TOPICS IN MATHEMATICAL PHYSICS ON SIERPINSKI CARPETS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Joe Po-Chou Chen

August 2013

TOPICS IN MATHEMATICAL PHYSICS ON SIERPINSKI CARPETS

Joe Po-Chou Chen, Ph.D.

Cornell University 2013

We study three topics in mathematical physics on fractal domains which are based on the Sierpinski carpet and its higher-dimensional analogs. First, we rigorously investigate the thermodynamics of the ideal massive and massless Bose gas, from which quantitative results about Bose-Einstein condensation, blackbody radiation, and the (zero- and finite-temperature) Casimir effect are obtained. Second, we prove the subsequential Mosco convergence of discrete Dirichlet forms on Sierpinski carpet graphs, and from there deduce the convergence of the discrete Green forms. Last but not least, we enumerate a collection of periodic billiard orbits in a planar self-similar Sierpinski carpet billiard table, which paves the way for future studies of billiard dynamics on fractal billiards.

BIOGRAPHICAL SKETCH

Joe Po-Chou Chen was born in Taipei, Taiwan on January 11, 1984. After completing the secondary education in Taiwan, he matriculated at Yale University, graduating *Phi Beta Kappa* with a major in Physics and minor in Mathematics in 2006. He then commenced graduate studies at Cornell, eventually choosing to specialize in mathematical physics, with a special interest in statistical mechanics on fractal spaces. In Summer 2012, he was invited for a 3-month-long research stay at the Hausdorff Research Institute for Mathematics in Bonn, Germany, as well as attendance at XVIIth International Congress on Mathematical Physics in Aalborg, Denmark. Effective late August 2013, he will become a Visiting Assistant Professor of Mathematics at the University of Connecticut.

Dedicated to my parents and family.

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Bob Strichartz for being a resourceful and gracious mentor. He taught me some of the best things from analysis and PDE, and never failed to offer me new ideas to explore. At the outset of my research four years ago, he suggested that I look into the more physically realistic three-dimensional fractals, as opposed to the low-dimensional ones. That decision proved pivotal in my investigation of statistical mechanics problems on fractals, and set the tenor for the current thesis. As I developed my own taste for problems, he gave me the room to formulate them and solve them, while doling out plenty of advice along the way. For that I'll be forever grateful. He is a man full of humor, and is truly devoted to the well-being of his students.

I also want to thank Veit Elser and Carl Franck for agreeing to serve on my committee, and suggesting several useful ideas and extensions of my thesis work.

I am extremely grateful to two "big siblings" who have guided me throughout my research work. Ben Steinhurst taught me plenty of the stochastic analysis on Sierpinski carpets, and gladly served as a sounding board for my various ideas. Naotaka Kajino, always precise and meticulous on everything (not just math related), provided the solid foundation upon which I could build my thesis work, and constantly offered incisive criticisms. I thank you both for setting great examples of an aspiring mathematician.

Thanks are also accorded to my co-authors, Robert Niemeyer (on billiards) and Baris Ugurcan (on free fields), for sharing your time and expertise on the joint projects. It's been a thrill working with and learning from you, and I hope the feeling is mutual.

The fractal compatriots at the University of Connecticut have been the most generous bunch. I'm especially grateful to Sasha Teplyaev for many useful suggestions ever since I started my research, and Luke Rogers, Gerald Dunne and Dan Kelleher for helpful discussions from time to time. I look forward to joining your group and collaborating with you come this September.

In the meantime, I want to thank members of the probability group at Cornell, past and present, for many inspiring conversations on various subjects, in particular Laurent Saloff-Coste, Nate Eldredge, Lionel Levine, Jon Peterson, Tianyi Zheng, and Janna Lierl. If you asked me four years ago what "transience" of random walk means, I probably couldn't give a straight answer. But thanks to you, I now know a little bit of everything from Dirichlet forms to percolation to sandpile models, which enables me to pursue new ideas and make unexpected connections.

Over my graduate career, I have learned a great deal of physics and mathematics from the best teachers at Cornell. Above all I would like to thank Len Gross, who taught an intensive yet through course on functional analysis. It was from his course that I properly learned about Dirichlet forms, as well as the essentials of mathematical physics on the level of Reed and Simon.

This thesis would not have existed without the prior contribution from the pioneers of fractal analysis, and I've been lucky to have known several of them on a personal level, including, but not limited to: Martin Barlow, Ben Hambly, Jun Kigami, Takashi Kumagai, and Michel Lapidus. Thank you for taking your time to share your wisdom and encourage me to do better.

Many younger fractologists deserve thanks as well. In no particular order, I thank Erin Pearse, John Rock, Machiel van Frankenhuijsen, Hung (Tim) Lu, Eva Curry, and Sabrina Kombrink, for being gracious hosts and companions at the various fractal conferences.

I'd also like to thank those who have facilitated or hosted my seminar/conference talks over the past 3 years: Eric Akkermans, Marek Biskup, Christian Benes, Daniel Lenz, Martina Zaehle, David Croydon, and Ka-Sing Lau.

In Summer 2012, I had the privilege of working at the Hausdorff Research Institute for Mathematics (HIM) in Bonn, Germany, under the Hausdorff Trimester Program

"Mathematical challenges of materials science and condensed matter physics." This was an all-expense-paid research stay, covering the round-trip air fare, a studio apartment by the Rhine, and free bike rental for 3 months! As a result I got to experience first-hand the exhilarating math-phys activity in Bonn, and Europe at large.

Several important project ideas germinated during my stay in Bonn, one being the entropic repulsion of Gaussian free field on fractal graphs. For this I'm grateful to Codina Cotar and Noemi Kurt for sharing their expertise and offering oodles of advice. Meanwhile I made preliminary progress on rigorously deriving Bose-Einstein condensation for dilute interacting Bose gas in random geometries. I wish to thank Benjamin Schlein for his many insights into the Hartree and Gross-Pitaevskii limits, and Daniel Ueltschi for suggesting potential directions.

To all the participants of the Trimester program, I thank you for your camaraderie. I especially want to give a shout-out to my fellow (now or soon-to-be) doctors in the basement office of HIM: Amartya Banerjee, Francesca Colasuonno, Elena Pulvirenti, and Mark Wilkinson. You've been the most awesome office mates!

To all my friends at Cornell and around the world, thank you for sharing your moments with me and enriching my life. And to my family, especially my parents, I owe a deep debt of gratitude for your unwavering support in good times and bad. Without your sacrifice, I would not be where I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This thesis consists of three topics in mathematical physics on *fractal* domains. The first topic focuses on the *Bose gas*, which is a familiar object in the study of quantum statistical mechanics. The second topic, which is slightly more technical, clarifies the mode of convergence of discrete resolvents on approximating fractal graphs to the continuum resolvent on the limiting fractal set. The third topic concerns the study of periodic *billiard trajectories* in planar fractal billiard tables. The novelty here is that the underlying geometry is a *fractal*, which, in our setting, possesses self-similarity but not translational invariance. As such the "conventional" notions of mathematical physics on Euclidean spaces need not immediately carry over to the fractal setting.

One of the distinguishing features of fractal spaces from smooth manifolds is that diffusion on the former is *anomalous* (more precisely, *subdiffusive*). This is because, intuitively speaking, the distribution of obstacles across all length scales impede the motion of the diffusing particle inside the fractal. This has important consequences for the underlying calculus, as the Laplacian—the infinitesimal generator of diffusion—does not behave as a second-order differential operator. Understanding the properties of the Laplacian, such as heat kernels and spectral asymptotics, is the mission behind the analysis on fractals project.

Analysis on fractals has its roots in the study of random walk in porous media during the early 80s (the "ant in the labyrinth" problem of de Gennes). As a way to obtain a first approximation, mathematical physicists proposed to study the same problem on deterministic self-similar fractals, and worked on identifying the scaling limit. Probabilists followed up with a rigorous construction of Brownian motion on the Sierpinski gasket (SG) [9], and later, the Sierpinski carpet (SC) [3,51]. In the 90s, Kigami made an impor-

tant breakthrough in building a purely analytic theory of calculus on post-critically finite (PCF) fractals (the Sierpinski gasket being the prime example), the details of which are described in the monograph [45].

Unlike Euclidean space or a smooth manifold, a fractal space has the following distinguishing properties:

♠ Typically it has a non-integer ***Hausdorff dimension*** d_h with respect to the Euclidean metric.

♠ It supports a diffusion (Brownian motion) which has anomalous *subdiffusive* space-time scaling: (time) $\propto$ (distance)^{d_w} for some ***walk dimension*** $d_w > 2$, whereas Brownian motion on manifolds satisfies (time) $\propto$ (distance)^2.

♠ As a result, the Laplacian on fractal spaces (defined as the infinitesimal generator of the Brownian motion) satisfies a Weyl asymptotic formula with effective dimension $d_s = 2(d_h/d_w)$, called the ***spectral dimension***. Equivalently, the heat kernel $p_t(x, y)$ scales with $t^{-d_s/2}$ asymptotically in the short-time limit $t \downarrow 0$.

As a mathematical physicist, I'm interested in how this irregular geometry affects the behavior of not just one particle, but many particles. The study of their collective behavior belongs to the realm of *statistical mechanics*. Roughly speaking, there are two classes of such problems: those associated with *discrete* spin space, such as percolation and ferromagnetic Ising model; and those with *continuous* spin space. In both classes, depending on the dimensionality of the space being considered, the model may exhibit different thermodynamic behaviors. (For example, does a nontrivial phase transition exist?) The main novelty on fractal spaces is that the relevant dimensions for the two classes are different—*isoperimetric dimension* for the former, and *spectral dimension* for the latter—and *neither* is equal to the Hausdorff dimension in general.

The bulk of this thesis focuses on models with *continuous* spin space, which has the *spectral dimension* as the relevant dimension. For concreteness, we will use the two-dimensional Sierpinski carpet and its higher-dimensional analogs as our model geometry. It is hoped that the techniques used to prove the results stated in this thesis can be extended to treat random geometries, such as percolation clusters and random trees.

1.1 Synopsis of topics covered

This thesis is organized as follows.

Chapter 2 is devoted to a mathematically rigorous investigation into the equilibrium thermodynamics of massless and massive *Bose gas* confined in generalized Sierpinski carpets (abbreviated GSCs). Due to the anomalous walk dimension $d_w > 2$ associated with Brownian motion on GSCs, all extensive thermodynamic quantities are shown to scale with the spectral volume with dimension $d_s = 2(d_h/d_w)$ rather than the Hausdorff volume. We prove that for a low-temperature, high-density ideal massive Bose gas in an unbounded GSC, Bose-Einstein condensation occurs if and only if $d_s > 2$, or equivalently, if the Brownian motion on the GSC is transient. We also derive explicit expressions for the energy of blackbody radiation in a GSC, as well as the Casimir pressure on the parallel plate of a fractal waveguide modelled after a GSC. Our proofs involve extensive use of the spectral zeta function, obtained via a sharp estimate of the heat kernel trace. We believe that our results can be verified through photonic and cold atomic experiments on fractal structures.

Chapter 3 gives a careful treatment of the convergence of discrete (graphical) Green forms to the continuum Green form on generalized Sierpinski carpets. To establish this result, we shall prove that there exists a subsequence such that the discrete Dirichlet

forms converge in the sense of Mosco. Furthermore, it will be shown that every Mosco limit point can be extended to a regular Dirichlet form, which is comparable to some local regular Dirichlet form that generates diffusion on the carpet. The results help clarify the precise manner in which the eigenvalues of the discrete Laplacians converge to that of the continuum Laplacian, and also establish a firm ground for studying Gaussian free fields on Sierpinski carpet graphs.

Chapter 4 focuses on the identification of billiard trajectories, *viz.* a collection of *periodic billiard orbits*, in a planar, self-similar Sierpinski carpet billiard table $\Omega(S_{\mathbf{a}})$. Based on our refinement of the result of Durand-Cartagena and Tyson regarding non-trivial line segments in $S_{\mathbf{a}}$, we construct what is called an eventually constant sequence of compatible periodic orbits of prefractal Sierpinski carpet billiard tables $\Omega(S_{\mathbf{a},n})$. The trivial limit of this sequence then constitutes a periodic orbit of $\Omega(S_{\mathbf{a}})$. We also determine the corresponding translation surface $\mathcal{S}(S_{\mathbf{a},n})$ for each prefractal table $\Omega(S_{\mathbf{a},n})$, and show that the genera $\{g_n\}_{n=0}^{\infty}$ of a sequence of translation surfaces $\{\mathcal{S}(S_{\mathbf{a},n})\}_{n=0}^{\infty}$ increase without bound. Various open questions and possible directions for future research are offered.

CHAPTER 2

STATISTICAL MECHANICS OF BOSE GAS IN SIERPINSKI CARPETS

The contents of this chapter are drawn from the preprint [17].

2.1 Introduction

This chapter is devoted to the study of the thermodynamics of quantum gases in fractal spaces. It was long recognized, by H. Lorentz and H. Weyl [80], that a deep connection exists between the thermodynamic properties of quantum gases and the underlying spectral geometry. One can probe the asymptotic behavior of elliptic linear differential operators in a given space by measuring the energy or pressure of a quantum gas in the said space. Since then much of this work has been done on Euclidean spaces. Here we wish to address the following question:

How does the fractal geometry affect the laws of quantum many-body physics? Conversely, what information about the fractal geometry can we obtain using probes made of quantum particles?

Below are some specific examples which are familiar to students of quantum and statistical physics, and which have wide ramifications in condensed matter physics and quantum field theory.

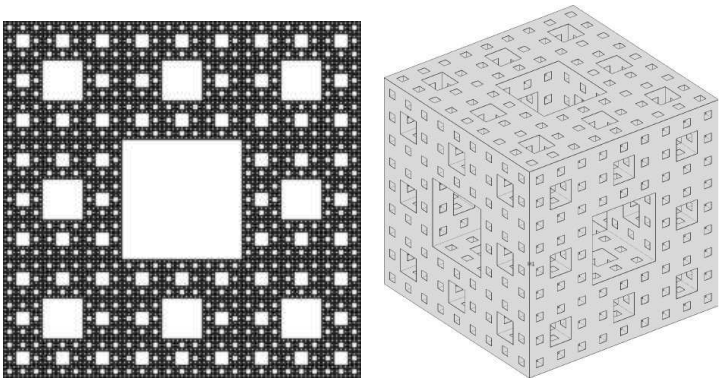
- Does Bose-Einstein condensation of atomic gas occur in non-integer dimension? If so, what dimension is it? And what is the critical density for condensation?
- What is the analog of the Stefan-Boltzmann law for blackbody radiation when the blackbody itself is a fractal?

- At zero temperature, what is the Casimir pressure produced by vacuum fluctuations inside a fractal?

We will answer all three questions in the case where the fractal is a *generalized Sierpinski carpet (GSC)*, whose representatives are the standard two-dimensional Sierpinski carpet and the three-dimensional Menger sponge (Fig. 2.1). These fractals have connected interior, are highly symmetric, and most of all, are infinitely ramified, which makes the analysis difficult. Conventional analytic tools, such as Fourier transform or spectral decimation, no longer apply. Essentially all the rigorous results known today originate from the study of Brownian motion. On the other hand, GSCs are embeddable in Euclidean space and resemble more realistic fractals found in nature. So we believe that a careful analysis of quantum gases in GSCs is warranted, because it provides us an avenue of attacking quantum many-body problems on general irregular spaces.

Our exposition is aimed at both physicists and mathematicians. It serves a dual purpose: to highlight the latest developments from the mathematical analysis on fractals, which have just begun to percolate through the physics community; and to illustrate how state-of-the-art potential theoretic results can be applied to answer physically inspired problems.

Notations. We write C and c for positive constants which may change from line to line. If a constant has a specific value, then we will add a numeral subscript to indicate this, *e.g.* C_1. Given two real-valued functions f and g, we say that f is comparable to g if there exist constants $c, C > 0$ such that $cg(x) \leq f(x) \leq Cg(x)$, denoted by $f(x) \asymp g(x)$ for short. For any $A \subset \mathbb{R}^d$ and $L > 0$, we write $LA = \{x \in \mathbb{R}^d : x/L \in A\}$. Finally, we use $\mathbb{N}_0$ and $\mathbb{N}$ to denote, respectively, the set of natural numbers with 0 and without 0.



(a) (b)

Figure 2.1: Examples of generalized Sierpinski carpet. (a) The standard Sierpinski carpet $SC(3,1)$, with $\ell_F = 3$ and $m_F = 8$. (b) The Menger sponge $MS(3,1)$, with $\ell_F = 3$ and $m_F = 20$.

2.1.1 Generalized Sierpinski carpet

Let $F_0 := [0,1]^d$ be the unit cube in $\mathbb{R}^d$, $d \geq 2$. Fix a *length scale factor* $\ell_F \in \mathbb{N}$, $\ell_F \geq 3$, and let $\mathcal{S}_n$ be the collection of closed cubes of side ℓ_F^{-n} with vertices in $\ell_F^{-n}\mathbb{Z}^d$. For $A \subset \mathbb{R}^d$, let $\mathcal{S}_n(A) = \{S \in \mathcal{S}_n : S \subset A\}$. Denote by Ψ_S the orientation-preserving affine map which maps F_0 to $S \in \mathcal{S}_n$.

Introduce a decreasing sequence $\{F_n\}_n$ of closed subsets of F_0 as follows. Fix $m_F \in \mathbb{N}$, $1 \leq m_F < \ell_F^d$, and let F_1 be the union of m_F distinct elements of $\mathcal{S}_1(F_0)$. In other words, we construct F_1 by retaining m_F of the cells of length ℓ_F^{-1}, and removing the rest. We will refer to m_F as the *mass scale factor*. Then by iteration we let

$$F_{n+1} = \bigcup_{S \in \mathcal{S}_n(F_n)} \Psi_S(F_1) = \bigcup_{S \in \mathcal{S}_1(F_1)} \Psi_S(F_n)\,,\ n \geq 1.$$

This iterated function system of contractions $\{\Psi_S\}$ has a unique fixed point $F = \bigcap_{n=0}^{\infty} F_n$. By standard arguments, the Hausdorff dimension of F is $d_h(F) = \log m_F / \log l_F$.

Definition 2.1.1. We say that F is a *generalized Sierpinski carpet (GSC)* if the following conditions on F_1 hold:

(H1) (Symmetry) F_1 is preserved under the isometries of the unit cube.

(H2) (Connectedness) $\mathrm{Int}(F_1)$ is connected, and contains a path connecting the hyperplanes $\{x_1 = 0\}$ and $\{x_1 = 1\}$.

(H3) (Non-diagonality) Let $m \geq 1$ and B be a cube of side length $2\ell_F^{-m}$, which is the union of 2^d distinct elements of $\mathcal{S}_m$. Then if $\mathrm{Int}(F_1 \cap B) \neq \emptyset$, it is connected.

(H4) (Borders included) F_1 contains the segment $\{x \in \mathbb{R}^d : x_1 \in [0,1],\ x_2 = \cdots = x_d = 0\}$.

We will denote by ∂F the boundary of F, and by $\partial_o F = \partial([0,1]^d) \cap \partial F$ the outer boundary.

In order to discuss the thermodynamic limit we need to consider unbounded versions of GSC. For each GSC F, we call $\tilde{F} := \bigcup_{n=0}^{\infty} l_F^n F_n$ the corresponding *pre-carpet*, and $F_\infty := \bigcup_{n=0}^{\infty} l_F^n F$ the *unbounded carpet*. Plainly speaking, the former is the infinite "blow-up" of the fractal, whereas the latter has both "blow-up" and "blow-down." Our focus will be on the unbounded carpet, where sharp results can be stated, though we believe many of them carry over to the pre-carpet modulo minor modifications.

Throughout the thesis, ν denotes the self-similar Borel probability measure on F which assigns mass m_F^{-n} to each $\Psi_S(F)$, $S \in \mathcal{S}_n$. Likewise, ν_∞ denotes the σ-finite self-similar measure on F_∞ which assigns mass m_F^n to $\ell_F^n F$.

2.1.2 Statement of the problem

Let F be a GSC, ν be the self-similar measure on F, and $\mathcal{H}_1 := L^2(F, \nu)$ be the single-particle Hilbert space. Given a self-adjoint, bounded-below Hamiltonian operator $H : \mathcal{H}_1 \to \mathcal{H}_1$, which captures the physics of a certain ideal Bose gas inside F, we wish

to compute its grand canonical partition function $\Xi_{\beta,\mu}$ at inverse temperature $\beta \in (0, \infty]$ and chemical potential $\mu \in (-\infty, \inf \mathrm{Spec}(H)]$:

$$\log \Xi_{\beta,\mu} = -\mathrm{Tr}_{\mathcal{H}_1} \log \left(1 - e^{-\beta(H-\mu)}\right). \tag{2.1}$$

Note that (2.1) already takes into account the Bose-Einstein statistics satisfied by the particles (see Section 2.4).

The Hamiltonian considered in this chapter involves the (nonnegative) Laplacian $-\Delta$ on the measure space (F, ν), constructed via either the Barlow-Bass approach (Brownian motions on the outer approximations) or the Kusuoka-Zhou approach (random walks on the associated graphs). Up to time change, both Laplacians generate the same (and the unique) Brownian motion which is invariant under the isometries of the carpet, as [8] has shown. We will concentrate on the following quantum gases, working in units where Planck's constant ($\hbar$), and either the speed of light or twice the atomic mass, are 1.

- **Massive Bose gas** (*e.g.* atoms), where $H = -\Delta$ and $\mu \leq \inf \mathrm{Spec}(H)$. The associated scalar field ψ satisfies the usual Schrödinger equation $i\partial_t \psi = -\Delta\psi$.
- **Massless Bose gas** (*e.g.* photons), where $H = \sqrt{-\Delta}$ and $\mu = 0$. Recall that the massless scalar field ψ satisfies the wave equation $-\partial_{tt}\psi = (-\Delta)\psi$. Since the Laplacian $-\Delta$ is a nonnegative self-adjoint operator, it has a unique square root $\sqrt{-\Delta}$. Thus we may take the formal square root of the wave equation to obtain a 1st-order-in-time PDE $i\partial_t\psi = \sqrt{-\Delta}\psi$, which is equivalent to a Schrödinger equation $i\partial_t\psi = H\psi$ with $H = \sqrt{-\Delta}$.

2.1.3 Methodology

An immediate obstacle to a rigorous thermodynamic calculation on GSCs is the lack of knowledge about the exact spectrum of the Laplacian. In fact, even the eigenvalue counting function (or the integrated density of states in physics parlance) in its optimal form cannot be proved (see Remark 2.2.3). Fortunately, many efforts which went into studying the short-time asymptotics of the heat kernel on GSC have produced sharp enough estimates, to the point that we can show that the corresponding spectral zeta function $\zeta_\Delta(s,\gamma) = \mathrm{Tr}_{\mathcal{H}_1}(-\Delta+\gamma)^{-s}$ admits a meromorphic extension (Section 2.3). This zeta function is then used effectively in computing the thermodynamic partition function, among other things.

We stress that the zeta function technology used on fractals applies equally well to Euclidean domains, manifolds, and graphs, and *a fortiori* produces the usual results on $\mathbb{R}^d$ or $\mathbb{Z}^d$. Furthermore, it unifies the treatments of massless and massive Bose gases: see [46] for a nice exposition in the $\mathbb{R}^d$ case.

2.1.4 Main results

Below is a summary of the major results on the thermodynamics of massive and massless Bose gases in Sierpinski carpets. All terminology will be explained in subsequent sections.

- Bose-Einstein condensation (Section 2.5). For a low-temperature, high-density ideal massive Bose gas in an unbounded GSC F_∞, Bose-Einstein condensation occurs at positive temperature if and only if the *spectral dimension* of the carpet $d_s(F_\infty) > 2$. In this case, whenever the Bose gas density exceeds $\overline{\rho}_c(\beta) :=$

$\frac{C_1(F_\infty)}{(4\pi\beta)^{d_s/2}}\zeta\left(\frac{d_s}{2}\right)$, where $C_1(F_\infty) \geq 1$ is a constant depending on F_∞ (see Proposition 2.5.3), then any excess density must condense in the lowest eigenfunction of the Laplacian. See Theorems 2.5.4 and 2.5.8.

- Blackbody radiation (Section 2.6). Let F be a GSC. At inverse temperature β, the energy per unit spectral volume of photons inside LF is

$$\mathcal{E}(\beta, L) = \beta^{-(d_s(F)+1)} H_1\left(-\log\left(\frac{\beta}{2L}\right)\right) + o(1) \quad \text{as} \quad L \to \infty,$$

  where H_1 is a periodic function of period $\frac{1}{2}d_w(F)\log l_F$ given in Proposition 2.6.3, and $d_w(F)$ is the *walk dimension* of the carpet.

- Casimir effect (Section 2.7). Consider a waveguide $\Omega_{a,b} = aF \times [0, b]$ modelled after a 2-dimensional GSC F, and impose Dirichlet conditions on the outer boundary. If one places a pair of parallel plates at $aF \times \{0\}$ and $aF \times \{b\}$, respectively, then the zero-temperature Casimir pressure on each plate is given by

$$P_{\text{Cas}}(a, b) = b^{-(d_s(\Omega)+1)} H_2\left(-\log\left(\frac{b}{a}\right)\right) + o(1) \quad \text{as} \quad a \to \infty,$$

  where $d_s(\Omega) = d_s(F) + 1$ is the spectral dimension of the waveguide $\Omega_{a,b}$, and H_2 is a periodic function of period $\frac{1}{2}d_w(F)\log l_F$ given in Proposition 2.7.2.

These are to be compared with the classical "textbook" results in Euclidean space:

- For a low-temperature, high-density ideal massive Bose gas in $\mathbb{R}^d$ or $\mathbb{Z}^d$, Bose-Einstein condensation occurs at positive temperature if and only if $d \geq 3$. In this case, whenever the Bose gas density exceeds $\rho_c(\beta) := \frac{1}{(4\pi\beta)^{d/2}}\zeta\left(\frac{d}{2}\right)$, any excess density must condense in the lowest eigenfunction of the Laplacian.

- At inverse temperature β, the energy per unit Euclidean volume of photons in a cube $[0, L]^d$ is

$$\mathcal{E}(\beta, L) = \frac{1}{\beta^{d+1}} \frac{d}{\pi^{(d+1)/2}} \Gamma\left(\frac{d+1}{2}\right) \zeta(d+1) + o(1) \quad \text{as} \quad L \to \infty.$$

When $d = 3$ we recover the familiar "T^4 law," $\mathcal{E}(\beta, L) = \pi^2/(30\beta^4) + o(1)$.

- Consider the rectangular waveguide $[0, a]^2 \times [0, b]$ with Dirichlet boundary conditions, with $[0, a]^2 \times \{0\}$ and $[0, a]^2 \times \{b\}$ being the two parallel plates. Then the zero-temperature Casimir pressure on each plate is given by

$$P_{\text{Cas}}(a, b) = -\frac{\pi^2}{240b^4} + o(1) \quad \text{as} \quad a \to \infty.$$

While our results are stated for true fractals, we believe that many of these behaviors can already be seen on finite-level approximations of the fractal, as indicated by our numerical work on the spectrum of the Laplacian [10, 19]. It would therefore be edifying if experimentalists can take on the challenge of constructing fractal-based quantum systems using, *e.g.* metamaterials, optical lattices, or superconducting qubits (for instance the proposal by [78]), and testing our results.

The chapter is organized as follows. In Section 2.2 we recapitulate several recently established results from the analysis on GSCs, namely the uniqueness of Brownian motion and the estimate of the heat kernel trace. In Section 2.3 we introduce the spectral zeta function on GSCs, show that it admits a meromorphic extension to $\mathbb{C}$, and give its poles and residues. After recalling the rudiments of quantum statistical mechanics in Section 2.4, we then present our thermodynamic computations of the massive and massless Bose gases in Sections 2.5, 2.6 and 2.7. We conclude with Section 2.8 by discussing a special case of interacting Bose gas on Sierpinski carpet graphs, and offering some open problems.

2.2 Established results on Sierpinski carpets

2.2.1 Existence and uniqueness of Brownian motion

In this subsection we give a quick account of how the Laplacian on GSC is constructed. The precise details are highly nontrivial and involve various techniques in potential theory (see *e.g.* [32] for background): we refer the reader to the original literature. For the purposes of this chapter, it is enough to recognize the following mathematical facts. A Laplacian Δ is in 1-to-1 correspondence with a nonnegative symmetric Markovian quadratic form, called the *Dirichlet form* $\mathcal{E}(u, v) = \int u(-\Delta v)$, on an appropriate Banach space. Moreover, the said Laplacian generates a Markov process, which in our setting is either a simple random walk (on a graph) or a Brownian motion (on a subset of $\mathbb{R}^d$).

The Barlow-Bass construction [3–6]. Let W_t^n be a reflecting Brownian motion on the nth approximating domain F_n of F. In order to produce a Brownian motion on the fractal F, one has to use the self-similarity of the carpet and take a suitable scaling limit. Barlow and Bass proved that there exists a family of time-scale factors $\{a_n\}_n$ satisfying

$$c_1 \left(\frac{\rho_F m_F}{l_F^2} \right)^n \leq a_n \leq c_2 \left(\frac{\rho_F m_F}{l_F^2} \right)^n , \tag{2.2}$$

for some $c_1, c_2, \rho_F \in (0, \infty)$ independent of n, such that the sequence of sped-up Brownian motions $X_t^n = W_{a_n t}^n$ on F_n has a subsequential limit. (The role of ρ_F will be discussed in the next subsection.) Any such limit process X_t, which respects the symmetry of F (henceforth referred to as F-symmetric), is called a Brownian motion on F. We denote its infinitesimal generator by $\mathcal{L}_{BB}$, the Barlow-Bass Laplacian (with Neumann conditions on $\partial_o F$).

One can also rephrase the above result in terms of Dirichlet forms. Let $\nu_n(dx) = (l_F^d / m_F)^n dx$ be the Borel probability measure on F_n which assigns equal weight to each

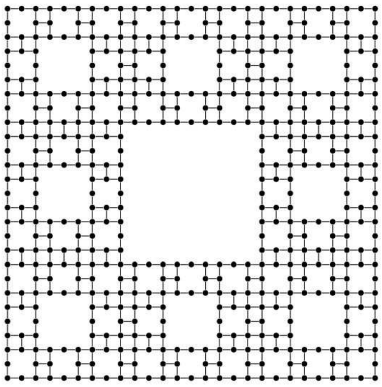


Figure 2.2: The level-3 graph approximation G_3 of the standard Sierpinski carpet $SC(3,1)$.

n-th level cell of F_n. (Note that ν_n converges weakly to a probability measure ν, which is a constant multiple of the $d_h(F)$-dimensional Hausdorff measure, on F.) We introduce the Dirichlet energy on (F_n, ν_n)

$$\mathcal{E}_n(u) = \int_{F_n} |\nabla u(x)|^2 \, \nu_n(dx) \tag{2.3}$$

for all $u \in L^2(F_n, \nu_n)$ such that $\mathcal{E}_n(u) < \infty$, and obtain the corresponding Dirichlet form by polarization: $\mathcal{E}_n(u,v) = \frac{1}{4}[\mathcal{E}_n(u+v) - \mathcal{E}_n(u-v)]$. Then the rescaled Dirichlet energies $\mathscr{E}_n(u) := a_n \mathcal{E}_n(u)$ converge in subsequence to

$$\mathscr{E}_{BB}(u) = \sup_{t>0} \frac{1}{t} \langle (1 - T_t)u, u \rangle_{L^2(F,\nu)}, \tag{2.4}$$

where $T_t = e^{t\mathcal{L}_{BB}}$ is the semigroup associated with X_t. The corresponding Dirichlet form is strongly local, regular, conservative, F-symmetric, and self-similar:

$$\mathscr{E}_{BB}(u,v) = \sum_{S \in \mathcal{S}_1(F_1)} \rho_F \mathscr{E}_{BB}(u \circ \Psi_S, v \circ \Psi_S). \tag{2.5}$$

The Kusuoka-Zhou construction [51]. Given F_n, let $G_n = (V_n, E_n)$ be the graph whose vertices lie at the centers of the cells $S \in \mathcal{S}_n(F_n)$, and whose edges connect vertices in nearest neighboring cells: see Fig. 2.2. Define the Dirichlet energy to be the

usual graph energy

$$\mathcal{E}_n(u) = \sum_{\langle xy\rangle \in E_n} [u(x) - u(y)]^2$$

for all continuous functions $u, v \in C(G_n; \mathbb{R})$, and obtain the Dirichlet form via polarization. By checking a series of geometric conditions for which Poincaré and Harnack inequalities hold, Kusuoka and Zhou were able to prove that $\{\rho_F^n \mathcal{E}_n\}_n$ converges in subsequence to a strongly local, regular, conservative, F-symmetric Dirichlet form $\mathscr{E}_{KZ}$ satisfying the self-similar identity

$$\mathscr{E}_{KZ}(u, v) = \sum_{S \in S_1(F_1)} \rho_F \mathscr{E}_{KZ}(u \circ \Psi_S, v \circ \Psi_S). \tag{2.6}$$

The corresponding Brownian motion has infinitesimal generator $\mathcal{L}_{KZ}$, which we call the Kusuoka-Zhou Laplacian.

A question which lingered for almost two decades was whether the two constructions yield the same limiting Laplacian on F. This was settled definitively by Barlow, Bass, Kumagai & Teplyaev.

Theorem 2.2.1 ([8, Theorem 1.2])**.** *Let F be a GSC equipped with the self-similar measure ν. Up to scalar multiples, the set of non-zero, local, regular, conservative, and F-symmetric Dirichlet forms on (F, ν) contains at most one element.*

An equivalent statement to Theorem 2.2.1 is that up to deterministic time change, $\mathcal{L}_{BB}$ and $\mathcal{L}_{KZ}$ both generate the unique F-symmetric Brownian motion. Henceforth we will denote this unique Laplacian by Δ.

It should be noted that through extensive numerical computations [CS,BKS], we can demonstrate that both versions of the Laplacian on *finite* approximations F_n of F coincide at the bottom of the spectrum, as Fig. 2.3 shows. Notice that upon removing the power-law growth from the eigenvalue counting function, the remainder exhibits

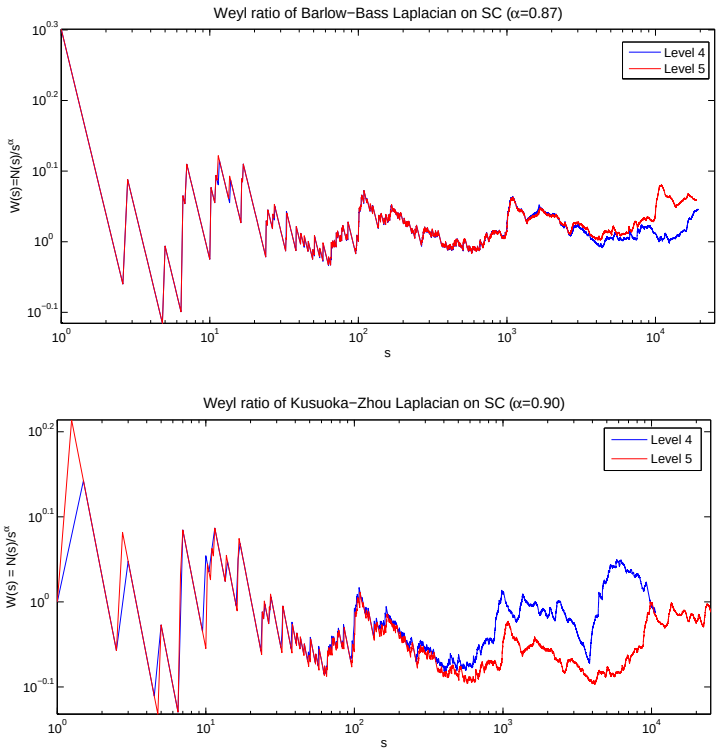


Figure 2.3: Numerical Weyl ratios $W(s) = s^{-d_s/2}N(s)$ associated with (a) the Barlow-Bass Laplacian [19] and (b) the Kusuoka-Zhou Laplacian [10] on approximations F_4 and F_5 of the standard two-dimensional Sierpinski carpet $F = SC(3,1)$, where $N(s) := \#\{\lambda \in \mathrm{Spec}(-\Delta) : \lambda/\lambda_1 < s\}$ is the eigenvalue counting function of the Neumann Laplacian, normalized by the lowest nonzero eigenvalue λ_1.

logarithmically periodic oscillation, which indicates the fractal nature of the geometry (see Section 2.2.3 for more details). For large enough n (typically $n \geq 3$), the first $(n-2)$ segments of the two spectra, and to a lesser extent the $(n-1)$th segment, agree very well. Then deviation creeps in the nth segment: the Barlow-Bass spectrum becomes more characteristic of the spectrum on $[0,1]^d$, while the Kusuoka-Zhou spectrum is truncated due to the finite cardinality m_F^n of the graph.

So despite our inability to state precise results rigorously on the finite approximations F_n, we believe strongly that all the asymptotic thermodynamic results presented in this chapter, which are stated for F but invariably tied to the bottom of the spectrum,

should be visible on F_n for sufficiently large n, beginning perhaps from $n = 3$.

2.2.2 Resistance and heat kernel estimates

The aforementioned coefficient ρ_F in (2.2) is called the *resistance scale factor* of the carpet F. When viewing each F_n as an electrical network, ρ_F gives the renormalization factor relating the resistance of F_n to that of F_{n+1}. To this date there is no known closed form expression for ρ_F. The best known bound, obtained by cutting and shorting resistances, is [6, Proposition 5.1]

$$\frac{l_F^2}{m_F} \leq \rho_F \leq 2^{1-d}\ell_F. \tag{2.7}$$

The connection between resistance and Brownian motion is as follows. Let $d_w(F) = \log(\rho_F m_F)/\log l_F$ be the **walk dimension** of the carpet: this is the time-to-space scaling exponent for Brownian motion. For instance, if $\tau(A) = \inf\{t > 0 : X_t \notin A\}$ is the exit time of a Brownian motion X_t from the set A, then d_w is defined through $\mathbb{E}^x[\tau(B(x,r))] \asymp r^{d_w}$, where $\mathbb{E}^x$ is the expectation with respect to the law of Brownian motions started at x.

A crucial point to make here is that while manifolds and Euclidean domains have $d_w = 2$, fractals have $d_w > 2$, as can be seen through *e.g.* (2.7). In a nutshell, this means that Brownian motion proceeds "more slowly" on fractals, due to the presence of obstacles at all length scales. One way to see this explicitly is through the **heat kernel** $p_t(x,y)$, which is the integral kernel of the Markov semigroup $T_t = e^{t\Delta}$, or equivalently, the transition density of Brownian motion. On fractals, regarded as a metric measure space (F, d, ν), the (short-time) heat kernel obeys sub-Gaussian rather than Gaussian bounds [5]:

$$p_t(x,y) \asymp Ct^{-d_h/d_w} \exp\left(-c\left(\frac{d(x,y)^{d_w}}{t}\right)^{\frac{1}{d_w-1}}\right), \quad t \in (0,1),\ x,y \in F. \tag{2.8}$$

The leading power-law term is t^{-d_h/d_w}, as opposed to the $t^{-d/2}$ on d-dimensional Euclidean domains. This observation leads us to introduce the **spectral dimension** of the carpet $d_s(F) := 2d_h(F)/d_w(F) = 2\log m_F/\log(\rho_F m_F)$. Note that for any GSC in $\mathbb{R}^d$, $1 \le d_s(F) < d_h(F) < d$.

The spectral dimension is a physically meaningful dimension because it is tied to the macroscopic behavior of Brownian motion. Recall that a Markov process on an unbounded state space is said to be *recurrent* if, with probability one, the process returns to the origin infinitely many times. Otherwise it is said to be *transient*, *i.e.,* with positive probability the process leaves for infinity. Roughly speaking, on a homogeneous space, $d_s = 2$ is the dimensional threshold above which Brownian motion is transient, and below which it is recurrent. Brownian motion at $d_s = 2$ is typically also recurrent, but its behavior is more subtle than the $d_s < 2$ case.

Indeed, for Brownian motion in an unbounded GSC F_∞, Barlow and Bass showed that the process is transient (resp. recurrent) if $d_s(F) > 2$ or $\rho_F < 1$ (resp. $d_s(F) \le 2$ or $\rho_F \ge 1$) [6, Theorem 8.1]. The same dichotomy holds on the pre-carpet $\tilde{F}$ [6, Theorem 8.7]. This distinction has important consequences for the thermodynamics of the massive Bose gas, as we will explain later.

