On the resistance of the Sierpiński carpet

BY M. T. BARLOW[1] AND R. F. BASS[2]

[1] *Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, U.K.*

[2] *Department of Mathematics, University of Washington, Seattle, Washington 98195, U.S.A.*

Let F_n be the nth stage in the construction of the Sierpiński carpet. Let R_n be the electrical resistance of F_n when the left and right sides are each short-circuited, and a voltage is applied between them. We prove that there exists a constant ρ such that $\frac{1}{4}\rho^n \leqslant R_n \leqslant 4\rho^n$. The motivation for this result came from the problem of establishing (*a*) the existence and (*b*) the value of the 'spectral dimension' of the Sierpiński carpet. In this and a subsequent paper, we settle (*a*) and give bounds for (*b*).

1. Introduction

The basic Sierpiński carpet (SC) is the set in $\mathbb{R}^2$ formed by a two-dimensional analogue of the construction of the Cantor set (see Sierpiński 1916; Mandelbrot 1982). Starting with the closed unit square $F_0 = (0,1)^2$, F_0 is divided into 9 equal squares, and the interior of the central one is removed, to obtain a set $F_1 = F_0 - (\frac{1}{3}, \frac{2}{3})^2$. This operation is then repeated on the eight closed squares that remain, and is continued indefinitely. We denote the set remaining at the nth stage by F_n: note that F_n consists of 8^n squares, each of side 3^{-n}. The Sierpiński carpet F is defined by

$$F = \bigcap_{n=0}^{\infty} F_n.$$

In this paper we examine the following problem. Consider F_n to be a thin sheet of conducting material, of unit conductivity. Let R_n be the effective resistance of F_n when two opposite edges of F_n are each short-circuited, and a voltage is applied between them. (The definition of effective resistance is made precise in §2.)

Theorem 1.1. *There exists a constant $\rho > 1$ such that*

$$\tfrac{1}{4}\rho^n \leqslant R_n \leqslant 4\rho^n, \quad n \geqslant 0. \tag{1.1}$$

Our proof uses a sub-additivity argument, and does not provide the value of ρ. However, elementary shorting and cutting arguments show that $\frac{7}{6} \leqslant \rho \leqslant \frac{3}{2}$ and the computer calculations in Barlow *et al.* (1990) suggest that $1.25147 < \rho < 1.25149$.

See Gefen *et al.* (1984) and Ben-Avraham & Havlin (1983) for previous theoretical work on the resistances of the F_n and Yuan & Tao (1986*a*, *b*) for experimental studies.

The motivation for Theorem 1.1 came from the problem of establishing (*a*) the existence and (*b*) the value, d_s, of the 'spectral dimension' of the SC. See Rammal & Toulouse (1983) and Havlin & Ben-Avraham (1987) for a discussion of the physics

Proc. R. Soc. Lond. A (1990) **431**, 345–360
Printed in Great Britain

aspects of spectral dimension, and see Ben-Avraham & Havlin (1983), Hattori *et al.* (1985), Osada (1990) and Watanabe (1985) for work on the spectral dimension of the SC and the connection between d_s and ρ.

In this and a subsequent paper (on transition densities for brownian motion on the Sierpiński carpet) we settle (*a*) and give bounds for (*b*). Our solution comes in two parts, of which the first, in this paper, is the proof of Theorem 1.1.

The second step is probabilistic and is based on our earlier work in Barlow & Bass (1989), in which we constructed a diffusion process X_t on F. Let W^n be brownian motion on F_n (with normal reflection on ∂F_n), and let α_n be the mean time it takes W^n to cross F_n. Then if $X^n_t = W^n_{\alpha_n t}$, the laws of the processes X^n are tight (Barlow & Bass 1989, Theorem 5.1), and the process X is obtained as the weak limit of X^{n_k} for suitable subsequences (n_k). In Barlow & Bass (1990) we proved that

$$c_1(\tfrac{8}{9})^n R_n \leqslant \alpha_n \leqslant c_2(\tfrac{8}{9})^n R_n. \tag{1.2}$$

The process X_t has a continuous transition density $p_t(x, y)$ with respect to Hausdorff $x^{\ln 8/\ln 3}$ measure on F, and the 'spectral dimension' d_s of X can be defined by

$$-\tfrac{1}{2}d_s = \lim_{t\downarrow 0} \frac{\ln p_t(x,x)}{\ln t}, \quad x \in F \tag{1.3}$$

provided that this limit exists, and is independent of x. (See Barlow & Perkins (1988) and Lindstrøm (1990) for proofs of the existence of d_s on other fractals.) It turns out that the numbers α_n are crucial in estimating $p_t(x, y)$, and that Theorem 1.1 (or a slightly weaker result) is necessary to establish the existence of the limit in (1.3). In a future paper we prove that d_s exists, that

$$d_s = 2\ln(8)/\ln(8\rho), \tag{1.4}$$

and that d_s is the 'density of states' of the infinitesimal generator of X.

There is no difficulty in considering more general carpets, and in fact we shall do so. Thus we fix $k \geqslant 3$, and, starting as before with the unit square F_0, we divide each square into k^2 subsquares, and at each stage remove the same fixed pattern of subsquares. The exact conditions we impose on this pattern are given in §3, but in essence the conditions ensure that $\operatorname{int}(F_1)$ is connected and symmetric.

The structure of this paper is as follows. In §2 we make precise our definitions of resistance, current, etc. This simply amounts to translating into the continuous case the work of Doyle & Snell (1984) on networks. The key result is a pair of variational principles. The first defines R_n^{-1} as the infimum of the energy of potentials on F_n with unit potential difference across F_n. The second gives R_n as the infimum of the energy of currents of unit flux flowing across F_n. Together, these provide a powerful technique for obtaining two-sided bounds on R_n, by constructing a feasible potential and a feasible current on F_n.

Let G_n be the network obtained from F_n by replacing each square of side k^{-n} by a horizontal and vertical crosswire of four linear resistors, each of resistance $\frac{1}{2}$. Let D_n be the network obtained from F_n by replacing each square of side k^{-n} by a diagonal crosswire of 4 resistors, each of resistance 1, and let R_n^G, R_n^D be the effective resistances of G_n and D_n when two opposite edges are short-circuited.

In §3 we use the optimum currents on F_n and G_m to construct a feasible current K on F_{n+m}. Using the variational principle this gives

$$R_{n+m} \leqslant R_n R_m^G, \quad n, m \geqslant 0. \tag{1.5}$$

In §4, using potentials we prove in a similar fashion that

$$\tfrac{1}{2}R_m^G \leqslant R_m^D \leqslant R_m^G, \tag{1.6}$$

and

$$\tfrac{1}{2}R_m^D R_n \leqslant R_{n+m}. \tag{1.7}$$

In §5 we combine these estimates to prove Theorem 1.1.

Some of the results of this paper were announced in Barlow *et al.* (1990).

2. Preliminaries on resistance

We begin by recalling (Doyle & Snell 1984) the definition of a wire network, and of effective resistance in the context of a wire network. See also Beurling & Deny (1958, 1959) for the very close connection with Dirichlet forms.

