GLOBAL ORDER FROM LOCAL SOURCES*

by
Charles Radin

Mathematics Department
University of Texas
Austin, TX 78712
USA
email: radin@math.utexas.edu

TO APPEAR IN THE RESEARCH-EXPOSITORY PAPERS SECTION

OF THE BULLETIN OF THE AMS, OCTOBER 1991

Subject classification: 47A35, 52A45, 82A55

*Research supported in part by NSF Grant No. DMS-9001475

1. INTRODUCTION

This article contains introductions to three open problems of significant research interest, taken from number theory, logic, and condensed matter physics. All three problems will be shown to have at their core special cases of one simply-stated optimization problem. Our goal is to use the intuition gained from these three perspectives to direct attention to this common core, which constitutes, in fact, one problem of remarkable depth and importance. We will also show that some of the tools developed in the separate problems are of real value in the others.

Since each of the three problems uses jargon peculiar to its field, we will give an informal introduction to each, together with all relevant definitions, in the following section. However it may be useful to include here a very brief description of each of them to give some idea of our eventual goal.

Our first problem is "sphere packing", in which we consider arrangements of infinitely many unit diameter spheres, each sphere having a variable position in $\mathbf{R}^3$, and try to determine those "optimal" arrangements in which the spheres are disjoint and yet occupy the largest possible fraction of space.

The problem from logic was originally concerned with the "decidability of AEA formulas". It then expanded to the area now known loosely as "tiling theory", in which one analyzes the tilings of the plane which are possible from a given set of tiles. Neighboring tiles must satisfy color (or matching) rules, and we try to optimize agreement of these rules.

¿From physics we consider the "crystal problem". Here the concern is to understand why real matter seems, experimentally, to have a strong tendency to have crystalline order at low temperature and high pressure. The problem is modeled using the equilibrium statistical mechanics of interacting particles, and a concise way to summarize that theory is that the state of the particles must be a minimum of the free energy.

A common theme for all three problems is that they involve a system of many objects (spheres, tiles, particles) each of whose position in space has some influence on the others nearby, and we seek those arrangements in space of the objects which optimize something (density, agreement of matching rules, free energy). One remarkable feature which generates interest in these problems is the conviction, based on a variety of evidence, that **in optimal arrangements the objects tend to be very regularly positioned in space – although the mechanism causing this tendency is completely unknown.** This will be formalized in section 3, leading to natural measures of "regularity" or "order".

In the above snapshots of the problems we have used the jargon of the relevant fields. In the next section we will flesh out the snapshots with the appropriate definitions needed by a nonspecialist.

2. THREE OPEN PROBLEMS

2a. Sphere Packing

In sphere packing we try to find that arrangement in space of unit diameter spheres or balls which are nonoverlapping and yet cover the highest possible fraction of space. (This is part of problem 18 in Hilbert's celebrated list of open problems [17]. A solution has recently been announced by Wu-Yi Hsiang, though no manuscript is generally available yet.)

First let us be more precise about the meaning of "cover the highest possible fraction of space". Take the balls to be open – that is, translates of $\{(x_1, x_2, x_3) \in \mathbf{R^3} : x_1^2+x_2^2+x_3^2 < 1/4\}$ – and, given an arrangement A of infinitely many disjoint balls, consider an increasing sequence $S(n)$, of spheres of radius n, all centered at the origin in $\mathbf{R^3}$. Let $d(A)$ (the "density of A") be the fraction of the volume of $S(n)$ which is contained in those balls of A which are inside $S(n)$, asymptotically in n, namely

$$d(A) = \liminf_{n\to\infty} \frac{\text{volume of balls of } A \text{ in } S(n)}{\text{volume of } S(n)}.$$

What is then desired is the supremum of $d(A)$ over all possible arrangements A, and those "optimal" arrangements A' such that $d(A') = \sup_A d(A)$. (It is not hard to see that such optimal arrangements exist: the supremum is really a maximum.)

The sphere packing problem is old (going back at least to Gauss [13]) and unsolved, though for many years it has been generally accepted [34] that one optimal arrangement (among others) has the centers of the balls at the points of a "face-centered cubic lattice", which, using a Euclidean basis of $\mathbf{R}^3$, can be given by: $\{a_1(0, s, s)+a_2(s, 0, s)+a_3(s, s, 0) : s = \sqrt{(1/2)},\ a_j \in \mathbf{Z}\}$. (This collection of points is called a "lattice" because it is of the form $\{a_1v_1+a_2v_2+\cdots+a_nv_n : a_j \in \mathbf{Z}\}$ for some basis $\{v_1, v_2, \cdots, v_n\}$ of $\mathbf{R}^n$; see Fig. 1a. Since we will frequently refer to the centers of balls, from now on we will use the word "configuration" for the set of points which are the centers of the balls in some "arrangement". Also, it is not hard to show that there are other ways to arrange spheres in $\mathbf{R}^3$ to obtain the same density as that of the "optimal" face-centered cubic lattice. This degeneracy is something we will need to discuss further in section 2d.) The simpler problem concerning disjoint unit diameter disks in $\mathbf{R}^2$ has a complicated history going back to Thue in 1910 [39], and has as an optimal configuration the "hexagonal lattice": $\{a_1(1,0) + a_2(\sqrt{(1/2)}, \sqrt{(3/2)})\ :\ a_j \in \mathbf{Z}\}$; see Fig. 1b.

Mathematical interest in sphere packing was originally due to a connection with minimization of quadratic forms. Specifically, there is a direct connection between the lattice $\{a_1v_1 + a_2v_2 + \cdots + a_nv_n : a_j \in \mathbf{Z}\}$ with basis $\{v_1, \cdots, v_n\}$, and the quadratic form

$$f(a_1, \cdots, a_n) = \sum_{j,k=1}^{n} a_j a_k\ v_j{\cdot}v_k$$

in the integer variables $a_1, \cdots, a_n$, where $v_j{\cdot}v_k$ refers to the inner product in $\mathbf{R}^n$. Forms are called "integrally equivalent" if they correspond to the same lattice under a change of basis. The density $d(A)$ of the corresponding lattices A is therefore an invariant for this equivalence, and the forms corresponding to the "densest lattice packing" have a natural place in this theory. (The difference between a "densest packing" and a "densest lattice packing" is that in the latter we only consider lattice configurations A when minimizing the density $d(A)$.) Sphere packing, especially when restricted to lattice packings, has also played a role in coding theory and several other allies of number theory. For an encyclopedic survey of the uses of lattice packings see [8].

As a separate subject the sphere packing problem owes its depth to the highly ordered or regular structure of the known optimal configurations; it is appropriate to restrict consideration to lattice configurations in number theory, but not here. To clarify the meaning of "optimal" then, it is convenient to reformulate the sphere packing problem in the following terms. We have at our disposal many variables (the centers of the balls in $\mathbf{R}^n$), summarized in the one quantity we call a configuration. For arbitrarily large bounded regions R in $\mathbf{R}^n$

we are trying to optimize certain functions F_R over all possible configurations A, where $F_R(A)$ is defined to be: $+\infty$ if there is a point of A which is both in R and also separated by a distance less than 1 from some other point of A, and otherwise $F_R(A)$ is defined to be the negative of the number of points of A which are in R. (F_R basically counts the contribution to the density of the spheres of A in R, subject to a "penalty" if two spheres overlap.) In these terms then, the interesting fact is that in order **to minimize the function F_R for large R we are lead to configurations A which are highly ordered (lattices)***.

