4

Protein Folding I:
Size Scaling of Time

Conceptual Outline

4.1 The simplest question about dynamics—how long does a process take?—becomes particularly relevant when the time may be so long that the process cannot happen at all. A fundamental problem associated with the dynamics of protein folding is understanding how a system of many interacting elements can reach a desired structure in a reasonable time. In this chapter, we discuss the parallel-processing idea for resolving this problem; kinetic pathways will be considered in the next chapter. Parallel processing and interdependence are at odds and must be balanced in the design of complex systems.

4.2 We use finite-size Ising type models to explore the nature of interactions that can allow a system to relax in a time that grows less than exponentially in the size of the system. These models illustrate various ways to realize the parallel-processing idea.

4.3 The simplest idealization of parallel processing is the case of completely independent spins. We discuss a two-spin model as a first example of how such a system relaxes.

4.4 Various homogeneous models illustrate some of the properties that enable systems to relax in a time that grows no more than a power law in the system size. These include ideal parallel processing, and nucleation and growth of a stable state from a metastable state. The models also illustrate cases where exponential growth in the relaxation time can prevent systems from relaxing.

4.5 Inhomogeneous models extend the range of possibilities for interaction architectures that still allow a reasonable relaxation time. Among these are space and time partitioning and preselected initial conditions. However, inhomogeneous long-range interactions generally lead to an exponential growth of relaxation time with system size.

4.1 The Protein-Folding Problem

One of the simplest questions we can ask about the dynamics of a complex system is, How long does a process take? In some cases this question presumes that we have an understanding of the initial and final state of the process. In other cases we are looking for a characteristic time scale of dynamic change. For a complex system, a particular process may not occur in any reasonable amount of time. The time that a dynamic process takes is of central importance when a system has an identifiable function or purpose. We will consider this in the context of proteins, for which this question is a fundamental issue in understanding molecular function in biological cells.

We begin by describing the structure of proteins, starting from their "primary structure." Proteins are molecules formed out of long chains of, typically, twenty different kinds of amino acids. Amino acids can exist as separate molecules in water, but are constructed so that they can be covalently bonded in a linear chain by removal of one water molecule per bond (Fig. 4.1.1). In general, molecules formed as long chains of molecular units are called polymers. Proteins, RNA and DNA, as well as other types of biological molecules (e.g., polysaccharides) are polymers. In biological cells, proteins are formed in a linear chain by transcription from RNA templates that are themselves transcribed from DNA. The sequence of amino acids forming the protein is called its primary structure (Fig. 4.1.2). The active form of proteins (more specifically

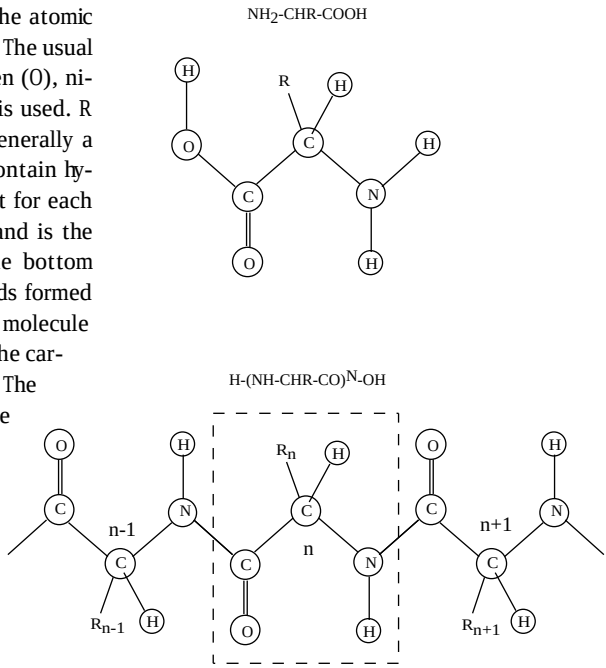

Figure 4.1.1 Illustration of the atomic composition of an amino acid. The usual notation for carbon (C), oxygen (O), nitrogen (N) and hydrogen (H) is used. R stands for a radical that is generally a hydrocarbon chain and may contain hydrocarbon rings. It is different for each of the distinct amino acids, and is the difference between them. The bottom figure is a chain of amino acids formed by removing a single water molecule and bonding one nitrogen to the carbon of the next amino acid. The sequence of amino acids is the primary structure of the protein (see Fig. 4.1.2). ∎

M R L N P G Q Q Q A V E F V T G P C L V L A G A G S G K T R –
V I T N K I A H L I R G C G Y Q A R H I A A V T F T N K A A –
R E M K E R V G Q T L G R K E A R G L M I S T F H T L G L D –
I I K R E Y A A L G M K A N F S L F D D T D Q L A L L K E L –
T E F L I E D D K V L L Q Q L I S T I S N W K N D L K T P S –
Q A A A S A I G E R D R I F A H C Y G L Y D A H L K A C N V –
L D F D D L I L K P T L L L Q A N E E V R K R W Q N K I R Y –
L L V D E Y Q D T M T S Q Y E L V K L L V G S R A R F T V V –
G D D D Q S I Y S W R G A R P Q N L V L L S Q D F P A L K V –
I K L E Q N Y R S S G R I L K A A N I L I A N N P H V F E K –
R L F S E L G Y G A E L K V L S A N N E E H E A E R V T G E –
L I A H H F V N K T Q Y K D Y A I L Y R G N H Q S R V F E K –
F L M Q N R I P Y L O S G G T S F F S R P E I K D L L A Y L –
R V L T N P D D D S A F L R I V N T P K R E I G P A T L K K –
L G E W A M T R N L S M F T A S F D M G L S Q T L S G R G Y –
E A L T R F T H W L A E I Q R L A E R E P I A A V R D L I H –
G M D Y E S W L Y E T S P S P K A A E M R M K N V N Q L F S –
W M T E M L E G S E L D E P M T L T Q V V T R F T L R D M M –
E R G E S E E E L D Q V Q L M T L H A S K G L E F P Y V Y M –
V G M E E G F L P J Q S S I D E D N I D E E R R L A Y V G I –
T R A Q K E L T F T L C K E R R Q Y G E L V R P E P S R F L –
L E L P Q D D L I W E Q E R K V V S A E E R M Q K G Q S H L –
A N L K A M M A A L R G K

Common Amino Acids

Name	Notation	Name	Notation
Glycine	(gly, G)	Cysteine	(cys, C)
Alanine	(ala, A)	Methionine	(met, M)
Valine	(val, V)	Asparagine	(asn, N)
Leucine	(leu, L)	Glutamine	(gln, Q)
Isoleucine	(ile, I)	Aspartic acid	(asp, D)
Phenylalanine	(phe, F)	Glutamic acid	(glu, E)
Tyrosine	(tyr, Y)	Lysine	(lys, K)
Tryptophan	(trp, W)	Arginine	(arg, R)
Serine	(ser, S)	Histidine	(his, H)
Threonine	(thr, T)	Proline	(pro, P)

Figure 4.1.2 Amino acid sequence of the protein acetylcholinesterase — its primary structure. A list of common amino acids and their commonly used three-letter and one-letter notation is attached. ∎

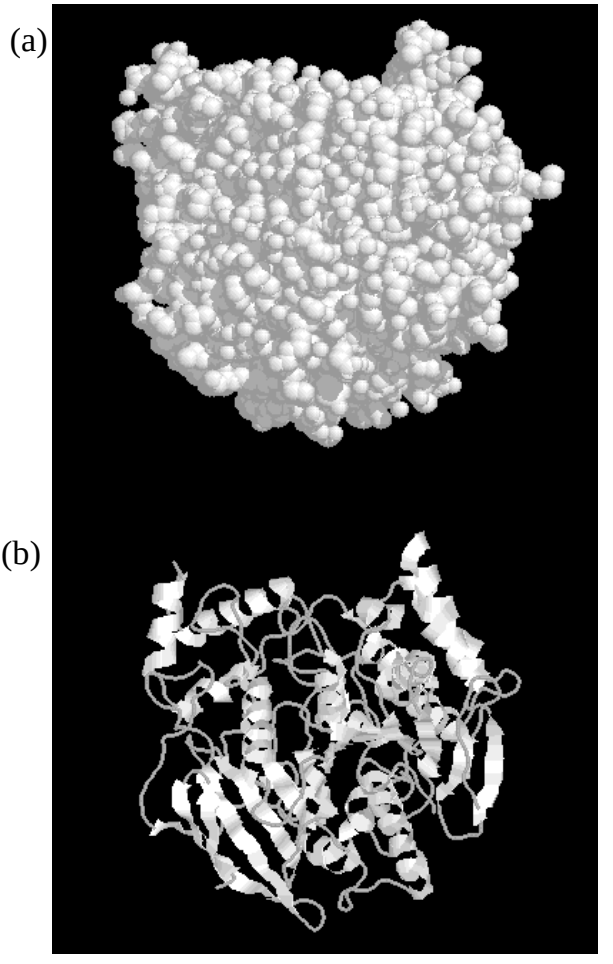

Figure 4.1.3 Three-dimensional structure of the protein acetylcholinesterase. The top picture is constructed using space-filling balls that schematically portray the electron density of each atom. The bottom illustration is a simplified version showing only the backbone of the protein. Helical segments (α-helices) and regions of parallel chains (β-sheets) are visible. They are illustrated as ribbons to distinguish them from the connecting regions of the chain (turns). The α-helices, β-sheets and turns constitute the secondary structure of the protein. (Rendered on a Macintosh using RasMol [developed by Roger Sayle] and a National Institutes of Health protein databank (PDB) file) ∎

globular proteins) is, however, a tightly bound three-dimensional (3-d) structure (Fig. 4.1.3) with active sites on the surface. The active sites serve enzymatic roles, controlling chemical reactions in the cell. The transformation of the linear protein chain to the enzymatically active 3-d structure is known as protein folding. The 3-d structure arises because of additional bonding between the amino acids of the chain. These bonds are characteristically weaker than the covalent bonds along the chain. They include hydrogen bonds, van der Waals bonds and a few covalent sulfur-sulfur (disulfide) bonds. The relative weakness of the bonds responsible for the 3-d structure makes the distinction between the primary and 3-d structure meaningful.

The 3-d structure of proteins can be further analyzed in terms of secondary, tertiary and,sometimes, quaternary structure. These describe levels of spatial organization between individual amino acids and the complete 3-d structure.A plot of the protein chain backbone in space (Fig. 4.1.3 (b)) generally reveals two kinds of amino acid

bonding structures known as α-helix and β-sheet. The α-helix consists of a single-chain helix, where each amino acid forms a hydrogen bond to the fourth amino acid along the chain. Each hydrogen bond attaches the N-H end of one amino acid with the C-OH end of another, resulting in 3.6 amino acids per helix turn. In this structure all such hydrogen bonds are formed, except at the ends of the helix. Thus, from the point of view of the primary chain (and without consideration of the radicals that distinguish different amino acids),this is a low-energy structure. There is a second natural way to provide hydrogen bonding. Placing two chains, or two segments of the same chain, parallel or antiparallel to each other allows a chain of hydrogen bonds. This can be extended on both sides by adding chains in a two-dimensional fashion to form a planar structure that provides complete hydrogen bonds everywhere, except at the edges.This is the β-sheet arrangement. In addition to the α-helix and β-sheet structures there are also segments of the protein,called turns, that connect different α-helix and β-sheet structures. The number of amino acids along a single α-helix typically ranges between ten and twenty-five (three to seven turns), and the number in a single strand of a β-sheet is less, only five to ten. The total number of amino acids in a region of β-sheet can be as high as fifty, divided into three to eight strands. The 3-d structure of a protein described in terms of segments of α-helix and β-sheet is known as the secondary structure of the protein. The number of different secondary-structure elements in a protein ranges from a few up to, possibly, fifty. When there are many secondary structural elements they are further grouped into intermediate structural elements. The complete 3-d structure of an individual amino acid chain is known as its tertiary structure. Several chains may be combined together to form a larger molecular aggregate that constitutes a functioning enzyme. The collective structure of the chains is the enzyme's quaternary structure. This describes the hierarchically subdivided structure of a protein. The number of components at each level of hierarchy is consistent with the generalized 7±2 rule discussed in Chapter 2. This rule is expected to apply to proteins or other complex systems that cannot be subdivided or modified locally without significant change in their global properties.

Protein folding is the transformation of a linear protein chain to the 3-d structure. The problem of understanding protein folding has achieved a separate existence from the problem of describing protein function in the cell. Many proteins can be unfolded and refolded reversibly in a test tube (*in vitro*) separate from other molecules that might otherwise be involved in the protein folding in the cell (*in vivo*). Various additives to the solution cause the protein to unfold or refold. Protein folding has attained a central significance in the effort to understand the molecular biology of the cell, because it is a key to understanding how the linear DNA code is converted into cellular function—as implemented by active enzymes. The 3-d structure of the protein is the form in which they perform enzymatic tasks.

Protein folding is an unsolved problem. What form will the solution of this problem take? One prospect is that it will be possible to predict the 3-d structure from a specified amino-acid sequence. The process of prediction may result from a complete set of rules that describe how particular sequences fold. Alternatively, the prediction may require a large-scale computer simulation of the dynamical process of folding.

Most researchers studying protein folding are concerned with determining or predicting the 3-d structure without describing the dynamics. Our concern is with the dynamics in a generalized context that applies to many complex systems.

From early on in the discussion of the protein-folding problem, it has been possible to separate from the explicit protein-folding problem an implicit problem that begs for a fundamental resolution. How, in principle, can protein folding occur? Consider a system composed of elements, where each element may be found in any of several states. A complete specification of the state of all the elements describes the conformation of the system. The number of possible conformations of the system grows exponentially with the number of elements. We require the system to reach a unique conformation—the folded structure. We may presume for now that the folded structure is the lowest energy conformation of the system. The amount of time necessary for the system to explore all possible conformations to find the lowest-energy one grows exponentially with system size. As discussed in the following paragraphs, this is impossible. Therefore we ask, How does a protein know where to go in the space of conformations to reach the folded structure?

We can adopt some very rough approximations to estimate how much time it would take for a system to explore all possible conformations, when the number of conformations grows exponentially with system size. Let us assume that there are 2^N conformations, where N is the size of the system—e.g., the number of amino acids in a protein. Assume further that the system spends only one atomic oscillation time in each conformation before moving on to the next one. This is a low estimate, so our result will be a reasonable lower bound on the exploration time. An atomic oscillation time in a material is approximately 10^{-12} sec. We should increase this by at least an order of magnitude, because we are talking about a whole amino acid moving rather than a single atom. Our conclusions, however, won't be sensitive to this distinction. The time to relax would be $2^N 10^{-12}$ sec, if we assume optimistically that each possible state is visited exactly once before the right arrangement is found.

A protein folds in, of order, 1 second. For conformation space exploration to work, we would have to restrict the number of amino acids to be smaller than that given by the equation:

$$2^N 10^{-12} \text{ sec} = 1 \text{ sec} \tag{4.1.1}$$

or $N = 40$. Real proteins are formed from chains that typically have 100 to 1000 amino acids. Even if we were to just double our limit from 40 to 80 amino acids, we would have a conformation exploration time of 10^{12} seconds or 32,000 years. The many orders of magnitude that separate a reasonable result from this simple estimate suggests that there must be something fundamentally wrong with our way of thinking about the problem as an exploration of possible conformations. Figuring out what is a reasonable picture, and providing justification for it, is the fundamental protein-folding problem.

The fundamental protein-folding problem applies to other complex systems as well. A complex system always has a large set of possible conformations. The dynamics of a complex system takes it from one type of conformation to another type of

conformation. By the argument presented above, the dynamics cannot explore all possible conformations in order to reach the final conformation. This applies to the dynamics of self-organization, adaptation or function. We can consider neural networks (Chapters 2 and 3) as a second example. Three relevant dynamic processes are the dynamics by which the neural network is formed during physiological development,the dynamics by which it adapts (is trained,learns) and the dynamics by which it responds to external information. All of these cause the neural network to attain one of a small set of conformations,selected from all of the possible conformations of the system. This implies that it does not explore all alternatives before realizing its final form. Similar constraints apply to the dynamics of other complex systems.

Because the fundamental protein-folding problem exists on a very general level, it is reasonable to look at generic models to identify where a solution might exist. Two concepts have been articulated as responsible for the success of biological protein folding—parallel processing and kinetic pathways. The concept of parallel processing suggests, quite reasonably, that more than one process of exploration may be done at once. This can occur ifand only ifthe processes are in some sense independent. If parallel processing works then, naively speaking, each amino acid can do its own exploration and the process will take very little time. In contrast to this picture,the idea of kinetic pathways suggests that a protein starts from a class of conformations that naturally falls down in energy directly toward the folded structure. There are large barriers to other conformations and there is no complete phase space exploration. In this picture there is no need for the folded structure to be the lowest energy conformation—it just has to be the lowest among the accessible conformations.One way to envisage this is as water flowing through a riverbed, confined by river banks, rather than exploring all possible routes to the sea.

Our objective is to add to these ideas some concrete analysis of simple models that provide an understanding of how parallel processing and kinetic pathways may work. In this chapter we discuss the concept of parallel processing, or independent relaxation, by developing a series of simple models. Section 4.2 describes the approximations that will be used. Section 4.3 describes a decoupled two-variable model. The main discussion is divided into homogeneous models in Section 4.4 and inhomogeneous models in Section 4.5.In the next chapter we discuss the kinetic aspects of polymer collapse from an expanded to a compact structure as a first test of how kinetics may play a role in protein folding. It is to be expected that the evolving biology of organisms will take advantage of all possible "tricks" that enable proteins to fold in acceptable time. Therefore it is likely that both parallel processing and kinetic effects do play a role. By understanding the possible generic scenarios that enable rapid folding, we are likely to gain insight into the mechanisms that are actually used.