To make comparisons, we mention in passing that all post-critically finite (pcf) fractals (for definition see [45, 74]), *e.g.* any d-dimensional Sierpinski gasket with $d \ge 2$, have $\rho_F > 1$ or $d_s < 2$, and support recurrent Brownian motion. Meanwhile, as an example of a random fractal, the incipient infinite cluster (IIC) at criticality of bond percolation on $\mathbb{Z}^d$ with $d \ge 19$, or of spread-out percolation with $d > 6$, has $d_s = 4/3$, which was proved recently by [47].

2.2.3 Sharp estimates of the heat kernel trace

In what follows we call $K(t) := \mathrm{Tr}(e^{t\Delta}) = \int_F p_t(x,x)d\nu(x)$ the *heat kernel trace*. Since the Laplacian on GSCs admits an eigenfunction expansion, one can write $K(t) = \sum_{j=0}^{\infty} e^{-t\lambda_j}$, where $\{\lambda_j\}$ are the eigenvalues of $-\Delta$. Equivalently, $K(t)$ represents the probability that Brownian motion started at any point x returns to the said point x at time t.

Using (2.8) we can already deduce that on GSC, $K(t) \asymp t^{-d_h/d_w}$ for small t. This estimate is typically enough on simple spaces, but on deterministic self-similar fractals there is logarithmically periodic modulation on top of the power-law dependence in t, which is attributed to the discrete scale invariance of the space. Indeed, using renewal theorem type arguments, Hambly and Kajino separately proved the following short-time asymptotics of the heat kernel trace.

Theorem 2.2.2 ([36,42])**.** *Let F be a GSC, ν be the self-similar measure on F, and Δ be the Laplacian on (F,ν) associated with either Dirichlet or Neumann condition on $\partial_o F$. Then*

$$K(t) = t^{-d_h/d_w}\left[G(-\log t) + o(1)\right] \quad \text{as } t \downarrow 0, \tag{2.9}$$

where G is $\log(\rho_F m_F)$-periodic, and is bounded away from 0 and ∞.

Remark 2.2.3. Let $N(s) := \#\{\lambda < s : \lambda \in \mathrm{Spec}(-\Delta)\}$ be the eigenvalue counting function (or the integrated density of states) of the Laplacian. Then $K(t) = \int_0^\infty e^{-ts}dN(s)$, *i.e.,* $K(\cdot)$ is the Laplace-Stieltjes transform of $N(\cdot)$. Using the estimate $K(t) \asymp t^{-d_s/2}$ one easily obtains the Weyl asymptotics $N(s) \asymp s^{d_s/2}$ for large s.

In view of (2.9), it is tempting to go a step further and claim that

$$N(s) = s^{d_s/2}\left[h(\log s) + o(1)\right] \quad \text{as } s \to \infty, \tag{2.10}$$

where h is $\log(\rho_F m_F)$-periodic and is bounded away from 0 and ∞. There is much numerical evidence that (2.10) holds on two-dimensional and three-dimensional

GSCs [12, CS, BKS], as Figure 2.3 demonstrates. Indeed, if (2.10) is true, then (2.9) follows immediately. However, as pointed out in [36, Section 4.2] and [42, Section 9], the inverse Laplace-Stieltjes transform involves sophisticated Tauberian theorems which are very difficult to prove. This explains why we have elevated the role of the heat kernel trace over that of the (integrated) density of states.

More recently, Kajino [43] proved a sharper estimate of the heat kernel trace than (2.9) by exploiting the full symmetry of the GSC and its boundaries.

Theorem 2.2.4. *Let F be a GSC, ν be the self-similar measure on F, and Δ be the Laplacian on (F, ν) with Dirichlet conditions on $\partial_o F$. Then there exist continuous,* $\log(\rho_F m_F)$*-periodic functions $G_k : \mathbb{R} \to \mathbb{R}$ for $k = 0, 1, \cdots, d$ such that*

$$K(t) = \sum_{k=0}^{d} t^{-d_k/d_w} G_k(-\log t) + O\left(\exp\left(-ct^{-\frac{1}{d_w - 1}}\right)\right) \quad \text{as } t \downarrow 0. \tag{2.11}$$

Here $d_k := d_h(F \cap \{x_1 = \cdots x_k = 0\})$. Moreover $G_0 > 0$ and $G_1 < 0$.

Remark 2.2.5. In [43] it is also proved that the same estimate holds on GSCs with Neumann conditions on $\partial_o F$, whereby $G_1 > 0$.

As the G_k are periodic, we can expand them in Fourier series: $G_k(x) = \sum_{p \in \mathbb{Z}} \hat{G}_{k,p} e^{2\pi p i x / \log R}$, where $R := \rho_F m_F = d_w(F) \log l_F$. We should note that very little is known about the functions G_k, except for the signs of G_0 and G_1. There is strong numerical evidence that G_0 is nonconstant, with $(\max G_0 - \min G_0)/\hat{G}_{0,0}$ typically on the order of 10^{-2} for GSCs in $\mathbb{R}^2$ or $\mathbb{R}^3$ [CS]; but this has not been rigorously established.

We point out three important features of the estimate (2.11) which are crucial to our thermodynamic computations in the sequel:

1. The polynomial terms all have nonpositive exponents $(-d_k/d_w)_k$, where d_k is the

Hausdorff dimension of the codimension-k outer boundary of F. In particular, $d_0 = d_h(F)$, $d_1 = d_h(\partial_o F)$, $d_{d-1} = 1$, and $d_d = 0$.

2. Since the G_k are real-valued bounded functions, we have $\hat{G}_{k,p} = \overline{\hat{G}_{k,-p}}$, and $|\hat{G}_{k,p}| \leq \frac{1}{2\pi}\int_0^{\log R} |G_k(x)|dx \leq C < \infty$ for all $p \in \mathbb{Z}$.

3. The remainder term decays exponentially in t, which is better than the power-law decay in the estimate (2.9).

The first two features guarantee that the thermodynamic quantities (energy, pressure, etc.) associated with GSCs will be finite and meaningful. The last feature, which is the most important of all, enables us to carry out the Casimir energy calculation unambiguously and without making any *ad hoc* regularization.

Now recall that the heat kernel trace on Riemannian manifolds $M \subset \mathbb{R}^d$ has the short-time asymptotics

$$K(t) = \frac{\mathrm{Vol}(M)}{(4\pi t)^{d/2}} + O\left(t^{-(d-1)/2}\right) \quad \text{as } t \downarrow 0. \tag{2.12}$$

Going back to (2.11), it is easy to check that if F is a GSC, then the heat kernel trace for LF ($L > 0$) is given by $K(L^{-2}t) = L^{d_s} G_0(-\log(L^{-2}t))t^{-d_s/2}$ + lower order terms as $t \downarrow 0$. Comparing this against (2.12), we see that LF has an effective spectral volume of $C_2 L^{d_s}(4\pi)^{d_s/2}$, where $C_2 \asymp G_0$. In fact, we can use Cesàro averaging to define more precisely the spectral content of a GSC, in the same manner that the Minkowski content of a fractal string is defined [56]. This definition also appeared in [1].

Definition 2.2.6. Let F be a GSC. Then we define the *spectral volume* of LF ($L > 0$) to be

$$V_s(LF) := L^{d_s}(4\pi)^{d_s/2} \lim_{n\to\infty} \frac{1}{n \log R} \int_{R^{-n}}^{1} G_0(-\log t)\frac{dt}{t} = (4\pi)^{d_s/2}\hat{G}_{0,0}L^{d_s}. \tag{2.13}$$

We emphasize that the anomalous walk dimension $d_w > 2$ of the GSC results in its spectral volume being distinct from its Hausdorff volume. As will be made clear in later sections, all "extensive" thermal observables (*e.g.* energy, particle number) scale with the spectral volume, not the Hausdorff volume.

Moreover we denote $d_{k,p} := 2\left(\frac{d_k}{d_w} + \frac{2p\pi i}{\log R}\right)$. These are the *complex dimensions* of the GSC in the sense of M. Lapidus [56], which reflect both the various spectral contents (via the real part of $d_{k,p}$) and the discrete scale invariance of the fractal (via the imaginary part). Using this notation, (2.11) can be rewritten as

$$K(t) = \sum_{k=0}^{d} \sum_{p\in\mathbb{Z}} \hat{G}_{k,p} t^{-d_{k,p}/2} + O\left(\exp\left(-ct^{-\frac{1}{d_w-1}}\right)\right) \quad \text{as} \quad t \downarrow 0. \tag{2.14}$$

2.3 Spectral zeta function on Sierpinski carpets

The spectral zeta function of a self-adjoint Laplacian Δ on a bounded domain is given by

$$\zeta_\Delta(s,\gamma) := \mathrm{Tr}\frac{1}{(-\Delta+\gamma)^s}.$$

This can be written as a Mellin transform of the heat kernel trace

$$\zeta_\Delta(s,\gamma) = \frac{1}{\Gamma(s)}\int_0^\infty t^s e^{-\gamma t} K(t)\frac{dt}{t}, \tag{2.15}$$

whenever the right-hand side is defined. In the case of a GSC with Dirichlet boundary, (2.11) implies that the absicssa of convergence for $\zeta_\Delta(\cdot,\gamma)$ is located at $\mathrm{Re}(s) = d_{0,0} = d_h(F)/d_w$. It is natural to ask whether this function can be extended to the entire complex plane, save for a countable (possibly infinite) number of poles. The following theorem, which was reported in [73], serves as the linchpin for all subsequent results in this chapter.

Theorem 2.3.1. *The spectral zeta function $s \mapsto \zeta_\Delta(s,\gamma)$ of the Laplacian on a GSC F, with Dirichlet conditions on $\partial_o F$, admits a meromorphic extension to $\mathbb{C}$.*

Proof. Without loss of generality we suppose that (2.14) holds for $t \in (0,1)$, while there exist $c_3, c_4 > 0$ such that $K(t) \le c_3 e^{-c_4 t}$ for $t > 1$. Then

$$\zeta_\Delta(s,\gamma)\Gamma(s) = I_1(s,\gamma) + I_2(s,\gamma) + I_3(s,\gamma),$$

where

$$\begin{aligned} I_1(s,\gamma) &= \int_0^1 t^s e^{-\gamma t} \sum_{k=0}^{d} \sum_{p\in\mathbb{Z}} \hat{G}_{k,p} t^{-d_{k,p}/2} \frac{dt}{t}, \\ I_2(s,\gamma) &= \int_0^1 t^s e^{-\gamma t} O\left(\exp\left(-ct^{-\frac{1}{d_w-1}}\right)\right) \frac{dt}{t}, \\ |I_3(s,\gamma)| &\le c_3 \left| \int_1^\infty t^s e^{-\gamma t} e^{-c_4 t} \frac{dt}{t} \right|. \end{aligned}$$

To compute I_1 we use the identity

$$\int_0^1 t^s e^{-\gamma t} t^{-p} \frac{dt}{t} = \int_0^1 t^s \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \gamma^n t^n t^{-p} \frac{dt}{t} = \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^n}{n!(s+n-p)}$$

for $\mathrm{Re}(p) \ge 0$. Using linearity of the polynomial terms we get

$$I_1(s,\gamma) = \sum_{k=0}^{d} \sum_{p\in\mathbb{Z}} \sum_{n=0}^{\infty} \frac{(-1)^n \gamma^n \hat{G}_{k,p}}{n!(s+n-d_{k,p}/2)}.$$

In deriving this expression we interchanged the order of p-summation and integration, which we justify as follows. For fixed $\gamma \neq 0$, the summand indicates that the simple poles are $d_{k,p}/2 - \mathbb{N}_0$. Away from the poles $s \mapsto I_1(s,\gamma)$ is holomorphic, since for each $k = 0, 1, \cdots, d$ and each $r \in \mathbb{C}$,

$$\left| \frac{\hat{G}_{k,p}}{r - d_{k,p}/2} + \frac{\hat{G}_{k,-p}}{r - d_{k,-p}/2} \right| \le C \left| \frac{2(r - \frac{d_k}{d_w})}{(r - \frac{d_k}{d_w})^2 + (\frac{2p\pi}{\log R})^2} \right|$$

is summable over $p \in \mathbb{N}$. When $\gamma = 0$, we find

$$I_1(s,0) = \sum_{k=0}^{d} \sum_{p\in\mathbb{Z}} \frac{\hat{G}_{k,p}}{s - d_{k,p}/2}.$$

Next we let $\alpha := 1/(d_w - 1)$, and make the bound

$$\left|\frac{I_2(s,\gamma)}{\Gamma(s)}\right| \le C\left|\frac{1}{\Gamma(s)}\int_0^1 t^s e^{-\gamma t} e^{-ct^{-\alpha}}\frac{dt}{t}\right| \le Ce^{-c}\left|\frac{\gamma^{-s}}{\Gamma(s)}\boldsymbol{\gamma}(s,\gamma)\right|,$$

where $\boldsymbol{\gamma}(s,x)$ is the lower incomplete gamma function. Since

$$\boldsymbol{\gamma}^*(s,\gamma) := \frac{\gamma^{-s}}{\Gamma(s)}\boldsymbol{\gamma}(s,\gamma) = e^{-\gamma}\sum_{m=0}^{\infty}\frac{\gamma^m}{\Gamma(s+m+1)}$$

is entire in both s and γ, $I_2(s,\gamma)/\Gamma(s)$ contributes an entire-in-s component to $\zeta_\Delta(s,\gamma)$.

For I_3, it suffices to know that for $\mathrm{Re}(\gamma) > -c_4$,

$$|I_3(s,\gamma)| \le c_3\left|(\gamma+c_4)^{-s}\int_{\gamma+c_4}^{\infty} u^s e^{-u}\frac{du}{u}\right| = c_3\left|(\gamma+c_4)^{-s}\Gamma(s,\gamma+c_4)\right|$$

is finite, thanks to the holomorphicity of the upper incomplete gamma function $s \mapsto \Gamma(s,x)$ when $x \neq 0$.

So finally we obtain, for $\mathrm{Re}(\gamma) > -c_4$, the Mittag-Leffler decomposition

$$\zeta_\Delta(s,\gamma) = \frac{1}{\Gamma(s)}\sum_{k=0}^{d}\sum_{p\in\mathbb{Z}}\sum_{n=0}^{\infty}\frac{(-1)^n\gamma^n\hat{G}_{k,p}}{n!\left(s+n-d_{k,p}/2\right)} + (\text{entire function in } s). \tag{2.16}$$

This proves the theorem. ■

Now we can enumerate all the possible poles and residues of the spectral zeta function on a GSC.

Corollary 2.3.2. *For a GSC F with Dirichlet condition on $\partial_o F$:*

1. *The poles of $\zeta_\Delta(\cdot,\gamma)$, where $\mathrm{Re}(\gamma) \ge -c_4$, are contained in the set*

$$\bigcup_{k=0}^{d}\bigcup_{p\in\mathbb{Z}}\bigcup_{n=0}^{\infty}\left\{-n+\frac{d_{k,p}}{2}\right\},$$

with corresponding residues

$$\mathrm{Res}\left(\zeta_\Delta(\cdot,\gamma), -n+\frac{d_{k,p}}{2}\right) = \frac{(-1)^n\gamma^n\hat{G}_{k,p}}{n!\Gamma\left(-n+d_{k,p}/2\right)}.$$

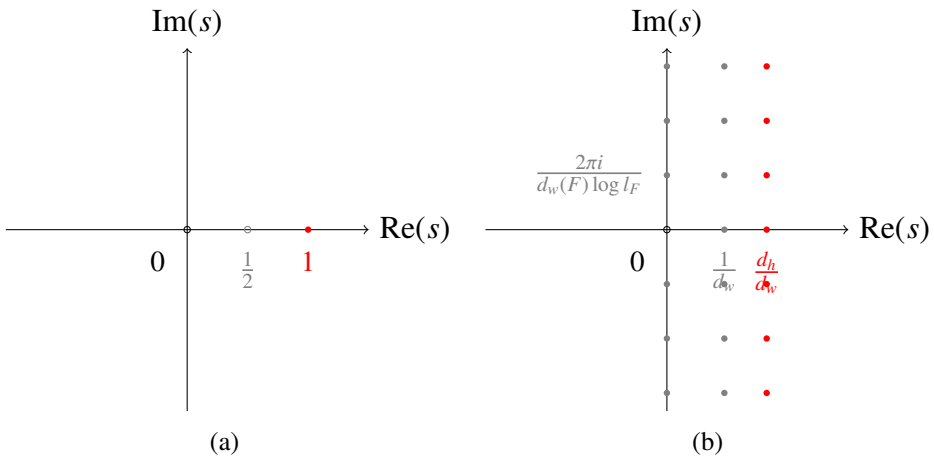


Figure 2.4: The singularities of $s \mapsto \zeta_\Delta(s, 0)$ associated to (a) a 2-dimensional parallelpiped, using the Epstein zeta function; (b) a 2-dimensional GSC, using Corollary 2.3.2. A hollow circle indicates that the singularity is either a removable singularity, or a pole with zero residue.

2. *The poles of* $\zeta_\Delta(\cdot, 0)$ *are contained in the set* $\bigcup_{k=0}^{d} \bigcup_{p\in\mathbb{Z}} \left\{ \frac{d_{k,p}}{2} \right\}$, *with corresponding residues*

$$\operatorname{Res}\left(\zeta_\Delta(\cdot, 0), \frac{d_{k,p}}{2}\right) = \frac{\hat{G}_{k,p}}{\Gamma\left(d_{k,p}/2\right)}.$$

In particular, $\zeta_\Delta(s, 0)$ *is analytic for* $\operatorname{Re}(s) < 0$.

3. $\zeta_\Delta(s, 0) = 0$ *for all* $s \in -\mathbb{N}$.

Proof. The poles and residues can be read off from (2.16), so the first two claims are immediate. Note that $d_{d,0}/2 - \mathbb{N} = -\mathbb{N}$ are poles of $\zeta_\Delta(\cdot, \gamma)$ with zero residue because $1/\Gamma(s) = 0$ for all $s \in -\mathbb{N}$. The last claim follows from the observation that when $\operatorname{Re}(s) < 0$, $\zeta_\Delta(s, 0)$ is a product of $1/\Gamma(s)$ and a holomorphic function. ■

Fig. 2.4 depicts the singularities of $\zeta_\Delta(\cdot, 0)$ for a typical GSC versus that for a parallelpiped. In both domains, the poles with the largest real part represent the spectral volume dimension. What distinguishes a GSC from an Euclidean domain is the "tower" of poles, or *complex dimensions*, along the imaginary direction at each $\operatorname{Re}(s) = d_{k,0}$, with

even spacing of $2\pi i/(d_w \log l_F)$. Thus the discrete scale invariance of the GSC manifests itself in two ways: it produces log-periodic modulations of the heat kernel trace and, via a Mellin transform, creates towers of complex-valued poles in the spectral zeta function.

2.4 Review of quantum statistical mechanics

Here we briefly review notions of quantum statistical mechanics that are needed for the rest of the chapter. Much of this is known to physicists, but we use the language of functional analysis to make notations precise. For a complete treatment please consult, e.g., [13, Chapter 5].

2.4.1 Gibbs state and partition function

The Bose (resp. Fermi) gas is a system of many quantum particles satisfying an appropriate rule of quantum statistics. Let us denote the Hilbert space for a single quantum particle confined to domain $F \subset \mathbb{R}^d$ by $\mathcal{H}_1 := L^2(F)$: the state of the particle is described by a wavefunction $\psi \in \mathcal{H}_1$. For a system of n such identical bosons (resp. fermions), the wavefunction $\psi_n(x_1, \cdots, x_n) \in L^2(F^n)$ is (anti)symmetric under particle exchange: $\psi_n(\cdots, x_i, \cdots, x_j, \cdots) = \pm\psi_n(\cdots, x_j, \cdots, x_i, \cdots)$. Therefore the appropriate n-body Hilbert space, denoted by $\mathcal{H}_n$, is the (anti)symmetric subspace of $\mathcal{H}_1^{\otimes n}$. By default we set $\mathcal{H}_0 = \mathbb{C}$.

To describe the equilibrium thermodynamics of the quantum many-body system, it is convenient to adopt the *grand canonical ensemble*, where the particle number is not kept fixed. Mathematically, we take the *Fock space* $\mathcal{F} = \overline{\bigoplus_{n=0}^{\infty} \mathcal{H}_n}$ as the underlying Hilbert space. For each one-body self-adjoint operator $W : \mathcal{H}_1 \to \mathcal{H}_1$, define $W_n : \mathcal{H}_n \to \mathcal{H}_n$

by

$$W_n P_\pm(\psi_1 \otimes \cdots \otimes \psi_n) = \begin{cases} 0, & \text{if } n = 0 \\ \sum_{j=1}^{n} P_\pm\left(\psi_1 \otimes \cdots \otimes (W\psi_j) \otimes \cdots \otimes \psi_n\right), & \text{if } n \geq 1 \end{cases}$$

for all $\psi_j \in \mathcal{H}_1$, where $P_\pm$ is the (anti)symmetric permutation on the components of the tensor product. Then we call $d\Gamma(W) := \overline{\bigoplus_{n=0}^{\infty} W_n}$ the *second quantization* of W on $\mathcal{F}$. An important example is where W is the identity operator $\mathbf{1}$ on $\mathcal{H}_1$: then $\mathbf{1}_n$ is n times the identity operator on $\mathcal{H}_n$, so its second quantization reads $d\Gamma(\mathbf{1}) = \oplus_{n=0}^{\infty} n =: \mathbf{N}$, also known as the *number operator* on $\mathcal{F}$.

For our purposes it suffices to consider a one-body Hamiltonian operator H which is bounded below. Let $\beta \in (0, \infty]$ be the inverse temperature, $\mu \in (-\infty, \inf \mathrm{Spec}(H)]$ be the chemical potential, and $\mathbf{H} := d\Gamma(H)$. Consequently $\rho_{\beta,\mu} := e^{-\beta d\Gamma(H-\mu\mathbf{1})}$ is a trace-class operator on $\mathcal{F}$, and its trace

$$\Xi_{\beta,\mu} := \mathrm{Tr}_{\mathcal{F}} \rho_{\beta,\mu} = \mathrm{Tr}_{\mathcal{F}} e^{-\beta(\mathbf{H}-\mu\mathbf{N})}$$

is called the *grand canonical partition function* of a thermal system with Hamiltonian H. The *free energy* of the system is

$$F_{\beta,\mu} = -\beta^{-1} \log \Xi_{\beta,\mu}.$$

The *Gibbs state* $\omega_{\beta,\mu}$ at (β, μ) is a linear functional over the C^*-algebra on $\mathcal{F}$ such that

$$\omega_{\beta,\mu}(\mathbf{A}) = \Xi_{\beta,\mu}^{-1} \mathrm{Tr}_{\mathcal{F}}\left(\mathbf{A}\rho_{\beta,\mu}\right).$$

Physically, if A belongs to the (*-)algebra of observables on $\mathcal{H}_1$, and $\mathbf{A} = d\Gamma(A)$, then $\omega_{\beta,\mu}(\mathbf{A})$ represents the expectation value of A in the grand canonical Gibbs state. To experts of operator algebra, we remark that the Gibbs state is a (τ, β)-Kubo-Martin-Schwinger (KMS) state [13, Definition 5.3.1]: namely, for all $\mathbf{A}, \mathbf{B}$ in the algebra of observables on $\mathcal{F}$,

$$\omega_{\beta,\mu}(\mathbf{A}\tau_{i\beta}(\mathbf{B})) = \omega_{\beta,\mu}(\mathbf{B}\mathbf{A}), \qquad \text{where } \tau_t(\mathbf{O}) := e^{it(\mathbf{H}-\mu\mathbf{N})}\mathbf{O}e^{-it(\mathbf{H}-\mu\mathbf{N})}.$$

Two observables we will be interested in the sequel are the expected number of bosons

$$\omega_{\beta,\mu}(\mathbf{N}) = \beta^{-1}\frac{1}{\Xi_{\beta,\mu}}\frac{\partial \Xi_{\beta,\mu}}{\partial \mu} = \beta^{-1}\frac{\partial}{\partial \mu}\log \Xi_{\beta,\mu},$$

and the expected energy

$$\omega_{\beta,\mu}(\mathbf{H}) = -\frac{1}{\Xi_{\beta,\mu}}\frac{\partial \Xi_{\beta,\mu}}{\partial \beta} = -\frac{\partial}{\partial \beta}\log \Xi_{\beta,\mu}.$$

For a non-interacting Bose gas the expression for the partition function can be reduced to a trace over the one-body Hilbert space $\mathcal{H}_1$. Assume that H has pure point spectrum with increasing eigenvalues $0 \leq E_0 < E_1 \leq \cdots \leq E_i \leq \cdots \uparrow \infty$. Then

$$\Xi_{\beta,\mu} = \prod_i \sum_{n_i=0}^{\infty} e^{-n_i\beta(E_i-\mu)} = \prod_i \frac{1}{1-e^{-\beta(E_i-\mu)}}.$$

Therefore

$$\begin{aligned}
\log \Xi_{\beta,\mu} &= -\mathrm{Tr}_{\mathcal{H}_1} \log(1-e^{-\beta(H-\mu)}),\\
\omega_{\beta,\mu}(\mathbf{N}) &= \mathrm{Tr}_{\mathcal{H}_1} \frac{1}{e^{\beta(H-\mu)}-1},\\
\omega_{\beta,\mu}(\mathbf{H}) &= \mathrm{Tr}_{\mathcal{H}_1} \frac{He^{\beta(H-\mu)}}{e^{\beta(H-\mu)}-1}.
\end{aligned}$$

The following fact is useful and will be invoked several times: $\mu \mapsto \omega_{\beta,\mu}(\mathbf{N})$ is a convex, monotone increasing function which maps $(-\infty, \inf \mathrm{Spec}(H)]$ to $(0, \infty]$.

2.4.2 Thermodynamic limit

Since we will deal with the delicate topic of phase transitions on fractals, it is important to discuss how we take the thermodynamic limit. As is customary, we exhaust an unbounded space S_∞ by an increasing family of finite subsets $\{S_n\}_n$ with suitable boundary

conditions. The thermodynamic limit of an "intensive" thermal observable (*e.g.* pressure, particle density) on S_∞ is then understood as the limit of a sequence of the said observables on S_n, subject to proper constraints.

For the unbounded carpet F_∞, we choose the natural exhaustion $\{\Lambda_n\}_n := \{\ell_F^n F\}_n$. This involves scaling up the size of the system by ℓ_F each time, which in effect scales down the Laplacian by ℓ_F^{-2}. To make this length dependence explicit, we will denote by $\underline{\Delta}$ the dimensionful Laplacian on a bounded domain not necessarily of unit characteristic length. If the characteristic length is L, then we write $\underline{\Delta}_L$, which is L^{-2} times the dimensionless Laplacian Δ (on the domain of unit length). In the same manner we will add a subscript L to any quantity which depends explicitly on the domain size L.

We will show shortly that, due to the log-periodic modulation inherited from the heat kernel trace, the thermodynamic limit along the natural exhaustion does not exist in the literal sense, but rather in the weaker limsup/liminf sense. Of course our choice of exhaustion is by no means unique. There might be other choices of exhaustion for which a stronger limit is attained, but we suspect that those would marginally improve the results to be presented below.

2.5 Massive Bose gas in Sierpinski carpets

In this section we focus on the Bose gas with $H = -\Delta$ and $\mu \leq \inf \mathrm{Spec}(H)$. This Hamiltonian captures the behavior of a gas of non-relativistic, non-interacting, and charge-neutral atoms. The partition function of the massive Bose gas in a general bounded domain is given by

$$\log \Xi_{\beta,\mu} = -\mathrm{Tr}_{\mathcal{H}_1} \log\left(1 - e^{-\beta(-\underline{\Delta}-\mu)}\right). \tag{2.17}$$

Its connection to the spectral zeta function is the following.

Proposition 2.5.1. *The grand canonical partition function $\Xi_{\beta,\mu}$ for an ideal massive Bose gas in a bounded domain F is given by*

$$\log \Xi_{\beta,\mu} = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \beta^{-t} \Gamma(t) \zeta(t+1) \zeta_{\underline{\Delta}}(t, -\mu)\, dt, \quad \sigma > \frac{d_s}{2}. \tag{2.18}$$

Proof. Using the Taylor expansion for $\log(1-x)$ and functional calculus, we expand the RHS of (2.17) in a power series:

$$\log \Xi_{\beta,\mu} = \mathrm{Tr}_{\mathcal{H}_1} \sum_{n=1}^{\infty} \frac{1}{n} e^{-n\beta(-\underline{\Delta}-\mu)}.$$

By the identity

$$e^{-a} = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} a^{-t} \Gamma(t) dt, \quad \sigma > 0,$$

we have

$$\log \Xi_{\beta,\mu} = \mathrm{Tr}_{\mathcal{H}_1} \frac{1}{2\pi i} \sum_{n=1}^{\infty} \frac{1}{n} \int_{\sigma-i\infty}^{\sigma+i\infty} (n\beta)^{-t} \Gamma(t) (-\underline{\Delta} - \mu)^{-t} dt.$$

It remains to interchange the order of integration and summations. To do so we need to pick σ such that both $\sum_{n=1}^{\infty} n^{-(\sigma+ic+1)}$ and $\mathrm{Tr}_{\mathcal{H}_1}(-\underline{\Delta} - \mu)^{-(\sigma+ic)}$ converge for any $c \in \mathbb{R}$. This happens when $\sigma > 0 \vee \frac{d_s}{2} = \frac{d_s}{2}$. ■

2.5.1 Bose-Einstein condensation in the unbounded carpet

For the massive Bose gas we are interested in the density of bosons, that is, the number of bosons per unit spectral volume:

$$\rho_L(\beta,\mu) := \frac{\omega_{L,\beta,\mu}(\mathbf{N})}{V_s(L)} = \frac{1}{V_s(L)} \mathrm{Tr}_{\mathcal{H}_1} \frac{1}{e^{\beta(-\underline{\Delta}_L-\mu)} - 1} = \frac{1}{\beta V_s(L)} \frac{\partial}{\partial \mu} \log \Xi_{L,\beta,\mu}. \tag{2.19}$$

Lemma 2.5.2. *Let F be a GSC. Then the density of ideal massive Bose gas in LF at (β,μ) is*

$$\rho_L(\beta,\mu) = \frac{1}{(4\pi)^{\frac{d_s}{2}} \hat{G}_{0,0}} \left(\sum_{k=0}^{d} \sum_{p\in\mathbb{Z}} \right)' \frac{L^{d_{k,p}-d_s}}{\beta^{d_{k,p}/2}} \hat{G}_{k,p} \sum_{m=0}^{\infty} \frac{(\beta\mu)^m}{m!} \zeta\left(-m + \frac{d_{k,p}}{2}\right), \tag{2.20}$$

where $\left(\sum_{k=0}^{d}\sum_{p\in\mathbb{Z}}\right)'$ *means the double sum excluding the term* $(k,p)=(d,0)$.

Proof. We evaluate the contour integral in (2.18) via a residue calculation. By Corollary 2.3.2, $\zeta_\Delta(\cdot,\gamma)$ has poles $\bigcup_{n\in\mathbb{N}_0}(d_{k,p}/2-n)$ with residues $\frac{(-1)^n\gamma^n\hat{G}_{k,p}}{n!\Gamma(-n+d_{k,p}/2)}$. Upon excluding the poles with zero residue, $d_{d,0}/2-\mathbb{N}_0$, one finds

$$\log\Xi_{L,\beta,\mu}=\left(\sum_{k=0}^{d}\sum_{p\in\mathbb{Z}}\right)'\left(\frac{L^2}{\beta}\right)^{d_{k,p}/2}\sum_{m=0}^{\infty}\frac{(\beta\mu)^m\hat{G}_{k,p}}{m!}\zeta\left(\frac{d_{k,p}}{2}-m+1\right).$$

The lemma follows from taking the μ-derivative of $\log\Xi_{L,\beta,\mu}$ and then dividing by the spectral volume $V_s(L)=(4\pi)^{\frac{d_s}{2}}\hat{G}_{0,0}L^{d_s}$. ∎

We are in a position now to take the thermodynamic limit of the particle density along $\Lambda_n\nearrow F_\infty$. To streamline the arguments, we introduce the *fugacity* $z=e^{\beta\mu}$, and state all quantities in terms of z rather than μ. We write $\rho_{\Lambda_n}(\beta,z)$ for the particle density in Λ_n at (β,z).

Proposition 2.5.3. *Let* $\overline{\rho}_c(\beta)=\limsup_{n\to\infty}\rho_{\Lambda_n}(\beta,1)$ *and* $\underline{\rho}_c(\beta)=\liminf_{n\to\infty}\rho_{\Lambda_n}(\beta,1)$ *be, respectively, the upper and lower critical density of the massive Bose gas in* F_∞. *Then* $\underline{\rho}_c(\beta)=\overline{\rho}_c(\beta)=\infty$ *if* $d_s\le 2$, *while* $0<\underline{\rho}_c(\beta)\le\overline{\rho}_c(\beta)<\infty$ *if* $d_s>2$. *In the latter case,*

$$\overline{\rho}_c(\beta)=\frac{1}{(4\pi\beta)^{\frac{d_s}{2}}}\frac{\max G_0}{\hat{G}_{0,0}}\zeta\left(\frac{d_s}{2}\right),\quad \underline{\rho}_c(\beta)=\frac{1}{(4\pi\beta)^{\frac{d_s}{2}}}\frac{\min G_0}{\hat{G}_{0,0}}\zeta\left(\frac{d_s}{2}\right).\tag{2.21}$$

Proof. First we note that as $L\to\infty$, only the volume terms ($k=0$) dominate:

$$\rho_L(\beta,\mu)=\frac{1}{(4\pi\beta)^{\frac{d_s}{2}}\hat{G}_{0,0}}\sum_{p\in\mathbb{Z}}\left(\frac{L^2}{\beta}\right)^{\frac{2\pi ip}{\log R}}\hat{G}_{0,p}\sum_{m=0}^{\infty}\frac{(\beta\mu)^m}{m!}\zeta\left(-m+\frac{d_{0,p}}{2}\right)+o(1).$$

By expanding the Riemann zeta function in a series and using some manipulation, we arrive at a more transparent expression

$$\rho_L(\beta,z)=\frac{1}{(4\pi\beta)^{\frac{d_s}{2}}\hat{G}_{0,0}}\sum_{m=1}^{\infty}z^m G_0\left(-\log\left(\frac{m\beta}{L^2}\right)\right)m^{-\frac{d_s}{2}}+o(1).$$

Replacing L with ℓ_F^n we conclude that as $n \to \infty$,

$$\rho_{\Lambda_n}(\beta, z) = \frac{1}{(4\pi\beta)^{\frac{d_s}{2}} \hat{G}_{0,0}} \sum_{m=1}^{\infty} z^m G_0\left(-\log\left(\frac{m\beta}{(l_F)^{2n}}\right)\right) m^{-\frac{d_s}{2}} + o(1),$$

from which the proposition follows. Finiteness comes from the convergence of the Riemann zeta function $\zeta(d_s/2) = \sum_{m=1}^{\infty} m^{-d_s/2}$. Note that we do not claim that $\underline{\rho}_c(\beta) = \overline{\rho}_c(\beta)$ since G_0 is probably nonconstant. ■

The consequence of a finite upper critical density is that any excess Bose gas must occupy the lowest eigenfunction of the Hamiltonian, a phenomenon known as *Bose-Einstein condensation (BEC)*. This is the content of the next theorem. We will state the result for a fixed-density Bose gas, *i.e.,* fix the particle density ρ_{tot} along $\Lambda_n \nearrow F_\infty$. This of course requires adjusting the fugacity z_n for each n. Moreover, we denote by $E_k(L)$ and $E_k(\Lambda_n)$ the $(k+1)$th eigenvalue of the Laplacian on, respectively, LF and Λ_n.

Theorem 2.5.4. *Assume $d_s > 2$. For each $\rho_{\text{tot}} > 0$, let z_n be the unique root of $\rho_{\Lambda_n}(\beta, z_n) = \rho_{\text{tot}}$.*

(i) If $\rho_{\text{tot}} \le \overline{\rho}_c(\beta)$, and $\overline{z}$ is the root of $\limsup_{n\to\infty} \rho_{\Lambda_n}(\beta, \overline{z}) = \rho_{\text{tot}}$, then $\liminf_{n\to\infty} z_n = \overline{z}$.

*(ii) If $\rho_{\text{tot}} > \overline{\rho}_c(\beta)$, then $\lim_{n\to\infty} z_n = 1$. Moreover if we denote the boson occupation density in the ground state, or the **condensate density**, by*

$$\rho^0_{\Lambda_n}(\beta, z) := \frac{1}{V_s(\Lambda_n)} \frac{1}{z^{-1} e^{\beta E_0(\Lambda_n)} - 1},$$

then $\lim_{n\to\infty} \left[\rho^0_{\Lambda_n}(\beta, z_n) + \rho_{\Lambda_n}(\beta, 1)\right] = \rho_{\text{tot}}$.

The proof, which mirrors that of [13, Theorem 5.2.30] with some modifications, relies upon extensive use of the convexity of $z \mapsto \rho_L(\beta, z)$ and the lemma below.

Lemma 2.5.5. *For any $z_1 > z_2$,*

$$\frac{\rho_L(\beta, z_2)}{z_2} \le \frac{\rho_L(\beta, z_1) - \rho_L(\beta, z_2)}{z_1 - z_2} \le \frac{\rho_L(\beta, z_1)}{z_1(1 - z_1 e^{-\beta E_0(L)})}. \tag{2.22}$$

Analogously, if we denote the boson occupation density above the $(m+1)$*th eigenfunction by*

$$\rho_L^{m+}(\beta, z) := \frac{1}{V_s(L)} \sum_{k>m} \frac{1}{z_L^{-1} e^{\beta E_k(L)} - 1}, \tag{2.23}$$

then for any $z_1 > z_2$,

$$\frac{\rho_L^{m+}(\beta, z_2)}{z_2} \leq \frac{\rho_L^{m+}(\beta, z_1) - \rho_L^{m+}(\beta, z_2)}{z_1 - z_2} \leq \frac{\rho_L^{m+}(\beta, z_1)}{z_1(1 - z_1 e^{-\beta E_m(L)})}. \tag{2.24}$$

Proof. Using the convexity of $z \mapsto \rho_L(\beta, z)$, we have for any $z_1 > z_2$,

$$\frac{\partial \rho_L}{\partial z}(\beta, z_2) \leq \frac{\rho_L(\beta, z_1) - \rho_L(\beta, z_2)}{z_1 - z_2} \leq \frac{\partial \rho_L}{\partial z}(\beta, z_1), \tag{2.25}$$

where the derivative is easily computed:

$$\begin{aligned} \frac{\partial \rho_L}{\partial z}(\beta, z) &= \frac{1}{V_s(L)} \mathrm{Tr}_{\mathcal{H}_1} \frac{e^{\beta \underline{\Delta}_L}}{(1 - z e^{\beta \underline{\Delta}_L})^2} \\ &= \frac{1}{V_s(L)} \mathrm{Tr}_{\mathcal{H}_1} \left[\frac{1}{z^{-1} e^{-\beta \underline{\Delta}_L} - 1} \cdot \frac{z^{-1}}{1 - z e^{\beta \underline{\Delta}_L}} \right]. \end{aligned} \tag{2.26}$$

Since $-\underline{\Delta}_L \geq E_0(L)$ and $z \leq e^{-\beta \underline{\Delta}_L}$ by assumption, we can make the bound

$$z^{-1} \leq \frac{z^{-1}}{1 - z e^{\beta \underline{\Delta}_L}} \leq \frac{z^{-1}}{1 - z e^{-\beta E_0(L)}}. \tag{2.27}$$

Applying (2.26) and (2.27) to (2.25) yields (2.22). The same argument carries over to ρ_L^{m+} by changing $E_0(L)$ to $E_m(L)$ in the upper bound. ∎

Proof of Theorem 2.5.4. Since $G_0 > 0$ and $G_1 < 0$, we have for all sufficiently large n,

$$\begin{aligned} \rho_{\Lambda_n}(\beta, z) &\leq \frac{1}{(4\pi\beta)^{\frac{d_s}{2}} \hat{G}_{0,0}} \sum_{m=1}^{\infty} z^m G_0 \left(-\log \left(\frac{m\beta}{(l_F)^{2n}} \right) \right) m^{-\frac{d_s}{2}} \\ &\leq \frac{1}{(4\pi\beta)^{\frac{d_s}{2}} \hat{G}_{0,0}} \sum_{m=1}^{\infty} z^m m^{-\frac{d_s}{2}} \max(G_0) = \limsup_{n \to \infty} \rho_{\Lambda_n}(\beta, z). \end{aligned}$$

Now we prove (i). By the preceding inequality (2.28) and the convexity of $\rho_L(\beta, \cdot)$, we deduce that $z_n \geq \overline{z}$ for all sufficiently large n. Then together with the first inequality in (2.22) of Lemma 2.5.5 with $z_1 = z_n$ and $z_2 = \overline{z}$, we get

$$0 \leq z_n - \overline{z} \leq \overline{z}\left(\frac{\rho_{\Lambda_n}(\beta, z_n)}{\rho_{\Lambda_n}(\beta, \overline{z})} - 1\right) = \overline{z}\left(\frac{\rho_{\text{tot}}}{\rho_{\Lambda_n}(\beta, \overline{z})} - 1\right).$$

Taking the lim inf of this inequality yields the result.