2b. AEA Formulas and Tiling

The relevant analysis of our logic problem was initiated by Hao Wang, and originally concerned the "decidability" of the class of all "AEA formulas" [43,44]. AEA formulas are sentences that begin "For every x there exists a y such that for every $z\ldots$", followed by a logical combination of predicates not containing the quantifiers "for every" or "there exists". (For example: For every subset x there exists a least upper bound y such that for every upper bound z it follows that $z \geq y$.) And a class of formulas is said to be decidable if (informally) there is an algorithm by which, given any member of the class, we can determine in finitely many steps if the formula is consistent, i.e. not self contradictory. Wang found a way to analyze this decision problem for the class of AEA formulas through a game he invented called "tiling".

In the game of tiling one uses (an unlimited number of) translates, called "tiles", of each of a finite set of "prototiles". A tile or prototile is a unit square in the plane, with each edge having some specified color, and with edges parallel to a fixed set of axes. For each color there is defined a "complementary" color. The object of the game is to cover the plane with tiles in checkerboard fashion (that is, abutting tiles touch full edge to full edge), with the condition – called the color (or matching) rules – that abutting edges must have complementary

---

* We repeat that the conjectured optimal configuration, the face centered cubic lattice, is not unique; there are other configurations, which are not lattices or even periodic, with the same density as the face centered cubic lattice [29,8]. This degeneracy is not central to our discussion however [29]. The role of degeneracy in our problem will be discussed below.

colors; see Figs. 2a,b. We will use the word "configuration" for a covering of the plane by tiles in checkerboard fashion, and reserve the term "tiling" for a configuration in which abutting tiles also satisfy the color rules; see Fig. 2c. (Given a set of prototiles there may or may not be any tilings of course – see Fig. 2d, but there are always configurations. Also, note that in placing tiles in the plane one may translate but not rotate or reflect them from the prototile.) One of the remarkable features of this tiling game is its versatility, and much of this is due to its relation with Turing machines, to which we now give an informal introduction.

A Turing machine is an imaginary computer consisting of a "tape", a "head" and a "program". The tape is one dimensional, oriented as to right and left, and consists of "cells" one of which is said to be "under" the head at any time. There are infinitely many cells to the left and right of this cell. Each cell is marked by one symbol from the finite set $\{s_0, s_1, \ldots, s_p\}$, and cells marked with symbol s_0 are called "blank". The program is a permanent, finite, labeled list $\{q_0, q_1, \ldots, q_r\}$ of "instructions" for the head, of which one is "current" at any given time. The instruction in operation at a given time is that of the "current" instruction of the program, and may depend in its effect on the mark of the cell under the head. Each single operation of the machine takes one second for completion and is of two types; either it replaces the contents of the cell under the head with a given mark, then moves the tape one cell to the left or right, and installs a specified instruction as the next "current" instruction, or else it causes the machine to Halt! (permanently). We can represent the nonhalting operations by quintuples of the form (q_i, s_j, s_k, M, q_t), which means that if the current instruction is q_i and the cell under the head is marked s_j then the mark in that cell is made to be s_k, the tape is moved M (a variable taking the values "left" or "right") and the next current instruction becomes q_t. Any pair (q_i, s_j) not the beginning of such a quintuple is assumed to make the machine halt. At the initial time the instruction is assumed to be q_0 and all but finitely many of the cells of the tape are blank. The "Halting problem" asks whether or not there is a Turing machine – i.e. an algorithm – which can determine precisely which Turing machines will eventually halt if the tape is blank initially. It was proven by Turing that there can be no such algorithm; the Halting problem is "undecidable" [40,20,35].

Now given any Turing machine one can construct a set of prototiles which mirrors the action of the machine by a special one of its tilings. This is done as follows. (This construction is due to Wang, with important generalizations by Berger [5], Robinson [33] and others.) We represent the state of the machine (that is, the current instruction, the markings on the tape cells, and which cell is under the head) at two consecutive seconds by colors on the top and bottom edges of an infinite row of tiles, one tile for each cell, the top representing the later time. Each of these top and bottom edges has a color representing the marking of the corresponding cell of the machine, and precisely one top edge and one bottom edge has an altered color denoting that it corresponds to the cell under the head, and also denoting the current instruction. Horizontal edges are also colored. We will denote the colors of the edges by a set of some or all of the following: a head or tail of an arrow perpendicular to the edge (an arrow head by definiton has its point touching the edge, as in the top of Fig. 3a), a cell marking and an instruction. The complement of a color is the set with the same marking and instruction, but opposite end of the arrow. There are several classes of prototiles. Those in Fig. 3a are for cells whose markings s_j do not change during the machine operation under consideration. Those in Figs. 3b,c represent a cell with marking s_j which is being put under the head, from the left or right, and will be subject to instruction q_i. And those in Figs. 3d,e represent cells with marking s_j which are being re-marked s_k under instruction q_i, moved left or right, and with next current instruction q_s. We only have prototiles with markings which correspond to quintuples of nonhalting operations. To represent the initial state of the machine which for convenience we assume includes a blank tape, we need the three prototiles of Figs. 3f-h (we will call the middle of these the "center prototile"), and the last prototile we require is that in Fig. 3i, which we will call the "blank" tile. It follows that in any (partial) tiling of the plane containing a center tile, the tiles making up the horizontal row containing that tile must be those of Figs. 3f,h. Each successive row above this one is then completely determined by the operation of the Turing machine being "mirrored", for as many rows as there are nonhalting operations of the machine. And the rows below the initial one must consist of blank tiles.

Therefore given any Turing machine we have specified a set of prototiles which can tile the plane **using the center prototile** if and only if the Turing machine never halts. However, the tiling game does not contain the condition that a given prototile (such as the center prototile) must be used, and without the condition we could tile the plane just using blank tiles. To eliminate this feature, R. Berger [5] developed a method, refined by Robinson [33] and others, by which given a Turing machine one extends the above construction to a more complicated set of prototiles which can tile the plane if and only if the Turing machine never halts; there is no longer a condition that some specified prototile must be used. The technique consists of mirroring at many points within **every possible** tiling of the plane arbitrarily large centered squares of the above construction, the only obstruction to tiling the plane being due to halting of the machine. (Thus it is possible, by adroit choice of the Turing machine being mirrored, to build sets of prototiles which can tile the plane but only with tilings having special features. This will be discussed further below.)

We will expand on this connection between Turing machines and the tiling game in various contexts, but first we want to clarify the original connection. By the above construction and the known solution of the Halting problem, it was proven that the game of tiling is undecidable; that is, there can be no algorithm by which, given a set of prototiles, one can determine in a finite number of steps if the corresponding tiles can tile the plane. On the other hand, given any finite set of prototiles which can tile the plane, Wang showed how to construct an AEA formula which specifies the positions of the tiles in the plane. It follows that if the class of AEA formulas were decidable, the tiling game would be; since the tiling game is not decidable, neither is the class of all AEA formulas. This is a rough outline of the role that the tiling game played in the original decidability problem for AEA formulas. At this point we will cease discussing the predicate calculus aspect of this circle of ideas, and concentrate on the tiling game and Turing machines.

Wang proved [43] that the tiling game is decidable if the following is true: that every set of prototiles which can tile the plane can also tile the plane in a "periodic" manner. (By "periodic" we mean that there is some "unit cell" consisting of a bounded collection of specified tiles in the plane, and the rest of the tiling consists of translations of this collection, in fact by a set of translations which constitute a lattice; see Fig. 2b.) We note that although it is easy to construct examples of sets of prototiles, as in Fig. 2a, which can tile the plane periodically, it is very hard to construct a set which can tile the plane but **only nonperiodically**. In fact the next big step was taken by Berger, who used the above ideas to construct such a set of prototiles, which can tile the plane, but only nonperiodically, in his proof that the tiling game is undecidable [5].