As we discuss various models of parallel processing we should keep in mind that we are not concerned with arbitrary physical systems, but rather with complex systems. As discussed in Section 1.3, a complex system is indivisible, its parts are interdependent. In the case of proteins this means that the complete primary structure—the sequence of amino acids—is important in determining its 3-d structure. The 3-d structure is sometimes, but not always,affected by changing a single amino acid. It is

likely to be affected by changing two of them. The resulting modifications of the 3-d structure are not localized at the position of the changed amino acids. Both the lack of effect of changing one amino acid,and the various effects of changing more amino acids suggest that the 3-d structure is determined by a strong coupling between the amino acids, rather than being solely a local effect. These observations should limit the applicability of parallel processing, because such a structural interdependence implies that the dynamics of the protein cannot be separated into completely independent parts. Thus, we recognize that the complexity of the system does not naturally lead to an assumption of parallel processing. It is this conflict of the desire to enable rapid dynamics through independence, with the need to promote interdependence, which makes the question of time scale interesting. There is a natural connection between this discussion and the discussion of substructure in Chapter 2. There we showed how functional interdependence arose from a balance between strong and weak interactions in a hierarchy of subsystems. This balance can also be relevant to the problem of achieving essentially parallel yet interdependent dynamics.

Before proceeding, we restate the formal protein-folding problem in a concrete fashion:the objective is to demonstrate that protein folding is consistent with a model where the basic scaling of the relaxation time is reduced from an exponential increase as a function of system size, to no more than a power-law increase. As can be readily verified, for 1000 amino acids, the relaxation time of a system where $\tau \quad N^z$ is not a fundamental problem when $z < 4$. Our discussion of various models in this chapter suggests a framework in which a detailed understanding of the parallel minimization of different coordinates can be further developed. Each model is analyzed to obtain the scaling of the dynamic relaxation (folding) time with the size of the system (chain length).

4.2 Introduction to the Models

We will study the time scale of relaxation dynamics of various model systems as conceptual prototypes of protein folding. Our analysis of the models will make use of the formalism and concepts of Section 1.4 and Section 1.6.A review is recommended. We assume that relaxation to equilibrium is complete and that the desired folded structure is the energy minimum (ground state) over the conformation space. The conformation of the protein chain is described by a set of variables $\{s_i\}$ that are the local relative coordinates of amino acids—specifically dihedral angles (Fig. 4.2.1). These variables, which are continuous variables, have two or more discrete values at which they attain a local minimum in energy. The local minima are separated by energy barriers. Formal results do not depend in an essential way on the number of local minima for each variable. Thus, it is assumed that each variable s_i is a two-state system (Section 1.4), where the two local minima are denoted by $s_i = \pm 1$.

A model of protein folding using binary variables to describe the protein conformation is not as farfetched as it may sound.On the other hand, one should not be convinced that it is the true protein-folding problem. Protein conformational changes do arise largely from changes in the dihedral angles between bonds (Fig. 4.2.1). The

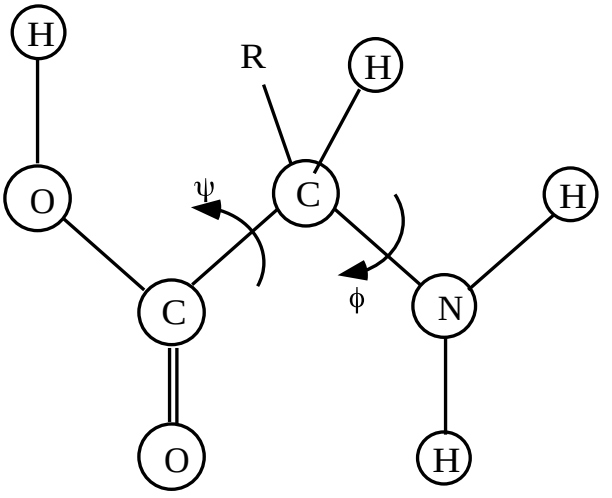

Figure 4.2.1 Illustration of the dihedral angles ψ and ϕ. These coordinates are largely responsible for the variation in protein chain conformation. Changing a single dihedral angle is achieved by rotating all of the protein from one end up to a selected backbone atom. This part of the protein is rotated around the bond that goes from the selected atom to the next along the chain. The rotation does not affect bond lengths or bond-to-bond angles. It does affect the relative orientation of the two bonds on either side of the bond that is the rotation axis. ∎

energy required to change the dihedral angle is small enough to be affected by the secondary bonding between amino acids. This energy is much smaller than the energy required to change bond lengths, which are very rigid, or bond-to-bond angles, which are less rigid than bond lengths but more rigid than dihedral angles. As shown in Fig. 4.2.1, there are two dihedral angles that specify the relative amino acid coordinates. The values taken by the dihedral angles vary along the amino acid chain. They are different for different amino acids, and different for the same amino acid in different locations.

It is revealing to plot the distribution of dihedral angles found in proteins. The scatter plot in Fig. 4.2.2 shows that the values of the dihedral angles cluster around two pairs of values. The plot suggests that it is possible, as a first approximation, to define the conformation of the protein by which cluster a particular amino acid belongs to. It might be suggested that the binary model is correct, by claiming that the variable s_i only indicates that a particular pair of dihedral angles is closer to one of the two aggregation points. However, this is not strictly correct, since it is conceivable that a protein conformation can change significantly without changing any of the binary variables defined in this way.

For our purposes, we will consider a specification of the variables $\{s_i\}$ to be a complete description of the conformation of the protein, except for the irrelevant ro-

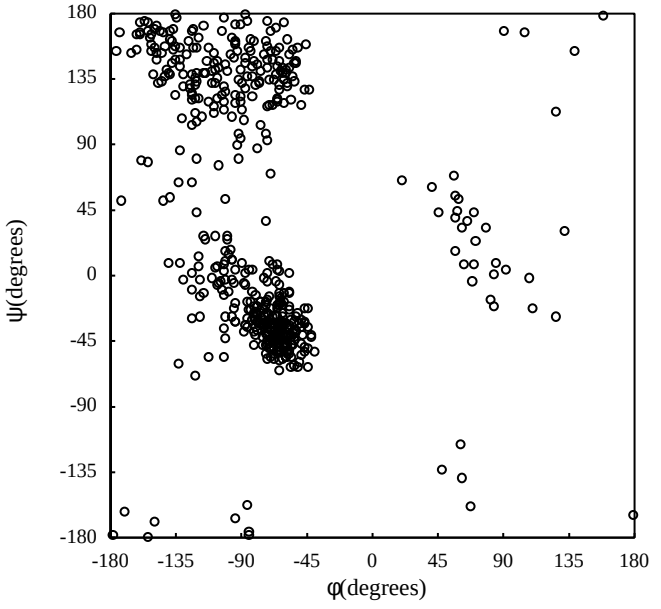

Figure 4.2.2 Scatter plot of the dihedral angle coordinates (Fig. 4.2.1) of each amino acid found along the protein acetylcholinesterase (Figs. 4.1.2–4.1.3). This is called a Ramachandran plot. The coordinates are seen to cluster in two groups. The clustering suggests that it is reasonable to represent the coordinates of the protein using binary variables that specify which of the two clusters a particular dihedral angle pair is found in. The two coordinates correspond to α-helix and β-sheet regions of the protein. The more widely scattered points typically correspond to the amino acid glycine which has a hydrogen atom as a radical and therefore has fewer constraints on its conformation. (Angles were obtained from a PDB file using MolView [developed by Thomas J. Smith]) ∎

tational and translational degrees of freedom of the whole protein. The potential energy, $E(\{s_i\})$, of the protein is a function of the values of all the variables. By redefining the variables $s_i \to -s_i$, when necessary, we let the minimum energy conformation be $s_i = -1$. Furthermore, for most of the discussion, we assume that the unfolded initial state consists of all $s_i = +1$. We could also assume that the unfolded conformation is one of many possible disordered states obtained by randomly picking $s_i = \pm 1$. The folding would then be a disorder-to-order transition.

The potential energy of the system $E(\{s_i\})$ models the actual physical energy arising from atomic interactions, or, more properly, from the interaction between electrons and nuclei, where the nuclear positions are assumed to be fixed and the electrons are treated quantum mechanically. The potential energy is assumed to be evaluated at the particular conformation specified by $\{s_i\}$. It is the potential energy rather than the total energy, because the kinetic energy of atomic motion is not included. Since a protein is in a water environment at non-zero temperature, the

potential energy is actually the free energy of the protein after various positions of water molecules are averaged over. Nevertheless, for protein folding the energy is closely related to the physical energy. This is unlike the energy analog that was used in Chapter 2 for the attractor neural network, which was not directly related to the physical energy of the system.

In addition to the energy of the system, $E(\{s_i\})$, there is also a relaxation time τ_i for each variable, s_i. The relaxation time is governed by the energy barrier E_{Bi} of each two-state system—the barrier to switch between values of s_i. The value of E_{Bi} may vary from variable to variable, and depend on the values of the other variables $\{s_j\}_{j\neq i}$. The model we have constructed is quite similar to the Ising model discussed in Section 1.6. The primary difference is the distinct relaxation times for each coordinate. Unless otherwise specified, we will make the assumption that the time for a single variable to flip is small. Specifically, the relaxation times will be assumed to be bounded by a small time that does not change with the size of the system. In this case the model is essentially the same as an Ising model with kinetics that do not take into account the variation in relaxation time between different coordinates. In specific cases we will address the impact of variation in the relaxation times. However, when there is a systematic violation of the assumption that relaxation times are bounded, the behavior is dominated by the largest barriers or the slowest kinetic processes and a different approach is necessary. Violation of this assumption is what causes the models we are about to discuss not to apply to glasses (Section 1.4), or other quenched systems. In such systems a variable describing the local structure does not have a small relaxation time. The assumption of a short single-variable relaxation time is equivalent to assuming a temperature well above the two-state freezing transition.

Our general discussion of protein folding thus consists of assigning a model for the energy function $E(\{s_i\})$ and the dynamics $\{\tau_i\}$ for the transition from $s_i = +1$ to $s_i = -1$. In this general prescription there is no assumed arrangement of variables in space, or the dimensionality of the space in which the variables are located. We will, however, specialize to fixed spatial arrays of variables in a space of a particular dimension in many of the models. It may seem natural to assume that the variables $\{s_i\}$ occupy a space which is either one-dimensional because of the chain structure or three-dimensional because of the 3-d structure of the eventual protein. Typically, we use the dimensionality of space to distinguish between local interactions and long-range interactions. Neither one nor three dimensions is actually correct because of the many possible interactions that can occur between amino acids when the chain dynamically rearranges itself in space. In this chapter, however, our generic approach suggests that we should not be overly concerned with this problem.

We limit ourselves to considering an expansion of the energy up to interactions between pairs of variables.

$$E(\{s_i\}) = -\sum h_i s_i - \sum J_{ij} s_i s_j \tag{4.2.1}$$

Included is a local preference field h_i determined by local properties of the system (e.g., the structure of individual amino acids), and the pairwise interactions J_{ij}.

Higher-order interactions between three or more variables may be included and can be important. However, the formal discussion of the scaling of relaxation is well served by keeping only these terms. Before proceeding we note that our assumptions imply $\sum h_i < 0$. This follows from the condition that the energy of the initial unfolded state is higher than the energy of the final folded state:

$$E(\{s_i = +1\}) - E(\{s_i = -1\}) = -2\sum h_i > 0 \tag{4.2.2}$$

Thus, in the lower energy state s_i tends to have the same sign as h_i. We will adopt the magnetic terminology of the Ising model in our discussions (Section 1.6). The variables s_i are called spins, the parameters h_i are local values of the external field, the interactions are ferromagnetic if $J_{ij} > 0$ or antiferromagnetic if $J_{ij} < 0$. Two spins will be said to be aligned if they have the same sign. Note that this does not imply that the actual microscopic coordinates are the same, since they have been redefined so that the lowest energy state corresponds to $s_i = -1$. Instead this means that they are either both in the initial or both in the final state. When convenient for sentence structure we use UP (↑) and DOWN (↓) to refer to $s_i = +1$ and $s_i = -1$ respectively. The folding transition between $\{s_i = +1\}$ and $\{s_i = -1\}$, from UP to DOWN, is a generalization of the discussion of first-order transitions in Section 1.6. The primary differences are that we are interested in finite-sized systems (systems where we do not assume the thermodynamic limit of $N \to \infty$) and we discuss a richer variety of models, not just the ferromagnet.

In this chapter we restrict ourselves to considering the scaling of the relaxation time, $\tau(N)$, in these Ising type models. However, it should be understood that similar Ising models have been used to construct predictive models for the secondary structure of proteins. The approach to developing predictive models begins by relating the state of the spins s_i directly to the secondary structure. The two choices for dihedral angles generally correspond to α-helix and β-sheet. Thus we can choose $s_i = +1$ to correspond to α-helix, and $s_i = -1$ to β-sheet. To build an Ising type model that describes the formation of secondary structure, the local fields, h_i, would be chosen based upon propensities of specific amino acids to be part of α-helix and β-sheet structures. An amino acid found more frequently in α-helices would be assigned a positive value of h_i. The greater the bias in probability, the larger the value of h_i. Conversely, for amino acids found more frequently in β-sheet structures, h_i would be negative. The cooperative nature of the α and β structures would be represented by ferromagnetic interactions J_{ij} between near neighbors. Then the minimum energy conformation for a particular primary structure would serve as a prediction of the secondary structure. A chain segment that is consistently UP or DOWN would be α-helix or β-sheet respectively. A chain that alternates between UP and DOWN would be a turn. Various models of this kind have been developed. These efforts to build predictive models have met with some, but thus far limited, success. In order to expand this kind of model to include the tertiary structure there would be a need to include interactions of α and β structures in three dimensions. Once the minimum energy conformation is determined, this model can be converted to a relaxation time model similar to the ones we will discuss, by redefining all of the spins so that $s_i = -1$ in the folded state.

4.3 Parallel Processing in a Two-Spin Model

Our primary objective in this chapter is to elucidate the concept of parallel processing in relaxation kinetics. Parallel processing describes the kinetics of independent or essentially independent relaxation processes. To illustrate this concept in some detail we consider a simple case of two completely independent spins—two independent systems placed side by side. The pair of spins start in a high energy state identified as (1,1), or $s_1 = s_2 = 1$. The low-energy state is (–1,–1), or $s_1 = s_2 = -1$. The system has four possible states: (1,1), (1,–1), (–1,1) and (–1,–1).

We can consider the relaxation of the two-spin system (Fig. 4.3.1) as consisting of hops between the four points (1,1), (1,–1), (–1,1), and (–1,–1) in a two-dimensional plane. Or we can think about these four points as lying on a ring that is essentially one-dimensional, with periodic boundary conditions.Starting from (1,1) there are two possible paths that might be taken by a particular system relaxing to (–1,–1),if we neglect the back transitions. The two paths are $(1,1)\rightarrow(1,-1)\rightarrow(-1,-1)$ and $(1,1)\rightarrow(-1,1)\rightarrow(-1,-1)$. What about the possibility of both spins hopping at once $(1,1)\rightarrow(-1,-1)$? This is not what is meant by parallel processing. It is a separate process,called a coherent transition. The coherent transition is unlikely unless it is enhanced by a lower barrier (lower τ) specifically for this process along the direct path from (1,1) to (–1,–1). In particular, the coherent process is unlikely when the two spins are independent. When they are independent,each spin goes over its own barrier without any coupling to the motion of the other. The time spent going over the barrier is small compared to the relaxation time τ. Thus it is not likely that both will go over at exactly the same time.