To prove (ii), suppose $\rho_{\text{tot}} > \overline{\rho}_c(\beta)$. If $z_n \leq 1$ for some large n, then $\overline{\rho}_c(\beta) < \rho_{\Lambda_n}(\beta, z_n) \leq \limsup_{n\to\infty} \rho_{\Lambda_n}(\beta, z_n) \leq \limsup_{n\to\infty} \rho_{\Lambda_n}(\beta, 1) =: \overline{\rho}_c(\beta)$, which is a contradiction. Thus $z_n > 1$. But since $z_n \leq e^{\beta E_0(\Lambda_n)}$ we have $\lim_{n\to\infty} z_n = 1$. Moreover $\rho^{m+}_{\Lambda_n}(\beta, z_L) > \rho^{m+}_{\Lambda_n}(\beta, 1)$ for any $m \in \mathbb{N}_0$, so in particular

$$\rho^{0+}_{\Lambda_n}(\beta, z_n) > \rho^{0+}_{\Lambda_n}(\beta, 1) \geq \rho_{\Lambda_n}(\beta, 1) - \frac{1}{V_s(\Lambda_n)\beta E_0(\Lambda_n)},$$

and

$$\begin{aligned} \rho^0_{\Lambda_n}(\beta, z_n) + \rho_{\Lambda_n}(\beta, 1) &\leq \rho^0_{\Lambda_n}(\beta, z_n) + \rho^{0+}_{\Lambda_n}(\beta, z_n) + \frac{1}{V_s(\Lambda_n)\beta E_0(\Lambda_n)} \\ &= \rho_{\text{tot}} + \frac{1}{V_s(\Lambda_n)\beta E_0(\Lambda_n)}. \end{aligned}$$

Noting that $[V_s(\Lambda_n)E_0(\Lambda_n)]^{-1} = Cn^{2-d_s} \to 0$ as $n \to \infty$, we find

$$\liminf_{n\to\infty} \left[\rho^0_{\Lambda_n}(\beta, z_n) + \rho_{\Lambda_n}(\beta, 1)\right] \leq \rho_{\text{tot}}.$$

It remains to show that $\limsup_{n\to\infty} \left[\rho^0_{\Lambda_n}(\beta, z_n) + \rho_{\Lambda_n}(\beta, 1)\right] \geq \rho_{\text{tot}}$. Here we utilize the fact that for any $m \in \mathbb{N}$ and $n \in \mathbb{N}_0$, the occupation density in the $(m+1)$th eigenfunction of the Laplacian on Λ_n

$$\begin{aligned} \frac{1}{V_s(\Lambda_n)} \frac{1}{z_n^{-1} e^{\beta E_m(\Lambda_n)} - 1} &\leq \frac{1}{V_s(\Lambda_n)} \frac{1}{e^{\beta(E_m(\Lambda_n) - E_0(\Lambda_n))} - 1} \\ &\leq \frac{1}{V_s(\Lambda_n)} \frac{1}{\beta(E_m(\Lambda_n) - E_0(\Lambda_n))}. \end{aligned}$$

The right-hand side scales with n^{2-d_s} and hence tends to 0 as $n \to \infty$, independently of m. This implies that

$$\rho_{\text{tot}} = \lim_{n\to\infty} \left[\rho^0_{\Lambda_n}(\beta, z_n) + \rho^{0+}_{\Lambda_n}(\beta, z_n)\right] = \lim_{n\to\infty} \left[\rho^0_{\Lambda_n}(\beta, z_n) + \rho^{m+}_{\Lambda_n}(\beta, z_n)\right]$$

for any $m \in \mathbb{N}$. By replacing L with Λ_n in the second inequality in (2.24) of Lemma 2.5.5, and setting $z_1 = z_n$ and $z_2 = 1$, we obtain

$$\rho^{m+}_{\Lambda_n}(\beta, z_n) \leq \rho_{\Lambda_n}(\beta, 1) \frac{z_n^{-1} - e^{-\beta E_m(\Lambda_n)}}{z_n^{-2} - e^{-\beta E_m(\Lambda_n)}}.$$

It follows that for all n and for all sufficiently large m,

$$\frac{\rho^{m+}_{\Lambda_n}(\beta, z_n)}{\rho_{\Lambda_n}(\beta, 1)} \leq \frac{z_n^{-1} - e^{-\beta E_m(\Lambda_n)}}{z_n^{-2} - e^{-\beta E_m(\Lambda_n)}} \leq \left(1 - 2\frac{E_0(\Lambda_n)}{E_m(\Lambda_n)}\right)^{-1} \leq \left(1 - 2Cm^{1-\frac{d_s}{2}}\right)^{-1},$$

where the Weyl asymptotics of the Laplacian was used in the last line. Thus

$$\rho_{\text{tot}} \leq \limsup_{n\to\infty} \left[\rho^0_{\Lambda_n}(\beta, z_n) + \rho_{\Lambda_n}(\beta, 1)\left(1 - 2Cm^{1-\frac{d_s}{2}}\right)^{-1}\right].$$

As the left-hand side is independent of m, we can take the limit $m \to \infty$ to complete the proof. ■

Remark 2.5.6. Here we encounter a subtlety not seen in the Euclidean setting, namely, that the length scale enters into the log-periodic modulation G_0 in the particle density. Therefore the thermodynamic limit needs to be handled with care. Theorem 2.5.4 says that in the lim sup density sense, condensation occurs if ρ_{tot} exceeds the upper critical density $\overline{\rho}_c(\beta)$. This appears to be the optimal threshold as our next result (Theorem 2.5.7) suggests. Notice also that neither the condensate density $\rho^0_{\Lambda_n}(\beta, z_n)$ nor the critical density $\rho_{\Lambda_n}(\beta, 1)$ likely has a limit as $n \to \infty$, but the sum of the two converges to ρ_{tot}.

One might ask whether a stronger sense of convergence could be used to characterize the thermodynamic limit on fractals. A possibility would be to use the Cesàro averaged density in place of the lim sup density. However incorporating the former into the convexity proof of Theorem 2.5.4 appears difficult.

We have now shown that for an ideal massive Bose gas in an unbounded GSC with $d_s > 2$, there is macroscopic occupation of particles in the ground state at sufficiently low temperatures. It is well known that Bose-Einstein condensation is a **quantum phase transition**. To illustrate this point, we will show that in the thermodynamic limit $L \to \infty$, the *free energy density* of the Bose gas

$$f_L(\beta, z) := \frac{F_L(\beta, z)}{V_s(L)} = -\frac{1}{\beta V_s(L)} \log \Xi_{L,\beta,z}$$

becomes non-analytic in an appropriate sense. This is a familiar criterion of phase transition in the physics literature. A more technical criterion, namely, the non-uniqueness of Gibbs (or KMS) states at low temperatures, will be discussed elsewhere.

A straightforward computation shows that as $L \to \infty$,

$$f_L(\beta, z) = -\frac{1}{(4\pi)^{\frac{d_s}{2}} \beta^{\frac{d_s}{2}+1} \hat{G}_{0,0}} \sum_{m=1}^{\infty} z^m G_0\left(-\log\left(\frac{m\beta}{L^2}\right)\right) m^{-\left(\frac{d_s}{2}+1\right)} + o(1).$$

Taking the sequence of free energy densities along $\{\Lambda_n\}_n$ yields the following result.

Theorem 2.5.7. *For each $\rho_{\text{tot}} > 0$, let z_n be the unique root of $\rho_{\Lambda_n}(\beta, z_n) = \rho_{\text{tot}}$.*

(i) If $\rho_{\text{tot}} \le \overline{\rho}_c(\beta)$, and $\overline{z}$ is the root of $\limsup_{n\to\infty} \rho_{\Lambda_n}(\beta, \overline{z}) = \rho_{\text{tot}}$, then

$$\liminf_{n\to\infty} f_{\Lambda_n}(\beta, z_n) = \liminf_{n\to\infty} f_{\Lambda_n}(\beta, \overline{z}) = -\frac{1}{(4\pi)^{\frac{d_s}{2}} \beta^{\frac{d_s}{2}+1}} \frac{\max G_0}{\hat{G}_{0,0}} \text{Li}_{\frac{d_s}{2}+1}(\overline{z}), \tag{2.28}$$

where $\text{Li}_s(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^s}$ is the polylogarithm.

(ii) If $\rho_{\text{tot}} > \overline{\rho}_c(\beta)$, then $\lim_{n\to\infty} [f_{\Lambda_n}(\beta, z_n) - f_{\Lambda_n}(\beta, 1)] = 0$.

Proof. The key is to realize that $-f_L(\beta, z) \ge 0$, and that $z \mapsto -f_L(\beta, z)$ is increasing and convex. From this it follows that for any $z_1 > z_2$,

$$\frac{f_L(\beta, z_2) - f_L(\beta, z_1)}{z_1 - z_2} \le -\frac{\partial f_L}{\partial z}(\beta, z_1) = \frac{1}{\beta z_1} \rho_L(\beta, z_1). \tag{2.29}$$

For (i) we replace L, z_1, and z_2 in (2.29) with Λ_n, z_n, and $\overline{z}$ respectively to find

$$0 \leq f_{\Lambda_n}(\beta, \overline{z}) - f_{\Lambda_n}(\beta, z_n) \leq \frac{\rho_{\text{tot}}}{\beta}\left(1 - \frac{\overline{z}}{z_n}\right).$$

Take the lim inf on all sides and use Part (i) of Theorem 2.5.4 to finish.

Similarly for (ii) we replace L, z_1, and z_2 in (2.29) with Λ_n, z_n, and 1 respectively. By virtue of Part (ii) of Theorem 2.5.4, we can take the limit this time to deduce the result.

■

Part (ii) of the preceding theorem says that in the condensation regime the thermodynamic free energy density is, in a proper sense, independent of the particle density ρ_{tot}, or of the (nonzero) condensate density. This is because in the thermodynamic limit, the condensate resides in the zero-energy eigenfunction and does not contribute to the free energy. For example, in the sense of Cesàro averaging, we have

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f_{\Lambda_k}(\beta, z_k) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f_{\Lambda_k}(\beta, 1) = -\frac{1}{(4\pi)^{\frac{d_s}{2}}\beta^{\frac{d_s}{2}+1}}\zeta\left(\frac{d_s}{2}+1\right).$$

On the other hand, in the liminf sense

$$\liminf_{n\to\infty} f_{\Lambda_n}(\beta, z_n) = \liminf_{n\to\infty} f_{\Lambda_n}(\beta, 1) = -\frac{1}{(4\pi)^{\frac{d_s}{2}}\beta^{\frac{d_s}{2}+1}}\frac{\max G_0}{\hat{G}_{0,0}}\text{Li}_{\frac{d_s}{2}+1}(1).$$

Combined with the result in Part (i), one may check that $\liminf_{n\to\infty} f_{\Lambda_n}(\beta, z_n)$, regarded as a function of $\rho_{\text{tot}} \in (0, \infty)$, is non-analytic at $\overline{\rho}_c(\beta)$, since the left derivative (a nonzero constant times $\zeta\left(\frac{d_s}{2}\right)$) does not equal the right derivative (0).

2.5.2 Connection to Brownian motion

Theorem 2.5.8. *Let F be a GSC. Then the following are equivalent in the unbounded carpet F_∞:*

1. *The spectral dimension $d_s > 2$.*
2. *Brownian motion is transient.*
3. *Bose-Einstein condensation occurs for a low-temperature, high-density ideal Bose gas.*

Proof. For F_∞, the equivalence between (1) and (2) is given in [6, Theorem 8.1]. The equivalence between (1) and (3) is implied by Lemma 2.5.2 and Theorem 2.5.4. ■

On general graphs, the connection between spectral dimension and BEC has been mentioned in previous literature [14, 15]. The equivalence between transient Brownian motion and BEC, on the other hand, appears to be a more general and robust criterion. To the author's knowledge, outside of $\mathbb{R}^d$ and $\mathbb{Z}^d$, an explicit statement about the latter has only been made by [29, 30, 52] in the context of inhomogeneous graphs. In these works, the authors considered the adjacency operator, as opposed to the Laplacian, on various exotic graphs (*e.g.* comb graphs, star graphs) and density-zero perturbations of Cayley trees. (We will discuss their model briefly in Section 2.8.1.) They gave conditions for BEC by analyzing the spectral properties of the adjacency operator, in particular its Perron-Frobenius eigenvector, and showed that the existence of "hidden spectrum" implies the finiteness of the critical density. These developments lead us to believe that the deep connection between Brownian motion and BEC holds on very general spaces, including graphs, manifolds, and fractals. A careful study of KMS states on these spaces may uncover more details.

We end this section by giving specific examples where one may probe BEC in non-integer dimensions. Recall that if F is a GSC embedded in $\mathbb{R}^d$, then $1 \leq d_s(F) < d_h(F) < d$. Also, an unbounded GSC whose cross-section contains a full copy of $\mathbb{R}^2_+$ supports transient Brownian motion, *i.e.,* the carpet has $d_s > 2$ [6]. So if one's goal is to find a

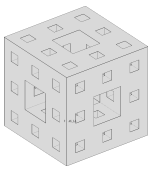
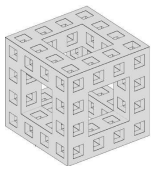
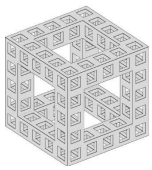
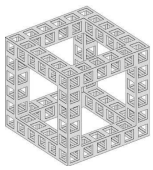
				
	$MS(3,1)$	$MS(4,2)$	$MS(5,3)$	$MS(6,4)$
d_h	$\log_3 20 \approx 2.73$	$\log_4 32 = 2.5$	$\log_5 44 \approx 2.35$	$\log_6 56 \approx 2.25$
Rigorous bounds on d_s [6]	$2.21 \sim 2.60$	$2.00 \sim 2.26$	$1.89 \sim 2.07$	$1.82 \sim 1.95$
Numerical d_s [CS]	2.51...	-	2.01...	-
BEC exists?	Yes	Yes	Yes (?)	No

Table 2.1: A sequence of Menger sponges whose spectral dimensions cross over 2. The criterion for existence of BEC in an ideal Bose gas is determined by Theorem 2.5.8.

GSC whose spectral dimension is close to 2, the most natural candidates are the Menger sponges in $\mathbb{R}^3$. For example, consider the family of Menger sponges $\{MS(a, a-2)\}_{a\geq 3}$, the first four members of which are shown in Table 2.1. All such sponges have $d_h > 2$, but only $MS(3,1)$, $MS(4,2)$, and possibly $MS(5,3)$ have $d_s > 2$, in which condensation of ideal Bose gas is possible. In particular $MS(5,3)$ lies just a hair above the critical dimension 2, as our numerics show. It would be very interesting if one could find a one-parameter family of fractal sponges whose spectral dimensions vary across 2, and study the onset of BEC in such spaces.

2.6 Blackbody radiation in Sierpinski carpets

In the next two sections we discuss the Bose gas with $H = \sqrt{-\Delta}$ and $\mu = 0$. This Hamiltonian corresponds to that of a gas of massless, non-interacting, relativistic particles, the prime example being photons. Here we have set $\mu = 0$ because the particle number cannot be kept finite in realistic experiments. The grand canonical partition function Ξ_β

in this case reads

$$\log \Xi_\beta = -\mathrm{Tr}_{\mathcal{H}_1} \log(1 - e^{-\beta \underline{D}}), \tag{2.30}$$

where $\underline{D} = \sqrt{-\underline{\Delta}}$ is the dimensionful Dirac operator, carrying units of inverse length.

Proposition 2.6.1. *The grand canonical partition function Ξ_β for an ideal photon gas in a bounded domain F is given by*

$$\log \Xi_\beta = \frac{1}{\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \beta^{-2t} \Gamma(2t) \zeta(2t+1) \zeta_{\underline{\Delta}}(t, 0) dt, \quad \sigma > \frac{d_s}{2}. \tag{2.31}$$

The proof is essentially identical to that of Proposition 2.5.1, so we omit it.

Lemma 2.6.2. *Let F be a GSC. Then the grand canonical partition function Ξ_β for an ideal photon gas confined in LF is given by*

$$\log \Xi_{L,\beta} = 2 \left(\sum_{k=0}^{d} \sum_{p \in \mathbb{Z}} \right)' \left(\frac{L}{\beta} \right)^{d_{k,p}} \frac{\Gamma\left(d_{k,p}\right)}{\Gamma\left(d_{k,p}/2\right)} \zeta\left(d_{k,p} + 1\right) \hat{G}_{k,p} + \frac{\beta}{2} \zeta_{\underline{\Delta}_L}\left(-\frac{1}{2}, 0\right), \tag{2.32}$$

where $\left(\sum_{k=0}^{d} \sum_{p \in \mathbb{Z}}\right)'$ means the double sum excluding the term $(k, p) = (d, 0)$.

Proof. We evaluate (2.31) via a residue calculation. The first term of (2.32) comes from the residues at the poles of $\zeta_{\underline{\Delta}}(s, 0)$, though we exclude $d_{d,0} = 0$ because it has zero residue (as well as being a simple pole of $\Gamma(2t)$ and of $\zeta(2t+1)$ simultaneously). Then there are contributions coming from the nonzero poles of $\Gamma(2t)$, which are $-\mathbb{N}/2$. But $\zeta_{\underline{\Delta}}(t, 0) = 0$ when $t \in -\mathbb{N}$ by Corollary 2.3.2, while $\zeta(2t+1) = 0$ when $t \in -\left(\mathbb{N} + \frac{1}{2}\right)$. Only the residue at $t = -1/2$ contributes and produces the second term of (2.32). ∎

The first term of (2.32), whose summands are all proportional to some nonnegative power of L/β, represent thermal (positive-temperature) contributions from the various codimension-k spectral volumes. On the other hand, the second term, which is linearly proportional to β/L, has no thermal origins. To tackle this issue, let us add to (2.32)

$$\log \Xi_{\mathrm{Cas}} := -\frac{\beta}{2} \zeta_{\underline{\Delta}}\left(-\frac{1}{2}, 0\right), \tag{2.33}$$

so we retain only the thermal contributions. Notice that this effectively adds a (free) energy

$$E_{\mathrm{Cas}} = -\beta^{-1} \log \Xi_{\mathrm{Cas}} = \frac{1}{2}\zeta_{\underline{\Delta}}\left(-\frac{1}{2}, 0\right). \tag{2.34}$$

to the existing vacuum energy H_0, which was 0 by default. (2.34) is the renowned *Casimir energy* representing the energy of vacuum fluctuations. It is half the regularized sum of the eigenvalues of the Dirac operator, where regularization should be understood as the analytic continuation of $\zeta_{\underline{\Delta}}(s, 0)$ to $s = -1/2$. By Theorem 2.3.1, such continuation is always possible on GSCs. The Casimir energy has measurable consequences and will be the focus of the next section.

The most important phenomenon associated with thermal photon gas is *blackbody radiation*. Here we are interested in the energy per spectral volume

$$\mathcal{E}(L, \beta) := \frac{\omega_{L,\beta,0}(\mathbf{H})}{V_s(L)} = -\frac{1}{V_s(L)} \frac{\partial}{\partial \beta} \log \Xi_{L,\beta,0},$$

as well as the isotropic pressure

$$P(L, \beta) := \frac{\partial}{\partial V_s(L)} F_{L,\beta,0} = -\frac{1}{\beta} \frac{\partial}{\partial V_s(L)} \log \Xi_{L,\beta,0}.$$

Proposition 2.6.3. *Let F be a GSC. As $L \to \infty$, the mean energy density and isotropic pressure of a thermal photon gas inside LF are*

$$\mathcal{E}(L, \beta) = \beta^{-(d_s+1)} H_1\left(-\log\left(\frac{\beta}{2L}\right)\right) + o(1), \tag{2.35}$$

$$P(L, \beta) = \frac{\mathcal{E}(L, \beta)}{d_s}, \tag{2.36}$$

where $H_1(\cdot)$ is $\frac{1}{2}\log R$-periodic with

$$\begin{aligned} H_1(x) = {} & \frac{d_s}{\pi^{\frac{d_s+1}{2}}} \Gamma\left(\frac{d_s+1}{2}\right) \zeta(d_s+1) \\ & + \frac{1}{\pi^{\frac{d_s+1}{2}} \hat{G}_{0,0}} \sum_{p \in \mathbb{Z}\setminus\{0\}} \left[e^{\frac{4\pi p i x}{\log R}} \Gamma\left(\frac{d_{0,p}+1}{2}\right) \zeta(d_{0,p}+1) d_{0,p} \hat{G}_{0,p} \right]. \end{aligned} \tag{2.37}$$

Proof. Straightforward differentiations of the thermal terms in (2.32), combined with the gamma function identity

$$\frac{\Gamma(z)}{\Gamma(z/2)} = \frac{\Gamma((z+1)/2)}{2^{1-z}\sqrt{\pi}},$$

yield the results. We note that for all $s \in \mathbb{R}$, $|\Gamma(s+it)| \sim \sqrt{2\pi}|t|^{s-1/2}e^{-\pi|t|/2}$ as $t \to \pm\infty$ by Stirling's formula, while for $s > 1$ and $t \in \mathbb{R}$ we have $|\zeta(s+it)| \leq \zeta(s)$. These observations, plus the boundedness of the $|\hat{G}_{0,p}|$, are sufficient to ensure the summability of (2.37). ■

We have intentionally split H_1 in order to make comparisons with the Euclidean case. If only the first term in (2.37) was considered, then (2.35) would be a direct extension of the Stefan-Boltzmann law from Euclidean to noninteger dimensions. But the discrete scale invariance of the GSC gives rise to log-periodic modulations depending on the temperature as well, *viz.* the second term in (2.37). Note also that the $\frac{1}{2}$ in the periodicity $\frac{1}{2}\log R$ of H_1 comes from the Dirac operator being the square root of the Laplacian.

Meanwhile, (2.36) represents the fractal version of the *equation of state* for the photon gas [1]. This is to be compared with the Euclidean version $P = \mathcal{E}/d$, where d is the Euclidean dimension (also the spectral dimension, since $d_w = 2$).

2.7 Casimir effect in Sierpinski carpet waveguide

In this section we derive the radiation pressure on some boundary faces of a compact fractal domain, due to either vacuum (zero-temperature, $\beta = \infty$) or thermal (nonzero-temperature, $\beta < \infty$) effects.

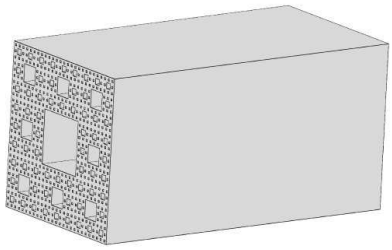


Figure 2.5: The Sierpinski carpet waveguide $a(GSC) \times [0, b]$.

2.7.1 The parallel-plate setup

The original calculation, due to H. Casimir in 1948, concerned the vacuum energy in $[0, a]^2 \times [0, b]$ for fixed $a \gg 1$, and the resultant pressure on the two parallel faces $[0, a]^2 \times \{0\}$ and $[0, a]^2 \times \{b\}$ as a function of b. The conventional method is to carry out a sum over the (exactly known) eigenvalues of the Dirac operator, take the $a \to \infty$ limit, then apply "zeta-regularization" to the formally divergent sum to produce a finite value. Similar calculations have also been done on other Euclidean domains or manifolds, such as spheres and cylinders.

In the spirit of the classical parallel-plate setup, we will work with a Sierpinski carpet waveguide $\Omega_{a,b} := a(GSC) \times [0, b]$, where we take $GSC \subset \mathbb{R}^2$ to be some 2-dimensional Sierpinski carpet and fix $a \gg 1$. See Figure 2.5. The parallel plates in this case correspond to $a(GSC) \times \{0\}$ and $a(GSC) \times \{b\}$. For consistency, we will impose Dirichlet conditions on the outer boundaries $\partial([0, a]^2) \times [0, b]$, $a(GSC) \times \{0\}$ and $a(GSC) \times \{b\}$, and Neumann conditions on the inner surfaces. The Hausdorff and spectral dimensions of the waveguide are, respectively, $d_h(\Omega) = d_h(GSC) + 1 \in (2, 3)$ and $d_s(\Omega) = d_s(GSC) + 1 \in (2, d_h(\Omega))$.

2.7.2 Casimir pressure at zero temperature

Our main objective is to compute the spectral zeta function $\zeta_\Omega\left(-\frac{1}{2},0\right)$, or twice the Casimir energy. Because the Laplacian on Ω is the direct sum of the Laplacian on GSC and the Laplacian on $[0,b]$, the associated heat kernel trace factorizes:

$$K_{\Omega_{a,b}}(t) = K_{a(GSC)}(t)K_{[0,b]}(t) = K_{GSC}\left(a^{-2}t\right)K_{[0,1]}\left(b^{-2}t\right).$$

Recalling that the spectrum of the Dirichlet Laplacian on the unit interval is $\bigcup_{j=1}^{\infty}\{j^2\pi^2\}$, we find

$$\begin{aligned} K_{[0,1]}(t) &= \sum_{j=1}^{\infty} e^{-j^2\pi^2 t} = \frac{1}{2}\left[\sum_{j\in\mathbb{Z}} e^{-j^2\pi^2 t} - 1\right] \\ &= \frac{1}{2}\left[\frac{1}{\sqrt{\pi t}}\sum_{j\in\mathbb{Z}} e^{-j^2/t} - 1\right] = \frac{1}{\sqrt{4\pi t}} - \frac{1}{2} + \frac{1}{\sqrt{\pi t}}\sum_{j=1}^{\infty} e^{-j^2/t}, \end{aligned}$$

where the Poisson summation formula was employed in the third equality. Then using (2.11) with $d = 2$ for $K_{GSC}(t)$, we get

$$\begin{aligned} K_{\Omega_{a,b}}(t) &= \left[\sum_{k=0}^{2}\sum_{p\in\mathbb{Z}} \hat{G}_{k,p}\left(\frac{t}{a^2}\right)^{-d_{k,p}/2} + O\left(\exp\left(-C\left(\frac{t}{a^2}\right)^{-\frac{1}{d_w-1}}\right)\right)\right] \\ &\quad\times\left[\frac{b}{\sqrt{4\pi t}} - \frac{1}{2} + \frac{b}{\sqrt{\pi t}}\sum_{j=1}^{\infty} e^{-j^2b^2/t}\right]. \end{aligned} \tag{2.38}$$

The spectral zeta function for the Laplacian on $\Omega_{a,b}$ is the meromorphic extension of

$$\zeta_{\Omega_{a,b}}(s,\gamma) = \frac{1}{\Gamma(s)}\int_0^\infty t^s K_{\Omega_{a,b}}(t)e^{-\gamma t}\frac{dt}{t}, \quad \mathrm{Re}(s) > \frac{d_s(\Omega)}{2}. \tag{2.39}$$

Theorem 2.7.1. *As $a \to \infty$,*

$$\begin{aligned} &\Gamma\left(-\frac{1}{2}\right)\zeta_{\Omega_{a,b}}\left(-\frac{1}{2},0\right) \\ &= \frac{1}{\sqrt{\pi}}\sum_{k=0}^{2}\sum_{p\in\mathbb{Z}} \hat{G}_{k,p}\left(\frac{a}{b}\right)^{d_{k,p}} b^{-1}\zeta\left(2+d_{k,p}\right)\Gamma\left(1+\frac{d_{k,p}}{2}\right) + o(1). \end{aligned} \tag{2.40}$$

Proof. Though the proof is similar to that of Theorem 2.3.1, we will be more precise in the computation here. Let $f(t) = O\left(\exp\left(-Ct^{-\frac{1}{d_w-1}}\right)\right)$. By definition, there exist $T, M \in (0,\infty)$ such that $|f(t)| \leq M\exp\left(-Ct^{-\frac{1}{d_w-1}}\right)$ for all $t \in (0,T)$. It follows that $|f(a^{-2}t)| \leq M\exp\left(-Ca^{\frac{2}{d_w-1}}t^{-\frac{1}{d_w-1}}\right)$ for all $t \in (0,a^2T)$. So when computing the Mellin integral we split the domain of integration $(0,\infty)$ into two intervals, $(0,a^2T)$ and $[a^2T,\infty)$.

Expanding the product in (2.38) and then plugging it into (2.39), we find

$$\zeta_{\Omega_{a,b}}\left(-\frac{1}{2},\gamma\right)\Gamma\left(-\frac{1}{2}\right) = \int_0^{a^2T} I(t)e^{-\gamma t}\frac{dt}{t} + \int_{a^2T}^{\infty} I(t)e^{-\gamma t}\frac{dt}{t},$$

where $I(t)$ contains terms of the following types:

1. Polynomial terms of the form t^{-p}, $\mathrm{Re}(p) > 0$.
2. Product of a polynomial term and an exponential term: $t^{-p}\exp(-c_5t^{-\eta})$, $\mathrm{Re}(p), \eta > 0$.
3. Product of a polynomial term and two exponential terms: $t^{-p}\exp(-c_6t^{-\eta})\exp(-c_7t^{-1})$, $\mathrm{Re}(p), \eta > 0$.

For the polynomial terms we have

$$\int_0^{a^2T} t^{-p}e^{-\gamma t}\frac{dt}{t} = \gamma^p\boldsymbol{\gamma}(-p,\gamma a^2T), \qquad \int_{a^2T}^{\infty} t^{-p}e^{-\gamma t}\frac{dt}{t} = \gamma^p\Gamma(-p,\gamma a^2T),$$

where $\boldsymbol{\gamma}(s,x)$ and $\Gamma(s,x)$ are, respectively, the lower and upper incomplete gamma functions. Then we need to use the asymptotics of the incomplete gamma functions near the branch point $x = 0$: $x^{-s}\boldsymbol{\gamma}(s,x) \to 1/s$ and, for $\mathrm{Re}(s) < 0$, $x^{-s}\Gamma(s,x) \to -1/s$. These imply that

$$\lim_{\gamma\to 0}\left(\int_0^{a^2T} t^{-p}e^{-\gamma t}\frac{dt}{t} + \int_{a^2T}^{\infty} t^{-p}e^{-\gamma t}\frac{dt}{t}\right) = 0.$$

Next we tackle terms of type (2). By making a change of variables we find

$$J_1(a,b,\gamma) := \int_0^{a^2T} t^{-p}e^{-c_5t^{-\eta}}e^{-\gamma t}\frac{dt}{t} = \frac{1}{\eta c_5^{p/\eta}}\int_{c_5/(a^2T)^{\eta}}^{\infty} u^{p/\eta}e^{-u}e^{-\gamma(c_5/u)^{1/\eta}}\frac{du}{u}.$$

Notice that for $u \in (c_5/(a^2T)^\eta, \infty)$ we have the upper bound $\left|e^{-\gamma(c_5/u)^{1/\eta}}\right| \le e^{|\gamma|a^2T}$, and the remaining integrand is integrable:

$$\left|\int_{c_5/(a^2T)^\eta}^{\infty} u^{p/\eta} e^{-u} \frac{du}{u}\right| = \left|\Gamma\left(\frac{p}{\eta}, \frac{c_5}{(a^2T)^\eta}\right)\right| < \infty.$$

So by the dominated convergence theorem,

$$\lim_{\gamma\to 0} J_1(a,b,\gamma) = \frac{1}{\eta c_5^{p/\eta}} \int_{c_5/(a^2T)^\eta}^{\infty} u^{p/\eta} e^{-u} \frac{du}{u} = \frac{1}{\eta c_5^{p/\eta}} \Gamma\left(\frac{p}{\eta}, \frac{c_5}{(a^2T)^\eta}\right).$$

The same argument applies to

$$J_2(a,b,\gamma) := \int_{a^2T}^{\infty} t^{-p} e^{-c_5 t^{-\eta}} e^{-\gamma t} \frac{dt}{t},$$

yielding

$$\lim_{\gamma\to 0} J_2(a,b,\gamma) = \frac{1}{\eta c_5^{p/\eta}} \boldsymbol{\gamma}\left(\frac{p}{\eta}, \frac{c_5}{(a^2T)^\eta}\right).$$

Thus

$$\begin{aligned}
&\lim_{\gamma\to 0}\left(\int_0^{a^2T} t^{-p} e^{-c_5 t^{-\eta}} e^{-\gamma t} \frac{dt}{t} + \int_{a^2T}^{\infty} t^{-p} e^{-c_5 t^{-\eta}} e^{-\gamma t} \frac{dt}{t}\right) \\
= &\ \frac{1}{\eta c_5^{p/\eta}}\left[\Gamma\left(\frac{p}{\eta}, \frac{c_5}{(a^2T)^\eta}\right) + \boldsymbol{\gamma}\left(\frac{p}{\eta}, \frac{c_5}{(a^2T)^\eta}\right)\right] \\
= &\ \frac{1}{\eta c_5^{p/\eta}} \Gamma\left(\frac{p}{\eta}\right).
\end{aligned}$$

Finally, for terms of type (3) we have

$$\begin{aligned}
&\left|\int_0^{a^2T} t^{-p} e^{-c_6 t^{-\eta}} e^{-c_7 t^{-1}} e^{-\gamma t} \frac{dt}{t}\right| \\
= &\ \left|\frac{1}{c_7^p} \int_{c_7/(a^2T)}^{\infty} u^p e^{-u} e^{-c_6 (u/c_7)^\eta} e^{-\gamma(c_7/u)} \frac{du}{u}\right| \\
\le &\ \frac{e^{|\gamma| c_7 a^2 T} e^{-c_7/(a^2T)}}{c_7^{\mathrm{Re}(p)}} \int_{c_7/(a^2T)}^{\infty} u^{\mathrm{Re}(p)} e^{-c_6 (u/c_7)^\eta} \frac{du}{u} \\
= &\ \frac{e^{|\gamma| c_7 a^2 T} e^{-c_7/(a^2T)}}{\eta c_6^{\mathrm{Re}(p)/\eta}} \Gamma\left(\frac{\mathrm{Re}(p)}{\eta}, \frac{c_6}{(a^2T)^\eta}\right).
\end{aligned}$$

Thus

$$\lim_{\gamma\to 0}\left|\int_0^{a^2T} t^{-p} e^{-c_6 t^{-\eta}} e^{-c_7 t^{-1}} e^{-\gamma t} \frac{dt}{t}\right| \le \frac{e^{-c_7/(a^2T)}}{\eta c_6^{\mathrm{Re}(p)/\eta}} \Gamma\left(\frac{\mathrm{Re}(p)}{\eta}, \frac{c_6}{(a^2T)^\eta}\right).$$

A similar calculation yields

$$\lim_{\gamma\to 0}\left|\int_{a^2T}^{\infty} t^{-p}e^{-c_6t^{-\eta}}e^{-c_7t^{-1}}e^{-\gamma t}\frac{dt}{t}\right| \leq \frac{1}{\eta c_6^{\mathrm{Re}(p)/\eta}}\gamma\left(\frac{\mathrm{Re}(p)}{\eta}, \frac{c_6}{(a^2T)^\eta}\right).$$

After putting in the numbers we obtain

$$\begin{aligned}\zeta_{\Omega_{a,b}}\left(-\frac{1}{2},0\right)\Gamma\left(-\frac{1}{2}\right) &= \frac{1}{\sqrt{\pi}}\sum_{k=0}^{2}\sum_{p\in\mathbb{Z}}\hat{G}_{k,p}\frac{a^{d_{k,p}}}{b^{d_{k,p}+1}}\sum_{j=1}^{\infty}\frac{1}{j^{(d_{k,p}+2)}}\Gamma\left(\frac{d_{k,p}}{2}+1\right)\\ &\quad +I_4(a,b),\end{aligned}$$

where

$$|I_4(a,b)| \leq \frac{M}{2}\frac{d_w-1}{aC^{(d_w-1)/2}}\Gamma\left(\frac{d_w-1}{2}, \frac{C}{T^{1/(d_w-1)}}\right) + I_5(a,b)$$

and $\lim_{a\to\infty} I_5(a,b) = 0$. This concludes the proof. ■

The Casimir pressure is defined to be the force per unit spectral area on each parallel plate as the separation b of the two plates is varied, that is,

$$P_{\mathrm{Cas}}(a,b) := -\frac{1}{V_s(a(GSC))}\frac{\partial}{\partial b}E_{\mathrm{Cas}}(a,b).$$

The following results are corollaries of the theorem.

Proposition 2.7.2. *Denote*

$$C_p := \zeta\left(1+d_s(\Omega)+\frac{4\pi pi}{\log R}\right)\Gamma\left(\frac{1}{2}+\frac{d_s(\Omega)}{2}+\frac{2\pi pi}{\log R}\right)\hat{G}_{0,p}. \tag{2.41}$$

As $a\to\infty$, the Casimir energy in a SC waveguide $\Omega_{a,b}$ is

$$\begin{aligned}E_{\mathrm{Cas}}(a,b) &:= \frac{1}{2}\zeta_{\Omega_{a,b}}\left(-\frac{1}{2},0\right)\\ &= -\frac{1}{4\pi}\frac{a^{d_s(GSC)}}{b^{d_s(GSC)+1}}\sum_{p\in\mathbb{Z}}C_p e^{\frac{4\pi pi}{\log R}\log\left(\frac{a}{b}\right)} + O(a^{\frac{2}{d_w(GSC)}}).\end{aligned} \tag{2.42}$$

The zero-temperature Casimir pressure on each parallel plate $a(GSC)\times\{0\}$ or $a(GSC)\times\{b\}$ is

$$\begin{aligned}P_{\mathrm{Cas}}(a,b) &= -\frac{b^{-(d_s(\Omega)+1)}}{(4\pi)^{\frac{d_s(\Omega)+1}{2}}\hat{G}_{0,0}}\sum_{p\in\mathbb{Z}}C_p\left[d_s(\Omega)+\frac{4\pi pi}{\log R}\right]e^{\frac{4\pi pi}{\log R}\log\left(\frac{a}{b}\right)}\\ &\quad +O\left(a^{-2\frac{d_h(GSC)-1}{d_w(GSC)}}\right).\end{aligned} \tag{2.43}$$

The key take-away message is that the Casimir pressure scales with the separation of the plates b as

$$P_{\mathrm{Cas}}(a,b) = b^{-(d_s(\Omega)+1)} H_2\left(\log\left(\frac{a}{b}\right)\right) + o(1) \quad \text{as } a \to \infty, \tag{2.44}$$

where $H_2(\cdot)$ is $\frac{1}{2}\log R$-periodic. We do not know the sign of P_{Cas} as it requires more information about the function G_0 than currently available.

2.7.3 Casimir pressure at positive temperature

To calculate the Casimir pressure due to thermal effects, we first derive the partition function, then find the free energy and take appropriate derivatives. Since most of the arguments resemble those in the blackbody radiation calculation (Section 2.6), we will omit the computational details.

Lemma 2.7.3. *The simple poles of the spectral zeta function $\zeta_{\Omega_{a,b}}(\cdot, 0)$ are contained in the set*

$$\bigcup_{k=0}^{2} \bigcup_{p\in\mathbb{Z}} \left(\left\{ \frac{d_{k,p}+1}{2} \right\} \bigcup \left\{ \frac{d_{k,p}}{2} \right\} \right), \tag{2.45}$$

where the residues at the two sets of poles read, respectively,

$$\frac{a^{d_{k,p}} b}{\sqrt{4\pi}} \frac{\hat{G}_{k,p}}{\Gamma\left(\frac{d_{k,p}+1}{2}\right)}, \quad -\frac{1}{2} a^{d_{k,p}} \frac{\hat{G}_{k,p}}{\Gamma(d_{k,p}/2)}. \tag{2.46}$$

Proof. It is straightforward to check that only the polynomial terms in the expansion of $K_{\Omega_{a,b}}(t)$ (2.38) produce singularities. The rest of the proof is as outlined in Theorem 2.3.1 and Corollary 2.3.2. ■

Lemma 2.7.4. *The thermal terms in the grand canonical partition function* $\Xi_{\Omega_{a,b},\beta}$ *for the photon gas in the Sierpinski carpet waveguide* $\Omega_{a,b}$ *are given by*

$$\begin{aligned}\log \Xi_{\Omega_{a,b},\beta} &= \frac{1}{\sqrt{\pi}} \sum_{k=0}^{2} \sum_{p\in\mathbb{Z}} \frac{a^{d_{k,p}} b}{\beta^{d_{k,p}+1}} \frac{\Gamma(d_{k,p}+1)}{\Gamma\left(\frac{d_{k,p}+1}{2}\right)} \zeta(d_{k,p}+2)\hat{G}_{k,p} \\ &\quad - \left(\sum_{k=0}^{2} \sum_{p\in\mathbb{Z}}\right)' \frac{a^{d_{k,p}}}{\beta^{d_{k,p}}} \frac{\Gamma(d_{k,p})}{\Gamma(d_{k,p}/2)} \zeta(d_{k,p}+1)\hat{G}_{k,p}, \end{aligned} \tag{2.47}$$

where $\left(\sum_{k=0}^{d} \sum_{p\in\mathbb{Z}}\right)'$ *means the double sum excluding the term* $(k,p) = (d,0)$.

