Language Extraction from ZnS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Dowman Parks Varn
December, 2001

Errata and Corrections

This thesis represents the first formal presentation of the algorithm for ϵ-machine reconstruction from spectral data called *ϵ-machine spectral reconstruction theory* or *ϵMSR*. As might be expected from a work that introduces a novel approach—let alone one that is written by a student to satisfy the requirements of a degree—there is significant attention given to both the theoretical developments that presage ϵMSR (Chapters 2-4) as well as the mechanical details of the theory itself (Chapters 6-8). In particular, the derivation of the *spectral equations*[†,‡] and the connection between ϵ-machines, correlation functions and power spectra is exhaustively explored by treating many worked examples (Chapter 5) that would be inappropriate for a research journal publication. For these reasons, this thesis retains much of its relevance in spite of subsequent expository papers.[†,‡,⋆]

However, since taking its final form in December 2001, it has become clear that one aspect of the exposition given herein, related to the interpretation of the reconstructed ϵ-machine, is erroneous. I concluded that the range of interaction between modular layers must be greater than or equal to three, since only ϵ-machines that have a *memory length*[†,‡,⋆] of at least three give reasonable agreement with experiment. Identifying the memory length of the ϵ-machine with the minimum range of interaction embodied in an interaction Hamiltonian that governs the coupling between modular layers is only valid under conditions of thermal equilibrium. The zinc sulphide polytypes studied here are almost certainly not in equilibrium. Indeed, they are trapped in a local minimum of the energy with insufficient thermal energy to escape. Hence the conclusion that the minimum range of interaction between modular layers be at least three is unsupportable. In fact, it can be shown* that a simple model of solid state transformations in these crystals—assuming only nearest neighbor interactions—can generate long range memories. However, this error in the interpretation of the memory length of the resulting ϵ-machine in no way invalidates the either the conceptual or computational development of ϵMSR given here. Further, the interpretation of weakly represented *causal state cycles* as faulting structures, (Chapter 8), is likewise sound.

Since I have felt the need to prepend this thesis with this explanatory note, I also take this opportunity to make a several changes to the thesis itself. It would certainly be disingenuous to make substantive, conceptual alterations, and so I confine my corrections to relatively minor, editorial matters. These are:

– A number of typographical errors have been corrected.

– A small change in notation has been made to avoid confusion and conform with subsequent work. Namely, the figure-of-merit α has been changed to β (§4.2) to agree with usage found elsewhere.[†,‡,⋆]

– While no new references have been added, I have updated those references that have subsequently been published.

– For several tables I discovered that there was no appropriate reference in the text alerting the reader to their presence. Such wording as necessary to inform the reader of useful tabulations has been added.

Perhaps some of us are fated to eternal dissatisfaction with our writing and wish for the luxury of continual, unfettered editorial control, however impractical that may be. I fear that describes me all too well. Indeed, it seems that with every page I find items that at best are phrased awkwardly and at worst are misleading, incomplete, or seem to cry out for clarification. I hope that any readers will forgive any shortcomings they may find.

Dowman P. Varn
Dresden, Germany
November, 2004

[†] Varn, D. P., G. S. Canright and J. P. Crutchfield (2002). "Discovering planar disorder in close-packed structures from x-ray diffraction: Beyond the fault model," *Physical Review B*, **66**: 174110.

[‡] Varn, D. P., G. S. Canright and J. P. Crutchfield (2003). "Inferring pattern and disorder in close-packed structures from x-ray diffraction studies, Part I: ϵ-Machine spectral reconstruction theory." [cond-mat/0302585].

[⋆] Varn, D. P., G. S. Canright and J. P. Crutchfield (2003) "Inferring pattern and disorder in close-packed structures from x-ray diffraction studies, Part II: Application to zinc sulphide." [cond-mat/0302587].

[∗] Varn, D. P. and J. P. Crutchfield (2004). "From finite to infinite range order via annealing: The causal architecture of deformation faulting in annealed close-packed crystals." *Physics Letters A*, **324/4**: 299-307.

for my parents,
Jim and Pat

Acknowledgments

It is not possible to list all the people to whom I owe an enormous debt and who have helped bring this work to fruition. At the risk of not including some, let me list a few that have been particularly meaningful.

My first introduction to physics came from my high school physics teacher Chris Wright. I thank her for her pedagogical style, which I have adopted as my own. H. Willingham Graben, Edward Gettys, and John McKelvey were inspiring instructors at Clemson University where my love of physics grew. I thank them for many lectures full of wonder and magic. Mickey Kutzner, James Boyle, Eric Graben, Daniel Himes, Mark Buchanan and Andrea Stout were fellow graduate students at the University of Virginia. Late nights, coffee, and their moral support prevented my enjoyment of physics from dying. I thank Hugh Kelly for teaching me about argon, I only wish there had been more time and energy to learn. Cathy Almquist at Trident Technical College and Robert Dukes at the College of Charleston gave me an opportunity to teach, and there I found a home. They were patient, and let me learn how to run a classroom on my own. The most remembered lessons are those that come with a bump. I thank the entire faculty and staff at the physics department of the University of Tennessee, who have truly gone out of their way to help me succeed. Deserving of special mention is Lee Riedinger, who tempered sternness with understanding at a critical juncture.

Part of my doctoral research has taken place at Indiana University in the form of repeated visits. I thank them for their hospitality. While there, I often stayed with Geoff Canright, Heidi Kjønsberg and Malin Kjønsberg. They welcomed me warmly into their home, fed me delicious foods of many origins, and provided entertainment in the evenings. I will not forget their kindness and friendship. I especially thank Malin for her services as banker during our monopoly games, and her cheerful and patient explanations of the Norwegian monetary system to a rather slow learner.

I want to particularly thank my doctoral committee, C.C. Shih, John Quinn, Charles Collins and Geoff Canright. They have endured drafts, complete and otherwise, graciously. Charles Collins provided help in the solutions of non-linear equations and John Quinn made helpful suggestions on the manuscript. Since my research advisor Geoff Canright and I have been *in absentia* from the University of Tennessee during my final year, Dr. Shih has taken on both the task of chairing my committee and being the "point man" at the the University of Tennessee. He has performed kindnesses great and small, and I thank him.

Some of the work and all of the writing of this dissertation were done at the Santa Fe Institute. The atmosphere and the people here are nothing less than magical. I especially want to thank my collaborator and mentor at the Institute, Jim Crutchfield, for sharing his insights into computational mechanics and providing guidance. Cosma Shalizi has answered innumerable questions both small and large, stupid and profound, with graciousness. I thank him for his willingness to entertain my often ill-conceived ideas without objection. Cris Moore has brought to my attention important facts concerning correlation functions and Dave Feldman has discussed his previous work on the computational mechanics of spin systems upon which this work is partially based. For numerous and helpful comments at a talk I gave, I thank Jim Crutchfield, Cosma Shalizi, Cris Moore, Michelle Girvan, Homayoun Bagheri-Chaichian, and Marcus Daniels. For additional help and friendship, I thank Joe Jun, Eric Lee, Paolo Patelli, Van Savage, Homayoun Bagheri-Chaichian, Michelle Girvan, Cosma Shalizi, Donatella Pasqualini, Marcus Daniels, Kris Klinkner, Mark Newman, D. Eric Smith, and Rob Haslinger. I thank Rebecca L. Halperin for helpful conversations.

Much of the work presented in this dissertation was produced using free software and I thank all those toil to make such software available. Of particular use in this work were the LaTeX typesetting program which was used to typeset this document; DataThief II which was used to digitize experimental data; gnuplot for nearly all of the graphs; xfig and dot for the machine diagrams; the C programming language and the gcc compiler; and the Linux operating system.

I wish to express my deep appreciation to my research advisor, Geoff Canright. He has provided unflagging support even when it may not have seemed the wise thing to do. I am inspired by him

and have immensely enjoyed our collaboration. My parents have been an unending source of support my entire life, and to them I dedicate this thesis. Finally, without the love, patience, and fortitude of my wife, Mary, I could not have flourished.

I acknowledge financial support for my studies at the University of Tennessee as a teaching assistant and later as a research assistant under grant DMR-9820816. I also acknowledge support under a grant to the Santa Fe Institute from the National Science Foundation, award PHY-9910217, "Pilot Program for NSF Physics Graduate Student Fellowships."

Abstract

Recent advances in the analysis of one-dimensional temporal and spacial series allow for detailed characterization of disorder and computation in physical systems. One such system that has defied theoretical understanding since its discovery in 1912 is polytypism. Polytypes are layered compounds, exhibiting crystallinity in two dimensions, yet having complicated stacking sequences in the third direction. They can show both ordered and disordered sequences, sometimes each in the same specimen. We demonstrate a method for extracting two-layer correlation information from ZnS diffraction patterns and employ a novel technique for epsilon-machine reconstruction. We solve a long-standing problem - that of determining structural information for disordered materials from their diffraction patterns - for this special class of disorder. Our solution offers the most complete possible statistical description of the disorder. Furthermore, from our reconstructed epsilon-machines we find the effective range of the interlayer interaction in these materials, as well as the configurational energy of both ordered and disordered specimens. Finally, we can determine the 'language' (in terms of the Chomsky Hierarchy) these small rocks speak, and we find that regular languages are sufficient to describe them.

Contents

List of Figures

List of Tables

Chapter 1

Prelude

> It is fascinating that Nature should produce such complicated structures with compounds such as SiC, CdI_2, and ZnS, all of which have a very simple proportion of atoms in them. If one just picks up at random one crystal of any of these materials, out of an as grown bunch of crystals, that tiny speck of the material can have one of several hundred structures - some ordered, some randomly faulted and some with non-random faulting. Which one will it have and why?
>
> *M.T. Sebastian and P. Krishna* [61]

Perhaps the most fundamental questions we can ask about a solid are "What is it made of?" and "How are the constituent parts assembled?" This is so elementary, and yet so basic to any detailed understanding of the thermal, electrical, magnetic, optical, and elastic properties of materials. At the beginning of the twenty-first century, concern over the placement of the atoms in a solid seems quaint and anachronistic, more suited to the dawn of the twentieth century. X-ray diffraction, electron diffraction, optical microscopy, x-ray diffraction tomography, to name a few, are powerful techniques to uncover structure in solids. With this arsenal of tools, and the efforts of many researchers, surely we can have nothing novel to say about the discovery and description of structure in solids, save perhaps the refinement of well-worn techniques or the analysis of particularly obstinate cases. But careful examination of present technology reveals that while we are quite good at finding and describing periodic order in nature, cases that lack such order are much more difficult. Certainly in the complete absence of structural order, as in a gas, statistical methods exist that permit a satisfying understanding of the properties of the system without knowing (or even wanting to know) the details of the microscopic placement of the constituents. But it is the in-between cases, where order and disorder coexist, that has proven so elusive to both analyze and describe. In this thesis, we will tackle these in-between cases for a special type of layered material, called polytypes. They exhibit disorder in one dimension only, making the analysis more tractable. We will give a method for determining the structure of these solids from experimental data and demonstrate how this structure, both the random and the non-random part, can be compactly expressed. From our solution, we will be able to calculate the effective range of the inter-layer interactions, as well as the configurational energies of the disordered stacking sequences.

Zinc sulfide would seem a rather unpretentious material. Its stoichiometry is simple enough, the proportion of zinc to sulphur atoms is 1:1. Each zinc atom is tetrahedrally coordinated to four sulphur atoms and *vice versa*. It is known to crystallize into the face-centered cubic (fcc) structure - alternately called cubic zinc sulfide, zinc blende or the sphalerite structure - at low temperatures. In this configuration, one can think of the zinc atoms as forming an fcc lattice and the sulphur atoms forming an inter-penetrating fcc lattice displaced from the latter one by one-quarter of a body diagonal [46]. For our purposes, it is useful to look along the [111] direction of the conventional unit cell. Taking a zinc atom at the origin of the this cell, we see that the zinc atoms are arranged hexagonally in the (111) plane with sulphur atoms residing a quarter body

diagonal length directly above them. We will call the configuration of a plane of zinc atoms in a hexagonal net with the sulphur atoms placed directly above a modular layer. We can think of the zinc atoms as ideally close-packed and the sulphur atoms as 'slaved' to sit atop the zinc atoms. The stacking of modular layers proceeds just as for the stacking of layers of close-packed spheres. We can identify three possible orientations of the modular layer, let us call them A, B and C. Just as for ideally close-packed spheres, the standard stacking constraints apply, namely that no two neighboring modular layers may have the same orientation. The zinc blende structure corresponds to the $...ABCABC...$ stacking sequence. Just as in the case of close-packed spheres, there can be other stacking sequences, for example $...ABABAB...$, which is known as the hexagonal close-packed (hcp) structure. It is not surprising then that there are an infinite number of possible stacking sequences, some of them periodic (as the zinc blende structure is) and some of them aperiodic. What is surprising is the frequency and apparent ease with which these alternate stacking sequences occur. Under virtually identical thermodynamic growth conditions, one can find crystals that have the fcc structure, the hcp structure, one of about 180 other known periodic structures and finally many disordered structures [61]. This phenomenon of solids built up from identical layers differing only in the manner of the stacking is called polytypism. The precise definition is somewhat more liberal than this, see Trigunayat [69] for a recent discussion. The various stacking structures of a particular solid are called polytypes. The unit cells for some polytypes of ZnS can extent over one hundred modular layers [61]. The mystery of polytypism then is two-fold: How can so many different structures exist under equivalent thermodynamic conditions and How can Nature orchestrate such ordered sequences extending over hundreds of angstroms, much farther than any known atomic interaction?

Were polytypism confined to just one, or at most a few, substances, then maybe we could be excused for treating this as a curiosity - interesting but perhaps not deserving of too much attention. Since the discovery of polytypism in SiC by Baumhauer [7] in 1912, however, it has been found in dozens of other materials. The most prominent polytypic materials are SiC, ZnS, and CdI_2 having about 150, 185, and 200 known periodic structures respectively. In addition to these, polytypism has been found in Si, PbI_2, SnS_2, GaS, micas, kaolins, metals such as Co and Li, various alloys like Cu-Al and Ag-Cd, oxides such as $BaRuO_3$, and halides like $RbNiF_3$. The list goes on. Interestingly, they have widely varying physical properties: some are metals like Co, others are insulators like CdI_2, and some are semiconductors like SiC. The nature of the inter-atomic forces also differs widely: some are mostly covalently bonded, like SiC; while in others, like CdI_2, the bonding is predominantly ionic. Polytypes are found to occur both naturally and in man-made samples. In the last fifty years, considerable effort has been expended to understand polytypism, with over a dozen theories having been proposed; but a general explanation is still lacking [61] [69]. Given the diversity of the different polytypic materials this does not seem surprising. Indeed, one wonders whether the phenomenon can be attributed to a single cause.

We will have nothing say about mechanisms of polytypism here; that is left to future researchers. Instead our concentration will be focussed on techniques to describe and detect the structure present in a particular polytypic material, ZnS. We feel that before a phenomenon can be explained, it must first be adequately described. In the past ten or fifteen years, significant progress has been made in pattern discovery and classification in one-dimensional systems. Growing out of an effort to understand nonlinear dynamical systems, computational mechanics combines ideas from symbolic dynamics [50] [38], the formal theory of languages [40] [54], and information theory as developed by Claude Shannon [64] [14] [1]. Computational mechanics provides a systematic method for finding and expressing the structure and organization in a one-dimensional, discrete sets of data. Using it, we can define properties like the entropy density, the complexity, and the excess entropy. Computational mechanics answers such questions as "What is the minimum memory the process must have to produce this data stream?", "How is information stored and used in the process?" and "How random is the process?" Computational mechanics does not supplant statistical mechanics, but rather is complementary to it. Computational mechanics allows us to construct a model of the

process, called the ϵ-machine, directly from the data itself. This model is the optimal, minimal and unique description of the process. So far, computational mechanics has been applied to both artificial systems such as cellular automata [37], the logistic map [18] [82], and the one-dimensional Ising model [23] [16], as well as more physical systems like the dripping faucet [30], atmospheric turbulence [51], and geomagnetic data [13].

It is worth mentioning that our technique for ϵ-machine reconstruction is novel. Instead of the usual situation of analyzing a data stream or long sequence of symbols, our information about the underlying process comes in the form of two-layer correlation functions. We will find that they provide only impoverished information about the process and impose severe restrictions on the classes of processes we will be able to detect. We adopt this method not for its superiority to previous methods, but out of necessity. Nonetheless, we will find that it gives a reasonable answer to our problem. While our method is specifically directed towards machine reconstruction in a particular polytypic material, ZnS, we hope that concepts and techniques explored here will prove useful in machine reconstruction from other kinds of power spectra.

The plan for the rest of this thesis is as follows. Since computational mechanics is not a common tool in most physicists' tool chest, we will devote Chapter 2 to explaining the basic ideas and providing definitions for some the quantities we will use subsequently. In Chapter 3 we will very briefly discuss the experimental details of the data we will use to infer the underlying process. In Chapter 4, we show how correlation information can be extracted from the experimental data. In Chapter 5, in order to build intuition, we give examples of the different diffraction patterns arising from different, fake processes. In Chapter 6, we discuss our technique for ϵ-machine reconstruction from two-layer correlation functions. In Chapter 7, we provide examples of ϵ-machine reconstruction from diffraction spectra of known processes. In Chapter 8, we give a discussion of previous models of disorder in layered solids and show how they relate to ϵ-machines. In Chapter 9, we employ this machinery on real experimental data for ZnS and give the models for the underlying process. In Chapter 10, we present our conclusions and possible directions for future work.

Chapter 2

A Brief Introduction to Computational Mechanics

Computational mechanics is not familiar to most physicists. We will therefore give a brief introduction to the ideas leading up to computational mechanics as well as an overview of the theory itself. There are several good references available on computational mechanics and the interested reader is urged to consult these for a much more detailed exposition. These references include Crutchfield and Feldman [16] [17], Feldman [23] [24], Feldman and Crutchfield [25] Shalizi [62], Shalizi and Crutchfield [63], Hanson [37] and Young [82].

2.1 The Intellectual Precursors of Computation Mechanics

Computational mechanics arises from the marriage of three distinct lines of thought: symbolic dynamics, language theory and information theory. Since we borrow terminology and concepts from each to formulate computational mechanics, let us consider each separately.

2.1.1 Symbolic Dynamics

In the study of nonlinear dynamical systems, discretizing the formalism can greatly facilitate the analysis. While some systems are naturally discrete, most are not and it is therefore necessary to introduce a some kind of discretization process. One way to do this is to describe a dynamical system by using a map [77] [67] [48]. Maps of course can be use to describe other systems, but the motivation for their study in physics is the connection to dynamical systems. Maps have the convenient property of being discrete in time, and we can additionally impose a partition $\mathsf{B} = \{B_1, B_2, ..., B_b\}$ on the other dynamical variables. Doing so, we associate some $B_i \in \mathsf{B}$ with the state of the system with each iteration of the map. In this way, we can build up a sequence of symbols describing the time evolution of the system. Symbolic dynamics [6] [8] [38] [45] is nothing more than the study of such a symbol sequence. From the sequence generated from an appropriate choice of B, it is possible to determine if the orbit of the trajectory is periodic or not, and put a lower bound on the entropy production of the system.

The slicing of a continuum into finite number of wedges may seem severe, and one may think that this is artificial in some way. Much of our experience in the macroscopic domain points to a continuous world. Our measuring instruments - our basic tools for empirical discovery - however are not. In the process of measurement, we must always face the fact that there is some finite resolution and we can not distinguish structure on a scale smaller than this. So partitioning the world into cells is really not so far from our practice.

Table 2.1: The Chomsky Hierarchy. The most powerful class of languages, those languages that have no restriction on word production, are called *unrestricted* languages. As one goes down the list, progressively stricter rules are imposed on word production.

Type	Language	Automaton
0	Unrestricted	Turing
1	Context-sensitive	Linear Bounded
2	Context-free	Pushdown
3	Regular	Finite

2.1.2 The Formal Theory of Languages

The analysis of strings of symbols is enhanced by a knowledge of the theory of languages [40] [54]. Let us first introduce some notation. A *symbol*, s, is some abstract entity that we accept intuitively as existing, but we do not define it. Examples of symbols are letters and digits. An *alphabet* $\mathcal{A}$ is the finite set of all allowed symbols, $\mathcal{A} = \{s^0, s^1, s^2, ..., s^{b-1}\}$, where s^i is the i^{th} symbol in the alphabet. We denote the size of $\mathcal{A}$ as $|\mathcal{A}|$, and this just the number of symbols in the alphabet. We also define the empty string as λ. A *word* is the juxtaposition of two or more symbols. For example, s^2s^4 is a word composed of the juxtaposition of the two symbols s^2 and s^4. The length of a word w is denoted by $|w|$, and is just the number of symbols in the word. For example $|s^1s^3| = 2$. We also see that $|\lambda| = 0$. A *concatenation* of two words is just their juxtaposition without any intervening space. If u and v are two words, their concatenation is uv. A word v is a *subword* of some word string S if $S = uvw$ with $|uw| \geq 1$.

With these definitions in place, we can now define a language. A *formal language*, $\mathcal{L}$, is the set of allowed words. For example, suppose we take our alphabet to be binary and represented by the symbols 0 and 1. Then $\mathcal{A} = \{0, 1\}$. We could define our set as all possible words. We would then have $\mathcal{L} = \{\lambda, 0, 1, 01, 10, 11, 000, 001, ..., 111, 0000, ..., 1111, ...\}$. A language can also be specified by the words it does not contain, called *forbidden words*. A forbidden word is called *irreducible* if it contains no forbidden subwords. The set of forbidden words is called the compliment of $\mathcal{L}$, which we will denote as $\mathcal{F}_{\mathcal{L}}$.

The classification of languages according to their allowed words leads to a rich hierarchical structure. In its most basic form this structure is called the Chomsky hierarchy, and is shown in table 2.1. The hierarchy is arranged such that the most powerful languages are at the top, and they decrease in sophistication as one moves down the table. We will not delve into the definitions of languages higher than those called *regular languages*; we simply wish to place them in context. Interested readers are urged to consult [40] [54] for more discussion. In order to define regular languages, we need to introduce the idea of a *grammar*. Languages can be classified according to restrictions placed on the production of words. These production rules are called grammars - because much like the rules for composing words in natural languages - they tell how words can be 'made' in formal languages.

Definition 2.1 *We define a generative grammar (or simply grammar) G as an ordered four-tuple $(\mathcal{A}_N, \mathcal{A}_T, S, F)$, where $\mathcal{A}_N$ and $\mathcal{A}_T$ are finite, non-intersecting alphabets, S is a distinguished symbol from $\mathcal{A}_N$ and F is a set of ordered pairs (P, Q) with $P, Q \in (\mathcal{A}_N \cup \mathcal{A}_T)$ and P containing at least one symbol from $\mathcal{A}_N$.*

Let us unpack this. The set $\mathcal{A}_N$ are called *nonterminal* symbols and we can think of them as intermediaries in the production of words. The set $\mathcal{A}_T$ are called *terminal* symbols and are the letters of the alphabet from which we wish to build the language. S is a special symbol that starts

the derivation of the word. Finally, F are the actual production rules we use make words. We write them as $P \to Q$, and this simply means we replace P with Q. We consider several examples.

Example 2.1 Let us construct the grammar for the language that allows every word over an alphabet $\mathcal{A} = \{0, 1\}$. So let us take $\mathcal{A}_T = \{0, 1\}$ and $\mathcal{A}_N = \{S\}$. The production rules F are just

$$\begin{aligned} S &\to \lambda \\ S &\to 0S \\ S &\to 1S \end{aligned}$$

To make a word, one starts with S and can use any of the substitution rules. We can quit when there are no more symbols from $\mathcal{A}_N$ in the word. So, to get the empty string λ we use the first rule and stop. To generate the string 11, we say $S \to 1S \to 11S \to 11\lambda \to 11$. In this fashion, these three rules allow us make all possible words composed from the alphabet $\mathcal{A}_T = \{0, 1\}$.

Example 2.2 Let us examine a more difficult example. Suppose we want to generate the language such that there are no two adjacent 0s, *i.e.*, '00' is a forbidden word. Let us take $\mathcal{A}_T = \{0, 1\}$ and $\mathcal{A}_N = \{S\}$ as before. We can accomplish this with the following production rules

$$\begin{aligned} S &\to \lambda \\ S &\to 01S \\ S &\to 1S \end{aligned}$$

The second rule requires that we always follow a 0 with a 1. This language is called the golden mean language, and we will discuss it in more detail in a later chapter.

Example 2.3 As a final example, let us consider the language such that between any two 0s there is an even number of 1s. Another way of saying this is that the set of forbidden words can be expressed as $\mathcal{F}_\mathcal{L} = \{01^{2k+1}0\}$ with k a non-negative integer. Let us once again take $\mathcal{A}_T = \{0, 1\}$ and $\mathcal{A}_N = \{S\}$ as before. We generate this language with the following production rules

$$\begin{aligned} S &\to \lambda \\ S &\to 11S \\ S &\to 0S \end{aligned}$$

The second rule insures that we always make 1s in pairs. This language is called the even language, and we will also discuss it in more detail in a later chapter.

We can now define a regular language:

Definition 2.2 *A grammar $G = (V_N, V_T, S, F)$ is said to be regular if every rule in F has the form either $A \to PB$ or $A \to P$ where $A, B \in V_N$ and $P \in V_T^*$.*

V_T^* is the set of all words over V_T. From the definition, we can see that the three examples of languages introduced above are all regular languages. Higher level languages are defined by relaxing the restrictions on word production.

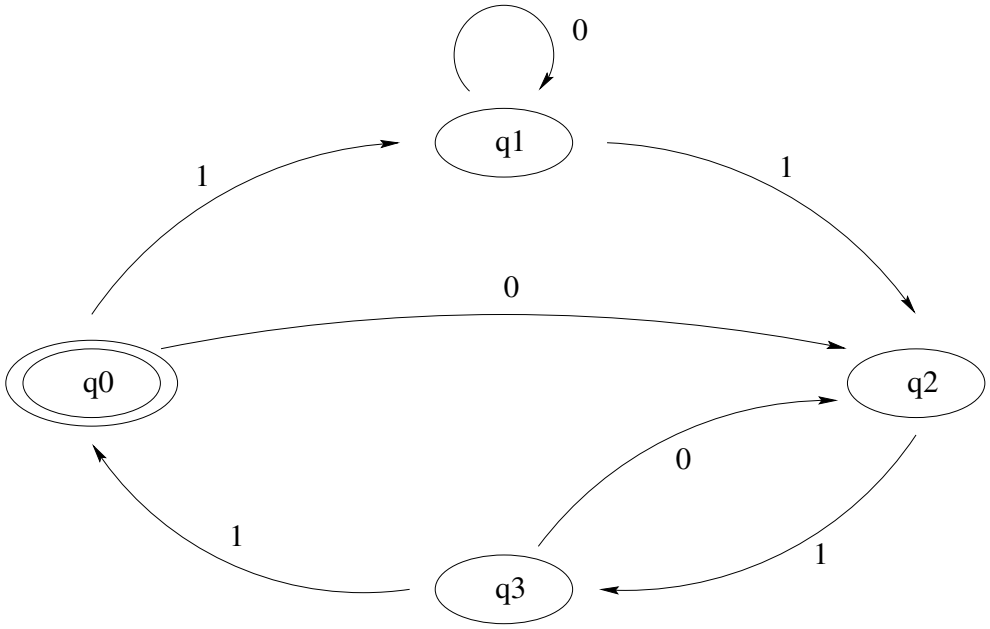

Figure 2.1: An example of a finite automaton. The states are given by ovals and are labeled q0, q1, q2 and q3. The double oval of q0 indicates that this is a start state, meaning that we must begin any string generated by the automaton with this state. The directed arcs connecting states label transitions between the states with the emission of the symbol dressing each arc.

All regular languages can be specified by giving a *regular expression* $\mathcal{R}$. Let us define a few operators and then we can proceed on to regular expressions. The notation * appended to a string means that we can repeat the string as many times as we wish, including no times. So, $a^* = \{\lambda, a, aa, aaa, ...\}$. The 'or' operator is given by +, and simply means we have a choice. So $(0 + 11)$ means either 0 or 11, not concatenation, which would be 011. So between the three operations of *, +, and concatenation, we can build up words. There is an order of precedence, when parentheses are not used, namely that * has higher priority than +, which has higher priority than concatenation. So the language that allows all words in the alphabet $\mathcal{A} = \{0, 1\}$ has a regular expression given by $\mathcal{R} = (0 + 1)^* = \{\lambda, 0, 1, 00, 01, 10, ...\}$. The regular expression for the golden mean language is given by $\mathcal{R} = (1 + 01)^*$. We really should define regular expressions over an alphabet in a more rigorous, recursive fashion, as is done in Hopcroft and Ullman [40]; but the notion seems intuitive enough that this will do. In later chapters, when we give a regular expression for a language, we will assume that a very long string from the set is taken for the purpose of calculation.

Yet another way to express a regular language is with a *finite automaton*. We can define a finite automaton as a finite set of states (also called vertices or nodes) connected by a arcs (also called edges). On transition from one state to the next, the finite automaton emits a symbol from the alphabet. Usually one state is designated as the start state, and other states are called final or accepting states. It is convenient to represent a finite automaton as a directed graph, as shown in figure 2.1. The ovals represent nodes and the line segments represent arcs. The states are identified by the labels q0, q1, q2, and q3. The double oval of q0 means that this is the start state. The edges have symbols attached to them, and these symbols are emitted on each transition from node to node. Alternately, one can think of this graph as a mechanism for determining whether a particular string belongs to a language. We start of course with the start state, and a symbol is read in. If there exists a transition that takes us to another state from our current state, then the symbol is

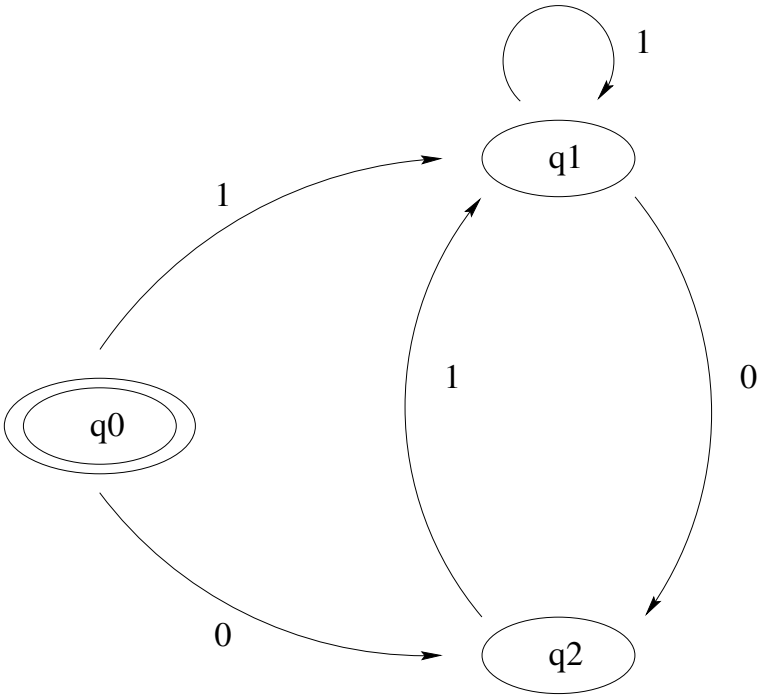

Figure 2.2: The finite automaton for the golden mean language. The golden mean language is the set of all words over $\mathcal{A} \in \{0, 1\}$ that do not contain two consecutive 0s. This automaton clearly does not generate two consecutive 0s, as emission of a 0 on the transition from state q1 to state q2 is always followed by the emission of a 1 back to state q2.

accepted. Otherwise, we reject it and determine that the proposed string is not part of the language. If, when the string is entirely read in, the final state is an accepting state, then we say that the string is in the language. For our purposes, we can treat all states as accepting states. Figure 2.2 shows the finite automaton for the golden mean language. Let us give a formal definition to a finite automaton.

Definition 2.3 *We define a finite automaton M as an ordered quintuple $(\mathcal{N}, \mathcal{A}, \delta, S, \mathcal{N}_F)$, where $\mathcal{N}$ is a finite set of states, $\mathcal{A}$ is a finite alphabet, δ is transition function, S is a start state, and $\mathcal{N}_F$ is a set of accepting states, $\mathcal{N}_F \subseteq \mathcal{N}$.*

The transition function δ just tells us how one state evolves into another on emission of a symbol a. That is, $\delta(\mathcal{N}_i, a)$ is a state for each state $\mathcal{N}_i \subseteq \mathcal{N}$, and symbol a. $\delta(\mathcal{N}_i, a)$ may not be defined for all $\mathcal{N}_i$ and a, in which case the transition is not allowed.

It is helpful to distinguish between different kinds of graphs. Let us call a finite automaton *deterministic* if the symbol emitted at each edge takes the system to a definite state. For example, suppose that the transition from q3 to q0 in figure 2.1 is labeled with a 0 instead of a 1. Then if we are at state q3, emission of a zero does not uniquely define the next state. Figure 2.1 would still be perfectly fine finite automaton, but we would call it *nondeterministic*. Determinism then here means something a little different than what a physicist might think. Determinism does not imply that each state has a unique successor state, but rather that each state on emission of a particular symbol has a unique successor state, provided such a transition is allowed. It can be shown that any nondeterministic finite automaton can be written as an equivalent deterministic finite automaton, although in general the deterministic version has exponentially more states.

Regular languages can be further divided into *subshifts of a finite type* and *strictly sofic systems*. Subshifts of a finite type (SFT) are those regular languages that have a largest irreducible forbidden word. The language composed of all strings, as in example 2.1, is one, as is the golden mean language. Certainly the golden mean language has forbidden words of indefinite length, but they are reducible because they all have '00' as a subword. The even language, however, has no largest irreducible forbidden word. In a sense, then this language must possess an infinite memory since it must be able to distinguish between indefinitely long sequences of 1s sandwiched between two 0s. The even language, then is an example of a strictly sofic language (SS). SFTs have the nice property that they represent r^{th}-order Markov processes. Recall that an r^{th}-order Markov process is one in which the admissibility (or probability in the case of stochastic processes) of the next symbol in a sequence depends at most on only the previous r symbols.

It turns out that all of these different ways of specifying a regular language are equivalent. We can give the list of allowed words, or alternatively the list of forbidden words; we can give the grammar; we can write out the regular expression; and finally we can draw the finite automaton. We have the freedom to use whichever is most convenient. In further developments, we will find that the graphical representation is most helpful. The formal theory of languages, which we have touched on in only the most cursory way here, is rich and leads to many avenues of inquiry, such as Gödel's Undecidability Theorem; but for the description of physical processes we find it lacking. The theory of languages only addresses the question of whether a particular string is allowed, not whether it is common, or rare, or somewhere in between. To treat the probabilities of seeing a word, we need to appeal to another discipline, that of information theory.

2.1.3 Information Theory

Information theory [1] [64] [14] was founded by Claude Shannon in 1940s to investigate how to better transmit and compress data for an increased efficiency in communication. It is often considered a part of communication theory, but it has applications beyond this field. We wish to introduce a few basic notions from information theory that we will find useful in the construction of computational mechanics.

The first idea is that of the *entropy* of a probability distribution. Suppose we have some random variable, X, that can take on the values $x \in \mathcal{A}$, where $\mathcal{A}$ is some finite discrete set. It is customary to take capital letters as random variables and the lower-case as their particular values. We will take $\Pr(x)$ as the probability of x. Then we define the entropy of the probability distribution X to be

$$H[X] \equiv -\sum_{x \in \mathcal{A}} \Pr(x) \log \Pr(x) \tag{2.1}$$

where the logarithm is understood here, and for the rest of this thesis, to be taken at base 2. With this convention, the units of H are *bits*. We can also see that $H \geq 0$ for any probability distribution. The form of equation 2.1 is not arbitrary, but rather it can be shown to follow, up to an overall multiplicative constant, from three quite reasonable assumptions.

1. H is a maximum for a uniform probability distribution. Since this represents the maximum uncertainty we can have about the value of X, it seems only reasonable that H should reflect this lack of knowledge.

2. H should be continuous in the probabilities. Certainly a small change in the probability distribution should result in only a small change in H.

3. Probabilities can be grouped in various ways. We insist that H be only a function of the distribution itself and not how events are grouped in the distribution.

There are several ways one can interpret equation 2.1. One is that H gives the average number of yes-no questions that one needs to ask to determine the value of X, if the questions are chosen optimally. Another is that we can interpret H as the expectation value of the surprise, where the surprise is defined as $(-\log \Pr(x))$. This makes some sense, in the following way. For very common events, $\Pr(x)$ is not too far from one, so $(-\log \Pr(x))$ is small indicating that we are not so 'surprised' when we see x. However, for rare events, $\Pr(x)$ is small making $(-\log \Pr(x))$ large, indicating we are 'surprised' to see x. To get the expectation value of the surprise, we multiply the surprise by the probability of seeing each event, $\Pr(x)$, and sum over events. This is just H. Of course we should not overlook the similarity of equation 2.1 to the thermodynamic entropy S written in terms of canonical probabilities [53] as

$$\mathsf{S} = -k \sum_i P_i \ln P_i \tag{2.2}$$

with $P_i = \frac{\exp^{-\beta E_i}}{Z}$ and $\sum_i P_i = 1$. (E_i, Z, β, and k have their usual thermodynamic meanings here.)

There are several other entropies defined in information theory, and we list them now. They involve two distributions, let us call them X and Y, and take values from the finite sets $\mathcal{A}$ and $\mathcal{B}$ respectively. As usual, we denote the variable with capital letters and the particular value it assumes in lower-case. Let us first fix some notation. We define the joint probability $\Pr(x, y)$ to be the probability that $X = x$ and $Y = y$. We define the conditional probability $\Pr(x|y) \equiv \frac{\Pr(x,y)}{\Pr(y)}$. With these definitions in place, we define the *joint entropy* of two variables as

$$H[X, Y] \equiv - \sum_{(x,y)\in\mathcal{A}\times\mathcal{B}} \Pr(x, y) \log \Pr(x, y). \tag{2.3}$$

We can also define the *conditional entropy* of one variable on another in terms of their joint entropy as

$$H[X|Y] \equiv H[X, Y] - H[Y]. \tag{2.4}$$

The interpretation of $H[X|Y]$ is simple enough. It represents the uncertainty remaining in X once we know Y.

Finally, let us define *mutual information* between two random variables X and Y.

$$I[X; Y] \equiv \sum_{(x,y)\in\mathcal{A}\times\mathcal{B}} \Pr(x, y) \log \frac{\Pr(x, y)}{\Pr(x) \Pr(y)}. \tag{2.5}$$

We can interpret I as the reduction in the uncertainty in one variable due to knowledge of another. We note that I is symmetric in its arguments, as $I[X; Y] = I[Y; X]$.

2.2 Computational Mechanics

Let us now give a brief account of computational mechanics, aiming at providing an intuitive understanding rather than formal discourse. For all of the technical details, proofs, and mathematical rigour, as well as a more detailed and complete disquisition, the reader is referred to [17] [63] [62]. The basic paradigm of computational mechanics is stated easily enough. We assume an observer has access to a one-dimensional data stream (often called a measurement channel) that is produced by some system. The data can be discrete, or we can apply some discretization process to make it

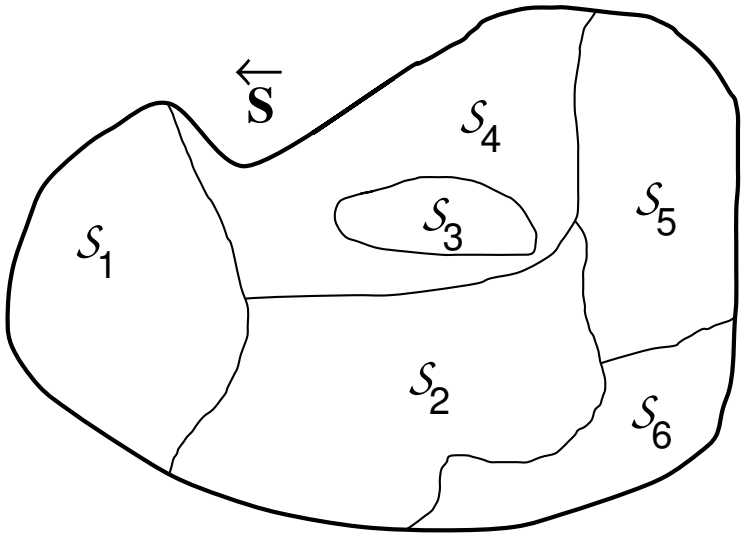

Figure 2.3: A pictorial representation of causal states. We show all possible histories $\overleftarrow{S}$ and their division into causal states $\mathcal{S}_i$. (Figure courtesy of C.R. Shalizi)

so in the case of a continuous source. Then the observer can build up a string of indefinite length of symbols s from some finite alphabet $\mathcal{A}$. This is the only information the observer has about the process producing the data. From this string, what, if any, conclusions or comments can the observer make? Can the observer, for instance, make a model of the process responsible for this data? Computational mechanics provides a technique for doing just this.

First we fix some notation. Let us denote a one-dimensional, bi-infinite chain as $\overleftrightarrow{S}$.

$$\overleftrightarrow{S} \equiv ...S_{-2}S_{-1}S_0S_1... \tag{2.6}$$

where the subscripts on S indicate their placement in the string. We further define semi-infinite chains as $\overrightarrow{S} \equiv S_0S_1....$ and $\overleftarrow{S} \equiv ...S_{-2}S_{-1}$. We call $\overleftarrow{S}$ *histories* or *pasts* and $\overrightarrow{S}$ *futures*. We may wish to speak of a block of variables, and we shall denote this as $S^L = S_0...S_{L-1}$.

We want to construct the minimal model that can explain the data. To do this let us consider the space of all histories. To make an optimal prediction of the future, many details of the past may be irrelevant. We want to forget as many of these irrelevant facts as possible and concentrate only on those details which have been shown to make some difference. Let us put aside how one actually determines what is important and what is not for the moment. We group all histories that have the same important facts together. Since they have everything in common that is important for predicting the future, each one will have the same future (in a statistical sense). So we say they have *equivalent futures*. This division is shown in figure 2.3. The $\mathcal{S}_i \in \mathcal{S}$ are called *causal states*. Simply put, these states keep track of everything that is useful for predicting the future and they forget everything else. It turns out that they can be constructed directly from the data stream itself, and there are two known algorithms for doing this. One is the Crutchfield-Young algorithm [37] for merging past histories, and the other is the Klinkner-Shalizi algorithm [47] that curiously takes the opposite view of splitting histories when something is found that that distinguishes them. Both of these, however, assume a data stream and as we will see, the experimental data used in our present work does not provide information in that form. So we will not discuss the determination the causal states from a data stream here.

We should mention a few properties of causal states. First, by the construction given, we see that they are *minimal*. By this we mean that they have no unnecessary structure in them. To eliminate any structure would be to throw away important predictive information. We could always add more structure, more complexity, but this would not improve the predictive power of the model. It would only mean that we are carrying around more information than we need. Secondly, also by construction, they are *maximally prescient*. Since we have kept all information that is shown to have predictive value, no other formulation can have more predictive power. The causal states have as much predictive power as the underlying process will allow. Lastly, causal states are *unique*. Up to a trivial relabeling of states, the causal states admit no reformulation.

We can now imagine that we have found the states as given in, say, figure 2.3. We have been observing the process for a very long time and let us say we know what causal state we are in. We observe another symbol. This new symbol becomes part of the history, and this new history must belong to some causal state we have listed. (This can be true in the limit of having seen some infinite past. Theorists are allow such luxuries.) So we say that, upon seeing a new symbol, we make a transition to another causal state. That is, we can think of transitions connecting causal states on the emission of a symbol. This is reminiscent of a finite automaton, and in fact we can treat causal states connected in this way as a finite automaton. When we do so, we have have an *ϵ-machine*. Graphically, an ϵ-machine looks just a finite automaton, except that there are conditional probabilities attached to the arcs.

Let us mention a few properties of ϵ-machines. First it turns out that they are deterministic. There is a unique successor state for each causal state on the emission of an allowed symbol. The machine is also *Markovian*. That is, knowledge of the current causal state is sufficient for optimal prediction. We do not need to know the history of past states. Finally, ϵ-machines are the maximally accurate predictors with the minimal statistical complexity (we define statistical complexity below). They are the best one can do, and they invoke the least complexity to do it. It is tempting to think of ϵ-machines as stochastic versions of finite automata, and indeed there are resemblances. But stochastic finite automata need not be minimal, maximally prescient, or deterministic. We do not interpret the nodes to be causal states.

Additionally, ϵ-machines have a unique start state, as do finite automata. Often, the start state is *transient*, that is, in the limit of an infinite string, transient states are causal states that are visited with probability zero. Of course, other states can be transient also. We differentiate these from *recurrent* states, which are those causal states that are visited with a probability greater than zero in the limit of an infinite string. All states in an ϵ-machine are accepting states. Also, we require that the recurrent states be strongly connected, that is, colloquially put, the graph is not allow split into two separate pieces.

The goal of computational mechanics is to reconstruct the ϵ-machine from data. Once done, the ϵ-machine is a model of the system that produced the data, with all of the nice aforementioned properties. It tells us how information is stored and processed by the system. From the ϵ-machine, we can calculate a number of quantities that describe these information storage and use features.

The Source Entropy Rate: Known variously as the thermodynamic entropy density and the metric entropy, h_μ specifies the irreducible randomness produced by a source after correlations are accounted for. It has units of *bits/symbol*. For a completely random source $h_\mu = \log_2 |\mathcal{A}|$ bits/symbol, where $|\mathcal{A}|$ is the size of the alphabet, and completely predictable processes have an $h_\mu = 0$ bits/symbol. We can define h_μ as,

$$h_\mu \equiv \lim_{L \to \infty} \frac{H[S^L]}{L}. \tag{2.7}$$

There are several alternate ways to find h_μ. We can take the limit in equation 2.7, or, we can say

$$h_\mu = H[S_1 | \overleftarrow{S}]. \tag{2.8}$$

If we know the ϵ-machine, we can say,

$$h_\mu = -\sum_{\{\mathcal{S}_i\}} \sum_{s\in\{\mathcal{A}\}} \Pr(s, \mathcal{S}_i) \log \Pr(s|\mathcal{S}_i). \tag{2.9}$$

Yet another way to find h_μ is through the relation

$$h_\mu = \lim_{L\to\infty} \Big(H[S^L] - H[S^{L-1}] \Big). \tag{2.10}$$

We can also define the entropy rate over a finite past as

$$h_\mu(L) \equiv H[S^L] - H[S^{L-1}],\ L \geq 1. \tag{2.11}$$

In general, we find $h_\mu(L)$ to be a monotonically decreasing function of L since more knowledge of the past should decrease the uncertainty of the future.

Total Predictability: When we make a single-symbol measurement, we can think of the information gain as having two contributions. The first is due to the randomness, h_μ, and the second comes from the order or redundancy in the process. We call the latter the total predictability, $\mathbf{G}$, and it is defined as

$$\mathbf{G} \equiv \log|\mathcal{A}| - h_\mu. \tag{2.12}$$

We see that the units of $\mathbf{G}$ must be the same as those of h_μ, namely *bits/symbol*. There are other interpretations of $\mathbf{G}$, and for a much more thorough discussion the reader is referred to reference [17].

Statistical Complexity: The statistical complexity, C_μ, is the minimum average amount of memory needed to statistically reproduce the configuration ensemble to which the sequence belongs. See ref [26]. We find this measure directly from the recurrent portion of the ϵ-machine. C_μ is defined as

$$C_\mu \equiv H[\mathcal{S}] \tag{2.13}$$

$$= -\sum_i \Pr(\mathcal{S}_i) \log \Pr(\mathcal{S}_i). \tag{2.14}$$

Being a measure of memory, C_μ has the units of *bits*.

Excess Entropy: Also called the effective measure of complexity [33], $\mathbf{E}$ measures the total *apparent* memory in a source. We can write this as

$$\mathbf{E} \equiv I[\overrightarrow{S}; \overleftarrow{S}]. \tag{2.15}$$

We can also write the excess entropy as,

$$\mathbf{E} = \sum_{L=1}^{\infty} [h_\mu(L) - h_\mu]. \tag{2.16}$$

The units of the excess entropy are *bits*. It can be shown that $\mathbf{E} \leq C_\mu$.

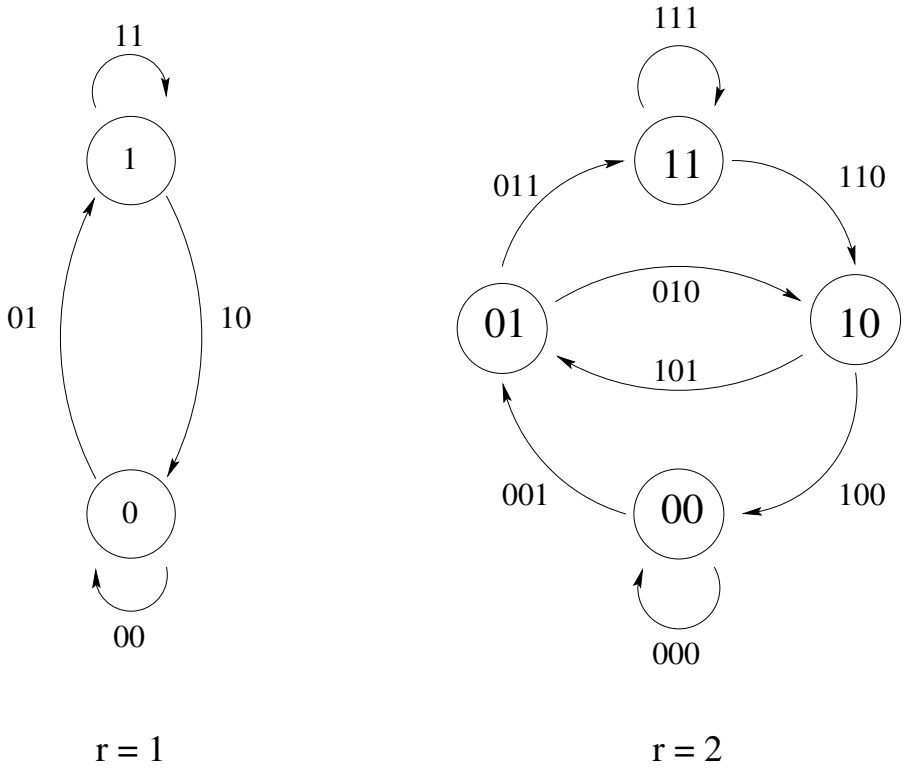

Figure 2.4: Examples of de Bruijn Graphs. The left most graph shows the $r = 1$ de Bruijn graph. This graph explicitly postulates a memory of one, so we can label the nodes by the last symbol seen. Transitions between nodes are labeled with two symbols. The $r = 2$ graph is shown on the left. Note that there are 2^r nodes and 2^{r+1} arcs.

Transient Information: We define $\mathbf{T}$ as the transient information. Basically, $\mathbf{T}$ is a measure of the amount of information an observer must extract from measurements to synchronize to a process. We can write the transient information as

$$\mathbf{T} = \sum_{L=1}^{\infty} L[h_\mu(L) - h_\mu]. \tag{2.17}$$

The units of $\mathbf{T}$ are *bits* $\times$ *symbol*.

Unless noted otherwise, we use equations 2.10, 2.12, 2.14, 2.16, and 2.17 to find the relevant computational quantities.

2.3 De Bruijn Graphs

De Bruijn graphs are **not** a part of computational mechanics, but they will form an important step in our method to reconstruct ϵ-machines. So we shall give an introduction here.

Introduced to provide a useful representation for one-dimensional ground states, de Bruijn graphs are directed graphs with 2^r nodes and 2^{r+1} arcs connecting nodes. To construct a graph, we write down all possible sequences of r spins as nodes and then connect two nodes, $\mathcal{N}_1$ and $\mathcal{N}_2$, with a

Table 2.2: The number of simple cycles N_{SC} in a de Bruijn graph of range r [68]. There is no general expression for the number of simple cycles as a function of r, the last two entries being found using a computer search. It is thought that $N_{SC} \propto r \times 2^{(2^{r-1}-r)}$.

r	N_{SC}
1	3
2	6
3	19
4	179
5	30176

directed arc from $\mathcal{N}_1$ to $\mathcal{N}_2$ if the rightmost $(r-1)$ spins of $\mathcal{N}_1$ are the same as the leftmost $(r-1)$ spins of $\mathcal{N}_2$. The arc is then uniquely labeled with the $(r+1)$ spin sequence. One can then associate a weight to each arc, which is usually taken as an energy. The assumption here is that there is an interaction among spins that has a range of at most r. The total energy of a string of spins may then be found by reading them in sequentially and allowing blocks of length $(r+1)$ to drive transitions between nodes. At each transition, the graph produces an energy, and the sum of all such energies gives the energy of the entire string (neglecting end effects). Since there are 2^r nodes, the longest path that one can see without visiting a node more than once is 2^r. This path is called a "Hamiltonian cycle". Unlike finite automata, there can be no possibility of rejection of a string, as, by construction, each graph allows all possible sequences of spins.

A useful notion arising from these graphs is that of *simple cycles*. We define a simple cycle as a closed, non-intersecting path in the graph. Any arbitrary string of indefinite length may then be written as a series of simple cycles. Each graph has a finite number of simple cycles, and hence we can determine the energy per symbol for each simple cycle, and the one, or if degenerate, few that have the lowest energy per symbol correspond to the ground state. Canright and Watson [12] used graphs of this form to argue that symmetry considerations can lead to degenerate ground states in one-dimensional Hamiltonians without the fine-tuning of parameters. Since each node in a de Bruijn graph is represented by the last r symbols read in, the graph clearly implies a finite memory of range r. We also see that de Bruijn graphs are deterministic in the sense given in §2.1.2. For our purposes then, de Bruijn graphs will give us a convenient way to represent r^{th}-order Markov processes. We shall no longer interpret each weight assigned to an arc as an energy, but as the probability of seeing the word that labels that arc in a string.

The advantages of using de Bruijn graphs are obvious. They resemble finite automata and therefore ϵ-machines, and so the bridge from de Bruijn graphs to ϵ-machines is not far. Each de Bruijn graph of memory r exhaustively represents all possible distinctions among strings of length r. Since we are considering a system which, on physical grounds, we expect to have a finite memory, the approximation of starting with small memory processes and progressively increasing that memory is appealing. Additionally, since crystal structures are just an indefinite repetition of a symbol sequence, we see that this corresponds to a simple cycle on some de Bruijn graph. The smallest graph on which we can express the simple cycle gives the range of the effective interaction that the system must have. The drawback to beginning with de Bruijn graphs, of course, is that we will be confining our search for processes to those which can expressed as part of a de Bruijn graph. If a process is strictly sofic, our attempt to find it this way may mislead us. We will see that a straightforward process such as the even system can be modeled by a finite r machine, but that it is only an approximation.

Chapter 3

Some Practical Details regarding Polytypes and Diffraction Patterns

Let us now abruptly switch gears and turn our attention to some experimental and notational matters concerning polytypes and diffraction patterns. In §3.1 we will discuss the various ways that the stacking sequences of close-packed structures are described; we will give a short account of the experiments we will analyze in §3.2; we will briefly provide some kinematic details in §3.3; we will show how to correct the experimental diffraction patterns for unwanted effects in §3.4; and finally in §3.5 we will list some assumptions necessary to analyze the data.

3.1 Notational Matters concerning Polytypes

There are many ways to express stacking sequences of closed-packed structures. These include Ramsdell notation, *ABC* notation, the Hägg or $\Delta-\nabla$ notation, Zhdanov notation, and the h-k notation [61]. To avoid confusion, we will minimize the notations we use to just few; but since our work is cross-disciplinary, it will be convenient to have several at our disposal. Let us list and discuss each.

ABC notation. An unambiguous way to specify the stacking sequence, the *ABC* notation is applicable for both ordered and disordered sequences. Simply put, this notation specifies the absolute position of each layer in the polytype. We will call the orientation of each layer its *spin* and since the *ABC* notation gives the absolute position for each layer, we will say that A, B or C gives the absolute spin of a particular layer.

Hägg notation. Due to stacking constraints, no two adjacent layers may have the same absolute spin, and we can take advantage of this by introducing relative spins. If the $(n+1)^{\text{th}}$-layer is related to the $(n)^{\text{th}}$-layer by a clockwise rotation ($A \to B \to C \to A$), Hägg denoted this interlayer spin by '+', and counterclockwise rotations ($A \to C \to B \to A$) are labeled with a '-'. We will find this useful, except that we prefer to use '1' and '0' in place of '+' and '-'. Up to a trivial overall rotation of the crystal, the *ABC* notation and the Hägg notation are completely interchangeable.

Ramsdell notation. Applicable only to ordered structures, the Ramsdell notation is a convenient short-hand for particular polytypes. The format is xZ, where x is the repeat length along the stacking direction and Z specifies the the symmetry present, either hexagonal(H), rhombohedral(R), cubic(C), or unknown(L). Since we only discuss in detail three ordered structures, it seems easy enough to just list them. The 3C structure is ...*ABCABC*..., the 2H structure is ...*ABABAB*... and the 6H structure is ...*ABCACB*.... For longer period polytypes, this notation can be quite useful.

ϵ-machine notation. Certainly a non-standard way to describe crystal structure, we can nevertheless specify ordered structures by giving the ϵ-machine or equivalently the finite automaton of the corresponding language. We interpret the output of the ϵ-machine or finite automaton as the Hägg notation for the sequence. Since $h_\mu = 0$ for ordered sequences, the ϵ-machine will precisely specify the structure. For disordered sequences, however, giving the ϵ-machine will reproduce members of an ensemble that all have the same statistical properties. Nonetheless, we will find this quite useful, since for disordered structures it turns out that the diffraction pattern is the same for virtually all members of the ensemble.

We also note that the 3C structure is also called the fcc lattice and the 2H structure is likewise also called the hcp lattice.

A short note on terminology. We use the term 'crystal' in a slightly different way than it is usually defined [46]. Typically, when one speaks of a crystal, the reference is to a basis of atoms attached to a lattice that forms a regular three dimensional array of points in space. The crystal is then invariant under a lattice translation operation. For the disordered samples we discuss, this is clearly not true. However, confining the disorder to one dimension preserves much of the symmetry one expects in a crystal. Lattice translation operations in the plane of the modular layers remain good symmetries, it is only along the stacking direction that this breaks down. And there, the modular layers can take only one of three orientations. The direction of the stacking axis is preserved throughout the specimen. Because of these facts, we can meaningfully define reciprocal lattice vectors. The use the term 'crystal' does provide an important distinction however. It is common to use powder samples in x-ray diffraction experiments, and we certainly are not considering those here.

3.2 A Brief Description of the Experiment

We will examine the diffraction pattern ZnS crystals along the $10.l$ row for seven spectra taken by M.T. Sebastian and P. Krishna [61] [59] [60] [55]. (We give a definition of the $10.l$ row in §3.3.) These crystals were grown in the presence of H_2S gas. Many revealed 2H, 3C, and other structures as well as stacking faults. The crystals were on the order of 0.5 - 2 mm in length and 0.1 - 0.4 mm in diameter. Many of the crystals were further annealed at temperatures ranging from 300 - 1100 C in vacuum for a duration of an hour before being quenched in cold water to arrest the solid transformations taking place. It was found that packing the crystals in ZnS powder helped prevent contamination of the crystals by the walls of the container. The crystals we will study did not show any kinking, a sign of transformations taking place via the periodic slip mechanism [61]. The shape of the crystals was unchanged by the annealing process. They report the intensity of the diffracted radiation along the $10.l$ row, using a MoK_α radiation source, which has a wave length of $\lambda = 0.7107$ Å. The diffracted intensity was recorded in steps of $\Delta l = 0.01$ using a computer controlled four circle single crystal diffractometer, with the crystal and counter held stationary during each observation. The crystals were oriented using the sharp reflections $h - k = 0 \bmod 3$, which are not smeared in the presence of disorder along the stacking direction. M.T. Sebastian and P. Krishna [60] report that the patterns suffer from a gradual change in the scale-factor so that relative intensities far removed from each other can not be compared accurately. The scale-factor changes slowly enough so that the intensity over the individual peak profiles is unaffected by this degradation. We will have more to say about this in later chapters.

3.3 Some Kinematic Details

Let us now write out a few of the equations necessary to describe the scattered wave. Let $\mathbf{k}$ be the wave vector of the incident radiation and $\mathbf{k}'$ be the wave vector of the scattered radiation. For

elastic scattering, we have $|\mathbf{k}'| = |\mathbf{k}|$. The de Broglie relation between the magnitude of the wave vector and the wave length is $|\mathbf{k}| = 2\pi/\lambda$. We write the change in the wave vector as

$$\Delta\mathbf{k} = \mathbf{k}' - \mathbf{k}. \tag{3.1}$$

We have the standard relation between the incident wave vector and the change in the wave vector, namely,

$$|\Delta\mathbf{k}| = 2|\mathbf{k}| \sin\theta. \tag{3.2}$$

where θ is the angle of incidence. We take the standard hexagonal net [71] in the plane of the modular layers, with the usual primitive translation vectors, $\mathbf{a}$ and $\mathbf{b}$ and let $a = |\mathbf{a}| = |\mathbf{b}|$. We define $\mathbf{c}$ to be perpendicular to the hexagonal net and take the magnitude of $|\mathbf{c}| = c$ to be the spacing between modular layers. Let us then write the 'reciprocal lattice vector' as

$$\Delta\mathbf{k} = \mathbf{G} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*, \tag{3.3}$$

where $\mathbf{a}^*$, $\mathbf{b}^*$, and $\mathbf{c}^*$ are primitive translation vectors of the reciprocal lattice. Clearly we do not have a reciprocal lattice any more than we have a lattice structure in real space, but, as in real space, there is crystallinity in the 'reciprocal modular layers.' So, in reciprocal space, as in real space, the disorder is confined to the stacking direction. Since we are interested in the diffraction pattern along the $10.l$ row we have

$$\Delta\mathbf{k} = \mathbf{G} = \mathbf{a}^* + l\mathbf{c}^*. \tag{3.4}$$

We find the magnitude of $\Delta\mathbf{k}$ to be

$$|\Delta\mathbf{k}| = \frac{2\pi}{c}\left[\frac{4}{3}\left(\frac{c}{a}\right)^2 + l^2\right]^{\frac{1}{2}}. \tag{3.5}$$

3.4 How to fix Diffraction Patterns

We want to extract information about the two layer correlation functions from the diffraction pattern, but there are other factors which affect the measured intensity and for which we must account [4] [5] [10] [29] [49] [79]. Let us call the *uncorrected* diffraction pattern $\mathcal{I}(l, Q)$ the raw data as reported by M.T. Sebastian and P. Krishna [61]. Q represents the correlation functions which we are trying to find. We can the write intensity of the uncorrected diffraction pattern as

$$\mathcal{I}(l, Q) = \mathcal{C}(l) \times I(l, Q). \tag{3.6}$$

where $\mathcal{C}(l)$ are diffraction effects not dependent on the correlation functions, and $I(l, Q)$ represents those factors which depend on the correlation functions. Let us call $\mathcal{C}(l)$ *correction* factors and $I(l, Q)$ the *corrected* diffraction pattern. We will detail the dependence of the diffraction pattern on the correlation functions in the next chapter. For now, let us concentrate on the correction factors. We have considered several factors, and we list them now.

Atomic Scattering Factor. The radiation is scattered from the electrons in the atom, and this scattering is dependent on their spacial distribution as well as the wavelength of the incident radiation and angle of reflection. Calculations of these effects are given in reference [36], and we employ them in our work. We call the atomic scattering factor for zinc f_{Zn} and that for sulphur f_{S}.

Structure Factor. We need to account for the fact that the basis is composed of two atoms, one zinc and the other sulphur. This is easily done using the structure factor [46]. Since the two atoms in the basis are separated only along the stacking direction, the structure factor is given by

$$f_{\mathrm{ZnS}}^2 = f_{\mathrm{Zn}}^2 + f_{\mathrm{S}}^2 + 2 f_{\mathrm{Zn}} f_{\mathrm{S}} \cos(3\pi l/2). \tag{3.7}$$

Anomalous Scattering Factors. The atomic scattering factors take into account the distribution of electrons in the atom, but they assume that the electronic binding energy is small compared to energy of the x-rays so that the electrons scatter radiation as if they were free [36]. Additionally, there may be changes in the phase of the scattered wave. These effects can accounted for by the anomalous scattering factors, also called dispersion factors. For our case, we find these to be small, but we have included them nonetheless.

Polarization Factor. The intensity of the radiation scattered from an electron depends on the polarization of the incident radiation and the angle at which the radiation is scattered. For unpolarized radiation, we average over the two polarization components and get a correction factor given by [74]

$$P(\theta) = \frac{1 + \cos^2(2\theta)}{2}. \tag{3.8}$$

Thermal Factors. Thermal vibrations cause the atoms in a crystal to vibrate about their mean position. Making the assumption of small, isotropic vibrations, one can derive the Debye-Waller factor, f_j^{DW} [79], for the j^{th} atom as,

$$f_j^{DW} = \exp^{-2B_j \sin^2\theta/\lambda^2} \tag{3.9}$$

where B_j is the temperature factor for the j^{th} atom. For ZnS, one can just take the average factor to find an approximation for the overall effect. Doing so, one finds $B_{\mathrm{ZnS}} = 0.79$. Typical values of $\sin\theta$ are 0.10 - 0.16 for the Sebastian and Krishna experiments, so the overall intensity is reduced by about 5%, with smaller values of l being less affected. (The actual range is $f_j^{DW} = 0.96 \rightarrow 0.93$ as $l = 0 \rightarrow 1$). We do not include this effect in our calculations.

Absorption Factor. Absorption can decrease the intensity of the x-rays as they pass through the sample. This decrease is an exponentially decreasing function of the distance through which the x-rays must pass. As the angle of incidence changes, this distance, and hence the absorption, changes. It is worthwhile to calculate the linear absorption coefficient, μ [79] [49], whose inverse measures the distance over which the intensity falls by a factor of $1/e$. This coefficient is given by

$$\mu = \rho \sum_i P_i(\mu_i/\rho_i) \tag{3.10}$$

where ρ is the mass density of the material, P_i is the fraction by mass of the i^{th} element in the material and μ_i/ρ_i is the mass absorption coefficient for the i^{th} element in the compound. Using $\rho \approx 4.1$ g/cm^3, $\mu_{\mathrm{Zn}}/\rho = 55.4$ cm^2/g, $\mu_{\mathrm{S}}/\rho = 9.55$ cm^2/g, we find $1/\mu \approx 6$ mm. Since the change in the angle is small in the interval $0 \leq l \leq 1$, the change in the path length and hence the absorption is also small for our crystals, therefore we consider this factor to be ignorable.

Combining the above, we get an overall correction factor $\mathcal{C}(l)$ for the diffraction pattern, and we show this in figure 3.1.

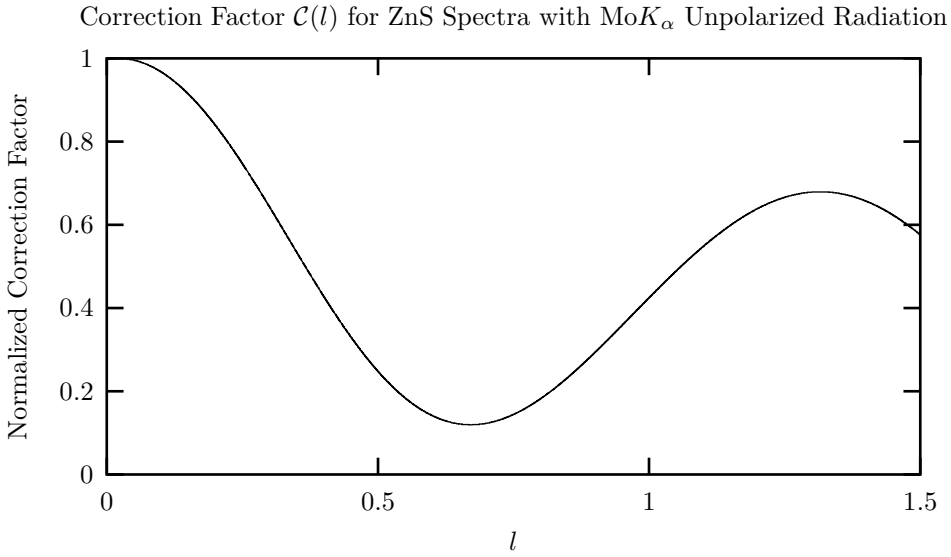

Figure 3.1: The correction factor $\mathcal{C}(l)$ for ZnS diffraction patterns assuming unpolarized radiation of wavelength $\lambda = 0.7107$ Å. The corrections included are for atomic form factors, structure factors, anomalous scattering factors and the polarization factor.

3.5 Some Assumptions

To make the analysis tractable, we employ the following assumptions.

Perfect Modular Layers. We assume the modular layers themselves are undefected. That is, each layer is crystalline in the strict sense, with no point defects, impurities, or distortions in the two dimensional lattice structure.