Although we will refer again to the use of Turing machines in constructing sets of prototiles whose tilings all have interesting properties, at this point we want to emphasize the simplest case, the existence of sets of prototiles which can tile the plane but only nonperiodically. See Fig. 4. This is reminiscent of our discussion of the sphere packing problem. To make the connection more transparent we assume given a set of prototiles and construct, for each bounded region R of the plane, a function F_R on the set of configurations A of the tiles. Define $F_R(A)$ to be the number of those pairs of abutting tiles in A that do not satisfy the color rules, and are also such that at least one of the pair is contained in the region R of the plane. In this notation we are interested in configurations A in the plane which minimize F_R for large R, and in fact for which $F_R(A) = 0$ for all R. Berger's example is then a solution of the same type of optimization problem as the sphere packing problem, but somewhat less ordered since the optimal A's are not lattices.

Although the motivating question of decidability has been answered, tiling is in fact a growing area of research. Now interest is basically due to the attempt to understand the contrast between the ease with which examples of prototiles can be given (such as those of Fig. 2a) which can tile the plane periodically, and the great difficulty of producing examples (such as those of Fig. 4) which can tile the plane but only nonperiodically. As we will discuss in section 3d, methods are being developed to produce prototiles which tile the plane but only with tilings of greater and greater "disorder". (There is also interest in the geometrical aspects of tiling, "five fold symmetry" and all that, spurred in large part by the examples due to Penrose [25,15] of tilings with nonsquare "tiles"; see Fig. 5. The connection between our problem of the "orderliness" of tilings and this geometrical subject – the study of the **symmetries** of tilings by nonsquare prototiles, is by no means clear, although the role of such symmetries in the study of quasicrystals has tended to mix the two [38].) Now we have used the words "order" and "disorder" so far in a colloquial sense, to refer for example to the intuitive feeling that the (periodic) configurations of the sphere packing solutions (Fig. 1b) are more orderly or regular than the (nonperiodic) tilings of some sets of prototiles (Fig. 4). However it will be one of our main goals to introduce and justify specific technical measures of order for problems of the sort we are discussing. Some of these measures will come from considerations in our next open problem, the crystal problem from condensed matter physics.

2c. Crystals

In the crystal problem we seek the fundamental mechanisms underlying the tendency of real bulk matter towards crystalline order at low temperature and high pressure. By crystalline order we mean that the atomic nuclei in the material, which can be roughly localized as fixed points in $\mathbf{R}^3$, form a point set which is either a lattice, or more generally, periodic (that is, there is a "unit cell" associated with each point of a lattice; see Fig. 6).

The crystal problem is old, and considered a major unsolved problem [7,41,42, 37,29,2]. There is a mathematical formalism (equilibrium statistical mechanics) in which to set up the problem, and we begin with a review of this formalism.

Assume we are modeling a material made of nuclei and electrons, and to be definite assume there are precisely $N-1$ different types of nuclei, N some fixed integer ($N=3$ for table salt, which contains the nuclei of sodium and chlorine). In statistical mechanics our model of the material contains several parameters, such as temperature, pressure, etc., appropriate to its macroscopic state – there are alternative choices for these parameters, and for convenience we will use the temperature T and the chemical potentials $M_1, \ldots, M_N$ of each of the constituent particle types – electrons and $N-1$ types of nuclei. (The chemical potential of a particle type measures the energy it would take to increase by one the number of that type of particle in the material.) At this point we need to decide on the mechanics we will use to describe the local state of each particle. For our purposes (understanding the origin of the crystalline state) it is convenient to use classical mechanics instead of the slightly more accurate quantum mechanics.

The state of the j^{th} particle then consists of the triple (x_j, s_j, p_j), where $x_j \in \mathbf{R}^3$ is its position, $s_j \in \{1, \ldots, N\}$ is its particle type and $p_j \in \mathbf{R}^3$ is its momentum. So associated with each particle there are several real variables. In **statistical** mechanics these are **random** variables, and their degree of dependence on one another is governed to a large extent (specified below) by the forces between the particles; that is, the formalism does not yield specific values for the variables, but the (joint) **probability** that particles will have specific positions, particle types and momenta. (The temperature and chemical potentials are parameters on which these probabilities will depend.) The forces between the particles enter the formalism as potential energies (from which the forces could be recovered by differentiation).*

Before specifying the potential energy that is appropriate (so as to determine the probabilities mentioned above), we need to make one further simplification wherein the local state of a particle has discrete values, and in particular the position varies through the cubic lattice $\mathbf{Z}^3$ or more generally $\mathbf{Z}^n$ instead of $\mathbf{R}^3$. So instead of variables such as the position x_j of the j^{th} particle, we will use variables A_j where $j \in \mathbf{Z}^n$ is a variable point or "site" of the lattice; A_j then represents whether or not there is a particle at j, and its local state (particle type etc.) if there is one. We assume A_j runs through some finite set $\{1, \ldots, m\}$, and associate a chemical potential M_α to each possible local state $\alpha \in \{1, \ldots, m\}$. By a "configuration" we will mean an element of $\{1, \ldots, m\}^{\mathbf{Z}^n}$.

We are now prepared to describe the type of potential energy that is appropriate for our model. We will use a "short range, two-body, translation invariant interaction potential", that is a real function $V(\cdot\,,\cdot)$ on $\big(\{1, \ldots, m\} \times \mathbf{Z}^n\big)^2$ such

* If we were using quantum mechanics we could use the static Coulomb potential energy of charged nuclei and electrons; some remarkable work has been performed in this way by Fefferman [10,11], Lieb [18,19] *et al*, but not for condensed matter, as we require. Since we chose to use classical mechanics we are now forced to use phenomenological forces to describe, not interacting nuclei and electrons but interacting ions and electrons, or even interacting atoms. What this means is that we still have N types of "particles" in our model, but now we need to chose some appropriate potential energy function.

that

$$\begin{aligned} &V(A_i, A_j)\exp(|i-j|) \to 0 \text{ as } |i-j| \to \infty \\ &V(A_{i+k}, A_{j+k}) = V(A_i, A_j), \text{ for all } i, j, k \in \mathbf{Z}^n. \end{aligned} \qquad 1)$$

Then given any bounded region R of $\mathbf{Z}^n$ (in the middle of some very large region $\hat{R}$), we know from statistical mechanics that the relative probability that the sites j of R are in states A_j is:

$$\sum_{j\in\hat{R}/R}\sum_{A_j=1}^{m}\exp\Big(-[E_{\hat{R}}(A) - M_1 n_1^{\hat{R}}(A) - \cdots - M_m n_m^{\hat{R}}(A)]/T\Big) \qquad 2)$$

where

$$E_{\hat{R}}(A) = \sum_{\substack{i,j\in Z^n;\ i\neq j \\ i \ and/or\ j\ \in\hat{R}}} V(A_i, A_j) \qquad 3)$$

and $n_\alpha^{\hat{R}}(A)$ is the number of the sites of $\hat{R}$ in state α*.

These probabilities simplify greatly in the limit where $T \downarrow 0$, but to describe this we need some further notation. The set $\{1,\ldots,m\}^{\mathbf{Z}^n}$ of all configurations A has, as a product, a conventional topology and Borel measure structure. So by taking $\hat{R}$ to be a variable cube centered at the origin and expanding to $\mathbf{Z}^n$, we can use 2) to define in the limit a probability measure $\mu^V_{T,M_1,\ldots,M_m}$ on $\{1,\ldots,m\}^{\mathbf{Z}^n}$ by its values on "cylinder sets"

$$C(A_{j_1},\ldots,A_{j_p}) \equiv \Big\{A' \in \{1,\ldots,m\}^{\mathbf{Z}^n} : {A'}_{j_k} = A_{j_k}, k = 1,\ldots,p\Big\}.$$

That is**,

$$\mu^V_{T,M_1,\ldots,M_m}[C(A_{j_1},\ldots,A_{j_p})] = \lim_{\hat{R}\to\mathbf{Z}^n}\frac{N(\hat{R})}{Z(\hat{R})} \qquad 4)$$

* This relative probability depends negligibly on sites outside $\hat{R}$, in a sense discussed below.

