THE BEHAVIOR OF PROCESSES WITH STATISTICAL MECHANICAL PROPERTIES

By

Lawrence Gray

IMA Preprint Series # 275

October 1986

INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

UNIVERSITY OF MINNESOTA

514 Vincent Hall

206 Church Street S.E.

Minneapolis, Minnesota 55455

Author(s) Title

165 **D. Arnold and R.S. Falk,** Continuous Dependence on the Elastic Coefficients for a Class of Anisotropic Materials
166 **I.J. Bakelman,** The Boundary Value Problems for Non-linear Elliptic Equation and the Maximum Principle for Euler-Lagrange Equations
167 **Ingo Muller,** Gases and Rubbers
168 **Ingo Muller,** Pseudoelasticity in Shape Memory Alloys - an Extreme Case of Thermoelasticity
169 **Luis Magalhaes,** Persistence and Smoothness of Hyperbolic Invariant Manifolds for Functional Differential Equations
170 **A. Damlamian and M. Vogelius,** Homogenization limits of the Equations of Elasticity in Thin Domains
171 **H.C. Simpson and S.J. Spector,** On Hadamard Stability in Finite Elasticity
172 **J.L. Vazquez and C. Yarur,** Isolated Singularities of the Solutions of the Schrodinger Equation with a Radial Potential
173 **G. Dal Maso and U. Mosco,** Wiener's Criterion and Γ-Convergence
174 **John H. Maddocks,** Stability and Folds
175 **R. Hardt and D. Kinderlehrer,** Existence and Partial Regularity of Static Liquid Crystal Configurations
176 **M. Nerukar,** Consruction of Smooth Ergodic Cocyles for Systems with Fast Periodic Approximations
177 **J.L. Ericksen,** Stable Equilibrium Configurations of Elastic Crystals
178 **Patricio Aviles,** Local Behavior of Solutions of Some Elliptic Equations
179 **S.-N. Chow and R. Lauterbach,** A Bifurcation Theorem for Critical Points of Variational Problems
180 **R. Pego,** Phase Transitions: Stability and Admissibility in One Dimensional Nonlinear Viscoelasticity
181 **Mariano Giaquinta,** Quadratic Functions and Partial Regularity
182 **J. Bona,** Fully Discrete Galerkin Methods for the Korteweg De Vries Equation
183 **J. Maddocks and J. Keller,** Mechanics of Robes
184 **F. Bernis,** Qualitative Properties for some nonlinear higher order
185 **F. Bernis,** Finite Speed of Propagation and Asymptotic Rates for some Nonlinear Higher Order Parabolic Equations with Absorption
186 **S. Reich eistein and S. Reiter,** Game Forms with Minimal Strategy Spaces
An Answer to Littlewood's Problem on Boundedness
188 **J. Rubinstein and R. Mauri,** Dispersion and Convection in Periodic Media
189 **W.H. Fleming and P.E. Souganidis,** Asymptotic Series and the Method of Vanishing Viscosity
190 **H. Beirao Da Veiga,** Existence and Asymptotic Behavior for Strong Solutions of Navier-Stokes Equations in the Whole Space
191 **L.A. Caffarelli, J.L. Vazquez, and N.I. Wolanski,** Lipschitz Continuity of Solutions and Interfaces of the N-Dimensional Porous Medium Equation
192 **R.Johnson,** m-Functions and Floquet Exponents for Linear Differential Systems
193 **F.V. Atkinson and L.A. Peletier,** Ground States and Dirichlet Problems for $-\Delta = F(U)$ In R
194 **G. Dal Maso, U. Mosco,** The Wiener Modulus of a Radial Measure
195 **H. A. Levine and H.F. Weinberger,** Inequalities between Dirichlet and Neumann Eigenvalues
196 **J. Rubinstein,** On the Macroscopic Description of Slow Viscous Flow Past a
197 **G. Dal Maso and U. Mosco,** Wiener Criteria and Energy Decay for Relaxed Dirichlet Problems
198 **V. Oliker and P. Waltman,** On the Monge-Ampere Equation Arising in the Reflect Mapping Problem
199 **M. Chipot, D. Kinderlehrer and L. Caffarelli,** Some Smoothness Properties of Linear Laminates
200 **Y. Giga and R. Kohn,** Characterizing Blow-up Using Similarity Variables
201 **P. Cannarsa and H. M. Soner,** On the Singularities of the Viscosity Solutions to Hamilton-Jacobi-Bellman Equations
202 **Andrew Majda,** Nonlinear Geometric Optics for Hyperbolic Systems of Conservation Laws
203 **G. Buttazzo, G. Dal Maso and U. Mosco,** A Derivation Theorem for Capacities with Respect to a Radon Measure
204 **S. Cowin, M. Mehrabadi,** On the Identification of Material Symmetry for Anisotropic Elastic Materials
205 **R.W.R. Darling,** Constructing Nonhomeomorphic Stochastic Flows.
206 **M. Chipot,** On the Reynolds Lubrication Equation
207 **R.V. Kohn and G.W. Milton,** On Bounding the Effective Conductivity of Anisotropic Composites
208 **I.J. Bakelman,** Notes Concerning the Torsion of Hardening Rods and Its N-Dimensional Generalizations
209 **I.J. Bakelman,**The Boundary Value Problems for Non-Linear Elliptic Equationsll
210 **Guangiu Gong & Minping Qian,**On the Large Deviation Functions of Markov Chains
211 **Arie Leizarowitz,**Control Problems with Random and Progressively Known Targets
212 **R.W.R. Darling,** Ergodicity of a Measure-Valued Markov Chain Induced by Random Transformations
213 **G. Gong, M. Qian & Zhongxin Zhao,** Killed Diffusions and Its Conditioning
214 **Arie Leizarowitz,** Controlling Diffusion Processes on Infinite Horizon with the Overtaking Criterion
215 **Millard Beatty,** The Poisson Function of Finite Elasticity
216 **David Terman,** Traveling Wave Solutions Arising From a Combustion Model
217 **Yuh-Jia Lee,** Sharp Inequalities and Regularity of Heat Semi-Group on Infinite Dimensional Spaces
218 **D. Stroock,** Lecture Notes
219 **Claudio Canuto,** Spectral Methods and Maximum Principle
220 **Thomas O'Brien,** A Two Parameter Family of Pension Contribution Functions and Stochastic Optimization
221 **Takeyuki Hida,** Analysis of Brownian Functionals
222 **Leonid Hurwicz,** On Informational Decentralization and Efficiency of Resource Allocation Mechanisms
223 **E.B. Fabes and D.W. Stroock,** A New Proof of Moser's Parabolic Harnack Inequality via the Old Ideas of Nash
224 **Minoru Murata,** Structure of Positive Solution to $(-\Delta+V)u = 0$ In R^n
225 **Paul Dupuis,** Large Deviations Analysis of Reflected Diffusion Constrained Stochastic Approximation Algorithms in Convex Sets
226 **F. Bernis,** Existence Results for Doubly Nonlinear Higher Order Parabolic Equations on Unbounded Domains.
227 **S. Orey and S. Pelikan,** Large Deviations Principles for Stationary Processes
228 **R. Gulliver and S. Hildebrandt,** Boundary Configurations Spanning Continua of Minimal Surfaces.
229 **J. Baxter, G. Dal Maso& U. Mosco,** Stopping Times and T-Convergence.
230 **Julio Boulliet,** Self-Similar Solutions, Having Jumps and Intervals of Constancy of a Diffusion-heat Conduction Equation
231 **R. Hardt, D. Kinderlehrer & F.-H. Lin,** A Remark About the Stability of Smooth Equilibrium Configurations of Static Liquid Crystal
232 **M. Chipot and M. Luskin,** The Compressible Reynolds Lubrication Equation
233 **J.H. Maddocks,** A Model for Disclinations in Nematic Liquid Crystal
234 **C. Foias, G.R. Sell and R. Temam,** Inertial Manifolds for Nonlinear Evolutionary Equations
235 **P.L. Chow,** Expectation Functionals Associated with Some Stochastic Evolution
236 **Guiseppe Buttazzo,** Reinforcement by a Thin Layer with Oscillating Thickness
237 **W.H. Fleming, S.J. Sheu and H.M. Soner,** On Existence of the Dominant Eigenfunction and Its Application to the Large Deviation Properties of an Ergodic Markov Process
238 **R. Jensen and P.E. Souganidis,** A Regularity Result for Viscosity Solutions of Hamilton-Jacobi Equations in one Space Dimension
239 **B. Boczar-Karaki, J.L. Bona and D.L. Cohen,** Interaction of Shallow-Water Waves and Bottom Topography
240 **F. Colonius and W. Kliemann,** Infinite Time Optimal Control and Periodicity
Dirichlet Problems

THE BEHAVIOR OF PROCESSES WITH STATISTICAL MECHANICAL PROPERTIES

Lawrence Gray
School of Mathematics
University of Minnesota
Minneapolis, MN 55455

