The application of non-linear dynamics to teletraffic modelling.

Samuel, L. G.

The Application of Non-linear Dynamics to Teletraffic Modelling

L. G. Samuel

Submitted for the Degree of Doctor of Philosophy

Department of Electronic Engineering
Queen Mary and Westfield College
University of London

July 1999

To my parents, wife and Anna Alexandra Isabella born 6/11/97.

Nature has a wonderful way of creating a true balance in life.

Abstract

Self-similar traffic has been observed in teletraffic networks over all time scales of engineering interest. This type of traffic has no characteristic time scale due to its burstiness and causes the network buffers to overflow affecting Quality of Service (QoS). Self-similar traffic has been modelled via stochastic methods. It has also been modelled using non-linear dynamics. In this thesis we use techniques from non-linear dynamics in the teletraffic modelling of modern packetised telecommunications networks. We develop a novel teletraffic framework for the modelling of self-similar traffic in a parsimonious, parameterisable and predictable manner based on the use of non-linear dynamics models in the form of a chaotic map family. This family consists of models, based on intermittency maps, for the accelerated simulation of self-similar behaviour of individual sources and aggregated traffic in such networks. We have significantly extended the characterisation of the individual source map models of Pruthi and Erramilli. The extension accounts for the impact of all five parameters (*e* and *m* for both states, and *d*) on *H*, the parameterisation for load via the invariant density, and the parameterisation for heavy tailed sojourn times in the *ON* and *OFF* states via the transit-time. These new aggregate traffic models provide up to two orders of magnitude speed-up over FBM/FGN and Pruthi's *N* one-step methods.

We perform mathematical analysis of the map family with respect to *H* proving the conjecture put forward by Pruthi that asymptotically *H* is only dependent on the dominant value of *m*. Numerical results show that convergence is slow and that for the coupled map *H* differs substantially from the theory. However the deviation from the theoretical is predictable and this leads to an empirical fit for the asymptotic dependence of *H* on *m*. These results also show that the underlying dynamics of the map persist in all of the map interpretations. The numerical results also show limitations of parameter ranges on *H*, particularly for *d* ($0.1 < d < 0.9$). However, this limit can be overcome practically to some degree by manipulating the time resolution of the iterates. Transit-time analysis of the map family highlights a further parameter limitation which stems from *e:* any value of $e > 0$ effectively limits the range of time-scales over which LRD occurs. These numerical limitations apply to all map interpretations.

We have developed a method of measuring *H* via the map's variance that is promising for measuring *H* on-line. We have also found, by comparing for accuracy this and the Abry-Veitch method for measuring *H*, that *H* by itself as a parameter for modelling self-similar traffic in a "parsimonious" manner may not be enough. This conclusion is drawn from observing different queueing behaviour with input traffic having the same *H*. This leads to the suggestion that reliance on a single method for determining *H* on-line may prove unwise. This comparison also shows the flexibility that these map models have in specifying key LRD behaviour that determines the impact on queueing. This flexibility is derived from the intuitive relationship that these map models have to their underlying physical *ON-OFF* process.

Acknowledgments

There are a number of people to who I am indebted. Undertaking a PhD is not easy especially for those connected indirectly to the task, in particular my wife Rosi: a constant source of encouragement and an outstanding display of patience well beyond the call of duty.

For those who are directly connected to the task I especially would like to thank Dr Raul Mondragón. His mathematical supervision, encouragement, enthusiasm and friendship has been a constant source of inspiration throughout this PhD. Dr Jonathan Pitts, his engineering supervision has always ensured that at least one leg be fixed to the ground, and above all ensured timely completion of the project. Professor David Arrowsmith and Dr John Schormans, for their enlivening conversations in, on and around the subject. Professor Laurie Cuthbert, paymaster general to the Telecoms Group, for that essential ingredient without which this project would surely not have started: finding funding. EPSRC and NORTEL for financial support. Dr Dragan Boscovic of Motorola for encouraging me to rise up to the task. My colleagues in the Lab: Arif Al-Hammadi, Husam Awadalla, Eliane Bodanese, Babul Miah, Sutha Sivagnanasundaram, Steven Winstanley, and Tijana Timotijevic, for hours of endless fun.

A special thank you goes to Rebecca Whiting and Arif Al-Hammadi, for without the use of their lap tops this project would certainly not have finished on time.

Finally, I would like to thank the Royal Navy: In their infinite wisdom they decreed that I did not posses the intelligence to complete a degree, let alone a PhD. It is comforting to know that such an august body can still make errors in judgement. Stand fast Captain Mike Bowker (RN).

Abbreviations and Acronyms

ABR	Available Bit Rate
ATM	Asynchronous Transfer mode
CML	Coupled Map Lattices
FBM	Fractional Brownian Motion
FGN	Fractional Gaussian Noise
FPDI	Floating Point Double Intermittency map
H	Hurst Parameter
IEA	Improved Exponential Approximation
IID	Identical and Independently Distributed
IVV	Indicator Variable Variance
LH	Left Hand
LRD	Long Range Dependence
MEA	Modified Exponential Approximation
MLE	Maximum Likelihood Estimator
RH	Right Hand
RNG	Renormalisation Group
R/S	Rescaled range Statistic
SIC	Sensitive dependence on Initial Conditions
SRD	Short Range Dependence
SBM	Standard Brownian Motion
TRAP	Trapsizoidal approximation
UPC	User Parameter Control

Contents

Figures

Tables

1 Introduction

Self-similar traffic has been mainly modelled via stochastic means, such as Fractional Gaussian Noise (FGN) and Fractional Brownian Motion (FBM). There is an alternative method for modelling such traffic, i.e. via non-linear dynamic models. In this thesis we study non-linear dynamic models. We have developed novel interpretations on the use of these models as prameterisable and predictable models for the accelerated simulation of self-similar behaviour of individual sources and aggregated traffic in telecommunications networks. Throughout this work we refer to them as "non-linear models", "non-linear map models" or "chaotic map models"; in the interests of brevity some times we simply call them "chaotic maps", "map models" or simply "maps"[1].

We focus on two theoretical areas of non-linear map models: their analytical tractability and their parameter interdependence with regard to the Hurst parameter *H*. We are interested in *H* because of its adverse affects on cell loss [NOR93, 95, ERR96]. We use results of the theoretical investigation to develop interpretations on the chaotic maps with respect to source aggregation that yield significant benefits for accelerated simulation models of self-similar traffic.

This thesis draws extensively on recent research in various fields but is primarily motivated by the observations on self-similar traffic initially made by Fowler and Leland [FOW91, LEL93, LEL94] who were following up on the measurement of Ethernet traffic initiated by Shoch and Hupp [SHO80]. Leland and Fowler were able to make precise measurements of Ethernet traffic over many years which showed that the variation in peak traffic load extended over many orders of magnitude. This called into question the Markov based models then currently used to predict network performance. These models aggregate to white noise i.e. the peaks in traffic load were smoothed out as traffic aggregation period was extended. This type of aggregation behaviour is not that witnessed in real traffic measurements. This type of behaviour is most strikingly exhibited by the "visual proof diagrams" (fig 4 in [LEL94]). Others have also questioned the validity of Markov based models, for example Paxson and Floyd [PAX95].

These observations led to the proposal of various stochastically self-similar models which render "realistic" traffic in simulators. Early contributors to self-similar models were most notably Norros [NOR93, 95], Veitch [VEI92] and Leland *et al* in [LEL94]. These models are inspired by the work undertaken by Mandelbrot and co-workers in the 1960's [MAN63, 65, 68a, 68b, 68c]. The family of models that Mandelbrot proposed are termed FBM models and are drawn from observations in hydrological records and error clusters in communication systems. As recently as 1995-96 Huang *et al* [HUA95], Slimane and Le-Ngoc [SLI95] and

[1] Strictly speaking there are classes of non-linear models that are not chaotic.

Lau *et al* [LAU95] had published papers on FBM algorithms for traffic simulation. Moreover Chen *et al* [CHEH96] has evaluated the relative performance of such algorithms in terms of accuracy and speed of sample generation. The main limitation with this type of model is that they generate their samples off-line i.e. their traces have to be generated separately form the simulation that is going to use them. The reason for this is mainly due to the algorithm's need for knowledge of all past samples in order to generate the next sample.

There are alternative models to FBM developed by Erramilli, Singh and Pruthi [ERR94a, 94b, 94d, 95a, PRU95a, 95b]. These alternative models are non-linear dynamic map models. As such these models use chaotic intermittency maps to model fractal-traffic. Pruthi [PRU95b] has shown that fractal traffic models using chaotic maps produce the required stretched exponential queue length distributions and that when these models aggregate they tend to FBM. These models can be use on-line because. the required self-similarity is inherent in the dynamics of the maps, allowing them to produce results in real time. However their behaviour has not been completely characterised in terms of the map's parameter interdependence on *H*. To exploit these models fully we need to understand the behaviour of the chaotic map models with regard to alterations in the model parameters. This thesis has contributed to the understanding of the map's parameter interdependence through contributions to the theoretical understanding of the map's transit time, which is the Long Range Dependent (LRD) component of the chaotic map. This line of investigation led to a proof on the asymptotic dependence of *H* on a single map parameter for a de-coupled map. Numerical experimentation supporting the proof highlighted the effect coupling has on the map resulting in the development of an empirical dependence on *H* for a single map parameter. An expression for the variance of the map output has also been developed which has lead to the development of a technique for measuring the *H* value of the map output on-line. Using the insight that these preliminary studies gave, aggregate map models were then developed which took advantage of the invariant densities of the maps in order to preserve the effects of self-similarity under aggregation of bursty traffic sources. However, the initial aggregate map models possessed limitations, principally in the type of LRD that could be modelled. This limitation motivated the development of the Bulk Property map which is an accelerated map that has an order of magnitude speed up over other map model techniques.

Analysis of queueing systems fed by self-similar traffic has been undertaken. The theoretical basis for the analysis of FBM source fed queues stems from Taqqu [TAQ86, 97] and Norros [NOR93]. Taqqu has theoretically shown that the aggregation of heavy tailed ON/OFF source models tends to FBM. Norros [NOR93] has shown that the queue length distributions are of the stretched exponential type. Taqqu's theoretical predictions have been evidenced in real traffic measurements by Willinger [WIL97]. These studies have motivated others to assess the impact of Long Range Dependence (LRD) exhibited by fractal traffic, on networking issues such as:

- queueing performance (Erramilli *et al* [ERR96a]),
- Quality of Service (Duffield *et al* [DUF95b])[2],
- bandwidth allocation and short term traffic prediction (Norros [NOR95]), and
- the length of time necessary for effective correlation analysis of traffic (Addie *et al* [ADD95]).

Others (Beran *et al* [BER95]) have examined particular types of traffic for LRD.