** In general the following limiting procedure depends on a choice of A_j for j outside $\hat{R}$, through $E_{\hat{R}}(A)$. However since we are trying to model a pure solid phase we are only interested in those values of $T, M_1, \ldots, M_m$ for which there is one and only one translation invariant measure that can be obtained by such a limit – which is therefore independent of such choices of A_j for j outside $\hat{R}$ – and we denote this measure by $\mu^V_{T,M_1,\ldots,M_m}$.

where

$$N(\hat{R}) = \sum_{\substack{j\in\hat{R} \\ j\neq j_1,\ldots,j\neq j_p}} \sum_{A_j=1}^{m} \exp -\Big([E_{\hat{R}}(A) - M_1 n_1^{\hat{R}}(A) - \cdots - M_m n_m^{\hat{R}}(A)]/T\Big)$$

and

$$Z(\hat{R}) = \sum_{j\in\hat{R}} \sum_{A_j=1}^{m} \exp -\Big([E_{\hat{R}}(A) - M_1 n_1^{\hat{R}}(A) - \cdots - M_m n_m^{\hat{R}}(A)]/T\Big).$$

(By "translation invariant" we mean that $\mu^V_{T,M_1,\ldots,M_m}[C(A_{j_1},\ldots,A_{j_n})]$ is unchanged as the sites $j_1,\ldots,j_n$ of $C(A_{j_1},\ldots,A_{j_n})$ are subjected to any fixed translation of $\mathbf{Z}^n$.) It is instructive to note that for the "free model" (otherwise known as a "Bernoulli shift" in probability theory), in which the interaction potential V is identically 0, the measure $\mu^V_{T,M_1,\ldots,M_m}$ is simply the product $\otimes_{j\in Z^n}\mu_j$ of copies of the local measure μ on $\{1,\ldots,m\}$ given by

$$\mu(\beta) = \frac{\exp(M_\beta/T)}{\sum_{\alpha=1}^{m}\exp(M_\alpha/T)}.$$

Thus the interaction is the source of the dependence among the random variables A_j. We now define the probability measure $\mu^V_{0,M_1,\ldots,M_m}$, called the "ground state for V", by the further limit $\mu^V_{T,M_1,\ldots,M_m} \to \mu^V_{0,M_1,\ldots,M_m}$ as $T \downarrow 0$, again by using the values of the measures on cylinder sets*. As a limit of $\mu^V_{T,M_1,\ldots,M_m}$ we know that $\mu^V_{0,M_1,\ldots,M_m}$ is also translation invariant. (We are really only interested in $\mu^V_{T,M_1,\ldots,M_m}$ for $T = 0$, but we introduced the distribution for $T > 0$ to emphasize the fact that the ground state $\mu^V_{0,M_1,\ldots,M_m}$ is a probability distribution on configurations, not just a configuration, a fact that is crucial for our problem.)

Eventually we will determine $\mu^V_{0,M_1,\ldots,M_m}$ for some interaction potentials V, but first we want to record one very useful fact about $\mu^V_{0,M_1,\ldots,M_m}$: the set of "ground state configurations for V" has measure 1 with respect to this measure, where a configuration A is a ground state configuration for V if, for every bounded region R in $\mathbf{Z}^n$,

$$F_R(A) = \inf\Big\{F_R(A') : A' \in \{1,\ldots,m\}^{\mathbf{Z}^n},\ \text{and } {A'}_j = A_j \text{ for all } j \in R\Big\} \quad 5)$$

* Here again one can question the existence of the limit, but again on physical grounds this limit should exist if the states $\mu^V_{T,M_1,\ldots,M_m}$ stay within a pure phase as $T \downarrow 0$. This is proven to hold for generic interactions in [28].

where the function F_R is defined by:

$$F_R(A) = E_R(A) - M_1 n_1^R(A) - \cdots - M_m n_m^R(A). \qquad 6)$$

It has required a lot of definitions, but finally we are at a useful point to contemplate what we have. The ground state $\mu^V_{0,M_1,\ldots,M_m}$ for the interaction V (and fixed chemical potential parameters M_α) is a translation invariant probability measure concentrated on the set of ground state configurations, and the ground state configurations are the solutions A of the set of optimization conditions 5) for all bounded regions R of $\mathbf{Z}^n$. The basic objective of the crystal problem consists in understanding why solutions A to this optimization problem tend to be periodic, or highly ordered – that is, crystals.

2d. Comparisons

We complete this section with a demonstration that the optimization formulations we gave for tilings and for sphere packing are closely related to the one for the crystal problem, 5).

Returning then to the game of tiling, we consider all possible finite sets of prototiles which can tile the plane. We are interested in certain features of the tilings of the plane which are possible for specific sets; for example, whether or not only nonperiodic tilings are possible.

Consider some set of m prototiles which can tile the plane, as in Fig. 2a. Construct a coordinate system in the plane containing the tilings, with the integer points $\mathbf{Z}^2$ at the centers of the tiles. Now define a discrete statistical mechanical model on $\mathbf{Z}^2$, with m states per site (and for configuration A, $A_j = \alpha$ means that the state at site j "is" a translate of the α^{th} prototile – see Fig. 7), with interaction potential V defined by:

$$\begin{aligned} V(A_i, A_j) &= 0, \text{ if } |i-j| \neq 1 \\ &= -1, \text{ if } |i-j| = 1 \text{ and tiles } A_i, A_j \text{ at } i,j \text{ satisfy the color rules} \qquad 7) \\ &= 0, \text{ if } |i-j| = 1 \text{ and tiles } A_i, A_j \text{ at } i,j \text{ don't satisfy the color rules.} \end{aligned}$$

(With the specific prototiles of Fig. 2a this statistical mechanical model is called the Ising antiferromagnet [36,14].) Now consider the set of ground state configurations for this statistical mechanical model, assuming for simplicity that all chemical potentials have value 0. It is easy to see that any configuration corresponding to a tiling of the plane will be a ground state configuration. The converse however is not quite true. For example, with respect to the model using the prototiles of Fig. 2a consider the configuration of tiles with a "fault line" as in Fig. 2c. It is easy to check that this is not a tiling but is a ground state configuration. There is however a sense in which this configuration is negligible, and that brings us back not just to ground state configurations, but to the ground state $\mu^V_{0,0,\ldots,0}$ itself.

At this point we need some further background material. First we note that for any collection of prototiles the corresponding set of tilings of the plane is a translation invariant, closed – and therefore compact – subset of the space of all configurations. Similarly, for any discrete statistical mechanical model the set of all ground state configurations is a translation invariant, closed – and therefore compact – subset of the space of all configurations. We define the "support" of a measure on a compact space to be the complement of the union of all open sets of measure zero, and, finally, we say a system consisting of a compact set, a group of homeomorphisms of the set, and a probability measure on the set invariant under the homeomorphisms, is "uniquely ergodic" if there is only one probability measure on the set invariant under the homeomorphisms. This implies the better known property of ergodicity of the system, which only requires that there be no other invariant probability measure **absolutely continuous** with respect to the given one. (A circle carrying Lebesgue measure is uniquely ergodic under iterates of an irrational rotation. Bernoulli shifts – noted above in the example of free models, and discussed further below – are ergodic but not uniquely ergodic; we can see the latter from the existence of those invariant probability measures which give full measure to a single constant configuration.) The best known result concerning unique ergodicity is its relation with pointwise ergodic theorems, as we discuss in connection with 11) below.