Let G be a finite set, and $g: G \times G \rightarrow \mathbb{R}$ be a function satisfying

$$g(x,y) = g(y,x), \quad g(x,y) \geqslant 0, \quad g(x,x) = 0, \quad x, y \in G. \tag{2.1}$$

We will call the pair (G, g) a wire network, and when there is no danger of confusion we will sometimes abuse terminology by referring to the network G. Associated with (G, g) is the Dirichlet form

$$\mathscr{E}_G(f,f) = \tfrac{1}{2} \sum_{x \in G} \sum_{y \in G} g(x,y)\,(f(x)-f(y))^2. \tag{2.2}$$

Let A, B be disjoint subsets of G. The effective resistance between A and B in G, denoted $R_G(A, B)$, is defined by

$$R_G(A,B)^{-1} = \inf\{\mathscr{E}_G(f,f): f = 0 \text{ on } A, f = 1 \text{ on } B\}. \tag{2.3}$$

The infimum in (2.3) is attained by a unique function u_G; we will refer to any function f which satisfies the constraints of (2.3) as a feasible potential for the potential problem pot (G, A, B), and the potential u_G as the optimum potential for the problem pot (G, A, B).

The set of wires in the network (G, g) is defined by

$$W(G) = \{(x,y) \in G \times G: g(x,y) > 0\}.$$

Set $N(x) = \{y: (x,y) \in W(G)\}$. A flow or current from A or B in G is a function $I: W(G) \rightarrow \mathbb{R}$ satisfying

$$I(x,y) = -I(y,x), \quad \sum_y I(x,y) = 0 \text{ if } x \in G-(A \cup B). \tag{2.4}$$

The total flux of I is given by

$$T(I; A, B) = \sum_{x \in A} \sum_{y \in N(x)} I(x,y) = -\sum_{x \in B} \sum_{y \in N(x)} I(x,y).$$

Define the energy dissipation of the current I by

$$E_G(I,I) = \tfrac{1}{2} \sum_{(x,y) \in W(G)} g(x,y)^{-1} I(x,y)^2. \tag{2.5}$$

Then (Doyle & Snell 1984, p. 63)

$$R_G(A,B) = \inf\{E_G(I,I)\colon I \text{ is a current from } A \text{ to } B \text{ in } G \text{ with total flux } 1\}. \quad (2.6)$$

We refer to any current I from A to B with total flux 1 as a feasible current for the current problem cur (G, A, B).

If $f\colon G \to \mathbb{R}$ define $\nabla f\colon W(G) \to \mathbb{R}$ by

$$\nabla f(x,y) = (f(y) - f(x))\, g(x,y),$$

and note that

$$\mathscr{E}_G(f,f) = \tfrac{1}{2} \sum_{(x,y)\in W} g(x,y)^{-1} |\nabla f(x,y)|^2. \quad (2.7)$$

The link between (2.3) and (2.6) arises from the fact that ∇u_G is a current from A to B with total flux $R_G(A,B)^{-1}$ (see Doyle & Snell 1984, p. 64) and $I_G = R_G(A,B)\,\nabla u_G$ is the optimum current for the problem cur (G, A, B). Note that

$$R_G(A,B)^{-1} = \mathscr{E}_G(u_G, u_G) = E_G(\nabla u_G, \nabla u_G) = R_G(A,B)^{-2} E_G(I_G, I_G).$$

We now turn to the continuum case. Let F be a Lipschitz domain in $\mathbb{R}^2$, and let B_0, B_1 be disjoint closed subsets of ∂F. We write σ for surface measure on ∂F, and n for the interior normal. Define the effective resistance $R_F(B_0, B_1)$ between B_0 and B_1 in F by

$$R_F(B_0,B_1)^{-1} = \inf\left\{\int_F |\nabla u|^2\,\mathrm{d}x : u(x) = i \text{ on } B_i,\ i = 0,1,\ u \in C(\bar{F}) \cap H^1(F)\right\}. \quad (2.8)$$

We write

$$\mathscr{E}_F(u,u) = \int_F |\nabla u|^2;$$

the Dirichlet form $\mathscr{E}_F(\cdot,\cdot)$ is simply the 'energy' of the function u. The space $H^1(F)$ is the set of functions on F with finite energy, endowed with the energy norm. We will refer to a function u satisfying the conditions of (2.8) as a feasible function, or feasible potential for the problem pot (F, B_0, B_1).

We will call a vector field $J = (J_1(x), J_2(x))$ on F a current or flow between B_0 and B_1 in F if

$$\left.\begin{array}{ll} \text{(i)} & J \in BV(F), \\ \text{(ii)} & \operatorname{div} J = 0 \text{ (in the sense of distributions) on } F, \\ \text{(iii)} & J\cdot n = 0\ \sigma\text{-a.e. on } \partial F - (B_0 \cup B_1). \end{array}\right\} \quad (2.9)$$

For the space $BV(F)$ see Maz'ja (1985, ch. 6) or Ziemer (1989, ch. 5); this is a space of functions of bounded variation. The energy of the flow J is given by

$$E_F(J,J) = \int_F |J|^2.$$

If J is a flow between B_0 and B_1 note that

$$\int_{B_0} J\cdot n\,\mathrm{d}\sigma = -\int_{B_1} J\cdot n\,\mathrm{d}\sigma = T(J, B_0, B_1), \quad \text{say},$$

and we call $T(J, B_0, B_1)$ the total flux of J from B_0 to B_1.

We will use Green's Theorem in the following form.

Lemma 2.1. ('Conservation of energy' (Doyle & Snell 1984, p. 61).) *Let J be a flow from B_0 to B_1 in F, and u be a feasible function for the problem* pot (F, B_0, B_1). *Then*

$$\int_F \nabla u \cdot J = -\int_{B_1} J \cdot n \, \mathrm{d}\sigma. \tag{2.10}$$

Proof. By Green's Theorem (see Maz'ja 1985, p. 340), and the hypotheses on u and J,

$$\begin{aligned}\int_F \nabla u \cdot J + \int_F u \operatorname{div} J &= -\int_{\partial F} u(J \cdot n)\, \mathrm{d}\sigma \\ &= -\int_{B_0} u(J \cdot n)\, \mathrm{d}\sigma - \int_{B_1} u(J \cdot n)\, \mathrm{d}\sigma \\ &= -\int_{B_1} J \cdot n \, \mathrm{d}\sigma.\end{aligned}$$

□

Now let v be the solution to the mixed boundary value problem

$$\left.\begin{aligned} \Delta v &= 0 && \text{in } F, \\ v &= i && \text{on } B_i \quad (i = 0, 1), \\ \partial v / \partial n &= 0 \text{ a.e. } \mathrm{d}\sigma && \text{on } F - (B_0 \cup B_1). \end{aligned}\right\} \tag{2.11}$$

Remark. The question of regularity of mixed boundary value problems such as (2.11) in Lipschitz domains appears to be still open. In what follows we will assume that F, B_0, B_1 are sufficiently regular so that a solution v to (2.11) exists with $v \in C(\bar{F}) \cap H^1(F)$ and $\nabla v \in L^2(\mathrm{d}\sigma)$. In the Appendix we will show that this is indeed the case for the approximations to Sierpiński carpets considered in §§3 and 4.