There are several ways to describe mathematically the relaxation of the two-spin system.One approach is to use the independence of the two systems to write the probability of each of the four states as a product of the probabilities of each spin:

$$P(s_1,s_2;t) = P(s_1;t)P(s_2;t) \tag{4.3.1}$$

The Master equation which describes the time evolution of the probability can be solved directly by using the solution for each of the two spins separately. We have solved the Master equation for the time evolution of the probability of a two state (one spin) system in Section 1.4. The probability of the spin in state s decays or grows exponentially with the time constant τ:

$$P(s;t) = (P(s;0) - P(s;\infty))e^{-t/\tau} + P(s;\infty) \tag{4.3.2}$$

which is the same as Eq.(1.4.45). The solution of the two-spin Master equation is just the product of the solution of each spin separately:

$$\begin{aligned} P(s_1,s_2;t) &= P(s_1;t)P(s_2;t) \qquad (4.3.3)\\ &= [P(s_1;0)e^{-t/\tau} + (1-e^{-t/\tau})P(s_1;\infty)][P(s_2;0)e^{-t/\tau} + (1-e^{-t/\tau})P(s_2;\infty)] \end{aligned}$$

For simplicity, it is assumed that the relaxation constant τ is the same for both. This equation applies to each of the four possible states. If the energy difference between

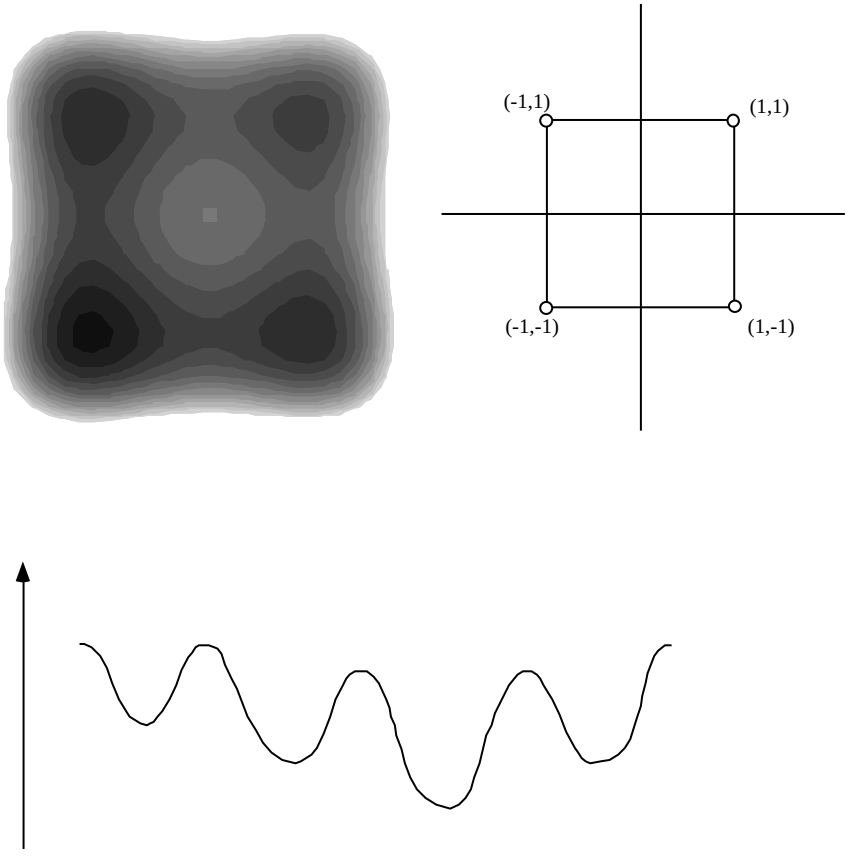

Figure 4.3.1 Illustration of a four-state (two-spin) system formed out of two independent two-state systems. The two-dimensional energy is shown on the upper left. The coordinates of the local energy minima are shown on the right. Below, a schematic energy of the system is shown on a one-dimensional plot, where the horizontal axis goes around the square in the coordinate space of the top right figure. ∎

the UP state and the DOWN state of each spin is sufficiently large,essentially all members of the ensemble will reach the (–1,–1) state. We can determine how long this takes by looking at the probability of the final state:

$$P(-1,-1;t) = [(1-e^{-t/\tau})P(-1;\infty)][(1-e^{-t/\tau})P(-1;\infty)] \approx (1-e^{-t/\tau})^2 \quad (4.3.4)$$

where we have used the initial and final values: $P(-1;0) = 0$, $P(-1;\infty) \approx 1$. Note that a key part of this analysis is that we don't care about the probability of the intermediate states. We only care about the time it takes the system to reach its final state. When does the system arrive at its final state? A convenient way to define the relaxation time of this system is to recognize that in a conventional exponential convergence, τ is the

time at which the system has a probability of only e^{-1} of being anywhere else. Applying this condition here we can obtain the relaxation time, $\tau(2)$, of two independent spins from:

$$1 - P(-1,-1;\tau(2)) = e^{-1} \approx 1 - (1 - e^{-\tau(2)/\tau})^2 \tag{4.3.5}$$

or

$$\tau(2) = \tau[-\ln(1 - (1 - e^{-1})^{1/2})] = 1.585\tau \tag{4.3.6}$$

which is slightly larger than τ. A plot of $P(-1,-1;t)$ is compared to $P(-1;t)$ in Fig. 4.3.2.

Why is the relaxation time longer for two systems? It is longer because we have to wait until the spin that takes the longest time relaxes. Both of the spins relax with the same time constant τ. However, statistically, one will take a little less time and the other a little more time. It is the longest time that is the limiting one for the relaxation of the two-spin system.

Where do we see the effect of parallel processing? In this case it is expressed by the statement that we can take either one of the two paths and get to the minimum energy conformation. If we take the path $(1,1) \rightarrow (1,-1) \rightarrow (-1,-1)$, we don't have to make a transition to the state (–1, 1) in order to see if it is lower in energy. In the two-spin system we have to visit three out of four conformations to get to the minimum energy conformation. If we add more spins, however, this advantage becomes much more significant.

There may be confusion on one important point. The ability to independently relax different coordinates means that the energies of the system for different states are correlated. For example, in the two-spin system, the energies satisfy the relationship

$$E(1,1) - E(1,-1) = E(-1,1) - E(-1,-1) \tag{4.3.7}$$

If we were to assume instead that each of the four energies, $E(\pm 1, \pm 1)$,can be specified independently, energy minimization would immediately require a complete exploration of all conformations. Independence of the energies of different conformations for a system of N spins would require the impossible exploration of all phase

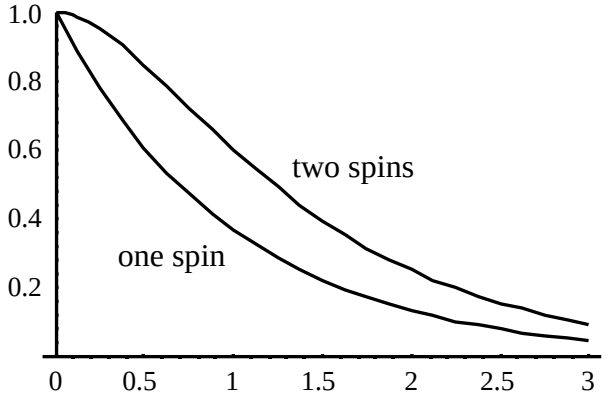

Figure 4.3.2 Comparison of the relaxation of a single spin with the relaxation of a two-spin system. The curves show the probability that the system is not in the ground state. The probability goes to zero asymptotically. The relaxation time is identified with the time when the probability is e^{-1}. ∎

space. It is the existence of correlations in the energies of different conformations that enables parallel processing to work.

4.4 Homogeneous Systems

The models we will consider for a system of N relaxing spins $\{s_i\}$ naturally divide into homogeneous models and inhomogeneous models. For homogeneous models a transformation can be made that maps any spin s_i onto any other spin s_j, where the transformation preserves the form of the energy. Specifically, it preserves the local fields and the interactions between spins.A homogeneous model is loosely analogous to assuming that all amino acids are the same. Such a polymer is called a homopolymer. Boundary conditions may break the transformation symmetry, but their effect can still be considered in the context of homogeneous models. In contrast, an inhomogeneous model is analogous to a heteropolymer where amino acids along the chain are not all the same. Inhomogeneities are incorporated in the models by varying local fields, relaxation times or interactions between spins in a specified way, or by assuming they arise from a specified type of quenched stochastic variable.

In the homogeneous case, all sites are equivalent, and thus the local fields h_i in Eq. (4.2.1) must all be the same. However, the interactions may not all be the same. For example, there may be nearest-neighbor interactions, and different second-neighbor interactions. We indicate that the interaction depends only on the relative location of the spins by the notation J_{i-j}:

$$E(\{s_i\}) = -h\sum s_i - \sum J_{i-j}s_i s_j \tag{4.4.1}$$

J_{i-j} is symmetric in $i-j$ and each pair i,j appears only once in the sum. Eq. (4.2.2) implies that h is negative.A further simplification would be to consider each spin to interact only with z neighbors with equal interaction strength J. This would be the conventional ferromagnet or antiferromagnet discussed in Section 1.6. When it is convenient we will use this simpler model to illustrate properties of the more general case. In the following sections, we systematically describe the relaxation in a number of model homogeneous systems. The results of our investigations of the scaling behavior of the relaxation time are summarized in Table 4.4.1.Each of the models illustrates a concept relevant to our understanding of relaxation in complex systems. This table can be referred to as the analysis proceeds.

4.4.1 *Decoupled*

The simplest homogeneous model is the decoupled case, where all spins are independent. Starting from Eq. (4.4.1) we have:

$$E = -h\sum s_i \tag{4.4.2}$$

This is the N spin analog of the two-spin system we considered in Section 4.3. The energetics are the same as the noninteracting Ising model. However, our interest here is to understand the dependence of kinetics on the number of spins N. The dynamics

Model	Scaling
Decoupled model	$O(\ln(N);1)$
Essentially decoupled model	$O(\ln(N);1)$
Nucleation and growth—with neutral boundaries	$O(a^{N^{(d-1)/d}};N^{-1};\ln(N);1)$
—with nucleating boundaries	$O(N^{1/d};\ln(N);1)$
Boundary-imposed ground state	$O(N^{2/d})$
Long-range interactions	$O(\ln(N);a^{N^2})$

Table 4.4.1 Summary of scaling behavior of the relaxation time of the homogeneous models discussed in Section 4.4. The notation indicates the different scaling regimes from smaller to larger systems. ∎

are defined by the individual two-state systems, where a barrier controls the relaxation rate. Relaxation is described by the exponential decay of the probability that each spin is +1.

We have to distinguish between two possible cases. When we analyzed the two-spin case we assumed that essentially all members of the ensemble reach the unique state where all $s_i = -1$. We have to check this assumption more carefully now. The probability that a particular spin is in the +1 state in equilibrium is given by the expression (Eq. (1.4.14)):

$$P(+1;\infty) = e^{-E_+/kT}/(1+e^{-E_+/kT}) \tag{4.4.3}$$

where $E_+ = -2h$ is the (positive) energy difference between $s_i = +1$ and $s_i = -1$. If we have N spins, the average number that are +1 in equilibrium is

$$N_+ = Ne^{-E_+/kT}/(1+e^{-E_+/kT}) \tag{4.4.4}$$

Because N can be large, we do not immediately assume that this number is negligible. However, we will assume that in equilibrium a large majority of spins are DOWN. This is true only when $E_+ >> kT$ and $e^{-E_+/kT} << 1$. In this case we can approximate Eq. (4.4.4) as:

$$N_+ = Ne^{-E_+/kT} \tag{4.4.5}$$

There are now two distinct possibilities depending on whether N_+ is less than or greater than one. If N_+ is less than one, all of the spins are DOWN in the final state. If N_+ is greater than one, almost all, but not all, of the spins are DOWN in the final state.

In the first case, $N_+ << 1$, we proceed as with the two-spin system to consider the growth of the probability of the final state:

$$P(\{s_i=-1\};t) = \prod_i [P(s_i=-1;0)e^{-t/\tau} + (1-e^{-t/\tau})P(s_i=-1;\infty)] = (1-e^{-t/\tau})^N \tag{4.4.6}$$

Defining the relaxation time as for the two-spin case we have:

$$1 - P(\{s_i=-1\};\tau(N)) = e^{-1} \approx 1-(1-e^{-\tau(N)/\tau})^N \tag{4.4.7}$$

or

$$\tau(N) = \tau[-\ln(1-(1-e^{-1})^{1/N})] \tag{4.4.8}$$

For large N we expand this using $a^{1/N} \quad 1 + (1/N)\ln(a)$ to obtain

$$\begin{aligned}\tau(N) &\sim \tau[-\ln(1-(1+(1/N)\ln(1-e^{-1})))] = \tau[-\ln(-(1/N)\ln(1-e^{-1}))] \\ &= \tau[\ln(N) - \ln(-\ln(1-e^{-1}))] = \tau[\ln(N) + 0.7794]\end{aligned} \tag{4.4.9}$$

Neglecting the constant term, we have the result that the time scales logarithmically with the size of the system $\tau(N) \quad \ln(N)$. We see the tremendous advantage of parallel processing, where the relaxation time grows only logarithmically with system size rather than exponentially.

In the second case, $N >> N_+ >> 1$, we cannot determine the relaxation time from the probability of a particular final state of the system. There is no unique final state. Instead, we have to consider the growth of the probability of the set of systems that are most likely—the equilibrium ensemble with N_+ spins $s_i = +1$. We can guess the scaling of the relaxation time from the divisibility of the system into independent groups of spins. Since we have to wait only until a particular fraction of spins relax, and this fraction does not change with the size of the system,the relaxation time must be independent of the system size or $\tau(N) \quad 1$. We can show this explicitly by writing the fraction of the remaining UP spins as:

$$\begin{aligned}N_+(t) &= \sum_i P(s_i = 1;t) = \sum_i [P(s_i = +1;0)e^{-t/\tau} + (1 - e^{-t/\tau})P(s_i = +1;\infty)] \\ &= N\left[e^{-t/\tau} + (1-e^{-t/\tau})e^{-E_+/kT}\right] \approx N\left[e^{-t/\tau} + e^{-E_+/kT}\right]\end{aligned} \tag{4.4.10}$$

where we use the assumption that $e^{-E_+/kT} << 1$. We must now set a criterion for the relaxation time $\tau(N)$.A reasonable criterion is to set $\tau(N)$ to be the time when there are not many more than the equilibrium number of excited spins,say $(1 + e^{-1})$ times as many:

$$N_+(\tau(N)) = (1 + e^{-1})N_+(\infty) \tag{4.4.11}$$

This implies that:

$$N\left[e^{-\tau(N)/\tau} + e^{-E_+/kT}\right] = (1 + e^{-1})Ne^{-E_+/kT} \tag{4.4.12}$$

or

$$\tau(N) = \tau(E_+/kT + 1) \equiv \tau_\infty \tag{4.4.13}$$

This relaxation time is independent of the size of the system or $\tau(N) \quad 1$; we name it τ_∞.

The $\tau(N) \quad 1$ scaling we found for this case is lower than the logarithmic scaling. We must understand more fully when it applies. In order for N_+ (Eq. (4.4.5)) to be greater than 1, we must have:

$$N > e^{+E_+/kT} \tag{4.4.14}$$

Thus N must be large in order for N_+ to be greater than 1. It may seem surprising that for larger systems the scaling is lower than for smaller systems. The behavior of the scaling is illustrated schematically in Fig. 4.4.1 (see Question 4.4.1).

There is another way to estimate the relaxation time for very large systems, τ_∞. We use the smaller system relaxation Eq.(4.4.9) at the point where we crossover into the regime of Eq.(4.4.14) by setting $N = e^{+E_+/kT}$. Because the relaxation time is a continuous function of N, at the crossover point it should give an estimate of τ_∞. This gives a similar result to that of Eq. (4.4.13):

$$\tau_\infty \sim \tau[E_+ / kT + 0.7794] \tag{4.4.15}$$

Question 4.4.1 Combine the analysis of both cases $N_+ << 1$ and $N_+ >> 1$ by setting an appropriate value of $N_+(\tau(N))$ that can hold in both cases. Use this to draw a plot like Fig. 4.4.1.

Solution 4.4.1 The time evolution of $N_+(t)$ is described by Eq. (4.4.10) for either case $N_+ << 1$ and $N_+ >> 1$. The difficulty is that when $N_+ >> 1$, the process stops when $N_+(t)$ becomes less than 1, and there is no more relaxation

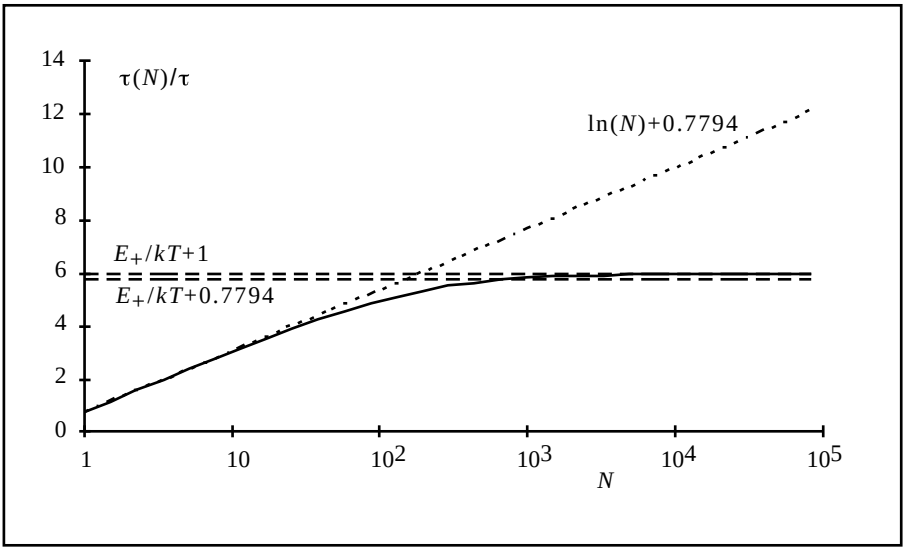

Figure 4.4.1 Relaxation time $\tau(N)$ of N independent spins as a function of N. For systems that are small enough, so that relaxation is to a unique ground state, the relaxation time grows logarithmically with the size of the system. For larger systems, there are always some excited spins, and the relaxation time does not change with system size. This is the thermodynamic limit. The different approximations are described in the text. A unified treatment in Question 4.4.1 gives the solid curve. In the illustrated example, the crossover occurs for a system with about 150 independent spins. This number is given by $e^{E_+/kT}$ so it varies exponentially with the energy difference between the two states of each spin. ∎

to be done. For this case we would like to identify the relaxation time as the time when there is less than one spin not UP. So we replace Eq. (4.4.11) by

$$N_+(\tau(N)) = (1 + e^{-1})N_+(\infty) + N_r$$

where N_r is a constant we can choose which is less than 1. When $N_+ >> 1$, the first term will dominate and we will have the same result as Eq. (4.4.13), when $N_+ << 1$ the second term will dominate. Eq. (4.4.16) leads instead of Eq. (4.4.13) to:

$$\tau(N) = \tau \ln(N / N_r + e^{-1}N_+)) \tag{4.4.17}$$

When $N_+ << 1$ this reduces to:

$$\tau(N) = \tau(\ln(N) - \ln(N_r)) \tag{4.4.18}$$

which is identical to Eq. (4.4.9) if we identify

$$N_r = -\ln(1 - e^{-1}) = 0.4587 \approx .5 \tag{4.4.19}$$

which shows that our original definition of the relaxation time is equivalent to our new definition if we use this value for the average number of residual unrelaxed spins.

