Application of Rayleigh's short-cut method to Polya's recurrence problem

Peter G. Doyle

Version dated 5 October 1998 *
GNU FDL†

Abstract

A method called *Rayleigh's short-cut method* from the classical theory of electricity is applied to prove and extend Polya's recurrence theorem for random walk on a lattice. The goals of the presentation are to explain "why" Polya's theorem is true and to develop techniques for applying Rayleigh's method. The main results make sense of the notion that if two graphs look alike then random walk is transient on one if and only if it is on the other.

*This work was the Dartmouth Ph.D. thesis of Peter Doyle, submitted in June 1982. Minor changes made 1994, 1998 by Peter Doyle.

Preface

In this work, I describe how a method from the classical theory of electricity can be used to prove and extend Polya's recurrence theorem. This theorem states that a point walking at random on an infinite 2-dimensional lattice is sure to come back to its starting point, but that on a 3-dimensional lattice it may never come back. The electrical method used is an offspring of Thomson's minimum dissipation theorem called *Rayleigh's short-cut method*. The relationship between the two comes about because Polya's theorem is equivalent to a statement about the resistance out to infinity of an infinite lattice of resistors.

Here's the plan. In part I, I present a simple proof of Polya's theorem using Rayleigh's method. This Part is meant mainly as a leisurely introduction to Part II, a way of getting a feeling for the kinds of problems we want to consider and the kinds of methods we mean to apply. The main action takes place in Part II, where I explore further applications of Rayleigh's method to Polya's theorem and related questions.

The reader hungry for new results will find them mainly in Section 3 of Part II, where I show how to make sense of the notion that if two graphs look alike then random walk is transient on one if and only if it is on the other.

My main concern is not so much with results, however, as with ideas and methods. In writing this work, my goals were to shed some light on the question of "why" Polya's theorem is true, and in the process, to develop techniques for applying Rayleigh's method.

The second goal is important because Rayleigh's method seems destined to play an increasingly important role in probability and analysis. As evidence of this, I give in the Appendix a sample application of Rayleigh's method to the classical type problem for Riemann surfaces. For further evidence, see the papers by Griffeath and Liggett [5], Kesten and Grimmett [10], and Lyons [12].

The first goal will make sense to anyone who, when told that one can get lost in three dimensions but not in two, asks "How come?"

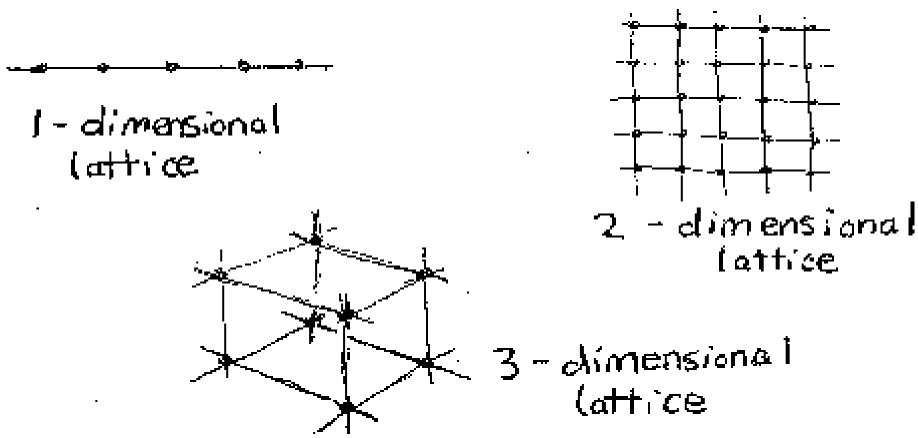

Figure 1: Lattices.

Part I
Electrical proof of Polya's theorem

1 The problem

1.1 Random walk on a lattice

In 1921 George Polya [19] investigated random walk on certain infinite graphs, or as he called them, "street networks." The graphs he considered, which we will refer to as "lattices," are illustrated in Figure 1.

To construct a d-dimensional lattice, we take as vertices those points $(x_l, \ldots, x_d)$ of R^d all of whose coordinates are integers, and we join each vertex by an undirected line segment to each of its $2d$ nearest neighbors. These connecting segments, which represent the edges of our graph, each have unit length and run parallel to one of the coordinate axes of R^d. We will denote this lattice by Z^d.

Now let a point walk around at random on this lattice. As usual, by walking at random we mean that upon reaching any vertex of the graph the probability of choosing any one of the $2d$ edges leading out of that vertex is

$1/(2d)$. To keep things specific, we will imagine that the point starts from the origin at time too and moves at unit speed. Thus the point arrives at a vertex only at integral times $t = 0, 1, 2, 3, \ldots$, at which times it is forced to choose between $2d$ equally attractive alternatives. (Note that there is nothing to prevent the point from returning the way it has just come. Indeed we expect this to happen about $1/(2d)$th of the time.)

When $d = 1$, our lattice is just an infinite line divided into segments of length one. We may think of the vertices of this graph as representing the fortune of a gambler betting on heads or tails in a fair coin tossing game. The random walk then represents the vicissitudes of his or her fortune, either increasing or decreasing by one unit after each round of the game.

When $d = 2$, our lattice looks like an infinite network of streets and avenues, which is why we describe the random motion of the wandering point as a "walk." When $d = 3$, the lattice looks like an infinite "jungle gym," so perhaps in this case we ought to talk about a "random climb," but we will not do so.

1.2 The question of recurrence

The question that Polya posed amounts to this: "Is the wandering point certain to return to its starting point during the course of its wanderings?" If so, we say that the walk is "recurrent." If not, that is, if there is a possibility that the point will never return to its starting point, then we say that the walk is "transient."

If we denote the probability that the point never returns to its starting point by p_{escape}, then the chain is

- recurrent iff $p_{\text{escape}} = 0$,
- transient iff $p_{\text{escape}} > 0$.

Hence, from our point of view, the question is not whether you can go home again—for indeed you can—but whether you can possibly avoid doing so.

1.3 Polya's original question

The definition of recurrence that we have given differs from Polya's original definition. Polya defined a walk to be recurrent if it is certain to pass through any specified point in the course of its wanderings. In our definition, we require only that the point return to its starting point. So we have to ask ourselves "Can the random walk be recurrent in our sense and fail to be recurrent in Polya's sense?"

The answer to this question is "no"—the two definitions of recurrence are equivalent. Why? Because if the point must return once to its starting point, then it must return there again and again, and each time it starts away from the origin it has a certain non- zero probability of making it over to a specified target vertex before returning to the origin. And since anyone can get a bull's-eye if he or she is allowed an infinite number of darts, eventually our wandering point will hit the target vertex.

1.4 Polya's theorem: recurrence In the plane, transience in 3-space

In the paper cited above, Polya proved the following theorem: Random walk on a d-dimensional lattice is recurrent for $d = 1, 2$ and transient for $d > 2$.

The rest of Part I is devoted to a proof of this theorem. Our approach will be to exploit the connection between questions about random walk on a graph and questions about electric currents in a corresponding network of resistors. The idea behind our proof is outlined in the next few subsections.

This proof is meant mainly as an indication of the benefits of applying electrical ideas to problems of random walk. These benefits will not be reaped until Part II; for now, the idea is to get a good feeling for what is going on.

1.5 The question is whether the effective resistance of a corresponding resistor network is infinite

As we will show later, random walk on a graph is recurrent if and only if a corresponding network of 1-ohm resistors has infinite resistance "out to infinity." (See Figure 2.) Since an infinite line of resistors obviously has infinite resistance, it follows that walk on the 1-dimensional lattice is recurrent, as stated by Polya's theorem. (See Figure 3.)

What happens in higher dimensions? We are asked to decide whether a d-dimensional lattice has infinite resistance out to infinity. The difficulty is that the d-dimensional lattice isn't nearly as symmetrical as the Euclidean space R^d in which it sits.

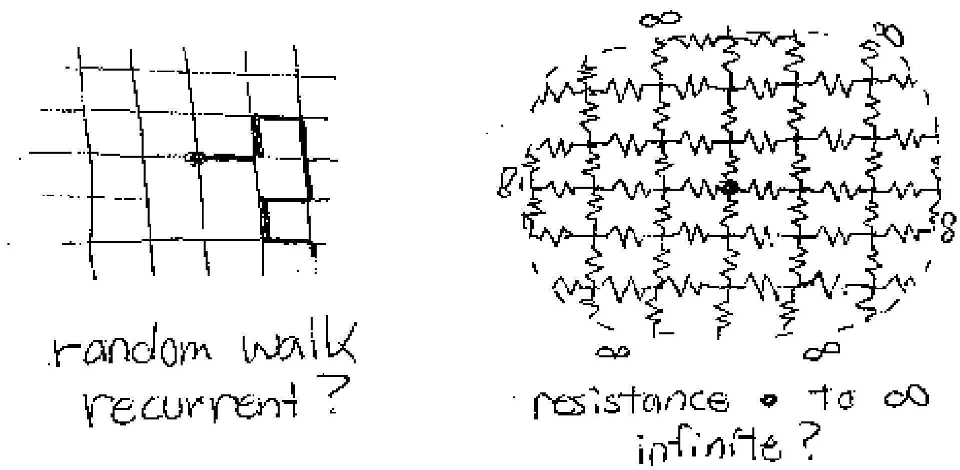

Figure 2: The equivalent resistance problem.

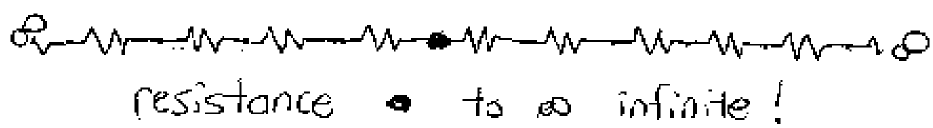

Figure 3: The 1-dimensional resistance problem.

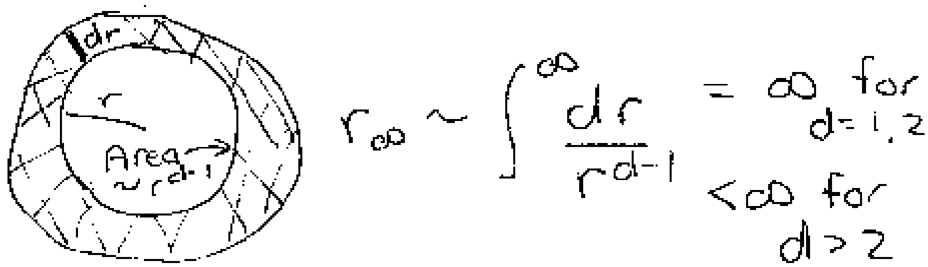

Figure 4: Can a physicist's scribblings prove Polya's theorem?

1.6 Getting around the asymmetry of the lattice

Suppose we replace our d-dimensional resistor lattice by a (homogeneous, isotropic) resistive medium filling all of R^d. and ask for the effective resistance out to infinity. Naturally we expect that the added symmetry will make the "continuous" problem easier to solve than the original "discrete" problem. If we took this problem to a physicist, he or she would probably produce something like the scribblings illustrated in Figure 4, and conclude that the effective resistance is infinite for $d = 1, 2$ and finite for $d > 2$. The analogy to Polya's theorem is obvious, but is it possible to translate these calculations for continuous media into information about what happens in the lattice?

This can indeed be done, as we will see in Part II. For the moment, we will take a different—though related—approach to getting around the asymmetry of the lattice. Our method will be to modify the lattice is such a way as to obtain a graph that is symmetrical enough so that we can calculate its resistance out to infinity. Of course, we will have to think carefully about what happens to that resistance when we make these modifications. The basic ideas are presented in the next subsection.

1.7 Shorting shows recurrence in the plane, cutting shows transience in 3-space

To take care of the case $d = 2$, we will modify the 2-dimensional resistor network by shorting certain sets of nodes together in such a way as to get a new network whose resistance is readily seen to be infinite. As the modifications made can only decrease the effective resistance of the network, the resistance of the original network must also have been infinite. Thus the walk is recurrent when $d = 2$.

To take care of the case $d = 3$, we will modify the 3-dimensional network by cutting out certain of the resistors so as to get a network whose resistance is readily seen to be finite. As the modifications made can only increase the resistance of the network, the resistance of the original network must have been finite. Thus the walk is transient when $d = 3$.

1.8 History of the method—Rayleigh and Nash-Williams

The method of applying shorting and cutting to get lower and upper bounds for the resistance of a resistive network was introduced by Lord Rayleigh in his paper *On the Theory of Resonance* [20]. For this reason, we refer to it as Rayleigh's short-cut method. This method is described in the classical treatises on electricity and magnetism (Maxwell [13], Jeans [9]). It was apparently first applied to random walk by C. St J. A. Nash-Williams [17], who used the shorting method to establish recurrence for random walk on the 2-dimensional lattice, and for "centrally-biased" random walk on higher dimensional lattices.

2 Formulation of the electrical analogy

2.1 Hitting probabilities are discrete harmonic functions

Let $V(i,j)$ denote the probability of reaching the origin for a random walk starting from (i,j). (To minimize notation, we are going to pretend that $d = 2$ throughout this subsection and the next.) For obvious reasons, we call $V(i,j)$ a hitting probability. We take $V(0,0) = 1$, considering a point starting from the origin as having reached it. The probability of starting from the origin and returning there is

$$\begin{aligned} p_{\text{return}} &= 1 - p_{\text{escape}} \\ &= \frac{1}{4}V(1,0) + \frac{1}{4}V(0,1) + \frac{1}{4}V(-1,0) + \frac{1}{4}V(0,-1) \\ &= \frac{1}{4}\sum V(\text{neighbors of } (0,0)). \end{aligned}$$

This formula is established by allowing the point to move over to one of its neighbors and then considering how likely the point is to get back to the origin from there. A similar argument shows that as long as $(i, j) \neq (0, 0)$ the function V satisfies the "discrete Laplace equation"

$$V(i, j) = \frac{1}{4} \sum V(\text{neighbors of } (i, j)).$$

This equation says that the value of the function V at (i, j) is the average of the values of V at the neighbors of (i, j). A function having this property is said to be "discrete harmonic," so we can sum up by saying that V is discrete harmonic everywhere except at $(0, 0)$.

Why the terms "discrete Laplace equation" and "discrete harmonic?" Because the equation

$$V(i, j) = \frac{1}{4} \sum V(\text{neighbors of } (i, j))$$

is the discrete analogue of Laplace's equation

$$\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} = 0.$$

A function $V(x, y)$ satisfying Laplace's equation, called a harmonic function, has the property that

$$V(x, y) = \text{average of } V \text{ over a circle with center } (x, y).$$

2.2 The electrical analogue

Let us regard the vertices of the lattice we are considering as the nodes of an infinite resistive network (network of resistors) whose branches are 1-ohm resistors corresponding to the edges of the lattice. (See Figure 5.) We want to translate the question of recurrence of the random walk into a question about the electrical properties of the resistive network.