Scattering Power same for all Layers. We assume that each layer diffracts x-rays with the same intensity. There is no reason to believe that this is not so, unless absorption effects are important or the geometry of the crystal is such that each layer does not have the same cross-sectional area.

Constant spacing between Modular Layers. The spacing between layers can change slightly between polytypes, but this is known to be quite small [58], perhaps about 0.3% between the 2C and the 3H modifications of ZnS. We therefore treat the separation between different modular layers as constant, regardless of the local stacking arrangement.

Entire Layer is Shifted with respect to Neighbors. We assume that the entire modular layer is shifted with respect to its neighbors. This is reasonable, since the stacking fault energy is quite small. Another way to say this, is that we assume that the stacking faults extend all through the crystal.

Spherical Atoms. To calculate the atomic form factors, we make the assumption of neutral, spherical atoms. The bonding in ZnS is at least partially covalent, so we know this is not completely realistic. But since only two of the electrons in a ZnS pair are directly involved in the bonding, the other forty-four being more or less bound by their respective nuclei, we do not consider this to be a large source of error.

Stationary Process. The requirement of stationarity is necessitated by our use of computational mechanics. Simply put, a stationary process is one in which the probability of finding a particular

stacking sequence is the same through out the crystal. We therefore are assuming that the faults are uniformly distributed though out the crystal.

Chapter 4

All about Two-Layer Correlation Functions and Diffraction Patterns

Now that we have corrected the diffraction pattern for the form factor, the polarization factor, etc., we can now examine that part of the diffraction pattern which depends only on the the number of layers and their arrangement. Let us visualize the picture we have. There are N hexagonal, close-packed layers, with each layer occupying one of three positions, A, B or C. These layers are then stacked, and we wish to find the relationship between the stacking order and the diffraction pattern. Let us introduce three quantities, $Q_c(n)$, $Q_r(n)$, and $Q_s(n)$, which we will call the two-layer correlation functions. Let us take c, r, s to stand for cyclic, reverse, and same respectively. $Q_c(n)$ is defined as the probability that any two layers at a separation of n are cyclically related. By cyclic, we mean that if the i^{th} layer is, say A, then the $(i+n)^{\text{th}}$ layer is B. $Q_r(n)$ and $Q_s(n)$ are defined in a similar fashion. Since these are probabilities, $0 \leq Q_\alpha(n) \leq 1$, where $\alpha \in \{c, r, s\}$. It is clear that

$$Q_c(n) + Q_r(n) + Q_s(n) = 1 \qquad \forall \quad n. \tag{4.1}$$

With these definitions, we can write an expression for the diffraction pattern, see Yi and Canright [80],

$$\begin{aligned} I(l) &= \frac{I_0}{N}\Big\{\frac{\sin^2(\pi N l)}{\sin^2(\pi l)} - 2\sqrt{3}\sum_{n=1}^{N}\Big[(N-n)[Q_c(n)\cos(2\pi n l + \frac{\pi}{6}) \\ &\quad + Q_r(n)\cos(2\pi n l - \frac{\pi}{6})]\Big]\Big\}. \end{aligned} \tag{4.2}$$

This expression for the diffraction pattern is only valid for a stacked sequence of two dimensional hexagonal layers. It is easy to see that $I(l)$ is periodic in l with period one, so we need only examine l over the unit interval.

It is instructive to rewrite equation 4.2 by expanding the cosine terms. Doing so, we get,

$$\begin{aligned} I(l) &= \frac{I_0}{N}\Big\{\frac{\sin^2(\pi N l)}{\sin^2(\pi l)} - 3\sum_{n=1}^{N}\Big\{(N-n)\big[(Q_c(n)+Q_r(n))\cos(2\pi n l) \\ &\quad + \frac{1}{\sqrt{3}}(-Q_c(n)+Q_r(n))\sin(2\pi n l)\big]\Big\}\Big\}. \end{aligned} \tag{4.3}$$

The diffracted intensity is now written in terms of orthogonal functions plus an additional term. The Qs then play the role of expansion coefficients. One can also see that if and only if $Q_c(n) = Q_r(n)$ for all n, then $I(l) = I(-l) = I(1-l)$.

We will find it useful to just consider the quantity in the large braces, so let us define a quantity $I'(l)$ such that

$$I(l) = \left(\frac{I_0}{N}\right) I'(l). \tag{4.4}$$

which gives for $I'(l)$

$$\begin{aligned} I'(l) &= \frac{\sin^2(\pi N l)}{\sin^2(\pi l)} - 2\sqrt{3}\sum_{n=1}^{N}\Big[(N-n)[Q_c(n)\cos(2\pi n l + \frac{\pi}{6}) \\ &\quad + Q_r(n)\cos(2\pi n l - \frac{\pi}{6})]\Big]. \end{aligned} \tag{4.5}$$

We can think of $I(l)$ as being the intensity diffracted per modular layer and $I'(l)$ as the intensity diffracted by N layers. Let us call $I'(l)$ the N-layer diffraction pattern and and $I(l)$ simply the diffraction pattern.

4.1 Q-extraction

The form of equation 4.3 suggests that the correlation functions can be pulled out by multiplying both sides by one of basis functions and integrating over a unit interval. Let us define integration of some function $f(l)$ over a unit interval as

$$\begin{aligned} \oint f(l)\,dl &= \int_{l_0}^{l_0+1} f(l)\,dl \\ &= \oint_{l_0} f(l)\,dl \end{aligned} \tag{4.6}$$

with $l = l_0$ specifying where the integration begins. In principle, of course, this does not matter, but for data containing errors, it turns out to make a difference. We will find this useful later on. Let us define $\mathsf{A_n}$ and $\mathsf{B_n}$ as

$$\mathsf{A}_n = \oint I(l)\cos(2\pi n l)\,dl \tag{4.7}$$

and

$$\mathsf{B}_n = \oint I(l)\sin(2\pi n l)\,dl. \tag{4.8}$$

We take N to be large and use the following identities

$$\lim_{N\to\infty}\frac{1}{N}\oint\frac{\sin^2 N\pi l}{\sin^2 \pi l}\,dl = 1,$$

$$\lim_{N\to\infty} \frac{1}{N} \oint \frac{\sin^2 N\pi l}{\sin^2 \pi l} \cos(2\pi n l)\, dl = 1,$$

and

$$\lim_{N\to\infty} \frac{1}{N} \oint \frac{\sin^2 N\pi l}{\sin^2 \pi l} \sin(2\pi n l)\, dl = 0.$$

Carrying out the integrations in equations 4.7 and 4.8, we find

$$\mathsf{A}_n = \frac{3I_0}{2}\left[\frac{2}{3} - Q_c(n) - Q_r(n)\right] \tag{4.9}$$

and

$$\mathsf{B}_n = \frac{3I_0}{2\sqrt{3}}\Big[Q_c(n) - Q_r(n)\Big]. \tag{4.10}$$

We can solve these for $Q_c(n)$ and $Q_r(n)$ to get,

$$Q_c(n) = \frac{1}{3} - \frac{1}{3I_0}\Big[\mathsf{A}_n - \sqrt{3}\mathsf{B}_n\Big] \tag{4.11}$$

and

$$Q_r(n) = \frac{1}{3} - \frac{1}{3I_0}\Big[\mathsf{A}_n + \sqrt{3}\mathsf{B}_n\Big]. \tag{4.12}$$

Finally, we need to find an expression for I_0. We can do this by integrating over the diffraction pattern. We find

$$I_0 = \oint I(l)\, dl. \tag{4.13}$$

4.2 Figures of Merit

We can exploit the fact that, due to stacking constraints, $Q_c(1) + Q_r(1) = 1$. Adding equations 4.11 and 4.12 gives

$$\mathsf{A}_1 = -\frac{I_0}{2}. \tag{4.14}$$

This is always true for error-free data, but we can use this as measure of how corrupted the data is over a particular interval. So let us define

$$\gamma = \frac{\oint I(l)\cos(2\pi l)\, dl}{\oint I(l)\, dl}. \tag{4.15}$$

Ideally, we should always have $\gamma = -\frac{1}{2}$.

We can also derive another constraint that must be obeyed by the correlation functions. It will be demonstrated in a later section that $Q_c(1) = p(1)$, $Q_r(1) = p(0)$, $Q_c(2) = p(00)$, and $Q_r(2) = p(11)$, where $p(x)$ is the probability of seeing the string x. From elementary probability theory, it is easy to show that

$$p(01) = p(10).$$

Using conservation of probability, we can write

$$p(00) + 2p(10) + p(11) = 1.$$

Now, from the relation $p(10) = p(1) - p(11)$, we can say,

$$p(00) + 2p(1) - p(11) = 1.$$

Substituting in the correlation functions for the word probabilities, we have

$$2Q_c(1) + Q_c(2) - Q_r(2) = 1.$$

We are therefore lead to define the quantity β as

$$\beta = 2Q_c(1) + Q_c(2) - Q_r(2). \tag{4.16}$$

We should have $\beta = 1$ for error free data.

So, γ and β are measures of the quality of the data over unit interval. It can be shown that $Q_s(1)$ is not independent of these, nonetheless we shall call these three 'figures of merit' for an interval. Of course, just because these quantities are near the theoretically expected value over some interval does not necessarily guarantee that the interval is free from error, but we take it as an indication. We know that intervals for which these figures deviate from the expected values are somehow 'corrupted.' It therefore seems sensible to choose for analysis an interval where the figures of merit most closely agree with the theoretical values.

4.3 Asymptotic Behavior of the Correlation Functions

Let us now consider the limit of large N. If we integrate $I'(l)$ over one unit interval, all of the terms in the summation vanish, giving a result of N. Since we divide the intensity of the diffraction pattern by N, the total integrated intensity for the diffraction pattern becomes independent of the number of layers. We can further normalize the diffraction pattern such that the total integrated intensity over a unit interval is one. This is useful, as we will be able to compare diffraction patterns for different processes not only qualitatively, but quantitatively. We also see that the total integrated intensity over a unit interval is *independent* of the Qs. In the large N limit we also see that the first term in the N-layer diffraction pattern becomes sharply peaked at $l =$ integer, with a height $\propto N^2$. We might then expect that every diffraction pattern has a strong reflection at $l =$ integer. Examination of the summation term, however, shows that it is possible to get contributions of order N^2 at integer l from it also. Should these contributions cancel, then we could conceivably see no sharp peak at integer l. In evaluating N-layer diffraction patterns, it is helpful to distinguish between terms that grow quadratically, linearly, or are constant in N. Terms that grow quadratically in N contribute to

what we will call Bragg peaks, while terms that grow linearly in N contribute to diffuse scattering. We ignore terms that do not scale with N, as they do not contribute in the many layer limit. We note that there must be some term which grows at least linearly in N, so that the integrated intensity will likewise grow linearly with N. Additionally, no term grows faster than quadratically. It is then helpful to understand under what conditions the summation gives rise to quadratic terms and under what conditions the quadratic terms may cancel. It is likely that experiment will only see structure proportional to the highest power of N. Let us then treat three cases, two of which are of empirical relevance. The first is the case where the Qs decay to some asymptotic value, and the second is the case where they approach some limit cycle of period T_q. As a final case, we mention those processes in which the Qs neither approach an asymptotic limiting value or become cyclic.

4.3.1 Qs decay to an asymptotic value of $\frac{1}{3}$

There seem to be many processes, mostly disordered ones, that can lead to Qs that decay to some limiting value. We prove in the appendix that if the correlation functions assume a constant asymptotic value, that value must be $\frac{1}{3}$. So let us specialize to cases where the Qs reach an asymptotic value of $\frac{1}{3}$ at some $n = n_c$. We will provide a more precise definition of n_c in a later section. We assume that $n_c \ll N$. Typical values of n_c are less than one-hundred.

Let us treat the special case of $l = 0$. The first term in equation 4.3 goes as N^2 and the sine term in the summation vanishes. The argument of the cosine term vanishes, giving a value of one. So we can write the N-layer diffraction intensity as

$$I'(0) = N^2 - 3\sum_{n=1}^{N}(N-n)\big[Q_c(n) + Q_r(n)\big]. \tag{4.17}$$

Since we are interested in how $I'(0)$ scales with N, it is permissible to replace the Qs by their asymptotic values. Carrying out the summation and using the notation Q_c and Q_r for the asymptotic value of $Q_c(n)$ and $Q_r(n)$ for n large, we get to order N^2

$$I'(0) = N^2\big[1 - \frac{3}{2}(Q_c + Q_r)\big]. \tag{4.18}$$

So when the asymptotic values of $Q_c(n)$ and $Q_r(n)$ are $\frac{1}{3}$, there will be no Bragg peak at $l = 0$. Therefore knowing then the asymptotic behavior of the Qs tells us the whether we see a Bragg peak at integer l.

We can also make some statements about the possibility of Bragg peaks for non-integer l. These peaks must come from the summation term in equation 4.3. Let us again restrict our attention to the many layer limit and examine the case of $l \neq$ integer. We can now approximately evaluate equation 4.3. The first term is of order one, so we neglect it. The sine term in the summation again cancels. We are then left with the cosine term. Setting $Q_c(n)$ and $Q_r(n)$ to their asymptotic values, we get,

$$\begin{aligned} I'(l \neq \text{integer}) &\approx -3\sum_{n=1}^{N}(N-n)\Big[\frac{1}{3} + \frac{1}{3}\Big]\cos(2\pi n l) \\ &\approx -3(\frac{2}{3})(-\frac{N}{2}) \\ &\approx N. \end{aligned} \tag{4.19}$$

So, if the Qs approach an asymptotic value of $\frac{1}{3}$, we see that $I'(l) \propto N$ for all l, showing that there are no Bragg peaks in the spectrum at all. This then implies that it is necessary to have Qs that

do not approach $\frac{1}{3}$ to get Bragg peaks, or equivalently Bragg peaks imply long range correlations as reflected in the Qs.

Our motivation then, is to define the diffraction pattern $I'(l)$ in such a way that we divide by the total number of layers N, is so that the integrated intensity is always $\propto I_0$, regardless of the number of layers. This, of course, is why we define $I(l)$ as we do. We can always divide the diffraction pattern by I_0 so that we have *normalized* diffraction patterns. For all corrected diffraction patterns in the rest of this thesis, we give such normalized diffraction patterns when we refer to or plot a unit interval.

4.3.2 Periodic Oscillation in the Qs

Let us now address the case where the asymptotic behavior of the Qs is such that they do not decay to a unique value, but rather approach some limiting cycle of period T_q. That is, we can write,

$$Q_\alpha(n + T_q) = Q_\alpha(n) \qquad \forall \quad n > n_s, \qquad \alpha \in \{c, s, r\}. \tag{4.20}$$

where n_s is the synchronization length. Such behavior is common in crystals. Yi and Canright [80] considered this case and found that one does, in general, get Bragg peaks. It is interesting to note that some disordered stacking sequences can also give rise to a limiting cycle in the Qs and thus have Bragg peaks [80].

4.3.3 Aperiodic Oscillation in the Qs

We have little to say about this case in general. We mention that of all the processes we have examined, only one seems to fall into this category, namely the Thue-Morse process. It seems unlikely that any natural processes will exhibit this behavior. But clearly this deserves more attention.

4.4 Measures of Correlation and Diffraction

We now discuss some measures that relate to the two-layer correlation functions.

Asymptotic Behavior: As we have seen, it is helpful to distinguish between the different kinds of behavior of the two-layer correlation functions in the asymptotic region. We divide the possible cases into three categories: asymptotic decay, periodic oscillation, and aperiodic oscillation.

Correlation Length of the Qs: The correlation length is usually defined as the distance, or in our case the n, at which a correlation function has decayed to $1/e$ of its maximum value. Since our correlation functions, the Qs, are in general oscillatory, and we have three of them, we will use the spirit of the definition but tailor it to our particular needs. Let us imagine first that the Qs decay to an asymptotic value of $1/3$. For those processes which do not, we will say that the correlation length, λ_q, is infinite. For those that do, we assume that the decay is exponential, and we define a quantity $\Psi_q(n)$ as

$$\Psi_q(n) = \sum_{\alpha \in \{c,s,r\}} \left| Q_\alpha(n) - \frac{1}{3} \right|. \tag{4.21}$$

$\Psi_q(n)$ gives some measure of the deviation of the Qs from their asymptotic value. We then say that

$$\Psi_q(n) = B \times 2^{-n/\lambda_q} \tag{4.22}$$

where B is some constant. We then find λ_q from the slope of the line of $\log_2 \Psi_q(n)$ vs. n. Due to the finite size of our sample, we find that we should only use those $\log_2 \Psi_q(n) \geq -5.5$. The statistics are no longer very reliable for $\Psi_q(n)$ smaller than this.

Period of Limit Cycle: Applicable only in the case of periodic oscillation, this quantity gives the length in terms of n, that we must go for the Qs to complete an oscillation. We call this quantity T_q.

Cutoff Length: While not really an intrinsic measure of the Qs, this quantity, which we give the symbol n_c, is the highest n for which we use the calculated values of the $Q_\alpha(n)$ in finding the diffraction pattern. We discuss and motivate a definition for this quantity in a later section.

Scattering Type: We classify diffraction patterns according to whether they exhibit Bragg, pure continuous, or singular continuous scattering. In the event of mixed scattering, we can determine the amount of energy diffracted into each type.

Chapter 5

Examples of Processes, Correlation Functions and their Diffraction Patterns

In this chapter, we attempt to expound on and flesh out the relationships between processes which generate binary sequences, the two-layer correlation functions they imply and the resulting diffraction patterns. All of the systems we consider can represented as a language, or their probabilistic versions as finite ϵ-machines, except the Thue-Morse process. We undertake this exercise to gain intuition into the interplay of these three. Ultimately, our task is to begin with a diffraction pattern and infer the underlying process. That is we want to go diffraction pattern $\Rightarrow$ Qs $\Rightarrow$ process (ϵ-machine). This procedure, which we can think of as a 'backwards procedure', is composed of two parts. First one finds the Qs from diffraction data and second one finds the ϵ-machine which can generate these Qs. The first is easy, while the second is more difficult. In contrast, the 'forward procedure', process $\Rightarrow$ Qs $\Rightarrow$ diffraction pattern, is easy and well defined. It is this forward procedure we address in this chapter. We then will consider about a dozen or so elementary processes and find the resulting correlation functions and diffraction patterns. These sample processes are chosen to give a natural cross-section of possible cases we might encounter in nature, as well as some more exotic cases of theoretical interest. They are, the fair coin toss, the 3C, the 2H, the golden mean process, the 4H, the 3C/2H/4H process, the 6H, the 3C/2H/6H process, a period 14 D-pair, the noisy period two, the even system, the sum zero process, and the Thue-Morse sequence. Many have the convenient property of only requiring at most two recurrent causal states. Except for the last four, all are expressible as SFT. The Thue-Morse sequence can not be written as a finite state machine, but it does give us a chance to examine a pattern that implies an infinite correlation between symbols and has a zero entropy density, while being aperiodic. We will calculate a number of properties for these systems, so that we may compare them.

For all of the following examples, unless stated otherwise, we take a sample of the process 10,000 characters in length to find the diffraction pattern using equation 4.3. All diffraction patterns are normalized to one over a unit interval. To find the correlation functions, we calculate them directly from a sample of the process 400,000 characters long. We take such a large sample to minimize the statistical fluctuations inherent in using a pseudo-random number generator. For completely predictable processes, where we can find the correlation functions analytically, we do so.

It is also perhaps worthwhile to make a comment on terminology. We have repeatedly used the term 'process' when referring to a spacial pattern or arrangement of 0s and 1s. This may indicate that there are some dynamics going on. Computational mechanics has roots in dynamical systems theory, where indeed this is a reasonable implication. For our work, we have used this term

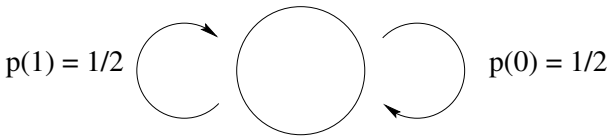

Figure 5.1: The ϵ-machine for an unbiased RNG or a fair coin toss. This is the most random process we can imagine. Every word is allowed, and each word of a given length appears with equal frequency. We see a single, unmarked causal state. We will adopt the convention of labeling states with either their 'memory' in the case of finite memory processes, or capital letters for those with an infinite memory. Since the fair coin toss is a memoryless process, we do not label the single state with a symbol.

for historical reasons, but we do not mean to imply that any dynamics is happening. All of the diffraction pattern we consider are of course static crystal structures. We may just as well substitute the term 'spacial pattern' for 'process' in most cases.

5.1 The Fair Coin Toss

The fair coin toss is a process that generates a completely disordered stacking sequence. We take this process as one in which a 0 or 1 can appear with equal probability, regardless of the the previous symbols. This system thus possesses no memory, and any 'order' or regularity we see in the correlation functions is solely an artifact arising from the stacking constraints. The regular expression for the corresponding language is $\mathcal{R} = (0+1)^*$. The ϵ-machine for the fair coin toss is shown in Figure 5.1.

The fair coin toss is so simple that it is possible to derive recursion relations for the two-layer correlation functions. It is easy to show that

$$\begin{aligned} Q_c(n+1) &= \frac{1}{2}Q_r(n) + \frac{1}{2}Q_s(n) \\ Q_r(n+1) &= \frac{1}{2}Q_s(n) + \frac{1}{2}Q_c(n) \\ Q_s(n+1) &= \frac{1}{2}Q_c(n) + \frac{1}{2}Q_r(n). \end{aligned} \tag{5.1}$$

We note that $Q_c(n) = Q_r(n)$. We can find all the Qs from the initial conditions $Q_c(1) = Q_r(1) = \frac{1}{2}$, $Q_s(1) = 0$. It is possible to solve for the Qs in closed form. It can be shown that

$$Q_c(n) = Q_r(n) = Q_s(n+1) = \frac{1}{3} + \frac{(-1)^{n+1}}{3 \times 2^n}. \tag{5.2}$$

Equation 5.2 shows that the correlation functions decay smoothly and exponentially to the value $\frac{1}{3}$. We can see that this decay is oscillatory, with a period of two in n. Had we biased the fair coin toss by making one of the symbols more likely than the other, the correlation functions would still have decayed to $\frac{1}{3}$, provided each symbol had some non-vanishing probability of appearing. These Equations 5.2 also show that even the Qs corresponding to the fair coin toss have some structure. We suspect that this structure is related to the stacking constraints, and hence is not 'interesting' in the sense of characterizing the order or lack thereof. We show the first few Qs for the fair coin toss in Table 5.1. A graph of $Q_s(n)$ for the fair coin toss is shown in Figure 5.2. A plot of $\log_2 \Psi_q(n)$ versus n is shown in Figure 5.3. The behavior of the $Q_c(n)$ and $Q_r(n)$ is similar to that of $Q_s(n)$.

The diffraction pattern generated by this set of Qs is shown in Figure 5.4. We can see that even for a completely disordered stacking, the diffraction pattern is not completely featureless. While there

Table 5.1: The first few values of the correlation functions for the fair coin toss. We see that even though this is a random process, the correlation functions do not assume their asymptotic value immediately. This is due to the stacking constraints for close-packed layers. We can see that they approach an asymptotic value of $\frac{1}{3}$.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	1/2	1/2	0
2	1/4	1/4	1/2
3	3/8	3/8	1/4
4	5/16	5/16	3/8
5	11/32	11/32	5/16
6	21/64	21/64	11/32

are no Bragg peaks in the spectrum, there is a broad diffuse concentration of scattered intensity centered at $l = 0.5$. The intensity at $l = 0.5$ is nine times that at $l = 0$. We also note that the diffraction pattern is symmetric about $l = 0.5$, as we expect from the equality $Q_c(n) = Q_r(n)$.

We can also consider the artificial case of a completely disordered sequence that is not required to obey any stacking constraints. This is, we allow ..*AA*.., etc. When we do so, we find that $Q_\alpha(n) = \frac{1}{3}$ for all n and α, where $\alpha \in \{c, r, s\}$. The diffraction pattern is just a flat line, that is $I(l) = 1$. So it is the stacking constraints that impose what feature we see in Figure 5.4.

We now make some computational remarks. It is found that the Qs do not settle down to their asymptotic value for large n. They seem to oscillate, and 'fidget' about. We might expect some noise superimposed on the asymptotic values of the Qs due to the finite size of the sample used to determine them. These fluctuations should be on the order of $1/\sqrt{N_s}$ where N_s is the length of the symbol sequence used. For our case, $N_s = 400,000$, giving fluctuations of about 0.002 in the correlation functions. These fluctuations can inflict havoc in the calculated diffraction pattern. It is not hard to see why. Failure of the correlation functions to reach their asymptotic values exactly will cause the summation term in equation 4.3 to become 'unbalanced'. The sum has the potential to be proportional to N^2, if a particular value of l favors Qs slightly higher/lower than $\frac{1}{3}$. Even a small bias in the sum can create terms that alter the 'correct' intensity noticeably. We show a diffraction pattern for the fair coin toss where the Qs have not been forced to their asymptotic value in Figure 5.5. In all subsequent diffraction patterns, where it is clear that the Qs tend to an asymptotic value, we impose that asymptotic value of $\frac{1}{3}$ on all $Q_\alpha(n)$ such that $n \geq n_c$. We define n_c as the smallest n for which $|Q_\alpha(n) - \frac{1}{3}| \leq \delta$. We take δ to be approximately 0.002. The exact value may change slightly with each diffraction pattern, and is somewhat subjective. We also note that not only do our Qs show some noise, but they also seem to have a low frequency oscillation imposed on them. We believe that this is unphysical and somehow a by-product in our random number generator.

Previous researchers have studied the diffraction pattern for randomly stacked close-packed structures. An expression for $I(l)$ in closed form is given by Guinier [34] and agrees with our Figure 5.4.

We make a few remarks concerning some properties of the fair coin toss. It is known to have a statistical complexity and an excess entropy of $C_\mu = \mathbf{E} = 0$ bits and an entropy rate of $h_\mu = 1$ bits/symbol. The total predictability is $\mathbf{G} = 0$ bits/symbol. The transient information is $\mathbf{T} = 0$ bit symbols, implying the an observer need extract no information to synchronize with the source. The Qs display asymptotic decay and have a cutoff length of approximately $n_c = 10$. We can calculate the correlation length exactly from equations 5.2, and obtain a value of 1. We can also determine λ_q from our computed Qs. Using the first six values of $\Psi_q(n)$, we find a correlation length of $\lambda_q = 0.994 \pm 0.004$. We see that our measure for the correlation length for the Qs is not zero, as one might expect for a random process. Indeed, the graph 5.4 implies a memoryless process. It does

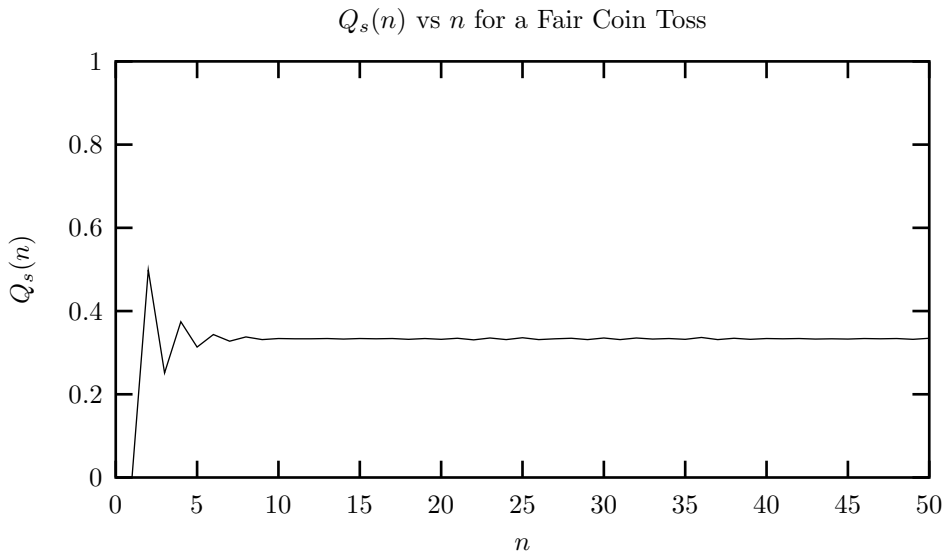

Figure 5.2: $Q_s(n)$ for the fair coin toss as a function of n. We see that even though the process is completely random, the correlation functions have some structure. This structure quickly dies out, though, as the correlation functions decay to their asymptotic value of $\frac{1}{3}$.

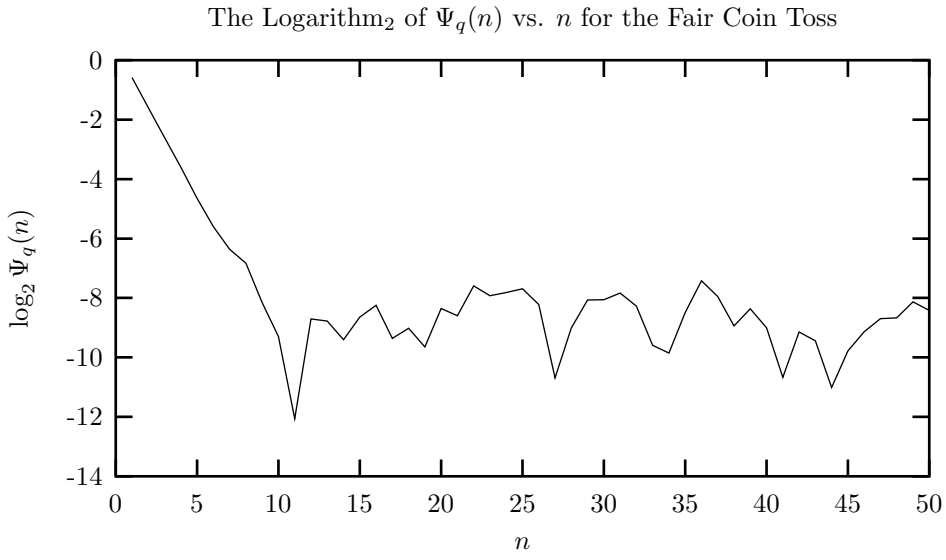

Figure 5.3: The logarithm of $\Psi_q(n)$ vs. n for the fair coin toss. Notice how the values of $\log_2 \Psi_q(n)$ become irregular for $n \geq 8$. Considering the first six n, we find a value for $\lambda_q = 0.994 \pm 0.004$.

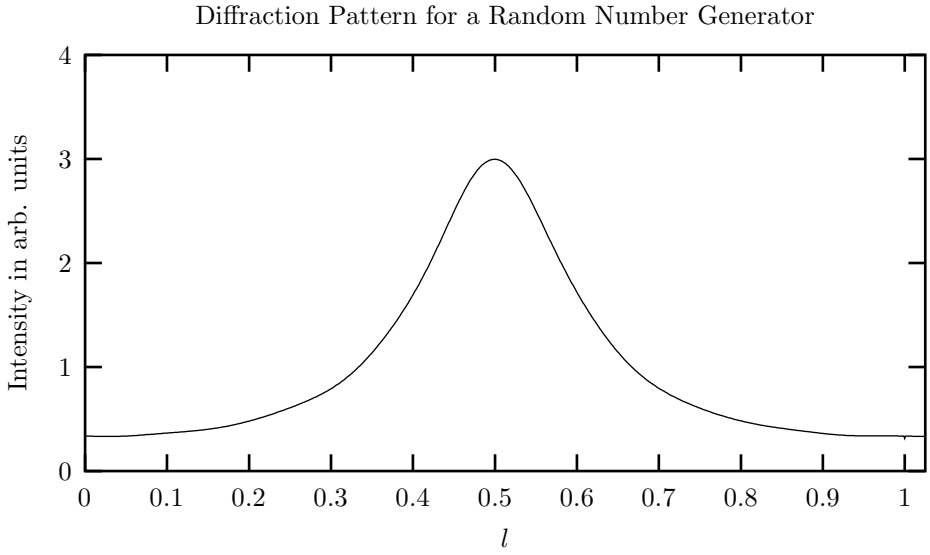

Figure 5.4: The diffraction pattern for a randomly stacked two dimensional hexagonal lattice using asymptotic values for the Qs. We force the correlation functions to their asymptotic value for $n \geq n_c$. Doing so, we get a smooth diffraction pattern.

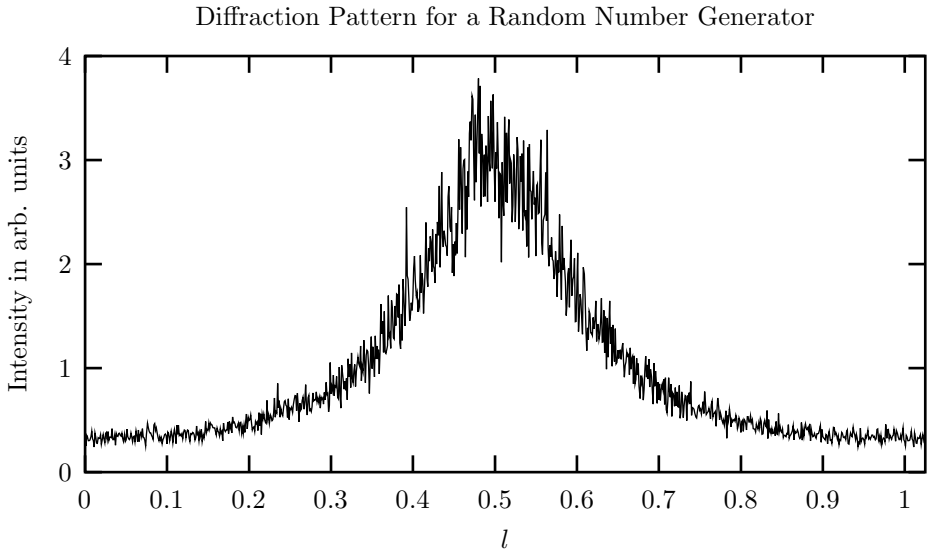

Figure 5.5: The diffraction pattern for a randomly stacked two dimensional hexagonal lattice using sequence calculated values for the Qs. The diffraction pattern is not smooth, and exhibits what appear to be random fluctuations. These fluctuations depend on the length of the sequence used to find the Qs. The longer the sequence, the smaller the fluctuations.

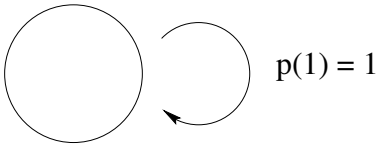

Figure 5.6: The ϵ-machine for the period one process. In contrast to the fair coin toss, this process is completely predictable. It produces the fcc lattice, also called the 3C structure. Like the fair coin toss, this process implies no memory, hence the single causal state is unlabeled.

show the structure imposed on correlation functions by the stacking constraints. The diffraction pattern is entirely diffuse.

5.2 The Period One or 3C Process

We next investigate a language at the other end of the order/disorder spectrum. Like the fair coin toss, the period one process has just one causal state, as shown in Figure 5.6. But this time the machine has no choice but to emit a 1 as each new symbol. In terms of the absolute position of each layer, this automaton generates the $...ABCABC...$ stacking sequence, which we recognize as the fcc crystal. The regular expression for this language is $\mathcal{R} = (1)^*$.

Finding the two-layer correlation functions for a crystal is usually easy. One immediately sees that

$$\begin{aligned} Q_c(n) &= 1 \text{ for } n = 1, 4, 7, \ldots \text{ and 0 otherwise,} \\ Q_r(n) &= 1 \text{ for } n = 2, 5, 8, \ldots \text{ and 0 otherwise,} \\ Q_s(n) &= 1 \text{ for } n = 3, 6, 9, \ldots \text{ and 0 otherwise.} \end{aligned} \tag{5.3}$$

For a crystal, the correlation functions do not decay asymptotically, but are instead periodic. For an 3C process, this period is three. The first few Qs are tabulated in table 5.2. The diffraction pattern for the 3C process is shown in Figure 5.7. We see that there is a single Bragg peak at $l = \frac{1}{3}$, otherwise the spectrum is empty. It can be shown that the Bragg peak at $l = 0$ vanishes. In fact, any process which results in an average value for $Q_s(n)$ of one-third can be shown to have no Bragg peaks at integer l. One can easily see that this is the case for the period one process. We can see then that the Bragg peaks really are quite bright. The diffraction pattern shown in Figure 5.7 has a maximum value of $I_{max} = 10,000$.

We mention in passing that had we chosen our language to be $\mathcal{R} = (0)^*$, the stacking would have had the opposite chirality, and this would resulted in the diffraction pattern having a single Bragg peak at $l = \frac{2}{3}$ instead of at $l = \frac{1}{3}$.

We finally remark on measures of computation, correlation and diffraction. The entropy rate is $h_\mu = 0$ bits/symbol giving a predictability $\mathbf{G} = 1$ bits/symbol. The statistical complexity and the excess entropy are both $C_\mu = \mathbf{E} = 0$ bits. The transient information is $\mathbf{T} = 0$ bit symbols. The Qs have a periodic oscillation with period $T_q = 3$. The correlation length is clearly infinite. The diffraction pattern shows only Bragg scattering with a single diffraction line.

5.3 The Period Two or 2H Process

We now turn our attention to another crystal, the period two process. This crystal corresponds to the the stacking sequence ...101010... which is represented by the regular language $\mathcal{R} = (01)^*$. The

Table 5.2: The first few values of the correlation functions for the period one process. We see that correlation functions are periodic in n with a period of three.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	1	0	0
2	0	1	0
3	0	0	1
4	1	0	0
5	0	1	0
6	0	0	1

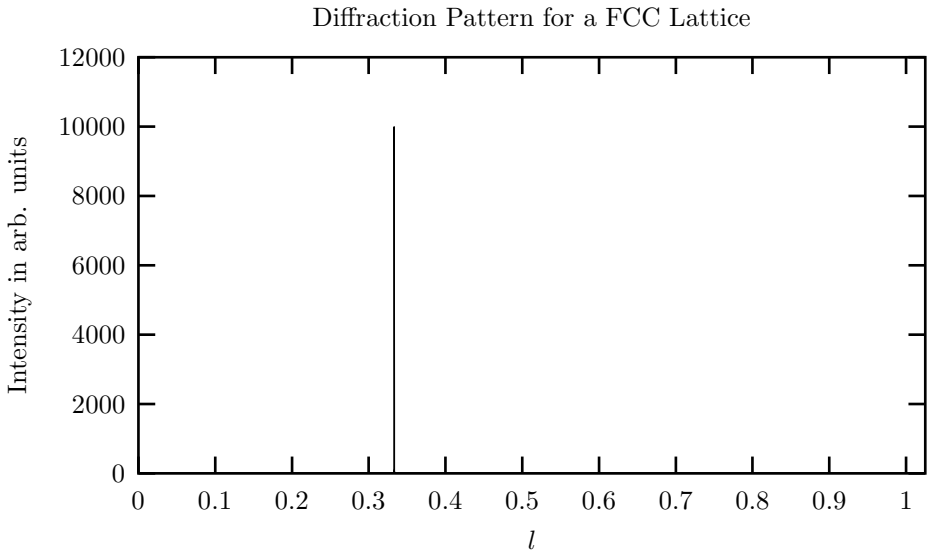

Figure 5.7: The diffraction pattern for the period one or 3C process. We observe a single Bragg peak at $l = \frac{1}{3}$. Everywhere else, including integer l the diffracted intensity is zero.

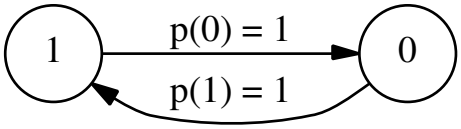

Figure 5.8: The recurrent portion of ϵ-machine for the period two process. We see that the machine just oscillates between the two causal states, alternately emitting a 1 followed by a 0. This process can be written as part of a de Bruijn graph with $r = 1$, so we label the causal states by the last symbol seen.

Table 5.3: The first values of the correlation functions for the period two process. We see that they are periodic in n with period two. Clearly the correlation length for this process is infinite.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	1/2	1/2	0
2	0	0	1
3	1/2	1/2	0
4	0	0	1
5	1/2	1/2	0
6	0	0	1

ϵ-machine for this process is shown in Figure 5.8. Only the recurrent portion of the ϵ-machine is displayed, as the transient state is not physically relevant. In terms of absolute layer positions, the period two lattice is specified by $...ABABAB...$, or equivalently, $...BCBCBC..$, as the naming of the layers is arbitrary. We recognize this as the hcp lattice.

The two-layer correlation functions are easily found to be given by:

$$\left.\begin{array}{ll} Q_c(n) = & Q_r(n) = \frac{1}{2} \\ Q_s(n) = & 0 \end{array}\right\} \text{ for } n \text{ odd,}$$

$$\left.\begin{array}{ll} Q_c(n) = & Q_r(n) = 0 \\ Q_s(n) = & 1 \end{array}\right\} \text{ for } n \text{ even.} \tag{5.4}$$

The first few of these are tabulated in Table 5.3. The diffraction pattern is shown in Figure 5.9. We observe two Bragg peaks in this spectrum, one at $l = 0$ and one at $l = \frac{1}{2}$. The ratio of their intensities is $I(\frac{1}{2})/I(0) = 3.00$. Again the symmetry of the cyclic and reverse Qs implies reflection symmetry in the diffraction pattern about $l = \frac{1}{2}$.

We find that entropy rate is $h_\mu = 0$ bits/symbol and the total predictability is $\mathbf{G} = 1$ bits/symbol. The period two has a statistical complexity and excess entropy of $C_\mu = \mathbf{E} = 1$ bit. The transient information is $\mathbf{T} = 1$ bit symbol. The Qs display periodic oscillation with period $T_q = 2$ and clearly have an infinite correlation length λ_q. The graph implies a minimum range of interaction of $r = 1$. The scattering is completely of the Bragg kind.

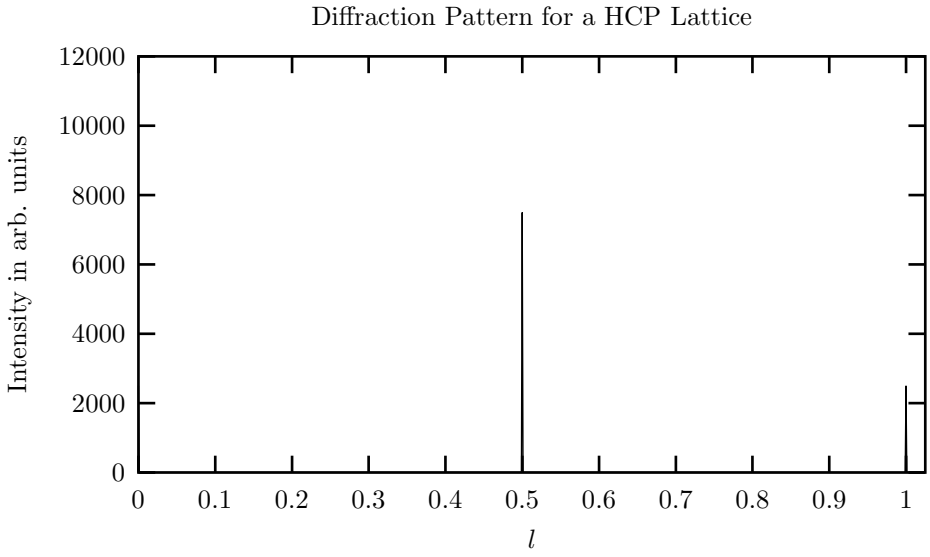

Figure 5.9: The diffraction pattern for a lattice stacked according to the period two rule. We see two Bragg peaks, one at $l = \frac{1}{2}$ and one at $l = 1$.

5.4 The Golden Mean Process

We next consider the diffraction pattern generated by a lattice stacked according to the the golden mean process. Stated simply, the golden mean system allows all sequences that do not contain two consecutive zeros, ie, '00' is an irreducible forbidden word (IFW). We call this the golden mean system because the logarithm of the total number of allowed sequences grows with the word length L at a rate given by the logarithm of the golden mean, $\phi = \frac{1}{2}(1 + \sqrt{5})$. See Crutchfield and Feldman [17]. We treat the probabilistic version of the the system here. We allow 0s and 1s with equal probability, except when the previous symbol is a 0. We insist that 0 always be followed by a 1. The recurrent portion of the ϵ-machine for the golden mean process is given in Figure 5.10. The regular expression for the corresponding language is $\mathcal{R} = (1 + 01)^*$. The first few Qs generated by this process are shown in Table 5.4. We note immediately that the $Q_c(n) \neq Q_r(n)$, suggesting that the diffraction pattern will not have reflection symmetry about $l = \frac{1}{2}$. We also see that $Q_c(2) = 0$. It can be demonstrated that this is a consequence of '00' being a forbidden word. An illustration of the $\log_2 \Psi_q(n)$ versus n is shown in Figure 5.12. One is tempted to think of the golden mean language as some sort of mixture between the period one and the period two languages. Somehow they are competing. We might expect then that the Qs are a compromise of the two. This interpretation seems validated when we examine the first few Qs. For $n = 1, 2$, indeed the Qs for the golden mean lie between those of the period one and period two. At $n = 3$, we see the first departure from this trend. For the golden mean, $Q_c(3)$ is greater than that of either of the other two. The $Q_r(4)$ for both the period one and the period two are zero, while the golden mean gives it a value of 0.500. So we conclude that the golden mean is not some sort of compromise between the two, at least not in such a simple-minded way.

When we examine the $Q_s(n)$ for the golden mean process as shown in Figure 5.11, we see another interesting feature. There appear to be two frequencies superimposed on each other. Let us define the frequency in a natural way as the number of oscillations in $Q_s(n)$ per n. We might expect then that there is a competition amongst frequencies, as the period one process has a natural frequency of one-third, while the period two has a natural frequency of one-half. We recall that the stacking

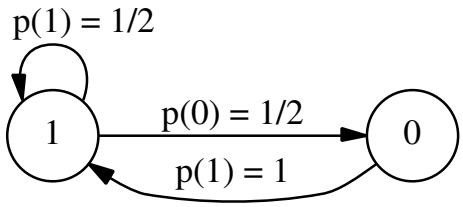

Figure 5.10: The recurrent portion of the ϵ-machine that generates the golden mean process. This is memory one process ($r = 1$), so we label the causal states by the last symbol seen. Notice that when the system is in state '0', the last symbol seen was a 0 and the next symbol must be a 1. In this way, the machine forbids the generation two consecutive zeros; *i.e.*, '00' is a forbidden word.

Table 5.4: The first few Qs generated by the golden mean process. We notice that $Q_c(2) = 0$. It can be demonstrated that this results from '00' being a forbidden word.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.667	0.333	0
2	0	0.334	0.665
3	0.666	0.166	0.167
4	0.084	0.500	0.415
5	0.539	0.126	0.335
6	0.210	0.519	0.271

constraints also impose a frequency of one-half on the $Q_s(n)$.

Looking at the diffraction pattern in Figure 5.13, we can observe that it rather resembles more the period two than the period one diffraction pattern. As suggested by careful comparison of the correlation functions, it certainly does not seem to be some sort of average of the two. We can understand this as follows. If we wish to see a stacking sequence that has properties of each somehow averaged, then we need to see sequences which are largely period one and sequences that are largely period two. To do this, we need a barrier between the simple cycles which represent each process. The two state graph is insufficient to allow this. We will consider a graph in a later section which will be able to accommodate both of these processes.

We note that the golden mean process generates entropy at a rate of $h_\mu = \frac{2}{3}$ bits/symbol. The statistical complexity is $C_\mu = 0.92$ bits and the excess entropy is $\mathbf{E} = 0.25$ bits. The transient information is $\mathbf{T} = 0.25$ bit symbols. Since the golden mean process can be represented by a de Bruijn graph of memory $r = 1$, we see that the minimum range of interaction is just 1. Nonetheless, we find a rather long correlation length of 4.48 ± 0.06. The diffraction pattern is completely diffuse and the Qs decay to an asymptotic value of $\frac{1}{3}$.

5.5 The 3C/2H Process

We now consider a process that will represent a genuine cross between the 3C and the 2H process. As we have seen with the golden mean process, in order to realize this, it is necessary to have long stretches of 3C and long stretches of 2H. To achieve this, there needs to be a barrier between the two cycles that limits capricious alteration between them. The graph shown in Figure 5.14 does just this. We note that since this process can be represented as part of de Bruijn graph of range $r = 2$,

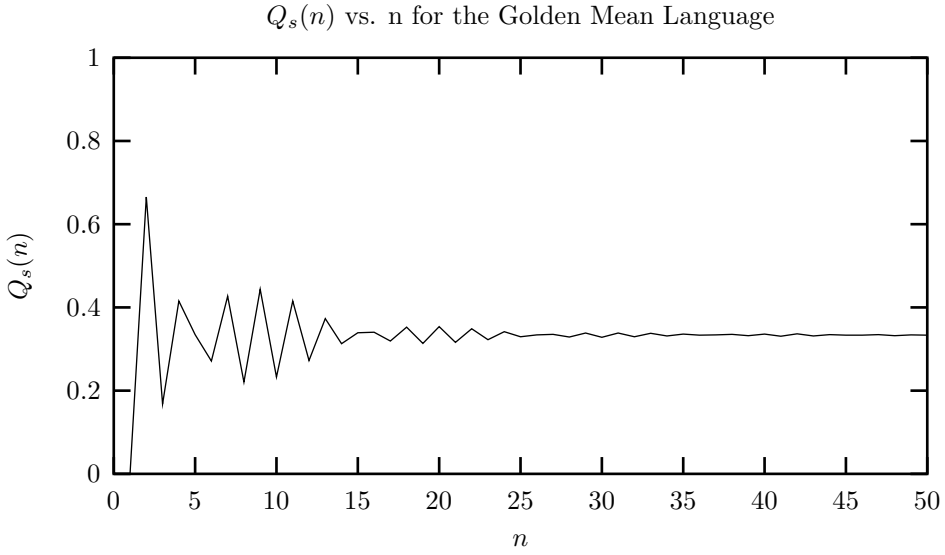

Figure 5.11: $Q_s(n)$ for the golden mean process a function of n. The correlation functions for the golden mean process decay to $\frac{1}{3}$ with a correlation length of $\lambda_q = 4.48 \pm 0.06$.

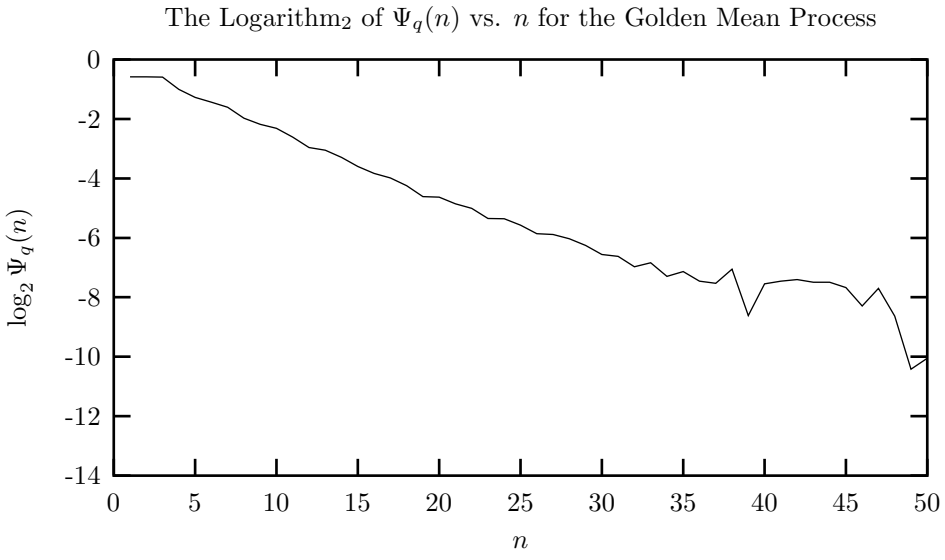

Figure 5.12: The logarithm of $\Psi_q(n)$ for the golden mean process as a function of n. Using the first twenty-five values of $\Psi_q(n)$, we get $\lambda_q = 4.48 \pm 0.06$.

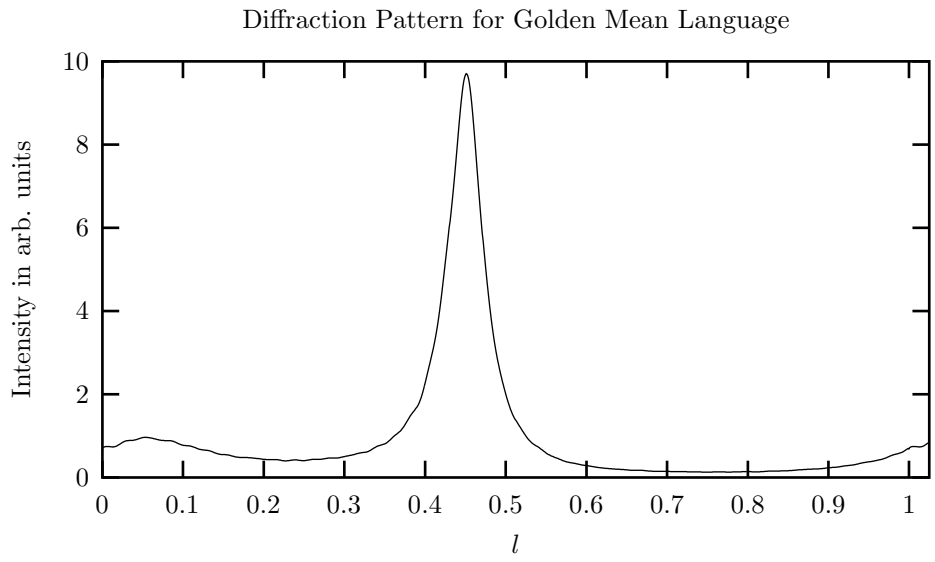

Figure 5.13: The diffraction pattern for a lattice stacked according to the golden mean process. We see that the scattering is entirely diffuse.

we can label the causal states by the last two symbols seen.

Let us examine the behavior of the process as the parameter q is changed from 0.01 to 0.40. The first few correlation function for the 3C/2H process with $q = 0.01$ are shown in Table 5.5. A plot of $Q_s(n)$ versus n for the first fifty n is given in Figure 5.15. We are tempted to make the following analysis of the graph. It can be shown that, regardless of the value of q, $p(11) = p(10) = p(01) = \frac{1}{3}$. For the particular value of $q = 0.01$, we have $p(111) = 0.330$, $p(101) = 0.333$, and $p(010) = 0.330$. Identifying $p(111) = 0.330$ as the 3C cycle and $p(101)+p(010) = 0.663$ as the 2H part, it is possible to associate one-third of the graph as 3C and the other approximately two-thirds as 2H. The remaining two arcs consume less than one percent of the probability weight. So we might expect that the resulting correlation functions and diffraction pattern will be a weighted average of the two. Indeed, we find that

$$Q_\alpha^{3C/2H}(n) = \frac{1}{3}Q_\alpha^{3C}(n) + \frac{2}{3}Q_\alpha^{2H}(n) \tag{5.5}$$

to a good approximation, where $Q_\alpha^{3C/2H}(n)$ are the correlation functions for the 3C/2H process with $q = 0.01$, $Q_\alpha^{3C}(n)$ are the correlation functions for the 3C process, and $Q_\alpha^{2H}(n)$ are the correlation functions for the 2H process, with $\alpha \in \{c, r, s\}$. Equation 5.5 is only valid for small n, where the correlation functions have not begun decaying to their asymptotic values. The diffraction pattern for the 3C/2H process with $q = 0.01$ is shown in Figure 5.17. We see that there are strong reflections at $l = \frac{1}{2}, 1$ which we associate with the 2H structure and a single reflection at $l = \frac{1}{3}$, which we can associate with the 3C structure. The picture, then, of two stacking sequences, each having long sections of their period structure interrupted by the other seems justified.

We can now ask what happens as q is increased. Figure 5.18 shows the diffraction pattern for the 3C/2H process with $q = 0.05$. We see an overall weakening of the peaks, but they remain separated and could reasonably be interpreted as a 3C crystal interspersed with a 2H crystal. At $q = 0.10$, the picture begins to break down. In Figure 5.19 we that the reflection at $l = \frac{1}{2}$ is slightly shifted to the left, and the there is noticeable diffuse scattering between the peaks at $l = \frac{1}{3}$ and $l = \frac{1}{2}$. At $q = 0.20$, the picture of two interspersed crystals seems to have broken down. Figure 5.20 shows the

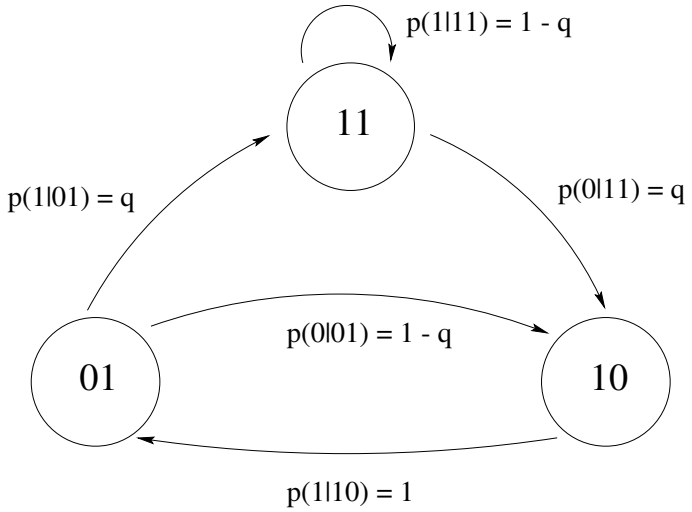

Figure 5.14: The recurrent portion of the ϵ-machine that generates the 3C/2H process. This process has memory of two, so we can represent it with a portion of an $r = 2$ de Bruijn graph. The causal states are labeled by the last two symbols seen.

Table 5.5: The first few Qs generated by the 3C/2H process, with $q = 0.01$. As for the golden mean process, a vanishing value of $Q_c(2)$ implies that '00' is a forbidden word, as can be seen from the ϵ-machine.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.668	0.332	0
2	0	0.336	0.664
3	0.339	0.329	0.332
4	0.329	0.013	0.658
5	0.336	0.651	0.013
6	0.013	0.013	0.973

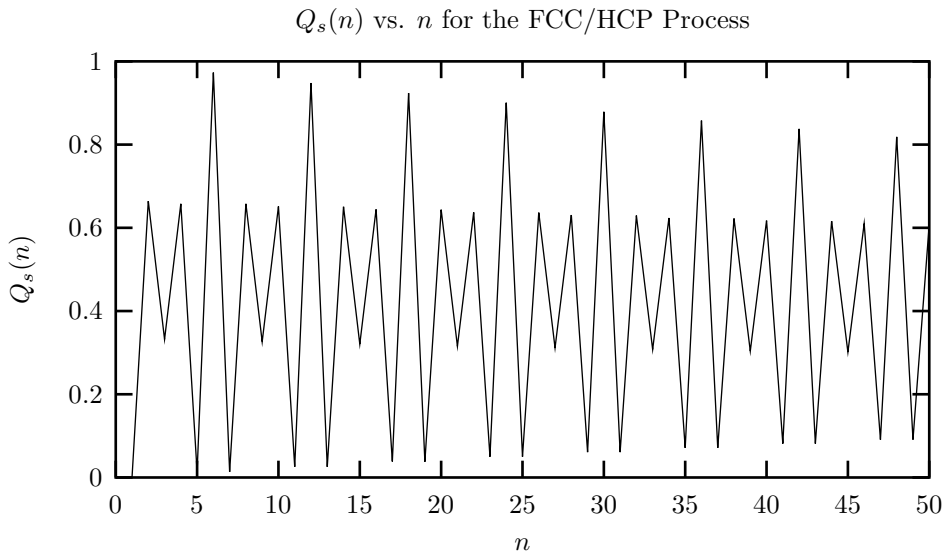

Figure 5.15: $Q_s(n)$ vs. n for 3C/2H process with $q = 0.01$. We see a slow decay in the correlation functions and find a correlation length of $\lambda_q = 123 \pm 2$. There seems to be a period six in correlation functions, which is not so unsuspected since this process has strong components from a period two and a period three process.

diffraction pattern for $q = 0.20$, and while we still see two peaks near the $l = \frac{1}{3}$ and $l = \frac{1}{2}$, there is no clear separation between them. The last two figures in this series, Figure 5.21 and Figure 5.22 show no recognizable 3C or 2H structure. This, is of course, not surprising, as there remains little barrier between the two simple cycles representing these two crystals. It seems dubious to try to interpret them as 3C or 2H structures any longer. The lesson we see from this series is that in order to preserve the picture of two interspersed crystal structures, we must have a barrier, or arcs with small probability, to separate the two structures. In this event, it is meaningful to think about interspersed crystal structure. But, as this barrier deteriorates, the picture breaks down. Also, simple examination of the diffraction pattern seemingly says little about how the sequence flips between the two crystal structures. There are any number of possible intermediate states between the two crystals. Sebastian and Krishna [61] claim that by studying the profile of the peaks, one can gain information about this. But, their analysis relies on the assumption of some sort of parent crystal, which, as we see in the event of even moderate disorder, may not be tenable.

A plot of the logarithm of $\Psi_q(n)$ versus n for the 3C/2H process with $q = 0.01$ is shown in Figure 5.16. We find a correlation length of $\lambda_q = 123 \pm 2$. Other computational and correlative measures are given in Table 5.15 and Table 5.16.

5.6 The 4H Process

We now treat a simple periodic system that occurs in mineral ZnS. This crystal is known as 4H in crystallography, and since there is no name for it in computational mechanics, we will call the process the 4H process. The regular expression for the corresponding language is $\mathcal{R} = (0011)^*$. It can be represented by part of an $r = 2$ de Bruijn graph as shown in Figure 5.23. In terms of a sequence of absolute layer positions, we would see a crystal described by $...ABCBABCBABCB...$.

The first few values of the correlation functions are given in Table 5.6. The correlation functions are periodic in n, with a period of $T_q = 4$. The diffraction pattern for this process is shown in

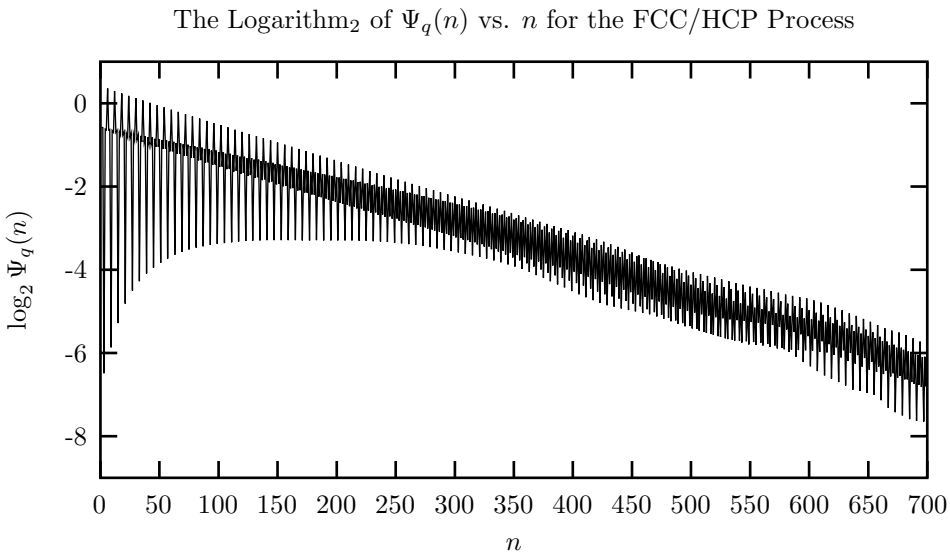

Figure 5.16: The logarithm of $\Psi_q(n)$ for the 3C/2H process as a function of n, with $q = 0.01$. We find a correlation length of 123 ± 2.

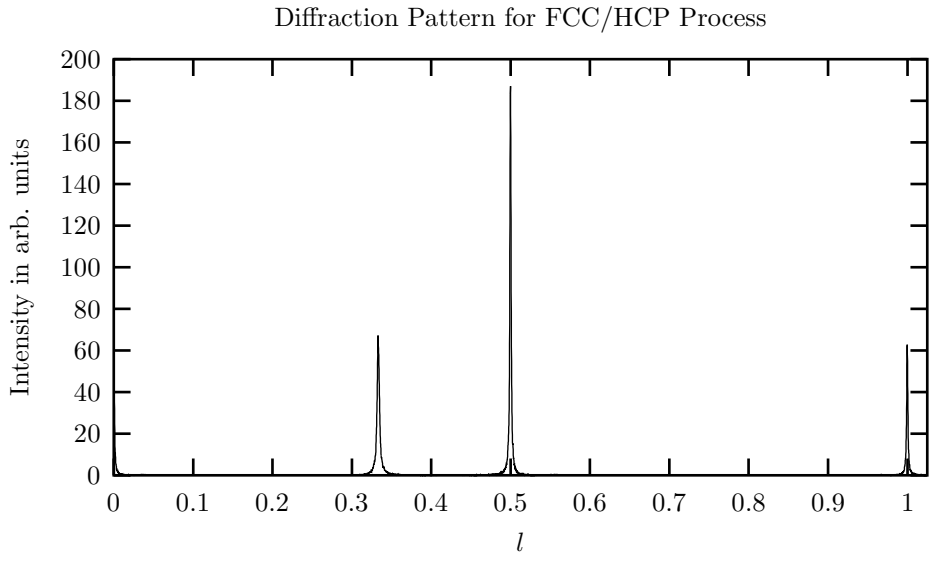

Figure 5.17: The diffraction pattern for a lattice stacked according to the 3C/2H process with $q = 0.01$. We see three sharp peaks in the spectrum, one at $l = \frac{1}{3}$ corresponding to the 3C structure and the other two at $l = \frac{1}{2}$ and 1 corresponding to the 2H structure.

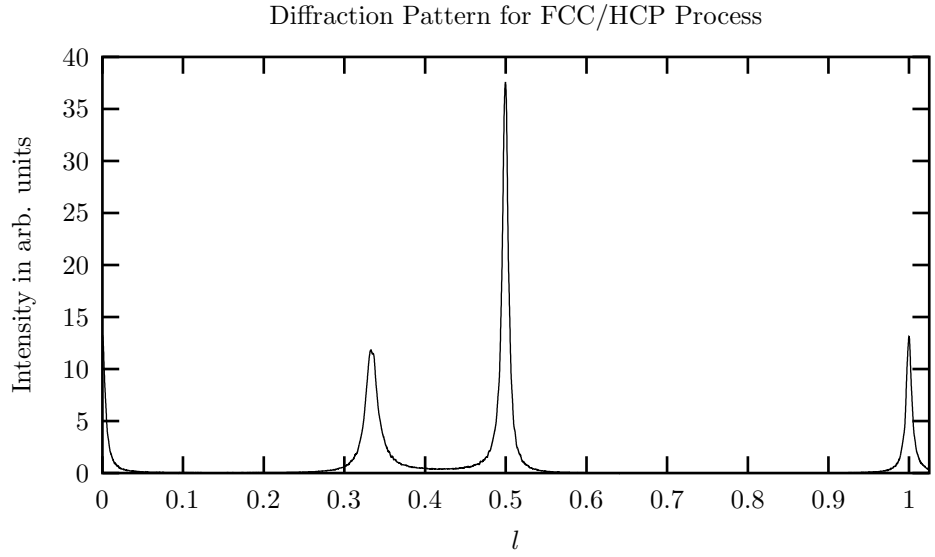

Figure 5.18: The diffraction pattern for a lattice stacked according to the 3C/2H process with $q = 0.05$. As we increase q, the disorder in the system increases causing shorter sequences of pure 3C and 2H. This results in a broadening of the peaks as well as well as diminishes their maximum intensity. They are still reasonably well separated though.

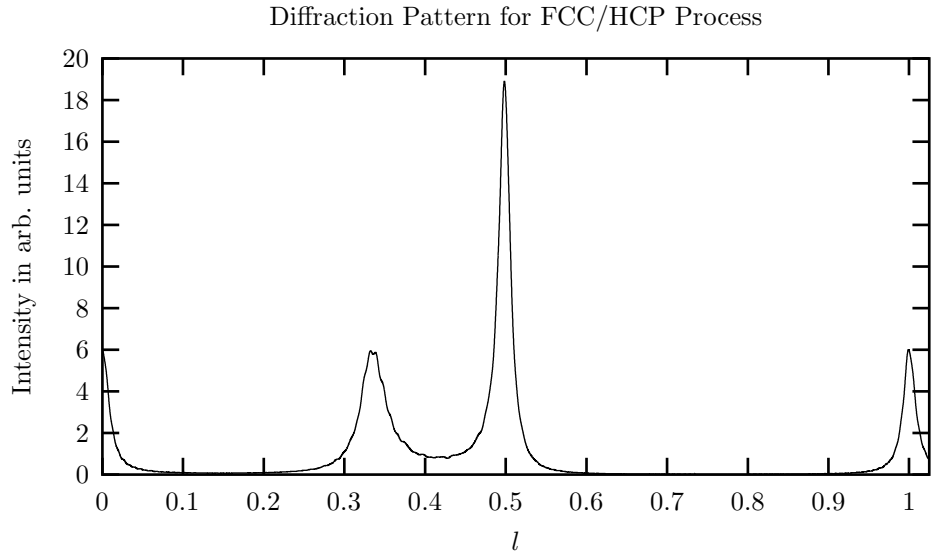

Figure 5.19: The diffraction pattern for a lattice stacked according to the 3C/2H process with $q = 0.10$. We see continued broadening as the q increases, and the picture of two interspersed crystal structures is breaking down.