The plot in Fig. 4.4.1 was constructed using a value of $E_+/kT = 5$ and Eqs. (4.4.9), (4.4.13), (4.4.15) and (4.4.17). ∎

The behavior for large systems satisfying Eq.(4.4.13) is just the thermodynamic limit where intrinsic properties, including relaxation times, become independent of the system size. In this independent spin model, the relaxation time grows logarithmically in system size until the thermodynamic limit is reached, and then its behavior crosses over to the thermodynamic behavior and becomes constant. To summarize the two regimes, we will label the scaling behavior of the independent system as O(ln(N);1) (the O is read "order").

While the scaling of the relaxation time in the thermodynamic limit is as low as possible, and therefore attractive in principle for the protein-folding problem, there is an unattractive feature—that the equilibrium state of the system is not unique. This violates the assumption we have made that the eventual folded structure of a protein is well defined and precisely given by $\{s_i = -1\}$. However, in recent years it has been found that a small set of conformations that differ slightly from each other constitute the equilibrium protein structure. In the context of this model, the existence of an equilibrium ensemble of the protein suggests that the protein is just at the edge of the thermodynamic regime. In the homogeneous model there is no distinction between different spins, and all are equally likely to be excited to their higher energy state. In the protein it is likely that the ensemble is more selective. For essentially all models we will investigate, for large enough systems, a finite fraction of spins must be thermally excited to a higher energy state. The crossover size depends exponentially on the characteristic energy required for an excited state to occur. This energy is just E_+ in the independent spin model. Because the fraction of excited

states also depends exponentially on the temperature, the structure of proteins is affected by physical body temperature. This is one of the ways in which protein function is affected by the temperature.

Either the logarithmic or the constant scaling of the independent spin model, if correct,is more than adequate to account for the rapid folding of proteins.Of course we know that amino acids interact with each other. The interaction is necessary for interdependence in the system. Is it possible to generalize this model to include some interactions and still retain the same scaling? The answer is yes, but the necessary limitations on the interactions between amino acids are still not very satisfactory.

4.4.2 *Essentially decoupled*

The decoupled model can be generalized without significantly affecting the scaling, by allowing limited interactions that do not affect the relaxation of any spins. To achieve this we must guarantee that at all times the energy of the spin s_i is lower when it is DOWN than when it is UP. For a protein, this corresponds to a case where each amino acid has a certain low-energy state regardless of the protein conformation. We specialize the discussion to nearest-neighbor interactions between each spin and z neighbors—a ferromagnetic or antiferromagnetic Ising model. We also assume the same relaxation time τ applies to all spins at all times. The more general case is deferred to Question 4.4.2.

When there are interactions,the change in energy upon flipping a particular spin s_i from UP to DOWN is dependent on the condition of the other spins $\{s_j\}_{j\neq i}$. We write the change as:

$$E_{+i}(\{s_j\}_{j\neq i}) = E(s_i = +1,\{s_j\}_{j\neq i}) - E(s_i = -1,\{s_j\}_{j\neq i}) = -2h - 2\sum_{j\neq i} J_{i-j}s_j \tag{4.4.20}$$

The latter expression is for homogeneous systems. For only nearest-neighbor interactions in both ferromagnet and antiferromagnet cases

$$E_{+i}(\{s_j\}_{j\neq i}) = -2h - 2J\sum_{nn} s_j \tag{4.4.21}$$

where the sum is over the z nearest neighbors of s_i. Note that this expression depends on the state of the neighboring spins,not on the state of s_i. For the spins to relax essentially independently, we require that the minimum possible value of Eq. (4.4.21)

$$E_{+\min} = -2h - 2z|J| \tag{4.4.22}$$

is greater than zero. To satisfy these requirements we must have

$$|h| > z|J| \tag{4.4.23}$$

which means that the local field $|h|$ is stronger than the interactions. When it is convenient we will also assume that $E_{+\min} >> kT$, so that the energy difference between UP and DOWN states is larger than the thermal energy.

The ferromagnetic case $J > 0$ is the same as the kinetics of a first-order transition (Section 1.6) when the local field is so large that nucleation is not needed and each spin can relax separately. Remembering that $h < 0$,the value of E_{+i} starts from its min-

imum value when all of the spins (neighbors of s_i) are UP, $s_j = +1$. E_{+i} then increases as the system relaxes until it reaches its maximum value everywhere when all the spins (neighbors of s_i) are DOWN, $s_j = -1$ (see Fig. 4.4.2). This means that initially the interactions fight relaxation to the ground state, because they are promoting the alignment of the spins that are UP. However, each spin still relaxes DOWN. The final state with all spins DOWN is self-reinforcing, since the interactions raise the energy of isolated UP spins. This inhibits the excitation of individual spins and reduces the probability that the system is out of the ground state. Thus, ferromagnetic interactions lead to what is called a cooperative ground state. In a cooperative ground state, interactions raise the energy cost of, and thus inhibit, individual elements from switching to a higher energy state. This property appears to be characteristic of proteins in their 3-d structure. Various interactions between amino acids act cooperatively to lower the conformation energy and reduce the likelihood of excited states.

In order to consider the relaxation time in this model, we again consider two cases depending upon the equilibrium number of UP spins, N_+. The situation is more complicated than the decoupled model because the eventual equilibrium N_+ is not necessarily the target N_+ during the relaxation. We can say that the effective $N_+(E_+)$ as

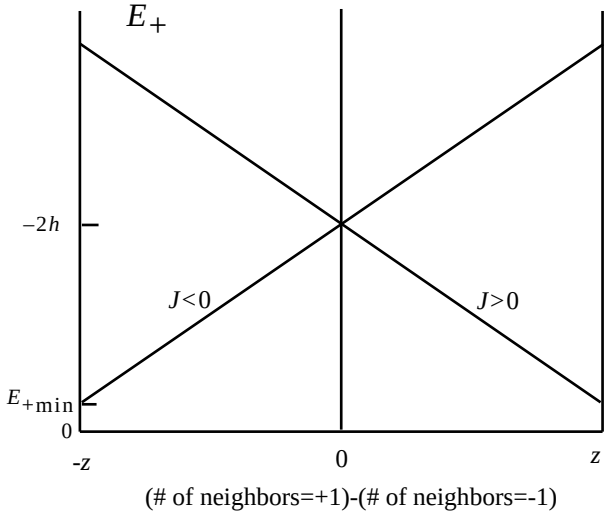

Figure 4.4.2 Illustration, for the essentially decoupled model, of the value of the single-spin energy E_+ as a function of the number of its neighbors (out of a total of z) that are UP and DOWN. At the right all of the neighbors are UP, and at the left all of the neighbors are DOWN. E_+ measures the energy preference of the spin to be DOWN. E_+ is always positive in the essentially decoupled model. The relaxation process to the ground state takes the system from right to left. For a ferromagnet, $J > 0$, the change reinforces the energy preference for the spin to be DOWN. For the antiferromagnet, $J < 0$, the change weakens the energy preference for the spin to be DOWN. Implications for the time scale of relaxation are described in the text. ∎

given by Eq. (4.4.5) changes over time. Because E_+ starts out small, it may not be enough to guarantee that all spins will be DOWN in the final state. But the increase in E_+ may be sufficient to guarantee that all spins will be DOWN at the end.

The situation is simplest if there is complete relaxation toward the ground state at all times. This means:

$$N < e^{+E_{+\min}/kT} \tag{4.4.24}$$

In this case, the relaxation time scaling is bounded by the scaling of the decoupled model. We can show this by going back to the equation for the dynamics of a single relaxing two-state system, as written in Eq. (1.4.43):

$$\dot{P}(1;t) = (P(1;\infty) - P(1;t))/\tau \tag{4.4.25}$$

The difficulty in solving this equation is that $P(1;\infty)$ (Eq. (4.4.3)) is no longer a constant. It varies between spins and over time because it depends on the value of E_+. Nevertheless, Eq. (4.4.25) is valid at any particular moment with the instantaneous value of $P(1;\infty)$. When Eq.(4.4.24) holds, $P(1;\infty)<1/N$ is always negligible compared to $P(1;t)$, even when all the spins are relaxed, so we can simplify Eq. (4.4.24) to be:

$$\dot{P}(1;t) = -P(1;t)/\tau \tag{4.4.26}$$

This equation is completely independent of E_+. It is therefore the same as for the decoupled model. We can integrate to obtain:

$$P(1;t) = e^{-t/\tau} \tag{4.4.27}$$

Thus each spin relaxes as a decoupled system,and so does the whole system with a relaxation time scaling of $O(\ln(N))$.

When Eq. (4.4.24) is not true, the difficulty is that we can no longer neglect $P(1;\infty)$ in Eq. (4.4.25). This means that while the spins are relaxing, they are not relaxing to the equilibrium probability. There are two possibilities. The first is that the equilibrium state of the system includes a small fraction of excited spins. Since the fraction of the excited spins does not change with system size,the relaxation time does not change with system size and is O(1).

The other possibility is that initially the relaxation allows a small fraction of spins to be excited. Then as the relaxation proceeds, the energy differences $E_{+i}(\{s_j\}_{j\neq i})$ increase. This increase in energy differences then causes all of the spins to relax. How does the scaling behave in this case? Since each of the spins relaxes independently, in O(1) time all except a small fraction χN will relax. The remaining fraction consists of spins that are in no particular relationship to each other; they are therefore independent because the range of the interaction is short. Thus,they relax in $O(\ln(\chi N))$ time to the ground state.The total relaxation time would be the sum of a constant term and a logarithmic term that we could write as $O(\ln(\chi N)+1)$, which is not greater than $O(\ln(N))$. This concludes the discussion of the ferromagnetic case.

For the antiferromagnetic case, the situation is actually simpler. Since $J < 0$, remembering that $h < 0$,the value of E_+ starts from its maximum value when all $s_j = +1$, and reaches its minimum value when all $s_j = -1$ (see Fig. 4.4.2). Thus $N_+(E_+)$ is largest in the ground state.Once again,if there is a nonzero fraction of spins at the end that

are UP then the relaxation must be independent of system size, O(1). If there are no residual UP spins in equilibrium,then in Eq.(4.4.25) $P(1;\infty)<1/N$ always,and the relaxation reduces directly to the independent case O(ln(N)).

The ferromagnetic case is essentially different from the antiferromagnetic case because we can continue to consider stronger values of the ferromagnetic interaction without changing the ground state. However, if we consider stronger antiferromagnetic interactions,the ground state will consist of alternating UP and DOWN spins and this is inconsistent with our assumptions (we would have redefined the spin variables). Thus,nearest-neighbor antiferromagnetic interactions,as long as they do not lead to an antiferromagnetic ground state, do not affect the relaxation behavior.

When there are spin-spin interactions, we would also expect the relaxation times τ_i to be affected by the interactions. The relaxation time depends on the barrier to relaxation, E_{Bi}, as shown in the energy curve of the two-state system Fig. 1.4.1. When the energy difference E_+ is higher, we might expect that the barrier to relaxation of the two-state system will become lower. This would be the case if we raise E_+ without raising the energy at the top of the barrier. On the other hand,if the energy surface is multiplied by a uniform factor to increase E_+, then the barrier would increase. These differences in the barrier show up in the relaxation times τ_i. In the former case the relaxation is faster, and in the latter case the relaxation is slower. For the nearest-neighbor Ising model, there would be only a few different relaxation times corresponding to the different possible states of the neighboring spins. We can place a limit on the relaxation time $\tau(N)$ of the whole system by replacing all the different spin relaxation times with the maximum possible spin relaxation time. As far as the scaling of $\tau(N)$ with system size,this will have no effect. The scaling remains the same as in the noninteracting case, O(ln(N);1).

Question 4.4.2 Consider the more general case of a homogeneous model with interactions that may include more than just nearest-neighbor interactions. Restricting the interactions not to affect the minimum energy of a spin,argue that the relaxation time scaling of the system is the same as the decoupled model. Assume that the interactions have a limited range and the system size is much larger than the range of the interactions.

Solution 4.4.2 As in Eq. (4.4.20), the change in energy on flipping a particular spin is dependent on the conditions of the other spins $\{s_j\}_{j\neq i}$.

$$E_{+i}(\{s_j\}_{j\neq i}) = -2h - 2\sum_{j\neq i} J_{i-j}s_j \tag{4.4.28}$$

We assume that $E_{+i}(\{s_j\}_{j\neq i})$ is always positive. Moreover, for relaxation to occur, the energy difference must be greater than kT. Thus the energy must be bounded by a minimum energy $E_{+\min}$ satisfying:

$$E_{+i}(\{s_j\}_{j\neq i}) > E_{+\min} >> kT \tag{4.4.29}$$

This implies that the interactions do not change the lowest energy state of each spin s_i. For the energy of Eq. (4.4.1), $E_{+\min}$ can be written

$$E_{+\min} = -2h - 2\sum_{j\neq i} |J_{i-j}| \tag{4.4.30}$$

Interactions may also affect the relaxation time of each spin $\tau_i\{s_j\}_{j\neq i}$, so we also assume that relaxation times are bounded to be less than a relaxation time $\tau_{\max}$.

We assume that the parameters $\tau_{\max}$ and $E_{+\min}$ do not change with system size. This will be satisfied, for example,if the interactions have a limited range and the system size is larger than the range of the interactions.

Together, the assumption of a bound on the energy differences and a bound on the relaxation times suggest that the equilibration time is bounded by that of a system of decoupled spins with $-2h = E_{+\min}$ and $\tau = \tau_{\max}$. There is one catch. We have to consider again the possibility of incomplete relaxation to the ground state. The scenario follows the same possibilities as the nearest-neighbor model. The situation is simplest if there is complete relaxation to the ground state at all times. This means:

$$N < e^{+E_{+\min}/kT} \tag{4.4.31}$$

which is a more stringent condition than Eq.(4.4.29). In this case the bound on $\tau(N)$ is straightforward because each spin is relaxing to the ground state faster than in the original case. Again using Eq. (1.4.43):

$$\dot{P}(1;t) = (P(1;\infty;t) - P(1;t))/\tau(t) \tag{4.4.32}$$

This equation applies at any particular moment, with the time-dependent values of $P(1;\infty;t)$ and $\tau(t)$, where the time dependence of these quantities is explicitly written. Since $P(1;\infty;t)$ is always negligible compared to $P(1;t)$, when Eq. (4.4.31) applies, this is

$$\dot{P}(1;t) = -P(1;t)/\tau(t) \tag{4.4.33}$$

We can integrate to obtain:

$$P(1;t) = e^{-\int_0^t \frac{dt}{\tau(t)}} < e^{-t/\tau_{\max}} \tag{4.4.34}$$

The inequality follows from the assumption that the relaxation time of each spin is always smaller than $\tau_{\max}$. Each spin relaxes faster than the decoupled system,and so does the whole system. The scaling behavior O(ln(N)) of the decoupled system is a bound for the increase in the relaxation time of the coupled system.

When Eq. (4.4.31) is not true, we can no longer neglect $P(1;\infty;t)$ in Eq.(4.4.32). This means that while the spins are relaxing faster, they are not relaxing to the equilibrium probability. There are two possibilities. The first is that the equilibrium state of the system includes a small fraction of excited spins. Since the range of the interactions is smaller than the system size, the

fraction of the excited spins does not change with system size and the relaxation time does not change with system size. The other possibility is that initially the values of $E_{+i}(\{s_j\}_{j\neq i})$ do not satisfy Eq.(4.4.31) and so allow a small fraction of spins to be excited. Then as the relaxation proceeds, the energy differences $E_{+i}(\{s_j\}_{j\neq i})$ increase. This increase in energy differences then causes all of the spins to relax. The relaxation time will not be larger than $O(\ln(N))$ as long as $E_{+\min} >> kT$ (Eq.(4.4.29)) holds. Because of this condition,each of the spins will almost always relax,and in O(1) time all except a small fraction χN will relax. The remaining fraction consists of spins that are in no particular relationship to each other; they are therefore independent, because the range of the interaction is short, and will relax in at most $O(\ln(\chi N))$ time to the ground state. The total relaxation time would be the sum of a constant term and a logarithmic term that we could write as $O(\ln(\chi N)+1)$, which is not greater than $O(\ln(N))$. ∎

We have treated carefully the decoupled and the almost decoupled models to distinguish between $O(\ln(N))$ and O(1) scaling. One reason to devote such attention to these simple models is that they are the ideal case of parallel processing. It should be understood, however, that the difference between $O(\ln(N))$ and O(1) scaling is not usually significant. For 1000 amino acids in a protein, the difference is only a factor of 7, which is not significant if the individual amino acid relaxation time is microscopic.

One of the points that we learned about interactions from the almost decoupled model is that the ferromagnetic interactions $J > 0$ cause the most problem for relaxation. This is because they reinforce the initial state before the effect of the field h acts to change the conformation. In the almost decoupled model, however, the field h dominates the interactions J. In the next model this is not the case.

The almost decoupled model is not satisfactory in describing protein folding because the interactions between amino acids can affect which conformation they are in. The next model allows this possibility. The result is a new scaling of the relaxation with system size, but only under particular circumstances.