There is another reason for studying chaotic map models. Fowler and Leland [FOW91] summarise clearly the congestion problems that bursty traffic inflicts on a network i.e. the traffic remains bursty over all times scales of engineering interest and that no amount of buffer space will ever prevent loss, and raise the question of how to combat the effects of bursty traffic. Non-linear dynamics and in particular the idea of chaotic control present an opportunity to move away from the present stochastic view of networks to a more dynamical systems view, enabling the problem of congestion to be combated using dynamical rather then stochastic techniques. This view is supported by recent advances in the chaotic control of Coupled Map Lattices (CML) [MON97b, OKE95, SEP95, YOU95]. CML's and high speed data networks have similar topologies. It is this similarity that makes the use of chaotic control as a method for the prevention of network congestion very appealing. In a chaotic system the chaotic controller exploits the "complexity" of the system dynamics such that a small change in the parameters can change the system to a required state. However, we should remark that in telecommunications networks the control must change the statistical behaviour of the traffic. For this reason a key stage in the development of a "statistical chaotic" control for networks is the identification of parameter adjustment in the statistical output of the chaotic map models. This thesis is a significant step towards developing theoretical understanding necessary for statistical chaotic control of networks since it addresses the stochastic aspects of traffic generation from chaotic map models that are used to depict network traffic.

The remainder of this thesis is organised as follows: Chapter 2 reports on network measurements carried out in the late 1980's and early 1990's, in which self-similar traffic was observed. It mentions the problems that this type of traffic brings and the various modelling methods employed to cope with it. Chapter 3 introduces the concepts behind stochastic self-similarity traffic. Chapter 4 explains the concepts of non-linear dynamics that are used to construct the non-linear models used in this thesis.

Chapter 5 presents theoretical work on intermittency map models which is used to develop an approximation of the intermittency map transit time in a perturbed system. We use this approximation to provide a proof of the dependence of H on the parameter m for the

[2] Note: In reality this is a large deviation theory approach. However it is included here because it attacks the same problem of an event occurring in the tail of a distribution.

decoupled map, and derive an empirical dependence on *H* for a coupled map from experimental data. We then develop a method of measuring the *H* parameter through the measurement of the variance of these map models and investigate the interdependence of parameters and their effect on the value of *H*.

Chapter 6 investigates source aggregation of a single intermittency map through the development of approximations on the single intermittency map's invariant density. This is done to provide a method of composition and de-composition of aggregate traffic at a node. Chapter 7 extends the work of Chapter 6 by introducing the Bulk Property map. This type of model is applicable to single and double intermittency maps. Bulk Property map is an accelerated map for use in on-line modelling. The Bulk Property map is devised to compose and decompose traffic at a node. We show that the underlying dynamics of the Bulk Property map persist in the map's output. We also show results on the Bulk Property map interdependence of parameters on *H*. We show the speed-up in simulation time over the single intermittency map. We also suggest an aggregated method for measuring the variance using these maps based on the technique developed in Chapter 5. Chapter 8 develops the method of measuring *H* via the map's variance (initially discussed in Chapter 5). We show that this method is promising for measuring *H* on-line.

In Chapter 9 we propose a new network control scheme based on chaotic control in which the aggregation techniques developed in Chapters 6 and 7 combined with the on-line measurement techniques of Chapter 8 form a pivotal part. This method is intended to control data flows (non-real time) in high speed networks. Chapter 10 lays out the conclusions.

2 Background

In this chapter we place into context the effects that self-similar traffic has on current networks and proceed to argue the case for viewing these problems in an alternative way: that of a non-linear dynamical systems approach to traffic modelling.

2.1 Introduction

Recent highly accurate traffic measurements of high speed networks, with time stamps accurate to within 100μs, have uncovered new phenomena in the time-series reflecting the behaviour of the network traffic [FOW91]. The phenomena present itself as bursty traffic over all time intervals of engineering interest. This observation caused great interest at the time since Markovian based models were being used to portray aggregated traffic in the network models and these models tended to white noise as the level of aggregation increased. Clearly this is in conflict with actual observations of network traffic. The engineering implications of using incorrect models in the design and planning stages of new networks is quite staggering since the robustness of a new design to congestion and its side effects on the systems when congestion occurs depends on the realism of the model used. The observations outlined above are particularly relevant to ATM. This is because ATM is becoming the transport vehicle for a wide variety of traffic streams, whether it is legacy traffic, LAN-LAN, Internet IP, multimedia, etc. ATM was created as a unifying transport mechanism. The mechanism provides the means to statistically multiplex variable and constant bit rate streams. One of the main features of ATM is its statistical multiplexing gain [SAI94]. Statistical multiplexing gain arrives out of multiplexing traffic streams where the sum of the individual peak bandwidths is greater than the capacity of a given link [CHET95]. This is possible because the peaks in the individual traffic streams seldom occur together. Therefore the statistical multiplexing effect relies on the condition that enough sources are multiplexed and that they are not correlated [PRY91].

The analysis of such a gain has been attempted under the assumptions of Poisson arrival processes and exponential distributed holding times [SAI94]. The implication of this type of analysis is that such traffic streams when aggregated tend to white Gaussian noise, i.e. the variation of the traffic would eventually smooth out (see Figure 2.1). However, traffic measurements carried out in the late 1980's and early 1990's revealed that whereas the correlations of the traffic were thought to decay exponentially fast (Markovian in structure) the traffic measured in real networks possessed correlation structures which decayed much slower than exponentially [FOW91]. This type of traffic has become known as Long Range Dependent (LRD). To resolve this issue new models and approaches have been sought which describe/analyse the network behaviour correctly. Broadly speaking the approaches come under two categories:

- Remain within Markovian framework, but characterise behaviour at the appropriate time scale of interest, an example being Large Deviation Theory as advocated by Duffield, Lewis and O'Connel [DUF94b, 95b] which assesses impact of rare events.
- Adopt a self-similar framework.

The former approach views how a population introduces non-stationary into teletraffic via service usage e.g. diurnal variations, which needs to be factored out for explicitly. The later approach views the LRD characteristics of individual services types as underlying cause for self-similarity but maintains the view that the traffic is stationary.

In a self-similar framework, stochastic self-similarity expresses how the probabilistic structure of a process varies with the time scale. There are two broad approaches to self-similar traffic modelling:

- Fractional Brownian/Gaussian Motion as advocated by Leland, Veitch and Norros [LEL94, VEI92, NOR93, NOR95]. These models have the attraction of parsimony i.e. they use a small number of parameters. These models use *H* as their principal parameter. However these models have a drawback in that they lack the intuitive relationship to underlying physical process, and the traces they produce may not have the specified *H*. It is this last point that prevents their use as an on-line model, or
- Chaotic maps as advocated by Erramilli and Pruthi [ERR94a, 94b, 94d, 95a,PRU95a, 95b]. Here the models utilise non-linear dynamical ideas to model traffic. These models were limited in application by analytical tractability and poor aggregate modelling. However they have the advantage of parameter parsimony and are predictable with respect to the *H* obtained. Furthermore these models have an intuitive relationship to the underlying physical *ON-OFF* process. These factors permit their on-line use and combined with dynamic parameter changing enables the self-similar framework to address non-stationary issues as well as LRD.

2.2 Characteristics of High Speed Network Traffic

Fowler [FOW91] reported on studies conducted at the end of the 1980's and early 1990's that packet traffic exhibited burstiness over a large number of time scales. There had been earlier studies that had also reported similar results [SHO80] and independent studies which had confirmed the scale of the observed burstiness and the failure of current models to portray this behaviour [PAX95]. Fowler's study is important because it was the first time that accurate measurements were made at fine time scales and thus the scaleability of a burst was reliably observed. These bursts existed at every time scale, from milliseconds to days and they looked similar independently of the time scale, i.e. the traffic is self-similar. One characteristic of this self-similar traffic is that it is correlated at all time scales of engineering interest, i.e. the traffic has LRD. The self-similarity and the LRD are quantified by the Hurst parameter H $(\frac{1}{2} \leq H < 1)$. Large values of *H* correspond to larger fluctuations on the burst size and stronger correlations in the traffic. The significance of this observation lies in the correlation structure of the LRD

traffic. A traffic source that possesses a sojourn distribution whose tail probabilities decay as a power rather than an exponential law is said to have a "heavy tail". It is this heavy tailedness, exhibited as non-negligible correlations in the traffic over large lags which is known as LRD. Heuristically, one can view the individual traffic stream correlation as overhanging each other when aggregated, causing an increase in the probability of the large aggregated bursts occurring. More importantly the aggregated traffic streams do not tend to white Gaussian noise (see Figure 2.1). In actual fact the aggregated traffic process tends towards a second-order statistically self-similar process which remains bursty over many time scales [TAQ97, WIL97]. We can state this more clearly in the following way:

$$\begin{pmatrix} LRD\ of\ individual \\ sources \end{pmatrix} \Rightarrow \begin{pmatrix} Slowly\ decaying\ correlation \\ structures \end{pmatrix} \Rightarrow \begin{pmatrix} Self-similarity\ in \\ aggregated\ traffic \end{pmatrix}.$$

This burstiness over many time scales combined with LRD is a characteristic of fractal time series. These time series are dominated by their low frequency behaviour. For various and more detailed explanations of this, the interested reader is directed to [TAQ86, TAQ97, BER94, BER95, ERR96a, KRI96]. This low frequency behaviour of the traffic poses problems for the traffic control schemes designed for ATM. In ATM preventative congestion control is preferred over reactive congestion control schemes. This is because the reactive control becomes inadequate in terms of response times for the high bit rates used in ATM [CHET95]. The preventative measures are concentrated in the connection admission control schemes (CAC) used to make decisions on the acceptance of calls into the system. Leland *et al* [LEL93, LEL94] have studied the effectiveness of CAC in the context of LRD traffic and found that CAC cannot minimise the congestion within the network and increasing buffer sizes appear to have no effect.

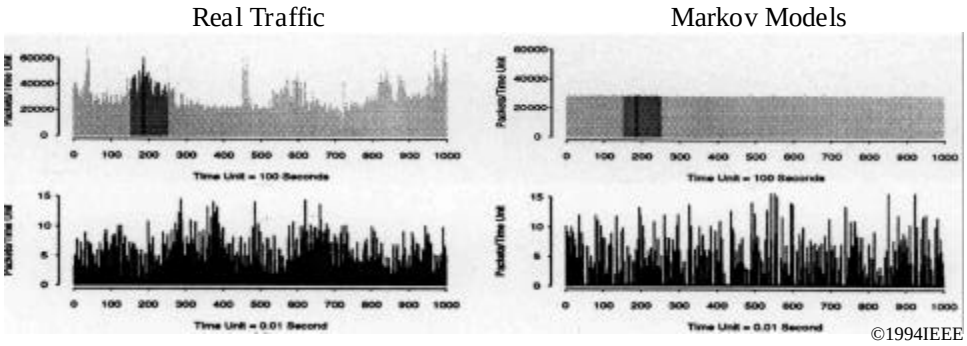

Figure 2.1 Real traffic trace against Markov model based trace for the same load. (The picture is taken from figure 4 in [LEL94]. Reproduced with permission).