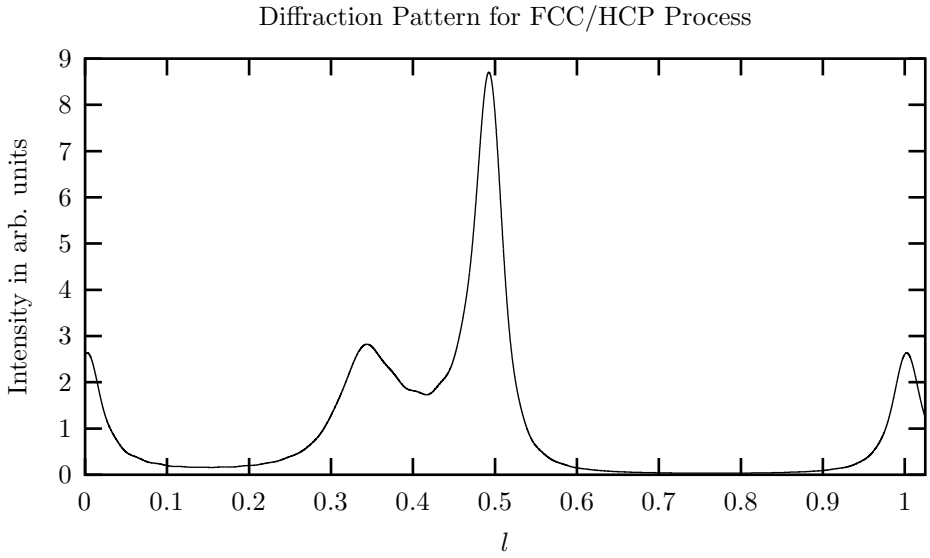

Figure 5.20: The diffraction pattern for a lattice stacked according to the 3C/2H process with $q = 0.20$. There is no longer clear separation between the two peaks at $l = \frac{1}{3}$ and $\frac{1}{2}$. We also see the peaks beginning to shift.

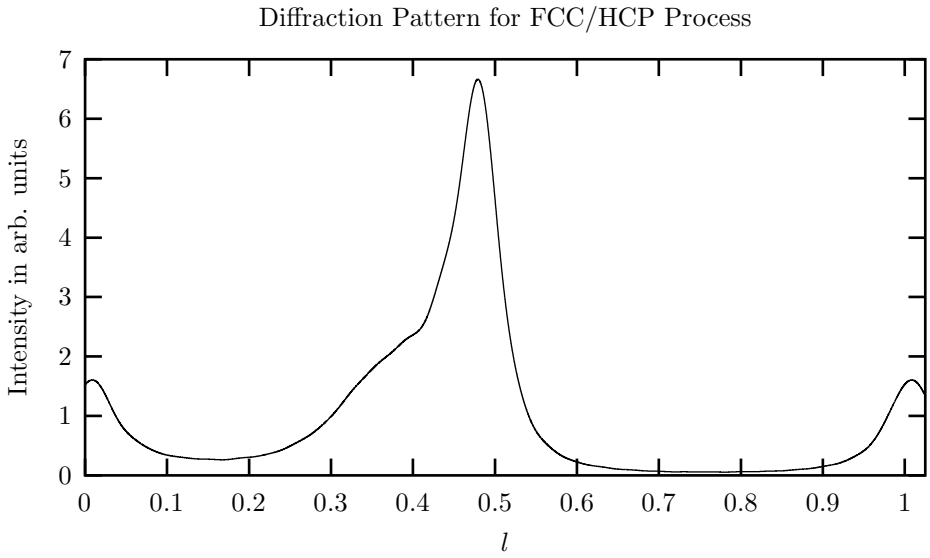

Figure 5.21: The diffraction pattern for a lattice stacked according to the 3C/2H process with $q = 0.30$. The peak corresponding to the 3C has almost completely disappeared, being absorbed into the peak at $l \approx \frac{1}{2}$, which has shifted noticeably to the left.

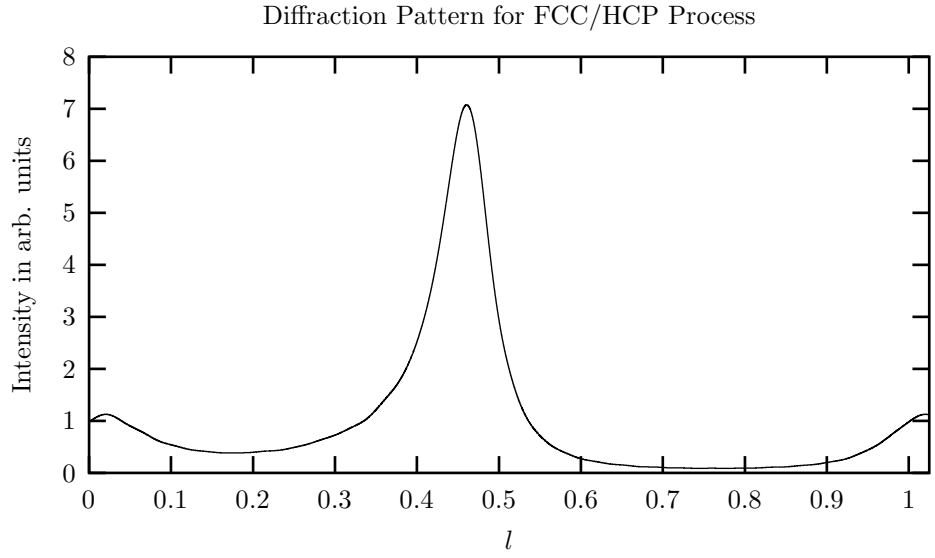

Figure 5.22: The diffraction pattern for the 3C/2H process with $q = 0.40$. Notice the resemblance of the spectrum to the golden mean spectrum. In chapter 6 we will explain why this is so.

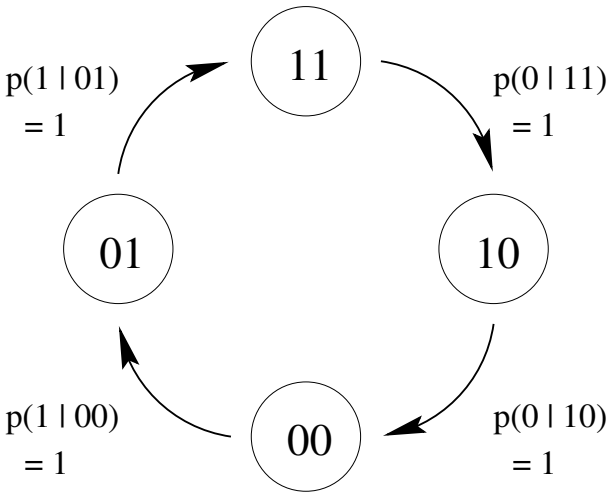

Figure 5.23: The recurrent portion of the ϵ-machine that generates the 4H process. We see four causal states labeled by a history of two. This graph is part of an $r = 2$ de Bruijn graph.

Table 5.6: The first few Qs generated by the 4H process. The correlation functions are periodic in n with period two. The correlation length is infinite.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.500	0.500	0
2	0.250	0.250	0.500
3	0.500	0.500	0
4	0	0	1.000
5	0.500	0.500	0
6	0.250	0.250	0.500

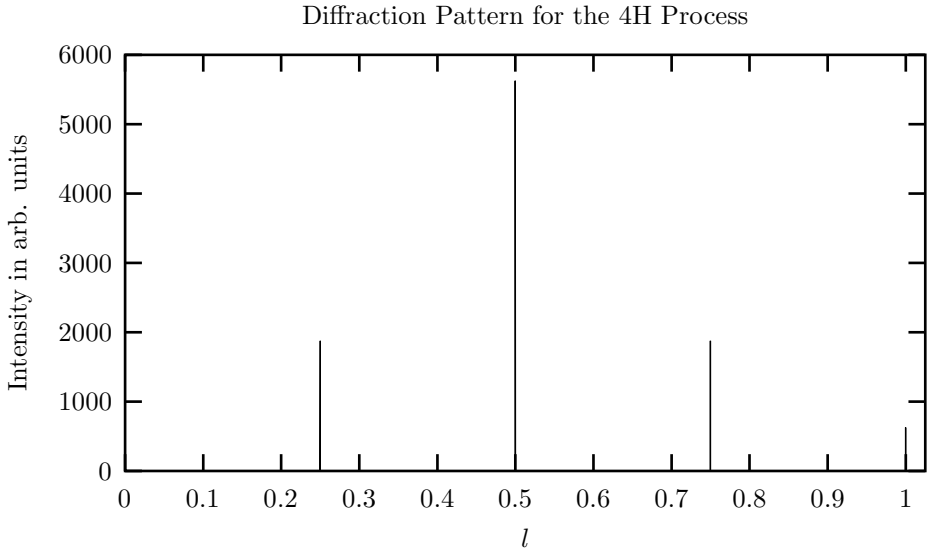

Figure 5.24: The diffraction pattern for a lattice stacked according to the 4H process. We see a completely point spectrum with four equally spaced Bragg peaks at $l = \frac{m}{4}$ with $m = 1, 2, 3, 4$.

Figure 5.24. We see four Bragg peaks, at $l = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, and 1. Since this a crystal, we have $h_\mu = 0$ and the predictability $\mathbf{G} = 1$ bits/symbol. We find that the excess entropy and the statistical complexity are $\mathbf{E} = C_\mu = 2$ bits. The transient information is $\mathbf{T} = 3$ bit symbols. The correlation length is infinite.

5.7 The 3C/2H/4H Process

We now consider a process that represents, at least roughly, three crystal structures interspersed. We treat this case to find out what the diffraction pattern for say, a crystal in the midst of a transformation from 2H to 3C via an intermediary 4H structure might look like. There is some experimental evidence that this might be important in the transition between 3C and 2H structures on annealing at sufficiently high temperatures [28]. The conditional probabilities attached to the states are not necessarily intended to be realistic, but we hope that they are not so different from what one might expect in Nature.

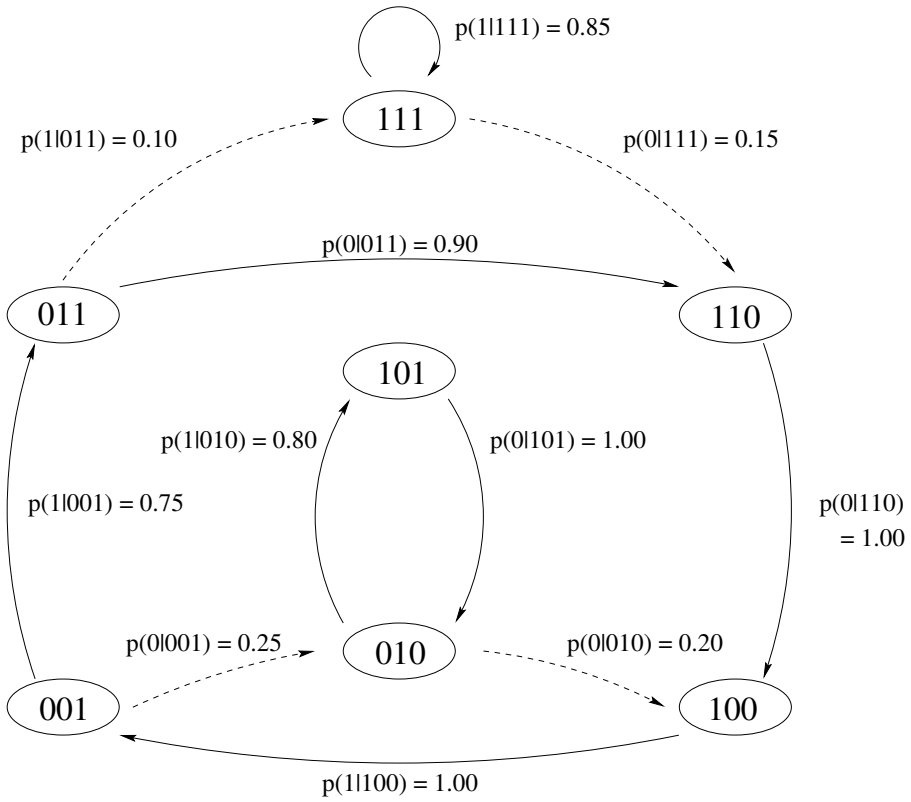

Figure 5.25: The recurrent portion of the ϵ-machine that generates the 3C/2H/4H process. This process is expressed as a portion of an $r = 3$ de Bruijn graph so that no two of the simple cycles corresponding to crystal structure share a causal state. The dashed lines indicate relatively weak transitions between crystal structures and the solid lines are transitions within a crystal structure.

The recurrent portion of the ϵ-machine for the 3C/2H/4H process is shown in Figure 5.25. The dashed arcs are meant to indicate weak transitions between the nodes they connect. We see clearly that there are three simple cycles denoted by nodes connected with solid lines. Even though the each of these simple cycles can be represented on a graph of $r = 2$ or less, to have barriers between cycles we need a graph of $r = 3$. The first few correlation functions for this process are shown in Table 5.7. A graph of $Q_s(n)$ versus n for the first fifty n is shown in Figure 5.26. We see that the correlation functions decay to an asymptotic value of $\frac{1}{3}$ and we find a correlation length of $\lambda_q = 19 \pm 1$.

The diffraction pattern for this process is shown in Figure 5.27. We see a strong reflection at $l \approx \frac{1}{2}$ and several smaller reflections at $l \approx \frac{1}{4}, \frac{1}{3}, \frac{3}{4}$, and 1. This is, of course, not surprising, since these are just the l values at which at least one of the parent structures has a Bragg peak. While this is only one example, the fact that there is a loose correspondence between some kind of underlying collection of crystal structures and the placement of small "bumps" in the spectrum is comforting. Simple examination of the diffraction pattern may provide some insight into possible candidates for parent crystals, but again, examination is not quantitative. Although one could integrate under the peaks associated with each crystal structure in an attempt to find the fraction of each parent crystal present, this only seems feasible in the limit of small disorder such that the peaks are well separated [27]. Also, in the absence of some kind of analysis, the mechanism of switching between crystal structures is not obvious.

We note measures of correlation and computation now. We find that the entropy density is

Table 5.7: The first few Qs generated by the 3C/2H/4H process. These correlation functions decay asymptotically to $\frac{1}{3}$ and have a correlation length $\lambda_q = 19 \pm 1$.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.556	0.444	0
2	0.148	0.259	0.593
3	0.370	0.481	0.149
4	0.200	0.044	0.755
5	0.389	0.566	0.045
6	0.203	0.168	0.629

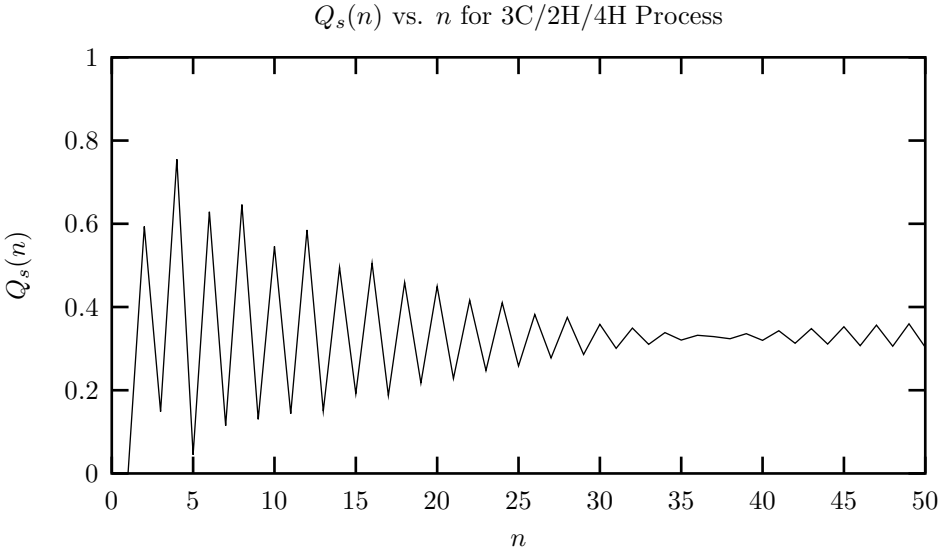

Figure 5.26: $Q_s(n)$ vs. n for 3C/2H/4H process.

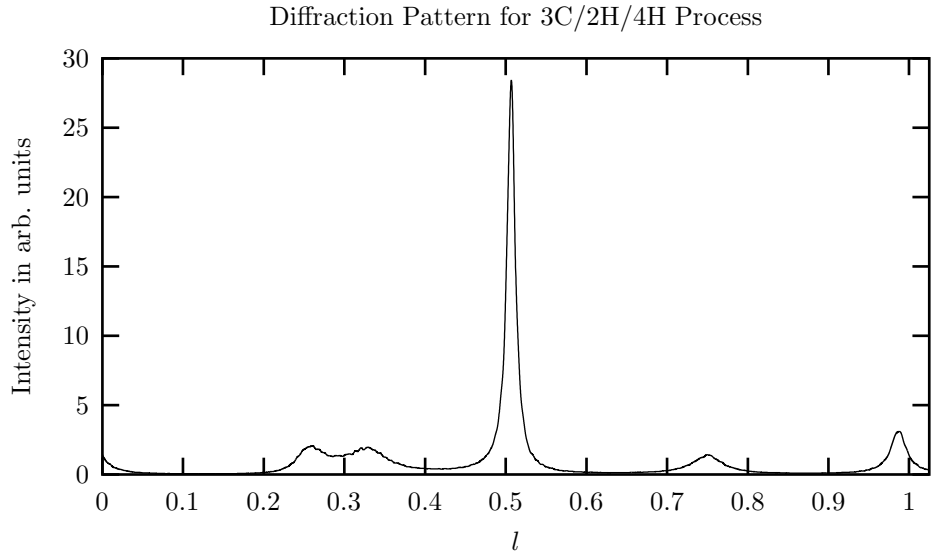

Figure 5.27: The diffraction pattern for a lattice stacked according to the 3C/2H/4H process. We see a strong, slightly shifted peak at $l \approx \frac{1}{2}$ and several smaller 'bumps' at $l \approx \frac{1}{4}, \frac{1}{3}, \frac{3}{4}$ and 1. We interpret the structure at $l = \frac{1}{4}$ and $\frac{3}{4}$ as due primarily to the 4H stacking present.

$h_\mu = 0.424$ bits/symbol, and the statistical complexity $C_\mu = 2.79$ bits. The excess entropy is $\mathbf{E} = 1.51$ bits and the transient information is $\mathbf{T} = 2.87$ bit symbols.

5.8 The 6H Process

We now treat a crystal structure that is expected to be important in ZnS. Researchers have reported that the transition from 2H to 3C upon annealing ZnS crystals proceeds via an intermediate 6H structure. It is therefore useful to have some experience with this crystal. The regular language for this process is $\mathcal{R} = (111000)^*$. In terms of the absolute layer positions, it can be written as $...ABCACB...$. The ϵ-machine for the 6H process is given in Figure 5.28. We immediately notice that the smallest de Bruijn graph that can accommodate this structure is $r = 3$. The first few values of the correlation functions for this process are shown in Table 5.8.

The diffraction pattern for the 6H process is shown in Figure 5.29. We notice the scattering is entirely of the Bragg kind, and there are five Bragg peaks, at $l = \frac{m}{6}$, where $m = 1, 2, 3, 4, 5$. There are no Bragg peaks at integer l, because the average value of $Q_s(n)$ is $\frac{1}{3}$. We also see that the spectrum is symmetric about $l = \frac{1}{2}$. We find that the entropy density is $h_\mu = 0$ and the statistical complexity and excess entropy are given by $C_\mu = \mathbf{E} = 2.59$ bits. The transient information is $\mathbf{T} = 4.84$ bit symbols.

5.9 The 3C/2H/6H Process

There is experimental evidence to suggest that solid state transformations between the 2H structure and the 3C structure in ZnS occur via an intermediary 6H phase [61]. We therefore examine a stacking sequence that, while having some significant disorder, has sections that be thought of as containing some of the 2H, 6H, and 3C character. The ϵ-machine for the 3C/2H/6H process is

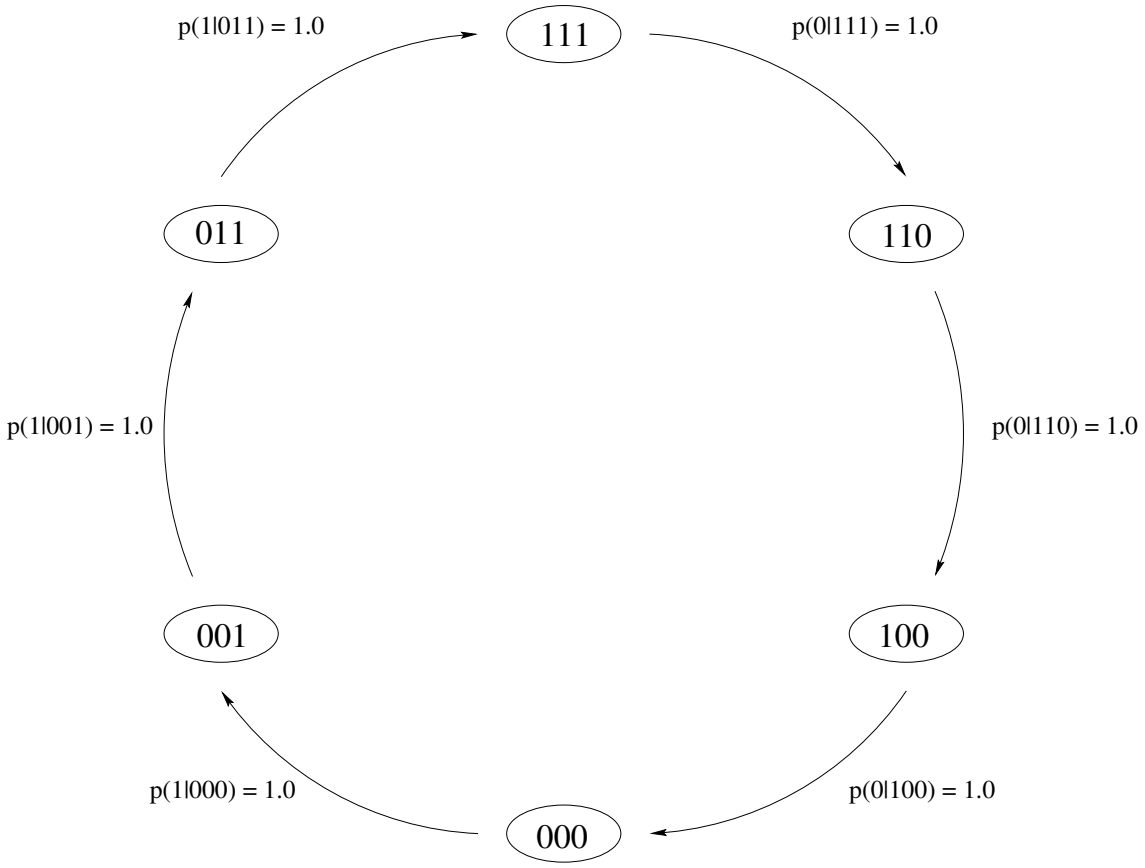

Figure 5.28: The recurrent portion of the ϵ-machine for the 6H process. This is a portion of an $r = 3$ de Bruijn graph, hence we label the causal states by the last three symbols seen.

Table 5.8: The first few Qs generated by the 6H process. The correlation functions are periodic in n with period six and have an infinite correlation length.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.500	0.500	0
2	0.333	0.333	0.333
3	0.333	0.333	0.333
4	0.333	0.333	0.333
5	0.500	0.500	0.000
6	0.000	0.000	1.000

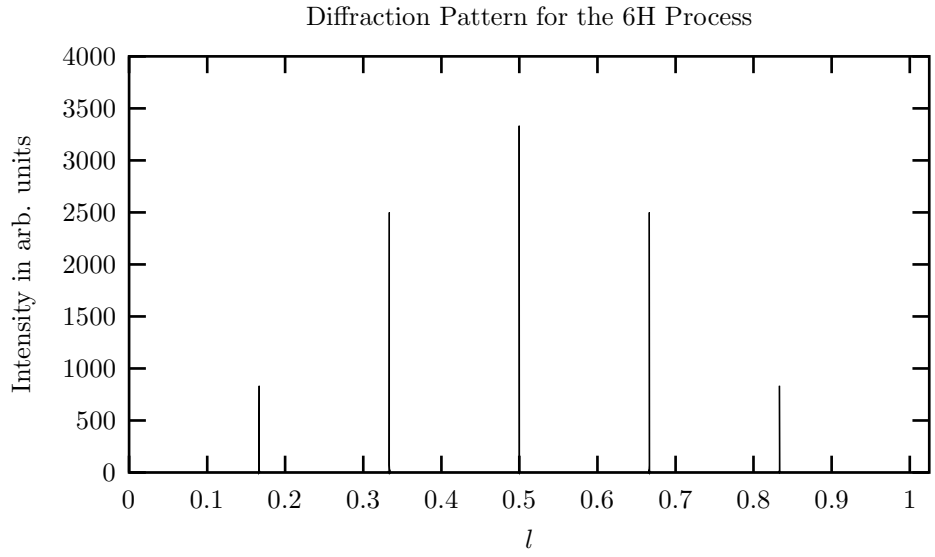

Figure 5.29: The diffraction pattern for a lattice stacked according to the 6H process. We see five Bragg peaks in the spectrum, at values of $l = \frac{m}{6}$ with $m = 1, 2, 3, 4, 5$. There is no Bragg peak at integer l.

shown in Figure 5.30. This machine is part of an $r = 4$ de Bruijn graph, the smallest graph that can accommodate both the 6H and the 3C cycles without having them sharing a causal state. The dashed lines connecting nodes indicate relatively infrequent transitions between cycles, while the solid lines indicate transitions between states within one of the three cycles. This graph has the minimal structure (fewest nodes and arcs) that can represent this architecture. The first few values of the correlation functions are tabulated in Table 5.9. A graph of the first few Qs generated by the 3C/2H/6H process is shown in Figure 5.9.

The diffraction pattern for the 3C/2H/6H process is displayed in Figure 5.32. We see a strong, sharp reflection at $l = \frac{1}{3}$ which we associate with the 3C structure. There are also several other broader, reflections of weaker intensity at $l = \frac{1}{6}, \frac{1}{2}, \frac{2}{3}$ and $\frac{5}{6}$. There seems to be a very slight increase in intensity at $l = 1$. Of most interest to us are the weak reflections at $l = \frac{1}{6}$ and $\frac{5}{6}$. These could only come from the 6H structure, and we see that the presence of both disorder and other stacking sequences does not seem to extinguish them. Again, it is difficult to draw general conclusions from

Table 5.9: The first few correlation functions generated by the 3C/2H/6H process. They decay asymptotically to $\frac{1}{3}$ and have a correlation length of $\lambda_q = 49 \pm 1$ calculated from the first one-hundred and fifty values of the correlation functions.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.604	0.396	0
2	0.224	0.433	0.344
3	0.308	0.271	0.420
4	0.408	0.273	0.319
5	0.383	0.519	0.099
6	0.066	0.140	0.793

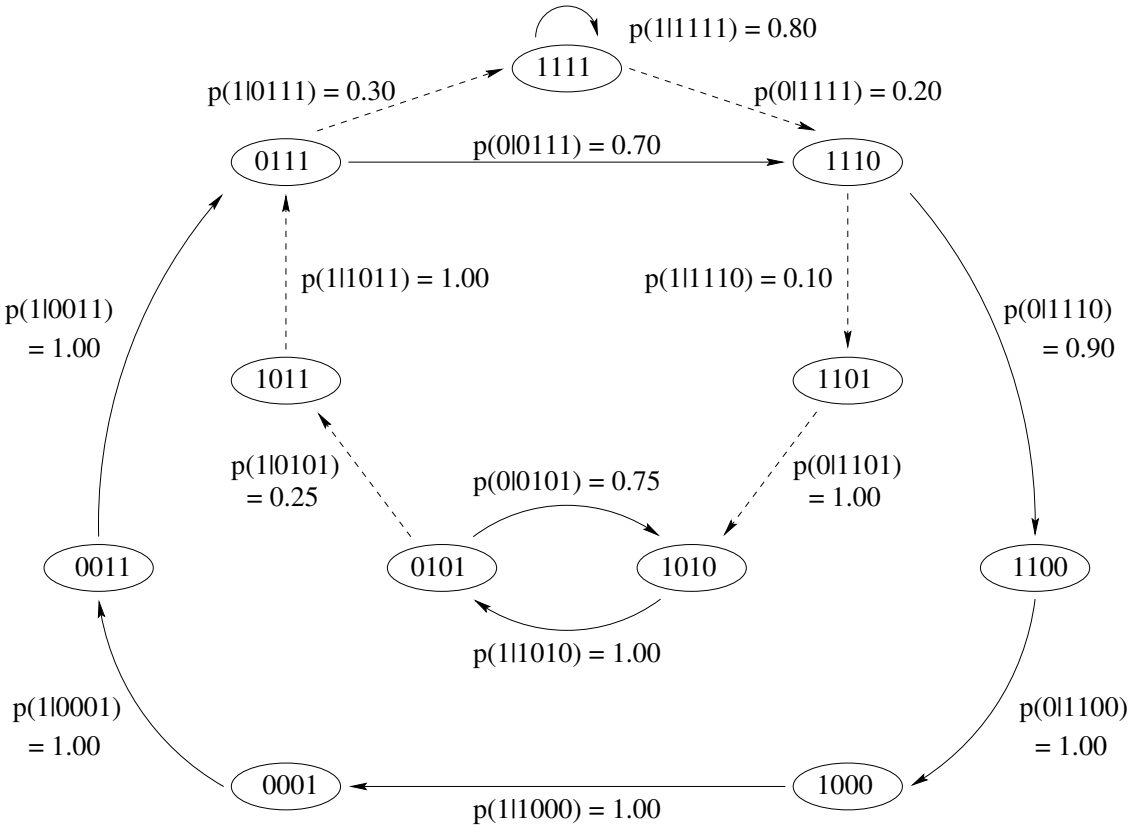

Figure 5.30: The recurrent portion of the ϵ-machine for the 3C/2H/6H process. In order to have a barrier between each simple cycle corresponding to some of the crystal structure present, we require that no two such simple cycles share a state. We need an $r = 4$ de Bruijn graph to accomplish this. The solid lines represent transitions between states associated with the simple cycles giving rise to some crystalline order, and the dotted lines indicate the relatively weak transitions between simple cycles.

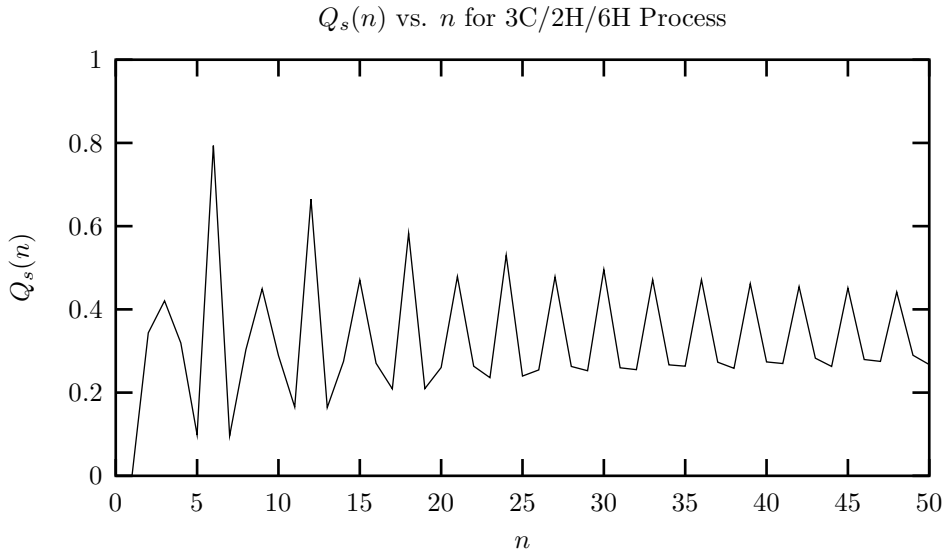

Figure 5.31: $Q_s(n)$ vs. n for 3C/2H/6H process. While they do decay asymptotically to $\frac{1}{3}$, they seem to have a quasiperiodic character of period six.

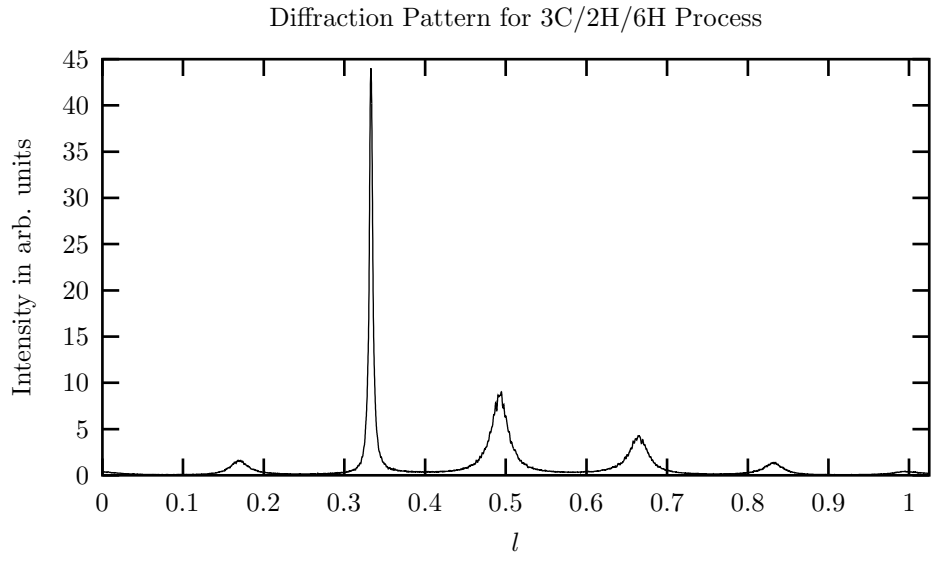

Figure 5.32: The diffraction pattern for a lattice stacked according to the 3C/2H/6H process. We see a strong reflection at $l = \frac{1}{3}$ which we associate primarily with the 3C structure, although the 6H undoubtedly contributes also. The small bumps in the spectrum at $l = \frac{1}{6}$ and $\frac{5}{6}$ are from the 6H structure present.

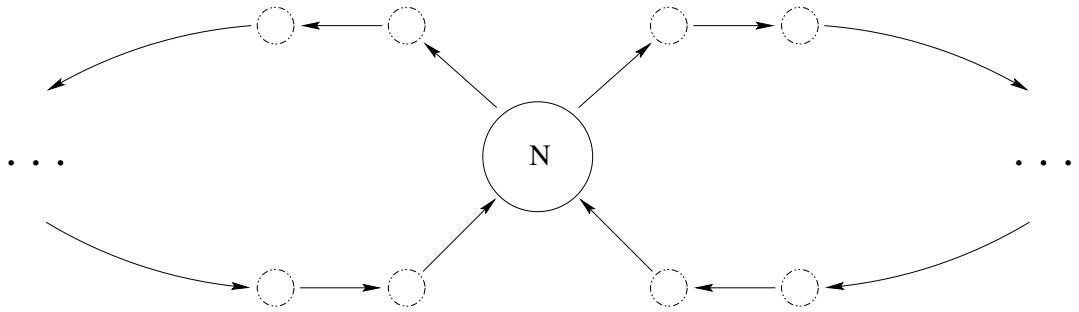

Figure 5.33: A pictorial representation of D-pair. Each cycle shares the node $\mathcal{N}$, which since this is a finite memory process with $r = 8$, can be labeled by the last eight symbols seen. So we label the node $\mathcal{N} = 11010100$. The two simple cycles which share the node $\mathcal{N}$ have their nodes represented by the smaller circles. Since each cycle is period fourteen, there should be thirteen smaller nodes in each cycle; but this is onerous to draw, so we have only explicitly shown four nodes in each cycle, with the other nine represented by the $\cdots$.

just one sample, but it is suggestive that enhanced intensity at these l might indicate some 6H structure present.

For this process, we find a correlation length of $\lambda_q = 49 \pm 1$. The entropy density is $h_\mu = 0.332$ bits/symbol and the statistical complexity is $C_\mu = 3.17$ bits. We find the excess entropy to be $\mathbf{E} = 1.84$ bits and the transient information to be $\mathbf{T} = 3.71$ bit symbols. This process implies a memory of range four layers.

5.10 A Period 14 D-pair

Canright and Watson [12], on the basis of elementary physical symmetries, proposed that certain simple cycles of a de Bruijn graph could be degenerate in energy and have a zero energy domain wall between them. They assumed a finite interaction between spins on a one-dimensional chain, and considered the case where each spin could assume only discrete values. They found it possible to find such pairs of symmetry related simple cycles that shared a node on a de Bruijn graph. The symmetry of the cycles insured that they had the same energy density (energy per spin) and the sharing of a node insured that there would be no energy cost in flipping from one cycle to the other. They called this pair a "D-pair" to indicate that a long string made from a series of these would be both disordered and degenerate. An illustration on this is shown in Figure 5.33.

The particular D-pair we will examine has a period of fourteen and is found on an $r = 8$ de Bruijn graph. The regular expression for this language is $\mathcal{R} = (11010100 + (001110 + 100011))^*$. The node that they share is $\mathcal{N} = 11010100$, which is invariant under simultaneous spacial and spin inversion. If the system is at the node $\mathcal{N}$ it has two options to proceed. Taking, say, the left path the series of spins would be 00111011010100 and taking the the right path would give a series 10001111010100. On either path, the cycle returns to the node $\mathcal{N}$ after visiting thirteen nodes in between. Each node in each cycle has a partner related by simultaneous spatial and spin inversion in the other cycle. Therefore, if the Hamiltonian that describes the system has these symmetries, then the two cycles will be degenerate and since they share a node, there is no energy cost to flip between them. If these simple cycles correspond to the ground state of the system, then we might see such disordered and degenerate states. The question then becomes, "What would the diffraction pattern for a lattice stacked according to this language look like?" This was the original motivation for treating this process. Yi and Canright [80] examined a number of these D-pairs, and we will repeat some of their analysis, but with an eye toward computational mechanics.

Table 5.10: The first few Qs generated by a period 14 D-pair. Except for the first few values of the correlation functions, these Qs are periodic with period fourteen.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.500	0.500	0
2	0.214	0.214	0.571
3	0.393	0.393	0.214
4	0.321	0.321	0.357
5	0.321	0.321	0.357
6	0.357	0.357	0.286

If we assume that at the node $\mathcal{N}$ the system has an equal probability of choosing either simple cycle, then we can find the correlation functions, the first few of which are given in Table 5.10. A graph of $Q_s(n)$ vs. n for the first fifty n is shown in Figure 5.34. We see that the correlation functions are periodic with period fourteen, which may seem odd at first. This process clearly has some randomness associated with it, and our past experience leads us to believe that this should result in correlation functions that decay to an asymptotic value of one-third. These do not. This is an example of long-range order without short-range order [44]. The process does have some indeterminism, but every fourteenth layer is required to be same since the process returns to the same state every fourteen layers and there is no net rotation of the crystal. Hence we see that there is long-range order which is reflected in the correlation functions, but a short-range there is a particular kind of constrained disorder. See Yi and Canright [80]. The diffraction pattern for this period 14 D-pair is shown in Figure 5.35. We see fourteen equally spaced Bragg peaks in the spectrum, although some have only a small intensity. Not obvious in the spectrum is the relatively weak diffuse scattering. Approximately 21% of the diffracted intensity is scattered into this background. The diffraction pattern, minus the fourteen Bragg peaks, is shown in Figure 5.36.

We find therefore, that the correlation length is infinite but the entropy density is $h_\mu = \frac{1}{14}$ bits/symbol. The statistical complexity is $C_\mu = 4.74$ bits and the excess entropy is $\mathbf{E} = 4.16$ bits. The transient information is $\mathbf{T} = 12.2$ bit symbols.

5.11 The Noisy Period Two Process

The noisy period two process is the first process we examine that does not have a finite memory. That it, the corresponding language is strictly sofic. We can write this language as $...1a1a1a1a1...$, where $a \in \{0, 1\}$. The regular expression for this language is $\mathcal{R} = (01 + 11)^*$. The recurrent portion of the ϵ-machine for this process is shown in figure 5.37.

The first few Qs for this process are shown in Table 5.11. Since $Q_c(n) \neq Q_r(n)$, we do not expect the diffraction pattern to be symmetric about $l = \frac{1}{2}$. A plot of $Q_s(n)$ versus n is given in Figure 5.38. The correlation functions are found to decay to asymptotic values of $\frac{1}{3}$, so we should see no Bragg peaks in the diffraction pattern, shown in Figure 5.40. Interestingly, the diffraction pattern is not too dissimilar to that of the fair coin toss. There is only one maximum, at $l = 0.413$, only a little displaced from the maximum in the fair coin toss at $l = \frac{1}{2}$. The intensity at this maximum is $I(0.413) = 5.678$, not so different from the maximum of the fair coin toss of $I_{max} = 3$. One might expect a more 'varied' pattern from a more sophisticated process as the noisy period two. Curiously, in comparison to the golden mean process which has a memory of one and a correlation length of $\lambda_q \approx 4.5$ we see that the correlation length of the noisy period two, $\lambda_q \approx 2$ is significantly small even though the noisy period two process has an infinite memory.

Another interesting feature of the diffraction pattern is the vanishing intensity at $l = \frac{5}{6}$. It is not

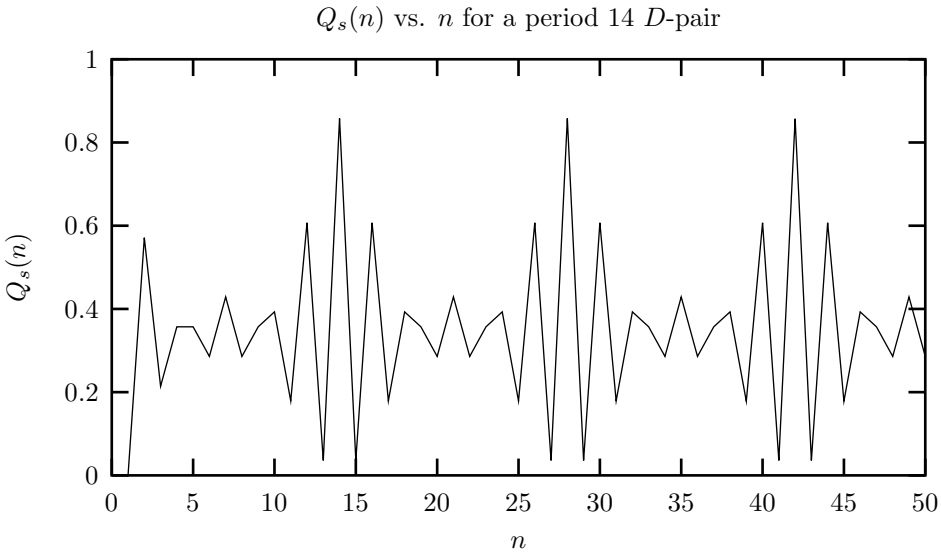

Figure 5.34: $Q_s(n)$ for period 14 D-pair. Except for the first few n, we see periodic correlation functions with period fourteen. This seems odd considering that this process contains some randomness. See text for details.

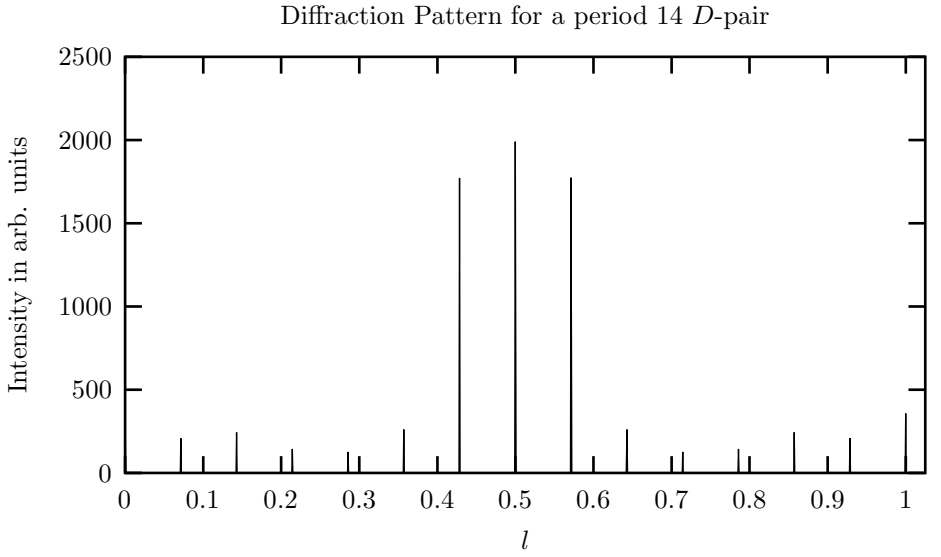

Figure 5.35: The diffraction pattern for a lattice stacked according to the period 14 D-pair. We see a total of fourteen equally spaced Bragg Peaks in the spectrum. Not clear in this figure is the diffuse background scattering.

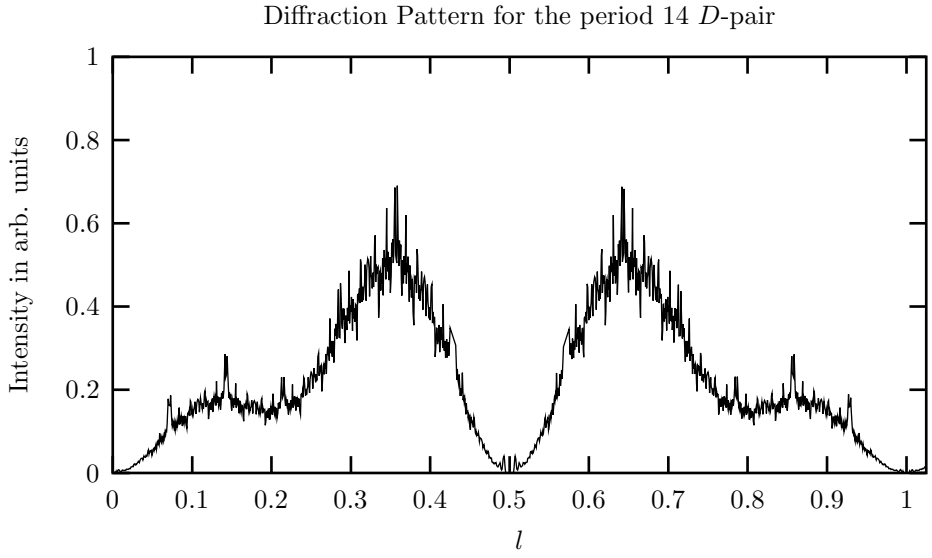

Figure 5.36: The diffraction pattern for a lattice stacked according to the period 14 D-pair with Bragg peaks removed. Approximately 21% of the diffracted intensity goes into this background scattering.

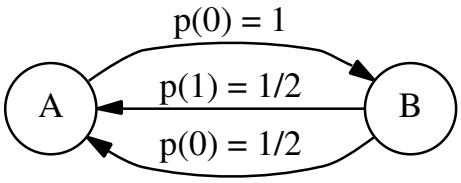

Figure 5.37: The recurrent portion of the ϵ-machine for the noisy period two process. This process is not describable in terms of a finite memory machine, so we have labeled the two causal states by A and B. We see that the machine oscillates between the two causal states. Every other symbol is destine to be a 1 but the symbols between these 1s can be either 0 or 1 with equal probability.

Table 5.11: The first few values of the correlation functions generated by the noisy period two process.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.750	0.250	0
2	0	0.500	0.500
3	0.500	0.125	0.375
4	0.250	0.500	0.250
5	0.312	0.250	0.438
6	0.375	0.375	0.250

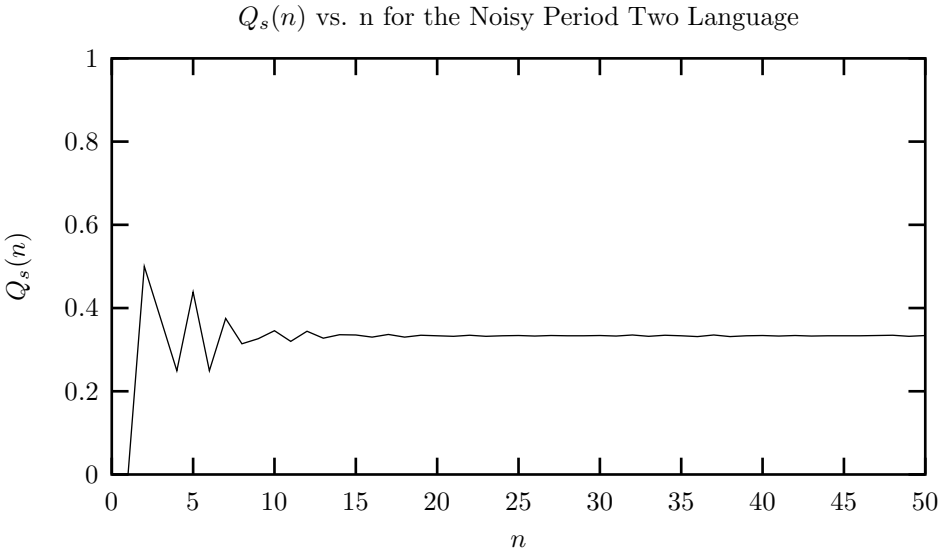

Figure 5.38: $Q_s(n)$ vs. n for the noisy period two process. We see that the correlation functions rather quickly decay to their asymptotic value of $\frac{1}{3}$. We find a correlation length of $\lambda = 2.04 \pm 0.03$.

clear why this is so.

The entropy production of this process is $h_\mu = \frac{1}{2}$ bits/symbol, and it has both a statistical complexity and excess entropy of $C_\mu = \mathbf{E} = 1$ bit. The transient information is $\mathbf{T} = 3.3$ bit symbols. We find an entirely diffuse diffraction pattern and the two-layer correlation functions all decay asymptotically to $\frac{1}{3}$, as they must. A graph of $\log_2 \Psi_q(n)$ vs. n is shown in Figure 5.39. We find a correlation length of $\lambda_q = 2.04 \pm 0.03$.

5.12 The Even System

Like the noisy period two process, the even system is a process that can not be represented as a portion of a de Bruijn graph. The regular expression for the corresponding language is $\mathcal{R} = (0{+}11)^*$. We can think of this language as the set of words such that there are always an even number of 1s sandwiched between any two 0s. We can also give the forbidden words, which are $\mathcal{F}_\mathcal{L} = \{01^{2k+1}0\}$ where k is a non-negative integer. In this sense, just as for the noisy period two, the language has an infinite memory, since for any string of 1s, however long, the language must remember whether there have been an even or odd number of 1s since the last 0. Another way of stating this is to say that there is no longest irreducible forbidden word. The recurrent part of the ϵ-machine for this process is given in Figure 5.41.

The first few Qs for this process are tabulated in Table 5.12 and a plot of $Q_s(n)$ vs. n for the first fifty n is given in Figure 5.42. We see that $Q_s(n)$ quickly approaches its asymptotic value of $\frac{1}{3}$. Indeed, $Q_s(n)$ shows little structure. A plot of $\log_2 \Psi_q(n)$ vs. n is given in Figure 5.43. We see that the approximation of exponential decay is not very good here. It is not at all clear why this is so. Using the first nine values of $\log_2 \Psi_q(n)$ versus n we can calculate the correlation length and we find it to be $\lambda_q = 1.7 \pm 0.2$. The diffraction pattern for this process is shown in Figure 5.44. We see completely diffuse pattern with two maxima, one at $l = 0.389$ and a second smaller maximum at $l = 0.772$. We also see that the diffraction pattern has a zero at $l = \frac{5}{6}$. This is curiously at the same l value as we saw a zero in the noisy period two.

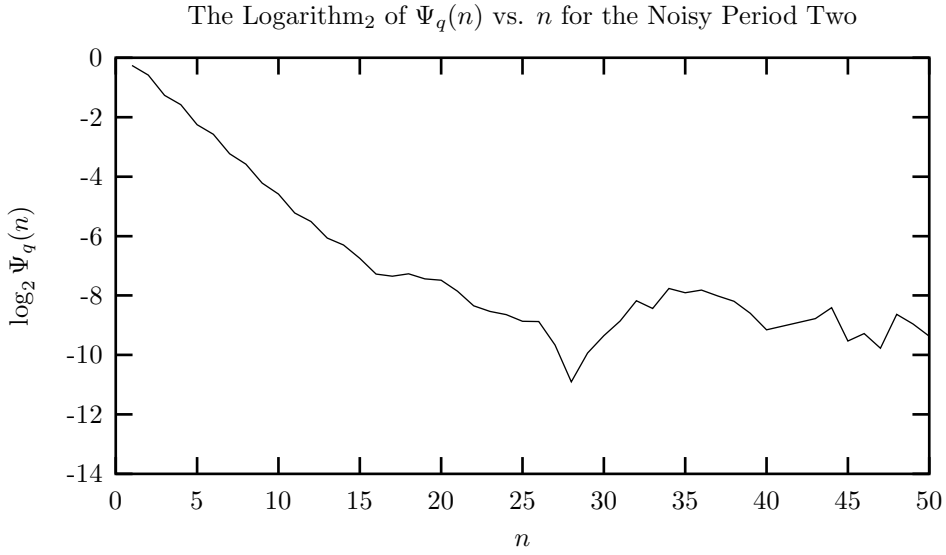

Figure 5.39: The logarithm of $\Psi_q(n)$ for the noisy period two process as a function of n. Considering the first twelve values of $\Psi_q(n)$, we find $\lambda_q = 2.04 \pm 0.03$.

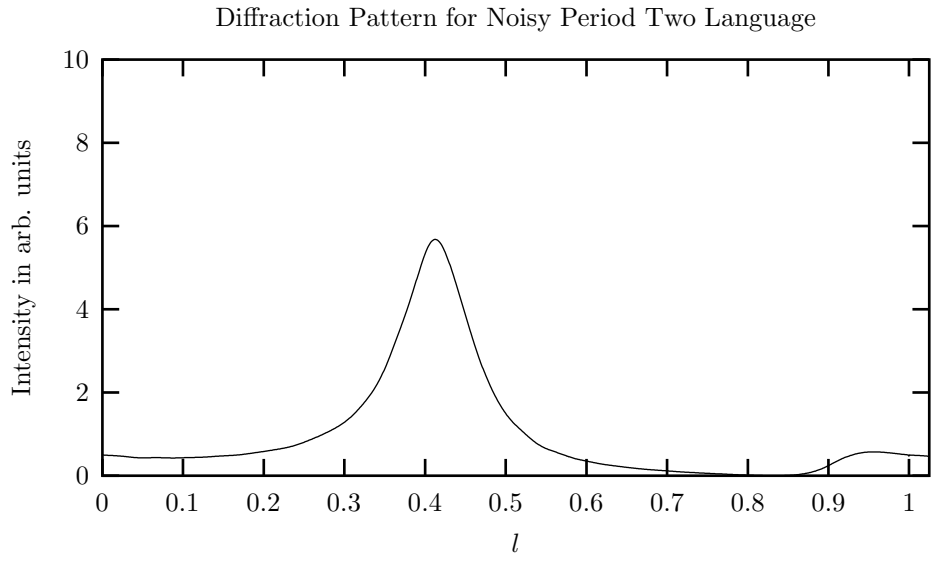

Figure 5.40: The diffraction pattern for a lattice stacked according to the noisy period two process. We see that the spectrum is entirely diffuse and relatively featureless, with a curious isolated zero at $l = \frac{5}{6}$.

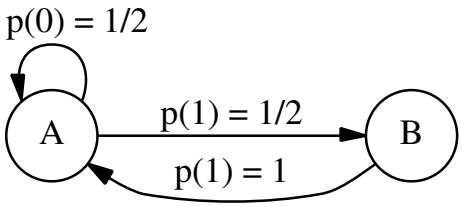

Figure 5.41: The recurrent portion of the ϵ-machine that generates the even process. Like the noisy period two process, this process can not be represented by a finite memory machine. We have labeled the two causal states by A and B. Superficially the machine resembles that of the golden mean process. We see that if state A emits a 1 the machine advances state B where it must always emit another 1. Thus this process can never generate a sequence that has an odd number of 1s sandwiched between two 0s.

Table 5.12: The first few values of the correlation functions generated by the even system.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.667	0.333	0
2	0.167	0.500	0.333
3	0.416	0.167	0.416
4	0.333	0.375	0.292
5	0.271	0.354	0.374
6	0.364	0.334	0.303

Turning our attention to computational measures, we see that the entropy rate is $h_\mu = \frac{2}{3}$ bits/symbol and the predictability $\mathbf{G} = \frac{1}{3}$ bits/symbol. The statistical complexity $C_\mu = 0.918$ bits and the excess entropy $\mathbf{E} = 0.913$ bits. The transient information is $\mathbf{T} = 3.09$ bit symbols.

5.13 The Sum Zero Process

We now consider a process that is similar in spirit to the period 14 D-pair in spirit, but is strictly sofic. That is, we wish to design a process that has both long range correlations as reflected in the Qs but also has some randomness such that $h_\mu \neq 0$. We need long range correlations to see Bragg scattering. So let us imagine, however artificial, a process that does that. Consider the stacking rule such that every other layer has the same orientation but the layers sandwiched between can have a randomly chosen orientation, subject of course to the stacking constraints. The sequence would then look like $...AxAxAxAxA...$, where $x \in \{B, C\}$. Clearly this will have long range correlations. In terms of a language, we can think of the sequence as being divided into doublets, such that $..aaaaaaaa..$ becomes $..(aa)(aa)(aa)(aa)..$ with $a \in \{0, 1\}$. The rule then is that each doublet must contain exactly one 0 and one 1, but the order is arbitrary. If we remind ourselves of the physical meaning of each symbol, that is a 1 gives a relative rotation of sixty degrees between adjacent layers about the stacking direction while a 0 gives a rotation in just the opposite sense, then in terms of operators they are inverses. Hence one followed by the other produces no net rotation and is therefore a 'sum zero' operation. Hence the name sum zero process. We can easily translate this into an ϵ-machine, the recurrent portion of which is shown in Figure 5.45. The regular expression for this language is $\mathcal{R} = (10 + 01)^*$.

The two-layer correlation functions are easily found to be given by

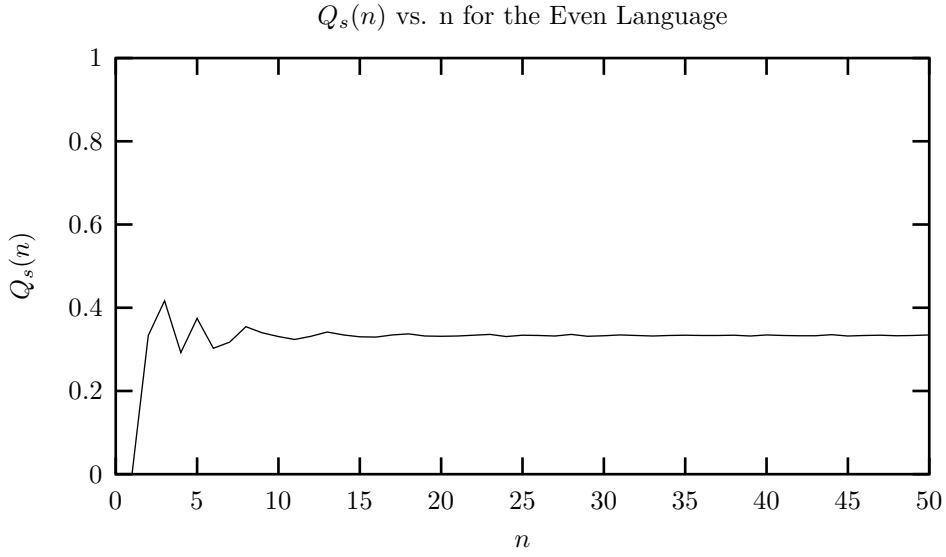

Figure 5.42: $Q_s(n)$ vs. n for the even process. As with the noisy period two, we see that the correlation functions are rather structureless and decay quickly to their asymptotic value of $\frac{1}{3}$.

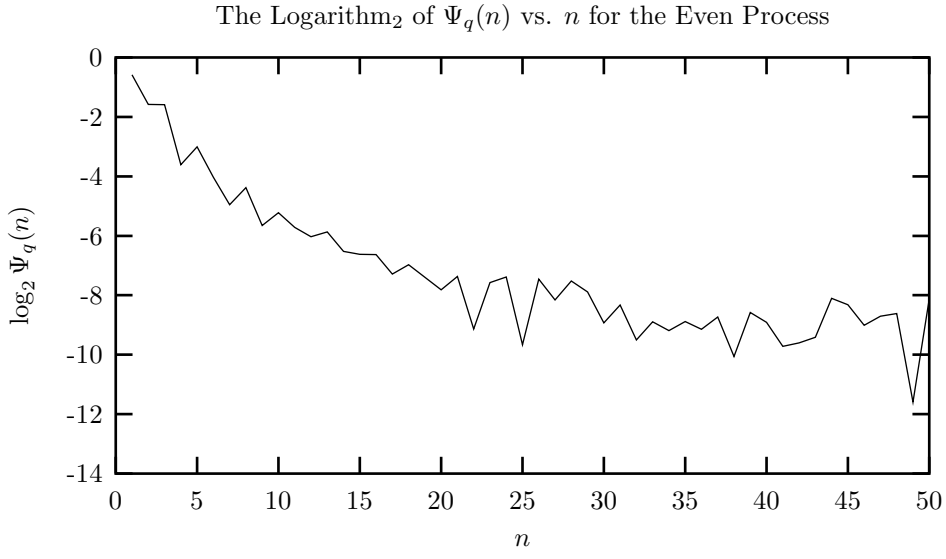

Figure 5.43: The logarithm of $\Psi_q(n)$ for the even process as a function of n. We find a correlation length of 1.7 ± 0.2. It does not seem that the approximation of exponential decay is very good here. Why?

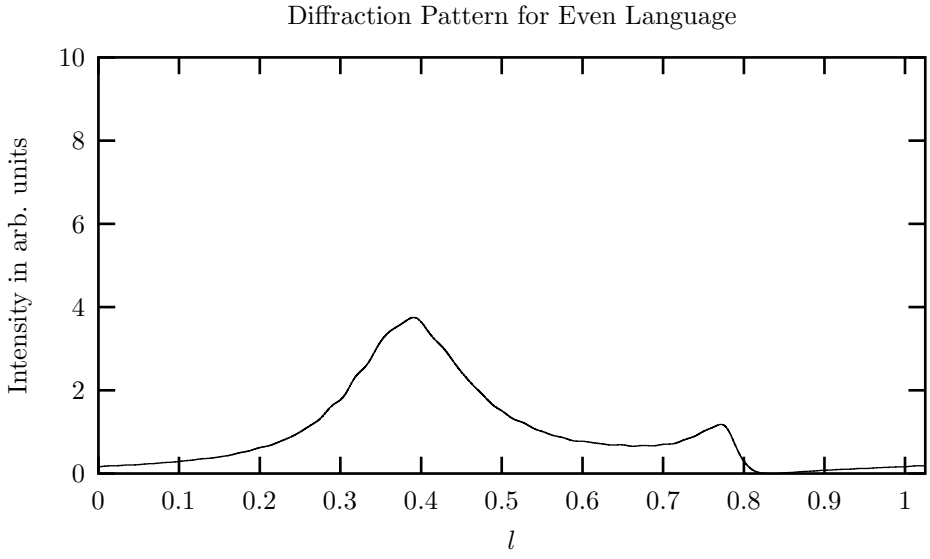

Figure 5.44: The diffraction pattern for a lattice stacked according to the even process. The spectrum is rather featureless, as with the noisy period two, and also has an isolated zero at $l = \frac{5}{6}$.

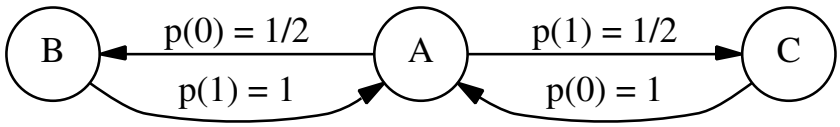

Figure 5.45: The recurrent portion of the ϵ-machine that generates the sum zero process. Since this is a strictly sofic process, it can not be described by a finite memory process. We label the causal states by the letters A, B, and C.

Table 5.13: The first few Qs generated by the sum zero process. They are periodic in n with period two. Note the similarity to the period two process.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.500	0.500	0
2	0.125	0.125	0.750
3	0.500	0.500	0
4	0.125	0.125	0.750
5	0.500	0.500	0
6	0.125	0.125	0.750

$$\left.\begin{array}{ll} Q_c(n) = & Q_r(n) = \frac{1}{2} \\ Q_s(n) = & 0 \end{array}\right\} \text{ for } n \text{ odd,}$$

$$\left.\begin{array}{ll} Q_c(n) = & Q_r(n) = \frac{1}{8} \\ Q_s(n) = & \frac{3}{4} \end{array}\right\} \text{ for } n \text{ even.} \tag{5.6}$$

The first few are tabulated in Table 5.13. Comparing these to the period two process we see a remarkable similarity. Even though the underlying process has some randomness, the Qs do not decay asymptotically at all. Instead they are quite periodic. The diffraction pattern as shown in Figure 5.46 is quite like that of the period two process. There are two Bragg reflections in the spectrum, one at $l = 0$ and the other at $l = 1/2$, but their relative intensities differ from the period two diffraction pattern. $I(0) = 625$ and $I(\frac{1}{2}) = 5625$, giving a ratio of $I(\frac{1}{2})/I(0) = 9.00$. We also find a constant background scattering of $I(\text{background}) = 0.375$. It turns out that nearly 37.5% of the scattered energy is diffracted into this constant background.

So we find that the entropy rate is given by $h_\mu = \frac{1}{2}$ bits/symbol and the total predictability is $\mathbf{G} = \frac{1}{2}$ bits/symbol. Grassberger [33] has considered this language and found that and $C_\mu = \mathbf{E} = \frac{3}{2}$ bits. We further find that the transient information is $\mathbf{T} = 4.7$ bit symbols. The Qs display periodic oscillations of period $T_q = 2$. The scattering is a mixture of Bragg peaks and a diffuse background.

5.14 The Thue-Morse Sequence

The Thue-Morse sequence is the only example which we consider that can not be expressed as a finite state process. The Thue-Morse sequence is an example of a process generated by substitution. The rule for production of a string is

$$\begin{aligned} \sigma(0) &\rightarrow 01, \\ \sigma(1) &\rightarrow 10. \end{aligned}$$

After four iterations, for example, beginning with the symbol 0 one gets the string

$$\sigma^4(0) = 0110100110010110.$$

One can define the Thue-Morse language as [17]

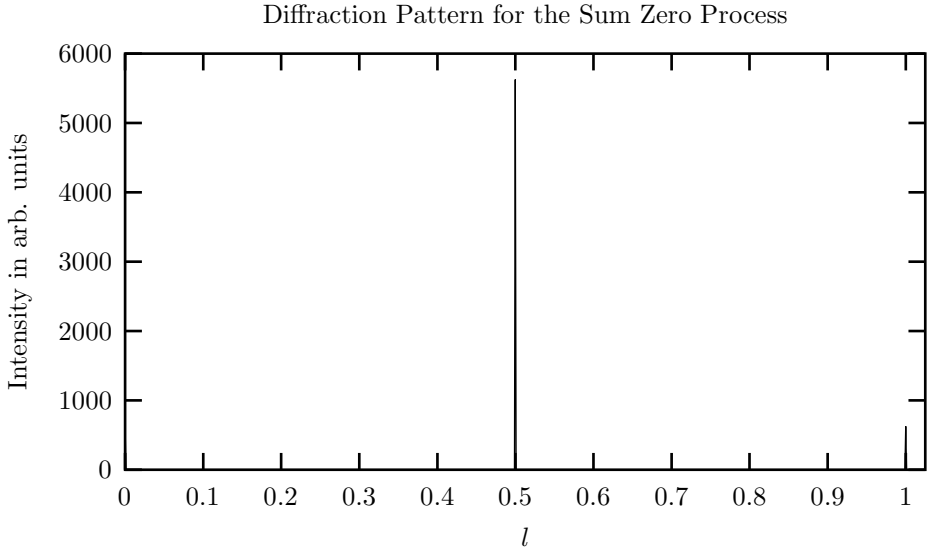

Figure 5.46: The diffraction pattern for a lattice stacked according to the sum zero process. Having two Bragg peaks in the spectrum, one at $l = \frac{1}{2}$ and the other at $l = 1$, this diffraction pattern resembles that of the period two process. The most easily recognizable difference is the the ratio of the intensities is not the same. For the period two we have $I(l = \frac{1}{2})/I(l = 1) = 3.00$ and for the sum zero process we have $I(l = \frac{1}{2})/I(l = 1) = 9.00$. Not seen in this plot is the constant background scattering.