4.4.3 *Nucleation and growth: relaxation by driven diffusion*

The next homogenous model results from assuming that the interactions are strong enough to affect the minimum energy conformation for a particular spin:

$$E_{+\min} < 0 \tag{4.4.35}$$

From Eqs.(4.4.20) and (4.4.21) we see that this implies that the total value of the interactions exceeds the local preference as determined by the field h. Eventually, it is h that ensures that all of the spins are DOWN in the ground state. However, initially when all of the spins are UP, due to the interactions the spins have their lowest energy UP rather than DOWN. During relaxation, when some are UP and some are DOWN, a particular spin may have its lowest energy either UP or DOWN. The effect of both the external field and the interactions together leads to an effective field, $h + \sum_j J_{i-j} s_j$, that determines the preference for the spin orientation at a particular time.

The simplest model that illustrates this case is the Ising ferromagnet in a d-dimensional space (Section 1.6). The interactions are all positive,and the spins try to align with each other. Initially the local preference is for the spins to remain UP; the global minimum of energy is for all of the spins to be DOWN. The resolution of this problem occurs when enough of the spins in a particular region flip DOWN using thermal energy, to create a critical nucleus. A critical nucleus is a cluster of DOWN spins that is sufficiently large so that further growth of the cluster lowers the energy of the system. This happens when the energy lowering from flipping additional spins is larger than the increase in boundary energy between the DOWN cluster and the UP surrounding region.Once a critical nucleus forms in an infinite system,the region of down spins grows until it encounters other such regions and merges with them to form the equilibrium state. In a finite system there may be only one critical nucleus that is formed, and it grows until it consumes the whole system.

The nucleation and growth model of first-order transitions is valid for quite arbitrary interactions when there are two phases, one which is metastable and one which is stable,if there is a well-defined boundary between them when they occur side by side. This applies to a large class of models with finite length interactions. For example, there could be positive nearest-neighbor interactions and negative second-neighbor interactions. As long as the identity of the ground state is not disturbed, varying the interactions affects the value of the boundary energy, but not the overall behavior of the metastable region or the stable region. We do not consider here the case where the boundaries become poorly defined. In our models, the metastable phase consists of UP spins and the stable phase consists of DOWN spins.A system with only nearest-neighbor antiferromagnetic interactions on a bipartite lattice is not included in this section. For $J < 0$ on a bipartite lattice, when Eq.(4.4.35) is satisfied,the ground state is antiferromagnetic (alternating $s_i = \pm 1$),and we would have redefined the spins to take this into consideration.

The dynamics of relaxation for nucleation and growth are controlled by the rate of nucleation and by the rate of diffusion of the boundary between the two phases. Because of the energy difference of the two phases,a flat boundary between them will move at constant velocity toward the metastable phase,converting UP spins to DOWN spins. This process is essentially that of driven diffusion down a washboard potential as illustrated in Fig. 1.4.5. The velocity of the boundary, v, can be measured in units of interspin separation per unit time.

During relaxation, once a critical nucleus of the stable phase forms, it grows by driven diffusion and by merging with other clusters. The number of spins in a particular region of the stable phase grows with time as $(vt)^d$. This rate of growth occurs because the region of the stable phase grows uniformly in all directions with velocity v. Every part of the boundary diffuses like a flat boundary (Fig. 4.4.3). This follows our assumption that the boundary is well defined. There are two parts to this assumption. The first is that the thickness of the boundary is small compared to the size of the critical nucleus. The second is that it becomes smooth, not rough, over time. When these assumptions are satisfied, the stable region expands with velocity v in all directions.

There are several cases that must be considered in order to discuss the scaling of the relaxation time of a finite system of N spins.First we must distinguish three different ranges for the system size. The system may be smaller than the size of a critical nucleus, N_{c0}. If the system is larger than a critical nucleus,then it may be smaller than the typical distance between critical nuclei. Third,it may be larger than this distance. Finally, we must also consider the properties of the boundary of the system, specifically whether or not it promotes nucleation.

Nonnucleating boundary

We start by considering the three system sizes when the boundary of the system is either neutral or suppresses nucleation. Under these circumstances, we can neglect the effect of the boundary because relaxation depends upon nucleation and growth from the interior. The spins near the boundary join the stable phase when it reaches them. We assume throughout that the number of spins in the boundary is negligible compared to the number in the interior.

The case of the system being smaller than the size of the critical nucleus, $N < N_{c0}$, is special because the energy barrier to relaxation grows as the system size increases. The energy may be seen schematically in Fig. 4.4.4 (or Fig. 1.6.10) as a function of cluster size. The washboard-like energy rises in the region below the critical nucleus size. When the system is smaller than the size of a critical nucleus, the energy necessary to form a region of DOWN spins of roughly the size of the system controls the rate of relaxation. Because the energy barrier to forming this region increases roughly linearly with system size, the relaxation time grows exponentially with system size. We can be more precise by using an expression for how the barrier energy grows with system size. The energy of a cluster in an infinite system grows with the number of spins in the cluster as (see also Question 1.6.14):

$$E_c(N_c) = 2hN_c + bN_c^{(d-1)/d} \tag{4.4.36}$$

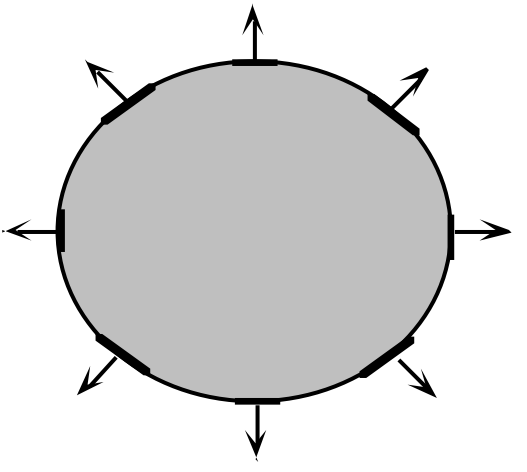

Figure 4.4.3 When a critical nucleus of a stable phase has formed in a metastable phase, the nucleus grows by driven diffusion. The motion of the boundary increases the volume of the equilibrium phase at the expense of the metastable phase. Each part of the boundary moves at a constant average velocity v. Thus, every dimension of the equilibrium phase grows at a constant rate. The number of spins in the equilibrium phase grows as $(vt)^d$ where d is the dimensionality of the space. ∎

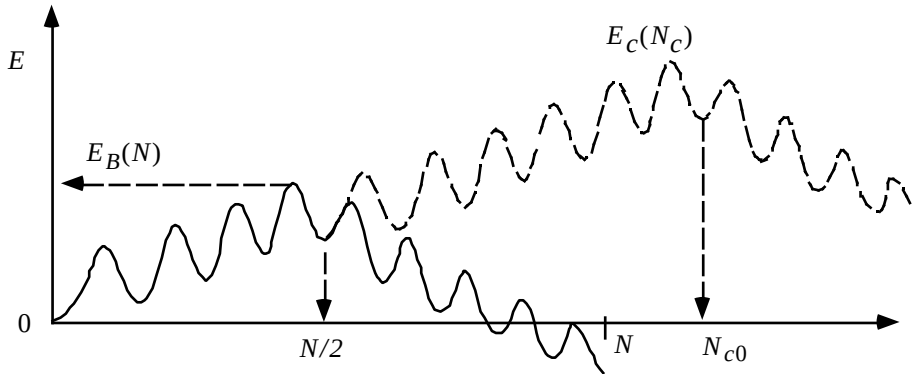

Figure 4.4.4 Schematic illustration of the energy of a cluster of DOWN spins in a metastable background of UP spins as a function of the number of spins in the cluster N_c. The corrugation of the line indicates the energy barrier as each spin flips from UP to DOWN. The dashed line illustrates the energy $E_c(N_c)$ of the cluster in an infinite system. The energy increases until it reaches the size of a critical nucleus N_{c0} and decreases thereafter as the cluster grows to become the stable phase. The solid line indicates the energy in a finite system of size $N < N_c0$. In this case the maximum energy, which is the barrier to relaxation, is located in the vicinity of $N/2$, as indicated. ∎

The first term is the bulk energy of the DOWN spins in the cluster as compared to metastable UP spins. The second term is the boundary energy, where b is a measure of the boundary energy per unit length. This expression is reasonable if the critical nucleus is large compared to the boundary width—the boundary is well defined. The critical nucleus for an infinite system is determined by the maximum value of $E_c(N_c)$. This is obtained setting its derivative with respect to N_c to zero. Aside from a factor of $(d-1)/d$, this means that both terms are equal in magnitude for the critical nucleus. If the system is smaller than the critical nucleus size, then the boundary energy must dominate the bulk energy of a cluster for all possible cluster sizes. Thus for a system with $N < N_{c0}$ we can neglect the first term in $E_c(N_c)$, leaving us with the energy $E_c(N_c) \approx bN_c^{(d-1)/d}$.

For a system with $N < N_{c0}$ that has periodic boundary conditions,the boundary of a cluster grows only as long as the cluster contains less than one-half of the spins in the system. Beyond this point, the boundary of the cluster shrinks. So the maximum cluster energy is reached when N_c is $N/2$. This is still true for a fixed boundary if the boundary is neutral. The relevant cluster may be identified by bisecting the system with UP spins on one side and DOWN spins on the other. If the boundary suppresses nucleation, then the maximum value of N_c may be greater than $N/2$, but it is not larger than N. As long as the maximum value of N_c is proportional to N, the results given below are essentially unaffected.

The cluster energy we have calculated is the energy at the bottom of a particular well in Fig. 4.4.4. It does not include the height of the corrugation E_{B0} which is the energy barrier to flipping a single spin. The energy barrier for nucleation in the system with $N < N_{c0}$ is thus given by

$$E_B(N) = E_c(N/2) + E_{B0} = b(N/2)^{(d-1)/d} + E_{B0} \tag{4.4.37}$$

The relaxation time is given approximately by the probability that the system will reach this barrier energy, as given by a Boltzmann factor of the energy. More specifically, it is given by Eq. (1.4.44), which gives the relaxation of a two-state system with the same barrier (we neglect the back transition rate):

$$\tau(N) = \nu^{-1} e^{E_B(N)/kT} = \nu^{-1} e^{(E_{B0} + bN^{(d-1)/d}/2^{(d-1)/d})/kT} = \tau e^{bN^{(d-1)/d}/2^{(d-1)/d}/kT} \tag{4.4.38}$$

This shows the exponential dependence of the relaxation time on system size in this small system limit when $N < N_{c0}$. We note that we have neglected to consider the many possible ways there are to form a cluster of a particular size, which may also affect the scaling of the relaxation time.

The existence of a region of exponential growth of the relaxation time should be understood in a context where we compare the nucleation time with the observation time. If the nucleation time is long compared to the observation time, we would not expect to see relaxation to the ground state.

If the size of the system is much larger than the size of a critical nucleus, $N >> N_{c0}$, we can consider each nucleus to be essentially a point object of no size, when it forms. A nucleus forms at a particular site according to a local relaxation process with a time constant we denote τ_{c0}—the nucleation time. The nuclei then grow, as discussed previously, with a constant velocity v in each direction. During the relaxation we either have one or many nuclei that form. Only one nucleus forms when the typical time for forming a nucleus in the system is longer than the time a single nucleus takes to consume the whole system. As soon as one nucleus forms, its growth is so rapid that no other nuclei form during the time it grows to the size of the whole system (Fig. 4.4.5(a)). The relaxation time is determined by the time that passes until the first nucleation event occurs in the system. For larger systems, the number of possible nucleation sites increases in direct proportion to N. Thus the time till the first nucleation event decreases, and the relaxation time actually decreases with system size. We will derive the result that $\tau(N) \sim N^{-1}$. To determine when this scaling applies we must find expressions for the nucleation time, and the time a nucleus takes to grow to the size of the system. Independent nuclei can form on every N_{c0} sites. The typical time to form a critical nucleus anywhere in the system, τ_{cN}, where $\tau_{cN} << \tau_{c0}$, is the time it takes any one of the possible N/N_{c0} sites to form a single critical nucleus:

$$(N/N_{c0}) e^{-\tau_{cN}/\tau_{c0}} = N/N_{c0} - 1 \tag{4.4.39}$$

expanding the exponential using $\tau_{cN}/\tau_{c0} << 1$ gives

$$\tau_{cN} = \tau_{c0} N_{c0}/N \tag{4.4.40}$$

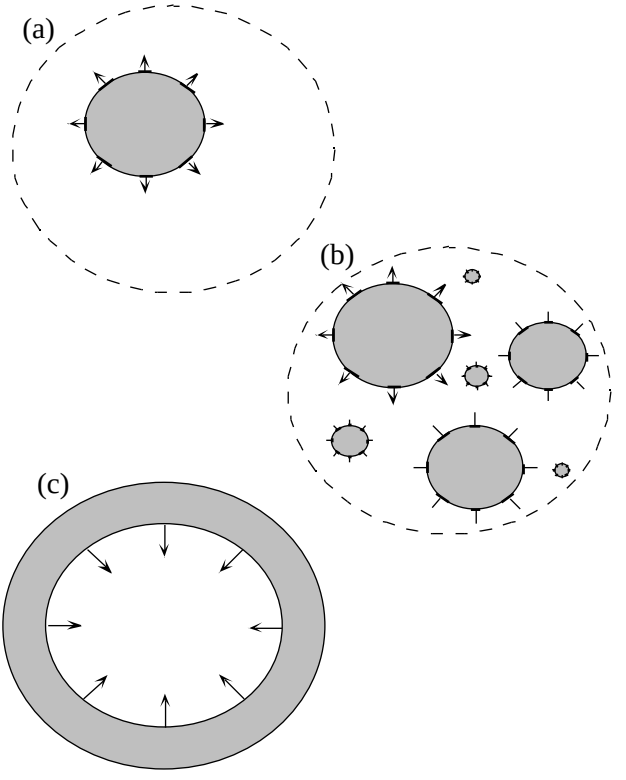

Figure 4.4.5 Several cases of the relaxation of systems by driven diffusion are illustrated. See the text for a detailed discussion. In (a) the system is larger than the size of a critical nucleus but small enough so that only one nucleation event occurs in the system. The boundary is nonnucleating. In (b) the system is large enough so that several nucleation events occur during the relaxation; the boundary is nonnucleating. In (c) the boundary nucleates the equilibrium phase. ∎

This result says that the time to form a single nucleus is inversely proportional to the size of the system. The time for a single nucleus to grow to the size of the system, τ_v, is given by

$$(v\tau_v)^d = N \tag{4.4.41}$$

or

$$\tau_v = N^{1/d} / v \tag{4.4.42}$$

We are neglecting numerical factors that reflect different possible locations the nucleus may form and their relationship to the boundary of the system.

The condition that a single nucleus will form $\tau_{cN} > \tau_v$ is given by combining Eq. (4.4.40) and Eq. (4.4.41) to obtain

$$(v\tau_{c0}N_{c0})^{d/(d+1)} > N >> N_{c0} \tag{4.4.43}$$

where we have also repeated our assumption that the size of the system is larger than the critical nucleus. Eq.(4.4.43) describes the bounds on the system size so that only one nucleus is important. Under these circumstances the relaxation time actually decreases with system size, because as the size of the system increases so do the opportunities for forming critical nuclei. The relaxation time is given by the sum of the nu-

cleation time and the time for consumption of the whole system. The latter has been assumed to be small compared to the former:

$$\tau(N) = \tau_{cN} + \tau_v \approx \tau_{cN} = \tau_{c0} N_{c0} / N \tag{4.4.44}$$

Thus the scaling of the relaxation time is $O(N^{-1})$.

If the system is large enough so that more than one nucleation event occurs (Fig. 4.4.5(b)),then different regions of the material may be treated as essentially decoupled. We expect from the analysis of independent systems that the scaling of the relaxation time is logarithmic.A more detailed analysis given as Question 4.4.3 shows the scaling is $O(\ln(N)^{1/(d+1)})$. While the analysis in Question 4.4.3 has interesting features, the difference between this and O(1) or $O(\ln(N))$ scaling is unlikely to be significant.Finally, as with the independent spin model,the relaxation time is independent of N if N_+ is greater than 1. For convenience we assume that this occurs after the transition between the regime of Fig. 4.4.5(a) and Fig. 4.4.5(b), i.e., for systems in which there are many nucleation events.

Question 4.4.3 Calculate the scaling of the relaxation time when there are many nuclei formed in a system with N spins with boundaries that do not affect the nucleation. Assume that all spins are DOWN at the end of the relaxation. Numerical factors that do not affect the dependence of $\tau(N)$ on N may be neglected.

Solution 4.4.3 Nucleation sites occur randomly through the system and then grow and merge together. In order to find the time at which the whole system will become DOWN, we calculate the probability that a spin at a particular site will remain UP. A particular spin s_i is UP at time t only if there has been no nucleation event in its vicinity that would have grown enough to reach its site.

The probability that no critical nucleus formed at a position r_j with respect to the site s_i until the time t is given by the probability of a two-state system with a time constant τ_{c0} remaining in its high energy state or

$$e^{-t/\tau_{c0}} \tag{4.4.45}$$

If we are looking at the spin s_i at time t, we must ask whether there was formed a nucleus at a distance r away prior to $t' = t - r_j / v$. If the nucleus formed before t' then the nucleus would arrive before time t at the site s_i. The maximum distance that can affect the spin s_i is $r_{\max} = \min(vt, R)$, where $R \quad N^{1/d}$ is the size of the system. When there are many nuclei in the system, then each nucleus is much smaller than the system and $R >> vt$, so that $r_{\max} = vt$. The probability that no nucleus formed within this radius at an early enough time is given by:

$$\prod_{r_j}^{r_{\max}} e^{-(t-r_j/v)/\tau_{c0}} = e^{-\sum (t-r_j/v)/\tau_{c0}} \tag{4.4.46}$$

where the product and the sum are over all possible nucleation sites within a distance $r_{\max}$.