Proof. The arguments are identical to those in the proof of Lemma 2.6.2. ■

The Casimir pressure is then defined to be

$$P_{\mathrm{Cas},\beta}(a,b) := -\frac{1}{V_s(a(GSC))} \frac{\partial}{\partial b} F_{\Omega_{a,b},\beta} = \frac{1}{\beta V_s(a(GSC))} \frac{\partial}{\partial b} \log \Xi_{\Omega_{a,b},\beta}.$$

Proposition 2.7.5. *As* $a \to \infty$*, the Casimir pressure on each parallel plate* $a(GSC) \times \{0\}$ *or* $a(GSC) \times \{b\}$ *at inverse temperature* β *reads*

$$P_{\mathrm{Cas},\beta}(a,b) = \beta^{-(d_s(\Omega)+1)} H_3\left(-\log\left(\frac{\beta}{2a}\right)\right) + o(1), \tag{2.48}$$

where $H_3(\cdot)$ *is* $\frac{1}{2}\log R$*-periodic with*

$$H_3(x) = \frac{\hat{G}_{0,p}/\hat{G}_{0,0}}{\pi^{\frac{d_s(\Omega)+1}{2}}} \sum_{p\in\mathbb{Z}} e^{\frac{4\pi i p x}{\log R}} \Gamma\left(\frac{1}{2} + \frac{d_s(\Omega)}{2} + \frac{2\pi p i}{\log R}\right) \zeta\left(1 + d_s(\Omega) + \frac{4\pi p i}{\log R}\right). \tag{2.49}$$

Notice that the dominant term in the positive-temperature Casimir pressure is independent of the separation of plates b: like in blackbody radiation, the scaling is with respect to temperature.

This concludes our discussions of various results regarding ideal massive or massless Bose gas in Sierpinski carpets. We stress once again the power of the heat kernel and the zeta function technique in quantum statistical mechanics. Not only does it apply to very general spaces, it also give unambiguous answers to major thermodynamic questions, such as the emergence of the Casimir energy.

2.8 Towards interacting Bose gas on Sierpinski carpet graphs

The above discussion of the ideal massive Bose gas is just the prelude to a full understanding of quantum many-body physics on fractals. This is because some of the most salient features of BEC (in $\mathbb{R}^d$ or $\mathbb{Z}^d$), such as superfluidity and soliton excitations, cannot be explained by the ideal Bose gas. So it behooves us to take the next step and consider models of interacting Bose gas on Sierpinski carpets. Mathematically speaking, proving BEC in an interacting Bose system is a challenging problem (see [57] for a careful exposition). In this section we introduce a particular interacting system which has been studied extensively on $\mathbb{Z}^d$, then focus on the so-called hardcore Bose gas, which we hope to study in more depth on Sierpinski carpet graphs.

2.8.1 The Bose-Hubbard model

In what follows we have in mind a connected graph $G = (V, E)$ with bounded degree. To describe an ideal Bose gas on G, we use the second-quantized Hamiltonian in the grand canonical ensemble $d\Gamma(-\Delta - \mu\mathbf{1})$, where Δ is the combinatorial graph Laplacian on G. To express this Hamiltonian, let $a_x^\dagger$ and a_x be the boson creation and annihilation operators at $x \in V$, which satisfy the canonical commutation relations (CCR): $a_x a_y^\dagger - a_y^\dagger a_x = \delta_{xy}$ and $a_x a_y - a_y a_x = 0$ for all $x, y \in V$. We also let $n_x = a_x^\dagger a_x$, which has the interpretation as the particle number at x. Then one may verify that

$$d\Gamma(-\Delta - \mu\mathbf{1}) = -\frac{1}{2}\sum_{\langle xy\rangle \in E}(a_x^\dagger a_y + a_y^\dagger a_x) - \sum_{x\in V}(\mu - \deg(x))n_x.$$

A slight variant of this Hamlitonian uses the adjacency operator in place of the graph Laplacian, *i.e.,*

$$\mathbf{H} = -\frac{1}{2}\sum_{\langle xy\rangle \in E}(a_x^\dagger a_y + a_y^\dagger a_x) - \mu\sum_{x\in V} n_x. \tag{2.50}$$

In a realistic experimental setup, multiple atoms tend not to occupy the same site $x \in V$ due to forbidden overlap of atomic orbitals. This can be modelled by introducing an on-site repulsion term to (2.50). The resultant Hamiltonian is that of the Bose-Hubbard model,

$$\mathbf{H}_{BH} = -\frac{J}{2} \sum_{\langle xy \rangle \in E} (a_x^\dagger a_y + a_y^\dagger a_x) - \mu \sum_{x \in V} n_x + U \sum_{x \in V} n_x(n_x - 1), \tag{2.51}$$

with parameters $\mu \in \mathbb{R}$ and $J, U > 0$. The Hamiltonian (2.51) is widely used by condensed matter theorists to model cold atomic systems in optical lattices. In the works of BEC on inhomogeneous graphs [29, 30, 52], the authors investigated a special case of (2.51) with $\mu = 0$ and $U = 0$.

2.8.2 Hardcore Bose gas and the XY model

Let us focus on another special case of (2.51) where the on-site repulsion $U = \infty$. In this scenario n_x for any $x \in V$ can only take the value 0 or 1, that is, particles interact as if they were hard cores. The Hilbert space of this hardcore Bose gas is then $(\mathbb{C}^2)^{\otimes G}$, which is spanned by the tensor product of the basis vectors $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}_x, \begin{pmatrix} 0 \\ 1 \end{pmatrix}_x \right\}$, representing respectively the occupation or non-occupation of vertex x. Naturally, the creation and annihilation operators at x read

$$a_x^\dagger = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad a_z = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Notice that they do not satisfy the CCR anymore: $a_x a_y^\dagger - a_y^\dagger a_x = (1 - 2n_x)\delta_{xy}$.

We can express the Hamiltonian of the hardcore Bose gas as

$$\mathbf{H}_{hc} = -\frac{J}{2} \sum_{\langle xy \rangle \in E} (a_x^\dagger a_y + a_y^\dagger a_x) - \mu \sum_{x \in V} n_x. \tag{2.52}$$

Introduce the Pauli matrices

$$S^1 = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad S^2 = \frac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad S^3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then up to an additive constant (2.52) becomes

$$\mathbf{H}_{hc} = -J \sum_{\langle xy \rangle \in E} \left(S^1_x S^1_y + S^2_x S^2_y \right) - \mu \sum_{x \in V} S^3_x, \tag{2.53}$$

which is the Hamiltonian of the **quantum (ferromagnetic) XY model**, with μ playing the role of the external magnetic field. Therefore showing that hardcore Bose gas condenses on GSC graphs amounts to showing that the corresponding quantum XY model has a phase transition.

In finding a point of attack to this problem, we review results which are known to hold on arbitrary graphs. Given any infinite graph G, [16] showed that if the simple random walk on G is recurrent, then for any $J > 0$, the quantum XY model at zero field ($\mu = 0$) on G does not exhibit spontaneous magnetization. More precisely, if $\{G_n\}_n = \{(V_n, E_n)\}_n$ exhausts the recurrent graph G, and $\omega_{n,\beta,\mu}$ is the Gibbs state on G_n at (β, μ), then for all $\beta > 0$,

$$\lim_{\mu \to 0} \lim_{n \to \infty} \frac{1}{|V_n|} \omega_{n,\beta,\mu} \left(\sum_{x \in V_n} S^1_x \right) = 0.$$

This result generalizes the Mermin-Wagner theorem [62] from translationally invariant lattices to arbitrary graphs. Note that $\mu = 0$ corresponds to bosons at half-filling (equal number of occupied and unoccupied vertices on average).

Specializing to the situation on GSC graphs, we deduce that on any GSC graph with $d_s \leq 2$, the quantum XY model at zero field does not undergo a phase transition. Under this setting, the hardcore Bose gas at half-filling does not Bose condense. Conversely, if condensation occurs on some GSC graph, then the graph is transient.

The harder problem is to prove sufficient conditions for which *spontaneous magnetization*,

$$\lim_{\mu\to 0}\liminf_{n\to\infty}\frac{1}{|V_n|}\omega_{n,\beta,\mu}\left(\left|\sum_{x\in V_n}S^1_x\right|\right)>0,$$

or *long-range order*,

$$\liminf_{n\to\infty}\frac{1}{|V_n|^2}\omega_{n,\beta,0}\left(\sum_{\langle xy\rangle\in E_n}\left(S^1_xS^1_y+S^2_xS^2_y\right)\right)>0,$$

occurs for large enough β. Either condition would imply the non-uniqueness of KMS states at low temperatures [13, Section 6.2.6].

Problem 1. *Given any GSC graph G, prove or disprove that the transience of simple random walk on G is a sufficient condition for the quantum XY model on G to exhibit spontaneous magnetization or long-range order at positive temperatures. In other words, determine whether the hardcore Bose gas at half-filling condenses on any GSC graph with $d_s > 2$.*

Recall that on $\mathbb{Z}^d$ this has been answered in the positive by [26]. However the use of Fourier transform was essential in their proof, which made the generalization to arbitrary graphs difficult.

Though outside the scope of quantum mechanics, we also mention the abelian version of the XY model on general graphs, which is better understood. It has the same Hamiltonian (2.53), but with $\mathbf{S}_x = (S^1_x, S^2_x, S^3_x)$ taking values in the unit 2-sphere, *i.e.,* $\sum_{j=1}^3 (S^j_x)^2 = 1$. It is known that there is no spontaneous rotational symmetry breaking at any temperature for recurrent graphs [16, 61]. On the flip side, it appears that transience of simple random walk is a necessary, but not sufficient, condition for phase transition in the classical XY model. Y. Peres conjectured that the sufficient condition is that there exists a $p \in (0, 1)$ such that upon performing an independent bond percolation with probability p, the resultant infinite cluster supports transient simple random walk [67, Conjecture 1.9].

For GSC graphs, which possess strong symmetry and connectivity, we suspect that transience alone is a sufficient condition, though we are not aware of any previous work on this matter. The key may be to prove a version of spin-wave condensation in the spirit of [31].

Problem 2. *Given any GSC graph G, prove or disprove that the transience of simple random walk on G is a sufficient condition for the classical XY model on G to exhibit spontaneous magnetization or long-range order at positive temperatures.*

Acknowledgments

I am grateful to Naotaka Kajino, Benjamin Steinhurst, Robert Strichartz, and Alexander Teplyaev for many useful discussions leading up to the writing of this work. Thanks also to Gerald Dunne for proposing the Sierpinski carpet waveguide example discussed in Section 2.7; Michel Lapidus for suggesting the use of Cesàro averaging in thermodynamical quantities; Daniele Guido for explaining his work on Bose-Einstein condensation on inhomogeneous graphs; and Matthew Begué for providing the raw spectral data of the Kusuoka-Zhou Laplacian, from which Fig. 3(b) is generated. Portions of this work were presented at the "Waves and Quantum Fields on Fractals" workshop at Technion, and at the 6th Prague Summer School on Mathematical Statistical Physics. I wish to thank Eric Akkermans and Marek Biskup for organizing the respective conference and giving me the opportunity to speak.

During the preparation of the manuscript, the author was partially supported by Robert Strichartz's NSF grant (DMS 0652440), as well as a graduate assistantship from the 2011 Research Experience for Undergraduates (REU) program at the Department of Mathematics, Cornell University.

CHAPTER 3

DIRICHLET FORMS AND GREEN FORMS ON SIERPINSKI CARPETS

This chapter is a moderately amended version of Section 2 and the Appendix from the paper *"Entropic repulsion of Gaussian free field on high-dimensional Sierpinski carpet graphs,"* [20], co-authored with Baris Ugurcan.

3.1 Introduction

As was mentioned in §2.2.1 of the previous chapter, the construction of Brownian motion on the Sierpinski carpet has a long and rich history. Barlow and Bass constructed a diffusion on the "pre-carpet" $\bigcup_{N=0}^{\infty} \ell_F^N F_N$ as the subsequential limit of reflecting Brownian motions on $\ell_F^N F_N$ [3, 5, 6]. Kusuoka and Zhou provided a different construction based on Dirichlet forms on approximating graphs [51] (see §3.2 for a brief discussion). For nearly two decades it was unclear whether the two approaches gave rise to the same *unique* Brownian motion. A recent seminal work of Barlow, Bass, Kumagai and Teplyaev [8] showed that up to constant multiples, the family $\mathfrak{E}_c$ of local, regular, nonzero, conservative Dirichlet forms on $L^2(F, \nu)$, which are moreover invariant under the *local symmetries* of the carpet (in the sense of [8, Definition 2.15]), contains only one element. In particular, the Dirichlet forms associated with the Barlow-Bass construction and the Kusuoka-Zhou construction both belong to $\mathfrak{E}_c$, and therefore the two diffusions are the same up to deterministic time change.

This leads to the following question: does there exist a more natural approach to constructing diffusion on the Sierpinski carpet?

As Barlow explained in his recent survey [2], a more natural approach, at least in the eyes of the original authors, would be to show that the sequence approximating Dirichlet

forms converges (in subsequence) in the sense of Mosco. In particular, one would like to show that the limit form is a closable Markovian quadratic form, which can then be extended to a *local*, *regular* Dirichlet form. Then by the theory expounded in Fukushima *et al.* [21, 32], one may associate with this Dirichlet form a *bona fide* diffusion (*i.e.,* a Hunt process) on the Sierpinski carpet. The difficulty with this approach is that one could not easily check whether the extended form is indeed *local* and *regular*.

The purpose of this chapter is to demonstrate that the aforementioned *Mosco convergence* indeed holds subsequentially along the sequence of approximating Sierpinski carpet graphs, and that each limit form has almost all the advertised properties (except *locality*).

Let us fix some notations apart from those already appearing in §2.1.1. For each generalized Sierpinski carpet F, we consider two associated graphs; see Figure 3.1.

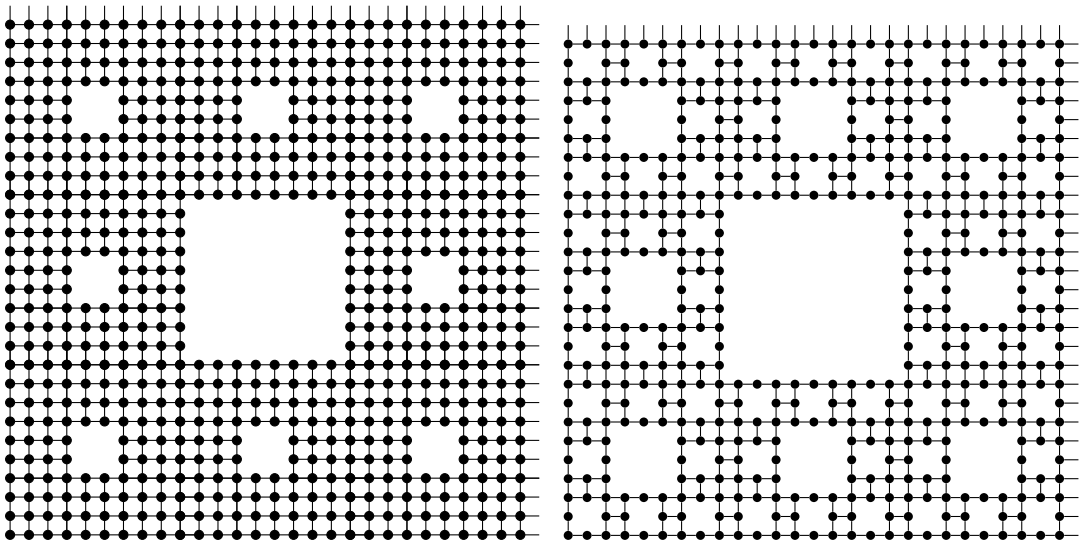
Let $V_N = \ell_F^N F_N \cap \mathbb{Z}^d$. Introduce the graph $\mathcal{G}_N = (V_N, \sim)$, where throughout the chapter, the edge relation "$\sim$" means that two vertices x, x' are connected by an edge if and only if their Euclidean distance $\|x - x'\| = 1$. Put $\mathcal{G}_\infty = \bigcup_{N\in\mathbb{N}} \mathcal{G}_N = (V_\infty, \sim)$, which we call the *outer* Sierpinski carpet graph. Observe that $\mathcal{G}_\infty$ is a subgraph of $(\mathbb{Z}_+)^d$.

Next, let $I_N = \ell_F^N F_N \cap \left(\mathbb{Z}^d + \left(\frac{1}{2}, \frac{1}{2}, \cdots, \frac{1}{2}\right)\right)$. Introduce the graph $\mathcal{I}_N = (I_N, \sim)$. Put $\mathcal{I}_\infty = \bigcup_{N\in\mathbb{N}} \mathcal{I}_N = (I_\infty, \sim)$, which we call the *inner* Sierpinski carpet graph. It is easy to see that $|I_N| = m_F^N$, and that there exist constants C and C', independent of N, such that $Cm_F^N \le |V_N| \le C'm_F^N$.

As is customary, we introduce the discrete Dirichlet form on graph $\mathcal{G}_\infty$ by

$$E_{\mathcal{G}_\infty}(f_1, f_2) = \frac{1}{2} \sum_{\substack{x,x'\in V_\infty \\ x\sim x'}} (f_1(x) - f_1(x'))(f_2(x) - f_2(x'))$$

for all f_1, f_2 in the natural domain $\mathcal{D}(E_{\mathcal{G}_\infty}) = \{f \in \ell^2(V_\infty) : E_{\mathcal{G}_\infty}(f, f) < \infty\}$. Similarly,



(a) (b)

Figure 3.1: The 3rd-level approximation of, respectively, the outer Sierpinski carpet graph $\mathcal{G}_\infty$ and the inner Sierpinski carpet graph $\mathcal{I}_\infty$, here shown for the standard 2-dimensional Sierpinski carpet. According to the conventions in the text, when embedded in $(\mathbb{R}_+)^d$, the least vertex of $\mathcal{G}_\infty$ is situated at the origin, while the least vertex of $\mathcal{I}_\infty$ is situated at $(\frac{1}{2}, \cdots, \frac{1}{2})$. All edges have Euclidean distance 1.

we define the discrete Dirichlet form on the graph $\mathcal{I}_\infty$ by

$$E_{\mathcal{I}_\infty}(f_1, f_2) = \frac{1}{2} \sum_{\substack{w,w' \in I_\infty \\ w \sim w'}} (f_1(w) - f_1(w'))(f_2(w) - f_2(w'))$$

for all f_1, f_2 in the natural domain $\mathcal{D}(E_{\mathcal{I}_\infty}) = \{f \in \ell^2(I_\infty) : E_{\mathcal{I}_\infty}(f, f) < \infty\}$. Furthermore, let $\mathcal{E}_N^{\mathcal{G}} = \rho_F^N E_{\mathcal{G}_\infty}$ and $\mathcal{E}_N^{\mathcal{I}} = \rho_F^N E_{\mathcal{I}_\infty}$ be the *renormalized* Dirichlet forms, where $\rho_F \in (0, \infty)$ is the unique *resistance scale factor* associated with the carpet F. It is known that simple random walk on either version of a Sierpinski carpet graph is transient if and only if $\rho_F < 1$ [7, 60].

The plan of this chapter is as follows:

- Prove that $(\mathcal{E}_N^{\mathcal{I}})_N$ Mosco converges in subsequence, and that each Mosco limit can be extended to a regular Dirichlet form on $L^2(F_\infty, \nu_\infty)$, which is then shown to be

comparable to some local regular Dirichlet form on $L^2(F_\infty, \nu_\infty)$ which is invariant under the local symmetries of the carpet (§3.3).

- Demonstrate the subsequential convergence of the so-called discrete **Green forms** (*i.e.* Dirichlet forms on smooth measures) using the Mosco convergence result (§3.3, §3.6.2).
- Compare the Dirichlet and Green forms $E_{\mathcal{I}_\infty}$ and $E_{\mathcal{G}_\infty}$ (and hence their renormalized versions) (§3.4).
- Prove the limsup convergence of discrete Green forms on $\mathcal{G}_\infty$ (§3.5).

The main results of this chapter are Theorem 3.3.1, Theorem 3.3.2, and Lemma 3.5.1. We relegate some general facts from the theory of Dirichlet forms and of Mosco convergence to §3.6.

Before marching into the details, we mention two implications of the results obtained in this chapter. First, it is a known fact [23, 63] that the Mosco convergence of Dirichlet forms is equivalent to the *strong resolvent convergence (in L^2)* of the associated self-adjoint operators (*i.e,* the discrete Laplacians $-\Delta_N$ on $\mathcal{I}_N$ in our setting). Let $-\Delta$ denote the limit operator, and let P_N (resp. P) denote the orthogonal projection from $L^2(F, \nu)$ to the domain of $-\Delta_N$ (resp. the domain of $-\Delta$). By [79, Theorem 2], we then have

$$E_N(\lambda)P_N \to E(\lambda)P \qquad \text{in subsequence for each } \lambda \notin \sigma(-\Delta),$$

where $E_N(\cdot)$ (resp. $E(\cdot)$) is the projection-valued spectral measure associated with $-\Delta_N$ (resp. $-\Delta$). In particular, this implies that for every $\lambda \in \sigma(-\Delta)$, there are $\lambda_N \in \sigma(-\Delta_N)$ such that $\lambda_N \to \lambda$ in subsequence. Hence we exhibit the precise mode in which the eigenvalues of the discrete Laplacians converge to the eigenvalue of the continuum Laplacian.

The second application pertains to the convergence of the discrete Green forms, which is otherwise known as *strong resolvent convergence in the energy norm*. This fact turns out to play an essential role in studying the Gaussian free field on Sierpinski carpet graphs, the details of which are reported in [20, §3-§4]. The reason for this connection is actually not surprising from the constructive QFT point of view, as described in E. Nelson's formulation [64, 65]. In an exposition by Röckner [68]), the massless free field on the measure space (X, m) is defined as a measure-indexed family of centered Gaussian random variables $\{\varphi_\mu\}_\mu$ with covariance $\mathrm{Cov}(\varphi_\mu, \varphi_\nu) = \mathcal{E}(U\mu, U\nu)$, where $\mathcal{E}$ is a local regular Dirichlet form on $L^2(X, m)$ which generates diffusion on X, and U is the 0-order potential operator associated with $\mathcal{E}$. Therefore, any quantitative information of the free field is encoded in the Dirichlet form on *smooth measures*, which has as its integral kernel the *Green's function*.

Notations. If (X, m) denotes a measure space, then $\langle f, \mu\rangle_X$ stands for $\int_X f d\mu$, pairing a function f on X with a Borel measure μ. As before, c, C and C' denote positive constants which may change from line to line. Specific constants will be denoted with a numeral subscript. Dependencies of the constant are indicated in parentheses. Finally, $C_c(X)$ denotes the space of continuous functions on X with compact support, and $\mathfrak{E}$ denotes the family of local, regular, nonzero, conservative Dirichlet forms on $L^2(F_\infty, \nu_\infty)$, which are moreover invariant under the *local symmetries* of the carpet in the sense of [8, Definition 2.15].

3.2 Kusuoka-Zhou construction of Dirichlet forms

Let F be a generalized Sierpinski carpet, and $\mathcal{I}_\infty = (I_\infty, \sim)$ be the inner Sierpinski carpet graph introduced in the previous subsection. For each $N \in \mathbb{N}$ and each $w \in I_\infty$, let $\Psi_w^{(N)}$

be the closed cube of side ℓ_F^{-N} centered at $\ell_F^{-N}w$. We define the *mean-value operator* $\tilde{P}_N : L^1(F_\infty, \nu_\infty) \to C(I_\infty; \mathbb{R})$ by

$$(\tilde{P}_N f)(w) = \frac{1}{\nu_\infty\left(\Psi_w^{(N)} \cap F_\infty\right)} \int_{\Psi_w^{(N)} \cap F_\infty} f(y)\nu_\infty(dy),$$

Similarly, if μ_∞ is a Radon measure on F_∞ such that $\mu_\infty \ll \nu_\infty$, then define $\tilde{P}_N\mu_\infty = \left(\tilde{P}_N \frac{d\mu_\infty}{d\nu_\infty}\right)\nu_N$, where $\nu_N = \frac{1}{m_F^N}\mathbb{1}_{I_\infty}$ is a self-similar measure on I_∞.

Let $\mathcal{F}_0 := \left\{ f \in L^2(F_\infty, \nu_\infty) : \sup_N \mathcal{E}_N^{\mathcal{I}}(\tilde{P}_N f, \tilde{P}_N f) < \infty \right\}$. The following convergence result for $(\mathcal{E}_N^{\mathcal{I}})_N$ is originally due to Kusuoka and Zhou [51, Proposition 5.2 & Theorem 5.4], and later generalized in [37, Lemma 4.1 & Theorem 4.3].

Proposition 3.2.1. *(i) There exists a constant $C_{3.1}$ such that for all $N, M \geq 1$ and all $f \in \mathcal{F}_0$,*

$$\mathcal{E}_N^{\mathcal{I}}(\tilde{P}_N f, \tilde{P}_N f) \leq C_{3.1}\mathcal{E}_{N+M}^{\mathcal{I}}(\tilde{P}_{N+M} f, \tilde{P}_{N+M} f).$$

(ii) There exists $(\mathcal{E}, \mathcal{F}_0) \in \mathfrak{E}$ and positive constants $C_{3.2}$ and $C_{3.3}$ such that for all $f \in \mathcal{F}_0$,

$$C_{3.2} \sup_N \mathcal{E}_N^{\mathcal{I}}(\tilde{P}_N f, \tilde{P}_N f) \leq \mathcal{E}(f, f) \leq C_{3.3} \varliminf_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}(\tilde{P}_N f, \tilde{P}_N f). \tag{3.1}$$

Remark 3.2.2. In their original work [51], Kusuoka and Zhou identified a family of Dirichlet forms, denoted $\mathcal{D}ch$, which are associated with cluster points of the sequence of suitably rescaled Markov processes on $\mathcal{I}_N$. Then they proved (3.1) for any $\mathcal{E} \in \mathcal{D}ch$, and showed that $(\mathcal{E}, \mathcal{F}_0)$ is a local regular Dirichlet form. Note that $\mathcal{D}ch \subset \mathfrak{E}$ by virtue of [8, Theorem 3.2].

Remark 3.2.3. We emphasize that Proposition 3.2.1 does not imply that the limit points of $\left(\mathcal{E}_N^{\mathcal{I}}\right)_N$ (in either the pointwise, Γ-, or the Mosco topology) belong to $\mathfrak{E}$. Rather, each of them is *comparable* to any element of $\mathfrak{E}$, in the sense that for any limit point $\bar{\mathcal{E}}$ and any $\mathcal{E} \in \mathfrak{E}$, there exist positive constants c and C such that $c\mathcal{E}(f, f) \leq \bar{\mathcal{E}}(f, f) \leq C\mathcal{E}(f, f)$ for all $f \in \mathcal{F}_0$. See Theorem 3.3.2 below.

3.3 Convergence of discrete Green forms

In this subsection we shall consider Dirichlet forms on a class of smooth measures (instead of functions), and derive a convergence result similar to Proposition 3.2.1. From now on let $\mathcal{M}_+(F)$ be the family of all nonnegative finite Borel measures on F, and let

$$\mathcal{M}_{0,\mathrm{ac}}^{(0)}(F) = \left\{\mu \in \mathcal{M}_+(F) : \mu \ll \nu,\ \frac{d\mu}{d\nu} \in \mathcal{F}_0\right\}.$$

Let $G_{\mathcal{I}_N} : V_\infty \times V_\infty \to \mathbb{R}$ be the Green's function for simple random walk on $\mathcal{I}_\infty$ killed upon exiting $\mathcal{I}_N$. By the reproducing property of Green's function, $E_{\mathcal{I}_\infty}(G_{\mathcal{I}_N}(w,\cdot),h) = h(w)$ for all $h \in \mathcal{D}(E_{\mathcal{I}_\infty})$ with $\mathrm{supp}(h) \subset I_N$. Therefore, denoting by $U_N^{\mathcal{I}}$ the 0-order potential operator associated with $\mathcal{E}_N^{\mathcal{I}}$, we have

$$\mathcal{E}_N^{\mathcal{I}}(U_N^{\mathcal{I}}\mu, h) = \langle h, \mu\rangle_{\mathcal{I}_N} = \frac{1}{m_F^N}\sum_{w\in I_N} h(w)\frac{d\mu}{d\nu_N}(w) = \mathcal{E}_N^{\mathcal{I}}\left(\rho_F^{-N}\frac{1}{m_F^N}\sum_{w\in I_N} G_{\mathcal{I}_N}(\cdot,w)\frac{d\mu}{d\nu_N}(w), h\right)$$

for all $h \in \mathcal{D}(E_{\mathcal{I}_\infty})$ with $\mathrm{supp}(h) \subset I_N$, and all nonnegative measures μ with support in I_N. It follows that

$$\begin{aligned}\mathcal{E}_N^{\mathcal{I}}(U_N^{\mathcal{I}}\mu, U_N^{\mathcal{I}}\mu) &= \mathcal{E}_N^{\mathcal{I}}\left(\rho_F^{-N}\frac{1}{m_F^N}\sum_{w\in I_N} G_{\mathcal{I}_N}(\cdot,w)\frac{d\mu}{d\nu_N}(w), \rho_F^{-N}\frac{1}{m_F^N}\sum_{w'\in I_N} G_{\mathcal{I}_N}(\cdot,w')\frac{d\mu}{d\nu_N}(w')\right)\\ &= \rho_F^{-N}\frac{1}{m_F^{2N}}\sum_{w,w'\in I_N} G_{\mathcal{I}_N}(w,w')\frac{d\mu}{d\nu_N}(w)\frac{d\mu}{d\nu_N}(w') \qquad (3.2)\end{aligned}$$

for all such measures μ. The expression in (3.2) is what we shall call the **Green form** corresponding to the Dirichlet form $\mathcal{E}_N^{\mathcal{I}}$. It has a kernel given by the (renormalized) Green's function $\rho_F^{-N}G_{\mathcal{I}_N}$, whence the name.

Our first main result of this chapter describes the convergence of the discrete Green forms.

Theorem 3.3.1. *There exist $(\mathcal{E},\mathcal{F}) \in \mathfrak{E}$ and constants $C_{3.4}(\mathcal{E}), C_{3.5}(\mathcal{E})$ such that*

$$\begin{aligned}C_{3.4}\mathcal{E}(U\mu, U\mu) &\le \varliminf_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}\left(U_N^{\mathcal{I}}\tilde{P}_N\mu, U_N^{\mathcal{I}}\tilde{P}_N\mu\right)\\ &\le \varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}\left(U_N^{\mathcal{I}}\tilde{P}_N\mu, U_N^{\mathcal{I}}\tilde{P}_N\mu\right) \le C_{3.5}\mathcal{E}(U\mu, U\mu) \qquad (3.3)\end{aligned}$$

for all $\mu \in \mathcal{M}^{(0)}_{0,\mathrm{ac}}(F)$, where U is the 0-order potential operator associated with $\mathcal{E}$.

While it is reasonable to expect that Theorem 3.3.1 follows from Proposition 3.2.1, the connection is *not* immediate. To fill in the necessary gap, we shall establish the (subsequential) *Mosco convergence* of the discrete Dirichlet forms $(\mathcal{E}^I_N)_N$, which is an equivalent condition to (subsequential) *strong resolvent convergence* in L^2 [23, 63]. This then implies (subsequential) *strong resolvent convergence in the energy norm* (Lemma 3.6.4). In addition, we will show that each Mosco limit point of $(\mathcal{E}^I_N)_N$ is comparable to some (and hence any) element of $\mathfrak{E}$, which finally leads to Theorem 3.3.1.

Our analysis of the Mosco limits draws from similar analysis of the Γ-limits by Sturm [75] and by Kumagai and Sturm [49], whose goal was to construct diffusions on an arbitrary metric measure space via Γ-limits. For the reader's convenience, we have collected some general notions about Γ-convergence and Mosco convergence in §3.6.2.

Theorem 3.3.2. *Let $\mathcal{E}^*(f,f) = \overline{\lim_{N\to\infty}} \mathcal{E}^I_N(\tilde{P}_N f, \tilde{P}_N f)$ and $\mathcal{F}^* := \{f \in C_c(F_\infty) : \mathcal{E}^*(f,f) < \infty\}$.*

(i) $\mathcal{F}^*$ is a Lipschitz space and is dense in $C_c(F_\infty)$.

(ii) There exists a subsequence $(r_N)_N \subset \mathbb{N}$ along which the Mosco limit $\mathcal{E}_M := M\text{-}\lim_{N\to\infty} \mathcal{E}^I_{r_N}$ exists. Moreover, given any two Mosco limit points $\mathcal{E}_M$ and $\mathcal{E}'_M$, there exist constants C and C' such that $C\mathcal{E}_M(f,f) \le \mathcal{E}'_M(f,f) \le C'\mathcal{E}_M(f,f)$ for all $f \in \mathcal{F}^*$.

For each Mosco limit point $\mathcal{E}_M$ of $(\mathcal{E}^I_N)_N$:

(iii) $(\mathcal{E}_M, \mathcal{F}^*)$ is a closable symmetric Markovian form on $L^2(F_\infty, \nu_\infty)$, and can be extended to a regular Dirichlet form $(\bar{\mathcal{E}}, \bar{\mathcal{F}})$ on $L^2(F_\infty, \nu_\infty)$ with core $\mathcal{F}^*$.

(iv) There exist $\mathcal{E} \in \mathfrak{E}$ and constants C and C' such that $C\mathcal{E} \le \bar{\mathcal{E}} \le C'\mathcal{E}$.

Remark 3.3.3. We do not, nor need not, assert that $\bar{\mathcal{E}} \in \mathfrak{E}$. One reason is that we cannot check whether $\bar{\mathcal{E}}$ is (strongly) local. In [63, Theorem 4.4.1] a sufficient condition was given for a sequence of (strongly) local regular Dirichlet forms on $L^2(X, m)$ to Γ-converge to a (strongly) local regular Dirichlet form. The condition states that the sequence of *energy measures* be bounded and absolutely continuous with respect to the reference measure m on X. However, in the fractal setting, the energy measure and the self-similar measure on the limiting fractal set are mutually singular [38] (see also [11, 50] for the version of this statement on post-critically finite fractals, such as the Sierpinski gasket).

On the other hand, since $(\bar{\mathcal{E}}, \bar{\mathcal{F}})$ is a regular Dirichlet form, one has $\bar{\mathcal{E}} = \bar{\mathcal{E}}^{(c)} + \bar{\mathcal{E}}^{(j)} + \bar{\mathcal{E}}^{(k)}$ by the Beurling-Deny formula (*cf.* [32, Section 3.2]), where $\bar{\mathcal{E}}^{(c)}$, $\bar{\mathcal{E}}^{(j)}$, $\bar{\mathcal{E}}^{(k)}$ stands respectively for the diffusion, jump, and killing part of $\bar{\mathcal{E}}$. In particular, $\bar{\mathcal{E}}^{(c)}$ is strongly local, and assuming Proposition 3.2.1, one can show that $\bar{\mathcal{E}}$ and $\bar{\mathcal{E}}^{(c)}$ are comparable (*cf.* [49, §2]).

Proof of Theorem 3.3.2. (i): By [6, Theorem 1.4], there is a $(\mathcal{E}, \mathcal{F}) \in \mathfrak{E}$ associated with a diffusion on F_∞ whose heat kernel satisfies the following estimate: there exist C_1, C_2, C_3 and C_4 such that for all $x, y \in F_\infty$ and $t > 0$,

$$C_1 t^{-d_h/d_w} \exp\left(-C_2 \left(\frac{\|x-y\|^{d_w}}{t}\right)^{\frac{1}{d_w-1}}\right) \le p_t(x,y) \le C_3 t^{-d_h/d_w} \exp\left(-C_4 \left(\frac{\|x-y\|^{d_w}}{t}\right)^{\frac{1}{d_w-1}}\right).$$

Then by [49, Theorem 4.1], $\mathcal{F} = \mathrm{Lip}_{\nu_\infty}(\frac{d_w}{2}, 2, \infty)(F_\infty)$, the Besov-Lipschitz space which consists of all $f \in L^2(F_\infty, \nu_\infty)$ such that

$$\sup_{N \in \mathbb{N} \cup \{0\}} \alpha^{N(d_w+d_h)} \int \int_{|x-y|<c_0\alpha^{-N}} |f(x) - f(y)|^2 d\nu_\infty(x) d\nu_\infty(y) < \infty$$

for some $\alpha > 1$ and $c > 0$. (See also [48] for an earlier derivation.) Using the inequality (3.1), we deduce that $\mathcal{F}^* = \mathcal{F}$ is a Lipschitz space and hence is dense in $C_0(F)$. Note that this is Assumption (B1) in [49, §3].

(ii): By [23, Proposition 8.10], it suffices to show that $L^2(F_\infty, \nu_\infty)$ has a separable dual (clear), and that there exists a function $\xi : L^2(F_\infty, \nu_\infty) \to \mathbb{R} \cup \{+\infty\}$ such that $\lim_{\|f\|_{L^2} \to +\infty} \xi(f) = +\infty$ and $\mathcal{E}_N^{\mathcal{I}}(\tilde{P}_N f, \tilde{P}_N f) \geq \xi(f, f)$ for all $f \in L^2(F_\infty, \nu_\infty)$ and all sufficiently large N. According to the right inequality in (3.1), one can fix $\xi(f) = C\mathcal{E}(f, f)$ for some $\mathcal{E} \in \mathfrak{E}$ and constant C.

As a consequence, $L^2(F_\infty, \nu_\infty)$ equipped with either the strong or the weak topology has a countable base, so by [23, Theorem 8.5], any (sub)sequence of $(\mathcal{E}_N^{\mathcal{I}})_N$ contains a Mosco- (resp. weak Γ-) convergent (sub)subsequence. Moreover, $\mathcal{E}_M(f, f) \leq \mathcal{E}^*(f, f) < \infty$ for all $f \in \mathcal{F}^*$ and for any Mosco limit point $\mathcal{E}_M$.

(iii): From Part (i) we have closability and regularity. Symmetry and the Markovian property can be verified easily.

(iv): Take any Mosco convergent subsequence $(\mathcal{E}_{r_N}^{\mathcal{I}})_N$ and denote its limit point (resp. the smallest closed extension thereof) by $\mathcal{E}_M$ (resp. $\bar{\mathcal{E}}$). By Proposition 3.2.1(i), for all $N \geq 1$ we have

$$\mathcal{E}_N^{\mathcal{I}}(\tilde{P}_N f, \tilde{P}_N f) \leq C_{2.1} \cdot \Gamma_w\text{-}\varliminf_{N' \to \infty} \mathcal{E}_{r_{N'}}^{\mathcal{I}}(\tilde{P}_{r_{N'}} f, \tilde{P}_{r_{N'}} f) = C_{2.1} \cdot \bar{\mathcal{E}}(f, f).$$

Taking $\sup_N$ on both sides and using Proposition 3.2.1(ii) yields $\bar{\mathcal{E}}(f, f) \geq C\mathcal{E}^*(f, f) \geq C'\mathcal{E}(f, f)$ for some $\mathcal{E} \in \mathfrak{E}$. Meanwhile $\bar{\mathcal{E}}(f, f) \leq \mathcal{E}^*(f, f) \leq C''\mathcal{E}(f, f)$ via Proposition 3.2.1(ii) again. This completes the proof. ■

Proof of Theorem 3.3.1.