Proposition 2.2. *The function v defined by* (2.11) *minimizes the variational problem* pot (F, B_0, B_1), *and is the unique function which does so. We have*

$$R_F(B_0, B_1)^{-1} = \mathscr{E}_F(v, v).$$

Proof. Let u be any feasible function for the problem pot (F, B_0, B_1). Note that $J = \nabla v$ is a flow between B_0 and B_1, so that by Lemma 2.1

$$\mathscr{E}_F(u, v) = \int_F \nabla u \cdot J = -\int_{B_1} J \cdot n \, \mathrm{d}\sigma.$$

The right-hand side is independent of u, so $\mathscr{E}_F(u, v) = \mathscr{E}_F(v, v)$, and therefore

$$\begin{aligned}\mathscr{E}_F(u, u) - \mathscr{E}_F(v, v) &= \mathscr{E}_F(u - v, u - v) + 2\mathscr{E}_F(u - v, v) \\ &= \mathscr{E}_F(u - v, u - v) \geqslant 0.\end{aligned}$$

Thus $\mathscr{E}_F(u, u) \geqslant \mathscr{E}_F(v, v)$, with equality only if $u = v$. □

Now consider the second variational problem

$$\mu = \inf\left\{\int_F |J|^2 : J \text{ is a flow between } B_0 \text{ and } B_1, \text{ with } T(J, B_0, B_1) = 1\right\}. \tag{2.12}$$

We call a current J which satisfies the constraint in (2.12) a feasible current for the problem $\operatorname{cur}(F, B_0, B_1)$.

Theorem 2.3. *The current* $I = R_F(B_0, B_1)\nabla v$ *is the unique flow which minimizes* $\operatorname{cur}(F, B_0, B_1)$. *We have* $\mu = R_F(B_0, B_1)$.

Proof. Note first that I is a flow between B_0 and B_1. The total flux of I is given by

$$\begin{aligned} T(I, B_0, B_1) &= -\int_{B_1} I\cdot n\,\mathrm{d}\sigma \\ &= \int_F \nabla v\cdot I\,\mathrm{d}\sigma \qquad \text{(by (2.10))} \\ &= R_F(B_0, B_1)\,\mathscr{E}_F(v, v) = 1. \end{aligned}$$

So I is feasible for the problem $\operatorname{cur}(F, B_0, B_1)$. Also,

$$E_F(I, I) = R_F(B_0, B_1)^2 \mathscr{E}_F(v, v) = R_F(B_0, B_1).$$

Let J be any other feasible flow. Then

$$\begin{aligned} E_F(I, J) &= R_F(B_0, B_1)\int_F \nabla v\cdot J \\ &= -R_F(B_0, B_1)\int_{B_1} J\cdot n\,\mathrm{d}\sigma = R_F(B_0, B_1). \end{aligned}$$

So $E_F(I, I) = E_F(I, J)$, and hence

$$E_F(J, J) - E_F(I, I) = E_F(I-J, I-J) \geqslant 0.$$

Thus I minimizes (2.12), and is the unique minimizing flow. □

Remarks

1. Note that if u, J are, respectively, a feasible potential and a feasible flow for the problems $\operatorname{pot}(F, B_0, B_1)$ and $\operatorname{cur}(F, B_0, B_1)$, then

$$\mathscr{E}_F(u, u)^{-1} \leqslant R_F(B_0, B_1) \leqslant E_F(J, J).$$

So Proposition 2.2 and Theorem 2.3 together enable us to obtain two-sided bounds on $R_F(B_0, B_1)$.

2. As the minimizing functions in (2.8) and (2.12) are unique, they will satisfy any symmetries of the set-up (B_0, B_1, F).

3. Let $\lambda > 0$, and for $A \subset \mathbb{R}^2$ set $\lambda A = \{\lambda x, x \in A\}$. Then $R_{\lambda G}(\lambda B_0, \lambda B_1) = R_G(B_0, B_1)$: resistance in two dimensions in 'scale invariant'. If we considered instead $\mathbb{R}^d$, we would have a scaling factor of λ^{2-d}.

3. Flows and the upper bound

Let $F_0 = [0, 1]^2$, and let $k \geqslant 3$ be fixed. Let $\mathscr{S}_n$ be the collection of closed squares of side k^{-n} with corners in $k^{-n}\mathbb{Z}^2$. Given a set $A \subseteq \mathbb{R}^2$, set

$$\mathscr{S}_n(A) = \{S : S \subset A, S \in \mathscr{S}_n\}.$$

For $S \in \mathscr{S}_n$, let Ψ_S be the orientation preserving linear map which maps F_0 onto S.

We now define a decreasing sequence (F_n) of closed subsets of F_0. Let $R \geqslant 1$, and let F_1 be the union of R distinct elements of $\mathscr{S}_1(F_0)$. We impose the following conditions on F_1:

$$\left.\begin{array}{c} \text{H1: (Symmetry) } F_1 \text{ is preserved by all the isometries that preserve} \\ \text{the unit square } F_0. \\ \text{H2: (Connectedness) int}\,(F_1) \text{ is connected, and contains a path} \\ \text{connecting the lines } \{x_1 = 0\} \text{ and } \{x_1 = 1\}. \\ \text{H3: (Non-diagonality) The boundary } \partial F_1 \text{ of } F_1 \text{ consists of a finite} \\ \text{number of disjoint Jordan curves.} \end{array}\right\} \tag{3.1}$$

We may think of F_1 as being derived from F_0 by removing the interiors of a number (k^2-R) of squares in $\mathscr{S}_1(F_0)$. Given F_1, F_2 is obtained by removing the same pattern from each of the squares in $\mathscr{S}_1(F_1)$. Iterating, we obtain a sequence (F_n), where F_n is the union of R^n squares in $\mathscr{S}_n(F_0)$. Formally, we have

$$F_{n+1} = \bigcup_{S \in \mathscr{S}_n(F_n)} \Psi_S(F_1) = \bigcup_{S \in \mathscr{S}_1(F_1)} \Psi_S(F_n). \tag{3.2}$$

The set $F = \bigcap F_n$ is a generalized Sierpiński carpet.

Remark. The conditions (3.1) on F_1 are slightly weaker than those we imposed in Barlow & Bass (1989), but they are in fact all that is needed there and in Barlow & Bass (1990).

It will be helpful to have some additional notation relating to the unit square. Let $C_0 = (0,0)$, and C_1, C_2, C_3 be the other three corners of F_0, labelled anticlockwise, and let $\bar{C} = (\frac{1}{2}, \frac{1}{2})$ be the centre of F_0. Throughout this paper we will use the notation $i' = i'(i) = i+1 \pmod 4$, $0 \leqslant i \leqslant 3$. Let L_i denote the line segment joining C_i and $C_{i'}$. Let $T_i = T_i(n)$ denote the intersection of F_n with the interior of the triangle with vertices C_i, $C_{i'}$, $\bar{C}$.

Now fix $n \geqslant 0$, and consider the two variational problems pot (F_n, L_3, L_1) and cur (F_n, L_3, L_1). Write

$$R_n = R_{F_n}(L_3, L_1), \tag{3.3}$$

and let u_n be the optimum potential for the problem pot (F_n, L_3, L_1) and $I_n = R_n \nabla u_n$ be the optimum flow for cur (F_n, L_3, L_1). Write $I_n^{31} = I_n$, and note that I_n^{31} is a flow from L_3 to L_1. By rotation we obtain flows $I_n^{02}, I_n^{13}, I_n^{20}$.