Now without loss of generality we can assume for the interactions we consider that the set of ground state configurations is uniquely ergodic under translations. (This is usually easy to check for any given interaction of interest, and it has been proven that this is true for "generic" interactions in some reasonable spaces of interactions. The proof [28], which is a variant of the classical theorem that a convex function is differentiable almost everywhere, shows that the condition of unique ergodicity is just a condition of nondegeneracy, a requirement that the optimization problem have a unique solution among invariant probability measures. This condition fails for sphere packing in three dimensions if the conjectured solution is correct, and this degeneracy means that the sphere packing problem is to that extent atypical, and consequently of less value as a guide to our intuition about such optimization problems. Because of the proof that unique ergodicity is generic, and other physical reasons, it is appropriate to require of any model of the ground state of a pure solid phase – as opposed to a state of coexistence of a fluid and solid phase say – that the condition hold that the set of ground state configurations must be uniquely ergodic; see [31,29,28].) For the interaction corresponding to Fig. 2a this is easy to check. In fact, if we denote by A and A' the (only) two tiling configurations allowed by the tiles of Fig. 2a, the one in Fig. 2b and its translate by one unit, then the ground state of the associated interaction must be $\mu^V_{0,0,\dots,0} = (\delta_A + \delta_{A'})/2$, the average of the point masses concentrated on the points A and A'. There are other ground state configurations for that interaction, such as that of Fig. 2c, but it is easy to show that the set of ground state configurations is uniquely ergodic, with this measure. So for this interaction (and others built from sets of prototiles) there is a one-to-one correspondence between the associated tilings of the plane and the "relevant" ground state configurations, the configurations in the support of the ground state $\mu^V_{0,0,\dots,0}$ [21].

We now see that the tiling problem can be considered a special case of the crystal problem in which: a) all chemical potentials are taken to be zero, and b) the interaction has the special form of equations 7).

To see sphere packing as a special case of the crystal problem requires more terminology since it relates to a continuum statistical mechanical model, not a discrete one as we have used. Physically, all that is needed is a force that prevents two particles from getting too close – what is called a "hard core repulsion" in the physics literature – together with either a chemical potential favoring a positive density or else a pressure, which would have the same effect. An added complication is that in fact unique ergodicity does not hold for sphere packing. For these reasons we take the cowards' way out and refer the reader to [29] for the messy details; nothing new is involved other than the technical definitions needed to deal with continuously distributed variables. We will only discuss the discrete version of our general optimization problem from here on.

To summarize what we have discussed so far, we have considered three different problems: sphere packing, tiling, and the crystal problem, and have seen that at heart they are all of the form of the optimization problem 5), where the functions F_R to be optimized are of the form given by 6) and 3). This type of optimization problem is interesting because there seems to be a strong tendency for solutions A to either be periodic (even a lattice), or at least highly ordered, though the reason for this is unknown.

3. ORDER AND THE ERGODIC THEORY OF SYMBOLIC DYNAMICS

3a. The general problem and solution in dimension 1

It is now time to analyze the concept of "order". To do that we will make use of the mathematical structure developed in section 2, the ergodic theory of symbolic dynamics.

In the ergodic theory of symbolic dynamics we have: some closed subset X of a product space $\{1,\dots,m\}^{\mathbf{Z}^n}$, invariant under the group $\mathbf{Z}^n$ of translations (sometimes called "shifts"), and a Borel probability measure μ on X which is invariant under the translations.

For convenience we review here the meaning of the symbol $X^V_{M_1,\dots,M_m}$. Consider the set of all $A \in \{1,\dots,m\}^{\mathbf{Z}^n}$ such that for every bounded region R in $\mathbf{Z}^n$,

$$F_R(A) = \inf\Big\{F_R(A') : A' \in \{1,\dots,m\}^{\mathbf{Z}^n},\ \text{and } {A'}_j = A_j \text{ for all } j \in R\Big\} \qquad 5)$$

where

$$F_R(A) = E_R(A) - M_1 n_1^R(A) - \cdots - M_m n_m^R(A), \qquad 6)$$

$$E_R(A) = \sum_{\substack{i,j\in Z^n;\ i\neq j\\ i \text{ and/or } j\in R}} V(A_i,A_j), \qquad 3)$$

the "chemical potentials" $M_\alpha \in \mathbf{R}$, $n_\alpha^R(A)$ is the number of the sites of R in state α, and the "interaction potential"

$$V : (A_i,A_j) \in \Big(\{1,\dots,m\} \times \mathbf{Z}^n\Big)^2 \to V(A_i,A_j) \in \mathbf{R}$$

satisfies

$$
\begin{aligned}
&V(A_i, A_j)\exp(|i-j|) \to 0 \text{ as } |i-j| \to \infty \\
&V(A_{i+k}, A_{j+k}) = V(A_i, A_j), \text{ for all } k \in \mathbf{Z}^n.
\end{aligned} \tag{1)}
$$

We assume that on the set of such A there is only one invariant probability measure (the "ground state") denoted $\mu^V_{0,M_1,\ldots,M_m}$, and define $X^V_{M_1,\ldots,M_m}$ as the support of $\mu^V_{0,M_1,\ldots,M_m}$.

So taking for X the set $X^V_{M_1,\ldots,M_m}$ of solutions A of our optimization problem 5), and for μ the ground state $\mu^V_{0,M_1,\ldots,M_m}$, we find that the structure of our problem refinements we have made are: a) the assumption that X be uniquely ergodic under translations, that is, that μ should be the **only** invariant probability measure on X, and b) that the support of μ should be X. In dynamical systems there is a more convenient way to express condition b). A closed subset Y of $\{1,\ldots,m\}^{\mathbf{Z}^n}$, invariant under translations, is called "minimal" if it contains no nontrivial closed invariant subset. It is easy to see that if a) holds then condition b) is just the restriction that X be minimal. Finally, the two conditions that X be uniquely ergodic and minimal are summarized in the one condition, that X be "strictly ergodic".

So in this new language our problem becomes:

The General Optimization Problem. Consider those strictly ergodic symbolic dynamical systems $X^V_{M_1,\ldots,M_m} \subset \{1,\ldots,m\}^{\mathbf{Z}^n}$ which arise as the solution sets of the above optimization problems. The goal is to "measure" the order in such systems, and determine the degree of order which is generic and the causes of the various degrees of order.

Now in most studies of strictly ergodic symbolic dynamical systems the dimension n is 1, which is unfortunate because that is the one dimension that is fairly trivial for us, at least if the interaction potential V is "of finite range", that is:

$$V(A_i, A_j) = 0 \quad \text{if } |i-j| > D \tag{8)}$$

for some $D \geq 1$. In that case we have:

Theorem 1 (C.R. and L. Schulman [27]). In dimension 1, if the interaction V is of finite range and $X^V_{M_1,\ldots,M_m}$ is strictly ergodic then $X^V_{M_1,\ldots,M_m}$ must be a finite set, namely the translates of some periodic configuration.

This is a positive solution of our problem for dimension 1: every 1 dimensional problem has a periodic ground state. (We will see that the problem is much more interesting in higher dimensions.) In particular, consider those "tiling dynamical systems" which define strictly ergodic $X^V_{0,\ldots,0}$ using V of the form 7). Since these V are of range 1, they (and the game of tiling itself) are relatively trivial in dimension 1. But as we will see this does not make irrelevant the known results of uniquely ergodic symbolic dynamics for dimension 1.

To get back to the question of "order", we are now ready to introduce three central ideas: entropy, recursive functions and spectrum. We will see that these measure different aspects of the "orderliness" of a dynamical system. In practice such a multifaceted approach seems to be the safest way to try to make precise the intuitive notion of order, though of course this still leaves open the use of other possible measures of order besides these three. We begin with entropy.