By Proposition 3.6.3 and Lemma 3.6.4, given each Mosco convergent subsequence

$(\mathcal{E}^{\mathcal{I}}_{r_N})_N$ with limit $\mathcal{E}_M$ and its smallest closed extension $\bar{\mathcal{E}}$, we have that for all $\mu \in \mathcal{M}^{(0)}_{0,\mathrm{ac}}(F)$,

$$\lim_{N\to\infty} \mathcal{E}^{\mathcal{I}}_{r_N}\left(U^{\mathcal{I}}_{r_N}\tilde{P}_{r_N}\mu, U^{\mathcal{I}}_{r_N}\tilde{P}_{r_N}\mu\right) = \bar{\mathcal{E}}\left(\bar{U}\mu, \bar{U}\mu\right).$$

According to Theorem 3.3.2(iv), there exist $(\mathcal{E}, \mathcal{F}) \in \mathfrak{E}$ and constants $C(\mathcal{E}), C'(\mathcal{E})$ such that for all $f \in \mathcal{F}$,

$$C\mathcal{E}(f,f) \le \bar{\mathcal{E}}(f,f) \le C'\mathcal{E}(f,f).$$

If f is a 0-order potential relative to $\mathcal{E}$, then we can write $f = U\mu$ for some $\mu \in S^{(0)}_0$. Using the identity

$$\mathcal{E}(U\mu, h) = \int \widetilde{h} d\mu = \bar{\mathcal{E}}(\bar{U}\mu, h) \qquad \text{for all } \mu \in S^{(0)}_0 \text{ and } h \in \mathcal{F}_e,$$

(where $\widetilde{h}$ is the quasi-continuous modification of h, *cf.* §3.6.1), we then apply the Cauchy-Schwarz inequality to $\mathcal{E}$, and deduce that for all $\mu \in \mathcal{M}^{(0)}_{0,\mathrm{ac}}(F)$,

$$[\bar{\mathcal{E}}(\bar{U}\mu, \bar{U}\mu)]^2 = [\mathcal{E}(U\mu, \bar{U}\mu)]^2 \le \mathcal{E}(U\mu, U\mu)\mathcal{E}(\bar{U}\mu, \bar{U}\mu) \le C^{-1}\mathcal{E}(U\mu, U\mu)\bar{\mathcal{E}}(\bar{U}\mu, \bar{U}\mu),$$

or $\bar{\mathcal{E}}(\bar{U}\mu, \bar{U}\mu) \le C^{-1}\mathcal{E}(U\mu, U\mu)$. Reversing the role of $\mathcal{E}$ and $\bar{\mathcal{E}}$ gives the opposite inequality, and hence

$$(C')^{-1}\mathcal{E}(U\mu, U\mu) \le \bar{\mathcal{E}}(\bar{U}\mu, \bar{U}\mu) \le C^{-1}\mathcal{E}(U\mu, U\mu).$$

Since this comparison holds when $\bar{\mathcal{E}}$ is the smallest closed extension of either the maximal and minimal cluster point of $(\mathcal{E}^{\mathcal{I}}_N)_N$ in the Mosco topology, we find (3.3). ■

3.4 Comparison of discrete Dirichlet & Green forms

In this section, we compare the discrete Dirichlet (and Green) forms on $\mathcal{G}_\infty$ and on $\mathcal{I}_\infty$.

Observe (from Figure 3.1) that for each "center vertex" $w \in I_\infty$, there is a unique set $C(w)$ of 2^d "corner vertices" in V_∞ which are nearest neighbors of w, *i.e.,* $C(w) = \{x \in$

$V_\infty : \|x - w\| = \sqrt{d}/2\}$. Let $\tilde{Q} : C(V_\infty; \mathbb{R}) \to C(I_\infty; \mathbb{R})$ be the projection operator given by

$$(\tilde{Q}f)(w) = \frac{1}{2^d} \sum_{x \in C(w)} f(x).$$

Let $G_{\mathcal{G}_N} : V_\infty \times V_\infty \to \mathbb{R}$ denote the Green's function killed upon exiting $\mathcal{G}_N$.

Lemma 3.4.1. *For all $f \in \mathcal{D}(E_{\mathcal{G}_\infty})$,*

$$E_{\mathcal{G}_\infty}(f, f) \geq E_{I_\infty}\left(\tilde{Q}f, \tilde{Q}f\right). \tag{3.4}$$

It follows that for all nonnegative functions f on V_N,

$$\sum_{x,x' \in V_N} G_{\mathcal{G}_N}(x, x') f(x) f(x') \leq 2^{2d} \sum_{w,w' \in I_N} G_{I_N}(w, w')(\tilde{Q}f)(w)(\tilde{Q}f)(w'). \tag{3.5}$$

Proof. Observe that for every $w_1, w_2 \in I_\infty$ with $w_1 \sim w_2$,

$$(\tilde{Q}f)(w_1) - (\tilde{Q}f)(w_2) = \frac{1}{2^d} \sum_{\substack{x_1 \in C(w_1) \setminus C(w_2) \\ x_2 \in C(w_2) \setminus C(w_1) \\ z \in C(w_1) \cap C(w_2) \\ x_1 \sim z \sim x_2}} \left[(f(x_1) - f(z)) + (f(z) - f(x_2))\right].$$

The sum is over the difference of f across 2^d edges in $\mathcal{G}_\infty$ which are parallel to the line segment $\overline{w_1 w_2}$, and whose vertices are nearest neighbors of either w_1 or w_2. Taking the square of both sides, and applying the inequality $(\sum_{k=1}^n a_k)^2 \leq n(\sum_{k=1}^n a_k^2)$, we obtain

$$\left[(\tilde{Q}f)(w_1) - (\tilde{Q}f)(w_2)\right]^2 \leq \frac{1}{2^d} \sum_{\substack{x_1 \in C(w_1) \setminus C(w_2) \\ x_2 \in C(w_2) \setminus C(w_1) \\ z \in C(w_1) \cap C(w_2) \\ x_1 \sim z \sim x_2}} \left[(f(x_1) - f(z))^2 + (f(z) - f(x_2))^2\right].$$

Upon summing over all $w_1, w_2 \in I_\infty$, we note that each edge in $\mathcal{G}_\infty$ contributes at most 2^d terms to the RHS, that is,

$$\begin{aligned} E_{I_\infty}(\tilde{Q}f, \tilde{Q}f) &= \frac{1}{2} \sum_{w_1 \sim w_2} \left[(\tilde{Q}f)(w_1) - (\tilde{Q}f)(w_2)\right]^2 \\ &\leq \frac{1}{2} \cdot \frac{1}{2^d} \cdot 2^d \sum_{\substack{x_1, x_2 \in V_\infty \\ x_1 \sim x_2}} (f(x_1) - f(x_2))^2 = E_{\mathcal{G}_\infty}(f, f). \end{aligned}$$

This proves (3.4).

Next we turn to the Green form inequality (3.5). Observe that for all nonnegative functions f, h on V_N,

$$\begin{aligned}\sum_{w\in I_N}(\tilde{Q}f)(w)(\tilde{Q}h)(w) &= \frac{1}{2^{2d}}\sum_{w\in I_N}\sum_{x,x'\in C(w)} f(x)h(x') \geq \frac{1}{2^{2d}}\sum_{w\in I_N}\sum_{x,x'\in C(w)} f(x)h(x)\\ &= \frac{1}{2^d}\sum_{w\in I_N}\sum_{x\in C(w)} f(x)h(x) \geq \frac{1}{2^d}\sum_{x\in V_N} f(x)h(x).\end{aligned}$$

By the reproducing property of Green's functions, we deduce that for all $h : V_\infty \to \mathbb{R}_+$ with support in V_N,

$$E_{\mathcal{I}_\infty}\left(\sum_{w\in I_N} G_{\mathcal{I}_N}(\cdot, w)(\tilde{Q}f)(w), \tilde{Q}h\right) \geq \frac{1}{2^d}E_{\mathcal{G}_\infty}\left(\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot, x)f(x), h\right).$$

Taking $h = \sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot, x)f(x)$ and again applying the reproducing property yields

$$E_{\mathcal{I}_\infty}\left(\sum_{w\in I_N} G_{\mathcal{I}_N}(\cdot, w)(\tilde{Q}f)(w), \sum_{x\in V_N}\tilde{Q}G_{\mathcal{G}_N}(\cdot, x)f(x)\right) \geq \frac{1}{2^d}\sum_{x,x'\in V_N} G_{\mathcal{G}_N}(x, x')f(x)f(x').$$

To simplify notations, we introduce shorthands for the functions

$$f_\alpha := \sum_{w\in I_N} G_{\mathcal{I}_N}(\cdot, w)(\tilde{Q}f)(w) \quad \text{and} \quad f_\beta := \sum_{x\in V_N}\tilde{Q}G_{\mathcal{G}_N}(\cdot, x)f(x),$$

on I_N. It is clear that

$$E_{\mathcal{I}_\infty}(f_\alpha, f_\alpha) = \sum_{w,w'\in I_N} G_{\mathcal{I}_N}(w, w')(\tilde{Q}f)(w)(\tilde{Q}f)(w').$$

Meanwhile, by the energy inequality (3.4) we just proved,

$$E_{\mathcal{I}_\infty}(f_\beta, f_\beta) \leq E_{\mathcal{G}_\infty}\left(\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot, x)f(x), \sum_{x'\in V_N} G_{\mathcal{G}_N}(\cdot, x')f(x')\right) = \sum_{x,x'\in V_N} G_{\mathcal{G}_N}(x, x')f(x)f(x').$$

So putting everything together,

$$\begin{aligned}0 &\leq E_{\mathcal{I}_\infty}\left(f_\alpha - \frac{1}{2^d}f_\beta,\ f_\alpha - \frac{1}{2^d}f_\beta\right)\\ &= E_{\mathcal{I}_\infty}(f_\alpha, f_\alpha) - \frac{2}{2^d}E_{\mathcal{I}_\infty}\left(f_\alpha, f_\beta\right) + \frac{1}{2^{2d}}E_{\mathcal{I}_\infty}(f_\beta, f_\beta)\\ &\leq \sum_{w,w'\in I_N} G_{\mathcal{I}_N}(w, w')(\tilde{Q}f)(w)(\tilde{Q}f)(w') - \frac{1}{2^{2d}}\sum_{x,x'\in V_N} G_{\mathcal{G}_N}(x, x')f(x)f(x'),\end{aligned}$$

which yields (3.5). ■

3.5 The main lemma

In this section, we establish the limsup convergence of discrete Green forms on $\mathcal{G}_\infty$, which play a crucial role in the main proofs of [20].

Let $G^{\square}_{\mathcal{G}_N} : V_N \times V_N \to \mathbb{R}$ be the restriction of $G_{\mathcal{G}_\infty}$ on $V_N \times V_N$; we have added a superscript $\square$ to distinguish it from the Green's function on $\mathcal{G}_\infty$ killed upon exiting $\mathcal{G}_N$. Also introduce the probability measure $\eta_N := \frac{1}{|V_N|}\mathbb{1}_{V_N}$ on V_N. Define, for any $h \in \ell^1(V_N;\mathbb{R}) \cap \ell^\infty(V_N;\mathbb{R})$,

$$U_N^{\mathcal{G}}(h\eta_N) := \rho_F^{-N}\frac{1}{|V_N|}\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot,x)h(x), \quad U_N^{\mathcal{G}^\square}(h\eta_N) := \rho_F^{-N}\frac{1}{|V_N|}\sum_{x\in V_N} G^{\square}_{\mathcal{G}_N}(\cdot,x)h(x).$$

Writing $\mathcal{E}_N^{\mathcal{G}} = \rho_F^N E_{\mathcal{G}_\infty}$ for the renormalized discrete Dirichlet form on $\mathcal{G}_\infty$, we have

$$\mathcal{E}_N^{\mathcal{G}}\left(U_N^{\mathcal{G}}(h\eta_N), U_N^{\mathcal{G}}(h\eta_N)\right) = \rho_F^{-N}\frac{1}{|V_N|^2}\sum_{x,x'\in V_N} G_{\mathcal{G}_N}(x,x')h(x)h(x')$$

by the reproducing property of $G_{\mathcal{G}_N}$. Meanwhile, let us abuse notations slightly and introduce the quadratic form

$$\mathcal{E}_N^{\mathcal{G}}(U_N^{\mathcal{G}^\square}(h\eta_N), U_N^{\mathcal{G}^\square}(h\eta_N)) := \rho_F^{-N}\frac{1}{|V_N|^2}\sum_{x,x'\in V_N} G^{\square}_{\mathcal{G}_N}(x,x')h(x)h(x'),$$

as it is suggestive of another Green form.

Lemma 3.5.1 (The main lemma)**.** *For every $h \in L^1(F,\nu)\cap L^\infty(F,\nu)$, define $h_N : V_N \to \mathbb{R}$ by $h_N(\cdot) = h(\ell_F^{-N}\cdot)$. Then the following hold:*

(i) $\overline{\lim_{N\to\infty}}\, \mathcal{E}_N^{\mathcal{G}}(U_N^{\mathcal{G}^\square}(h_N\eta_N), U_N^{\mathcal{G}^\square}(h_N\eta_N)) = \overline{\lim_{N\to\infty}}\, \mathcal{E}_N^{\mathcal{G}}(U_N^{\mathcal{G}}(h_N\eta_N), U_N^{\mathcal{G}}(h_N\eta_N))$.

For some $(\mathcal{E},\mathcal{F}) \in \mathfrak{E}$:

(ii) There exists a constant $C_{3.6}(\mathcal{E})$ such that

$$\overline{\lim_{N\to\infty}}\, \mathcal{E}_N^{\mathcal{G}}\left(U_N^{\mathcal{G}^\square}\left(\left(\frac{d\mu}{d\nu}\right)_N \eta_N\right), U_N^{\mathcal{G}^\square}\left(\left(\frac{d\mu}{d\nu}\right)_N \eta_N\right)\right) \le C_{3.6}\mathcal{E}(U\mu, U\mu)$$

for all $\mu \in \mathcal{M}^{(0)}_{0,\mathrm{ac}}(F)$.

(iii) There exists a constant $C_{3.7}(\mathcal{E})$ such that

$$\overline{\lim_{N\to\infty}}\,\rho_F^N\left\langle \mathbb{1}_{V_N}, \sum_{x\in V_N}(G^{\square}_{\mathcal{G}_N})^{-1}(\cdot,x)\mathbb{1}_{V_N}(x)\right\rangle_{V_N} \le C_{3.7}\mathrm{Cap}_{\mathcal{E}}(F),$$

where $\left(G^{\square}_{\mathcal{G}_N}\right)^{-1}$ denotes the matrix inverse of $G^{\square}_{\mathcal{G}_N}$, and $\mathrm{Cap}_{\mathcal{E}}(F)$ denotes the 0-capacity of F with respect to $\mathcal{E}$.

For some general facts about the 0-order capacity (which will be used in the proof of Part (iii)), please see §3.6.1.

Proof. (i): We need two facts. The first is the observation that $\uparrow \lim_{N\to\infty} G_{\mathcal{G}_N}(x,x') = G_{\mathcal{G}_\infty}(x,x')$ for all $x,x'\in V_\infty$, since the first exit time from $\mathcal{G}_N$ increases unboundedly with N. The second is the following two-sided Green's function estimate on $\mathcal{G}_\infty$, proved in [7, Theorem 5.3]: there exist constants C_5, C_6 such that for all $x,x'\in V_\infty$ with $d_{\mathcal{G}_\infty}(x,x')\ge 1$,

$$C_5\cdot d_{\mathcal{G}_\infty}(x,x')^{d_w-d_h} \le G_{\mathcal{G}_\infty}(x,x') \le C_6\cdot d_{\mathcal{G}_\infty}(x,x')^{d_w-d_h}.$$

Note that $d_w-d_h<0$ as $\mathcal{G}_\infty$ is a transient Sierpinski carpet graph. Therefore

$$\epsilon_N(x,x') := \frac{G_{\mathcal{G}_\infty}(x,x')-G_{\mathcal{G}_N}(x,x')}{d_{\mathcal{G}_\infty}(x,x')^{d_w-d_h}}$$

is bounded above by C_6 (and bounded below by 0), and $\lim_{N\to\infty}\epsilon_N(x,x')=0$ pointwise.

Let us now estimate the difference between the two sides of the equation in Part (i):

$$\begin{aligned}
0 &\le \overline{\lim_{N\to\infty}}\left[\mathcal{E}_N^{\mathcal{G}}(U_N^{\mathcal{G}\square}(h_N\eta_N),U_N^{\mathcal{G}\square}(h_N\eta_N)) - \mathcal{E}_N^{\mathcal{G}}(U_N^{\mathcal{G}}(h_N\eta_N),U_N^{\mathcal{G}}(h_N\eta_N))\right]\\
&= \overline{\lim_{N\to\infty}}\,\rho_F^{-N}\frac{1}{|V_N|^2}\sum_{x,x'\in V_N}\left[G_{\mathcal{G}_\infty}(x,x')-G_{\mathcal{G}_N}(x,x')\right]h_N(x)h_N(x')\\
&= \overline{\lim_{N\to\infty}}\,\rho_F^{-N}\frac{1}{|V_N|^2}\sum_{x,x'\in V_N}\epsilon_N(x,x')\cdot d_{\mathcal{G}_\infty}(x,x')^{d_w-d_h}h_N(x)h_N(x')\\
&\le C\,\overline{\lim_{N\to\infty}}\,\rho_F^{-N}\frac{1}{|V_N|^2}\sum_{y,y'\in\ell_F^{-N}V_N}\rho_F^N\epsilon_N(\ell_F^N y,\ell_F^N y')\|y-y'\|^{d_w-d_h}h(y)h(y')\\
&= C\,\overline{\lim_{N\to\infty}}\int_{F\times F}\epsilon_N(\ell_F^N y,\ell_F^N y')\frac{h(y)h(y')d\mathfrak{m}_N(y)d\mathfrak{m}_N(y)}{\|y-y'\|^{d_h-d_w}},
\end{aligned}\tag{3.6}$$

where $\mathfrak{m}_N = \dfrac{1}{|V_N|}\mathbb{1}_{\ell_F^{-N}V_N}$ is a probability measure on F, and $\mathfrak{m}_N$ converges weakly to ν. Now

$$\int_{F\times F} \frac{h(y)h(y')d\nu(y)d\nu(y')}{\|y-y'\|^{d_h-d_w}} \le \|h\|_\infty^2 \int_{F\times F} \frac{d\nu(y)d\nu(y')}{\|y-y'\|^{d_h-d_w}} < \infty,$$

where we use a fact from geometric measure theory (see *e.g.* [59, Ch. 8]) that since ν is a d_h-dimensional Hausdorff measure with respect to the Euclidean norm $\|\cdot\|$ on the compact metric space $(F, \|\cdot\|)$,

$$\int_{F\times F} \frac{d\nu(y)d\nu(y')}{\|y-y'\|^{\alpha}} < \infty$$

for any $\alpha < d_h$. So by the reverse Fatou's lemma for weakly converging measures (*cf.* [69, 71]; see also [28, Theorem 1.1] for the statement and proof), the RHS of (3.6) is bounded above by

$$C\int_{F\times F} \varlimsup_{\substack{N\to\infty\\ y_1\to y, y_2\to y'}} \left(\frac{\epsilon_N(\ell_F^N y_1, \ell_F^N y_2)h(y_1)h(y_2)}{\|y_1-y_2\|^{d_h-d_w}}\right) d\nu(y)d\nu(y') = 0.$$

The result in (i) then follows.

(ii): By Part (i) and then (3.5), there exists a constant $C(d)$ such that

$$\begin{aligned}\varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{G}}\left(U_N^{\mathcal{G}\square}(h_N\eta_N), U_N^{\mathcal{G}\square}(h_N\eta_N)\right) &= \varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{G}}\left(U_N^{\mathcal{G}}(h_N\eta_N), U_N^{\mathcal{G}}(h_N\eta_N)\right)\\ &\le C\varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}\left(U_N^{\mathcal{I}}\left(\tilde{Q}h_N\nu_N\right), U_N^{\mathcal{I}}\left(\tilde{Q}h_N\nu_N\right)\right).\end{aligned}$$

We claim that $\tilde{Q}h_N$ can be replaced by $\tilde{P}_N h$ in the above inequality. Indeed, for any continuous function h on F,

$$\begin{aligned}&\lim_{N\to\infty} \|\tilde{Q}h_N - \tilde{P}_N h\|_{L^1(I_N,\nu_N)}\\ \le\ &\lim_{N\to\infty} \frac{1}{m_F^N}\sum_{w\in I_N} \frac{1}{\left|\nu(\Psi_w^{(N)}\cap F)\right|}\int_{\Psi_w^{(N)}\cap F}\left|h(y) - \frac{1}{2^d}\sum_{x\in C(w)} h\left(\frac{x}{\ell_F^N}\right)\right| d\nu(y)\\ \le\ &\lim_{N\to\infty}\sup_{w\in I_N}\sup_{y\in\Psi_w^{(N)}\cap F}\left|h(y) - \frac{1}{2^d}\sum_{x\in C(w)} h\left(\frac{x}{\ell_F^N}\right)\right| = 0.\end{aligned}$$

By the density of $C_c(F)$ in $L^1(F,\nu)$, we can extend this result to all $h \in L^1(F,\nu)\cap L^\infty(F,\nu)$. Hence

$$\begin{aligned}
&\varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}\left(U_N^{\mathcal{I}}\left(\tilde{Q}h_N\nu_N\right), U_N^{\mathcal{I}}\left(\tilde{Q}h_N\nu_N\right)\right) = \varlimsup_{N\to\infty}\left\langle U_N^{\mathcal{I}}\left(\tilde{Q}h_N\nu_N\right), \tilde{Q}h_N\nu_N\right\rangle_{\mathcal{I}_N}\\
=\ &\varlimsup_{N\to\infty}\left\langle U_N^{\mathcal{I}}\left((\tilde{Q}h_N+\tilde{P}_Nh)\nu_N\right), (\tilde{Q}h_N-\tilde{P}_Nh)\nu_N\right\rangle_{\mathcal{I}_N} + \varlimsup_{N\to\infty}\left\langle U_N^{\mathcal{I}}\left(\tilde{P}_Nh\nu_N\right), \tilde{P}_Nh\nu_N\right\rangle_{\mathcal{I}_N}\\
=\ &\varlimsup_{N\to\infty}\left\langle U_N^{\mathcal{I}}\left(\tilde{P}_Nh\nu_N\right), \tilde{P}_Nh\nu_N\right\rangle_{\mathcal{I}_N} = \varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}\left(U_N^{\mathcal{I}}\left(\tilde{P}_Nh\nu_N\right), U_N^{\mathcal{I}}\left(\tilde{P}_Nh\nu_N\right)\right).
\end{aligned}$$

Now put $h = \frac{d\mu}{d\nu}$ for some $\mu \in \mathcal{M}_{0,\mathrm{ac}}^{(0)}(F)$. It follows from the preceding discussions and Theorem 3.3.1 that

$$\begin{aligned}
\varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{G}}\left(U_N^{\mathcal{G}\square}\left(\left(\frac{d\mu}{d\nu}\right)_N\eta_N\right), U_N^{\mathcal{G}\square}\left(\left(\frac{d\mu}{d\nu}\right)_N\eta_N\right)\right) &\le C\varlimsup_{N\to\infty}\mathcal{E}_N^{\mathcal{I}}\left(U_N^{\mathcal{I}}(\tilde{P}_N\mu), U_N^{\mathcal{I}}(\tilde{P}_N\mu)\right)\\
&\le C'\mathcal{E}(U\mu, U\mu),
\end{aligned}$$

where $C' = CC_{2.5}$.

(iii): We recognize that for all $h : V_N \to \mathbb{R}$,

$$\left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}^{\square}(\cdot,x)(h\eta_N)(x), h\eta_N\right\rangle_{V_N} \ge \left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot,x)(h\eta_N)(x), h\eta_N\right\rangle_{V_N}. \tag{3.7}$$

Fixing $f : V_N \to \mathbb{R}$, we let $\eta = \rho_F^N\sum_{x\in V_N}(G_{\mathcal{G}_N})^{-1}(\cdot,x)f(x)$ and $\eta^{\square} = \rho_F^N\sum_{x\in V_N}(G_{\mathcal{G}_N}^{\square})^{-1}(\cdot,x)f(x)$. Then upon applying the Cauchy-Schwarz inequality and (3.7), we find

$$\begin{aligned}
\left\langle f, \rho_F^N\sum_{x\in V_N}(G_{\mathcal{G}_N}^{\square})^{-1}(\cdot,x)f(x)\right\rangle_{V_N}^2 &= \left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot,x)\eta(x), \eta^o\right\rangle_{V_N}^2\\
&\le \left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot,x)\eta(x), \eta\right\rangle_{V_N}\left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot,x)\eta^o(x), \eta^o\right\rangle_{V_N}\\
&\le \left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot,x)\eta(x), \eta\right\rangle_{V_N}\left\langle \rho_F^{-N}\sum_{x\in V_N} G_{\mathcal{G}_N}^{\square}(\cdot,x)\eta^o(x), \eta^o\right\rangle_{V_N}\\
&= \left\langle f, \rho_F^N\sum_{x\in V_N}(G_{\mathcal{G}_N})^{-1}(\cdot,x)f(x)\right\rangle_{V_N}\left\langle f, \rho_F^N\sum_{x\in V_N}(G_{\mathcal{G}_N}^{\square})^{-1}(\cdot,x)f(x)\right\rangle_{V_N}.
\end{aligned}$$

Hence for all $f : V_N \to \mathbb{R}$,

$$\left\langle f, \rho_F^N \sum_{x\in V_N} (G^{\square}_{\mathcal{G}_N})^{-1}(\cdot, x) f(x) \right\rangle_{V_N} \leq \left\langle f, \rho_F^N \sum_{x\in V_N} (G_{\mathcal{G}_N})^{-1}(\cdot, x) f(x) \right\rangle_{V_N}.$$

In particular,

$$\begin{aligned} \varlimsup_{N\to\infty} \rho_F^N \left\langle \mathbb{1}_{V_N}, \sum_{x\in V_N} (G^{\square}_{\mathcal{G}_N})^{-1}(\cdot, x)\mathbb{1}_{V_N}(x) \right\rangle_{V_N} &\leq \varlimsup_{N\to\infty} \rho_F^N \left\langle \mathbb{1}_{V_N}, \sum_{x\in V_N} (G_{\mathcal{G}_N})^{-1}(\cdot, x)\mathbb{1}_{V_N}(x) \right\rangle_{V_N} \\ &= \varlimsup_{N\to\infty} \langle \mathbb{1}_{V_N}, \mu_{V_N} \rangle_{V_N}, \end{aligned} \tag{3.8}$$

where μ_{V_N} is the equilibrium measure on V_N with respect to $\mathcal{E}_N^{\mathcal{G}}$. The equality in the second line is just a direct calculation by

$$U_N^{\mathcal{G}} \mu_{V_N} = \rho_F^{-N} \frac{1}{|V_N|} \sum_{x\in V_N} G_{\mathcal{G}_N}(\cdot, x) \frac{d\mu_{V_N}}{d\eta_N}(x) = 1 \quad \text{on } V_N,$$

cf. Proposition 3.6.2(ii). From this it also follows that $\rho_F^{-N}\mu_{V_N}(V_N) \leq 1$, and

$$\begin{aligned} U_N^{\mathcal{I}} \tilde{Q}\mu_{V_N} &= \rho_F^{-N} \frac{1}{m_F^N} \sum_{w\in I_N} G_{\mathcal{I}_N}(\cdot, w) \frac{1}{2^d} \sum_{x\in C(w)} \left(\frac{d\mu_{V_N}}{d\eta_N} \right)(x) \\ &\leq \frac{|V_N|}{m_F^N} \|G_{\mathcal{I}_\infty}\|_\infty \rho_F^{-N} \mu_{V_N}(V_N) \leq \frac{|V_N|}{m_F^N} \|G_{\mathcal{I}_\infty}\|_\infty \quad \text{on } I_N. \end{aligned}$$

So if we let $\hat{\mu}_{V_N} := \left(\frac{|V_N|}{m_F^N} \|G_{\mathcal{I}_\infty}\|_\infty \right)^{-1} \mu_{V_N}$, then $U_N^{\mathcal{I}} \tilde{Q} \hat{\mu}_{V_N} \leq 1$ on I_N. Hence by Proposition 3.6.2(iv),

$$\begin{aligned} \langle \mathbb{1}_{V_N}, \mu_{V_N} \rangle_{V_N} &\leq 2^d \frac{m_F^N}{|V_N|} \langle \mathbb{1}_{I_N}, \tilde{Q}\mu_{V_N} \rangle_{I_N} = 2^d \|G_{\mathcal{I}_\infty}\|_\infty \langle \mathbb{1}_{I_N}, \tilde{Q}\hat{\mu}_{V_N} \rangle_{I_N} \\ &\leq 2^d \|G_{\mathcal{I}_\infty}\|_\infty \langle \mathbb{1}_{I_N}, \mu_{I_N} \rangle_{I_N} = 2^d \|G_{\mathcal{I}_\infty}\|_\infty \mathcal{E}_N^{\mathcal{I}}(e_{I_N}, e_{I_N}), \end{aligned}$$

where μ_{I_N} and e_{I_N} are, respectively, the equilibrium measure and equilibrium potential of I_N with respect to $\mathcal{E}_N^{\mathcal{I}}$. Putting $C_7 := 2^d \|G_{\mathcal{I}_\infty}\|_\infty$, and applying Proposition 3.6.2(i) and Proposition 3.2.1(ii), we find

$$\varlimsup_{N\to\infty} \langle \mathbb{1}_{V_N}, \mu_{V_N} \rangle_{V_N} \leq C_7 \varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}(e_{I_N}, e_{I_N}) \leq C_7 \varlimsup_{N\to\infty} \mathcal{E}_N^{\mathcal{I}}\left(\tilde{P}_N e_F, \tilde{P}_N e_F \right) \leq C_7 C_{2.2}^{-1} \mathcal{E}(e_F, e_F). \tag{3.9}$$

Inequalities (3.8) and (3.9) together imply the result. ■

3.6 Appendix

3.6.1 Smooth measures and capacity

In this subsection we introduce the concepts of smooth measures and (0-order) capacity with respect to a (transient) regular Dirichlet form. Much of this can be found in [32, Chapter 2] and [21, Chapter 2].

Suppose $(\mathcal{E}, \mathcal{F})$ is a regular Dirichlet form on $L^2(X, m)$. Let $\mathfrak{O}$ denote the family of all open subsets of X, and for each $A \in \mathfrak{O}$, define $\mathcal{L}_A = \{u \in \mathcal{F} : u \geq 1 \ m\text{-a.e. on } A\}$. The 1-capacity of the set $A \in \mathfrak{O}$ with respect to $\mathcal{E}$ is given by

$$\mathrm{Cap}_{\mathcal{E},1}(A) = \begin{cases} \inf_{f \in \mathcal{L}_A} \left[\mathcal{E}(f,f) + \|f\|^2_{L^2}\right], & \mathcal{L}_A \neq \emptyset \\ \infty, & \mathcal{L}_A = \emptyset \end{cases}. \tag{3.10}$$

If $A \subset X$ is an arbitrary subset, then put $\mathrm{Cap}_{\mathcal{E},1}(A) = \inf_{B \in \mathfrak{O}, A \subset B} \mathrm{Cap}_{\mathcal{E},1}(B)$. A statement is said to hold *quasi-everywhere* (*q.e.*) on A if and only if there exists a set $U \subset A$ with $\mathrm{Cap}_{\mathcal{E},1}(U) = 0$ such that the statement holds everywhere on $A \backslash U$. A function $f : X \to \mathbb{R}$ is said to be *quasi-continuous* if for every $\epsilon > 0$, there exists an open set Ω with $\mathrm{Cap}_{\mathcal{E},1}(\Omega) < \epsilon$ such that f is continuous on $X \backslash \Omega$. We say that v is a *quasi-continuous modification* of f if v is quasi-continuous and $v = f$ m-a.e, and denote v by $\widetilde{f}$.

A positive Radon measure μ on X is called a *measure of finite energy integral* (with respect to $\mathcal{E}$) if there exists a constant $C_\mu > 0$ such that for all $f \in \mathcal{F} \cap C_c(X)$,

$$\int_X |f| d\mu \leq C_\mu \left[\mathcal{E}(f,f) + \|f\|^2_{L^2}\right]^{1/2}. \tag{3.11}$$

We denote by S_0 the family of all measures of finite energy integral.

If furthermore $(\mathcal{E}, \mathcal{F})$ is transient, then one may complete $\mathcal{F}$ in the $\mathcal{E}$-norm, and

$(\mathcal{F}_e := \overline{F}^{\mathcal{E}}, \mathcal{E})$ is a Hilbert space called the *extended Dirichlet space*. Then we have the following 0-order counterparts of the above notions: the 0-*capacity* of a set $A \in \mathfrak{O}$, denoted by $\mathrm{Cap}_{\mathcal{E}}(A)$, is given by (3.10) with $\mathcal{F}$ and $\mathcal{E}(f,f) + \|f\|^2_{L^2}$ replaced respectively by $\mathcal{F}_e$ and $\mathcal{E}(f,f)$. The 0-capacity of an arbitrary set A then follows similarly. Likewise, a positive Radon measure μ on X is called a *measure of finite* 0*-order energy integral* if (3.11) holds with the same replacements. Denote by $S_0^{(0)}$ the family of all measures of finite 0-order energy integral.

There is an important connection between $S_0^{(0)}$ and $\mathcal{F}_e$, which is based on the Riesz representation theorem. For every $\mu \in S_0^{(0)}$, there exists a unique $U\mu \in \mathcal{F}_e$ such that $\mathcal{E}(f, U\mu) = \langle \widetilde{f}, \mu \rangle_X$ for all $f \in \mathcal{F}_e$. We shall refer to $U : S_0^{(0)} \to \mathcal{F}_e$ as the 0-order *potential operator* associated with $\mathcal{E}$. Any $h \in \mathcal{F}_e$ which can be written in the form $h = U\mu$ for some $\mu \in S_0^{(0)}$ is called a 0-order *potential* relative to $\mathcal{E}$.

Let us remark that $S_0^{(0)} \subset S_0 \subset S$, where S is the family of *smooth measures* consisting of all positive Borel measures μ on X such that:

- μ charges no set of zero 1-capacity.
- There exists an increasing sequence $(F_n)_n$ of closed sets such that $\mu(F_n) < \infty$ for all n, and that $\lim_{n\to\infty} \mathrm{Cap}_{\mathcal{E},1}(K \backslash F_n) = 0$ for any compact set K.

In general, elements of S_0 need not be absolutely continuous with respect to m, but each of them can be approximated by a sequence of absolutely continuous measures, *cf.* [32, Lemma 2.2.2]. Here we give the 0-order version of this statement.

Proposition 3.6.1. *Let $(\mathcal{E}, \mathcal{F})$ be a transient regular Dirichlet form on $L^2(X, m)$, and let G_β and U denote respectively the β-resolvent and the* 0*-order potential operator associated with $\mathcal{E}$. Given each $\mu \in S_0^{(0)}$, let $h_\beta := \beta(U\mu - \beta G_\beta(U\mu))$ for each $\beta \in \mathbb{N}$. Then as $\beta \to \infty$, $h_\beta \cdot m$ converges vaguely to μ.*

Proof. This is the Yosida approximation (*cf.* [32, (1.3.18)]): $h_\beta \geq 0$ m-a.e., and for all $f \in \mathcal{F}$,

$$(h_\beta, f)_{L^2(m)} = (\beta(U\mu - \beta G_\beta(U\mu)), f)_{L^2(m)} \underset{\beta\to\infty}{\longrightarrow} \mathcal{E}(U\mu, f).$$

Therefore $\lim_{\beta\to\infty}\langle f, h_\beta \cdot m\rangle_X = \langle f, \mu\rangle_X$ for all $f \in \mathcal{F} \cap C_c(X)$. ■

Last but not least, let us record several equivalent characterizations of the 0-capacity.

Proposition 3.6.2. *Let $(\mathcal{E}, \mathcal{F})$ be a transient regular Dirichlet form on $L^2(X, m)$. Fix an arbitrary set $B \subset X$ and suppose $\mathcal{L}_B \neq \emptyset$.*

(i) There exists a unique element e_B in $\mathcal{L}_B$ minimizing $\mathcal{E}(\cdot,\cdot)$. In particular, $\mathrm{Cap}_\mathcal{E}(B) = \mathcal{E}(e_B, e_B)$.

(ii) e_B is the unique element of $\mathcal{F}_e$ satisfying $\widetilde{e}_B = 1$ q.e. on B and $\mathcal{E}(e_B, f) \geq 0$ for any $f \in \mathcal{F}_e$ with $\widetilde{f} \geq 0$ q.e. on B.

(iii) There exists a unique measure $\mu_B \in S_0^{(0)}$ supported in B such that $e_B = U\mu_B$. In particular,

$$\mathrm{Cap}_\mathcal{E}(B) = \mathcal{E}(U\mu_B, U\mu_B) = \langle \widetilde{U\mu_B}, \mu_B\rangle_X.$$

(iv) If B is a compact set, then

$$\begin{aligned}\mathrm{Cap}_\mathcal{E}(B) \;=\; \langle \mathbb{1}_B, \mu_B\rangle_X \;&=\; \sup\left\{\mathcal{E}(U\mu, U\mu) : \mu \in S_0^{(0)},\ \mathrm{supp}(\mu) \subset B,\ \widetilde{U\mu} \leq 1 \text{ q.e.}\right\} \\ &=\; \sup\left\{\frac{\langle \mathbb{1}_B, \mu\rangle_X^2}{\mathcal{E}(U\mu, U\mu)} : \mu \in S_0^{(0)},\ \mathrm{supp}(\mu) \subset B\right\}.\end{aligned}$$

Proof. The first two items are the 0-order version of [32, Theorem 2.1.5], as explained on [32, p. 74]. Item (iii) is proved in conjunction with [32, Lemma 2.2.10]. The first two equalities in Item (iv) follow directly from (ii) and (iii), while the third equality can be obtained by a variational argument. ■

The function e_B and the measure μ_B are known as, respectively, the 0-order *equilibrium potential* and *equilibrium measure* of the set B (with respect to $\mathcal{E}$).

3.6.2 Γ-convergence and Mosco convergence

In this subsection we collect some elementary notions of Γ-convergence and Mosco convergence; see [23, 63] for more details.

Let $((\mathcal{E}_N, \mathcal{F}_N))_N$ be a sequence of symmetric quadratic forms on $\mathcal{H} = L^2(X, m)$, where for each N, $\mathcal{F}_N = \{f \in \mathcal{H} : \mathcal{E}_N(f, f) < \infty\}$ denotes the natural domain of $\mathcal{E}_N$. Also we shall fix a sequence $(P_N)_N$ of orthogonal projections $P_N : \mathcal{H} \to \mathcal{F}_N$. Finally, for each $f \in \mathcal{H}$, let $\mathcal{N}(f)$ be the collection of all open neighborhoods of f with respect to the usual topology on $\mathcal{H}$.

Define

$$(\Gamma\text{-}\varliminf_{N\to\infty} \mathcal{E}_N)(f, f) = \sup_{U\in\mathcal{N}(f)} \varliminf_{N\to\infty} \inf_{h\in U} \mathcal{E}_N(P_N h, P_N h), \tag{3.12}$$

$$(\Gamma\text{-}\varlimsup_{N\to\infty} \mathcal{E}_N)(f, f) = \sup_{U\in\mathcal{N}(f)} \varlimsup_{N\to\infty} \inf_{h\in U} \mathcal{E}_N(P_N h, P_N h). \tag{3.13}$$

If the liminf (3.12) coincides with the limsup (3.13) for all $f \in \mathcal{H}$, then we say that the sequence $(\mathcal{E}_N)_N$ **Γ-*converges***, and denote the limit by $\Gamma\text{-}\lim_{N\to\infty} \mathcal{E}_N$.

One may also consider Γ-convergence with respect to the *weak* topology on $\mathcal{H}$. In particular put

$$(\Gamma_w\text{-}\varliminf_{N\to\infty} \mathcal{E}_N)(f, f) = \sup_{U\in\mathcal{N}_w(f)} \varliminf_{N\to\infty} \inf_{h\in U} \mathcal{E}_N(P_N h, P_N h), \tag{3.14}$$

where $\mathcal{N}_w(f)$ denotes the collection of all open neighborhoods of f with respect to the weak topology on $\mathcal{H}$. If the weak liminf 3.14) coincides with the limsup (3.13), then we say that the sequence $(\mathcal{E}_N)_N$ ***converges in the sense of Mosco***, and denote the limit by $M\text{-}\lim_{N\to\infty} \mathcal{E}_N$.

The above notions of convergence can be alternatively characterized in terms of sequences in $\mathcal{H}$. It can be shown (*cf.* [23, Ch. 8]) that $\mathcal{E} = \Gamma\text{-}\lim_{N\to\infty} \mathcal{E}_N$ if and only if the following two conditions hold:

(Γ1) For every sequence $(u_N)_N$ in $\mathcal{H}$ converging to u, $\varliminf_{N\to\infty} \mathcal{E}_N(P_N u_N, P_N u_N) \geq \mathcal{E}(Pu, Pu)$.