1. Introduction.

Ever since Spitzer's famous paper in 1970, there has been interest in a class of Markov processes which have as time-reversible stationary measures certain special distributions from the theory of statistical mechanics. The state space for these processes is $\Xi = \{-1, +1\}^{\mathbf{Z}^d}$, which is the space of configurations of + and − spins on the sites of the lattice $\mathbf{Z}^d$. Transitions occur when there is a "flip" at a site $x \in \mathbf{Z}^d$, or in other words, a change of sign in the spin at x. The probability that a flip occurs at x in a short time interval $(t, t + h]$, given the history of the process up to time t, is $c_x(\xi_t)h + o(h)$, where ξ_t is the state of the process at time t, and c_x is a non-negative function defined on Ξ, called the flip rate at x. Simultaneous flips at two different sites do not occur. A system of Markov processes with this description, one process for each possible initial state, is often called a "spin-flip system" with rates $\{ c_x \}$. Spitzer pointed out that for certain kinds of interaction potentials commonly used in statistical mechanics, one can always find a set of rates $\{ c_x \}$ such that the corresponding spin-flip system has as time-reversible equilibria the Gibbs states that correspond to the interaction potential. (Spitzer's results required a certain uniqueness hypothesis that was later verified for a large class of systems by Liggett (1972).) This is illustrated in the following example:

Example 1. (The stochastic Ising model) We start with a pair potential with range R: let $J: (0,\infty) \to [0,\infty)$ be a non-increasing function such that $J(r) = 0$ for $r > R$, and then we define the energy at x in state ξ by

$$E_x(\xi) = -\sum_{y:\, y\neq x} \xi(x)\xi(y)J(\|x-y\|) \ .$$

We have used $\|\cdot\|$ for the usual Euclidean distance in $\mathbf{Z}^d$, and $\xi(x)$ stands for the value of the spin (+1 or −1) at x in the configuration ξ. We have restricted our attention to pair potentials whose strength decreases with distance merely for convenience. The assumption that J is non-negative is more significant. It ensures certain monotonicity properties that seem to be almost indispensable in the study of spin-flip systems.

Next we define a <u>Gibbs state with potential J</u> to be any probability measure μ on the Borel sets of Ξ such that for all $\eta \in \Xi$ and all $x \in \mathbf{Z}^d$,

$$\mu\,(\xi:\, \xi(x) = \eta(x) \mid \xi(y) = \eta(y) \text{ for } y \neq x) = Z^{-1}(\eta)\exp(-E_x(\eta)), \qquad (1)$$

where $Z^{-1}(\cdot)$ is the normalizing constant that makes the total conditional probability equal to one,

$$Z^{-1}(\eta) = \exp(-E_x(\eta)) + \exp(E_x(\eta)) \ .$$

To get a spin system that has all the Gibbs states with potential J as its time-reversible equilibria, we have many choices for the rates, but we choose $c_x(\xi) = \exp(E_x(\xi))$, which is one of the choices that makes the system symmetric with respect to interchange of + and − spins. For a detailed explanation as to why a system with these rates has time-reversible equilibria as claimed, see Spitzer's original paper, or

more recently, Ligget's book on interacting particle systems, both listed in the references.

Next we introduce a parameter $\varepsilon > 0$ (corresponding to a constant multiple of the temperature in statistical mechanics) which we call the noise parameter. Let

$$c_x^{\varepsilon}(\xi) = \exp(E_x(\xi)/\varepsilon),$$

which defines rates for a system corresponding to the interaction potential $J^{\varepsilon} = J/\varepsilon$. The idea is that as the noise decreases, the interaction becomes stronger.

One final parameter: let H be a real constant, which we call the bias parameter (the external field strength in statistical mechanics), and define flip rates for a biased system by

$$c_x^{\varepsilon,H}(\xi) = \exp(E_x(\xi)/\varepsilon)\exp(-H\xi(x)). \qquad (2)$$

These biased systems also have Gibbs states as their time-reversible equilibria. They are defined by replacing $E_x(\xi)$ in (1) by $(E_x(\xi)/\varepsilon) - H\xi(x)$.

The expression in (2) ensures that the flip rate at x is high if E_x would be lowered by a flip, and the flip rate is low if the energy would be raised. A quick glance at the expression for the energy shows that there is a positive contribution to it for each site y within range R of x such that the spins at x and y disagree. In biased systems, there is also a positive contribution to the energy if the spin at x disagrees with the sign of H. Conversely, sign agreement between such x-y spin pairs or between H and the spin at x contribute negatively to the energy. Thus the rates are set up so that the system tends to spend more time in configurations of low overall energy. In the unbiased case, there are two configurations of minimum energy, which we call ξ^+ and ξ^-, the two configurations in which all the spins have the same sign. These are called

ground states. In the biased case, only one of these states has minimum energy, namely the one that agrees in sign with H. Under certain circumstances which will be discussed in the next section, the extreme equilibria of the system become more and more concentrated near the ground states as the noise parameter decreases to 0. This might be expected from the form of the flip rates.□

One of the purposes in defining a Markovian system with Gibbs states as equilibria is that it makes available the tools and techniques of Markov processes to the study of statistical mechanics. Semigroups, generators, martingales, coupling methods, etc., can now be brought to bear on problems that originally had no time dependence. Furthermore, information about rates of convergence to equilibrium, dependence of the process on the initial state and other details about the time evolution can be translated into information about uniqueness and mixing properties of the Gibbs states. As a simple example of how this can work, we take one of the most basic questions in the study of Gibbs states, namely the question of whether there exists more than one Gibbs state corresponding to a given potential J. If the corresponding spin system is ergodic (or in other words, has a unique equilibrium), then one can conclude that there is only one Gibbs state for J, and if the system converges exponentially fast to this equilibrium, then the Gibbs state must have exponential mixing properties (see Liggett's book for these and other results).

To a certain limited extent, this carryover from the study of spin-flip to statistical mechanics has occurred. Some nice contributions to various questions are to be found in the work of Holley and Stroock (1976a and 1976b). There has also been some interesting work about the increase of free energy in the time evolution which adds to one's understanding of the concept of free energy in statistical mechanics —

see Holley and Stroock (1977). But these instances are the exceptions, rather than the rule. Most of what is known about the equilibrium behavior of the stochastic Ising model is derived from already known facts about Gibbs states. This is putting the cart before the horse if one is trying to find applications for spin-flip systems. (Of course, a great deal of interesting and important work has been done on other kinds of spin-flip systems which have nothing directly to do with statistical mechanics, such as those which are related to population or genetics models like the contact process or the voter model.)

Based on what has been said so far, the problem seems to be that most techniques for studying the time evolution of the stochastic Ising model depend on *a priori* information about the equilibria. It is proposed here that methods should be developed which deal directly with the time evolution. These methods should be robust, in the sense that they should apply to systems whose rates are qualitatively like the rates defined in Example 1, but which do not necessarily have the same precise form, systems which do not necessarily have Gibbs states as equilibria. Once such methods are developed, then it should be possible to isolate those essential properties of a time evolution that lead to "statistical mechanical-like" behavior (we will call this SML behavior for short), thus leading to a better understanding of time evolutions of actual statistical mechanical models.

In the next section, we will propose a program of research that has these goals in mind. We will do this by first defining a class of flip rates that we conjecture have the features necessary for SML behavior, then by giving some examples of such rates, and finally by listing some of the properties that we feel are a part of typical SML behavior. In the third section of the paper, we will "set a good example" by outlining a proof that a certain well-known system which is not a stochastic Ising model (the ma-

jority vote model) has SML behavior in one dimension for all sufficiently small values of the noise parameter ε. (The SML behavior I have in mind here is ergodicity, or as the people in statistical mechanics would say, "no phase transitions in one-dimensional, finite range systems".) This is a partial solution to an open problem of several years (one would like to remove the restriction that ε be small). Since this is only a research report, the proof will only be indicated for the nearest neighbor case. An extension to the arbitrary finite range case is forthcoming in a paper which is in preparation (Gray (1986)). It should even be possible to extend the proof to cover the wider class of systems defined in Section 2; the proof can certainly be adapted to cover one-dimensional stochastic Ising models. In Section 3 we will also discuss briefly an unexpected benefit of the proof, namely a sort of invariance principle for one-dimensional systems as $\varepsilon \to 0$. At the time of the writing of this report, I don't quite know what to make of this last material, but it is hoped that it has some consequences for the theory of statistical mechanics.