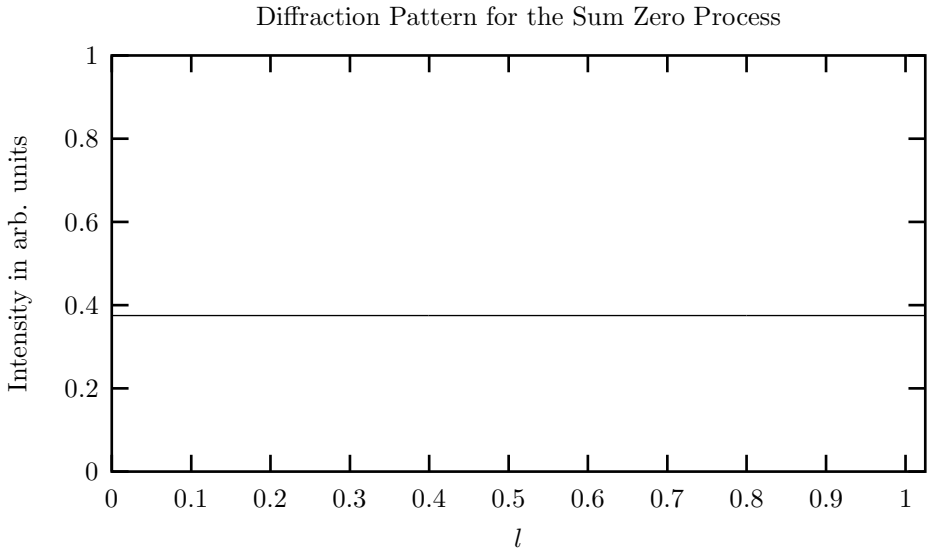

Figure 5.47: The background diffraction pattern for a lattice stacked according to the sum zero process. We see a constant background intensity. Approximately 37.5% of the diffracted intensity falls into this diffuse scattering.

Table 5.14: The first few values of the correlation functions for the Thue-Morse process.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.500	0.500	0
2	0.167	0.167	0.667
3	0.500	0.500	0
4	0.167	0.167	0.667
5	0.500	0.500	0
6	0.083	0.083	0.833

$$\mathcal{L}_{TM} = \mathrm{sub}\Big(\lim_{t\to\infty} \sigma^t(1)\Big),$$

where 'sub(s)' means that we take all sub-words of the string s. Another example of a language generated by substitution is the Fibonacci sequence. These languages are the called substitution languages and are formally known as D0L languages. (The "D" stands for deterministic, the "L" acknowledges A. Lindenmeyer who studied them, and the "0" indicates that the substitutions occur independent of the context of the symbol being substituted.) D0L languages are a subset of the so-called index languages which sit between the context-free and context-sensitive languages on the Chomsky Hierarchy [6]. We can further define the the Thue-Morse process [17] by assigning the natural measure (or the frequency of occurrence in $\sigma^\infty(1)$) to the words in $\mathcal{L}_{TM}$. The Thue-Morse sequence is completely predictable, but aperiodic. In order to make a prediction, one, in general, needs to know an infinite history. In terms of computational mechanics, then, we say that $h_\mu = 0$ and C_μ diverges as the $\log_2 L$, where L is the length of the sequence [17].

In contrast to the other processes we have considered, we use a much shorter string from the Thue-Morse sequence to find the diffraction pattern. Since it is known that the power spectrum for the limit words of the Thue-Morse sequence is singular continuous, we expect that the diffraction pattern will show essential properties of a singular continuous measure. Indeed, super-lattice heterostructures of GaAs-AlAs have been fabricated in the laboratory and x-ray diffraction studies have been performed [3] [52]. There singular continuous properties were found. Since the spectrum is continuous but not differentiable everywhere, it is more instructive to examine smaller samples. We use a chain of length 1024 symbols. We generate a sequence of approximately one-half million symbols and select our 1024 symbol chain from a randomly chosen consecutive sequence in the middle.

The first few values of the correlation functions are shown in Table 5.14. $Q_s(n)$ vs. n for the first fifty n is plotted in Figure 5.48. We see that the correlation functions appear to show some periodicity in the odd n, namely that $Q_c(n) = Q_r(n) = \frac{1}{2}$, $Q_s(n) = 0$ for odd n. This is almost always true, but an examination of the odd n values shows that occasionally the $Q_c(n)$, $Q_r(n)$ differ from $\frac{1}{2}$ by a small amount. This behavior of the odd n correlation functions appears to be independent of the size of the string we examine. The correlation functions at even n however, do not seem to decay to $\frac{1}{3}$ or become periodic. Instead they fidget about, with $Q_s(n) \approx \frac{3}{4}$ and $Q_c(n) \approx Q_r(n) \approx \frac{1}{8}$.

The diffraction pattern for the Thue-Morse sequence is shown in Figure 5.49. We notice that there are two strong reflections, one at $l = \frac{1}{2}$ and the other at $l = 1$. We should not be surprised at this given the similarity of the correlation function to the period two process. (Recall that the period two process has two Bragg peaks, one at $l = \frac{1}{2}$ and the other at $l = 1$.) Approximately 58% of the scattered intensity goes into the peak at $l = \frac{1}{2}$ and 6% into the peak at $l = 0$. The remaining 36% is diffracted into the background. These peaks seem to scale as N^2, suggesting that they are

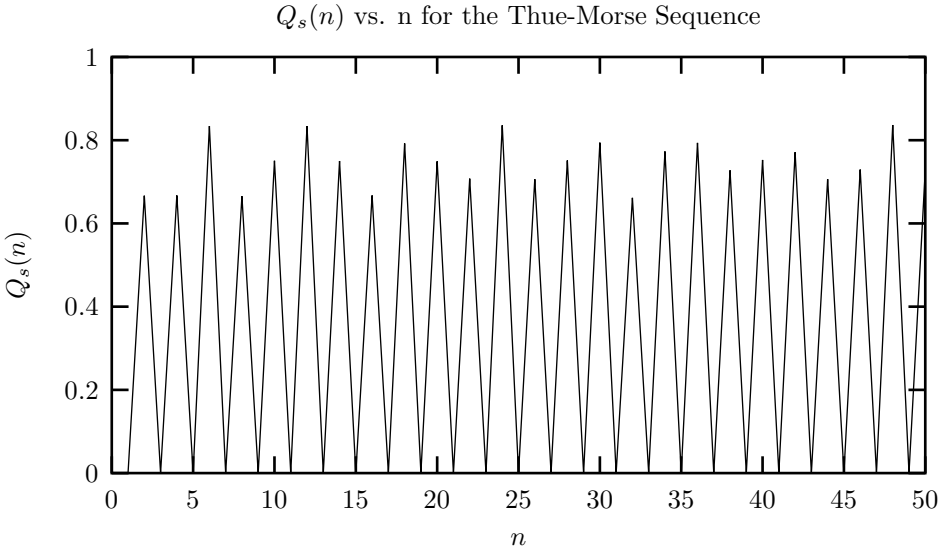

Figure 5.48: $Q_s(n)$ vs. n for the Thue-Morse sequence. For odd n, $Q_s(n)$ is zero, but for the even values of n, $Q_s(n)$ do not approach an asymptotic values, but seem to oscillate in what appears to be a random fashion. Of course the Thue-Morse sequence is completely predictable, so the oscillations are not random.

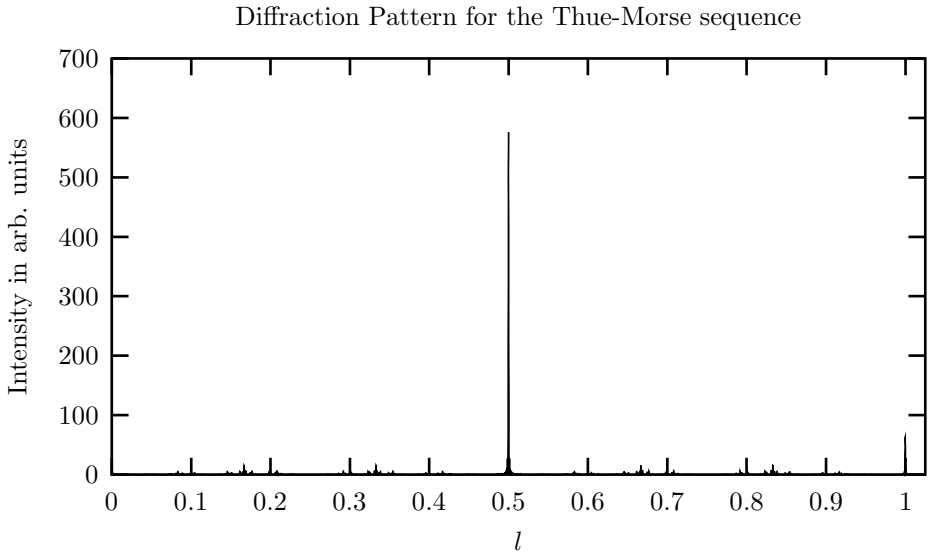

Figure 5.49: The diffraction pattern for a lattice stacked according to the Thue-Morse sequence. We use 1024 layers for this pattern and calculate the intensity at 10,000 equally spaced points. We see two Bragg peaks, one at $l = \frac{1}{2}$ and the other at $l = 1$. This is similar to the period two process and the sum zero process. We find the ratio of the intensities of the two Bragg peaks to be $I(l = \frac{1}{2})/I(l = 1) = 8.95$.

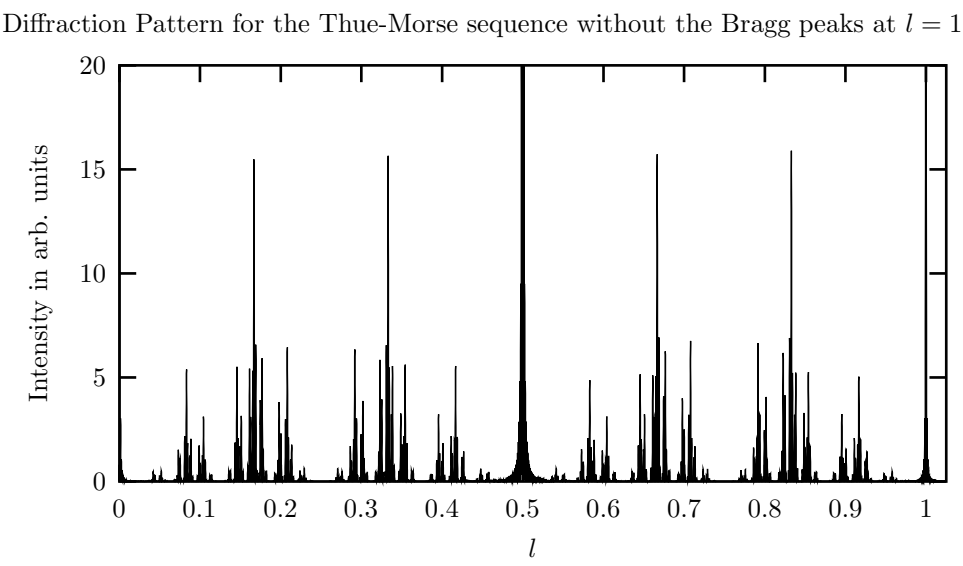

Figure 5.50: The diffraction pattern for a lattice stacked according to the Thue-Morse sequence scaled to see the background scattering. This background is singular continuous, and approximately 36% of scattered intensity falls into it.

Bragg peaks. Experiments on heterostructures have not observed this behavior, but we must keep in mind that there is a fundamental difference in how we stack modular layers as compared to how they stacked in a heterostructure. For hexagonal close-packed layers, there are *three* kinds of layers, if we treat orientation as the defining characteristic. The Thue-Morse sequence specifies the stacking of the relative spins. In contrast, for heterostructures, there are only *two* kinds of layers, and the Thue-Morse sequence specifies the stacking of these layers. The background intensity is shown in Figure 5.50. We can see that there is structure in the background not apparent in the Figure 5.49. The power spectrum of the Thue-Morse sequence is known to be fractal [6], so we should not be surprised at the self-similar features present. We finally show the logarithm of the diffracted intensity in Figure 5.51.

Let us mention a few of the measures of computation and correlation here. As stated before, the Thue-Morse sequence has an entropy density of $h_\mu = 0$, and a statistical complexity of $C_\mu \propto \log L$. The excess entropy $\mathbf{E}$ is also proportional to the logarithm of the length of the sequence L [17]. The transient information $\mathbf{T}$ is proportional to the length of the sequence. The correlation functions are aperiodic so the correlation length λ_q is infinite.

5.15 Summary

Let us pause and consider what we have discovered. Although implicit in our discussions, we should perhaps state explicitly that correlation length and range of interaction measure very different aspects of a system. The memory or the range of the interaction gives the minimum distance, in terms of previous layers or spins seen, that is necessary for an optimal prediction of the next spin. The correlation length can be thought of as an average decay rate. It gives the distance in terms of layers over which correlation information is halved. By correlation, of course, we mean what the average knowledge of a particular spins tells us about a spin n layers away, regardless of the values of the intervening spins. This point is made clear when we recall a few processes. For a completely

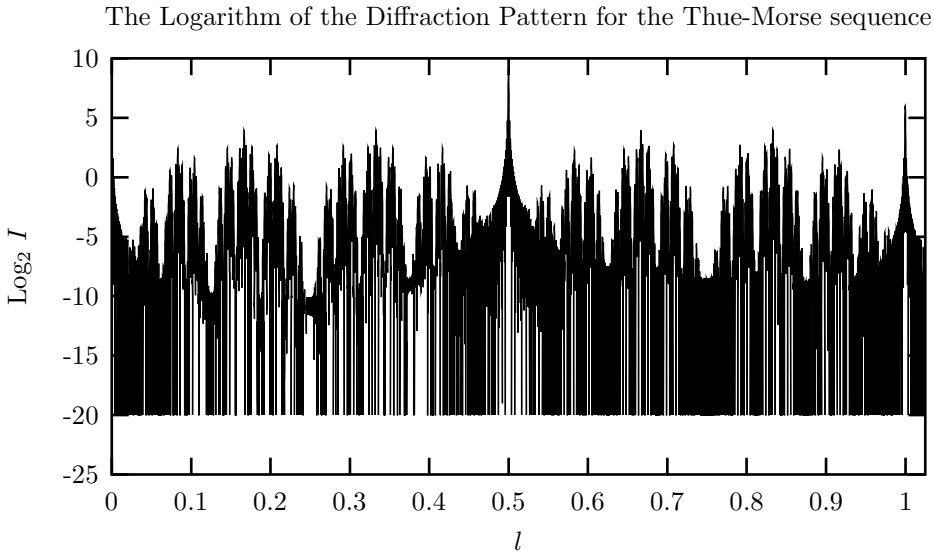

Figure 5.51: The logarithm of the diffraction pattern for a lattice stacked according to the Thue-Morse sequence. We find points in the spectrum where the scattered intensity is zero, and we have cut off the lower portion of the graph at -20.

predictable process, such as the period one, no memory of the previous spin is necessary to make an optimal prediction of the next spin. The period one does, however, have an infinite correlation length. Knowing one spin and the stacking rule provides information about spins infinitely far away. The Thue-Morse sequence is an example of a system with both an infinite correlation length and memory. The noisy period two has an infinite memory yet correlation information about the absolute positions of the stacking layers decays.

Another interesting point to consider is that diffuse scattering does not preclude the underlying process from being SS. Indeed, we found that two of three SS systems we examined had relatively featureless spectra. There is the interesting phenomenon of the isolated, vanishing intensity at $l = \frac{5}{6}$ for these two spectra, but it is not known how general this phenomenon is.

We also see that Bragg peaks do not preclude some randomness in a process. We have two examples of this, in the period 14 D-pair and the sum zero process. Also, we see that a completely predictable process, such as the Thue-Morse, can have at least part of the spectrum continuous.

The measures of computation for all of the system considered in this chapter are given in Table 5.15 and measures for correlative and diffraction effects are likewise tabulated in Table 5.16.

Table 5.15: Computational Results. In the first column we give the process we are considering. In the second column, we specify the language type, subshift of a finite type (SFT), strictly sofic (SS), or D0L, see §5.14. The third column gives the memory, or minimum effective range of interaction for the process. The last five columns list values for the measures of computation as given in §2.2.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
Fair Coin Toss or RNG	SFT	0	1	0	0	0	0
Biased RNG, $q = 0.01$	SFT	0	0.081	0.919	0	0	0
Biased RNG, $q = 0.02$	SFT	0	0.142	0.858	0	0	0
Biased RNG, $q = 0.05$	SFT	0	0.286	0.714	0	0	0
Biased RNG, $q = 0.10$	SFT	0	0.468	0.532	0	0	0
Biased RNG, $q = 0.20$	SFT	0	0.721	0.279	0	0	0
Biased RNG, $q = 0.30$	SFT	0	0.881	0.119	0	0	0
3C	SFT	0	0	1	0	0	0
2H	SFT	1	0	1	1	1	1
Golden Mean	SFT	1	2/3	1/3	0.918	0.252	0.252
3C/2H, $q = 0.01$	SFT	2	0.054	0.95	1.58	1.48	2.10
3C/2H, $q = 0.05$	SFT	2	0.190	0.81	1.58	1.21	1.68
3C/2H, $q = 0.10$	SFT	2	0.313	0.687	1.58	0.96	1.32
3C/2H, $q = 0.20$	SFT	2	0.483	0.571	1.58	0.62	0.83
3C/2H, $q = 0.30$	SFT	2	0.587	0.413	1.58	0.41	0.49
3C/2H, $q = 0.40$	SFT	2	0.648	0.352	1.58	0.29	0.31
4H	SFT	2	0	1	2	2	3
3C/2H/4H	SFT	3	0.424	0.576	2.79	1.51	2.87
6H	SFT	3	0	1	2.59	2.59	4.84
3C/2H/6H	SFT	4	0.340	0.660	3.19	1.83	3.72
Period 14 D-pair	SFT	8	1/14	13/14	4.74	4.16	12.2
Noisy Period Two	SS	∞	1/2	1/2	1	1.00	3.31
Even	SS	∞	2/3	1/3	0.918	0.913	3.09
Sum Zero	SS	∞	1/2	1/2	3/2	3/2	4.80
Thue-Morse	D0L	∞	0	1	$\propto \log N$	$\propto \log N$	$\propto N$

Table 5.16: Correlative and diffractive results. In the second column, we give the asymptotic behavior of the correlations, see §4.3. In the third column, we give the correlation length, λ_q, for each process, see §4.4. The fourth column shows the period of the limit cycle for periodic processes, see §4.4. The last column shows the scattering type for the diffraction pattern. We see that there are three possibilities: absolute continuous (AC), singular continuous (SC), and pure point (PP) or for physicists, δ-function or Bragg scattering.

System	Q Behavior	λ_q	T_q	Scattering Type
Fair Coin Toss	*decays*	1	-	AC
Biased RNG, $q = 0.01$	*decays*	50.4	-	AC
Biased RNG, $q = 0.02$	*decays*	23.7	-	AC
Biased RNG, $q = 0.05$	*decays*	9.57	-	AC
Biased RNG, $q = 0.10$	*decays*	4.49	-	AC
Biased RNG, $q = 0.20$	*decays*	2.14	-	AC
Biased RNG, $q = 0.30$	*decays*	1.43	-	AC
3C	*periodic*	∞	3	PP
2H	*periodic*	∞	2	PP
Golden Mean	*decays*	4.48	-	AC
3C/2H, $q = 0.01$	*decays*	123	-	AC
3C/2H, $q = 0.05$	*decays*	25.7	-	AC
3C/2H, $q = 0.10$	*decays*	13.1	-	AC
3C/2H, $q = 0.20$	*decays*	5.5	-	AC
3C/2H, $q = 0.30$	*decays*	3.5	-	AC
3C/2H, $q = 0.40$	*decays*	3.0	-	AC
4H	*periodic*	∞	4	PP
3C/2H/4H	*decays*	19.2	-	AC
6H	*periodic*	∞	6	PP
3C/2H/6H	*decays*	49	-	AC
Period 14 D-pair	*periodic*	∞	14	PP/AC
Noisy Period Two	*decays*	2.04	-	AC
Even	*decays*	1.70 (?)	-	AC
Sum Zero	*periodic*	∞	2	PP/AC
Thue-Morse	*aperiodic*	∞	-	PP/SC

Chapter 6

The Finite r Approximation to ϵ-Machine Reconstruction from Two-Layer Correlation Functions

Now that we have extracted the correlation functions or $\{Q_\alpha(n)\}$ from the diffraction data, we can proceed to reconstruct the ϵ-machine. We note that there has been work done in the area of relating correlations among symbols in a data stream to word probabilities [2], but our case is somewhat different. Our correlation information is with respect to the absolute spins of the stacking sequence and we want the word probabilities of the relative spin sequence. Our task, then, is to relate $\{Q_\alpha(n)\}$ to word probabilities $\{p(\omega)\}$, where $\omega \in \mathcal{A}^L$, and $\mathcal{A}^L$ is the set of all words of length L over the alphabet $\mathcal{A}$, with $\mathcal{A} \in \{0, 1\}$. We do this in successive approximations, by considering a machine of finite memory r and then writing down equations which relate the $\{Q_\alpha(n)\}$ to the $\{p(\omega)\}$. This is equivalent to approximating the process by an r^{th}-order Markov process, which we can graphically represent by an r^{th}-order de Bruijn graph. It is known that such a graph has 2^r nodes and 2^{r+1} arcs connecting nodes. Since each node has a memory of r, transitions between nodes are labeled by symbol sequences of length $r + 1$, or words of length L. An r^{th}-order Markov process can be completely specified by assigning a probability to each arc. We note, however, only 2^r of these probabilities are independent. The other 2^r probabilities are then constrained by conservation of probability.

We find the r^{th}-order approximation by writing down the de Bruijn graph that corresponds to the r^{th}-order process. We require conservation of probability at each node, which gives 2^r equations, of which $2^r - 1$ are independent. We can express this conservation principle mathematically as

$$p(0u) + p(1u) = p(u0) + p(u1) \qquad \forall \quad u, \tag{6.1}$$

where u is a sequence of symbols corresponding to a particular node. We additionally require that the total probability to see a word of length L be unity, ie,

$$\sum_{\omega \in \mathcal{A}^L} p(\omega) = 1. \tag{6.2}$$

This then give 2^r equations for 2^{r+1} variables. We find the other 2^r equations by relating $\{Q_\alpha(n)\}$ to $\{p(\omega)\}$. First let us define $\xi(\omega)$ as

$$\xi(\omega) = \nu_1(\omega) - \nu_0(\omega) \bmod 3, \tag{6.3}$$

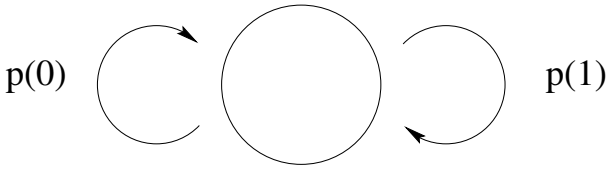

Figure 6.1: The most general $r = 0$ machine. Since this process has no memory, $r = 0$, we do not label the single state.

where $\nu_1(\omega)$ is the number of 1s in the word ω and $\nu_0(\omega)$ is the number of 0s in ω.

We now define the sets $\Upsilon_s(L)$, $\Upsilon_c(L)$, and $\Upsilon_r(L)$ as

$$\begin{aligned} \Upsilon_s(L) &= \{\omega \in \mathcal{A}^L : \xi(\omega) = 0\}, \\ \Upsilon_c(L) &= \{\omega \in \mathcal{A}^L : \xi(\omega) = 1\}, \\ \Upsilon_r(L) &= \{\omega \in \mathcal{A}^L : \xi(\omega) = -1\}. \end{aligned} \tag{6.4}$$

The interpretation of the $\Upsilon_\alpha(L)$ where $\alpha \in \{s, c, r\}$ is straight forward. $\Upsilon_c(L)$, for instance, is the set of all words of length L such that any two layers related by a stacking sequence described by a member of this set are cyclically related. Similarly for $\Upsilon_r(L)$ and $\Upsilon_s(L)$. We can now say

$$Q_\alpha(L) = \sum_{\omega \in \Upsilon_\alpha(L)} p(\omega). \tag{6.5}$$

We take as many of the latter relations as needed to obtain a complete set of equations that we can solve. We note that at each L (or n) there are at most two independent $Q_\alpha(n)$, since $\sum_\alpha Q_\alpha(n) = 1$.

Finding a de Bruijn graph which approximates the process does not however constitute finding the ϵ-machine. For that, we need to discover the minimal number of causal states. We do this by collapsing nodes that have equivalent futures.

6.1 The $r = 0$ approximation

As an example, let us consider $r = 0$ case. Figure 6.1 shows the most general machine with this memory. Since we take $L = r + 1$, we see that equations 6.1 provide no information. Further we expect $2^{L-1} = 2^0 = 1$ independent parameters. Applying equation 6.2 gives $p(0) + p(1) = 1$, showing that $p(0)$ and $p(1)$ are not independent. Using the relations 6.5 we finally get,

$$\begin{aligned} p(1) &= Q_c(1), \\ p(0) &= Q_r(1). \end{aligned} \tag{6.6}$$

There is no contradiction here, as $Q_c(1) + Q_r(1) = 1$ as dictated by stacking constraints. We see that the $\{Q_\alpha(n)\}$ are linearly related to the $\{p(w)\}$. We can now us this $r = 0$ machine to generate a sequence and find the resulting $\{Q_\alpha(n)\}$ and diffraction pattern. We call this resulting diffraction pattern the $r = 0$ approximation. We can compare this approximate diffraction pattern to the experimental one to determine whether we need to examine higher order graphs.

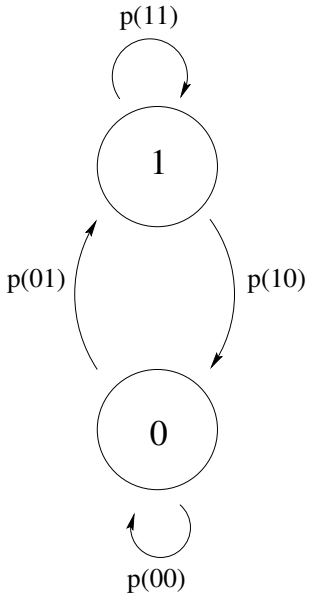

Figure 6.2: The most general $r = 1$ machine. This graph implies a memory of one, so we label the states by the last symbol seen. This graph can be described by two independent parameters. Two of the four variables, $p(11)$, $p(10)$, $p(01)$ and $p(00)$ are constrained by conservation of probability, so we need to use two values from the correlation functions to fix the graph. We use $Q_c(2)$ and $Q_r(2)$ for this purpose.

6.2 The $r = 1$ approximation

We now work out the $r = 1$ approximate machine. We need a total of four equations, two of which are constraints among $\{p(\omega)\}$ and two of which relate $\{p(\omega)\}$ and $\{Q_\alpha(n)\}$. Let us find the $\{p(\omega)\}$ in terms of $Q_c(2)$ and $Q_r(2)$. Using equation 6.1 we get $p(10) + p(00) = p(01) + p(00)$, giving $p(01) = p(10)$. Requiring the sum of the probabilities to be unity gives $p(11) + 2p(01) + p(00) = 1$. We now use equations 6.5 to get $Q_c(2) = p(00)$ and $Q_r(2) = p(11)$. This is a system of linear equations which are easily solved to give

$$\begin{aligned} p(11) &= Q_r(2), \\ p(01) &= p(10) = \frac{1}{2}[1 - Q_c(2) - Q_r(2)], \\ p(00) &= Q_c(2). \end{aligned} \tag{6.7}$$

The most general $r = 1$ ϵ-machine is shown in Figure 6.2.

6.3 The $r = 2$ approximation

We proceed analogously to find the $r = 2$ approximation to the process. The most general $r = 2$ machine is shown in Figure 6.3. There are a total of eight word probabilities at $r = 2$, so we need four equations relating constraints among the $\{p(\omega)\}$ and four equations relating the $\{Q_\alpha(n)\}$ to the $\{p(\omega)\}$. We can write the first three equations as

$$p(011) = p(110),$$

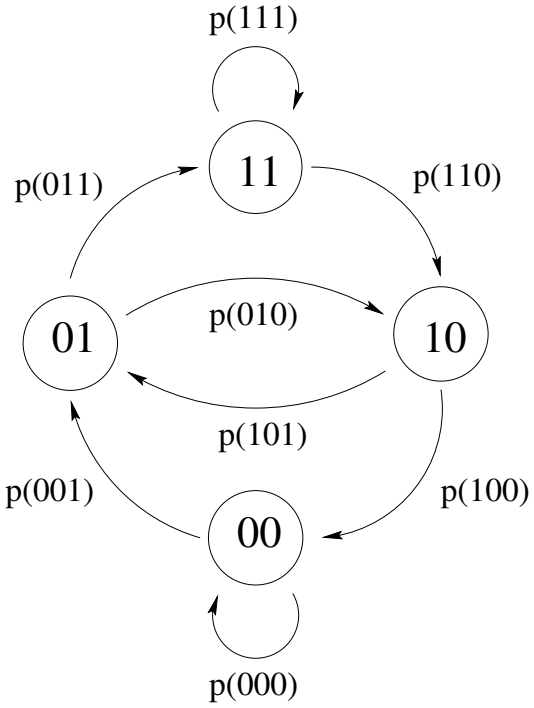

Figure 6.3: The most general $r = 2$ machine. Each of the four nodes is labeled by its unique history of the past two symbols seen. There are eight arcs, thus eight variables. There are four constraints arising from conservation of probability, so we need four additional conditions to fix the graph. We use $Q_c(2)$, $Q_r(2)$, $Q_c(3)$ and $Q_r(3)$ as shown in equations 6.11.

$$\begin{aligned} p(001) &= p(100), \\ p(110) + p(010) &= p(100) + p(101). \end{aligned} \tag{6.8}$$

We also require conservation of probability,

$$\begin{aligned} p(000) + p(001) + p(010) + p(011) + p(100) \\ + p(101) + p(110) + p(111) = 1. \end{aligned} \tag{6.9}$$

The four equations relating the $\{Q_\alpha(n)\}$ to the $\{p(\omega)\}$ are

$$\begin{aligned} Q_c(2) &= p(00) = p(000) + p(001), \\ Q_r(2) &= p(11) = p(110) + p(111), \\ Q_c(3) &= p(011) + p(101) + p(110), \\ Q_r(3) &= p(001) + p(010) + p(100). \end{aligned} \tag{6.10}$$

These equations are linear and can be solved to give

$$p(000) = \frac{1}{6}[+3Q_c(2) - 2Q_c(3) - 3Q_r(2) - 4Q_r(3) + 3],$$

$$
\begin{aligned}
p(001) &= \frac{1}{6}[+3Q_c(2) + 2Q_c(3) + 3Q_r(2) + 4Q_r(3) - 3], \\
p(010) &= \frac{1}{3}[-3Q_c(2) - 2Q_c(3) - 3Q_r(2) - Q_r(3) + 3], \\
p(011) &= \frac{1}{6}[+3Q_c(2) + 4Q_c(3) + 3Q_r(2) + 2Q_r(3) - 3], \\
p(100) &= \frac{1}{6}[+3Q_c(2) + 2Q_c(3) + 3Q_r(2) + 4Q_r(3) - 3], \\
p(101) &= \frac{1}{3}[-3Q_c(2) - Q_c(3) - 3Q_r(2) - 2Q_r(3) + 3], \\
p(110) &= \frac{1}{6}[+3Q_c(2) + 4Q_c(3) + 3Q_r(2) + 2Q_r(3) - 3], \\
p(111) &= \frac{1}{6}[-3Q_c(2) - 4Q_c(3) + 3Q_r(2) - 2Q_r(3) + 3].
\end{aligned}
\tag{6.11}
$$

6.4 The $r = 3$ approximation

Lastly we treat the case of $r = 3$. The most general $r = 3$ machine is shown in Figure 6.4. A total of 16 word probabilities dress the $r = 3$ graph, so we need eight constraints among the $\{p(\omega)\}$ and eight equations relating $\{p(\omega)\}$ to $\{Q_\alpha(n)\}$. The first seven constraints between the $\{p(\omega)\}$ can be written as

$$
\begin{aligned}
p(0111) &= p(1110), \\
p(0001) &= p(1000), \\
p(0011) + p(1011) &= p(0111) + p(0110), \\
p(0101) + p(1101) &= p(1011) + p(1010), \\
p(0010) + p(1010) &= p(0101) + p(0100), \\
p(0001) + p(1001) &= p(0011) + p(0010), \\
p(0100) + p(1100) &= p(1001) + p(1000).
\end{aligned}
\tag{6.12}
$$

We still require the overall probability of seeing a word of length four to be unity, so we have,

$$
\begin{aligned}
&p(0000) + p(0001) + p(0010) + p(0011) + p(0100) \\
&p(0101) + p(0110) + p(0111) + p(1000) + p(1001) \\
&p(1010) + p(1011) + p(1100) + p(1101) + p(1110) \\
&\qquad\qquad p(1111) = 1.
\end{aligned}
\tag{6.13}
$$

We now write out the equations which relate the $\{Q_\alpha(n)\}$ and the $\{p(\omega)\}$ and find,

$$
\begin{aligned}
Q_c(2) &= p(0000) + p(0001) + p(0010) + p(0011), \\
Q_r(2) &= p(1100) + p(1101) + p(1110) + p(1111), \\
Q_c(3) &= p(0110) + p(0111) + p(1010) + p(1011) \\
&\quad + p(1100) + p(1101), \\
Q_r(3) &= p(0010) + p(0011) + p(0100) + p(0101) \\
&\quad + p(1000) + p(1001),
\end{aligned}
$$

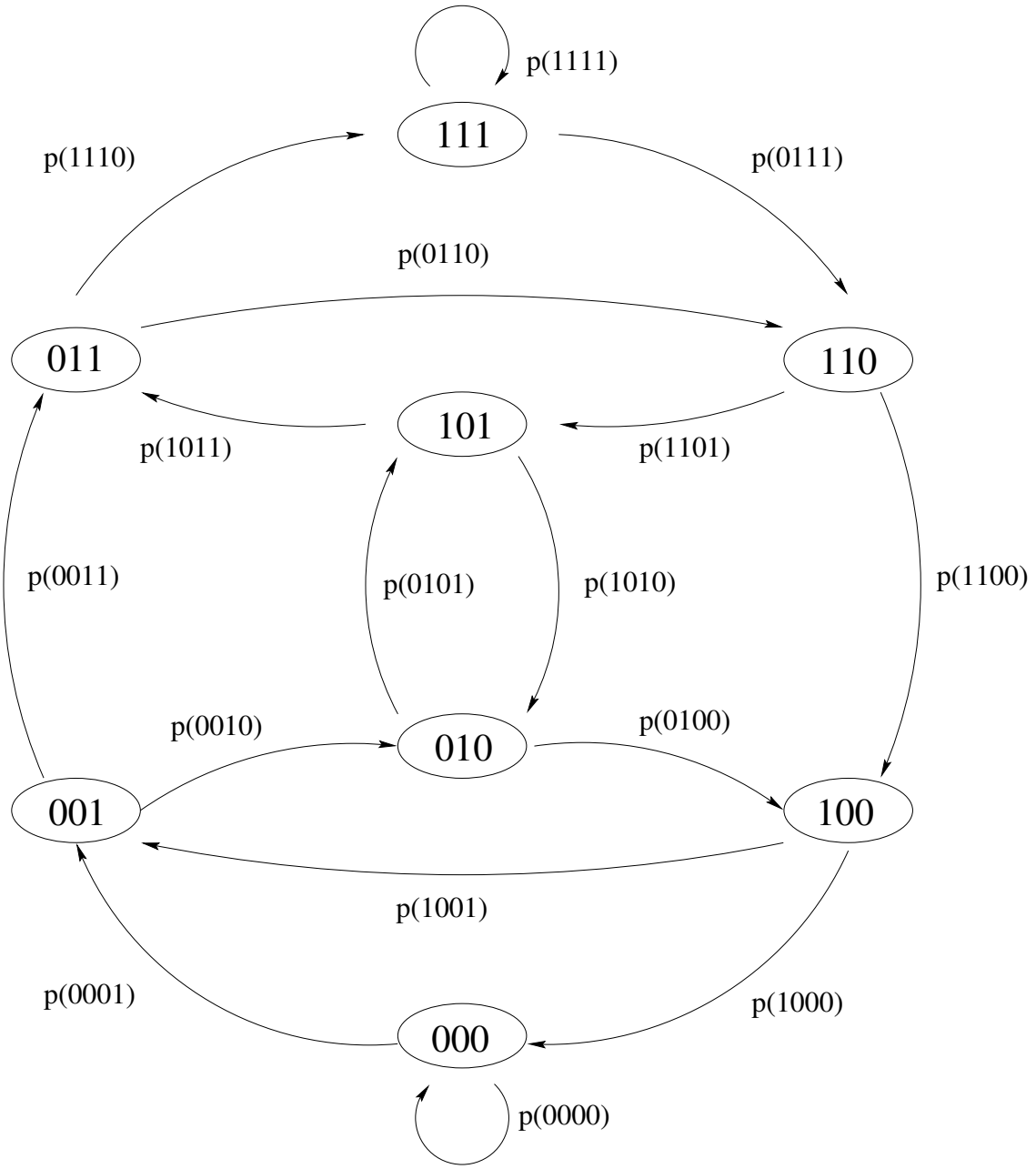

Figure 6.4: The most general $r = 3$ machine. We now have eight nodes and sixteen arcs at $r = 3$. Each node is labeled by the last three symbols seen. Not all of the word probabilities are independent, there are eight constraints coming from conservation of probability. The other eight constraints come using $Q_\alpha(n)$ where $\alpha \in \{c, r\}$ and $n \in \{2, 3, 4, 5\}$.

$$
\begin{aligned}
Q_c(4) &= p(1111) + p(1000) + p(0100) + p(0010) \\
&\quad + p(0001), \\
Q_r(4) &= p(0000) + p(0111) + p(1011) + p(1101) \\
&\quad + p(1110), \qquad (6.14) \\
Q_c(5) &= \frac{p^2(0000)}{p(0000)+p(0001)} + \frac{p(0011)p(0111)}{p(0111)+p(0110)} \\
&\quad + \frac{p(0101)p(1011)}{p(1011)+p(1010)} + \frac{p(0110)p(1101)}{p(1101)+p(1100)} \\
&\quad + \frac{p(0111)p(1110)}{p(1110)+p(1111)} + \frac{p(1001)p(0011)}{p(0011)+p(0010)} \\
&\quad + \frac{p(1010)p(0101)}{p(0101)+p(0100)} + \frac{p(1011)p(0110)}{p(0110)+p(0111)} \\
&\quad + \frac{p(1100)p(1001)}{p(1001)+p(1000)} + \frac{p(1101)p(1010)}{p(1010)+p(1011)} \\
&\quad + \frac{p(1110)p(1100)}{p(1100)+p(1101)}, \\
Q_r(5) &= \frac{p^2(1111)}{p(1111)+p(1110)} + \frac{p(1100)p(1000)}{p(1000)+p(1001)} \\
&\quad + \frac{p(1010)p(0100)}{p(0100)+p(0101)} + \frac{p(1001)p(0010)}{p(0010)+p(0011)} \\
&\quad + \frac{p(1000)p(0001)}{p(0001)+p(0000)} + \frac{p(0110)p(1100)}{p(1100)+p(1101)} \\
&\quad + \frac{p(0101)p(1010)}{p(1010)+p(1011)} + \frac{p(0100)p(1001)}{p(1001)+p(1000)} \\
&\quad + \frac{p(0011)p(0110)}{p(0110)+p(0111)} + \frac{p(0010)p(0101)}{p(0101)+p(0100)} \\
&\quad + \frac{p(0001)p(0011)}{p(0011)+p(0010)}.
\end{aligned}
$$

The last two relations of equations 6.14 require some explanation. At $L = 5$, a typical term in the sum to find $Q_c(5)$ might look like $p(00111)$. We want to express this probability of seeing a length five word as some function of the probability of seeing length four words. So we say that $p(00111) = p(0011)p(1|0011)$, where $p(1|0011)$ is the conditional probability of seeing a 1 having already seen a 0011. But looking at the Figure 6.4 it is clear that having last seen a 0011 puts us at the node 011. The probability of seeing another 1 is just the branching ratio at node 011, which we can write as $\frac{p(0111)}{p(0111)+p(0110)}$. Therefore, we find that $p(00111) = \frac{p(0011)p(0111)}{p(0110)+p(0111)}$.

It is reasonable to ask why we need the de Bruijn graphs to write down these relations. Some seem to follow directly from probability theory. Indeed, the de Bruijn graphs are a convenient and pictorial way to describe r^{th}-order Markov processes. Up to $r = 2$, the de Bruijn graphs impose no additional constraints on the probabilities and one can derive the same $r = 2$ equations from elementary probability theory. At $r = 3$, they do alter the equations slightly. By making the assumption of a finite range, the conditional probabilities are truncated to look only at a depth r. Consider the term $p(00111)$. We found this term to be equal to $\frac{p(0011)p(0111)}{p(0110)+p(0111)}$ which can be rewritten

as $p(1|011)p(0011)$. From elementary probability theory, we obtain $p(00111) = p(1|0011)p(0011)$. So we have effectively approximated the term $p(1|0011)$ by $p(1|011)$.

We immediately notice that the last two of these equations are nonlinear. We therefore have not found an analytical solution and have instead resorted to numerical solutions. At first glance these equations might seem easy to solve, after all fourteen of them are linear and the nonlinearities are algebraic. However, in practice it is the case that denominators of many of the terms in the last two equations become quite small, especially if there are forbidden words. Let us now turn to the task of solving these equations.

6.5 Solutions to the $r = 3$ Equations

Let us require $0 \leq p(\omega) \leq 1$ for all ω. Let the space occupied by this 16-dimensional hypercube be denoted as $\Re_s$. In general, nonlinear equations must be solved using numerical techniques. Several standard procedures are available, such as Newton's method, secant methods such as Broyden's method, and steepest descent techniques [11] [20]. Newton's method requires finding the derivatives of the equations and while it does converge quadratically for initial starting points close to the answer, if one starts far from a root, convergence is not guaranteed. There is also the necessity of finding and inverting the Jacobian matrix at each iteration. The method fails if the Jacobian matrix becomes singular. Broyden's method avoids having to calculate the Jacobian by approximating it. It converges only super-linearly. The steepest descent method converges linearly and it is not as sensitive to the initial starting point. The main drawback to using any of these three is that they are intended for *unconstrained* problems. That is, no precondition, save perhaps the reality of the solution, is imposed *a priori*. The solutions to our equations can only be physically meaningful for real word probabilities lying on the unit interval. We expect the data will yield correlation functions that contain considerable error, and may take the solution $\{p(\omega)\}$ out of $\Re_s$. It is perhaps better, then, to think of this as a constrained optimization problem. The constraint is that whatever we call a solution must lie in $\Re_s$, and the optimization is the recognition that there might not be $\{p(\omega)\}$ satisfying the equations which lies in $\Re_s$ and therefore we might have to settle for some approximate, 'best' solution. As such, we forgo mathematically exact solutions lying outside $\Re_s$ and instead seek $\{p(\omega)\} \in \Re_s$ which best satisfy the equations, where best is defined as the $\{p(\omega)\} \in \Re_s$ which minimize the fitness function, $\mathcal{F}$. We further define the fitness function as

$$\mathcal{F}(\{p(\omega)\}) = \sum_{i}^{N_{eq}} a_i^2 f_i^2(\{p(\omega)\}), \tag{6.15}$$

where N_{eq} is the number of equations, and a_i is a weight factor. The $f\Big(\{p(\omega)\}\Big)$ are a simple rewriting of the equations 6.12, 6.13, and 6.14. Explicitly, they are

$$\begin{aligned}
f_1 &= p(1110) - p(0111), \\
f_2 &= p(1000) - p(0001), \\
f_3 &= p(0111) + p(0110) - p(0011) - p(1011), \\
f_4 &= p(1011) + p(1010) - p(0101) - p(1101), \\
f_5 &= p(0101) + p(0100) - p(0010) - p(1010), \\
f_6 &= p(0011) + p(0010) - p(0001) - p(1001),
\end{aligned}$$

$$
\begin{aligned}
f_7 &= p(1001) + p(1000) - p(0100) - p(1100), \\
f_8 &= p(0000) + p(0001) + p(0010) + p(0011) + p(0100) \\
&\quad + p(0101) + p(0110) + p(0111) + p(1000) + p(1001) \\
&\quad + p(1010) + p(1011) + p(1100) + p(1101) + p(1110) \\
&\quad + p(1111) - 1, \\
f_9 &= p(0000) + p(0001) + p(0010) + p(0011) - Q_c(2), \\
f_{10} &= p(1100) + p(1101) + p(1110) + p(1111) - Q_r(2), \\
f_{11} &= p(0110) + p(0111) + p(1010) + p(1011) \\
&\quad + p(1100) + p(1101) - Q_c(3), \\
f_{12} &= p(0010) + p(0011) + p(0100) + p(0101) \\
&\quad + p(1000) + p(1001) - Q_r(3), \\
f_{13} &= p(1111) + p(1000) + p(0100) + p(0010) \\
&\quad + p(0001) - Q_c(4), \\
f_{14} &= p(0000) + p(0111) + p(1011) + p(1101) \\
&\quad + p(1110) - Q_r(4), \qquad (6.16) \\
f_{15} &= \frac{p^2(0000)}{p(0000) + p(0001)} + \frac{p(0011)p(0111)}{p(0111) + p(0110)} \\
&\quad + \frac{p(0101)p(1011)}{p(1011) + p(1010)} + \frac{p(0110)p(1101)}{p(1101) + p(1100)} \\
&\quad + \frac{p(0111)p(1110)}{p(1110) + p(1111)} + \frac{p(1001)p(0011)}{p(0011) + p(0010)} \\
&\quad + \frac{p(1010)p(0101)}{p(0101) + p(0100)} + \frac{p(1011)p(0110)}{p(0110) + p(0111)} \\
&\quad + \frac{p(1100)p(1001)}{p(1001) + p(1000)} + \frac{p(1101)p(1010)}{p(1010) + p(1011)} \\
&\quad + \frac{p(1110)p(1100)}{p(1100) + p(1101)} - Q_c(5), \\
f_{16} &= \frac{p^2(1111)}{p(1111) + p(1110)} + \frac{p(1100)p(1000)}{p(1000) + p(1001)} \\
&\quad + \frac{p(1010)p(0100)}{p(0100) + p(0101)} + \frac{p(1001)p(0010)}{p(0010) + p(0011)} \\
&\quad + \frac{p(1000)p(0001)}{p(0001) + p(0000)} + \frac{p(0110)p(1100)}{p(1100) + p(1101)} \\
&\quad + \frac{p(0101)p(1010)}{p(1010) + p(1011)} + \frac{p(0100)p(1001)}{p(1001) + p(1000)} \\
&\quad + \frac{p(0011)p(0110)}{p(0110) + p(0111)} + \frac{p(0010)p(0101)}{p(0101) + p(0100)} \\
&\quad + \frac{p(0001)p(0011)}{p(0011) + p(0010)} - Q_r(5).
\end{aligned}
$$

To find $\{p(\omega)\}$ which best satisfy the equations, we choose to employ a Monte Carlo method. Let us begin with a random choice for $\{p(\omega)\} \in \Re_s$ and evaluate $\mathcal{F}$. We call the resulting numerical value for the fitness function $\mathcal{F}_{old}$ and the set of word probabilities $\{p_{old}(\omega)\}$. One member of the $\{p_{old}(\omega)\}$ is incremented/decremented to get $\{p_{new}(\omega)\}$ and a new value of the fitness function $\mathcal{F}_{new}$ is calculated. The incrementation/decrementation procedure is never allowed to take $\{p_{new}(\omega)\}$ out of $\Re_s$. The standard Monte Carlo procedure is followed to decide whether to accept $\{p_{new}(\omega)\}$. Namely, if $\mathcal{F}_{new} \leq \mathcal{F}_{old}$ the $\{p_{new}(\omega)\}$ is accepted. If $\mathcal{F}_{new} > \mathcal{F}_{old}$, then the $\{p_{new}(\omega)\}$ is accepted with a probability $\exp^{-(\mathcal{F}_{new}-\mathcal{F}_{old})/\tau}$, where τ is a "fake" temperature. The procedure is then repeated for a maximum of 10^9 iterations, or until $\mathcal{F}_{new}$ falls below some preset minimum value. Initially, τ is set high, at 10^{-2}, so that most choices of $\{p_{new}(\omega)\}$ are accepted, but then it is gradually lowered so that the process becomes increasingly selective. This hopefully eliminates any bias that the initial guess may give and permits the $\{p(\omega)\}$ to freely explore much of $\Re_s$. τ is never allowed fall below one one-hundredth of the smallest value of $\mathcal{F}$ that the algorithm has found. a_i is always taken as unity here, although we have tried to using different values, to reflect the fact that equations 6.12 and 6.13 should be satisfied identically. Using larger a_i for $i = 1, 2, ..., 8$ penalizes failure to satisfy conservation of probability relations more heavily than the equations that relate the $\{p(\omega)\}$ to the $\{Q_\alpha(n)\}$. However we find the resulting $\{p(\omega)\}$ to be much the same. The entire algorithm is repeated twenty times.

A few tricks are also used to help find a minimum of $\mathcal{F}$. Randomly, two of the $\{p_{old}(\omega)\}$ are incremented simultaneously. This seems to accelerate convergence to the minimum. Also, the amount by which a $p(\omega)$ is incremented (the step size) is adjusted during search. Initially, it is large, on the order of 10^{-1}, but as τ falls, it too is reduced to a minimum value of 10^{-8}. Periodically, the step size is increased and allowed to fall again. This helps the prevent $\{p(\omega)\}$ from getting stuck in a local minimum.

It is often possible to find more than one $\{p(\omega)\}$ which give nearly the same $\mathcal{F}$. One possibility is that the solutions are symmetry related. The equations 6.16 are of the form that if $\{p(\omega)\}$ is a solution, then $\{p'(\omega)\}$ is also, where we define $p'(\omega) = p(\overline{\omega})$, where the overline indicates spacial inversion. (For example, $\overline{0010} = 0100$). We can not distinguish between these two roots. Another possibility is that there are distinct roots, not related by symmetry. In this event, other criteria must be used to select among the possible $\{p(\omega)\}$.

We have tested this for sample data, to see if it works. The first process we examine is a completely random one. It is chosen because analytical solutions are available. We want to see if the method works for exact data and also for some data with noise. The correlation functions for the random number generator are given in Table 6.1. The noisy correlation functions differ from the exact values by about ± 0.0013 on average, a value much smaller than one expects from experiment. The solution to the equations is shown in Table 6.2. Using the exact correlation functions, the method finds the correct $\{p(\omega)\}$ result to within ± 0.0001 for each $p(\omega)$. For the noisy correlation functions, we see an error in the average $p(\omega)$ of about ± 0.0013.

The correlation functions for a second process we examine are shown in Table 6.3. The minimal graph that describes this process is $r = 3$. The numerical solutions to equations (5) for these correlation functions is shown in Table 6.4. First one notices that there are two solutions. These solutions are symmetry related, and either is a reasonable answer. The algorithm converged to solution #1 nine times, and found solution #2 eleven times. From the smallness of the fitness function, as well as examining the $\{p(\omega)\}$, we see that the algorithm has found an excellent solution. The noisy correlation functions differ from their exact values by ± 0.00079 on average. Again, this error is much smaller than we expect from experiment. Curiously, we find that the two solutions for the exact correlation functions have merged into one solution for the noisy Q's, which seems to just be the average of the two. The fitness is again quite good.

Table 6.1: Exact and noisy correlation functions for the random number process. These are the correlation functions used in solving equations 6.16. The exact $\{Q_\alpha(n)\}$ are found from an analytical solution for the random number generator, and the noisy version is found by considering a finite sample of the process.

n	$Q_c(n)$ exact	$Q_r(n)$ exact	$Q_c(n)$ noisy	$Q_r(n)$ noisy
2	0.25000	0.25000	0.2539	0.2489
3	0.37500	0.37500	0.3744	0.3732
4	0.31250	0.31250	0.3127	0.3142
5	0.34375	0.34375	0.3444	0.3407

Table 6.2: Solution at $r = 3$ for the random number generator. Using the exact correlation functions, we get a fitness of $\mathcal{F} = 1.26 \times 10^{-14}$. With the noisy correlation functions, the fit is not nearly as good, $\mathcal{F} = 1.72 \times 10^{-6}$.

Word	Exact Answer	Numerical Solution with exact Q's	Numerical Solution with noisy Q's
$p(0000)$	0.0625	0.0626	0.0668
$p(0001)$	0.0625	0.0624	0.0619
$p(0010)$	0.0625	0.0625	0.0630
$p(0011)$	0.0625	0.0625	0.0622
$p(0100)$	0.0625	0.0626	0.0631
$p(0101)$	0.0625	0.0624	0.0599
$p(0110)$	0.0625	0.0626	0.0648
$p(0111)$	0.0625	0.0624	0.0607
$p(1000)$	0.0625	0.0624	0.0618
$p(1001)$	0.0625	0.0626	0.0634
$p(1010)$	0.0625	0.0625	0.0599
$p(1011)$	0.0625	0.0625	0.0633
$p(1100)$	0.0625	0.0624	0.0622
$p(1101)$	0.0625	0.0626	0.0633
$p(1110)$	0.0625	0.0624	0.0608
$p(1111)$	0.0625	0.0626	0.0625

Table 6.3: Exact and noisy correlation functions for the second process. These are the correlation functions used in solving equations 6.16. The exact $\{Q_\alpha(n)\}$ are found from direct analytical calculation from the process. Only four digits of accuracy are reported here for convenience. The noisy version is found by considering a finite sample of the process.

n	$Q_c(n)$ exact	$Q_r(n)$ exact	$Q_c(n)$ noisy	$Q_r(n)$ noisy
2	0.2300	0.1300	0.2279	0.1289
3	0.3600	0.4500	0.3611	0.4504
4	0.2900	0.1200	0.2891	0.1191
5	0.2969	0.5254	0.2981	0.5252

Table 6.4: Solution at $r = 3$ for for a second process. Using the exact correlation functions, (accurate to 10^{-11}), I find two solutions. The first has a fitness of $\mathcal{F} = 1.21 \times 10^{-16}$ and the second has a fitness of $\mathcal{F} = 1.62 \times 10^{-16}$. The two solutions are symmetry related. The numerical solution for the noisy correlations has a fitness of $\mathcal{F} = 2.31 \times 10^{-9}$.

Word	Exact Answer	Numerical Solution #1 with exact Q's	Numerical Solution #2 with exact Q's	Numerical Solution with noisy Q's
$p(0000)$	0.050	0.050	0.050	0.050
$p(0001)$	0.080	0.080	0.080	0.079
$p(0010)$	0.040	0.050	0.040	0.045
$p(0011)$	0.060	0.050	0.060	0.054
$p(0100)$	0.050	0.040	0.050	0.045
$p(0101)$	0.200	0.210	0.200	0.206
$p(0110)$	0.050	0.050	0.050	0.051
$p(0111)$	0.020	0.020	0.020	0.019
$p(1000)$	0.080	0.080	0.080	0.079
$p(1001)$	0.020	0.020	0.020	0.021
$p(1010)$	0.210	0.200	0.210	0.206
$p(1011)$	0.010	0.020	0.010	0.016
$p(1100)$	0.050	0.060	0.050	0.054
$p(1101)$	0.020	0.010	0.020	0.016
$p(1110)$	0.020	0.020	0.020	0.019
$p(1111)$	0.040	0.040	0.040	0.041

6.6 The $r = 4$ and higher approximations

As we reconstruct processes of larger and larger r, we notice that finding a solution is becoming more and more difficult. It is therefore reasonable to ask why this is so and to try to determine just how complicated the equations are for a given r. The number of variables we need to solve for grows exponentially in r, given by the relation $N_V(r) = 2^{r+1}$. Just from this we see that the computational requirements become onerous fast. But we also notice that as we increment r, the additional number of constraints in the form of Qs only increases by a constant, namely two. This is a disaster. It necessitates using equations of higher n to generate enough constraints to solve for the word probabilities. The problem is even worse than this however. We need to go out exponentially far in n to find enough constraints, but the number of terms in the highest $Q_\alpha(n)$ equation grows exponentially in n. So, the maximum number of terms in the equations we need to solve for a given r grows like the exponential of an exponential, ie, $N_{terms} \propto 2^{(2^r+1)}$. The problem is, however, even more difficult than this. It is these equations that have so many terms that turn out to be nonlinear! All of the terms are added, so we can expect no fortuitous cancellations. The problem has become intractable. It is probably true that $r = 4$ is doable, though requiring much work. It is likely $r = 5$ can not be done, at least in general, and higher r are almost certainly prohibitively difficult. These results are tabulated in Table 6.5. For an illustration of the $r = 4$ graph, see Figure 6.5.

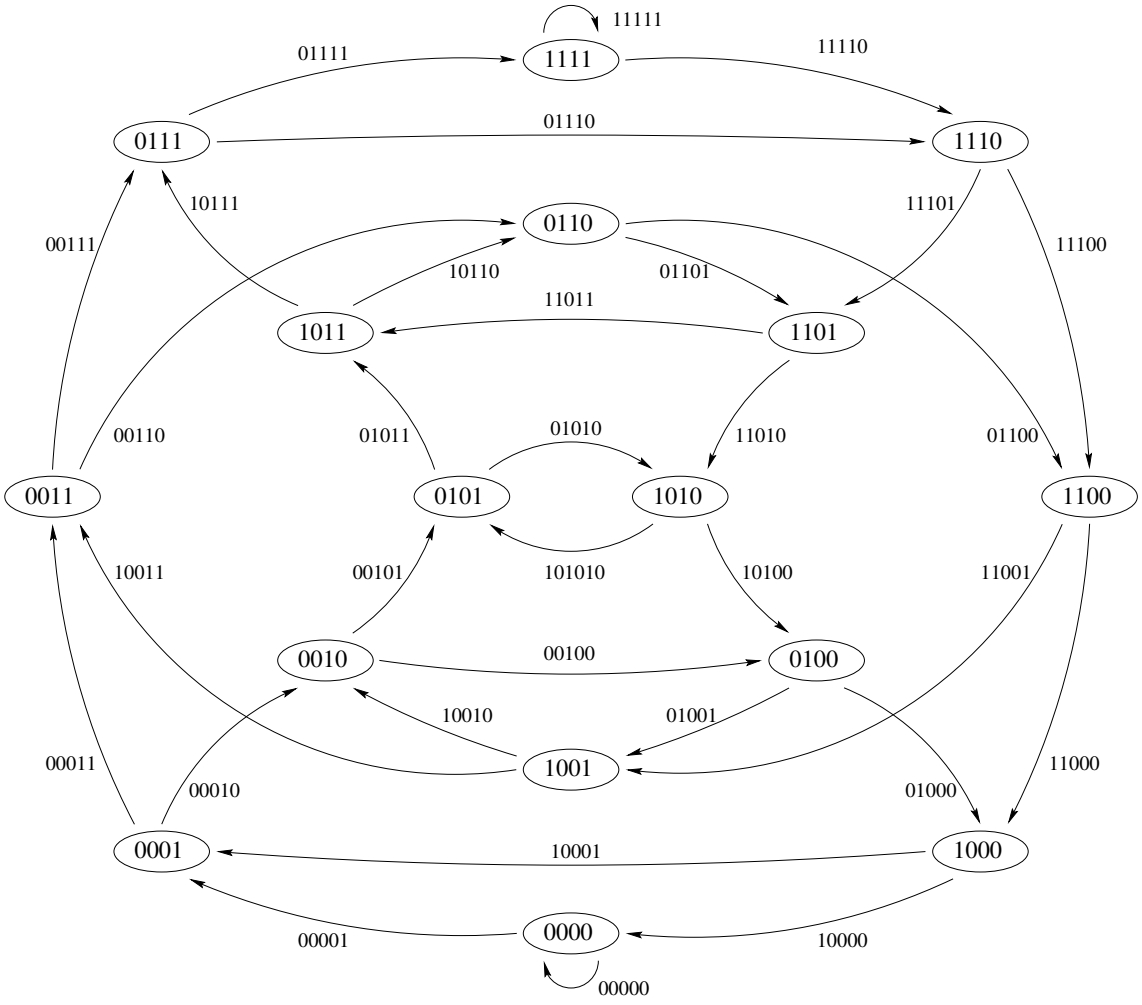

Figure 6.5: The most general $r = 4$ machine. The number of nodes has grown to sixteen and now there are thirty-two arcs. Conservation of probability provides sixteen constraints among the arcs. The remaining sixteen constraints come from the correlation functions, and we would use $Q_\alpha(n)$ with $\alpha \in \{c, r\}$ and $n \in \{2, 3, 4, 5, 6, 7, 8, 9\}$. To solve for this system we would need to solve thirty-two simultaneous algebraic equations, eight of which would be non-linear. At $n = 9$, the two equations relating correlation functions to word probabilities would have 171 terms.

Table 6.5: Mathematical requirements for general solution of equations as a function of r. N_V is the total number of variables, N_{IV} is the total number of independent variables after constraints among the probabilities are taken into account, N_{NLEq} is the number of nonlinear equations, and finally the killer is N_{terms}, the maximum number of terms in any one equation. We can see that this number becomes prohibitively large fast.

r	n	N_V	N_{IV}	N_{NLEq}	N_{terms}
0	1	2	1	0	1
1	2	4	2	0	4
2	3	8	4	0	8
3	5	16	8	2	16
4	9	32	16	8	171
5	17	64	32	22	43690
6	33	128	64	52	2.86×10^9
7	65	256	128	114	1.23×10^{19}

Chapter 7

Examples of ϵ-Machine Reconstruction from Known Processes

We will now consider four examples of machine reconstruction from processes that can not be represented on an $r = 3$ de Bruijn graph. For any process that is describable in terms of a third-order Markov process, our reconstruction procedure will find the underlying process. It is for those cases where the structure is not third-order Markovian that we wish to determine how the reconstruction algorithm works. So, we will treat the $r = 4$ process given in §5.9, another $r = 4$ process not so far discussed, the noisy period two in §5.11, and the even system, §5.12.

7.1 Machine Reconstruction for the 3C/2H/6H Process

Let us begin with the 3C/2H/6H process described in §5.9. We see that this process has both a strong fcc component and 6H cycle. We might therefore expect that this will give our algorithm some difficulty as the simultaneous existence of these two is possible only on an $r = 4$ graph. The machine reconstruction results are given in Table 7.1. We also calculate the the correlation functions and the diffraction patterns of the various r approximations. Since the $r = 0$ approximation at best corresponds to a biased random number generator, and such a process clearly does not represent the structure seen in the 3C/2H/6H process, we do not calculate it here.

Figure 7.1 shows $Q_s(n)$ versus n for both the 3C/2H/6H process and the $r = 1$ approximation to it. We see that with exception of $n = 1, 2$ and 3, the correlations die out far too fast for the $r = 1$ approximation. Indeed, there appears to be some significant long range structure that the $r = 1$ is failing to capture. A comparison of the diffraction pattern of the $r = 1$ approximation to that of the 3C/2H/6H process is shown in Figure 7.2. We see that Bragg peak at $l = \frac{1}{3}$, save for a small bump, is missing in the $r = 1$ approximation. The only structure in the diffraction pattern that the approximation models moderately well is that at $l = \frac{2}{3}$. The other features in the spectrum, at $l = \frac{1}{6}, \frac{1}{2}$ and $\frac{5}{6}$, are totally absent in the $r = 1$ approximation to the diffraction pattern.

Figure 7.3 compares the $Q_s(n)$ for the $r = 2$ approximation to that of the 3C/2H/6H process. We see that very little, if any, additional structure has been discovered by increasing r. Examining the diffraction patterns in Figure 7.4 paints an even more dismal picture. We see a smooth diffuse background for the $r = 2$ spectrum, giving scant notice to the diffraction maximum in the 3C/2H/6H process. It seems that no progress has been made in increasing r from 1 to 2. We can also compare the computation measures for the two approximations, as shown in Table 7.2. We observe that the entropy density h_μ only decreases slightly (from $0.910 \rightarrow 0.904$) as r goes $1 \rightarrow 2$. The excess entropy

Table 7.1: Machine reconstruction results for the 3C/2H/6H Process. We show the word probabilities resulting from the $r = 0, 1, 2, 3$ approximations to the 3C/2H/6H process. For comparison, the exact word probabilities from the 3C/2H/6H process are given in the last column. Comparing the last two columns, we see that the $r = 3$ approximation does not do a very good job. It forbids some words the exact process allows, $p(0001)$ for example, and *vice versa*. Most seriously, the graph is not strongly connected as $p(0001) = p(1000) = 0$ while $p(0000) \neq 0$.

r values	Words	Word Probabilities	r values	Words	Reconstructed WPs	Exact WPs
$r = 0$	$p(0)$	0.394	$r = 3$	$p(0000)$	0.123	0.000
	$p(1)$	0.606		$p(0001)$	0.000	0.111
				$p(0010)$	0.034	0.000
$r = 1$	$p(00)$	0.223		$p(0011)$	0.064	0.111
	$p(01)$	0.172		$p(0100)$	0.068	0.000
	$p(10)$	0.172		$p(0101)$	0.000	0.049
	$p(11)$	0.434		$p(0110)$	0.098	0.000
				$p(0111)$	0.009	0.124
$r = 2$	$p(000)$	0.111		$p(1000)$	0.001	0.111
	$p(001)$	0.111		$p(1001)$	0.099	0.000
	$p(010)$	0.048		$p(1010)$	0.035	0.049
	$p(011)$	0.124		$p(1011)$	0.045	0.012
	$p(100)$	0.111		$p(1100)$	0.033	0.111
	$p(101)$	0.060		$p(1101)$	0.084	0.012
	$p(110)$	0.124		$p(1110)$	0.009	0.124
	$p(111)$	0.311		$p(1111)$	0.304	0.185

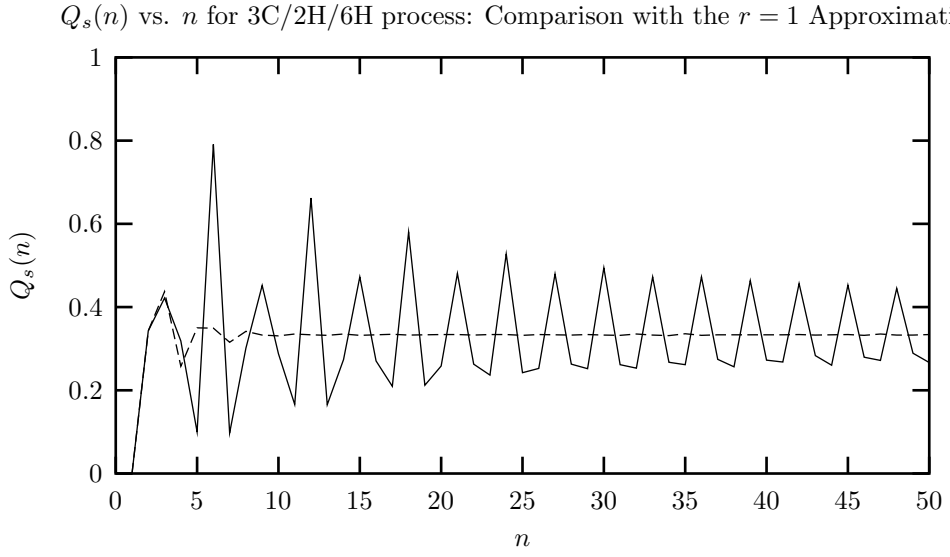

Figure 7.1: $Q_s(n)$ vs. n for the 3C/2H/6H Process (solid line) and the $r = 1$ approximation to the process (dashed line). We can see that the correlation functions for the $r = 1$ approximation die out entirely too fast.

