,2}(x;0) + Q_{L-1,0}(x;0)) \ , \tag{8.33}$$

where $Q_{L-1,0}(x;\epsilon) = 1$ by Eq. (8.16), consistent with Eq. (8.27). In fact, the calculation above can be generalised to $n \geq 1$:

$$\begin{aligned} Q_{L,n}(x;\epsilon) = & \sum_{i=0}^{2^{L-1}-1} \sum_{j=1}^{n} \binom{n}{j} \epsilon^{n-j} (x\delta\lambda_{L-1,i})^j \\ & \times \langle 0|i\rangle_{L-1} \langle 0|e_\lambda(x\lambda_{L-1,i})\rangle_1 \langle i|0\rangle_{L-1} \langle e_\lambda(x\lambda_{L-1,i})|0\rangle_1 \\ & + \sum_{i=0}^{2^L-1} \epsilon^n \langle 0|i(x)\rangle_L \langle i(x)|0\rangle_L \end{aligned} \tag{8.34}$$

Again, all sums can be written in terms of $Q_{L-1,n}(x;\epsilon)$ plus ϵ^n:

$$\begin{aligned} Q_{L,n}(x;\epsilon) \quad = \quad & \sum_{j=1}^{n} \binom{n}{j} \delta^j \epsilon^{n-j} \\ & \times \Big((p^2+q^2)x^j Q_{L-1,j}(x;\epsilon) + pqx^{j+1}Q_{L-1,j+1}(x;\epsilon) + pqx^{j-1}Q_{L-1,j-1}(x;\epsilon) \Big) \\ & + \epsilon^n \end{aligned} \tag{8.35}$$

For vanishing dissipation this simplifies to the central result

$$Q_{L+1,n}(x;0) = x^n \left(\tilde{D}Q_{L,n}(x;0) + D(xQ_{L,n+1}(x;0) + x^{-1}Q_{L,n-1}(x;0)) \right) \tag{8.36}$$

with $D = pq$ and $\tilde{D} = p^2 + q^2 = 1 - 2D$. Eq. (8.36) is closely related to a diffusion equation. The boundary conditions are $Q_{L,0}(x;\epsilon) \equiv 1$ for $L \geq 1$ as mentioned above and $Q_{L=0,n}(x;\epsilon) \equiv 1$. The latter comes from a direct evaluation of (8.27) for $L = 1$, which is identical to (8.35) for $Q_{L=0,n}(x;\epsilon) \equiv 1$.

8.2.3 Solving $Q_{L,n}$

There is no general solution for (8.36) known to the author. However, one can solve it order by order in derivatives by x at $x = 1$, *i.e.* calculate every individual moment,

[4] If $\lim_{\epsilon\to 0} Q_{L-1,j}(x;\epsilon) = Q_{L-1,j}(x;0)$ finite, then by (8.35) $\lim_{\epsilon\to 0} Q_{L,j}(x;\epsilon) = Q_{L,j}(x;0)$ finite for finite x; the initial condition of the induction is given by the boundary conditions mentioned after (8.36).

see (8.17). In the following, the notation

$$\begin{aligned} Q_{L,n} &= Q_{L,n}(1;0) && \text{(8.37a)} \\ Q'_{L,n} &= \left.\frac{d}{dx}\right|_{x=1} Q_{L,n}(x;0)\ , && \text{(8.37b)} \end{aligned}$$

etc. is used. One finds for $n \geq 1$

$$Q_{L+1,n} = \tilde{D}Q_{L,n} + D(Q_{L,n+1} + Q_{L,n-1}) \tag{8.38}$$

which is solved with the boundary conditions introduced above by $Q_{L,n} = 1$. Of course, this is just normalisation. Using this result the next derivative is

$$Q'_{L+1,n} = n + \tilde{D}Q'_{L,n} + D(Q'_{L,n+1} + Q'_{L,n-1}) \tag{8.39}$$

with the boundary conditions $Q'_{L,n=0} = 0$ and $Q'_{L=0,n} = 0$. The solution of (8.39) can easily be guessed as

$$Q'_{L,n} = nL\ . \tag{8.40}$$

This is not surprising, because it says that the average number of topplings occurring in the system per n kicks is nL. That is obviously true, because every unit added must leave the system by travelling through the entire lattice.

The next order is the first non-trivial one. The difference equation then reads

$$Q''_{L+1,n} = (n^2 + 2D)(2L+1) - n + \tilde{D}Q''_{L,n} + D(Q''_{L,n+1} + Q''_{L,n-1})\ . \tag{8.41}$$

The formal, general solution of the difference equation

$$Q''_{L+1,n} = g(L+1,n) + \tilde{D}Q''_{L,n} + D(Q''_{L,n+1} + Q''_{L,n-1}) \tag{8.42}$$

with a general source term $g(L,n)$ with $g(0,n) = 0$ is

$$Q''_{L,n} = \sum_{l=1}^{L} \sum_{m=-\infty}^{\infty} g(l,m)\phi_{L-l,n,m}\ , \tag{8.43}$$

where the propagator $\phi_{L,n,m}$ obeys

$$\phi_{L+1,n,m} = \tilde{D}\phi_{L,n,m} + D(\phi_{L,n+1,m} + \phi_{L,n-1,m})\ . \tag{8.44}$$

with $\phi_{L=0,n,m} = \delta_{n,m}$ and appropriate boundary conditions.

The source term $(n^2+2D)(2L+1)-n$ in (8.41) can be simplified without changing the homogenous boundary conditions, $Q''_{L,n=0} = 0$ and $Q''_{L=0,n} = 0$. One defines

$$Q''_{L+1,n} = S_{L+1,n} + \tilde{Q}''_{L+1,n} \tag{8.45}$$

with

$$S_{L,n} = \sum_{i=0}^{L-1} n^2 - n + 2in^2 = -nL + n^2L^2 \tag{8.46}$$

which has the useful property $S_{L+1,n} - S_{L,n} = -n + 2n^2L + n^2$ and $S_{L,n+1} + S_{L,n-1} = 2S_{L,n} + 2L^2$, so that

$$\tilde{Q}''_{L+1,n} = 2D(L+1)^2 + \tilde{D}\tilde{Q}''_{L,n} + D(\tilde{Q}''_{L,n+1} + \tilde{Q}''_{L,n-1}) \,. \tag{8.47}$$

with a source term, which is now independent of n. The solution is therefore

$$\tilde{Q}''_{L,n} = \sum_{l=1}^{L} 2Dl^2 \sum_{m=1}^{\infty} \phi_{L-l,n,m} \tag{8.48}$$

with the propagator $\phi_{L,n,m}$ obeying $\phi_{L,n=0,m} = 0$ and the sources only being placed at $m \geq 1$, as the whole problem lives only at $n \geq 0$ and for $n = 0$ the boundary condition $\tilde{Q}''_{L,n=0} = 0$ applies (and (8.47) does not). It is not even necessary to construct this propagator explicitly by a mirror charge trick. One simply discards those regions of sources, which are cancelled by the mirror charge, as shown in Fig. 8.1. Therefore

$$\tilde{Q}''_{L,n} = \sum_{l=1}^{L} 2Dl^2 \sum_{m=1}^{2n} \phi_{L-l,n,m} \,. \tag{8.49}$$

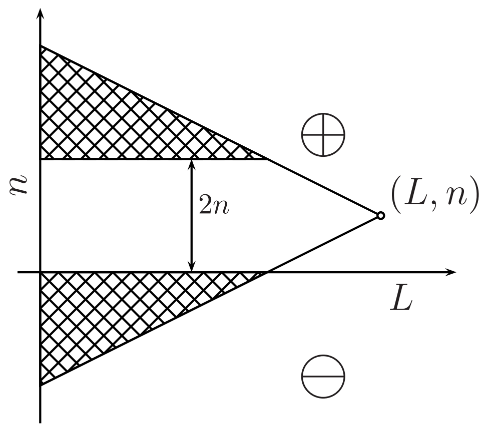


Figure 8.1: The "light cone" of the position marked as (L, n) contains positive and negative sources. However, the negative sources exactly cancel out some of the positive ones.

The propagator, now without the boundary condition $\phi_{L,n=0,m} = 0$, is translationally invariant, $\phi_{L,n,m} = \phi_{L,n-m,0}$. It can be constructed by FOURIER transforming (8.44) using $\tilde{D} = 1 - 2D$, so that

$$\tilde{\phi}_{L+1,k} = \tilde{\phi}_{L,k}(D(e^{ik} + e^{-ik} - 2) + 1) \tag{8.50}$$

where $\tilde{\phi}_{L,k} = \sum_{n=-\infty}^{\infty} \phi_{L,n,0} e^{-ink}$. Therefore

$$\tilde{\phi}_{L,k} = \left[D(e^{ik} + e^{-ik}) + \tilde{D}\right]^L = \sum_{m=0}^{L} \binom{L}{m} D^m \tilde{D}^{L-m} \sum_{l=0}^{m} \binom{m}{l} e^{ikl} e^{-ik(m-l)} \tag{8.51}$$

using $\tilde{\phi}_{L=0,k} = 1$ from $\phi_{L=0,n,0} = \delta_{n,0}$. Transforming back via $(1/(2\pi)) \int_{-\pi}^{\pi} dk e^{ikn}$ gives

$$\phi_{L,n,0} = \sum_{m=0}^{L} \binom{L}{m} D^m \tilde{D}^{L-m} \sum_{l=0}^{m} \binom{m}{l} \delta_{m-2l,n} = \sum_{m=0}^{L} \binom{L}{m} \binom{m}{\frac{m-n}{2}} D^m \tilde{D}^{L-m} \tag{8.52}$$

where the second binomial is defined as 0 for odd or negative differences $m - n$ and for $m - n > 2m$.

There is, however, another way to find the same propagator in a more useful form, namely by writing $\tilde{D} = p^2 + q^2$ and $D = pq$ again and therefore according to (8.50)

$$\tilde{\phi}_{L,k} = \left[(pe^{ik/2} + qe^{-ik/2})(pe^{-ik/2} + qe^{ik/2})\right]^L \tag{8.53}$$

which gives after a similar calculation as above

$$\phi_{L,n,0} = \sum_{m=0}^{L} \binom{L}{m} \binom{L}{m-n} p^{2L-2m+n} q^{2m-n} \tag{8.54}$$

For $p = q = 1/2$ this simplifies to the well known solution of the diffusion equation on the lattice of

$$\binom{2L}{L+n} 2^{-2L} \tag{8.55}$$

using "VANDERMODE's convolution" (Knuth, 1997)

$$\sum_{m=0}^{L} \binom{L}{m} \binom{L}{m-n} = \binom{2L}{L+n} . \tag{8.56}$$

Eq. (8.54) can now be plugged into (8.49) to give the full solution for $\tilde{Q}''_{L,n}$. For a single kick this gives

$$\tilde{Q}''_{L,1} = \sum_{l=1}^{L} 2Dl^2(\phi_{L-l,0,0} + \phi_{L-l,1,0}) \tag{8.57a}$$

$$= \sum_{l=0}^{L-1} 2D(L-l)^2 \sum_{m=0}^{l} \binom{l}{m} p^{2l-2m} q^{2m} \left(\binom{l}{m} + \binom{l}{m+1} \frac{q}{p} \right) \tag{8.57b}$$

In order to analyse the asymptotic behaviour for $L \to \infty$, one writes

$$\tilde{Q}''_{L,1} = \sum_{l=0}^{L-1} 2D(L-l)^2 \phi^*(l) \tag{8.58}$$

with

$$\phi^*(l) = \sum_{m=0}^{l} \binom{l}{m} p^{2(l-m)} q^{2m} \left(\binom{l}{m} + \binom{l}{m+1} \frac{q}{p} \right) . \tag{8.59}$$

The binomial can be approximated by

$$\binom{l}{m} p^{l-m} q^m \to \frac{1}{\sqrt{2\pi pql}} e^{-\frac{(m-lq)^2}{2pql}} \tag{8.60}$$

and the summation can be written as an integral:

$$\phi^*(l) \to \int_0^l dm \frac{1}{\pi pql} e^{-\frac{2(m-lq)^2}{2pql}} \tag{8.61}$$

$$= \frac{1}{2\sqrt{lpq\pi}} \left(\mathcal{E}(\sqrt{lp/q}) + \mathcal{E}(\sqrt{lq/p}) \right) \tag{8.62}$$

where $\mathcal{E}(x) \equiv 2\int_0^x dz \exp(-z^2)/\sqrt{\pi}$. The two binomials have been treated identically, so that finally

$$\tilde{Q}''_{L,1} \to \int_0^{L-1} 2D(L-l)^2 \frac{1}{2\sqrt{lpq\pi}} \left(\mathcal{E}(\sqrt{lp/q}) + \mathcal{E}(\sqrt{lq/p}) \right) \tag{8.63}$$

which, to leading order, turns out to be

$$\tilde{Q}''_{L,1} \to \frac{32}{15\sqrt{\pi}} \sqrt{pq} L^{5/2} . \tag{8.64}$$

This is perfectly confirmed by numerical simulations of the model.

The two exponents $\gamma_1 = 1$ (see Eq. (8.40)) and $\gamma_2 = 5/2$ (8.64) lead together with (8.3) to $\tau = 4/3$ and $D = 3/2$.

8.3 Reaction-Diffusion Mapping

It is possible to map the model onto a very simple reaction-diffusion process of the form $A + A \to A$ (ben-Avraham and Havlin, 2000). To this end, the configuration of the lattice is described by the thick line shown in Fig. 8.2. The line consists of segments, which can either point up or down by an angle of $45°$. If the line corresponding to the ith site goes up, it indicates that the ith site is in state $z = 1$, otherwise the line goes down indicating the state of the site to be $z = 2$. According to (8.28b) the configuration of the lattice after an avalanche is a product state, where a site is in state $z = 1$ with probability p and in state $z = 2$ with probability q. Thus, the thick line is in fact the trajectory of a random walker with drift corresponding to the difference $p - q$.

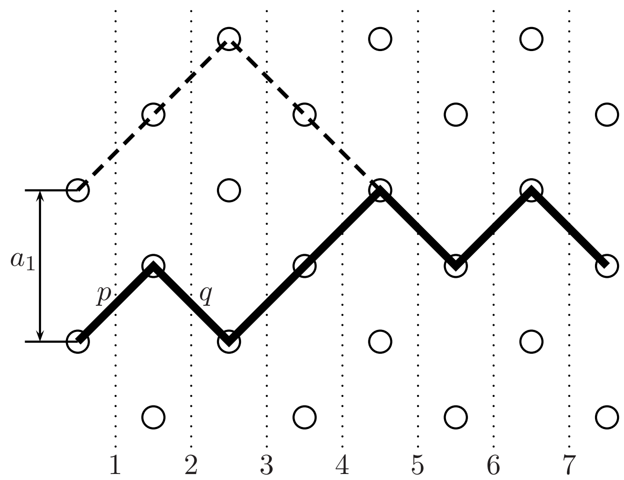


Figure 8.2: The thick, full line shows the configuration of the lattice after an avalanche has passed through. Each up or down-pointing segment corresponds to a single site, the position label of which is shown under the dotted line. A segment pointing upwards corresponds to a site being in state $z = 1$ (with probability p, see Eq. (8.28b)), a segment pointing downwards corresponds to state $z = 2$ (probability q), as indicated. The dashed line corresponds to a "toppling trajectory" as explained in the text.

The avalanche itself, on the other hand, is a random walk with the same probabilities. One can see that by considering the activity a_i, which is the number charges received at site i during an update-sweep as described in Sec. 8.1.2, page 318. The activity can either remain constant or change by 1 up or down. Apparently $a_1 = 1$ is the driving. If a site receives a_i charges and changes state by $\Delta z_i = z_i(t) - z_i(t+1)$, then its right neighbour receives $a_{i+1} = a_i + \Delta z_i$ charges. If $\Delta z_i = 0$, then the vertical distance between two consecutive configuration trajectories (as shown as thick and dashed lines in Fig. 8.2) does not change. If, however, the new configuration of site i has an increased value $z_i(t+1) > z_i(t)$, the activity goes down, because $\Delta z_i < 0$. The only way to increase z_i is to go from state 1 to state 2, *i.e.* the line segment of the former configuration points up (probability p), while the line segment of the new configuration points down (probability q), so that the gap between the two trajectory decreases, see Fig. 8.2 at the dotted line 3. Similarly, if the activity goes up, then $\Delta z_i > 0$ and the gap increases.

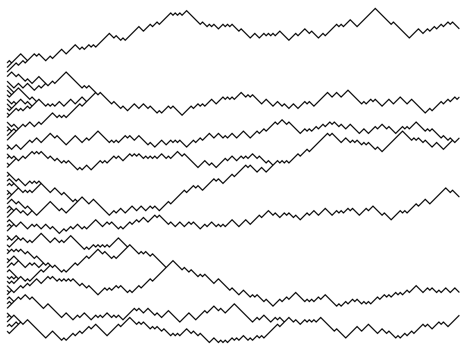
After the activity vanishes, the profile of the new configuration remains unchanged compared to the former, *i.e.* the gap between the two configurations is a constant. In fact, if the gap was initially 1 and goes up and down by 1 as described above, then the gap will be 0 as soon as the activity vanishes. This is exactly what is shown in Fig. 8.2: *The thick dashed line shows the new configuration and its distance to the old configuration is the activity during the avalanche.* This avalanche occurs within the configuration shown as a thick line, initiated by a single kick. Initially the gap is $a_1 = 1$. If the dashed line would go down immediately on site 1, the site would have "absorbed" the initial unit and would be in state $z = 2$ (*i.e.* a segment pointing down). Instead, in the example, it goes up twice; first just like in the old configuration so that the activity does not increase, and then in the opposite direction to the old configuration so that the activity increases by 1. On site 3 and 4 it goes down twice; the toppling on site 4 is particularly interesting. Here, initially the activity is 1, *i.e.* the site has received one unit. But the site is in state $z = 1$, so it absorbs the unit with probability q, corresponding to the probability of the dashed line segment to point downwards.

The activity is measured half a unit left of each site as the distance between old and new trajectories, which, in turn, is measured in such units that the vertical distance between two circles is 1. The reason for the shift is that one wants to measure how many charges have *arrived* at a site, not affected by the value of the resulting activity.

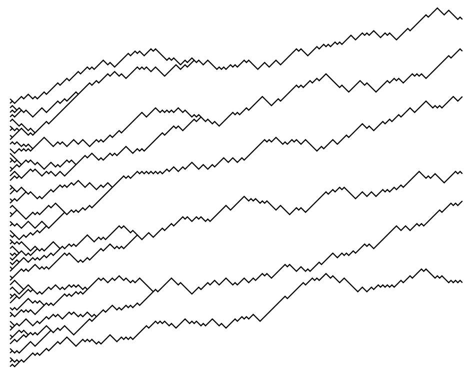
To repeat this important point, the trajectory of an avalanche becomes the configuration for the next avalanche, *i.e.* the thick dashed line in Fig. 8.2 becomes the thick solid line for the next avalanche.

One can calculate the probability of the changes of activity explicitly: The new segment goes up with probability p and down with probability q, the same applies to the old segment. Thus, they point in the same direction (no change of activity) with probability $p^2 + q^2$, the gap widens with probability pq and shrinks with qp. Hence, the gap between the two trajectories is in fact a symmetric random walk, even though the individual trajectories might have a bias, according to $p - q$. Fig. 8.3(a) shows a sequence of trajectories for $p = q = 1/2$ and Fig. 8.3(b) for $p = 0.6$.

As described above (see Sec. 8.1.1), the avalanche size is measured as the number of topplings. For convenience, one can define it as the number of charges, which makes hardly any difference, because the number of topplings of site i is identical to the number of charges on site $i + 1$, unless $i = L$, simply because there is no site



(a) $p = q = 1/2$



(b) $p = 0.6$

Figure 8.3: Examples for sequences of trajectories resulting from different values of p.

$L + 1$; similarly for $L = 1$. Moreover, the number of charges on site $i = 1$ is exactly 1 per avalanche, while there is only one toppling *on average* on site L. More explicitly, if $t_{i,i+1}$ denotes the number of units transported from site i to site $i + 1$ during an avalanche, the total number of topplings is

$$\sum_{i=1}^{L} t_{i,i+1} \tag{8.65}$$

where $t_{L,L+1}$ is the number of units leaving the system. If, correspondingly, $t_{0,1}$ is the initial kick, then

$$\sum_{i=1}^{L} t_{i-1,i} = t_{0,1} + \sum_{i=1}^{L} t_{i,i+1} - t_{L,L+1} \tag{8.66}$$

is the total number of charges. In the following, we will consider the number of charges as the avalanche size, because the total number of charges is simply the area between two of those trajectories described above, namely the sum over all activities. From this it is also clear that the avalanche size *is* actually uniquely determined by the initial and the final configuration, see Sec. 8.1.1, page 316, especially page 318.

8.3.1 Relation to other models

Before the above identification of the process as a random walker is cast into an continuum problem and subsequently solved, it is worth pointing out other models which are closely linked to the present one.

8.3.1.1 Anisotropic BTW model

DHAR and RAMASWAMY (Dhar and Ramaswamy, 1989) developed an anisotropic variant of the well-known BTW sandpile model (Bak *et al.*, 1987), which is now known as the directed sandpile model. This model, however, is situated on a $1+1$-dimensional lattice and the annihilating random walkers represent the contours of the compact area covered by an avalanche. The randomness here comes solely from the randomness of whether a site charged by particles from toppling sites topples in turn. An equivalence to a variant of directed percolation has been already pointed out in (Domany and Kinzel, 1984), see also (Mohanty and Dhar, 2002).

MORTON KLOSTER, SERGEI MASLOV and CHAO TANG (Kloster *et al.*, 2001) have studied a stochastic directed sandpile model, which was originally proposed by PASTOR-SATORRAS and VESPIGNANI (Pastor-Satorras and Vespignani, 2000b). This model is closely related to the one presented in this paper, even though it is also situated on a $1+1$-dimensional lattice. The authors find the same exponents by scaling arguments. The mapping to the two-dimensional reaction-diffusion process presented above, questions their assertion that their model is in a different universality class than the model by DHAR and RAMASWAMY.

For these models it is fairly obvious how to extend them systematically to higher dimensions. Using scaling arguments in conjunction with some simplifying assumptions, PACZUSKI and BASSLER (Paczuski and Bassler, 2000) arrive at a general expression for the value of the exponents of this model in higher dimensions. Unfortunately, it is not so clear how to generalise the model studied in this paper to higher dimensions, because it is unclear how to generalise the driving and what boundary conditions to apply.

8.3.2 Continuum solution

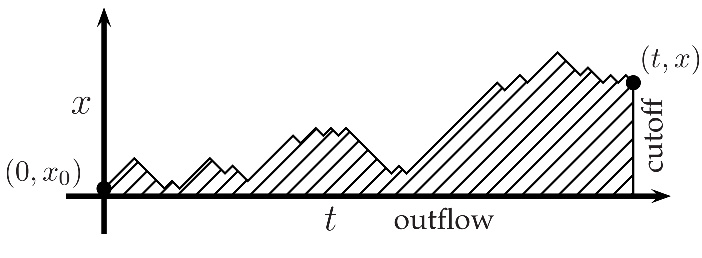


Figure 8.4: The area under the trajectory (hatched) is the avalanche size. The two filled circles mark the starting point $(0, x_0)$ and the end point (t, x)

Having mentioned already the mapping to an annihilating random walk, the continuum description is straight forward. To this end, the quantity $\psi_n(t, x; x_0)$ is introduced. It quantifies the properties of a random walker along an absorbing wall. For $n = 0$ it is the probability density of random walkers at time t and height x over the absorbing wall, starting at height x_0, which is $x_0 = 1$ for a single initial kick. Here, t takes on the rôle of the horizontal (continuous) position between $t = 0$ and $t = L$ in a picture like Fig. 8.2. To motivate the following calculation, one imagines a large set of trajectories of random walkers along the absorbing wall from $t = 0$, $x = x_0$ to t and x. The set of areas under the trajectories, as exemplified in Fig. 8.4, is then $\{s_i(t, x; x_0)\}$, where i is indexing the elements in the set. $\langle\{s^n(t, x; x_0)\}\rangle$ is the average of the nth moment over this set. Now one can express the time-evolution of this average as the sum of three contributions of the three processes of up, down or straight movement of the random walker. Thus, up to terms of order $\Delta t \Delta x$ (see caption of Fig. 8.5)

$$\psi_0(t + \Delta t, x; x_0)\langle\{s^n(t + \Delta t, x; x_0)\}\rangle$$

$$= pq\psi_0(t, x + \Delta x; x_0)\langle\{(s(t, x + \Delta x; x_0) + x\Delta t)^n\}\rangle \tag{8.67a}$$

$$+(p^2 + q^2)\psi_0(t, x; x_0)\langle\{(s(t, x; x_0) + x\Delta t)^n\}\rangle \tag{8.67b}$$

$$+pq\psi_0(t, x - \Delta x; x_0)\langle\{(s(t, x - \Delta x; x_0) + x\Delta t)^n\}\rangle \tag{8.67c}$$

where each term corresponds to a process like the one shown in Fig. 8.5. The multiplication by $\psi_0(t, x; x_0)$ is necessary in order to weight each of the ensembles for each contribution properly. For example, there might a much larger contribution from below, even though on average the moment at this position is smaller than at the other positions.