2. SML flip rates and SML behavior.

It seems (to the author at least) that there is nothing special about the exponential function in (1) and (2). Its presence is dictated by certain considerations from statistical mechanics, but from a purely dynamical viewpoint, it is hard to see why its use in the definition of the flip rates should be crucial. To best see what are the essential properties of this function, we consider a state in which there is a flat <u>interface</u> between the two types of spins. To simplify things, let $H = 0$. Let π be any $(d-1)$-dimensional hyperplane that does not contain any of the lattice points in $\mathbb{Z}^d$. Assume that a normal vector is attached to π to give it some orientation, and let ξ^π be the state in which the sites on the positive side of π have spin +1 and the sites on

the negative side have spin −1. In this configuration, the energy $E_x(\xi)$ at any x is less than or equal to 0. If we change the sign of the spin at some site x, then the energy at x becomes strictly positive. Thus, if for any configuration ξ, we define the configuration ${}^x\xi$ by

$$ {}^x\xi(y) = \xi(y) \quad \text{for } y \neq x, $$
$$ -\xi(y) \quad \text{for } y = x, $$

then the ratio $c_x(\xi^\pi)/\, c_x({}^x\xi^\pi)$ is strictly less than 1. When we introduce the noise parameter, we find that

$$ c_x^\varepsilon(\xi^\pi) \,/\, c_x^\varepsilon({}^x\xi^\pi) \to 0 \quad \text{as} \quad \varepsilon \to 0. \tag{3} $$

(Note that it is the ratio of the rates that is important, since we can always rescale time without changing the equilibrium behavior. Also note that (3) still holds for the biased rates $c_x^{\varepsilon,H}$.) Thus, the smaller the noise, the more the system attempts to maintain a flat interface where one exists. If we define $\varphi^\varepsilon(E) = \exp(E/\varepsilon)$, then for all $E > 0$, $\varphi^\varepsilon(E)/\varphi^\varepsilon(0)$ converges to ∞ as ε goes to 0, and it is this property of the exponential that implies (3). We feel that this property is also the reason that stochastic Ising models behave the way that they do, and that other non-decreasing positive functions φ^ε with this property would do just as well. This leads us to the following:

Definition of SML flip rates. Let $E_x(\xi)$ and $J(\,\cdot\,)$ be as in Example 1. We will say that a parameterized family $\{c_x^{\varepsilon,H}\,,\ \varepsilon > 0\}$ is a family of SML flip rates with potential J if there are positive, non-decreasing functions $\{\varphi^\varepsilon,\ \varepsilon > 0\}$ defined on $\mathbb{R}$ such that

$$ c_x^{\varepsilon,H}(\xi) = \varphi^\varepsilon(E_x(\xi))\exp(-H\xi(x))\,, \tag{4} $$

and such that for all $E > 0$,

$$\varphi^{\varepsilon}(E)/\varphi^{\varepsilon}(0) \to \infty \quad \text{as} \quad \varepsilon \to 0. \tag{5}$$

We further impose the regularity condition that the ratios $\varphi^{\varepsilon}(E)/\varphi^{\varepsilon}(0)$ be non-decreasing for $E > 0$ and non-increasing for $E < 0$ as ε decreases to 0, so that a decrease in the noise parameter corresponds to an increase in the strength of the potential. □

In our definition, we have retained the way in which the rates depend on H and we have not done anything to generalize the energy function. Presumably more generality is possible, but we feel that we have done enough violence to the *original* model already. Actually, the class of SML rates is even larger than it first appears, due to the flexibility with which φ_{ε} and J may be chosen. This point will be partially illustrated by the following examples.

Example 2. Perturbations of the stochastic Ising model. We simply let $\varphi(E)$ be some increasing positive function that is uniformly close to $\exp(x)$ and define $\varphi^{\varepsilon}(E) = \varphi(E/\varepsilon)$. It seems incredible, but it is true, that as soon as a perturbation like this is made, virtually all the facts known about the stochastic Ising model become open questions.

Example 3. Sums of Ising model rates. Let $\{a_x^{\varepsilon,H}\}$ and $\{b_x^{\varepsilon,H}\}$ be two sets of rates for stochastic Ising models, corresponding to two different potential energies, and define $c_x^{\varepsilon,H} = a_x^{\varepsilon,H} + b_x^{\varepsilon,H}$. As in Example 2, most of the usual results are no longer known to apply. For example, suppose both of the original systems have phase transitions (non-ergodicity for small ε). It is not known whether the hybrid system with rates $\{c_x^{\varepsilon,H}\}$ has a phase transition. (It is not immediately clear that this example

fits into our class of systems with SML flip rates. It turns out, however, that by choosing J and φ^{ε} properly, the rates in this example can always be realized as SML flip rates. I am indebted to R. Schonmann and J. Lebowitz for pointing out this example.)

Example 4. The majority vote model (continuous time). Define $J \equiv 1$ and

$$\varphi^{\varepsilon}(E) = \varepsilon \quad \text{if } E \leq 0$$
$$= 1 \quad \text{otherwise.}$$

With $H = 0$, the flip rate at a site x is 1 if $\xi(x) \neq \xi(y)$ for more than half the sites y within range R of x (these sites are called the neighbors of x). When the spin at x agrees with at least half the spins at neighbors of x, the rate is ε. This is the simplest continuous time example of SML rates, yet its behavior is only known in the one-dimensional case with $R = 1$, in which case it just happens to be a stochastic Ising model! □

We are also interested in models in discrete time. Such models have often been avoided by people interested in statistical mechanics, because there is no simple way to define such a model so that it has Gibbs states as equilibria, as one can in continuous time. From our point of view, however, they are just as worthy of study as their continuous time relatives.

In a discrete time model, there are no rates. Instead, if the systems is in state ξ_t at time $t \in \{0, 1, 2, \ldots\}$, there is a flip probability $c_x(\xi_t)$ that the spin at x changes sign during the next time unit, in which case it takes on the new value at time $t + 1$. During any single step of the time evolution, the flips are made independently at the different sites. We will use the same notation for flip probabilities as we use for flip rates (of course, flip probabilities must be between 0 and 1). An analogue to the notion of SML flip rates is:

Definition of SML flip probabilities. Let $E_x(\xi)$ and $J(\cdot)$ be as in Example 1. We will say that a parameterized family $\{c_x^{\varepsilon,H}, \varepsilon > 0\}$ is a family of SML flip probabilities with potential J if there are non-decreasing, positive functions $\{\varphi^\varepsilon, \varepsilon > 0\}$ defined on $(-\infty, 0]$ such that

$$c_x^{\varepsilon,H}(\xi) = \varphi^\varepsilon(E_x(\xi)) \exp(-H\xi(x)) \quad \text{if} \quad E_x(\xi) \le 0 \tag{4'}$$
$$= 1 - (\varphi^\varepsilon(-E_x(\xi)) \exp(H\xi(x))) \quad \text{otherwise,}$$

and such that for all $E < 0$,

$$\varphi^\varepsilon(E) \to 0 \quad \text{as} \quad \varepsilon \to 0. \tag{5'}$$

As before, we make a regularity assumption, namely that $\varphi^\varepsilon(E)$ be non-increasing for all $E < 0$ as ε decreases. Note that we have formulated (4') in such a way that for all states ξ, $c_x^{\varepsilon,H}(\xi) + c_x^{\varepsilon,H}({}^x\xi) = 1$. This is not a necessary assumption, but it does save us from being forced to pay attention to several annoying details that come up in discrete time models. We are now ready to define what we consider to be the simplest SML model of all:

Example 5. The majority vote model (discrete time). Define $J \equiv 1$ and

$$\varphi^\varepsilon(E) = \varepsilon \quad \text{for all} \quad E \le 0.$$

Thus the spin at x changes with probability ε if it is in agreement with at least half the spins at neighbors of x, otherwise it changes with probability $1-\varepsilon$. Until now, nothing was known about this model for small $\varepsilon > 0$, even in the one-dimensional, nearest neighbor case (compare with the continuous time model above). It seems that the simpler the model gets, the less one knows about its SML behavior. In the next section,

we will alleviate this situation a little by discussing the very simplest of all models, namely the one–dimensional nearest neighbor discrete time majority vote model. For this model, we will sketch the proof that the equilibrium is always unique for sufficiently small $\varepsilon > 0$. Our methods can be adapted to the arbitrary finite range case in both discrete and continuous time (see Gray (1986)) , and probably even to the general one–dimensional SML model. □

We conclude this section with a discussion of what, in our view, constitutes basic SML behavior. We have picked out a few properties that we feel characterize the behavior of a fairly general class of statistical mechanical systems. We are aware that there are many interesting examples of systems which do not conform to this picture (which is one reason that they are so interesting). For example, we are ignoring the variety of behavior that occurs in systems with infinite range interactions, or in systems with infinitely many "spin" values possible at each site.

We start by defining the magnetization at x, which is merely the expected value of $\xi_t(x)$. For a fixed family of SML flip rates, this quantity depends on the initial state ξ_0, the site x, the time t, and on the parameters ε and H. We will only be interested in the initial states ξ^+, ξ^-, and ξ^π defined earlier. Let $M_t^+(\varepsilon, H)$, $M_t^-(\varepsilon, H)$, and $M_t^\pi(x, \varepsilon, H)$ be the corresponding magnetizations at x.

Note that by translation invariance, $M_t^+(\varepsilon, H)$ and $M_t^-(\varepsilon, H)$ do not depend on x, and that by symmetry, $M_t^+(\varepsilon, H) = -M_t^-(\varepsilon, H)$. It follows from our assumption that J be a non–negative function that the flip rates are "attractive" (see Liggett's book in the references), which in turn implies that $M_t^+(\varepsilon, H)$ decreases as $t \to \infty$. We will write $M_\infty^+(\varepsilon, H)$ and $M_\infty^-(\varepsilon, H)$ for the limits of $M_t^+(\varepsilon, H)$ and $M_t^-(\varepsilon, H)$ as $t \to \infty$. It is not clear that $M_t^\pi(x, \varepsilon, H)$ has a limit as $t \to \infty$, so we will use the notation $M_\infty^\pi(x, \varepsilon, H)$ for the lim sup if x is on the positive side of π, otherwise it will stand for the lim inf.

Using the notion of magnetization, we can now list four kinds of behavior that we conjecture are exhibited by all systems with SML flip rates in which the range R is finite:

SML Behavior.