2.3 Effects of Self-similarity on Queues

The burstiness in the traffic measurements stems from the fluctuations in the heavy tailed distributions of the individual traffic sources [LEL94, TAQ97, WIL97]. The probability of higher buffer occupancy has been linked to heavy tail distributions [NOR93, 95]. Practically the increase in probability has a drastic effect on the buffer occupancy since providing more buffer space is not a solution to buffer saturation [ERR96]. Eventually the buffer will fill up.

The implication of an increasing value of *H* is that it ultimately leads to higher buffer state occupancy [NOR93, ERR96] and hence increased probability of network congestion. Analysis of queueing systems fed by self-similar traffic has been undertaken. The theoretical basis for the analysis of FBM source fed queues stems from Taqqu [TAQ86, 97] and Norros [NOR93]. Taqqu has theoretically shown that the aggregation of heavy tailed ON/OFF source models tends to FBM. Norros [NOR93] has shown that the queue length distributions are of the stretched exponential type. Taqqu's theoretical predictions have been evidenced in real traffic measurements by Willinger [WIL97]. The effect of self-similar traffic on the queue blocking probability using Norros' blocking formula [NOR93] for a given load and varying *H*.can be seen in Figure 2.2.

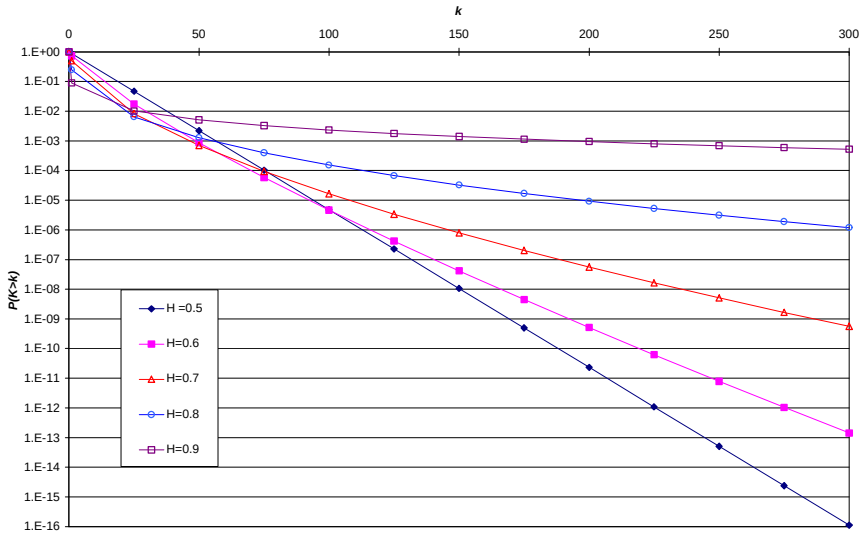

Figure 2.2 Queue state occupancy *P(K>k)* for 50 homogenous sources, with LRD (*a =1.5*) and mean load *r = 0.7*, for different self-similar traffic. *H=0.5* equates to Poisson traffic

One approach which one would have thought would reduce the impact of self-similar traffic on the network buffers would have been traffic shaping. One would have thought that spreading the burstiness of the individual traffic sources would have altered the characteristics of the traffic sufficiently to the point where individual traffic streams did not become a problem. Unfortunately this is not the case [LEL94]. Work undertaken recently by Molonár [MOL97] shows that shaping will not alter greatly the self-similarity present in the traffic. A robust indication of this could be implied form the work of Erramilli [ERR96] where experiments on reshuffled LRD data were undertaken. Essentially the entire order of a data stream had to be shuffled randomly before the LRD nature in the stream was lost. If all the LRD streams are shaped then all that is achieved is an extension over the period over which the self-similar traffic is present. This is because shaping still preserves the order of the data and the queue acts as a low pass filter. Naturally, the effects of self-similarity has motivated others to assess the impact of LRD in terms of queueing performance [ERR96a], Quality of

Service [DUF95b][3], bandwidth allocation and short term traffic prediction [NOR95], and the length of time necessary for effective correlation analysis of traffic [ADD95]. While others [BER95] have examined particular types of traffic for LRD.

2.4 Traffic Modelling

Conventional stochastic source models have a very intuitive interpretation in the form of simple *ON-OFF* models. These models have been extended to reflect aggregate traffic behaviour in the form of Markov Modulated models. Unfortunately these models are hard to parameterise. The conventional models are also restrictive since they describe real traffic only over a single time scale. They do not have LRD (see Figure 2.1).

There exist alternative stochastic traffic models known as FBM, and its incremental process FGN [MAN68a, 68b, 68c, NOR93, 95]. These models describe the traffic characteristics of real traffic, the self-similarity and LRD. These models were developed as a modification to standard Brownian Motion (SBM). SBM can be interpreted as the limit of a random walk. SBM is an attractive model for computer networks because the first order approximation of the arrival behaviour in such a network can be viewed as independent events and can thus be viewed as a form of random walk. For introductory reading on SBM see Ross [ROS87]. The problem with SBM is that it does not describe the traffic witnessed by Leland and Fowler in as much as the variance does not scale as the sample size increases. What is required is a stochastic process that is similar in definition to SBM but with additional scaling properties. FBM and FGN have correlation structures that scale and which are parameterised by the Hurst parameter *H*. Various models have been proposed using this process as the basis of the models for network traffic which exhibit burstiness (see for example [VEI92, NOR93, LEL94, HUA95, LAU95]). These models have the additional appeal that for aggregate traffic modelling they can be parameterised in a parsimonious way (have few parameters). However, these models do not have the intuitive appeal that the conventional models have and the traffic traces produced by these models are generated off-line.

There are traffic models based on non-linear chaotic maps that reproduce the properties of real traffic [ERR94a, 94b, 95a, 95b, I, II]. The chaotic map approach stems from an iterated dynamical systems approach to modelling[4]. The type of chaotic map that we are interested in is one that displays intermittency. Intermittency in this context means that the orbit of the map has protracted episodes of smooth behaviour interrupted by transitory bursts of activity. Intermittency is one of the transitory paths to chaos [POM80]. A map which has orbits that are intermittent also has LRD. It has been shown that when such source models are aggregated they produce self-similar traffic [PRU95b]. These models are attractive because

[3] Note: In reality this is a large deviation theory approach. However it is included here because it attacks the same problem of an event occurring in the tail of a distribution.
[4] The interested reader is directed to [COL80, GUL92, DEV94, MAC95, SCHU95] for good introductions into this subject of iterated dynamical systems.

they retain the intuitive feel of the conventional *ON-OFF* model, possess the required self-similarity and can be used on-line. This topic will be covered in greater depth in Chapter 4, and throughout the rest of this thesis. However these models are difficult to analyse and have not been characterised fully.

2.5 Alternative Approaches to Traffic Modelling

Although we have concentrated on a self-similar interpretation of the traffic traces - an interpretation which assumes stationarity, we can equally interpret the traffic as not being stationary [DUF94a, VAT98]. Under this interpretation we can use large deviation theory to assess the impact of bursty traffic on the system queues. In this approach the adequacy of Markovian models is accepted and large deviation techniques are used to assess the impact of the rare event "large bursts" on a queueing system. The large deviation principle comes from risk theory and was developed by the Swedish Mathematician Harald Cramér. The theory predicts the probability of rare events. We are interested in the tail of queue length distributions and more importantly the probability of buffer overflow. Large deviation theory states that the tail of a queue decays as

$$P(Q > q) \approx e^{-d\,q}, \tag{2.1}$$

where $-d$ is the asymptotic slope of the queue length distribution as the queue state becomes large, Q is the queue length and q is a given queue state. What this states is that for higher queue states the queue length decays linearly on a lin-log plot. This information can also be used to accept/reject calls entering a network. This has led to the formulation of CAC algorithms based on this principle [DUF94b, 94c, 95a, 95b]. The main point of contention between self-similar traffic modelling and large deviation assessment of traffic is that for the large deviation approach to hold, equ.(2.1) implies that the traffic is not LRD. This assumption questions that validity of self-similarity which requires LRD to be present in the traffic in order for the traffic to exhibit self-similarity.

Naturally there have been approaches which combine self-similarity and large deviations theory in order to arrive at some qualitative characterisation of the effects that self-similar traffic has on network buffering systems. These approaches have led to the notion of the cross-over effect [KRI96, FAN97]. This effect describes an increase in the multiplexing gain in the buffering system when streams of self-similar traffic are multiplexed. The point in this approach is that while the traffic is in the cross-over region, Markovian models (those with H=0.5) provide good (conservative) estimates for the buffer size required by the systems to cope with self-similar traffic streams. However, doubt has recently been cast as to the validity of cross-over effects in high speed networks [III].

The practicalities of modelling and/or assessing the effects of the traffic ultimately influence the chosen method. For example, the large deviation method assumes that there is quite a lot

of knowledge about the sources feeding a network. Conversely, the LRD approach assumes very little about the source other than it is an *ON-OFF* source that has protracted sojourn times in either state [TAQ97]. This last assumption is quite attractive since it cuts down on the complexity of the problem at hand and hence may influence the choice of technique.

2.6 Chaotic Maps and Traffic Modelling - Putting Things into Context

The generally accepted view on traffic measurements and modelling expressed during the debate held in the Hot Topics session on self-similar traffic of Performance96 conference held in Lausanne Switzerland in October of 1996, was as follows.

There was a great deal of discussion on whether the traffic measured was self-similar or not. Basically, the competing camps polarise into either LRD and self-similarity, or into large deviations. The conflict between the two lies in the interpretation of the measured traces as being stationary or not. In the large deviation approach the traces are assumed not to be stationary or at least to be stationary and mixing (i.e. the theory takes into account the causal effects) whereas the self-similar approach assumes that the traffic is stationary but with stepped increments. In order to solve this quandary, effort was placed on the search for causal effects. Again the causal effects are divided into two camps. The large deviationists are looking at network behavioural changes in order to justify their approach. An example of this is the correlation between changes in network activity and network user idiosyncrasies, i.e. breaks for lunch time, coffee, etc. This type of correlation is being given as a reason for the network not being stationary and therefore not self-similar. The long range dependencists are looking at the traffic type as the cause of self-similarity. For example, they are looking at long file transfers which are independent of the time of day and which therefore cause the traffic to be stationary but with increments.