3b. Entropy

Using the above framework of a symbolic dynamical system $X \subseteq \{1,\ldots,m\}^{\mathbf{Z}^n}$ with invariant measure μ, the (topological) entropy $h(X)$ is given by:

$$h(X) = \lim_{N\to\infty} \frac{1}{(2N+1)^n} \sum_{C\in C(N)} \eta[\mu(C)], \tag{9}$$

where

$$\eta[t] = -t\log(t), \text{ if t>0}$$
$$= 0, \text{ if t} = 0$$

and $\mathbf{C}(N)$ is the collection of all cylinder sets based in the cube $\{(j_1,\ldots,j_n) \in \mathbf{Z}^n : |j_k| \le N\}$. (Since X is strictly ergodic, it is known that this topological entropy coincides with "measure theoretic entropy" [23]. It is also relevant that it coincides with the entropy for ground states that is used in physics [1].)

It is instructive to compute $h(X)$ for the following two examples. If X is the finite set of translations of some periodic configuration of period p – see Fig. 2b, then $\mu(C)$ is only nonzero for about p^n cylinder sets $C \in \mathbf{C}(N)$, and since this bound is uniform in N, $h(X) = 0$. The other example, called a "Bernoulli shift on $\mathbf{Z}^n$", is defined by taking $\mu = \otimes_{j\in Z^n}\mu_j$ on $X = \{1,\ldots,m\}^{\mathbf{Z}^n}$ with $\mu_j = (1/m)\sum_{\alpha=1}^m \delta_\alpha$. It is easy to check that $h(X) = \log(m)$. (Note the connection between Bernoulli shifts and the free models of section 2c.)

Now entropy has been widely used as a measure of disorder, for example in information theory [6] and physics [36]. In physics, the Third Law of Thermodynamics says that the ground state of any macroscopic material has zero entropy, and this is usually "understood" theoretically to be an expression of the small number of configurations contributing to the distribution 2) at low temperature: basically, translations and small perturbations of some periodic configuration. Of course this **assumes** that the ground state is a perfect crystal, corresponding to a periodic configuration. The closest there is to a proof of this law is the recent:

Theorem 2 (J. Miękisz and C.R. [31]) If the interaction potential V is of finite range (and $X^V_{M_1,\dots,M_m}$ is strictly ergodic), then the entropy $h(X^V_{M_1,\dots,M_m}) = 0$.

(There is a stronger version of this result that holds for interactions of longer range [30].) This proves that, at least for finite range interactions, our optimal configurations must be orderly as measured by entropy. The next idea we will use to measure the regularity or complexity of solutions of our optimization problem is that of "recursive function".

3c. Recursive functions

A function in $\mathbf{Z}^{\mathbf{Z}^n}$ is said to be recursive if there is a Turing machine program which can, with variable tapes corresponding to the points in the domain of the function, produce output (when the machine halts) interpretable as the corresponding values of the function [35]. We can use this notion to catagorize the tilings of the plane for a given set of prototiles (the tilings being functions in $\{1,\dots,m\}^{\mathbf{Z}^2} \subset \mathbf{Z}^{\mathbf{Z}^2}$), or the ground state configurations of a given statistical mechanical model. It is clear that a periodic configuration is recursive, as it is easy to give an algorithm to specify uniquely, step by step, each local state in such a configuration. What is not so obvious is that there exists a set of prototiles (as was proven by Myers [22] using the technique noted in section 2b), which can tile the plane, but such that **all** such tilings of the plane are **nonrecursive**; **none** of the tilings can be uniquely specified by an algorithm. (We note that this is an existence result – no such set of prototiles has been exhibited by Myers or anyone else.) In this sense these tilings are "complex". It is debatable whether or not we should call such tilings irregular or disordered, but this is perhaps more a question of semantics than of substance. It seems reasonable to say that such a set of prototiles, or corresponding interaction potential, is relevant in our search for the structure of solutions to our optimization problem. (It is possible to measure the degree by which a function, in particular a tiling, fails to be recursive. See [22].) The next measure of order that we will use is the spectrum.

3d. Spectrum

Given a symbolic dynamical system $X \subseteq \{1, \ldots, m\}^{\mathbf{Z}^n}$, with invariant probability measure μ, the spectrum we are interested in is associated with the unitary operators $T(j)$ representing translations by $j \in \mathbf{Z}^n$ on the complex Hilbert space $L^2(X, \mu)$. (So $[T(j)f](x) = f(y)$ where $y = x - j$.) We refer to the projection valued measure $dE(\lambda)$ on $[0, 1)^n$ such that for any vector $f \in L^2(X, \mu)$,

$$< f, T(j)f > = \int_{[0,1)^n} \exp(i\, j \cdot \lambda) d < f, E(\lambda) f >, \quad j \in \mathbf{Z}^n, \qquad 10)$$

where $<,>$ refers to the inner product in $L^2(X, \mu)$. The vector f belongs to the pure point, singular continuous or absolutely continuous subspace of $L^2(X, \mu)$ if the real valued measure $< f, E(\lambda)f >$ is pure point, singular continuous or absolutely continuous on $[0, 1)^n$. Finally, since the subspace of $L^2(X, \mu)$ consisting of the constant functions is invariant under translations, we can and will restrict attention to f in its orthogonal complement in $L^2(X, \mu)$. See [32].

There is an intuitive feeling that the more smooth these measures $< f, E(\lambda)f >$ are, the more disordered the configurations in X are. For example, in the "most ordered" case where X is the finite set of translations of some periodic configuration, the measure $< f, E(\lambda)f >$ consists of a finite number of point masses for any f. At the other extreme, for the Bernoulli shift on $\mathbf{Z}^n$ discussed above, the typical element of X with respect to μ is intuitively as disordered as possible, and any measure $< f, E(\lambda)f >$ is absolutely continuous.

Our objective is to analyze the degree of order which occurs in strictly ergodic symbolic dynamical systems of the form $X^V_{M_1,\ldots,M_m}$. We know from Theorem 2 that, for interaction potentials V of finite range, disorder is small in the sense that the entropy is zero. (And this has been generalized somewhat to longer range interactions [30].) On the other hand we know from the tiling examples of Myers [22] noted above that there exist finite range interactions for which $X^V_{M_1,\ldots,M_m}$ is disordered in the sense of being nonrecursive. We now want to consider the possible degree of disorder using the notion of spectrum, but we make a short digression first.

Fixing the number m of states per site, and a bound D on the range of the interaction (as defined by 8)), the space of interactions is finite dimensional. It is therefore natural to ask whether some specific sort of order is "generic" (either of full Lebesgue measure, or contains the countable union of dense open sets). Note that this approach would not be available if we restricted attention to tilings, or statistical mechanical models which come from tilings through 7), since the corresponding space of interactions would then contain only finitely many points, and there could be no useful notion of genericity. This is one reason why the generalization from tilings to statistical mecahnics can be useful: it effectively changes the mathematics from algebra to analysis. Unfortunately we do not know whether generic models are recursive.

Before we continue with this analysis we need to emphasize another aspect of the strictly ergodic systems we are studying. As we alluded above, an n-dimensional ergodic symbolic dynamical system X with invariant measure μ is uniquely ergodic if and only if

$$1/(2N+1)^n \sum_{|j|\leq N} [T(j)f](x) \to \int_X f \; d\mu(x), \tag{11}$$

as $N \to \infty$, **uniformly in** $x \in X$, for every continuous function f on X [24]. Intuitively, this means that all the points $x \in X$ are in some sense equivalent, and studying the disorder of X is the same as studying the disorder of any one configuration x of X. Note the essential difference between this situation and that of the Bernoulli shift mentioned above. For the latter at best we could only hope to get one type of "order" for **generic** configurations of X (in either the measure theoretic or topological sense) but certainly not for **all** configurations. Also, the **uniform** convergence in 11) contrasts sharply with the **pointwise** almost-everywhere convergence of Birkhoff's ergodic theorem, which applies to ergodic dynamical systems (such as the Bernoulli shift) which are not strictly ergodic. For these reasons, as well as others, the restriction we have made to strictly ergodic dynamical systems means that we are trying to determine the degree of order of a fairly well defined "pattern", (in one interpretation the ground state configuration of a model of a solid), as opposed to the degree of order of a measure (which, again in this interpretation, could be that of a model of a fluid – to which one does not usually attribute any one "pattern").