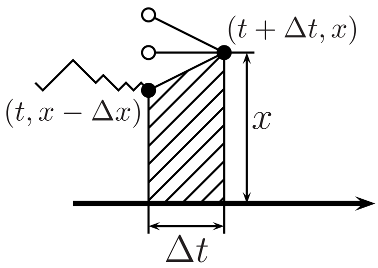


Figure 8.5: A new segment (hatched area) is added to the currently considered path, increasing all areas in the ensemble $\{s_i(t, x - \Delta x; x_0)\}$ by $x\Delta t + \mathcal{O}(\Delta t \Delta x)$ and producing a new ensemble $\{s_i(t + \Delta t, x; x_0)\}$. The example shown corresponds to (8.67c), which starts at $(t, x - \Delta x)$. Starting points of other contributions are shown as empty circles. The coordinates of the two black points are given in the form (t, x).

Defining

$$\psi_n(t, x; x_0) \equiv \psi_0(t, x; x_0)\langle\{s(t, x; x_0)^n\}\rangle \tag{8.68}$$

one finds in the continuum limit of (8.67) (keeping $D\Delta t/\Delta x^2$ constant)

$$\partial_t \psi_n(t,x;x_0) = D\partial_x^2 \psi_n(t,x;x_0) + xn\psi_{n-1}(t,x;x_0) \tag{8.69}$$

where $D = pq$ again[5]. The boundary conditions for $n = 0$ are observed immediately and transfered to ψ_n using (8.68) by noting that $\langle \{s(t,x;x_0)^n\}\rangle$ is non-divergent,[6] so

$$\lim_{t\to 0} \psi_n(t,x;x_0) = \delta_{n,0}\delta(x - x_0) \tag{8.70a}$$
$$\psi_n(t,0;x_0) = 0 \tag{8.70b}$$

and the PDE (8.69) is to be solved for $x \in [0, \infty[$.

The avalanche sizes are measured from avalanche trajectories which have died out or reached the end of the system. Thus, the averages measured in the model are taken from the random walkers which have reached the absorbing wall or did not do so until a cutoff time t. Therefore the nth moment observed is

$$\langle s^n\rangle(t;x_0) = \int_0^t dt' j_n(t';x_0) + \int_0^\infty dx' \psi_n(t,x';x_0) \tag{8.71}$$

where the first integral runs over the "outflow", $j_n(t, x = 0; x_0) \equiv D\partial_x|_{x=0}\psi_n(t,x;x_0)$ and the second over the contributions at cutoff time (see marks in Fig. 8.4). $\langle s^n\rangle(t;x_0)$ denotes the nth moment of the avalanche size (measured as the number of charges) for a system of size t starting with x_0 initial charges. Using (8.69) one has

$$\langle s^n\rangle(t;x_0) = \int_0^t dt' \int_0^\infty dx' x' n\psi_{n-1}(t',x';x_0) \tag{8.72}$$

The dimensionless form of ψ is given by

$$\psi_n(x,t;x_0) = \frac{1}{x_0}\left(\frac{x_0^3}{D}\right)^n \widetilde{\psi}_n(y,\tau) \tag{8.73}$$

with $y = x/x_0$ and $\tau = t/(x_0^2/D)$. The propagator $G(y,\tau;y_0)$ is easily obtained from a mirror-charge trick,

$$G(y,\tau;y_0) \equiv \frac{1}{\sqrt{4\tau\pi}}\left(e^{-\frac{(y-y_0)^2}{4\tau}} - e^{-\frac{(y+y_0)^2}{4\tau}}\right) , \tag{8.74}$$

[5]It is interesting to note that this can be written using a generating function $\Psi(t,x;x_0,\xi)$ with $\partial_t\Psi = x\xi\Psi + D\partial_x^2\Psi$, so that indeed $\left.\frac{d^n}{d\xi^n}\right|_{\xi=0} \Psi(t,x;x_0,\xi) = \psi_n(t,x;x_0)$

[6]Especially, $\langle\{s(t,x;x_0)^n\}\rangle > 0$ for any $t > 0$ gives (8.70a) and $\lim_{t\to 0}\langle\{s(t,x;x_0)^n\}\rangle = 0$ gives (8.70b).

and $\widetilde{\psi}_0(y,\tau) = G(y,\tau;1)$, *i.e.*

$$\widetilde{\psi}_0(y,\tau) = \frac{1}{\sqrt{\tau\pi}} e^{-\frac{y^2+1}{4\tau}} \sinh\left(\frac{y}{2\tau}\right) . \tag{8.75}$$

One might be inclined to transfer the problem into k-space, which, however, does not simplify the problem because of the boundary condition (8.70b). The expression

$$\widetilde{\psi}_n(y,\tau) = \int_0^\tau d\tau' \int_0^\infty dy' n y' \widetilde{\psi}_{n-1}(y',\tau') G(y,\tau-\tau';y') \tag{8.76}$$

is the formal solution. Rescaling the arguments of $\widetilde{\psi}_n$ by powers of μ one finds

$$\widetilde{\psi}_n(\sqrt{\mu}y,\mu\tau) = \mu^{3/2} \int_0^\tau d\tau' \int_0^\infty dy' n y' \widetilde{\psi}_{n-1}(\sqrt{\mu}y',\mu\tau') G(y,\tau-\tau';y') . \tag{8.77}$$

From $\widetilde{\psi}_{n-1}(\sqrt{\mu}y,\mu\tau) = \mu^{\alpha_{n-1}}\widetilde{\psi}_{n-1}(y,\tau)$, then follows $\widetilde{\psi}_n(\sqrt{\mu}y,\mu\tau) = \mu^{\alpha_{n-1}+3/2}\widetilde{\psi}_n(y,\tau)$. Thus, starting with $\widetilde{\psi}_0(\sqrt{\mu}y,\mu\tau) = \mu^{\alpha_0}\widetilde{\psi}_0(y,\tau)$ one has apparently

$$\widetilde{\psi}_n(\sqrt{\mu}y,\mu\tau) = \mu^{\frac{3}{2}n+\alpha_0}\widetilde{\psi}_n(y,\tau) . \tag{8.78}$$

Unfortunately the scaling behaviour of $\widetilde{\psi}_0$ is a bit more complicated. Nevertheless, it can be expanded for large μ, or actually large $\mu\tau$, as

$$\widetilde{\psi}_0(\sqrt{\mu}y,\mu\tau) = \frac{1}{\mu}\frac{1}{\sqrt{\tau\pi}} e^{-\frac{y^2}{4\tau}} \left(\frac{y}{2\tau} + \frac{1}{\mu}\left(\frac{y^3}{48\tau^3} - \frac{y}{8\tau^2}\right) + \dots\right) . \tag{8.79}$$

Bearing in mind the necessity of large $\mu\tau$ one can now apply the scaling argument (8.78) order by order in μ since Eq. (8.69) and its dimensionless counterpart are linear. From (8.79) it is $\alpha_0 = -1$ for the leading order, $\alpha_0 = -2$ for the first sub-leading order and so on.

Eq. (8.78) immediately translates to $\langle s^n \rangle$ using (8.72) and (8.73); to leading order one finds

$$\langle s^n \rangle(\mu t; x_0) = \mu^{(3/2)n+1/2+\alpha_0} \langle s^n \rangle(t; x_0) \tag{8.80}$$

Assuming (8.1), from (8.2) with t taking the rôle of L it follows that $D = 3/2$ and $D(1-\tau) = 1/2 + \alpha_0$, *i.e.* for $\alpha_0 = -1$ one has $\tau = 4/3$. The next order correction is $D = 3/2$ and $\tau' = 2$.

Of course, it is also possible to calculate the leading orders of $\langle s^n \rangle$ exactly. Because of (8.80), one needs to calculate $\langle s^n \rangle(\mu t; x_0)$ for one value of t only. The simplest choice is to set $t = x_0^2/D$, which gives $\langle s \rangle(x_0^2/D; x_0) = x_0^3/D$ for $n = 1$, *i.e.*

$$\langle s \rangle(t; x_0) = x_0 t \tag{8.81}$$

which is exactly (8.40) (n in (8.40) corresponds to x_0 here and L in (8.40) to t). This is actually surprising, because (8.81) is only the *leading* order and corrections are expected from higher orders. However, it turns out that in fact all higher order corrections cancel. In fact, remarkably

$$\int_0^\infty dx x e^{-\frac{x^2}{4t}} \left(\frac{x}{t} - 2e^{-\frac{1}{4t}} \sinh\left(\frac{x}{2t}\right)\right) = 0 \,. \tag{8.82}$$

even though x/t is only the leading order of $2\exp(-1/(4t))\sinh(x/(2t))$. Especially

$$\int_0^\infty dx x e^{-\frac{x^2}{4t}} \left(\frac{x^3}{48t^3} - \frac{x}{8t^2}\right) = 0 \,. \tag{8.83}$$

According to (8.72) the next moment is

$$\langle s^2 \rangle(\mu t; x_0) = 2\mu^{5/2} \int_0^{tD/x_0^2} d\tau \frac{x_0^2}{D} \int_0^\infty dy \frac{x_0^3}{D} y x_0 \widetilde{\psi}_1(y,\tau) \tag{8.84}$$

the leading order of which can be determined using the leading order of $\widetilde{\psi}_1$,

$$\widetilde{\psi}_1(y,\tau) = \int_0^\tau d\tau' \int_0^\infty dy' \frac{1}{\sqrt{\tau'\pi}} e^{-\frac{y'^2}{4\tau'}} \frac{y'^2}{2\tau'} \frac{1}{\sqrt{4\pi(\tau-\tau')}} \left(e^{-\frac{(y-y')^2}{4(\tau-\tau')}} - e^{-\frac{(y+y')^2}{4(\tau-\tau')}} \right) + \dots \tag{8.85}$$

which gives the leading order of $\langle s^2 \rangle$

$$\langle s^2 \rangle(t; x_0) = \frac{32}{15\sqrt{\pi}} t^{5/2} \sqrt{D x_0^2} + \mathcal{O}(t^{3/2}) \tag{8.86}$$

identical to (8.64). Higher orders become very tedious, so that numerical evaluation seems to offer the better option.

8.4 Discussion and Conclusion

The results above represent some of the few exact result for sandpile-like models: Eq. (8.40) and Eq. (8.64) are the exact leading orders of the first two moments of the avalanche size distribution without making any assumptions about scaling behaviour. The conclusion that $\tau = 4/3$ and $D = 3/2$ can only be drawn by either assuming (8.1), or by accepting the continuum result (8.80) and then making assumptions about the uniqueness of the distribution inferred from its moments.

One important result is that in this exactly solvable case the relation $D = 1 + \chi$, see Sec. 6.1.4, page 253, is confirmed, since the roughness exponent of a random walk along an absorbing wall is $\chi = 1/2$, a result that is certainly quite straight-forward to verify rigorously.

Open task

The method introduced in Sec. 8.2 is not restricted to sandpile-like models. The underlying idea is to use a MARKOV matrix not only to evolve the state distribution, but also to calculate the moment generating function of the relevant observable. In order to obtain the finite-size scaling behaviour, its set of eigenvectors is generated recursively. From this recursion relation one can then develop a (discrete) PDE like (8.36), which can subsequently be used as a starting point for other techniques. In a two-dimensional variant of the present model, this recursion relation is much more complicated to obtain and might require the use of a matrix product state ansatz (Derrida and Evans, 1997). Nevertheless, it seems promising to apply the approach to more complicated processes, such as the TASEP and recent variants (Parmeggiani *et al.*, 2003), for which there is no solution known yet.

8.4.1 Summary

In this chapter an operator approach has been developed, closely related to the one presented in Sec. 5.3, page 225. The focus here, however, was rather on a recursive expression for the eigenvectors than on an expression for the matrix representation of the operator itself. The sections cover the following topics:

- The model is introduced and defined in Sec. 8.1. In particular, the Abelian property is illustrated.
- The totally asymmetric Oslo model is then cast into a Markov-matrix, the properties of which are calculated in Sec. 8.2. By "dressing" the Markov-matrix with powers of a dummy variable, it can be used to calculate the moment generating function of the avalanche size distribution. The result of this effort is a hierarchy of generating functions on a lattice which can be solved moment by moment. This has been done to leading order up to the second moment.
- In Sec. 8.3, the model is mapped onto a reaction-diffusion problem. That way, links to many closely related models can be identified. Subsequently, a continuum approach to the original problem on the lattice can be formulated and tackled with standard techniques. The leading order in this approach fully agrees with the leading order on the lattice.

Chapter 9

The Thermal EDWARDS-WILKINSON Equation with Drift

The effect of a drift term in the presence of fixed boundaries is studied for the one-dimensional EDWARDS-WILKINSON *equation to reveal a general mechanism that causes a change of exponents for a very broad class of growth processes. This mechanism represents a relevant perturbation and therefore is important for the interpretation of experimental and numerical results. In effect, the mechanism leads to the roughness exponent assuming the same value as the growth exponent. In the case of the* EDWARDS-WILKINSON *equation this implies exponents deviating from those expected by dimensional analysis.*

In the course of deriving various properties of the EDWARDS-WILKINSON *equation with periodic and fixed boundary conditions as well as with or without drift, some general results obtained in the theory of growth processes are reviewed.*

9.1 Introduction

The EDWARDS-WILKINSON (EW) equation (Edwards and Wilkinson, 1982), as it is discussed below, is probably the best-studied equation describing surface growth processes. Due to its linearity it is solvable by standard methods and has been studied analytically as well as numerically in great detail (Krug and Spohn, 1991; Halpin-Healy and Zhang, 1995; Krug, 1997). The equation is very well-behaved, so that the outcome of these investigations are usually quite predictable. There is no reason to suspect that well-accepted methods, such as dimensional analysis and coarse graining, produce wrong results, even if applied to the EW equation with an extra drift term (EWd), which still represents a linear problem.

However, it is shown below that, depending on the boundary conditions, such a drift term changes the exponents dramatically to anomalous values, which apparently have been missed in the literature. The drift in conjunction with the boundary condition poses a relevant perturbation to the original equation. While it is not possible to capture its effect by the simple methods mentioned above, it can be understood using physical arguments. The mechanism turns out to be very powerful and extends far beyond the EW problem.

9.2 The EW Equation with Periodic Boundary Conditions

The EDWARDS-WILKINSON equation (Edwards and Wilkinson, 1982) describes the temporal evolution of an interface characterised by its height $h(x,t)$ over a substrate of length L, $x \in [0, L]$, at time t under the influence of a thermal noise $\eta(x,t)$. In one dimension it reads

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) + \eta(x,t) \ , \tag{9.1}$$

with a diffusion constant or surface tension D. The initial conditions are usually (Nattermann and Tang, 1992; Krug, 1997) chosen to be $h(x, t = 0) \equiv 0$ and periodic boundary conditions (PBC) are applied.[1] Correspondingly, the EDWARDS-WILKINSON with drift (or convection) is

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) + \mathsf{v}\partial_x h(x,t) + \eta(x,t) \ . \tag{9.2}$$

In the following, two statistical properties of (9.1) are of special interest: the correlator of h and the width. While the definition of the former needs some discussion, the latter is defined as

$$w^2(t, L) = \left\langle \overline{h^2} \right\rangle - \left\langle \overline{h}^2 \right\rangle , \tag{9.3}$$

where $\overline{A}$ denotes the spatial average,

$$\overline{A} = \frac{1}{L}\int_0^L dx A(x) \tag{9.4}$$

and $\langle \ \rangle$ is the ensemble average, averaging over all realizations of the noise η. In order to determine w^2, the only property of η which enters is

$$\left\langle \eta(x,t)\eta(x',t') \right\rangle = \Gamma^2 \delta(x - x')\delta(t - t') \ , \tag{9.5}$$

where Γ parameterises the strength of noise. To fully specify the noise, the higher

[1] Periodic boundary conditions are realised by imposing $h(x = 0, t) = h(x = L, t)$ and $h(x, t)$ analytic at $x = 0, L$.

order correlations are usually chosen to be those of Gaussian white noise and the average is set to $\langle\eta(x,t)\rangle = 0$.

The first step to solve the EDWARDS-WILKINSON equation[2] (9.2), is to write it in dimensionless form:

$$\partial_\tau \varphi(y,\tau) = \partial_y^2\varphi + q\partial_y\varphi + \xi(y,\tau) \tag{9.6}$$

where $y = x/L$, $\tau = t/(L^2/\mathsf{D})$, $q = \mathsf{v}L/\mathsf{D}$ and

$$\varphi(y,\tau;q) = \frac{1}{\Gamma\sqrt{L/\mathsf{D}}} h(x,t;L,\mathsf{D},\Gamma,\mathsf{v}) \tag{9.7}$$

are all dimensionless quantities. The same applies to the noise $\xi = \eta/(\Gamma\sqrt{\mathsf{D}}/L^{3/2})$, so that

$$\big\langle \xi(y,\tau)\xi(y',\tau')\big\rangle = \delta(y-y')\delta(\tau-\tau') \quad \text{and} \quad \langle\xi(y,\tau)\rangle = 0\ . \tag{9.8}$$

The propagator φ_0 of this problem is a Gaussian "wrapped around the circle and evaluated at $y + q\tau$", *i.e.* essentially Jacobi's ϑ_3 function (Farkas and Fülöp, 2001; Magnus *et al.*, 1966)

$$\varphi_0(y,\tau) = \frac{1}{\sqrt{4\pi\tau}} \sum_{n=-\infty}^{\infty} e^{-\frac{(y+q\tau+n)^2}{4\tau}} = \sum_{n=-\infty}^{\infty} e^{ik_n(y+q\tau)} e^{-k_n^2\tau} \tag{9.9}$$

with $k_n = 2\pi n$ and $\tau \geq 0$. Plugging this into

$$\varphi(y,\tau) = \int_0^1 dy' \int_0^\tau d\tau' \varphi_0(y-y',\tau-\tau')\xi(y',\tau')\ , \tag{9.10}$$

and using the properties of the noise (9.8), one has

$$\begin{aligned}\langle\varphi\varphi\rangle = \sum_{n_1,n_2=-\infty}^{\infty} \int_0^{\tau_1} d\tau' \int_0^1 dy' e^{ik_{n_1}(y_1-y'+q(\tau_1-\tau'))} e^{-k_{n_1}^2(\tau_1-\tau')} \\ \times\ e^{ik_{n_2}(y_2-y'+q(\tau_2-\tau'))} e^{-k_{n_2}^2(\tau_2-\tau')}\ ,\end{aligned} \tag{9.11}$$

where the short-hand notation $\langle\varphi\varphi\rangle$ for $\langle\varphi(y_1,\tau_1)\varphi(y_2,\tau_2)\rangle$ has been used. In the same spirit we note explicitly

$$\langle hh\rangle = \frac{\Gamma^2 L}{D}\langle\varphi\varphi\rangle\ . \tag{9.12}$$

[2]The solution of the standard (thermal) EDWARDS-WILKINSON equation with periodic boundaries is ubiquitous in the literature and is carried out here only for completeness. It is performed in real-space simply because k-space does not help much and the FOURIER-transforms are trivialities anyway.

Moreover, in (9.11) the convention $\tau_2 \geq \tau_1$ has been applied.[3]

The integration over y' produces a KRONECKER $\delta_{n_1,-n_2}$, so that after integration over τ' the correlator (9.11) becomes

$$\langle \varphi\varphi \rangle = \tau_1 + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{2k_n^2} \left(e^{-k_n^2(\tau_2-\tau_1)} - e^{-k_n^2(\tau_1+\tau_2)} \right) e^{ik_n(y_1-y_2+q(\tau_1-\tau_2))} \tag{9.14}$$

where the leading τ_1 comes from the integration at $n = 0$. The time order enters the exponentials when the integration runs over the δ-function of the correlator. The reason why $\langle \varphi\varphi \rangle$ is a function of the *sum* of τ_1 and τ_2 and the difference, is that the substrate is flat initially.

9.2.1 Exponents from the correlator

Before deriving the roughness exponents from the width of the interface (defined below), it is very instructive to obtain the scaling of the correlator. Using the notation $\langle\rangle_{L\to\infty}$ for $\lim_{L\to\infty} \langle\rangle$, the definition of the exponents is based on the *equal-time* correlator in the thermodynamic limit (Krug and Spohn, 1991; Meakin, 1998):

$$\begin{aligned} & \left\langle (h(x_1,t) - h(x_2,t))^2 \right\rangle_{L\to\infty} \\ = \; & \left\langle h(x_1,t)^2 \right\rangle_{L\to\infty} + \left\langle h(x_2,t)^2 \right\rangle_{L\to\infty} - 2\langle h(x_1,t)h(x_2,t)\rangle_{L\to\infty} && (9.15) \\ = \; & |x_2 - x_1|^{2\chi} \mathcal{G}\left(\frac{t}{|x_2-x_1|^z} \right) && (9.16) \end{aligned}$$

with a universal scaling function $\mathcal{G}(z)$, which converges to a non-zero constant for $z \to \infty$ and behaves like $z^{2\beta}$ for small arguments. It is important to do the thermodynamic limit before considering exponents, otherwise there would be an additional spatial parameter in addition to x_1 and x_2. Clearly, because of translational invariance, the correlator can only depend on the *difference* $x_2 - x_1$, which is therefore the only purely spatial parameter. The exponent χ is obtained as the behaviour in $|x_2-x_1|$ when taking t to infinity, and β as the behaviour in t when taking $|x_2 - x_1|$ to infinity. As one expects spatial independence in the short time limit, $\beta = \chi/z$.

To characterise asymptotic behaviour in the following, a rather sloppy but useful notation of limits will be used, for example $\lim_{z\to 0} \mathcal{G}(z) = az^{2\beta}$ to indicate $\lim_{z\to 0} \mathcal{G}(z) z^{-2\beta} = a$.

[3] This convention comes into play when evaluating integrals of the form

$$\int_0^{\tau_1} d\tau_1' \int_0^{\tau_2} d\tau_2' \delta(\tau_1' - \tau_2') f(\tau_1', \tau_2') = \int_0^{\tau_1} d\tau_1' f(\tau_1', \tau_1') \,. \tag{9.13}$$

Evaluating Eq. (9.14) for *equal times*, $\tau_1 = \tau_2 = \tau$, one has

$$\langle \varphi(y_1,\tau)\varphi(y_2,\tau)\rangle = \tau + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{2k_n^2}\left(1 - e^{-2k_n^2\tau}\right) e^{ik_n(y_1-y_2)} . \tag{9.17}$$

It is not surprising to see that at equal times the drift q disappears: The system is translational invariant, so the drift can be compensated by observing the interface in a comoving frame, *i.e.* a Galilean transformation. If φ obeys the equation for the problem at $q = 0$, $\partial_\tau \varphi = \partial_y^2 \varphi + \eta$, then a transformed $\varphi'(y,\tau) = \varphi(y + q\tau, \tau)$ obeys[4]

$$\partial_\tau \varphi' = \partial_y^2 \varphi' + q\partial_y \varphi' + \eta \tag{9.18}$$

According to (9.17), at equal times, the movement of the moving frame becomes invisible to an observer of the correlator.

The leading τ indicates that h is actually divergent for $t \to \infty$. This is not surprising as well; there is no mechanism which binds the interface to the substrate, *i.e.* the interface floats freely above (or below) it. The time integrated influx per site due to the noise leads to an average height of variance τ. This can be seen directly in Eq. (9.53) below.

Based on the observation that

$$g(y) \equiv \sum_{\substack{n=-\infty \\ n\neq 0}} \frac{1}{2k_n} e^{ik_n y} \tag{9.19}$$

obeys $g(y)'' = (\delta(y) - 1)/2$ and $g(y = 0) = 1/24$ one easily finds

$$g(y) = \frac{1}{24} - \frac{1}{4}\left(|y| - y^2\right) . \tag{9.20}$$

Thus, for a correlator of the form (9.15) one finds in the stationary state *without* taking the thermodynamic limit

$$\lim_{t\to\infty} \left\langle (h(x_1,t) - h(x_2,t))^2 \right\rangle = \frac{\Gamma^2}{2\mathrm{D}} \left(|x_1 - x_2| - \frac{|x_1 - x_2|^2}{L} \right) , \tag{9.21}$$

which obeys translational invariance for $0 \leq x_{1,2} \leq L$

$$|\Delta x| - \frac{\Delta x^2}{L} = |\Delta x - L| - \frac{(\Delta x - L)^2}{L} \quad \text{for} \quad 0 \leq \Delta x \leq L \tag{9.22a}$$

$$|\Delta x| - \frac{\Delta x^2}{L} = |\Delta x + L| - \frac{(\Delta x + L)^2}{L} \quad \text{for} \quad -L \leq \Delta x \leq 0 \tag{9.22b}$$

[4]Note the remarks on the correlator of η, Eq. (7.80), page 310.