Notwithstanding the comments on stationary/non-stationary traffic, what the practitioners of network design are looking for is a reasonable general modelling/traffic assessment tool and so are open to both ideas (large deviations and self-similarity). Presently there is a movement away from the current Black Box approach to traffic modelling (Box-Jenkins). This stems from the following argument. The traffic traces taken, for example, from a network at midday on Tuesday and which are subsequently used to characterise traffic in a simulator will bear no resemblance to the traffic traces taken at the same time the next day on the same network, i.e. the traffic traces are too diurnally dependent. This diurnal dependence favours the large deviation approach proposed by Lewis, Duffield, O'Connell, Crosby and co-workers [CRO-A, CRO-B, DUF94a, 94b, 95a, 95b, OCO], provided that the causal effects are known. The counter argument to this is that the traffic traces look self-similar and that they can be shown to have Hurst parameters greater than ½, indicating that the traces are positively correlated and that the self-similarity is due to independent increments.

From the preceding paragraph we can see that the traffic traces of Fowler and Leland have proved to be very interesting. The traces called into question:

- the suitability of Markov based models;
- the assumption of stationarity;
- the causal effects of the behaviour observed in the traces; which in turn,
- questioned the generality of the modelling methods used to study the observed behaviour.

The position we take in choosing to study non-linear map models that produce self-similarity we are agreeing implicitly with those who say that data traffic is self-similar. There are two very good reasons for agreeing with this position:

- all the causal effects (infinite knowledge) of a system can never be known; and
- in order to begin to understand the complexity of the behaviour that has been witnessed we need to abstract out as much of the complexity as possible but at the same time retain the complex behaviour. Using self-similar models goes some way to achieving this.

Moreover, this type model can also address the non-stationary aspect traffic modelling since these models can generate traffic and have their parameters altered on-line. This is in contrast to the FBM/FGN approach to traffic modelling since the traffic traces are generated off-line with their parameters fixed for the duration of the trace generation.

Finally, it is preferable to have an intuitive element in any approach used to modelling self-similar traffic. This intuitive appeal can have far reaching consequences in as much as it can lead to insight that permits problems to be approached in fresh ways. Non-linear models permit this kind of approach because there is an intuitive connection to the underlying physical *ON-OFF* process, which may render a new perspective to traffic modelling/control problems, i.e. that of chaotic dynamics and chaotic control. This thesis sets out to exploit the potential that these non-linear models have by furthering their analysis and the development of source aggregation methods for chaotic map models.

3 Stochastic Self-similarity and Teletraffic Modelling

Stochastic self-similarity was first alluded to by Mandelbrot in the late 1950's and early 1960's. The usefulness of his approach to measuring highly variable traffic only came to light and popularity in the late 1980's and early 1990's through the results presented by Fowler [FOW91]. Stochastic self-similarity as applied to time series is the preservation of the probabilistic structure of the time series as the time scale over which measurements are taken increases. The degree of stochastic self-similarity in the probabilistic structure is expressed through the Hurst parameter H $(0 \le H \le 1)$ and can be seen in the scaling of the sample variance. A value of $H > \frac{1}{2}$ implies positive correlations and affects convergence of the sample variance and the correlation decay. This effect is termed long range dependence (LRD). Formulations exist which link the sample variance, the correlation decay and LRD to *H*. Stochastic models such as FBM/FGN are based on the definition of *H*. The estimation of *H* becomes increasingly difficult as *H* approaches its upper limit (1) because of its slow convergence, which in turn makes parameterisation of self-similar models troublesome. For this reason knowledge of the error bound on *H* is essential. An alternative to stochastic self-similar models is to aggregate *ON-OFF* models that have LRD in at least one of its states. The aggregation of this type of *ON-OFF* model tends to FBM. Chaotic map formulations of the *ON-OFF* models with LRD will be covered in the next chapter. This chapter introduces the fundamentals required to understand self-similarity, long range dependence, the Hurst parameter, and self-similar processes.

3.1 Traditional 1/N Sample Convergence

The difference between "traditional" time series and self-similar time series lies in the dependence of sample estimates to converge on their true values as a function of the number of samples. In "traditional" time series this convergence goes as N^{-1}, where *N* is the number of samples. In self-similar time series the convergence is much slower and goes as N^{-b} where $0 < b < 1$.

Traditionally we say that the sample variance decays as the inverse of the sample size. This occurs under the following conditions. Suppose we have taken a series of samples $\{X_1, X_2, \cdots, X_N\}$ which are independent and identically distributed (IID) and where

- the mean $m \equiv E(X_i)$ exists and is finite, and
- the variance $s^2 \equiv \mathrm{var}(X_i)$ exists and is finite.

Under these conditions, the sample mean is given by

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \,. \tag{3.1}$$

The variance of the sample mean is

$$\mathrm{var}\left(\overline{X}\right) = \mathrm{var}\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right) = \frac{1}{N}\sum_{i=1}^{N}\mathrm{var}\left(X_i\right) \tag{3.2}$$

and since by definition $s^2 = \mathrm{var}\left(X_i\right)$ then we can see that the variance of the mean decays as a function of the sample size $\mathrm{var}\left(\overline{X}\right) = \frac{s^2}{N}$ [†]. The type of behaviour leading from this assumption is given by the right hand side of Figure 2.1.

3.2 Stationarity and Ergodicity

The analysis of stochastic time series depends on the stationarity of the realisation process. Moreover, stationarity implies that the realisation process of the traffic is Ergodic. Both concepts are central to the analysis of stochastic time series.

A stationary process can be loosely defined as one in which the statistical properties of the process do not change over time. By this we mean that the system has reached equilibrium in the sense that realisations of the process at a particular time look as the same other realisation of the process at some time displaced from the original instance of measurement. Priestly [PRI94] terms this "Statistical Equilibrium". There are two types of stationarity that we are interested in. These are "*Completely stationary***"** and "*Stationary up to order m"*. As we shall see the former is a much stricter definition of stationarity. In practice the latter definition is more usually used.

Definition - Completely Stationary [PRI94]. A stochastic process {***X(t)***} is said to be completely stationary if for any admissible $t_1,t_2,\ldots,t_n$ and any *k*, the joint probability distribution of $\{\boldsymbol{X}(t_1),\boldsymbol{X}(t_2),\ldots,\boldsymbol{X}(t_n)\}$ is identical to the joint probability distribution of $\{\boldsymbol{X}(t_1+k),\boldsymbol{X}(t_2+k),\ldots,\boldsymbol{X}(t_n+k)\}$. Priestly further clarified this point by stating the following property

$$F_{X(t_1),\cdots X(t_n)}\left(x_1,\cdots x_n\right) \equiv F_{X(t_1+k),\cdots X(t_n+k)}\left(x_1,\cdots x_n\right) \tag{3.3}$$

where *F(.)* denotes the distribution function of the random variables $\boldsymbol{X}(t_i)$. What equ.(3.3) states is that in probability the process structure of a completely stationary process is invariant under time, i.e. any shift in the time origin does not affect the distribution function.

[†] We can say this because the definition of $\mathrm{var}\left(\overline{X}\right) = E\left[\overline{X}\right]^2 - \left(E\left[\overline{X}\right]\right)^2$, and because of the independence between samples we can say $\mathrm{var}\left(\overline{X}\right) = E\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right]^2 - \left(E\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right]\right)^2$. This reduces to $\mathrm{var}\left(\overline{X}\right) = \frac{1}{N^2}\sum_{i=1}^{N}\left\{E\left[X_i\right]^2 - \left(E\left[X_i\right]\right)^2\right\} = \frac{N\,\mathrm{var}\left(X_i\right)}{N^2} = \frac{\mathrm{var}\left(X_i\right)}{N}$ and hence $\frac{s}{\sqrt{N}}$.

The stationarity definition given above is, for practical purposes, a little too stringent. It is very difficult to obtain a real random process especially those realised in network traffic which do not posses fluctuations which would prevent them from being considered completely stationary.

Definition - Stationary up to order m [PRI94]. A stochastic process $\{\boldsymbol{X}(t)\}$ is said to be stationary up to order m if for any admissible $t_1,t_2,\ldots,t_n$ and any k, the joint moments up to order m of $\{\boldsymbol{X}(t_1),\boldsymbol{X}(t_2),\ldots,\boldsymbol{X}(t_n)\}$ exist and equal the corresponding joint moments up to order m of $\{\boldsymbol{X}(t_1+k),\boldsymbol{X}(t_2+k),\ldots,\boldsymbol{X}(t_n+k)\}$. That is to say

$$E\left[\left\{X(t_1)\right\}^{m_1}\left\{X(t_2)\right\}^{m_2}\cdots\left\{X(t_n)\right\}^{m_n}\right]=E\left[\left\{X(t_1+k)\right\}^{m_1}\left\{X(t_2+k)\right\}^{m_2}\cdots\left\{X(t_n+k)\right\}^{m_n}\right] \tag{3.4}$$

What equ.(3.4) states is that all moments of the process up to order m are independent of time. In particular what we seek is stationarity up to order $m=2$. This implies that the mean and variance of the process are independent of time i.e.

- the mean of the process: $E\left[X(t)\right]=m$ is a constant (i.e. independent of time)
- the second moment of the process: $E\left[X^2(t)\right]=m_2'$ is also a constant
- hence, the variance: $\mathrm{var}\left(X(t)\right)=m_2'-m^2=s^2$ is also independent of time.

Furthermore, we are in a position to simplify the time dependence of the covariance structures of the process so as to include the time of measurement only, i.e.

- $\mathrm{cov}\left\{X(t)X(s)\right\}\equiv E\left[X(t)X(s)\right]-m^2$. Because the mean is independent of time then the only time dependence that exists in the structure is due to t and s which amounts to the lag involved in the measurement.

Ergodicity is an important concept that was originally used in statistical mechanics (G. Birkoff). It deals with the behaviour of time averages of a system. If a system is ergodic then the time average over a limited number of finite points, N say, will produce an estimate of the average which as $N\to\infty$ will converge on the true average value. The implication of ergodicity is that the system does not have any "strange" states in which the system can remain indefinitely. Even if the system starts in a "strange" state it will eventually escape this state and behave in an average (equilibrium) like manner.[5] This is a difficult concept to prove in many real systems. For this reason ergodicity is often assumed (a reasonable assumption often borne out of measurement). In our case the system is the stochastic process which yields the behaviour of network traffic and we assume it to be ergodic. Under this assumption we expect measurements of network traffic taken over a set of measurement days to be representative of the network traffic behaviour.

[5] For more of an explanation of this see [GAR95] pp 61-63.

3.3 Self-similar Processes

There are many forms of self-similarity. The most obvious forms that one can usually imagine are geometric fractal shapes where the shape is repeated at all scales. As the observer zooms in and out on the shape he observes recursively the same shape. An example of this is a Koch curve (see [SCHR91] p8). In 1941 Kolmogorov introduced self-similar processes in the context of turbulence in fluids. Much later Mandelbrot and co-workers put self-similarity into a statistical context. Their motivation was to find a model that captured the high variability observed by Hurst in measurements of reservoir water levels. In stochastic processes the recursive nature of self-similarity cannot be readily seen. The fractal behaviour in a self-similar stochastic process lies in the relationship of the probability distribution to the number of measurements required to determine the distribution and can be described as preserving the probabilistic structure of a process regardless of the time scale.