The sum can be directly evaluated to give:

$$\sum_j (t - r_j/v) \quad \frac{1}{N_{c0}} \int_0^{r_{\max}=vt} (t - r/v) r^{d-1} dr \quad (vt)^{d+1}/(vN_{c0}) \qquad (4.4.47)$$

where we divided by the volume of a nucleation site and neglected constants. The number of sites that remain UP is given by N times Eq.(4.4.46) with Eq. (4.4.47) substituted in:

$$N_+ = Ne^{-\chi(vt)^{d+1}/(v\tau_{c0}N_{c0})} \qquad (4.4.48)$$

The coefficient χ accounts for the numerical prefactors we have neglected. Requiring that this is a number less than 1 when $t = \tau(N)$ gives the relaxation time $\tau(N) \sim \ln(N)^{1/(d+1)}$ as indicated in the text.

If we consider this same derivation but do not substitute for $r_{\max}$ in Eq. (4.4.47) then we arrive at the expression:

$$\sum_j (t - r_j/v) \quad \frac{1}{N_{c0}} \int_0^{r_{\max}} (t - r/v) r^{d-1} dr \quad (t - (d/d+1) r_{\max}/v) r_{\max}^d / N_{c0} \qquad (4.4.49)$$

and

$$N_+ = Ne^{-\chi(t-(d/d+1)r_{\max}/v) r_{\max}^d/(\tau_{c0}N_{c0})} \qquad (4.4.50)$$

This more general expression also contains the behavior when we have only one nucleation event. We can recover this case by substituting a constant value of $r_{\max} = R \quad N^{1/d}$. Then the time dependence of N_+ is given by the simple exponential dependence with the relaxation constant $\tau(N) = \tau_{c0}N_{c0}/\chi r_{\max}^d \quad 1/N$. ∎

Nucleating boundary

If the boundary of the system promotes nucleation,the nucleus formed at the boundary will increase by driven diffusion. If there are no other nucleation events (Fig. 4.4.5(c)) then the relaxation-time scales as $\tau(N) \sim N^{1/d}$. Since the critical nucleus forms at the boundary, the system does not have to be larger than a critical nucleus for this to occur. If the system is large enough so that there are many nucleation events,then the behavior of the boundary is irrelevant and the same scaling found before applies.

We have found an anomaly in the intermediate regime characterized by Eq. (4.4.43). In this regime the relaxation time of a system with a nonnucleating boundary decreases, while that with a nucleating boundary increases. It should be understood that for the same microscopic parameters (except at the boundaries),the relaxation time is longer in the former case than in the latter.

Summary

In summary, a system of finite size with a driven-diffusion relaxation has a scaling of the relaxation time with system size as $O(a^{N^{(d-1)/d}}, N^{-1}; \ln(N); 1)$ for a nonnucleating boundary, and $O(N^{1/d}; \ln(N); 1)$ for a nucleating boundary. The results are illustrated in Fig. 4.4.6.

One interesting conclusion from the results in this section is that we do not have to create a very complicated model in order to find a relaxation time that grows exponentially with system size. A ferromagnetic Ising model with a large critical nucleus is sufficient. What is the significance of this result? The size of the critical nucleus N_{c0} and the nucleation time τ_{c0} are both controlled by the magnitude of h compared to the interaction strength J. When h is large the critical nucleus is small and the nucleation time is small. In this model h is the driving force for the relaxation; when this driving force is weak, the relaxation may take arbitrarily long.

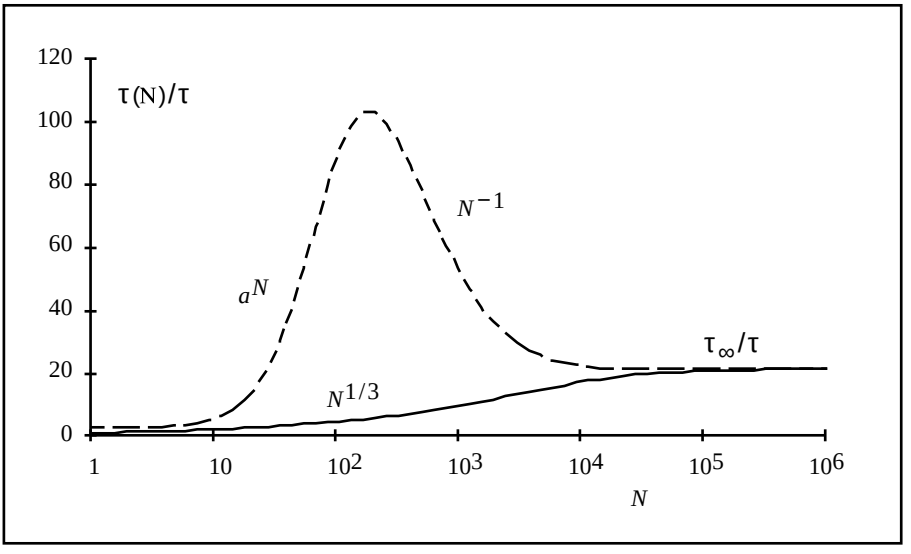

Figure 4.4.6 Schematic plot of the relaxation-time behavior for a system that equilibrates by driven diffusion (see Figs. 4.4.3–4.4.5). Two cases are shown, the solid line is for a system with a boundary that nucleates the stable phase; the dashed line is for a system with a nonnucleating boundary. When the boundary nucleates the stable phase, the stable phase grows by driven diffusion. It consumes the whole system in a time that scales with system size as $N^{1/d}$. For this plot, d is taken to be 3. When the boundary does not nucleate the stable phase, nucleation becomes harder as the system increases in size until it reaches the size of a critical nucleus. For larger systems, the relaxation time decreases because it becomes easier to form a critical nucleus somewhere. Independent of the boundary behavior, when a system becomes so large that the nucleation time, τ_{cN}, becomes equal to the time it takes for driven diffusion to travel the distance between one nucleus and another, τ_v, then the system reaches the large size (thermodynamic) limit and the relaxation time becomes constant. Logarithmic corrections that may arise in this regime have been neglected in the figure. ∎

Our assumption that the relaxation time of an individual spin is rapid should be discussed in this context. We have seen that the nucleation time can become longer than the experimental time. In overcoming the nucleation barrier, the formation of a nucleus is like the relaxation of a two-state system. What we have done,in effect,is to group together a region of spins that is the size of a critical nucleus,and treat them as if they were a single spin. This is a process of renormalization as discussed in Section 1.10. The nucleation time becomes the effective local relaxation time. Thus we see that if the field h is small enough,the effective local relaxation time increases. Even though the ultimate behavior of the system is that of growth by driven diffusion of the stable phase, the relaxation is inhibited locally. This leads to the persistence of the metastable phase.One example of a system where equilibration is inhibited by a long local relaxation time is diamond. Diamond is a metastable phase under standard conditions. The stable phase is graphite.

The second interesting conclusion is the importance of the boundary conditions for the scaling behavior. It is particularly interesting that the new scaling behavior, $N^{1/d}$, arises only for the case of nucleation by the boundary of the system. The scaling behavior of a system with nonnucleating boundaries is quite different, as discussed above.

The model of nucleation and growth of the stable phase has played an important role in conceptual discussions of protein folding. Various theoretical and experimental efforts have been directed at identifying how nucleation and growth of secondary and tertiary structure of proteins might occur. Of particular significance is that it allows interdependence through interactions, and yet can allow relaxation to proceed in a reasonable time. From our discussion it is apparent that nucleating boundaries are beneficial. Our treatment of nucleating boundaries is a mechanism for including the effect of certain system inhomogeneities. While we will consider nucleation and growth more generally in the context of inhomogeneous systems, we will not gain much further insight. The central point is that when there are predetermined nucleation sites, at a boundary or internally in the system, the relaxation of a system into the stable state can proceed rapidly through driven diffusion. This behavior occurs when the interactions in the system are cooperative,so that locally they reinforce both the metastable and stable phases. It is easy to imagine particular amino acids or amino acid combinations serving as nucleation sites around which the 3-d structure of the protein forms. In particular, the formation of an α-helix or β-sheet structure may nucleate at a particular location and grow from there.

Our discussion of nucleation and growth takes care of almost all cases of relaxation in homogeneous systems when the interactions are short-ranged and there is a well-defined ground state in the bulk—away from the boundaries. We can,however, have a well-defined ground state of the system even if the bulk ground state is not well-defined,if the boundary conditions impose the ground state. This is the subject of the next section.

4.4.4 *Boundary-imposed relaxation*

We have been careful to consider cases in which the energy of the state with all spins DOWN, $\{s_i = -1\}$, is lower in energy than any other state,and in particular of the ini-

tial state with all spins UP, $\{s_i = +1\}$. In a system with ferromagnetic interactions,if the energies of the initial and final states are equal, then there are two ground states. In the homogeneous model this is the case where $h = 0$. In general, the existence of two ground states is counter to our assumptions about relaxation to a unique ground state. However, we can still have a unique ground state if the boundaries impose $s_i = -1$. Such boundaries mean that the ground state is uniquely determined to be $\{s_i = -1\}$, even though $h = 0$.

In the absence of additional nucleation events, such a system would equilibrate by inward diffusion of the interface between the UP interior and the DOWN border, as in Fig. 4.4.5(c). There is no bulk driving force that locally causes the UP region to shrink. The only driving force is the interface energy (surface tension) that causes the interface to shrink. We can treat the system as performing a driven random walk in the number of UP spins. However, we must treat each part of the boundary as moving essentially independently. The rate of change of the average number of UP spins $N_+(t)$ is given by the boundary velocity times the boundary length:

$$\frac{dN_+(t)}{dt} \quad vN_+(t)^{(d-1)/d} \tag{4.4.51}$$

The velocity of a driven random walk is (from Eq. (1.4.58) and (1.4.60))

$$v = a\nu(e^{-\Delta E_+/kT} - e^{-\Delta E_-/kT}) = a\nu e^{-\Delta E_+/kT}(1 - e^{(\Delta E_+ - \Delta E_-)/kT}) \tag{4.4.52}$$

From Fig. 1.4.5 we can see that $(\Delta E_+ - \Delta E_-)$ is the energy difference between steps. A single step changes the number of UP spins by one, so

$$(\Delta E_+ - \Delta E_-) = E(N_+(t)) - E(N_+(t) - 1) \approx \frac{dE(N_+(t))}{d(N_+(t))} \quad N_+(t)^{-1/d} \tag{4.4.53}$$

where $E(N_+(t)) \quad N_+(t)^{(d-1)/d}$ is the average surface energy for a cluster with $N_+(t)$ UP spins. Since the single-step energy difference decreases with the number of UP spins, we can assume it is small compared to kT. We can then expand the exponential inside the parenthesis in Eq.(4.4.52) and substitute the resulting expression for the velocity into Eq. (4.4.51) to obtain

$$\frac{dN_+(t)}{dt} \approx -\frac{a\nu}{kT} e^{-\Delta E_+/kT}(\Delta E_+ - \Delta E_-) N_+(t)^{(d-1)/d} \quad -N_+(t)^{-(d-2)/d} \tag{4.4.54}$$

The negative sign is consistent with the decreasing size of the region of UP spins. We integrate Eq.(4.4.54) to find the dependence of the relaxation time on the size of the system:

$$\tau(N) \quad \int_N^0 -N_+(t)^{(2-d)/d} dN_+(t) = N^{2/d} \tag{4.4.55}$$

The final expression is valid even in one dimension, where the boundary executes a random walk because there is no local preference of the boundary to move in one direction or the other and $\tau(N) \quad N^2$.

In this discussion we have ignored the possible effect of the nucleation of regions of DOWN spins away from the boundary of the system.One way to understand this is to note that the size of a critical nucleus is infinite when $h = 0$. Nucleation may only change the relaxation behavior when the interface between UP and DOWN spins is not well-defined. Otherwise, nucleation does not help the relaxation, since any region of DOWN spins inside the region of UP spins will shrink. Thus, for the case where the boundary determines the ground state, the relaxation is $O(N^{2/d})$.

It is also possible to consider the case where h is positive and the bulk preference is to have all of the spins UP. However, the boundary imposes all $s_i = -1$,and the competition between the bulk and boundary energies still results in all of the spins in the ground state being DOWN. This occurs for systems where h is much smaller than $J > 0$, so that the influence of the boundary can extend throughout the system. The energy at the interface is $bN_+(t)^{d/d+1}$ where $b \quad J$ is a local measure of the boundary energy. The bulk energy is $-2hN_+(t)^d$. The latter must be smaller in magnitude than the former. As $N_+(t)$ becomes smaller, the bulk energy becomes still smaller compared to the interface energy. Thus we can neglect the bulk energy in calculating the relaxation time, which scales with N as if $h = 0$.

4.4.5 *Long-range interactions*

When interactions have a range comparable to the system size,the possibility of defining interior and exterior to a domain does not generally exist. If we assume a long-range ferromagnetic interaction between spins so that $J_{ij} = J$, for all i and j, the energy of the system is

$$E(\{s_i\}) = -h\sum_i s_i - J\frac{1}{2}\sum_{i\neq j} s_i s_j \tag{4.4.56}$$

There is a difficulty with this expression because the energy is no longer extensive (proportional to N) since the second term grows as N^2 when all the spins are aligned. As discussed in Section 1.6, for many calculations the long-range interactions are scaled to decrease with system size, $J \sim 1/N$, so that the energy is extensive. However, it is not obvious that this scaling should be used for finite systems. If we keep h and J fixed as the system size increases,then,as shown below, one term or the other dominates in the energy expression.

We can solve Eq. (4.4.56) directly by defining the collective variable

$$M = \sum_i s_i \tag{4.4.57}$$

Substituting this into the energy gives:

$$E(\{s_i\}) = -hM - \tfrac{1}{2}JM^2 + NJ/2 \tag{4.4.58}$$

The term $NJ/2$, which is independent of M, accounts for the missing $i = j$ terms in Eq.(4.4.56). It does not affect any of the results and can be neglected. Adding the entropic contribution from Section 1.6 to obtain the free energy as a function of M we obtain

$$F(M) = -hM - \tfrac{1}{2}JM^2 - TNs_0(M/N) \tag{4.4.59}$$

$$s_0(x) = k\{\ln(2) - \tfrac{1}{2}(1+x)\ln(1+x) + (1-x)\ln(1-x)\} \tag{4.4.60}$$

The exact substitution of the collective variable M for the many variables s_i indicates that this system reduces to a single-variable system. The maximum value of M increases linearly with N. In the following we show self-consistently that M itself grows linearly with N, and obtain the relaxation-time scaling.

Assuming that M grows linearly with N, the first and third terms in the free energy grow linearly with N. The second term $\frac{1}{2}JM^2$, describing interactions, grows quadratically with N. For small enough N the interaction term will be insignificant compared to the other terms, and the system will become essentially decoupled. For a decoupled system M must grow linearly with N. The relaxation time is also the same as the relaxation time of a decoupled system.

For N larger than a certain value,the terms that are linear in N become negligible. Only the interaction term is important. All of the spins must be either UP or DOWN in order to minimize the free energy. This also implies M grows linearly with N. There is a small energy difference between UP and DOWN that is controlled by the value of h. However, to switch between them the system must pass through a conformation where half of the spins are UP and half are DOWN. The energy barrier, $F(M=0) - F(M=N) = JM^2/2$,scales as N^2. Because the barrier grows as N^2 the relaxation time grows as e^{N^2}. Thus the system is frozen into one state or the other. We can still consider raising the temperature high enough to cause the system to flip over the barrier. In this case, however, the difference in energy between UP and DOWN is not enough to force the system into the lower energy state.

Including the small system regime, where the long-range interactions are not relevant, and the large system regime, gives a relaxation-time scaling of $O(\ln(N), e^{N^2})$. We see that even simple models with long-range interactions have a relaxation time that scales exponentially with system size. Another conclusion from this section is that in the presence of long-range interactions, the relaxation-time scaling does not decrease as the system size increases. This behavior was characteristic of systems that have short-range interactions.

It is interesting to consider what would happen if we scale the interactions $J \sim 1/N$. Since all the energies are now extensive, the free-energy barrier would grow linearly in the size of the system and the relaxation time would grow exponentially with the size of the system.Starting from all of the spins UP, the system would rapidly relax to a metastable state consisting of most of the spins UP and a fraction of DOWN spins as determined by the local minimum of the free energy. This relaxation is fast and does not scale with the system size. However, to flip to the ground state with most of the spins DOWN would require an $O(e^N)$ relaxation time.

We could also consider decaying interactions of the form

$$E(\{s_i\}) = -h\sum s_i - \sum J(|r_i - r_j|)s_i s_j \tag{4.4.61}$$

$$J(x) \quad x^p \tag{4.4.62}$$

For $p < -1$ this is essentially the same as short-range interactions, where there is a well-defined boundary and driven diffusion relaxation. For $p > 1$ this is essentially the same as long-range interactions with exponential relaxation time. $p = 1$ is a crossover case that we do not address here.

4.5 Inhomogeneous Systems

In the general inhomogeneous case, each spin has its own preference for orientation UP or DOWN, determined by its local field, h_i, which may be positive or negative. This preference may also be overruled by the interactions with other spins. We begin, however, by reconsidering the decoupled or essentially decoupled model for inhomogeneous local fields and relaxation times.