(Γ2) For every $u \in \mathcal{H}$, there is a sequence $(u_N)_N$ in $\mathcal{H}$ converging to u such that $\varlimsup_{N\to\infty} \mathcal{E}_N(P_N u_N, P_N u_N) \leq \mathcal{E}(Pu, Pu)$.

Replace (Γ1) by the stronger condition

(wΓ1) For every sequence $(u_N)_N$ in $\mathcal{H}$ converging *weakly* to u, $\varliminf_{N\to\infty} \mathcal{E}_N(P_N u_N, P_N u_N) \geq \mathcal{E}(Pu, Pu)$.

Then $\mathcal{E} = M\text{-}\lim_{N\to\infty} \mathcal{E}_N$ if and only if (wΓ1) and (Γ2) hold. Here P is the orthogonal projection onto $\mathcal{D}(\mathcal{E})$.

The relevance of Mosco convergence to our setting is due to the following fact.

Proposition 3.6.3. *For each $\beta > 0$, let G_β^N and G_β respectively denote the β-resolvent associated with $\mathcal{E}_N$ and $\mathcal{E}$. Then the following are equivalent:*

(1) (Mosco convergence) $\mathcal{E} = M\text{-}\lim_{N\to\infty} \mathcal{E}_M$.

(2) (Strong resolvent convergence in L^2) For every $\beta > 0$ and every $f \in \mathcal{H}$, $\lim_{N\to\infty} G_\beta^N P_N f = G_\beta P f$ in L^2.

Proof. See, for example, [23, Theorem 13.6] or [63, Theorem 2.4.1]. ■

Let us now specialize to the case where $(\mathcal{E}_N)_N$ is a sequence of regular Dirichlet forms.

Lemma 3.6.4. *Let $((\mathcal{E}_N, \mathcal{F}_N))_N$ be a Mosco convergent sequence of regular Dirichlet forms on $\mathcal{H}$, and assume that the limit form can be extended to a regular Dirichlet form*

$(\mathcal{E}, \mathcal{F})$ *on* $\mathcal{H}$. *Then for all* $\beta > 0$ *and all* $h \in \mathcal{F}$ *with* $h \geq 0$ m*-a.e.,*

$$\lim_{N\to\infty} \mathcal{E}_N \left(G_\beta^N P_N h, G_\beta^N P_N h\right) = \mathcal{E}(U_\beta(h \cdot m), U_\beta(h \cdot m)). \tag{3.15}$$

Furthermore, if all the $((\mathcal{E}_N, \mathcal{F}_N))_N$ *and* $(\mathcal{E}, \mathcal{F})$ *are transient regular Dirichlet forms, then for all* $h \in \mathcal{F}_e \cap L^1(X, m)$ *with* $h \geq 0$ m*-a.e.,*

$$\lim_{N\to\infty} \mathcal{E}_N \left(G^N P_N h, G^N P_N h\right) = \mathcal{E}(U(h \cdot m), U(h \cdot m)). \tag{3.16}$$

Proof. By the identity

$$\mathcal{E}_N(G_\beta^N P_N f, P_N h) + \beta(G_\beta^N P_N f, P_N h)_{L^2} = (P_N f, P_N h)_{L^2} \quad \text{for all } f, h \in \mathcal{F}, \tag{3.17}$$

Proposition 3.6.3, and the conditions (wΓ1) and (Γ2), one finds that for every $\beta > 0$ and all $h \in \mathcal{F}$ with $h \geq 0$ m-a.e.,

$$\begin{aligned}
0 \;\leq\; & \overline{\lim_{N\to\infty}} \mathcal{E}_N \left(G_\beta^N P_N h - P_N U_\beta(h \cdot m),\; G_\beta^N P_N h - P_N U_\beta(h \cdot m)\right) \\
\leq\; & \overline{\lim_{N\to\infty}} \mathcal{E}_N \left(G_\beta^N P_N h, G_\beta^N P_N h\right) - 2 \varlimsup_{N\to\infty} \mathcal{E}_N \left(G_\beta^N P_N h, P_N U_\beta(h \cdot m)\right) \\
& + \overline{\lim_{N\to\infty}} \mathcal{E}_N(P_N U_\beta(h \cdot m), P_N U_\beta(h \cdot m)) \\
\leq\; & \overline{\lim_{N\to\infty}} \left[\left(G_\beta^N P_N h, P_N h\right)_{L^2} - \beta \left\|G_\beta^N P_N h\right\|_{L^2}^2\right] \\
& -2 \varlimsup_{N\to\infty} \left[\left(P_N U_\beta(h \cdot m), P_N h\right)_{L^2} - \beta \left(G_\beta^N P_N h, P_N U_\beta(h \cdot m)\right)_{L^2}\right] \\
& +\mathcal{E}(U_\beta(h \cdot m), U_\beta(h \cdot m)) \\
\leq\; & \left(G_\beta h, h\right)_{L^2} - \beta \left\|G_\beta h\right\|_{L^2}^2 - 2(U_\beta \mu, h)_{L^2} + 2\beta \left(G_\beta h, U_\beta(h \cdot m)\right)_{L^2} \\
& +\mathcal{E}(U_\beta(h \cdot m), U_\beta(h \cdot m)) \;=\; 0.
\end{aligned}$$

(In this case $Ph = h$.) To attain the last equality, notice first that since $(\mathcal{F}, \mathcal{E}_\beta)$ is a Hilbert space, and

$$\mathcal{E}_\beta \left(G_\beta h, f\right) = \langle \widetilde{f}, \mu \rangle_X = \mathcal{E}_\beta(U_\beta(h \cdot m), f) \qquad \text{for all } f, h \in \mathcal{F},\ h \geq 0 \ m\text{-a.e.},$$

deduce that $G_\beta h = U_\beta(h \cdot m)$. Then apply the identity (3.17) with $\mathcal{E}_N$ replaced by $\mathcal{E}$. The upshot is

$$\lim_{N\to\infty} \mathcal{E}_N \left(G_\beta^N P_N h, G_\beta^N P_N h\right) = \lim_{N\to\infty} \mathcal{E}_N(P_N U_\beta(h\cdot m), P_N U_\beta(h\cdot m)) = \mathcal{E}(U_\beta(h\cdot m), U_\beta(h\cdot m)).$$

Next, assuming the transience of the Dirichlet forms, we have that in the limit $\beta \downarrow 0$, $\left(G_\beta^N P_N h\right)$ $\mathcal{E}_N$-converges to $G^N P_N h$, and $(U_\beta(h \cdot m))$ $\mathcal{E}$-converges to $U(h \cdot m)$ by [32, Lemma 2.2.11]. A standard limiting argument yields (3.16). ■

CHAPTER 4

PERIODIC BILLIARD ORBITS OF SELF-SIMILAR SIERPINSKI CARPETS

This chapter reports the author's joint work with Robert G. Niemeyer [18].

4.1 Introduction

The subject of polygonal billiards is well-developed; see, e.g., [33, 34, 40, 44, 58, 72, 76, 77, 81] and the pertinent references therein for excellent surveys on the subject and the current state of the art. There is a particular family of fractal sets, each of which can be viewed as the limit of a sequence of polygonal approximations. M. L. Lapidus and R. G. Niemeyer have begun investigating the billiard dynamics on the Koch snowflake fractal billiard table and the so-called T-fractal billiard table; see [53–55]. Besides determining periodic orbits of these two fractal billiard tables, these articles proposed a possible framework in which one could begin investigating the billiard dynamics on an arbitrary billiard table with fractal boundary.

In this article, we proceed to identify a collection of periodic billiard orbits in the self-similar Sierpinski carpet billiard table $\Omega(S_{\mathbf{a}})$, where $\mathbf{a} := \{a_i\}_{i=1}^{\infty}$ is a constant sequence of positive odd integers with $a_i \geq 3$ for every $i \geq 1$; see Figure 4.1. To fix notations, we regard $S_{\mathbf{a}}$ as a subset of the unit square Q in $\mathbb{R}^2$, and fix the lower-left corner of Q to be the origin $(0, 0)$. The boundary of an open square removed in the construction of $S_{\mathbf{a}}$ is referred to as a *peripheral square* of $S_{\mathbf{a}}$.

The basis for our work is the recent result of [25] by E. Durand-Cartagena and J. Tyson, who identified all the possible slopes of nontrivial line segments in a given Sierpinski carpet $S_{\mathbf{a}}$. Up to isometries of the square, this set of slopes, denoted by $\text{Slopes}(S_{\mathbf{a}})$, is a finite set of rational values, and equals the disjoint union of $A_{\mathbf{a}}$ and $B_{\mathbf{a}}$,

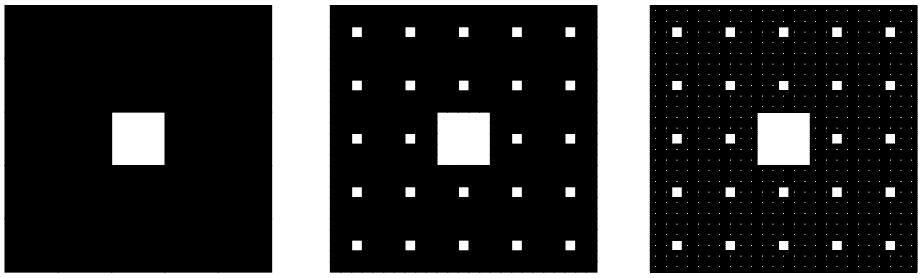


Figure 4.1: The first three approximations of the self-similar Sierpinski carpet $S_{\mathbf{5}}$. The explicit construction of a Sierpinski carpet is given in §4.2.1.

where (see Theorem 4.1 of [25])

$$A_{\mathbf{a}} = \left\{ \frac{p}{q} \ : \ p+q \le a,\ 0 \le p < q \le a-1,\ p,q \in \mathbb{N} \cup \{0\},\ p+q \text{ is odd} \right\},$$
$$B_{\mathbf{a}} = \left\{ \frac{p}{q} \ : \ p+q \le a-1,\ 0 \le p \le q \le a-2,\ p,q \in \mathbb{N},\ p \text{ and } q \text{ are odd} \right\}.$$

We note that $A_{\mathbf{a-2}} \subsetneq A_{\mathbf{a}}$ and $B_{\mathbf{a-2}} \subsetneq B_{\mathbf{a}}$. Roughly speaking, a nontrivial line segment of slope $\alpha \in A_{\mathbf{a}}$ can emanate from the origin $(0,0)$, and a nontrivial line segment of slope $\alpha \in B_{\mathbf{a}}$ can emanate from $(\frac{1}{2},0)$, although the association between the slope set and starting point is not absolute. For example, in $S_{\mathbf{7}}$, the line segments emanating from, respectively, $(0,0)$ and $(\frac{1}{2},0)$, with slope $\frac{2}{3} \in A_{\mathbf{7}}$ are both nontrivial line segments; see the examples in Items (1)–(4) in §4.3 and the detailed discussion of such examples in 4.4.

Key to our description of particular family of periodic orbits of a self-similar Sierpinski carpet is analyzing whether nontrivial line segments hit or avoid peripheral squares of $S_{\mathbf{a}}$. It has already been shown in [25] that nontrivial line segments of slope $\alpha \in B_{\mathbf{a}}$ starting from $(\frac{1}{2},0)$ avoid all peripheral squares of $S_{\mathbf{a}}$. We show below in Theorem 4.3.4 an important distinction between nontrivial line segments of slope $\alpha \in A_{\mathbf{a-2}}$ and those of slope $\alpha \in A_{\mathbf{a}} \backslash A_{\mathbf{a-2}}$. Our result is as follows. Suppose that the line segment starts from $(0,0)$. If the segment has slope $\alpha \in A_{\mathbf{a-2}}$, then it avoids all peripheral squares of $S_{\mathbf{a}}$. If the segment has slope $\alpha \in A_{\mathbf{a}} \backslash A_{\mathbf{a-2}}$, then it must intersect with the corner of some

peripheral square of $S_{\mathbf{a}}$. This distinction was not stated in Theorem 4.1 of [25], and turns out to play an essential role in identifying particular periodic billiard orbits of a self-similar Sierpinski carpet billiard.

We then apply the results about nontrivial line segments to obtain information about the billiard orbits. This is achieved via a sequence of arguments regarding the scaling of cells and reflected-unfolding of particular orbits. We view a Sierpinski carpet as the unique fixed point attractor of a sequence of rational polygonal approximations (or prefractal approximations). As such, we investigate the limiting behavior of particular sequences of compatible periodic orbits of prefractal billiard tables.[1] Our main result, stated more precisely in Theorem 4.4.2, says that the following orbits in $\Omega(S_{\mathbf{a}})$ are periodic:

- Those starting from $(0, 0)$ with slope $\alpha \in A_{\mathbf{a-2}}$.
- Those starting from $(\frac{r}{2a^n}, 0)$ with slope $\alpha \in B_{\mathbf{a}}$, for any $n \in \mathbb{N}$ and any positive odd integer $r < 2a^n$.

A priori, this list does not exhaust all of the initial conditions giving rise to periodic orbits of $\Omega(S_{\mathbf{a}})$. In addition to describing particular periodic orbits, we can understand what constitutes the larger collection of closed orbits of $\Omega(S_{\mathbf{a}})$. In Proposition 4.4.3, we show that an orbit starting from $(\frac{k}{a^n}, 0)$ with slope $\alpha \in A_{\mathbf{a-2}}$, some $n \in \mathbb{N}$, and some positive integer $k < a^n$ is always closed, but not always periodic.

We believe that investigating the self-similar Sierpinski carpet billiard table may inspire further studies on related billiard tables with corresponding translation surfaces of infinite genus. Indeed, while we can identify the translation surface of the pre-Sierpinski carpet billiard table $S_{\mathbf{a},n}$, and show that their genera grow unboundedly as $n \to \infty$ (see

[1] We present the necessary concepts and definitions from the subject of fractal billiards in §4.2.3.

§4.4.1), it is unclear how one may define a translation surface for the fractal billiard table $\Omega(S_{\mathbf{a}})$, since $S_{\mathbf{a}}$ has zero Lebesgue area. In §4.5, we propose to investigate two related billiard tables which share similar features with $S_{\mathbf{a}}$, but have nonzero Lebesgue area.

This chapter is organized as follows: In §4.2, we provide some background on the Sierpinski carpet, mathematical billiards, and translation surfaces. Our result on non-trivial line segments in $S_{\mathbf{a}}$, which clarifies part of Theorem 4.1 in [25], is described in §4.3. The core of this article is contained in §4.4, where we identify the two types of periodic billiard orbits in $S_{\mathbf{a}}$, and describe the translation surface corresponding to the pre-Sierpinski carpet billiard table. Open problems are addressed in §4.5.

4.2 Background

In this section, we provide the necessary background for understanding the remainder of the article. We will give the necessary definitions from the field of fractal geometry and mathematical billiards, as well as the pertinent terms from [25]. In addition to this, we will draw upon the vocabulary of the emerging field of fractal billiards when discussing a self-similar Sierpinski carpet billiard $\Omega(S_{\mathbf{a}})$ in §4.4; see [53–55].

4.2.1 Self-similarity and Sierpinski carpets

Definition 4.2.1 (A self-similar Sierpinski carpet via an IFS)**.** Let $a \geq 3$ be a positive odd integer and $\phi_i : \mathbb{R}^2 \to \mathbb{R}^2$ a similarity contraction defined as:

$$\phi_i(x) := \frac{1}{a}(x + (u_i, v_i)), \tag{4.1}$$

where $u_i, v_i \in \mathbb{N} \cup \{0\}$, $0 \leq u_i \leq a-1$, $0 \leq v_i \leq a-1$ and $(u_i, v_i) \neq (\frac{a-1}{2}, \frac{a-1}{2})$. Then, $\{\phi_i\}_{i=1}^{a^2-1}$ is an iterated function system. The unique fixed point attractor of the map $\Phi(\cdot) := \bigcup_{i=1}^{a^2-1} \phi_i(\cdot)$ is called a *self-similar Sierpinski carpet*.

We recall the notation used in constructing Sierpinski carpets in [25]. Let

$$\mathbf{a} = (a_1^{-1}, a_2^{-1}, ...) \in \left\{\frac{1}{3}, \frac{1}{5}, \frac{1}{7}, ...\right\}^{\mathbb{N}}.$$

Let Q be the unit square. Beginning with $S_0 := Q$, partition S_0 into a_1^2 equal squares of side-length a_1^{-1} and remove the middle open square, leaving $a_1^2 - 1$ many squares. We denote the union of the remaining squares by $S_{\mathbf{a},1}$ and refer to $S_{\mathbf{a},1}$ as the first level approximation of the Sierpinski carpet $S_{\mathbf{a}}$. Then, partition each remaining square with side-length a_1^{-1} into a_2^2 many squares with side-length $a_1^{-1} \cdot a_2^{-1}$. From each square with side-length a_1^{-1}, remove the middle open square of side-length $a_1^{-1} \cdot a_2^{-1}$. The union of the remaining squares of side-length $a_1^{-1} \cdot a_2^{-1}$ constitutes the second level approximation $S_{\mathbf{a},2}$ of a Sierpinski carpet $S_{\mathbf{a}}$. Continuing this process ad infinitum, we construct the Sierpinski carpet $S_{\mathbf{a}}$. More precisely,

$$S_{\mathbf{a}} = \bigcap_{n=0}^{\infty} S_{\mathbf{a},n}, \tag{4.2}$$

where $S_{\mathbf{a},0} = S_0$.

Definition 4.2.2 (A self-similar Sierpinski carpet)**.** If $\mathbf{a} = \{a_i^{-1}\}_{i=0}^{\infty}$, with $a_i = 2k_i + 1$ and $k_i \in \mathbb{N}$, is a periodic sequence, then the Sierpinski carpet $S_{\mathbf{a}}$ is called a *self-similar Sierpinski carpet*.

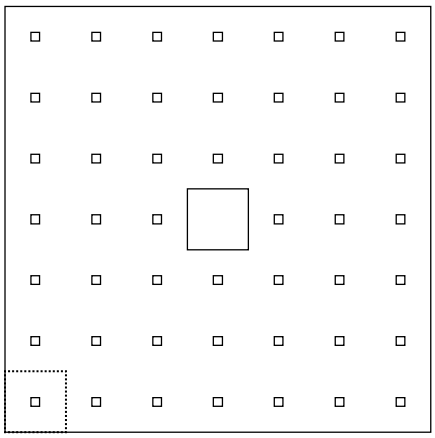


Figure 4.2: We see here an example of a cell $C_{1,7}$ of $S_{\mathbf{7},1}$. While this cell is clearly shown as a subset of $S_{\mathbf{7},2}$, it has a side-length that indicates it is called a cell of $S_{\mathbf{7},1}$.

Both Definitions 4.2.1 and 4.2.2 are equivalent. More precisely, when **a** is a constant sequence, then it is clear how the two definitions are equivalent. When **a** is a nonconstant periodic sequence, then there exists a' such that $S_{\mathbf{a}}$ is the unique fixed point attractor of an appropriately defined iterated function system $\{\phi_i\}_{i=1}^{a'^2-1}$. For the purposes of this chapter, it may be advantageous to use one definition over another, depending on the context of the situation.[2]

Definition 4.2.3 (A cell of $S_{\mathbf{a},n}$)**.** Let $a_0 = 2k_0 + 1$, $k_0 \in \mathbb{N}$. Consider a partition of the unit square $Q = S_0$ into a_0^2 many squares of side-length a_0^{-1}. A subsquare of the partition is called a *cell of* S_0 and is denoted by C_{0,a_0}. Let $S_{\mathbf{a}}$ be a Sierpinski carpet and $n \geq 0$. Consider a partition of the prefractal approximation $S_{\mathbf{a},n}$ into subsquares with side-length $(a_0 \cdot a_1 \cdots a_n)^{-1}$. A subsquare of the partition of $S_{\mathbf{a},n}$ is called a *cell of* $S_{\mathbf{a},n}$ and is denoted by $C_{n,a_0a_1\cdots a_n}$ and has side-length $(a_0 \cdot a_1 \cdots a_n)^{-1}$.

Let **a** be a constant sequence of odd, positive integers such that $S_{\mathbf{a}}$ is a self-similar

[2]For a more detailed discussion of self-similar sets and iterate function systems, see [41]. For a more complete treatment of the subject of fractal geometry, see [27].

Sierpinski carpet and $\{\phi_i\}_{i=1}^{a^2-1}$ an iterated function system for which $S_{\mathbf{a}}$ is the unique fixed point attractor. If C_{n,a^n} is a cell of $S_{\mathbf{a},n}$, then $C_{n,a^n} = \phi_{i_n} \circ \cdots \circ \phi_{i_1}(Q)$ for suitably chosen $i_1, i_2, ..., i_n$.

We now define the important notion of a *peripheral square*.

Definition 4.2.4 (Peripheral square)**.** In accordance with the convention established in [25], the boundary of an open square removed in the construction of $S_{\mathbf{a}}$ is called a *peripheral square* of $S_{\mathbf{a}}$. By convention, the unit square $Q = S_0$ is not a peripheral square.

Example 4.2.5 (A cell of the prefractal $S_{\mathbf{7},1}$)**.** In Figure 4.2, we see an example of a cell of the prefractal approximation $S_{\mathbf{7},1}$. We wish to emphasize the fact that the cell with the middle open square removed constitutes the scaling of $S_{\mathbf{7},1}$ (relative to the origin) by $\frac{1}{7}$. At times, we may make reference to a cell C_{m,a^m}, even when such a region is considered as a subset of $S_{\mathbf{a},n}$ for $n > m$. In such a case, a cell C_{m,a^m} will contain peripheral squares of $S_{\mathbf{a},n}$ with side length a^{-k} for all k such that $m < k \leq n$.

Also of great importance is the notion of a *nontrivial line segment*.

Definition 4.2.6 (Nontrivial line segment of $S_{\mathbf{a}}$)**.** A *nontrivial line segment of a Sierpinski carpet* $S_{\mathbf{a}}$ is a (straight-line) segment of the plane contained in $S_{\mathbf{a}}$ that has nonzero length.

Notation 4.2.7. Let $S_{\mathbf{a}}$ be a Sierpinski carpet. We denote by Slope($S_{\mathbf{a}}$) the set of slopes, with values in $[0, 1]$, of nontrivial line segments in $S_{\mathbf{a}}$. There is no loss of generality, since applying an isometry of the square to a slope $\alpha \in$ Slope($S_{\mathbf{a}}$) yields a slope of $\frac{1}{\alpha}$, $-\alpha$, or $-\frac{1}{\alpha}$.

Theorem 4.1 of [Du-CaTy]. *Let* $\mathbf{a} = (\frac{1}{a}, \frac{1}{a}, \frac{1}{a}, ...)$ *be a constant sequence. Then the set of slopes of nontrivial line segments Slope*($S_{\mathbf{a}}$) *is the union of the following two sets*:

$$A_{\mathbf{a}} = \left\{ \frac{p}{q} \ : \ p+q \le a,\ 0 \le p < q \le a-1,\ p,q \in \mathbb{N} \cup \{0\},\ p+q \text{ is odd} \right\}, \tag{4.3}$$

$$B_{\mathbf{a}} = \left\{ \frac{p}{q} \ : \ p+q \le a-1,\ 0 \le p \le q \le a-2,\ p,q \in \mathbb{N},\ p \text{ and } q \text{ are odd} \right\}. \tag{4.4}$$

Moreover, if $\alpha \in A_{\mathbf{a}}$, then each nontrivial line segment in $S_{\mathbf{a}}$ with slope α touches vertices of peripheral squares, while if $\alpha \in B_{\mathbf{a}}$, then each nontrivial line segment in $S_{\mathbf{a}}$ with slope α is disjoint from all peripheral squares. For each $\alpha \in A_{\mathbf{a}} \cup B_{\mathbf{a}}$, there exist maximal line segments in $S_{\mathbf{a}}$ with slope α. Finally, if $b < a$, then any maximal nontrivial line segment in $S_{\mathbf{b}}$ is also contained in $S_{\mathbf{a}}$. In particular, Slope$(S_{\mathbf{b}}) \subseteq$ Slope$(S_{\mathbf{a}})$.

Observe that $A_{\mathbf{a-2}} \subsetneq A_{\mathbf{a}}$ and $B_{\mathbf{a-2}} \subsetneq B_{\mathbf{a}}$. For concreteness, we give Slope$(S_{\mathbf{a}})$ for the first four self-similar Sierpinski carpets:

$$A_{\mathbf{3}} = \left\{ \mathbf{0}, \mathbf{\frac{1}{2}} \right\} \quad , \quad B_{\mathbf{3}} = \{\mathbf{1}\},$$

$$A_{\mathbf{5}} = \left\{ 0, \mathbf{\frac{1}{4}}, \frac{1}{2}, \mathbf{\frac{2}{3}} \right\} \quad , \quad B_{\mathbf{5}} = \left\{ \mathbf{\frac{1}{3}}, 1 \right\},$$

$$A_{\mathbf{7}} = \left\{ 0, \mathbf{\frac{1}{6}}, \frac{1}{4}, \mathbf{\frac{2}{5}}, \frac{1}{2}, \frac{2}{3}, \mathbf{\frac{3}{4}} \right\} \quad , \quad B_{\mathbf{7}} = \left\{ \mathbf{\frac{1}{5}}, \frac{1}{3}, 1 \right\},$$

$$A_{\mathbf{9}} = \left\{ 0, \mathbf{\frac{1}{8}}, \frac{1}{6}, \frac{1}{4}, \mathbf{\frac{2}{7}}, \frac{2}{5}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \mathbf{\frac{4}{5}} \right\} \quad , \quad B_{\mathbf{9}} = \left\{ \mathbf{\frac{1}{7}}, \frac{1}{5}, \frac{1}{3}, \mathbf{\frac{3}{5}}, 1 \right\}.$$

We have presented the slopes in color in order to represent those in, respectively, $A_{\mathbf{a-2}}$, $\mathbf{A_a} \setminus \mathbf{A_{a-2}}$, $B_{\mathbf{a-2}}$, and $\mathbf{B_a} \setminus \mathbf{B_{a-2}}$. Note that the elements of $A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$ (resp., $B_{\mathbf{a}} \setminus B_{\mathbf{a-2}}$) are bolded while the elements of $A_{\mathbf{a-2}}$ (resp., $B_{\mathbf{a-2}}$) are not bolded. For example, the elements $\frac{1}{4}$, $\frac{1}{3}$ and $\frac{2}{3}$ in Slope$(S_{\mathbf{5}})$ listed above are bolded, while the elements 0, $\frac{1}{2}$ and 1 in Slope$(S_{\mathbf{5}})$ are not. The distinction between slopes in $A_{\mathbf{a-2}}$ and slopes in $A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$ will be crucial to our refinement of Theorem 4.1 of [25] stated below as Theorem 4.3.4,

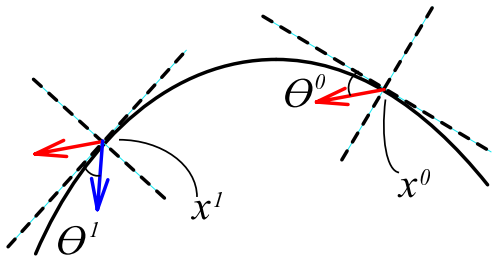


Figure 4.3: A demonstration of how a billiard ball reflects in the sufficiently smooth boundary of a billiard table.

as well as our main results concerning the billiard $\Omega(S_{\mathbf{a}})$ stated in Theorem 4.4.2 and Proposition 4.4.3.

4.2.2 Mathematical billiards

Consider a point mass making a collision in a boundary subject to the Law of Reflection. That is, the angle of reflection is equal to the angle of incidence. Alternatively (and equivalently), one may consider the incoming vector reflected through the tangent at the point of collision; see Figure 4.3. Furthermore, we require, prior to colliding in the boundary, that the point mass be traveling in a straight line. For the remainder of the article, we assume that a mathematical billiard table $\Omega(D)$ with boundary D is a path-connected region in the plane with sufficiently piecewise smooth boundary[3] that is also simple and connected; see Figure 4.4 for an example of such a billiard table.

An initial condition of the billiard flow is given by (x^0, θ^0), where we take x^0 to be on the boundary D of the billiard table $\Omega(D)$ and θ^0 to be an angle measured relative to the tangent line at x^0. The next point of collision x^1 in the boundary is determined by the billiard map f_D. That is, $f_D(x^0, \theta^0) = (x^1, \theta^1)$, where θ^1 is an inward pointing

[3]In the case of a polygonal billiard table, there are finitely many vertices where a well-defined tangent does not exist.

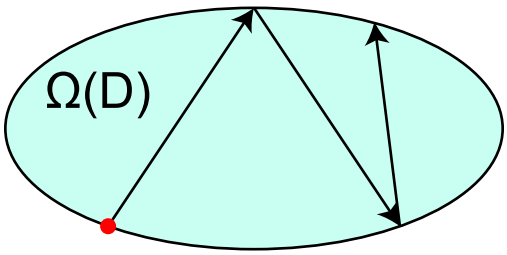


Figure 4.4: The billiard table with a boundary that is an ellipse constitutes an example of a mathematical billiard.

vector based at x^1. Then, successive iterates f^k, $k \geq 1$, of the billiard map f determine the collision point x^k of the billiard ball in the boundary D of the billiard table. To be clear, the angle of reflection θ^k, determined from the Law of Reflection, is the angle made by the vector based at x^k pointing inward towards the interior of $\Omega(D)$, measured relative to the tangent based at x^k. Then, one can describe an equivalence relation $\sim$ on $\Omega(D) \times S^1$ that amounts to identifying the outward pointing vector based at a point x^k with the inward pointing vector based at x^k.[4]

Then $f_D([(x^0, \theta^0)]) = [(x^1, \theta^1)]$, where θ^1 is the angle made by the inward pointing vector, measured relative to the tangent at x^1. For now, we shall denote by (x, θ) the equivalence class of $[(x, \theta)]$, relative to the equivalence relation $\sim$. That is, (x, θ), such that θ is the angle determined from the inward pointing vector, is the representative element of the equivalence class $[(x, \theta)]$. In order to reduce the complexity of the phase space, we only consider the space $(D \times S^1)/\sim$.

Definition 4.2.8 (An orbit of $\Omega(D)$)**.** Let f_D be the billiard map describing the discrete billiard flow in the phase space $(\Omega(D) \times S^1)/\sim$. An orbit $\mathcal{O}(x^0, \theta^0)$ of $\Omega(D)$ is then given by $\{f_D^n(x^0, \theta^0)\}_{n=0}^{\infty}$. Equivalently, an orbit may be written as $\{f_D^n(x^0, \theta^0)\}_{n=0}^{-\infty}$.

[4]For a more complete treatment of the billiard map f_D, the billiard flow ϕ^t in the phase space and an equivalence relation placed on $\Omega(D) \times S^1$, see [72].

We frequently make an abuse of notation and say that an orbit is the path traversed by the billiard ball connecting the base points x^i of $f_D^i(x^0, \theta^0)$, $0 \leq i \leq N$.

Definition 4.2.9 (A closed orbit of $\Omega(D)$)**.** An orbit $\mathscr{O}(x^0, \theta^0)$ of $\Omega(D)$ is said to be *closed* if the orbit consists of finitely many elements.

Definition 4.2.10 (A periodic orbit of $\Omega(D)$)**.** Let $\mathscr{O}(x^0, \theta^0)$ be a closed orbit of $\Omega(D)$. If there exists a least integer $m \geq 1$ such that $f^m(x^0, \theta^0) = (x^0, \theta^0)$, then we say the orbit $\mathscr{O}(x^0, \theta^0)$ is a *periodic* orbit.

Remark 4.2.11. One does not call an orbit *closed* simply because one decides to arbitrarily terminate the billiard flow. If, under the billiard flow, the billiard ball intersects a point of the boundary for which reflection cannot be determined in a well-defined manner, then the billiard ball trajectory terminates at that point. Such an orbit is then called *singular*. If $\mathscr{O}(x^0, \theta^0)$ is a singular orbit and there exists $k \in \mathbb{N}$ such that the base point x^{-k} of $f^{-k}(x^0, \theta^0)$ is a point of D not admitting a well defined derivative, then we say $\mathscr{O}(x^0, \theta^0)$ forms a *saddle connection*. In the context of this chapter, $\mathscr{O}(x^0, \theta^0)$ is also a closed orbit.

Definition 4.2.12 (A dense orbit of $\Omega(D)$)**.** If the path traversed by a billiard ball is dense in $\Omega(D)$ (forward or backward in time), then the orbit $\mathscr{O}(x^0, \theta^0)$ is said to be dense in $\Omega(D)$.

Definition 4.2.13 (Rational polygonal billiard)**.** A *rational polygon* is a polygon where each interior angle v_i is of the form $\frac{p_i}{q_i}\pi$, $p_i, q_i \in \mathbb{Z}$, $q_i \neq 0$. A *rational polygonal billiard* is a billiard table with boundary D, where D is a rational polygon.

Translation surfaces via rational polygonal billiard tables

Consider a rational polygonal billiard $\Omega(D)$ with interior angles $\frac{p_1}{q_1}\pi, ..., \frac{p_k}{q_k}\pi$. If $N = \text{lcm}(q_1, ..., q_k)$, then, by appropriately identifying the sides of $2N$ many copies of $\Omega(D)$,

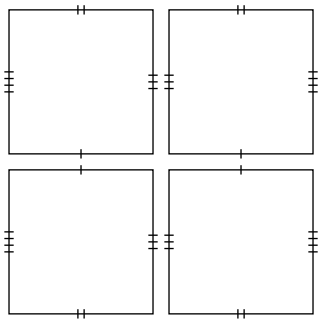


Figure 4.5: A translation surface constructed from the unit square billiard $\Omega(R)$. Opposite and parallel sides are identified.

we can construct a compact, connected orientable surface with finitely many conic singularities, with the coordinate changing functions given by translations.[5] In addition, if the conic angle about a conic singularity is 2π, then the singularity is called a *removable singularity*; if the conic angle is greater than 2π, then the singularity is called a *non-removable singularity*.

Consider the unit square Q. Then, $N = \text{lcm}(2, 2, 2, 2) = 2$. In Figure 4.5, we indicate how the sides of $2N = 4$ many copies of $\Omega(Q)$ must be identified so as to produce a translation surface. In this example, each corner is identified with three other corners to produce a singularity of the translation surface. In each case, such a singularity has a conic angle of 2π. This implies that the billiard flow at each vertex can be made well defined. We note that, in this case of the square, the translation surface is topologically equivalent to the flat torus.

If $\{u_1, u_2\}$ is a basis for $\mathbb{R}^2$, then a vector $z \in \mathbb{R}^2$ is called *rational with respect to* $\{u_1, u_2\}$ if $z = mu_1 + nu_2$, for some $m, n \in \mathbb{Z}$. Combining the results of [35] with Theorem 3 of [34], we can state the following result, which we do not claim as a new theorem,

[5] For a more complete definition of translation surface as well as descriptions of closely related types of surfaces, please see [34, 40, 58, 81] and the relevant references therein.

but which we rephrase in a way that is suitable for our purposes.[6]

Theorem 4.2.14 ([34])**.** *Let $\mathcal{S}(D)$ be a translation surface determined from a rational polygonal billiard $\Omega(D)$. If $\mathcal{S}(D)$ is a branched cover of a singly punctured torus, then a geodesic on $\mathcal{S}(D)$ is periodic or forms a saddle connection if and only if the geodesic has a direction that is rational. In addition, a geodesic on $\mathcal{S}(D)$ is dense if and only if the geodesic has a direction that is irrational.*

Theorem 4.2.14 is equivalent to the following, similarly attributed to E. Gutkin.

Theorem 4.2.15 ([34])**.** *Let $\Omega(D)$ be a rational billiard table that is tiled by an integrable billiard[7] table $\Omega(P)$. Then an orbit on $\Omega(D)$ is closed if and only if the orbit has an initial direction that is rational. In addition, an orbit on $\mathcal{S}(D)$ is dense if and only if the orbit has an initial direction that is irrational.*

If $\Omega(D)$ is a rational polygon with k many sides and interior angles $\frac{p_i}{q_i}\pi$, $i \leq k$ and $N = \text{lcm}\{q_1, q_2, ..., q_k\}$, then the genus g of a translation surface constructed from a rational polygonal billiard $\Omega(D)$ is given by[8]

$$g = 1 + \frac{N}{2}\left(k - 2 - \sum_{i=1}^{k} \frac{1}{n_i}\right). \tag{4.5}$$

Unfolding a billiard orbit and equivalence of flows

Consider a rational polygonal billiard $\Omega(D)$ and an orbit $\mathcal{O}(x^0, \theta^0)$ of $\Omega(D)$. Reflecting the billiard $\Omega(D)$ and the orbit in the side of the billiard table containing the base point x^1

[6] This same statement is also given in [54] and is similarly attributed to E. Gutkin.

[7] If $\Omega(D)$ is a billiard table that tiles the plane, then $\Omega(D)$ is called an *integrable billiard table*.

[8] See [58] for an explanation of the formula describing the genus of a translation surface determined from a rational billiard table $\Omega(D)$.

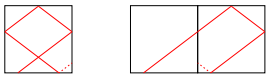


Figure 4.6: Partially unfolding an orbit of the square billiard $\Omega(Q)$.

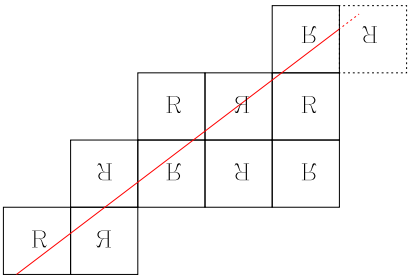


Figure 4.7: Unfolding an orbit of the square billiard $\Omega(Q)$.

of the orbit (x^1 being the first point of collision after starting from x^0) partially unfolds the orbit $\mathscr{O}(x^0, \theta^0)$; see Figure 4.6 for the case of the square billiard. Continuing this process until the orbit is a straight line produces as many copies of the billiard table as there are base points of the orbit; see Figure 4.7. That is, if the period of an orbit $\mathscr{O}(x^0, \theta^0)$ is some positive integer p, then the number of copies of the billiard table in the unfolding is also p. We refer to such a straight line as the *unfolding of the billiard orbit*.

We note that the construction of a translation surface from a rational billiard $\Omega(D)$ amounts to letting a group of symmetries act on $\Omega(D)$; see §4.2.2. That is, a rational billiard $\Omega(D)$ can be acted on by a dihedral group $\mathscr{D}_N$ to produce a translation surface in a way that is similar to unfolding the billiard table. Hence, we can quickly see how the billiard flow is dynamically equivalent to the geodesic flow on the associated translation surface; see Figure 4.8 and the corresponding caption.

One may modify the notion of "reflecting" so as to determine orbits of a billiard table tiled by a rational polygonal billiard $\Omega(D)$. As an example, we consider the unit-square billiard table. An appropriately scaled copy of the unit-square billiard table can be tiled

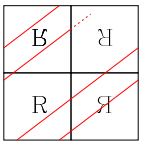


Figure 4.8: Rearranging the unfolded copies of the unit square from Figure 4.7 and correctly identifying sides so as to recover the flat torus, we see that the unfolded orbit corresponds to a closed geodesic on the translation surface.

by the unit-square billiard table by making successive reflections in the sides of the unit square. One may then unfold an orbit of the unit-square billiard table into a larger square billiard table. When the unfolded orbit of the original unit-square billiard intersects the boundary of the appropriately scaled (and larger) square, then one continues unfolding the billiard orbit in the direction determined by the Law of Reflection.[9] We will refer to such an unfolding as a *reflected-unfolding*.

We may continue this process in order to form an orbit of a larger scaled square billiard table. Suppose that an orbit $\mathcal{O}(x^0, \theta^0)$ has period p. If s is a positive integer, then we say that $\mathcal{O}^s(x^0, \theta^0)$ is an orbit that traverses the same path as $\mathcal{O}(x^0, \theta^0)$ s-many times. For sufficiently large $s \in \mathbb{N}$, an orbit that traverses the same path as an orbit $\mathcal{O}(x^0, \theta^0)$ s-many times can be reflected-unfolded in an appropriately scaled square billiard table to form an orbit of the larger billiard table; see Figure 4.9. Such a tool is useful in understanding the relationship between the billiard flow on a rational polygonal billiard $\Omega(D)$ and a billiard table tiled by $\Omega(D)$, and will be particularly useful in understanding the material presented in the sequel.

[9] That is, assuming the unfolded orbit is long enough to reach a side of the larger square.