(i) Monotonicity in the parameters: for any fixed potential J and for any time $0 \le t \le \infty$, the magnetizations $M_t^+(\varepsilon, H)$, $M_t^-(\varepsilon, H)$, and $M_t^\pi(x, \varepsilon, H)$ are monotone functions of the range R, the dimension d, the noise ε and the bias H. Note: the monotonicity in H — like the monotonicity in t —is a consequence of the fact that the kinds of flip rates that we are concerned with are attractive. Monotonicity in the remaining parameters does not come so easily.

(ii) No spontaneous magnetization in biased systems: If $H \neq 0$, then the limits $M_\infty^+(\varepsilon, H)$, $M_\infty^-(\varepsilon, H)$, and $M_\infty^\pi(x, \varepsilon, H)$ are all equal for all $\varepsilon > 0$.

(iii) No spontaneous magnetization in one dimension: If $d = 1$, then the limits $M_\infty^+(\varepsilon, 0)$, $M_\infty^-(\varepsilon, 0)$, and $M_\infty^\pi(x, \varepsilon, 0)$ are all equal to 0 for all $\varepsilon > 0$. Furthermore, the convergence is exponentially fast in t.

(iv) Spontaneous magnetization in dimensions greater than 1: If $d > 1$, then $M_\infty^+(\varepsilon, 0) > M_\infty^-(\varepsilon, 0)$ for all sufficiently small $\varepsilon > 0$.

(v) No tight interface in one or two dimensions: If $d = 1$ or 2, then $M_\infty^\pi(x, \varepsilon, 0) = 0$ for all $\varepsilon > 0$ and all sites x.

(vi) Existence of tight interfaces in dimensions greater than 2: If $d > 2$, then for all sufficiently small $\varepsilon > 0$, $M_\infty^\pi(x, \varepsilon, 0)$ is dependent on x. □

This list is based on known behavior of Gibbs states. There is a great deal that can be said about these properties — a proper discussion of them is beyond the scope of this report. The interested reader should consult Liggett's book, and also the very readable monograph of Kindermann and Snell (1980). We will merely note here that we

feel that property (i) is of the most immediate importance. It corresponds to various correlation inequalities in statistical mechanics which are used over and over again (for example, the monotonicity in ε corresponds to the "Griffith's inequalities"). In particular, it allows one to talk about "critical" values of the various parameters.

3. SML behavior in the one–dimensional majority vote model.

In this section, we will outline the proof that property (iii) of SML behavior holds for the one–dimensional nearest neighbor discrete time majority vote model, at least for all sufficiently small $\varepsilon > 0$. If one could prove that property (i) also holds for this model, then we could remove the restriction that ε be sufficiently small.

Theorem. Let (ξ_t), $t = 0, 1, 2, \ldots$, be the discrete time system in Example 5, with $H = 0$ and $R = 1$. There exists $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0]$, $M_t^+(\varepsilon, 0)$, $M_t^-(\varepsilon,0)$ and $M_t^{\pi}(x, \varepsilon, 0)$ converge exponentially fast to 0 as $t \to \infty$. (As noted earlier, this result will be extended, using essentially the same proof, to all finite range majority vote models in discrete and continuous time in Gray (1986).)

Outline of Proof. This is not a complete proof. We will be very detailed for certain parts of the argument and very sketchy in others. Our choice will be based on whether we consider the part of the argument to contain new ideas or not. See Gray (1986) for more details.

We start by giving an explicit construction of the process. For integer times $t \geq 0$ and sites $x \in \mathbf{Z}$, let $e(x, t)$ be i.i.d. random variables, with distribution determined by

$$P(e(x, t) = +1) = P(e(x, t) = -1) = \varepsilon \quad \text{and} \quad P(e(x, t) = 0) = 1 - 2\varepsilon.$$

We will construct the entire system of processes (one process for each initial state) on the probability space associated with the random variables $e(x, t)$. Assuming that an initial state ξ_0 has been chosen, then we inductively define ξ_t for times $t > 0$ by

$$\xi_t(x) = e(x, t) \qquad \text{if} \quad e(x, t) \neq 0$$

$$= \text{the majority spin value in}$$

$$\text{the set } \{\xi_{t-1}(y),\ |y - x| \leq 1\} \qquad \text{if} \quad e(x, t) = 0.$$

Thus, to determine the value at x at time t, a vote is taken at time t − 1 at the sites within distance R = 1 of x (including x itself). The outcome of this vote determines $\xi_t(x)$, provided the "error variable" e(x, t) is 0. If this error variable is not 0, then the vote is ignored, and $\xi_t(x)$ is determined by e(t, x). This clearly agrees with the description of the system in Example 5. Note that if e(x, t) happens to agree with the outcome of the majority vote, then the value of $\xi_t(x)$ is the same as if e(x, t) had been 0. We will say that an error occurs at x at time t if e(x, t) is not 0 and if e(x, t) does not agree with the outcome of the majority vote taken at time t − 1 at the sites within distance 1 of x. Thus, an error occurs independently with probability ε at any given point (x, t) in space–time.

We can now describe the general strategy of the proof. Let ξ_t^+ and ξ_t^- be the processes with initial states ξ^+ and ξ^- respectively (these are the "all +1's" and "all −1's" initial states defined earlier). As noted in the previous paragraph, these two processes are jointly defined on the same probability space, so we can define the joint process $\gamma_t = (\xi_t^+, \xi_t^-)$. It is easy to check that $\xi_t^+(x) \geq \xi_t^-(x)$ for all t and x, so that $\gamma_t(x)$ can only take three values: (+1, +1), (+1, −1), and (−1, −1). We will call these values +, 0, and − respectively. Note that

$$\tfrac{1}{2} P(\gamma_t(0) = 0) = M_t^+(\varepsilon, 0) - M_t^-(\varepsilon, 0).$$

Since M_∞^π is sandwiched between M_∞^+ and M_∞^-, it is enough to prove that

For all sufficiently small $\varepsilon > 0$, $P(\gamma_t(0) = 0) \to 0$ exponentially fast as $t \to \infty$. (6)

(Of course, the exponential rate is allowed to depend on ε.)

The basic idea behind the proof of (6) is simple. Since there is a positive probability ε of an error occurring at any given site, and since these errors occur independently, during each time step there will be infinitely many places where at least 2 consecutive sites have simultaneous errors of the same sign. Such a block will show up as a string of consecutive +'s or consecutive −'s in the process γ_t. As long as no errors occur near the endpoints of such a string, no 0's can appear within the block, so we have a somewhat stable string of non-zeroes. This string of non-zeroes will change in size as errors occur near its endpoints. If ε is small, the most likely event is that a single error will eventually occur near an endpoint (rather than simultaneous multiple errors), and the symmetry of the model ensures that, for strings of length 4 or more, such an error is equally likely to result in an increase of the size of the string by 1 unit as it is to result in a decrease by 1 unit. One should try to envision infinitely many such strings of non-zeroes whose endpoints are essentially doing independent simple random walks. Of course, this picture is only approximate, for several reasons. First, simultaneous multiple errors do occur. Secondly, a single error can simultaneously affect both endpoints of a short string (of length 3 or less). Finally, when the endpoints of different strings get close to each other, various movements can occur which are not due to errors. If we could ignore these difficulties, the rest of the proof would be easy. Simple computations would show that with infinitely many strings of non-zeroes appearing in each time step, all of which change size like random walks, the probability would go to 1 exponentially fast as $t \to \infty$ that a given site be contained in such a string, implying (6).

The crux of our proof is to show that for small enough ε, the behavior of strings of non-zeroes in the process γ_t is sufficiently like the naive description given in the preceding paragraph to obtain (6). We begin by investigating the behavior of a single

string of +'s. Let $\zeta_t^n = (\xi_t^n, \xi_t^+)$, where ξ_t^n is the process with initial state defined by

$$\xi_0^n(x) = +1 \quad \text{if} \quad |x| \le n$$
$$= -1 \quad \text{otherwise.}$$

As with the process γ_t, $\zeta_t^n(x)$ can take three values, +, 0, and −. It starts with the value + at x inside the interval [−n, n] and 0 at all other x. We wish to analyze the movement of the endpoints of this interval of non-zero values due to two different causes, namely due to errors and due to collisions with other intervals of non-zero values.

There is one more point that we wish to make before we start our analysis. It will be seen when we make our estimates that events that have probability $O(\varepsilon^3)$ can be completely ignored. Thus we do not need to worry about the occurrence of clusters of more than two errors in space-time. More precisely, we will eventually be working on some block of space-time $[-N, N] \times [0, T]$, where N is $O(1/\sqrt{\varepsilon})$ and T is $O(\varepsilon^{-2})$. It follows that the probability that somewhere in the $[-N, N] \times [0, T]$ rectangle, three or more errors occur in any space-time block of size 10×10, say, will be small as $\varepsilon \to 0$. One should keep this in mind when checking the various claims that will be made — in particular, we will never concern ourselves with checking for the effects of clusters of three or more errors.

Movement of the endpoints due to errors. The most important effects arise from a single error. Let us assume that at time 1 a single error occurs at some site x within the interval $[n-1, n+2]$, and that no other errors occur anywhere near the site n for several time units. Also, for simplicity, assume that n is at least 2. If $x \le n$ (the error will produce a − at x in this case) then it is easy to check that at time 2, there will be a 0 at the site n in the process ζ_t, while all other sites near n will retain their original value. Thus, the right endpoint of the interval of +'s has

moved one unit to the left. Similarly, if $x > n$, and if the error at x produces a $-$, then at time 2 there will be a $+1$ at the site $n + 1$ and the endpoint will have moved 1 unit to the right. Examples of this type of movement at the right endpoint are illustrated below in "Movie #1". An analogous description applies to the left endpoint of the string, with all the directions reversed.