As a preliminary measure, we temporarily reduce the network to a finite network by grounding all points "far away" from the origin. (See Figure 6.) The points we have grounded are now at fixed voltage zero. Using a battery, maintain the origin at voltage 1, and measure the voltage $U(i, j)$ at every node (i, j). (See Figure 7.)

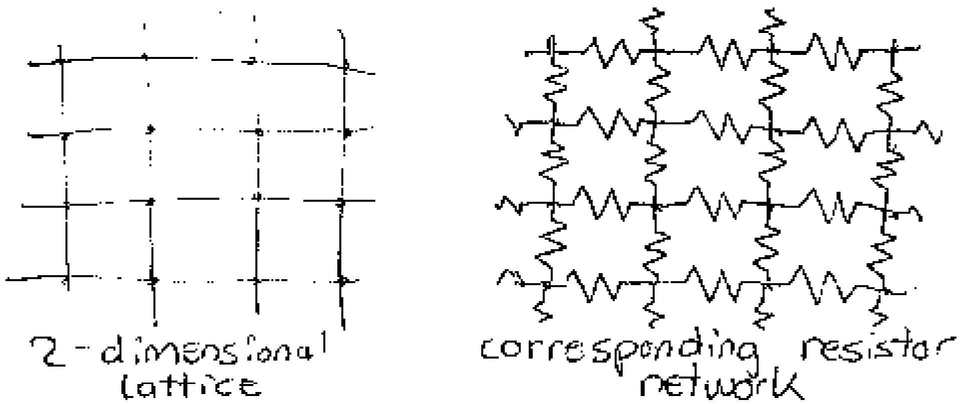

Figure 5: Making a lattice out of resistors.

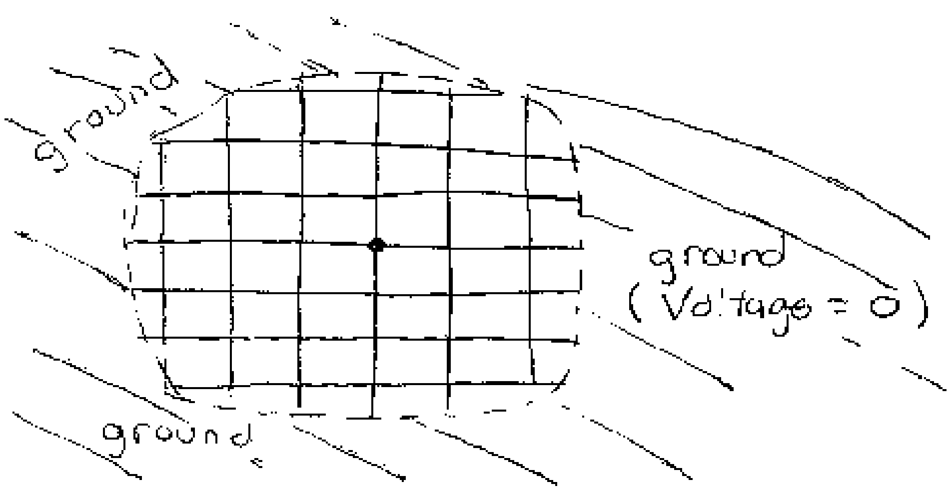

Figure 6: Grounding points far away from the origin.

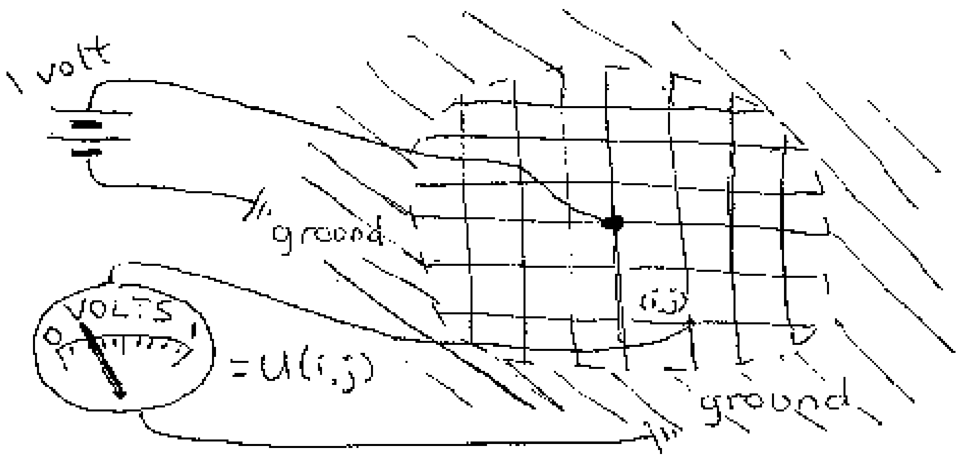

Figure 7: Hooking up a battery.

Evidently U is 1 at $(0,0)$ and 0 at nodes in the grounded region. Its values at the remaining nodes are determined by the equations of circuit theory. If you think about it, you will realize that these equations amount to the statement that U is discrete harmonic, that is, that U satisifies the discrete Laplace equation

$$U(i,j) = \frac{1}{4} \sum U(\text{neighbors of } (i,j))$$

at all nodes (i,j) other than $(0,0)$ and nodes in the grounded region.

In fact, it is not hard to show that U is the only function satisfying this equation and having boundary values

$$U(0,0) = 1, U(i,j) = 0 \text{ for nodes } (i,j) \text{ in the grounded region.}$$

(The proof involves a "maximum principle" argument.) But the probability that a random walk starting from (i,j) will reach the origin before reaching a point in the grounded region also has these properties, so we conclude that

$$U(i,j) = \text{Prob starting from } (i,j)(\text{reach 0 before reaching grounded region}).$$

If we now let the grounded region recede to infinity, then it is obvious that $U(i,j)$ will increase until in the limit it takes on the value $V(i,j)$, the probability that a point starting from (i,j) will eventually reach the origin.

Because of this, we may interpret $V(i,j)$ as the potential (voltage) measurable at (i,j) when a potential difference of 1 volt is maintained between the origin and a grounded region "at infinity." In the pages to follow we will often use this kind of loose but suggestive terminology in order to avoid a long series of straight-forward limiting arguments.

We have now established a correspondence that will allow us to turn questions about random walk into questions about the conduction of electricity. The implications of this correspondence for the recurrence question are worked out in the next section. This correspondence between hitting probabilities and voltages was established by showing that both satisfy the discrete Laplace equation, and then applying the principle that "the same equations have the same solutions." Here's how Maxwell ([13], p.70) expressed this principle:

> In many parts of physical science, equations of the same form are found applicable to phenomena which are certainly of quite different natures, as, for instance, electric induction through dielectrics, conduction through conductors, and magnetic induction. In all these cases the relation between the intensity and the effect produced is expressed by a set of equations of the same kind, so that when a problem in one of these subjects is solved, the problem and its solution may be translated into the language of the other subjects and the results in their new form will still be true.

Further remarks about this principle, and its application to systems satisfying the "indiscrete" Laplace equation, can be found in Feynman's Lectures ([4], Vol. II, Ch. 12).

2.3 Currents and resistances

We want next to shift gears and talk about currents and resistances rather than voltages. If we let p_{escape} denote as before the probability starting from the origin that we will never return there, then we obtain the following expression for the current C_∞ flowing out of the origin (and out to infinity):

$$\begin{aligned} C_\infty &= [V(0,0)-V(1,0)]+[V(0,0)-V(0,1)] \\ &\quad +[V(0,0)-V(-1,0)]+[V(0,0)-V(0,-1)] \\ &= 4-\sum V(\text{neighbors of }(0,0)) \end{aligned}$$

$$
\begin{aligned}
&= 4\left(1 - \frac{1}{4}\sum V(\text{neighbors of } (0,0))\right) \\
&= 4p_{\text{escape}}
\end{aligned}
$$

where we have used the expression for p_{escape} derived earlier. Thus the problem of recurrence of the walk, in terms of currents, becomes to determine whether $C_\infty = 0$ (recurrent walk) or $C_\infty > 0$ (transient walk). In other words, the presence of a current out to infinity in the electrical network corresponds to a transient walk, which agrees well with our intuition.

Now the effective resistance of the network "out to infinity" is inversely proportional to the current C_∞ (in fact it is exactly $1/C_\infty$, since we are applying a 1 volt potential difference). So in terms of resistances, our task is to determine whether the effective resistance R_∞ of the network between the origin and the grounded region at infinity is infinite (recurrent walk) or finite (transient walk). Again, that recurrent walk should correspond to infinite resistance out to infinity is intuitively appealing.

It is this version of the recurrence problem, phrased in terms of the effective resistance of the network, to which Rayleigh's method applies. We will introduce that method in the next chapter.

recurrent	$p_{\text{escape}} = 0$	$C_\infty = 0$	$R_\infty = \infty$
transient	$p_{\text{escape}} > O$	$C_\infty > 0$	$R_\infty < \infty$

Summary of recurrence criteria

3 Rayleigh's short-cut method

3.1 Shorting and cutting

As mentioned above, Rayleigh's method involves modifying the network whose resistance we are interested in so as to get a simpler network. We consider two kinds of modifications, shorting and cutting. Cutting involves nothing more than clipping some of the branches of the network, or what is the same, simply deleting them from the network. Shorting involves connecting a given set of nodes together with perfectly conducting wires, so that current can pass freely between them. In the resulting network, the nodes that were shorted together behave as if they were a single node, so they all attain the same potential. (See Figure 8.)

The process of grounding described earlier may be thought of as shorting the nodes that we want to ground together with a reference node whose

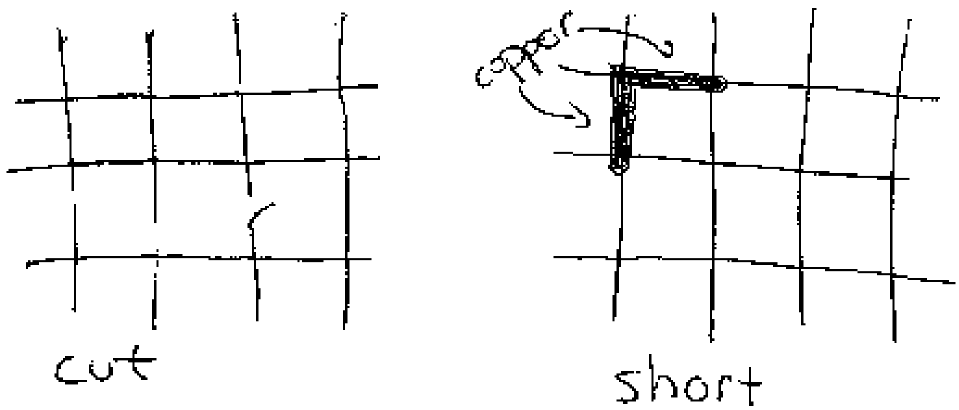

Figure 8: Shorting and cutting.

potential is fixed at zero. (Indeed this is precisely what we would do in practice.)

3.2 The monotonicity law

The usefulness of these two procedures (shorting and cutting) stems from the following observations:

Shorting Law: Shorting certain sets of nodes together can only decrease the effective resistance of the network between two given nodes.

Cutting Law: Cutting certain branches can only increase the effective resistance between two given nodes.

Monotonicity Law: More generally, the effective resistance between two given nodes is monotonic in the branch resistances.

Here's another statement of the monotonicity law, as it appears in Maxwell's *Treatise on Electricity and Magnetism* ([13], p. 427).

> If the specific resistance of any portion of the conductor be changed, that of the remainder being unchanged, the resistance of the whole conductor will be increased if that of the portion is increased, and diminished if that of the portion is diminished.

3.3 Why it's true

Maxwell goes on to assert that the monotonicity law "may be regarded as self-evident." Here's the intuitive idea: To begin with, although the monotonicity law is apparently stronger than the other two versions, it is a simple matter to derive it from either one of them, so it will be sufficient to convince ourselves that when we cut a resistor, the effective resistance increases. Now our feeling is that the electrons take the easiest route through the network. Cutting a branch can only make it more difficult for the electrons to get through the network—after all, no one was forcing them to pass along that branch if they didn't want to. Thus the effective resistance, which measures how hard it is for the electrons to go through the network, can only increase when we cut a branch.

This is admittedly a rather teleological view of the behavior of electrons. A real proof of these laws will be given in Part II. The proof relies on Thomson's minimum dissipation theorem, which is the precise way of expressing the "easiest route" idea.

Rayleigh's idea was to use the shorting and cutting laws to get lower and upper bounds for the resistance of a network. In the next section we apply this method to solve the recurrence problem for random walk in dimensions 2 and 3.

4 Application of Rayleigh's method

4.1 The plane is a snap

When $d = 2$, we apply the shorting law as follows: Short together nodes on squares about the origin. (See Figure 9.) The network we obtain is equivalent to the network shown if Figure 10. Now as P 1-ohm resistors in parallel are equivalent to a single resistor of resistance $1/P$ ohms, the modified network is equivalent to the network shown in Figure 11. The resistance of this network out to infinity is

$$\sum_{n=1}^{\infty} \text{Order}(\frac{1}{n}) = \infty.$$

As the resistance of the old network can only be bigger, we conclude that it, too, must be infinite, so that the walk is recurrent when $d = 2$. As

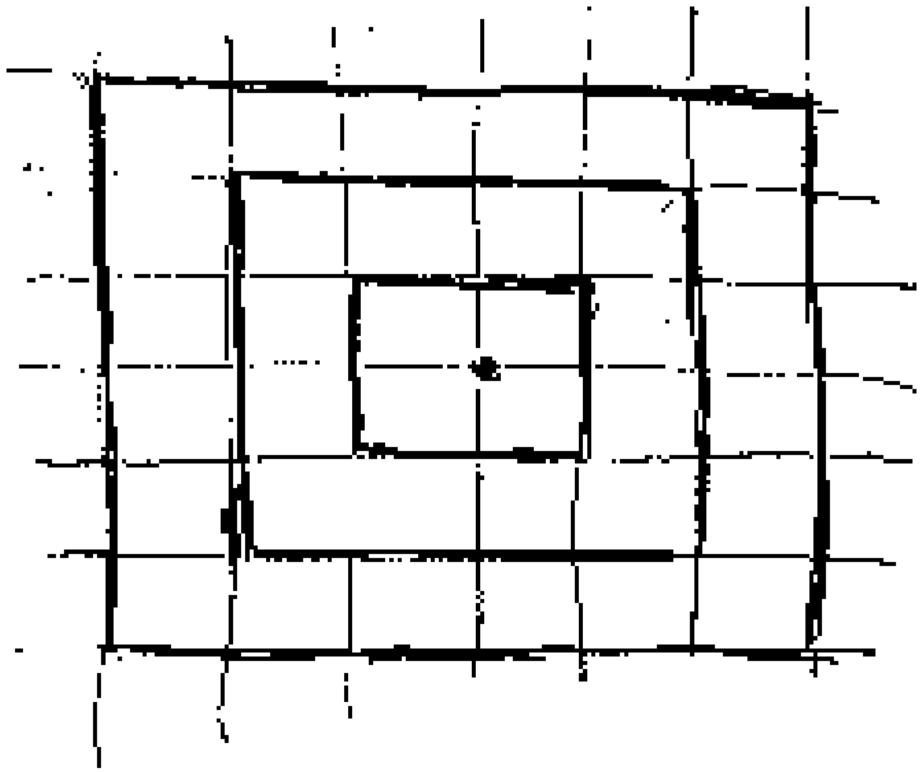

Figure 9: Shorting squares together.