Now define flows U^i, V^i on T_i by

$$U^i = I_n^{ii''}|_{T_i}, \quad \text{where } i'' = i+2 \pmod 4, \quad 0 \leqslant i \leqslant 3,$$

$$V^i = I_n^{jj''}|_{T_i}, \quad \text{where } j = i-1 \pmod 4, \quad j'' = i+1 \pmod 4, \quad 0 \leqslant i \leqslant 3.$$

By the symmetry of F_n, and I_n, the energy $\int_{T_i} |U^i|^2$ takes the same value for each i, and similarly for V_i. Set

$$E(U) = \int_{T_i} |U^i|^2, \quad E(V) = \int_{T_i} |V^i|^2,$$

and note that

$$R_n = \int_{F_n} |I_n^{31}|^2 = 2E(U) + 2E(V). \tag{3.4}$$

Now set

$$I_n^{30} = U^3 1_{T_0} - U^0 1_{T_0} - V^1 1_{T_1} - V^2 1_{T_2},$$

and similarly define the remaining $I_n^{ii'}$, $I_n^{i'i}$ by rotation and reflection. The flow I_n^{30} may be thought of as the flow obtained from I_n^{31} by reflecting across the diagonal $x_1 = x_2$ (cf. Barlow *et al.* 1990, fig. 1). Using the symmetry of the U^i and V^i it is straightforward to check that I_n^{30} is a flow from L_3 to L_0 in F_n.

Lemma 3.1. $\int_{T_i} U^i \cdot V^i = 0$.

Proof. It is enough to consider the case $i = 0$. For $x = (x_1, x_2) \in T_0$ write $x' = (1 - x_1, x_2)$, and set $B_2 = T_0 \cap \{x_1 < \frac{1}{2}\}$. Now writing $U_j^0, V_j^0, j = 1, 2$, for the components of U^0, V^0, we have by the symmetry across the line $x_1 = \frac{1}{2}$

$$V^0(x') = (V_1^0(x), -V_2^0(x)), \quad U^0(x') = (-U_1^0(x), U_2^0(x)).$$

So

$$U^0(x) \cdot V^0(x) + U^0(x') \cdot V^0(x') = 0,$$

and hence

$$\int_{T_0} U^0(x) \cdot V^0(x)\, dx = \int_{B_2} (U^0(x) \cdot V^0(x) + U^0(x') \cdot V^0(x'))\, dx = 0. \qquad \square$$

Now consider the crosswire formed by joining the mid-points of each of the lines $L_0, \ldots, L_3$ to $\bar{C}$ by a wire of resistance $\frac{1}{2}$. Let H be a flow on the crosswire, and write H_i, $0 \leqslant i \leqslant 3$ for the current flowing along wire i, with a positive value corresponding to an inward flow. Then $\Sigma_{i=0}^3 H_i = 0$, and if $E_+(H)$ is the energy dissipation in the crosswire then

$$E_+(H) = \tfrac{1}{2} \sum_{i=0}^{3} H_i^2. \tag{3.5}$$

Set $h_i = |H_i|$, and write $h = \frac{1}{2}\Sigma h_i$ for the total flux through the crosswire.

We now construct a matching flow $J (= J_n(H))$ on F_n by

$$J = \sum_{i=0}^{3} \sum_{j \neq i} h^{-1} H_i^+ H_j^- I_n^{ij}. \tag{3.6}$$

Here the superscripts $+$, $-$ denote the positive and negative parts respectively.

Proposition 3.2. (*a*) *The total flux of J through the edge L_i is H_i.* (*b*) $\int_{F_n} |J|^2 \leqslant R_n E_+(H)$.

Proof. By reflection, rotation and inversion of currents we can reduce the situation to one of three cases:

(i) $H_0, H_1 \geqslant 0$, $H_2, H_3 \leqslant 0$, and $J = h^{-1}(h_0 h_3 I_n^{03} + h_0 h_2 I_n^{02} + h_1 h_2 I_n^{12} + h_1 h_3 I_n^{13})$;

(ii) $H_0, H_2 \geqslant 0, H_1, H_3 \leqslant 0$, and $J = h^{-1}(h_0 h_1 I_n^{01} + h_0 h_3 I_n^{03} + h_2 h_1 I_n^{21} + h_2 h_3 I_n^{23})$;

(iii) $H_0 \geqslant 0, H_1, H_2, H_3 \leqslant 0, h_1 < h_3$, and $J = h_1 I_n^{01} + h_2 I_n^{02} + h_3 I_n^{03}$.

(*a*) This follows immediately by examining each of the cases (i)–(iii).

(*b*) In each case we have, by Lemma 3.1,

$$\int_{T_i} |J|^2 \leqslant h_i^2 E(U) + (h - h_i)^2 E(V),$$

and so
$$E_{F_n}(J,J) = \int_{F_n} |J|^2 \leqslant E(U)\sum_0^3 h_i^2 + E(V)\sum_0^3 (h-h_i)^2.$$

From the definition of h we always have
$$\Sigma (h-h_i)^2 = \Sigma h_i^2,$$
and thus by (3.4) and (3.5)
$$E_{F_n}(J,J) \leqslant (E(U)+E(V))\,\Sigma h_i^2 = R_n E_+(H). \qquad \square$$

We now define two wire networks, G_n and D_n, associated with F_n. G_n is the network obtained by replacing each square in $\mathscr{S}_n(F_n)$ by a crosswire of four wires each of resistance $\frac{1}{2}$, parallel to the axes. For D_n the wires are each of resistance 1, and connect the centre of each square to its corners. Set

$$X_0 = \{(\tfrac{1}{2},\tfrac{1}{2})\}, \quad Y_0 = \{(\tfrac{1}{2},0),(1,\tfrac{1}{2}),(\tfrac{1}{2},1),(0,\tfrac{1}{2})\},$$
$$Z_0 = \{(0,0),(0,1),(1,1),(1,0)\},$$
$$X_n = \bigcup_{S\in\mathscr{S}_n(F_n)} \Psi_S(X_0), \quad Y_n = \bigcup_{S\in\mathscr{S}_n(F_n)} \Psi_S(Y_0),$$
$$Z_n = \bigcup_{S\in\mathscr{S}_n(F_n)} \Psi_S(Z_0), \quad G_n = X_n \cup Y_n, \quad D_n = X_n \cup Z_n.$$

(The points in X_n are centres of squares in $\mathscr{S}_n(F_n)$, while the points in Y_n are the midpoints of sides, and Z_n are the corners.) Define $g: G_n \times G_n \rightarrow \mathbb{R}$ by
$$g_n(x,y) = \begin{cases} \frac{1}{2} & \text{if} \quad x\in X_n, y\in Y_n \quad \text{and} \quad |x-y| = \frac{1}{2}k^{-n}, \\ 0 & \text{otherwise,} \end{cases}$$
and $d_n: D_n \times D_n \rightarrow \mathbb{R}$ by
$$d_n(x,z) = \begin{cases} 1 & \text{if} \quad x\in X_n, z\in Z_n \quad \text{and} \quad |x-z| = 2^{-\frac{1}{2}}k^{-n} \\ 0 & \text{otherwise.} \end{cases}$$
Let $\mathscr{E}_{G_n}$, E_{G_n}, $\mathscr{E}_{D_n}$, E_{D_n} be the energy forms for the networks (G_n, g_n), (D_n, d_n) defined by (2.2) and (2.5).