This function, (9.21), can actually be related to the width calculated below, see Eq. (9.56).

From (9.21) one can already derive the exponent $\chi = 1/2$ according to (9.16).

When taking the thermodynamic limit before taking $t \to \infty$, one has to bear in mind the L-dependent pre-factor in Eq. (9.12). Taking it into account, the thermodynamic limit renders the sum in (9.11) a Riemann sum with mesh $1/L$, while the leading time gets removed and one arrives at

$$\begin{aligned} &\left\langle (h(x_1,t) - h(x_2,t))^2 \right\rangle_{L\to\infty} \\ =\; & \frac{\Gamma^2}{2\pi \mathsf{D}} \left\{ \sqrt{8\pi \mathsf{D}t} \left[1 - e^{(\Delta x)^2/(8Dt)}\right] + \pi|\Delta x| \left[1 - \mathcal{E}\left(\frac{|\Delta x|}{\sqrt{8\mathsf{D}t}}\right)\right]\right\} \end{aligned} \tag{9.23}$$

where $\mathcal{E}(x) \equiv 2/\sqrt{\pi}\int_0^x dz \exp(-z^2)$ is the error function and $\Delta x \equiv x_2 - x_1$. The two relevant limits are:

$$\lim_{t\mathsf{D}/\Delta x^2 \to \infty} \left\langle (h(x_1,t) - h(x_2,t))^2 \right\rangle_{L\to\infty} = \frac{\Gamma^2}{2\mathsf{D}}|\Delta x| \tag{9.24a}$$

$$\lim_{t\mathsf{D}/\Delta x^2 \to 0} \left\langle (h(x_1,t) - h(x_2,t))^2 \right\rangle_{L\to\infty} = \frac{2\Gamma^2}{\sqrt{2\pi \mathsf{D}}}\sqrt{t} \tag{9.24b}$$

Equivalently the first limit, Eq. (9.24a), could be replaced by $\lim_{t\to\infty}$ or $\lim_{\Delta x\to 0}$ and similarly for the second.

The first limit, (9.24a), is equivalent to the limit $t \to \infty$ in (9.21) and confirms $\chi = 1/2$. The second limit implies $\beta = 1/4$ and therefore $z = 2$. To summarise the exponents:

$$\chi = 1/2 \quad \text{and} \quad \beta = 1/4 \quad \text{and} \quad z = 2 \quad \text{for any v in the periodic case} \tag{9.25}$$

9.2.1.1 The two-time correlator

In order to understand the literature, it is helpful to consider the two-time correlation function. As the drift can be gauged away in the periodic case, one expects an expression of the form $|\Delta x + v\Delta t| g(\Delta t/(\Delta x + \mathsf{v}\Delta t)^2)$. Indeed, from (9.14) one has

$$\left\langle (\varphi(y_1,\tau_1) - \varphi(y_2,\tau_2))^2 \right\rangle \tag{9.26a}$$

$$= \left\langle \varphi(y_1,\tau_1)^2 \right\rangle + \left\langle \varphi(y_2,\tau_2)^2 \right\rangle - 2\langle \varphi(y_1,\tau_1)\varphi(y_2,\tau_2)\rangle \tag{9.26b}$$

$$\begin{aligned} =\; & \tau_1 + \tau_2 - 2\tau_1 + \sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} \frac{1}{2k_n^2}\left(2 - e^{-2k_n^2\tau_1} - e^{-2k_n^2\tau_2}\right) \\ & -2\sum_{\substack{n=-\infty\\ n\neq 0}}^{\infty} \frac{1}{2k_n^2}\left(e^{-k_n^2(\tau_2-\tau_1)} - e^{-k_n^2(\tau_1+\tau_2)}\right) e^{ik_n(y_1-y_2+q(\tau_1-\tau_2))} \end{aligned} \tag{9.26c}$$

for which one can take the limit $\tau_{1,2} \to \infty$ keeping $\Delta\tau = \tau_2 - \tau_1$ fixed. Then the correlator becomes

$$\lim_{\tau_{1,2}\to\infty} \left\langle (\varphi(y_1,\tau_1) - \varphi(y_2,\tau_2))^2 \right\rangle = \tau_2 - \tau_1 + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{k_n^2} \left(1 - e^{-k_n^2 \Delta\tau} e^{ik_n(y_1 - y_2 - q\Delta\tau)}\right) \tag{9.27}$$

and by taking into account the pre-factor (9.12) one can take the thermodynamic limit:

$$\begin{aligned} &\left\langle (h(x_1,t_1) - h(x_2,t_2))^2 \right\rangle_{L\to\infty, t\to\infty} \\ &\qquad = \frac{\Gamma^2}{D} \int_{-\infty}^{\infty} dz \frac{1}{(2\pi)^2 z^2} \left(1 - e^{-(2\pi)^2 z^2 \Delta t \mathsf{D} - 2\pi i z (\Delta x + \mathsf{v}\Delta t)}\right) \end{aligned} \tag{9.28}$$

where the subscript on the LHS denotes the limits $\lim_{L\to\infty} \lim t_1, t_2 \to \infty$, again with constant $\Delta t = t_2 - t_1$, so that the second limit represents only a single limit. The integral can be written as

$$g(x,q) = \int_{-\infty}^{\infty} dz \frac{1}{z^2} \left(1 - e^{-izx - z^2 q}\right) \tag{9.29a}$$

$$= |x| \int_{-\infty}^{\infty} dz' \frac{1}{z'^2} \left(1 - e^{-iz' - qz'^2/x^2}\right) \tag{9.29b}$$

where the second line contains a leading term $|x|$ because of the transformation of the boundaries when doing $z' = zx$; if $x < 0$ the boundaries of the integral change sign. This second line can be denoted as $|x|\tilde{g}(q/x^2)$, so that one arrives at

$$\left\langle (h(x_1,t_1) - h(x_2,t_2))^2 \right\rangle_{L\to\infty, t\to\infty} = \frac{\Gamma^2}{2\pi\mathsf{D}} g\left(\Delta x + \mathsf{v}\Delta t, \Delta t \mathsf{D}\right) \tag{9.30a}$$

$$= \frac{\Gamma^2}{2\pi\mathsf{D}} |\Delta x + \mathsf{v}\Delta t| \, \tilde{g}\left(\frac{\Delta t \mathsf{D}}{(\Delta x + \mathsf{v}\Delta t)^2}\right) \tag{9.30b}$$

Finally this reveals the structure of the correlator: Just as predicted, it is the correlator for $\mathsf{v} = 0$ with Δx replaced by $\Delta x + \mathsf{v}\Delta t$. The various limits of this correlator depend on the behaviour of $\tilde{g}(s)$ or $g(x,q)$. For $q = 0$ one has for the latter

$$g(x, q=0) = \int_{-\infty}^{\infty} dz \frac{1 - e^{-izx}}{z^2} \tag{9.31}$$

and therefore $\partial_x^2 g(x, q = 0) = 2\pi\delta(x)$. This differential equation is solved using $g(x, q = 0) = g(-x, q = 0)$ and $g(x = 0, q = 0) = 0$, so that $g(x, q = 0) = \pi|x|$. Since $g(x,q) = |x|\tilde{g}(q/x^2)$, it is $\tilde{g}(0) = \pi$.

For $x = 0$ one has

$$g(x=0, q) = \int_{-\infty}^{\infty} dz \frac{1}{z^2} \left(1 - e^{-z^2 q}\right) \tag{9.32}$$

and therefore $\partial_q g(x = 0, q) = \sqrt{\pi/q}$, so that $g(x = 0, q) = 2\sqrt{\pi q}$ using again $g(x = 0, q = 0) = 0$. This transfers to $\tilde{g}(u)$, *i.e.*

$$\tilde{g}(u) = \begin{cases} \pi & \text{for } u \to 0 \\ 2\sqrt{u\pi} & \text{for } u \to \infty \end{cases} \tag{9.33}$$

Together with (9.30b), it is now easy to determine the various limits of the two time correlation function, always assuming that the other argument remains finite. One finds

$$\left\langle (h(x_1, t_1) - h(x_2, t_2))^2 \right\rangle_{L\to\infty, t\to\infty}$$
$$= \frac{\Gamma^2}{D} \begin{cases} \frac{1}{2}|\Delta x| & \text{for} \quad \Delta t \to 0 & (9.34a) \\ \sqrt{\frac{\Delta t \mathsf{D}}{\pi}} & \text{for} \quad \Delta t \to \infty \quad \text{and} \quad \mathsf{v} = 0 & (9.34b) \\ \frac{1}{2}|\Delta x + \mathsf{v}\Delta t| & \text{for} \quad \Delta t \to \infty \quad \text{and} \quad \mathsf{v} \neq 0 & (9.34c) \\ \sqrt{\frac{\Delta t \mathsf{D}}{\pi}} & \text{for} \quad \Delta x \to 0 \quad \text{and} \quad \mathsf{v} = 0 & (9.34d) \\ \frac{1}{2}|\mathsf{v}\Delta t| & \text{for} \quad \Delta x \to 0 \quad \text{and} \quad \mathsf{v} \neq 0 \quad \text{and} \quad \Delta t \to \infty & (9.34e) \\ \sqrt{\frac{\Delta t \mathsf{D}}{\pi}} & \text{for} \quad \Delta x \to 0 \quad \text{and} \quad \mathsf{v} \neq 0 \quad \text{and} \quad \Delta t \to 0 & (9.34f) \\ \frac{1}{2}|\Delta x + \mathsf{v}\Delta t| & \text{for} \quad \Delta x \to \infty & (9.34g) \end{cases}$$

Again, if a variable appears which has actually been taken to a limit, it is there to show how the equation behaves in the limit.

9.2.1.2 Interpretation of the correlator

It is very helpful to develop a physical understanding of the behaviour of the correlator.

- Eq. (9.34a) is exactly (9.21) for $L \to \infty$.
- Eq. (9.34b) is the correlator of the following procedure: Take a measurement at t_1, then wait a very long time and take another measurement somewhere else. Repeat this procedure for increasing temporal distances and observe the asymptotic behaviour. The two measurements are asymptotically *not* independent for a given distance Δx, because at long time the diffusion covers the entire substrate. The variance is dominated by the time span between the two measurements and the growth of the interface in the meantime. This behaviour corresponds to initial growth.
- Eq. (9.34c) would behave like (9.34b), but as Δt diverges, the distance of the (effective) position of the second measurements, diverges much faster than the diffusion could cover it. Thus, the two measurements are not correlated by the processes happening *after* the first measurement, but only due to the history of

the interfaces. Based on the idea of asymptotic independence, one might think this is simply

$$\left\langle (h_1 - h_2)^2 \right\rangle \to \left\langle h_1^2 \right\rangle - 2\langle h_1\rangle\langle h_2\rangle + \left\langle h_2^2 \right\rangle \tag{9.35}$$

which differs from the correct form by the cross term being $\langle h_1\rangle\langle h_2\rangle$ instead of $\langle h_1 h_2\rangle$. Here, the short-hand notation $h_{1,2} = h(x_{1,2}, t_{1,2})$ has been used. Thus Eq. (9.35) should be dominated by (9.52). However, the latter expression is divergent, as the limit $t \to \infty$ has been taken *before* the thermodynamic limit. Thus, for any finite temporal distance Δt there are apparently correlations encoded in the interface due to its history, which violate $\langle h_1 h_2\rangle \to \langle h_1\rangle\langle h_2\rangle$.

- Eq. (9.34d) is essentially (9.34b), just that now the relation $\sqrt{\Delta t \mathsf{D}} \gg \Delta x$ is realised by small Δx.

- Eq. (9.34e) and Eq. (9.34f) are $\Gamma^2/(2\pi)|\mathsf{v}\Delta t|\tilde{g}(D/(v^2\Delta t))$, keeping Δt finite. If Δt becomes large compared to D/v^2 (note the second timescale), the correlator behaves like (9.34c), otherwise one sees initial growth, like (9.34b). The similar distinction has been made implicitly in (9.34c) by effectively comparing the timescales Δt and $\Delta x/\mathsf{v}$. In (9.34e) and (9.34f), however, Δx has already disappeared.

- Eq. (9.34g) is essentially again (9.34a).

9.2.1.3 Exponents in the literature

In the literature (Biswas *et al.*, 1998), one finds the following exponents:

$$\chi = 1/2 \quad \text{and} \quad \beta = 1/4 \quad \text{and} \quad z = 2 \quad \text{for} \quad \mathsf{v} = 0 \tag{9.36a}$$

$$\chi = 0 \quad \text{and} \quad \beta = 0 \quad \text{and} \quad z = 1 \quad \text{for} \quad \mathsf{v} \neq 0 \tag{9.36b}$$

However, these exponents were based on the two-time correlator,

$$\langle h(x_1, t_1) h(x_2, t_2)\rangle_{L\to\infty, t\to\infty} \ ,$$

where the thermodynamic limit has to be taken first, then the stationary state is considered by taking $t_2 + t_1$ to ∞ while keeping $\Delta t = t_2 - t_1$ constant, denoted by a subscript $t \to \infty$. If the limits are taken the other way around, one is directly confronted with a divergence, as $h(x,t)$ simply floats away (diverges) as $t \to \infty$. BISWAS *et al.* try to fix that by subtracting $\langle h\rangle$ or $\langle \overline{h}\rangle$ which, however, both vanish.[5]

[5] What they actually would need to do is either subtracting $\overline{h}$, *i.e.* write $\left\langle (h(x_1,t_1) - \overline{h}(t_1))(h(x_2,t_2) - \overline{h}(t_2)) \right\rangle$ or use the correlator Eq. (9.26a), (Kardar, 2001).

Similar to Eq. (9.16), one might expect (Biswas *et al.*, 1998) this two-time correlator to scale like:

$$\langle h(x_1,t_1)h(x_2,t_2)\rangle_{L\to\infty,t\to\infty} = \begin{cases} |\Delta x|^{2\chi} & \text{for large } \Delta x \text{ and } \Delta t = 0 \\ |\Delta t|^{2\beta} & \text{for large } \Delta t \text{ and } \Delta x = 0 \end{cases} \tag{9.37}$$

It is very important to note that this is *not* the standard definition of the exponents: For example RAMASCO, LÓPEZ and RODRÍGUEZ (Ramasco *et al.*, 2000) or GALLUCCIO and ZHANG (Galluccio and Zhang, 1995) use the equal time correlation function, and so does KRUG (Krug, 1997), LÄSSIG (Lässig, 1998) as well as KRUG and SPOHN (Krug and Spohn, 1991) as well as HALPIN-HEALEY and ZHANG (Halpin-Healy and Zhang, 1995). BISWAS *et al.* go even further and assume a scaling behaviour like

$$\langle h(x_1,t_1)h(x_2,t_2)\rangle_{L\to\infty,t\to\infty} \approx |\Delta x|^{2\chi}\mathcal{F}\left(\frac{|\Delta t|}{|\Delta x|^z}\right) \tag{9.38}$$

which is according to (9.30b) simply impossible: Eq. (9.30b) is not divergent in the limit $t \to \infty$, while $\langle h^2 \rangle$ is so [see (9.52), (9.47)]. However, according to Eq. (9.26b), then $\langle h_1 h_2 \rangle$ must be divergent as well, hence Eq. (9.38) is ill-defined on the LHS.

If $\langle h(x_1,t_1)h(x_2,t_2)\rangle$ was not divergent, one might think that a reasonable choice for the exponents in (9.38) is $\chi = 0$ and $z = 1$, in order to allow the correlator for $\mathsf{v} \neq 0$ to be expressible in terms of the correlator for $\mathsf{v} = 0$:

$$\langle h(x_1,t_1)h(x_2,t_2)\rangle_{\mathsf{v}\neq 0} = \langle h(x_1 + \mathsf{v}\Delta t, t_1)h(x_2,t_2)\rangle \tag{9.39}$$

The exponent $z = 1$ would make the Δt cancel in the argument of $\mathcal{F}$. These exponents are exactly those shown in (9.36).

9.2.1.4 A critique

However, BISWAS *et al.* derive their exponents in k-space, suggesting a structure factor, which, however, is defined as the modulus squared of the FOURIER transform of $h(x,t) - \langle h(x,t)\rangle$, see footnote 5 on page 349. That is probably the reason why they do not encounter the divergence of their correlator. They perform a FOURIER transform of the correlator and consider its scaling for small k and ω. All limits in t and x have already been taken in the FOURIER transform. If one wants to understand their results from Eq. (9.34), one has to mimic the limits they take in k and ω in real-space. Of course, these arguments must remain handwaving.

Based on the correlator calculated in Eq. (9.34) and the new definition of the exponents in (9.37), one has $\chi = 1/2$ from (9.34a) for any v, but $\beta = 1/4$ for $\mathsf{v} = 0$ from (9.34d) and $\beta = 1/2$ for $\mathsf{v} \neq 0$ from (9.34e). This is, apart from z, not what they find, (9.36). The reason is their FOURIER transform, which is dominated by the behaviour of the correlator for large arguments.

In this context an interesting line is (9.34g). Here, the correlator seems to indicate a linear space dependence, *i.e.* $\chi = 1/2$, however, for large $\mathrm{v}\Delta t$ it is divergent, indicating an asymptotic independence from Δx, so that $\chi = 0$. Thus, it seems as if there was a crossover, governed by the timescale $1/\mathrm{v}$. Below the crossover, Δx dominates and one sees the standard exponent, above the crossover, the correlator diverges and $\chi = 0$.

One other aspect in which (Biswas *et al.*, 1998) differs from standard treatments is that the growth starts at $t = -\infty$. Thus, there is no initial configuration, *i.e.* the growth exponent β can only be seen in local, temporal correlation, not in the development of the roughness. It remains an open issue to fully understand the causes for the exponents (9.36). Open issue

9.2.2 Exponents from the interface width

Still considering the case of PBC, the exponents characterising the EW equation with drift are now derived from the width.

The width has actually already been defined in Eq. (9.3) — we repeat its properties here for further reference:

$$w^2(t, L) = \left\langle \overline{h(x,t)^2} \right\rangle - \left\langle \overline{h(x,t)}^2 \right\rangle \tag{9.40}$$

with $\overline{A}$ denoting the spatial average in dimensionful space as well as dimensionless space:

$$\begin{aligned} \overline{A(x)} &= \frac{1}{L}\int_0^L dx A(x) & (9.41)\\ \overline{A'(y)} &= \int_0^1 dy A'(y)\ . & (9.42) \end{aligned}$$

Assuming a FAMILY-VICSEK scaling behaviour (Family and Vicsek, 1985), three exponents, χ, β and z, are defined for the asymptotic behaviour of w^2

$$w^2(t, L) = aL^{2\chi}\mathcal{G}\left(\frac{t}{bL^z}\right)(1 + c(t, L; \dots))\ , \tag{9.43}$$

with appropriate, system dependent parameters a and b (metric factors) which make the universal scaling function $\mathcal{G}(x)$ a dimensionless function of a dimensionless argument. The scaling function behaves asymptotically like

$$\begin{aligned} \lim_{x\to\infty} \mathcal{G}(x) &= a_L & (9.44a)\\ \lim_{x\to 0} \mathcal{G}(x)x^{-2\beta} &= a_t & (9.44b) \end{aligned}$$

The term $c(t, L; \dots)$ in Eq. (9.43), which is usually omitted, accounts for corrections and vanishes in the appropriate limits. To be more specific

$$\lim_{t\to\infty} w^2(t,L) = a_L L^{2\chi}(1 + c_L(L;\dots)) \tag{9.45a}$$

$$\lim_{L\to\infty} w^2(t,L) = a_t t^{2\beta}(1 + c_t(t;\dots)) \tag{9.45b}$$

where a_L and a_t are constants and $c_L(L; \dots) = \lim_{t\to\infty} c(t, L; \dots)$ is a dimensionless function, which vanishes as L diverges. As a dimensionless function it must be representable as a function of dimensionless parameters only. Those additional parameters required are included in .... Similarly, the correction $c_t(t; \dots) = \lim_{L\to\infty} c(t, L; \dots)$ vanishes for small t.

In Eq. (9.43) the exponent z is independent of χ and β. However, for large L one has

$$\lim_{L\to\infty} w^2(t,L) = \lim_{L\to\infty} L^{2\chi}\left(\frac{t}{L^z}\right)^{2\beta} a_t(1 + c_t(t;\dots)) \, . \tag{9.45c}$$

If the RHS of this equation is finite, *i.e.* independent of L, as suggested in (9.45b), then $\chi = z\beta$. It is important to keep in mind that generalised homogeneity, (9.43), even without corrections, does *not* entail $\chi = z\beta$. It can only be derived from an additional assumption such as (9.45b).

9.2.2.1 Relation between width and correlator

In principle, the correlator contains much more information than the width. However, in order to see scaling behaviour in the correlator, various limits had to be taken. For example, Eq. (9.34) includes already the thermodynamic limit and describes only the behaviour in the stationary state.

From the definition of the width (9.40) one has by pulling the spatial average out the ensemble average

$$\begin{aligned} 2w^2(t,L) &= \frac{1}{L}\int_0^L dx_1 \left\langle h(x_1,t)^2\right\rangle + \frac{1}{L}\int_0^L dx_2 \left\langle h(x_2,t)^2\right\rangle \\ &\quad -2\frac{1}{L^2}\int_0^L dx_1 \int_0^L dx_2 \langle h(x_1,t)h(x_2,t)\rangle &\text{(9.46a)} \\ &= \frac{1}{L^2}\int_0^L dx_1 \int_0^L dx_2 \left\langle (h(x_1,t) - h(x_2,t))^2\right\rangle \, . &\text{(9.46b)} \end{aligned}$$

By translational invariance one has $\left\langle h(x_1,t)^2\right\rangle = \left\langle h(x_2,t)^2\right\rangle$ and therefore

$$\left\langle h(x_1,t)^2\right\rangle = \left\langle \overline{h(x_2,t)^2}\right\rangle \, . \tag{9.47}$$

More general, in case of translational invariance one has for (9.46b)

$$2w^2(t,L) = \frac{1}{L}\int_0^L dx \langle (h(0,t) - h(x,t))^2 \rangle \ . \tag{9.48}$$

For small x the integrand is always finite, while for $x^2 \gg \mathrm{D}t$ it converges to a function independent of x, see for example (9.24b). Thus, in the thermodynamic limit, the value of the integral is dominated solely by the large x behaviour and the averaging simply reproduces its large x value (Krug, 1997, Eqn. (3.46)):

$$2 \lim_{L\to\infty} w^2(t,L) = \lim_{\Delta x\to\infty} \lim_{L\to\infty} \langle (h(0,t) - h(\Delta x,t))^2 \rangle \ . \tag{9.49}$$

This is confirmed by comparing (9.24b) and (9.55).

Similarly, for $t \to \infty$ the correlator $\langle (h(0,t) - h(x,t))^2 \rangle$ is expected to be a power law in Δx, so that the averaging (9.48) should only change the pre-factor. Doing simply an integration over (9.24a) like[6]

$$\frac{2}{L}\int_0^{L/2} d\Delta x \frac{\Gamma^2}{2D}|\Delta x| = \frac{\Gamma^2 L}{8D} \tag{9.50}$$

gives the correct scaling behaviour, namely the same exponent as (9.24a), $\chi = 1/2$, but compared to (9.56) the wrong pre-factor, because of the second term in the bracket of (9.21), which disappears in the thermodynamic limit taken in (9.24a). Subtracting

$$\frac{2}{L}\int_0^{L/2} d\Delta x \frac{\Gamma^2}{2D}\frac{\Delta x^2}{L} = \frac{\Gamma^2 L}{24L} \tag{9.51}$$

reproduces (9.56) doubled, as predicted by (9.48).