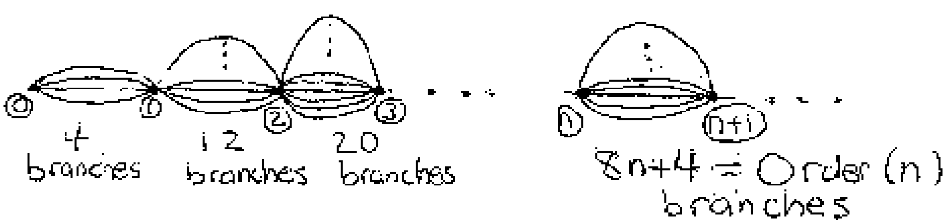

Figure 10: An electrical equivalent.

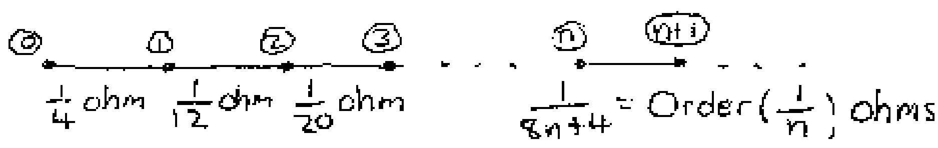

Figure 11: Another equivalent network.

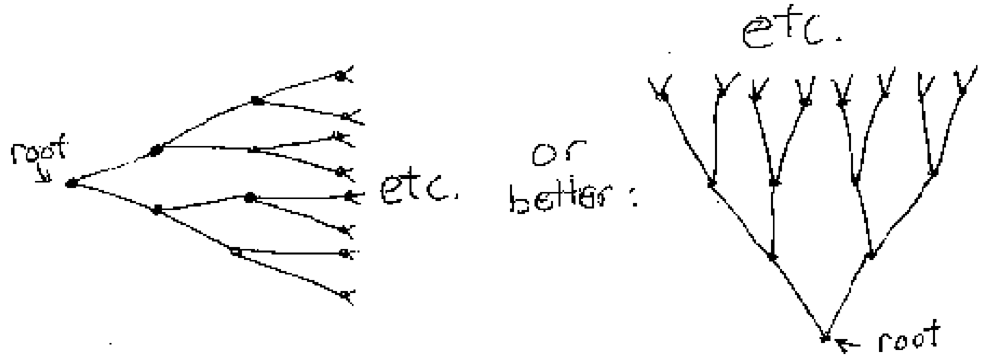

Figure 12: The full binary tree.

noted before, this method of showing recurrence in the plane was used by Nash-Williams [17].

4.2 3-Space: searching for a residual network

When $d = 3$, what we want to do is to delete certain of the branches of the network so as to leave behind a residual network having manifestly finite resistance. Obviously, the problem is to reconcile the "manifestly" with the "finite." We want to cut out enough edges so that the effective resistance of what is left is easy to calculate, while leaving behind enough edges so that the result of the calculation is finite.

4.3 Trees are easy to analyze, especially if they are pretty symmetrical, like the full binary tree

Trees—that is, graphs without loops—are undoubtedly easiest to deal with. For instance, consider the "full binary tree" shown in Figure 12. Notice that sitting inside this tree just above the root are two copies of the tree itself. This "self-similarity" property can be used to compute the effective resistance R_∞ from the root out to infinity. It turns out that $R_\infty = 1$. We will demonstrate this below by a more direct method.

To begin with, let us determine the effective resistance R_n between the root and the set of nth generation branch points. (See Figure 13.) To do this, we should ground the set of branch points, hook the root up to a 1-volt battery, and see how much current flows. (See Figure 14.) In the

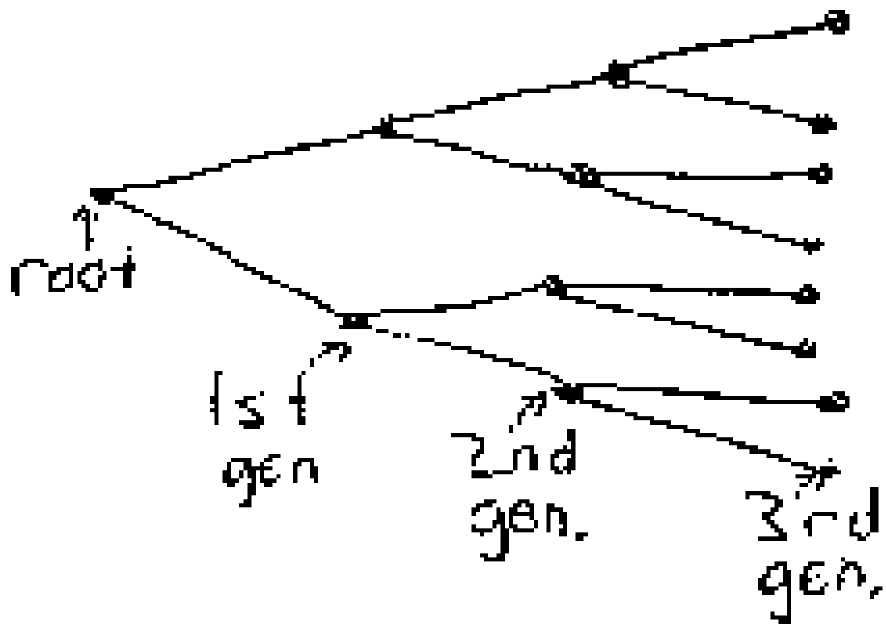

Figure 13: Generations in a tree.

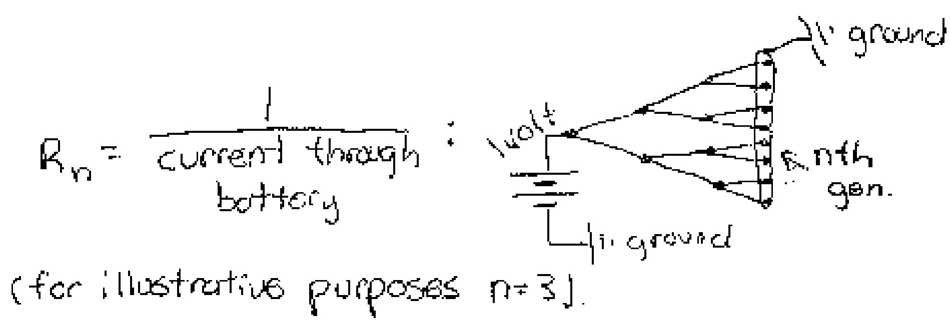

Figure 14: Computing the resistance unto the nth generation.

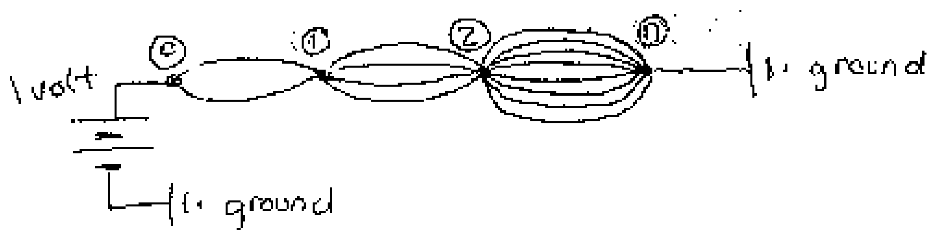

Figure 15: The network with equivalent nodes shorted together.

resulting circuit, all branch points of the same generation are at the same potential (by symmetry). Nothing happens when you short together nodes that are at the same potential. Thus shorting together branch points of the same generation will not affect the distribution of currents in the branches. (See Figure 15.) In particular, this modification will not affect the current through the battery, and we conclude that

$$\begin{aligned} R_n &= \frac{1}{\text{current in original circuit}} \\ &= \frac{1}{\text{current in modified circuit}} \\ &= \text{effective resistance(the network in Figure 15 above)} \\ &= \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^n} \\ &= 1 - \frac{1}{2^n}. \end{aligned}$$

Letting $n \to \infty$, we get

$$R_\infty = \lim_{n\to\infty} R_n = \lim_{n\to\infty} 1 - \frac{1}{2^n} = 1.$$

A closely related tree is the "tree homogeneous of degree three." (See Figure 16.) Note that all nodes of this tree are similar—there is no intrinsic way to distinguish one from another. To find the resistance out to infinity, we notice that this tree is "$1\frac{1}{2}$ copies" of the full binary tree, so its resistance must be 2/3.

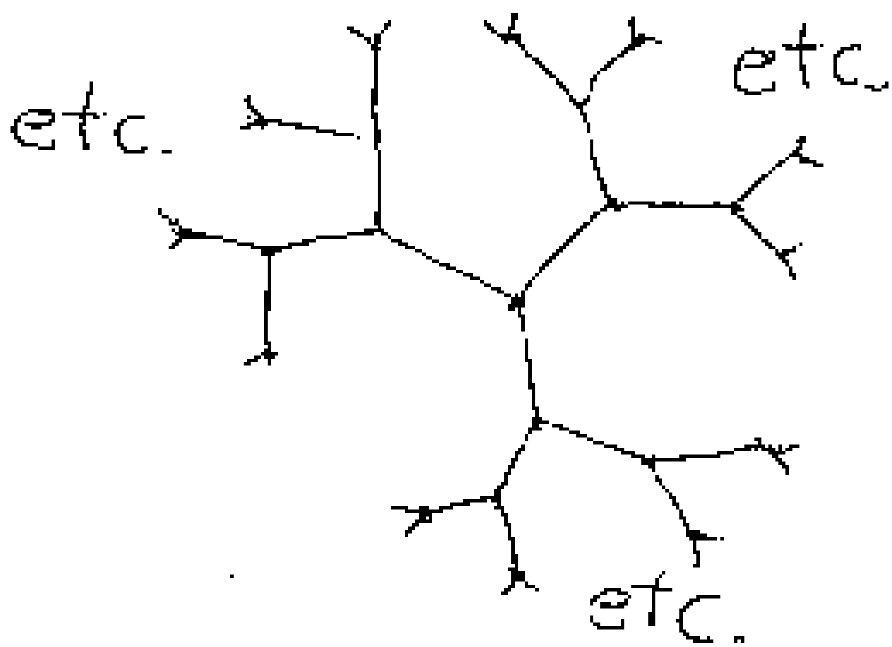

Figure 16: The tree homogeneous of degree three.

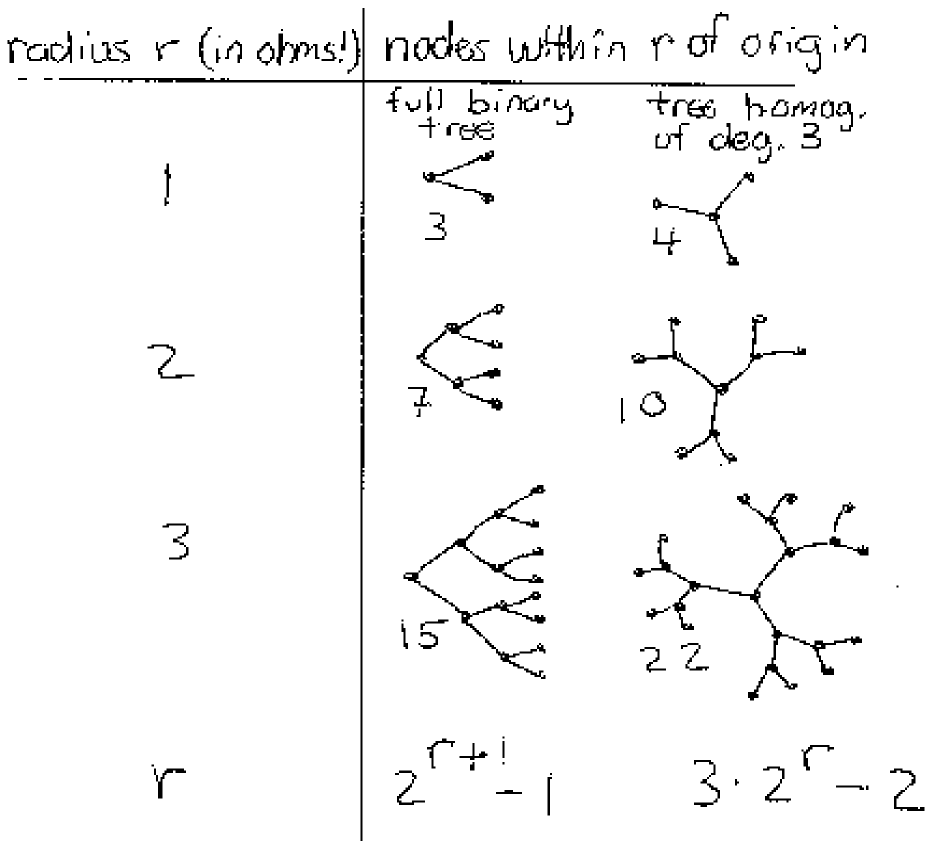

Figure 17: Exponential growth of balls.

4.4 The full binary tree is way too big

Nothing could be nicer than these two trees. They are the prototypes of networks having manifestly finite resistance out to infinity. Unfortunately, we can't even come close to finding either of these trees as a subgraph of the three-dimensional lattice. For in these trees, the number of nodes within a given radius r of the origin grows exponentially with r, whereas in a d-dimensional lattice, it grows as r^d. (See Figure 17.)