We will now describe the types of spectrum which are known to occur for our dynamical systems $X^V_{M_1,\ldots,M_m}$. Of course it is easy to obtain isolated pure point spectrum, corresponding to the highest possible order, in any perfectly periodic $X^V_{M_1,\ldots,M_m}$ as in Fig. 2b. The spectrum corresponding to the nonperiodic tiling models such as in Fig. 4 has not been worked out in detail, though it is again expected to be pure point, but now dense in $[0, 2\pi)^2$ (for appropriate f), as seems to be the case of real quasicrystals [38]. Recently an important new method for constructing prototiles with spectral disorder has been developed by Mozes [21], and we will discuss this after some preliminary background.

First we outline the notion of (1 dimensional) "substitution dynamical systems". Such a system $X \subseteq \{1, \ldots, m\}^{\mathbf{Z}}$ is defined as the set of all points $x = \big\{x_j : x_j \in \{1, \ldots, m\},\ j \in \mathbf{Z}\big\}$ for which all "blocks" $\{x_j : J \le j \le K\}$ in x satisfy certain restrictions. These restrictions are defined through "substitution rules" by which, for each element $\alpha \in \{1, \ldots, m\}$, we associate a finite sequence $\{\alpha_1, \cdots, \alpha_k\}$ where $\alpha_j \in \{1, \ldots, m\}$ and k (the so-called "length" of the rule) may depend on α. As an example with $m = 2$, the "Morse", or "Thue-Morse" substitution rules:

$$0 \ \to\ 01, \quad 1 \to 10 \qquad 12)$$

are both of length 2. Now given any finite sequence B of elements of $\{1, \ldots, m\}$, we define $D(B)$ as the finite sequence obtained by replacing each of the elements of B using its rule. (So 01 becomes 0110 using 12).) Next define $W_0 = \{1, \ldots, m\}$, $W_{n+1} = \bigcup_{B \in W_n} D(B)$ and $W = \bigcup_{n \ge 0} W_n$. Then X is the set of all sequences x for which every subblock of x is a subblock of an element of W. (Once we have defined X we have a dynamical system, with discrete "time" represented by translation in $\{1, \ldots, m\}^{\mathbf{Z}}$. And since under very mild conditions substitution dynamical systems are uniquely ergodic [26], they automatically come equipped with a canonical measure.) Finally, given a set of substitution rules defining the set X, we say the dynamical system has "unique derivation" if for every $x \in X$ there is a unique $y \in X$ (unique only up to translation) such that x is obtained from y when the substitution rules are used to replace the elements of y. For example, the dynamical system $X \subset \{0,1\}^{\mathbf{Z}}$ defined [9] by

$$0 \;\to\; 001, \quad 1 \;\to\; 11100 \tag{13)}$$

has unique derivation, as does the better known Thue-Morse system, defined by 12). An example of a dynamical system without unique derivation is $X \subset \{0,1\}^{\mathbf{Z}}$ with the rules $0 \to 010,\ 1 \to 101$. See [21,26]. One other definition we need is that of the 2 dimensional symbolic dynamical system "associated with" a pair X_1, X_2 of 1 dimensional symbolic dynamical systems, with translations $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$. The associated 2 dimensional dynamical system consists of the cartesian product $X_1 \times X_2$ with translations

$$\begin{aligned} T_1 \times I &: (x_1, x_2) \in X_1 \times X_2 \to (T_1 x_1, x_2) \in X_1 \times X_2 \\ I \times T_2 &: (x_1, x_2) \in X_1 \times X_2 \to (x_1, T_2 x_2) \in X_1 \times X_2. \end{aligned} \tag{14)}$$

(For completeness we note that substitution dynamical systems can be generalized to higher dimensions [21].) It is perhaps noteworthy that Thue introduced the system 12) in one of a series of works devoted to combinatorial modeling of the "antithesis" of periodicity. Specifically, the sequences defined by 12) are precisely those sequences of 0's and 1's which have the following "BBb property": for no block $B = \{b_1, b_2, \ldots, b_m\}$ of 0's and 1's does the block

$$BBb \equiv \{b_1, b_2, \ldots, b_m, b_1, b_2, \ldots, b_m, b_1\}$$

appear in the sequence. (See [16] for a discussion of Thue's work in this direction, and [12] for a version of the BBb characterization tailored to our problem.)

Now that we have the notion of a substitution dynamical system, we can discuss the following recent result of Mozes, which relates substitutional dynamical systems with the 2 dimensional "tiling" dynamical systems considered above (where X is the subset of $\{1, \ldots, m\}^{\mathbf{Z}^2}$ consisting of all the tilings associated with some set of m tiles).

Theorem 3 (S. Mozes [21]). Given any pair of 1 dimensional substitutional dynamical systems with unique derivation and with substitution rules of length at least 2, one can exhibit a (2 dimensional) tiling dynamical system which is measure theoretically isomorphic to the 2 dimensional dynamical system associated with the pair of 1 dimensional substitution dynamical systems.

Although we will not reproduce Mozes' proof, it is worth noting that it depends heavily on the techniques, discussed above, invented for constructing a set of prototiles which mirrors the action of a Turing machine in all its tilings. And just as Myers, who used these methods to prove the existence of tiling dynamical systems which are disordered in the sense that all the tilings are nonrecursive, Mozes uses the methods to exhibit tiling dynamical systems which are disordered in the sense of their spectrum. (This result is constructive; the prototiles can easily be exhibited.) An interesting aplication in this direction [31] uses the 1 dimensional system defined by 13), and generates a tiling dynamical system for which the measures $< f, E(\lambda)f >$ on $[0, 2\pi)^2$ are singular continuous. Unfortunately this method cannot yield measures $< f, E(\lambda)f >$ which are absolutely continuous, because this already fails for the 1 dimensional substitutional dynamical systems used as components [9]. So it is an open question how smooth the measures can be for tiling dynamical systems, or more generally for $X^V_{M_1,\ldots,M_m}$ defined by V of finite range.

3e. Comparison of various measures of order

We have now considered three ways to measure the order (or complexity) of our strictly ergodic $X^V_{M_1,\ldots,M_m}$, using entropy, recursive functions, and spectrum. By any of these measures the "crystalline" case, where $X^V_{M_1,\ldots,M_m}$ is the finite set of translates of some periodic configuration, has the highest possible order. For finite range V we found that the entropy was always zero, but that we could get some disorder in the sense of recursive functions and spectrum. However in neither of these cases could we determine the generic situation.

If we do not require that V be of finite range, and also drop the requirement that V be a **two-body** interaction potential, we can (using a technique due to Aubry [3,4]) easily construct V with very disordered $X^V_{M_1,\ldots,M_m}$ [31]. In particular we can then obtain $X^V_{M_1,\ldots,M_m}$ with positive entropy, and absolutely continuous spectrum. Such results seem to belong to a different class from the ones we have been discussing.

Finally, we should mention the notion of "order parameter" which has been used extensively in physics to help understand the order properties of models **at positive temperatures** [2]. Unfortunately this has not yet proven to be of much use for ground states; see however [30].

4. SUMMARY

We have considered a general class of optimization problems which contains as special cases classic open problems in number theory and condensed matter physics. Informally, in such a problem there are many variable points in space, with points near each other contributing, in pairs, to some global quantity that we want to minimize. The motivating feature is that there seems to be a strong tendency for the solution of such an optimization problem to consist of an array of points very regularly positioned in space, while the reason behind this tendency is unknown. Therefore we seek the degree of order of the solution of a generic problem of the class.