It is possible to give a similar but more involved analysis of the possible effects of two simultaneous errors. The result is that there is again a balance between movement to the right and movement to the left. However, in the case with $R = 1$, there is no need to make the effort. We will find that the direct effects of two simultaneous errors are negligible, as far as the movement of endpoints is concerned. Of course, the occurrence of two simultaneous errors at some distance from the endpoint can indirectly affect the movement of the endpoints through eventual collisions, as will be seen below. Also, remember that we are ignoring clusters of three or more errors, as explained earlier.

Thus we see that, at least to the extent that we can ignore interactions with other strings and the occurrence of two simultaneous errors near the endpoints of the string of +'s, they are seen to behave like independent symmetric random walks. Each endpoint moves 1 unit in either direction with probability ε during each unit of time. Of course, the same is true of a string of $-$'s. (The situation in the general finite range case is very similar. The proper statements are somewhat more difficult to formulate and require more machinery to justify, but behind all the technicalities lie just two factors: symmetry and the fact that we can ignore clusters of more than R errors.)

Collisions. Let us suppose that we have in some way defined the positions of the endpoints of an interval of +'s or $-$'s up to some time t. Let E_t^ℓ and E_t^r stand for the

positions of the left and right endpoints at time t. Initially, each endpoint behaves like a random walk as described above. Now we consider the problem of what happens when one endpoint comes close to the endpoint of another interval. For example, suppose that we are dealing with a string of +'s, and that somewhere between E_t^ℓ and E_t^r, two errors that produce −1's occur simultaneously at neighboring sites. This produces an "inner" string of −'s which will move about with the same kind of random walk behavior exhibited by the "outer" string. After some time, its right endpoint may come within three units of E_t^r. Let the three sites that separate E_t^r from the right endpoint of the interval of −'s be called x, y, and z. The process ζ_t^n has +'s at these sites. If an error now occurs at y that produces a −, then after the dust settles (in two time units), the process ζ_t^n will have values −, 0, and 0 at the sites x, y, and z respectively (see Movie #2; note in the movie that a single inner string can be the cause of two collisions — once when the string first gets close to the endpoint, and then later when the inner string becomes too short). Such collisions with inner strings cause us a few problems. One is that the value + at the end of the outer string seems to have changed to −. Up until now, E_t^r has always marked a dividing line between +'s and 0's in the process ζ_t^n. We will be forced to drop this interpretation and will have to move E_t^r two sites to the left when such a collision occurs (for a more precise definition, see below). Our interpretation is now that the endpoints mark positions between which there are <u>essentially</u> no zeroes in the process ζ_t^n. We will be more explicit later.

There is still another more serious difficulty that the collision causes for us. The endpoint has moved 2 units to the left, but only a single error has occurred. One may say that the error caused a movement of 1 unit and that the collision caused a further movement of 1 unit. This introduces a drift into the random walk behavior of the endpoints. Of course, one might hope that this drift is balanced out by the possibility

that a symmetrically opposite collision occurs on the right side of E_t^r. That is, two consecutive errors that produce +'s could occur to the right of E_t^r, then this string of +'s could eventually come within 3 units of the endpoint, then an error could occur at the middle site separating the string of +'s from E_t^r, and then E_t^r would move at least 2 units to the right. This is probably correct, but rigorous proof seems very difficult. For example, it is hard to say anything about the independence of increments due to collisions in the movement of E_t^r. These collisions are events that involve large chunks of space–time and are highly dependent. They constitute the main problem that has prevented any progress on this model for so many years.

The way out of the difficulty is to prove that the effects of collisions can be ignored when ε is sufficiently small. We will show that when ε is sufficiently small, E_t^ℓ and E_t^r behave enough like a random walks so that the heuristics given near the beginning of the proof can be carried out. Before we can do this, we must get the right definition of E_t^ℓ and E_t^r. But first, here are the movies promised earlier.

Note: In all movies, time progresses down the page.

Movie #1

. . . + + + + + + + + 0 0 0 0 0	These are the sites near the right endpoint.
. . . + + + + + + + + 0 ⊕ 0 0 0	An error (circled) occurs . . .
. . . + + + + + + + + + 0 0 0 0	moving the endpoint to the right.
. . . + + + + + + + + ⊖ 0 0 0 0	Another error occurs . . .
. . . + + + + + + + + 0 0 0 0 0	and the endpoint moves left.

Movie #2

. . . + + + + + + + + + 0 0 0 0	The sites near the endpoint.
. . . + + + − − + + + + 0 0 0 0	An inner string of −'s appears.
. . . + + + − − − + + + 0 0 0 0	Later, the inner string has moved.

. . . + + + − − − + ⊖ + 0 0 0 0	An error occurs and . . .
. . . + + + − − − − + 0 0 0 0 0	in two time steps . . .
. . . + + + − − − − 0 0 0 0 0 0	the 0's have gained two sites.
. . . + + + + − − − 0 0 0 0 0 0	Later, the string of −'s shortens.
. . . + + + + + − ⊕ − 0 0 0 0 0 0	An error occurs and . . .
. . . + + + + + + 0 0 0 0 0 0 0 0	two time units later, the 0's gain two more sites

Definition of the endpoints. We will explain how to construct the right endpoint process E_t^r. The left endpoint is defined analogously. We start by defining $E_0^r = n + {}^1/_2$. This position marks the midpoint between the rightmost + and its neighboring 0 in the initial state of the process ζ_t^n. Next let us assume inductively that E_t^r has been defined to be some half integer value for some $t \geq 0$. We will include in our inductive assumption the requirement that the value assigned by ζ_t^n to the site $E_t^r - {}^3/_2$ is nonzero and agrees either with the value assigned to the site $E_t^r - {}^1/_2$ or with the value assigned to $E_t^r - {}^5/_2$. Thus we assume that there is an interval of at least two consecutive +'s or two consecutive −'s which includes the site $E_t^r - {}^3/_2$. Note that this assumption is satisfied at $t = 0$ as long as $n > 0$. We call the common value at the two consecutive sites the (right) endpoint sign. In the case that the endpoint sign does not agree with the value assigned by ζ_t^n to the site $E_t^r - {}^1/_2$, we make the further assumption that value assigned to the site $E_t^r + {}^1/_2$ does agree with the endpoint sign. To make it easier to visualize all this, we note that if the endpoint sign is +, then according to the assumptions made, the values at the four sites in $[E_t^r - 3, E_t^r + 1]$ are either ? + + ? or + + ? + , where the symbol "?" stands for any of the three possible values +, −, or 0.

We will now define the increment $E_{t+1}^r - E_t^r$. Following our previous discussion,

this increment is built from two pieces, the first of which may be considered to be the result of errors, and the second the result of collisions. The part due to errors will be called X^r_{t+1} and the part due to collisions will be called Y^r_{t+1}, with the increment being the sum of the two.

We start by defining X^r_{t+1}. Consider the values of the error variables e(x, t+1) for sites x in the interval $[E^r_t - 2, E^r_t + 2]$. If all of these have the value 0, then set X^r_{t+1} equal to 0. Next we define X^r_{t+1} in the case that exactly one of these error variables is not 0. Let x' be the site where this occurs. If $x' < E^r_t$ and if e(x', t+1) does <u>not</u> agree with the endpoint sign, then let X^r_{t+1} equal -1. If $x' > E^r_t$ and if e(x', t+1) <u>does</u> agree with the endpoint sign, then let X^r_{t+1} equal +1. In all other cases, let X^r_{t+1} equal 0. It will be noted that this definition of X^r_{t+1} exactly corresponds to the jumps described in the discussion of movement of the endpoint due to the occurrence of a single error.

Now that X^r_{t+1} has been defined, we let Y^r_{t+1} be the largest integer ≤ 0 such that the choice

$$E^r_{t+1} = E^r_t + X^r_{t+1} + Y^r_{t+1}$$

satisfies the inductive assumption at time t+1. (If no such integer Y^r_{t+1} exists, we let $E^r_u = -\infty$ for all $u \geq t+1$. It is easy to see that if $\varepsilon > 0$, such an integer will exist with probability one at all times.) We may describe our definition of the movement of the right endpoint in this way: first move the endpoint in accordance with the description given in the paragraphs which discuss movement due to a single error; then, if necessary, move it to the left until the values assigned by ζ^n_{t+1} to the four sites in $[E^r_{t+1} - 3, E^r_{t+1} + 1]$ are $?++?$ or $?--?$ or $++?+$ or $--?-$, with "?" representing any of the three possible values 0, + or − as before. A little investigation of the possibilities will reveal that the extra movement to the left represented by Y^r_{t+1} can only occur when there is a string of sites close to the endpoint which have the opposite value

from the endpoint sign (a collision occurs), or when multiple simultaneous errors occur near the endpoint.