Here's a way to think of it. Take any old infinite (connected) graph and select one node to be called the origin. For instance we might be talking about one of our trees, in which case we would refer to the root as the origin. Let us call the set of nodes within a distance r of the origin as a "ball of radius r." (By the distance of a point to the origin, we mean the length

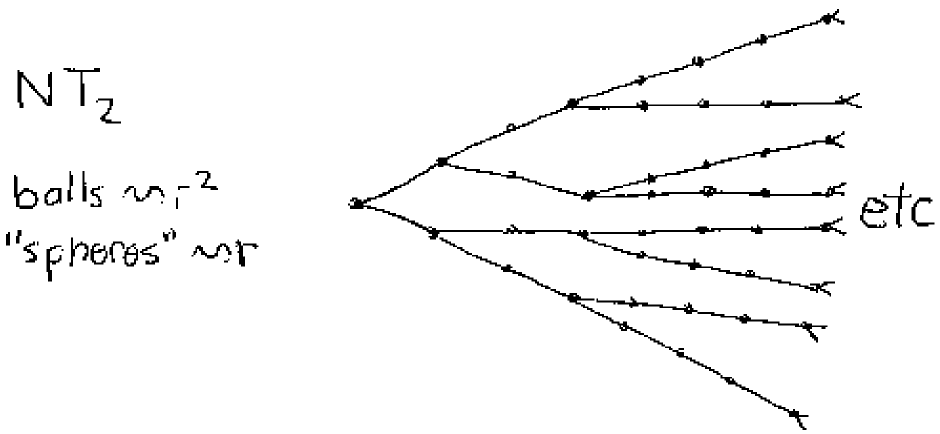

Figure 18: NT_2: A two-dimensional tree.

of the shortest path between that point and the origin.) Of course we need only consider balls of radius n, where n is an integer. Now if it is possible to embed the graph we are considering in a d-dimensional lattice, then it is absolutely necessary that the size of a ball of radius r in our graph be no bigger than the size of a ball of radius r in the lattice (the number of nodes of the lattice which can be gotten to from the origin along a path in the lattice of length at most r). Now without looking too closely at this question, it is clear that there are going to be something like r^d points in a ball of radius r in the lattice, just as the volume of a ball of radius r in d-dimensional space is $\text{constant}_d r^d$.

4.5 NT_3: A "three-dimensional" tree

These observations suggest that we would do well to look around for a nice tree NT_3 where the number of nodes within a radius r of the root is on the order of r^3. For we might hope to find something resembling NT_3 in the three-dimensional lattice, and if there is any justice in the world, this tree would have finite resistance out to infinity, and we would be done.

Before describing NT_3, let's tackle NT_2, our choice for the tree most likely to succeed in the 2-dimensional lattice. It is shown in Figure 18. The idea here is that since "balls of radius r" ought to contain something like r^2 points, the "spheres of radius r" ought to contain something like r points, so

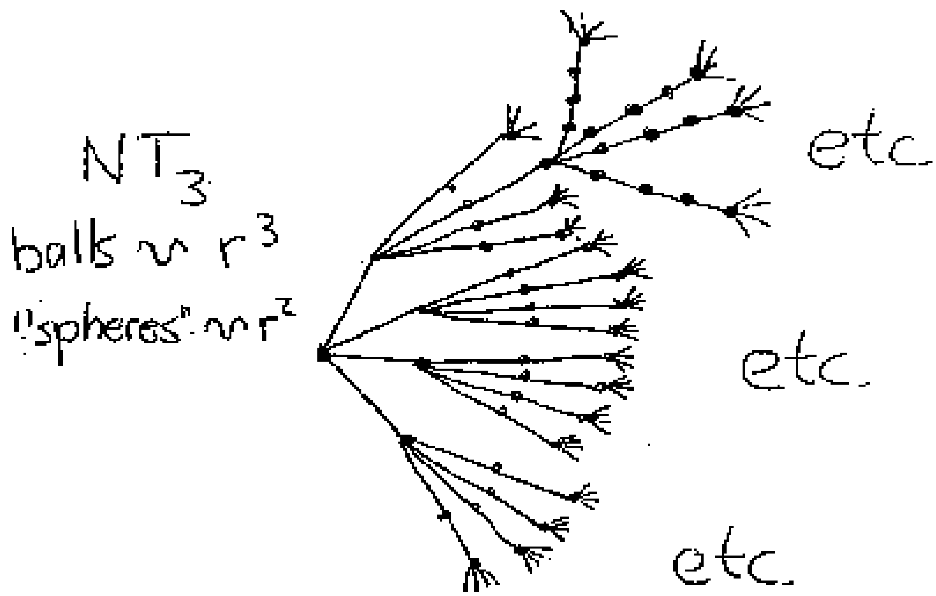

Figure 19: NT_3: A three-dimensional tree.

the number of points in a “sphere” should roughly double when the radius of the sphere is doubled. For this reason, we make the branches of our tree split in two every time the distance from the origin is (roughly) doubled.

Similarly, in a 3-dimensional tree, when we double the radius the size of a sphere should roughly quadruple. The result is the tree NT_3, shown in Figure 19. Obviously, NT_3 is none too happy about being drawn in the plane. Nor for that matter were the first two trees we discussed, which are in a certain sense infinite-dimensional.

4.6 NT_3 has finite resistance

To see if we're on the right track, let's work out the resistance of our new trees. The computations are shown in Figures 20 and 21. As expected, we find that NT_2 has infinite resistance to infinity and NT_3 has finite resistance to infinity.

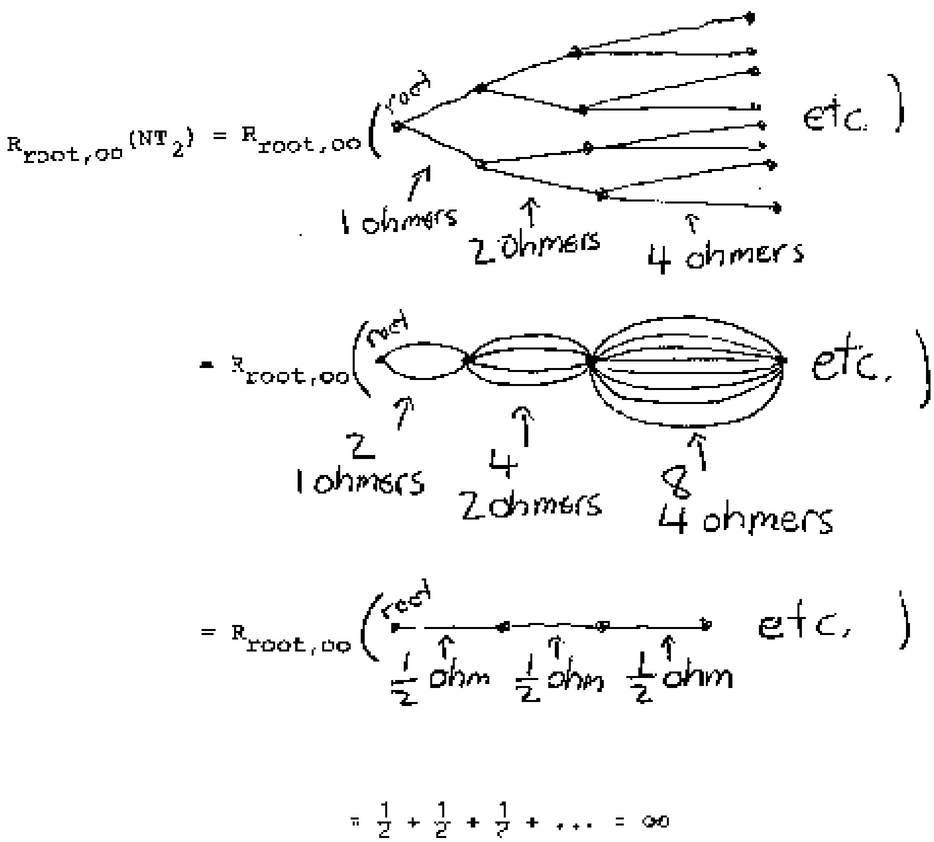

Figure 20: Computing the resistance of NT_2.

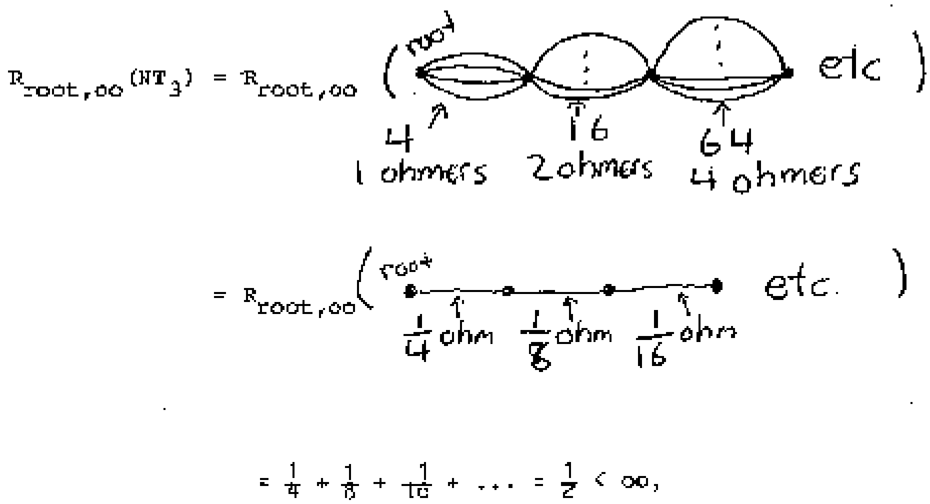

Figure 21: Computing the resistance of NT_3.

4.7 But does NT_3 fit in the 3-dimensional lattice?

Things are looking good. If we can find something like NT_3 in the 3-dimensional lattice, we will be done. Before we go looking for it, however, it is probably wise to begin by finding something like NT_2 in the two-dimensional network. The desired subgraph is shown in Figure 22. Figure 23 shows a caricature of this subgraph, distorted so as to display its structure: This graph differs from NT_2 only in that corresponding to each 1-ohm resistor of NT_2 we have here a series of three 1-ohmers. These three resistors behave together like a single 3-ohm resistor, so we are off by a factor of three. In other words, any effective resistance that we compute in the embedded network will be three times larger than the corresponding resistance computed in NT_2. But as our investigations concern only whether certain resistances approach infinity or not, this factor of 3 is of no concern to us.

How was this graph constructed? We can visualize this process as follows (see Figure 24): Two particles leave the origin at speed 3, one heading north, one east. At time $t = 1$, these two particles are located at the points $(3, 0)$ and $(0, 3)$. Both split into two particles moving at speed 3, one heading north, one east (obviously momentum is not conserved). Two of these particles are headed for a collision at $t = 2$. Instead, we let them "bounce" so that at $t = 3$, we have particles at (9,0),(6,3),(3,6), and (0,9). These particles then

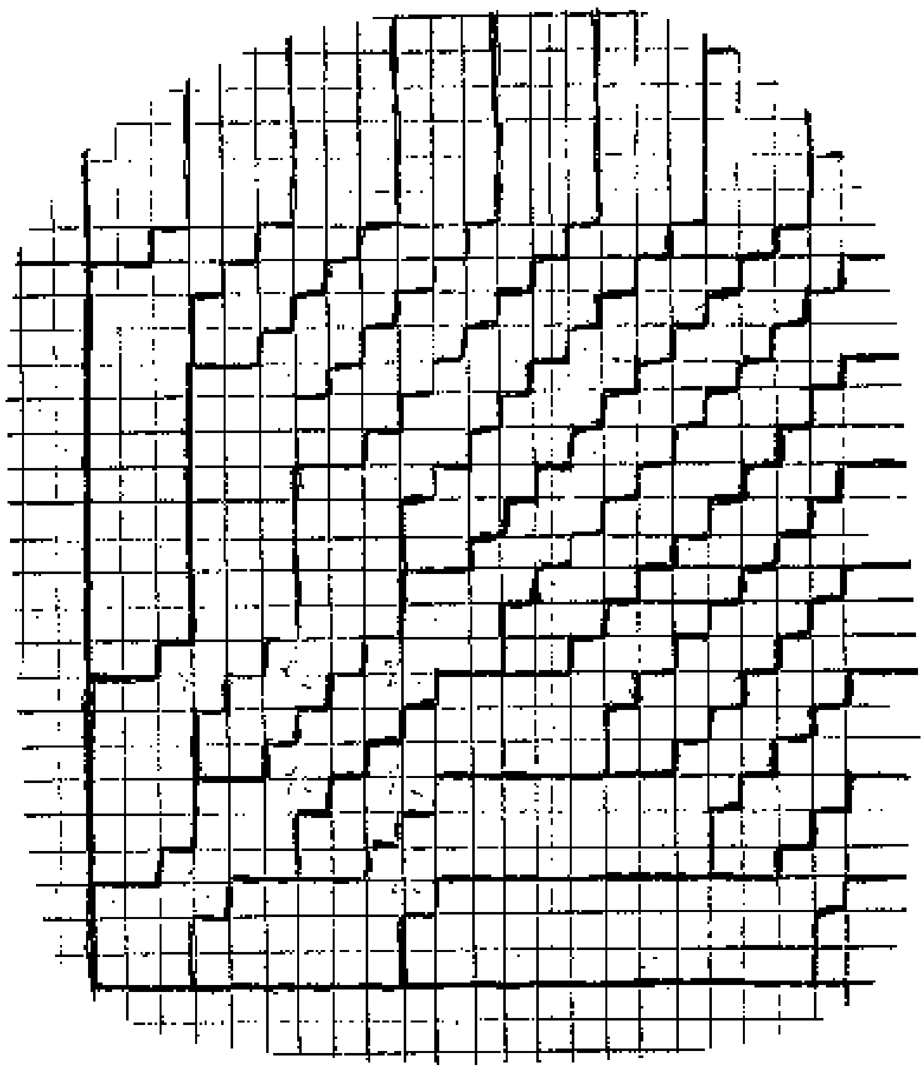

Figure 22: Embedding NT_2.

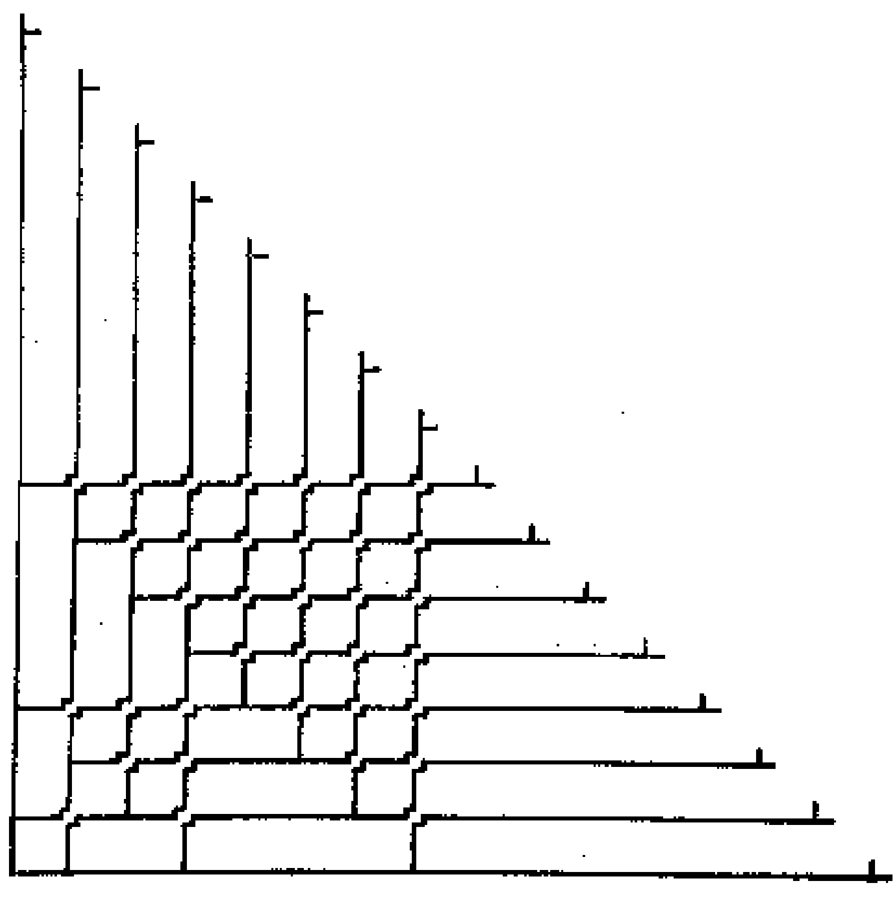

Figure 23: A caricature of the embedded tree.

split in two and move along, bouncing off each other whenever they would collide until at time t=7 we have particles at [(0,21),(3,18),...,(21,0)]. These now split, and so forth. The trails left by the moving particles represent our embedded graph.