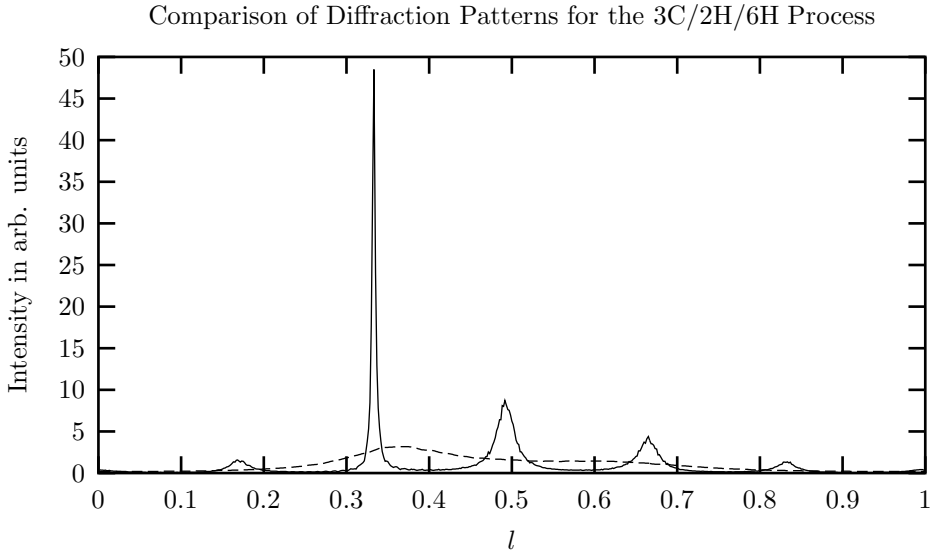

Figure 7.2: The diffraction pattern for a lattice stacked according to the 3C/2H/6H process (solid line) and the $r = 1$ approximation (dashed line). The diffraction pattern for the $r = 1$ approximation is clearly missing all of the structure present. Of particular note is the fact that the small rises at $l = \frac{1}{6}$ and $\frac{5}{6}$ are completely absent in the diffraction pattern for the $r = 1$ approximation.

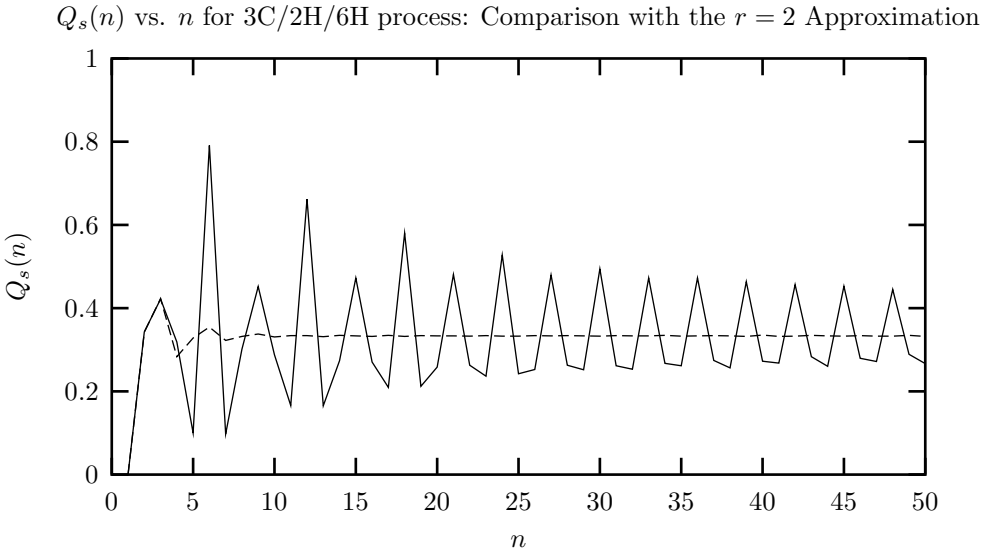

Figure 7.3: $Q_s(n)$ vs. n for the 3C/2H/6H Process (solid line) and the $r = 2$ approximation to the process (dashed line). Again, as with the $r = 1$ approximation, we see that the correlation functions die out too quickly for the $r = 2$ approximation.

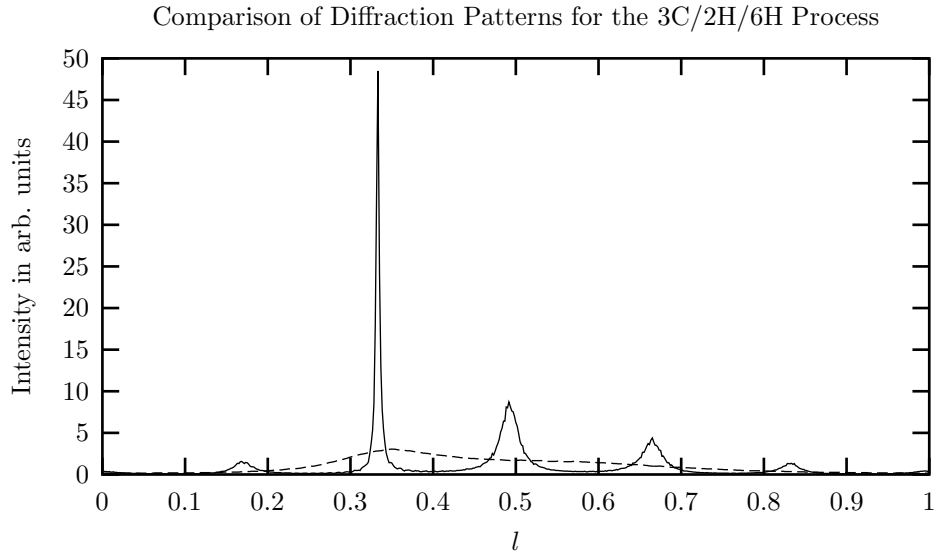

Figure 7.4: The diffraction pattern for a lattice stacked according to the 3C/2H/6H process (solid line) and the $r = 2$ approximation (dashed line). The $r = 2$ approximation seems to do little better than the $r = 2$ approximation at matching the diffraction pattern.

and the transient information likewise only show small changes. We note that the $r = 2$ machine collapses into a three state machine because the '11' node and the '01' node have equivalent futures. We can see that $p(1|11) \approx p(1|01) \approx 0.72$.

The machine reconstruction results for $r = 3$ are shown in the next to last column of Table 7.1. We find a fitness of $\mathcal{F} = 1.35 \times 10^{-4}$. From the table we also see that there are some forbidden words appearing. We find that '0001' and '0101' have vanishing probabilities and '1000' is nearly so. (Of course, $p(0001)$ and $p(1000)$ should be identically equal, but the relatively large value of the fitness indicates that some equations are not well satisfied.) This is problematic because '0000' has a reasonably large probability of occurrence at 0.123. In fact, it is the second most common word of length four in our reconstruction, behind only '1111'. We find that the graph is no longer strongly connected. So, machine reconstruction at $r = 3$ has failed. We cannot calculate a sample of the language or find the diffraction pattern. We can however compare our reconstructed word probabilities with those of the 3C/2H/6H process, which are given in the last column of Table 7.1. We

Table 7.2: Computational results for 3C/2H/6H process and the $r = 0, 1$ and 2 approximations. Since machine reconstruction has failed at $r = 3$, we can not calculate measures of computation for it. Notice that at even at $r = 2$, the entropy density remains well above that of the actual process.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
3C/2H/6H process, Example A	SFT	4	0.341	0.659	3.19	1.82	3.71
$r = 0$ Approximation	SFT	0	0.967	0.000	0.00	0.000	0.000
$r = 1$ Approximation	SFT	1	0.910	0.090	0.967	0.056	0.056
$r = 2$ Approximation	SFT	2	0.904	0.096	1.36	0.070	0.076
$r = 3$ Approximation	SFT	3	—	—	—	—	—

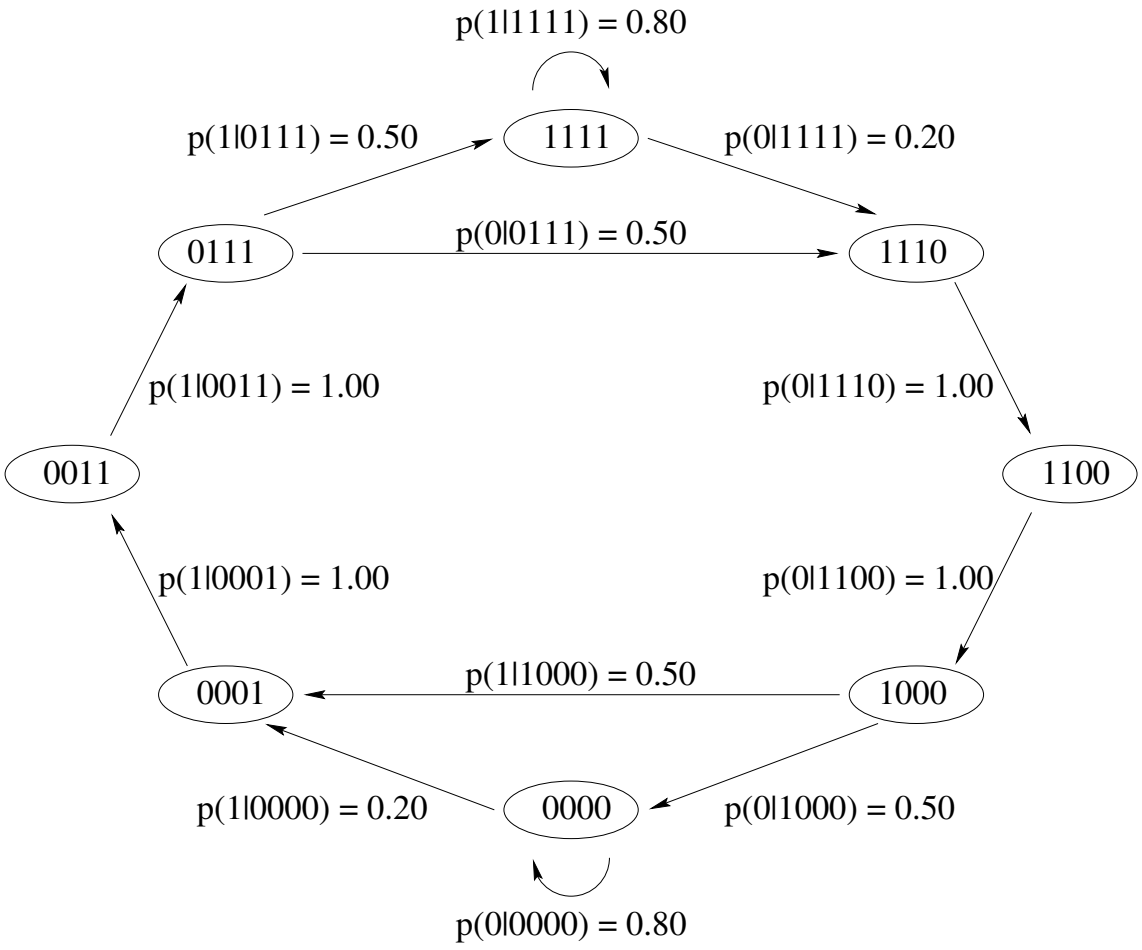

Figure 7.5: The recurrent portion of the ϵ-machine for the 3C/6H process. Since we wish to have a barrier between the two cycles representing the 3C and the 6H, we must use a graph with an $r = 4$ memory. So this graph is a portion of a $r = 4$ de Bruijn graph.

see that actually '0000' does not occur at all in the 3C/2H/6H process, and the reconstructed machine significantly overestimates $p(1111)$. Similarly, other word probabilities are not well-represented. We understand these results as stemming from the inability of an $r = 3$ graph to simultaneously support both 3C and 6H structure.

7.2 Machine Reconstruction for the 3C/6H Process

We now apply machine reconstruction to another process which is only describable by a machine with a memory of at least $r = 4$. We expect that some ZnS crystals will incompletely transform to a twinned 3C structure and contain remnants of the 6H structure. We call this the 3C/6H process, and the recurrent portion of the ϵ-machine is shown in Figure 7.5.

The results for machine reconstruction up to $r = 3$ are shown in Table 7.3. We find the correlation functions for the $r = 1$ approximation and the first fifty values of $Q_s(n)$ from the $r = 1$ approximation are compared with those of the 3C/6H process in Figure 7.6. Perhaps not surprisingly, there is only good agreement for the first few n, after which the $Q_s(n)$ for the $r = 1$ approximation decay too

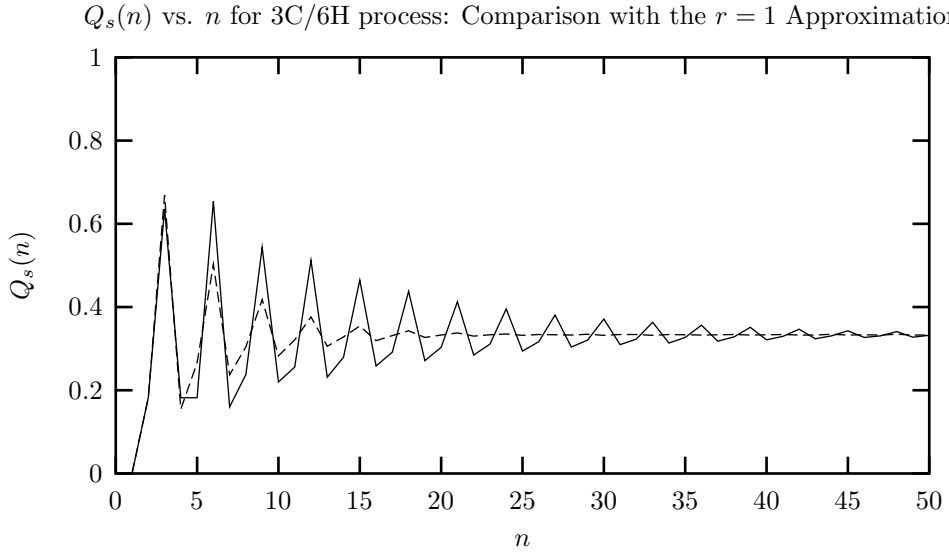

Figure 7.6: $Q_s(n)$ vs. n for the 3C/6H process (solid line) and the $r = 1$ approximation to the process (dashed line). The correlation functions for the $r = 1$ approximation do not show nearly as much structure as those of the actual process.

rapidly to their asymptotic value. Comparing the diffraction patterns as shown in Figure 7.7, we see that the $r = 1$ approximation does show some of the structure evident in the original spectrum at $l = \frac{1}{3}$ and $\frac{2}{3}$, but it is too diffuse. The bumps in the spectrum of the 3C/6H process at $l = \frac{1}{6}, \frac{1}{2}$ and $\frac{5}{6}$ are not evident at all in the the spectrum of the $r = 1$ process. We should expect this, as the structure at these l comes from the 6H cycles, which are not present in the $r = 1$ graph.

We find that the $r = 2$ gives only slight, if any, improvement. Again we should expect this, as there are no new cycles in the underlying process to be discovered upon moving from $r = 1$ to $r = 2$. A comparison of the $Q_s(n)$ is shown in Figure 7.8. It seems that they reflect a little more structure, but not much. A comparison of the diffraction patterns for the 3C/6H process and the $r = 2$ approximation is shown in Figure 7.9. Arguably, the agreement has become worse. It seems that the peaks at $l = \frac{1}{3}$ and $\frac{2}{3}$, while remaining too diffuse, have also shifted slightly to the left and right, respectively. The structure in the 3C/6H spectrum at $l = \frac{1}{6}, \frac{1}{2}$ and $\frac{5}{6}$ remains absent in the $r = 2$ approximation.

Finally, we compare the $r = 3$ approximation to the 3C/6H process. In Figure 7.10 we see the first fifty $Q_s(n)$ versus n plotted for each. The agreement is quite good. The peaks at $l = \frac{1}{3}$ and $\frac{2}{3}$ in the 3C/6H spectrum are reproduced well in the $r = 3$ approximation. The $r = 3$ approximation even seems to be picking up the small rise in the spectrum at $l = \frac{1}{2}$. The bumps at $l = \frac{1}{6}$ and $\frac{5}{6}$ are missing however. At $r = 3$, for the first time we have the possibility of modeling the 6H portion of the process. It is surprising that it does so well. We should view this as cautionary, however, because fitting the peaks and ignoring the diffuse background scattering can be misleading. We have found a process that generates correlation functions similar to those of the 3C/6H process, but it is different. Namely, it implies a range of interaction of one less than the true process. We can also compare the word probabilities for each process. This is shown in the last two columns of Table 7.3. The next to last column is the $r = 3$ approximation and the last column has the word probabilities generated by the 3C/6H process. The $r = 3$ approximation does identify several of the forbidden words from the 3C/6H process, such as '0101' and '1010', but it does miss quite a few others, such as '0010', '1001', and '1101', to name a few. It does, however, assign relatively small probabilities to

Table 7.3: Machine reconstruction results for the 3C/6H process. Comparing the last two columns, we see that the agreement is not too bad, but there is still a troublingly large discrepancy between the reconstructed word probabilities (WPs) and the those of the exact process.

r values	Words	Word Probabilities	r values	Words	Reconstructed WPs	Exact WPs
$r = 0$	$p(0)$	0.498	$r = 3$	$p(0000)$	0.296	0.227
	$p(1)$	0.502		$p(0001)$	0.026	0.091
				$p(0010)$	0.030	0.000
$r = 1$	$p(00)$	0.407		$p(0011)$	0.046	0.091
	$p(01)$	0.091		$p(0100)$	0.026	0.000
	$p(10)$	0.091		$p(0101)$	0.000	0.000
	$p(11)$	0.410		$p(0110)$	0.045	0.000
				$p(0111)$	0.025	0.091
$r = 2$	$p(000)$	0.316		$p(1000)$	0.027	0.091
	$p(001)$	0.091		$p(1001)$	0.049	0.000
	$p(010)$	0.000		$p(1010)$	0.000	0.000
	$p(011)$	0.091		$p(1011)$	0.027	0.000
	$p(100)$	0.091		$p(1100)$	0.052	0.091
	$p(101)$	0.000		$p(1101)$	0.029	0.000
	$p(110)$	0.091		$p(1110)$	0.024	0.091
	$p(111)$	0.319		$p(1111)$	0.300	0.227

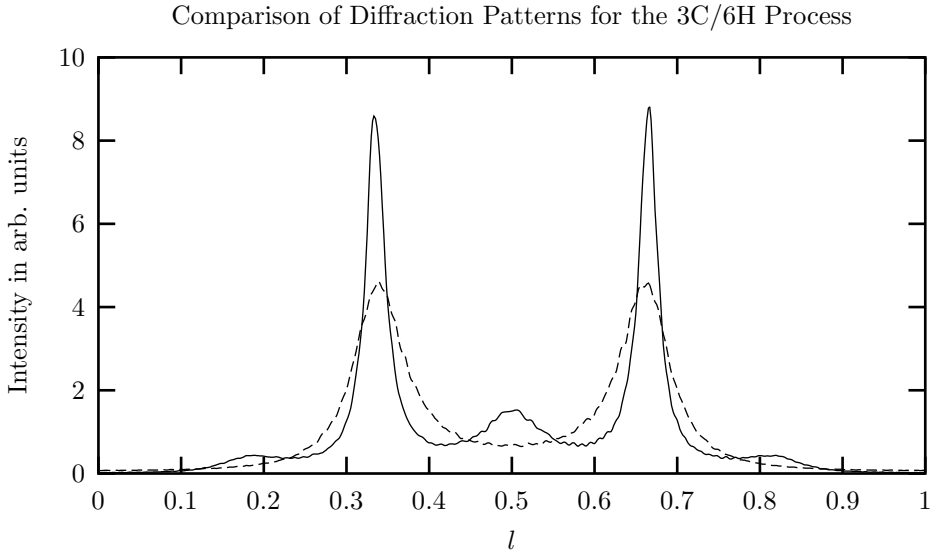

Figure 7.7: The diffraction pattern for a lattice stacked according to the 3C/6H process (solid line) and the $r = 1$ approximation (dashed line). We see that the rise at $l = \frac{1}{2}$ in the diffraction pattern for the 3C/6H process is completely absent in the $r = 1$ approximation.

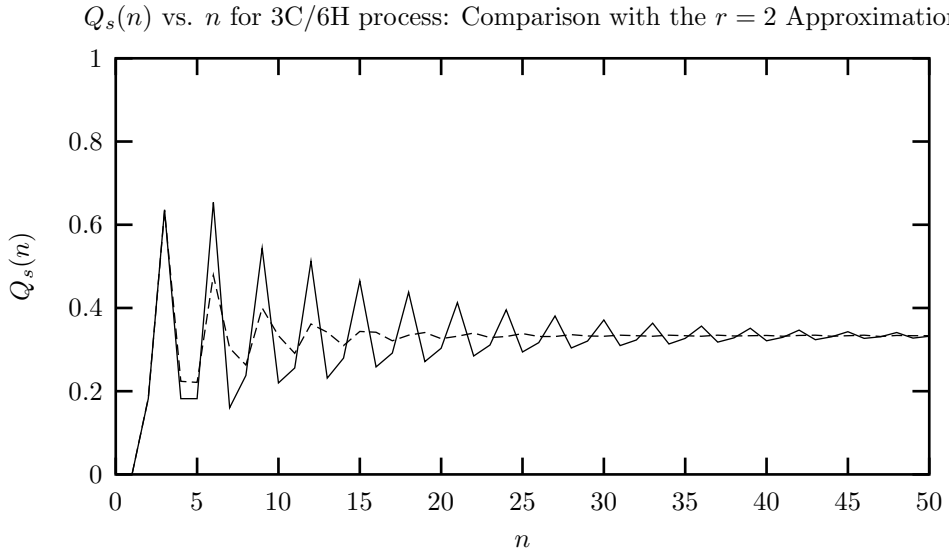

Figure 7.8: $Q_s(n)$ vs. n for the 3C/6H process (solid line) and the $r = 2$ approximation to the process (dashed line). As with the correlation functions for the $r = 1$ approximation, those of the $r = 2$ approximation fall of far too quickly as compared with actual process.

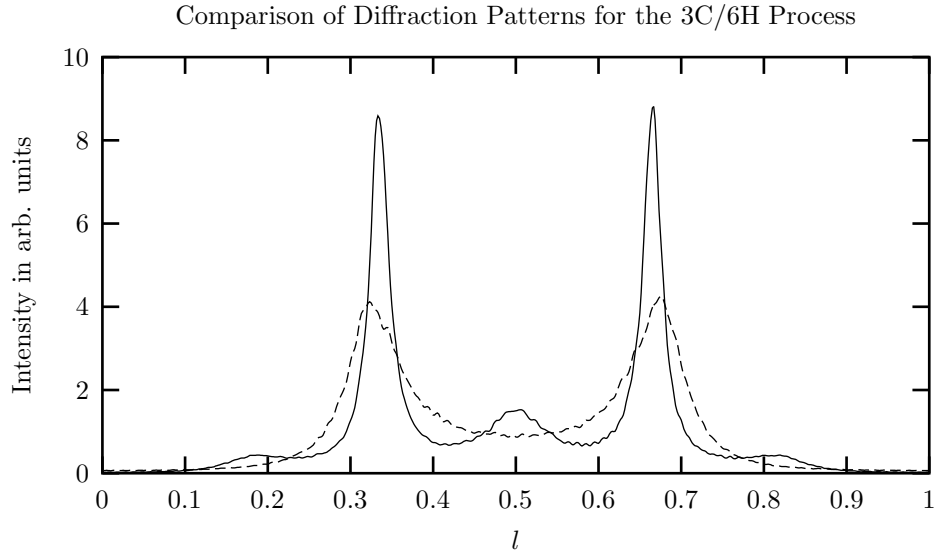

Figure 7.9: The diffraction pattern for a lattice stacked according to the 3C/6H process (solid line) and the $r = 2$ approximation (dashed line). There seems to be little improvement compared with the $r = 1$ approximation.

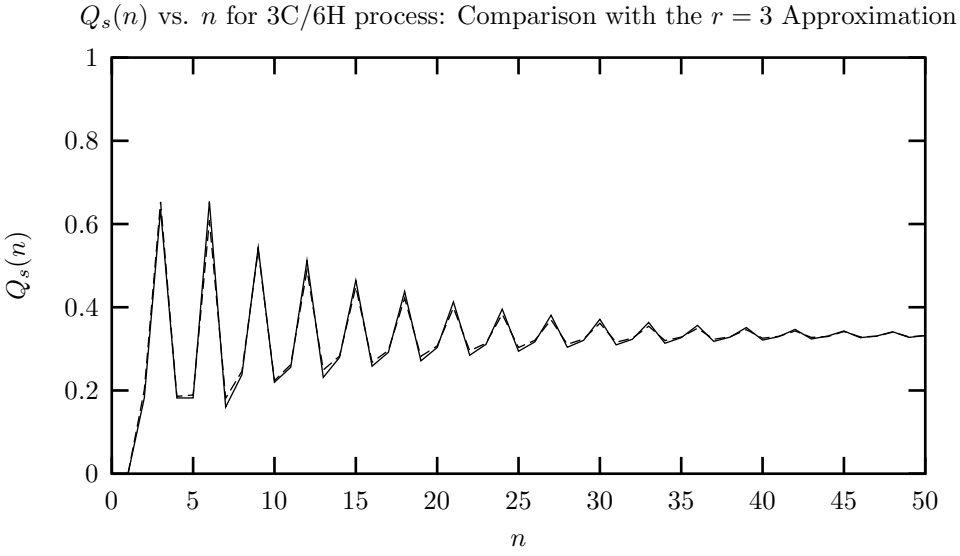

Figure 7.10: $Q_s(n)$ vs. n for the 3C/6H process (solid line) and the $r = 3$ approximation to the process (dashed line). Here we see quite good agreement between the correlation functions for $r = 3$ approximation to the process and those of the actual process.

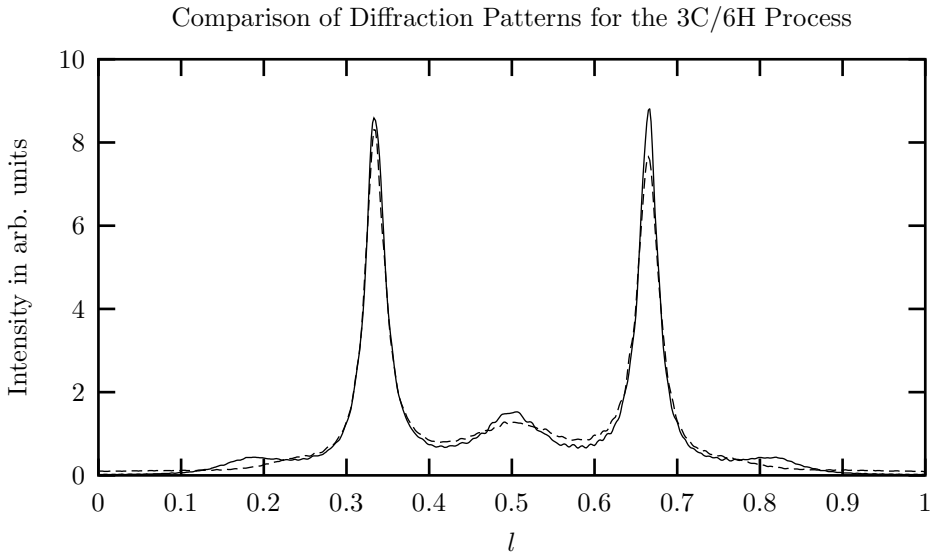

Figure 7.11: The diffraction pattern for a lattice stacked according to the 3C/6H process (solid line) and the $r = 3$ approximation (dashed line). As with the $r = 3$ correlation functions, the agreement between the $r = 3$ approximation to the diffraction pattern and the diffraction pattern of the actual process is quite good, except for the small rises at $l = \frac{1}{6}$ and $\frac{5}{6}$. We interpret this as resulting from the fact that at $r = 3$ there is not a sufficient structure in the graph to simultaneously support both 3C and 6H simple cycles.

Table 7.4: Computational results for 3C/6H process and the $r = 0, 1, 2$ and 3 approximations. We notice that the entropy density decreases with increasing r. This seems reasonable, since as r increases, we see more of the past and should therefore be less surprised at each new symbol.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
3C/6H process	SFT	4	0.510	0.490	2.86	0.819	1.342
$r = 0$ Approximation	SFT	0	1.000	0.000	0.00	0.000	0.000
$r = 1$ Approximation	SFT	1	0.685	0.315	1.00	0.315	0.315
$r = 2$ Approximation	SFT	2	0.626	0.374	1.69	0.433	0.492
$r = 3$ Approximation	SFT	3	0.537	0.463	2.44	0.831	1.380

these words. The $r = 3$ approximation also over estimates the word probabilities associated with the ferromagnetic arcs, '1111' and '0000'. It assigns these a value of about 0.30, where as the underlying process gives them a value of about 0.23.

Examining Table 7.4, we see how the computational values change with each approximation and how the compare with those of the true process. Notably, the entropy density falls with increasing r. This is reasonable, as remembering more of the past should decrease our 'surprise' on seeing each new symbol, so h_μ should decrease and finally approach that of the underlying process. We also see that the statistical complexity increases and approaches that of the underlying process as we increase r. The excess entropy and the transient information actually slightly *overestimate* the real values at $r = 3$.

7.3 Machine Reconstruction for Noisy Period Two Process

Let us now attempt machine reconstruction for a process that is simple to describe (in the sense that it requires only two recurrent causal states), but nonetheless is not depict-able on a portion of a de Bruijn graph. The recurrent portion of the ϵ-machine for the noisy period two process is shown in Figure 5.37. The machine reconstruction results for the $r = 0, 1, 2$ and 3 are given in Table 7.5.

$Q_s(n)$ versus n for both the $r = 0$ approximation and the noisy period two are shown in the Figure 7.12. Both approach their asymptotic values of relatively quickly, but there is some noticeable disagreement for $n \leq 10$. The diffraction pattern for this approximation is compared with the diffraction pattern for the noisy period two in Figure 7.13. The agreement is fair, with the broad diffuse increase in intensity at $l \approx 0.4$ being somewhat modeled by the $r = 0$ approximation, although shifted. The curious zero $l = \frac{5}{6}$ in the noisy period two spectrum is not captured by the $r = 0$ approximation.

At $r = 1$, we pick up our first forbidden word, '00'. The reconstructed $r = 1$ machine does recognize this. $Q_s(n)$ versus n comparing the $r = 1$ approximation and the noisy period two is shown in Figure 7.14. The $r = 1$ approximation seems to over-estimate the strength of the correlations for the smaller n, especially in the region of $5 \leq n \leq 15$. The diffraction pattern for these two is shown in Figure 7.15. The placement of the broad diffuse peak at $l \approx 0.4$ seems fine, but the $r = 1$ approximation over-estimates its sharpness. The zero in the spectrum at $l = \frac{5}{6}$ is again not well-represented.

$Q_s(n)$ versus n for the $r = 2$ approximation to the process and the noisy period two is shown in Figure 7.16. We see that now the correlation functions are *under-represented* by the approximation. The diffraction patterns for the $r = 2$ approximation and the noisy period two are shown in Figure 7.17. The broad diffuse central maximum is also now under-represented. The zero at $l = \frac{5}{6}$ is more closely approximated than before.

The machine reconstruction results for the $r = 3$ approximation are shown in the next to last

Table 7.5: Machine reconstruction results for the noisy period two system. Comparing the last two columns we see that the same difficulty that arose in the 3C/6H/2H process is present here, namely that the recurrent portion of the ϵ-machine is not strongly connected since $p(0000) \neq 0$ even though $p(0001) = p(1000) = 0$. The problem is less severe here due to the smallness of the weight attached to $p(0000)$. The other word probabilities are reasonably well-represented by the $r = 3$ approximation.

r values	Words	Word Probabilities	r values	Words	Reconstructed WPs	Exact WPs
$r = 0$	$p(0)$	0.250	$r = 3$	$p(0000)$	0.007	0.000
	$p(1)$	0.750		$p(0001)$	0.000	0.000
				$p(0010)$	0.000	0.000
$r = 1$	$p(00)$	0.000		$p(0011)$	0.000	0.000
	$p(01)$	0.250		$p(0100)$	0.005	0.000
	$p(10)$	0.250		$p(0101)$	0.116	0.125
	$p(11)$	0.500		$p(0110)$	0.010	0.000
				$p(0111)$	0.115	0.125
$r = 2$	$p(000)$	0.000		$p(1000)$	0.000	0.000
	$p(001)$	0.000		$p(1001)$	0.005	0.000
	$p(010)$	0.125		$p(1010)$	0.122	0.125
	$p(011)$	0.125		$p(1011)$	0.124	0.125
	$p(100)$	0.000		$p(1100)$	0.002	0.000
	$p(101)$	0.250		$p(1101)$	0.132	0.125
	$p(110)$	0.125		$p(1110)$	0.116	0.125
	$p(111)$	0.375		$p(1111)$	0.249	0.250

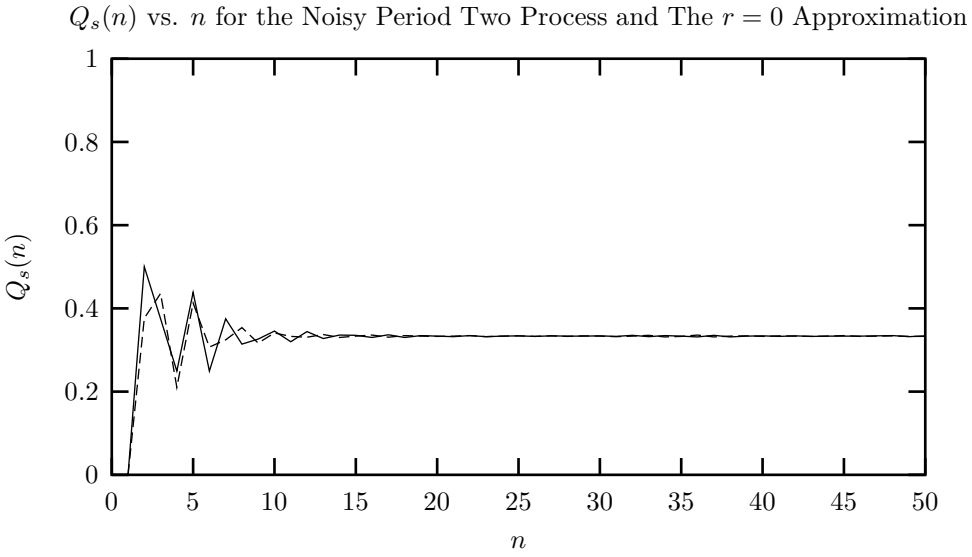

Figure 7.12: $Q_s(n)$ vs. n for the noisy period two process (solid line) and the $r = 0$ approximation to the process (dashed line). We see that both decay quickly to the asymptotic value of $\frac{1}{3}$, but there is some small difference for small n.

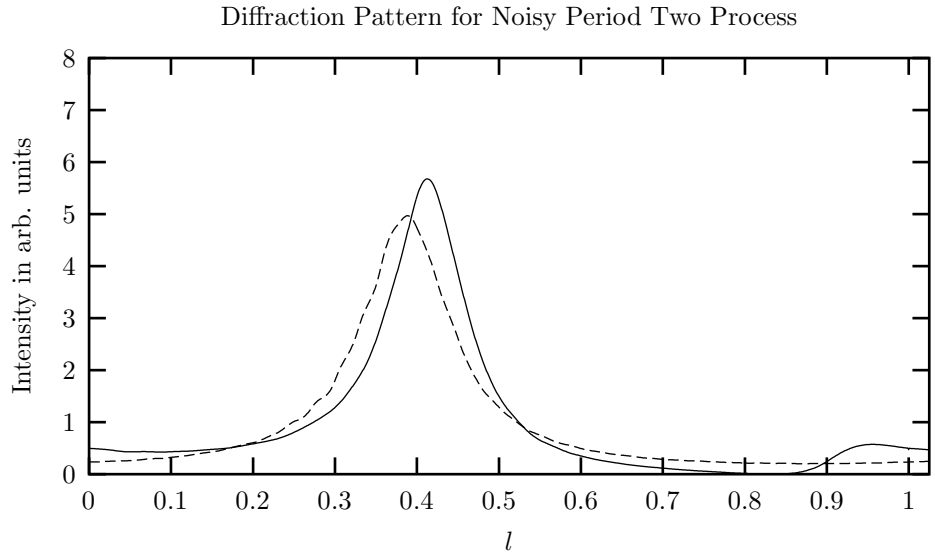

Figure 7.13: The diffraction pattern for a lattice stacked according to the noisy period two process (solid line) and the $r = 0$ approximation (dashed line).

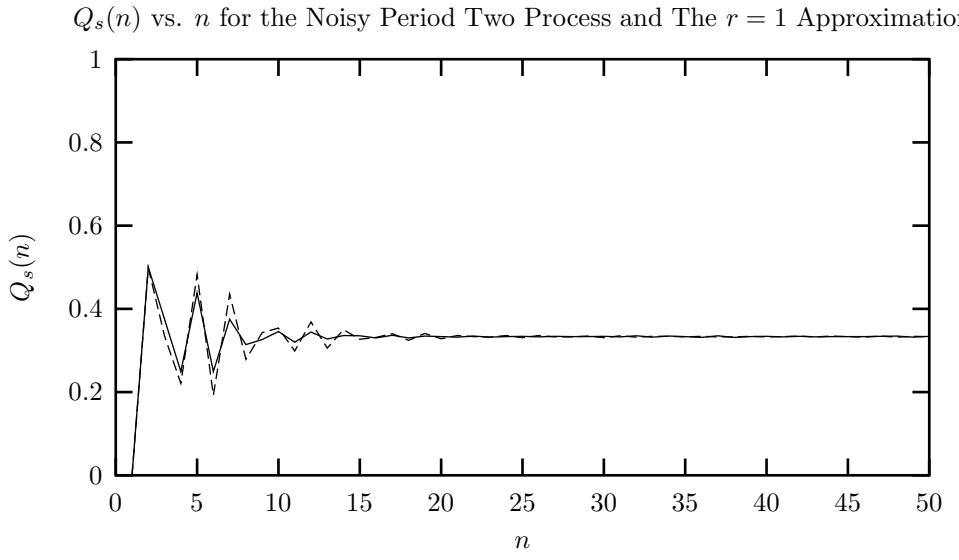

Figure 7.14: $Q_s(n)$ vs. n for the noisy period two process (solid line) and the $r = 1$ approximation to the process (dashed line).

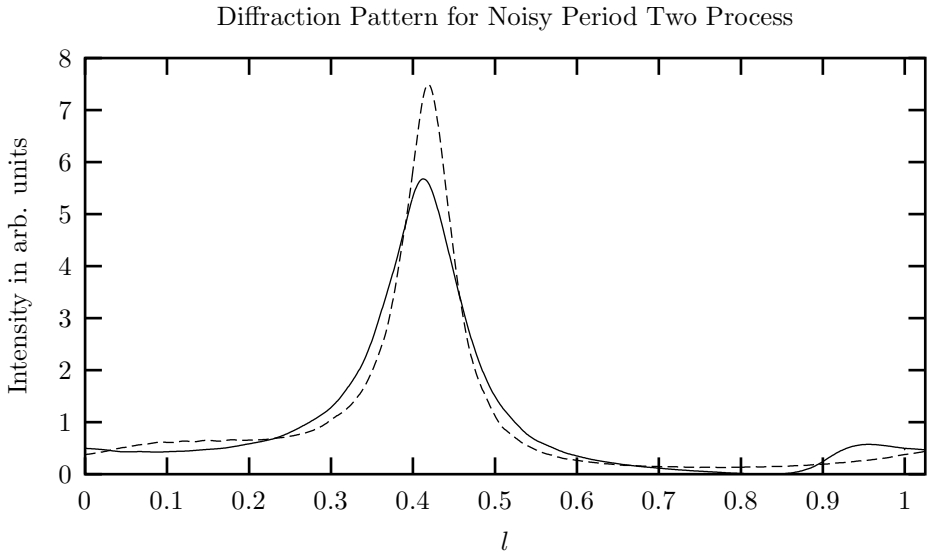

Figure 7.15: The diffraction pattern for a lattice stacked according to the noisy period two process (solid line) and the $r = 1$ approximation (dashed line).

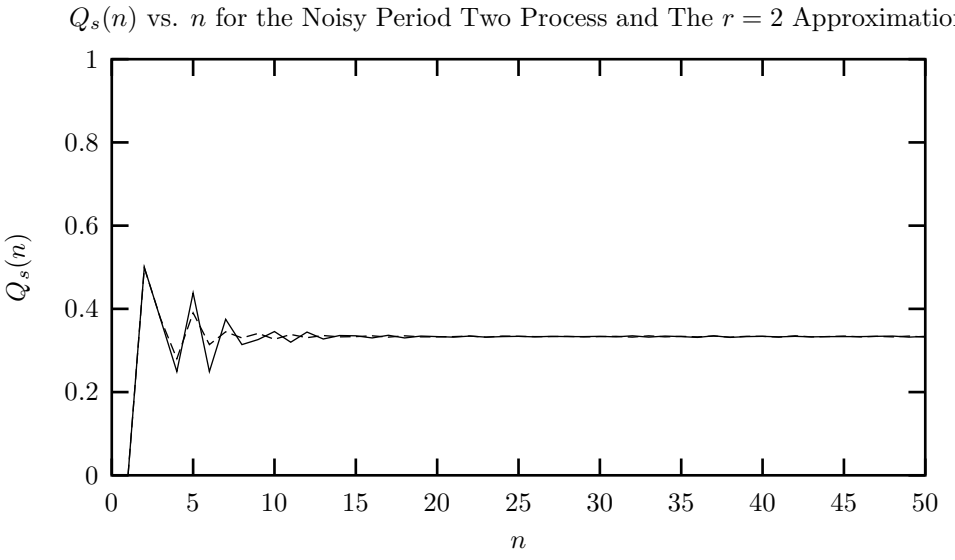

Figure 7.16: $Q_s(n)$ vs. n for the noisy period two process (solid line) and the $r = 2$ approximation to the process (dashed line).

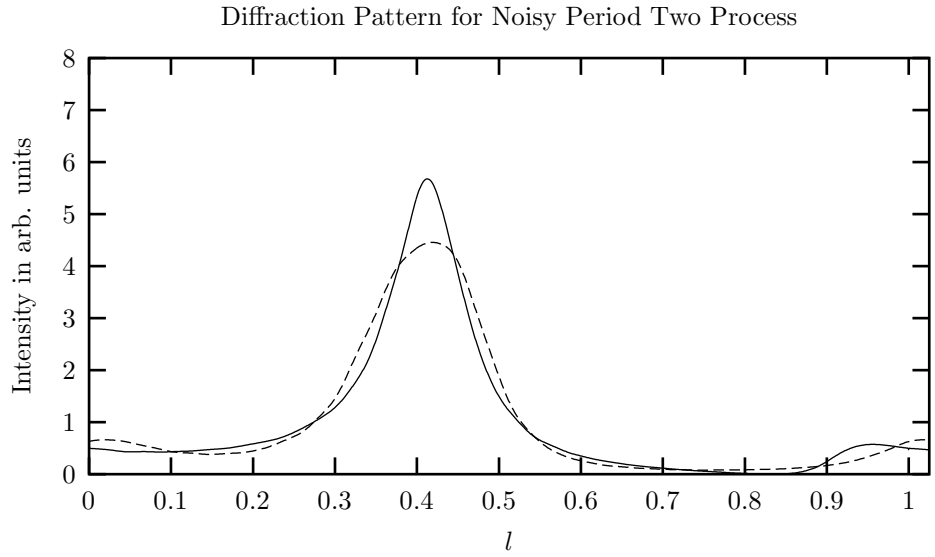

Figure 7.17: The diffraction pattern for a lattice stacked according to the noisy period two process (solid line) and the $r = 2$ approximation (dashed line).

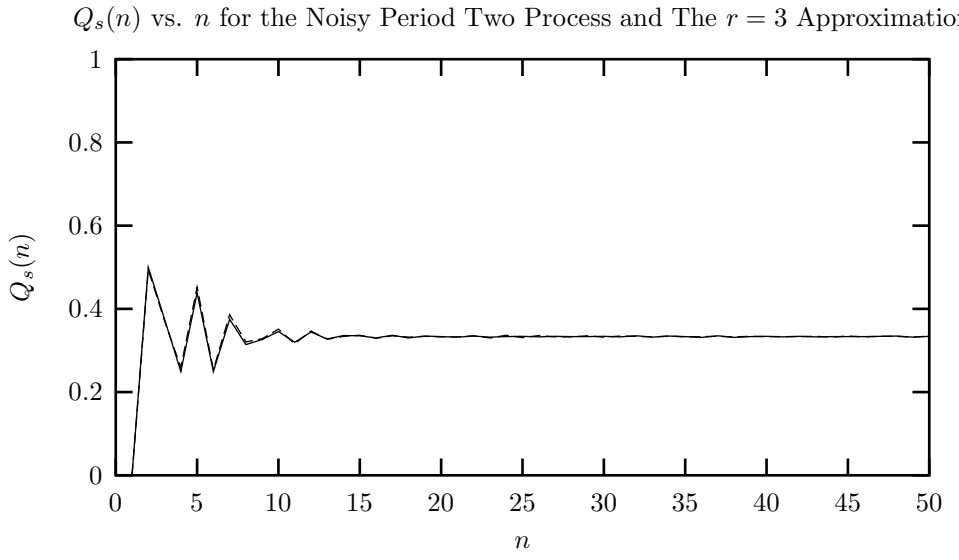

Figure 7.18: $Q_s(n)$ vs. n for the noisy period two process (solid line) and the $r = 3$ approximation to the process (dashed line).

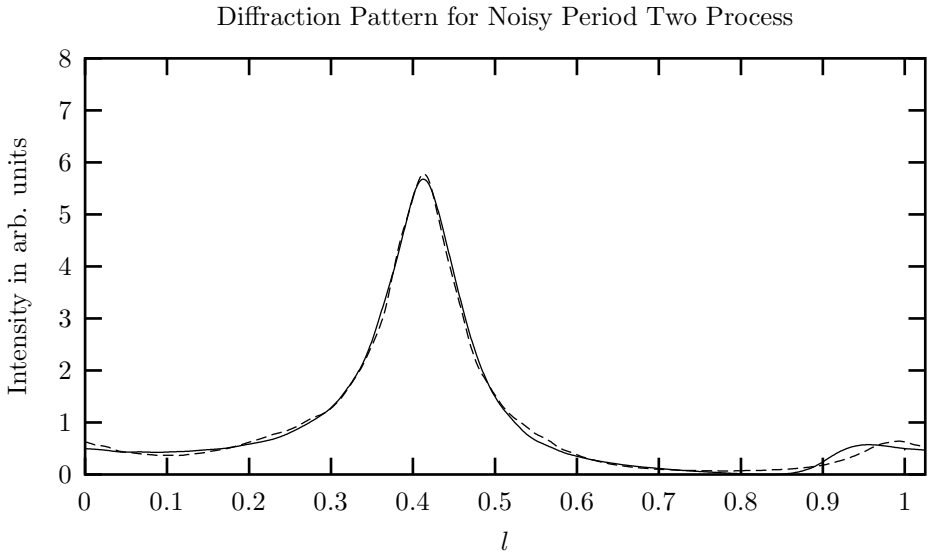

Figure 7.19: The diffraction pattern for a lattice stacked according to the noisy period two process (solid line) and the $r = 3$ approximation (dashed line).

column of Table 7.5, along with the word probabilities for the noisy period two in the last column. There are several curiosities. We found that the lower approximations were detecting the forbidden word '00' and words constructed from it. At $r = 3$, some of these words have become 'unforbidden.' For example, '0100', containing '00' as a subword is forbidden by the noisy period two process, but $r = 3$ approximation assigns it a small probability weight of 0.005. We do acquire a new irreducible forbidden word at $r = 3$, '0110', but the $r = 3$ approximation gives this a small probability weight of 0.010. Most troubling is the isolated node '0000'. While assigning only a small weight of 0.007, it is nonetheless isolated from the rest of the graph as $p(0001) = p(1000) = 0.000$. The same difficulty that plagued us in the 3C/2H/6H process arises here. Given the small weight, though, we chose to ignore it and proceed as if $p(0000) = 0.000$. A plot of $Q_s(n)$ versus n for both the $r = 3$ approximation and the noisy period two is given in Figure 7.18. We see excellent agreement. There is similar excellent agreement between the diffraction patterns as shown in Figure 7.19. The broad diffuse maximum is well-represented by the $r = 3$ approximation, but the minimum at $l = \frac{5}{6}$ is slightly off. There is a rise in the spectrum of the noisy period two that is not accounted for well in the $r = 3$ approximation. From the example in §7.2, we see that the details of the diffuse background scattering are important, and the point is reinforced here.

Table 7.6 gives the computational results for the noisy period two and the four r approximations. As we might expect, the entropy density decreases with increasing r, except for the $r = 2 \rightarrow 3$ step. There is a slight rise in h_μ. This is counter-intuitive, and it is not known why h_μ increases.

7.4 Machine Reconstruction for the Even Process

As a final example, let us consider another process that we will not be able to find using a finite r procedure. Instead, we expect to once again generate graphs of increasing size that approximate the process. The recurrent portion of the ϵ-machine for this process is given in Figure 5.41. The machine reconstruction results for the first four r approximations are shown in Table 7.7.

Figure 7.20 shows the comparison between the $r = 0$ approximation and the even system for the

Table 7.6: Computational results for The Noisy Period Two System and the $r = 0, 1, 2$ and 3 approximations. Unlike the 3C/6H process, h_μ does not steadily decrease as r increases. There is a slight rise in r as r goes from 2 to 3. This is counterintuitive, and it is not known why this is so.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
Noisy Period Two	SS	∞	0.500	0.500	1.000	1.000	3.31
$r = 0$ Approximation	SFT	0	0.811	0.189	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.689	0.311	0.811	0.123	0.123
$r = 2$ Approximation	SFT	2	0.655	0.345	1.50	0.189	0.222
$r = 3$ Approximation	SFT	3	0.671	0.329	2.25	0.242	0.370

Table 7.7: Machine reconstruction results for the even system. Comparing the last two columns, we see reasonable agreement between the word probabilities for the even system and the $r = 3$ approximation.

r values	Words	Word Probabilities	r values	Words	Reconstructed WPs	Exact WPs
$r = 0$	$p(0)$	0.333	$r = 3$	$p(0000)$	0.035	0.042
	$p(1)$	0.667		$p(0001)$	0.045	0.042
				$p(0010)$	0.007	0.000
$r = 1$	$p(00)$	0.167		$p(0011)$	0.078	0.083
	$p(01)$	0.167		$p(0100)$	0.002	0.000
	$p(10)$	0.167		$p(0101)$	0.000	0.000
	$p(11)$	0.500		$p(0110)$	0.073	0.083
				$p(0111)$	0.089	0.083
$r = 2$	$p(000)$	0.083		$p(1000)$	0.044	0.042
	$p(001)$	0.083		$p(1001)$	0.040	0.042
	$p(010)$	0.000		$p(1010)$	0.000	0.000
	$p(011)$	0.167		$p(1011)$	0.084	0.083
	$p(100)$	0.083		$p(1100)$	0.083	0.083
	$p(101)$	0.083		$p(1101)$	0.087	0.083
	$p(110)$	0.167		$p(1110)$	0.089	0.083
	$p(111)$	0.333		$p(1111)$	0.240	0.250

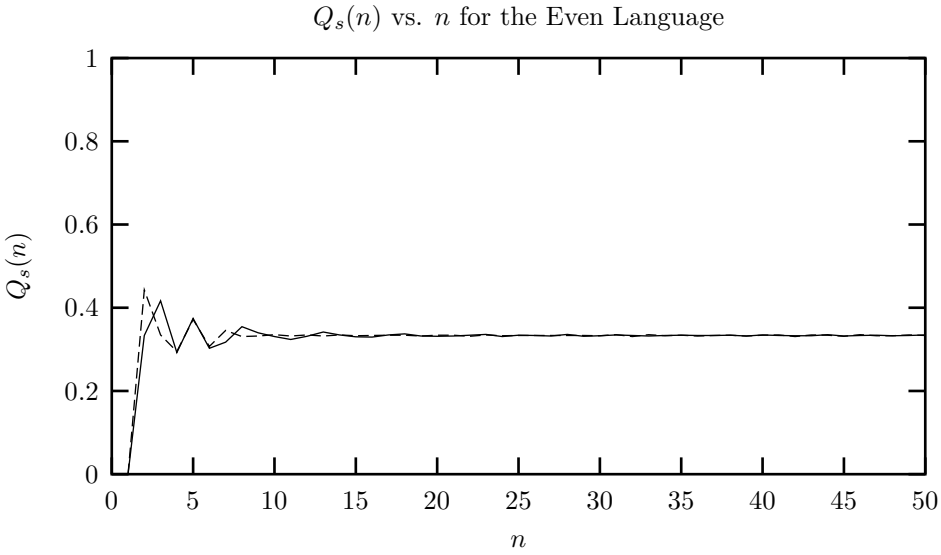

Figure 7.20: $Q_s(n)$ vs. n for the even process (solid line) and the $r = 0$ approximation to the process (dashed line).

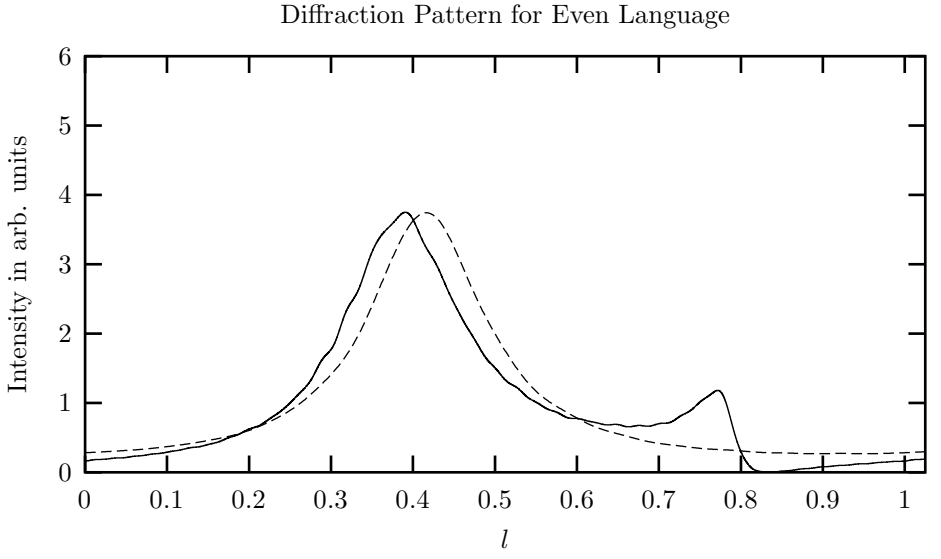

Figure 7.21: The diffraction pattern for a lattice stacked according to the even process (solid line) and the $r = 0$ approximation (dashed line).

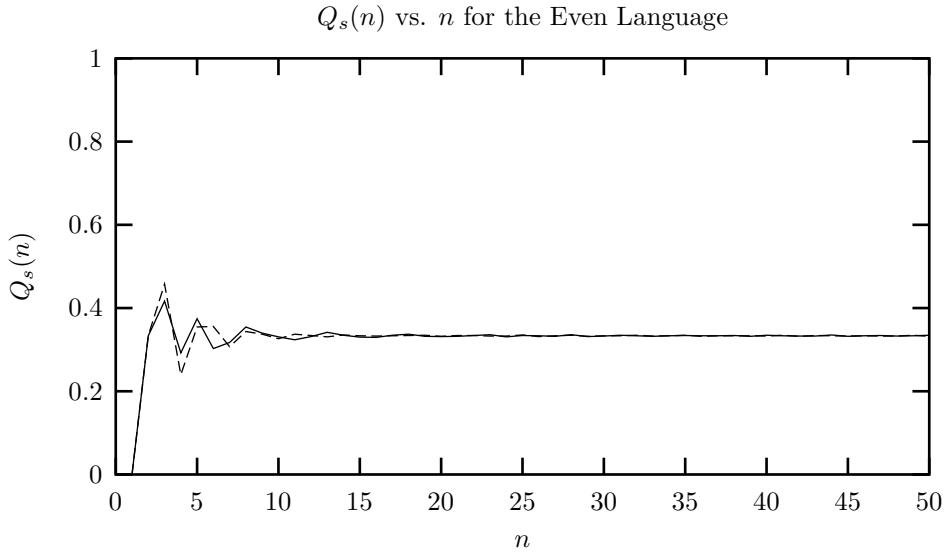

Figure 7.22: $Q_s(n)$ vs. n for the even process (solid line) and the $r = 1$ approximation to the process (dashed line).

first fifty correlation functions $Q_s(n)$. Both quickly approach their asymptotic value, but there is some disagreement at small n. The diffraction patterns are compared in Figure 7.21. The spectrum for the even system has two broad diffuse peaks, one of which the $r = 0$ approximation is seems to be trying to fit. It is slightly shifted though. The second peak at $l \approx 0.78$ is not represented at all in the $r = 0$ approximation.

A plot of $Q_s(n)$ versus n for the for the $r = 1$ approximation and the even process is shown in Figure 7.22. Again, the agreement between the correlation functions for small n is not as good as one would hope, but it is not bad. A comparison of the diffraction patterns for the $r = 1$ approximation and the even process is shown in Figure 7.23. Again the broad background peak at $l \approx 0.4$ is slightly shifted in the $r = 1$ approximation, but this time to the left. The secondary peak at $l \approx 0.78$ is again not well-represented in this approximation.

At $r = 2$, we find the first word forbidden by the even system, '010'. A comparison between the first fifty values of the correlation functions for the even system and the $r = 2$ approximation are given in Figure 7.24. We see reasonable agreement, but there are still discrepancies for $4 \leq n \leq 15$. A comparison of the diffraction patterns is given in Figure 7.25, where we see improved reproduction of the broad rise at $l \approx 0.4$, but the peak at $l \approx 0.78$ is shifted to the left.

We lastly consider the $r = 3$ reconstruction to the even process. The word probabilities for the reconstructed process are shown in the next to last column of Table 7.7 and the length four word probabilities for the even process are in the last column of Table 7.7. We see the same phenomenon of words becoming 'unforbidden' at larger r, namely '0010' and '0100' have small probabilities even though they contain '010', a forbidden sequence, as a subword. Unlike the previous two examples, the recurrent portion of the graph is strongly connected. We see the $r = 3$ approximation does a quite reasonable job of reproducing the length four word probabilities. Comparing the $Q_s(n)$ generated by the even process and the $r = 3$ approximation in Figure 7.26, we see good agreement, except for some small discrepancy in the region $5 \leq n \leq 10$. The diffraction pattern generated by the $r = 3$ approximation also shows good agreement in region $0 \leq l \leq 0.6$. The rise at $l \approx 0.78$ and the subsequent zero in the diffracted intensity at $l = 0.83$ are not well modeled here. Computational

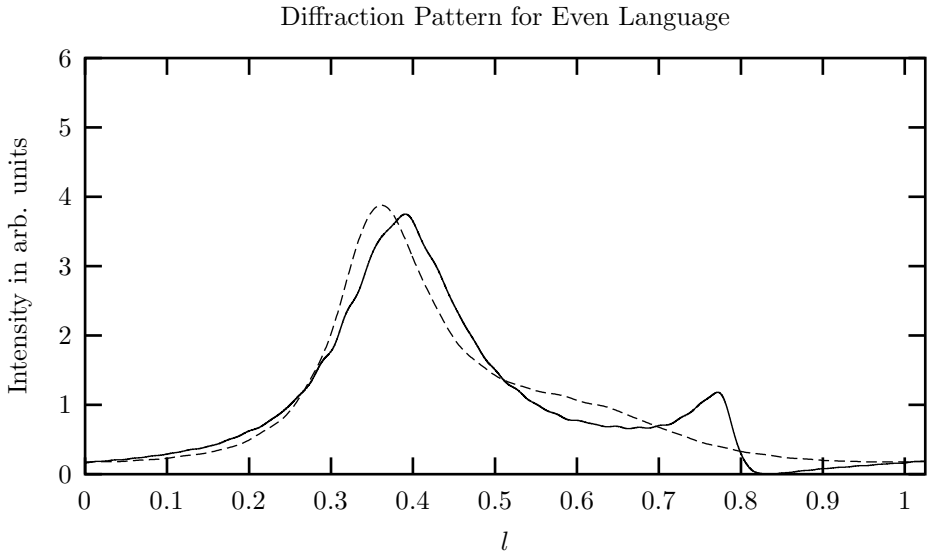

Figure 7.23: The diffraction pattern for a lattice stacked according to the even process (solid line) and the $r = 1$ approximation (dashed line).

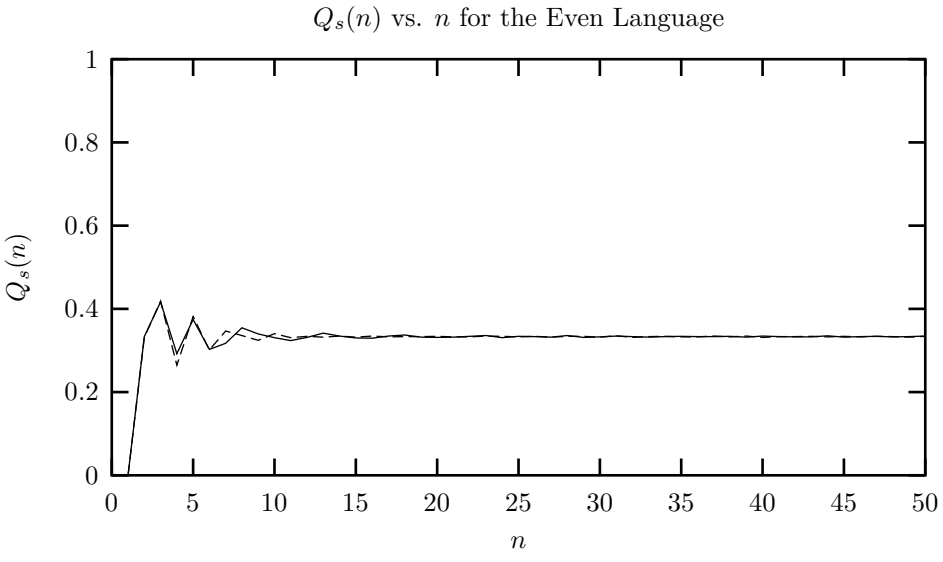

Figure 7.24: $Q_s(n)$ vs. n for the even process (solid line) and the $r = 2$ approximation to the process (dashed line).

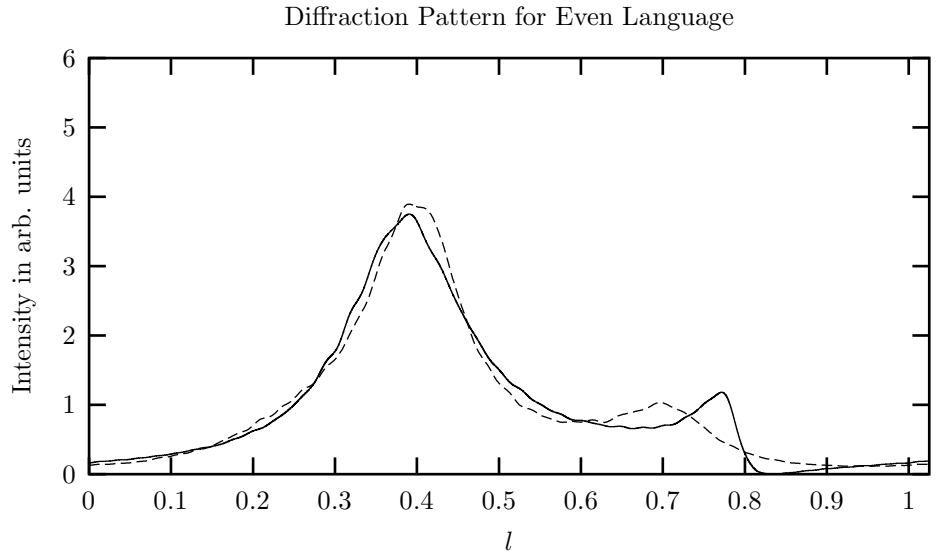

Figure 7.25: The diffraction pattern for a lattice stacked according to the even process (solid line) and the $r = 2$ approximation (dashed line).

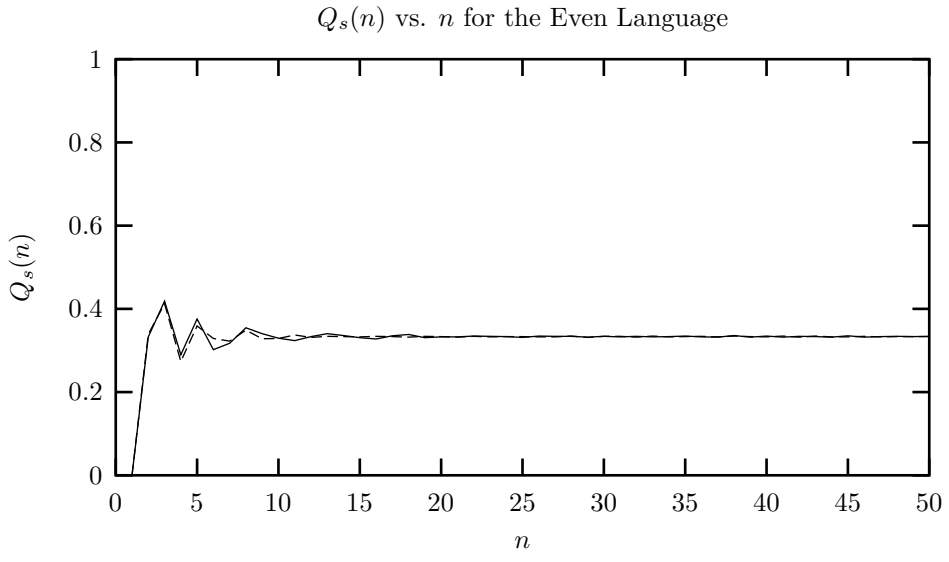

Figure 7.26: $Q_s(n)$ vs. n for the even process (solid line) and the $r = 3$ approximation to the process (dashed line).

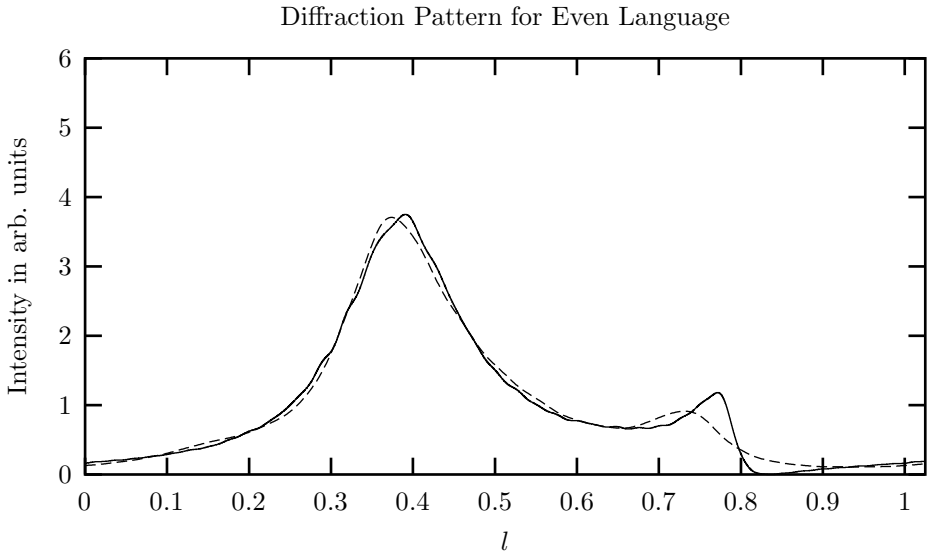

Figure 7.27: The diffraction pattern for a lattice stacked according to the even process (solid line) and the $r = 3$ approximation (dashed line).

Table 7.8: Computational results for the even process and the $r = 0, 1, 2$ and 3 approximations. As with the noisy period two process, there is a slight rise in the entropy density as r goes from 2 to 3.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
Even	SS	∞	0.667	0.333	0.918	0.913	3.09
$r = 0$ Approximation	SFT	0	0.918	0.082	0.0	0.0	0.000
$r = 1$ Approximation	SFT	1	0.874	0.126	0.918	0.044	0.044
$r = 2$ Approximation	SFT	2	0.792	0.208	1.79	0.208	0.290
$r = 3$ Approximation	SFT	3	0.803	0.197	2.63	0.222	0.351

results are given in Table 7.8. We see the same decrease in entropy density as r increases except at $r = 2 \rightarrow 3$, where there is a slight increase in h_μ.

Chapter 8

Previous Classifications of Disorder in Layered Materials

8.1 Overview

A zeroth order attempt to 'classify' disorder in physical systems is to acknowledge that the disorder exists, but give no further details. This is the approach taken in a recent paper on determination of the polytype distribution in SiC [27]. There the authors perform x-ray diffraction experiments on powder samples of SiC. For several samples, they discovered diffuse background scattering which they can not associate with any crystalline polytype. They can, however, determine the fraction of the scattered intensity diffracted into this background, and simply refer to this fraction as the percentage of disorder present. It is certainly an honest approach, akin to the admission of ignorance. We believe much more desirable though, is a statistical description of this disorder. Another approach with a long history is the assumption that there is a crystal structure present but that there are stacking errors, or faults present, which lead to the disorder. Typically, often guide by intuition, and ease of implementation, one assumes a certain number and kind of possible ways for the disorder to exist and calculates what effect this disorder can have on the diffraction pattern. Often this analysis is confined to considering only the effect of the assumed disorder on the Bragg peaks. We saw in §7.2 that this can be misleading. Proposed structures may account relatively well for the placement, intensity, shape, etc. of Bragg peaks but still not represent the underlying mechanism. It is important to take into account the intensity distribution over an entire unit interval. Recent work by Gosk [31] [32] does just this, but he is still adheres to a picture of imposing *a priori* a select number and kind of possible faults. We have several difficulties with these approaches. The first is not specific enough, and the second requires assumptions that in general are not supportable and in fact may be misleading. A more serious objection of the second approach is our contention that the fundamental picture of faulting is deeply flawed. We discuss this in detail in §9.2. A final quibble is perhaps philosophical. Instead of needing to assume some underlying mechanism, we would rather make fewer initial assumptions and let the data more directly tell us about the underlying mechanism.