9.2.3 Calculation of the width

The behaviour of the width is much easier to calculate than the behaviour of the correlator: Contrary to the correlation function, the width depends only on two parameters, namely L and t. Having obtained already the correlator in Eq. (9.17), it is straight forward to calculate the two contributions:

$$\left\langle \overline{\varphi(x,t)^2} \right\rangle = \tau + \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{2k_n^2}\left(1 - e^{-2k_n^2\tau}\right) \ , \tag{9.52}$$

[6]The idea to integrate only over[0,L/2] and to multiply by 2 is due to translational invariance. The same result is obtained, when averaging (9.21) over x_1 for any fixed x_2, namely $(1/6)L\Gamma^2/(2D)$.

and

$$\left\langle \overline{\varphi(x,t)}^2 \right\rangle = \tau \ . \tag{9.53}$$

This equation can also be derived by considering the time and space integrated inflow due to the noise, the variance of which is $\Gamma^2 tL$. A spatial normalisation produces a factor $1/L^2$, leading to $\Gamma^2 t/L$ altogether. Comparing (9.53) to (9.12) confirms this.

Therefore

$$w^2(t,L) = \frac{\Gamma^2 L}{D} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{2k_n^2} \left(1 - e^{-2k_n^2 t\mathsf{D}/L^2}\right) \ . \tag{9.54}$$

Again, the form of this solution actually follows directly from the physical covariance principle, as discussed in Sec. 9.4. Similar to Eq. (9.17) and Eq. (9.23), the thermodynamic limit converts the sum into an integral and one arrives at

$$\lim_{L\to\infty} w^2(t,L) = \frac{\Gamma^2}{D}\sqrt{\frac{Dt}{2\pi}} \quad \text{periodic boundaries.} \tag{9.55}$$

The limit $t \to \infty$ is even more trivial, because the exponential in (9.54) simply disappears and with $\sum_{n=1}^{\infty} 1/n^2 = \pi^2/6$ one has

$$\lim_{t\to\infty} w^2(t,L) = \frac{\Gamma^2 L}{D}\frac{1}{24} \quad \text{periodic boundaries.} \tag{9.56}$$

Once more, the resulting exponents are, according to Eq. (9.45), the standard exponents already listed in Eq. (9.25),

$$\chi = 1/2 \quad \text{and} \quad \beta = 1/4 \quad \text{and} \quad z = 2 \quad \text{for any v with periodic boundaries} \tag{9.57}$$

As a test for consistency one integrates (9.21) according to (9.46b) and finds

$$\frac{1}{L^2}\left(\int_0^L dx_1 \int_0^L dx_2 |\Delta x| - \int_0^L dx_1 \int_0^L dx_2 \frac{\Delta x^2}{L}\right) = \frac{1}{6}L \tag{9.58}$$

and therefore again (9.56). It is worth noting that only taking the spatial average of (9.34a) does not work, because all expressions in Eq. (9.34) contain already a thermodynamic limit, as mentioned in Sec. 9.2.2.1. However, even though Eq. (9.34) is valid only in the stationary state, one can recover (9.55) from Eq. (9.34), by a spatial integral of (9.34b) and replacing Δt by t (as well as dividing by 2, according to (9.46a)).

The equations above, (9.55) and (9.56), are exact and do not contain any of those corrections mentioned in (9.45). This is because there are simply no independent scales available which could define a reference for t in (9.55) or L in (9.56). This is expected to change, as soon as v remains as parameter in w^2, because D/v^2 provides a timescale and D/v provides a length scale. More details on these issues are presented in Sec. 9.4.

One can compare the pre-factor in (9.56) to the one cited in (Foltin *et al.*, 1994), which is $1/12$. The reason for the discrepancy is that the authors of (Foltin *et al.*, 1994) start from a random walker representation of the interface, rather than growing it "by hand", *i.e.* directly from the LANGEVIN equation. Of course, the roughness of the interface depends on the diffusion constant of the random walker, which is apparently *not* the diffusion constant in the LANGEVIN equation (9.1).

9.3 The EW Equation with Fixed Boundary Conditions

The original motivation of these calculation was to understand the mechanism behind the fact that adding a drift term to a quenched EDWARDS-WILKINSON equation with *fixed* boundary conditions, makes it solvable or at least makes the model fall into a universality class, which can be represented by an exactly solvable model. So far, the effect of a drift term has only been studied for the EDWARDS-WILKINSON equation with thermal noise in case of periodic boundary conditions. In this section it is extended to the EW equation with *fixed* boundary conditions,

$$h(x=0,t) = h(x=L,t) \equiv 0 \tag{9.59}$$

In that case, the propagator φ_0 for (9.2) becomes a bit more complicated:

$$\begin{aligned} \varphi_0(y,\tau;y_0,q) &= \frac{1}{\sqrt{4\pi\tau}} \sum_{n=-\infty}^{\infty} \left(e^{-\frac{(y-y_0+2n)^2}{4\tau}} - e^{-\frac{(y+y_0+2n)^2}{4\tau}} \right) \\ &\quad \times e^{-\frac{1}{2}(y-y_0)q - \frac{1}{4}\tau q^2} &\text{(9.60a)} \\ &= 2\sum_{n=1}^{\infty} \sin(k_n y)\sin(k_n y_0) e^{-k_n^2\tau} e^{-\frac{1}{2}(y-y_0)q-\frac{1}{4}\tau q^2} &\text{(9.60b)} \end{aligned}$$

where y_0 is the "starting point", *i.e.* $\lim_{\tau\to 0}\varphi(y,\tau,y_0,q) = \delta(y-y_0)$. One can easily check that the propagator is invariant under mirroring, *i.e.* applying $q \to -q$, $y \to 1-y$, $y_0 \to 1-y_0$ leaves φ_0 unchanged.

There is another symmetry, which can be immediately spotted, but is physically less intuitive, $\varphi_0(y,\tau;y_0,q) = \varphi_0(y_0,\tau;y,-q)$. It is caused by the linearity of (9.1); it makes it impossible to find out from local information at y only the sign of the velocity without knowing the relative position (to the right or to the left) of the source.

The propagator (9.60) is essentially a mirror charge version of Eq. (9.9), where on a circle of circumference 2 a positive charge sits at y_0 and a negative one at $-y_0$. Without drift term the field would vanish at $y=0$ and $y=1$. The drift makes the additional factor $\exp(-(1/2)(y-y_0)q-(1/4)\tau q^2)$ necessary. Just like in Eq. (9.9), the second line, (9.60b), comes from properties of ϑ_3 (Farkas and Fülöp, 2001; Magnus *et al.*, 1966), or, equivalently, from a Poisson summation. It is very important to note that in contrast to (9.9), now $k_n = n\pi$, because of the circle now having a circumference of 2.

In the subsequent integration with $\tau_2 \geq \tau_1$

$$\langle \varphi(y_1, \tau_1)\varphi(y_2, \tau_2)\rangle = \int_0^1 dy' \int_0^{\tau_1} d\tau' \varphi_0(y_1, \tau_1 - \tau'; y', q)\varphi_0(y_2, \tau_2 - \tau'; y', q) \qquad (9.61)$$

the dimensionless velocity q remains in the problem even for equal times $\tau_1 = \tau_2$. Consequently, the integral cannot be calculated in a closed form. At the heart of this problem lies the circumference of the circle used now. While the propagator (9.9) has the nice property

$$\int_0^1 dy' \sum_{n=-\infty}^{\infty} e^{\frac{-(x+n)^2}{4\tau}} f(x) = \int_{-\infty}^{\infty} e^{\frac{-x^2}{4\tau}} f(x) \qquad (9.62)$$

for any periodic function $f(x)$, the propagator for fixed boundaries has a factor 2 in front of the n and a negative sign in front of the exponential representing the other half of the circle $[-1, 1]$, where the mirror charge resides, even for vanishing q.

9.3.1 The case $\mathsf{v} = 0$

For vanishing velocity, the roughness should in the thermodynamic limit tend to the bulk values, because "it cannot see" the fixed boundaries. Nevertheless, for fixed boundaries the interface cannot float away and $\left\langle \overline{h}^2 \right\rangle$ cannot diverge as $t \to \infty$.

It is important to note again that the corrections suggested in Eq. (9.45) cannot exist for $\mathsf{v} = 0$, because they require an extra scale, which does not exist, see Sec. 9.4.5.1. For example, to obtain a correction $c_L(L; \dots)$, one needs a "counter-length-scale" by which L can be divided. Otherwise, c_L cannot be dimensionless. Since this correction applies in the limit $t \to \infty$, there is no such scale left, so that $c_L(L; \dots)$ must be independent of L and is therefore absorbed in the constant.

For $q = 0$ the propagator (9.60b) adapts a very simple form,

$$\varphi_0(y, \tau; y_0, q) = 2 \sum_{n=1}^{\infty} \sin(k_n y) \sin(k_n y_0) e^{-k_n^2 \tau} , \qquad (9.63)$$

so that

$$\left\langle \overline{\varphi^2} \right\rangle = \frac{1}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \left(1 - e^{-2n^2\pi^2\tau}\right) \qquad (9.64a)$$

$$\left\langle \overline{\varphi}^2 \right\rangle = \frac{4}{\pi^4} \sum_{n=1,3,5,\dots}^{\infty} \frac{1}{n^4} \left(1 - e^{-2n^2\pi^2\tau}\right) \qquad (9.64b)$$

In the limit of divergent time, these sums give

$$\lim_{\tau\to\infty} \left\langle \overline{\varphi^2} \right\rangle = \frac{1}{12} \quad \text{for fixed boundaries} \tag{9.65a}$$
$$\lim_{\tau\to\infty} \left\langle \overline{\varphi}^2 \right\rangle = \frac{1}{24} \quad \text{for fixed boundaries .} \tag{9.65b}$$

Plugging this into (9.40) one arrives at exactly the same result as (9.56). *Prima facie* it looks surprising that an interface fixed to the substrate at its two ends has the same roughness as a freely floating, periodic interface. However, by choosing an appropriate reference height, it is alway possible to transform a freely floating interface into one, which intersects this reference line at least once and can therefore be seen as an interface fixed to the substrate at these intersections. This is of course only a handwaving argument which gives a good idea why the *roughness* of the two models might be the same, but it is not sufficient to show that the two *ensembles* are identical. Of course, $\left\langle \overline{\varphi^2} \right\rangle$ are different in the two models.

Taking the thermodynamic limit in (9.64) means to investigate its behaviour for small τ. Clearly, the expressions have to vanish at least as fast as $1/L$, otherwise w^2 diverges because of the extra pre-factor $\Gamma^2 L/\mathsf{D}$ in (9.40). One finds in a sloppy notation

$$\lim_{\tau\to 0} \left\langle \overline{\varphi^2} \right\rangle = \sqrt{\frac{\tau}{2\pi}} \tag{9.66a}$$
$$\lim_{\tau\to 0} \left\langle \overline{\varphi}^2 \right\rangle = \tau - \frac{8}{3}\sqrt{\frac{2\tau}{\pi}}\tau \tag{9.66b}$$

It is very important to stress that the corrections on the right-hand-side of (9.66b) are in τ. Only the first term, Eq. (9.66a), does not vanish, if the thermodynamic limit with a pre-factor $\Gamma^2 L/\mathsf{D}$ is taken. A correction of the form $c_t(t; \dots)$, as described in (9.45b) cannot occur, because there is no additional timescale.

Thus, one arrives for the initial growth at

$$\lim_{L\to\infty} w^2(t, L) = \frac{\Gamma^2}{D}\sqrt{\frac{Dt}{2\pi}} \tag{9.67}$$

This is again the same result as obtained for the freely moving interface with periodic boundaries, Eq. (9.55). It simply means that the interface cannot see the fixation of the boundaries initially and any small influence of the boundary simply disappears in the thermodynamic limit.

Conclusively, the exponents of the EDWARDS-WILKINSON equation with fixed boundary conditions is

$$\chi = 1/2 \quad \text{and} \quad \beta = 1/4 \quad \text{and} \quad z = 2 \quad \text{for } \mathsf{v} = 0 \text{ with fixed boundaries} \tag{9.68}$$

9.3.2 The case $\mathsf{v} \neq 0$

The case of non-vanishing drift is what actually interests us in this chapter. From the form of the two point correlation function (9.61) and the propagator ((9.60a), (9.60b)) it is clear that a non-vanishing velocity v remains in the problem even in the equal-time correlation function. The velocity gives rise to an extra length scale D/v, which makes possible critical exponents deviating from the results of dimensional analysis.

Moreover, the extra length scale makes corrections possible (see Sec. 9.4.5.1) as described in Eq. (9.45), which make the analysis significantly more complicated. When plugging (9.60a) into (9.61) at $\tau_1 = \tau_2 = \tau$, the two correlators give rise to four terms, over which one has to integrate. The arguments of the exponentials are

$$
\begin{aligned}
(n_1+n_2)q \quad &- \quad \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1-y_2)+n_1-n_2\right)^2 && (9.69a)\\
&- \quad \frac{1}{2\tilde{\tau}}\left(y'-\left[\frac{1}{2}(y_1+y_2)+n_1+n_2+\tilde{\tau}q\right]\right)^2 &&\\
(n_1-n_2-y_2)q \quad &- \quad \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1+y_2)+n_1+n_2\right)^2 && (9.69b)\\
&- \quad \frac{1}{2\tilde{\tau}}\left(y'-\left[\frac{1}{2}(y_1-y_2)+n_1-n_2+\tilde{\tau}q\right]\right)^2 &&\\
(-n_1+n_2-y_1)q \quad &- \quad \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1+y_2)+n_1+n_2\right)^2 && (9.69c)\\
&- \quad \frac{1}{2\tilde{\tau}}\left(y'-\left[\frac{1}{2}(y_1-y_2)-n_1+n_2+\tilde{\tau}q\right]\right)^2 &&\\
(-n_1-n_2-y_1-y_2)q \quad &- \quad \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1-y_2)+n_1-n_2\right)^2 && (9.69d)\\
&- \quad \frac{1}{2\tilde{\tau}}\left(y'-\left[\frac{1}{2}(-y_1-y_2)-n_1-n_2+\tilde{\tau}q\right]\right)^2 &&
\end{aligned}
$$

where $\tilde{\tau} = \tau - \tau'$. It is instructive to identify the mirror symmetry $q \to -q$, $x_1 \to 1-x_1$ and $x_2 \to 1-x_2$ in all four terms, which requires the following transformations of the variables:

$$
\begin{array}{llll}
n_1 \to -n_1 & n_2 \to -n_2 & y' \to 1-y' & \text{for (9.69a)}\\
n_1 \to -n_1 & n_2 \to -(n_2+1) & y' \to 1-y' & \text{for (9.69b)}\\
n_1 \to -(n_1+1) & n_2 \to -n_2 & y' \to 1-y' & \text{for (9.69c)}\\
n_1 \to -(n_1+1) & n_2 \to -(n_2+1) & y' \to 1-y' & \text{for (9.69d)}
\end{array}
\qquad (9.70)
$$

Clearly, these transformations become irrelevant, when n_1 and n_2 are summed from $-\infty$ to ∞ and y' integrated from 0 to 1 (see Eq. (9.60a) and Eq. (9.61)). Therefore, one can fix the sign of q to be positive without restrictions.

In a saddle point approximation (Dingle, 1973) the minima of these polynomials in y' are supposed to lie within the interval $[0,1]$, for the exponential to contribute $\sqrt{2\pi\tilde{\tau}}$.

It is worth rewriting (9.69) in the form

$$\tilde{y}q - \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1 - y_2) + n_1 - n_2\right)^2 - \frac{1}{2\tilde{\tau}}\left(y' - \left[\frac{1}{2}(y_1 + y_2) + \tilde{y} + \tilde{\tau}q\right]\right)^2 \quad \text{where } \tilde{y} = n_1 + n_2 \tag{9.71a}$$

$$\tilde{y}q - \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1 + y_2) + n_1 + n_2\right)^2 - \frac{1}{2\tilde{\tau}}\left(y' - \left[\frac{1}{2}(y_1 + y_2) + \tilde{y} + \tilde{\tau}q\right]\right)^2 \quad \text{where } \tilde{y} = n_1 - n_2 - y_2 \tag{9.71b}$$

$$\tilde{y}q - \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1 + y_2) + n_1 + n_2\right)^2 - \frac{1}{2\tilde{\tau}}\left(y' - \left[\frac{1}{2}(y_1 + y_2) + \tilde{y} + \tilde{\tau}q\right]\right)^2 \quad \text{where } \tilde{y} = -n_1 + n_2 - y_1 \tag{9.71c}$$

$$\tilde{y}q - \frac{1}{2\tilde{\tau}}\left(\frac{1}{2}(y_1 - y_2) + n_1 - n_2\right)^2 - \frac{1}{2\tilde{\tau}}\left(y' - \left[\frac{1}{2}(y_1 + y_2) + \tilde{y} + \tilde{\tau}q\right]\right)^2 \quad \text{where } \tilde{y} = -n_1 - n_2 - y_1 - y_2 \tag{9.71d}$$

First we derive the expression for $\langle\varphi^2(y,t)\rangle$, *i.e.* $y_1 = y_2 = y$ in (9.69), (9.71). The minimum in y' of the last term of (9.71a) is at $y + \tilde{y} + \tilde{\tau}q$. Whether we want to derive the exponent χ or β, in any case we are interested in the large L behaviour. Since $q = \mathsf{v}L/\mathsf{D}$, in this limit[7] $\tilde{y} = n_1 + n_2$ must be non-negative because of the first term $\tilde{y}q$ in (9.71a). Moreover $q\tilde{\tau} \geq 0$, so that $0 \leq y + \tilde{y} + \tilde{\tau}q \leq 1$ requires[8] $\tilde{y} = n_1 + n_2 \leq 0$. Thus, in the first expression, the exponential of (9.69a) contributes only if $n_1 + n_2 = 0$ and if $0 \leq y + \tilde{\tau}q \leq 0$.

Similarly, the minimum in the last term of (9.69b) requires[9] $n_1 - n_2 \leq 0$. However, then the first term is $-yq$ or smaller, which suppresses the contribution of (9.69b) for divergent q. The same argument applies to (9.69c).

The last expression, (9.69d) or alternatively (9.71d), is a bit more complicated. For this term to contribute one needs for the minimum $0 \leq -y - n_1 - n_2 + \tilde{\tau}q \leq 1$ and at the same time $-n_1 - n_2 - 2y \geq 0$ for the first term, $\tilde{y}q$. Since $y \geq 0$, this first term requires a negative $n_1 + n_2$. If that sum is -1, the difference must be of the form $n_1 - n_2 = -1 - 2n_2 \neq 0$. That means, the second term in (9.69d), $-(n_1 - n_2)^2/(2\tilde{\tau}) = -1/(2\tilde{\tau})$ contributes as well. This deserves somewhat more discussion, but we first need to check, whether even more negative sums $n_1 + n_2$ contribute. The next one is $n_1 + n_2 = -2$, but then $-y - n_1 - n_2 + \tilde{\tau}q = 2 - y + \tilde{\tau}q \geq 1$ is not accepted anymore as a valid position of the minimum in a saddle point approximation. The same applies for

[7]It is worth noting that even if q is divergent, $\tilde{\tau}q$ still covers the whole positive real line.

[8]Note that $\tilde{y}$ is integer in (9.71a), so that the next larger choice for $\tilde{y}$ obeying $0 \leq y + \tilde{y} + \tilde{\tau}q \leq 1$ is $\tilde{y} = 1$, which is marginal since $y, \tilde{\tau}q \geq 0$.

[9]Again, the case $n_1 - n_2 = 1$ is marginal.

more negative sums. Thus, the only possible contribution comes from $n_1 + n_2 = -1$.

The rôle second term in (9.69d) can be understood as soon as $\tilde{\tau}$ can be determined. All four terms (9.69) require $0 \leq \tilde{\tau} q \leq 1$ because of the position of the minimum in the last term, so that

$$0 \leq \tilde{\tau} \leq q^{-1} . \tag{9.72}$$

That means the second term in (9.69d) contributes at least $-q$, which is (negatively) divergent in the thermodynamic limit. This is because $n_1 + n_2 = -1$ entails

$$\left(\frac{1}{2}(y_1 - y_2) + n_1 + n_2 \right)^2 \geq 1 \tag{9.73}$$

for $y_1 = y_2 = y$ and $n_1, n_2 \in \mathbb{Z}$. Thus, (9.69d) does not contribute either.

The only term to contribute to $\left\langle \varphi^2(y,t) \right\rangle$ in a saddle point approximation is therefore the first term, (9.69a). Because of its second term, $-(n_1 - n_2)^2/(2\tilde{\tau})$, one needs $n1 - n_2 = 0$, which makes together with $n_1 + n_2 = 0$ from above $n_1 = n_2 = 0$. Therefore

$$\left\langle \varphi^2(y,t) \right\rangle = \int_0^\tau d\tilde{\tau} \left(\frac{1}{\sqrt{4\pi\tilde{\tau}}} \right)^2 \sqrt{2\pi\tilde{\tau}}\, I\left(y + \tilde{\tau} q\right) \tag{9.74}$$

where $I(x)$ is an indicator function

$$I(x) \begin{cases} 1 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{9.75}$$

For $\tau \geq q^{-1}$ Eq. (9.74) can be written as

$$\left\langle \varphi^2(y,t) \right\rangle = \int_0^{q^{-1}} d\tilde{\tau} \frac{1}{2\sqrt{2\pi\tilde{\tau}}} I\left(y + \tilde{\tau} q\right) = 2\sqrt{\frac{1-y}{2\pi q}} \tag{9.76}$$

so that

$$\left\langle \overline{\varphi^2(y,\tau)} \right\rangle = \frac{2}{3\sqrt{2\pi q}} \text{ for } \tau > q^{-1} \tag{9.77}$$

$$\left\langle \overline{h^2(x,t)} \right\rangle = \frac{2\Gamma^2}{3} \sqrt{\frac{L}{2\pi \mathsf{v} \mathsf{D}}} \text{ for } t > L/\mathsf{v} \tag{9.78}$$

(9.79)

For $\tau < q^{-1}$ the upper limit of the integral is actually given by τ, so that the indicator function does not necessarily impose any condition on the integral over $\tilde{\tau}$.

Doing the integral one finds

$$\left\langle \overline{\varphi^2(y,\tau)} \right\rangle = \sqrt{\frac{\tau}{2\pi}} \left(1 - \frac{1}{3}\tau q\right) \text{ for } \tau < q^{-1} \tag{9.80}$$

$$\left\langle \overline{h^2(x,t)} \right\rangle = \Gamma^2 \sqrt{\frac{t}{2\pi \mathsf{D}}} \left(1 - \frac{1}{3}\frac{\mathsf{v}t}{L}\right) \text{ for } t < L/\mathsf{v} \tag{9.81}$$

The second average required for w^2 is $\left\langle \overline{\varphi(y,\tau)}^2 \right\rangle$. For long times it can be derived quite easily from physical considerations. The interface can now be thought of as moving from the right to the left (positive q), being fixed at $\varphi = 0$ at both ends. Thus, it looks like "freshly initialised" close to $y = 1$, but "very old" towards $y = 0$. The integrated inflow of noise depends on the time the interface has spend under the influence of the noise, which apparently depends on the position, see Fig. 9.1. The variance of the total volume of noise is then a triangle of height $\Gamma^2 L/\mathsf{v}$ and length L. Thus, in the large t limit the spatial average, which gives another factor L^{-2}, is given by

$$\left\langle \overline{h(x,t)}^2 \right\rangle = \frac{\Gamma^2}{2v} \text{ for } t > L/\mathsf{v} \; . \tag{9.82}$$

For short times, the interface grows homogeneously, where it has not been "reinitialised" by the right boundary. In the latter region it grows like a wedge. According to Fig. 9.1 the area is $\Gamma^2 t(L - \frac{1}{2}\mathsf{v}t)$, so that

$$\left\langle \overline{h(x,t)}^2 \right\rangle = \frac{\Gamma^2 t}{L}(1 - \frac{\mathsf{v}t}{2L}) \text{ for } t < L/\mathsf{v} \; . \tag{9.83}$$

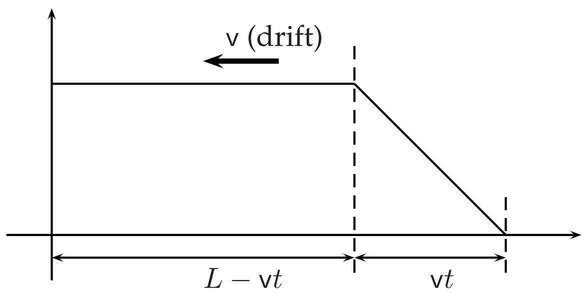


Figure 9.1: The initial growth of the interface is homogeneous, where it has not yet been reached by "reinitialisation". The latter region has length $\mathsf{v}t$.