So much for two dimensions. What if we imagine the same procedure taking place in three dimensions? That is, let 3 particles start from the origin at speed 3, heading north, east, and up. At time $t = 1$, they split, bounce (in pairs), split again,... What we end up with is not NT_3! Instead, it is a subgraph of the three-dimensional lattice resembling—again up to a factor of 3—the tree $\mathrm{NT}_{2.5849\ldots}$ shown in Figure 25. We call this tree $\mathrm{NT}_{2.5849\ldots}$ because it is $2.5849\ldots$ dimensional in the sense that when you double the radius of a "ball," the size of the ball gets multiplied roughly by

$$6 = 2^{\log_2 6} = 2^{2.5849\ldots}.$$

So we haven't come up with our embedded NT_3 yet. But why bother? The resistance of $\mathrm{NT}_{2.5849\ldots}$ out to infinity is

$$\begin{aligned} R_{\mathrm{root},\infty}(\mathrm{NT}_{2.5849\ldots}) &= \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \ldots \\ &= \frac{1}{3}(1 + \frac{2}{3} + \frac{4}{9} + \ldots) \\ &= \frac{1}{3}\frac{1}{1 - \frac{2}{3}} \\ &= 1. \end{aligned}$$

The resistance of our embedded subgraph is three times larger, namely 3 ohms. Thus we have found a subgraph of the 3-dimensional lattice having finite resistance out to infinity, and we are done.

5 How we did it

We have at last carried through our plan of how to prove Polya's theorem. Our success was based on our ability to relate the question of recurrence to a question concerning a physical system about which we had pretty well developed intuitions. In part II we will see just how much this relationship can do for us.

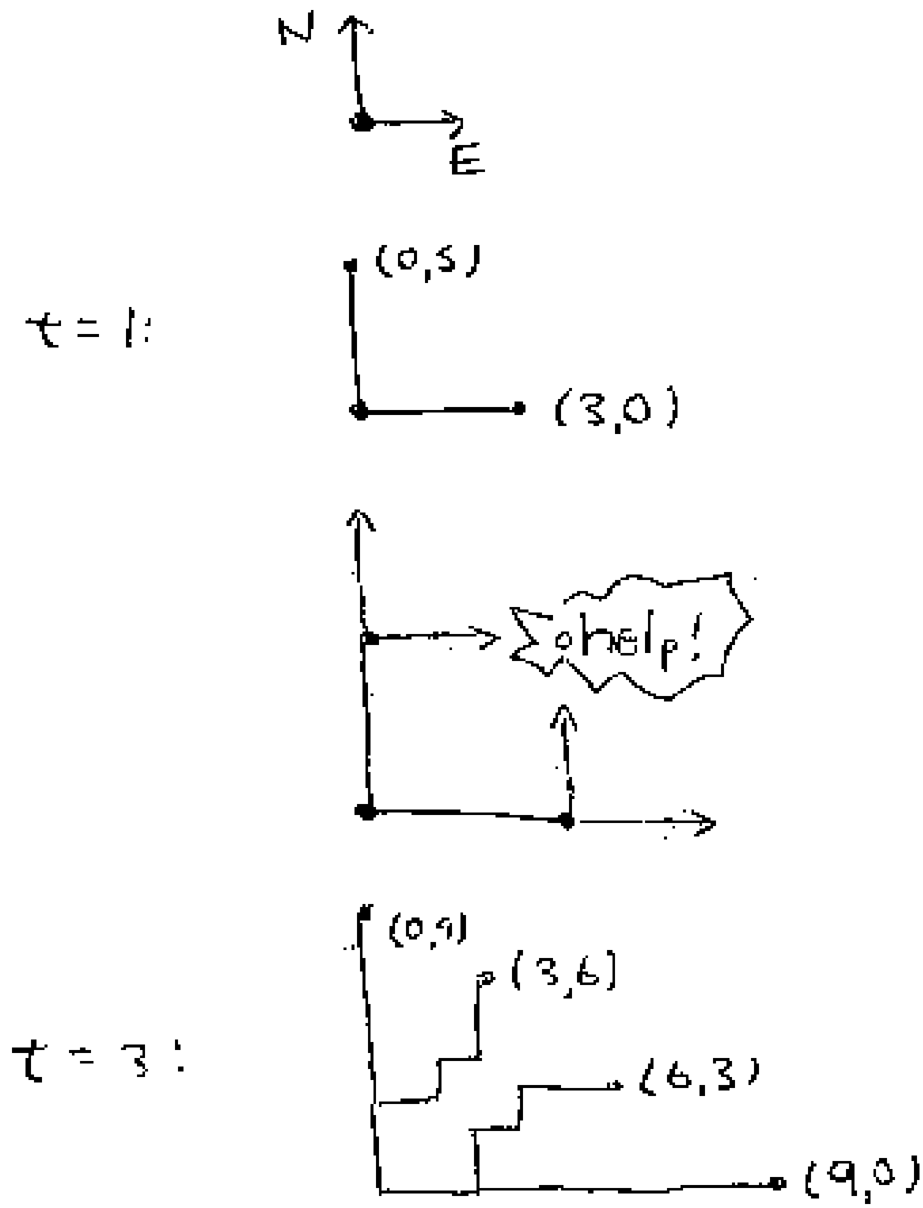

Figure 24: A dynamical picture of the construction.

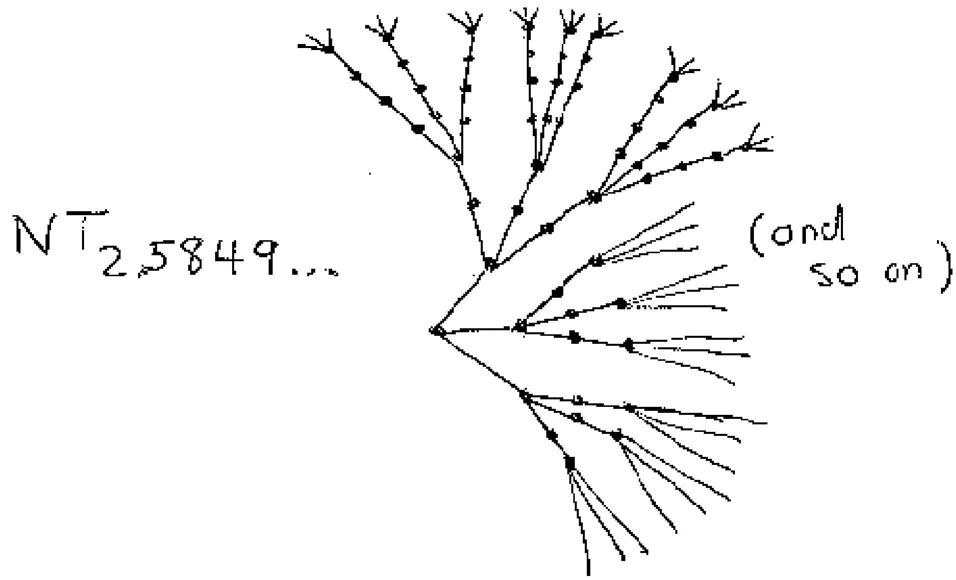

Figure 25: $NT_{2.5849\ldots}$: A 2.5849...-dimensional tree.

Part II

Questions raised by Polya's theorem

In Part II, I am going to describe a number of applications of Rayleigh's short-cut method to Polya's recurrence theorem. The idea will be to try to get to the point where we have an intuitive feeling for "why" Polya's theorem is true. Our approach will be to ask ourselves a bunch of natural questions about Polya's theorem, the kind of questions likely to occur to anyone who contemplates it, and then show how Rayleigh's methods can be used to answer these questions.

Before starting in on this, however, I would like to fill in a little of the background on Rayleigh's method, and explain the various guises in which the method will appear.

6 Rayleigh and resonance

Rayleigh introduced the short-cut method for estimating the effective resistance of an electrical conductor in a paper entitled "On the Theory of Resonance" [20], in which he considered the problem of determining the resonant frequency of a hollow body. You may wonder why an electrical method should make its debut in a paper on acoustics, so let me say a word or two about this.

Rayleigh's idea was to imagine the air inside the hollow body acting like a spring, alternately drawing the air into the body and pushing it out. The stiffness of the spring was to depend on the volume of the body, a smaller body corresponding to a stiffer spring and hence to a higher resonant frequency; the mass to be attached to the spring was to depend on the effective size of the mouth of the vessel, a larger mouth corresponding to a smaller mass and hence to a higher resonant frequency.

To calculate the effective size of the mouth, he imagined the motion of air through the mouth to be incompressible, i.e. potential flow, and came in that way to a problem of potential theory whose electrical analog was that of determining the effective resistance from the inside of the vessel—through the mouth—to the outside.

Now except in the very simplest of cases it was not possible to compute this effective resistance explicitly, so the question became how to get good approximate values for this resistance. The method Rayleigh hit upon was the short-cut method. We introduced this method in Part I, justifying it by an appeal to physical intuition. Let us now take a closer look at this method.

As indicated in Part I, the short-cut method is closely related to Thomson's minimum dissipation theorem, the intuitive content of which is that a current passing through a resistive medium takes the easiest route. As an illustration of Thomson's theorem, let us consider a network of resistors with distinguished nodes A and B. (See Figure 26.)

If we pass a 1-amp current through the network by shoving charge into node A and pulling it out of node B. there are many ways for the current to distribute itself through the branches of the network. Thomson's theorem says that the actual distribution of the currents will be that distribution that minimizes the *dissipation rate*

$$D = \sum_b R_b I_b^2,$$

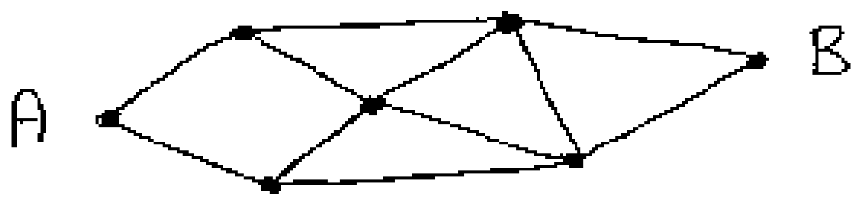

Figure 26: A network with two distinguished nodes A and B.

where the sum runs over all branches (i.e. resistors) of the network, and where R_b is the resistance of branch b and I_b the current through it. D represents the rate at which heat is being generated, i.e. the rate at which free energy is being dissipated, hence the statement that the current takes the easiest route.

Now since the current we are shoving through the network is 1 amp, the actual rate at which heat is dissipated is equal to the effective resistance of the network ($D = RI^2$). Thus any set of currents representing a flow of 1 amp from A to B gives us an upper bound for the effective resistance of the network.

To see how to get lower bounds, and to see the relationship to shorting and cutting, recall the shorting and cutting laws:

Shorting Law: Shorting out a branch of the network decreases the effective resistance from A to B.

Cutting Law: Cutting a branch of the network increases the effective resistance from A to B.

These are both limiting cases of the

Monotonicity Law: Increasing the resistance of any branch increases the effective resistance from A to B.

All three laws are equivalent, in the sense that it is easy to derive any one of them from either of the others, and all are easy consequences of Thomson's theorem. For instance, to prove the cutting law, consider that the effective resistance from A to B is the minimum of the dissipation rate D over a certain class of flows, and cutting a branch of the network has the effect of cutting down on that class of flows.

On the other hand, it is possible to view Thomson's theorem as a consequence of the cutting law, and this was apparently the attitude taken by Rayleigh [20]; see also Maxwell ([13], pp. 429-431) and Onsager [18].

To appreciate this attitude, consider the flow of electricity in a continuous medium. For a continuous medium, cutting may be interpreted as introducing surfaces through which we don't allow the current to pass. Given any 1-amp flow through the medium, we can think of the medium as being divided up into the infinitesimal tubes of flow of the given flow. We now imagine cutting the medium along all these tubes of flow. The effect will be to specify completely the streamlines of any admissible flow through the medium. This doesn't quite restrict the admissible flows to the flow we started out with, since there may be other flows having the same streamlines, but let's ignore this point. Then the dissipation of the flow we started with is equal to the effective resistance of the medium after the cutting, whereas the dissipation of the true 1-amp flow is equal to the effective resistance of the medium before the cutting. Applying the cutting law yields Thomson's theorem.

This explains how we can regard Thomson's theorem as being more or less equivalent to the short-cut method. This is good, because we may want to apply Thomson's theorem directly and still use the catchy name "short-cut method." This argument also explains how we can get from the original form of Thomson's theorem, which looks like it should only yield upper bounds for the resistance of a conductor, to a form yielding both upper and lower bounds.

As an aside, let me mention that there is a more direct way to get lower bounds out of Thomson's theorem. The route is through the duality that exists between currents and voltages. This duality, long familiar to electricians, is known to mathematicians as Hodge duality [8]. Under this duality resistance corresponds to conductance—i.e., inverse resistance—so to a principle yielding upper bounds for resistance there corresponds a principle yielding upper bounds for conductance, that is, lower bounds for resistance. This dual principle is what is mistakenly known as Dirichlet's principle; its true discoverer was Thomson, which isn't too surprising.

The duality referred to in the last paragraph is the same duality that arises in the study of planar graphs. It finds its most natural expression in Whitney's matroid theory [22], specifically in the theory of regular matroids, the class of matroids for which an electrical theory can be developed. For more about this, see the survey articles by Duffin [3] and Minty [16] .

7 Questions about Polya's theorem

This section records an imaginary dialog on the subject of Polya's theorem. The idea is to ask natural questions about Polya's theorem, seeing how many we can answer as we go, and saving up those we can't.

First, some conventions. In addition to random walks on the lattices $\mathbf{Z}^d$, we will want to talk about random walks on more general infinite graphs. We will restrict ourselves to *connected graphs* having *no loops*, though they may occasionally have *multiple edges*. We will further restrict ourselves to graphs where every vertex has *finite valence*. By random walk on such a graph we will always mean *simple random walk*, whereby upon reaching a vertex the walker choose randomly one of the edges leading out of that vertex, giving each edge equal weight.