Let u_n^G be the optimum potential for the problem $\operatorname{pot}(G_n, G_n \cap L_3, G_n \cap L_1)$ and write $R_n^G = R_{G_n}(G_n \cap L_3, G_n \cap L_1)$. Then the current
$$I_n^G = R_n^G \nabla u_n^G$$
is the optimum current for the problem $\operatorname{cur}(G_n, G_n \cap L_3, G_n \cap L_1)$. Write u_n^D, R_n^D and I_n^D for the analogous quantities for the network D_n.

Theorem 3.3. *For* $n, m \geqslant 0$, $\quad R_{n+m} \leqslant R_n R_m^G$.

Proof. We use the current I_m^G to construct a flow K on F_{n+m}. Let S be any square in $\mathscr{S}_n(F_n)$, and let H^S be the restriction of the flow I_m^G to the crosswire contained in S. Write H_i^S, $0 \leqslant i \leqslant 3$ for the flow in each of the four wires of the crosswire. Let $J_n(H^S)$ be the flow in F_n defined by (3.6), and define vector fields K^S on $S \cap F_{n+m}$ for $S \in \mathscr{S}_n(F_n)$, and K on F_{n+m} by
$$K^S = J_n(H^S) \circ \Psi_S^{-1}, \quad K = \sum_{S\in\mathscr{S}_n(F_n)} 1_S K^S.$$

Let S_1 and S_2 be two adjacent squares in $\mathscr{S}_n(F_n)$, with, say, S_2 lying to the right of S_1. Then $H_1^{S_1}+H_3^{S_2}=0$, and it follows from the construction of the $J_n(H^{S_i})$ that the flows K^{S_1} and K^{S_2} match on the border $S_1 \cap S_2$. So K is a flow on F_{n+m}, and we have

$$-\int_{L_3 \cap F_{n+m}} K \cdot n \,\mathrm{d}\sigma = \int_{L_1 \cap F_{n+m}} K \cdot n \,\mathrm{d}\sigma = 1,$$

$$K \cdot n = 0 \quad \text{on} \quad (L_2 \cup L_0) \cap F_{n+m}.$$

So K is a feasible flow for the variational problem cur(F_{n+m}, L_3, L_1) and thus

$$\int_{F_{n+m}} |K|^2 \geqslant R_{n+m}.$$

However,

$$\int_{F_{n+m}} |K|^2 = \sum_{S \in \mathscr{S}_n(F_n)} \int_{F_{n+m} \cap S} |K^S|^2 = \sum_{S \in \mathscr{S}_n(F_n)} \int_{F_n} |J_n(H^S)|^2$$

$$\leqslant \sum_{S \in \mathscr{S}_n(F_n)} R_n \cdot \tfrac{1}{2} \sum_i |H_i^S|^2 = R_n E(I_m^G, I_m^G) = R_n R_m^G. \qquad \square$$

4. Potentials and the lower bound

We begin by relating R_m^G and R_m^D. Let $\mathscr{H}(D_m)$ be the set of $f: D_m \to \mathbb{R}$ which satisfy

$$f(x) = \tfrac{1}{4} \sum_{z \in Z_m} \mathrm{d}(x,z) f(z), \quad x \in X_m. \tag{4.1}$$

(So $f \in \mathscr{H}(D_m)$ is harmonic at the centre of any square in $\mathscr{S}_m(F_m)$.)

Proposition 4.1. $\tfrac{1}{2} R_m^G \leqslant R_m^D \leqslant R_m^G$.

Proof. We begin with the right-hand inequality. Define $f: G_m \to \mathbb{R}$ by

$$f(x) = u_m^D(x), \qquad x \in X_m,$$
$$f(y) = \tfrac{1}{2}(u_m^D(z_1) + u_m^D(z_2)), \quad y \in Y_m,$$

where z_1, z_2 are the two points in Z_m such that y is the midpoint of the line segment connecting z_1 and z_2, and $|z_i - y| = \tfrac{1}{2} k^{-m}$.

Let $S \in \mathscr{S}_m(F_n)$, and write $a_0, \ldots, a_3$ for the values of u_m^D at the four corners of S, $\bar{a} = \tfrac{1}{4} \Sigma_0^3 a_i$. Then if b_i is the value of f at the midpoint of the line segment connecting $\Psi_S(C_i)$ and $\Psi_S(C_{i'})$, we have $b_i = \tfrac{1}{2}(a_i + a_{i'})$. The energy of the functions u_m^D and f in the square S are given by

$$\mathscr{E}_S(u_m^D) = \sum_i (a_i - \bar{a})^2, \quad \mathscr{E}_S(f) = 2 \sum (b_i - \bar{a})^2.$$

But writing $x_i = a_i - \bar{a}$

$$\mathscr{E}_S(f) = 2 \sum_i (b_i - \bar{a})^2 = \tfrac{1}{2} \sum_{i=0}^{3} (x_i + x_{i'})^2$$
$$= \mathscr{E}_S(u_m^D) + 2(x_0 + x_2)(x_1 + x_3).$$

As $\Sigma x_i = 0$ the final term is negative, and thus $\mathscr{E}_S(f) \leqslant \mathscr{E}_S(u_m^D)$. Summing over S

$$\mathscr{E}_{G_m}(f, f) \leqslant \mathscr{E}_{D_m}(u_m^D, u_m^D),$$

and, as f is feasible for the problem pot $(G_m, G_m \cap L_3, G_m \cap L_1)$, we have

$$\begin{aligned}(R_m^G)^{-1} &\leqslant \mathscr{E}_{G_m}(f,f),\\ &\leqslant \mathscr{E}_{D_m}(u_m^D, u_m^D) = (R_m^D)^{-1}.\end{aligned}$$

To prove the left-hand inequality, we begin by noting the following elementary inequalities, which may be verified by expanding the right-hand side. Let $x_0, x_1, x_2, x_3 \in \mathbb{R}$; then

$$\sum_{i=0}^{3} x_i^2 \leqslant \frac{1}{2} \sum_{i=0}^{3} (x_i - x_{i'})^2 \quad \text{if} \quad \sum_{i=0}^{3} x_i = 0, \tag{4.2}$$

$$\sum_{i=0}^{2} x_i^2 \leqslant \sum_{i=0}^{1} (x_i - x_{i+1})^2 \quad \text{if} \quad \sum_{i=0}^{2} x_i = 0, \tag{4.3}$$

$$\sum_{i=0}^{1} x_i^2 \leqslant \tfrac{1}{2}(x_1 - x_0)^2 \quad \text{if} \quad \sum_{i=0}^{1} x_i = 0. \tag{4.4}$$

Now define $h: D_m \rightarrow \mathbb{R}$ by setting

$$\begin{aligned} h(x) &= u_m^G(x), \quad x \in X_m,\\ h(z) &= 0, \quad z \in Z_m \cap L_3,\\ &= 1, \quad z \in Z_m \cap L_1,\\ h(z) &= \sum_{x \in X_m} \mathrm{d}(z,x)\, h(x) \Big/ \sum_{x \in X_m} \mathrm{d}(z,x), \quad z \in Z_m - (L_1 \cup L_3).\end{aligned}$$