The zeroth order attempt at describing disorder is simple enough and needs no further exposition. We do wish, however, to examine the faulting picture more closely so that we may compare our results with previous work.

8.2 'Faultology'

Here we discuss various layer faults found in the 2H and 3C close-packed structures. By fault, one means an error or mistake in the stacking sequence that deviates from a crystalline structure. This naturally presupposes the existence of a parent crystal. This terminology is unfortunate because it implies that the structure should have been a crystal (and therefore there has been some sort of "mistake") and perhaps limits our thinking of how one more systematically describe a disordered system. Indeed, one of results of this work is to demonstrate that one-dimensionally disordered lattices can be described more generally by ϵ-machines. So let us briefly list and describe several of the faulting structures proposed in the literature. This list is compiled largely from reference [61].

8.2.1 Growth Faults in the 2H and 3C structures of ZnS

The growth of crystals often proceeds by a layer addition process. Suppose a layer is added which can not be thought of as a continuation of the previous crystal structure, but the layers added subsequent to that layer return to the original stacking rule. Such a layer inserted into the sequence is called a growth fault. For the 2H structure, the rule is that the added layer is the same as the next to last layer. For example, imagine that the unfaulted 2H crystal is $..ABABAB...$ Then a growth fault in this structure would be a C layer following a B layer. The remaining layers would continue to follow the 2H stacking rule, giving a sequence like,

$$...A\ B\ A\ B\ A\ B\ \underline{C}\ B\ C\ B\ C\ B...$$

where underlining indicates the fault plane. Notice that the original crystal is composed of alternating A and B layers, while after the fault this becomes a sequence of alternating B and C layers. In terms of a relative spin sequence, a growth fault for the 2H crystal is just the insertion of a single 0 or 1 into the sequence. For example, ...01010101... becomes ...0101$\underline{1}$0101... upon insertion of a 1. The underlined character is the inserted spin. The smallest de Bruijn graph on which this can be represented is $r = 1$. The $r = 3$ de Bruijn graph showing this fault is shown in Figure 8.1.

In the 3C structure, the stacking rule is that the added layer is different from the previous two layers. There are, of course, two distinct, symmetry related 3C structures, one being the $...ABCABC...$ and the other its spacial inversion $...CBACBA...$. The relative spin sequences for these are (1)* and (0)* respectively. A growth fault for this crystal gives a sequence like,

$$...A\ B\ C\ A\ B\ \underline{C}\ B\ A\ C\ B\ A...$$

where underlining indicates the fault plane. It is conventional to take this as the fault plane because it is the only atomic plane in the sequence that is hexagonally related to its neighbors. In terms of relative spins, the sequence is ...11111|00000..., where the the vertical line indicates the fault plane. The effect of a growth fault in a 3C structure is then to switch from one 3C structure to another, or to flip all of the relative spins after the fault plane. This fault is also known as a twin fault of the 3C structure, because it produces a crystal containing both kinds of 3C sequences. The smallest de Bruijn graph that can represent this fault is the $r = 1$ graph. The $r = 3$ graph is shown in Figure 8.2.

8.2.2 Deformation Faults in the 2H and 3C structure of ZnS

Other faults can occur after a crystal structure has been formed. Caused by external stresses or inhomogeneous temperature distributions within the crystal, deformation faults are the result of one plane in the crystal slipping past another. Only slips that result in a structure where the layers are

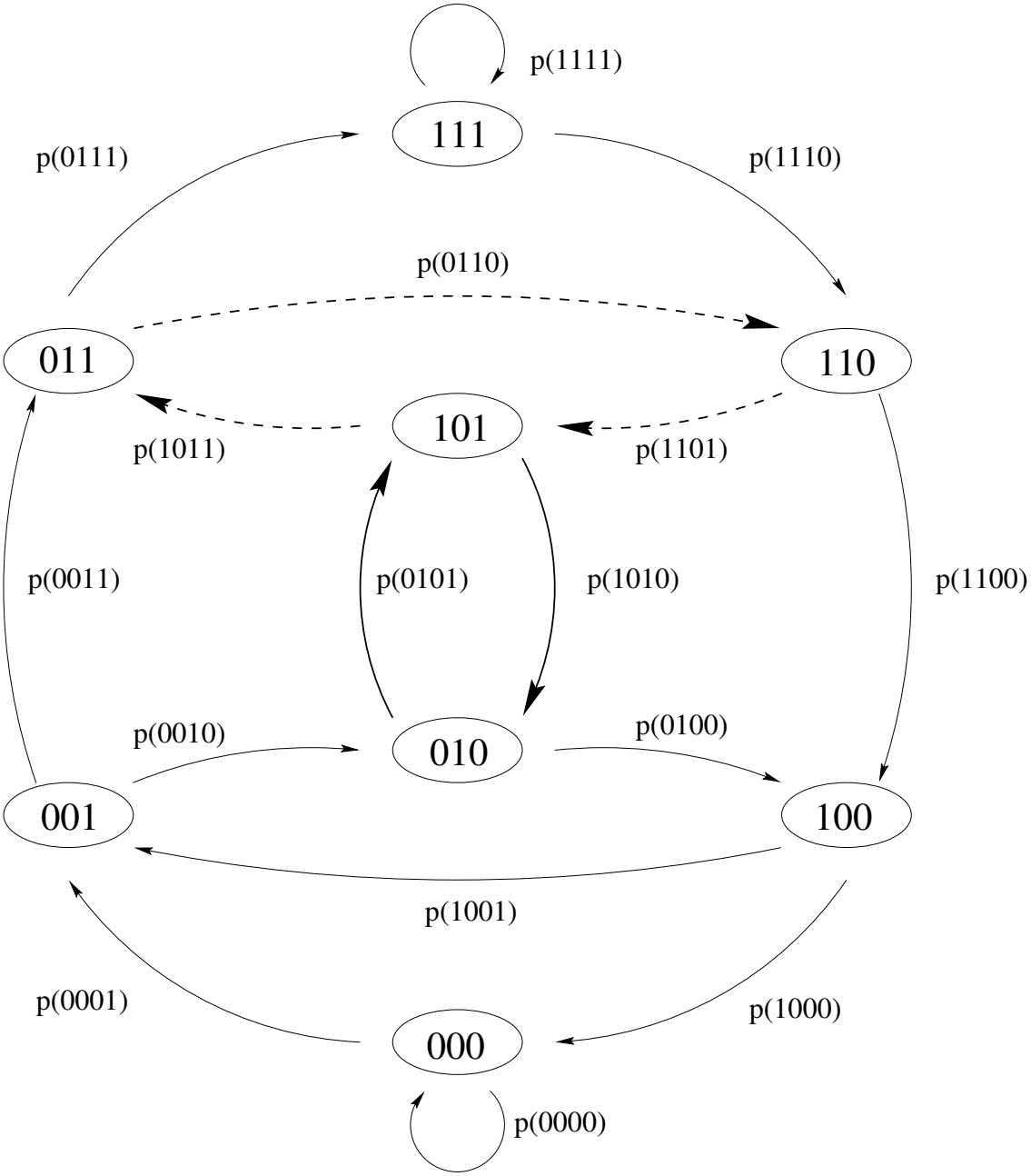

Figure 8.1: Growth faults in the 2H structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 2H structure, the dashed lines are the growth faults, and the thin solid lines are the remainder of the graph. For convenience, only the faults in the upper portion of the graph are shown, corresponding to an insertion of a 1. In general, there are of course the spin inverse of these present (insertion of a 0), and these faults occupy the spin symmetric portion in the lower part of the graph. Growth faults can be seen on a $r = 1$ graph.

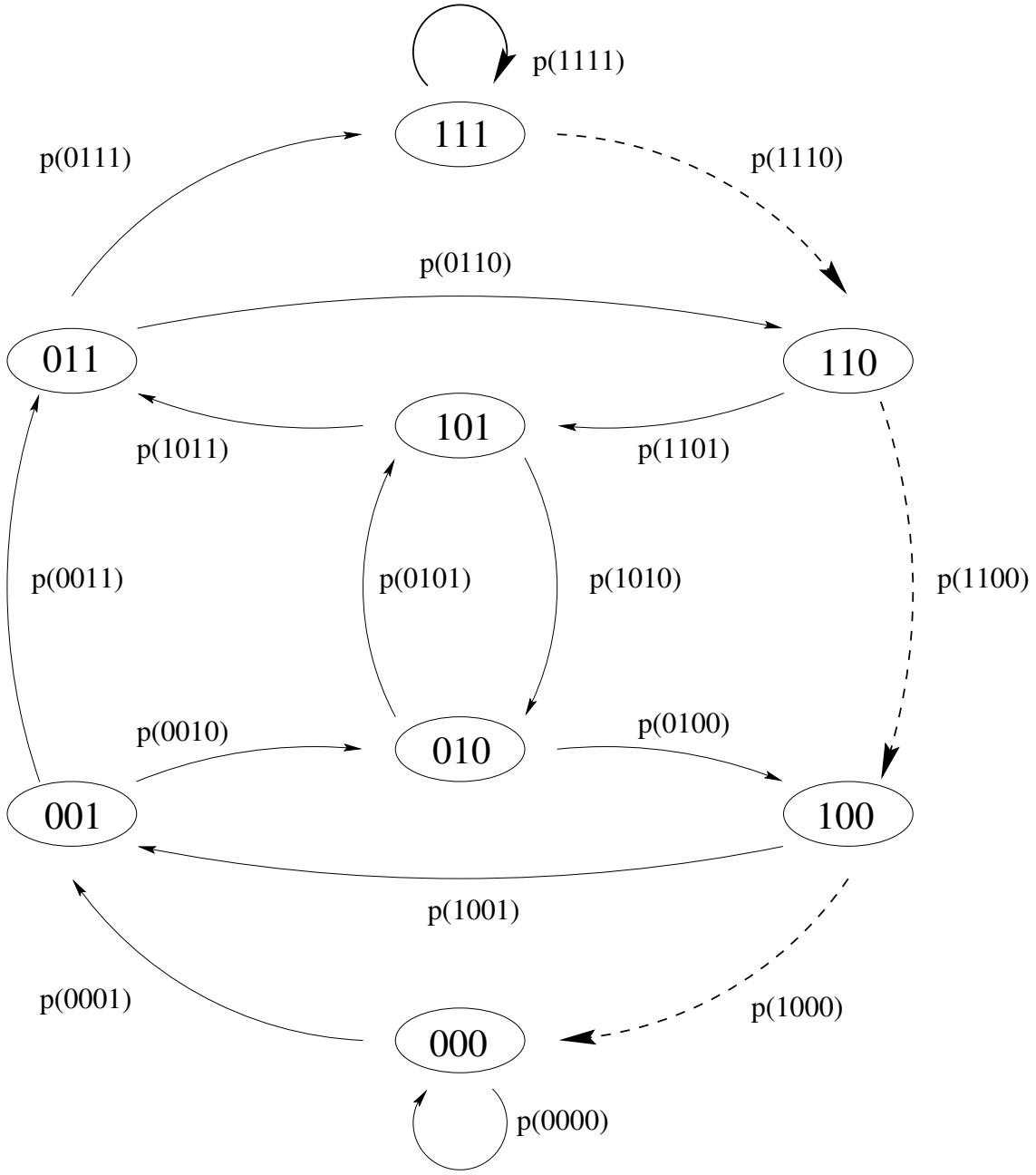

Figure 8.2: Growth faults in the 3C structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 3C structure (only the positive chirality - (1)* shown), the dashed lines are the growth faults, and the thin solid lines are the remainder of the graph.

describable in terms of A, B, or C positions are considered. An example of deformation faulting in the 2H structure is the following:

$$...A\ B\ A\ B\ |\ C\ A\ C\ A\ C\ A...$$

The vertical bar indicates the plane across which the slip occurred. In terms of relative spins, a deformation fault in the 2H structure is realized by flipping a spin. In this example, the unfaulted sequence ...10101010... transforms to ...101$\underline{1}$1010..., where again the underlined character demarcates the flipped spin. The minimum size de Bruijn graph on which this fault can be demonstrated is $r = 3$. This is shown in Figure 8.3.

In the 3C structure, deformation faults appear much the same. An example of a deformation fault in a 3C structure is

$$...A\ B\ C\ A\ B\ C\ |\ B\ C\ A\ B\ C\ A...$$

The vertical bar again indicates the slip plane. Expressed in relative spins, the unfaulted 3C crystal, ...11111111..., becomes ...1111$\underline{0}$111..., giving a single spin flip. This can be expressed on a $r = 1$ de Bruijn graph; it is shown on a $r = 3$ de Bruijn graph in Figure 8.4.

8.2.3 Layer Displacement Faults in the 2H and 3C structure of ZnS

Layer displacement faults are characterized by a shifting of one or two layers in the crystal, while leaving the remainder of the crystal undisturbed. As such, these faults do not interrupt the long range order present in a structure. They are thought to be introduced at high temperatures by diffusion of the atoms through the crystal. Sebastian and Krishna [61] give a nice discussion of the possible mechanisms. In the 2H structure, an example of a layer displacement fault is:

$$...A\ B\ A\ B\ \underline{C}\ B\ A\ B\ A...$$

where the underlined layer is the faulted layer. Written as relative spins, ...10101010... becomes ...101$\underline{1}$0010..., the underlined characters indicating the the relative spins that have flipped. The minimal de Bruijn graph necessary to show this structure is $r = 3$, and this is displayed in Figure 8.5.

Layer displacement faults in 3C structures are more difficult to realize, since each layer is sandwiched between two unlike layers and changing its orientation would violate stacking constraints. It is therefore necessary for two adjacent layers to shift. Consequently one might expect that these are more rare. An example of layer displacement is the following:

$$...A\ B\ C\ A\ B\ C\ \underline{B\ A}\ C\ A\ B\ C\ A...$$

where the underlined layers are faulted. The relative spin sequence changes from a series of all 1s to one where three consecutive spins have been flipped to 0. The minimal graph on which this can be demonstrated is $r = 3$, and this is shown in Figure 8.6.

8.2.4 Additional Faulting Structures

The previous examples by no means exhaust all the possible faulting structures known or postulated to be important in the polytypism of close-packed lattices. Additionally, one finds the double deformation fault for 3C structures, shown in Figure 8.7. Another fault, unnamed and requiring a de Bruijn graph of $r = 4$ is postulated in 2H structures. There are also faults that are believed to correspond to the removal or insertion of entire layers in the crystal. Called extrinsic faults, one mechanism for their creation is irradiation. They have high fault energies, and are thus rare. None of the samples considered in this work have been irradiated, so it is not discussed further here.

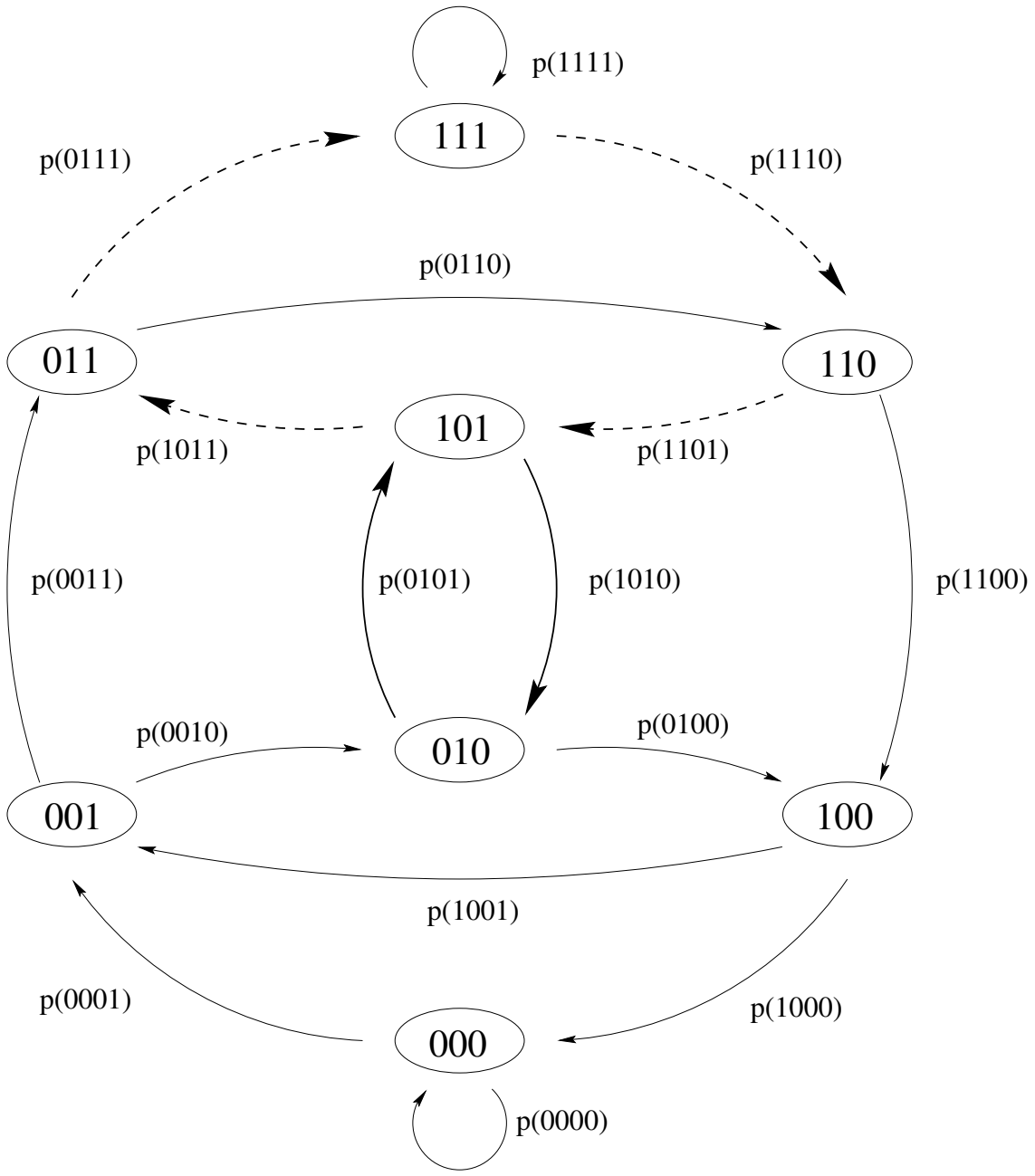

Figure 8.3: Deformation faults in the 2H structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 2H structure, the dashed lines are the deformation faults, and the thin solid lines are the remainder of the graph. For convenience, only the faults in the upper portion of the graph are shown, corresponding to a spin flip of $0 \rightarrow 1$. In general, there are of course the spin inverse of these present ($1 \rightarrow 0$), and these faults occupy the spin symmetric portion in the lower part of the graph.

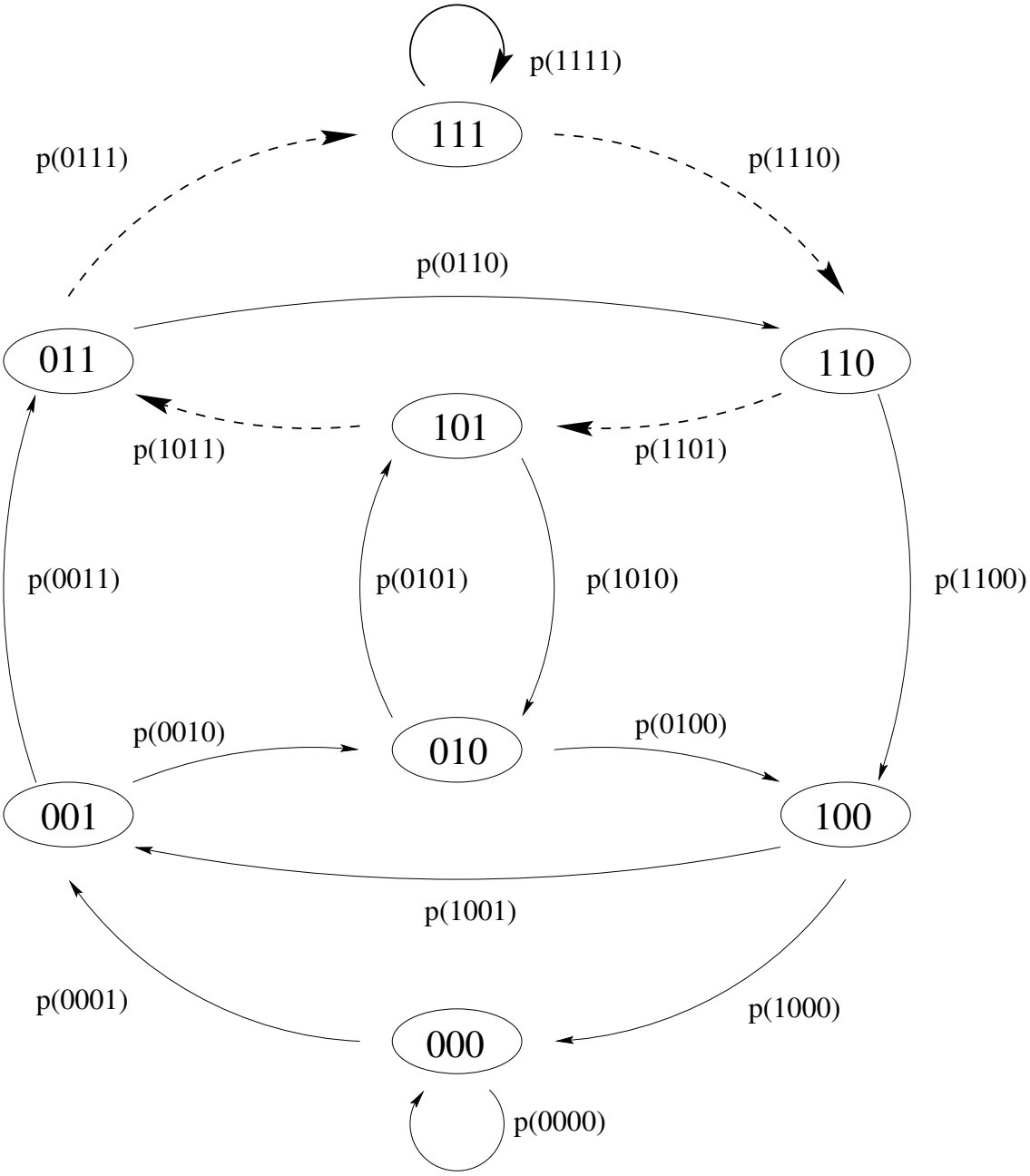

Figure 8.4: Deformation faults in the 3C structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 3C structure (only the positive chirality structure (1)* shown), the dashed lines are the deformation faults, and the thin solid lines are the remainder of the graph.

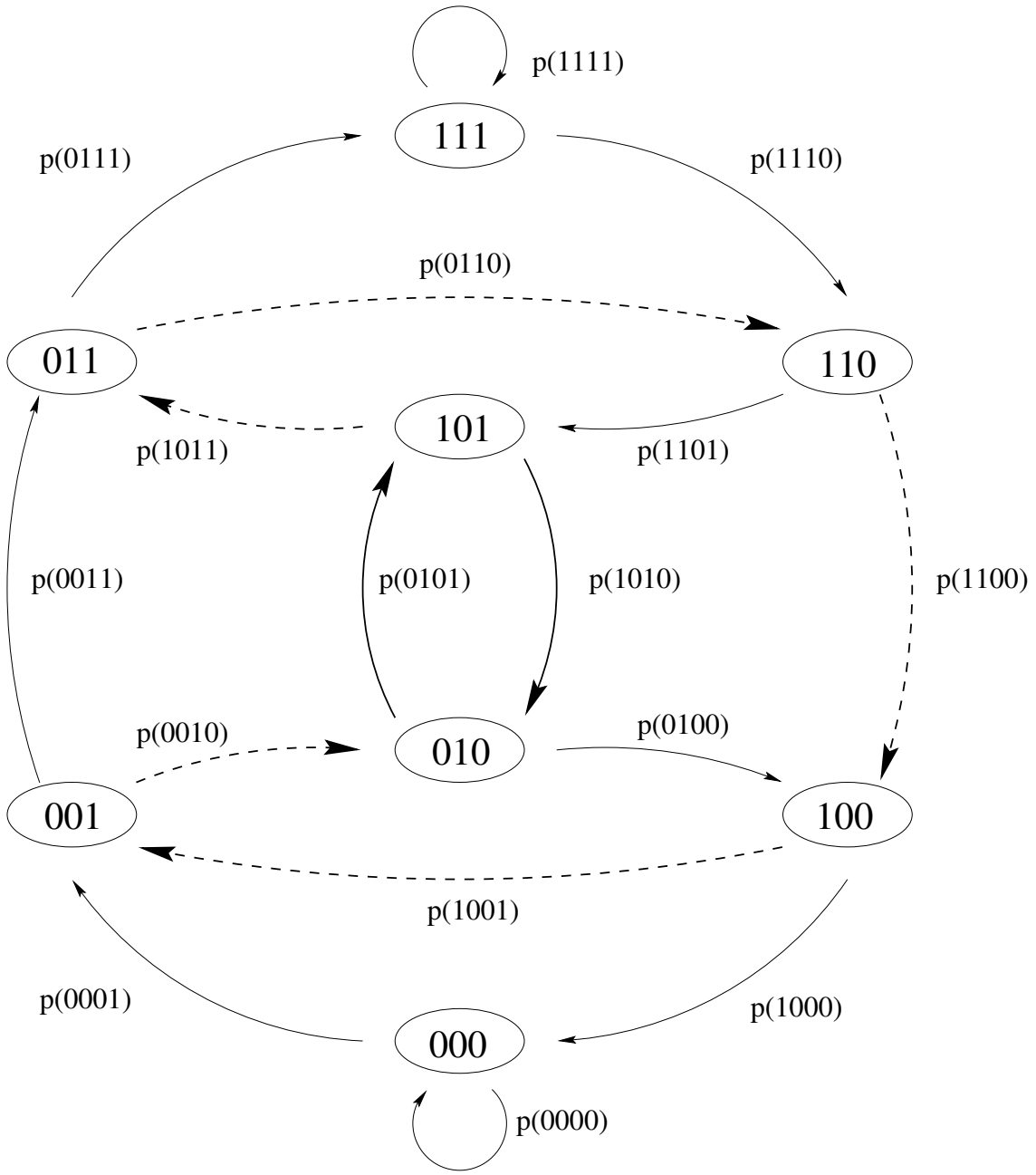

Figure 8.5: Layer displacement faults in the 2H structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 2H structure, the dashed lines are the layer displacement faults, and the thin solid lines are the remainder of the graph. For convenience, only the fault that begins with 1011 is shown.

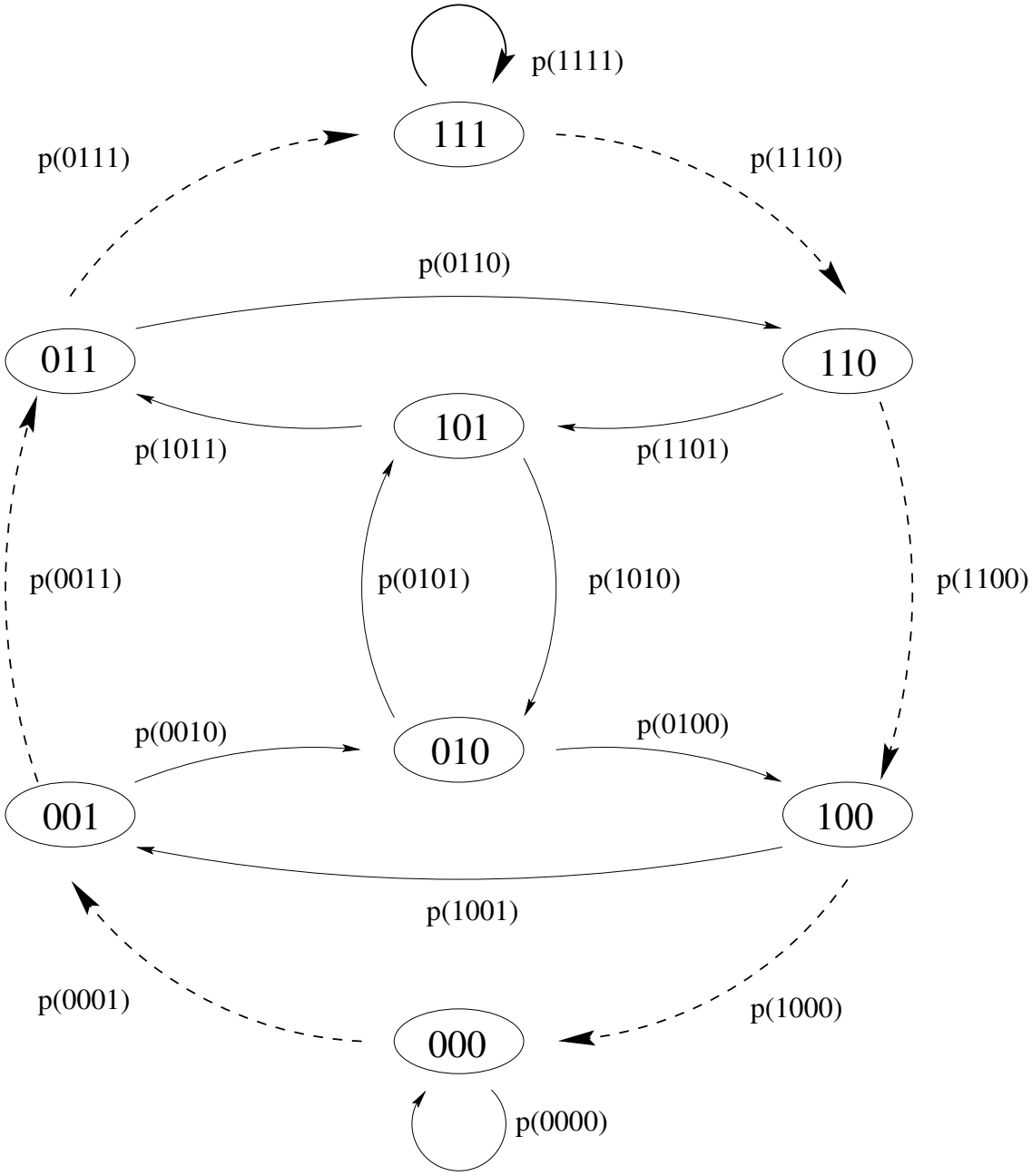

Figure 8.6: Layer displacement faults in the 3C structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 3C structure (only the positive chirality structure (1)* shown), the dashed lines are the layer displacement faults, and the thin solid lines are the remainder of the graph.

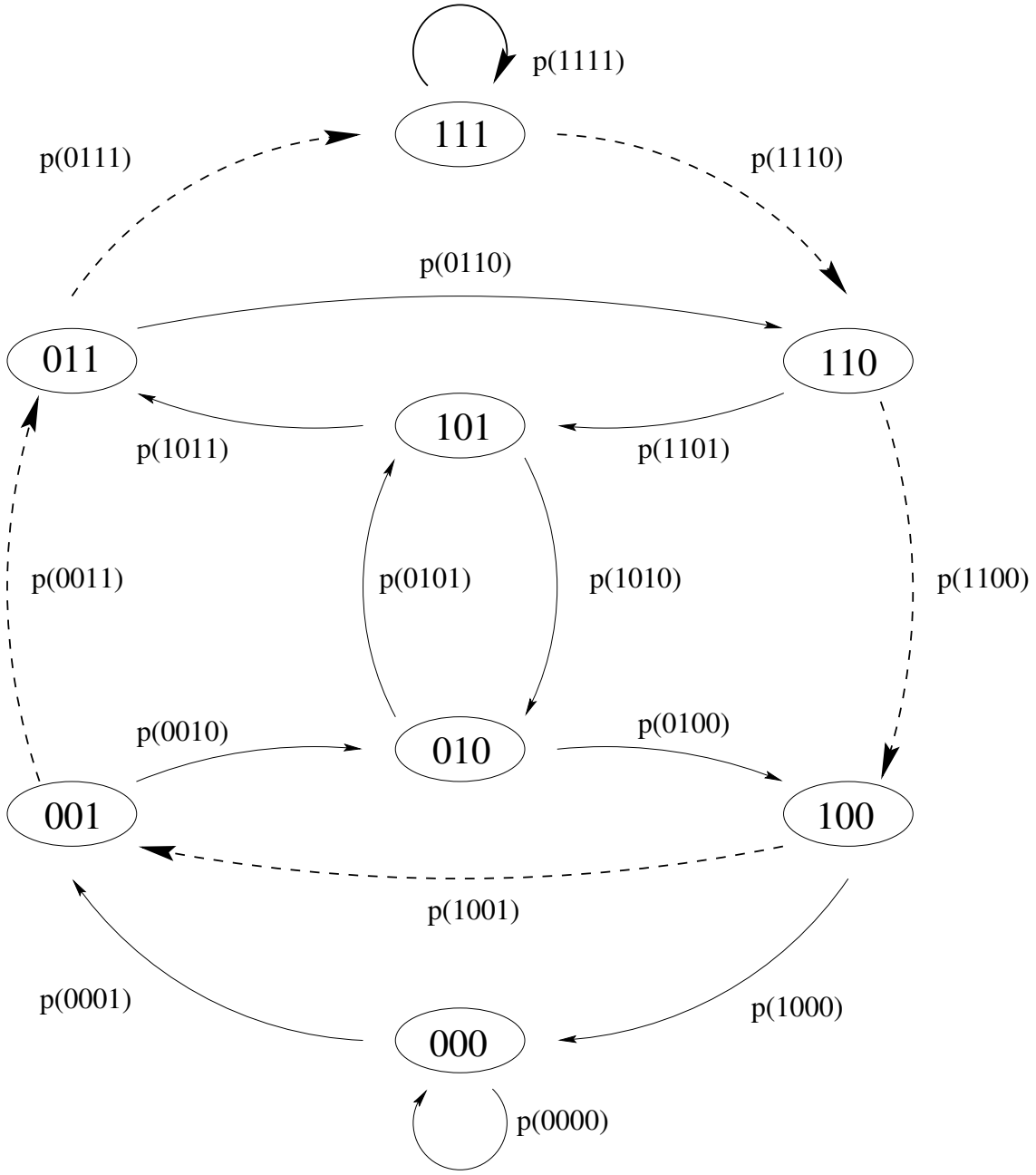

Figure 8.7: Double deformation faults in the 3C structure of ZnS depicted on an $r = 3$ de Bruijn graph. The broad solid lines represent the 3C structure (only the positive chirality structure (1)* shown), the dashed lines are the double deformation faults, and the thin solid lines are the remainder of the graph.

Chapter 9

ϵ-Machine Reconstruction from Experimental ZnS Diffraction Patterns

Now that the work of previous researchers has been discussed and new theoretical tools and procedures have been introduced, we are equipped to examine experimental spectra and discover the underlying process which gives rise to them. We will examine the diffraction patterns for seven single crystal ZnS found in reference [61]. They will be referred to by the page number on which they appear in [61]. Even though the data was taken in the mid 1980s, it unfortunately is only available in graphical form [56]. It was therefore necessary to scan the graphs and digitize the data.

A brief word about significant figures in the following. While most quantities are calculated to a precision of three decimal places, this in no way implies that any quantity is this well known. Recall that the spirit of this work is exploratory, and we are attempting to expound a new technique for the characterization and discovery of patterns in the stacking order of ZnS. As such, we do not want the clarity and power of the exposition to be lost in too little precision. No error analysis has been attempted, but if one were performed, the error would undoubtedly be large.

9.1 Machine Reconstruction from Experimental Diffraction Pattern SK229

The diffraction pattern along the $10.l$ row for an as grown 2H ZnS crystal is shown in Figure 9.1 and the diffraction pattern corrected for $\mathcal{C}(l)$ is displayed in Figure 9.2. We immediately notice in Figure 9.2 that the pattern is not periodic in l, but instead suffers from variations in the intensity. We see that the peaks at $l = -\frac{1}{2}$ and $\frac{1}{2}$, are of similar intensity, but the peak at $l = \frac{3}{2}$ seems to have about roughly one-half their brightness. The peaks at $l = 0$ and 1 also differ in their intensity. So we can be sure that this spectrum contains substantial error, and we will find this to plague the other spectra as well. As discussed in §4.2, there are criteria we can use to help select a unit interval over which the intensity seems not to vary too much. Looking at the spectrum in Figure 9.2, we might expect this interval to be between $l = -\frac{1}{2}, \frac{1}{2}$, and we indeed find that choosing $l_0 = -0.33$ gives reasonable figures of merit, namely $Q_s(1) = 0.008$, $\gamma = -0.489$ and $\beta = 1.004$. The first few correlation functions found by integrating over this interval are shown in Table 9.1. Values near one-half for the $Q_c(n)$ and $Q_r(n)$ with odd n are what we expect for a disordered 2H crystal. Calculating the correlation length for the Qs up to $n = 40$, we find it to be $\lambda_q = 19 \pm 2$. In Figure 9.3 a plot of $Q_s(n)$ vs. n is given that shows the oscillatory behavior in $Q_s(n)$ with period two over the first

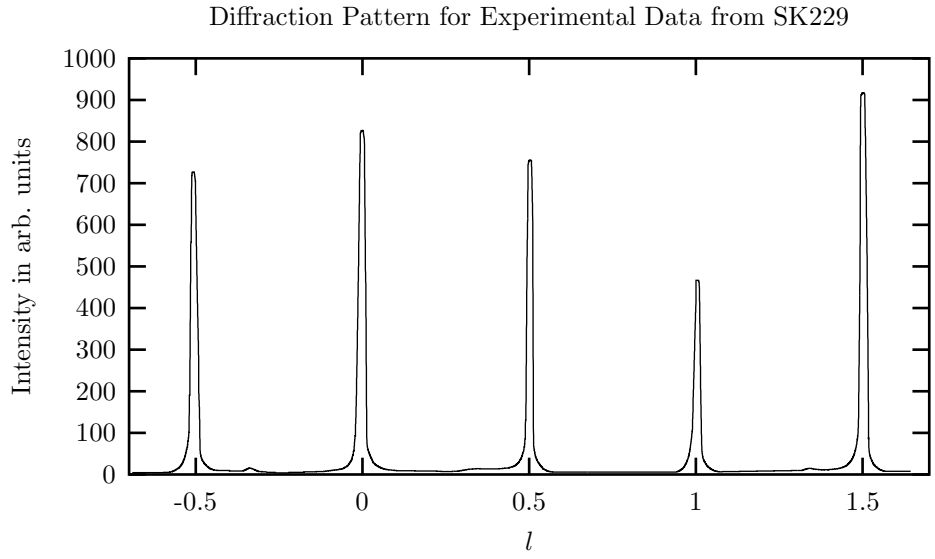

Figure 9.1: The uncorrected diffractometer pattern of intensity vs. l along the 10.l row for a single crystal of as grown ZnS. Experimental Data SK229.

Table 9.1: The first few correlation functions for experimental data SK229.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.487	0.505	0.008
2	0.066	0.037	0.897
3	0.442	0.490	0.068
4	0.088	0.067	0.845
5	0.438	0.488	0.074
6	0.095	0.067	0.838

forty-five n, which is again what we would expect for a crystal that can be thought of as largely 2H.

The results for machine reconstruction are displayed in Table 9.2. We see that the probability to visit words 1010 and 0101 have a combined total about 82.5%. These words of course correspond to the 2H cycle, as we might expect. The remaining probability is distributed among the other fourteen words. The $r = 3$ graph in Figure 9.4 shows the the strong 2H cycle in bold arcs. The dashed arcs for the words 1100 and 0011 are used to indicate their relative weakness. It is tempting to interpret the remaining arcs as faults, and indeed it seems we can. Let us treat the dashed arcs as though they are vanishing. Then in the upper half of the graph there appears to be a 2H deformation fault with probability weight 0.040 $[p(1011)+p(0111)+p(1110)+p(1101)]$ and in the bottom half there likewise seems to be a 2H deformation fault with weight 0.049 $[\frac{1}{2}p(0100)+p(1000)+p(0001)+\frac{1}{2}p(0010)]$. In the bottom portion there also seems to be growth fault with weight 0.036 $[\frac{1}{2}p(0100)+p(1001)+\frac{1}{2}p(0010)]$. The ferromagnetic arcs at the poles (1111 and 0000) have a combined weight of 0.041. A possible interpretation of this graph is a crystal structure with faults in the the following proportions:

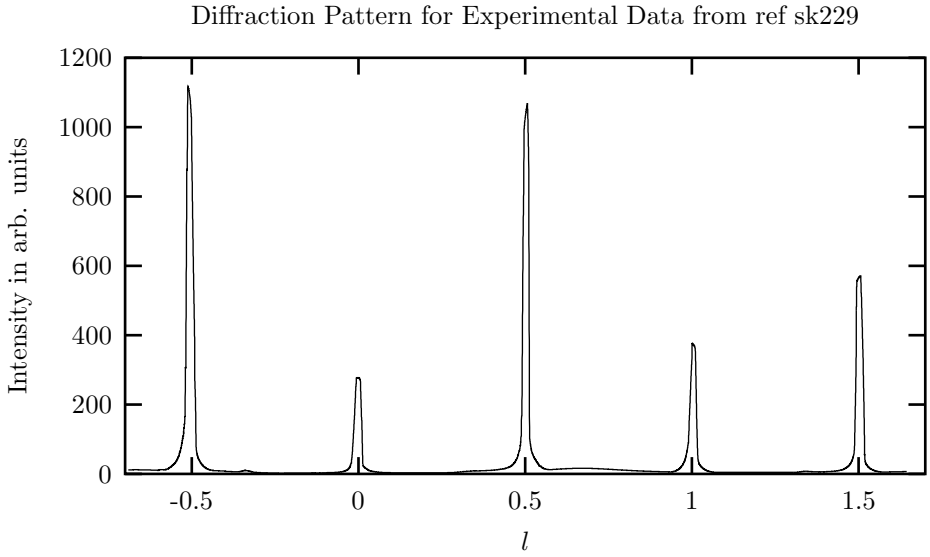

Figure 9.2: The diffraction pattern for Experimental Data SK229 corrected for $\mathcal{C}(l)$.

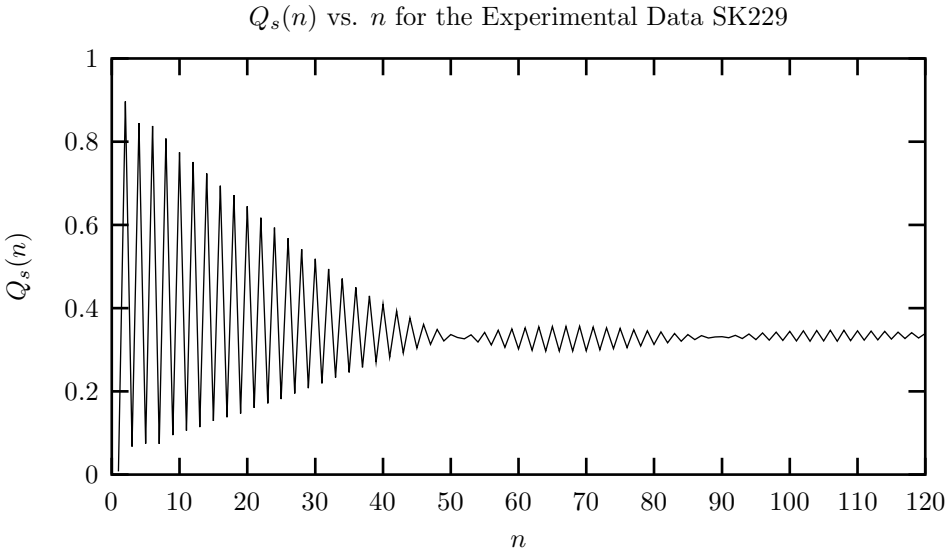

Figure 9.3: $Q_s(n)$ vs. n for Experimental Data SK229. We use $l_0 = -0.330$ and get a value of $\gamma = -0.489$ and $\beta = 1.004$. We find a correlation length of $\lambda_q = 19 \pm 2$ over the first forty layers.

Table 9.2: Machine reconstruction results for the experimental diffraction pattern SK229. The fitness for the $r = 3$ is $\mathcal{F} = 2.04 \times 10^{-6}$.

$r = 0$	$p(0)$	0.509	$r = 3$	$p(0000)$	0.026
	$p(1)$	0.491		$p(0001)$	0.013
				$p(0010)$	0.025
$r = 1$	$p(00)$	0.066		$p(0011)$	0.002
	$p(01)$	0.449		$p(0100)$	0.022
	$p(10)$	0.449		$p(0101)$	0.414
	$p(11)$	0.037		$p(0110)$	0.000
				$p(0111)$	0.012
$r = 2$	$p(000)$	0.040		$p(1000)$	0.013
	$p(001)$	0.026		$p(1001)$	0.013
	$p(010)$	0.438		$p(1010)$	0.411
	$p(011)$	0.010		$p(1011)$	0.010
	$p(100)$	0.026		$p(1100)$	0.004
	$p(101)$	0.423		$p(1101)$	0.006
	$p(110)$	0.010		$p(1110)$	0.012
	$p(111)$	0.027		$p(1111)$	0.015

Table 9.3: Computational Results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK229.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	1.000	0.000	0.000	0.000	0.00
$r = 1$ Approximation	SFT	1	0.473	0.527	1.000	0.526	0.53
$r = 2$ Approximation	SFT	2	0.308	0.692	1.47	0.856	1.02
$r = 3$ Approximation	SFT	3	0.301	0.699	1.79	0.887	1.09

2H	82.5%
3C	4.1%
deformation fault	8.9%
growth fault	3.6%
other disorder	0.9%

This decomposition is sensible because clearly there is some underlying crystal structure present, and the smaller, 'faulting' arcs aren't too large or complex. This need not always be the case.

Figure 9.5 compares the experimental $Q_s(n)$ and the $Q_s(n)$ obtained from the $r = 3$ approximation. The agreement is excellent, differences in the two only becoming apparent at about $n = 40$. We observe that the theoretical fit has stronger oscillations in $Q_s(n)$ above $n = 40$. There is similarly excellent agreement between in the diffraction patterns as shown in Figure 9.6. In fact, it is difficult to distinguish between the experimental and the $r = 3$ approximation in Figure 9.6, except perhaps in the region of $-0.4 \leq l \leq -0.3$. There we see that the $r = 3$ approximation slightly over estimates the diffuse scattering.

Table 9.3 presents a comparison of different measures of computation for the four different r approximations to the process. The general trends seem obvious enough. Increasing r implies more complex processes (as measured both by C_μ and **E**) as well as a decrease in the measure of the randomness, as measured by h_μ.

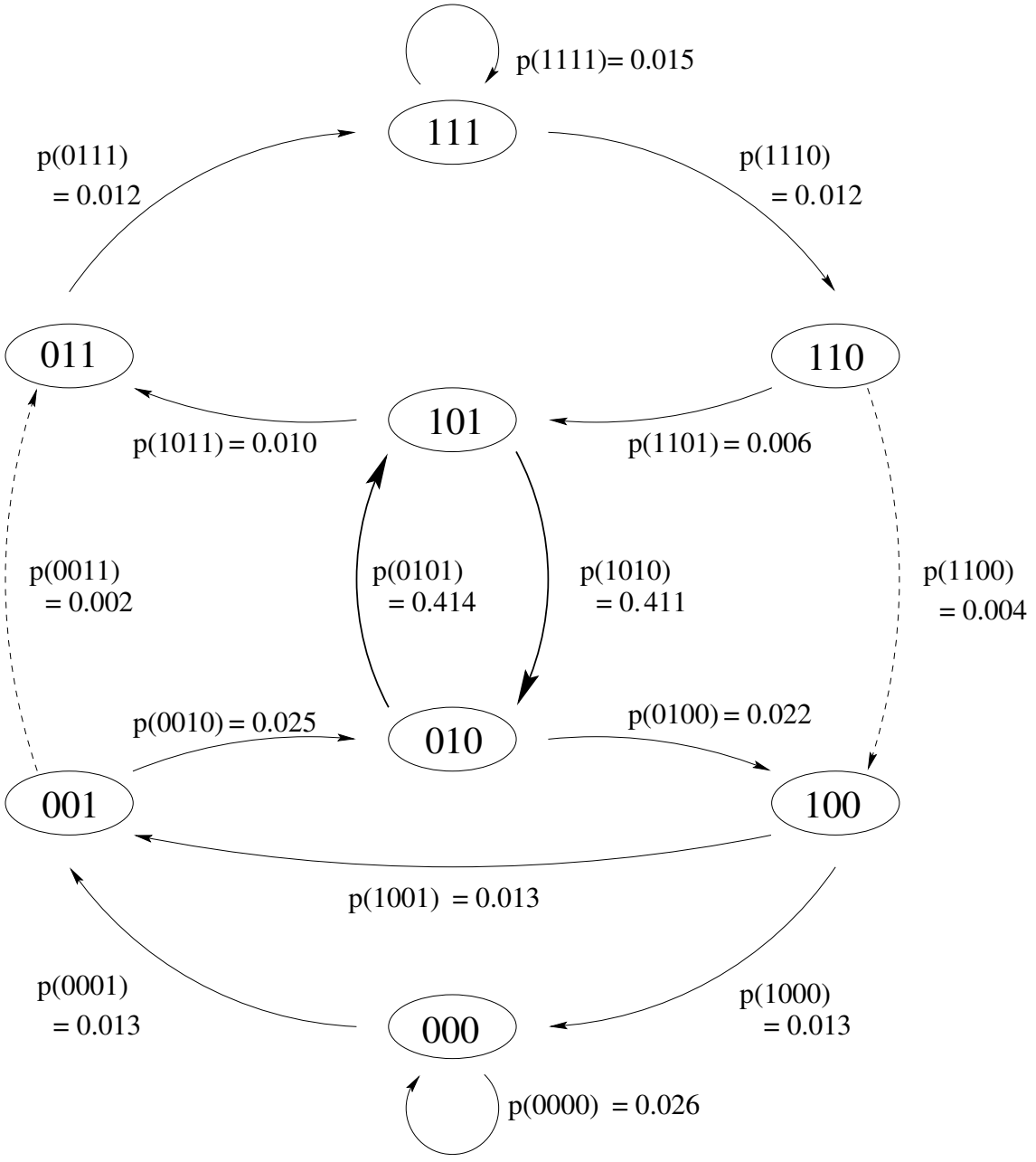

Figure 9.4: The $r = 3$ machine for diffraction pattern SK229. The bold arcs correspond to the 2H crystal structure and the dashed arcs have so little probability that we can take them to be zero for the purpose of decomposing the graph into a crystal and faults. The missing 0110 arc indicates that this word is absent.

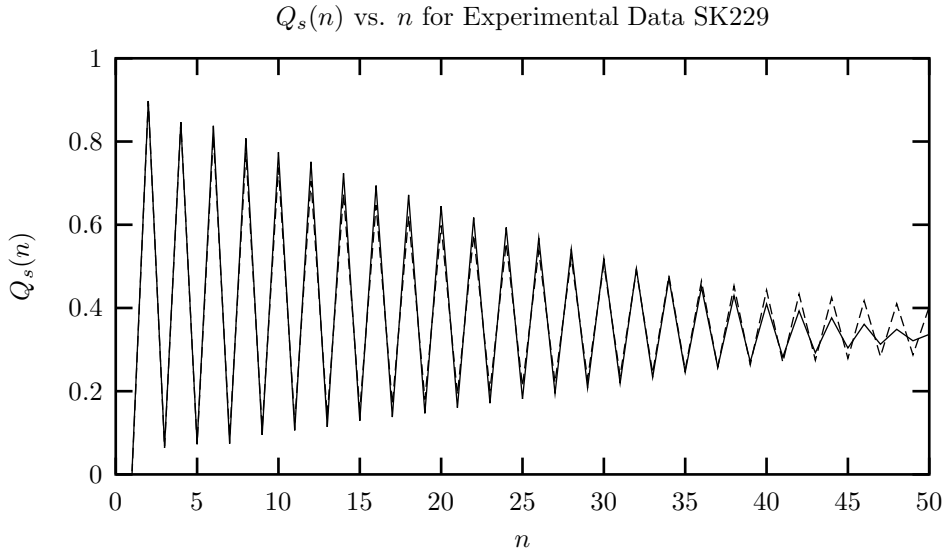

Figure 9.5: $Q_s(n)$ *vs.* n for Experimental Data SK229 (solid line) and the $r = 3$ approximation (dashed line). We see excellent agreement between the correlation functions for the $r = 3$ approximation and those of the process. It is only after $n = 35$ that differences become apparent.

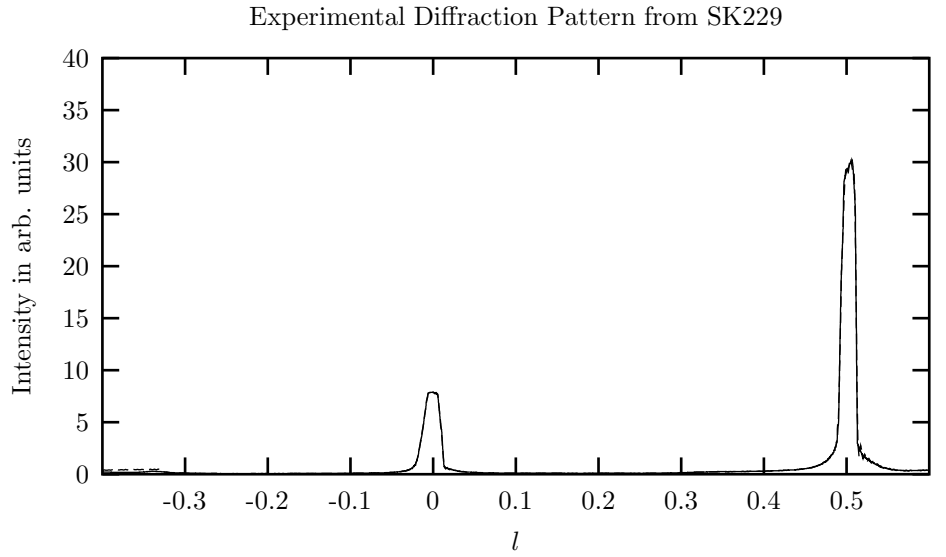

Figure 9.6: The diffraction pattern for Experimental Data SK229 (solid line) and the $r = 3$ approximation (dashed line). The agreement between the $r = 3$ approximation and experiment is so close that it is difficult to tell that there are two curves plotted.

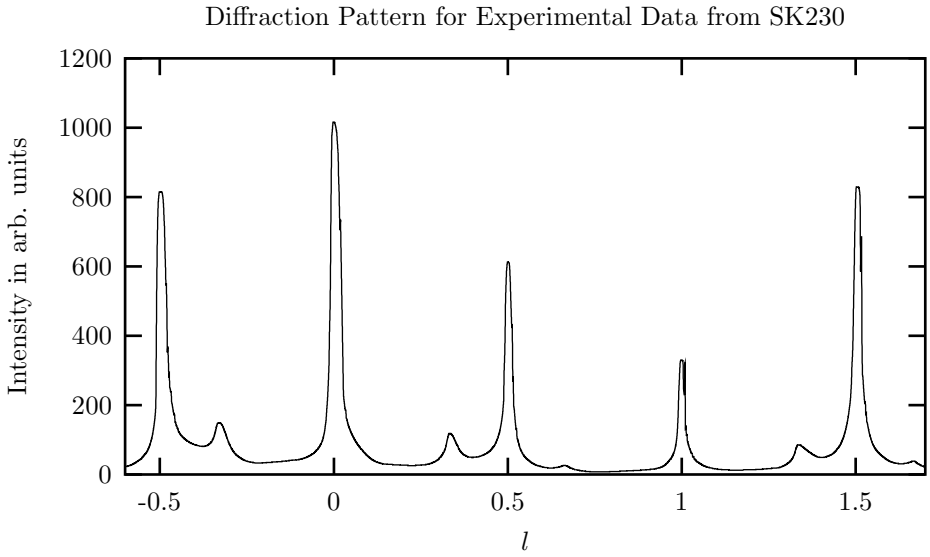

Figure 9.7: The uncorrected diffractometer pattern of intensity vs. l along the 10.l row for a perfect 2H crystal of ZnS after annealing for one hour at 300 C. Experimental Data SK230.

Table 9.4: The first few Qs for experimental data SK230.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.467	0.524	0.009
2	0.164	0.077	0.759
3	0.398	0.439	0.163
4	0.214	0.124	0.662
5	0.396	0.450	0.154
6	0.242	0.106	0.652

9.2 Machine Reconstruction from Experimental Diffraction Pattern SK230

Figure 9.7 shows the diffraction pattern along the 10.l row of a perfect 2H ZnS crystal that has been annealed at 300 C for one hour. The same diffraction pattern corrected for $\mathcal{C}(l)$ is shown in Figure 9.8. Again we notice that the intensity is not periodic in l, but appears to fall off slowly as l increases. The intensity seems to have been reduced by a factor of two on the interval $-0.5 \leq l \leq 1.5$. Integrating over the interval $-0.63 \leq l \leq 0.37$ we find the figures of merit to be $Q_s(n) = 0.009$, $\gamma = -0.486$ and $\beta = 1.022$. The first few correlation functions found from integrating over this interval are shown in Table 9.4. We can see values of $Q_c(n), Q_r(n)$ near one-half for n odd, suggesting that the original 2H structure is not too corrupted. This is reasonable considering strong peaks at integer and half integer l. Considering the first forty layers we get a correlation length of $\lambda_q = 8.4 \pm 0.4$.

The results for machine reconstruction appear in table 9.5 and Figure 9.10. We notice immediately two missing arcs, 0011 and 1001. Again the probability weights on the arcs corresponding to the 2H cycle are large, the sum of $p(0101)$ and $p(1010)$ being 60.9%. The remaining 30% arc weight can be

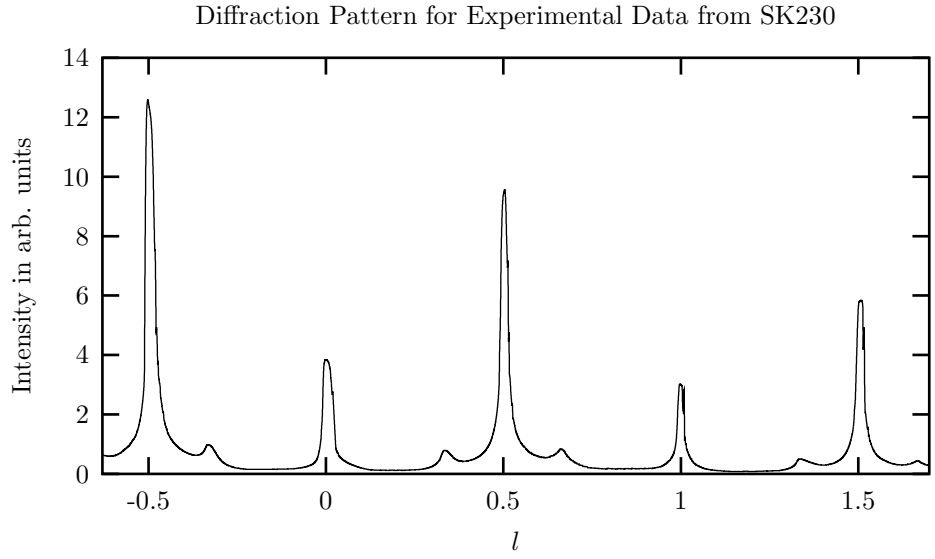

Figure 9.8: The diffraction pattern for Experimental Data SK230 corrected for $\mathcal{C}(l)$.

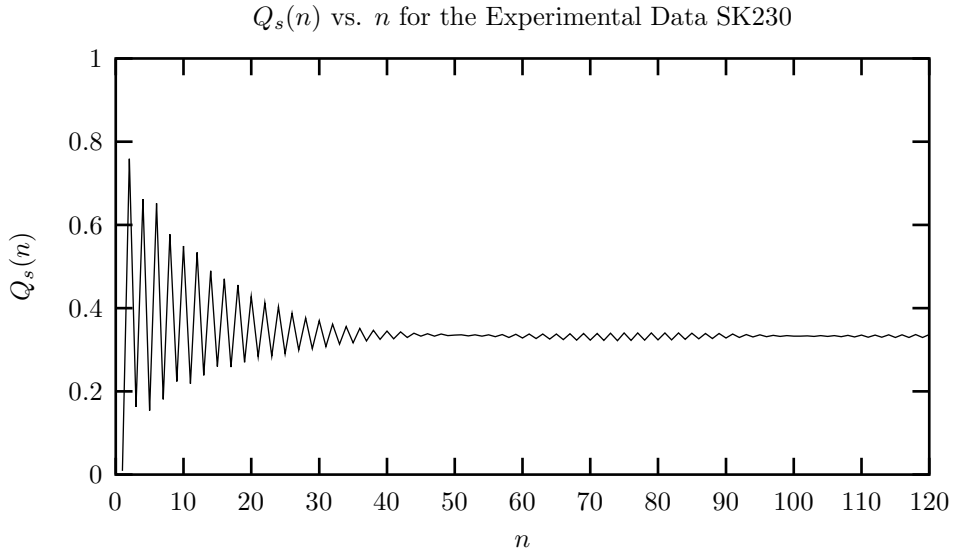

Figure 9.9: $Q_s(n)$ vs. n for Experimental Data SK230. We use $l_0 = -0.630$ and get a value of $\gamma = -0.486$ and $\beta = 1.022$. We find a correlation length of $\lambda_q = 8.4 \pm 0.4$ over the first forty layers.

Table 9.5: Machine reconstruction results for the experimental diffraction pattern SK230. The fitness for the $r = 3$ solution is $\mathcal{F} = 1.01 \times 10^{-5}$.

$r = 0$	$p(0)$	0.528	$r = 3$	$p(0000)$	0.066
	$p(1)$	0.472		$p(0001)$	0.049
				$p(0010)$	0.049
$r = 1$	$p(00)$	0.164		$p(0011)$	0.000
	$p(01)$	0.380		$p(0100)$	0.026
	$p(10)$	0.380		$p(0101)$	0.316
	$p(11)$	0.077		$p(0110)$	0.031
				$p(0111)$	0.005
$r = 2$	$p(000)$	0.118		$p(1000)$	0.049
	$p(001)$	0.046		$p(1001)$	0.000
	$p(010)$	0.347		$p(1010)$	0.293
	$p(011)$	0.032		$p(1011)$	0.036
	$p(100)$	0.046		$p(1100)$	0.022
	$p(101)$	0.334		$p(1101)$	0.011
	$p(110)$	0.032		$p(1110)$	0.005
	$p(111)$	0.045		$p(1111)$	0.040

relegated to disorder. Again we can attempt to understand this disorder in terms of stacking faults. One possible break down is to say that the simple cycle $1011 \to 0111 \to 1110 \to 1101$ is a deformation fault and assign to each arc the same weight. This can be done if each is given the weight 0.005. We might then want to identify the cycle $1011 \to 0110 \to 1101$ as a growth fault, and assign the weight of 0.006 to each arc. If we then follow the cycle $1011 \to 0110 \to 1100 \to 1000 \to 0001 \to 0010$ this appears to be a deformation fault, with each arc bearing a weight of approximately 0.022. We can finally round out the faults by taking the cycle $0100 \to 1000 \to 0001 \to 0010$ as a deformation fault with each arc taking a weight of about 0.026. This is a consistent breakdown of the graph into crystalline and faulted cycles, but it is not the only possible such decomposition.

We could, for instance, take the position that there is no (single) deformation fault on the upper part of the graph by instead treating the cycle $1011 \to 0111 \to 1110 \to 1100 \to 1000 \to 0001 \to 0010$ as a fault in its own right. One can find a logically consistent distribution of stacking faults under this assumption that differs from the previous analysis. We then are faced with a situation of a single graph giving rise to two different faulting configurations, and should well question the validity and usefulness of the fault picture. It seems rather that the graph is the thing, being concise and unique at each r. It is perhaps not as intuitive as thinking of a crystal permeated with a certain fraction of stacking 'errors', but it is unambiguous. We consider this no mere semantic quibble either. Where the faulting picture becomes less tenable as the fraction of stacking faults increases, the graphical picture seamlessly handles any amount of disorder, from a near perfect crystal to complete randomness and everything in between. Indeed, the graph, or ϵ-machine, provides the minimum structure needed to specify the statistics of the stacking. In short, it is the answer. Our interpretation of the graph is then a matter of convenience and perhaps psychology, but a description of disordered structures based on 'faultology' is not fundamental.

This point can be made more quantitative by a careful treatment of our procedure for assigning stacking faults. It is always possible to break down an infinite sequence generated by de Bruijn graph into a sequence of simple cycles. This decomposition however is not unique. Since a simple cycle is just the indefinite repetition of a finite length sequence, we see that this just corresponds to some crystal structure. A graph, of course, may have more that one simple cycle. If there are two such cycles sharing, say, a node, and the conditional probability to branch at this node strongly

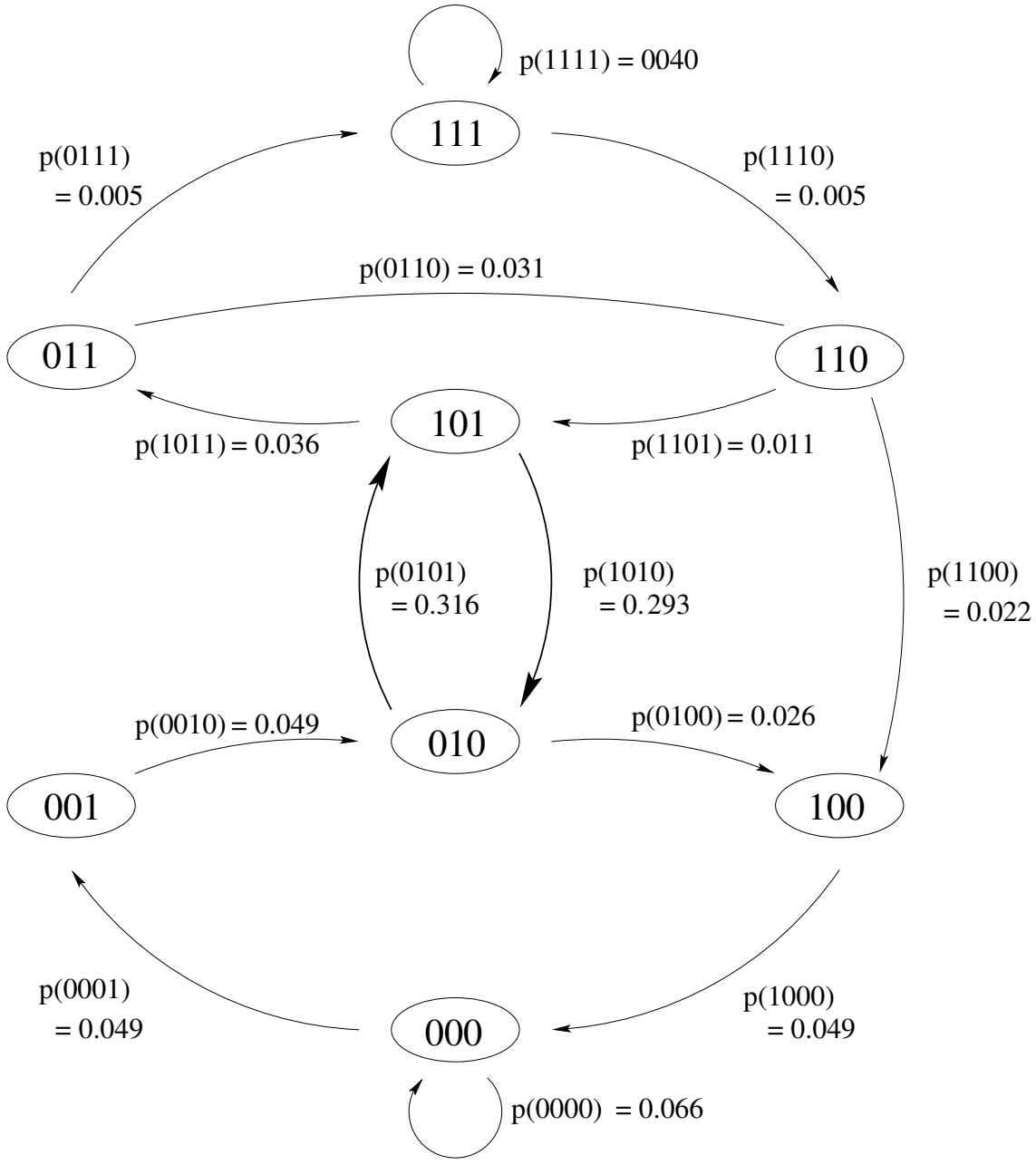

Figure 9.10: The $r = 3$ machine for diffraction pattern SK230.