We will refer to the problem of determining whether random walk on a given graph is recurrent or transient as the *type problem*. This terminology is borrowed from the theory of Riemann surfaces (cf. the Appendix). Just as in the case of a lattice, the type problem for a general infinite graph is equivalent to determining whether in a corresponding network of 1-ohm resistors the resistance out to infinity is infinite (recurrent type) or finite (transient type); the resistance to infinity is again defined as a limit of resistances of finite graphs.

Warning: On a graph whose valence varies from vertex to vertex, simple random walk is not quite as congenial as it is on a lattice. This has to do with the fact that for simple random walk on a finite graph, the limiting probability of being at a given vertex will be proportional to the valence of that vertex. In the discussions to follow we will make sure to steer clear of this problem, but it is well to be aware of it.

The most obvious question to ask about Polya's theorem is: Why does the type change from recurrent to transient when you raise the dimension? Presumably because in higher dimensions there is more room to get lost. But how could we make this precise? And what about the obvious generalization that a graph G is always more likely to be transient than any of its subgraphs, in the sense that if any subgraph is transient then G must also be transient?

Having made it through Part I we know right away how to answer these questions. All we have to do is to view the type problem as an electrical problem and appeal to Rayleigh's method, in the form of the cutting law. This is already quite a triumph for Rayleigh's method, as you will appreciate if you try to prove this monotonicity property probabilistically.

The second most obvious question to ask is: Why 2? I mean, granted that $\mathbf{Z}^d$ may change from recurrent to transient when d becomes large enough, why should the change occur between 2 and 3?

To answer this, recall an argument that was given back in Part I: Looked at from far enough away the lattice $\mathbf{Z}^d$ looks a lot like the surrounding space $\mathbf{R}^d$. Thus instead of trying to determine the resistance of the lattice we may try to determine the resistance out to infinity of a resistive medium filling all of $\mathbf{R}^d$. Because of the rotational symmetry of $\mathbf{R}^3$, we can express this as an integral

$$\int_\alpha^\infty \frac{1}{r^{d-1}} dr.$$

The value $d = 2$ is where this integral changes from divergent to convergent. This is the reason, we say, that the change from recurrent to transient should occur at $d = 2$.

The effect of this argument is to replace the original question by two questions, one serious and one not-so-serious. The serious question, already raised back in Part I, is this:

First Main Question: How can we relate the discrete problem to its continuous analog?

This is, to my mind, easily the most important question raised by Polya's theorem. We will run up against it again in a slightly different guise a little later in the section. Here the analogy is between currents on the lattice and currents in space, i.e. between the discrete Laplace equation and the honest-to-goodness Laplace equation; there the analogy will be between the discrete and continuous forms of the heat equation. In either case, the question is the same: How can we make mathematical sense of our feeling that from far away you can't distinguish $\mathbf{Z}^d$ from $\mathbf{R}^d$? As we will see, Rayleigh's method provides a simple way of answering this question.

The not-so-serious question is this: So we have reduced the question of the critical value of d in Polya's theorem to a question about the divergence of the integral

$$\int_\alpha^\infty \frac{1}{r^{d-1}} dr.$$

Why should this integral choose $d = 2$ as the spot where it decides to change its manners? Believe it or not, we will have something to say about this question also.

In order to come up with further questions, let's think about how we might go about proving Polya's theorem by probabilistic methods. We'll begin in dimension 1.

We want to convince ourselves that the probability p_{escape} of walking off and never coming back is 0. This is equivalent to saying that the expected number of times at 0 is infinite, since

$$\mathbf{E}(\text{number of times at } 0) = \frac{1}{p_{\text{escape}}}.$$

To see that $\mathbf{E}(\text{number of times at } 0) = \infty$ we write

$$\mathbf{E}(\text{number of times at } 0) = \sum_{n=0}^{\infty} P_n(0),$$

where

$$P_n(x) = \mathbf{P}(\text{at } x \text{ at time } n).$$

Since

$$P_n(0) = \begin{cases} 0, & n \text{ odd} \\ \text{Order}(\frac{1}{\sqrt{n}}), & n \text{ even} \end{cases}$$

we have

$$\begin{aligned} \mathbf{E}(\text{number of times at } 0) &= \sum_{n=0}^{\infty} \text{Order}(\frac{1}{\sqrt{n}}) \\ &= \infty. \end{aligned}$$

Now how did we know that

$$P_n(0) = \text{Order}(\frac{1}{\sqrt{n}})?$$

The simplest way out of this is to appeal to Stirling's formula

$$n! \approx \sqrt{2\pi n} e^{-n} n^n.$$

To be back at 0 after n steps we must have taken $n/2$ steps to the left and $n/2$ steps to the right, and the probability of this is

$$\begin{aligned} \binom{n}{\frac{n}{2}} \frac{1}{2^n} &= \frac{n!}{(\frac{n}{2})!^2} \frac{1}{2^n} \\ &\approx \sqrt{2\pi n} e^{-n} n^n \left(\sqrt{\pi n} e^{-\frac{n}{2}} \left(\frac{n}{2} \right)^{\frac{n}{2}} \right)^{-2} \frac{1}{2^n} \\ &= \sqrt{\frac{2}{\pi n}}. \end{aligned}$$

This last argument doesn't do much to dispel the mystery of the square root. One way to explain this is in terms of the fact that when you add two independent random variables their variances add. If we let X_i be independent identically-distributed random variables taking on the values +1 and -1 with probability 1/2 , and set

$$S_n = \sum_{i=1}^{n} X_i$$

then

$$P_n(0) = \mathbf{P}(S_n = 0).$$

The variance of the X_i is 1 , so the variance of S_n is n, i.e. its standard deviation is $\sqrt{n}$. Now look at $P_n(x)$ as a function of x. This is a probability density on the integers, stretched out to have standard deviation $\sqrt{n}$, so it seems reasonable that the height at the middle, namely $P_n(0)$, should be on the order of $\frac{1}{\sqrt{n}}$.

This reduces the mystery of the square root to the mystery of why when you add two independent random variables their variances add. This is the Second Fundamental Mystery of probability theory, the First being the mystery of why when you add two random variables, independent or not, their expected values add.

Of course none of this has anything to do with the real reason that we expect to see the square root. The real reason is that we remember the fundamental limit theorem of probability, the so-called "Central Limit Theorem," which says that for n large enough the measure on $\mathbf{R}^1$ associated to the function $P_n(x)$ is approaching in an appropriate sense the measure with density

$$\frac{1}{\sqrt{2\pi n}} e^{-\frac{x^2}{n}}.$$

There's the square root all right, but where on earth did this formula come from? The answer is, from the heat equation

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}.$$

This is the equation that governs the flow of heat in a one-dimensional bar of infinite length; under this interpretation $u(t, x)$ represents temperature. More to the point, this is also the equation that governs diffusion in a one-dimensional fluid; under this interpretation $u(t, x)$ represents concentration.

Since a diffusing particle carries out the continuous analog of a random walk, it is no wonder that we run into the heat equation here. Nor is it any wonder that the formula above is obtained by substituting n for t in the "fundamental solution of the heat equation"

$$u(t,x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{t}},$$

since this represents the concentration that would be measured if a unit mass (delta function) of the diffusing material were released at the origin at time $t = 0$.

What we are saying, then, is that we expect to see the square root because it is there in the fundamental solution of the heat equation. Once again we are depending on the relationship between our discrete problem and a continuous analog, only now we are concerned with the heat equation rather than the Laplace equation, its steady- state companion. So we meet again our

First Main Question: How can we relate the discrete problem to its continuous analog?

Needless to say, we now have some new candidates for methods of answering this question, namely whatever methods are used to establish the fundamental limit theorem. One such method uses Fourier analysis (see Lamperti [11]); applied to Polya's theorem, this yields a very nice proof, along with extensions to sums of independent identically-distributed random variables having distributions far more general than our Bernoulli variables X_i (See Chung and Fuchs [1]). Unlike Rayleigh's method, however, the Fourier method cannot be applied to problems where there is not a good deal of symmetry present. Each method has its own advantages and disadvantages, and each gives its own perspective on Polya's theorem. I will not discuss the Fourier approach here, as it would take us too far out of our way.

Moving to higher dimensions, the fundamental solution of the heat equation in $\mathbf{R}^d$ is

$$u(t, x_1, \ldots, x_d) = \frac{1}{(\sqrt{2\pi t})^d} e^{-\frac{(x_1^2+\ldots+x_d^2)}{t}},$$

the "dth power" of the one-dimensional solution. This suggests that

$$\begin{aligned} \mathbf{Z}^d \text{ is recurrent} \quad &\iff \quad \sum \frac{1}{n^{d/2}} < \infty \\ &\iff \quad d \leq 2. \end{aligned}$$

Could we make this transition from dimension 1 to dimension d in the discrete case, without making use of the continuous analog? This would give us a proof of Polya's theorem relying only on the fact that

$$P_n(0) = \mathrm{Order}(\frac{1}{\sqrt{n}}),$$

which we had no trouble deducing using only Stirling's formula.

That the fundamental solution to the heat equation in d dimensions is the d'th power of the one-dimensional solution reflects the fact that a particle diffusing in $\mathbf{R}^d$ may be thought of as performing simultaneously d independent one-dimensional diffusions in each of the d coordinate directions. So we want to think about what happens if instead of walking on the lattice $\mathbf{Z}^d$ we have our walker carry out simultaneously d independent one-dimensional random walks. Such a walker performs a simple random walk on a new lattice having the same vertices as $\mathbf{Z}^d$, but where every vertex is connected to the points gotten from it by adding one of the $2d$ vectors

$$(\pm 1, \ldots, \pm 1).$$

We denote this new lattice by $(\mathbf{Z}^1)^d$.

Now $(\mathbf{Z}^1)^d$ has recurrent type if and only if $d \leq 2$, since

$$\begin{aligned} \mathbf{E}(\text{number of times at } 0) &= \sum \mathrm{Order}(\frac{1}{\sqrt{n}})^d \\ &= \infty \text{ if and only if } d \leq 2. \end{aligned}$$

This is great, but how can we get from here to Polya's theorem? For although it is just barely true that $\mathbf{Z}^2 = (\mathbf{Z}^1)^2$, the sad fact is that $\mathbf{Z}^d \neq (\mathbf{Z}^1)^d$ for $d > 2$. This brings us to our

Second Main Question: If two graphs are essentially the same, do they have the same type?

Enough of questions! In the next section we will show how Rayleigh's method allows us to deal with this Second Main Question. We will come back to our First Main Question in the section after that.

8 Graphs that look alike

In this section we will use Rayleigh's method to attack our

Second Main Question: If two graphs are essentially the same, do they have the same type?

Our approach will be by way of a number of Observations, each meant to extend the classes of graphs that we can relate. To start out, our main concern will be to show that $(\mathbf{Z}^1)^d$ and $\mathbf{Z}^d$ have the same type, since this is what inspired our question. After that, if we can come up with a reasonable definition of what it means for two graphs to look alike, and show that any two such graphs have the same type, so much the better.

The idea behind the Rayleigh's method approach to this question is easy to state. Define a transient flow to be a flow out to infinity having finite dissipation rate.

Idea: To show that two graphs have the same type, show how a transient flow on either one could be converted into a transient flow on the other.

This Idea relies on the following extension to infinite graphs of

Thomson's theorem: A graph is transient if and only if it has a transient flow.

For a proof, see Doyle and Snell [2].

Neither this Idea nor the version of Thomson's theorem on which it relies will be mentioned explicitly in the rest of the section. The reason for this is that wherever we might have been tempted to base an argument on Thomson's theorem we will always base it instead on the somewhat earthier shorting and cutting laws. Nevertheless this Idea will be there, lurking beneath the surface, silently influencing everything we do.

Observation 1: Let H be a subgraph of G. Then G is more likely to be transient than H, in the sense that if H is transient then so is G.

This Observation, really just a version of the cutting law, is old hat by now. Equally old hat is the associated idea of showing that a graph is recurrent by embedding it in a recurrent graph, or showing that it is transient by embedding a transient graph in it.

The latter was the approach taken in Part I to prove that $\mathbf{Z}^3$ is transient. If you look back at that proof you will see that it succeeded more or less by accident, or rather by excess of cleverness (not mine—see the Acknowledgments). We started out trying to embed a nice three-dimensional tree NT_3 in $\mathbf{Z}^3$, and ended up embedding a 2.5849...-dimensional one. A more pedestrian solution to this embedding problem can be found by way of

Observation 2: Start with a graph G and construct a new graph H by replacing each edge e of G by a series of edges of length $l_e \geq 1$. If there is an upper bound to the numbers l_e then G and H have the same type.

This Observation is a simple consequence of the monotonicity law. Combining it with Observation 1 shows that in trying to embed G in H we don't have to look for an exact copy of G, only a *boundedly-broken* one. This gives us a way of showing that $\mathbf{Z}^3$ is more likely to be transient than NT_3. The trick is to draw NT_3 on a scale much larger than that of $\mathbf{Z}^3$, and then muscle NT_3 around so that its vertices lie on top of vertices of the lattice and its edges run along paths of bounded length in the lattice without running into each other. (See the Proposition below for a similar argument.)

This seems like a promising start. At last we have NT_3 safely tucked away in $\mathbf{Z}^3$, giving us yet another proof of Polya's theorem. Moreover, the same method allows us to show that there is a boundedly-broken copy of $\mathbf{Z}^3$ sitting inside $(\mathbf{Z}^1)^3$, so that $(\mathbf{Z}^1)^3$ is more likely to be transient than $\mathbf{Z}^3$. However if we try to go the other way and look for a boundedly-broken copy of $(\mathbf{Z}^1)^3$ inside $\mathbf{Z}^3$ we run into difficulties, related to the fact that there isn't any. The problem is that no matter where we try to put a vertex of $(\mathbf{Z}^1)^3$ in $\mathbf{Z}^3$, there will only be 6 edges leading out of the target vertex and we need 8.

This is vexing. For one thing, we have proved the wrong half of the equivalence between $\mathbf{Z}^3$ and $(\mathbf{Z}^1)^3$, since after all $(\mathbf{Z}^1)^3$ was the easier to prove transient. For another, this business about the valences seems like a pretty stupid reason for our method to fail. It seems like we ought to be able to fuzz the graph out a little without affecting its type.

Observation 3: Define the *k-fuzz* $F_k(G)$ of a graph G as the graph gotten from G by throwing in edges between all pairs of points whose distance in G lies between 2 and k (inclusive). Assume that G has bounded valence, i.e. that there is an upper bound for the valences of the vertices of G. Then for any k, G and $F_k(G)$ have the same type.

Remark. This does not hold for graphs of unbounded valence—see Doyle and Snell [2].

Proof of Observation 3. Of course $F_k(G)$ is at least as likely to be transient as G, by the cutting law. To show the other direction, we are going to short around with $F_k(G)$ until we come up with something that looks like G. To simplify the proof, let's assume that G has no multiple edges. This is OK because of

Observation 4: Removing redundant edges from a graph of bounded valence doesn't change its type.