3.4 Stochastic Self-similarity

Definition - Stochastic self-similarity. Let *Z(t)* be a stochastic process with continuous time parameter t. *Z(t)* is called self-similar with self-similarity parameter *H*, if for any positive stretching factor α, the re-scaled process with time scale at, $a^{-H}Z(a\,t)$ is equal in distribution to the original process *Z(t)*

$$Z(t) \underset{d}{=} a^{-H} Z(a\,t). \tag{3.5}$$

By rearranging this we can restate this as

$$a^{H} Z(t) \underset{d}{=} Z(a\,t) \tag{3.6}$$

where *H* is the Hurst parameter and, $H \in (0,1)$ and $\underset{d}{=}$ means equal in distribution. The discrete version of this for discrete samples, Y_n, of a self-similar process is

$$Y_t \underset{d}{=} t^{H} Y_1. \tag{3.7}$$

The Hurst parameter describes the degree of self-similarity. For *H>0.5* the realisation of the process has positive correlations. For *H< 0.5* the realisation of the process has negative correlations. For *H=0.5* the realisation of the process is totally random in the sense that the realisations produce uncorrelated random variates. An example of a process with *H=0.5* is traditional Brownian motion.

In practice, stochastic self-similarity means that the traces of *Z(t)* and *Z(a t)* look qualitatively the same up to the scaling /stretching factor used. However, it should be noted that, unlike the self-similarity present in fractal sets such as Mandelbrot and Julia sets in which the detail of a picture is preserved exactly as we zoom in and out from the picture, stochastic self-similarity preserves the probabilistic structure of the picture as we zoom in and out in terms of time scales.

3.5 Partial Sums

An important factor in determining whether a time series is stochastically self-similar lies in the behaviour of its partial sums. In this respect we expect to see some form of scaling of the stochastic attributes of the time series over subsets of the complete time series. In order to explain this further we define stationary increments of a time series.

Definition - Stationary Increments of a Time Series [BER94]. If for any $k \geq 0$ and any k time points, $t_1, t_2, \cdots, t_k$, the distribution of $\left(Z(t_1+c)-Z(t_1+c-1), \cdots, Z(t_k+c)-Z(t_k+c-1)\right)$ does not depend on $c \in R$, then we say that *Z(t)* has stationary increments.

The above definition is needed so that we can state the following theorem which links partial sums which we have used earlier in the magnification of the time series to the stochastic self-similarity parameter *H*. The theorem is stated without proof.

Theorem - Relationship of normalised partial sums to the Hurst parameter H [BER94]. Suppose that *Z(t)* is a stochastic process such that $Z(1) \neq 0$ with positive probability and *Z(t)* is the limit in distribution of the sequence of normalised partial sums

$$\frac{S_{Nt}}{a_N} = \frac{1}{a_N}\sum_{i=1}^{[Nt]} X(i), \quad n = 1,2,\cdots . \tag{3.8}$$

Where $[Nt]$ indicates the integer part of $N t$, $X(1), X(2), \cdots$ is a stationary sequence of random variables and $a_1, a_2, \cdots$ is a sequence of positive normalising constants such that $\ln(a_N) \to \infty$. Then there exists a $H > 0$ such that for any $b > 0$,

$$\lim_{N\to\infty} \frac{a_{Nb}}{a_N} = b^H . \tag{3.9}$$

This is an important theorem since it links the scaling through partial sums of a time series $X(t)$ to the self-similarity parameter *H*. All self-similar processes with stationary increments and $H>0$ can be obtained by partial sums from times series of the type defined above.

3.6 Symptoms of Self-similarity

We can assess heuristically whether data gathered from a system is self-similar. There are certain symptoms that prevail in systems which are stochastically self-similar. These symptoms usually include the following:

- The variance of the sample mean appears to decay to zero at a slower rate than $1/N$ where N is the sample size. Moreover the rate of decay is more likely to be of the form $1/N^{\alpha}$ where a exits in the interval $a \in (0,1)$.

- The sample correlations $\hat{r}(k) = \hat{g}(k)/\hat{g}(0)$ decay at a rate proportional to the lag k i.e. $\hat{r}(k) \approx 1/k^{a}$ again for some a in the interval $a \in (0,1)$.
- The periodogram and spectral profiles of the sample data are divergent as the frequency tends to zero, $f(w) \approx w^{-a}$. More specifically, on a doubly-logarithmic plot of amplitude against the frequency we observe a decay with a negative slope for some a in the interval $a \in (0,1)$.

Figure 2.1 shows typical traffic traces that display stochastic self-similarity and 1/*N* sample convergence. The left-hand side of the figure displays stochastic self-similarity, i.e. the upper and lower traces appear the same regardless of the aggregation level. The right-hand side of the figure shows a good example of a stochastic process in which the assumption that the variance has an inverse dependence on the sample size holds.

3.7 Long Range Dependence

LRD can be thought of in traffic measurements as there still being a high degree of correlations between measurements long after we expected there to be none. Unfortunately there is no absolute dividing line between LRD and Short Range Dependence (SRD). As we shall see later Taqqu [TAQ86,97] and Willinger [WIL97] have connected the aggregation of traffic sources which display LRD to stochastic self-similarity. LRD can be defined in the time and frequency domains. Both definitions are equivalent.

Definition - Correlation decay [BER94]. Let $X(t)$ be a stationary process for which the following holds. There exists a real number α in the interval $\alpha \in (0,1)$ and some constant $C_\rho > 0$ such that

$$\lim_{k\to\infty} r(k) = \frac{1}{C_r k^{a}}, \tag{3.10}$$

where $C_r = \frac{C_g}{s^2}$ and $C_g = 2C_f \Gamma\left(2(1-H)\right)\sin\left(p\left(H-\tfrac{1}{2}\right)\right)$. Then $X(t)$ is called a stationary process with long memory or LRD or strong dependence, or a stationary process with slowly decaying correlations or long range correlations.

Definition - Spectral decay [BER94]. Let $X(t)$ be a stationary process for which the following holds. There exists a real number α in the interval $\alpha \in (0,1)$, and a constant $C_f > 0$ such that

$$\lim_{w\to\infty} f(w) = \frac{1}{C_f |w|^{a}}, \tag{3.11}$$

where $C_f = \frac{s^2}{p} C_r \Gamma(2H-1)\sin\left(p(1-H)\right)$ and $s^2 = \mathrm{var}\left(X(t)\right)$. Then $X(t)$ is called a stationary process with long memory or LRD or strong dependence, or a stationary process.

The definitions given above merely state what has been observed in terms of decay with increasing lag and the shape of the spectral decay as the frequency tends to zero.

3.8 Hurst Parameter's Relationship to Decay Rates

What the preceding definitions have not said is how these decays relate to the Hurst parameter. What we now give are two theorems, given without proof, of how these decays relate to the Hurst parameter.

Theorem - Hurst parameter's relation to Spectral decay [BER94]. Suppose that the definition of correlation decay holds with the following constraint on a: $0 < a = 2(1-H) < 1$. Then the spectral density exists and asymptotically has the following form

$$\lim_{w\to\infty} f(w) = \frac{1}{C_f(H)|w|^{1-2H}} \tag{3.12}$$

where $C_f = \frac{s^2}{p} C_r \Gamma(2H-1)\sin\left(p(1-H)\right)$ and $s^2 = \mathrm{var}\left(X(t)\right)$.

Theorem - Hurst parameters relation to Correlation decay [BER94]. Suppose that the definition of the spectral decay holds with the following constraint on a: $0 < a = 2H-1 < 1$. Then the correlation decay exists and asymptotically has the following form

$$\lim_{k\to\infty} r(k) = \frac{1}{C_r k^{2(H-1)}} \tag{3.13}$$

where $C_r = \frac{C_g}{s^2}$ and $C_g = 2C_f \Gamma(2(1-H))\sin\left(p\left(H-\frac{1}{2}\right)\right)$.

The two previous theorems rely on the correlation and spectral decays for their relationship to the Hurst parameter. Another statistical property of a time series which it is convenient to relate to the Hurst parameter is the variance.

Theorem - Hurst parameters relation to the variance of the time series [BER94]. Let $X(t)$ be a stationary process with LRD. Then

$$\lim_{N\to\infty} \mathrm{var}\left(\sum_{i=1}^{N} X(i)\right) = \frac{C_g N^{2H}}{H(2H-1)}, \tag{3.14}$$

where N is the number of samples and $C_g = 2C_f \Gamma(2(1-H))\sin\left(p\left(H-\frac{1}{2}\right)\right)$.

Observation: Here we note that all of the relationships to H are of an asymptotic nature. This has a practical significance which is most ably demonstrated by the spectral decay. In order to measure high values of H we have to wait a long time for the very rare events to "cycle" by.

This means that for high H we have to wait for a (very) long time for estimates of H to converge.

3.9 Stochastic Self-similar Models

The two most common stochastic self-similar models were developed by Mandelbrot and co-workers in the 1960's [MAN63, 65, 68a, 68b, 68c]. These models are FBM and FGN. FBM is a non-stationary stochastic process that was developed as a generalisation of the standard Brownian motion model. FGN is a stationary process. FGN is related to FBM since FGN is produced by taking the differences in FBM realisations.

FBM is a type of Gaussian random function based on standard Brownian motion. If $B(t)$ is standard Brownian Motion and H is some parameter in the range $0<H<1$ then FBM of exponent H is a weighted moving average process of $dB(t)$ in which past increments are weighted by a kernel function $(t-s)^{H-\frac{1}{2}}$, i.e.

$$B_H(t) = \frac{1}{\Gamma\left(H+\frac{1}{2}\right)} \int_{-\infty}^{t} (t-s)^{H-\frac{1}{2}} dB(s) . \tag{3.15}$$

We can see from this structure how LRD arises to give self-similar behaviour. The realisation of the self-similar process takes into account all past realisations.

Definition - Fractional Brownian Motion [MAN68a]. Fractional Brownian Motion of Hurst exponent $0 < H < 1$ is a zero mean Gaussian process $\left\{B_H(t), t \in R\right\}$ such that

1. $B_H(0) = 0$
2. $B_H(t) \underset{d}{=} N\left(0, s|t|^H\right)$
3. $B_H(t+d) - B_H(t) \underset{d}{=} N\left(0, s|d|^H\right)$.

That is to say: normal distributed non-stationary zero mean process that has a variance that scales exponentially with H. FGN is the incremental process of FBM and as such it is a stationary process.