Let $z \in Z_m$. Then z is at the corner of four squares in $\mathscr{S}_m$, of which at least one is in $\mathscr{S}_m(F_m)$. Suppose first that all four are in $\mathscr{S}_m(F_m)$, and write x_0, x_1, x_2, x_3 for their centres, $b_i = h(x_i) - h(z)$. The energy dissipation related to the function h in the diagonal crosswires connecting z and x_i is

$$\sum_{i=0}^{3} (h(x_i) - h(z))^2 = \sum_{i=0}^{3} b_i^2.$$

The energy dissipation of u_m^G in the wires connecting x_i and $x_{i'}$, $0 \leqslant i \leqslant 3$, in the network G_m is

$$\sum_{i=0}^{3} (h(x_i) - h(x_{x'}))^2 = \sum_{i=0}^{3} (b_i - b_{i'})^2,$$

which is greater than $\Sigma\, b_i^2$ by the inequality (4.2). We obtain a similar bound in the other cases. Summing over $z \in Z_m$, and noting that each wire in the network G_m is counted twice we deduce that

$$\mathscr{E}_{D_m}(h,h) \leqslant 2\mathscr{E}_{G_m}(u_m^G, u_m^G),$$

and so $(R_m^D)^{-1} \leqslant 2(R_m^G)^{-1}$. □

Recall from §3 the definition of the potential u_n, the points C_i and the sets T_i. Let u_n^i, $i = 0, \ldots, 3$ be the rotation of u_n by $i\pi/2$: so for example we have $u_n^1(C_0) = u_n^1(C_1) = 0$, $u_n^1(C_2) = u_n^1(C_3) = 1$. Now let n be fixed, and define functions v, w on T_i by

$$v^i = u_n^{i'}|_{T_i} \quad (i' = i+1 \pmod 4), \quad w^i = u_n^i|_{T_i}.$$

Thus $v^0(C_0) = v^0(C_1) = 0$, $v^0(\bar{C}) = \frac{1}{2}$, $w^0(C_0) = 0$, $w^0(C_1) = 1$ and $w^0(\bar{C}) = \frac{1}{2}$.

Let $\mathscr{E}_v = \int_{T_i} |\nabla v_i|^2$, $\mathscr{E}_w = \int_{T_i} |\nabla w_i|^2$, and note that $2\mathscr{E}_v + 2\mathscr{E}_w = R_n^{-1}$.

Lemma 4.2. $\int_{T_i} \nabla v_i \cdot \nabla w_i = 0$, $0 \leqslant i \leqslant 3$.

Proof. It is enough to prove this in the case $i = 0$. For $x = (x_1, x_2)$, let $x' = (1-x_1, x_2)$. Writing $\nabla v_i = (v_{i1}, v_{i2})$, we have

$$\nabla v_i(x') = (-v_{i1}(x), v_{i2}(x)), \nabla w_i(x') = (w_{i1}(x), -w_{i2}(x)).$$

Hence

$$\nabla v_i(x) \cdot \nabla w_i(x) + \nabla v_i(x') \cdot \nabla w_i(x') = 0,$$

and integrating over $T_0 \cap \{x_1 < \frac{1}{2}\}$ the result follows. □

We now compare the energy dissipation of matching potentials on D_0 and F_n. Let $\varphi \in \mathscr{H}(D_0)$, write $x_i = \varphi(C_i)$, and write $\bar{x} = \frac{1}{4}\Sigma x_i$. Define $f_n^\varphi : F_n \to \mathbb{R}$ by

$$f_n^\varphi = \sum_{i=0}^{3} [x_i + (2\bar{x} - (x_i + x_{i'}))\, v_i + (x_{i'} - x_i)\, w_i]\, 1_{T_i}, \tag{4.5}$$

with f_n^φ defined on $F_n - \bigcup_{i=0}^{3} T_i$ by continuity. Note that f_n^φ agrees with φ on D_0, and that f_n^φ is continuous. Then

$$\mathscr{E}_{F_n}(f_n^\varphi, f_n^\varphi) = \sum_{i=0}^{3} [(x_{i'} - x_i)^2\, \mathscr{E}_w + ((x_i - \bar{x}) + (x_{i'} - \bar{x}))^2 \mathscr{E}_v].$$

Now

$$\sum_i (x_i - x_{i'})^2 \leqslant 4 \Sigma (x_i - \bar{x})^2,$$

$$\sum_i ((x_i - \bar{x}) + (x_{i'} - \bar{x}))^2 \leqslant 4 \Sigma (x_i - \bar{x})^2,$$

so

$$\mathscr{E}_{F_n}(f_n^\varphi, f_n^\varphi) \leqslant 4(\mathscr{E}_v + \mathscr{E}_w)\, \mathscr{E}_{D_0}(\varphi, \varphi) = 2R_n^{-1} \mathscr{E}_{D_0}(\varphi, \varphi). \tag{4.6}$$

Theorem 4.3. $\frac{1}{2}R_n R_m^D \leqslant R_{n+m}$, *for* $n, m \geqslant 0$.

Proof. For $S \in \mathscr{S}_m(F_m)$ let $\varphi_S = u_m^D \circ \Psi_S$, and set

$$f = \sum_{S \in \mathscr{S}_m(F_m)} (f_n^{\varphi_S} \circ \Psi_S^{-1})\, 1_S.$$

From (4.5) we see that, on the line L_i, f_n^φ depends only on the values of φ at C_i and $C_{i'}$. So f is continuous, and is feasible for the potential problem pot (F_{n+m}, L_3, L_1). So using (4.6),

$$R_{n+m}^{-1} \leqslant \mathscr{E}_{F_{n+m}}(f, f) = \sum_{S \in \mathscr{S}_m(F_m)} \mathscr{E}_{F_n}(f^{\varphi_S}, f^{\varphi_S})$$

$$\leqslant 2R_n^{-1} \sum_{S \in \mathscr{S}_m(F_m)} \mathscr{E}_{D_0}(\varphi_S, \varphi_S) = 2R_n^{-1} \mathscr{E}_{D_m}(\varphi, \varphi) = 2R_n^{-1} (R_m^D)^{-1}. \quad □$$

5. Main results and concluding remarks

Theorem 5.1. *Let* (F_n) *be a sequence of approximations to the generalized Sierpiński carpet* F, *satisfying the hypotheses* (3.1). *Let* R_n *be the resistance of* F_n *when two opposite sides are each shorted. Then there exists a constant* ρ *such that*

$$\tfrac{1}{4}\rho^n \leqslant R_n \leqslant 4\rho^n, \quad n \geqslant 0. \tag{5.1}$$

Proof. From Theorem 3.3, Proposition 4.1 and Theorem 4.3 we have

$$\begin{aligned} \tfrac{1}{2}R_n R_m^D \leqslant R_{n+m} \leqslant R_n R_m^G, \\ \tfrac{1}{2}R_m^G \leqslant R_m^D \leqslant R_m^G. \end{aligned} \tag{5.2}$$

Taking $n = 0$ in (5.2) gives

$$\tfrac{1}{2}R_m^D \leqslant R_m \leqslant R_m^G.$$

So

$$\tfrac{1}{4}R_n R_m \leqslant R_{n+m} \leqslant 4R_n R_m.$$

Let $a_n = 4R_n$, $b_n = \frac{1}{4}R_n$. Then $a_{n+m} \leqslant a_n a_m$, $b_{n+m} \geqslant b_n b_m$: the sequence (a_n) is submultiplicative, and (b_n) is supermultiplicative. Using the standard properties of subadditive sequences there exist θ_0, θ_1 such that

$$\theta_0 = \lim_{n\to\infty} \frac{\ln a_n}{n} = \inf_{n\geqslant 0} \frac{\ln a_n}{n}, \quad \theta_1 = \lim_{n\to\infty} \frac{\ln b_n}{n} = \sup_{n\geqslant 0} \frac{\ln b_n}{n}.$$

As a_n/b_n is constant, $\theta_0 = \theta_1 = \ln\rho$ say, and we have $a_n \geqslant \rho^n \geqslant b_n$ for $n \geqslant 0$; (5.1) now follows. □

The proof of Theorem 5.1 gives no information on the value of ρ. We do have the following lower bound.