The proof is by monotonicity, as in the proof of Observation 2.

To any edge e of $F_k(G)$ there corresponds a path in G of length $l_e \leq k$. Choose one such path and call it P_e. Replace every edge e of $F_k(G)$ by a series of l_e edges, and call the resulting graph $\overline{F_k(G)}$. $\overline{F_k(G)}$ is a boundedly-broken copy of $F_k(G)$, so by Observation 2 they have the same type.

Now for every edge e of $F_k(G)$ look at the corresponding series of resistors in $\overline{F_k(G)}$. Its endpoints lie in the original graph G. Take its intermediate vertices and short them to vertices of G along the path P_e. Do this for every e and call the resulting graph $\overline{G}$. By the shorting law, $\overline{G}$ is more likely to be transient than $F_k(G)$.

$\overline{G}$ differs from G only in that where there was a single edge of G there may now be a number of edges in parallel, one for every path P_e that traverses the edge in question. This number is no bigger than the total number of paths of length $\leq k$ in G that traverse that edge, which is bounded since G has bounded degree. Hence $\overline{G}$ has bounded degree. Applying Observation 4 shows that G and $\overline{G}$ have the same type, and we are done.

Confession: Honesty compels me to confess that the origin of the Observation 3 was probabilistic. If a random walk is recurrent, allowing the walker to take as many as k steps at a time isn't going to make the walk transient. Combining this observation with an electrical argument based on the monotonicity law yields a proof of the Observation, and this is exactly how the Observation was made—cf. Doyle and Snell [2]. This alternative proof is not necessarily simpler or more intuitive than the electrical proof just given, but it did come first.

Combining Observations 1 and 3, we can now show that $\mathbf{Z}^d$ and $(\mathbf{Z}^1)^d$ have the same type, for each is a subgraph of a k-fuzz of the other. More generally, any two d-dimensional lattices have the same type, where by a lattice I mean any connected graph that can be gotten by taking as vertices the vertices of $\mathbf{Z}^d$, and connecting every vertex to the vertices gotten from it by adding one of the $2k$ vectors

$$\pm v_1, \pm v_2, \ldots, \pm v_k$$

where $v_1, v_2, \ldots, v_k$ have integer coordinates.

We have accomplished out initial goal of showing that $\mathbf{Z}^d$ and $(\mathbf{Z}^1)^d$ have the same type. In the process we have developed a number of techniques for relating graphs that look alike, and with a little more work these techniques can be extended far enough to produce a satisfactory general answer to our Second Main Question. To see this, we begin with a proposition extending the argument that NT_3 can be embedded in $\mathbf{Z}^3$.

Proposition: Any graph G that can be drawn in a civilized manner in $\mathbf{R}^d$ can be embedded in some k-fuzz of any d-dimensional lattice L.

To make this true we make the following definition: A graph G (of bounded valence) can be *drawn in* $\mathbf{R}^d$ *in a civilized manner* if there is a function $f : V(G) \to \mathbf{R}^d$ defined on the vertices of G and $\epsilon > 0, M < \infty$ such that

$$g_1 \neq g_2 \;\Rightarrow\; \text{distance between } f(g_1) \text{ and } f(g_2) > \epsilon$$

and

$$g_1 \text{ adjacent to } g_2 \;\Rightarrow\; \text{distance between } f(g_1) \text{ and } f(g_2) < M.$$

In other words, no points are too close together and no edges are too long. Note that we don't care at all if the edges cross each other; we feel that we can draw many a non-planar graph in $\mathbf{R}^2$ in a civilized manner. On the other hand, whereas topologists may feel that they can embed the full binary tree in the plane, we consider it to be infinite-dimensional, since it can't be drawn in a civilized manner in any $\mathbf{R}^d$.

Proof of the Proposition. Draw the lattice L on a scale much smaller than the minimum distance between vertices in our civilized drawing of G, shift each point in the drawing so it coincides with a nearby point of L, and choose k big enough so that any two points that are adjacent in G are adjacent in $F_k(L)$—this can be done because the length of edges in the original drawing was bounded. G now appears as a subgraph of $F_k(L)$.

Note. Taking a careful look at this proof, we uncover the hidden assumption that G has no multiple edges. So we should define a *multiple* of a graph to be a graph gotten by replacing each edge of the original graph by a fixed number of edges in parallel, and go back and insert "some multiple of" before "some k-fuzz of" in the statement of the Proposition.

Corollary: Any two d-dimensional lattices have the same type.

Proof. Each can be drawn in $\mathbf{R}^d$ in a civilized manner.

This last result was proven earlier by arguing (implicitly) that when the vertices of both lattices are taken to coincide with the vertices of $\mathbf{Z}^d$, as they may be according to the definition of a lattice that I gave, then each appears as a subgraph of a fuzz of the other. Thus that proof depends on the definition of a lattice in an essential way. The virtue of this second proof is that it can easily be modified to show that $\mathbf{Z}^d$ has the same type as any graph that looks enough like it, so that e.g. $\mathbf{Z}^2$ has the same type as the infinite hexagonal not-quite-a-lattice illustrated in Figure 27.

But if we now have the techniques to prove that $\mathbf{Z}^d$ has the same type as any graph that looks enough like it, the proof that we would end up giving

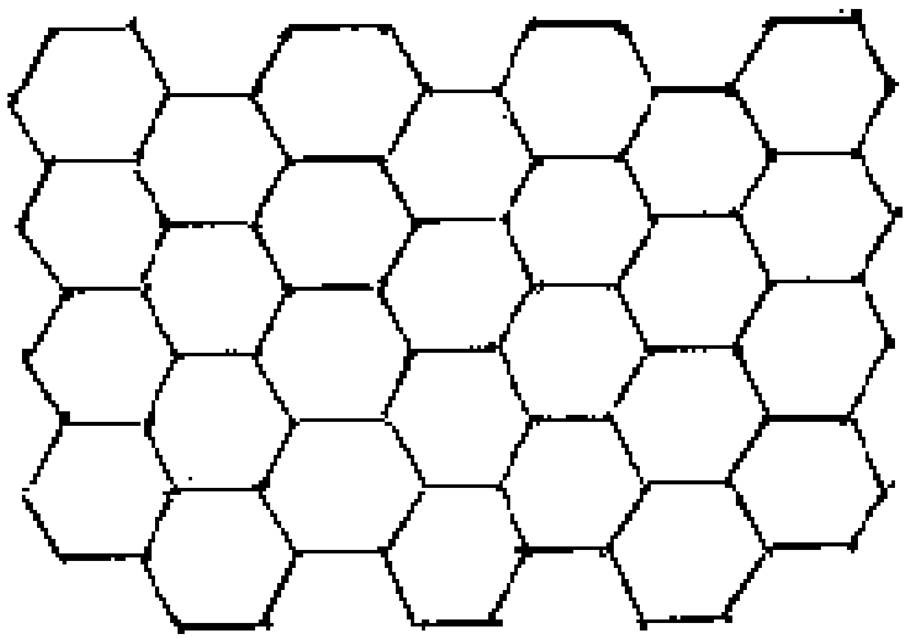

Figure 27: The hexagonal grid.

would not truly reflect the idea that the two graphs look alike. Tracing through the two applications of the Proposition that go into a proof of the Corollary, we see that what we are really saying there is that the two lattices have the same type, not because they look alike, but because they both look like $\mathbf{R}^d$! It would be nice to be able to give a more direct argument.

Observation 5: Let G be a graph of bounded valence. Let H be obtained from G by shorting. If for some N any two nodes that get shorted together in H have distance $\leq N$ in G then H has the same type as G. (In other words, if R is an equivalence relation on the vertices of G and if the equivalence classes of R have bounded diameter in G, then G and G/R have the same type.)

Proof. H is certainly more likely to be transient than G by the shorting law. To go the other way, our idea will be to try to exhibit H as a subgraph of a fuzz of G.

To each vertex h of H there corresponds a finite set S_h of vertices of G. Map the vertices of H to the vertices of G by sending every vertex h to some vertex $f(h)$ in S_h. To an edge from h_1 to h_2 in H there corresponds a path in G from $f(h_1)$ to $f(h_2)$ of length $\leq N + 1 + N = 2N + 1$.

If H had no multiple edges, we could conclude from this that H occurs as a subgraph of $F_{2N+1}(G)$. Putting it another way, if we delete the redundant edges of H and call the resulting graph $\overline{H}$, then f extends to an embedding of $\overline{H}$ in $F_{2N+1}(G)$. Hence G is more likely to be transient than $\overline{H}$. But it is easy to verify that $\overline{H}$ has bounded valence, so H and $\overline{H}$ have the same type, and we are done.

Taken together, our five Observations form the basis for the following definition:

Definition: We say that two graphs G and H of bounded valence *look alike* if there are two functions

$$f : V(G) \to V(H),$$

$$g : V(H) \to V(G)$$

defined on the vertices of G and H such that:

1. For both f and g, the distance between the images of any two adjacent vertices is bounded.

2. For both f and g, the distance between any two points that have the same image is bounded.

3. For both of the composed functions $g \circ f : V(G) \to V(G)$, $f \circ g : V(H) \to V(H)$, the distance between any point and its image is bounded.

Remark. Condition 2 is redundant. It is included for the sake of clarity.

This definition has been carefully phrased so as not to depend on the presence or absence of multiple edges in G and H; adding or removing redundant edges will not affect whether the two graphs look alike, as long as both graphs continue to have bounded valence.

Theorem: Two graphs that look alike have the same type.

Proof. By symmetry, it will be enough to show that H is more likely to be transient than G. Let $\overline{G}$ be the graph gotten by shorting together vertices that get sent to the same place by f, and then eliminating redundant edges. $\overline{G}$ has the same type as G by condition 2 and Observations 5 and 4. By condition 1, f induces an embedding of $\overline{G}$ in some fuzz of H, so H is more likely to be transient than $\overline{G}$. But G and $\overline{G}$ have the same type, so we are done.

This Theorem encompasses everything we have done in this section, and pretty much polishes off our Second Main Question. If one wanted to continue in this vein, the thing to do would be to consider other potential-theoretic questions in place of the type problem, and ask—for instance—whether graphs that look alike have the same Martin boundary.

9 Relating the discrete to the continuous

In this section I will show how we can use Rayleigh's method to approach our

First Main Question: How can we relate the discrete problem to its continuous analog?

Let me warn you in advance that in contrast to the last section, this one isn't going to have a happy ending. All we will do will be to answer this question in the context in which it was raised, i.e. in the case of relating $\mathbf{Z}^d$ to $\mathbf{R}^d$. What one really wants is a definition of what it means for two spaces to look alike—a definition strong enough to allow us to say that $\mathbf{Z}^d$ looks like $\mathbf{R}^d$—together with a theorem saying that two spaces that look alike have the same type. I am convinced that Rayleigh's method will eventually be found to provide a sound basis for such a theory, but time will have to tell.

If we look back at the proof of the recurrence of $\mathbf{Z}^2$ that was given in Part I, we see that in this case we have in effect already used the analogy between the discrete and continuous problems. That proof relied on the divergence of the sum

$$\sum \text{Order}(\frac{1}{n}),$$

whereas the fact that $\mathbf{R}^2$ has infinite resistance to infinity comes down to the divergence of the integral

$$\int_{\alpha}^{\infty} \frac{1}{r} dr.$$

It seems reasonable to regard this as an answer to the First Main Question in the case of $\mathbf{Z}^2$. Thus the case we want to consider is that of $\mathbf{Z}^d$ for $d > 2$. To simplify things, let's talk first about the case $d = 3$.

That the resistance of $\mathbf{R}^3$ out to infinity is finite means that there is some flow out to infinity having finite dissipation rate. This would follow from a suitable version of Thomson's theorem, but there is no need to be that fancy. We can exhibit such a flow without any difficulty: Just take the good old inverse-square radial flow field.

Can we somehow use this flow field to produce a transient flow on the lattice? As soon as we ask ourselves this question we recognize that it is the right question, and it won't be long before we can show that the answer is yes. The whole difficulty was to get to the point where we would think to ask it.

To produce a transient flow on the lattice from the inverse-square flow field in $\mathbf{R}^3$, proceed as follows. Surround each node of $\mathbf{Z}^3$ by a unit cube, so that together these cubes fill up all of $\mathbf{R}^3$. Every edge of $\mathbf{Z}^3$ is now bisected by the unit square that separates the two cubes surrounding the endpoints of that edge. Direct along each edge an amount of fluid equal to the flux across the bisecting square under the inverse-square flow field. This yields a flow on the lattice whose dissipation rate is readily shown to be finite, and we are done.

I believe that this argument can truly be said to show why the fact that $\mathbf{R}^3$ has finite resistance to infinity entails the transience of $\mathbf{Z}^3$, and of course the same argument works for any $d > 2$. Together with the shorting argument used to prove the recurrence of $\mathbf{R}^2$, this answers our First Main Question in the case of the lattices $\mathbf{Z}^d$.

Still, one could certainly ask for better, if not for the sake of understanding Polya's theorem then with an eye to extending the classes of look-alike spaces that we can relate. There are two related drawbacks to the approach we have taken. First, we have used two separate arguments in relating $\mathbf{Z}^d$ to $\mathbf{R}^d$. one showing recurrence for $d \leq 2$, and another showing transience for $d > 2$. Second, both arguments depend on specific knowledge of the spaces involved. We would prefer to be able to argue directly that if $\mathbf{Z}^d$ had finite resistance to infinity then $\mathbf{R}^d$ would have to also, and vice versa.

In what remains of this section, I am going to describe an approach to relating $\mathbf{Z}^d$ to $\mathbf{R}^d$ based on this idea. I have yet to extend this to a satisfactory general method for relating spaces that look alike, and even in the case at hand the argument must be considered conjectural, in the sense that I am not sure precisely what regularity conditions to demand of flows, etc., in order to make these arguments work out.

To begin with, $\mathbf{R}^d$ is more likely to be transient than $\mathbf{Z}^d$, for if there were a transient flow on $\mathbf{Z}^d$ we could transfer it to $\mathbf{R}^d$ by fattening up the wires, so to speak, and the resulting flow would still be transient.