Let E_t^ℓ, X_t^ℓ and Y_t^ℓ be defined in the obvious way. Then it is easy to check inductively that

All sites in the interval $[E_t^\ell + 1, E_t^r - 1]$ are assigned non-zero values by ζ_t^n. (7)

In fact, the statement in (7) is essentially true for the larger interval $[E_t^\ell, E_t^r]$: any zeroes that occur at the end sites of this interval are temporary (they last only one time unit) because of the assumption made about the signs at sites near the endpoints. (Such zeroes can arise at, say, the right endpoint, when the right endpoint moves one unit to the right due to an error that occurs at the site $E_t^r + 3/2$ at time t+1.) Thus for all practical purposes, we may consider the interval $[E_t^\ell, E_t^r]$ to be free from 0's in the process ζ_t^n.

Inner strings. One of the goals in this proof is to obtain good upper bounds on the amount that the endpoint moves due to collisions with inner strings. Thus we need to analyze the way in which these strings move about and interact with one another. We will separate the inner strings into two types. The first type, called a <u>simple string</u>, is an inner string which has never joined together with another inner string which has the same sign and which was originally disjoint from it. When two simple strings with the same sign join together, the larger string formed is called a <u>complex string</u>. We will also insist that the occurrence of multiple simultaneous errors not figure into the movements of the endpoints of a simple string. If more than one error occurs simultaneously within four sites of either side of the endpoint of a simple string, it becomes a complex string.

Let us describe the typical life of a simple string. It is born when two simultaneous errors with the same sign occur at adjacent sites at some time t in the interval $[E_t^\ell, E_t^r]$, forming a block of two neighboring sites which are assigned the same non-zero value by ζ_t^n. As single errors occur near the endpoints of this block, it grows or shrinks as described in the paragraphs on the movement of endpoints due to errors. If at any time one of these endpoints comes too close to another block of sites containing the same sign as the endpoint of the simple string, or if multiple simultaneous errors occur near one of these endpoints, then its life as a simple string ends and it becomes a complex string (see Movie # 3 below). It can also happen that an "inner, inner" string is formed within the first simple string, with the opposite sign. If, say, the right endpoint of this inner, inner string eventually joins up with the right endpoint of the simple string, then we will consider the simple string to have shrunk by an amount equal to the width of the inner, inner string (move the right endpoint of the simple string to coincide with the left endpoint of the inner, inner string). The inner, inner string then "breaks out" of the simple string and is then absorbed, so to speak, by whatever is outside of the original simple string (see Movie #4). Note that we are treating simple inner strings differently here than we treated the large interval $[E_t^\ell, E_t^r]$. A simple inner string always retains its integrity as a string of +'s or a string of −'s. It does not change its endpoint sign like the interval $[E_t^\ell, E_t^r]$ does. Incidentally, an inner, inner string is considered to be a separate inner string itself, which may be either simple or complex, and which may cause its own collisions with the endpoints E_t^ℓ and E_t^r. Once it breaks out, it either disappears as a string, or joins together with another string with the same sign, so in any case, it will at that time cease its life as a separate string. To summarize, the endpoints of simple strings move in two ways: the usual random walk kind of movement due to single errors, and larger jumps that

shrink the size of the simple string and allow inner, inner strings to break out.

Movie #3

. . .+++++++++++++. . .	Some sites in the interval $[E_t^\ell, E_t^r]$.
. . .+++++−−++++++. . .	An inner string is born.
. . .+++−−−+++++++. . .	The inner string has grown and moved.
. . .+++−−−+++−−++. . .	A second inner string appears.
. . .+++−−−+⊖+−−++. . .	An error (circled) occurs between the strings . . .
. . .+++−−−−−−−−++. . .	and a complex string is formed in 2 time units.

Movie #4

. . .+++++++++++++. . .	Some sites in the interval $[E_t^\ell, E_t^r]$.
. . .+++++−−++++++. . .	An inner string appears and . . .
. . .+++−−−−−−−−++. . .	grows after some time has passed.
. . .+++−−−+++−−++. . .	An inner, inner string appears.
. . .+++−⊕−++−−−++. . .	An error (circled) occurs, allowing. . .
. . .+++++++−−−++. . .	the inner, inner string to break out, causing the string of −'s to be diminished after 2 time units.

We are interested in upper bounds on the number of times that a collision occurs between one of the endpoints E_t^ℓ or E_t^r and one of the inner strings. We will be able to estimate the probability that such a collision occurs when the inner string remains simple, but estimates for complex strings are troublesome (they <u>can</u> be done, but we prefer not). We find that it is sufficient to be quite crude with complex strings: it is enough to note that the number of collisions between complex strings and E_t^ℓ or E_t^r is bounded by the twice number of times that a new complex string is formed (recall from Movie #2 that an inner string can cause two collisions). In other words, once a complex string is first formed, we will count it as if it has already collided twice

with E_t^ℓ or E_t^r, whether it actually eventually does so or not. The formation of complex strings in the interval $[E_t^\ell, E_t^r]$ is rare enough that this overestimate does not get us into trouble. Note that we do not need to keep track of instances where simple strings join up with complex strings, because such events do not increase the number of complex strings. We simply count complex strings once they are formed, and then we ignore them. Note also that we do not exclude the possibility that the same set of simultaneous errors that destroyed one simple string could also start an new simple string. Our estimates will allow for this possibility. To summarize, we have the following:

> Assume that $E_t^r > E_t^\ell + 2$ for all $t \in [0, T]$. Then the number of collisions between inner strings and E_t^ℓ or E_t^r during the time interval $[0, T]$ is bounded above by twice the number of simple inner strings that collide with E_t^ℓ or E_t^r during $[0, T]$ plus twice the number of collisions between two disjoint simple inner strings during $[0, T]$ plus twice the number of simple strings that become complex through the occurrence simultaneous errors. (8)

Thus, according to (8), we only need to work with simple strings in our estimates. This fact is quite useful, because for all practical purposes, we can treat simple strings as if their endpoints moved like random walks. The deviation from random walk behavior occurs when an inner, inner string breaks out. This causes a decrease in the size of the simple string, making it harder for it to collide with another disjoint simple string or with the endpoints E_t^ℓ or E_t^r. If we ignore such decreases, we are only making it easier for complex strings to be formed and for collisions to occur between

simple strings and E_t^ℓ or E_t^r. Thus we are justified in our estimates in treating simple strings as if the movements of their endpoints were only of the random walk type that results from single errors.

Estimates. Our goal is to show that the endpoint processes E_t^ℓ and E_t^r behave like random walks for small $\varepsilon > 0$. Define

$$\sum_{s \le t} Y_s^r = C_t^r \qquad \text{and} \qquad \sum_{s \le t} X_s^r = B_t^r$$

so that $E_t^r - E_0^r = B_t^r + C_t^r$. The increments X_t^r were constructed to be independent identically distributed symmetric random variables, so the B_t^r process is just a symmetric random walk. Thus we would like to show that the decreasing process C_t^r is negligible in comparison to B_t^r as $\varepsilon \to 0$ (and similarly for the left endpoint process). Recall that Y_t^r is 0 unless there is some movement of the right endpoint E_t^r which is not due to the occurrence of a single error. Multiple simultaneous errors near E_t^r and collisions between an inner string and the right endpoint can cause Y_t^r to be negative. When such events occur, $-Y_t^r$ measures the distance that the right endpoint must move toward the left in order that the assumption about the values assigned by the process ζ_t^n to sites near the right endpoint remain valid. This distance is roughly the distance to the nearest pair of neighboring sites that are both assigned the same non-zero value. Recall that none of the sites in the interval $[E_t^\ell + 1, E_t^r - 1]$ can be assigned the value 0, so as long as E_t^r remains larger than $E_t^\ell + 2$, the sites in the interval $[E_t^r + Y_t^r, E_t^r - 1]$ must contain alternating +'s and −'s. It is not hard to check that intervals of alternating +'s and −'s of length N in the process ζ_t^n can only be produced by clusters of at least $N/2 - 1$ errors occurring near each other in space–time. As mentioned earlier, we can assume that clusters of three or more errors do not occur (they can be ignored if ε is small), so it is safe to assume that $-Y_t^r$ is always less than or equal to 6. If

we combine this bound with (8), we obtain the following, which is valid for small $\varepsilon>0$:

Assume that $E_s^r > E_s^\ell + 2$ for all $s \in [0, t]$. Then

$$
\begin{aligned}
C_t^r \;\le\; & 6\,[(\text{the number of collisions between the right endpoint } E_s^r \text{ and} \\
& \qquad \text{inner strings during } [0, t]) \;+ \\
& \quad (\text{the number of times } s \in [0, t] \text{ that two simultaneous} \\
& \qquad \text{errors occur near } E_s^r)] \qquad (9) \\
\le\; & 12\,[(\text{the number of simple inner strings that collide} \\
& \qquad \text{with the right endpoint } E_s^r \text{ during } [0, t]) \;+ \\
& \quad (\text{the number of complex strings formed between} \\
& \qquad E_s^\ell \text{ and } E_s^r \text{ at some time } s \in [0, t]) \;+ \\
& \quad (\text{the number of times } s \in [0, t] \text{ that two simultaneous} \\
& \qquad \text{errors occur near } E_s^r)].
\end{aligned}
$$

We will estimate the expected values of each of the three terms on the right of (9). It is easy to see that

The expected number of times $s \in [0, t]$ that two simultaneous errors occur near E_s^r is $O(t\varepsilon^2)$, (10)

provided we understand "near" to mean "within some fixed number of sites". Estimates for the other two terms will take more effort. Let

$F(x, s)$ = the event that $x \in [E_s^\ell, E_s^r]$ and $e(x, s) = e(x + 1, s) \neq 0$.