In unifying the problem in terms of the ergodic theory of symbolic dynamical systems we were lead to concentrate not on a single (optimal) array of points in space, but on its orbit closure under translations, and in place of periodic arrays we found that uniquely ergodic orbit closures are the natural setting. We considered a variety of ways of analyzing the degree of order of solutions, including entropy, recursive functions and spectrum. The calculation of entropy is complete, and turns out to be too coarse a measure of order for this problem. There are unexpected examples of disorder in terms of recursive functions and spectrum.

It is hoped that this exposition will both alert experts in specific aspects of the problem to advances on other fronts, and enlist fresh troups to the fray. In particular, consider how we used nonperiodic tilings to construct very unexpected and enlightening examples in statistical mechanics. Benefit also flows the other way. The above presentation develops a shift in emphasis from the individual configurations of a problem, say a tiling problem, to the unique probability measure on the set of configurations; and this in turn lead to a shift from the original question of "periodic versus nonperiodic" to consideration of various aspects of order. (Questions of periodicity were then restated in terms of the discrete part of the spectrum of the measure.) Roughly, this constitutes the introduction of Analysis into the traditional Algebraic field of tiling. One example of the benefits of this reorientation is the set of examples due to Mozes described in section

3d. Finally, we note that Theorem 2 above, which implies that strictly ergodic tilings have zero entropy, is a natural application of statistical mechanical ideas to tiling.

5. REFERENCES

[1] M. Aizenman and E.H. Lieb, The third law of thermodynamics and the degeneracy of the ground state for lattice systems, *J. Stat. Phys.* 24 (1981), 279-297.

[2] P.W. Anderson, *Basic Notions of Condensed Matter Physics*, Benjamin/Cummings, Menlo Park, 1984.

[3] S. Aubry, Weakly periodic structures and example, *J. Phys. (Paris), Coll. C3, supp. no. 3* 50 (1989), 97-106.

[4] S. Aubry, "Weakly periodic structures with a singular continuous spectrum", preprint, Laboratoire Léon Brillouin, Gif-sur-Yvette, 1989.

[5] R. Berger, The undecidability of the domino problem, *Mem. Amer. Math. Soc., no. 66*, (1966).

[6] P. Billingsley, *Ergodic Theory and Information*, John Wiley, New York, 1965.

[7] S.G. Brush, *Statistical Physics and the Atomic Theory of Matter, from Boyle and Newton to Landau and Onsager*, Princeton University Press, Princeton, 1983, p. 277.

[8] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, Springer-Verlag, New York, 1988.

[9] F.M. Dekking and M. Keane, Mixing properties of substitutions, *Z. Wahrsch. Verw. Geb.* 42 (1978), 23-33.

[10] C. Fefferman, The atomic and molecular nature of matter, *Rev. Math. Iberoamericana*, 1 (1985), 1.

[11] C. Fefferman, The N-body problem in quantum mechanics, *Commun. Pure Appl. Math. Suppl.* 39 (1986), S67-S109.

[12] C. Gardner, J. Miękisz, C. Radin and A. van Enter, Fractal symmetry in an Ising model, *J. Phys. A.: Math. Gen.* 22 (1989), L1019-1023.

[13] C.F. Gauss, Untersuchungen über die Eigenschaften der positiven ternären quadratischen Formen von Ludwig August Seeger, Göttingische gelehrte Anzeigen,

1831 Juli 9 = Werke, II (1876), 188-196.

[14] J. Ginibre, in *Systèmes à un nombre infini de degrés de liberté*, CNRS, Paris, 1969, pp. 163-175.

[15] B. Grünbaum and G.C. Shephard, *Tilings and Patterns*, Freeman, New York, 1986.

[16] G.A. Hedlund, Remarks on the work of Axel Thue on sequences, *Nordisk Mat. Tidskrift* 15 (1967), 148-150.

[17] D. Hilbert, Mathematische Probleme, *Archiv. Math. Phys.* 1 (1901), 44-63 and 213-237, translated in *Bull. Amer. Math. Soc.* 8 (1902), 437-479.

[18] E.H. Lieb and H.-T. Yau, The stability and instability of relativistic matter, *Commun. Math. Phys.* 118 (1988), 177-213.

[19] E.H. Lieb and H.-T. Yau, Many-body stability implies a bound on the fine-structure constant, *Phys. Rev. Letts.* 61 (1988), 1695-1697.

[20] M.L. Minsky, *Computation: Finite and Infinite Machines*, Prentice Hall, Englewood Cliffs, New Jersey, 1967.

[21] S. Mozes, Tilings, substitution systems and dynamical systems generated by them, *J. d'Analyse Math.* 53 (1989), 139-186.

[22] D. Myers, Nonrecursive tilings of the plane, II, *J. Symbolic Logic* 39 (1974), 286-294.

[23] W. Parry, Symbolic dynamics and transformations of the unit interval, *Trans. Amer. Math. Soc.* 122 (1966), 368-378.

[24] W. Parry, *Topics in Ergodic Theory*, Cambridge tracts in mathematics, Vol. 75, Cambridge University Press, 1981, pp. 14-15.

[25] R. Penrose, The role of aesthetics in pure and applied mathemtical research, *Bull. Inst. Math. Appl.* 10 (1974), 266-271.

[26] M. Queffélec, *Substitution Dynamical Systems - Spectral Analysis*, Lecture Notes in Mathematics, Vol. 1294, Springer-Verlag, Berlin, 1987.

[27] C. Radin and L.S. Schulman, Periodicity of classical ground states, *Phys.*

Rev. Lett. 51 (1983) 621-622.

[28] C. Radin, Correlations in classical ground states. *J. Stat. Phys.* 43 (1986), 707-712.

[29] C. Radin, Low temperature and the origin of crystalline symmetry, *Int. J. Mod. Phys. B* 1 (1987), 1157-1191 .

[30] C. Radin, Ordering in lattice gases at low temperature, *J. Phys. A: Math. Gen.* 22 (1989), 317-319.

[31] C. Radin, Disordered ground states of classical lattice models, *Rev. Math. Phys.* to appear.

[32] F. Riesz and B. Sz-Nagy, *Functional Analysis*, tr. by L.F. Boron, Frederick Ungar, New York, 1955, pp. 392-393.

[33] R.M. Robinson, Undecidability and nonperiodicity for tilings of the plane, *Invent. Math.* 12 (1971), 177-209.

[34] C.A. Rogers, *Packing and Covering*, Cambridge University Press, 1964, p. 14.

[35] H. Rogers, Jr., *Theory of Recursive Functions and Effective Computabiilty*, McGraw-Hill, New York, 1967.

[36] D. Ruelle, *Statistical Mechanics; Rigorous Results*, Benjamin, New York, 1969.

[37] B. Simon, in *Perspectives in Mathematics: Anniversary of Oberwolfach 1984,* Birkhauser Verlag, Basel, 1984, p. 442.

[38] P. J. Steinhrdt and S. Ostlund, *The Physics of Quasicrystals*, World Scientific, Singapore, 1987.

[39] A. Thue, Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene, *Skr. Vidensk-Selsk., Christ., No. 1*, (1910), 1-9.

[40] A.M. Turing, On computable numbers, with an application to the Entscheidungsproblem, *Proc. London Math. Soc., ser. 2,* 42 (1936), 230-265; 43 (1936), 544-546.

[41] G.E. Uhlenbeck, in *Statistical Mechanics; Foundations and Applications,* ed. T.A. Bak, Benjamin, New York, 1967, p. 581.

[42] G.E. Uhlenbeck, in *Fundamental Problems in Statistical Mechanics II,* ed. E.G.D. Cohen, Wiley, New York, 1968, pp. 16-17.

[43] H. Wang, Proving theorems by pattern recognition II., *Bell Systs. Tech. J.* 40 (1961), 1-41.

[44] H. Wang, Notes on a class of tiling problems, *Fund. Math.* 82 (1975), 295-305.