Proposition 5.2. *Let F_n, ρ be as above. Then*

$$\rho \geqslant k^2/(k^2-R). \tag{5.3}$$

Proof. Set $A_i = [(i-1)/k, i/k] \times [0,1]$, and let a_i be the number of squares of side k^{-1} in $F_1 \cap A_i$. We have $\Sigma a_i = k^2 - R$. Let R'_n be the resistance of F_n (between L_3 and L_1) when 'shorts' are applied along the lines $x_1 = i/k$, $1 \leqslant i \leqslant k-1$. R'_n is the solution to variational problem

$$(R'_n)^{-1} = \inf\left\{\int_{F_n} |\nabla u|^2 \,\mathrm{d}x : u = 0 \text{ on } L_3, u = 1 \text{ on } L_1 \right.$$
$$\left. \text{and } u(i/k,\cdot) \text{ is constant for } 1 \leqslant i \leqslant k-1\right\},$$

and from this it is clear that $R_n \geqslant R'_n$.

From the standard electrical circuit rules for combining resistors in series and parallel we have

$$R'_n = R_{n-1}\sum_{i=1}^{k} a_i^{-1} \geqslant R_{n-1}\left(k^2\Big/\sum_1^k a_i\right) = R_{n-1}\frac{k^2}{k^2-R}.$$

Therefore $R_n \geqslant (k^2/(k^2-R))^n R_0$, and hence $\rho \geqslant k^2(k^2-R)$. □

The Hausdorff dimension of the limiting Sierpiński carpet F is given by

$$d_f(F) = \ln(k^2-R)/\ln k.$$

In a future paper we will prove that the spectral dimension of F is

$$d_s(F) = \frac{2\ln(k^2-R)}{\ln(\rho(k^2-R))}.$$

An immediate consequence of Proposition 5.2 is

Corollary 5.3. *Let F_1 satisfy* (3.1) *and let $F = \bigcap_n F_n$ be a generalized Sierpiński carpet. Then $d_s(F) \leqslant d_f(F)$.*

Remarks

1. The bound in Proposition 5.2 was obtained in Ben-Avraham & Havlin (1983) for a more restricted class of carpets.

2. Write $C_n = R_n^{-1}$ for the conductivity of F_n and A_n for the area of F_n. Note that

$$C_n \approx \rho^{-n}, \quad A_n = ((k^2-R)/k^2)^n.$$

The proposition may be interpreted as stating that the removal of the squares in F_n causes the conductivity to decrease at least as rapidly as the area.

3. See Kusuoka (1987) for a result, for graphs, similar to Corollary 5.3.

4. The methods of Theorems 3.3 and 4.3 also yield the inequalities $R^G_{m+n} \leqslant R^G_m R^G_n$, $R^D_{n+m} \leqslant R^D_n R^G_m$, $R^G_{n+m} \geqslant \frac{1}{2} R^G_n R^D_m$, and $R^D_{n+m} \geqslant \frac{1}{2} R^D_n R^D_m$. Using these we deduce

$$\tfrac{1}{2}\rho^n \leqslant R^D_n \leqslant 2\rho^n, \quad \rho^n \leqslant R^G_n \leqslant 4\rho^n. \tag{5.4}$$

5. Theorem 5.1 remains true if the hypothesis (H3) is dropped. The proof of Theorem 3.3 goes through exactly as before, but for §4 the network D_m needs to be modified. It is necessary to split into two points those $z \in Z_m$ which lie in two diagonally adjacent squares in $\mathscr{S}_m(F_m)$, and where the remaining squares in the 2×2 block are not in $\mathscr{S}_m(F_m)$.

6. The inequalities in Theorem 3.3 and 4.3 provide a method of obtaining rigorous numerical bounds on ρ. For if we have calculated R^G_m, R^D_m for some $m \geqslant 1$ then

$$R_{nm} \leqslant R^G_m R_{(n-1)m} \leqslant (R^G_m)^n,$$

so that $\rho \leqslant (R^G_m)^{1/m}$ for any $m \geqslant 1$. Similarly $\rho \geqslant (\frac{1}{2}R^D_m)^{1/m}$ for any $m \geqslant 1$. Unfortunately these bounds are not very good; the sequence $2^{1/m}$ converges very slowly, and the computation of R_m requires an order of magnitude more memory and computing time for each successive m. For the standard Sierpiński carpet, for example, Barlow *et al.* (1990) gives $R^G_7 \approx 5.524280$, while additional computations by one of us give $R^D_5 \approx 2.903291$. These prove that

$$1.07738 \leqslant \rho \leqslant 1.27656. \tag{5.5}$$

On the other hand, a glance at the sequence of ratios R^G_n/R^G_{n-1}, R^D_n/R^D_{n-1} suggests that each sequence is converging. The limit must be ρ and hence it seems probable that

$$1.25147 \leqslant \rho \leqslant 1.25149. \tag{5.6}$$

(We also remark that the ratios of the R^D_n appear to converge more rapidly than the R^G_n.)

7. Let $S \in \mathscr{S}_m$, and let A, B be the two opposite sides of S. Then the distance between A and B is $a = k^{-m}$, while

$$R_{F_n \cap S}(A,B) \approx \rho^{n-m} = \rho^n a^{(\ln\rho/\ln k)}.$$

The quantity $\tilde{\zeta} = \ln\rho/\ln k$ is referred to in the physics literature as the resistance exponent.

We conclude this paper by mentioning two open problems.

Problem 1. Does the limit

$$\lim_{n\to\infty} \rho^{-n} R_n \tag{5.7}$$

exist? This is closely related to problems concerning the uniqueness of the diffusion X_t on F constructed in Barlow & Bass (1989).

Problem 2. Find a method of calculating ρ. The approach we used in Barlow *et al.* (1990), which was to look at ratios R_n^G/R_{n-1}^G, works reasonably well for Sierpiński carpets with small k but already required significant amounts of computing time and memory.

The research of both authors was partly supported by the National Science Foundation.

Appendix

Let n be fixed, $F = \text{int}(F_n)$, $B_1 = L_1 \cap F_n$, and $B_0 = L_3 \cap F_n$. Write $E = \partial F - (B_0 \cup B_1)$. We want to show the existence and uniqueness of a harmonic function v in $C(\bar{F}) \cap H^1(F)$ with $\nabla v \in L^2(\partial F)$, $v = 0$ on B_0, $v = 1$ on B_1, and $\partial v/\partial n = 0$ a.e. ($d\sigma$) on E.

Let (W_t^n, P^x) be brownian motion in F_n with normal reflection on ∂F_n. For the construction of this process, see Barlow & Bass (1990, §2). Define $v(x) = P^x(W_t^n$ hits B_1 before $B_0)$. It is standard (see, for example, Port & Stone 1978) that v is harmonic in $\text{int}(F_n)$.

First we show $v \in C(\bar{F})$. The continuity of v in neighbourhoods of points in $\text{int}(F_n)$ and in $B_i - \bar{E}$, $i = 0, 1$, is well known, as is the fact that $v = i$ on B_i. For the continuity of v in neighbourhoods of points in E, one can use Bass & Hsu (1991), §3 (easily extended to the case of two dimensions) or else a conformal mapping and reflection argument (cf. below).