4.5.1 *Decoupled model—barrier and energy difference variation*

There are two ways in which inhomogeneity affects the decoupled model. Both the spin relaxation time, τ_i, and the energy difference $E_{+i} = -2h_i$, may vary between spins. Analogous to Eq. (4.4.10), the average number of UP spins is given by:

$$N_+(t) = \sum_i P(s_i = -1; t) \approx \sum_i \left[e^{-t/\tau_i} + e^{-E_{+i}/kT} \right] = \sum_i e^{-t/\tau_i} + N_+(\infty) \quad (4.5.1)$$

For a distribution of relaxation times $P(\tau)$ this can be written as:

$$N_+(t) = N \int d\tau \, P(\tau) e^{-t/\tau} + N_+(\infty) \quad (4.5.2)$$

We are assuming that $P(\tau)$ does not depend on the number of spins N. The relaxation time of the system is defined so that all spins relax to their ground state. It might seem natural to define the system relaxation time $\tau(N)$ as before by Eq. (4.4.11) or Eq. (4.4.16):

$$N_+(\tau(N)) = (1 + e^{-1}) N_+(\infty) + N_r \quad (4.5.3)$$

However, allowing an additional factor of e^{-1} spins that are unrelaxed can cause problems in the inhomogeneous model that were not present in the homogeneous case. When there is only one microscopic relaxation time, the existence of nonequilibrium residual UP spins can be considered as a small perturbation on the structure, if they are a smaller fraction of spins than the equilibrium UP spins. There is no special identity to the spins that have not yet relaxed. In the present case, however, the spins with longest relaxation times are the last to relax. It is best not to assume that the structure of the system is relaxed when there are specific spins that have not relaxed. This leads us to adopt a more stringent condition on relaxation by leaving out the e^{-1} in Eq.(4.5.3), $N_+(\tau(N)) = N_+(\infty) + N_r$. Combining this with Eq. (4.5.2) we have:

$$N_r = N \int d\tau \, P(\tau) e^{-\tau(N)/\tau} \quad (4.5.4)$$

where N_r is a number that should be less than one, or for definiteness we can take $N_r \approx 0.5$, as in Eq. (4.4.19).

One way to understand Eq.(4.5.4) is to let all of the spins except one have a relaxation time τ_1. The last one has a relaxation time of τ_2. We ask how does the relaxation time of the final spin affect the relaxation time of the whole system. The relaxation of the spins with τ_1 is given by the usual relaxation time of a system of N spins (Eq. (4.4.17)). If the relaxation time of the final spin is shorter than this, it does not affect the system relaxation time. If it is longer, then the system relaxation time will be determined by τ_2. Thus spins with long relaxation times,almost as long as the relaxation of the whole system,can exist and not effect the relaxation time of the system. The conclusion is more general than the decoupled model.A realization of this in protein folding is the amino acid proline. Experimental studies indicate that proline has two conformations that correspond to cis and trans isomers. The conversion of one form to the other has been found to limit the time of folding of particular proteins. We note that the temperature at which the folding is done can play a role in the relative importance of a single long relaxation time as compared to the relaxation of the rest of the system. When a single relaxation time is large in comparison to the relaxation of other spins,it becomes proportionately even larger as temperature is lowered (Question 4.5.1). The existence of the long proline relaxation time is consistent with a rule of thumb that nature takes advantage of all possibilities. Since it is possible for such a long relaxation time to exist, it does.

Question 4.5.1 Assume all of the spins in a system except one have a relaxation time of τ_1 and the last one has a relaxation time of τ_2. Show that if the last spin has the same relaxation time as the rest of the spins together, at a particular temperature,then it is slower at lower temperatures and faster at higher temperatures.

Solution 4.5.1 The key point is that the relaxation times depend exponentially on the temperature and the large relaxation time will change more rapidly with temperature than the smaller one. The ratio of relaxation times of individual spins as a function of temperature is given by:

$$\tau_2(T)/\tau_1(T) = e^{-E_{B2}/kT} / e^{-E_{B1}/kT} = e^{-(E_{B2}-E_{B1})/kT} \tag{4.5.5}$$

where E_{B1} and E_{B2} are the barrier energies for the respective relaxation processes. In order for the relaxation time of the last spin to be relevant we must have $E_{B2} > E_{B1}$. As a function of temperature,the ratio increases exponentially with decreasing temperature:

$$\tau_2(T)/\tau_1(T) = \left(\tau_2(T_0)/\tau_1(T_0)\right)^{T_0/T} \tag{4.5.6}$$

where T_0 is a reference temperature.

We are interested in comparing the relaxation time of $N-1 \approx N$ spins whose individual relaxation time is $\tau_1(T)$, with the relaxation of one spin

whose individual relaxation time is $\tau_2(T)$. Thus we are concerned with the quantity:

$$\tau_2(T)\ /\tau_1(T,N) \approx \tau_2(T)/(\tau_1(T)\ln(N)) \tag{4.5.7}$$

where we write $\tau_1(T,N)$, as the relaxation time of N spins whose individual relaxation time is τ_1. We have used the expression for this relaxation time obtained from the decoupled spin model of Section 4.4.1. This is not essential; as discussed below the result really only depends on having $\tau_2(T_0)/\tau_1(T_0) >> 1$.

Since we are given that the last spin has the same relaxation time as the rest of the spins together at the reference temperature T_0, i.e., $\tau_2(T_0)/\tau_1(T_0,N) = 1$ evaluating Eq. (4.5.7) at $T = T_0$ we have that:

$$\left(\tau_2(T_0)\ /\tau_1(T_0)\right) = \ln(N) \tag{4.5.8}$$

Considering this relaxation time ratio as an expression for $\ln(N)$, we substitute Eq. (4.5.8) and Eq. (4.5.6) into Eq. (4.5.7) to find that:

$$\tau_2(T)\ /\tau_1(T,N) = \left(\tau_2(T_0)\ /\tau_1(T_0)\right)^{(T_0/T)-1} \tag{4.5.9}$$

which implies the desired result. For $T > T_0$ this ratio is less than one, and for $T < T_0$ this ratio is greater than one.

For the decoupled model, because the relaxation time increases only slowly with the number of spins, the ratio of the relaxation times in Eq. (4.5.8) is not very large, so that the temperature dependence of the ratio of relaxation times will also not be strong, even though it is exponential. However, Eq. (4.5.9) is more general. We can understand this by allowing the rest of the system to interact, except for the individual spin. Our conclusions hold as long as the relaxation of the interacting spins depends on a large number of hops over barriers. These interacting spins give rise to a relaxation time $\tau_1(T,N)$ that depends on the number of spins as some function of N. The consequence in the above equations would only be to replace $\ln(N)$ with this function of N. Eq. (4.5.9) would be unaffected. The ratio of individual spin relaxation times at a reference temperature, $\left(\tau_2(T_0)/\tau_1(T_0)\right)$, could even be determined empirically. Moreover, if a single barrier has a relaxation time of the same duration as the rest of the protein, the conclusion is immediate. Since microscopic relaxation times of a single degree of freedom can be as small as 10^{-10} seconds, and that of the protein is of order 1 second, the ratio between the two relaxation times is large and Eq. (4.5.9) would imply a rapid dependence of the relaxation time ratio with temperature. ∎

The more general case of an arbitrary distribution of individual spin relaxation times $P(\tau)$ in Eq. (4.5.4) can lead to arbitrary scaling of the total relaxation time with the number of spins. Intuitively, there appears to be a problem with this statement, since the spins are independent. How can the relaxation time grow arbitrarily if we

only, say, double the system size? The reason that the relaxation time can grow arbitrarily is that when we increase the system size,there is a greater chance for spins with longer relaxation times to occur. It is the addition of spins in the tail of the distribution of probabilities $P(\tau)$ that controls the scaling of the relaxation time of the system. However, if we only have a few different relaxation times corresponding to a limited number of types of amino acids,then increasing the system size cannot change the relaxation time more than logarithmically with the system size. Thus,if the distribution of spin relaxation barriers is relatively narrow or is composed of a number N_τ of narrow distributions, where $N_\tau << N$, then we will still have the characteristic scaling $O(\ln(N);1)$. This will be assumed for the remaining inhomogeneous models.

From Eq. (4.5.4) we see that variations in E_{+i}, while keeping τ_i fixed, do not affect the scaling of the relaxation time in the decoupled model. If we return to a consideration of the basic properties of relaxation there are two points that imply this conclusion. The first is the effect of E_{+i} on the relaxation rate of an individual spin. The relaxation rate of an individual spin can be affected only if the difference in energy between the two states becomes very small. Even in this case, the change can be at most a factor of 2 (see Eq.(1.4.44)). A factor of 2 is not particularly important when we consider relaxation-time scaling. The second point is that in general we do not allow the value of E_{+i} to become very small because of our assumption that almost all of the spins relax to their ground state. Thus the impact of variations in E_{+i} should be negligible.

Our discussion in this section of the effect of variations in τ_i and E_{+i} is valid also in the case of the essentially decoupled model, where interactions are allowed between spins as long as the interactions do not affect which of the states of a particular spin is the lowest energy. In addition to allowing variations in τ_i and E_{+i}, we can also allow inhomogeneous interactions between spins. In Section 4.4.2, in the homogeneous case, it was natural to assume that the parameters $\tau_{\max}$ and $E_{+\min}$ do not change with system size. In the inhomogeneous case this assumption is less natural. However, once this assumption is made, the arguments presented in Question 4.4.2 proceed as before.

More significant for our interests is that the inhomogeneous case provides new models that retain the same relaxation-time scaling as the decoupled model. Specifically, it is possible for interactions to affect the minimum energy conformation of particular spins without changing the relaxation-time scaling. This is the topic of the next section.

4.5.2 *Space and time partition (decoration of the decoupled model)*

The next inhomogeneous model includes interactions that change the minimum energy state of particular spins. In the homogeneous case this led immediately to models with relaxation controlled by nucleation and growth. In the inhomogeneous case there is a richer analysis.Our first objective is to construct a generalization of the decoupled model that still relaxes with the same scaling. This can happen because, even if a few spins start out with their local equilibrium being UP, as long as the other spins have their equilibrium as DOWN the few spins will relax DOWN once the rest have. We can generalize this systematically. The idea that we will develop in this section is that

an inhomogeneous system may be constructed so that it can be partitioned in space and time. The partitioning results in a finite collection of subsystems. We can then relate the relaxation of the whole system to the relaxation of each of the subsystems, and to the behavior of the subsystems as N increases. Partitioning the system in space and time is closely related to the discussion of subnetworks in Chapter 2. Partitioning in space is directly related to the discussion of subdivision in attractor networks, while partitioning in time is more loosely analogous to the discussion of feedforward networks.

It is useful to consider again the conceptual framework of renormalization discussed in Section 1.10. In essence the subsystems that we will construct are decoupled relaxing variables. They act like individual spins in a decoupled model. We can think about renormalizing the system by grouping together all of the spins in each subsystem. Each subsystem is then replaced by a single spin, with a relaxation time equal to the relaxation time of the original subsystem. The result of the renormalization is a decoupled system of spins. Another way to think about this is to invert the process of renormalization. This inverse process is called decoration. Starting from the decoupled model, we decorate it by replacing each spin with a subsystem formed out of many spins.

Space partitioning is the separation of the whole system into subsystems. We impose a much more stringent form of separation than that in Chapter 2. Within each subsystem the values of the spins may affect each other's minimum energy state but they do not affect the minimum energy state of spins in other subsystems. This does not mean that there are no interactions between spins in different subsystems, only that they are not strong enough to matter. The whole system then relaxes according to the combination of relaxation times of each subsystem combined as in the decoupled case, specifically Eq. (4.5.4). However, the distribution of relaxation times $P(\tau)$ may now depend directly upon N.

As N increases, either the number of subsystems or the size of subsystems grows. If the size of the subsystems does not grow with N, the internal behavior of each subsystem does not affect the scaling of the relaxation time of the whole system. The relaxation of the system depends only on the distribution of relaxation times of the subsystems, exactly as Eq.(4.5.4) describes the relaxation in terms of individual spins. If the number of subsystems does not change and the subsystems grow linearly with N, then the relaxation-time scaling of the whole system follows the relaxation-time scaling of the subsystem with the longest relaxation time. Unless special circumstances apply, this would correspond to the highest scaling. There are other possible ways for the growth of the system with N to be distributed between subsystem growth and growth of the number of subsystems. They can be analyzed in a similar manner.

Time partitioning implies that some spins know their equilibrium conformation from the start. When they are equilibrated, their effect on the remainder causes some of the remaining spins to relax. Then a third set of spins relax. The dynamics is like a row of dominoes. This can be illustrated first by considering only two subsystems. Let

$$W_1 = \{s_i \mid \min(E_{+i}(\{s_j\}_{j\neq i})) > 0\}$$
$$W_2 = \{s_i \mid \min(E_{+i}(\{s_j\}_{j\neq i})) \leq 0\} \tag{4.5.10}$$

Thus, W_2 is the set of s_i such that $E_{+i}(\{s_j\}_{j\neq i})$ can be negative. If all s_i in W_2 are in some sense independent of each other, then the relaxation of the system will still scale as $O(\ln(N);1)$. This is because the spins in W_1 relax first, then the spins in W_2 relax. The condition of independence of spins in W_2 that we need to impose has to do with which spins can affect the sign of their energy $E_{+i}(\{s_j\}_{j\neq i})$. Specifically, the spins whose state can affect the sign of $E_{+i}(\{s_j\}_{j\neq i})$ must all be in W_1, not W_2. This implies that only relatively weak interactions exist between two spins in W_2. If this is true, then consider all spins in W_1. These spins satisfy the conditions of the essentially independent model, so their relaxation takes at most $O(\ln(N);1)$ time. Once these have flipped DOWN, the remaining UP spins, all of which must be in W_2, are decoupled and therefore must satisfy $E_{+i}(\{s_j\}_{j\neq i})>0$. Since they satisfy the conditions of the essentially independent model, they also relax in $O(\ln(N);1)$. The total relaxation is (at most) the sum of these relaxation times and so is also $O(\ln(N);1)$. In summary, the relaxation scaling does not depend on spins that prefer to be UP for some arrangements of their neighbors, if none of their neighbors have this property at the same time as they do.

The partitioning of the system into two subsystems that relax sequentially can be generalized to a finite number of subsystems. If the spins of a system can be partitioned into a finite set of subsystems $\{W_k\}$, such that for a spin s_i of set W_k, $E_{+i}(\{s_j\}_{j\neq i})>0$ when all the $s_j = -1$ in sets $W_1,\ldots,W_{k-1}$, then the system relaxes in $O(\ln(N);1)$. This follows because the subsystems relax sequentially, each in $O(\ln(N);1)$.One may think about the subsystems as illustrated in Fig. 4.5.1.Each successive circle denotes the boundary of a subsystem. The smallest region relaxes first, followed by the next larger one. The scaling $O(\ln(N);1)$ for the whole system follows from the scaling of each subsystem in $O(\ln(N);1)$, when the number of subsystems is assumed to be independent of N. It is also possible to construct models where the number of subsystems grows with N. For specific assumptions about how the number of subsystems changes with N, the relaxation-time scaling can be determined.

A better way to describe the partitioned model uses a concept of the neighborhood of a spin. (The definition of "neighborhood" used in this section does not satisfy the conditions necessary to give a topology on the space.) For statistical fields, the physical distance is not particularly significant; it is the magnitude of the interaction between spins that determines the effective proximity. For the nearest-neighbor Ising models in Section 1.6, we determine interactions by using a spatial arrangement of spins and assign equal interactions to the nearest neighbors. For a cubic lattice, the number of nearest neighbors is directly related to the dimensionality ($z = 2d$). Other lattices give different numbers of neighbors. More generally, we can define the neighbors of a spin s_i as the spins s_j that can change the minimum energy state of the spin s_i.

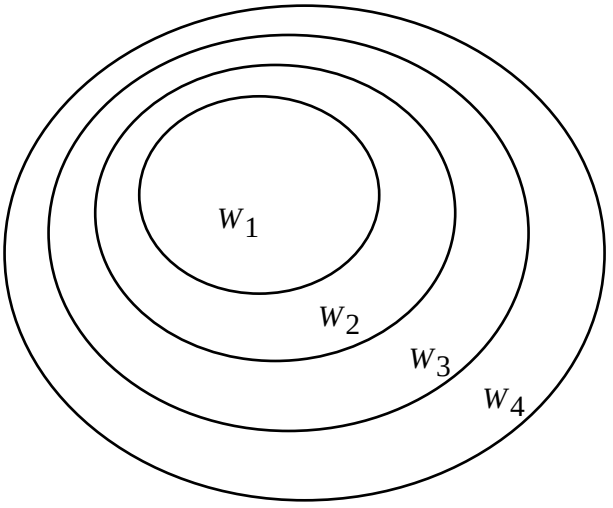

Figure 4.5.1 Illustration of the time-partitioning model. A system of N spins is partitioned so that each of the subsystems $W_1,\ldots,W_k$ relaxes sequentially. In order for the partitions to relax sequentially, the interaction between spins must satisfy a constraint on their interactions described in the text. For example, the relaxation of a spin in W_2 can only depend on spins in W_1 and not on any others. Under these circumstances, once the spins in W_1 relax, so can the spins in W_2. There is no implied spatial arrangement of spins in this model. ∎

Let a neighbor s_j of a spin s_i be a spin that can affect the minimum energy conformation of s_i. Let the neighborhood U_i of s_i be the set of its neighbors. Then the neighborhood of an arbitrary set of spins is the union of the neighborhoods of all its members. We can summarize the results of time partitioning by recognizing that the definition of W_k implies that a spin s_i in W_k must have all of its neighbors in the subsystems $W_1,\ldots,W_{k-1}$. Thus, time partitioning corresponds to a set of subsystems W_k such that the neighborhood of W_k is contained entirely in $W_1,\ldots,W_{k-1}$. For such a system the relaxation time is O(ln(N);1).