In other words, $F(x, s)$ is the event that an inner string of two +'s or two −'s starts at time s at the sites x and $x + 1$. We are ignoring strings that start in other ways,

since all other ways require clusters of three or more errors, which we have claimed all along are negligible. We are also including some situations in which a true string is not formed, namely those cases where the signs of the error random variables e(x, s) and e(x + 1, s) agree with the values of the process ζ^n_{s-1} at the sites x and x + 1. This overcounting of strings will only increase our estimates, so it is justified. Next let

G(x, r, s) = the event that F(x, r) occurs , and the resulting inner string eventually collides <u>as a simple string</u> at time s with the right endpoint E^r_s

H(x, r, s) = the event that F(x, r) occurs , and the resulting inner string becomes a complex string at time s, either by colliding with another simple string or through the occurrence of multiple simultaneous errors near one of its ends.

I(N, t) = the event that $[-N, N] \supset [E^\ell_s, E^r_s]$ and $E^r_s \geq E^\ell_s + 2$ for all $s \in [0, t]$.

According to (9) and (10),

$$E(C^r_t;\, I(N, t)) \leq 12 \sum_{r<s<t} \sum_{x \in [-N, N]} \left(P(G(x, r, s)) + P(H(x, r, s)) \right) + O(t\varepsilon^2) \qquad (11)$$

We will first obtain an estimate for P(H(x, r, s)). There are two ways in which H(x, r, s) can occur. One is that an inner string appears at the sites x and x + 1 at time r, this inner string survives as a simple string until time s, and then at time s, two simultaneous errors occur near one of its ends. The probability that the inner string appears is ε^2. The length of a simple inner string behaves essentially like a symmetric random walk which starts at 2 and has an absorbing barrier at 1, and which has jumps at rate 2ε. Thus the probability that it survives for s − r time units is bounded above by a constant times $1/\sqrt{(s-r)\varepsilon}$. The probability that two errors occur simultaneous near one of its ends at time s is a constant times ε^2. There is enough

independence around so that we can multiply these probabilities, yielding a bound of $C(\sqrt{(s-r)\varepsilon}\,)^{-1}\varepsilon^4$ for the probability that H(x, r, s) occurs in this way, where C is some constant independent of ε. The second way that H(x, r, s) can occur is that the inner string appears at time r and survives for s − r time units, as before, and then a second inner string appears at time s at a distance d from the first inner string, and the two inner strings collide <u>as simple strings</u> some time later. We claim that the probability that this later collision occurs is bounded by a constant times d^{-1}. Again there is enough independence so that we are justified in multiplying probabilities, so we obtain a bound of $C\varepsilon^4(d\sqrt{(s-r)\varepsilon})^{-1}$, where the factor of ε^4 comes from the probabilities of the appearances of the two inner strings, and the factor of $(\sqrt{(s-r)\varepsilon})^{-1}$ arises as before. Of course, we have ignored the possibility that the collision occurred with a second inner string which appeared <u>before</u> time s. However, this only introduces a factor of 2 in our estimates (the collision will be counted in H(x', r', s) for some other site x' and some time r' < r), so we will not worry about it. Since all inner strings must appear somewhere in the interval [−N, N], d must be less than 2N. Summing over all such d we find that P(H(x, r, s)) is $O((s-r)^{-1/2}\varepsilon^{7/2}\ln N)$, so that

$$\sum_{r \le s \le t} \sum_{x \in [-N, N]} P(H(x, r, s)) \quad \text{is} \quad O(t^{3/2}\varepsilon^{7/2} N \ln N). \tag{12}$$

We will now justify our claim about the probability of collision being a constant times d^{-1}. In order to picture the situation, let us assume that the second inner string appears to the right of the first. We will concentrate on the movements of the right endpoint of the first inner string and both endpoints of the second inner string. They essentially all move like independent random walks until two of them come close enough together to collide. If the two endpoints of the second inner string come too

close to one another, the second inner string will disappear and no collision between the two inner strings will occur. Thus the two inner strings can only collide if the right endpoint of the first inner string and the left endpoint of the second inner string come close to each other before the two endpoints of the second inner string get too close together. This is a variation on the classical gambler's ruin problem: the right endpoints of the two strings play the roles of the absorbing barriers, and the left endpoint of the second inner string plays the role of the gambler's fortune. This comparison makes the claim quite plausible. One can easily rigorously justify the claim by using the same kind of martingale arguments that one uses in the gambler's ruin problem. In fact, with a little work, one can prove the following: let $x_1(t)$, $x_2(t)$ and $x_3(t)$ be martingales with respect to the same sequence of σ-algebras (the martingales are not assumed to be independent). Assume that $x_1(0) = x_2(0) = x_3(0) = 0$, and for simplicity also assume that the increments of each of the three martingales are bounded in size by some fixed constant. Finally assume that the expected value of either $x_1(t)^2$ or $x_2(t)^2$ grows at a non-zero rate. Let τ_1 be the first time t that $x_1(t) > x_2(t) + 1$ and let τ_2 be the first time t that $x_2(t) > x_3(t) + d$. Then $P(\tau_1 > \tau_2) < Cd^{-1}$ for some constant C independent of d.

We now turn to the estimate of P(G(x, r, s)). The event G(x, r, s) occurs as follows. First, the inner string appears at x and x + 1 at time r. This happens with probability ε^2. Then the two ends of this inner string move about like symmetric random walks, jumping at rate ε. At time s, the right one of these two random walks collides with E^r_s. The two endpoints of the inner string must not collide during [r, s], since the inner string would disappear if they did. Thus we can consider them to be moving like independent random walks during that time. These two random walks are also moving independently of the process B^r_u, which is itself moving like a random walk. If we

could identify E_u^r with B_u^r, then we would again have a gambler's ruin type problem involving the three random walks. There would be a probability of the form p(r, s) that the collision at time s would occur between E_s^r and the right end of the inner string before the inner string disappeared. The sum over $s > r$ of these probabilities would be the probability that the inner string would eventually collide with E_u^r before disappearing. If we let d be the distance between x and the right endpoint E_r^r, then as before, we would obtain an upper bound for this sum of the form Cd^{-1}. It is easy to see that the probabilities p(r, s) depend only on $s - r$, so we would also have the same bound for the sum over all values of r less than s. Summing over $d < 2N$, we would then have

$$\sum_{r:\, r<s} \quad \sum_{x \in [-N,\, N]} P(G(x, r, s)) \qquad \text{is} \qquad O(\varepsilon^2 \ln N) .$$

Unfortunately, we cannot identify E_u^r with B_u^r. We must take into account the effect of the process C_u^r. The simplest way to do this is to replace the event G(x, r, s) by a different event. Let x(t) be the position of the right endpoint of the inner string formed at x and x + 1 at time r and define

$\tilde{G}(x, r, s)$ = the event that F(x, r) occurs and the resulting inner string is still alive and simple at time s, and s is the smallest time such that either $B_s^r - B_r^r + E_r^r - x(s) < (E_r^r - x - 2)/2$ or $-(C_s^r - C_r^r) > (E_r^r - x - 2)/2$.

The event $\tilde{G}(x, r, s)$ does not contain the event G(x, r, s), but the union over x, r, s of the $\tilde{G}$ events does contain the union of the G events, and more importantly, the number of $\tilde{G}$ events that occur is greater than or equal to the number of G events that occur. The reason for this is that in order for a collision between an inner string and E_u^r to occur, the (distance − 2) between the two must be reduced at least half way either through the drift caused by the process C_u^r or by the wandering about of the random walk components of the various endpoints, and all this must happen before the inner

string disappears. If we let d equal this distance (i.e., $d = (E_r^r - x - 2)$), then the probability that half this distance is covered by the wandering of the random walks is bounded above by a quantity p(r, s) which sums to Cd^{-1} as before. The probability that it is covered by the process C_u^r is equal to $P(C_r^r - C_u^r \le d/2$ for all $u \in [r, s-1]$; $C_r^r - C_s^r > d/2)$. Since the process C_u^r is decreasing , we have a bound of $P(-C_s^r > d/2)$ for the sum of these terms over $r < s$. Summing over $d < 2N$, we obtain

$$\sum_{r:\, r<s} \quad \sum_{x \in [-N, N]} P(\tilde{G}(x, r, s)) \;<\; 2\varepsilon^2 E(-C_s^r) \;+\; O(\varepsilon^2 \ln N)\,.$$

Thus we have the following inequality:

$$E(-C_t^r;\, I(t, N)) \;\le\; C\varepsilon^2\Big(t \ln N \;+\; \varepsilon^{3/2} t^{3/2} N \ln N \;+ \sum_{s \le t} E(-C_s^r;\, I(s, N))\Big) \qquad (13)$$

(We have glossed over a technical point here, namely the appearance of the event I(s, N) on the right side of (13). This can be justified by using a stopping time that stops the process E_u^r at the first time that the inequalities in the definition of I(s, N) become violated. We will spare the reader the details.) The constant C in (13) is independent of ε, t and N. Let us assume that $t < c/\varepsilon^2$ and $N < c/\sqrt{\varepsilon}$ for some constant c independent of ε, and define $f^\varepsilon(s) = \sqrt{\varepsilon}\, E(-C_s^r;\, I(s, N))$. Then (13) becomes

$$f^\varepsilon(t) \;<\; C\Big(\sqrt{\varepsilon}\,\big|\ln\varepsilon\big|\varepsilon^2 t \;+\; \varepsilon^2 \sum_{s \le t} f^\varepsilon(s)\Big). \qquad (14)$$