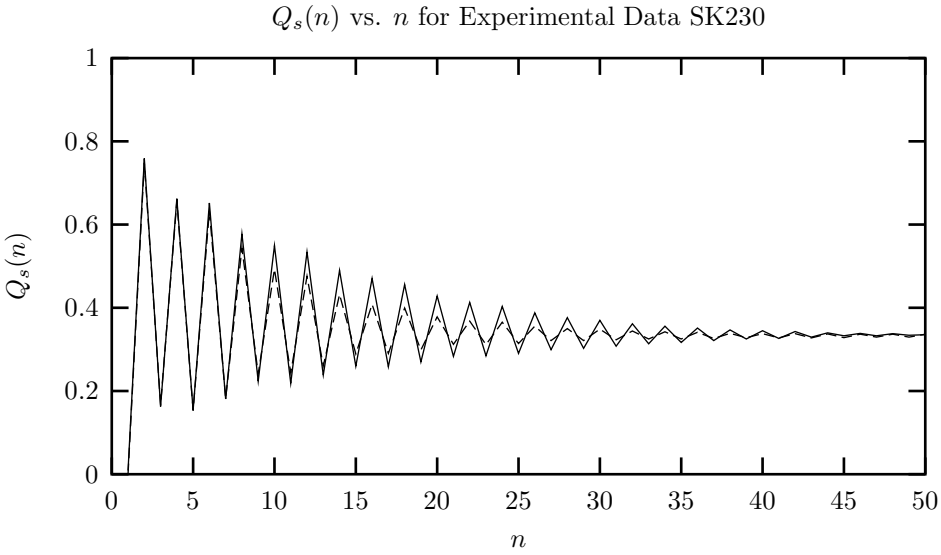

Figure 9.11: $Q_s(n)$ vs. n for Experimental Data SK230 (solid line) and the $r = 3$ approximation (dashed line).

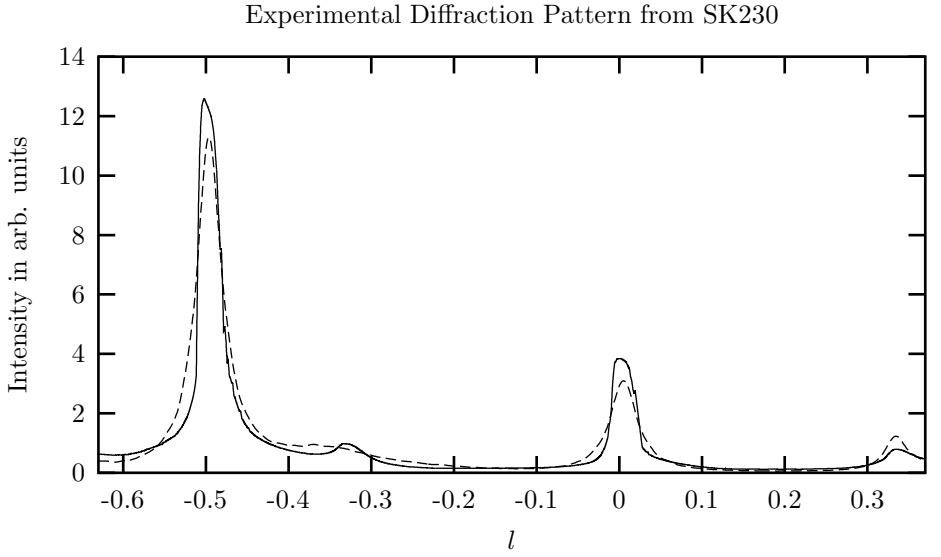

Figure 9.12: The diffraction pattern for Experimental Data SK230 and the $r = 3$ approximation.

Table 9.6: Computational Results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK230.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	0.998	0.002	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.779	0.221	0.995	0.216	0.22
$r = 2$ Approximation	SFT	2	0.577	0.423	1.774	0.620	0.82
$r = 3$ Approximation	SFT	3	0.487	0.513	2.41	0.945	1.51

Table 9.7: Resolution of arc weights into crystal structures and faulting structures for experimental data. We see that most spectra can not be thought of profitably in this way as it is not possible to resolve a large portion of the arc weights.

Spectrum	3C Weight	2H Weight	Other Known Weight	Unresolved Weight
SK229	0.041	0.822	0.137	0.000
SK230	0.106	0.586	0.000	0.308
SK231	0.407	0.128	0.000	0.445
SK232	0.445	0.000	0.000	0.555
SK134	0.080	0.644	0.000	0.276
SK135	0.541	0.000	0.000	0.459
SK137	0.440	0.058	0.000	0.502

favors taking one of the simple cycles, then a reasonable picture is that the oft-taken branch is the 'parent' crystal structure and less-taken branch is the faulting. An important point here is that both the crystal structure and the fault can be thought of as simple cycles. This picture works well in this simple case. Even in the presence of a strongly favored simple cycle and a few much weaker, non-intersecting simple cycles, the view makes sense. But we are lead to ask how many possible simple cycles are there on a r^{th}-order de Bruijn graph. In general this is not known, but it is thought to grow as the exponential of an exponential in r [68]. Table 2.2 shows the number of simple cycles on a de Bruijn graph of range r. Since it takes only 2^r parameters to specify all the probability weights on a graph, we see that there are usually many more simple cycles the free parameters. This means that, for a given structure as specified by the 2^r parameters of a de Bruijn graph, there are on the order of $3r \times 2^{f(r)}$ simple cycles, where $f(r) = 2^{r-1} - r$ [68]. The problem of relating the probability weights on a graph to the frequency of simple cycles is then under-determined. There are, in general, infinitely many decompositions. In this light, any such decomposition becomes dubious unless there are enough vanishing arcs to uniquely specify the simple cycle decomposition. We find in the reconstructed ϵ-machines of the seven spectra we have analyzed, that only the the first, SK229, admits a reasonable decomposition into simple cycles. For the other six, decomposition into simple cycles is not unique, and we maintain, not useful. Table 9.7 shows how each spectrum can be broken down into 3C and 2H structure, along with contributions from other known cycles and finally probability weights to which we cannot meaningfully assign either a faulting or crystal structure. For the spectrum SK232 for instance, we find that a full 55% of the probability weight can not be meaningfully assigned to either some crystal or specific fault. The faulting picture has failed.

Lastly, we compare the diffraction pattern and correlation functions generated by our $r = 3$ reconstructed machine with experiment. In Figure 9.11 we see good agreement with experiment in the $Q_s(n)$ up to about $n = 10$, and thereafter the magnitude of the oscillations in the experimental $Q_s(n)$ is underrepresented in the machine generated $Q_s(n)$. From our experience with fake data,

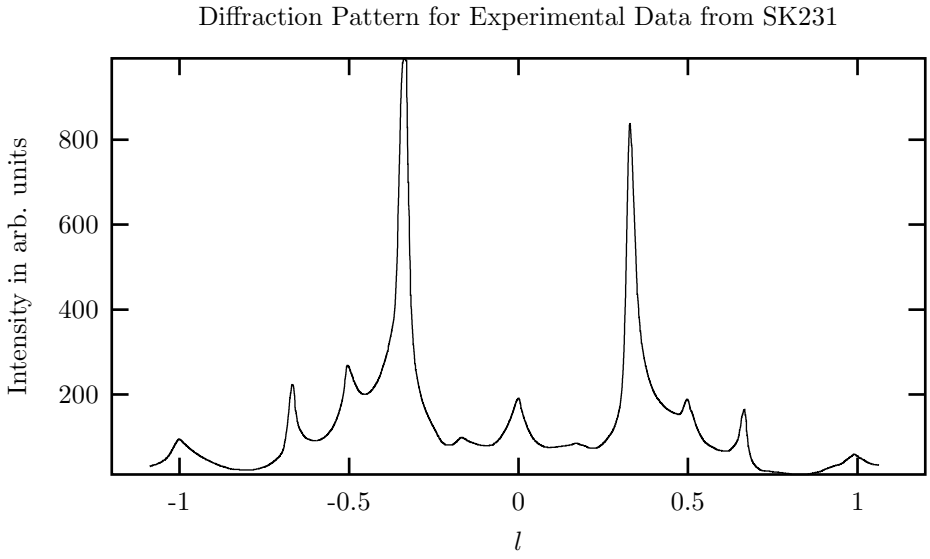

Figure 9.13: The uncorrected single crystal diffractometer reading along the 10.l reciprocal lattice row after annealing a 2H ZnS crystal at 600 C for one hour. SK231 experimental data.

a likely cause is undiscovered computation in the process. Such order might be manifest in higher r diagrams. Indeed there is speculation that the transformation to the 6H structure $(111000)^*$ is important in annealed crystals. In order to detect competition between the 3C and 6H structures it is important that there be a 'barrier' between them, that is their cycles must not share a node. For $r = 3$, they share the 111 node; so, by construction, the proposed process lacks the necessary richness to model this. There is such a barrier in $r = 4$ graph, as seen in §7.1, 7.2, and we have hopes that the competition between these polytypes can be detected there. That is, however, beyond the scope of this current work. In Figure 9.12, we observe that the agreement between the experimental diffraction pattern and the diffraction pattern from the $r = 3$ approximation to it is less than satisfactory. Again, smeared out peaks is what we expect if the process does not contain sufficient structure to model the computation present.

9.3 Machine Reconstruction from Experimental Diffraction Pattern SK231

In Figure 9.13 we see the diffraction pattern along the 10.l row for a ZnS crystal annealed at 600 C for one hour and Figure 9.14 shows this same pattern after correcting for $\mathcal{C}(l)$. This pattern also suffers from a lack of periodicity in l although the symmetry present leads us to hope that the gradual falling off of intensity is not too great over either the interval $-1.0 \leq l \leq 0.0$ or $0.0 \leq l \leq 1.0$. Indeed, the figures of merit over the former interval are rather good, being $Q_s(n) = 0.000$, $\gamma = -0.500$, and $\beta = 1.024$. The first few correlation functions found by integration over this interval are given in Table 9.8. We notice that these first few Qs, save those at $n = 1$, are confined to a rather narrow interval of $0.244 \leq Q \leq 0.487$. Compared with previous patterns, this one doesn't seem to have as much variation, the Qs having more in common with a random number generator. We might guess that the original 2H structure is largely obliterated, and in fact the total weight assigned to the 1010 and 0101 arc is 14.8%. This is not much more than one sees in a random number generator

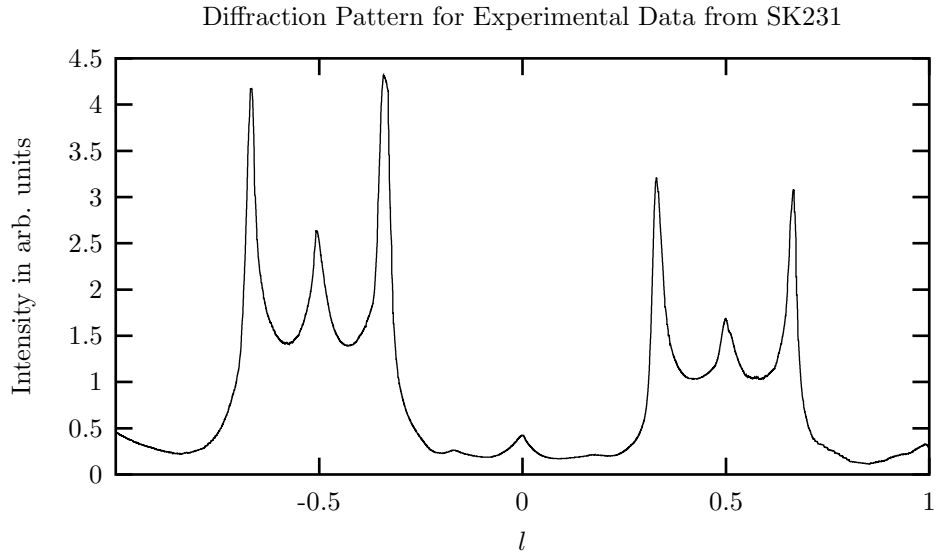

Figure 9.14: The diffraction pattern for Experimental Data SK231 corrected for $\mathcal{C}(l)$.

Table 9.8: The first few Qs for experimental data SK231.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.502	0.498	0.000
2	0.324	0.304	0.373
3	0.255	0.260	0.485
4	0.373	0.369	0.258
5	0.385	0.348	0.267
6	0.244	0.269	0.487

(12.5%). The machine reconstruction results are displayed in Table 9.9 and the $r = 3$ machine is in Figure 9.16. We see that no arcs are missing.

There is enhancement in the ferromagnetic arcs, $p(1111) = 0.193$ and $p(0000) = 0.214$. These are by a factor of about two and one-half the most prominent arcs on the graph. Since they have nearly the same weight, this crystal is referred to as a twinned crystal.

While the figures of merit are excellent for this spectrum, the fitness function is not nearly as small as for the previous two spectra. We get a value of $\mathcal{F} = 2.52 \times 10^{-4}$ for SK231, which is one hundred times the $\mathcal{F}$ of SK229 and twenty times the $\mathcal{F}$ value of SK230. It is near the fitness we found for the noisy period two and the even system. Recall that neither of these systems is describable as a finite memory process, so we expect the fitness at $r = 3$ not to be good. Looking at the comparison between the experimental $Q_s(n)$ and the $r = 3$ generated $Q_s(n)$ in Figure 9.17, we see that their agreement only extends out to about $n \approx 8$ or 9. After that the $r = 3$ generated $Q_s(n)$ decay to the asymptotic value too quickly. Looking at the comparison between the diffraction patterns in Figure 9.18 we observe that the peaks at $l = -\frac{2}{3}$ and $-\frac{1}{3}$ are underrepresented by the $r = 3$ approximate process. The peak at $l = -\frac{1}{2}$ is too broad in the $r = 3$ generated diffraction pattern. We also notice that at $r = 3$, the value for the entropy production per layer remains large at $h_\mu = 0.785$. All of these facts together imply that there is important computation not being

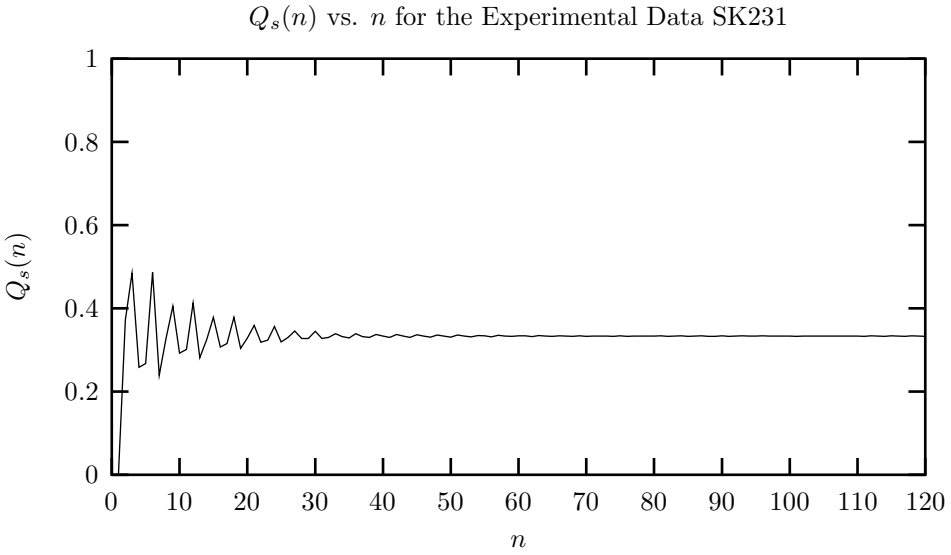

Figure 9.15: $Q_s(n)$ vs. n for Experimental Data SK231. We use $l_0 = -1.000$ and get a value of $\gamma = -0.500$ and $\beta = 1.024$. We find a correlation length of $\lambda_q = 6.8 \pm 1.7$ over the first twenty n.

Table 9.9: Machine reconstruction results for the experimental diffraction pattern SK231. We find a fitness of $\mathcal{F} = 2.52 \times 10^{-4}$ for the $r = 3$ solution.

$r = 0$	$p(0)$	0.498	$r = 3$	$p(0000)$	0.214
	$p(1)$	0.502		$p(0001)$	0.041
				$p(0010)$	0.056
$r = 1$	$p(00)$	0.324		$p(0011)$	0.011
	$p(01)$	0.186		$p(0100)$	0.038
	$p(10)$	0.186		$p(0101)$	0.084
	$p(11)$	0.304		$p(0110)$	0.040
				$p(0111)$	0.028
$r = 2$	$p(000)$	0.251		$p(1000)$	0.043
	$p(001)$	0.072		$p(1001)$	0.024
	$p(010)$	0.116		$p(1010)$	0.064
	$p(011)$	0.071		$p(1011)$	0.057
	$p(100)$	0.072		$p(1100)$	0.029
	$p(101)$	0.114		$p(1101)$	0.040
	$p(110)$	0.071		$p(1110)$	0.036
	$p(111)$	0.233		$p(1111)$	0.193

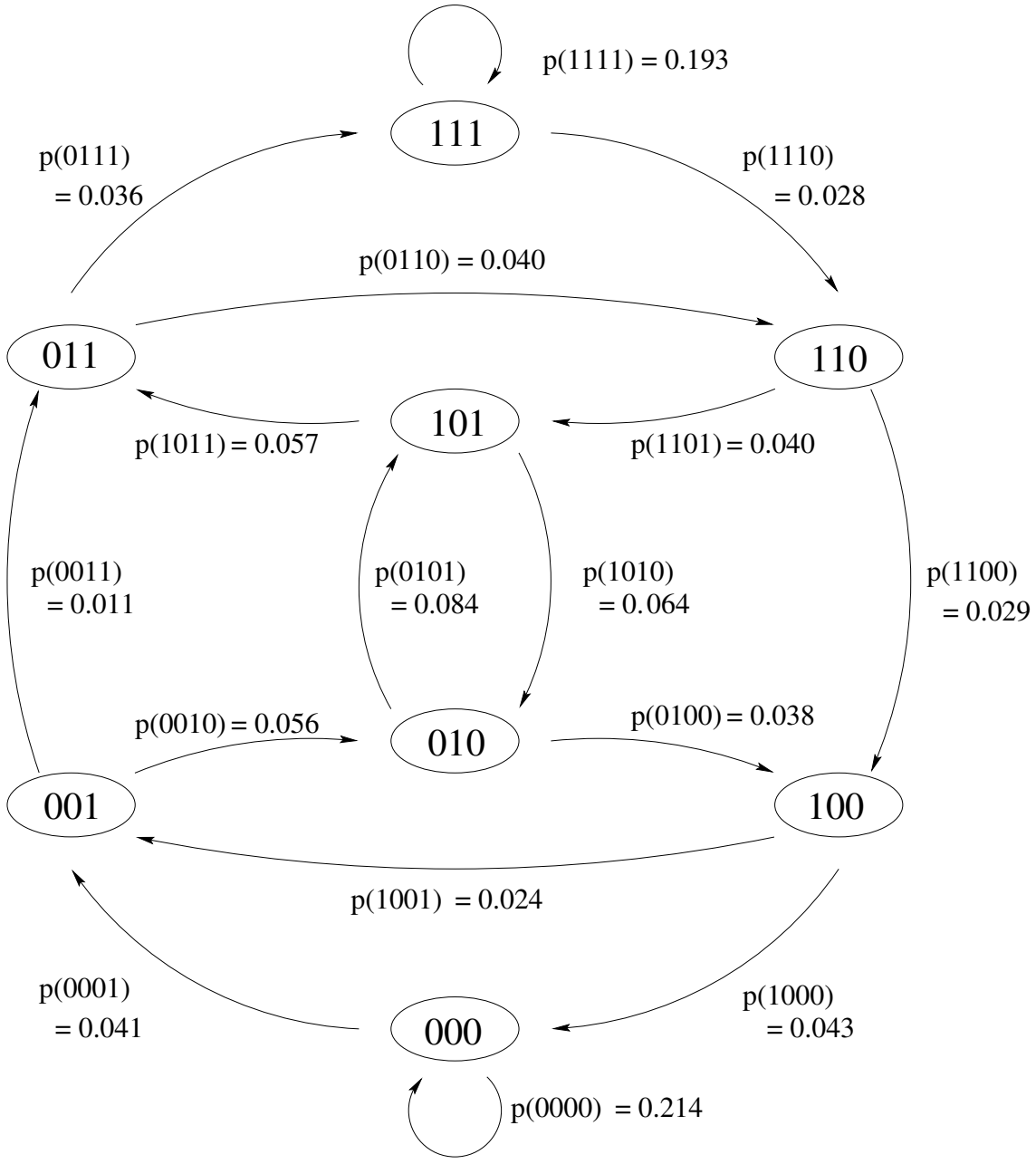

Figure 9.16: The $r = 3$ machine for diffraction pattern SK231.

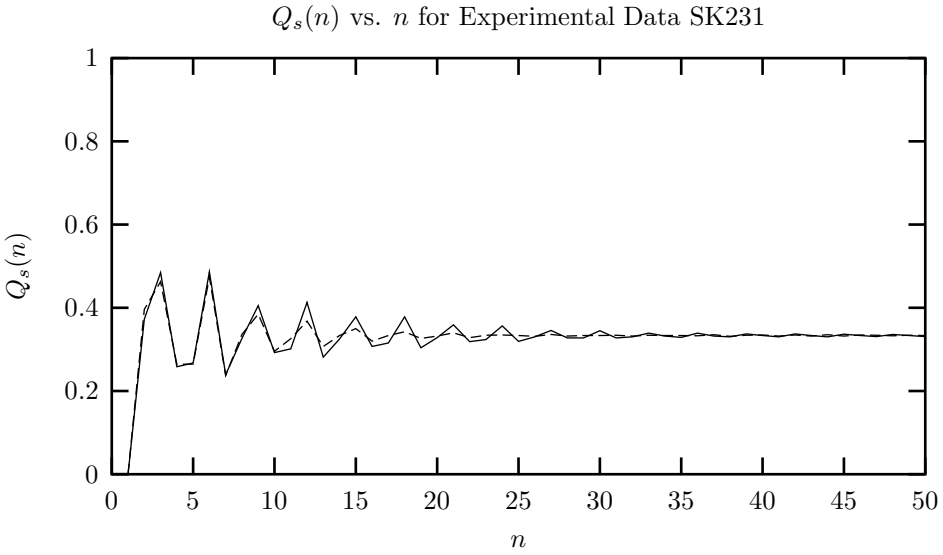

Figure 9.17: $Q_s(n)$ vs. n for Experimental Data SK231 (solid line) and the $r = 3$ approximation (dashed line).

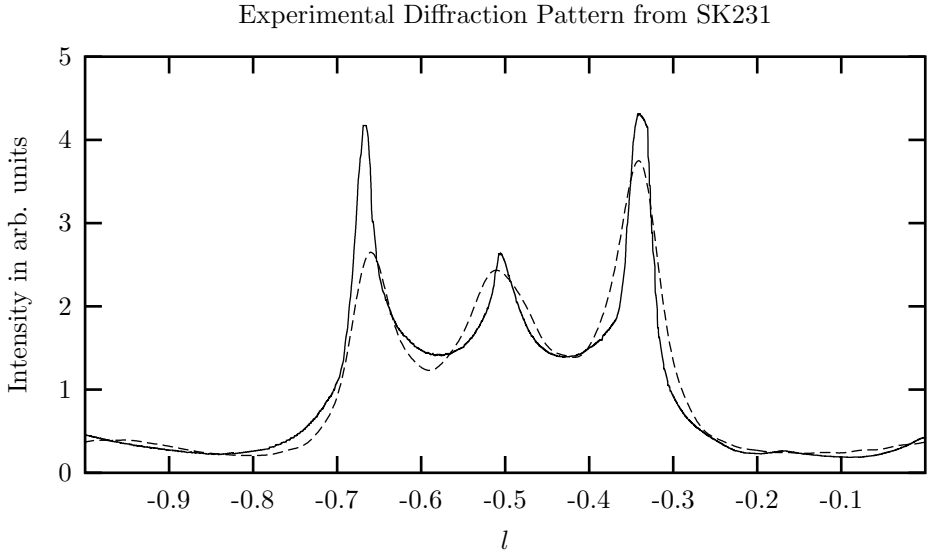

Figure 9.18: The diffraction pattern for Experimental Data SK231 and the $r = 3$ approximation.

Table 9.10: Computational results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK231.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	1.000	0.000	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.952	0.048	1.000	0.048	0.048
$r = 2$ Approximation	SFT	2	0.844	0.156	1.953	0.265	0.373
$r = 3$ Approximation	SFT	3	0.785	0.215	2.804	0.449	0.747

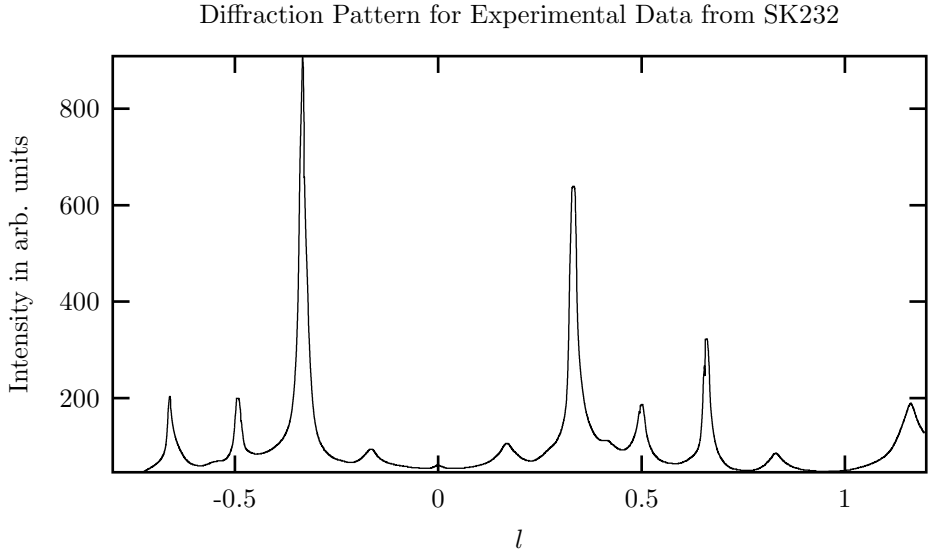

Figure 9.19: The uncorrected single crystal diffractometer reading along the $10.l$ reciprocal lattice row after annealing a 2H ZnS crystal at 700 C for one hour. SK232 experimental data.

modeled well in the $r = 3$ approximation. We can only speculate as to what this missing structure is. It seems likely that there is competition between the 3C and the 6H structures present, which can not effectively be modeled in an $r = 3$ graph.

9.4 Machine Reconstruction from Experimental Diffraction Pattern SK232

We show the diffraction pattern along the $10.l$ row for a ZnS crystal annealed at 700 C for one hour in Figure 9.19 and this same pattern corrected for $\mathcal{C}(l)$ in Figure 9.20. As before, the corrected diffraction pattern is not periodic in l. We find reasonable figures of merit, however by integrating over the interval $-0.72 \leq l \leq 0.28$. They are, $Q_s(1) = 0.014$, $\gamma = -0.480$ and $\beta = 0.966$. The first few correlation functions are shown in Table 9.11. A plot of $Q_s(n)$ vs. n is given in Figure 9.21 and we find a correlation length for the Q's to be $\lambda_q = 9.2 \pm 1.6$ considering the first twenty-eight layers.

The results for machine reconstruction are displayed in Table 9.12. We get a fitness of $\mathcal{F} = 2.70 \times 10^{-5}$ for the $r = 3$ solution. The graph for the $r = 3$ approximation is shown in Figure 9.22. We notice that the ferromagnetic arcs are of about equal probability and together comprise about 45% of the probability weight, making this another disordered twinned crystal. The remaining probability

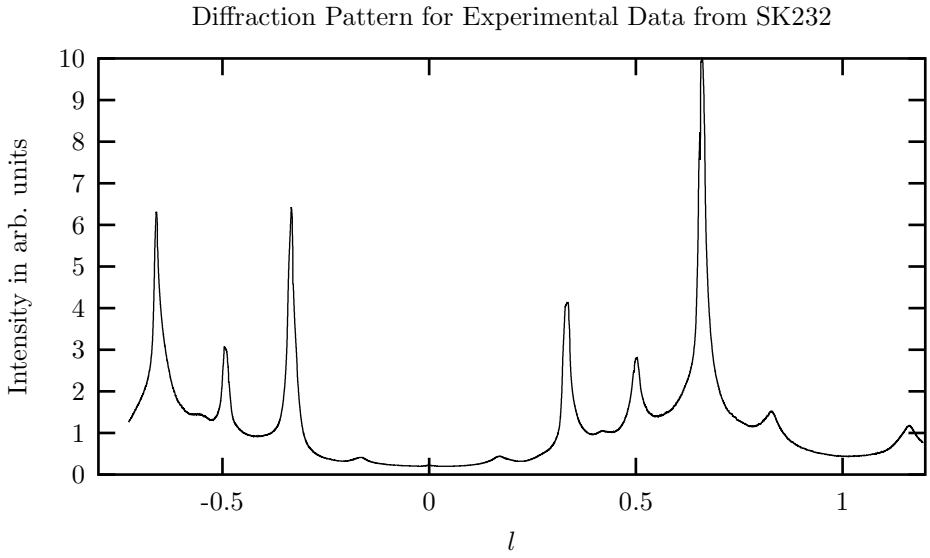

Figure 9.20: The diffraction pattern for Experimental Data SK232 corrected for $\mathcal{C}(l)$.

Table 9.11: The first few Qs for experimental data SK232 found by integration over the interval $-0.72 \leq l \leq 0.28$.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.542	0.445	0.014
2	0.289	0.406	0.305
3	0.274	0.211	0.515
4	0.388	0.387	0.225
5	0.358	0.397	0.245
6	0.260	0.214	0.526

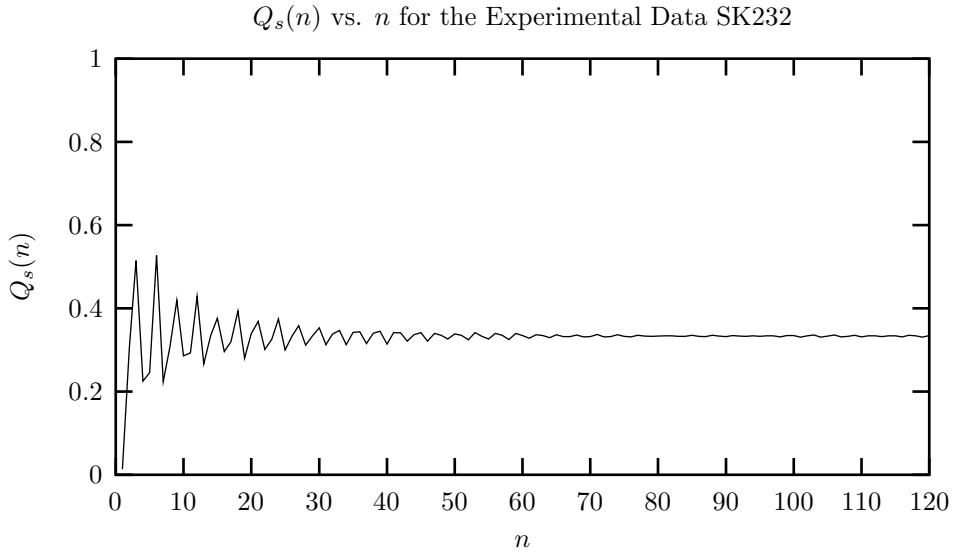

Figure 9.21: $Q_s(n)$ vs. n for Experimental Data SK232. We use $l_0 = -0.720$ and get a value of $\gamma = -0.480$ and $\beta = 0.966$. We find a correlation length of $\lambda_q = 9.2 \pm 1.6$ over the first twenty-eight n.

Table 9.12: Machine reconstruction results for the experimental diffraction pattern SK232. We find a fitness of $\mathcal{F} = 2.70 \times 10^{-5}$ for the $r = 3$ solution.

$r = 0$	$p(0)$	0.451	$r = 3$	$p(0000)$	0.183
	$p(1)$	0.549		$p(0001)$	0.030
				$p(0010)$	0.061
$r = 1$	$p(00)$	0.289		$p(0011)$	0.014
	$p(01)$	0.152		$p(0100)$	0.008
	$p(10)$	0.152		$p(0101)$	0.051
	$p(11)$	0.406		$p(0110)$	0.048
				$p(0111)$	0.045
$r = 2$	$p(000)$	0.210		$p(1000)$	0.030
	$p(001)$	0.080		$p(1001)$	0.046
	$p(010)$	0.052		$p(1010)$	0.000
	$p(011)$	0.100		$p(1011)$	0.081
	$p(100)$	0.080		$p(1100)$	0.068
	$p(101)$	0.073		$p(1101)$	0.030
	$p(110)$	0.100		$p(1110)$	0.046
	$p(111)$	0.305		$p(1111)$	0.262

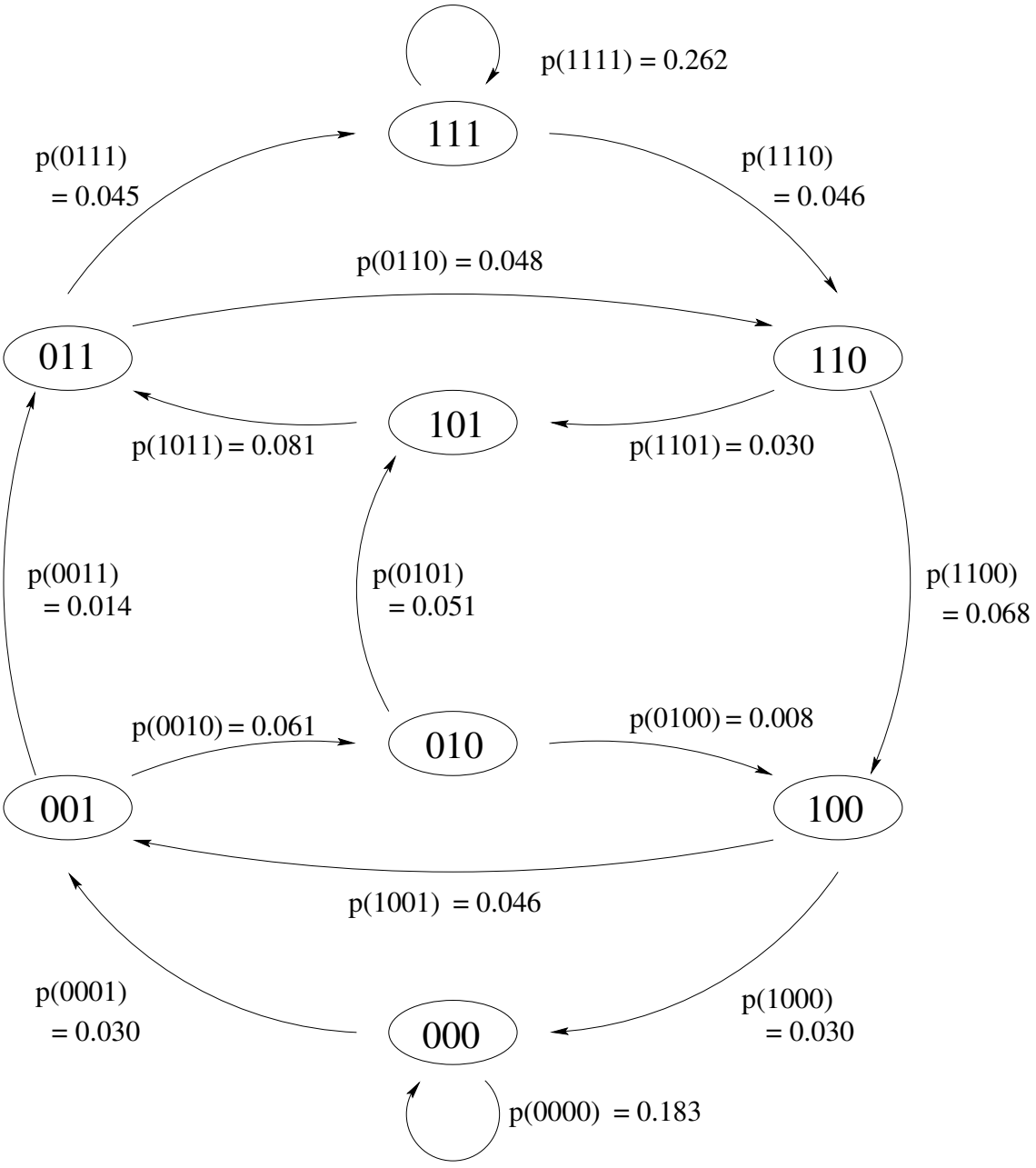

Figure 9.22: The $r = 3$ machine for diffraction pattern SK232.

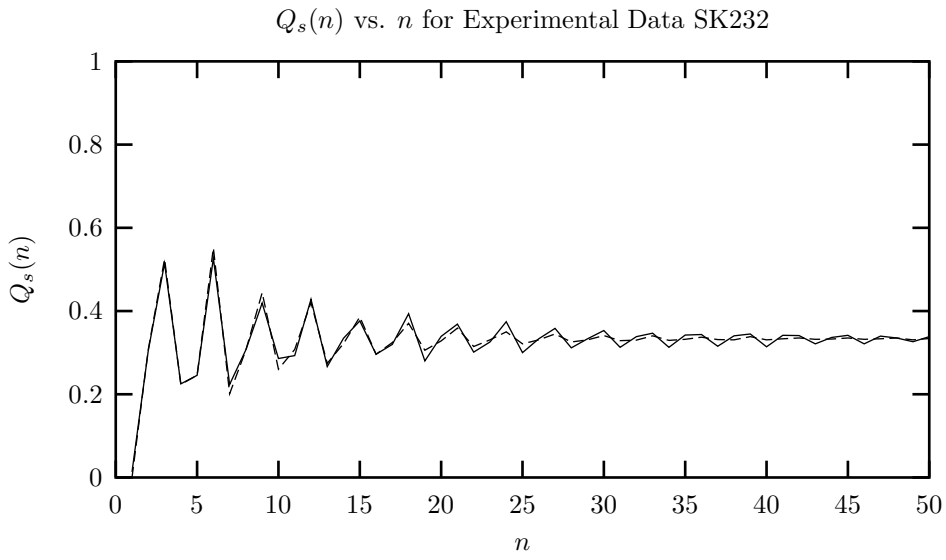

Figure 9.23: $Q_s(n)$ vs. n for Experimental Data SK232 (solid line) and the $r = 3$ approximation (dashed line).

Table 9.13: Computational Results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK232.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	0.993	0.007	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.882	0.118	0.990	0.108	0.108
$r = 2$ Approximation	SFT	2	0.866	0.134	1.872	0.140	0.156
$r = 3$ Approximation	SFT	3	0.644	0.356	2.726	0.793	1.451

is sprinkled among the other thirteen non-vanishing arcs in what appears to be a haphazard fashion. Since there is no unique decomposition into faults, and no obviously important "minor" arcs present, we do not attempt to identify a particular faulting mechanism here. Sebastian and Krishna [61] report that the major mechanism is deformation. It is interesting to note that the only vanishing arc in the graph is the 1010 arc. Since this began as a 2H crystal, we see that the annealing process has wiped out the original structure. It therefore seems ill-conceived to think of this as a faulted 2H crystal. It is simply a highly disordered one, with $h_\mu = 0.644$, that has some sections showing 3C structure.

The plot of $Q_s(n)$ vs. n for both the experimental data and the $r = 3$ process are shown in Figure 9.23. We notice that they begin to disagree as early as $n = 7$, but the disagreement doesn't become too bad until about $n = 18$. For smaller n, the $Q_s(n)$ overestimate the correlations and after $n = 18$ they underestimate it. This is novel, since in the previous examples the correlation functions derived from approximation did one or the other. Examining the experimental and theoretical diffraction patterns in Figure 9.24 we see that the $r = 3$ approximation has difficulty reproducing the sharpness of the reflection peaks at $l = -\frac{1}{2}$ and $-\frac{1}{3}$. It seems likely that there exists undiscovered structure in the natural process that the $r = 3$ approximation is not picking up.

Finally, the computational results for the $r = 0, 1, 2$ and 3 approximations is displayed in Table 9.13.

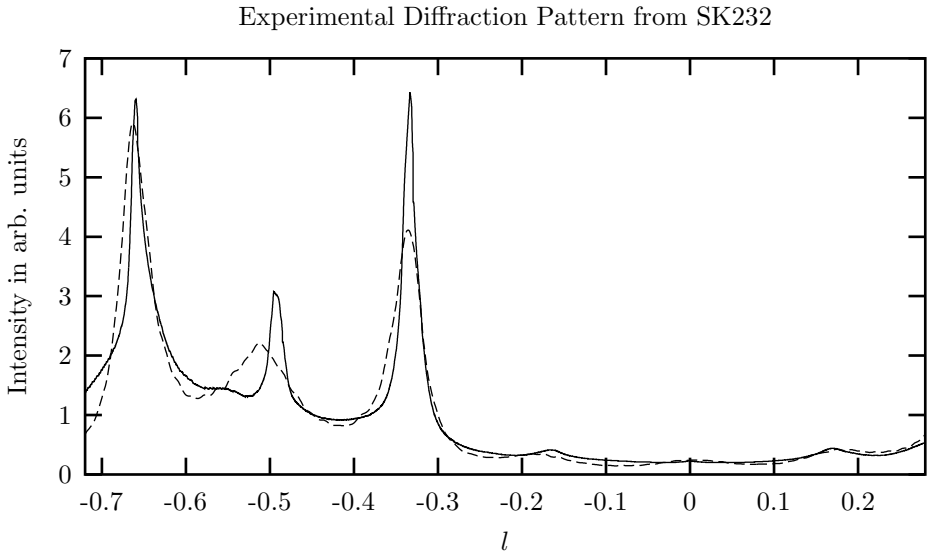

Figure 9.24: The diffraction pattern for Experimental Data SK232 and the $r = 3$ approximation.

9.5 Machine Reconstruction from Experimental Diffraction Pattern SK134

Figure 9.25 shows the diffraction pattern along the $10.l$ row of a ZnS crystal obtained from annealing a perfect 2H crystal at 300 C for one hour. The same diffraction pattern is shown in Figure 9.26 after correcting for $\mathcal{C}(l)$. Like the other diffraction patterns from this series, there is no strict periodicity in l but rather a slow degradation in the intensity as one moves from left to right across the spectrum. We use the same standard criteria to help select a unit interval to analyze, and we find that integration over the interval $0.04 \leq l$ 1.04 gives the figures of merit to be $Q_s(1) = -0.006$, $\gamma = -0.509$, and $\beta = 0.948$. The first few correlation functions are shown in Table 9.14 and a plot of $Q_s(n)$ vs. n is shown in Figure 9.27. The correlation functions show large and roughly equal values for the $Q_c(n), Q_r(n)$ for n odd, suggesting that this crystal retains much of its original 2H character. Indeed, examining the machine reconstruction results in Table 9.15 we see that arcs 0101 and 1010 together comprise 65.5.% of the probability weight for the total graph. The fitness for this machine reconstruction is $\mathcal{F} = 5.25 \times 10^{-5}$. We find a correlation length of $\lambda_q = 9.5 \pm 0.5$ over the first forty layers.

In most respects we see that this spectrum is much like SK230. Both began as perfect 2H crystals and have been annealed at 300 C for one hour. Both retain much of their much original 2H structure, (60.9% and 65.5% respectively) and have similar correlation lengths (8.4 and 9.5 respectively). The rate of entropy production per layer is similar ($h_\mu = 0.487$ and 0.501 respectively). This speaks well of the consistency of our technique.

Figure 9.29 shows a comparison of the $Q_s(n)$ obtained from the experimental diffraction pattern and that obtained from the $r = 3$ approximation. Differences become apparent around $n \approx 16$, with the approximate machine underestimating the correlations for larger n. This is behavior similar to that which we saw in SK230, except there the disagreement began at at $n \approx 10$. Sebastian and Krishna [61] attribute the disorder in both cases to deformation faulting, but in this present case they give a specific probability for this faulting, namely 5%. They arrive at this value by considering in some detail the change in the shape, placement, etc of the peaks. They examine several candidate

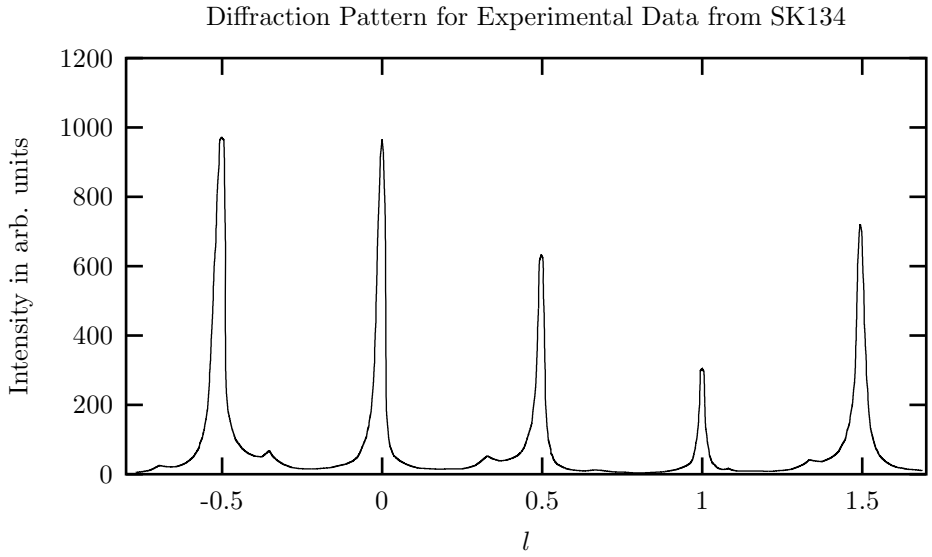

Figure 9.25: The uncorrected single crystal diffractometer reading along the $10.l$ reciprocal lattice row after annealing a perfect 2H ZnS crystal at 300 C for one hour. SK134 experimental data.

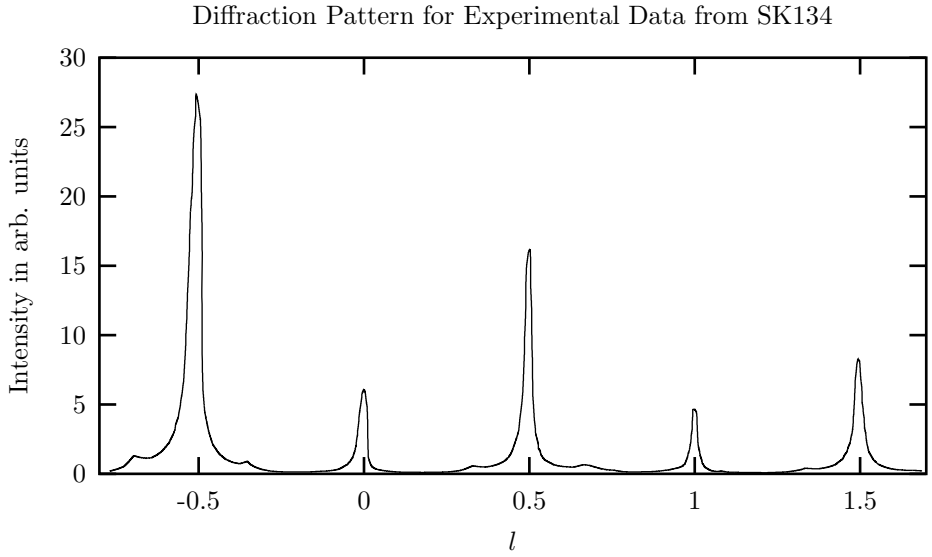

Figure 9.26: The diffraction pattern for Experimental Data SK134 corrected for $\mathcal{C}(l)$.

Table 9.14: The first few Qs for experimental data SK134 found by integration over the interval $0.04 \leq l \leq 1.04$.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.502	0.504	-0.006
2	0.077	0.133	0.790
3	0.475	0.408	0.117
4	0.093	0.197	0.710
5	0.478	0.402	0.120
6	0.116	0.211	0.673

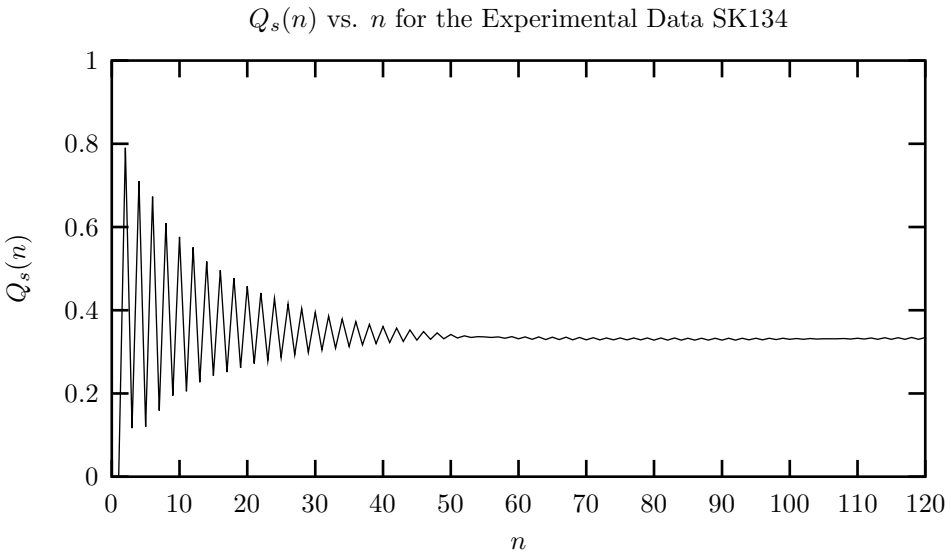

Figure 9.27: $Q_s(n)$ vs. n for Experimental Data SK134. We use $l_0 = 0.040$ and get a value of $\gamma = -0.509$ and $\beta = 0.948$. We find a correlation length of $\lambda_q = 9.5 \pm 0.5$ over the first forty n.

Table 9.15: Machine reconstruction results for the experimental diffraction pattern SK134. We find a fitness of $\mathcal{F} = 5.25 \times 10^{-5}$ for the $r = 3$ solution.

$r = 0$	$p(0)$	0.501	$r = 3$	$p(0000)$	0.039
	$p(1)$	0.499		$p(0001)$	0.004
				$p(0010)$	0.029
$r = 1$	$p(00)$	0.077		$p(0011)$	0.000
	$p(01)$	0.395		$p(0100)$	0.017
	$p(10)$	0.395		$p(0101)$	0.332
	$p(11)$	0.133		$p(0110)$	0.012
				$p(0111)$	0.036
$r = 2$	$p(000)$	0.042		$p(1000)$	0.004
	$p(001)$	0.035		$p(1001)$	0.027
	$p(010)$	0.337		$p(1010)$	0.322
	$p(011)$	0.058		$p(1011)$	0.048
	$p(100)$	0.035		$p(1100)$	0.016
	$p(101)$	0.360		$p(1101)$	0.039
	$p(110)$	0.058		$p(1110)$	0.036
	$p(111)$	0.075		$p(1111)$	0.041

Table 9.16: Computational results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK134.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	1.000	0.000	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.733	0.267	0.998	0.265	0.265
$r = 2$ Approximation	SFT	2	0.617	0.383	1.732	0.498	0.614
$r = 3$ Approximation	SFT	3	0.501	0.499	2.256	0.753	1.068
Fault Approximation	SFT	3	0.250	0.750	1.768	1.018	1.377

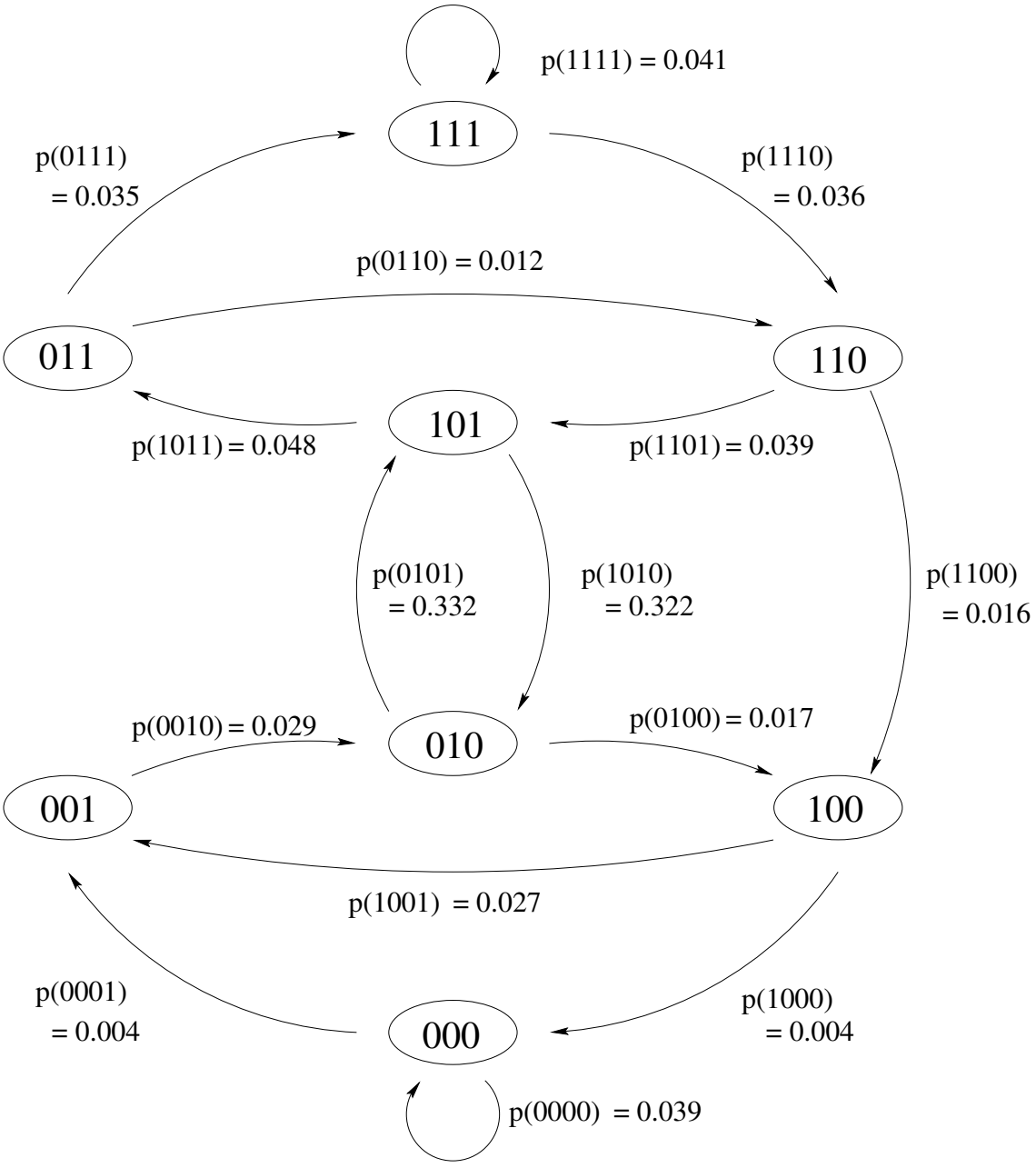

Figure 9.28: The $r = 3$ machine for diffraction pattern SK134.

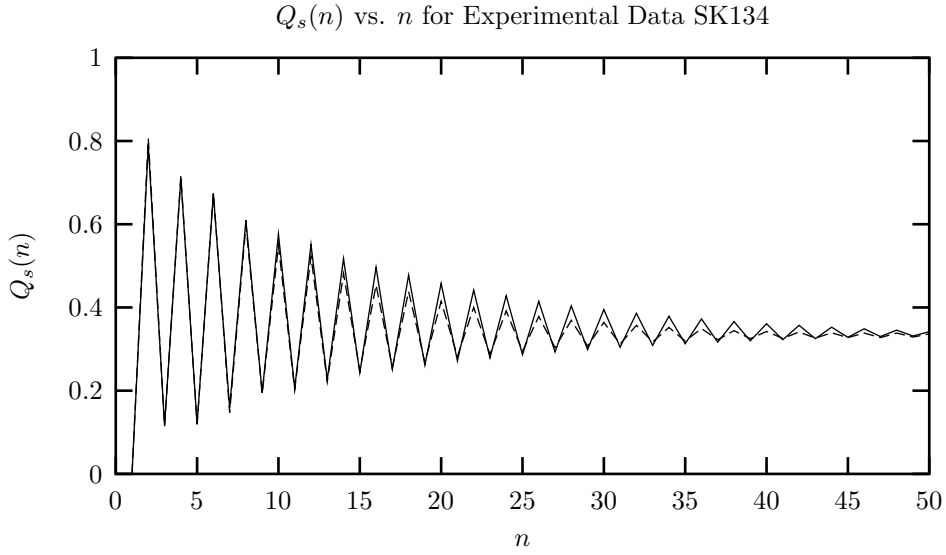

Figure 9.29: $Q_s(n)$ vs. n for Experimental Data SK134 (solid line) and the $r = 3$ approximation (dashed line).

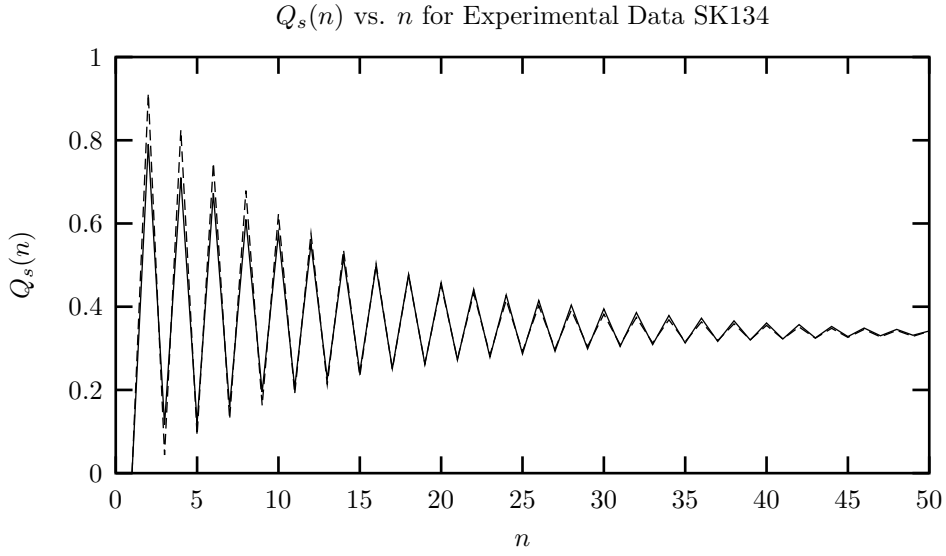

Figure 9.30: $Q_s(n)$ vs. n for Experimental Data SK134 (solid line) and the fault model (dashed line).

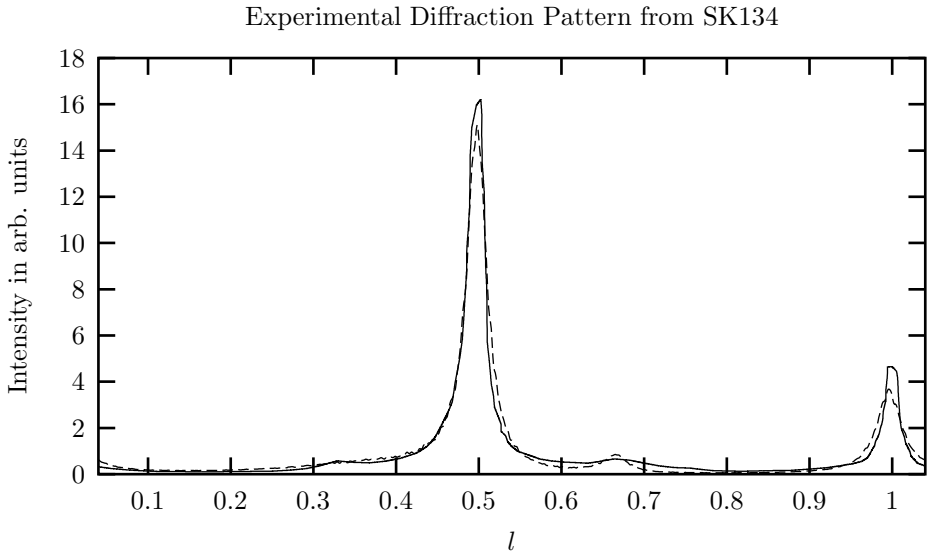

Figure 9.31: The diffraction pattern for Experimental Data SK134 (solid line) and the $r = 3$ approximation (dashed line).

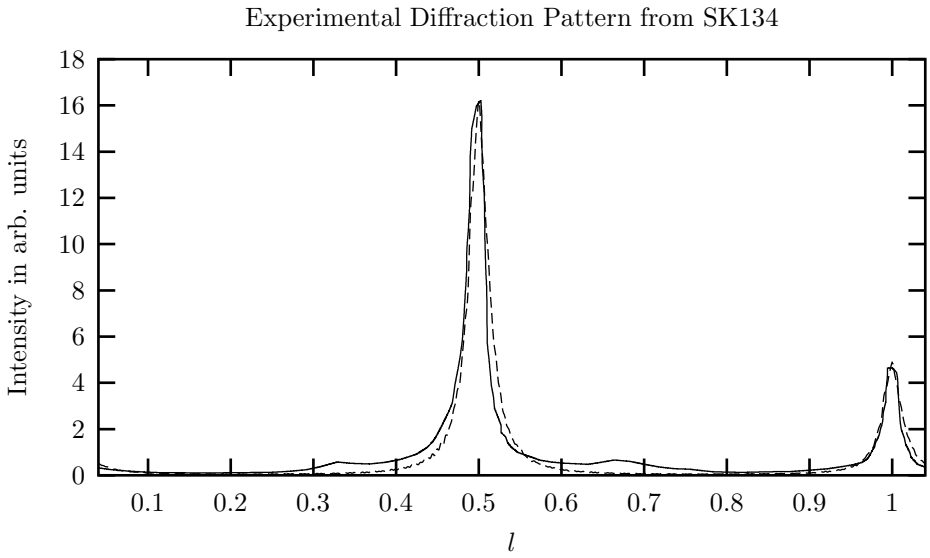

Figure 9.32: The diffraction pattern for Experimental Data SK134 (solid line) and the fault model (dashed line).

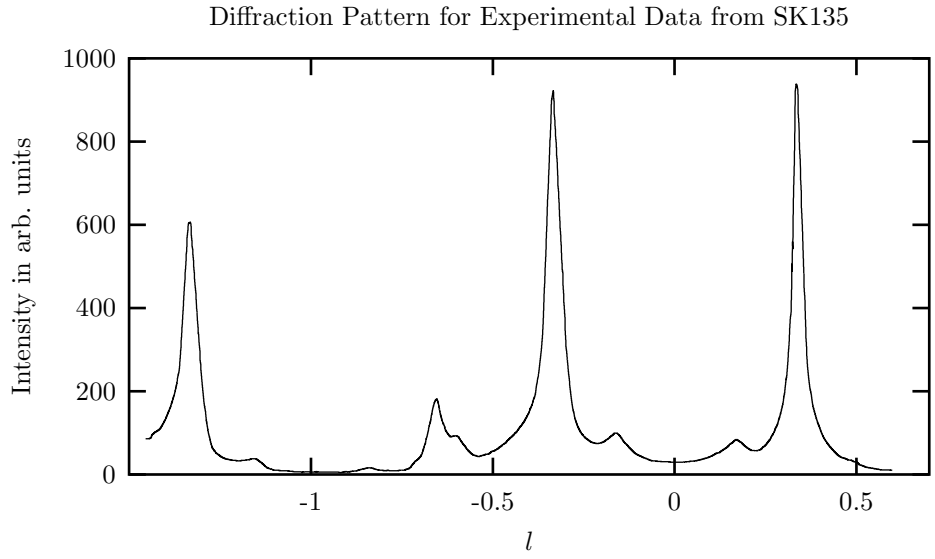

Figure 9.33: The uncorrected single crystal diffractometer reading along the 10.l reciprocal lattice row after annealing a 2H ZnS crystal at 500 C for one hour. SK135 experimental data.

faulting structures and find this deformation faulting most consistent with the observed spectrum. We can take their faulting mechanism and express it as an arc weighted de Bruijn graph (see §8.2.2) and find the $Q_s(n)$ to compare with experiment. This is shown in Figure 9.30. We notice that they reproduce the correlation functions well for $n \geq 15$, but overestimate them for smaller n. We can also compare the resulting diffraction patterns. Figure 9.31 shows the experimental diffraction pattern and our $r = 3$ approximation to it. We notice that the peak at $l = \frac{1}{2}$ seems well represented in shape and placement by the $r = 3$ approximation, except that it lacks a little in intensity at the strongest peak. The peak at $l = 1$ is reasonably well reproduced by the $r = 3$ approximation, but it is not as sharp as the experimental peak. The diffuse scattering is also in fair agreement, although clearly we miss the small rise in the spectrum at $l = \frac{2}{3}$. The enhancement of scattering in the vicinity of $l = \frac{m}{6}$ with m an integer is usually attributed to some 6H structure, which as we have already commented, can not be well represented by an $r = 3$ graph in the presence of 3C structure. Examining Figure 9.32, we see the comparison of the 5% deformation fault model with experiment. The model also reproduces the peaks well, doing an excellent job at $l = 1$. It does however, miss the small rise in intensity at $l = \frac{1}{3}$ and $\frac{2}{3}$. We should not be surprised at this, as Sebastian and Krishna have made no attempt to build in the necessary structure into their model to account for this scattering.

Lastly we can compare measures of computation between the two approximations as shown in Table 9.16. We see that the two models differ by a factor of two in the rate of entropy production per layer, with the faulting model having the lesser rate.

9.6 Machine Reconstruction from Experimental Diffraction Pattern SK135

The uncorrected diffraction pattern along the 10.l row for a ZnS crystal obtained by annealing a 2H crystal for one hour at 500 C is shown in Figure 9.33. The Sebastian and Krishna [61] report this

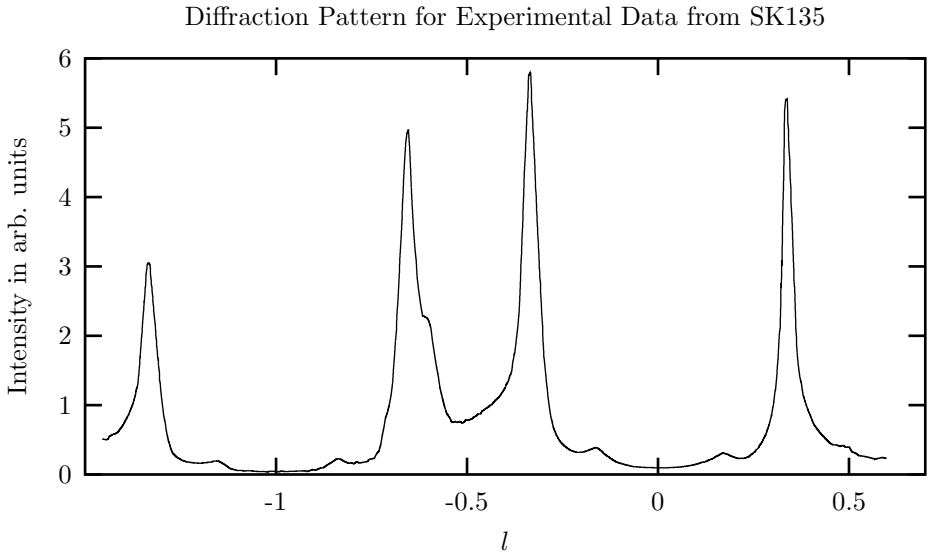

Figure 9.34: The corrected diffraction pattern for Experimental Data SK135.

Table 9.17: The first few Qs for experimental data SK135.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.475	0.523	0.002
2	0.376	0.395	0.229
3	0.228	0.139	0.633
4	0.401	0.480	0.120
5	0.355	0.354	0.291
6	0.283	0.198	0.519

as a disordered 3C crystal, due to the presence of asymmetrically broaden peaks and the absence of peak shifts. This same diffraction pattern corrected for $\mathcal{C}(l)$ is shown in Figure 9.34. Again there is no strict periodicity in the spectrum which we attribute to experimental error. Employing our criteria for selecting a suitable interval to analyze, we find that the interval $-0.80 \leq l \leq 0.20$ gives figures of merit to be $Q_s(1) = 0.002$, $\gamma = -0.498$ and $\beta = 0.932$. Using this interval, we find the correlation functions and the first few of them are shown in Table 9.17. Since $Q_c(3) \approx 0.23 \neq 0.5$, and $Q_r(3) \approx 0.14 \neq 0.5$, we expect that the original 2H structure has largely been eliminated. A plot of $Q_s(n)$ vs. n for this data is given in Figure 9.35. We find a correlation length for this crystal to be $\lambda_q = 4.4 \pm 0.7$.