This result is easily verified. The integral

$$\int_0^1 dy_1 \int_0^1 dy_2 \langle \varphi(y_1,\tau)\varphi(y_2,\tau)\rangle \tag{9.84}$$

has two independent spatial variables, y_1 and y_2. However, very similar arguments apply, so that $n_1 + n_2 = 0$ in (9.69a). The integration over y' gives therefore a contribution of the form $\sqrt{2\pi\tilde{\tau}}\ I\left(\frac{1}{2}(y_1+y_2)+\tilde{\tau}q\right)$. When applying a saddle point approximation for the second term of (9.69a), the integration over y_1 requires $0 \le \frac{1}{2}y_2 - n_1 + n_2 \le \frac{1}{2}$ and therefore again $n_1 - n_2 = 0$. The total contribution of (9.69a) therefore is

$$\sqrt{2\pi\tilde{\tau}}\ I\left(\frac{1}{2}(y_1+y_2)+\tilde{\tau}q\right) e^{-\frac{1}{8\tilde{\tau}}(y_1-y_2)^2} \tag{9.85}$$

where the integration over y_1 is not performed yet, even though we derived from it the condition $n_1 - n_2 = 0$ already.

For Eq. (9.69b) one needs $\tilde{y} = n_1 - n_2 - y_2 \ge 0$ because of the first term and $0 \le \frac{1}{2}(y_1+y_2)+\tilde{y}+\tilde{\tau}q \le 1$ in the third term. The only choice which gives a non-vanishing contribution to the integral is $n_1 - n_2 = 1$. The second term requires $0 \le -(n_1 + n_2 + \frac{1}{2}y_2) \le \frac{1}{2}$ for the integration over y_1, which is satisfied for no choice of $n_1 + n_2$. Thus, the second term (9.69b) does not contribute. The same applies to the third term, Eq. (9.69c).

The analysis of (9.69d) gives: $n_1 + n_2 < 0$ because of the first term. As for (9.69a), the second term gives $0 \le \frac{1}{2}y_2 - n_1 + n_2 \le 1/2$ and therefore $n_2 - n_1 = 0$, thus the sum of n_1 and n_2 must be an even, negative integer. However, if $n_1 + n_2 \le -2$, then the third term has the minimum at $y'_{\min} \ge 2 - \frac{1}{2}(y_1+y_2)+\tilde{\tau}q$, which is unreachable. So, this term does not contribute neither.

After applying a saddle point approximation for the y_1 integration of (9.85), one finds

$$\begin{aligned}
\left\langle \overline{\varphi(y,\tau)}^2 \right\rangle &= \int_0^\tau d\tilde{\tau} \int_0^1 dy_2 \left(\frac{1}{\sqrt{4\pi\tilde{\tau}}}\right)^2 \sqrt{2\pi\tilde{\tau}} I\left(y_2+\tilde{\tau}q\right)\sqrt{8\pi\tilde{\tau}} &\quad (9.86)\\
&= \int_0^\tau d\tilde{\tau} \int_0^1 dy_2 I\left(y_2+\tilde{\tau}q\right)\ , &\quad (9.87)
\end{aligned}$$

where the $(4\pi\tilde{\tau})^{-1}$ comes from (9.60). For $\tau q \ge 1$ the indicator function truncates the integration over $\tilde{\tau}$ at q^{-1}, so that

$$\begin{aligned}
\left\langle \overline{\varphi(y,\tau)}^2 \right\rangle &= \frac{1}{2q} \text{ for } \tau > q^{-1} &\quad (9.88)\\
\left\langle \overline{h(x,t)}^2 \right\rangle &= \frac{\Gamma^2}{2v} \text{ for } t > L/\mathsf{v} &\quad (9.89)
\end{aligned}$$

as already stated in (9.82), while for $\tau q < 1$ one has

$$\begin{aligned}
\left\langle \overline{\varphi(x,t)}^2 \right\rangle &= \tau(1-\frac{1}{2}\tau q)\ \tau < q^{-1} &\quad (9.90)\\
\left\langle \overline{h(x,t)}^2 \right\rangle &= \frac{\Gamma^2 t}{L}(1-\frac{\mathsf{v}t}{2L}) \text{ for } t < L/\mathsf{v} &\quad (9.91)
\end{aligned}$$

as found in (9.83).

Thus, the roughness calculated from a saddle point approximation is[10]

$$w^2(t,L) = \Gamma^2 \times \begin{cases} \sqrt{\frac{t}{2\pi \mathsf{D}}} - \frac{t}{L} - \frac{|\mathsf{v}|t^{3/2}}{3L\sqrt{2\pi\mathsf{D}}} + \frac{t^2|\mathsf{v}|}{2L^2} & \text{for } t \ll \frac{L}{|\mathsf{v}|} \quad (9.92a) \\ \frac{2}{3\sqrt{2\pi}}\sqrt{\frac{L}{|\mathsf{v}|\mathsf{D}}} - \frac{1}{2|\mathsf{v}|} & \text{for } t \gg \frac{L}{|\mathsf{v}|} \quad , (9.92b) \end{cases}$$

where the crossover apparently takes place at

$$t_{\mathsf{X}} = \frac{L}{|\mathsf{v}|} \ . \tag{9.93}$$

or by equating the leading terms in (9.92a) and (9.92b), $t_{\mathsf{X}} = (4/9)L/|\mathsf{v}|$.

As we shall see below (Sec. 9.4.5.1, Eq. (9.137), page 375), all corrections comply to the form presented in (9.45a) and (9.45b). Therefore the exponents found are

$$\chi = 1/4 \text{ and } \beta = 1/4 \text{ and } z = 1 \text{ for } v \neq 0 \text{ fixed boundaries} \tag{9.94}$$

This is a very remarkable result, because these exponents are "classically" impossible, as dimensional analysis allows in general only exponents according to (9.127), see below. Therefore the exponents found in (9.94) are *anomalous*. The details of dimensional analysis are discussed in the next section.

9.4 Dimensional Analysis and Coarse Graining Arguments

The aim of this section is to show how and why standard arguments used in conjunction with dimensional analysis seem to fail. To this end, it is necessary to develop a clear understanding of the way dimensional analysis works. The following discussion is focused on w^2, rather than the correlation function, but can similarly be applied to that as well.

To motivate this section, we will first look at some general misconceptions and a specific misleading argument, which produces the *wrong* answer for the exponents of the EW equation with drift.

[10] The modulus of v appears because of the symmetry consideration Eq. (9.70). Based on them, we would repeat our analysis for negative q and $-q$ would appear where q appears above.

9.4.1 Misconception 1: self-affinity

Considering the initial equation (9.2)

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) + \eta(x,t) , \tag{9.95}$$

the standard "trick" (Barabási and Stanley, 1995; Kardar, 2001) is to impose that the solution h is self-affine, *i.e.* that there are exponents χ and z such that

$$h(x,t) = b^{-\chi} h(bx, b^z t) . \tag{9.96}$$

Plugging this into (9.95), one arrives at

$$b^{z-\chi}\dot{h}(bx, b^z t) = b^{2-\chi}\mathsf{D}h''(bx, b^z t) + \eta(x,t) \tag{9.97}$$

where $\dot{h}(bx, b^z)$ means "derivative of h by time, *evaluated* at $bx, b^z t$" and similarly for h''. Taking into account the self-affinity of the noise[11] (see also Eq. (9.102)),

$$\eta(x,t) = b^{(1+z)/2}\eta(bx, x^z t) \tag{9.98}$$

one arrives at the condition $z - \chi = 2 - \chi = (1+z)/2$ for (9.97) to be equivalent to (9.95). The resulting exponents in one spatial dimension are

$$\chi = 1/2 \quad \text{and} \quad \beta = 1/4 \quad \text{and} \quad z = 2 \tag{9.99}$$

where $\beta = \chi/z$ has been used.

There are several important points of criticism:

- The procedure above has simply no mathematical footing whatsoever. The statement (Barabási and Stanley, 1995, p. 51) "[t]he growth equation [(9.95)] must be invariant under" $x \to x' = bx$ and $h \to h' = b^\chi h$ is simply wrong. The equation is in all non-trivial cases not invariant under this transformation and *a priori* there is no reason for that. However, it might be invariant under this transformation asymptotically.

- There is *a priori* no reason why the noise transforms according to (9.98).

- The procedure above is simply *not dimensional analysis*, even though it is sometimes called that way, see for example (Barabási and Stanley, 1995, p. 315).

[11] In spatial dimension d, the exponent $1+z$ has to be replaced by $d+z$

9.4.1.1 Self-affinity as transformation

There is a way to understand dimensional analysis in the spirit above. As a *result* of dimensional analysis, rather than as the analysis itself, one might be able to show for some *very simple cases*[12] such as (9.95) that the following statement holds: If $h(x,t)$ solves (9.95) for a given $\eta(x,t)$, D and L, then

$$h'(x',t') = b^{\chi} h(x,t) \tag{9.100a}$$
where
$$x' = bx \tag{9.100b}$$
$$t' = b^z t \tag{9.100c}$$

is a solution of (9.95) for $\eta'(x',t')$, D$'$ and L' where

$$\eta'(x',t') = b^{-(1+z)/2}\eta(x,t) \tag{9.101a}$$
where
$$\mathsf{D}' = \mathsf{D} \tag{9.101b}$$
$$L' = bL\ , \tag{9.101c}$$

provided that $\chi = 1/2$ and $z = 2$. So, in fact, the self-affinity is to be understood as a transformation from one solution to another. The point is of course, that one is often looking for exactly that form of transformation: one is looking for the solution of (9.95) for all (in particular large) L, fixed D and given noise correlator. The latter is in fact conserved under the transformation (9.101a) if the noise is δ-correlated:

$$\begin{aligned}\left\langle \eta'(x_1',t_1')\eta'(x_2',t_2')\right\rangle = b^{-(1+z)}\left\langle \eta(x_1,t_1)\eta(x_2,t_2)\right\rangle = \\ b^{-(1+z)}\delta(x_1-x_2)\delta(t_1-t_2) = \delta(bx_1-bx_2)\delta(b^z t_1 - b^z t_2)\end{aligned} \tag{9.102}$$

If that was not the case, (9.100) and (9.101) would merely map families of solutions onto each other, without much use for the problem one is actually interested in, namely (9.95) with fixed noise correlator as said above. Regarding L, the underlying assumption is that the boundary condition transforms properly, for example $h(x = L,t) = h(x=0,t)$ (and h analytical[13] at $x = 0, L$) or $h(x=L,t) = h(x=0,t) = 0$.

Sometimes it is said that h needs to be *statistically* independent under the transformation, without further specifying what that actually means. The proper underlying

[12] In most interesting cases, there is no such simple transformation — for example there is no such relation for the EW equation with drift and periodic boundary conditions, even though its exponents are trivial; if h obeys (9.100), then $\tilde{h}(x,t) = h(x + \mathsf{v}t, t)$ does certainly not, unless $z = 1$. See also the remarks around Eq. (9.18), page 345.

[13] Otherwise the periodic boundary condition could be met by *defining* h correspondingly, $h(x = L,t) = h(x=0,t)$.

idea is that there is a set of solutions $\{h(x,t)\}$ of a PDE like (9.95) for an ensemble of noise realisations $\{\eta(x,t)\}$. Eq. (9.100) then provides us with a transformation rule from one set of solutions to another as L is rescaled to bL, provided that the ensemble of noise terms transforms according to (9.101a).

It is worth stressing that this interpretation of dimensional analysis as a transformation prescription simply misses the point of dimensional analysis: Dimensional analysis is much more fundamental and does in general not produce a simple transformation rule like (9.100).

9.4.1.2 From self-affinity of h to the scaling of the width

The exponents used above, for example in (9.100), have the same symbols as the width (for example (9.43)) for a good reason. From (9.100) follows for the roughness of h', defined in Eq. (9.40), page 351

$$w_{h'}^2(b^z t, bL) = b^{2\chi} w_h^2(t, L) \tag{9.103}$$

where the subscript indicates the underlying field-variable. The roughness of h' is, as mentioned above, the roughness on the rescaled lattice. Thus, w^2 is a generalised homogenous function,

$$w^2(t, L) = aL^{2\chi}\mathcal{G}\left(\frac{t}{bL^z}\right) , \tag{9.104}$$

with roughness exponent χ and dynamical exponent z, see Eq. (9.43).

9.4.2 Misconception 2: coarse-graining arguments

While the first misconception (Sec. 9.4.1) can be salvaged by an interpretation like (Sec. 9.4.1.1), there is another much more dangerous misunderstanding of dimensional analysis. It has developed into one of the standard arguments about relevant and irrelevant terms[14] in LANGEVIN equations, and runs under the label "dimensional analysis", "scaling arguments" or "coarse-graining", even though the argument itself is only loosely related to dimensional analysis.

To illustrate it, one needs another term in (9.95), say the EW equation with drift:

$$\partial_t h(x,t) = \mathsf{D}\partial_x^2 h(x,t) + \mathsf{v}\partial_x h(x,t) + \eta(x,t) \tag{9.105}$$

[14]This misunderstanding is, for example, very explicit in the notion of $[n] < [x]^2$ in (Paczuski and Bassler, 2000), which simply does not make any sense. There is no ordering relation on dimensions themselves; it makes sense to impose $n < x^2$, implying that both objects are measured in the same units. But what could $[n] < [x]^2$ mean? Maybe that the units $[n]$ is measured in is smaller than the unit x^2 is measured in, compared after transforming both to the same units?

Plugging in (9.96) now leads to

$$b^{z-\chi}\dot{h}(bx, b^z) = b^{2-\chi}\mathsf{D}h''(bx, b^z) + b^{1-\chi}\mathsf{v}h'(bx, b^z) + \eta(x,t) \tag{9.106}$$

rather than (9.97). Together with (9.98), the "resulting equation" (compare to (Barabási and Stanley, 1995, appendix B)) is

$$z - \chi = 2 - \chi = 1 - \chi = (1+z)/2 \tag{9.107}$$

which cannot possibly be correct, for example, because $2 - \chi = 1 - \chi$ cannot be obeyed. However, by solving only

$$z - \chi = 1 - \chi = (1+z)/2 \tag{9.108}$$

one arrives at

$$\chi = 0 \quad \text{and} \quad \beta = 0 \quad \text{and} \quad z = 1 \tag{9.109}$$

(see Eq. (9.36b)) suggesting that the term $\mathsf{v}\partial_x h$ in (9.106) is "relevant compared to the diffusion term $\mathsf{D}\partial_x^2 h$, because on large scales, $b \to \infty$, the choice of exponents (9.109) will rescale all terms apart from $\mathsf{D}\partial_x^2 h$ in (9.106) by a factor b, while the latter remains constant." By dividing the whole equation by b it becomes clear that for $b \to \infty$ the diffusion term vanishes. Therefore the convection or drift term is relevant compared to the diffusion term.

9.4.2.1 The conundrum

The latter line of arguments, which is widely accepted, leads to a bizarre conundrum: The "dimensional analysis" in Sec. 9.4.1 gives the correct exponents if no drift is present — Eq. (9.99) agrees with Eq. (9.25). On the other hand, *with* drift, the result Eq. (9.109) obtained in Sec. 9.4.2 is wrong (compared to (9.25) in case of periodic boundaries and (9.94) in case of fixed boundaries, respectively). One might argue now[15] that the reason for the apparent failure of coarse-graining is that the drift term disappears from the problem due to an *additional symmetry*, namely translational invariance (see page 345). So, does coarse-graining prevail if this symmetry is broken, for example by fixing the boundaries? Apparently not, compare Eq. (9.109) to Eq. (9.94). So, why then does coarse-graining fail?

The answer to this question is very simple: There is *a priori* simply no reason why coarse-graining actually *should* work at all. In order to understand that, the following section presents dimensional analysis in a (hopefully) more profound fashion.

[15] That was my line of arguments, initially.

9.4.3 The method of dimensional analysis[16]

The first step in dimensional analysis is to identify the physical dimension of the variables in a problem. The key to the analysis is a "physical covariance principle" (Barenblatt, 1996): The same physical observable might be expressed in different units (a length, for example, in meters or inches) and therefore change its numerical value, but not its physical meaning.

The solution of the EDWARDS-WILKINSON equation with drift (9.2), for example is

$$h(x, t; L, \mathsf{D}, \mathsf{v}, \Gamma) , \tag{9.110}$$

where a δ-correlator (9.5) for η has been assumed. The physical dimension of the variables, denoted by the bracket [ ], are

$$\begin{aligned}
[x] &= \mathbf{L} && (9.111a)\\
[t] &= \mathbf{T} && (9.111b)\\
[h] &= [\eta]\mathbf{T} && (9.111c)\\
[\mathsf{v}] &= \mathbf{L}/\mathbf{T} && (9.111d)\\
[\mathsf{D}] &= \mathbf{L}^2/\mathbf{T} && (9.111e)\\
[\eta^2] &= [\Gamma^2]/(\mathbf{TL}) && (9.111f)
\end{aligned}$$

where the last line comes from the correlator, (9.5). The variable $\mathbf{T}$ stands for a time unit (say second), the variable $\mathbf{L}$ for a unit of length etc. Because the problem is linear, the dimension of Γ is not fixed. If $[h]$ has dimension $\mathbf{H}$, then $[\Gamma^2] = \mathbf{H}^2\mathbf{L}/\mathbf{T}$, so that we can only say that h must be linear in Γ. Changing now the units of length by a factor b (say, going from meters to centimetres, $b = 100$), x is transformed like $x' = bx$, D by $\mathsf{D}' = \mathsf{D}b^2$ etc., but h does not change because it has an independent dimension:

$$h(bx, t; bL, b^2\mathsf{D}, b\mathsf{v}, b^{1/2}\Gamma) = h(x, t; L, \mathsf{D}, \mathsf{v}, \Gamma) \tag{9.112}$$

Instead of following each of the symmetries individually, one can equivalently express h as a function of dimensionless quantities only. These quantities do not change under a change of units:

$$h(x, t; L, \mathsf{D}, \mathsf{v}, \Gamma) = \sqrt{\frac{\Gamma^2 L}{\mathsf{D}}}\varphi\left(\frac{x}{L}, \frac{t\mathsf{D}}{L^2}; \frac{\mathsf{v}}{\mathsf{D}/L}\right) , \tag{9.113}$$

where φ is now also dimensionless. For example $L \to bL$, $\mathsf{D} \to b^2\mathsf{D}$ does not change

[16] A mathematically profound derivation of all results in this section can be found in (Barenblatt, 1996).

$t\mathsf{D}/L^2$.

Eq. (9.113) contains a fairly arbitrary choice of ratios, other choices are possible. However, in the end, h will always be written as a dimensionful pre-factor times a dimensionless function, which depends on exactly three dimensionless quantities. Of course, (9.113) entails (9.112). In fact, (9.113) is simply a statement about the symmetries of a solution of (9.2). Without physical reasoning, it can be obtained by identifying variables, which can be rescaled independently.

The key of dimensional analysis is the identification of dimensionless ratios of dimensionful parameters. Dimensional analysis reduces a problem depending on n independent parameters with $m \leq n$ independent dimensions to a problem with $n-m$ independent variables. With three independent units, L, T and H, the original problem h depending on 6 parameters becomes a problem depending on 3 parameters.

This is all dimensional analysis does, in contrast to what has been anticipated in Sec. 9.4.1; dimensional analysis prescribes symmetries like (9.100), but in a much more general form, (9.113). In fact for $\mathsf{v} = 0$, the latter entails the former as

$$h(bx, b^2t; bL, \mathsf{D}, 0, \Gamma) = \sqrt{b}h(x, t; L, \mathsf{D}, 0, \Gamma) \ . \tag{9.114}$$

Dimensional analysis suggests a much wider, exact symmetry:

$$h(x, t; bL, \mathsf{D}, \mathsf{v}, \Gamma) = b^{\frac{z-1}{2}} h(x/b, t/b^z; L, \mathsf{D}b^{z-2}, \mathsf{v}b^{z-1}, \Gamma) \tag{9.115}$$

for *any* exponent z.

In particular, now it becomes clear what happened in Sec. 9.4.2: Based on (9.115), the limit $b \to \infty$ seems to suggest again, that v is "more important" than D, because the latter has a smaller exponent. But no exact statement is possible. Only for $\mathsf{v} = 0$ (9.115) states that $z = 2$ entails a self-affine h like (9.100),

$$h(bx, b^2t; bL, \mathsf{D}, 0, \Gamma) = b^{1/2}h(x, t; L, \mathsf{D}, 0, \Gamma) \tag{9.116}$$

which, in turn, entails the scaling of the width, as discussed in Sec. 9.4.1.2. The advantage of Eq. (9.116) over Eq. (9.115) is that it allows us to answer the question for a solution of the EW equation on large or actually any spatial scale L, keeping the other parameters, D, $\mathsf{v} = 0$ and Γ, fixed.

While the symmetry relation derived from dimensional analysis predict the exponents properly for $\mathsf{v} = 0$ by self-affinity, there is no such argument for $\mathsf{v} \neq 0$, neither for periodic nor for fixed boundary conditions; for $\mathsf{v} \neq 0$ the solution is *not self-affine*.

It is worth noting (thanks to ANDY PARRY for pointing that out) that D constitutes a dangerously irrelevant variable. Starting from a model with $\mathsf{D} = 0$ but $\mathsf{v} \neq 0$, one observes a divergent width, see next section. Eq. (9.115) suggests that the diffusion term is irrelevant compared to the drift. However, any $\mathsf{D} \neq 0$ leads, in the limit $L \to \infty$, to the exponents (9.94), provided that fixed boundary conditions are applied. With periodic boundary conditions, the standard exponents are (9.57) recovered. Of course, without the drift term, the diffusion term is not irrelevant and then it is not surprising that any $\mathsf{D} \neq 0$ leads to exponents different from $\mathsf{D} = 0$.

9.4.3.1 EW without diffusion

What makes the "apparent result" in Sec. 9.4.2 even worse is the fact that it predicts the wrong exponent for the EW equation without *diffusion* term. If the diffusion term disappears, the process corresponds to random deposition. In fact, the propagator of

$$\partial_t \phi = \mathsf{v} \partial_x \phi + \eta(x, t) \tag{9.117}$$

is just a δ-function in x and the correlator therefore reads

$$\langle \phi(x_1, t_1) \phi(x_2, t_2) \rangle = \Gamma^2 t_1 \delta(x_1 - x_2) \tag{9.118}$$

which leads to a divergent width for any L and t. The value $\beta = 1/2$ found in BARABASI and STANLEY (Barabási and Stanley, 1995) is only due to a finite lattice constant. The correct discussion can be found in KRUG's review (Krug, 1997, Sec. 3.2.2). With an appropriate noise correlator (9.102) and periodic boundary conditions,[17] one can again transform the drift term away, $\phi'(x + vt, t) = \phi(x, t)$ and one is left with

$$\partial_t \phi = \eta(x, t) \tag{9.119}$$

and therefore $z = 0$ in KRUG's terminology which entails that the lattice constant dominates the physics for all dimensions $d > 0$. There is no real roughening as the interface size is increased. [18]

9.4.3.2 Scales in η

It is important to highlight the rôle of η, the correlator of which has not been investigated carefully above. How is it possible that (9.113) entails (9.100) without actually referring to the scaling behaviour of η? The answer is that Eq. (9.113) is of course only

[17] PBC are required for having $\sum_n \delta(x + nL)$ as propagator rather than just $\delta(x)$.

[18] The results of this section can also be obtained by considering the propagator of $\partial_\tau = q\partial_y$, which is $\delta(y + q\tau)$, so that

$$h(x, t) = \int_0^L dx' \int_0^t dt' \eta(x', t') \delta(x - x' + v(t - t'))) = \int_0^t dt' \eta(x + vt', t')$$

which is statistically identical to $\int_0^t dt' \eta(x, t')$.

correct, if the original equation (9.2) is parametrised by exactly the variables listed on the LHS of (9.113). Especially, the only *scale* characterising η is Γ — no complicated correlation function is allowed. However, (9.113) does not make use of the fact that there is a whole ensemble of η-functions, it only states that *if the problem is parametrised by x, t, L, D, v and Γ alone, then a certain symmetry relation, (9.113), holds.* Here, Γ enters only to characterise the dimension of the noise, not to characterise its correlator. Fortunately, one is usually interested in the behaviour of the solutions of Eq. (9.2) for an ensemble of η-functions, which is characterised solely by Γ, see (9.5), and can be transformed from one length scale to the other without involving any additional scales, see (9.100), (9.101). For example, if η was

$$\eta(x,t;\Gamma) = \Gamma \frac{1}{\sqrt{xt}} \tag{9.120}$$

the "roughness exponent" would be the same, provided that w^2 is finite. But it would possibly change for

$$\eta(x,t;\Gamma) = \Gamma \frac{1}{\sqrt{(x_0+x)t}} \tag{9.121}$$

because a new scale, x_0, would enter.