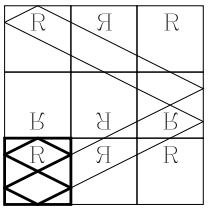


Figure 4.9: Unfolding the orbit of the unit-square billiard in a (larger) scaled copy of the unit-square billiard. This constitutes an example of a reflected-unfolding of an orbit. The edges of the original unit-square billiard table and the segments comprising the orbit have been thickened to provide the reader with a clearer view of what is the orbit of the unit square and what constitutes the reflected-unfolding of the orbit of the larger square.

The billiard table $\Omega(S_{\mathbf{a},n})$

We note that $S_{\mathbf{a},n}$ does not have a connected boundary. However, the boundary of S_0 and the boundary of each omitted square is connected. In addition, the notation "$\Omega(D)$" indicates that D is the boundary of the billiard table. To be technically correct, $\Omega(\partial S_{\mathbf{a},n})$ is the proper way of referring to the prefractal billiard table. In order to simplify notation, we refer to the prefractal billiard table by $\Omega(S_{\mathbf{a},n})$.

Clearly, each prefractal billiard $\Omega(S_{\mathbf{a},n})$ can be interpreted as a square billiard with open subsquares removed. In fact, every prefractal billiard $\Omega(S_{\mathbf{a},n})$ is a rational billiard table.

Notation 4.2.16. Due to the fact that Theorem 4.1 in [25] refers to the slope of a non-trivial line segment and we frequently refer to the main result of [25], we will denote the initial condition (x_n^0, θ_n^0) of an orbit of $\Omega(S_{\mathbf{a},n})$ by (x_n^0, α_n^0), where $\alpha_n^0 = \tan(\theta_n^0)$.

Definition 4.2.17 (An orbit of the cell C_{k,a^k} of $\Omega(S_{\mathbf{a},k})$)**.** Consider the boundary of a cell C_{k,a^k} of $\Omega(S_{\mathbf{a},k})$ as a barrier to the billiard flow.[10] Then an orbit determined by reflecting

[10]Here, C_{k,a^k} is a cell of the kth prefractal approximation $S_{\mathbf{a},k}$, as given in Definition 4.2.3 with each a_j equal to a.

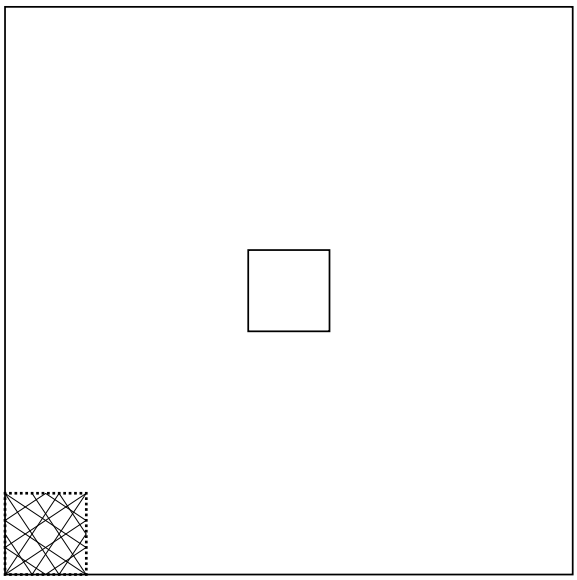


Figure 4.10: An example of an orbit of a cell of $S_{\mathbf{7},1}$.

in the boundary of the cell and with an initial condition contained in the cell is called an *orbit of the cell* C_{k,a^k} *of* $\Omega(S_{\mathbf{a},k})$.

Example 4.2.18. In Figure 4.2, we saw an example of a cell of $S_{\mathbf{7},1}$. In Figure 4.10, we see an example of an orbit of the same cell. The orbit of the cell shown has an initial condition of $((0,0), \frac{2}{3})$.

Remark 4.2.19. So as to be clear, the boundary of the cell C_{k,a^k} does not form a barrier to the billiard flow on $\Omega(S_{\mathbf{a},k})$. Rather, we are treating the cell C_{k,a^k} as a billiard table in its own right, embedded in the larger prefractal approximation $\Omega(S_{\mathbf{a},k})$. Our motivation for doing so is found in the fact that we can proceed to reflect-unfold an orbit of a cell in $\Omega(S_{\mathbf{a},k})$.

Recall from Definition 4.2.1 that a self-similar Sierpinski carpet $S_{\mathbf{a}}$ is the unique fixed point attractor of a suitably chosen iterated function system $\{\phi_j\}_{j=1}^{a^2-1}$ consisting of similarity contractions. In light of this, an orbit of a cell C_{k,a^k} of $\Omega(S_{\mathbf{a},k})$ is the image of an orbit $\mathcal{O}_0(x_0^0, \alpha_0^0)$ of the unit-square billiard $\Omega(S_0)$ under the action of a composition of

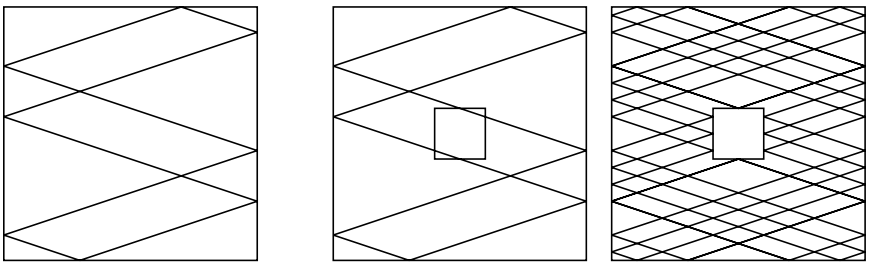


Figure 4.11: In the first image on the left, we see an orbit of $\Omega(S_0)$ that has an initial condition of $((\frac{3}{10}, 0), \frac{1}{3})$. In the second image, we see that the same orbit would intersect the omitted square in the first level approximation of $\Omega(S_{\mathbf{5}})$. The third image is an orbit of $\Omega(S_{\mathbf{5},1})$ with the same initial condition as the orbit shown in the first image.

contraction mappings $\phi_{m_k} \circ \cdots \circ \phi_{m_1}$, with $1 \leq m_i \leq a^2 - 1$ and $1 \leq i \leq k$, determined from the iterated function system $\{\phi_j\}_{j=1}^{a^2-1}$ for which $S_{\mathbf{a}}$ is the unique fixed point attractor.

4.2.3 Fractal billiards

We recall various definitions from [54]. So that the reader may find the definitions more accessible, we phrase the following in terms of a self-similar Sierpinski carpet $S_{\mathbf{a}}$ and a prefractal approximation $S_{\mathbf{a},n}$.

Definition 4.2.20 (Compatible initial conditions)**.** Without loss of generality, suppose that n and m are nonnegative integers such that $n > m$. Let $(x_n^0, \alpha_n^0) \in (\Omega(S_{\mathbf{a},n}) \times S^1)/\sim$ and $(x_m^0, \alpha_m^0) \in (\Omega(S_{\mathbf{a},m}) \times S^1)/\sim$ be two initial conditions of the orbits $\mathscr{O}_n(x_n^0, \alpha_n^0)$ and $\mathscr{O}_m(x_m^0, \alpha_m^0)$, respectively, where we are assuming that α_n^0 and α_m^0 are both inward pointing. If $\alpha_n^0 = \alpha_m^0$ and if x_n^0 and x_m^0 lie on a segment determined from α_n^0 (or α_m^0) that intersects $\partial S_{\mathbf{a},n}$ only at x_n^0, then we say that (x_n^0, α_n^0) and (x_m^0, α_m^0) are *compatible initial conditions*.

Remark 4.2.21. When two initial conditions (x_n^0, α_n^0) and (x_m^0, α_m^0) are compatible, then

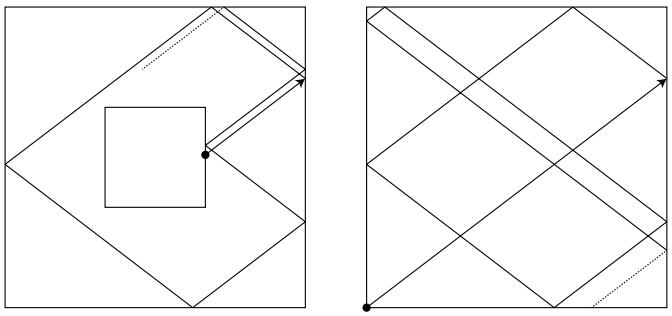


Figure 4.12: In this figure, we see on the left an orbit of $\Omega(S_{\mathbf{3},1})$ beginning at a point on the side of the peripheral square indicated by the small disc. On the right, we see an orbit of the square billiard $\Omega(S_0)$ beginning at the origin $(0,0)$ (again, indicated by a small disc) with the same initial direction as the orbit on the left. While both initial base points are collinear in the direction dictated by the initial direction of either orbit, these two orbits are not compatible because their initial conditions are not compatible. Specifically, the line connecting the two initial base points intersects an additional point of the boundary of the billiard table, thereby preventing the initial conditions from being compatible initial conditions.

we simply write each as (x_n^0, α^0) and (x_m^0, α^0). If two orbits $\mathscr{O}_m(x_m^0, \alpha_m^0)$ and $\mathscr{O}_n(x_n^0, \alpha_n^0)$ have compatible initial conditions, then we say such orbits are *compatible*.

It may be the case that an initial condition (x_n^0, α_n^0) is not compatible with (x_m^0, α_m^0), for any $m < n$. As such, in Definitions 4.2.22 and 4.2.23, we consider sequences beginning at $i = N$, for some $N \geq 0$; see Figure 4.12 and the corresponding caption.

Definition 4.2.22 (Sequence of compatible initial conditions)**.** Let $\{(x_i^0, \alpha_i^0)\}_{i=N}^{\infty}$ be a sequence of initial conditions, for some integer $N \geq 0$. We say that this sequence is a *sequence of compatible initial conditions* if for every $m \geq N$ and for every $n > m$, we have that (x_n^0, α_n^0) and (x_m^0, α_m^0) are compatible initial conditions. In such a case, we then write the sequence as $\{(x_i^0, \alpha^0)\}_{i=N}^{\infty}$.

Definition 4.2.23 (Sequence of compatible orbits)**.** Consider a sequence of compatible initial conditions $\{(x_n^0, \alpha^0)\}_{n=N}^{\infty}$. Then the corresponding sequence of orbits

$\{\mathscr{O}_n(x_n^0, \alpha^0)\}_{n=N}^{\infty}$ is called *a sequence of compatible orbits*.

If $\mathscr{O}_m(x_m^0, \alpha_m^0)$ is an orbit of $\Omega(S_{\mathbf{a},m})$, then $\mathscr{O}_m(x_m^0, \alpha_m^0)$ is a member of a sequence of compatible orbits $\{\mathscr{O}_n(x_n^0, \alpha^0)\}_{n=N}^{\infty}$ for some $N \geq 0$. It is clear from the definition of a sequence of compatible orbits that such a sequence is uniquely determined by the first orbit $\mathscr{O}_N(x_N^0, \alpha^0)$. Since the initial condition of an orbit determines the orbit, we can say without any ambiguity that a sequence of compatible orbits is determined by an initial condition (x_N^0, α^0).

Definition 4.2.24 (A sequence of compatible $\mathcal{P}$ orbits)**.** Let $\mathcal{P}$ be a property (resp., $\mathcal{P}_1, ..., \mathcal{P}_j$ a list of properties). If every orbit in a sequence of compatible orbits has the property $\mathcal{P}$ (resp., a list of properties $\mathcal{P}_1, ..., \mathcal{P}_j$), then we call such a sequence *a sequence of compatible $\mathcal{P}$ (resp., $\mathcal{P}_1, ..., \mathcal{P}_j$) orbits*.

Considering the fact that $\Omega(S_{\mathbf{a},n})$ is tiled by a square (i.e., an integrable billiard), Theorems 4.2.14 and 4.2.15 allow us to construct sequences of compatible closed orbits and sequences of compatible dense orbits. Furthermore, under the right conditions, we will be able to construct sequences of compatible periodic orbits. That is, we will be able to construct sequences of compatible orbits in which each orbit is a nonsingular, closed orbit in its respective billiard table $\Omega(S_{\mathbf{a},n})$.

4.3 A refinement of Theorem 4.1 in [25]

We now discuss Theorem 4.1 in [25] and provide various examples indicating the necessity for correcting the latter half of their statement.

In Part 1a of the proof of Theorem 4.1 of [25], it is determined that nontrivial line segments with slope $\alpha \in A$ necessarily avoid omitted squares of $S_{\mathbf{a}}$ (recall that the

omitted squares are the open squares removed in the construction of $S_{\mathbf{a}}$). However, in stating their result, the authors of [25] say that such nontrivial line segments necessarily intersect corners of peripheral squares, which is more restrictive than what is concluded in their proof. We provide an example where a nontrivial line segment beginning at $(0, 0)$ (in accordance with their proof) with a slope $\alpha \in A$ does not intersect any corner of any peripheral square. Furthermore, in the proof of Theorem 4.1 in [25], the authors assume in Part 1a that a nontrivial line segment with slope $\alpha \in A$ starts from the origin $(0, 0)$ and in Part 1b that a nontrivial line segment with slope $\alpha \in B$ starts from $(\frac{1}{2}, 0)$, but they do not state this in the assumptions of their theorem. This does not invalidate their characterization of self-similar Sierpinski carpets, but can be included in the statement of the Theorem 4.1 in [25] and further expanded upon. As our examples will show, such assumptions are necessary to state and can actually be made more robust to allow for an explicit description of other nontrivial line segments (some of which are not maximal,[11] but are nontrivial nonetheless).

In Remark 4.3 of [25], the authors indicate that one can translate the initial base point of a nontrivial line segment however one chooses (so long as the translation of the base point remains within the base of the unit square and not outside the unit square) and still maintain that the line segment starting from the new base point and having the same slope as before remains as a nontrivial line segment of $S_{\mathbf{a}}$. Their statement can be interpreted as implying that a variety of nontrivial line segments can be constructed in this manner, but not all. We provide two examples for which translating a nontrivial line segment does not produce a different nontrivial line segment, lending credence to this particular interpretation.

We summarize the above discussion by listing the examples indicating that the latter

[11] A maximal nontrivial line segment is a nontrivial line segment that connects two segments of the boundary of $S_{\mathbf{a}}$ corresponding to the boundary unit square S_0.

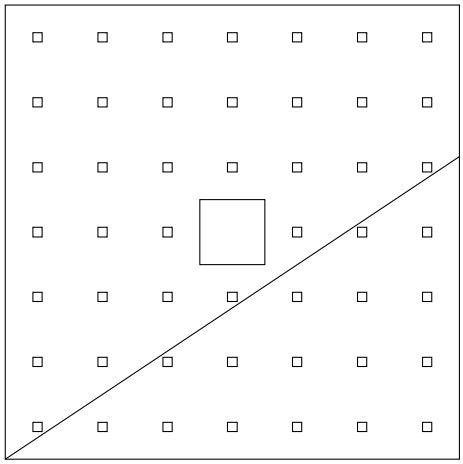


Figure 4.13: In this figure, we see a nontrivial line segment beginning from the origin $(0,0)$ and having a slope of $\frac{2}{3}$. Such a segment will avoid every peripheral square of $S_{\mathbf{7}}$.

half of the statement in Theorem 4.1 in [25] must be refined or stated separately as an additional theorem. These examples will be discussed in greater detail in §4.4.

1. A nontrivial line segment of $S_{\mathbf{7}}$ with slope $\alpha = \frac{2}{3} \in A$ beginning at $(0,0)$ that avoids corners of every peripheral square.
2. A nontrivial line segment of $S_{\mathbf{7}}$ with slope $\alpha = \frac{2}{3} \in A$ beginning at $(\frac{1}{2},0)$ that avoids corners of every peripheral square.
3. A nontrivial line segment of $S_{\mathbf{5}}$ with slope $\alpha = \frac{1}{3} \in B$ beginning from $(\frac{1}{2},0)$ that, when translated to the origin $(0,0)$, constitutes a trivial line segment of $S_{\mathbf{5}}$.
4. A nontrivial line segment of $S_{\mathbf{7}}$ with slope $\alpha = \frac{3}{4} \in A$ beginning from $(0,0)$ that, when translated to $(\frac{1}{2},0)$, constitutes a trivial line segment of $S_{\mathbf{7}}$.

The above examples indicate the necessity for distinguishing between slopes in $A_{\mathbf{a-2}}$ and slopes in $A_{\mathbf{a}} \backslash A_{\mathbf{a-2}}$. This is the key point of our Theorem 4.3.4.

In preparation for our result, we give three lemmas.

Lemma 4.3.1. *Let $S_{\mathbf{a}}$ be a self-similar Sierpinski carpet. Then,*

$$A_{\mathbf{a}} \setminus A_{\mathbf{a-2}} = \left\{ \frac{p}{q} : 1 \leq p < q,\ p+q = a,\ p \text{ and } q \text{ are coprime} \right\}.$$

Proof. We first note that if $\alpha \in A_{\mathbf{a}}$, then $\alpha = \frac{p}{q}$ such that

$$p + q \leq a \text{ and } 0 \leq p < q \leq a - 1.$$

Likewise, if $\frac{p}{q} \in A_{\mathbf{a-2}}$, then

$$p + q \leq a - 2 \text{ and } 0 \leq p < q \leq a - 3.$$

Hence, $\frac{p}{q} \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$ satisfies either

$$a - 2 < p + q \leq a \text{ and } 1 \leq p < q \leq a - 1, \tag{4.6}$$

or

$$p + q \leq a,\ 1 \leq p < q \leq a - 1 \text{ and } q > a - 3, \tag{4.7}$$

in addition to $p + q$ being odd ***and*** p, q being coprime. It can be deduced from the inequalities given on line (4.6) that $p + q = a$. Similarly, since p and q are nonnegative integers, it can be deduced from the inequalities on line (4.7) that $q = a-2$ or $q = a-1$, which forces $p = 2$ or $p = 1$, respectively. (Otherwise, $p + q$ would be even or exceed the value a.) The union of these two sets comprises all $\frac{p}{q}$ for which $p+q = a$, $1 \leq p < q$, and p and q being coprime. The inclusion in the other direction is straight forward. ■

Lemma 4.3.2. *Let $S_{\mathbf{a}}$ be a self-similar Sierpinski carpet. If $\alpha \in B_{\mathbf{a}}$, $n \geq 1$, and r is an odd positive integer with $r \leq 2a^n$, then the line $y = \alpha(x - \frac{r}{2a^n})$ avoids points of the plane of the form $(\frac{u}{a^l}, \frac{v}{a^m})$, $u, v, l, m \in \mathbb{Z}$.*

Proof. Let $S_{\mathbf{a}}$ be a self-similar Sierpinski carpet billiard, $\alpha \in B_{\mathbf{a}}$, $n \geq 1$, and r a positive odd integer with $r \leq 2a^n$. Suppose there exist $u, v, l, m \in \mathbb{Z}$, $l, m \geq 0$ such that $y = \alpha(x - \frac{r}{2a^n})$ intersects the point $(\frac{u}{a^l}, \frac{v}{a^m})$. Then,

$$\frac{v}{a^m} = \alpha\left(\frac{u}{a^l} - \frac{r}{2a^n}\right) \tag{4.8}$$

$$2va^{n+l} = \alpha a^m(2ua^n - ra^l). \tag{4.9}$$

Since $\alpha = \frac{p}{q}$, $p, q \in \mathbb{Z}$, p and q both odd and positive, we have that

$$2qva^{n+l} = pa^m(2ua^n - ra^l). \tag{4.10}$$

The left-hand side of Equation (4.10) is even and the right-hand side is odd, so we have an immediate contradiction. Therefore, no line of the form $y = \alpha(x - \frac{r}{2a^n})$ will ever intersect any point of the form $(\frac{u}{a^l}, \frac{v}{a^m})$, $u, v, l, m \in \mathbb{Z}$, $m, l \geq 0$. ■

Lemma 4.3.3. *Let $b, c \in \mathbb{N}$, $d = \gcd(b, c)$, and $k \in \mathbb{Z}$. If d divides k, then the equation $bx + cy = k$ has integer solutions (x, y). If (x_0, y_0) is one such integer solution, then all integer solutions can be expressed in the form*

$$(x, y) = \left(x_0 + m\frac{c}{d},\ y_0 - m\frac{b}{d}\right) \tag{4.11}$$

for some $m \in \mathbb{Z}$.

Proof. This is an application of the elementary fact (see *e.g.* Theorem 1.3 of [66]) that there exist $x_1, y_1 \in \mathbb{Z}$ such that $\gcd(b, c) = bx_1 + cy_1$. ■

We are now ready to state our refinement of Theorem 4.1 of [25], which corrects the latter half of their statement.

Theorem 4.3.4 (A refinement of Theorem 4.1 in [25])**.** *Let* $\mathbf{a} = (\frac{1}{a}, \frac{1}{a}, \frac{1}{a}, ...)$ *be a constant sequence. Then the set of slopes Slope$(S_{\mathbf{a}})$ is the union of the following two sets*:

$$A_{\mathbf{a}} = \left\{ \frac{p}{q} \ : \ p+q \leq a,\ 0 \leq p < q \leq a-1,\ p,q \in \mathbb{N} \cup \{0\},\ p+q \text{ is odd} \right\}, \tag{4.12}$$

$$B_{\mathbf{a}} = \left\{ \frac{p}{q} \ : \ p+q \leq a-1,\ 0 \leq p \leq q \leq a-2,\ p,q \in \mathbb{N},\ p \text{ and } q \text{ are odd} \right\}. \tag{4.13}$$

Moreover, if $\alpha \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$, then each nontrivial line segment in $S_{\mathbf{a}}$ with slope α beginning from $(0,0)$ touches vertices of peripheral squares, while if $\alpha \in A_{\mathbf{a-2}}$, then each nontrivial line segment in $S_{\mathbf{a}}$ with slope α beginning from $(0,0)$ is disjoint from all peripheral squares. If $\alpha \in B_{\mathbf{a}}$, then each nontrivial line segment in $S_{\mathbf{a}}$ with slope α beginning at $(\frac{r}{2a^n}, 0)$, $n \geq 1$ and $r < 2a^n$ being a positive odd integer, is disjoint from all peripheral squares with side-length $< \frac{1}{a^n}$.

In addition, for each $\alpha \in A_{\mathbf{a}} \cup B_{\mathbf{a}}$, there exist maximal line segments in $S_{\mathbf{a}}$ with slope α. Finally, if $b < a$, then any maximal nontrivial line segment in $S_{\mathbf{b}}$ is also contained in $S_{\mathbf{a}}$. In particular, Slope$(S_{\mathbf{b}}) \subseteq$ Slope$(S_{\mathbf{a}})$.

Proof. We know from the proof of Theorem 4.1 in [25] that the set Slope$(S_{\mathbf{a}})$ can be partitioned as described in Equations (4.12) and (4.13). Moreover, the proof of Theorem 4.1 in [25] is valid for the case when $\alpha \in B_{\mathbf{a}}$ and the nontrivial line segment with slope α begins at $(\frac{1}{2}, 0)$; the proof also remains valid for when $\alpha \notin$ Slope$(S_{\mathbf{a}})$. Hence, we focus on the cases where $\alpha \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$ and a nontrivial line segment with slope α begins at $(0,0)$; $\alpha \in A_{\mathbf{a-2}}$ and a nontrivial line segment with slope α begins at $(0,0)$; and $\alpha \in B_{\mathbf{a}}$ and a nontrivial line segment with slope α begins at $(\frac{r}{2a^n}, 0)$, for some $n \geq 1$, $r \neq a$ and r an odd positive integer.

Case 1: $\alpha \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$. First note that the inferior right corner of a peripheral square of $S_{\mathbf{a}}$ has coordinate

$$\left(\frac{a+1}{2a^{n+1}}, \frac{a-1}{2a^{n+1}} \right) + \left(\frac{u}{a^n}, \frac{v}{a^n} \right) \tag{4.14}$$

for some u,$v \in \mathbb{N} \cup \{0\}$, with $0 \leq u, v \leq a^n - 1$ and $u \neq v$. We wish to show that any line segment emanating from $(0,0)$ with slope $\alpha = \frac{p}{q} \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$ hits the inferior right corner of some peripheral square: that is, there exist $0 \leq u \leq a^n - 1$ and $0 \leq v \leq a^n - 1$ such that the equation

$$\frac{1}{q}\left(u + \frac{a+1}{2a}\right) = \frac{1}{p}\left(v + \frac{a-1}{2a}\right) \tag{4.15}$$

holds.

By Lemma 4.3.1, $p+q = a$ for any $\frac{p}{q} \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$, so Equation (4.15) can be rewritten as

$$2(a-q)u - 2qv = 2q - a - 1. \tag{4.16}$$

Geometrically speaking, the next step is to find, in the uv-plane (regarding u and v as real-valued rather than integer-valued), the intersections between the line defined by Equation (4.16) and the lattice $\{(u,v) : u, v \in \mathbb{Z} \cap [0, a^n - 1], u \neq v\}$. Observe that the line defined by Equation (4.16) has v-intercept $\frac{a+1}{2q} - 1$, which is between $-\frac{1}{2} + \frac{1}{a-1}$ and 0 (because $\frac{a+1}{2} \leq q \leq a - 1$ by Lemma 4.3.1). If $a + 1 - 2q = 0$ (or $p = q - 1$), then immediately $(u,v) = (0,0)$ is an intersection point. If not, we move along the line in integer increments of u, and check whether or not the v-coordinate is an integer. The successive v-coordinates can be written as

$$v = \left(\frac{a+1}{2q} - 1\right) + u\left(\frac{a-q}{q}\right) \tag{4.17}$$

$$= \frac{(a+1) + 2au}{2q} - (u+1), \tag{4.18}$$

with $u \in \mathbb{Z} \cap [0, a^n - 1]$. Thus it is enough to find a $u \in \mathbb{Z} \cap [0, a^n - 1]$ such that q divides $\left(\frac{a+1}{2} + au\right)$ (note that $a + 1$ is even). An equivalent way to say this is that we are looking for integer solutions (u, r) of the equation

$$au + qr = -\frac{a+1}{2}. \tag{4.19}$$

Using the fact that $\gcd(a, q) = \gcd(p, q) = 1$ by Lemma 4.3.1, we apply Lemma 4.3.3 to deduce that there is a unique pair of integers $(u^*, r^*) \in \{0, 1, \cdots, q-1\} \times \mathbb{Z}$ which solves Equation (4.19). Then, q divides $\left(\frac{a+1}{2} + au^*\right)$. Denote by v^* the quantity $\left(\frac{a+1}{2q} - 1\right) + u^*\left(\frac{a-q}{q}\right)$ and note that $v^* < u^*$. It follows that the line segment emanating from $(0, 0)$ with slope $\frac{a-q}{q} \in A_{\mathbf{a}} \setminus A_{\mathbf{a-2}}$ hits the inferior right corner of a peripheral square whose coordinate is $\frac{1}{a^n}(u^* + \frac{a+1}{2a}, v^* + \frac{a-1}{2a})$.

Case 2: $\alpha \in A_{\mathbf{a-2}}$**.** Consider a nontrivial line segment with slope α beginning at $(0, 0)$. We suppose the nontrivial line segment intersects a peripheral square at a corner, whose coordinate is given by the expression on line (4.14). Then, if $\alpha = \frac{p}{q} \in A_{\mathbf{a-2}}$, we have

$$\frac{1}{q}\left(u + \frac{a+1}{2a}\right) = \frac{1}{p}\left(v + \frac{a-1}{2a}\right), \tag{4.20}$$

which reduces to

$$q + 2vq - p - 2up = \frac{p+q}{a}. \tag{4.21}$$

While the left-hand side of Equation (4.21) is an integer, the right-hand side is not, since $p + q \leq a - 2 < a$. This contradiction then implies that the nontrivial line segment cannot intersect any lower-right corner of any peripheral square (and, by symmetry, any upper-left corner of any peripheral square).

Case 3: $\alpha \in B_{\mathbf{a}}$**.** Consider a nontrivial line segment of $S_{\mathbf{a}}$ with slope $\alpha \in B_{\mathbf{a}}$ emanating from $(\frac{1}{2}, 0)$, the existence of which follows from the proof of Theorem 4.1 in [25]. We claim that a line segment emanating from $(\frac{r}{2a^n}, 0)$ with slope α will also be a nontrivial line segment of $S_{\mathbf{a}}$ and never intersect the sides of any peripheral square with side-length $\frac{1}{a^k}$, $k \geq n$.

To such end, consider $n \geq 0$, and r an odd positive integer with $r \leq 2a^n$. We scale

the nontrivial line segment emanating from $(\frac{1}{2}, 0)$ by $\frac{1}{a^n}$. Then the scaled nontrivial line segment is a nontrivial line segment of the cell. Then, there exists $(c \leq a^n, 0)$ such that translating the line segment of the cell by $\frac{c}{a^n}$ results in a nontrivial line segment of a cell with $(\frac{r}{2a^n}, 0)$ as the midpoint of the base of the cell. Moreover, the translated nontrivial line segment emanates from $(\frac{r}{2a^n}, 0)$ with slope α.

We now consider a tiling of the plane by the cell having $(\frac{r}{2a^n}, 0)$ as the midpoint of the base. Then, a line $y = \alpha(x - \frac{r}{2a^n})$ avoids every peripheral square in the tiling. If not, then the resulting line must also intersect a side of a peripheral square contained in the cell that does not correspond to a corner of the peripheral square by virtue of Lemma 4.3.2. Since $S_{\mathbf{a}}$ is a self similar Sierpinski carpet, any line segment with slope $\alpha \in B_{\mathbf{a}}$ beginning from $(\frac{r}{2a^n}, 0)$ must, by construction, be a trivial line segment in $S_{\mathbf{a}}$.

■

4.4 Eventually constant sequences of compatible periodic orbits of $S_{\mathbf{a}}$

In this section, we further elaborate on the examples described in Items (1)–(4) in §4.3 and discuss how exactly each example necessitates the correction of the latter half of the statement made in Theorem 4.1 in [25] and refining various comments made throughout. Even though Theorem 4.3.4 makes Items (1)–(4) non-issues, we verify such examples constitute counterexamples for the latter half of Theorem 4.1 of [25] and justification for refining comments made in and around their proof. In so doing, we will provide proper motivation for our main results stated in Theorem 4.4.2 and Proposition 4.4.3. Moreover, as the title of this section suggests, we make considerable use of the terminology and

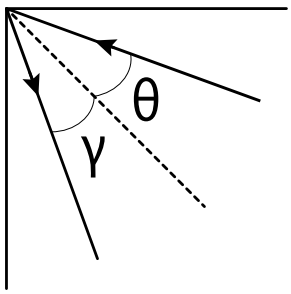


Figure 4.14: The corners of a square billiard table constitute removable singularities of the billiard flow on the square billiard table $\Omega(S_0)$; see §4.2.2 for a discussion of removable and nonremovable singularities of a translation surface. In this figure, $\gamma = \theta$. Consequently, every corner of a prefractal approximation of a Sierpinski carpet $S_{\mathbf{a}}$ corresponding to a corner of the original square constitutes a removable singularity of the billiard flow on $\Omega(S_{\mathbf{a},n})$.

results from §4.2 regarding mathematical billiards and fractal billiards. Finally, we will determine a family of periodic orbits of self-similar Sierpinski carpets.

We first note that the corners of the square billiard table constitute removable singularities of the billiard flow. That is, reflection can be defined in the corners of the square in a well-defined manner. Moreover, in every prefractal approximation of a Sierpinski carpet, reflection in the corners of the original square boundary can be made well-defined, as well. Specifically, a billiard ball entering into a corner of the original square of any prefractal approximation of a Sierpinski carpet will exit the corner in a direction that makes an angle γ that constitutes the reflection of the incoming angle θ as reflected through the angle bisector of the vertex; see Figure 4.14.

In Item (1) of §4.3, we stated that a nontrivial line segment of S_7 with an initial starting point of $(0, 0)$ and a slope $\alpha = \frac{2}{3}$ avoids corners of every peripheral square of S_7. Consider an orbit $\mathcal{O}_0((0, 0), \frac{2}{3})$ of the square billiard table $\Omega(S_0)$. We claim that this orbit avoids corners of every peripheral square of $\Omega(S_7)$.

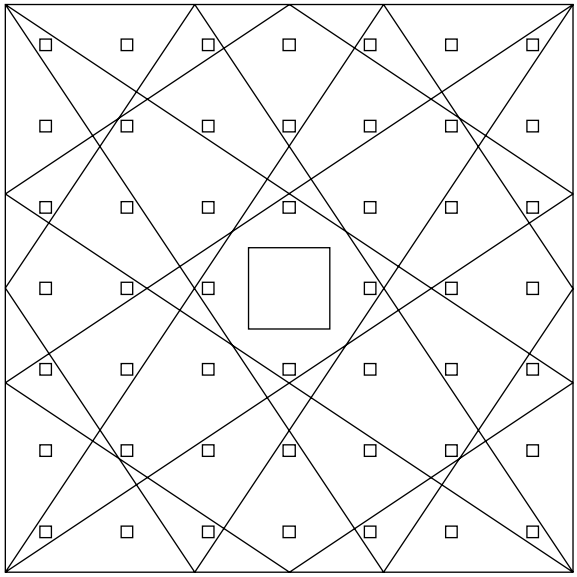


Figure 4.15: In this figure, we see an orbit that avoids peripheral squares of $\Omega(S_{7,2})$, the second level approximation of the self-similar Sierpinski carpet billiard table $\Omega(S_7)$. As noted in the text, the fact that the orbit intersects corners does not pose a problem for determining the trajectory in a well-defined way.

To such end, consider the orbit shown in Figure 4.15, which has an initial condition $((0,0),\frac{2}{3})$. It is clear that this orbit avoids peripheral squares of $\Omega(S_{7,2})$. Then, scaling $\Omega(S_{7,2})$ and the orbit together, relative to the origin, we see that $\frac{1}{7}\mathcal{O}((0,0),\alpha)$ avoids peripheral squares of the cell $C_{1,a}$ (i.e., the cell with side length $\frac{1}{7}$ and lower left corner corresponding to the origin); see Figure 4.16. Not only does the orbit of the cell avoid the peripheral square with side-length $\frac{1}{49}$, it also avoids the peripheral squares with side length $\frac{1}{7^3}$, since the orbit $\mathcal{O}((0,0),\alpha)$ avoided every peripheral square of length $\frac{1}{49}$ in $\Omega(S_{7,2})$.

Now, the reflected-unfolding of the orbit of the cell $C_{1,a}$ must follow the same path as $\mathcal{O}((0,0),\alpha)$ on account of both orbits having the same initial condition and reflection being an isometry; see Figure 4.17. This then implies that $\mathcal{O}((0,0),\alpha)$ is an orbit of $\Omega(S_{7,3})$ that avoids all peripheral squares of side-length $\frac{1}{7^3}$.

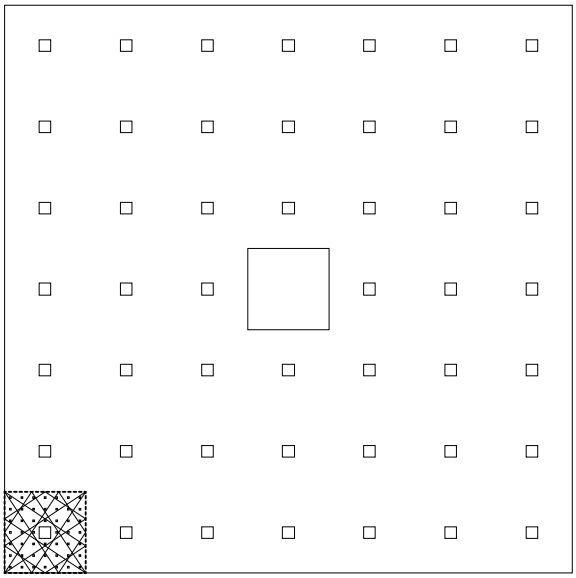


Figure 4.16: We see here the billiard table $\Omega(S_{7,2})$ and the orbit scaled by $\frac{1}{7}$. Such a scaled orbit is then an orbit of the cell $C_{1,7}$. We then proceed to reflect-unfold this orbit to recover the orbit before scaling; see Figure 4.17.

Suppose there exists $N \in \mathbb{N}$, $N \geq 3$ such that for every $n \leq N$, the orbit $\mathcal{O}((0,0),\alpha)$ avoids corners of peripheral squares with side-length $\frac{1}{7^n}$. Then, scaling $\mathcal{O}((0,0),\alpha)$ together with $\Omega(S_{7,N})$ by $\frac{1}{7}$, we see that the orbit of the cell $C_{1,7^N}$ avoids the peripheral square of side-length $\frac{1}{7^N}$ since $\mathcal{O}((0,0),\alpha)$ avoids peripheral squares of side length $\frac{1}{7^n}$, for all $n \leq N$. In addition, $\frac{1}{7}\mathcal{O}((0,0),\alpha)$ avoids peripheral squares of side-length $\frac{1}{7^{N+1}}$ contained in the cell, since $\mathcal{O}((0,0),\alpha)$ avoided peripheral squares of side-length $\frac{1}{7^N}$ (recall that peripheral squares of side-length $\frac{1}{7^N}$ are scaled by $\frac{1}{7}$ when constructing $S_{7,N+1}$ from $S_{7,N}$ via an iterated function system comprised of contraction mappings).

We may then reflect-unfold the orbit $\frac{1}{7}\mathcal{O}((0,0),\alpha)$ of the cell in $\Omega(S_{7,N+1})$. The reflected-unfolded orbit $\frac{1}{7}\mathcal{O}((0,0),\alpha)$ must then retrace the exact path of $\mathcal{O}((0,0),\alpha)$ in $\Omega(S_{7,N})$, but now in $\Omega(S_{7,N+1})$. Since reflection is an isometry, the reflected-unfolding of the orbit of the cell must then avoid peripheral squares of $\Omega(S_{7,N+1})$.

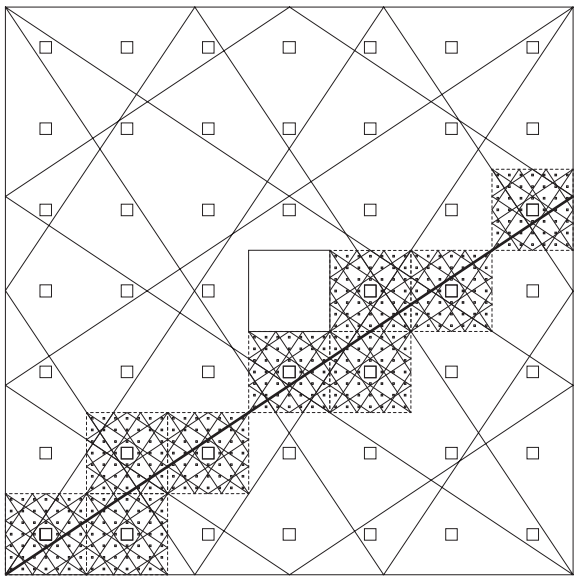


Figure 4.17: In this figure, we see the scaled orbit unfolded. Rather, we see $\frac{1}{7}\mathscr{O}^s((0,0),2/3)$ unfolded, for sufficiently large s. Indeed, we see that we can recapture the original orbit $\mathscr{O}((0,0),2/3)$ by reflecting-unfolding the scaled orbit. Moreover, reflecting-unfolding is an isometry, so no peripheral squares are intersected.

Therefore, we have concluded that $\mathscr{O}((0,0),\alpha)$ must avoid all peripheral squares of $\Omega(S_{7,n})$, for every $n \geq 1$.

In verifying that the example in Item (1) of §4.3 constitutes a counter example to one aspect of the latter half of Theorem 4.1 of [25], we see how we can quickly build upon it to verify that Item (2) of §4.3 is also a counter example to another aspect of the latter half of the main result in [25]. If $\alpha = \frac{2}{3} \in A_{\mathbf{5}} \subseteq A_{\mathbf{7}}$, then a nontrivial line segment of $S_{\mathbf{7}}$ beginning at $(\frac{1}{2},0)$ with slope α necessarily avoids all peripheral squares of $S_{\mathbf{7}}$. A similar reasoning as above can be used to demonstrate the validity of this statement. In fact, the path traversed by the orbit $\mathscr{O}_n((0,0),\frac{2}{3})$ is identical to the path traversed by the orbit $\mathscr{O}_n((\frac{1}{2},0),\frac{2}{3})$, for every $n \geq 0$.