Definition - Fractional Gaussian Noise [MAN68a]. Fractional Gaussian Noise of Hurst exponent $0 < H < 1$ is a zero-mean Gaussian process $\left\{G_{H,d}(t), (t,d) \in R \times R_+\right\}$ defined by

$$G_{H,d}(t) = \frac{1}{d}\left(B_H(t+d) - B_H(t)\right). \tag{3.16}$$

Having defined stochastic self-similar processes we now go on to examine the covariance structures of such processes.

3.10 Auto-Covariance Structure of Self-similar Processes

The covariance, $\gamma(.)$, is first product moment about the means of two random variables and is defined as

$$g(t_1,t_2) \equiv E\left\{\left[X(t_1)-m\right]\left[X(t_2)-m\right]\right\}.$$

(3.17)

We could compare displaced samples of the same series. This is termed the auto-covariance and is defined as

$$g(t_1,t) \equiv E\left\{\left[X(t)-m\right]\left[X(t+t)-m\right]\right\}.$$

(3.18)

The auto-correlation, $r(.)$, is defined in terms of the autocovariance,

$$r(t) \equiv \frac{g(t)}{g(0)}.$$

(3.19)

We recall that FBM is a zero-mean process with stationary increments. Consider discrete samples Y_t, Y_s of an FBM process with $s<t$. The covariance function is

$$g_Y(t,s) = \mathrm{cov}\left(Y_t, Y_s\right).$$

(3.20)

Because FBM is a zero-mean process we can write the variance as

$$s^2 = E\left[\left(Y_t - Y_{t-1}\right)^2\right] = E\left[Y_1^2\right],$$

(3.21)

and because of stationary increments we can say

$$E\left[\left(Y_t - Y_s\right)^2\right] = E\left[\left(Y_{t-s} - Y_0\right)^2\right] = E\left[Y_{t-s}^2\right].$$

(3.22)

Using the self-similar property, see equ.(3.7), and the result of equ.(3.21) then we write

$$E\left[Y_t^2\right] = t^{2H} E\left[Y_1^2\right] = t^{2H} s^2$$

(3.23)

and equ.(3.22) as

$$E\left[Y_{t-s}^2\right] = (t-s)^{2H} E\left[Y_1^2\right] = (t-s)^{2H} s^2.$$

(3.24)

The covariance can be written in an alternative manner

$$E\left[\left(Y_t - Y_s\right)^2\right] = E\left[Y_t^2\right] + E\left[Y_s^2\right] - 2g_Y(t,s).$$

(3.25)

By substitution of equations (3.23) and (3.24) in the above we obtain

$$(t-s)^{2H} s^2 = t^{2H} s^2 + s^{2H} s^2 + 2g_Y(t,s).$$

(3.26)

Hence

$$g_Y(t,s) = \frac{s^2}{2}\left\{t^{2H} + s^{2H} - (t-s)^{2H}\right\},$$

(3.27)

which is the auto-covariance structure of FMB.

Recall that FGN is the incremental process of FBM, $X_t = Y_t - Y_{t-1}$, which is stationary and has a zero-mean. The auto-covariance function is then

$$g_X(k) = \mathrm{cov}(X_t, X_{t+k}) = \mathrm{cov}(X_1, X_{1+k}) = E[X_t X_{t+k}]$$

(3.28)

because of stationarity. We can write the product term $X_t X_{t+k}$ as

$$\begin{aligned} X_t X_{t+k} &= X_1 X_{1+k} \\ &= \left(\sum_{j=1}^{k+1} X_j\right)^2 + \left(\sum_{j=2}^{k} X_j\right)^2 - \left(\sum_{j=1}^{k} X_j\right)^2 - \left(\sum_{j=2}^{k+1} X_j\right)^2. \end{aligned}$$

(3.29)

By using equ.(3.29) and substituting X_j by $Y_j - Y_{j-1}$ in the summations and expanding we can write the auto-covariance as

$$g_X(k) = \frac{1}{2}\left\{E\left[\left(Y_{k+1} - Y_0\right)^2\right] + E\left[\left(Y_{k-1} - Y_0\right)^2\right] - E\left[\left(Y_k - Y_0\right)^2\right] - E\left[\left(Y_k - Y_0\right)^2\right]\right\}.$$

(3.30)

We use equ.(3.24) in the above and obtain

$$g_X(k) = \frac{s}{2}\left\{(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right\}.$$

(3.31)

We are interested in the behaviour as $k \to \infty$. To do this we write equ.(3.31) as

$$g_X(k) = \frac{sk^{2H}}{2}\left\{\left(1+\tfrac{1}{k}\right)^{2H} - 2 + \left(1-\tfrac{1}{k}\right)^{2H}\right\}$$

(3.32)

and define the function *g(.)*

$$g(x) \equiv (1+x)^{2H} - 2 + (1-x)^{2H}.$$

(3.33)

The covariance is then

$$g_X(k) = \frac{sk^{2H}}{2} g\left(\frac{1}{k}\right).$$

(3.34)

We apply Taylor's expansion to the function *g(.)* and obtain

$$g_X(k) \approx sH(2H-1)k^{2H-2}.$$

(3.35)

This is the asymptotic auto-covariance structure of FGN. We can see clearly form equ.(3.27) that FBM is not stationary and from equ.(3.35) that FGN is stationary. For more details on the derivation of the covariance structures see Appendix A.

3.11 Estimators of the Hurst Parameter

The study of statistical estimators of *H* is a subject in it self. There are many estimators of the Hurst parameter. Some are easier to understand than others. The first statistical method for estimating *H* was developed by Hurst in an attempt to capture the scaling of the variability seen in hydrographic records. The method he developed is called the Rescaled Range Statistic, R/S, method. This method is a good example of the partial sum approach to estimating *H* (see section 3.5). As we have seen earlier there are many relationships in terms of decay to the Hurst parameter (for example the variance, correlation and spectral decays), all of which can be adapted to give an estimate of *H*. The interested reader is referred to [BER94] Chapter 4 for more detail and discussion of the subject. However not all these methods are Maximum Likelihood Estimators (MLE). We are primarily concerned with MLE because some form of error bound on the estimate can be developed. Until recently there was only one reliable estimator of *H*. This was Whittles MLE (see [BER94] Chapters 5 and 6). The main drawback with Whittles MLE was its evaluation time. Recently newer methods have been developed for estimating *H* based on wavelet transforms [ABR95, 98]. These methods are more of a natural measure of *H* since wavelet transformations actually rely on scaling to work, see for example Figure 1 in [ABR95 p19]. In this thesis we have used the wavelet method developed by Abry *et al* for the measurement of *H*.

For a given octet range j_1 to j_2 the wavelet based estimator of *H* is given by the following components. A weight term given by

$$S_j = \frac{n \ln^2 2}{2^{j+1}} \tag{3.36}$$

where *n* is the number of samples (must be a power of 2) and *j* is the octave in question, i.e. S_j is the weighted sum of the number of samples of the octave in question. $d_x(j,k)$ is the k^{th} wavelet coefficient for the octave *j*. The squared average coefficient wavelet value in $\log_2$ form given by

$$h_j = \log_2\left(\frac{1}{n_j}\sum_{k=1}^{n_j}\left|d_x(j,k)\right|^2\right). \tag{3.37}$$

The estimator is then defined as

$$\hat{H}(j_1, j_2) \equiv \frac{1}{2}\left[\frac{\sum_{j=j_1}^{j_2} S_j \sum_{j=j_1}^{j_2} j h_j S_j - \sum_{j=j_1}^{j_2} j S_j \sum_{j=j_1}^{j_2} h_j S_j}{\sum_{j=j_1}^{j_2} S_j \sum_{j=j_1}^{j_2} j^2 S_j - \left(\sum_{j=j_1}^{j_2} j S_j\right)^2} + 1\right]. \tag{3.38}$$

The variance is given by

$$\operatorname{var} \hat{H}(j_1, j_2) = \frac{2}{n_{j_1} \ln^2 2} \frac{1 - 2^{(j_2 - j_1)}}{\left[1 - 2^{-(j_2 - j_1)}\left((j_2 - j_1)^2 + 4\right) + 2^{-2(j_2 - j_1)}\right]}. \tag{3.39}$$

The confidence intervals are then

$$\hat{H} - Z_b \sqrt{\operatorname{var} \hat{H}(j_1, j_2)} \le H \le \hat{H} + Z_b \sqrt{\operatorname{var} \hat{H}(j_1, j_2)}\ , \tag{3.40}$$

where Z_b is the $1-b$ quantile of the standard Gaussian distribution.

The estimator is speedy and can be programmed up relatively easily. See [PRE94] for routines which render the wavelet transform. This estimator has been used as the preferred method of measuring *H* used in this thesis.

3.12 ON-OFF Self-similar Traffic Modelling

Recently a connection has been established between *ON-OFF* sources and the appearance of self-similar traffic in aggregated traffic streams [TAQ97, WIL97]. The theorem developed by Taqqu *et al* [TAQ97] can be stated as follows:

> *"The superposition of many strictly alternating independent and identically distributed ON/OFF sources (packet trains) each of which exhibits the Noah effect*[6] *results in self- similar traffic"*

The implication of this to teletraffic is that the *ON-OFF* model has sojourn lengths in the *ON* and/or *OFF* states that can be very large with non-negligible probability. This means that if we are to model this type of traffic we require source models to exhibit sojourn characteristics over a wide range of time scales. In effect the sojourn characteristic of the stochastic model has to be governed by the Noah effect. The Noah effect can be modelled by heavy tailed distributions with infinite variance or truncated state distributions. The importance of the Noah effect cannot be understated. In fact Taqqu identifies the Noah effect as:

> *"essential point of departure from traditional to self-similar traffic modelling"*.

This means that the intensity of the tail of the sojourn distribution controls the variance of the distribution. Moreover there is a relationship between the parameter which describes the tail intensity and the Hurst parameter, *H*. In reality the tail intensity describes the LRD and *H* describes the overall self-similarity[7].

[6] This is an analogous reference to the variability in Biblical flood levels witnessed by Noah.
[7] Occasionally *H* is termed the Joseph effect. This is another analogous reference to the Biblical frequency of abundance and famine witnessed by Joseph.

The practical significance of Taqqu's results is that when there is a large enough number of *ON-OFF* sources, the aggregated traffic converges to a second order self-similar process within the bounds of natural cut off limits. These natural limits are determined by the following:

- lower cut off - this is a reflection of the behaviour of the medium access control (MAC) protocol,
- upper cut off - this is a reflection of the natural diurnal variations.