We follow a chain of seemingly natural definitions. The interior W^I of a set W is the set of spins whose neighborhoods are entirely contained in W. The exterior W^E of a set W is the set of spins whose neighborhoods do not intersect W. The boundary W^B of a set W is the set of spins that are not in its interior or exterior (spins whose neighborhoods intersect but are not entirely contained within W). For the time-partitioned model,all subsystems W_k are contained in their own exterior, $W_k \quad W_k^E$. This unusual conclusion points to the difference between our neighborhoods and the usual concept of neighborhood. It is rooted in a fundamental asymmetry in our definition of "neighbor".

Time partitioning depends on an asymmetric neighbor relationship. If s_j is a neighbor of s_i, then s_i does not have to be a neighbor of s_j. This arises through inhomogeneity of the local fields h_i that make J_{ij} have a different significance for s_i than for

s_j. The spins with the largest values of h_i tend to be in W_k with lower values of k. A spin in W_1 must have a large enough h_i so that it dominates all of the interactions and there are no spins in its neighborhood.

The definition of "neighborhoods" enables us also to summarize space partitioning. The partitioning of space corresponds to a partitioning of the system into disjoint neighborhoods. The neighborhood of each subsystem does not intersect any other subsystem. Thus, in this case, we can say that each subsystem is the same set of spins as its own interior. Space partitioning can result from both inhomogeneous interactions and fields.

The model of decorated independent relaxation with both spatial and temporal subsystems is attractive as a model of the relaxation in protein folding. The existence of secondary structure, with limitations on the size of secondary-structure elements, suggests that secondary-structure elements may form first. Moreover, each of them may form essentially independently of the others. This would correspond to space partitioning. Each set of coordinates that change and cause the formation of a particular secondary-structure element would be a single subsystem. All of these together would be included in the same time partition. Then there is a second stage of relaxation that forms the tertiary structure. The coordinates that control the formation of the tertiary structure would constitute the second time partition. It is possible, however, and even likely, that during the second stage in which tertiary structure is formed, some of the secondary structure also changes.

4.5.3 *Nucleation and growth in inhomogeneous fields*

Diffusive equilibration can be generalized to the inhomogeneous case. General conclusions can be reached by relatively simple considerations; a complete analysis is difficult. Nucleation and growth is a model that applies when nucleation is a relatively rare event, so that only one critical nucleus forms in a large region. After the critical nucleus is formed, the region of the stable phase grows by driven diffusion of the boundary between the stable and metastable phases. In order to have a diffusive inhomogeneous system, the interactions between spins J_{ij} must be important compared to the variation in the local field, h_i, and the interactions must be essentially local and uniform. Inhomogeneities tend to enhance nucleation and inhibit diffusion of the boundaries between stable and metastable phases. Thus, increasing the inhomogeneity tends to reduce the relevance of nucleation and growth. We will discuss more specifically the effect of variations in h_i and J_{ij}, and then the effect of inhomogeneity in general, on the scaling of the relaxation time.

Inhomogeneities of the local fields h_i cause variations in the local strength of preference for the stable and metastable phases. Regions that have a larger average negative h_i will tend to nucleate before regions of a smaller average negative h_i. Since the problem is to form a nucleus somewhere, in contrast to the rare nucleation in a homogeneous system, this variation increases the rate of nucleation. The effect of variation in h_i on diffusion of a boundary between stable and metastable phases occurs through local variation in the driving force. Sites that have a larger than average negative h_i tend to increase the boundary velocity v, while sites of lower than average

negative h_i tend to decrease the boundary velocity. The boundary must sweep through every site. Moreover, there is no bound on how long the boundary can be delayed, so the sites that slow it tend to trap it. Thus, on average the velocity is reduced.

Inhomogeneities of the interactions J_{ij} cause similar variations in nucleation and diffusion. Smaller values of J_{ij} make nucleation easier and the boundary diffusion slower. Conversely, larger values of J_{ij} make nucleation harder and the boundary diffusion faster. Since nucleation can occur anywhere while diffusion must sweep through everywhere, again the nucleation rate is increased while the diffusion rate is reduced.

For the case of nonnucleating boundaries, the effect on relaxation time is particularly significant. The time necessary to form a critical nucleus is apparent in the relaxation-time scaling behavior as a peak in Fig. 4.4.6. With the introduction of inhomogeneities, the peak will decrease in height. For the case of nucleating boundaries, the relaxation time is controlled by the diffusion rate and so the relaxation time will increase. For both cases, the transition to the thermodynamic limit, where the relaxation time is independent of N, will occur at smaller system sizes. This occurs because the increasing nucleation rate and decreasing diffusion rate causes the typical size to which one nucleus grows—which is the size of independently relaxing parts of the system—to decrease.

Another consideration in the discussion of diffusive relaxation in inhomogeneous fields is the structure of the boundary. In the presence of inhomogeneities, the moving boundary becomes rougher due to the inhomogeneities that slow and speed its motion. As long as the bulk energy dominates the boundary energy, it will remain smooth; however, when the variation in boundary energy becomes large enough, the boundary will become rough and the dynamic behavior of the system will change. Since we have limited ourselves to considering smooth boundaries, our discussion does not apply to this regime.

As briefly discussed in Section 4.4.3, the model of diffusion in variable fields is likely to be of relevance to understanding the local properties of protein folding in the nucleation and growth of the secondary structure. If this applies locally to each of the segments of secondary structure separately, then the scaling of this relaxation is not necessarily relevant to the folding as a whole. However, it is relevant to our understanding of the local kinetics by which secondary structural elements are formed.

4.5.4 *Spin glass*

There have been some efforts to describe the problem of protein folding in terms of a spin glass model and spin glass dynamics. Spin glasses are treated using models that have long-range random interactions between all spins (Section 1.6.6):

$$E[\{s_i\}] = -\frac{1}{2N}\sum_{ij} J_{ij} s_i s_j \tag{4.5.11}$$

The difficulty with this model is that many of the properties of spin glasses do not apply to proteins. Spin glasses have many degenerate ground states, the number of which grows with the size of the system. This means that there is no unique conformation

that can be identified with the folded state of the protein. Choose any conformation, the system will spend much more time in dramatically different conformations because of the essential degeneracy of ground states. Moreover, the barriers between low-lying states also grow with the size of the system. Thus, the relaxation time between any of the low-lying states grows exponentially with system size. Even the concept of equilibration must be redefined for a low temperature spin glass, since true equilibration is not possible. What is possible is a descent into one of the many low-lying energy states. If we model a particular set of interactions J_{ij} as being specified by the primary structure of a protein, there would be no correlation between low-lying states reached by different proteins with the same primary structure. This is in direct contrast to protein folding, where a unique (functional) structure of the protein must be reached.

Despite the great discrepancy between the phenomenology of spin glasses and the protein-folding problem,there are reasons for considering this model. The use of a spin glass model for protein folding is based on the understanding that many possible bonding arrangements between amino acids are possible. For a sufficiently long chain there are many compact conformations of the chain where different bonding arrangements are found. There is always an inherent frustration in the competition between different possible bonding arrangements of the amino acids. This frustration is similar to the frustration that is found in a spin glass. Because of this, in the very long chain limit, the spin glass model should become relevant. In this limit the frustration and multiple ground states are likely to be the correct description of the chain.

However, as discussed in Section 4.4.5, even when there are long-range interactions, the local fields, h_i, can be more important than the interactions, J_{ij}, for small enough systems. In an inhomogeneous system we can expand the term "local field" to include the effect of local interactions:

$$E(\{s_i\}) = -\sum h_i s_i - \frac{1}{2}\sum_{<ij>} J_{ij} s_i s_j - \frac{1}{2}\sum_{ij} J'_{ij} s_i s_j \tag{4.5.12}$$

where the second sum describes the near-neighbor interactions and the third describes the long-range interactions. Long-range interactions that give rise to frustration may not dominate over local interactions. There are many different energies in the problem of protein folding. The analog of local interactions in Eq.(4.5.12) are the interactions between amino acids near each other along the chain, not interactions that are local in space. Hydrogen bonding between different parts of the chain, even though it is local in space,can give rise to frustration. Note that the α-helix structure is constructed entirely out of short-range interactions, while the β-sheet structure is formed out of a combination of short-range and long-range interactions.

There is a difference between bonding between different parts of the amino acid chain and long-range interactions in an Ising model. Although there are many possible hydrogen bond interactions between amino acids,these interactions are quite restricted. The number of amino acids that can interact with a particular amino acid at any one time is limited. Moreover, the chain structure restricts which combinations

of amino acids can interact at one time. These limitations do not eliminate the problem of frustration for very long chains. They do, however, increase the chain length at which crossover occurs,from the regime in which local interactions dominate, to the regime in which long-range interactions dominate. It is the latter regime which is a candidate for the spin glass model.

Our discussion implies that proteins are fundamentally restricted in their length, and that a treatment of their dynamics should include this finite length restriction. From experiment we know that each element of the secondary structure has a limited number of amino acids, and the number of secondary-structure elements in the protein is also limited. These observed limitations on protein size are consistent with our discussion of the relative importance of local fields and long-range interactions.Structural frustration due to long-range interactions must limit the size of proteins to the regime in which local fields, or more generally local interaction energies, are important. It should be assumed that proteins extend up to their maximal possible size. Thus, the largest proteins are likely to be at the crossover point when both short-range and long-range interactions compete. This competition should then play an important role in the relaxation-time scaling.

The assumption of frustration in the long-range interactions appears to be the opposite of the cooperative bonding that has been found in proteins. Cooperative bonding is equivalent to long-range ferromagnetic interactions that enhance the stability of the ground state. Frustration implies that different bonds are competing with each other. It is possible to argue that the low-energy states of the spin glass represent cooperative action of many bonds and therefore constitute cooperative bonding. On the other hand, proteins are engineered, so that we would expect that bonds are designed to reinforce each other and cooperatively lower the energy of the folded state to increase its stability. This is unlike the random spin glass model. This notion of engineered cooperativity leads us to consider the engineered spin glass, which is more typically used as a model of neural network memory.

4.5.5 *Engineered spin glass—neural network*

Neural networks (Chapter 2) have been modeled as engineered spin glass systems (the attractor network) where energy minima of the system are specified. This might be considered to be analogous to the engineering of the 3-d structure of a folded protein by selection of the amino acid sequence. In the attractor network, the interactions J_{ij} determine the minimum energy states. In our discussion of protein folding in this chapter, it is largely the local fields h_i that determine the minimum energy state.One of the differences is that the attractor network cannot have a unique ground state—the inverse of a state has the same energy.

The simplest way to model the engineered spin glass is through the Mattis model (Question 1.6.12). In this model a particular state is determined to be a ground state using only interactions J_{ij} and no local fields h_i. We can redefine all of the spins in the ground state to be $s_i = -1$. Then the Mattis model is equivalent to the long-range ferromagnetic Ising model with no external field, $h = 0$,and all $J_{ij} = J$. Since it is the in-

teractions that determine the ground state, both $s_i = -1$ and its inverse $s_i = +1$ are ground states.

Under these circumstances we cannot consider the folding transition to be from $s_i = +1$ to $s_i = -1$. We can recognize, however, that the essential point of this model is to consider the impact of the initial conditions. We therefore abandon our insistence on starting from a state where all of the spins are UP. The system will relax to the desired ground state if the initial conditions are favorable, specifically, if more of the spins in the initial state are DOWN than are UP.

One way to think about this is to look at the transition in terms of changing suddenly the interaction parameters. Indeed, this is a physically meaningful analogy, since the actual folding of proteins is achieved by changing the interaction energies of the real system. Fig. 4.5.2 illustrates several different transitions on a phase diagram of the ferromagnet that includes both the interaction J and the field h. The transition we have been considering thus far in this chapter is the transition across the first-order transition boundary shown as (A). In this section we are considering the disorder-to-order transition that is represented by (B). As long as there are a majority of DOWN

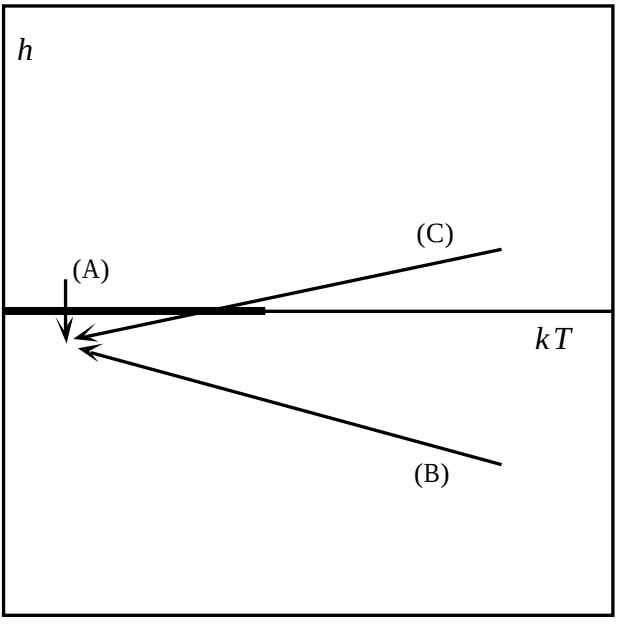

Figure 4.5.2 Illustration of transitions in a ferromagnetic Ising model that start with different initial conditions. The transitions, indicated by arrows, are superimposed on the Ising model phase diagram. The final state in each case corresponds to having all spins DOWN. (A) is a first-order transition starting from all spins UP. (B) and (C) both start from a largely random arrangement of spins but (B) starts from a majority of DOWN spins. (C) starts from a majority of UP spins. ∎

spins in the initial state,there is no need for the process of nucleation and growth to occur. The relaxation is local, and the system reduces to the decoupled model.

We can generalize the Mattis model to the attractor neural network models discussed in Chapter 2. In these models, there may be more than one energy minimum. As with the random spin glass,an arbitrary initial condition leads to any one of these low-energy states. Therefore, we cannot talk about a unique folded state in equilibrium. However, there is a difference in this case. The neural network can be designed to have only a limited number of low-energy states. Each energy state has a basin of attraction that consists of all of the states of the system that will naturally fall toward the low-energy state. The basin of attraction of a particular minimum energy state consists of initial states that have more than a certain overlap with the minimum energy state.Within this basin of attraction,the dynamics that updates the spins reaches the ground state in a finite number of steps. This can be seen to be equivalent to the time-partitioned decoupled model (Section 4.5.2). The spins that flip in a particular update correspond to a particular subsystem. The time scale for relaxation is again $O(\ln(N);1)$.

To make use of the neural network model for protein folding, we can choose an initial conformation that has a finite fraction of spins overlapping with the desired ground state. There is a lesson to be learned from this model regarding the importance of the initial conformation in protein folding. Recently there have been suggestions that the initial conformation is not arbitrary, but instead assumes one of a restricted family of conformations that are either partially folded or are related in some way to the eventual folded conformation. This would be consistent with the concept of a basin of attraction. The introduction of a limited phase space exploration, where the protein does not explore all possible conformations but is restricted from the beginning to the basin of attraction of the folded conformation,also brings us to the second mechanism for reducing the relaxation time—kinetic effects. We will discuss kinetic effects more generally in the next chapter.

The attractor neural network model may also be useful for understanding more complex protein dynamics than just protein folding. Proteins act as enzymes. However, their enzymatic efficiency may be influenced by chemical or other influences that control their function.One mechanism for this control is a change in conformation that affects the active enzymatic site. Thus a protein may respond to a variety of controlling influences by changing its conformation. This suggests that there may be two or more well-defined folded conformations that are each relevant under particular external conditions. If a change in conformation due to a particular external influence is maintained for some time after the external influence is removed,then a description of the protein in terms of multiple minimum energy conformations may become useful.

Missing from attractor neural networks is the incorporation of propagative structures, specifically, interactions that can support driven diffusion or diffusion. Thus, the equilibration of neural network spin glass systems corresponds to the decoupled model and not to any of the models that include driven diffusion or diffusion. The absence of propagative structures is not realistic either for protein folding

or for the general description of neural networks. Feedforward networks are a simple approach to incorporating propagation in neural networks. More complex propagative structures are likely both in proteins and the brain.

4.6 Conclusions

In this chapter we have considered a variety of models that display a range of scaling behavior of the relaxation time with system size. There are diverse individual features of these models that can be related to properties observed in protein-folding experiments. The models also provide some insight into the nature of the relaxation time and its relationship to inter-amino-acid interactions. All of these models, however, are missing the chain structure and its relaxation in space. When a chain is spread out in space, there is an inherent scaling of the relaxation time with chain length, due to the travel time of amino acids through the space before they can bond with other amino acids. In the following chapter we show that this travel time leads to a characteristic relaxation time that scales approximately as N^2 for an expanded chain.

While the models in this chapter are general enough that they cannot be used directly as models of the kinetics of protein folding, this investigation does allow us to relate our findings to other complex systems. There are some general conclusions that can be made. First, it is not difficult to design models that cannot relax in any reasonable time. Long-range interactions, in particular, lead to exponential scaling of the relaxation time. A weak driving force for the transition may also cause problems. There are, however, systematic approaches to interactions that give rise to relaxation in a time that scales as a low power of the size of the system. One approach is partitioning in space and time; another is diffusion or driven diffusion of boundaries; a third is predisposing the system by its initial state; a fourth is dominance of local interactions. All of these are likely to occur in protein folding as well as in the dynamics of other complex systems. It should be apparent that creating a complex system where interactions cause interdependence and yet allow dynamics to proceed in a reasonable time requires a careful design. Complex systems have specific properties that are not generic to physical systems. The issues of how complex systems arise will be discussed in Chapter 6.