The implications of dimensional analysis are considerable. In the following, dimensional analysis will be used to analyse the width w^2. Since space is integrated out, there is one less variable, so (compare to (9.113))

$$w^2(t,L;\mathsf{D},\mathsf{v},\Gamma) = \frac{\Gamma^2 L}{\mathsf{D}} \tilde{\mathcal{G}} \left(\frac{t\mathsf{D}}{L^2}, \frac{\mathsf{v}}{\mathsf{D}/L} \right) , \tag{9.122}$$

where $\tilde{\mathcal{G}}$ denotes an unknown, dimensionless function.

9.4.4 The case $\mathsf{v} = 0$

For $\mathsf{v} = 0$ — not in the limit of vanishing v, but based on (9.1) — the problem reduces further, because then w^2 is a function only of t, L, D and Γ. Thus

$$w^2(t,L;\mathsf{D},\mathsf{v}=0,\Gamma) = \frac{\Gamma^2 L}{D} \tilde{\mathcal{G}} \left(\frac{t\mathsf{D}}{L^2} \right) \tag{9.123}$$

where the pre-factor $\Gamma^2 L/\mathsf{D}$ can be fixed up to a constant by imposing that it is independent of t, just like as it is imposed in FAMILY-VICSEK scaling, Eq. (9.43). If w^2 is finite for divergent t, it *must necessarily* scale linearly in L, *i.e.* $\chi = 1/2$:

$$\lim_{t\to\infty} w^2(t,L;D,\mathsf{v}=0,\Gamma) = \frac{\Gamma^2 L}{D} \lim_{x\to\infty} \tilde{\mathcal{G}}(x) \; . \tag{9.124}$$

Corrections of the form proposed in Eq. (9.45a) are not possible, because there are simply not enough parameters for a dimensionless function $c_L(L; D, \Gamma)$, because it is not possible to form a dimensionless argument from them; but in order for a dimensionless function to depend on anything, the argument must be dimensionless as well, otherwise it changes value if physical units are changed. Therefore Eq. (9.124) and Eq. (9.56) are necessarily exact.

Similarly, if the thermodynamic limit of w^2 exists and does not vanish, according to (9.123) $\tilde{\mathcal{G}}(x)$ must behave like $\sqrt{x}$ for small arguments. Otherwise there is a non-vanishing power of L left in front of $\tilde{\mathcal{G}}$, so w^2 would either diverge or vanish in the thermodynamic limit. Thus

$$\lim_{L\to\infty} w^2(t, L; D, \Gamma) = \frac{\Gamma^2}{\mathsf{D}} \sqrt{t\mathsf{D}} \lim_{x\to 0} \frac{\tilde{\mathcal{G}}(x)}{\sqrt{x}} \tag{9.125}$$

corresponding to (9.55) where, again, no corrections are possible.

Thus, dimensional analysis is in full agreement with the results for $\mathsf{v} = 0$ above [Eq. (9.57) and Eq. (9.68)], for periodic and fixed boundaries, respectively.

All what was needed in the analysis above was the form (9.123) and in particular that $\tilde{\mathcal{G}}$ depends in only one argument, so that the limit $x \to \infty$ and $x \to 0$ can be taken, producing a particular scaling behaviour only be assuming the very existence of these limits. However, the form (9.123) solely comes from dimensional analysis.[19]

In that sense, the FAMILY-VICSEK scaling of the form

$$w^2(t, L) = aL^{2\chi}\mathcal{G}\left(\frac{t}{bL^z}\right) \tag{9.126}$$

is too naïve — if the number of independent arguments is so small, the exponents are fixed by dimensional analysis only, see Sec. 9.4.3. One can prove, that any exponent χ derivable from dimensional analysis must obey (Krug, 1997)

$$\chi = \frac{z-1}{2} \,. \tag{9.127}$$

[19]For example, w^2 is linear in L only because of dimensional requirements. Even if one accepts an alternative form to (9.123),

$$w^2(t, L) = \frac{\Gamma^2 L}{\mathsf{D}} \frac{t\mathsf{D}}{L^2} \tilde{\tilde{\mathcal{G}}} \left(\frac{t\mathsf{D}}{L^2}\right)$$

the existence of the limit $t \to \infty$ then requires $\tilde{\mathcal{G}} \propto x^{-1}$, so that $\chi = 1/2$ again.

9.4.5 The case $\mathsf{v} \neq 0$

In case of $\mathsf{v} \neq 0$ things are not that simple any more. The roughness now acquires an extra argument:

$$w^2(t, L; \mathsf{D}, \mathsf{v}, \Gamma) = \frac{\Gamma^2 L}{\mathsf{D}} \tilde{\mathcal{G}} \left(\frac{t\mathsf{D}}{L^2}, \frac{\mathsf{v}L}{\mathsf{D}} \right) \tag{9.128}$$

The arguments like (9.124) cannot be applied anymore - it is unknown, how the function $\tilde{\mathcal{G}}$ behaves in the various limits. If, for example[20]

$$\lim_{x\to\infty} \tilde{\mathcal{G}}(x, y) = y^{2\mu} \tag{9.129}$$

which is dimensionally completely consistent, because $\tilde{\mathcal{G}}$, x and y are dimensionless, then

$$\lim_{t\to\infty} w^2(t, L; \mathsf{D}, \mathsf{v}, \Gamma) = \frac{\Gamma^2 L}{\mathsf{D}} \left(\frac{\mathsf{v}L}{\mathsf{D}} \right)^{2\mu} \tag{9.130}$$

and therefore[21] $\chi = \mu + 1/2$. The same problems appear for $L \to \infty$; it is simply *impossible* to determine the exponents from dimensional analysis alone in the presence of *additional length scales*.

9.4.5.1 Additional length scales

One cannot stress enough the rôle of such an additional length scale: If an additional, independent length enters the problem, it can form a dimensionless ratio in conjuction with L. Similarly for other parameters. For example, if there was a further length scale l_0 in the problem, then the relation (9.128) would read

$$w^2(t, L; \mathsf{D}, \mathsf{v}, \Gamma, l_0) = \frac{\Gamma^2 L}{\mathsf{D}} \tilde{\mathcal{G}} \left(\frac{t\mathsf{D}}{L^2}, \frac{\mathsf{v}L}{\mathsf{D}}, \frac{L}{l_0} \right) \tag{9.131}$$

and $\tilde{\mathcal{G}}(x, y, z)$ could behave like

$$\lim_{x\to\infty} \tilde{\mathcal{G}}(x, y, z) = y^{2\mu} z^{2\lambda} , \tag{9.132}$$

leading to an exponent $\chi = \mu + \lambda + 1/2$

It is worth stressing that a change of boundary conditions or initial conditions might introduce new length scales or remove others. For example, if the noise would have a net contribution[22], $\langle\eta\rangle \neq 0$, that would give rise to new scale, $\langle\eta\rangle/(\Gamma\sqrt{L\mathsf{D}})$.

---

[20] Note, however, that this limit exists and is analytical.

[21] The behaviour of the function $\tilde{\mathcal{G}}$ is probably not completely arbitrary, because one would expect some crossover from $\mathsf{v} = 0$ to $\mathsf{v} \neq 0$, depending on the dimensionless parameter $\mathsf{v}L/\mathsf{D}$ as discussed above (see Eq. (9.93)).

[22] The scales due to the noise are obtained only when taking ensemble averages. Otherwise, a specific realisation has in general infinitely many independent parameters, namely the entire function $\eta(x, t)$.

However, in the case of $\mathsf{v} = 0$ the exponents remain unchanged, if periodic boundaries are replaced fixed boundaries, $h(x = 0, t) = h(x = L, t) = 0$; the exact solution changes, but as the new boundary condition does not contain any new non-zero parameter and therefore cannot introduce any new scale, the scaling form Eq. (9.123) *must* necessarily remain unchanged; there are just not enough (independent) parameters for the form Eq. (9.123) to change.

The reason for the change of the exponents as fixed boundaries are introduced in case of $\mathsf{v} \neq 0$ is *not* that the fixed boundaries introduce a further length scale. They only avoid the additional length scale, D/v, to disappear from the problem.

In fact, *only in the presence of additional length scales such as* D/v, *a behaviour different from trivial dimensional analysis is actually possible.* There are, in principle, three different scenarios if further length scales (or in general further independent parameters) enter the problem.

i) The exponents do not change, *i.e.* the new parameter does not actually change the problem.

ii) More importantly, the resulting behaviour could be *non-critical*; for example, if

$$\lim_{x\to\infty} \tilde{\mathcal{G}}(x, y) = e^{-y} \tag{9.133}$$

then

$$\lim_{t\to\infty} w^2(t, L; \mathsf{D}, \mathsf{v}, \Gamma) = \frac{\Gamma^2 L}{\mathsf{D}} e^{-\frac{\mathsf{v}L}{\mathsf{D}}} \ . \tag{9.134}$$

where D/v is a saturation length for the roughness. Note that this is perfectly compatible to $\chi = 1/2$ at $\mathsf{v} = 0$.

iii) A new exponent appears, which differs from the one obtained by dimensional analysis, as has been discussed in (9.129). This is, in fact, what happened in Eq. (9.92a): For large $x = t\mathsf{D}/L^2$, the scaling function $\tilde{\mathcal{G}}(x, y)$ behaves like $y^{-1/2}$, provided that y is large enough, *i.e.* in the limit[23] $y \to \infty$. Conclusively, $\mu = -1/4$ in the notation of (9.129) and the new roughness exponent is $\chi = 1/4$.

Moreover, an additional scale allows corrections of the form discussed in (9.45). It is very instructive to rewrite (9.92) in the form (9.128)

$$\lim_{L\to\infty} w^2(t, L) = \Gamma^2 \frac{L}{\mathsf{D}} \left(\frac{1}{\sqrt{2\pi}} \sqrt{\frac{t\mathsf{D}}{L^2}} \right) \tag{9.135a}$$

$$\lim_{t\to\infty} w^2(t, L) = \Gamma^2 \frac{L}{\mathsf{D}} \left(\frac{2}{3\sqrt{2\pi}} \sqrt{\frac{\mathsf{D}}{L|\mathsf{v}|}} - \frac{\mathsf{D}}{2L|\mathsf{v}|} \right) \tag{9.135b}$$

That is the reason, why we had to say above (first paragraph of Sec. 9.4.3.2) that η is solely parametrised by Γ. See also the remark at the end of Sec. 9.4.6.

[23] This has been used explicitly in the saddle point approximation, $q \gg 1$, Sec. 9.3.2.

so that the scaling function in (9.128) has the following two limits

$$\lim_{x\to 0} \tilde{\mathcal{G}}(x, y' x^{-1/2}) x^{-1/2} = \frac{1}{\sqrt{2\pi}} \tag{9.136a}$$

$$\lim_{x\to\infty} \tilde{\mathcal{G}}(x, y) = (\frac{2}{3\sqrt{2\pi}}|y|^{-1/2} - \frac{1}{2}|y|^{-1} \tag{9.136b}$$

where according to (9.128) $x = t\mathsf{D}/L^2$ and $y = \mathsf{v}L/\mathsf{D}$. In (9.136a) the new parameter is $y' = \mathsf{v}\sqrt{t/\mathsf{D}}$, but in the limit $\tilde{\mathcal{G}}$ is independent of it anyway; in the initial growth phase, the drift is invisible and therefore (9.55) (initial growth in the thermodynamic limit with periodic boundaries) is identical to (9.135a) (initial growth in the thermodynamic limit fixed boundaries).

According to (9.136b), w^2 acquires some extra corrections in the stationary state of order $y^{-1} \propto 1/L$. They have been included already in (9.45a), so that apparently

$$\lim_{t\to\infty} w^2(t, L) = \underbrace{\frac{2}{3\sqrt{2\pi}} \frac{\Gamma^2}{\sqrt{\mathsf{vD}}}}_{a_L} \quad \underbrace{L^{1/2}}_{L^{2\chi}} \quad \underbrace{\left(1 - \frac{3\sqrt{2\pi}}{4} \frac{\mathsf{D}}{L\mathsf{v}}\right)}_{1+c_L\left(\frac{L\mathsf{v}}{\mathsf{D}}\right)} \tag{9.137}$$

It might look surprising that the expression does not seem to suggest a proper crossover to $\chi = 1/2$ as v vanishes. The reason for that is twofold: First of all, the derivation is based on a saddle point approximation, which becomes exact only as $q = L\mathsf{v}/\mathsf{D}$ diverges. So, the limit $\mathsf{v} \to 0$ produces a correct result if the thermodynamic limit is taken first, which allows access to the initial growth phase, see Eq. (9.135a), identical to (9.55).

9.4.6 Summary

To summarise this section, dimensional analysis is an exact method, which helps to reduce the number of free parameters in a problem. However, one must take great care to include *all* scales present in a problem, such as all parameters in

- in the original PDE
- the boundary conditions
- the initial conditions
- the noise term

The latter includes all parameters characterising the noise, which relates to the discussion on page 270. The only reason why the noise itself does not enter as a scale (or actually the entire ensemble), is that it is *assumed* to be parametrised *solely* by the

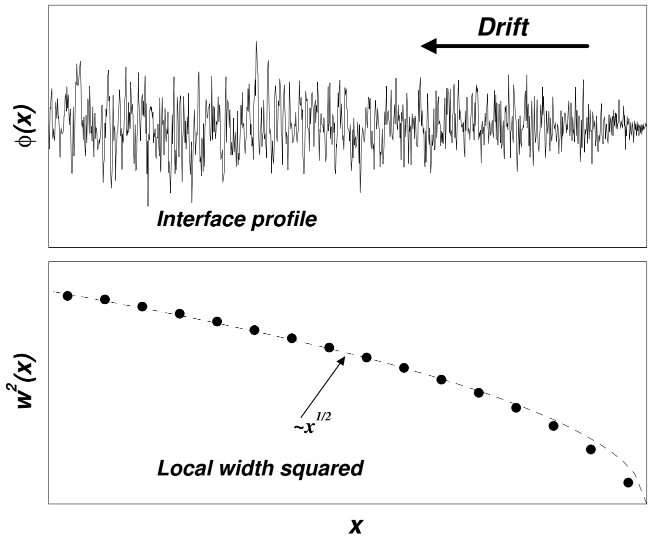


Figure 9.2: A qualitative picture of an interface snapshot with its "local roughness"; scales are irrelevant. Upper panel: An example of an interface profile with fixed boundaries and drift term. Lower panel: The ensemble averaged *local* width squared (numerical data, circles) is proportional to the square root of the position where measured, $(L-x)^{1/2}$ (fitted, dashed line).

scales introduced in the last point. For example, $\eta(0,0)$ could in principle provide a valid scale.

If there are sufficiently few parameters in a problem, dimensional analysis predicts the exponents. Otherwise, it does not make any statement, even if the functions involved have a proper, analytic, smooth asymptotic behaviour.

9.5 Physical Explanation

After this detour, we want to continue to understand the physical origin of the anomalous exponent found in the EW equation with drift.

The fact that the results in Sec. 9.3 cannot be derived directly from dimensional analysis and require tedious algebra, might suggest that the mechanism leading to these exponents is very subtle. However, it turns out that it can be understood quite easily.

The drift term makes the entire interface configuration move from one boundary to the other. Without noise, a peak starting somewhere in the bulk gets slowly moved by the drift to one of the boundaries, while diffusively broadening. It eventually disappears at the boundary. The time it spends between the boundaries depends on the starting position and the direction of the movement. The maximum time is L/v, which is also the maximum time, any noise-generated structure has to develop. However, as known from the model without drift, it takes time L^2/D in order to fully develop the roughness. Thus, for $L^2/\mathsf{D} \gg L/\mathsf{v}$, *i.e.* $L \gg \mathsf{D}/\mathsf{v} = L_\times$, the interface will remain in its initial growth phase; the characteristic length scale $L_\times$ represents an effective cutoff for correlations. At the same time it enables the system to display anomalous exponents. Depending on the direction of the drift, the interface "comes out" of the right boundary initialised to $h = 0$ and moves to the left boundary, as shown in Fig. 9.2. The average "age" is proportional to $t_\times = L/\mathsf{v}$, so that according to (9.43) $w^2 \propto t_\times^{1/2} \propto L^{1/2}$, therefore $\chi = 1/4$. Regarding β, the interface cannot "see" the drift initially, so that $\beta = 1/4$ just like for the case without drift. Indeed, even the amplitude of the leading term in (9.92a) corresponds to the amplitude obtained for the problem without drift, see (9.67) and also the problem with periodic boundary conditions (9.55). The identity $\chi = \beta$ already indicates $z = 1$, which can also be derived from the fact that saturation should be reached as soon as the interface has swiped through the system once, *i.e.* after $t_\times = L/\mathsf{v}$.

The mechanism is illustrated in Fig. 9.2: The upper panel shows a snapshot of an interface configuration. The "local age" of the interface can be read off the local roughness (in an appropriate *ad hoc* definition, as the width within a small window around a particular position x) as shown in the panel below, because spending more time between the boundaries increases the local roughness according to $w^2 \propto t^{1/2} = ((L - x)/\mathsf{v})^{1/2}$, with x being the position where the roughness has been measured.

9.6 Discussion and Conclusion

The physical explanation presented above goes beyond the EW equation; provided that the crossover time of the original model without drift scales faster in L than $t_\times$, *i.e.* $z > 1$, the argument should apply, so that at sufficiently large system sizes χ obtains the value of β and therefore $z = 1$. It is a very efficient mechanism, which works under very general circumstances even in the most simple, linear case. It therefore speaks a clear warning as to the interpretation of numerical and experimental studies: the true value of χ might have been "washed away" by a very small drift.

Especially, one expects the KPZ (Kardar *et al.*, 1986; Krug and Spohn, 1991; Halpin-Healy and Zhang, 1995; Krug, 1997) equation ($z = 3/2$) to show this behaviour. Using one of the standard methods to analyse it (COLE-HOPF transformation) (Halpin-Healy and Zhang, 1995) the problem boils down to an equation very similar to those discussed above. However, a quick numerical check did not unambiguously confirm this. What makes this case particularly interesting is the fact that the non-linearity of the standard KPZ equation does not lead to a stationary ensemble being qualitatively different from the standard EW one; the non-linearity is only important during the initial growth phase ($\beta = 1/3$) and becomes insignificant in the stationary regime ($\chi = 1/2$). However, by adding an additional drift term, according to the physical explanation above, the equation would effectively remain in the initial growth phase, thereby providing a stationary state which is strongly influenced by the presence of the non-linearity.

It is not yet completely clear how to generalise this arguments to higher dimensions. For two dimensions it is tempting to speculate whether exponents observed in experimental molecular beam epitaxy are related to such a drift term, for example when $\chi \approx \beta$ (You *et al.*, 1993) or when χ is close to $\beta_{\mathrm{KPZ}} \approx 0.24$ (Kardar, 2001; Eklund *et al.*, 1991). Interestingly, only one boundary needs to be fixed in order to observe the phenomenon, namely the boundary perpendicular to the velocity. In one dimension, it is the boundary the velocity points away from, in Fig. 9.2 the right boundary.

The mechanism stresses once more the relevance of boundary conditions as prominently pointed out by DAVID LANDAU and KURT BINDER (Landau and Binder, 1988). However, it is worth emphasising that in the present case, the change of boundary conditions leads to a change of the *bulk* critical exponents.

The argument does not extend to models with quenched noise in an obvious way, because the interface might be pinned locally so that the horizontal movement does not apply uniformly to the entire interface, thereby possibly introducing non-linearities and interactions, which are not present originally. Moreover, as different parts of the interface will experience the *same* noise when sweeping over the surface horizontally, effectively the noise acquires some extra correlations also not present in the model without drift term. For example, in the qEW, the growth exponent without drift is $\beta = 0.88(2)$ (Leschhorn, 1993; Leschhorn *et al.*, 1997), quite different from the roughness exponent observed in the qEW with drift, $\chi = 1/2$. See also the discussion below.

In conclusion we have presented a remarkably simple mechanism which reduces the roughness exponent to the value of the growth exponent for any small amount of drift in the LANGEVIN equation in the presence of fixed boundary conditions, provided that in the original model, the dynamical exponent z is larger than unity. On sufficiently large scale, this mechanism should be visible in many experimental and numerical systems. Most unexpectedly, it can even be found in the EDWARDS-WILKINSON equation, which consequently shows anomalous exponents, depending on the boundary conditions imposed.

9.6.1 Relation to the qEW

The behaviour of the quenched EW (qEW) equation under drift was the initial motivation (see Chapter 7, especially Sec. 7.7, page 310) of the study presented above. However, it is now clear that the mechanisms responsible for the exponents in the qEW equation must be different from those obtained here for the standard EW equation with thermal noise. Most significantly, the drift term affects the roughness ex-

Boundaries	Drift	χ for EW	χ for qEW	
Periodic boundaries	$v = 0$	$1/2$	$1.25(1)$	(Leschhorn, 1993)
Periodic boundaries	$v \neq 0$	$1/2$	$1/2$	(Tang *et al.*, 1995)
Fixed boundaries	$v = 0$	$1/2$	$1.25(2)$	Chapter 5
Fixed boundaries	$v \neq 0$	$1/4$	$1/2$	Chapter 7

Table 9.1: The roughness exponents observed for the EW equation with thermal and with quenched noise. Most remarkably, the drift term changes the roughness exponent to $1/2$ in the qEW equation independent from the boundary condition. In case of periodic boundary conditions, it is identical to the value for thermal noise.

ponent of the qEW equation with FBC *and* PBC (see Tab. 9.1) in the same way, very different from the thermal EW equation.

As discussed already in Sec. 7.7, from the values of the exponents obtained, it might seem reasonable to argue that the drift term converts the quenched noise effectively to a thermal noise. This conversion can be obtained by a Galilean transformation. However, apart from those presented in Sec. 7.7, page 309, there are several other reasons, why this does not explain the exponents, as discussed in the following.

First of all, as already stated above, such a transformation cannot be done for the case of FBC without introducing highly non-trivial boundary conditions. So, if the drift would make the noise in the qEW equation thermal, this could only explain the behaviour of the qEW with PBC. However, as mentioned above, the same exponents are obtained for PBC and FBC. In case of PBC the conversion of quenched noise would in turn suggest that the roughness of the interface changes if the observer changes to a comoving frame, which does not make much sense. Secondly, if nevertheless one accepts that the noise in the qEW equation with FBC becomes indeed thermal by a drift, one would expect the exponents of the EW equation with FBC and drift. This, however, does not correspond to the exponents observed. So, conclusively, for PBC, the drift term in the qEW must work completely differently compared to the drift term in the EW equation.

Ironically, as seen above, the roughness exponent χ of the EW equation with drift corresponds to the growth exponent β of the EW equation without drift. Unfortunately, no such correspondence can be found in case of a quenched noise.

9.6.2 Summary

Motivated by the observation made in Chapter 7 that an additional drift term in the quenched EW equation leads to a model, the universal properties of which can be obtained exactly (Chapter 8), in this chapter the effect of a drift term to the EW equation with thermal noise has been studied.

- After the introduction, Sec. 9.2 discusses the effect of a drift term in the EW equation with periodic boundary conditions. Correlator and width are calculated in detail. It transpires that the drift term does not change the scaling of the width.

- In Sec. 9.3 the effect of a drift term in a system with fixed boundary conditions is discussed. It becomes clear that the exponents, characterising the scaling of the roughness, are now affected by a drift term.
- To understand the apparent failure of dimensional analysis, Sec. 9.4 discusses this technique in detail and points out some possible misconceptions. Relating back to the preceeding sections, it becomes evident why the EW equation with drift is not accessible to dimensional analysis.
- The effect of the drift term can be understood on physical grounds quite easily, as is shown in Sec. 9.5. The drift leads to a continuous reinitialisation of the interface, so that it remains in its initial growth phase with an age proportional to $t_{\mathsf{X}} = L/\mathsf{v}$. Therefore, the width saturates at $w^2 \propto (L/\mathsf{v})^{2\beta}$, effectively converting the growth exponent β into the roughness exponent χ.
- Finally, Sec. 9.6 contains a summary of the chapter and links back to the quenched EDWARDS-WILKINSON equation as discussed in Chapter 6.

Summary and Outlook

An investigation of SOC models can only be as good as the underlying methods of analysis. In this thesis, a comparison has been made with classical critical phenomena in order to benefit from established techniques in that field. It also makes clearer in which sense SOC is really different from traditional criticality. Furthermore, SOC itself might help to illuminate the hitherto elusive problems of non-equilibrium statistical mechanics.

Starting from the basic question concerning whether there is anything that could justifiably be called SOC, the first step was to find a robust model which shows all the expected features. While the Forest Fire model might look very appealing because of its close relation to standard percolation, it turns out to be yet another instance of a class of ill-behaved models, showing no sign of convergence towards simple scaling.