On the other hand, if there were a transient flow in $\mathbf{R}^d$ we we could transfer it to the lattice just as we did the inverse-square flow in $\mathbf{R}^3$. Unfortunately, the resulting flow would not necessarily have finite dissipation. This has to do with what happens where the boxes into which we are dividing $\mathbf{R}^d$ come together. To see how this causes problems, consider that if you take the $1/(rlog\frac{1}{r})$ flow around the origin in $\mathbf{R}^2$ and restrict it to the disk $r \leq 1/2$, you get a flow having finite dissipation rate but infinite flux through any radius. (See Figure 28.) Indeed, for the dissipation we have

$$\begin{aligned}
\text{dissipation} &= 2\pi \int_0^{\frac{1}{2}} \frac{1}{r^2(\log\frac{1}{r})^2} r\,dr \\
&= 2\pi \int_0^{\frac{1}{2}} \frac{1}{r(\log r)^2} dr \\
&= 2\pi \left. \frac{-1}{\log r} \right|_0^{\frac{1}{2}} \\
&= \frac{2\pi}{\log 2},
\end{aligned}$$

while for the flux through a radius we have

$$\text{flux} = \int_0^{\frac{1}{2}} \frac{1}{r\log\frac{1}{r}} dr$$

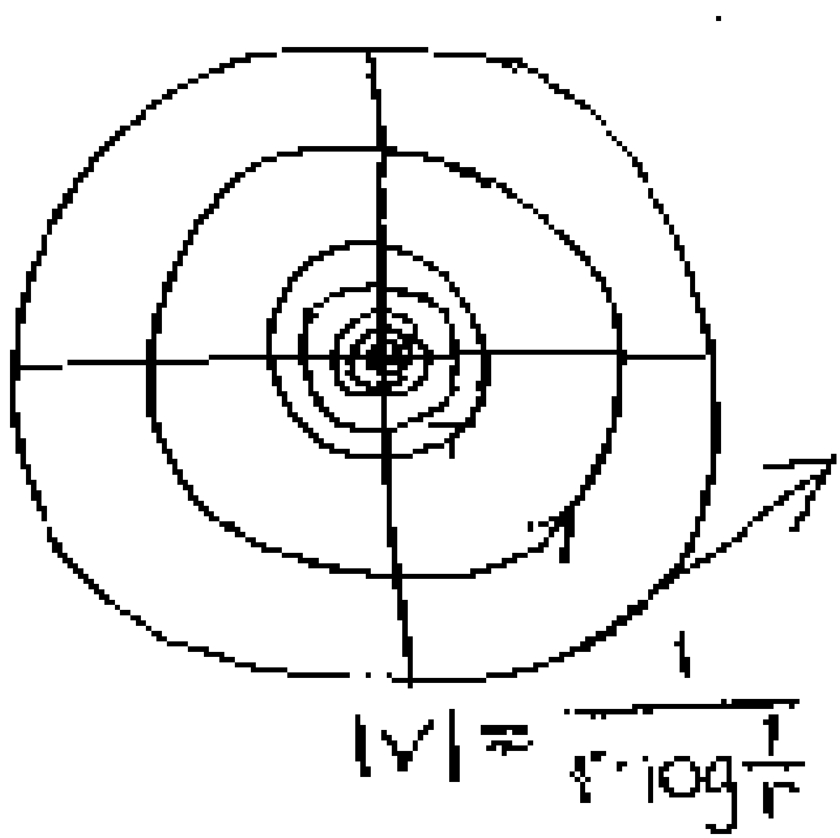

Figure 28: A finite energy flow with infinite flux.

$$
\begin{aligned}
&= \int_0^{\frac{1}{2}} \frac{-1}{r\log r} dr \\
&= -\log\log\frac{1}{r}\Big|_0^{\frac{1}{2}} \\
&= \infty.
\end{aligned}
$$

There are (at least) two ways around this difficulty. The first is to incorporate some hysteresis into our method of transferring the flow to the lattice, so that instead of declaring that a fluid particle has moved from one box to another as soon as it crosses the boundary between them we wait until it has gone some prescribed small distance out of the box it was last in. This yields a flow on the d-fuzz of $\mathbf{Z}^d$, Its dissipation rate is finite, for we can now find a lower bound for the dissipation required to pass from one box for another, and use this to show that if the lattice flow had infinite dissipation then the original flow must have had infinite dissipation also.

The second way around this difficulty involves modifying the flow (before transferring it to the lattice) by eliminating all flow along streamlines that do not eventually go off to infinity. We then transfer the flow to the lattice as we did the inverse-square flow in $\mathbf{R}^3$. This way we get a flow on $\mathbf{Z}^3$, not on a fuzz of it, but to show that it is transient involves an argument closely resembling the proof that a graph and any fuzz of it have the same type, so we are really no better off using this second approach.

The idea behind the approach just described is an obvious generalization of the Idea we used to approach our Second Main Question:

Idea: To show that two spaces have the same type, show how a transient flow on either one could be converted into a transient flow on the other.

To develop this Idea into the general theory needed to give this section the happy ending it deserves, what we will need are techniques for relating flows on related spaces. It is in this spirit that I offer the foregoing method of relating $\mathbf{Z}^d$ to $\mathbf{R}^d$.

10 Why 2?

By now you may feel that it is high time to stop this business of trying to understand Polya's theorem, to "declare we won and B home." Before we do, however, let us have a last look at the question of what is so special about 2. Back when we raised this question we decided that $d = 2$ is distinguished as the point where the integral

$$\int_{\alpha}^{\infty} \frac{1}{r^{d-1}} dr.$$

changes from divergent to convergent, but this merely replaced one silly question with another: Why should $d = 2$ be the spot where this integral decides to converge? For some time we have had the tools to answer this question, and it seems silly to leave it up in the air.

The real reason, then, that the integral

$$\int_{\alpha}^{\infty} \frac{1}{r^{d-1}} dr.$$

starts to converge at $d = 2$ and not somewhere else, say $d = 7$ or $d = 14.3$, is that this is the value of d for which random walk changes from recurrent to transient! No, this does not put us back where we started, for we have another description of this critical value, namely the place where the integral

$$\int^{\infty} \frac{1}{\sqrt{r}^{d}} dr.$$

changes from divergent to convergent. Comparing this with the original integral shows that the critical value of d is that for which

$$d - 1 = \frac{d}{2},$$

i.e.

$$d = 2.$$

Needless to say, it would be a mistake to attribute too much significance to the foregoing argument. My real reason for harking back to the "Why 2?" question is quite different. If one thinks for a while about the significance of dimension 2 is Polya's theorem, one may make the following conjectures:

Conjecture 1: To any sufficiently regular graph, say one whose automorphism group has only a finite number of orbits, one can assign a dimension d, either a positive integer or ∞, such that the number of polnts in a ball of radius r grows like r^d (if $d < \infty$) or like e^{ar} for some a (if $d = \infty$).

This is a version of a conjecture of Milnor [14].

Conjecture 2: Such a graph is recurrent if $d \leq 2$ and transient if $d > 2$.

This is a version of a conjecture of Kesten (cf. [7], p. vi).

Large chunks of both conjectures have now been proven—cf. Gromov [6] and Guivarc'h et al. [7]. I am not sufficiently *au courant* [Ed.: that's French for 'smart'] to tell what their exact status is. In any case, the challenge remains to give a simple proof of Conjecture 2 using Rayleigh's method (or, Heaven forbid, some other method). It appears to me that to answer this question is an intuitively satisfying way would do a lot to remove whatever nagging doubts we may have as to why Polya's theorem is true.

Appendix

In this appendix I present an application of Rayleigh's method to the classical type problem for Riemann surfaces. My aim in doing so is to call attention to this domain of application of Rayleigh's method, for while this method is already known to complex analysts (cf. Royden [21]), I am convinced that its true importance for an understanding of the type problem has not yet been appreciated. At some point I mean to carry on at some length in order to justify this statement. For the moment, however, I will let Rayleigh's method speak for itself.

The problem to which we will apply Rayleigh's method was raised by Milnor [15].. Consider a complete Riemannian 2-manifold having a global geodesic polar coordinate system (r, θ) about some point p. In this coordinate system the metric takes the form

$$dr^2 + g(r,\theta)^2 d\theta^2$$

and the Gaussian curvature takes the form

$$\kappa = -\frac{\frac{\partial^2 g}{\partial r^2}(r,\theta)}{g(r,\theta)}.$$

Milnor wanted to relate the conformal type of the surface to the Gaussian curvature κ. Specifically, he proposed the following two conjectures:

> **P**: If $\kappa \geq -1/(r^2 \log r)$ for large r then the surface is conformally *parabolic*, i.e. can be mapped conformally onto the complex plane.
>
> **H**: If $\kappa \leq -(1+\epsilon)/(r^2 \log r)$ for large r, and if $G(r) = \int_0^{2\pi} g(r,\theta) d\theta$ is unbounded, then the surface is conformally *hyperbolic*, i.e. can be mapped conformally onto the unit disk.

Milnor established these criteria for the case of a surface symmetric about p (that is, for a surface where $g(r,\theta)$ depends only on r) by constructing explicitly a conformal mapping of the surface onto a disk or the whole complex plane. He then asked for a proof in the more general case. Evidently Robert Osserman pointed out to him that by applying a method of Ahlfors we can conclude that

$$\int_\alpha^\infty \frac{dr}{\int_0^{2\pi} g(r,\theta) d\theta} = \infty \;\Rightarrow\; \text{conformally parabolic.}$$

The criterion **P** is a simple consequence of this parabolicity criterion. Hence what was missing was a proof of the criterion **H**.

Now to someone familiar with Rayleigh's method, the expression appearing in Ahlfors's criterion has a particularly concrete significance. If we think of the Riemann surface as being made of an isotropic resistive material of "constant thickness," then the type problem can be interpreted as the problem of determining whether the resistance of the surface out to infinity is infinite (parabolic type) or finite (hyperbolic type). The expression

$$\int_{\alpha}^{\infty} \frac{dr}{\int_0^{2\pi} g(r,\theta)d\theta}$$

represents the resistance out to infinity of the electrical system obtained from the Riemann surface by "shorting the points of the surface together along the circles r = constant." As the resistance of this new system is $\leq$ the resistance of the old system, we conclude that

$$\int_{\alpha}^{\infty} \frac{dr}{\int_0^{2\pi} g(r,\theta)d\theta} = \infty \ \Rightarrow \ \text{conformally parabolic.}$$

To someone familiar with Rayleigh's method, the obvious next step is to get an upper bound for the resistance of the surface by making some cuts in it. In the case we are looking at, the most natural thing to do is to cut along the rays θ = constant. The conductance (i.e. inverse resistance) of the resulting system out to infinity is

$$\int_0^{2\pi} \frac{d\theta}{\displaystyle\int_{\alpha}^{\infty} \frac{dr}{g(r,\theta)}}.$$

This is $\leq$ the conductance of the original Riemann surface out to infinity, so we conclude that

$$\int_0^{2\pi} \frac{d\theta}{\displaystyle\int_{\alpha}^{\infty} \frac{dr}{g(r,\theta)}} > 0 \ \Rightarrow \ \text{conformally hyperbolic.}$$

This hyperbolicity criterion obviously resembles the parabolicity criterion just obtained, and just as **P** followed from the earlier criterion, so **H** follows from this one.

Acknowledgements

The work described here owes most of its inspiration to Laurie Snell, who has always advocated the electrical approach to random walk problems. In addition, he supplied much of the energy needed to turn what began as vague notions into concrete results. My debt of gratitude to him is larger than I can say. I dedicate this thesis to him.

The rest of the inspiration for this work comes from Bill Doyle, my father. It was he who first suggested to me the idea of applying Rayleigh's method to probabilistic questions, and while I have jealously tried to deal him out of the ensuing developments, it was really all his idea. That is, he said it first. I had the idea at the same time, you understand, but just didn't get the words out in time...

Further help came from David Griffeath, who conceived on his own the notion of proving Polya's theorem in dimension three by a cutting argument, and suggested the particularly elementary approach taken in Part I. Thanks also to Alfred Huber for his help with the application of Rayleigh's method described in the Appendix, and to the many others who through lectures, correspondence, and conversations have helped me in this endeavor. Thanks especially to my advisor Reese Prosser, whose friendly advice and intimidation never failed to guide me back on course when I seemed to be drifting.

Thanks to our magnificent librarians Monique Cleland and Rebecca Lear; to Marie Slack, who maneuvered Part I through the computer; and to Dartmouth's Kiewit Computation Center, whose Avatar terminal and text processing software made the preparation of Part II remarkably easy.

Thanks, finally, to Mom, Dad, Jeff, and Mart, who were always there with the money, and the animal crackers.

References

[1] K. L. Chung and W. H. J. Fuchs. On the distribution of values of sums of random variables. *Mem. Amer. Math. Soc.*, 6, 1951.

[2] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks*. Mathematical Association of America, Washington, D. C., 1984.

[3] R. J. Duffin. Electrical network models. In D. R. Fulkerson, editor, *Studies in Graph Theory*, pages 94–138. Mathematical Association of America, 1975.

[4] R. P. Feynmann. *The Feynmann Lectures on Physics*. Addison-Wesley, 1964.

[5] D. Griffeath and T. M. Liggett. Critical phenomena for Spitzer's reversible nearest particle systems. *Ann. Probab.*, 10:881–895, 1982.

[6] M. Gromov. Groups of polynomial growth and expanding maps. *Inst. Hautes Études Sci. Publ. Math.*, 53:53–78, 1981.

[7] Y. Guivarc'h, M. Keane, and B. Roynette. *Marches aléatoires sur les groupes de Lie (Lecture Notes in Math 624)*. Springer, 1977.

[8] W. V. D. Hodge. *The Theory and Applications of Harmonic Integrals (2nd edition*. Cambridge, 1951.

[9] J. Jeans. *The MAthematical Theory of Electricity and Magnetism, 5th edition*. Cambridge, 1966.

[10] H. Kesten and ? Grimmett. Silver and gold? *?*, ?:?, ?

[11] J. Lamperti. *Probability: A Survey of the Mathematical Theory*. Benjamin, 1966.

[12] T. J. Lyons. A simple criterion for transience of a reversible Markov chain. *Ann. Probab.*, 11:393–402, 1983.

[13] J. C. Maxwell. *Treatise on Electricity and Magnetism*. Clarendon, Oxford, third edition, 1891.

[14] J. Milnor. Problem 5603. *Amer. Math. Monthly*, 75:685–686, 1968.

[15] J. Milnor. On deciding whether a surface is parabolic or hyperbolic. *Amer. Math. Monthly*, 84:43–46, 1977.

[16] G. J. Minty. On the axiomatic foundations of the theories of directed linear graphs, electrical networks and network-programming. *J. Math. Mech.*, 15:485–520, 1966.

[17] C. St. J. A. Nash-Williams. Random walk and electric currents in networks. *Proc. Camb. Phil. Soc.*, 55:181–194, 1959.

[18] L. Onsager. Reciprocal relations in irreversible processes I. *Phys. Rev.*, 37:405–426, 1931.

[19] G. Polya. Über eine Aufgabe betreffend die Irrfahrt im Strassennetz. *Math. Ann.*, 84:149–160, 1921.

[20] J. W. S. Rayleigh. On the theory of resonance. In *Collected scientific papers*, volume 1, pages 33–75. Cambridge Univ. Press, Cambridge, 1899.

[21] H. L. Royden. Harmonic functions on open Riemann surfaces. *Trans. Amer. Math. Soc.*, 73:40–94, 1952.

[22] H. Whitney. On the abstract properties of linear dependence. *Amer. J. Math.*, 57:509–533, 1935.