This leaves points in $\bar{E} \cap B_i$, $i = 0, 1$. By the geometry of F_n, if $x \in \bar{E} \cap B_0$, there exists $\epsilon > 0$ and an isometry φ from $F \cap B(x, \epsilon)$ to $\{(x_1, x_2): x_1, x_2 < 0\} \cap B(0, \epsilon)$, where $B(y, \epsilon)$ is the ball of radius ϵ about y. By defining $\bar{v}$ on $\{(x_1, x_2): x_2 < 0\} \cap B(0, \epsilon)$ by $\bar{v}(x_1, x_2) = v \circ \varphi^{-1}(|x_1|, x_2)$, i.e. by reflecting across the line $x_1 = 0$, we see that $\bar{v}$ is harmonic in $\{x_2 > 0\} \cap B(0, \epsilon)$, and the continuity of $\bar{v}$, hence of v, may be obtained as for points in B_0. The points in $E \cap B_1$ are treated similarly.

Next we show $\partial v/\partial n \in L^2(\partial F)$. For points $x \in E$ that are not corners of squares we can show that $\partial v/\partial n$ restricted to a neighbourhood of x is in $L^2(\partial F)$ by using a reflection argument similar to the preceding paragraph. This argument also shows that $\partial v/\partial n = 0$ at these points. If $x \in E$ is the corner of a square, we can again use the reflection argument provided we first use a conformal mapping to map $B(x, \epsilon) \cap F_n$ into $B(0, \epsilon) \cap \{x_2 > 0\}$.

Suppose now that $x \in B_0 - \bar{E}$. Let x_1 be a point in the interior of F_n and let $g(y)$ be the value at y of the Green function (for W_t^n with absorption at $B_0 \cup B_1$) with a pole at x_1. Since F is Lipschitz, it is a regular domain, and g vanishes on B_0. Since v also vanishes on B_0, by the boundary Harnack principle (Jerison & Kenig 1982, Theorem 5.25), $v(y) \leqslant cg(y)$ for y in a neighbourhood of x, c a constant. Hence $0 \leqslant \partial v/\partial n \leqslant c\,\partial g/\partial n$ in a neighbourhood of x. But $\partial g/\partial n$ is in $L^2(\partial F)$ (see Jerison & Kenig (1982), Theorem 3.1). The same argument applied to $1 - v$ works for points in $B_1 - \bar{E}$, while points in $(B_0 \cup B_1) \cap \bar{E}$ are treated similarly, after first using a reflection argument as above. Hence, as ∂F is compact, $\partial v/\partial n \in L^2(\partial F)$. By Jerison & Kenig (1982, Theorem 4.1 (extended to non-starlike Lipschitz domains)) $\nabla v \in L^2(\partial F)$ and $v \in H^1(F)$.

To get uniqueness, let v, v' be two solutions and look at their difference. By the maximum principle, the maximum is taken at some point of ∂F. That point cannot be in E by the reflection argument we have used repeatedly above (at corners of squares, first use a conformal mapping). So the maximum must be in $B_0 \cup B_1$, where $v - v' = 0$. Hence $v - v' \leqslant 0$ on F, and similarly $v - v' \geqslant 0$.

We thank K. Burdzy, E. B. Davies, and D. Jerison for comments related to this appendix.

References

Barlow, M. T. & Bass, R. F. 1989 The construction of Brownian motion on the Sierpiński carpet. *Annls Inst. H. Poincaré* **25**, 225–257.

Barlow, M. T. & Bass, R. F. 1990 Local times for Brownian motion on the Sierpiński carpet. *Probab. Theor. Rel. Fields* **85**, 91–104.

Barlow, M. T., Bass, R. F. & Sherwood, J. D. 1990 Resistance and spectral dimension of Sierpiński carpets. *J. Phys.* A **23**, L253–L258.

Barlow, M. T. & Perkins, E. A. 1988 Brownian motion on the Sierpiński gasket. *Probab. Theor. Rel. Fields* **79**, 543–623.

Bass, R. F. & Hsu, P. 1991 Some potential theory for reflecting Brownian motion in Hölder and Lipschitz domains. *Ann. Probab.* (In the press.)

Ben-Avraham, D. & Havlin, S. 1983 Exact fractals with adjustable fractal and fracton dimensionalities. *J. Phys.* A **16**, L559–563.

Beurling, A. & Deny, J. 1958 Espaces de Dirichlet, I, le cas élementaire. *Acta Math.* **99**, 203–224.

Beurling, A. & Deny, J. 1959 Dirichlet spaces. *Proc. natn. Acad. Sci. U.S.A.* **45**, 208–215.

Doyle, P. G. & Snell, J. L. 1984 *Random walks and electrical networks.* Washington, D.C.: Mathematics Association America.

Gefen, Y., Aharony, A. & Mandelbrot, B. B. 1984 Phase transitions on fractals: III. Infinitely ramified lattices. *J. Phys.* A **17**, 1277–1289.

Hattori, K., Hattori, T. & Watanabe, H. 1985 New approximate renormalisation method on fractals. *Phys. Rev.* A **32**, 3730–3733.

Havlin, S. & Ben-Avraham, D. 1987 Diffusion in disordered media. *Adv Phys.* **36**, 695–798.

Jerison, D. S. & Kenig, C. E. 1982 Boundary value problems in Lipschitz domains. In *Studies in partial differential equations* (ed. W. Littman), pp. 1–68. Washington, D.C.: Mathematics Association America.

Kusuoka, S. 1987 A diffusion process on a fractal. In *Probabilistic methods in mathematical physics* (ed. K. Ito & N. Ikeda), pp. 251–274. Boston: Academic Press.

Lindstrøm, T. 1990 Brownian motion on nested fractals. *Mem. Am. math. Soc.* **420**.

Mandelbrot, B. B. 1982 *The fractal geometry of nature.* San Francisco: W. H. Freeman.

Maz'ja, V. G. 1985 *Sobolev spaces.* Berlin: Springer.

Osada, H. 1990 Isoperimetric dimension and estimates of heat kernels of pre-Sierpiński carpets. *Probab. Theor. Rel. Fields.* (In the press.)

Port, S. C. & Stone, C. J. 1978 *Brownian motion and classical potential theory.* New York: Academic Press.

Rammal, R. & Toulouse, G. 1983 Random walks on fractal structures and percolation clusters. *J. Physique Lett.* **44**, L13–L22.

Sierpiński, W. 1916 Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. *C. r. hebd. Śeanc. Acad. Sci., Paris* **162**, 629–632.

Watanabe, H. 1985 Spectral dimension of a wire network. *J. Phys.* A **18**, 2807–2823.

Yuan, L.-Y. & Tao, R. 1986*a* Experimental study of the conductivity exponent for some Sierpiński carpets. *Phys. Lett.* A **116**, 284–286.

Yuan, L.-Y. & Tao, R. 1986*b* Studies on the scaling exponents of conductivity for Sierpiński carpets. *J. Phys.* C **21**, 401–409.

Ziemer, W. P. 1989 *Weakly differentiable functions: Sobolev spaces and functions of bounded variation.* New York: Springer-Verlag.

Submitted by D. Williams; received and accepted 11 July 1990