The third example referred to in Item (3) of §4.3 consists of a nontrivial line segment

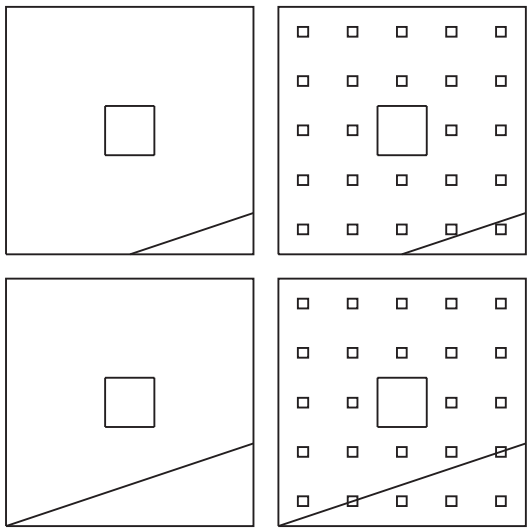


Figure 4.18: An example of a nontrivial line segment of $S_{\mathbf{5}}$ with slope $\alpha = \frac{1}{3} \in B_{\mathbf{5}}$ starting from $\left(\frac{1}{2}, 0\right)$, which becomes trivial when translated to the base point $(0, 0)$.

of $S_{\mathbf{5}}$ beginning at $(\frac{1}{2}, 0)$ with a slope $\alpha = \frac{1}{3} \in B_{\mathbf{5}}$. As illustrated in Figure 4.18, when a nontrivial line segment beginning at $(\frac{1}{2}, 0)$ with slope $\alpha = \frac{1}{3}$ is translated to $(0, 0)$ and extended so as to intersect the right side of the portion of the boundary corresponding to the unit square $S_0 = Q$, we see that the new segment intersects omitted squares. Since $S_{\mathbf{7}}$ is self-similar, such a segment will intersect infinitely many omitted squares. Hence, the new segment must be a trivial line segment of $S_{\mathbf{5}}$.

Our fourth example referred to in Item (4) of §4.3 consists of a nontrivial line segment of $S_{\mathbf{7}}$ beginning at $(0, 0)$ with a slope $\alpha = \frac{3}{4}$. Consider the translation of the nontrivial line segment with slope $\alpha = \frac{3}{4}$ from $(0, 0)$ to $(\frac{1}{2}, 0)$. As illustrated in Figure 4.19, such a segment intersects omitted squares of $S_{\mathbf{7}}$. Since $S_{\mathbf{7}}$ is self-similar, it must be the case that the translated segment intersects infinitely many omitted squares of $S_{\mathbf{7}}$. Hence, a segment starting from $(\frac{1}{2}, 0)$ with a slope $\alpha = \frac{3}{4}$ must be a trivial line segment.

Definition 4.4.1 (Constant sequence of compatible orbits)**.** Given a nonnegative integer

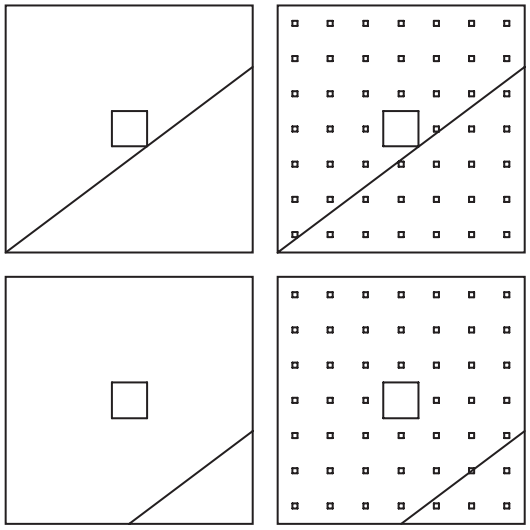


Figure 4.19: An example of a nontrivial line segment of S_7 with slope $\alpha = \frac{3}{4} \in A_7$ starting from $(0,0)$, which becomes trivial when translated to the base point $\left(\frac{1}{2}, 0\right)$.

N, we say that a sequence of compatible orbits $\{\mathscr{O}_n(x_n^0, \alpha^0)\}_{n=N}^{\infty}$ is a *constant sequence of compatible orbits* if the path traversed by $\mathscr{O}(x_n^0, \alpha_n^0)$ is identical to the path traversed by $\mathscr{O}(x_n^0, \alpha_n^0)$, for every $n \geq N$. Furthermore, we say that a sequence of compatible orbits $\{\mathscr{O}_n(x_n^0, \alpha^0)\}_{n=0}^{\infty}$ is *eventually constant* if there exists a nonnegative integer N such that $\{\mathscr{O}_n(x_n^0, \alpha^0)\}_{n=N}^{\infty}$ is constant, in the above sense.

Theorem 4.4.2. *Let $\alpha \in Slope(S_{\mathbf{a}})$.*

1. *If $\alpha \in A_{\mathbf{a}-\mathbf{2}}$, then the sequence of compatible orbits $\{\mathscr{O}_n((0,0), \alpha)\}_{n=0}^{\infty}$ is a sequence of compatible periodic orbits.*

2. *If $\alpha \in B_{\mathbf{a}}$ and $r < 2a^n$ is an odd positive integer, then the sequence of compatible orbits $\{\mathscr{O}_n((\frac{r}{2a^n}, 0), \alpha)\}_{n=0}^{\infty}$ is a sequence of compatible periodic orbits.*

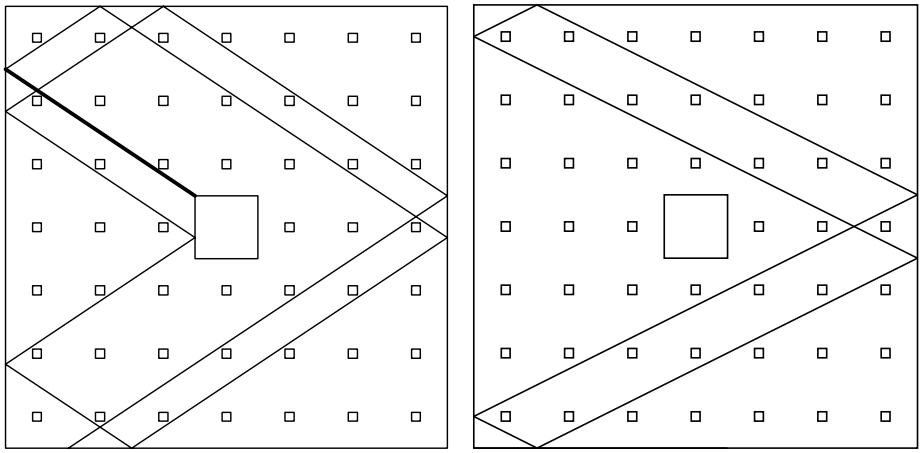
Furthermore, in each case, the sequence of compatible periodic orbits is eventually constant, and its trivial limit constitutes a periodic orbit of a self-similar Sierpinski

carpet billiard table $\Omega(S_{\mathbf{a}})$.

Proof. Part 1: Suppose $\alpha \in A_{\mathbf{a-2}}$ and an orbit of $\Omega(S_0)$ has an initial condition $((0,0),\alpha)$. Consider an orbit of $\mathscr{O}_1((0,0),\alpha)$ of $\Omega(S_{\mathbf{a},1})$. We claim that such an orbit avoids the peripheral square of $\Omega(S_{\mathbf{a},1})$. If not, then unfolding the orbit in the tiling of the plane by $\Omega(S_{\mathbf{a},1})$ must intersect a peripheral square in the tiling. Then, appropriately scaling the unfolded orbit and the tiling by a factor of $\frac{1}{a^n}$ results in a nontrivial line segment intersecting a peripheral square. In Theorem 4.3.4, we showed that a nontrivial line segment with a slope $\alpha \in A_{\mathbf{a-2}}$ cannot intersect any point of a peripheral square of any approximation $S_{\mathbf{a},m}$, $m \geq 1$. Now suppose there exists $N > 0$ such that an orbit $\mathscr{O}_n((0,0),\alpha)$ avoids corners of peripheral squares of $\Omega(S_{\mathbf{a},n})$, for all $n \leq N$. Then, scaling $\Omega(S_{\mathbf{a},N})$ and the orbit by $\frac{1}{a}$ results in an orbit of a cell with side length $\frac{1}{a}$. By construction, the orbit of the cell avoids all peripheral squares contained in the cell, meaning the orbit avoids peripheral squares with side-length $\frac{1}{a^k}$, for all $k \leq N+1$. Then, reflecting-unfolding the orbit of the cell, as well as the peripheral squares of the cell, we see that the reflected-unfolded orbit $\mathscr{O}^s_{N+1}((0,0),\alpha)$ of $\Omega(S_{\mathbf{a},N+1})$ avoids all peripheral squares. Hence, for all $n \geq 1$, the orbit $\mathscr{O}_n((0,0),\alpha)$ avoids peripheral squares of $\Omega(S_{\mathbf{a},n})$.

Part 2: Let $\alpha \in B_{\mathbf{a}}$ and $(x_0^0, y_0^0) = (\frac{1}{2}, 0)$. Then, the orbit $\mathscr{O}_1((\frac{1}{2},0),\alpha)$ avoids the peripheral square of $\Omega(S_{\mathbf{a},1})$. This follows from the fact that the unfolded orbit must avoid peripheral squares of the tiling of the plane by $\Omega(S_{\mathbf{a},1})$. Moreover, for every $m \geq 1$, $\mathscr{O}_n((\frac{1}{2},0),\alpha)$ must avoid peripheral squares of $\Omega(S_{\mathbf{a},m})$ and $\mathscr{O}_m((\frac{1}{2},0),\alpha)$ is identical to $\mathscr{O}_0((\frac{1}{2},0),\alpha)$. Therefore, the compatible sequence of periodic orbits $\{\mathscr{O}_n((\frac{1}{2},0),\alpha)\}_{n=0}^{\infty}$ is an eventually constant sequence of compatible periodic orbits.

Let $n \geq 0$, r a positive, odd integer and $r \leq 2a^n$. Consider again the orbit $\mathscr{O}_0((\frac{1}{2},0),\alpha)$. We know that $\mathscr{O}_0((\frac{1}{2},0),\alpha)$ avoids peripheral squares of $\Omega(S_{\mathbf{a},n})$ for every $n \geq 1$. Therefore, scaling $\mathscr{O}_0((\frac{1}{2},0),\alpha)$ relative to $(\frac{1}{2},0)$ by $\frac{1}{a^n}$ and making a sufficient



(a) (b)

Figure 4.20: In $\Omega(S_{7,2})$, $\mathscr{O}_2((\frac{1}{7},0),\frac{2}{3})$ is a singular orbit (the line segment connecting with the singularity is shown as a segment with a greater weight than the other segments in the orbit), while $\mathscr{O}_2((\frac{1}{7},0),\frac{1}{2})$ is a periodic orbit.

translation so that $x_0^0 = \frac{r}{2a^n}$ results in an orbit of a cell with side length $\frac{1}{a^n}$ that avoids all peripheral squares contained in the cell. By Theorem 4.3.4, we know that the unfolding of the orbit of the cell must not intersect any peripheral squares of a tiling of the plane by the cell (as well as the scaled copy of $\Omega(S_{\mathbf{a}})$ contained in the cell). Hence, the reflected-unfolding of the orbit of the cell must then also avoid peripheral squares with side-length $\frac{1}{a^k}$, for all $k \geq n$. Moreover, by Lemma 4.3.2, the reflected-unfolding of the orbit of the cell cannot intersect any corners of any peripheral squares of $\Omega(S_{\mathbf{a}})$ with side-length $\frac{1}{a^j}$, $j < n$. Therefore, $\{\mathscr{O}_n((\frac{r}{2a^n},0),\alpha)\}_{n=0}^{\infty}$ is an eventually constant sequence of compatible periodic orbits. ■

For an orbit with an initial direction $\alpha \in A_{\mathbf{a-2}}$ starting from a corner of some cell C_{m,a^m}, but not from $(0,0)$, we have the following result.

Proposition 4.4.3. *Consider a self-similar Sierpinski carpet billiard table $\Omega(S_{\mathbf{a}})$. Let $\alpha \in A_{\mathbf{a-2}}$, $m \in \mathbb{N}$, and $k \in \{1,2,\cdots,a^m-1\}$. Then there exists an $N \in \mathbb{N} \cup \{0\}$ such that $\{\mathscr{O}_n((\frac{k}{a^m},0),\alpha)\}_{n=N}^{\infty}$ is a constant sequence of compatible closed orbits.*

Proof. By Theorem 4.3.4, the orbit unfolded onto the tiling of the plane by $a^{-m}\Omega(S_{\mathbf{a},1})$ (that is, $\Omega(S_{\mathbf{a},1})$ scaled by a^{-m}) must avoid all peripheral squares of the tiling. Upon scaling and reflected-unfolding as in the proof of Theorem 4.4.2, Part 1, we deduce that for every $n \in \mathbb{N}$, $\mathcal{O}_n((\frac{k}{a^m}, 0), \alpha)$ avoids all peripheral squares of $\Omega(S_{\mathbf{a},n})$ whose side-length is less than a^{-m}. The claim then follows. ■

Remark 4.4.4. Note that Proposition 4.4.3 does not say whether $\mathcal{O}_n((\frac{k}{a^m}, 0), \alpha_n^0)$ is a periodic or singular orbit of $\Omega(S_{\mathbf{a},n})$. The focus of Proposition 4.4.3 is on determining a sequence of compatible *closed* orbits, which can include both periodic *and* singular orbits of prefractal billiards. If it hits a corner of some larger peripheral square (which has conic angle 6π), then we must terminate the billiard trajectory at the corner and declare the orbit $\mathcal{O}_n((\frac{k}{a^m}, 0), \alpha_n^0)$ to be singular. On the other hand, if the orbit avoids corners of peripheral squares, then it will constitute a periodic orbit. See Figure 4.20 for examples of both types of orbits.

4.4.1 The translation surface $\mathcal{S}(S_{\mathbf{a},n})$

Recall from §4.2.2 that a rational billiard $\Omega(D)$ can be used to construct a translation surface by appropriately identifying $2N$ many copies of $\Omega(D)$, where $N = \text{lcm}\{q_1, q_2, ..., q_k\}$. Suppose $\Omega(S_{\mathbf{a}})$ is a self-similar Sierpinski carpet billiard table and $\Omega(S_{\mathbf{a},n})$ is an approximation of $\Omega(S_{\mathbf{a}})$. Then, each interior angle has a denominator 2. Therefore, $N = \text{lcm}\{2, ..., 2\} = 2$, meaning that, when four copies of $\Omega(S_{\mathbf{a},n})$ are appropriately identified, we recover the corresponding translation surface $\mathcal{S}(S_{\mathbf{a},n})$; see Figure 4.21 for the case of a translation surface constructed from $\Omega(S_{\mathbf{7},2})$, as well as an illustration of how a billiard ball would traverse the corresponding translation surface.

As previously shown in Figure 4.14, reflection in a corner with an angle measuring

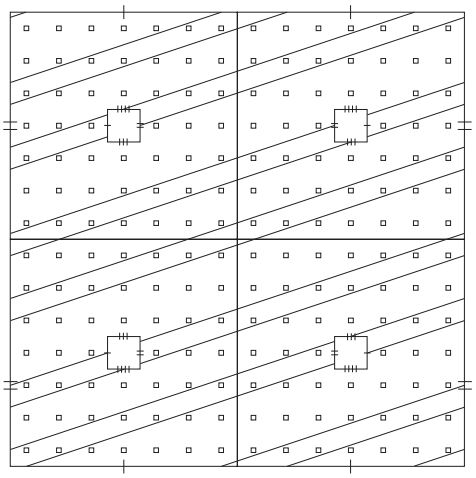


Figure 4.21: We see here four copies of the prefractal approximation $\Omega(S_{\mathbf{7},2})$ properly identified in such a way that the resulting translation surface is no longer a torus, but a higher genus surface. The geodesic shown here further indicates exactly how sides of $\Omega(S_{\mathbf{7},2})$ are identified in the construction of $\mathcal{S}(S_{\mathbf{7},2})$. Furthermore, the way in which sides of $\Omega(S_{\mathbf{7},2})$ are identified makes the geodesic equivalent to a billiard orbit under the action of the group of symmetries acting on the billiard table to produce the translation surface. In the terminology introduced in §4.2.2, one could consider the geodesic in this figure to be the billiard orbit unfolded in the translation surface.

$\frac{\pi}{2}$ in the square can be made well-defined. The justification for this is provided by the fact that in the associated translation surface, a vertex of the unit square is identified with three other vertices in such a way that the conic angle about the conic singularity is 2π. In Figure 4.22, we see the effect that a removable singularity (i.e., a singularity with conic angle 2π) has on the straight-line flow. Specifically, each point of $\mathcal{S}(S_{\mathbf{a},n})$ corresponding to a corner of the zeroth level approximation $\Omega(S_0)$ (or, the unit square billiard table), is a removable singularity of the translation surface; each point of $\mathcal{S}(S_{\mathbf{a},n})$ corresponding to a corner of a peripheral square of $\Omega(S_{\mathbf{a},n})$ is a nonremovable singularity of the translation surface, with the conic angle about each nonremovable singularity measuring 6π.

Using the formula given in Equation (4.5), the genus g_n of a translation surface

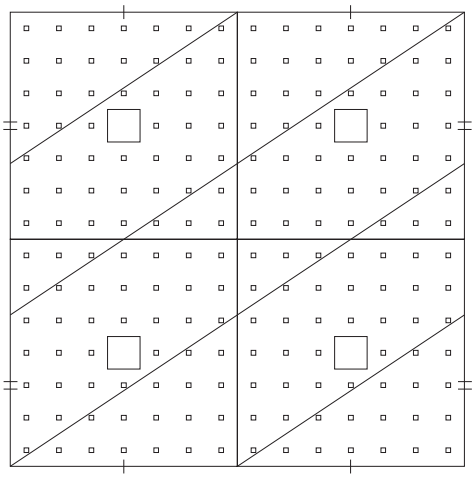


Figure 4.22: In this figure, we see a geodesic that of the translation surface $\mathcal{S}(S_{\mathbf{7},2})$ that intersects removable conic singularities of the surface. As such, the straight-line flow can continue unimpeded. Since the geodesic is equivalent to a billiard orbit in $\Omega(S_{\mathbf{u},2})$, reflection in the corners with angles measuring $\frac{\pi}{2}$ can be determined from how an equivalent geodesic passes through the corresponding removable singularity.

$\mathcal{S}(S_{\mathbf{a},n})$ is calculated as follows.

$$g_n = 1 + 2\sum_{m=1}^{n} a^{m-1}. \tag{4.22}$$

Then, for a sequence of rational (prefractal) billiard tables $\{\Omega(S_{\mathbf{a},n})\}_{n=0}^{\infty}$ approximating $\Omega(S_{\mathbf{a}})$, we have that $\lim_{n\to\infty} g_n = \infty$. Hence, if there is a translation surface associated with the (yet to be properly defined) self-similar billiard table $\Omega(S_{\mathbf{a}})$, then such a translation surface presumably has infinite genus. On the other hand, if such a surface is given as a suitable limit of translation surfaces $\mathcal{S}(S_{\mathbf{a},n})$, then such a surface will presumably have no surface area. In other words, it is reasonable to expect that a suitable translation surface cannot be constructed as a suitable limit of a sequence of translation surfaces $\mathcal{S}(S_{\mathbf{a},n})$.

4.5 Concluding remarks and open questions

In this chapter we have made progress in identifying a family of periodic orbits in the self-similar carpet billiard $\Omega(S_{\mathbf{a}})$. As of now, an orbit of a self-similar Sierpinski carpet billiard is the trivial limit of an eventually constant sequence of compatible periodic orbits. As an interim goal, we would like to identify sequences of compatible periodic orbits that are not eventually constant, but converge (in a suitable sense) to a set that constitutes a periodic orbit of a self-similar Sierpinski carpet billiard table $\Omega(S_{\mathbf{a}})$. Ultimately, we would like to construct a well-defined (topological) dynamical system on a self-similar Sierpinski carpet billiard table and determine whether or not there exists a topological dichotomy for the billiard flow.

Question 4.5.1. *If one can construct a well-defined notion of reflection for $\Omega(S_{\mathbf{a}})$, is it possible to show that, for a fixed direction α^0, the billiard flow in the direction α^0 is either closed or dense in $\Omega(S_{\mathbf{a}})$?*

In constructing the translation surface corresponding to $\Omega(S_{\mathbf{a},n})$, we have come up with a concrete example of a sequence of translation surfaces whose genera increase to infinity as $n \to \infty$. However, because $S_{\mathbf{a}}$ has zero Lebesgue area, it is unclear exactly how to describe the limit of the sequence of translation surfaces as a translation surface, let alone a surface with any specific structure.

Question 4.5.2. *Can one apply a suitable notion of limit to a sequence of translation surfaces $\{\mathcal{S}(S_{\mathbf{a},i})\}_{i=0}^{\infty}$ so as to recover a 'surface' with some structure that is analogous to that of a translation surface?*

If Question 4.5.2 can be answered in the affirmative, a natural question to ask is whether one can determine a flow on $\mathcal{S}(S_{\mathbf{a}})$, and whether such a flow can be shown to be equivalent to the (yet to be determined) billiard flow on $\Omega(S_{\mathbf{a}})$.

A related (and interesting) problem involves investigating the behavior of a sequence of compatible orbits of billiard tables which share a similar construction with the self-similar Sierpinski carpet billiard table $\Omega(S_{\mathbf{a}})$, but occupy nonzero area. A clear candidate is the non-self-similar Sierpinski carpet $S_{\mathbf{a}}$ with $\mathbf{a} = (a_1^{-1}, a_2^{-1}, \cdots)$ being a sequence in l^2 (*i.e.,* $\sum_{j=1}^{\infty} a_j^{-2} < \infty$). Since $\limsup \mathbf{a} = 0$ in this setting, [25, Theorem 5.3(a)] implies that $S_{\mathbf{a}}$ contains nontrivial line segments of every rational slope, and contains no nontrivial line segments of any irrational slope. One may then begin to discuss and analyze the behavior of a sequence of compatible orbits with each orbit having an initial direction given by $\alpha^0 \in \text{Slope}(S_{\mathbf{a}}) = \mathbb{Q}$.

A related Sierpinski carpet billiard we may consider investigating is a so-called *fat Sierpinski carpet* billiard table, constructed as follows. Begin by fixing a positive odd integer $a \geq 3$, and partition the unit square into a^2 many squares of side-length a^{-1}. Rather than removing the open middle square of side-length a^{-1} as in the first-stage of the construction of a self-similar Sierpinski carpet, one removes an open square of side-length $(1-\delta)a^{-1}$, $0 < \delta < 1$. In the second stage of the construction, a square of side-length $(1-\delta)a^{-2}$ centered within each of the remaining $a^2 - 1$ squares of side-length a^{-1} is removed. This process is then repeated ad infinitum. In effect, each removed square of side-length $(1-\delta)a^{-n}$ is surrounded by a solid square annulus of width $(\delta/2)a^{-n}$. Such a carpet admits nontrivial line segments of irrational slopes, and there will be additional base points from which a nontrivial line segment can emanate. We can then ask the following question.

Question 4.5.3. *How does the existence and behavior of a sequence of compatible dense orbits depend on δ? Moreover, can one recover a dense orbit of the self-similar Sierpinski carpet billiard $\Omega(S_{\mathbf{a}})$ in the limit as $\delta \to 0$?*

The development of our understanding of the Sierpinski carpet billiard table may be

possibly aided by studying the results for other types of billiard tables. In particular, the so-called wind-tree billiard tables (and generalizations of them) investigated in [22, 24, 39] are very reminiscent of a Sierpinski carpet billiard table.

Alternatively, we may be able to approach the problem of determining billiard dynamics on $\Omega(S_{\mathbf{a}})$ by taking an algebraic perspective. For each translation surface $\mathcal{S}(S_{\mathbf{a},n})$, there is an associated *Veech group*. A Veech group of a translation surface $\mathcal{S}(D)$ determined from a rational billiard table $\Omega(D)$ is the stabilizer of $\mathcal{S}(D)$. The work of [70] may prove useful in determining a well-defined translation structure on a suitable limit of a sequence of translation surfaces $\mathcal{S}(S_{\mathbf{a},n})$. In particular, analyzing the inverse limit of an inverse limit sequence of Veech groups $\{\Gamma(S_{\mathbf{a},n})\}_{n=0}^{\infty}$ (assuming such an inverse limit sequence can be properly constructed) may yield a group that is the stabilizer of some surface. With such a surface and an associated stabilizer defined, we may begin to investigate whether or not the proposed surface contains saddle connections. The directions of such saddle connections may then indicate directions for which closed geodesics are occurring.

Acknowledgements

We wish to thank Jeremy Tyson for several useful discussions regarding his work with Estibalitz Durand-Cartagena. We also thank our respective Ph.D. advisors, Robert S. Strichartz and Michel L. Lapidus, for their support of the project.

BIBLIOGRAPHY

[1] E. Akkermans, G. V. Dunne, and A. Teplyaev, *Thermodynamics of photons on fractals*, Phys. Rev. Lett. **105** (2010), 230407.

[2] M. T. Barlow, *Analysis on the Sierpinski Carpet*, Analysis and Geometry of Metric Measure Spaces: Lecture Notes of the 50th Seminaire De Mathematiques Superieures (SMS), Montreal, 2011 (CRM Proceedings & Lecture Notes), 2013, pp. 27–53.

[3] M. T. Barlow and R. F. Bass, *The construction of Brownian motion on the Sierpiński carpet*, Ann. Inst. H. Poincaré Probab. Statist. **25** (1989), no. 3, 225–257 (English, with French summary). MR1023950 (91d:60183)

[4] ———, *Local times for Brownian motion on the Sierpiński carpet*, Probab. Theory Related Fields **85** (1990), no. 1, 91–104, DOI 10.1007/BF01377631. MR1044302 (91j:60129)

[5] ———, *Transition densities for Brownian motion on the Sierpiński carpet*, Probab. Theory Related Fields **91** (1992), no. 3-4, 307–330. MR1151799 (93k:60203)

[6] ———, *Brownian motion and harmonic analysis on Sierpinski carpets*, Canad. J. Math. **51** (1999), no. 4, 673–744. MR1701339 (2000i:60083)

[7] ———, *Random walks on graphical Sierpinski carpets*, Sympos. Math., XXXIX, Cambridge Univ. Press, Cambridge, 1999. MR1802425 (2002c:60116)

[8] M. T. Barlow, R. F. Bass, T. Kumagai, and A. Teplyaev, *Uniqueness of Brownian motion on Sierpiński carpets*, J. Eur. Math. Soc. (JEMS) **12** (2010), no. 3, 655–701. MR2639315 (2011i:60146)

[9] M. T. Barlow and E. A. Perkins, *Brownian motion on the Sierpiński gasket*, Probab. Theory Related Fields **79** (1988), no. 4, 543–623. MR966175 (89g:60241)

[10] M. Begué, T. Kalloniatis, and R. S. Strichartz, *Harmonic functions and the spectrum of the Laplacian on the Sierpinski carpet*, Fractals **21** (2013), no. 1, 1350002, 32. MR3042410

[11] O. Ben-Bassat, R. S. Strichartz, and A. Teplyaev, *What is not in the domain of the Laplacian on Sierpinski gasket type fractals*, J. Funct. Anal. **166** (1999), no. 2, 197–217. MR1707752 (2001e:31016)

[12] T. Berry, S. M. Heilman, and R. S. Strichartz, *Outer approximation of the spectrum of a fractal Laplacian*, Experiment. Math. **18** (2009), no. 4, 449–480. MR2583544 (2010m:28006)

[13] O. Bratteli and D. W. Robinson, *Operator algebras and quantum statistical mechanics. 2*, 2nd ed., Texts and Monographs in Physics, Springer-Verlag, Berlin, 1997. Equilibrium states. Models in quantum statistical mechanics. MR1441540 (98e:82004)

[14] R. Burioni, D. Cassi, M. Rasetti, P. Sodano, and A. Vezzani, *Bose-Einstein condensation on inhomogeneous complex networks*, J. Phys. B: At. Mol. Opt. Phys. **34** (2001), 4697.

[15] R. Burioni, D. Cassi, and A. Vezzani, *Topology, hidden spectra and Bose-Einstein condensation on low-dimensional complex networks*, J. Phys. A **35** (2002), no. 5, 1245–1252. MR1891638 (2003b:82005)

[16] D. Cassi, *Phase transitions and random walks on graphs: a generalization of the Mermin-Wagner theorem to disordered lattices, fractals, and other discrete structures*, Phys. Rev. Lett. **68** (1992), 3631–3634.

[17] J. P. Chen, *Statistical mechanics of Bose gas in Sierpinski carpets*, arXiv:1202.1274 (2012).

[18] J. P. Chen and R. G. Niemeyer, *Periodic billiard orbits of self-similar Sierpinski carpets*, arXiv:1303.4032 (2013).

[19] J. P. Chen and R. S. Strichartz, *Spectral asymptotics and heat kernels on three-dimensional fractal sponges*, preprint (2013).

[20] J. P. Chen and B. E. Ugurcan, *Entropic repulsion of Gaussian free field on high-dimensional Sierpinski carpet graphs*, arXiv:1307.5825 (2013).

[21] Z.-Q. Chen and M. Fukushima, *Symmetric Markov processes, time change, and boundary theory*, London Mathematical Society Monographs Series, vol. 35, Princeton University Press, Princeton, NJ, 2012. MR2849840 (2012m:60176)

[22] J.-P. Conze and E. Gutkin, *On recurrence and ergodicity for geodesic flows on non-compact periodic polygonal surfaces*, Ergodic Theory Dynam. Systems **32** (2012), no. 2, 491–515. MR2901357

[23] G. Dal Maso, *An introduction to* Γ*-convergence*, Progress in Nonlinear Differential Equations and their Applications, 8, Birkhäuser Boston Inc., Boston, MA, 1993. MR1201152 (94a:49001)

[24] V. Delecroix, *Divergent directions in some periodic wind-tree models*, arXiv:1107.2418 (2011).

[25] E. Durand-Cartagena and J. T. Tyson, *Rectifiable curves in Sierpiński carpets*, Indiana Univ. Math. J. **60** (2011), no. 1, 285–309. MR2952419

[26] F. J. Dyson, E. H. Lieb, and B. Simon, *Phase transitions in quantum spin systems with isotropic and nonisotropic interactions*, J. Statist. Phys. **18** (1978), no. 4, 335–383. MR0496246 (58 #14819)

[27] K. Falconer, *Fractal geometry*, 2nd ed., John Wiley & Sons Inc., Hoboken, NJ, 2003. Mathematical foundations and applications. MR2118797 (2006b:28001)

[28] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk, *Fatou's lemma for weakly converging probabilities*, arXiv:1206.4073 (2012).

[29] F. Fidaleo, *Harmonic analysis on perturbed Cayley trees*, J. Funct. Anal. **261** (2011), no. 3, 604–634. MR2799573

[30] F. Fidaleo, D. Guido, and T. Isola, *Bose-Einstein condensation on inhomogeneous amenable graphs*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **14** (2011), no. 2, 149–197. MR2813486

[31] J. Fröhlich, B. Simon, and T. Spencer, *Infrared bounds, phase transitions and continuous symmetry breaking*, Comm. Math. Phys. **50** (1976), no. 1, 79–95. MR0421531 (54 #9530)

[32] M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet forms and symmetric Markov processes*, extended, de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 2011. MR2778606 (2011k:60249)

[33] E. Gutkin, *Billiards on almost integrable polyhedral surfaces*, Ergodic Theory Dynam. Systems **4** (1984), no. 4, 569–584. MR779714 (86m:58123)

[34] ———, *Billiards in polygons: survey of recent results*, J. Statist. Phys. **83** (1996), no. 1-2, 7–26, DOI 10.1007/BF02183637. MR1382759 (97a:58099)

[35] E. Gutkin and C. Judge, *The geometry and arithmetic of translation surfaces with applications to polygonal billiards*, Math. Res. Lett. **3** (1996), no. 3, 391–403. MR1397686 (97c:58116)

[36] B. M. Hambly, *Asymptotics for functions associated with heat flow on the Sierpinski carpet*, Canad. J. Math. **63** (2011), no. 1, 153–180. MR2779136 (2012f:35548)

[37] B. M. Hambly, T. Kumagai, S. Kusuoka, and X. Y. Zhou, *Transition density estimates for diffusion processes on homogeneous random Sierpinski carpets*, J. Math. Soc. Japan **52** (2000), no. 2, 373–408. MR1742797 (2001e:60158)

[38] M. Hino, *On singularity of energy measures on self-similar sets*, Probab. Theory Related Fields **132** (2005), no. 2, 265–290. MR2199293 (2007g:28010)

[39] P. Hubert, S. Lelièvre, and S. Troubetzkoy, *The Ehrenfest wind-tree model: periodic directions, recurrence, diffusion*, J. Reine Angew. Math. **656** (2011), 223–244. MR2818861 (2012f:37079)

[40] P. Hubert and T. A. Schmidt, *An introduction to Veech surfaces*, Handbook of dynamical systems. Vol. 1B, Elsevier B. V., Amsterdam, 2006, pp. 501–526. MR2186246 (2006i:37099)

[41] J. E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), no. 5, 713–747. MR625600 (82h:49026)

[42] N. Kajino, *Spectral asymptotics for Laplacians on self-similar sets*, J. Funct. Anal. **258** (2010), no. 4, 1310–1360, DOI 10.1016/j.jfa.2009.11.001. MR2565841 (2011j:31010)

[43] ———, *Log-periodic asymptotic expansion of the spectral partition function for self-similar sets*, to appear in Comm. Math. Phys. (2013).

[44] A. B. Katok and A. N. Zemljakov, *Topological transitivity of billiards in polygons*, Mat. Zametki **18** (1975), no. 2, 291–300 (Russian). MR0399423 (53 #3267)

[45] J. Kigami, *Analysis on fractals*, Cambridge Tracts in Mathematics, vol. 143, Cambridge University Press, Cambridge, 2001. MR1840042 (2002c:28015)

[46] K. Kirsten, *Basic zeta functions and some applications in physics*, A window into zeta and modular physics, Math. Sci. Res. Inst. Publ., vol. 57, Cambridge Univ. Press, Cambridge, 2010, pp. 101–143. MR2648362 (2012a:11135)

[47] G. Kozma and A. Nachmias, *The Alexander-Orbach conjecture holds in high dimensions*, Invent. Math. **178** (2009), no. 3, 635–654. MR2551766 (2011a:60350)

[48] T. Kumagai, *Brownian motion penetrating fractals: an application of the trace theorem of Besov spaces*, J. Funct. Anal. **170** (2000), no. 1, 69–92. MR1736196 (2001a:31009)

[49] T. Kumagai and K.-T. Sturm, *Construction of diffusion processes on fractals, d-sets, and general metric measure spaces*, J. Math. Kyoto Univ. **45** (2005), no. 2, 307–327. MR2161694 (2006i:60113)

[50] S. Kusuoka, *Dirichlet forms on fractals and products of random matrices*, Publ. Res. Inst. Math. Sci. **25** (1989), no. 4, 659–680. MR1025071 (91m:60142)

[51] S. Kusuoka and X. Y. Zhou, *Dirichlet forms on fractals: Poincaré constant and resistance*, Probab. Theory Related Fields **93** (1992), no. 2, 169–196. MR1176724 (94e:60069)

[52] T. Matsui, *BEC of free bosons on networks*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **9** (2006), no. 1, 1–26, DOI 10.1142/S0219025706002202. MR2214499 (2007h:82009)

[53] M. L. Lapidus and R. G. Niemeyer, *Towards the Koch snowflake fractal billiard: computer experiments and mathematical conjectures*, Gems in experimental mathematics, Contemp. Math., vol. 517, Amer. Math. Soc., Providence, RI, 2010, pp. 231–263. MR2731085 (2012b:37101)

[54] ———, *Towards the Koch snowflake fractal billiard: computer experiments and mathematical conjectures*, Gems in experimental mathematics, Contemp. Math., vol. 517, Amer. Math. Soc., Providence, RI, 2010, pp. 231–263. MR2731085 (2012b:37101)

[55] ———, *The current state of fractal billiards*, to appear in Contemporary Mathematics (2013).

[56] M. L. Lapidus and M. van Frankenhuijsen, *Fractal geometry, complex dimensions and zeta functions*, Springer Monographs in Mathematics, Springer, New York, 2006. Geometry and spectra of fractal strings. MR2245559 (2007j:11001)

[57] E. H. Lieb, R. Seiringer, J. P. Solovej, and J. Yngvason, *The mathematics of the Bose gas and its condensation*, Oberwolfach Seminars, vol. 34, Birkhäuser Verlag, Basel, 2005. MR2143817 (2006e:82001)

[58] H. Masur and S. Tabachnikov, *Rational billiards and flat structures*, Handbook of dynamical systems, Vol. 1A, North-Holland, Amsterdam, 2002, pp. 1015–1089. MR1928530 (2003j:37002)

[59] P. Mattila, *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics, vol. 44, Cambridge University Press, Cambridge, 1995. Fractals and rectifiability. MR1333890 (96h:28006)

[60] I. McGillivray, *Resistance in higher-dimensional Sierpiński carpets*, Potential Anal. **16** (2002), no. 3, 289–303. MR1885765 (2003e:31006)

[61] F. Merkl and H. Wagner, *Recurrent random walks and the absence of continuous symmetry breaking on graphs*, J. Statist. Phys. **75** (1994), no. 1-2, 153–165. MR1273056 (95b:82013)

[62] N. D. Mermin and H. Wagner, *Absence of ferromagnetism or antiferromagnetism in one-or two-dimensional isotropic Heisenberg models*, Phys. Rev. Lett. **17** (1966), 1133–1136.

[63] U. Mosco, *Composite media and asymptotic Dirichlet forms*, J. Funct. Anal. **123** (1994), no. 2, 368–421. MR1283033 (95d:47088)

[64] E. Nelson, *Construction of quantum fields from Markoff fields*, J. Functional Analysis **12** (1973), 97–112. MR0343815 (49 #8555)

[65] ———, *The free Markoff field*, J. Functional Analysis **12** (1973), 211–227. MR0343816 (49 #8556)

[66] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An introduction to the theory of numbers*, 5th ed., John Wiley & Sons Inc., New York, 1991. MR1083765 (91i:11001)

[67] R. Pemantle and J. E. Steif, *Robust phase transitions for Heisenberg and other models on general trees*, Ann. Probab. **27** (1999), no. 2, 876–912. MR1698979 (2000e:60184)

[68] M. Röckner, *Generalized Markov fields and Dirichlet forms*, Acta Appl. Math. **3** (1985), no. 3, 285–311. MR790552 (86k:60091)

[69] M. Schäl, *Average optimality in dynamic programming with general state space*, Math. Oper. Res. **18** (1993), no. 1, 163–172. MR1250112 (94h:90068)

[70] G. Schmithüsen, *An algorithm for finding the Veech group of an origami*, Experiment. Math. **13** (2004), no. 4, 459–472. MR2118271 (2006b:30080)

[71] R. Serfozo, *Convergence of Lebesgue integrals with varying measures*, Sankhyā Ser. A **44** (1982), no. 3, 380–402. MR705462 (84g:28003)

[72] J. Smillie, *The dynamics of billiard flow in rational polygons*, Dynamical systems, ergodic theory and applications, 2000, pp. xii+459. Edited and with a preface by Sinai; Translated from the Russian; Mathematical Physics, I. MR1758456 (2001k:37004)

[73] B. Steinhurst and A. Teplyaev, *Existence of a Meromorphic Extension of Spectral Zeta Functions on Fractals*, to appear in Lett. Math. Phys. (2013).

[74] R. S. Strichartz, *Differential equations on fractals*, Princeton University Press, Princeton, NJ, 2006. A tutorial. MR2246975 (2007f:35003)

[75] K.-T. Sturm, *How to construct diffusion processes on metric spaces*, Potential Anal. **8** (1998), no. 2, 149–161. MR1618442 (99c:60175)

[76] S. Tabachnikov, *Billiards*, Panor. Synth. **1** (1995), vi+142 (English, with English and French summaries). MR1328336 (96c:58134)

[77] ———, *Geometry and billiards*, Student Mathematical Library, vol. 30, American Mathematical Society, Providence, RI, 2005. MR2168892 (2006h:51001)

[78] D. I. Tsomokos, S. Ashhab, and F. Nori, *Using superconducting qubit circuits to engineer exotic lattice systems*, Phys. Rev. A **82** (2010), 052311.

[79] J. Weidmann, *Strong operator convergence and spectral theory of ordinary differential operators*, Univ. Iagel. Acta Math. **34** (1997), 153–163. MR1458041 (98d:47053)

[80] H. Weyl, *Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung)*, Math. Ann. **71** (1912), no. 4, 441–479 (German). MR1511670

[81] A. Zorich, *Flat surfaces*, Frontiers in number theory, physics, and geometry. I, Springer, Berlin, 2006, pp. 437–583. MR2261104 (2007i:37070)