Fortunately, it transpires that the Oslo model is an almost ideal model of SOC: It is very simple, resembles features of a sandpile and possesses stochasticity. More importantly, it shows truly universal behaviour and this thesis has demonstrated how to link it to the quenched EDWARDS-WILKINSON (qEW) equation. The latter is a known problem in the field of non-equilibrium critical phenomena and thus by the relation to the Oslo model it can be turned into a self-organised system.

The universal features of the Oslo model allow us to introduce anisotropy and to observe the crossover behaviour in the thermodynamic limit. Due to the link between the Oslo model and the qEW equation, any result obtained for the Oslo model can also be used in the examination of this equation. As discussed in this thesis, the exact solution of the totally asymmetric Oslo model represents an entire universality class, which in particular includes the qEW equation with drift.

Studying the effect of a drift term in the thermal EDWARDS-WILKINSON equation then allows us to understand the rôle of such a term. Ironically, in the thermal case, there is a close relationship between the equation with and without drift, but no such direct connection exists in the quenched case.

Two very promising perspectives have emerged from this thesis:
i) It seems worthwhile to push the exact solutions a bit further. The MANNA model has a major advantage and a major disadvantage in this respect: it has a simpler matrix representation and might be tractable as a fermionic problem. At the same time,

it has a much more complicated algebraic structure.
ii) There are many interesting open questions regarding the boundary driven qEW equation, *i.e.* the SOC version of the qEW. Developing the field-theory of the qEW a bit further, the Oslo model, and in particular the exact solution of the totally asymmetric Oslo model. seems to be the perfect toy model for such a theory.

SOC is a complex problem — in many ways. It would not by very wise to withdraw from the ill-behaved models completely: they *are* abundant and do have something in common, most notably distribution functions which show intermediate, effective power laws. The significance of these power laws might be debatable, yet they cover many orders of magnitude. So, it is important to develop a clearer understanding of those ill-behaved models. However, critical phenomena, as currently understood, might not be the most suitable framework.

References

The numbers in brackets at the end of each entry refer to the page where the entry is cited. That way, it can also be used as an index.

Aegerter, C. M., R. Günther, and R. J. Wijngaarden, 2000, *Avalanche dynamics, surface roughening, and self-organized criticality: Experiments on a three-dimensional pile of rice*, Phys. Rev. E **67**, 051306-1–6, `cond-mat/0305592`. [205]

Aizenman, M., 1997, *On the Number of Incipient Spanning Clusters*, Nucl. Phys. B **485**, 551–582, `cond-mat/9609240v4`. [60]

Aji, V., and N. Goldenfeld, 2001, *Fluctuations in Finite Critical and Turbulent Systems*, Phys. Rev. Lett. **86**(6), 1007–1010. [81]

Akhiezer, N. I., 1965, *The Classical Moment Problem and Some Related Questions in Analysis* (Oliver and Boyd, Edinburgh), engl. transl. [75]

Alava, M., 2002, *Scaling in self-organized criticality from interface depinning?*, J. Phys.: Condens. Matter **14**, 2352–2360, `cond-mat/0204226`. [247, 249]

Alava, M., 2003, *Self-organized criticality as a phase transition*, `cond-mat/0307688`, (unpublished). [223, 276, 277]

Alava, M. J., and K. B. Lauritsen, 2001, *Quenched noise and over-active sites in sandpile dynamics*, Europhys. Lett. **53**(5), 563–569, `cond-mat/0002406v2`. [275]

Aluru, S., G. M. Prabhu, and J. Gustafson, 1992, *A random number generator for parallel computers*, Parallel Computing **18**, 839–847. [172]

Amaral, L. A. N., A. L. Barabási, H. A. Makse, and H. E. Stanley, 1995, *Scaling properties of driven interfaces in disordered media*, Phys. Rev. E **52**(4), 4087–4104. [267]

Anderson, T. W., 1964, *The Statistical Analysis of Time Series* (John Wiley & Sons, London, New York, Sydney). [101, 163]

Anisimov, M. A., E. Luijten, V. A. Agayan, J. V. Sengers, and K. Binder, 1998, *Shape of Crossover Between Mean-Field and Asymptotic Critical Bheavior in a Three-Dimensional Ising Lattice*, `cond-mat/9810252v2`, (unpublished). [36]

Avnir, D., O. Biham, D. Lidar, and O. Malcai, 1998, *Is the Geometry of Nature Fractal?*, Science **279**, 39–40. [38]

Bak, P., 1996, *how nature works* (Copernicus, New York, NY, USA). [39, 40, 57, 253]

Bak, P., K. Chen, and C. Tang, 1990, *A forest-fire model and some thoughts on turbulence*, Phys. Lett. A **147**, 297–300. [47, 140]

Bak, P., K. Christensen, L. Danon, and T. Scanlon, 2002, *Unified Scaling Law for Earthquakes*, Phys. Rev. Lett. **88**, 178501-1–178501-4. [39]

Bak, P., and K. Sneppen, 1993, *Punctuated Equilibrium and Criticality in a Simple Model of Evolution*, Phys. Rev. Lett. **71**(24), 4083–4086. [52]

Bak, P., C. Tang, and K. Wiesenfeld, 1987, *Self-Organized Criticality: An Explanation of 1/f Noise*, Phys. Rev. Lett. **59**(4), 381–384. [19, 40, 208, 249, 276, 315, 334]

Bak, P., C. Tang, and K. Wiesenfeld, 1988, *Self-Organized Criticality*, Phys. Rev. A **38**(1), 364–374. [40]

Ball, P., 2001, *Long waits will always try patients*, published in Nature science update (online journal), 5 April 2001, comments (Smethurst and Williams, 2001), URL `http://www.nature.com/nsu/010404/010404-16.html`. [39]

Barabási, A.-L., and H. E. Stanley, 1995, *Fractal concepts in surface growth* (Cambridge University Press, Cambridge, UK). [38, 266, 364, 367, 370]

Barber, M. N., 1978, *Non-universality in the Ising model with nearest and next-nearest neighbour interactions*, J. Phys. A: Math. Gen. **12**(5), 679–688. [218]

Barenblatt, G. I., 1996, *Scaling, self-similarity, and intermediate asymptotics* (Cambridge University Press, Cambridge, UK). [368]

Barrat, A., A. Vespignani, and S. Zapperi, 1999, *Fluctuations and Correlations in Sandpile Models*, Phys. Rev. Lett. **83**(10), 1962–1965. [61]

ben-Avraham, D., and S. Havlin, 2000, *Diffusion and Reactions in Fractals and Disordered Systems* (Cambridge University Press, Cambridge, UK). [331]

Bengrine, M., A. Benyoussef, A. E. Kenz, M. Loulidi, and F. Mhirech, 1999a, *A sharp transition between a trivial 1D BTW model and self-organized critical rice-pile model*, Eur. Phys. J. B **12**, 129–133. [217, 246, 249]

Bengrine, M., A. Benyoussef, F. Mhirech, and S. D. Zhang, 1999b, *Disorder-induced phase transition in a one-dimensional model of rice-pile*, Physica A **272**, 1–11. [217, 246, 249]

Bentley, R. A., and H. D. G. Maschner, 1999/2000, *Subtle nonlinearity in popular album charts*, Adv. Complex Syst. **2**(3), 197–208, preprint from `http://www.sit.wisc.edu/~rabentley/Bentley&Maschner_1999.pdf`. [19]

Berg, B. A., 1992, *Double Jackknife bias-corrected estimators*, Comp. Phys. Com. **69**, 7–14. [103]

Bhattacharjee, S. M., and F. Seno, 2001, *A Measure of data collapse for scaling*, J. Phys. A: Math. Gen. **34**(33), 6375–6380. [43, 93]

Binney, J. J., N. J. Dowrick, A. J. Fisher, and M. E. J. Newman, 1998, *The Theory of Critical Phenomena* (Clarendon Press, Oxford). [140]

Biswas, P., A. Majumdar, A. Metha, and J. K. Bhattacharjee, 1998, *In search of smooth sandpiles: The Edwards-Wilkinson equation with flow*, Physica A **248**, 379–392. [349, 350, 351]

Botet, R., and M. Płoszajczak, 2000, *Universal features of the order-parameter fluctuations: Reversible and irreversible aggregation*, Phys. Rev. E **62**(2), 1825–1841. [81]

Boulter, C. J., and G. Miller, 2003, *Nonuniversality and scaling breakdown in a nonconservative earthquake model*, Phys. Rev. E **68**, 056108-1–6. [111, 139, 315]

Brandt, S., 1998, *Data Analysis* (Springer-Verlag, Berlin Heidelberg New York). [101, 103]

Bröker, H.-M., and P. Grassberger, 1997, *Random neighbour theory of the Olami-Feder-Christensen earthquake model*, Phys. Rev. E **56**(4), 3944–3952. [47, 113, 114, 116, 123]

Bruinsma, R., and G. Aeppli, 1984, *Interface Motiona dn Nonequilibrium Properties of the Random-Field Ising Model*, Phys. Rev. Lett. **52**(17), 1547–1550. [246, 271]

Cafiero, R., V. Loreto, L. Pietronero, A. Vespignani, and S. Zapperi, 1995, *Local rigidity and Self-Organized Criticality for Avalanches*, Europhys. Lett. **29**(2), 111–116. [249]

Cardy, J., 1992, *Critical percolation in finite geometries*, J. Phys. A: Math. Gen. **25**, L201-L206, `hep-th/9111026`. [60]

Cardy, J., 1997, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, UK). [66, 75, 77]

Cardy, J. L. (ed.), 1988, *Finite-Size Scaling* (North-Holland, Amsterdam). [73]

Carreras, B. A., D. E. Newman, I. Dobson, and A. B. Poole, 2002, *Evidence for self-organized criticality in a time series of electric power system blackouts*, submitted to IEEE Transactions on Circuits and Systems, Part 1, May 2002, preprint from `http://eceserv0.ece.wisc.edu/~dobson/PAPERS/carrerasCAS02preprint.pdf`, (unpublished). [19]

Chabanol, M.-L., and V. Hakim, 1997, *Analysis of a dissipative model of self-organized criticality with random neighbors*, Phys. Rev. E **56**(3), R2343–R2346. [47, 114]

Chen, Y., and M. E. H. Ismail, 1997, *Ladder operators and differantial equations for orthogonal polynomials*, J. Phys. A: Math. Gen. **30**, 7817–7829. [132]

Chen, Y., and N. Lawrence, 1999, *Small eigenvalues of large Hankel matrices*, J. Phys. A: Math. Gen. **32**, 7305–7315. [75]

Christensen, K., A. Corral, V. Frette, J. Feder, and T. Jøssang, 1996, *Tracer Dispersion in a Self-Organized Critical System*, Phys. Rev. Lett. **77**(1), 107–110. [205, 206, 246, 283, 306, 315, 316, 317]

Christensen, K., H. Flyvbjerg, and Z. Olami, 1993, *Self-Organized Critical Forest-Fire Model: Mean-Field Theory and Simulation Results in 1 to 6 Dimensions*, Phys. Rev. Lett. **71**(17), 2737–2740. [51, 113, 125, 140, 189]

Christensen, K., H. C. Fogedby, and H. J. Jensen, 1991, *Dynamical and Spatial Aspects of Sandpile Cellular Automata*, J. Stat. Phys. **63**(3/4), 653–684. [40, 147]

Christensen, K., D. Hamon, H. J. Jensen, and S. Lise, 2001, *Comment on "Self-Organized Criticality in the Olami-Feder-Christensen Model"*, Phys. Rev. Lett. **87**, 039801. [119]

Christensen, K., N. Moloney, O. Peters, and G. Pruessner, 2004, *Avalanche Behavior in an Absorbing State Oslo Model*, `cond-mat/0405454`, (unpublished). [31, 64]

Chua, A., and K. Christensen, 2002, *Exact enumeration of the Critical States in the Oslo Model*, `cond-mat/0203260v2`, (unpublished). [239]

Clar, S., B. Drossel, and F. Schwabl, 1994, *Scaling laws and simulation results for the self-organized critical forest-fire model*, Phys. Rev. E **50**(2), 1009–1018. [51, 126, 140, 141, 142, 144, 145, 147, 150, 173, 176, 180, 189, 192]

Clar, S., B. Drossel, and F. Schwabl, 1996, *Forest Fires and other examples of self-organized criticality*, J. Phys. C: Condens. Matter **8**, 6803–6824. [49, 50, 62, 125, 145, 146, 149, 173, 176, 190]

Coddington, P. D., 1996, *Random Number Generators for Parallel Computers*, NHSE Review ™ (2), URL `http://www.crpc.rice.edu/NHSEreview/RNG/`. [172]

Cormen, T. H., C. E. Leiserson, and R. L. Rivest, 1990, *Introduction to Algorithms* (The MIT Press, Cambridge, MA, London, England). [143, 154, 233]

Corral, A., and M. Paczuski, 1999, *Avalanche Merging and Continuous Flow in a Sandpile Model*, Phys. Rev. Lett. **83**(3), 572–575. [61, 246]

Datta, A. S., K. Christensen, and H. J. Jensen, 2000, *On the physical relevance of extremal dynamics*, Europhys. Lett. **50**(2), 162–168. [54, 246, 315]

De Menech, M., A. L. Stella, and C. Tebaldi, 1998, *Rare events and breakdown of simple scaling in the Abelian sandpile model*, Phys. Rev. E **58**(3), R2677–R2680. [100, 186]

Derrida, B., E. Domany, and D. Mukamel, 1992, *An Exact Solution of a One-Dimensional Asymmetric Exclusion Model with Open Boundaries*, J. Stat. Phys. **69**(3/4), 667–687. [316]

Derrida, B., and M. R. Evans, 1997, in *Nonequilibrium Statistical Mechanics in One Dimension*, edited by V. Privman (Cambridge University Press, Cambridge, UK), pp. 277–304. [226, 244, 316, 339]

Derrida, B., M. R. Evans, V. Hakim, and V. Pasquier, 1993, *Exact solution of a 1D asymmetric exclusion model using a matrix formulation*, J. Phys. A: Math. Gen. **26**, 1493–1517. [316]

Dhar, D., 1990, *Self-Organized Critical State of Sandpile Automaton Models*, Phys. Rev Lett. **64**(14), 1613–1616. [43, 208]

Dhar, D., 1999a, *The Abelian sandpile and related models*, Physica A **263**, 4–25. [223]

Dhar, D., 1999b, *Some results and a conjecture for Manna's stochastic sandpile model*, Physica A **270**, 69–81. [223]

Dhar, D., 1999c, *Studying Self-Organized Criticality with Exactly Solved Models*, preprint cond-mat/9909009, (unpublished). [43, 206, 208, 225, 318, 320]

Dhar, D., 2003, *Steady State and Relaxation Spectrum of the Oslo Rice-pile*, `cond-mat/0309490`, (unpublished). [230, 239]

Dhar, D., and R. Ramaswamy, 1989, *Exactly Solved Model of Self-Organized Critical Phenomena*, Phys. Rev Lett. **63**(16), 1659–1662. [280, 281, 283, 284, 334]

Díaz-Guilera, A., 1992, *Noise and dynamics of self-organized critical phenomena*, Phys. Rev. A **45**(2), 8551–8558. [247, 249]

Dickman, R., 2003, *private communication*. [59]

Dickman, R., M. A. Muñoz, A. Vespignani, and S. Zapperi, 2000, *Paths to Self-Organized Criticality*, Braz. J. Phys. **30**, 27, `cond-mat/9910454v2`. [57, 64, 109, 114, 221, 223]

Dickman, R., A. Vespignani, and S. Zapperi, 1998, *Self-organized Criticality as an absorbing state phase transition*, Phys. Rev. E **57**(5), 5095–5105. [61]

Dingle, R. B., 1973, *Asymptotic Expansions: Their Derivation and Interpretation* (Academic Press, London, UK and New York, NY). [358]

Domany, E., and W. Kinzel, 1984, *Equivalence of Cellular Automata to Ising Models and Directed Percolation*, Phys. Rev. Lett. **53**(4), 311–314. [334]

Dong, M., M. C. Marchetti, A. A. Middleton, and V. Vinokur, 1993, *Elastic String in a Random Potential*, Phys. Rev. Lett. **70**(5), 662–665. [249]

Dorn, P. L., D. S. Hughes, and K. Christensen, 2001, *On the avalanche size distribution in the BTW model*, preprint available from `http://www.cmth.ph.ic.ac.uk/kim/papers/preprints/preprint_btw.pdf`, (unpublished). [43, 106, 246]

Dowd, K., and C. Severance, 1998, *High Performance Computing* (O'Reilly, Sebastopol, CA), 2nd edition edition. [156, 159]

Drossel, B., 2002, *The complex scaling behavior of non-conserved self-organized critical systems*, `cond-mat/0205658`, (unpublished). [112]

Drossel, B., S. Clar, and F. Schwabl, 1993, *Exact Results for the One-Dimensional Self-Organized Critical Forest-Fire Model*, Phys. Rev. Lett. **71**(23), 3739–3742. [51]

Drossel, B., and F. Schwabl, 1992, *Self-Organized Critical Forest-Fire Model*, Phys. Rev. Lett. **69**(11), 1629–1632. [47, 105, 114, 115, 139, 140]

Edwards, S. F., and D. R. Wilkinson, 1982, *The surface statistics of a granular aggregate*, Proc. R. Soc. Lond. A **381**, 17–31. [341, 342]

Efetov, K. B., and A. I. Larkin, 1977, *Charge-density wave in a random potential*, Sov. Phys. JETP **45**, 1236, Zh. Eksp. Teor. Fiz. **72**, 2350–2361 (June 1971), translated by J. G. Adashko. [272]

Efron, B., 1982, *The Jackknife, the Bootstrap and Other Resampling Plans* (SIAM, Philadelphia, PA). [103, 165, 191]

Eklund, E. A., R. Bruinsma, J. Rudnick, and R. S. Williams, 1991, *Submicron-Scale Surface Roughening Induced by Ion Bombardment*, Phys. Rev. Lett. **67**(13), 1759–1762. [378]

Entacher, K., 1999, *On the CRAY-System Random Number Generator*, Simulation **72**(3), 163–169. [172]

Family, F., and T. Vicsek, 1985, *Scaling of the active zone in the Eden process on percolation networks and the ballistic deposition model*, J. Phys. A: Math. Gen. **18**(2), L75-L81. [266, 351]

Farkas, Z., and T. Fülöp, 2001, *One-dimensional drift-diffusion between two absorbing boundaries: application to granular segregation*, J. Phys. A: Math. Gen. **34**(15), 3191–3198. [299, 343, 355]

Feder, J., 1988, *Fractals* (Plenum, London, New York). [39]

Ferrenberg, A. M., D. P. Landau, and K. Binder, 1991, *Statistical and Systematic Errors in Monte Carlo Sampling*, J. Stat. Phys. **63**, 867–882. [102, 163]

Fisher, M. E., 1967, *The Theory of equilibrium critical phenomena*, Rep. Prog. Phys. **30**, 615–730. [73]

Fisher, M. E., 1984, *Walks, Walls, Wetting, and Melting*, J. Stat. Phys. **34**(5/6), 667–729. [127]

Fisher, M. E., 1986, *Interface Wandering in Adsorbed and Bulk Phases, Pure and Impure*, J. Chem. Soc., Faraday Trans. 2 **82**, 1596–1603 (Faraday Symposium 20), (Faraday Symposium 20). [266]

Fisher, M. E., S.-K. Ma, and B. G. Nickel, 1972, *Critical Exponents for Long-Range Interactions*, Phys. Rev. Lett. **29**, 917–920. [218]

Foltin, G., K. Oerding, Z. Rácz, R. L. Workman, and R. K. P. Zia, 1994, *Width distribution for random-walk interfaces*, Phys. Rev. E **50**, R639–R642. [355]

Frette, V., K. Christensen, A. Malthe-Sørenssen, J. Feder, T. Jøssang, and P. Meakin, 1996, *Avalanche dynamics in a pile of rice*, Nature **379**, 49–52. [205, 245, 280, 305, 306, 316]

Galluccio, S., and Y. Zhang, 1995, *Driven interfaces in quenched disorder at critical depinning*, Phys. Rev. E **51**(3), 1686–1689. [254, 350]

Garrido, P. L., J. L. Lebowitz, C. Maes, and H. Spohn, 1990, *Long-range correlations and conservative dynamics*, Phys. Rev. A **42**(4), 1954–1969. [108]

Goldenfeld, N., 1985, *Lectures on Phase Transition and the Renormalization Group* (Perseus Books, Reading, Massachusetts). [36]

Gorshenev, A. A., and Y. M. Pis'mak, 2003, *Punctuated Equilibrium in Software Evolution*, `cond-mat/0307201`, (unpublished). [19]

Grassberger, P., 1993, *On a self-organized forest-fire model*, J. Phys. A: Math. Gen. **26**, 2081–2089. [140, 142, 145, 146, 147, 176, 189]

Grassberger, P., 1994, *Efficient large-scale simulations of a uniformly driven system*, Phys. Rev. E **49**(3), 2436–2444. [55, 109]

Grassberger, P., 1995, *The Bak-Sneppen model for punctuated evolution*, Phys. Lett. A **200**, 277–282. [54, 56]

Grassberger, P., 2002, *Critical Behaviour of the Drossel-Schwabl Forest Fire Model*, New J. Phys. **4**, 17, `cond-mat/0202022`. [52, 183, 189, 190, 200, 315]

Grassberger, P., and H. Kantz, 1991, *On a forest-fire model with supposed self-organized criticaility*, J. Stat. Phys. **63**, 685–700. [47]

Grimmett, G. R., and D. R. Stirzaker, 1992, *Probability and Random Processes* (Oxford University Press, New York), second edition. [299]

Grinstein, G., D.-H. Lee, and S. Sachdev, 1990, *Conservation Laws, Anisotropy, and "Self-Organized Criticality" in Noisy Nonequilibrium Systems*, Phys. Rev. Lett. **64**(16), 1927–1930. [37, 38, 46, 105, 108, 113, 139, 279]

Gropp, W., E. Lusk, and A. Skjellum, 1999, *Using MPI* (The MIT Press, Cambridge, MA, London, England), second rdition edition. [166]

Gutenberg, B., and C. F. Richter, 1956, *Magnitude and energy of earthquakes*, Ann. Geofis. **9**(1), 1–15, annali di Geofisica. [39]

Halpin-Healy, T., and Y.-C. Zhang, 1995, *Kinetic roughening phenomena, stochastic growth, directed polymers and all that. Aspects of multidisciplinary statistical mechanics*, Phys. Rep. **254**, 215–414. [341, 350, 377]

Harris, T. E., 1963, *The Theory of Branching Processes* (Springer-Verlag, Berlin Göttingen Heidelberg). [119, 125, 126, 140]

Hasenbusch, M., 1999, *Habilitationsschrift: Monte Carlo Studies of the three dimensional Ising Model*, Preprint, HU-Berlin, Institut für Physik. [70, 71]

Hasenbusch, M., 2001, *Eliminating leading corrections to scaling in the 3-dimensional* $O(N)$*-symmetric* ϕ^4 *model:* $N = 3$ *and* 4, J. Phys.A: Math. Gen. **34**, 8221–8236, `cond-mat/0010463`. [219]

Henkel, M., 1999, *Conformal Invariance and Critical Phenomena* (Springer-Verlag, Berlin Heidelberg New York). [60]

Henley, C. L., 1993, *Statics of a "Self-Organized" Percolation Model*, Phys. Rev. Lett. **71**(17), 2741–2744. [142, 163, 189]

Hinrichsen, H., 2000, *Non-equilibrium critical phenomena and phase transitions into absorbing states*, Adv. Phys. **49**, 815–958, `cond-mat/0001070v2`. [57, 58, 226, 254, 318]

Honecker, A., 1997, *Program for simulating the two-dimensional forest-fire model*, URL `http://www-public.tu-bs.de:8080/~honecker/software/forest2d.html`. [150, 156, 161, 162]

Honecker, A., and I. Peschel, 1997, *Length scales and power laws in the two-dimensional forest-fire model*, Physica A **239**, 509–530. [142, 150, 163, 173, 176, 180, 189, 193, 197, 198]

Hoshen, J., and R. Kopelman, 1976, *Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm*, Phys. Rev. B **14**(8), 3438–3445. [151, 199]

Hughes, D., and M. Paczuski, 2002, *Large Scale Structures, Symmetry, and Universality in Sandpiles*, Phys. Rev. Lett. **88**, 054302. [207, 209, 281]

Hwa, T., and M. Kardar, 1989, *Dissipative Transport in Open Systems: An Investgation of Self-Organized Criticality*, Phys. Rev. Lett. **62**(16), 1813–1816. [38, 105, 106, 139, 279]

Ising, E., 1925, *Beitrag zur Theorie des Ferromagnetismus*, Z. Phys. **31**, 253–258. [19]