Examining the $r = 3$ machine reconstruction results for this process in Table 9.18 we see that the antiferromagnetic arcs (0101 and 1010) have a relatively small combined weight of only about 4%. In fact, the probability weight for the 0101 arc is zero. So our original suspicion that the 2H structure has largely been eliminated proves correct. In its place we see large ferromagnetic arcs of nearly equal weight occupying a total of 54% of the weight on the graph. So we agree with Sebastian and Krishna that this is a disordered, twinned crystal. In addition to the 0101 arc, we also find the 1001 and 0010 arcs missing. Looking at the $r = 3$ graph in Figure 9.36, this implies that the twinning fault mechanism is important, as Sebastian and Krishna found, but also the remnant of the

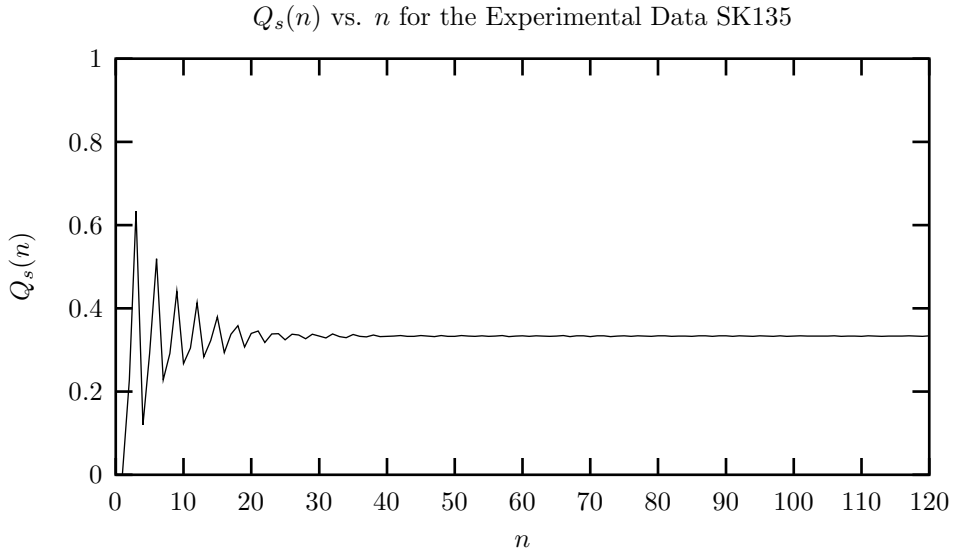

Figure 9.35: $Q_s(n)$ vs. n for Experimental Data SK135. We use $l_0 = -0.800$ and get a value of $\gamma = -0.498$ and $\beta = 0.932$. Using the first twenty n, we find a correlation length of $\lambda_q = 4.4 \pm 0.7$.

Table 9.18: Machine reconstruction results for the experimental diffraction pattern SK135. We find a fitness $\mathcal{F} = 4.3 \times 10^{-5}$.

$r = 0$	$p(0)$	0.524	$r = 3$	$p(0000)$	0.278
	$p(1)$	0.476		$p(0001)$	0.050
				$p(0010)$	0.000
$r = 1$	$p(00)$	0.376		$p(0011)$	0.048
	$p(01)$	0.114		$p(0100)$	0.037
	$p(10)$	0.114		$p(0101)$	0.000
	$p(11)$	0.395		$p(0110)$	0.027
				$p(0111)$	0.051
$r = 2$	$p(000)$	0.322		$p(1000)$	0.051
	$p(001)$	0.054		$p(1001)$	0.000
	$p(010)$	0.030		$p(1010)$	0.037
	$p(011)$	0.084		$p(1011)$	0.030
	$p(100)$	0.054		$p(1100)$	0.015
	$p(101)$	0.060		$p(1101)$	0.068
	$p(110)$	0.084		$p(1110)$	0.049
	$p(111)$	0.311		$p(1111)$	0.263

Table 9.19: Computational results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK135.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	0.998	0.002	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.775	0.225	0.999	0.224	0.224
$r = 2$ Approximation	SFT	2	0.727	0.273	1.775	0.320	0.367
$r = 3$ Approximation	SFT	3	0.590	0.410	2.478	0.705	1.112
Fault Approximation	SFT	1	0.529	0.471	1.000	0.471	0.471

1010 arc has some role. Instead of faulting ...1111|0000... where the vertical line indicates the fault plane, we see that the path ...1111|01000... has nearly twice as much probability weight associated with it. In the lower portion of the graph, we see that twinned faulting is largely responsible for the $(0)^*$ fcc cycle converting to the $(1)^*$ fcc cycle and we also observe that double deformation faulting important. Perhaps it is interesting to mention that, while a modular layer of ZnS has spin inversion symmetry [72] and thus the one-dimensional Hamiltonian describing the energetics of the stacking is also spin invariant, in general these graphs are not spin inverse invariant. That is, the probability of seeing a word and its spin inverse is not the same. By spin inverse, of course we mean just flipping all the spins in a word, ie $1101 \rightarrow 0010$. There is of course no reason why we should expect spin inversion; after all, then one could never have a crystal of purely one fcc structure or the other. We note that the fault picture always assumes spin inversion symmetry. Sebastian and Krishna (1994) attribute the faulting to the mechanism of twinned faulting and assign a probability of 12% for this to occur.

Examining the $Q_s(n)$ derived from experiment with those found from the $r = 3$ machine in Figure 9.37, we find reasonable agreement up to $n \approx 15$, and slight deviations thereafter. Looking at the $Q_s(n)$ found from the faulting picture and comparing it with experimentally derived ones, Figure 9.38, we find that the fault picture reproduces the general form of the plot, but overestimates the magnitude of the oscillations. We can further examine the diffraction patterns. In Figure 9.39, the diffraction pattern found from the $r = 3$ approximation is compared with experiment. We see a reasonable fit, except perhaps at a shoulder in the experimental spectrum at $l = -0.6$ and the small rise at $l = -0.16$. We can speculate that there is some minor competition between the 3C and 6H cycles that is not being well modeled here. Comparison of the fault derived diffraction pattern with that from experiment, Figure 9.40, reveals good agreement with the peak at $l = -\frac{1}{3}$ and fair agreement with the one at $l = -\frac{2}{3}$. However, the diffuse scattering between peaks is not at all well represented. Additionally, the small rise in diffuse scattering at $l = \pm\frac{1}{6}$ is likewise absent in the fault model diffraction pattern.

The computational results for the various machine approximations to the process generating the diffraction pattern SK135 are shown in Table 9.19. While the fault approximation has a similar but slightly lower rate of entropy production per layer as compared to the $r = 3$ approximation, the other measures of computation are uniformly lower. This seems to indicate that the fault approximation is missing some important computational aspects of the stacking.

9.7 Machine Reconstruction from Experimental Diffraction Pattern SK137

The last experimental spectrum we analyze is show in Figure 9.41. The intensity versus l in reciprocal space for the $10.l$ row of an as grown crystal is shown. This same diffraction pattern corrected for $\mathcal{C}(l)$ is shown in Figure 9.42. We see that the spectrum again is not strictly periodic in l, so we need to

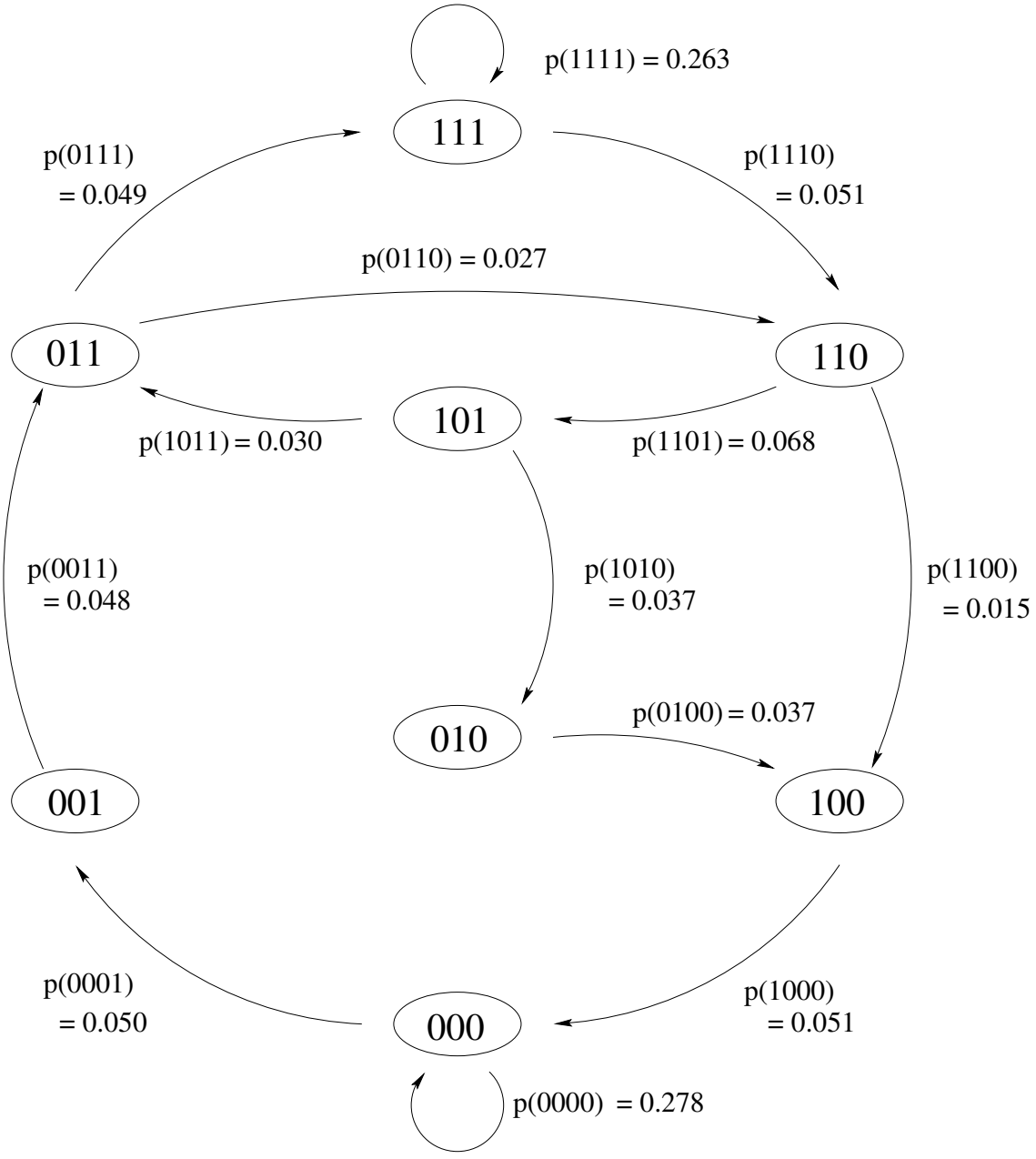

Figure 9.36: The $r = 3$ machine for diffraction pattern SK135.

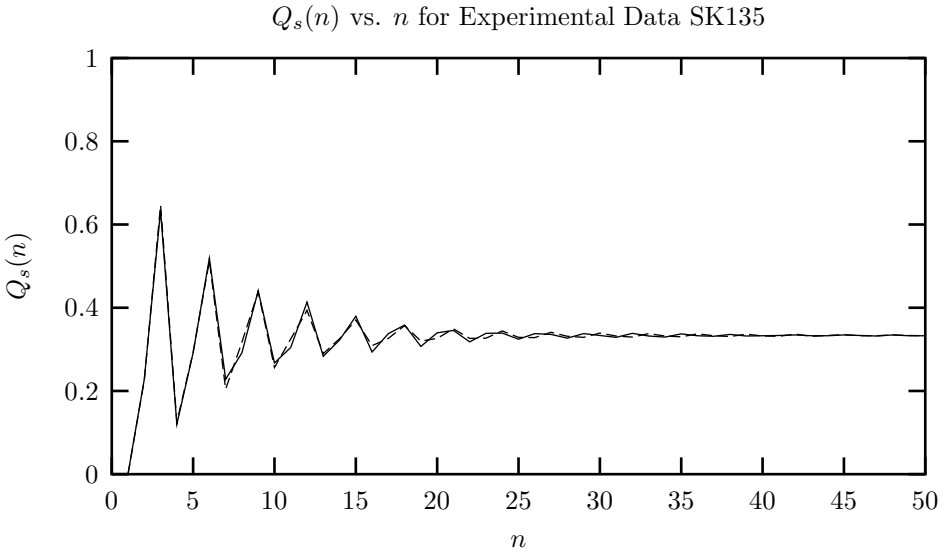

Figure 9.37: $Q_s(n)$ vs. n for Experimental Data SK135 (solid line) and the $r = 3$ approximation (dashed line).

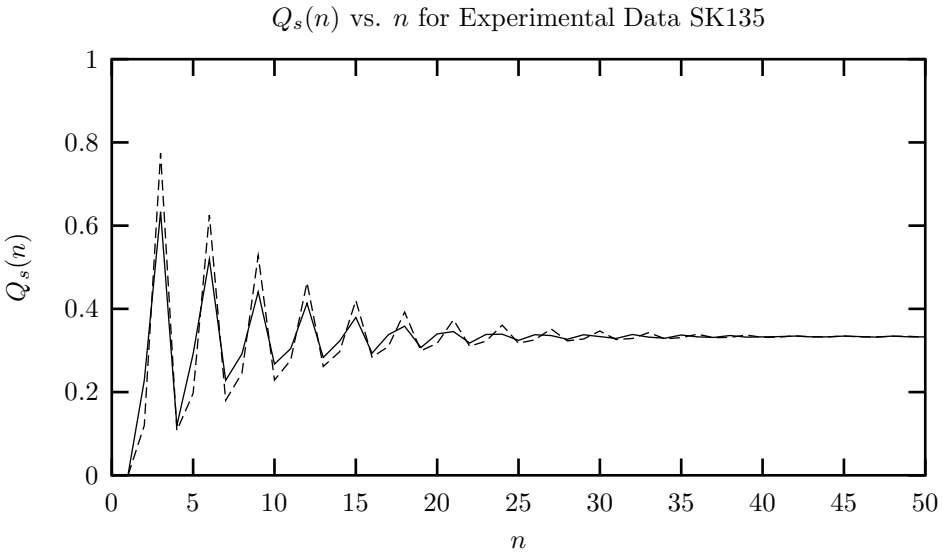

Figure 9.38: $Q_s(n)$ vs. n for Experimental Data SK135 (solid line) and the fault approximation (dashed line).

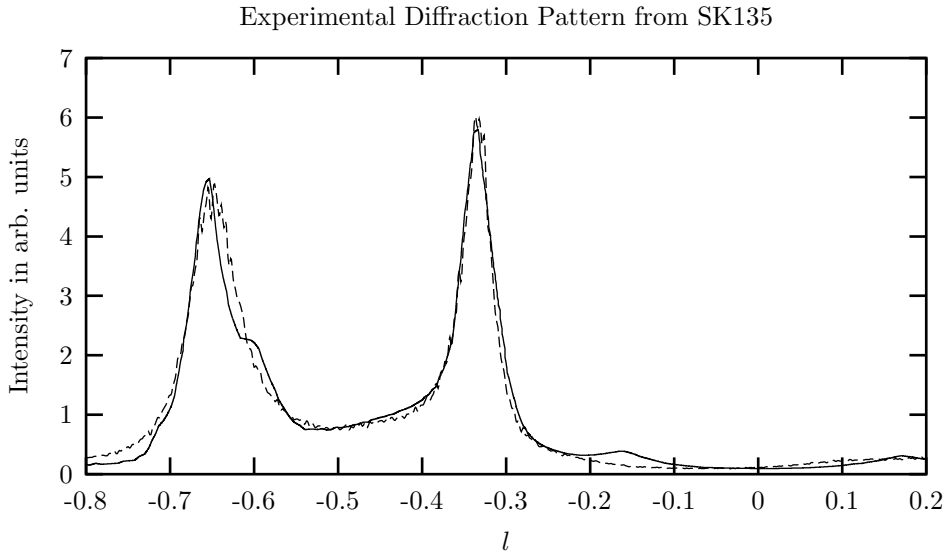

Figure 9.39: The diffraction pattern for Experimental Data SK135 (solid line) and the $r = 3$ approximation (dashed line).

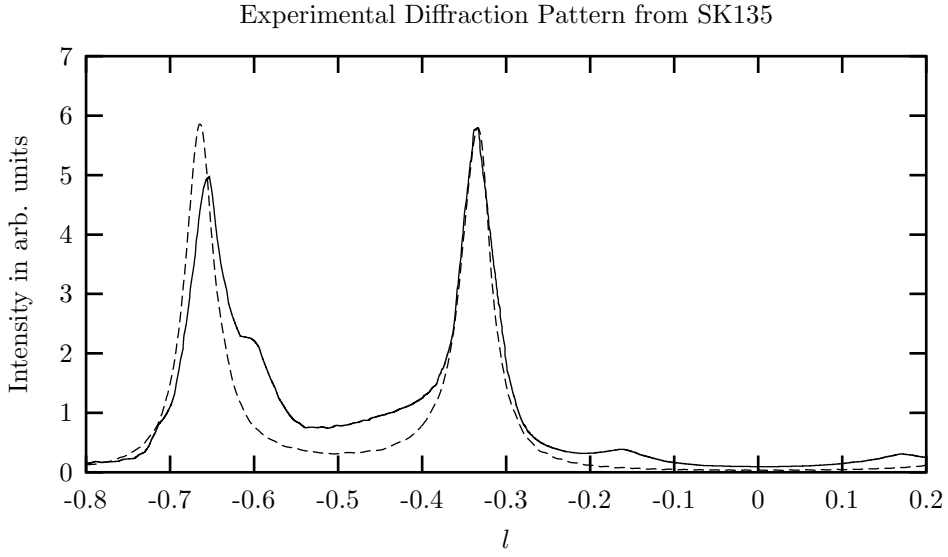

Figure 9.40: The diffraction pattern for Experimental Data SK135 (solid line) and the fault approximation (dashed line).

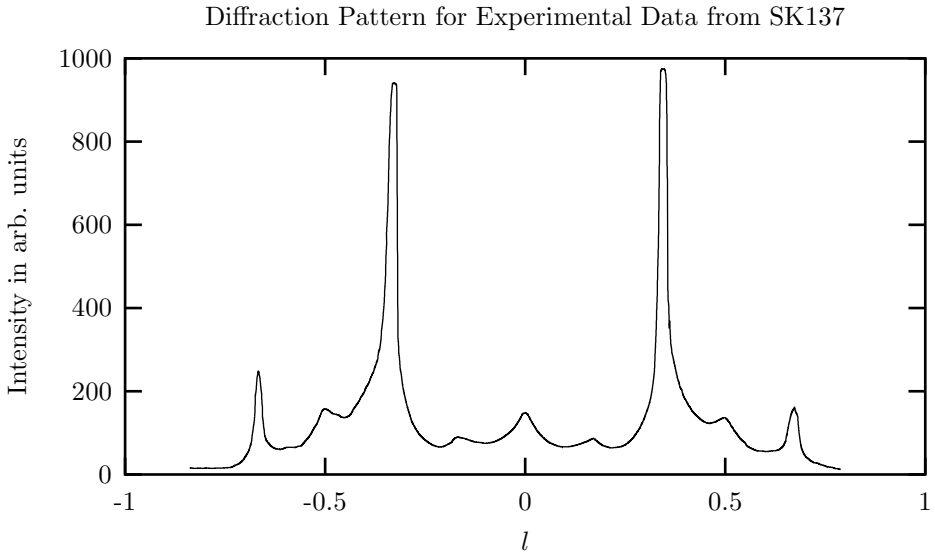

Figure 9.41: he uncorrected diffraction pattern for Experimental Data SK137.

Table 9.20: The first few Qs for experimental data SK137.

n	$Q_c(n)$	$Q_r(n)$	$Q_s(n)$
1	0.484	0.506	0.010
2	0.338	0.324	0.339
3	0.230	0.245	0.525
4	0.386	0.386	0.228
5	0.377	0.367	0.255
6	0.231	0.246	0.523

choose an interval which gives the best figures of merit. We find this interval to be $-0.80 \leq l \leq 0.20$, and the figures of merit turn out to be $Q_s(1) = 0.010$, $\gamma = -0.485$ and $\beta = 0.982$. The first few correlation functions are tabulated in Table 9.20 and a graph of $Q_s(n)$ vs. n is shown in Table 9.43. We find a correlation length of $\lambda_q = 12 \pm 3$.

Machine reconstruction results can be found in Table 9.21 and the the $r = 3$ machine is displayed in Figure 9.44. Sebastian and Krishna [61] report this as a disordered 3C crystal, which we can confirm, as the ferromagnetic arcs consume about 44% of the weight in the graph. The faulting mechanism is not so clear. There is only one forbidden word, 0011, and the remaining words, save 1101, all appear at about the 3% to 9% level. We will not attempt a fault analysis since it certainly is not unique and most faulting mechanisms seem to play at least some role. Sebastian and Krishna [61] report that this crystal is well described by a random distribution of twin faults with a 6.8% of occurrence.

A comparison of the $Q_s(n)$ derived from experiment and the $Q_s(n)$ obtained from the $r = 3$ reconstructed machine is shown in Figure 9.45. There is reasonable agreement until about $n \approx 10$, after which the reconstructed results fall off to the asymptotic value too soon. For n between 10 and 40, the $Q_s(n)$ from the $r = 3$ approximate machine only weakly echo the oscillations in $Q_s(n)$ from experiment. This provides a hint that there is significant computation missing at $r = 3$. The fault

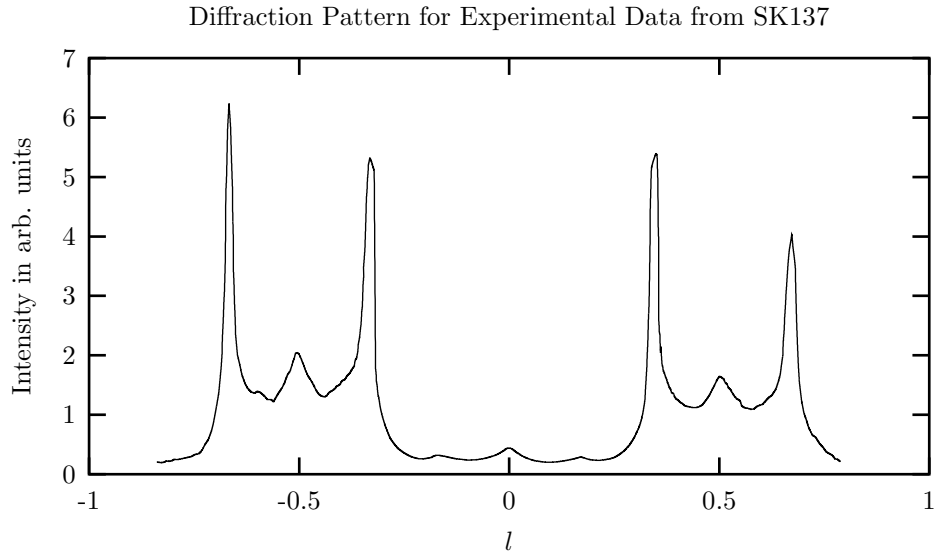

Figure 9.42: The corrected diffraction pattern for Experimental Data SK137.

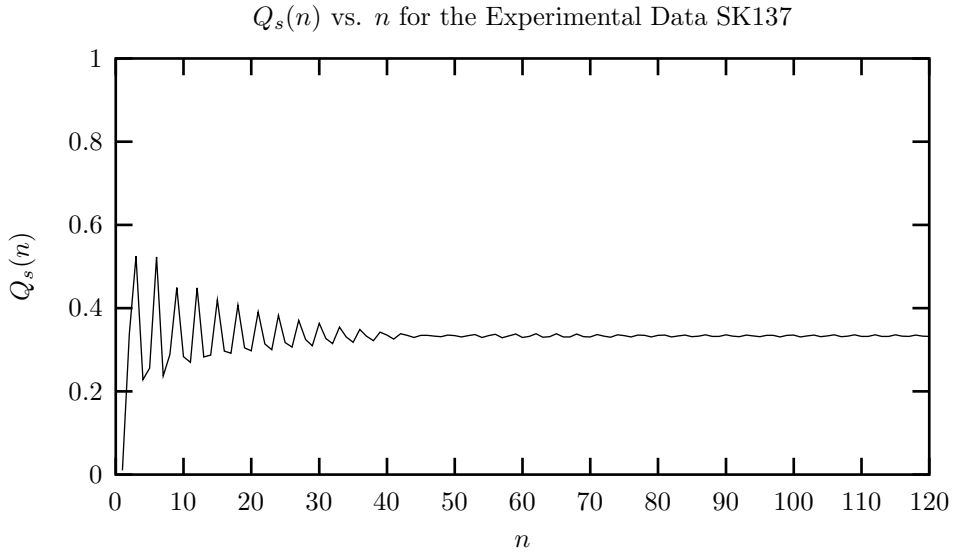

Figure 9.43: $Q_s(n)$ vs. n for Experimental Data SK137. We use $l_0 = -0.800$ and get a value of $\gamma = -0.485$ and $\beta = 0.982$. We find a correlation length of $\lambda_q = 12 \pm 3$.

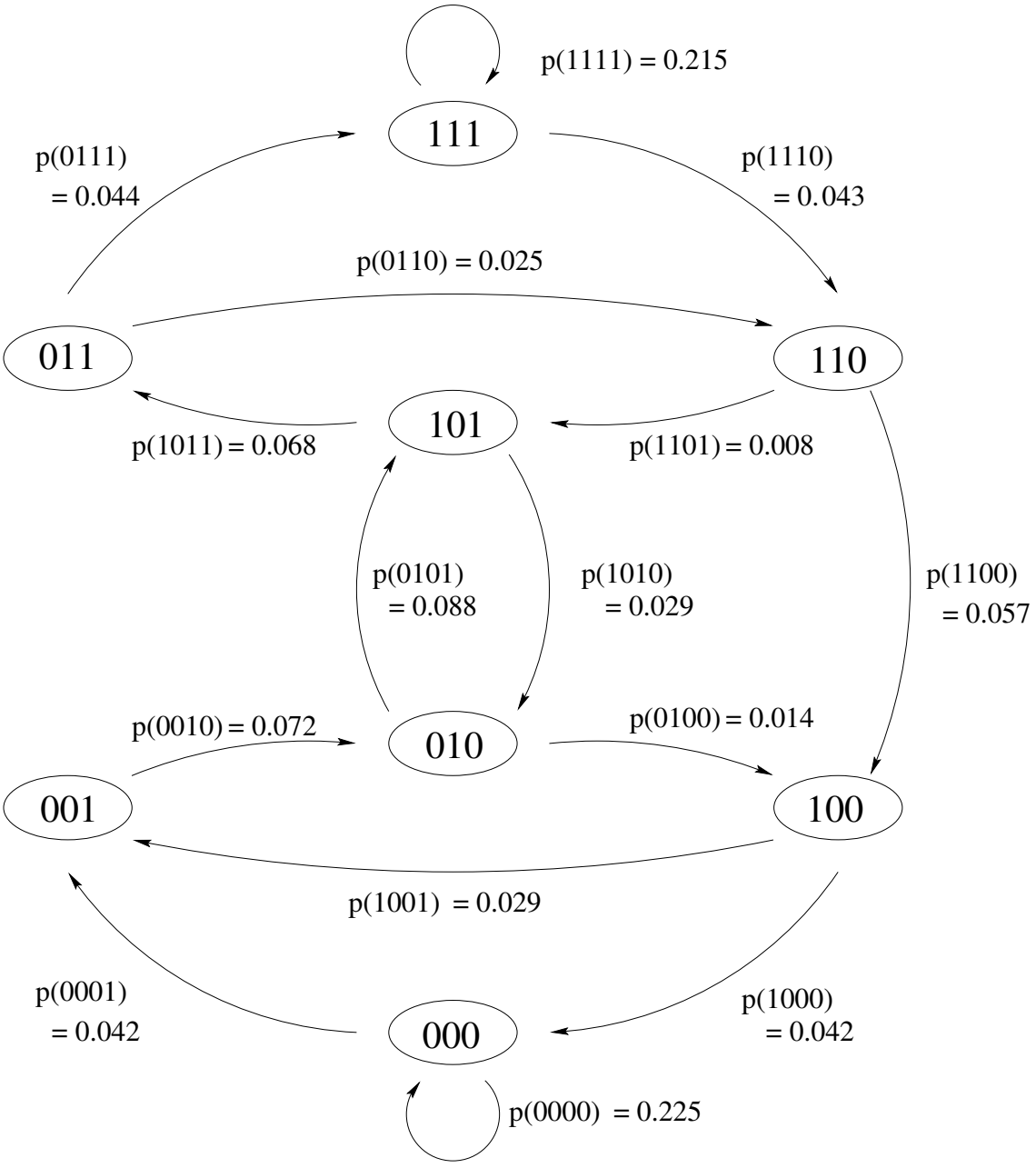

Figure 9.44: The $r = 3$ machine for diffraction pattern SK137.

Table 9.21: Machine reconstruction results for the experimental diffraction pattern SK137. We find a fitness $\mathcal{F} = 8.2 \times 10^{-6}$.

$r = 0$	$p(0)$	0.511	$r = 3$	$p(0000)$	0.225
	$p(1)$	0.489		$p(0001)$	0.042
				$p(0010)$	0.072
$r = 1$	$p(00)$	0.338		$p(0011)$	0.000
	$p(01)$	0.170		$p(0100)$	0.014
	$p(10)$	0.170		$p(0101)$	0.088
	$p(11)$	0.324		$p(0110)$	0.025
				$p(0111)$	0.044
$r = 2$	$p(000)$	0.267		$p(1000)$	0.042
	$p(001)$	0.071		$p(1001)$	0.029
	$p(010)$	0.103		$p(1010)$	0.029
	$p(011)$	0.066		$p(1011)$	0.068
	$p(100)$	0.071		$p(1100)$	0.057
	$p(101)$	0.098		$p(1101)$	0.008
	$p(110)$	0.066		$p(1110)$	0.043
	$p(111)$	0.258		$p(1111)$	0.215

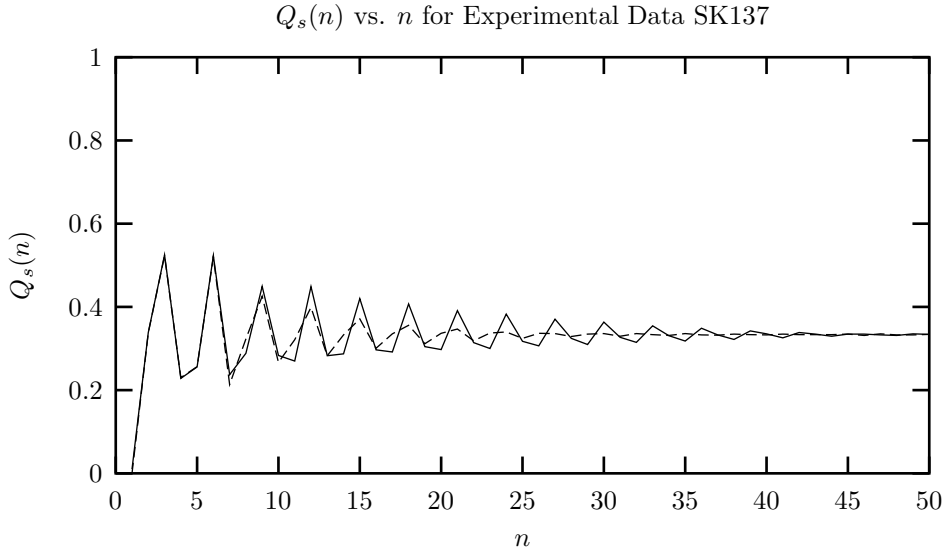

Figure 9.45: $Q_s(n)$ vs. n for Experimental Data SK137 (solid line) and the $r = 3$ approximation (dashed line).

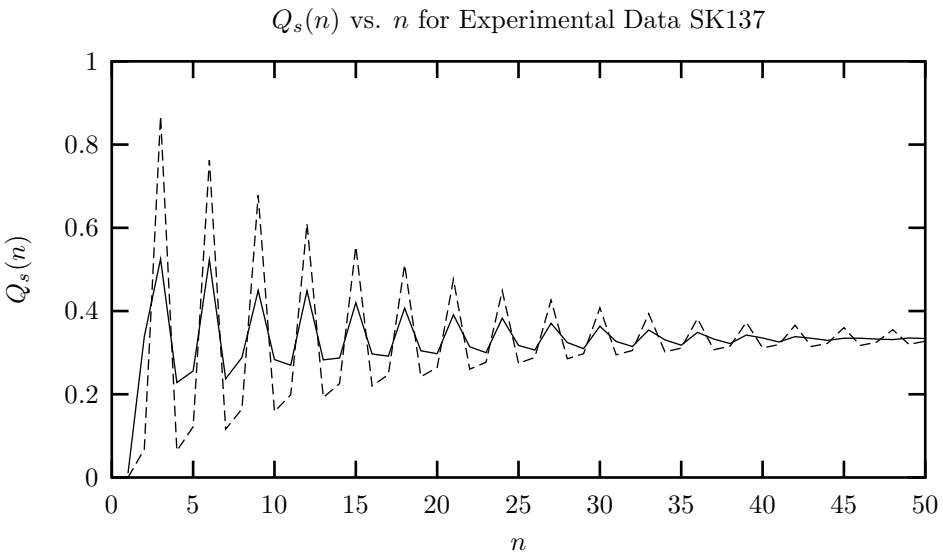

Figure 9.46: $Q_s(n)$ vs. n for Experimental Data SK137 (solid line) and the fault approximation (dashed line).

Table 9.22: Computational results for the $r = 0, 1, 2$ and 3 approximations to experimental data SK137.

System	Language Type	r	h_μ	**G**	C_μ	**E**	**T**
$r = 0$ Approximation	SFT	0	1.000	0.000	0.000	0.000	0.000
$r = 1$ Approximation	SFT	1	0.924	0.076	1.000	0.076	0.076
$r = 2$ Approximation	SFT	2	0.816	0.184	1.922	0.291	0.398
$r = 3$ Approximation	SFT	3	0.651	0.349	2.744	0.792	1.408
Fault Approximation	SFT	1	0.359	0.641	1.000	0.641	0.641

model with a 6.8% twinned fault probability fares worse. It over estimates the magnitude of the oscillations in the $Q_s(n)$ significantly for $n \leq 50$. This simple model (it can after all be expressed as an $r = 1$ machine), seems to insert too much correlation into the $Q_s(n)$. A comparison of diffraction patterns is also possible. For the $r = 3$ reconstruction, Figure 9.47 compares the diffraction pattern with the experimental one. The diffuse scattering is well represented, but the two peaks at $l = -\frac{2}{3}$ and $-\frac{1}{3}$ lack sharpness. Comparing the diffraction pattern generated from the fault model with experiment as shown in Figure 9.48, we see that the peaks are reasonably well reproduced but the diffuse scattering is almost completely absent in the fault model. The rise in intensity at $l = -\frac{1}{2}$ is also missing in the fault model.

Table 9.22 shows the computational quantities for the various r approximations and the fault model. The fault model misses much of the complexity present, as measured by C_μ and **E**. It is likely that neither the $r = 3$ approximation or the fault approximation is modeling the computation present very well, but clearly the fault picture falls far shorter of representing reality here.

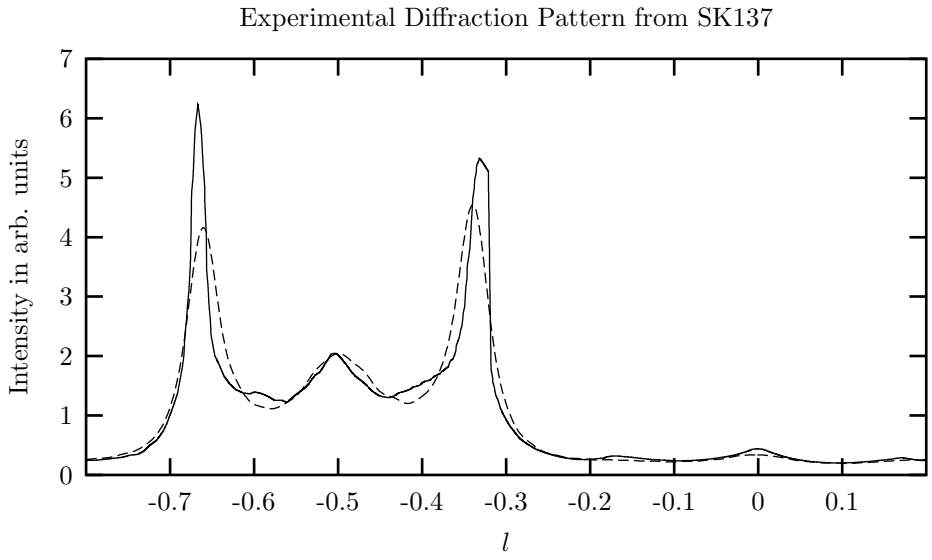

Figure 9.47: The diffraction pattern for Experimental Data SK137 and the $r = 3$ approximation.

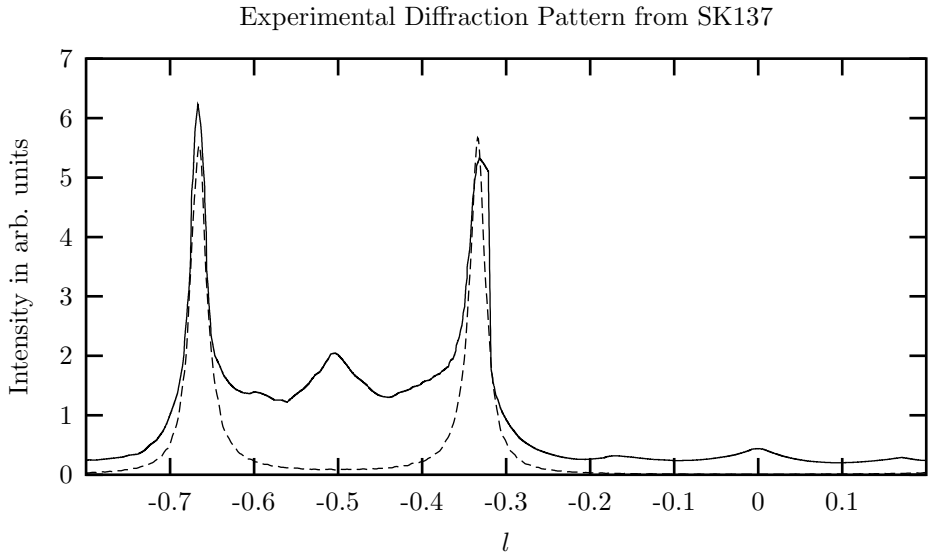

Figure 9.48: The diffraction pattern for Experimental Data SK137 and the fault approximation.

9.8 Configurational Energies of Polytypes

Now that we have a statistical model for the stacking of the layers, we can calculate physical parameters that depend on this stacking. One such quantity amenable to calculation is the difference in the configurational energies of the particular polytypes. Numerical calculations find that the configurational energy depend only the nearest and the next nearest neighbors in the stacking arrangement. Engel and Needs [22] have done a first-principles pseudopotential calculation of the total energy of five ZnS polytypes, from which they can determine the strength of the interactions up to the third nearest layer. The most general expression possible for inter-layer interactions up the third nearest neighbors is given by [65]

$$E = E_0 - J_1 \sum_i s_i s_{i+1} - J_2 \sum_i s_i s_{i+2} - J_3 \sum_i s_i s_{i+3} - K \sum_i s_i s_{i+1} s_{i+2} s_{i+3}. \tag{9.1}$$

Terms with an odd number of spins do not appear because of symmetry considerations. We take the $s_i = \pm 1$ here. Engel and Needs [22] have found that

$$\begin{array}{ll} J_1 = 0.00187 \text{ eV per ZnS pair}, & J_3 \text{ negligible}, \\ J_2 = -0.00008 \text{ eV per ZnS pair}, & K \text{ negligible}. \end{array}$$

Let us rewrite equation 9.1 in terms of the energy per ZnS pair, and take the zero of the energy such that $E_0 = 0$. We have then,

$$\tilde{E} = -J_1 \langle s_i s_{i+1} \rangle - J_2 \langle s_i s_{i+2} \rangle, \tag{9.2}$$

where $\langle \ldots \rangle$ means the expectation value of ' $\ldots$ '. We can find the expectation values directly from word probabilities,

$$\langle s_i s_{i+1} \rangle = p(11) + p(00) - 2p(01), \tag{9.3}$$

$$\langle s_i s_{i+2} \rangle = p(111) + p(101) + p(000) + p(010) - 2p(110) - 2p(100). \tag{9.4}$$

We show the configurational energy in terms of meV per ZnS pair in Table 9.23 for both the crystalline structures considered by Engels and Needs as well as the seven disordered polytype structures on which we have performed machine reconstruction. We see that the two of the disordered samples, SK232 and SK135, have energies not too different from the lowest energy crystal, 3C. As we might expect from the relative magnitudes of J_1 and J_2, the contribution from the J_1 term completely dominates the energy.

This is one example of a quantity that can be calculated once the statistical nature of the stacking is known. It is obviously desirable to calculate other measurable, physical parameters from the reconstructed ϵ-machine. Examples of such quantities would be the specific heat and the transmission of electrons through a disordered, layered sample.

Table 9.23: Relative configurational energies of experimental polytypes and some crystalline polytypes. In the last column we give the history of each sample, where PC stands for perfect crystal, AG as grown and D disordered. The configurational energies for the five crystalline polytypes were calculated from first principles by Engels and Needs [22]. We have not encountered the 6Ha polytype structure before, its relative stacking sequence is given by 110010. We use the energy coupling constants calculated by Engels and Needs, J_1 and J_2, along with our reconstructed ϵ-machine for the disordered process to find the configurational energy of the disordered structures.

System	$\langle s_i s_{i+1} \rangle$	$\langle s_i s_{i+2} \rangle$	$\tilde{E}$ in meV/ZnS pair	Sample History
2H	-1.000	1.000	1.95	PC
SK229	-0.795	0.856	1.56	AG 2H
SK134	-0.580	0.628	1.13	D 2H, 300 C for 1h
SK230	-0.519	0.688	1.03	PC 2H, 300 C for 1h
6Ha	-0.333	-0.333	0.60	PC
4H	0.000	-1.000	-0.08	PC
SK231	0.255	0.428	-0.44	2H, 600 C for 1h
SK137	0.324	0.452	-0.57	AG D 3C
6H	0.333	-0.333	-0.65	PC
SK232	0.389	0.280	-0.71	2H, 700 C for 1h
SK135	0.563	0.447	-1.02	2H, 500 C for 1h
3C	1.000	1.000	-1.79	PC

Chapter 10

Finale

Understanding disorder in three dimensions is a difficult task [76]. Physicists often resort to descriptions that use correlation information, but this usually does not provide insight into the underlying mechanisms of the disorder or indeed, even provide a detailed picture of the disorder. Recent progress has been made, however, in the analysis and description of disorder in one-dimensional systems. These new methods provide a detailed way to describe, classify and quantify systems both simple, such as periodic structure, and complex. We can meaningfully discuss the entropy density in the one-dimensional system, as well as specify quantities that describe computation and memory. There have, however, been relatively few applications of these methods to physical systems. We bridge this gap for the case of polytypism. Since polytypes can be treated as a one-dimensional spin system and also have interesting physical properties not yet explained, they provide an ideal system to explore the usefulness of these theories. From the physical side, the phenomenon of polytypism has been known for nearly ninety years, but continues to defy theoretical explanation or, even in the case of disordered sequences, adequate description - until now. We have demonstrated that one can describe both the ordered and disordered stacking of polytypes using ϵ-machines. Further, we have introduced a new technique that takes as input spectral information instead of the specific sequences for machine reconstruction. Since data from physical systems often comes in this form, we are hopeful that this will prove useful for pattern discovery elsewhere. In the process of applying these new ideas to the old mystery of polytypism, we have given a critical examination of the concept of faulting in polytypes. We have shown that, in the case where the faulting picture is meaningful, ϵ-machines provide an equivalent description, as in SK229. But we have discredited the general fault picture as inadequate to describe the disorder in polytypes, and have shown that ϵ-machines provide a unique description, at each range of memory r, of the underlying architecture. Additionally, we demonstrate the usefulness of the ϵ-machine description, in that we can calculate physical quantities from it, such as the range of interaction between modular layers and the stacking energy. We could also consider other quantities, such as the local density of phonon states, the specific heat and the transmission coefficient though layered structures [35]. Indeed, any quantity that can be expressed in statistical terms is amenable to calculation once a model of the underlying process (the ϵ-machine) is known. For those that are not, a specific sample of the language can be used to numerically calculate quantities. This has important applications in terms of, say, heterostructures. There one is interested in how physical quantities change as the stacking rules are changed. Indeed, such heterostructures can be made artificially in the laboratory and understanding how the band gap changes with the stacking rules has important consequences for semiconductor technology. It is known that the bad gap in ZnS is sensitive to the period of the polytype [61] and that there is an anomalous photovoltaic effect of up to several hundred volts/cm found only in disordered ZnS [21] [57] [66].

Let us now discuss our machine reconstruction results for experimental data. From the ϵ-machine reconstruction procedure, we see that $r = 3$ captures most of the structure present. We can not rule

out the possibility that sofic processes play a role, we can only say that 3^{rd}-order Markov processes seem adequate to explain much of the structure seeing we are seeing in the diffraction patterns. However, several of the diffraction patterns, such as SK232 and SK137, suggest that we are still missing some of the structure in our reconstructed graphs. There is theoretical and experimental evidence to believe that new structure awaits discovery at $r = 4$. Indeed, coexistence of 3C and 6H polytypes in a single sample requires a graph of a minimum memory of $r = 4$. As a future project, this needs investigation. The experimental situation is still too murky to allow us to distinguish between sofic and finite memory processes, if both produce diffuse spectra. We hope this will be a call to experimentalists to take careful measurements of both the Bragg peaks and the diffuse background intensity in future polytype diffraction experiments. Let us again emphasize that there is valuable information in the diffuse scattering. We have shown in §7.2 that only fitting the peaks in the spectrum can lead one astray in discovering the underlying pattern. Our technique uses the whole spectrum over a unit interval in l, so that we discard no information in the spectrum. We have also shown how one can quantify and describe the quality of the data over a unit interval in l and use this to determine the best interval for machine reconstruction. We find the experimental data to be fairly corrupted, in agreement with the experimentalists' assessment. Since this is so, we feel that this work, even though at present providing the most comprehensive description of disorder in polytypic ZnS available, should be treated as exploratory. Again, better data is demanded for a more thorough treatment.

So we are finding, perhaps not surprisingly, that the 'language' these little ZnS crystals speak is of a particularly simple kind. Glancing back at the Chomsky hierarchy, we see that ϵ-machine reconstruction as presented here involved only the lowest of the four tiers of languages. (Although, one can define machine reconstruction for higher languages too. See [15].) And within this tier, the simplest languages seem to suffice. Indeed, it is difficult to imagine a less sophisticated family of processes in which we could look. Our analysis clearly implies that the effective range of interaction in ZnS must extend at least three modular layers in contrast to the calculations of Engels and Needs [22]. We can offer no explanation for this discrepancy, but instead hope that having a better description of the disorder present can aid future researchers explaining why this is so.

Finally, we return to our original question and ask "Where are the atoms?" We have spent some time and effort in trying to answer this question and it is fair to ask if we have succeeded. From diffraction data, we have introduced a technique for both discovering and describing the stacking rules for close-packed polytypic materials. For disordered crystals, this description is necessarily statistical. As such, we never were able to determine the precise location of any modular layer, let alone the atoms. Have we failed? It is known that without making the assumption of crystallinity, the power spectrum loses information so that the exact structure can not be uniquely determined. So it is not surprising that we have not found a particular stacking sequence for any of these polytypes. It therefore seems that we do not know where any of the layers are. So in a strict sense, we have failed. But let us pause and contemplate what we have learned. We know the correlation functions, which give a statistical answer to the placement of the atoms, and we have a model of the underlying process. But suppose we could have found the absolute position of each of the, say, one million layers, then our conclusion would be a list one million bits long. Would this be useful? Could we then say we understand the stacking in any sense? Certainly we would have described it unambiguously, but what would we have gained? No, we would want to express this answer in a meaningful way such that we can gain insight. What we would want, is a model of the process. And that is what we have. So I suggest that we have succeeded, and better perhaps than our initial strivings would lead us to believe. We have found a model that gives the most compact possible description of the stacking in a statistical sense, and have inferred that model directly from experimental data with minimal assumptions. This is the best one can hope for. And from this model, the ϵ-machine, many physical properties are in principle forthcoming and it is only a matter of performing the calculations.

Bibliography

Bibliography

[1] Ash, Robert B. (1990). *Information Theory*. New York: Dover Publications, Incorporated.

[2] Attard, Phil, Owen G. Jepps and Stjepan Marčelja (1997). "Information Content of Signals using Correlation Function expansions of the Entropy." *Physical Review E*, **56**: 4052-4067.

[3] Axel, Françoise and Hikaru Terauchi (1991). "High-Resolution X-Ray-Diffraction Spectra of Thue-Morse GaAs-AlAs Heterostructures: Towards a Novel Description of Disorder." *Physical Review Letters*, **66**: 2223-2226.

[4] Azároff, Leonid V. (1968). *Elements of X-Ray Crystallography*. New York: McGraw-Hill Book Company.

[5] Azároff, Leonid V., Roy Kaplow, N. Kato, Richard J. Weiss, A.J.C. Wilson and R.A. Young (1974). *X-Ray Diffraction*. New York: McGraw-Hill, Incorporated.

[6] Badii, Remo and Antonio Politi (1997). *Complexity: Hierarchical Structures and Scaling and Physics*, volume 6 of *Cambridge Nonlinear Science Series*. Cambridge: Cambridge University Press.

[7] Baumhauer, H. (1912). "Uber die Kristalle des Carborundums." *Zeitschrift fur Kristallographie*, **50**: 33-39.

[8] Beck, Christian and Friedrich Schlögl (1993). *Thermodynamics of Chaotic Systems*, volume 4 of *Cambridge Nonlinear Science Series*. Cambridge: Cambridge University Press.

[9] Berliner, R. and S.A. Werner (1986). "Effects of stacking faults on diffraction: The structure of lithium metal." *Physical Review B*, **34**: 3586-3603.

[10] Buerger, M.J. (1942). *X-Ray Crystallography: An Introduction to the Investigation of Crystals by Their Diffraction of Monochromatic X-Radiation* New York: John Wiley & Sons, Incorporated.

[11] Burden, Richard L. and J. Douglas Faires (1989). *Numerical Analysis, Fourth Edition* Boston: PWS-Kent Publishing.

[12] Canright, Geoffrey S. and Greg Watson (1996). "Disordered Ground States for Classical Discrete-State Problems in One Dimension." *Journal of Statistical Physics*, **84**: 1095-1131.

[13] Clarke, Richard W., Mervyn P. Freeman and Nicholas W. Watkins (2003). Application of computational mechanics to the analysis of natural data: An example in geomagnetism." *Physical Review B*, **67** 016203.

[14] Cover, Thomas M. and Joy A. Thomas (1991). *Elements of Information Theory*. New York: John Wiley & Sons, Incorporated.

[15] Crutchfield, James P. (1994). "The Calculi of Emergence: Computation, Dynamics, and Induction." *Physica D*, **75**: 11-54.

[16] Crutchfield, James P. and David P. Feldman (1997). "Statistical Complexity of Simple One-dimensional Spin Systems." *Physical Review E*, **55**: R1239-R1242.

[17] Crutchfield, James P. and David P. Feldman (2003). "Regularities Unseen, Randomness Observed: Levels of Entropy Convergence." *Chaos*, **15**: 25-54.

[18] Crutchfield, James P. and Karl Young (1989). "Inferring Statistical Complexity." *Physical Review Letters*, **63**: 105-108.

[19] Crutchfield, James P. and Karl Young (1990). "Computation at the onset of Chaos." in W. H. Zurek, editor, *Complexity, Entropy, and the Physics of Information*, volume VIII of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 223-269. Reading, Massachusetts. Addison-Wesley.

[20] Dennis, J.E. and Robert B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Englewood Cliffs, New Jersey: Prentice-Hall, Incorporated.

[21] Ellis, S.G., F. Herman, E.E. Loebner, W.J. Merz, C.W. Struck and J.G. White (1958). "Photovoltages Larger than the Band Gap in Zinc Sulphide Crystals." *Physical Review*, **109**: 1860.

[22] Engel, G.E. and R.J. Needs (1990). "Total Energy Calculations on Zinc Sulphide Polytypes." *Journal of Physics: Condensed Matter*, **2**: 367-376.

[23] Feldman, David Polant (1998). *Computational Mechanics of Classical Spin Systems*, Ph.D. thesis, University of California, Davis.

[24] Feldman, David Polant (1998). "A Brief Introduction to: Information Theory, Excess Entropy, and Computational Mechanics." *Unpublished.*

[25] Feldman, David P. and James P. Crutchfield (1998). "Discovering Non-critical Organization: Statistical Mechanical, Information Theoretic, and Computational Views of Patterns in One-Dimensional Spin Systems." Santa Fe Institute Working Paper, 98-04-26.

[26] Feldman, David P. and James P. Crutchfield (1998). "Measures of Statistical Complexity: Why?" *Physics Letters A*, **238**: 244-252.

[27] Frevel, L.K., D.R. Petersen, and C.K. Saha (1992). "Polytype distribution in Silicon Carbide." *Journal of Materials Science*, **27**: 1913-1925.

[28] Frey, Friedrich, Heinz Jagodzinski and Georg Steger (1986). "On the Phase transformation Zinc Blende to Wurtzite." *Bulletin de Minéralogie*, **109**: 117-129.

[29] Glasser, L.S. Dent (1977). *Crystallography and its Applications.* New York: Van Nostrand Reinhold Company Limited.

[30] Gonçalves, W.M., R.D. Pinto, J.C. Sartorelli, and M.J. de Oliveira (1998). "Inferring Statistical Complexity in the Dripping Faucet Experiment." *Physica A*, **257**: 385-389.

[31] Gosk, J.B. (2000). "Investigation of One-Dimensionally Disordered Structures of $A^{II}B^{VI}$ Crystals by Monte Carlo Techniques. I. The 3C Disordered Structure and the 3C Structure with Different Kinds of Stacking Faults." *Crystal Research and Technology*, **35**: 101-116.

[32] Gosk, J.B. (2001). "Investigation of One-Dimensionally Disordered Structures of $A^{II}B^{VI}$ Crystals by Monte Carlo Techniques. II. 2H Structure with Different Kinds of Stacking Faults." *Crystal Research and Technology*, **36**: 197-213.

[33] Grassberger, Peter (1986). "Toward a Quantitative Theory of Self-Generated Complexity." *International Journal of Theoretical Physics*, **25**: 907-938.

[34] Guinier, André (1963). *X-Ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies.* San Francisco: W.H. Freeman and Company.

[35] Gumps, Godfrey, Girija Dubey, A. Salman, B.S. Mahmoud and Danhong Huang (1995). "Statistical and Transport Properties of Quasiperiodic Layered Structures: Thue-Morse and Fibonacci." *Physical Review B*, **52**: 210-219.

[36] Hahn, Theo, A.J.C. Wilson and U. Shmueli (1992). *International Tables for Crystallography,* 3^{rd}*, revised edition.* Boston: Kluwer Academic publishers.

[37] Hansen, James E. (1993). *Computational Mechanics of Cellular Automata.* Ph.D. thesis, University of California, Berkeley.

[38] Hao, Bai-Lin and Wei-Mou Zheng (1998). *Applied Symbolics Dynamics and Chaos.* Singapore: World Scientific Publishing Company.

[39] Hendricks, Sterling, and Edward Teller (1942). "X-Ray Interference in Partially Ordered Layer Lattices." *Journal Of Chemical Physics*, **10**: 147-167.

[40] Hopcroft, John E. and Jeffrey D. Ullman (1979). *Introduction to Automata Theory, Languages, and Computation.* Reading, Massachusetts: Addison-Wesley Publishing Company.

[41] Jagodzinski, Heinz (1949). "Eindimensionale Fehlordnung in Kristallen und ihr Einfluss auf die Röntgeninterferenzen. I. Berechnung des Fehlordnungsgrades aus den Röntgenintensitäten." *Acta Crystallographica*, **2**: 201-207.

[42] Jagodzinski, Heinz (1949). "Eindimensionale Fehlordnung in Kristallen und ihr Einfluss auf die Röntgeninterferenzen. II. Berechnung der fehlgeordneten dichtesten Kugelpackungen mit Wechselwirkungen der Reichweite 3." *Acta Crystallographica*, **2**: 208-214.

[43] Jagodzinski, Heinz (1987). "Diffuse X-ray Scattering from Crystals." *Progress in Crystal Growth and Characterization of Materials*, **14**: 47-102.

[44] Kabra, V.K. and Dhananjai Pandey (1988). "Long-Range Ordered Phases without Short-Range Correlations." *Physical Review Letters*, **61**: 1493-1496.

[45] Kitchens, Bruce P. (1998). *Symbolic Dynamics One-sided, Two-sided and Countable State Markov Shifts.* Berlin: Springer-Verlag.

[46] Kittel, Charles (1996). *Introduction to Solid State Physics,* 7^{th} *Edition.* New York: John Wiley & Sons, Incorporated.

[47] Klinkner, Kristina Lisa, Cosma Rohilla Shalizi and James P. Crutchfield (2001). "An Algorithm for Pattern Discovery in Time Series." *Journal of Machine Learning Research*, submitted.

[48] McCauley, Joseph L. (1993). *Chaos, Dynamics and Fractals: an Algorithmic approach to Deterministic Chaos* volume 2 of *Cambridge Nonlinear Science Series.* Cambridge: Cambridge University Press.

[49] Milburn, G.H.W. (1973). *X-ray Crystallography: An Introduction to the Theory and Practice of Single-crystal Structure Analysis.* London: Butterworth & Company Limited.

[50] Packard, N.H., J.P. Crutchfield, J.D. Farmer and R.S. Shaw (1980). "Geometry from a Time Series." *Physical Review Letters*, **45**: 712-716.

[51] Palmer, A.J., C.W. Fairall, and W.A. Brewer (2000). "Complexity in the Atmosphere." *IEEE Transactions on Geoscience and Remote Sensing*, **38**: 2056-2063.

[52] Peyrière, Jacques, Eric Cockayne and Françios Axel (1995). "Line-Shape Analysis of High Resolution X-Ray Diffraction Spectra of Finite Size Thue-Morse GaAs-AlAs Multilayer Heterostructures." *Journal de Physique I*, **5**: 111-127.

[53] Reif, F. (1965). *Fundamentals of Statistical and Thermal Physics*. New York: McGraw-Hill Book Company.

[54] Révész, György E. (1983). *Introduction to Formal Languages*. New York: McGraw-Hill Book Company.

[55] Sebastian, M.T. (1988). "X-ray Diffraction from faulted Zinc Sulphide Crystals undergoing Phase Transitions." *Journal of Materials Science*, **23**: 2014-2020.

[56] Sebastian, M.T. (2001). *Private Communication.*

[57] Sebastian, M.T. and P. Krishna (1983). "Anomalous photovoltaic effect and disorder in ZnS Crystals." *Bulletin of Material Science*, **5**: 257-266.

[58] Sebastian, M.T. and P. Krishna (1984). "An X-ray Diffraction study of Faulting in Single Crystal of cubic ZnS grown from the Vapour Phase." *Philosophical Magazine A*, **49**: 809-821.

[59] Sebastian, M.T. and P. Krishna (1987). "X-ray Investigation of the Mechanism of Phase Transformation in Single Crystals of ZnS, Zn_xCd_{1-x} and $Zn_xMn_{1-x}S$ (I) Calculation of Diffraction Effects by a Three Parameter Model." *Crystal Research and Technology*, **22**: 929-941.

[60] Sebastian, M.T. and P. Krishna (1987). "X-ray Investigation of the Mechanism of Phase Transformation in Single Crystals of ZnS, Zn_xCd_{1-x} and $Zn_xMn_{1-x}S$ (II) Comparison of Observed and Calculated Diffraction Effects." *Crystal Research and Technology*, **22**: 1063-1072.

[61] Sebastian, M.T. and P. Krishna (1994). *Random, Non-Random and Periodic Faulting in Crystals.* Langhorne, Pennsylvania: Gordon and Breach Science Publishers.

[62] Shalizi, Cosma Rohilla (2001). *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata.* Ph.D. thesis, University of Wisconsin, Madison.

[63] Shalizi, Cosma Rohilla and James P. Crutchfield (2001). "Computational Mechanics: Pattern and Prediction, Structure and Simplicity." *Journal of Statistical Physics*, **104**: 819-881.

[64] Shannon, Claude E. and Warren Weaver (1963). *The Mathematical Theory of Communication.* Chicago: University of Illinois Press.

[65] Shaw, J.J.A. and Volker Heine (1990). "The Nature of interplanar interactions in SiC Polytypes." *Journal of Physics: Condensed Matter*, **2**: 4351-4361.

[66] Steinberger, I.T. (1983). "Polytypism in Zinc Sulphide," in *Crystal Growth and Characterization of Polytype Structure*. Ed. P. Krishna. Oxford: Pergamon Press.

[67] Strogatz, Steven H. (1994). *Nonlinear Dynamics and Chaos*. Cambridge, Massachusetts: Perseus Publishing.

[68] Teubner, Max (1990). "Ground States of Classical One-Dimensional Lattice Models." *Physica A*, **169**: 407-420.

[69] Trigunayat, G.C. (1991). "A Survey of the phenomenon of polytypism in Crystals." *Solid State Ionics*, **48**: 3-70.

[70] Upper, Daniel Ray (1997). *Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models.* Ph.D. thesis, University of California, Berkeley.

[71] Vainshtein B.K. (1981). *Modern Crystallography I: Symmetry of Crystals Methods of Structural Crystallography.* Berlin: Springer-Verlag.

[72] Varn, Dowman P. and Geoffrey S. Canright (2001). "The Crystal Problem for Polytypes." *Acta Crystallographica A*, **57**: 4-19.

[73] Verma, Ajit Ram and P. Krisha (1966). *Polymorphism and Polytypism in Crystals.* New York: John Wiley & Sons, Incorporated.

[74] Warren, B.E. (1969). *X-Ray Diffraction.* Reading, Massachusetts: Addison-Wesley Publishing Company.

[75] Watson, Greg, Geoffrey S. Canright and Frank L. Somer, Jr. (1997). "Reasonable and Robust Hamiltonians violating the Third Law of Thermodynamics." *Physical Review E*, **56**: 6459-6465.

[76] Welberry, T.R. (1985). "Diffuse x-ray Scattering and Models of Disorder." *Reports on Progress in Physics*, **48**: 1543-1593.

[77] Wiggins, Stephen (1990). *Introduction to Applied nonlinear Dynamical Systems and Chaos.* New York: Springer-Verlag.

[78] Wilson, A.J.C. (1942). "Imperfection in the Structure of Cobalt II. Mathematical Treatment of the Proposed Structure." *Proceedings of the Royal Society Series A: Mathematical and Physical Sciences*, **180**: 277-285.

[79] Woolfson, Michael M. (1997). *An Introduction to X-ray Crystallography.* Cambridge: Cambridge University Press.

[80] Yi, Jaichul, and Geoffrey S. Canright (1996). "Possible disordered ground states for layered solids and their Diffraction Patterns." *Physical Review B*, **53**: 5198-5210.

[81] Yoemens, Julia (1988). "The Theory and Application of Axial Ising Models." *Solid State Physics*, **41**: 151-200.

[82] Young, Karl (1991). *The Grammar and Statistical Mechanics of Complex Physical Systems.* Ph.D. Thesis, University of California, Santa Cruz.

Appendix

Appendix A

Conjecture

We state and prove a conjecture in this section.

Conjecture 1 *When a process results in two-layer correlation functions that decay to an asymptotic value, that value is 1/3.*

We begin by making some definitions. Suppose we have a sample consisting of N layers, where the orientation of the i^{th} layer is denoted by $x_i \in \{A, B, C\}$. Let us define a triad of quantities f_α such that

$$f_c(x_i, x_j) = \begin{cases} 1 & \text{if } x_i \text{ is cyclically related to } x_j \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

with $f_r(x_i, x_j)$ and $f_s(x_i, x_j)$ defined in an obviously similar way. It is clear that the following identity must hold,

$$f_c(x_i, x_j) + f_r(x_i, x_j) + f_s(x_i, x_j) = 1 \qquad \forall \quad i, j. \tag{A.2}$$

Using the fs, we can define the Qs.

$$Q_\alpha(n) = \frac{1}{(N-n)} \sum_{k=1}^{N-n} f_\alpha(x_{k+n}, x_k) \tag{A.3}$$

where $\alpha \in \{c, r, s\}$. We now write the identity

$$\frac{1}{(N-n)} \sum_{k=1}^{N-n} \sum_{\alpha} f_\alpha(x_{k+n}, x_k) = 1. \tag{A.4}$$

Recognizing the last term in the inner sum as $Q_s(n)$, and inserting an identity, we have

$$\begin{aligned} Q_s(n) + \frac{1}{(N-n-1)} \sum_{k=1}^{N-n-1} \{ & f_r(x_{k+n}, x_k)[f_s(x_{k+n+1}, x_k) + f_c(x_{k+n+1}, x_k)] \\ & + f_c(x_{k+n}, x_k)[f_s(x_{k+n+1}, x_k) + f_r(x_{k+n+1}, x_k)] \} = 1. \end{aligned}$$

Regrouping terms, and summing one term, we get

$$Q_s(n) + Q_s(n+1) + \frac{1}{(N-n-1)} \sum_{k=1}^{N-n-1} \{f_r(x_{k+n}, x_k) f_c(x_{k+n+1}, x_k) + f_c(x_{k+n}, x_k) f_r(x_{k+n+1}, x_k)\} = 1. \tag{A.5}$$

Looking at the summation term in equation A.5 we see that is non-negative. Further it has a maximum value of one. This then constrains the possible values of $Q_s(n) + Q_s(n+1)$ to lie between zero and one for all n. This holds not only for $Q_s(n)$, but for the cyclic and reverse Qs also so that we may write

$$0 \leq Q_\alpha(n) + Q_\alpha(n+1) \leq 1. \tag{A.6}$$

For all the processes we have examined, this inequality holds.

Let us now examine the sum in equation A.5. We define a new quantity, $g(x_i, x_j, x_l)$, as

$$g(x_i, x_j, x_l) = f_r(x_j, x_l) f_c(x_i, x_l) + f_c(x_j, x_l) f_r(x_i, x_l). \tag{A.7}$$

It is easy to see that $g(x_i, x_j, x_l)$ may also be written as

$$g(x_i, x_j, x_l) = \begin{cases} 1 & \text{if } x_i, x_j, x_l \text{ are all different} \\ 0 & \text{otherwise} \end{cases} \tag{A.8}$$

This then implies that $g(x_i, x_j, x_l)$ is invariant under exchange of any two of its arguments. We write equation A.5 as

$$Q_s(n) + Q_s(n+1) + \left\langle g(x_{k+n+1}, x_{k+n}, x_k) \right\rangle_k = 1 \tag{A.9}$$

where the $\left\langle ... \right\rangle_k$ indicates an average over all layers in the sample. For a fcc stacking, we see that $bigl\langle g(x_{n+2}, x_{n+1}, x_n) \right\rangle_k = 1$ while $\left\langle g(x_{n+2}, x_{n+1}, x_n) \right\rangle_k = 0$ for the hcp stacking sequence. By exchanging the second and third indices of g in equation A.9, and writing out the average, one can show,

$$\left\langle g(x_{k+n+1}, x_{k+n}, x_k) \right\rangle_k = \left\langle f_c(x_{k+n+1}, x_{k+n}) f_c(x_{k+n}, x_k) \right\rangle_k + \left\langle f_r(x_{k+n+1}, x_{k+n}) f_r(x_{k+n}, x_k) \right\rangle_k. \tag{A.10}$$

Let us now specialize to the case where Qs have reached their asymptotic values, that is $n \gg \lambda_q$, and the layers are by definition uncorrelated. Let P_X be the probability of occurrence of the layer X, $X \in \{A, B, C\}$. Then $Q_s = P_A^2 + P_B^2 + P_C^2$. Similarly, $Q_c = P_A P_B + P_B P_C + P_C P_A = Q_r$ since multiplication is commutative. We also identify $\langle f_\alpha(x_{k+n}, x_k) \rangle = Q_\alpha(n)$. Then in the asymptotic region

$$\left\langle g(x_{k+n+1}, x_{k+n}, x_k) \right\rangle_k = Q_c Q_c(1) + Q_r Q_r(1) = Q_c. \tag{A.11}$$

Therefore we see that $2Q_s + Q_c = 1$ and $Q_s + 2Q_c = 1$. The solution $Q_c = Q_r = Q_s = 1/3$ immediately follows.

Vita

Dowman Parks Varn was born on 11 March 1964. He was graduated from Clemson University with a Bachelor of Science degree in Physics in 1986. He was graduated from the University of Virginia in 1994 with a Master of Arts in Physics. He is married to Mary Varn, and they are the proud parents of three dogs.