Taqqu's theory relies on the superposition of alternating renewal processes which have retention probability distributions which can be characterised by slow varying functions. The aggregation of many such processes leads in the limit to FBM. Each reward process represents a cell/packet train. The reward process is defined in a strict alternating sense, i.e. *ON* periods are always followed by *OFF* periods and viceverca. This strict alternation conforms to the accepted *ON-OFF* traffic source models. The reward process for an individual source is defined as:

$$\{W(t), t \geq 0\} \quad W(t) = \begin{cases} 1, & \textit{while source is in ON state} \\ 0, & \textit{while source is in OFF state} \end{cases}. \tag{3.41}$$

W(t) as the reward at time *t*. *W(t)* applies to the whole period in which the source is either *ON* or *OFF*. A further simplification is to assume that the lengths of the *ON* periods are IID as are the lengths of the *OFF* periods[8]. Taqqu then examines the time series of the aggregated reward process of *M* traffic sources $\{W^{(M)}(t), t \geq 0\}$. This is the superposition of the cumulative call/packet counts from the *M* sources in the given time *t*,

$$W_M(t) \equiv \sum_{m=1}^{M} W^{(m)}(t). \tag{3.42}$$

If we assume that the aggregated process scales (i.e. has self-similar properties), then the cumulative cell/packet count can be re-scaled by a factor *T*,

$$W_M^*(Tt) = \int_0^{Tt} \left(\sum_{m=1}^{M} W^{(m)}(t) \right) du. \tag{3.43}$$

This is the aggregated cumulative packet count in interval [0,*Tt*). It is the stochastic behaviour of $W^*_M(Tt)$ as *T* and *M* get large, i.e. as the time interval and the number of aggregated sources increases, that is of interest. The interested reader is referred to [FEL66] Chapter 11 for a general treatment of renewal theory and [COX67] Chapter 6 for the superposition renewals.

[8] This condition can be relaxed without affecting the end result in that *ON-OFF* periods may have different distributions.

Taqqu's theorem requires the *ON-OFF* distributions to be chosen such that the cumulate traffic converges to FBM or the aggregated traffic (without cumulating) converges to FGN. In order to satisfy these conditions the *ON-OFF* distributions must be chosen such that as $M \to \infty, T \to \infty$ the aggregated process $\left\{W_M^*(Tt), t \geq 0\right\}$ adequately normalised becomes $\left\{s_{\lim} B_H(t), t \geq 0\right\}$, where $B_H(t)$ is FBM and has covariance structure dependent on H, and s_{lim} is a finite normalising constant determined by the behaviour of the sojourn distributions in the *ON* and *OFF* states. The normalising constant required by Taqqu's theorem depends on the complementary distribution of the sojourn time having the characteristic of a slow varying function at infinity. A function $L_j(t)$ is slowly varying at infinity if for any multiplying factor $T > 1$ the following limit holds, $\lim_{t\to\infty} \frac{L_j(Tt)}{L_j(t)} = 1$. The complementary distribution of the sojourn time is the probability of remaining in the *ON* or *OFF* state. In the general case of the *ON* or *OFF* distributions, if the probability density is denoted by $f_j(t)$, then its distribution is given by $F_j(t) = \int_0^t f_j(x)dx$ and the complementary distribution is given by $F_{jc}(t) = 1 - F_j(t)$.

It is $F_{jc}(t)$ which must be heavy tailed and slowly varying at infinity. For $t \to \infty$ this function can be approximated as $F_{jc}(t) \approx \frac{l_j}{t^{a_j}} L_j(t)$. Here we note that $l_j > 0$ is a constant and a_j is the tail decay rate. The subscript j, j=1,2 refers to the *ON* and *OFF* states respectively.

For $1 < a < 2$ the variance s_j is infinite and we define the constant a_j to be $a_j = \frac{l_j \Gamma(2 - a_j)}{(a_j - 1)}$.

For $a > 2$, s_j is finite and we set a_j=2, l_j=1 and $a_j = s_j^2/2$. The normalisation constant $\sigma_{\lim}$ then depends on the ratio of the slow varying function in the *ON* and *OFF* states. If we define b as

$$b = \lim_{t\to\infty} \frac{t^{a_1}}{t^{a_2}} \frac{L_1(t)}{L_2(t)} \tag{3.44}$$

then if $a_1 = a_2$ and $0 < b < \infty$, in this case $a_{\min} = a_1 = a_2$. Note $\alpha_{\min}$ sets the largest value of H and hence the dominant term is given by

$$s_{\lim}^2 = \frac{2\left(m_2^2 a_1 b + m_1^2 a_2\right)}{(m_1 + m_2)^3 \Gamma(4 - a_{\min})} \text{ and } L = L_2. \tag{3.45}$$

On the other hand if $a_1 > a_2$ then b=0 or if $a_1 < a_2$ then b=∞, then we have a dominant a term and

$$s_{\lim}^{2} = \frac{2m_{\max}^{2} a_{\min}}{(m_1+m_2)^3 \Gamma(4-a_{\min})} \text{ and } L=L_{min}.$$

(3.46)

where *min* is the index 1 if $b=\infty$, i.e. *ON* has the dominant sojourn, and *min* is the index 2 if $b=0$, i.e. *OFF* has the dominant sojourn.

Theorem - Homogenous Case [TAQ97]. For large *T* and *M*, the aggregated cumulative traffic process behaves statistically like

$$W_M^*(Tt) = TtM\frac{m_1}{m_1+m_2} + T^H s_{\lim}\sqrt{L(t)M}.B_H(t)$$

(3.47)

where $H=(3-a_{\min})/2$ and $s_{\lim}$ is as outlined at equs.(3.45) and (3.46). In fact the theorem depends on convergence in probability as the limits of *T* and *M* are taken in the right order, that is

$$\lim_{T\to\infty}\lim_{M\to\infty} \frac{\left(W_M^*(Tt) - TtM\frac{m_1}{m_1+m_2}\right)}{T^H\sqrt{L(t)M}} = s_{\lim}B_H(t).$$

(3.48)

In the heterogeneous case we suppose that there are *R* distinctive types of *r* sources being aggregated, i.e. $r=1\cdots R$ types of sources. The characteristics of each source of type *r* are denoted $F_j^{(r)}, a^{(r)}, s^{(r)}, L^{(r)}$. If there is a total of *M* sources then there is a proportion $M^{(r)}/M$ of type *r* of the total number being aggregated and if we assume that this proportion is not negligible as $M\to\infty$, then under these assumptions we can modify the homogenous theorem in the following way.

Theorem - Heterogeneous Case [TAQ97]. For large $M^{(r)}$, $r=1\cdots R$ and large *T, the* aggregated cumulative packet/cell traffic behaves statistically as

$$W_M^*(Tt) = Tt\left(\sum_{r=1}^{R} M^{(r)}\frac{m_1^{(r)}}{m_1^{(r)}+m_2^{(r)}}\right) + \sum_{r=1}^{R} T^{H^{(r)}} s_{\lim}^{(r)}\sqrt{L^{(r)}(T)M^{(r)}}.B_{H^{(r)}}(t),$$

(3.49)

where $H^{(r)}=(3-a_{\min}^{(r)})/2$ and $B_{H^{(r)}}$ are independent fractional Brownian motions.

The implications of this are the following:

- Behaviour - the limit is a superposition of independent fractional Brownian motions with different Hurst parameters $H^{(r)}$.
- Fluctuations - the term with the highest $H^{(r)}$ (or smallest $a^{(r)}$) ultimately dominates as $T\to\infty$.

- The contribution of sources with finite variance is simply that of ordinary Brownian motion.

Here we note that these theorems were arrived at by taking the limits in the correct order (see equ.(3.48)). Taqqu has extended the theorems by taking the limits in the reverse order. However he points out the practical significance of this is (at the moment) unclear. For this reason we have only included the main result in its homogenous and heterogeneous forms.

As we have seen it is the aggregation of *ON-OFF* sources with LRD in at least one of the states that leads to self-similar traffic. *ON-OFF* sources can also be modelled via chaotic maps, moreover chaotic maps can have LRD built into the dynamics of the map. It is these attributes of chaotic maps that make them an attractive base for self-similar traffic modelling and will be the topic of the next chapter.

4 Teletraffic Modelling Using Non-linear Dynamical Maps

The application of non-linear dynamics theory to teletraffic modelling draws heavily on the theories of intermittency developed in the investigation of turbulence in thermodynamic systems. Teletraffic modelling is not only a new application for the existing theory, but has also required extensions to analyse the statistical characteristic behaviours relevant to a teletraffic interpretation of the models.

The initial investigations on non-linear teletraffic models were carried out by Pruthi, Erramilli and Singh [ERR94a, 94b, 95a, PRU95a, 95b]. They used non-linear rather than stochastic models in their search for alternative parsimonious models that displayed 2^{nd} order self-similarity. Pruthi and Erramilli advanced the use of non-linear teletraffic models by developing techniques for the determination of the models' invariant density, and through the study of the models' aggregate behaviour. The true value of non-linear models lies in their simplicity and speed of sample generation. However these models have their drawback in terms of their analytical tractability, in particular, for their behaviour near the point of bifurcation (for transit times when $e > 0$), the prediction of H in a coupled map, and the behaviour of the map under aggregation. These drawbacks are addressed in the chapters that follow, where contributions are made to transit time analysis, H prediction and aggregate map modelling.

Effective teletraffic models must have the following teletraffic properties: capacity to set the state sojourn time, the traffic load, and the degree of self-similarity (i.e. their scaling factor). The properties of non-linear dynamical models that relate to these teletraffic properties are transit time analysis, the invariant density and the spectral decay. These model properties also relate to the applicability of non-linear dynamical models as self-similar teletraffic models in the following way.

The transit time (and therefore its physical *ON/OFF* behaviour) can be made to have long run correlations that are heavy tailed. The heavy tailedness in its correlations leads to LRD, which is then reflected in self-similar behaviour dependent on H. Moreover there is an intuitive appeal in those non-linear dynamical models that link their underlying physical process to standard teletraffic *ON/OFF* model interpretations. The link between those non-linear dynamical maps and standard teletraffic models can be explained via analogy between a single intermittency map and a single *ON/OFF* teletraffic model. In the single intermittency map the physical meaning of transit time probability is the probable duration (in terms of the number of iterates) that an orbit of the map takes to leave the influence of one of its fixed points. This equates to the sojourn time of an *ON/OFF* teletraffic model, i.e. the time it takes for a source to leave the influence of a particular state. The invariant density of the single intermittency map captures the long run orbital behaviour of the map (in effect the long-term influence of

the fixed points on the map's orbit) and hence reflects the probability of the map being in a particular state i.e. its load. This is the equivalent of the probability density in an *ON/OFF* teletraffic model. Finally, the LRD nature of the single intermittency map is related to its transit time, and is reflected in its spectral density and correlation structures that are dependent on *H*. The same is true for the sojourn time of teletraffic *ON/OFF* models.

In this chapter we refer to non-linear dynamical models as "non-linear models", "non-linear map models" or "chaotic map models"; in the interests of brevity some times we simply call them "chaotic maps", "map models" or simply "maps".