Again, C is a constant which is independent of ε, N and t. It follows from Gronwall's inequality (see Coddington and Levinson, Chapter 1, Exercise 1) that

$$f^\varepsilon(t) \;<\; C\sqrt{\varepsilon}\,\big|\ln\varepsilon\big|\varepsilon^2 t \qquad \text{for all } \varepsilon > 0\,,\ t < c/\varepsilon^2 \text{ and } N < c/\sqrt{\varepsilon}\,. \qquad (15)$$

This last inequality has been the main goal of all our estimates, so let us take a moment here to understand what it gives us. By the Central Limit Theorem , $\sqrt{\varepsilon}\, B_t^r$ is approximately normally distributed, with mean 0 and variance equal to a constant times $\varepsilon^2 t$. On the other hand, we have just shown that if I(t, N) occurs, $-\sqrt{\varepsilon}\, C_t^r$ has an expected value which is small in comparison with $\varepsilon^2 t$, at least when ε is small. Typically, the size of $\sqrt{\varepsilon}\, B_t^r$ will be the same order of magnitude as the square root of its variance, while $-\sqrt{\varepsilon}\, C_t^r$ will be comparable to its expected value, so for $t < c/\varepsilon^2$ and $\varepsilon > 0$ sufficiently small, the drift part of E_t^r, namely C_t^r, will be small in comparison to the random walk part, which is B_t^r . Of course, there are still some technicalities to overcome, but the worst is over. We will not give these details here, but the remaining work on this part of the proof involves using the fact that there is sufficient independence built into the $\tilde{G}$ and H events to make the phrases "typically" , "comparable to" and "small in comparison to" sufficiently precise to prove the following statement:

Let $E_t = E_t^r - E_t^\ell$ and define $\tau = \inf\{t > 0: E_t < 2 \text{ or } E_t \geq d/\sqrt{\varepsilon}\}$,
for some $d > 0$. Then there is a constant C independent of d such that (16)
for all sufficiently small $\varepsilon > 0$ (depending on d)

$$P(E_\tau \geq d/\sqrt{\varepsilon}\,) > C\sqrt{\varepsilon}\, n/d.$$

(Recall that n is the number that determines the length of the original interval of +'s in the process ζ_t^n , so that $E_0 = 2n + 1$.) The lower bound on $P(E_\tau \geq d)$ is just what we would expect if there were no movement of the endpoints E_t^r and E_t^ℓ due to collisions. In other words, if C_t^r and C_t^ℓ were both 0, then E_t would equal $B_t^r - B_t^\ell$, which is just a random walk. If such a random walk starts at $2n + 1$, then the probability is $\sim n/d$ that it will hit d before hitting 1. Thus (16) is a way of expressing the statement

that the movement of endpoints due to collisions is negligible in comparison with the movement of endpoints due to single errors. We will use (16) to complete the proof in the next subsection.

Conclusion of proof. After our long excursion concerning the movement of endpoints, we are finally ready to return to an analysis of the process γ_t and the proof of (6). Recall that the initial state of γ_t has 0's at all sites. There is a probability of ε^2 that simultaneous errors with the same sign will occur at any fixed pair of neighboring sites at any given time. One can imagine that an interval of non-zeroes has appeared, with endpoints 2 units apart. One can define the subsequent positions of these endpoints in the same manner that E_t^r and E_t^ℓ were defined (the values at sites outside this interval are not relevant — we never assumed anything about these values in our actual definitions of the endpoints). Estimate (16) will apply. In particular, the probability that the interval will achieve a length of at least $d/\sqrt{\varepsilon}$ before disappearing is greater than $C\sqrt{\varepsilon}/d$, as long as $\varepsilon > 0$ is sufficiently small (depending on d).

Now consider a $c/\sqrt{\varepsilon} \times c^2/\varepsilon^2$ rectangle of space-time. Within such a rectangle, we would expect about $c^3/\sqrt{\varepsilon}$ intervals to appear as described in the last paragraph. Thus the expected number of these intervals to achieve a length of $d/\sqrt{\varepsilon}$ before disappearing is greater than Cc^3/d, at least for sufficiently small $\varepsilon > 0$. Choose c large enough so that $Cc^2 > 1$ and set $d = c$. Then we have the following:

> The expected number of intervals originating in a $c/\sqrt{\varepsilon} \times c^2/\varepsilon^2$ space-time rectangle which achieve length $c/\sqrt{\varepsilon}$ before disappearing is at least 1 for sufficiently small $\varepsilon > 0$, where c is independent of ε. (17)

There are now two remaining steps. These are both non-trivial, but they contain no

new ideas, so we will only briefly describe them and refer the reader to the paper referred to at the beginning of the proof for details. The first step is to convert (17) into the following:

> Let c be as above. Then there exists $\delta > 0$ such that for sufficiently small $\varepsilon > 0$ and for any $c/\sqrt{\varepsilon} \times c^2/\varepsilon^2$ space–time rectangle R, (18)
>
> $$P(\text{an interval originates in R which achieves length } c/\sqrt{\varepsilon} \text{ in time } c^2/\varepsilon^2) > 1 - \delta.$$

As in other parts of the proof, we claim that there is sufficient independence around to get from (17) to (18). The dimension of the rectangles in the time direction has been chosen so that any interval that achieves length $c/\sqrt{\varepsilon}$ has a minimum positive probability of doing it within a time interval equal to the length of the time side of the rectangle. This is important in order to maintain a certain amount of independence for the rest of the argument.

In the final step of the proof, we finally use a fact that we have carefully avoided until now, namely the fact that two intervals of non–zeroes can join to form a larger interval. The statement in (18) guarantees that there is a minimum positive density of intervals of non–zeroes with a certain minimum length in rescaled space–time. If we focus on one of these intervals, we find that its endpoints will continue to move about like random walks (or Brownian motion as $\varepsilon \to 0$). There will be a certain rate at which the interval will join with another interval. The endpoints of this new longer interval will also move about like a random walk until a further coalescence with yet another interval. It is not too hard to show that we therefore will get an interval which grows linearly in length. By choosing ε sufficiently small, this description can

be made accurate for arbitrarily long periods of rescaled time. There are certain percolation techniques (fast becoming standard in this area) which can now be used to show that for all sufficiently small $\varepsilon > 0$, with probability 1 these growing intervals will link together to form an infinite region in space–time, and every site will eventually become a part of this region forever. The statement in (6) follows easily from this. We do not have space here to give the details of the percolation argument. For an example, see Durrett and Griffeath (1983). This and other missing parts of our proof will be found in Gray (1986). □

A concluding note. In the proof we rescaled space and time in such a way that the endpoint processes behaved like Brownian motions as $\varepsilon \to 0$. The same thing can be done in other models with SML flip probabilities or flip rates. It is natural to ask whether there is some limiting distribution in space–time which is independent of the particular model. I believe that there is. In other words, I conjecture that there is an invariance principle in operation which applies to all finite range SML models as $\varepsilon \to 0$. The chief difficulty with this conjecture is not to prove that it applies to a large class of models. Instead, I have difficulty trying to make sense out of it even in the special case of the nearest neighbor majority vote model. The problem lies in trying to identify the limiting object. One must try to imagine a process which lives on $\mathbb{R}$ in which swarms of infinitesimally small intervals of two different colors are appearing. The endpoints of these intervals move like Brownian motions, so most of them disappear immediately. However, so many of them are produced that some of them become long enough to be "visible". Collisions produce no effects except to join two intervals into one. One is vaguely reminded here of the production of so–called virtual particles in physics, most of which immediately disappear, but a few of which have a relatively long life. It is hoped that if it is really possible to make sense out of such

a process, some insights into the nature of Gibbs states at low temperatures will be attained. Then our goal of applying the stochastic theory to equilibrium theory would be realized.

References

1. Coddington, E. and Levinson, N. (1955). Theory of Ordinary Differential Equations. McGraw–Hill. New York.
2. Durrett, R. and Griffeath, D. (1983). Supercritical contact processes on $\mathbf{Z}$. Z. Wahrsch. verw. Gebiete **61** 389–404.
3. Gray, L. (1986). Finite range majority vote models in one dimension. In preparation.
4. Holley, R. and Stroock, D. (1976). Applications of the stochastic Ising model to the Gibbs states. Communications in Math. Physics **48** 249–266.
5. Holley, R. and Stroock, D. (1976). L_2 Theory for the stochastic Ising model. Z. Wahrsch. verw. Gebiete **35** 87–101.
6. Holley, R. and Stroock, D. (1977). In one and two dimensions, every stationary measure for a stochastic Ising model is a Gibbs state. Communications in Math. Physics **55** 37–45.
7. Kindermann, R. and Snell, J. L. (1980). Markov Random Fields and their Applications. Contemporary Mathematics Volume 1. Amer. Math. Soc.
8. Liggett, T. (1972). Existence theorems for infinite particle systems. T.A.M.S. **165** 471–481.
9. Liggett, T. (1985). Interacting Particle Systems. Springer–Verlag. New York.
10. Spitzer, F. (1970). Interaction of Markov processes. Adv. in Math. **5** 246–290.