Jensen, H. J., 1995, *The fate of the elastic string: roughening near the depinning threshold*, J. Phys. A: Math. Gen. **28**, 1861–1877. [266]

Jensen, H. J., 1998, *Self-Organized Criticality* (Cambridge University Press, New York, NY). [37, 41, 42, 45, 48, 113, 139, 143, 246, 249, 315]

Jensen, I., 1999, *Low-density series expansions for directed percolation: I. A new efficient algorithm with applications to the square lattice*, J. Phys. A: Math. Gen. **32**, 5233–5249. [60]

Johnston, A. C., and S. Narva, 1985, *Recurrence Rates and Probability Estimates for the New Madrid Seismic Zone*, J. Geophys. Res. **90**, 6737. [39]

Kadanoff, L. P., S. R. Nagel, L. Wu, and S. min Zhou, 1989, *Scaling and universality in avalanches*, Phys. Rev. A **39**(12), 6524–6537. [280]

van Kampen, N. G., 1992, *Stochastic Processes in Physics and Chemistry* (Elsevier Science B. V., Amsterdam, The Netherlands), third impression 2001, enlarged and revised edition. [58, 87, 91, 226, 228, 263, 318, 319, 321]

Kardar, M., 1998, *Nonequilibrium dynamics of interfaces and lines*, Phys. Rep. **301**, 85–112. [254]

Kardar, M., 2001, in *Annual Reviews of Computational Physics*, edited by D. Stauffer (World Scientific, Singapore), Volume VIII, pp. 1–47. [349, 364, 378]

Kardar, M., G. Parisi, and Y.-C. Zhang, 1986, *Dynamic Scaling of Growing Interfaces*, Phys. Rev. Lett. **56**(9), 889–892. [377]

Kaulke, M., 1999, *Anwendung der Dichtematrix-Renormierung auf nichthermitesche Probleme*, Ph.D. thesis, Fachbereich Physik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany/EU. [239]

Kirchner, J. W., and A. Weil, 1998, *No fractals in fossil extinction statistics*, Nature **395**(6700), 337–338. [40]

Kloster, M., S. Maslov, and C. Tang, 2001, *Exact solution of a stochastic directed sandpile model*, Phys. Rev. E **63**, 026111-1–026111-4. [280, 281, 283, 334]

Knuth, D. E., 1997, *Fundamental Algorithms*, Volume 1 of *The Art of Computer Programming* (Addison Wesley, Reading, Massachusetts), third edition. [66, 329]

Koba, Z., H. B. Nielsen, and P. Olesen, 1972, *Scaling of multiplicity distributions in high energy hadron collisions*, Nucl. Phys. B **40**, 317–334. [81]

Koplik, J., and H. Levine, 1985, *Interface moving through a random background*, Phys. Rev. B **32**, 280–292. [246, 270]

Krug, J., 1997, *Origins of scale invariance in growth processes*, Adv. Phys. **46**(2), 139–282. [38, 271, 341, 342, 350, 353, 370, 372, 377]

Krug, J., and H. Spohn, 1991, in *Solids far from Equilibrium*, edited by C. Godrèche (Cambridge University Press, Cambridge, UK), pp. 479–582. [266, 341, 344, 350, 377]

Ktitarev, D. V., S. Lübeck, P. Grassberger, and V. B. Priezzhev, 2000, *Scaling of waves in the Bak-Tang-Wiesenfeld sandpile model*, Phys. Rev. E **61**(1), 81–92. [43]

Kunz, H., and F. Y. Wu, 1978, *Site percolation as a Potts model*, J. Phys. C: Solid State Phys. **11**, L1–L4. [82]

Landau, D. P., and K. Binder, 1988, *Surface and size effects in magnetic phase transitions (invited)*, J. Appl. Phys. **63**(8), 3077–3081. [378]

Landau, D. P., and K. Binder, 2000, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, UK). [70, 74, 88, 89, 162]

Langlands, R., C. Pichet, P. Pouliot, and Y. Saint-Aubin, 1992, *On the Universality of Crossing Probabilities in Two-Dimensional Percolation*, J. Stat. Phys. **67**(3/4), 553–574. [60]

Lässig, M., 1998, *On growth, disorder, and field theory*, J. Phys.: Condens. Matter **10**, 9905–9950. [38, 350]

Le Bellac, M., 1991, *Quantum and Statistical Field Theory [Phenomenes critiques aux champs de jauge, English]* (Oxford University Press, New York, NY). [263]

Le Doussal, P., K. J. Wiese, and P. Chauve, 2002, *2-Loop Functional Renormalization Group Theory of the Depinning Transition*, Phys. Rev. B **66**, 174201, `cond-mat/0205108`. [267, 270, 271]

Le Doussal, P., K. J. Wiese, and P. Chauve, 2003, *Functional Renormalization Group and the Field Theory of Disordered Elastic Systems*, `cond-mat/0304614`, (unpublished). [270]

Leschhorn, H., 1993, *Interface depinning in a disordered medium — numerical results*, Physica A **195**, 324–335, `cond-mat/9302039`. [267, 378, 379]

Leschhorn, H., T. Nattermann, S. Stepanow, and L.-H. Tang, 1997, *Driven interface depinning in a disordered medium*, Ann. Physik **6**, 1–34, `cond-mat/9603114`. [249, 252, 259, 260, 261, 264, 265, 266, 271, 378]

Lise, S., and H. J. Jensen, 1996, *Transitions in Nonconserving Models of Self-Organized Criticality*, Phys. Rev. Lett. **76**(13), 2326–2329. [47]

Lise, S., and M. Paczuski, 2001a, *Scaling in a nonconservative earthquake model of self-organized criticality*, Phys. Rev. E **64**, 046111. [111, 139, 246]

Lise, S., and M. Paczuski, 2001b, *Self-organized criticality and universality in a nonconservative earthquake model*, Phys. Rev. E **63**, 036111. [111, 114, 139, 199]

Lübeck, S., 2003a, *Mean-field theory of self-organized critical sandpile models*, private communication, (unpublished). [65]

Lübeck, S., 2003b, *Universal behavior of crossover scaling functions for continuous phase transitions*, Phys. Rev. Lett. **90**, 210601-1–4. [304]

Madras, N., and A. D. Sokal, 1988, *The Pivot Algorithm: A Highly Efficient Monte Carlo Method for the Self-Avoiding Walk*, J. Stat. Phys. **50**, 109–186. [101]

Magnus, W., F. Oberhettinger, and R. P. Soni, 1966, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer-Verlag, Berlin Heidelberg New York). [343, 355]

Mahan, G. D., 1990, *Many-Particle Physics* (Plenum Press, New York, NY), second edition edition. [244]

Malamud, B. D., G. Morein, and D. L. Turcotte, 1998, *Forest Fires: An Example of Self-Organized Criticial Behavior*, Science **281**, 1840–1842. [201]

Malcai, O., D. A. Lidar, O. Biham, and D. Avnir, 1997, *Scaling range and cutoffs in empirical fractals*, Phys. Rev. E **56**, 2817–2828. [38]

Malthe-Sørenssen, A., 1999, *Tilted sandpiles, interface depinning and earthquake models*, Phys. Rev. E **59**(4), 4169–4174. [221, 222, 249, 276]

Mandelbrot, B. B., 1983, *The fractal geometry of nature* (Freeman, New York). [39]

Manna, S. S., 1991, *Two-state model of self-organized criticality*, J. Phys. A: Math. Gen. **24**, L363–L369. [109, 223]

Manna, S. S., L. B. Kiss, and J. Kertész, 1990, *Cascades and Self-Organized Criticality*, J. Stat. Phys. **61**(3/4), 923–932. [61]

Maslov, S., and Y.-C. Zhang, 1995, *Exactly Solved Model of Self-Organized Criticality*, Phys. Rev. Lett. **75**(8), 1550. [281]

Meakin, P., 1998, *Fractals, scaling and growth far from equilibrium* (Cambridge University Press, Cambridge, UK). [251, 344]

Meester, R., F. Redig, and D. Znamenski, 2001, *The Abelian Sandpile: a Mathematical Introduction*, Markov Proc. Rel. Fields **7**(4), 509–523, preprint from `http://www.cs.vu.nl/~rmeester/preprints/sandpile.ps`. [320]

Meester, R., and D. Znamenski, 2001, *Non-triviality of the discrete Bak-Sneppen evolution model*, preprint from `http://www.cs.vu.nl/~rmeester/preprints/bs.ps`, (unpublished). [56]

Meng Ta-chung, R. Rittel, and Zhang Yang, 1999, *Inelastic Diffraction and Color-Singlet Gluon Clusters in High-Energy Hadron-Hdron and Lepton-Hadron Collisions*, Phys. Rev. Lett. **82**(10), 2044–2047. [39]

Mertens, S., 1990, *Lattice Animals: A Fast Enumeration Algorithm and New Perimeter Polynomials*, J. Stat. Phys. **58**(5/6), 1095–1108. [182, 196]

Middleton, A. A., 1992, *Asymptotic Uniqueness od the Sliding State for Charge-Density Waves*, Phys. Rev. Lett. **68**(5), 670–673. [251]

Middleton, A. A., and C. Tang, 1995, *Self-Organized Criticality in Nonconserved Systems*, Phys. Rev. Lett. **74**(5), 742–745. [46, 112, 113, 114]

Mohanty, P. K., and D. Dhar, 2002, *Generic Sandpile Models Have Directed Percolation Exponent*, Phys. Rev. Lett. **89**(10), 104303–1 - 104303–4. [334]

Moloney, N. R., and G. Pruessner, 2003, *Asynchronously parallelized percolation on distributed machines*, Phys. Rev. E **67**, 037701-1–4, `cond-mat/0211240`. [22, 32, 91, 150]

Moßner, W. K., B. Drossel, and F. Schwabl, 1992, *Computer simulations of the forest-fire model*, Physica A **190**, 205–217. [47]

Müller-Krumbhaar, H., and K. Binder, 1973, *Dynamic Properties of the Monte Carlo Method in Statistical Mechanics*, J. Stat. Phys **8**(1), 1–24. [101, 162]

Narayan, O., 2000, *Anomalous scaling in depinning transitions*, Phys. Rev. E **62**, R7563–R7566. [252, 259]

Narayan, O., and D. S. Fisher, 1993, *Threshold critical dynamics of driven interfaces in random media*, Phys. Rev. B **48**(10), 7030–7042. [269, 270, 271]

Nattermann, T., S. Stepanow, L.-H. Tang, and H. Leschhorn, 1992, *Dynamics of interface depinning in a disordered medium*, J. Phys. II France **2**, 1483–1488. [245, 247, 249, 251, 252, 260, 270, 271, 280]

Nattermann, T., and L.-H. Tang, 1992, *Kinetic surface roughening. I. The Kardar-Parisi-Zhang equation in the weak-coupling regime*, Phys. Rev. A **45**(10), 7156–7161. [259, 342]

Newman, M. E. J., and R. M. Ziff, 2000, *Efficient Monte Carlo Algorithm and High-Precision Results for Percolation*, Phys. Rev. Lett. **85**(19), 4104–4107. [150, 152, 183, 195, 196]

Newman, M. E. J., and R. M. Ziff, 2001, *Fast Monte Carlo Algorithm for site and bond percolation*, Phys. Rev. E **64**, 016706-1–16, `cond-mat/0101295`. [150, 161, 165]

Nicolaides, D., and A. D. Bruce, 1988, *Universal configurational structure in two-dimensional scalar models*, J. Phys. A: Math. Gen. **21**(1), 233–243. [81]

Nordhagen, H., 2003, *Stick-Slip Friction*, Master's thesis, Department of Physics, University of Oslo, thesis submitted for the degree of Candidatus Scientiarum. [99]

Novotny, M. A., 1996, *Cross-over and effective scaling dimension*, Int. J. Mod. Phys. C **7**, 361–370. [36]

Olami, Z., H. J. S. Feder, and K. Christensen, 1992, *Self-Organized Criticaility in a Continuous, Nonconservative Cellular Automaton Modeling Earthquakes*, Phys. Rev. Lett. **68**(8), 1244–1247. [13, 39, 44, 46, 61, 105, 139, 199]

Onsager, L., 1944, *Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition*, Phys. Rev. **65**, 117–149. [19]

Paczuski, M., and K. E. Bassler, 2000, *Theoretical Results for Sandpile Models of SOC with Multiple Topplings*, preprint cond-mat/0005340v2, (unpublished). [208, 280, 281, 282, 334, 366]

Paczuski, M., and S. Boettcher, 1996, *Universality in Sandpiles, Interface Depinning, and Earthquake Models*, Phys. Rev. Lett. **77**(1), 111–114. [210, 246, 247, 253, 267, 276, 283, 286]

Paczuski, M., S. Maslov, and P. Bak, 1996, *Avalanche dynamics in evolution, growth, and depinning models*, Phys. Rev. E **53**(1), 414–443, `adap-org/9510002`. [53, 56, 221, 223, 253, 276]

Parmeggiani, A., T. Franosch, and E. Frey, 2003, *Phase Coexistence in Driven One Dimensional Transport*, Phys. Rev. Lett. **90**(8), 086601-1–4, `cond-mat/0301475`. [339]

Pastor-Satorras, R., and A. Vespignani, 2000a, *Corrections to scaling in the forest-fire model*, Phys. Rev. E **61**, 4854–4859. [73, 89, 147, 186, 189]

Pastor-Satorras, R., and A. Vespignani, 2000b, *Universality classes in directed sandpile models*, J. Phys. A: Math. Gen. **33**, L33–L39. [73, 101, 280, 282, 284, 334]

Peschel, I., X. Wang, M. Kaulke, and K. Hallberg (eds.), 1999, *Density-Matrix Renormalization*, Volume 528 of *Lecture notes in physics* (Springer-Verlag, Berlin Heidelberg New York). [239]

Peters, O., C. Hertlein, and K. Christensen, 2002, *A Complexity View of Rainfall*, Phys. Rev. Lett. **88**(1), 018701-1–018701-4, `cond-mat/0201468`. [39]

Pfeuty, P., and G. Toulouse, 1977, *Introduction to the Renormalization Group and to Critical Phenomena* (John Wiley & Sons, Chichester). [69, 76, 187]

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992, *Numerical Recipes in C* (Cambridge University Press, New York, NY), 2nd edition edition. [101, 171, 225, 267]

Priezzhev, V. B., 1994, *Structure of Two-Dimensional Sandpile. I. Height Probilities*, J. Stat. Phys. **74**(5/6), 955–979. [43]

Priezzhev, V. B., E. V. Ivashkevich, A. M. Povolotsky, and C.-K. Hu, 2001, *Exact Phase Diagram for an Asymmetric Avalanche Process*, Phys. Rev. Lett. **87**(8), 084301-1–084301-4. [280, 281]

Priezzhev, V. B., D. V. Ktitarev, and E. V. Ivashkevich, 1996, *Formation of Avalanches and Critical Exponents in an Abelian Sandpile Model*, Phys. Rev. Lett. **76**(12), 2093–2096. [43]

Privman, V., and M. E. Fisher, 1984, *Universal critical amplitudes in finite-size scaling*, Phys. Rev. B **30**(1), 322–327. [77, 91, 317]

Privman, V., P. C. Hohenberg, and A. Aharony, 1991, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and J. L. Lebowitz (Academic Press, New York), Volume 14, chapter 1, pp. 1–134. [77, 88]

Pruessner, G., 2003a, *Drift causes anomalous exponents in growth processes*, `cond-mat/0404007v2`, accepted for publication in Phys. Rev. Lett., (unpublished). [23, 31]

Pruessner, G., 2003b, *Exact solution of the totally asymmetric Oslo model*, `cond-mat/0402564`, accepted for publication in JPA, (unpublished). [23, 31, 281, 282]

Pruessner, G., 2003c, *Oslo rice pile model is a quenched Edwards-Wilkinson equation*, Phys. Rev. E **67**, 030301(R)-1–030301(R)-4, `cond-mat/0209531`. [22, 32, 280]

Pruessner, G., and H. J. Jensen, 2002a, *Broken scaling in the forest-fire model*, Phys. Rev. E **65**, 056707–1 - 056707–8, `cond-mat/0201306`. [22, 32, 52, 93, 140, 172, 188, 246, 315]

Pruessner, G., and H. J. Jensen, 2002b, *A solvable non-conservative model of Self-Organised Criticality*, Europhys. Lett. **58**(2), 250–256, `cond-mat/0104567`. [21, 32, 105, 139]

Pruessner, G., and H. J. Jensen, 2003a, *Anisotropy and universality: The Oslo model, the rice pile experiment and the quenched Edwards-Wilkinson equation*, Phys. Rev. Lett. **91**, 244303-1–4, `cond-mat/0307443`. [22, 31, 294, 315]

Pruessner, G., and H. J. Jensen, 2003b, *A new, efficient algorithm for the Forest Fire Model*, `cond-mat/0309173`, accepted for publication in Phys. Rev. E, (unpublished). [22, 31]

Pruessner, G., D. Loison, and K.-D. Schotte, 2001, *Monte Carlo Simulation of an Ising Model on a Sierpiński Carpet*, Phys. Rev.B **64**, 134414–1 - 134414–10. [165, 191]

Pruessner, G., and N. R. Moloney, 2003, *Numerical results for crossing, spanning and wrapping in two-dimensional percolation*, J. Phys. A: Math. Gen. **36**(44), 11213–11228, `cond-mat/0309126`. [22, 31, 150]

Pruessner, G., and N. R. Moloney, 2004, *Winding Clusters in Percolation on the Torus and the Möbius strip*, J. Stat. Phys. **115**(3/4), 839–853, `cond-mat/0310361`. [22, 31, 150]

Ramasco, J. J., J. M. López, and M. A. Rodríguez, 2000, *Generic Dynamic Scaling in Kinetic Roughening*, Phys. Rev. Lett. **84**(10), 2199–2202. [350]

Redner, S., 2001, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK). [241]

Rittel, R., 2000, *Selbstorganisierte Kritikalität auf der Ebene von Quarks und Gluonen und deren Manifestation in inelastisch diffraktiven hochenergetischen Streuprozessen*, Ph.D. thesis, Fachbereich Physik, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany/EU. [39, 53, 56]

Roberts, D. C., and D. L. Turcotte, 1998, *Fractality and Self-Organized Criticality of Wars*, Fractals **6**(4), 351–357. [19]

Salas, J., and A. D. Sokal, 2000, *Universal Amplitude Ratios in the Critical Two-Dimensional Ising Model on a Torus*, J. Stat. Phys. **98**, 551–588, `cond-mat/9904038v2`. [91]

Schenk, K., B. Drossel, S. Clar, and F. Schwabl, 2000, *Finite-size effects in the self-organized critical forest-fire model*, Eur. Phys. J. B **15**, 177–185. [142, 147, 150, 180, 182]

Schenk, K., B. Drossel, and F. Schwabl, 2002, *Self-organized critical foerst-fire model on large scales*, Phys. Rev. E **65**, 026135-1–8, `cond-mat/0105121`. [52, 173, 180, 189, 195, 197, 198]

Sepkoski Jr., J. J., 1993, *Ten Years in the Library: New Data Confirm Paleontological Patterns*, Paleobiology **19**, 43. [19, 40]

Sethna, J. P., K. Dahmen, S. Kartha, J. A. Krumhansl, B. W. Roberts, and J. D. Shore, 1993, *Hysteresis and Hierarchies: Dynamics of Disorder-Driven First-Order Phase Transformations*, Phys. Rev. Lett. **70**(21), 3347–3350. [251]

Sinha-Ray, P., and H. J. Jensen, 2000, *Forest-fire models as a bridge between different paradigms in self-organized criticality*, Phys. Rev. E **62**, 3215–3218. [199]

Sire, C., 2002, *A precise approximation for directed percolation in $d = 1 + 1$*, `cond-mat/0205490`, (unpublished). [60]

Smethurst, D. P., and H. C. Williams, 2001, *Are hospital waiting lists self-regulating?*, Nature **410**, 652–653. [39, 384]

Sornette, D., 2001, *Mechanism for Powerlaws without Self-Orgnization*, `cond-mat/0110426`, (unpublished). [39]

Stanley, H. E., 1971, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York). [35, 140, 148, 315]

Stauffer, D., and A. Aharony, 1994, *Introduction to Percolation Theory* (Taylor & Francis, London). [35, 66, 78, 83, 140, 146, 182, 196, 197]

Sykes, M. F., and M. Glen, 1976, *Percolation processes in two dimensions I. Low-density series expansions*, J. Phys. A: Gen. Phys. **9**(1), 87–95. [182, 196]

Szegö, G., 1975, *Orthogonal Polynomials* (Americal Mathematical Society, Providence, RI), fourth edition. [136]

Tadić, B., and D. Dhar, 1997, *Emergent Spatial Structures in Critical Sandpiles*, Phys. Rev. Lett. **79**(8), 1519–1522. [280, 284]

Tang, C., and P. Bak, 1988, *Critical Exponents and Scaling Relations for Self-Organized Critical Phenomena*, Phys. Rev. Lett. **60**(23), 2347–2350. [57]

Tang, L.-H., M. Kardar, and D. Dhar, 1995, *Driven Depinning in Anisotropic Media*, Phys. Rev. Lett. **74**(6), 920–923. [287, 379]

Tebaldi, C., M. De Menech, and A. L. Stella, 1999, *Multifractal Scaling in the Bak-Tang-Wiesenfeld Sandpile and Edge Events*, Phys. Rev. Lett. **83**(19), 3952–3955, `cond-mat/9903270`. [100, 186, 198, 283]

The British Library Science Technology and Business (STB), 2001, *Welfare Reform on the Web: NATIONAL HEALTH SERVICE — REFORM — GENERAL*, `http://www.bl.uk/services/stb/spiswlfr/issue22/nhs-rfrm.html`. [39]

Tsuchiya, T., and M. Katori, 1999a, *Effect of anisotropy on the self-organized critical states of Abelian sandpile models*, Physica A **266**, 358–361. [280]

Tsuchiya, T., and M. Katori, 1999b, *Exact results for the directed Abelian sandpile models*, J. Phys. A: Math. Gen. **32**, 1629–1641. [280, 303]

Turcotte, D. L., 1999, *Self-organized criticality*, Rep. Prog. Phys. **62**, 1377–1426. [19]

Vazquez, A., 2000, *Numerical simulations of directed sandpile models*, `cond-mat/0003420`, (unpublished). [280, 281]

Vespignani, A., R. Dickman, M. A. Muñoz, and S. Zapperi, 1998, *Driving, Conservation and Absorbing States in Sandpiles*, Phys. Rev. Lett. **81**(25), 5676–5679. [57]

Vespignani, A., R. Dickman, M. A. Muñoz, and S. Zapperi, 2000, *Absorbing-state phase tranistions in fixed-energy sandpiles*, Phys. Rev. E **62**(4), 4564–4582, `cond-mat/0003285`. [223, 249, 276, 277]

Vespignani, A., and S. Zapperi, 1997, *Order Parameter and Scaling Fields in Self-Organized Criticality*, Phys. Rev. Lett. **78**(25), 4793–4796. [61]

Vespignani, A., and S. Zapperi, 1998, *How self-organized criticality works: A unified mean-field picture*, Phys. Rev. E **57**(6), 6345–6362. [57, 61, 65, 113, 114, 139, 145, 146]

Wannier, G. H., 1987, *Statistical Physics* (Dover (Reprint, Originally published, New York, Wiley, 1966), Mineola, NY). [244]

Wegner, F. J., 1972, *Corrections to Scaling Laws*, Phys. Rev. B **5**(11), 4529–4536. [88, 147, 187, 283]

Wiese, K. J., 2002, *Disordered Systems and the functional renormalization group, a pedagogical introduction*, Acta Physica Slovaca **52**(4), 341–351, `cond-mat/0205116`. [266, 272]

You, H., R. P. Chiarello, H. K. Kim, and K. G. Vandervoot, 1993, *X-Ray Reflectivity and Scanning-Tunneling-Microscope Study of Kinetic Roughening of Sputter-Deposited Gold Films during Growth*, Phys. Rev. Lett. **70**(19), 2900–2903. [378]

Zhang, S., 1997, *On the universality of a one-dimensional model of a rice pile*, Phys. Lett. A **233**, 317–322. [217, 221, 246, 249]

Zhang, Y.-C., 1989, *Scaling Theory of Self-Organized Criticality*, Phys. Rev. Lett. **63**(5), 470–473. [109, 114, 276]

Colophon

This thesis was typeset using the $\LaTeX 2_\epsilon$ typesetting system created by Leslie Lamport, based on Donald Knuth's $\TeX$, using $\mathcal{AMS}$-$\TeX$, PSTricks and many other packages. The body text is set in Palatino designed by Hermann Zapf, 1948.