4.1 Justification of Non-linear Map Models

An important observation of Fowler [FOW91] was that the bursts of traffic activity had no characteristic scale (they appeared scale invariant). Leland [LEL94] noted that the spectral decay of the traffic was divergent at the origin. Schuster [SCHU95 p94] noted that there have been many observations, in a variety of physical systems, of divergent spectra at the origin. This phenomenon is called *1/f* noise. Schroeder [SCHR91] points out that in many physical solid-state systems the 1/*f* power spectrum can be explained through their relaxation times[9]. In an analogy with solid-state physics, one could try to explain the 1/*f* spectrum observed in network traffic by equating the solid-state systems'

- energy barriers to the emission rates of the traffic sources, and
- excited state to the on probability of the traffic source.

Divergent spectra and scale invariance are also typical of critical phenomena. Willson first proposed the investigation of critical phenomena through Renormalization Group Theory (RNG) in the analysis of thermodynamic systems [WILS74, 75]. Garrod [GAR95 pp 268-269] states that the fundamental assumptions of Willson's initial analysis on the renormalization group treatment of critical phenomena were:

- at the critical point, the probability distribution of the fluctuations in the order parameter is scale invariant; and
- the scale invariant distribution can be obtained as a non-trivial (unstable) fixed point of a properly formulated renormalization transformation.

Non-trivial (unstable) fixed points describe the critical points of the associated thermodynamic system. At the critical point the correlation function has a power law decay. The stability properties of the fixed points are closely associated with the decay of the correlation function in the states described by those fixed points, i.e. the states have long range dependence (LRD). Unstable fixed points give power law decays, while stable fixed points describe states whose correlation function has an exponential decay, i.e. the states have short range dependence (SRD).

[9] An example of this is given by Schroeder in [SCHR91, p124].

A fundamental requirement for renormalization to work is self-similarity; this is because many critical phenomena in physics show self-similar behaviour near the critical point. These phenomena are amenable to renormalization theoretic treatment, yielding the critical exponents for the correlation length and other important parameters.

A striking characteristic of critical phenomena is the fact that certain detailed quantitative measures of a system's behaviour near a critical point are quite independent of the interactions between the parameters that go to make up the system. This characteristic is called universality. The chaotic maps that we use exhibit structural universality. Structural universality occurs where the pattern of the chaotic maps functions is retained as the function describing the dynamics is composed with itself.

The physical phenomena of turbulence displays scale invariance and 1/*f* noise. Turbulence has been extensively studied by non-linear dynamicists and theories of intermittency as explanations for turbulence have been put forward [POM80, HIR82a, KLA93, YAN94, HEA94, LUS96] some of which use RNG [HIR82b, PRO83] as pivotal point of their explanation. It is this theory which has been adapted to provide the non-linear models of bursty traffic used in this thesis.

For further explanations on intermittency and turbulence including various discussions and treatments on intermittency and the onset of chaos, the interested reader is referred to Chapter 4 in Schuster [SCHU95], section 8.2 in Ott [OTT94] and Chapter 10 in McCauley [MCC95].

4.2 Iterated systems

Central to the chaotic map models that we use is the notion of an iterated system. Iterated systems involve functions that take initial values as their argument and yield the next initial values as their results, i.e. $x_{n+1} = f(x_n)$. The function that operates on the argument is well described and confined to a given interval. The whole system is deterministic since the behaviour of the functions is known. For a good general treatment of iterated systems the interested reader is referred to [COL80].

4.2.1 Some Essential Definitions

These following definitions are after Gulick [GUL92].

Definition - Iterates of a function. Let *f* be function and let x_0 be in the domain of *f* (i.e. x_0 is a value in the interval over which the function exists) then

$$f(x_0) = f^1(x_0) \text{ = the first iterate of } x_0 \text{ for } f$$

$$f(f(x_0)) = f^2(x_0) \text{ = the second iterate of } x_0 \text{ for } f.$$

More generally if n is any positive integer, and x_n is the n[th.] iterate of x_0 for f, then $x_{n+1} = f(x_n)$ is the $(n+1)^{th.}$ iterate of x_0. This is usually denoted

$$x_{n+1} = f^{n+1}(x_0).$$

Definition - Orbit of a function. The infinite sequence of iterates of the function f, from an initial condition x_0, is termed the orbit of x_0, i.e.

$$\left\{f^n(x_0)\right\}_{n=0}^{\infty} = \text{orbit of } x_0 .$$

Definition - Fixed point. Let x_p be in the domain of f. Then x_p is a fixed point of f if

$$x_p = f(x_p).$$

Definition - Attracting fixed point. Let x_p be a fixed point of f. The point x_p is an attracting fixed point of f provided that there is an interval $(x_p-e\ ,\ x_p+e)$ containing x_p such that if x is in the domain of f and in $(x_p-e\ ,\ x_p+e)$ then $f^n(x) \to x_p$ as n ® ¥.

Definition - Repelling fixed point. Let x_p be a fixed point of f. The point x_p is a repelling fixed point of f provided that there is an interval $(x_p-e\ ,\ x_p+e)$ containing x_p such that if x is in the domain of f and in $(x_p-e\ ,\ x_p+e)$ but $x \neq x_p$ then $\left|f(x)-x_p\right| > \left|x-x_p\right|$.

The following theorem deals with the identification of the fixed point through the function f.

Theorem - fixed points. Suppose the function f is differentiable at the fixed point x_p. Then

1. if $\left|f'(x_p)\right| < 1$ then x_p is attracting.
2. if $\left|f'(x_p)\right| > 1$ then x_p is repelling.
3. if $\left|f'(x_p)\right| = 1$ then x_p can be either attracting, repelling or neither.

Definition - Periodic orbit. Let x_0 be in the domain of f. Then x_0 has period n if $f^n(x_0) = x_0$ and the iterates are distinct, i.e. if x_0 has period n then the orbit x_0 is

$$\left\{x_0, f^1(x_0), f^2(x_0), \cdots, f^n(x_0)\right\}.$$

This is occasionally termed an n-cycle periodic orbit.

We can see from these definitions that in a chaotic map the fixed points are where $x_n = x_{n+1}$. To make fixed points easier to spot a "reflection line" (where $x_n = x_{n+1}$) is introduced into the map diagram, see Figure 4.1. This also makes the orbit of the map easier to follow.

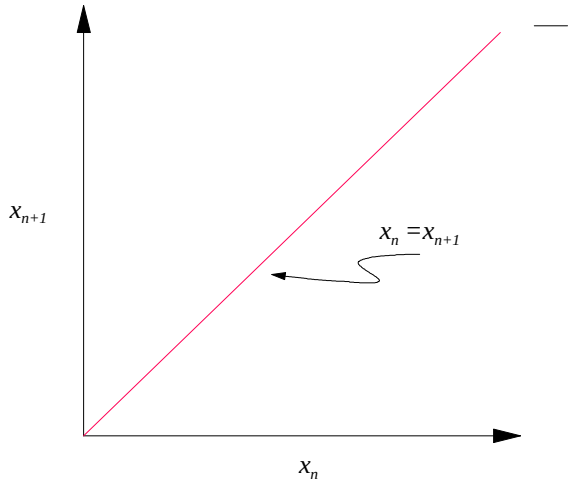

Figure 4.1 The "reflection line", $x_n = x_{n+1}$

4.2.2 Required Ingredient for Chaos in Iterated Systems

Generally for an iterated system to be chaotic it must fulfil the following criteria [DEV94]:

- the system has sensitive dependence on initial conditions,
- the periodic orbits of the system are dense in the space in which the system exists, and
- there exists an orbit of the system which is dense.

It is now widely accepted that the last two criteria imply the first. Note that by "dense" we mean that an orbit of the system will approach every point in the space within which the system exits with arbitrary accuracy.

4.2.3 Sensitive Dependence on Initial Conditions (SIC)

Sensitive Dependence on Initial Conditions (SIC) means that points which are initially very close together but separated by some small error, ε, will diverge exponentially as the system evolves with time. Mathematically this can be expressed as the difference between iterates of a chaotic map with starting points separated by some small error

$$e\, e^{Nb(x_0)} = \left| f^N(x_0 - e) - f^N(x_0) \right|. \tag{4.1}$$

The Lyupanov exponent, b, gives the rate of divergence of the points, and this can be expressed mathematically as

$$b = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln\left| f'(x_i) \right| = \frac{1}{n} \ln\left| f^{(n)'}(x_i) \right|. \tag{4.2}$$

All chaotic systems exhibit SIC.

4.3 Intermittency as the Basis for Map Models

The types of chaotic map chosen for investigation are those which display intermittency. Intermittency as defined by Schuster is:

"the occurrence of a signal that alternates randomly between regular (laminar) phases and relatively short irregular bursts"

Such signals have been detected in a large number of physical systems. Intermittency is connected to the transition to chaos in dynamical systems. Pomeau and Manneville [POM80] have classified intermittency in chaotic maps. The classification relies on the type of bifurcation that leads to chaos from the intermittent behaviour at the fixed point. The type of bifurcation depends on the manner in which their eigenvalues cross the unit circle (how they become unstable). Their classification is obtained in the following way. Suppose that there is an iterated function system that is dependent on some parameter *r*, i.e. $x_{n+1} = f_r(x_n)$. Furthermore the parameter *r* takes on some critical value r_c. Stable oscillations, $r < r_c$, correspond to stable fixed points. Above r_c this fixed point can be come unstable. There are three ways in which a fixed point can be come unstable. In all three the modulus of the eigenvalue becomes larger than the unit circle and the eigenvalue leads to the classification of the type of intermittency:

- Real Pair of eigenvalues cross the unit circle at +1 → TYPE I intermittency.
- Complex conjugate pair of eigenvalues cross the unit circle (at any point not ±1) → TYPE II intermittency.
- Real Pair of eigenvalues cross the unit circle at -1 → TYPE III intermittency.

Type I intermittency is associated with tangent node bifurcation (saddle node bifurcation). The $f^3(x)$ logistic map (see Figure 4.2) displays this type of bifurcation behaviour. The bifurcation occurs when two fixed points - one stable and the other unstable - merge. This occurs at the point of tangency (hence its name - see Figure 4.3). For:

- $r < r_c$ - the map has a stable fixed point
- $r = r_c$ - tangency is achieved and we have a single fixed point that is both attracting and repelling
- $r > r_c$ - there are no stable fixed points.

The width of the critical region *e*, $e = r - r_c$, determines how fast the orbit passes through the restriction. The narrower the restriction the longer the orbit takes to pass through the critical region. This type of behaviour is reminiscent of memory since the orbit appears to remember the fixed point being there. In this case we can see that long run correlations in the orbit values can build up and LRD behaviour appears for sufficiently small values of *e*.

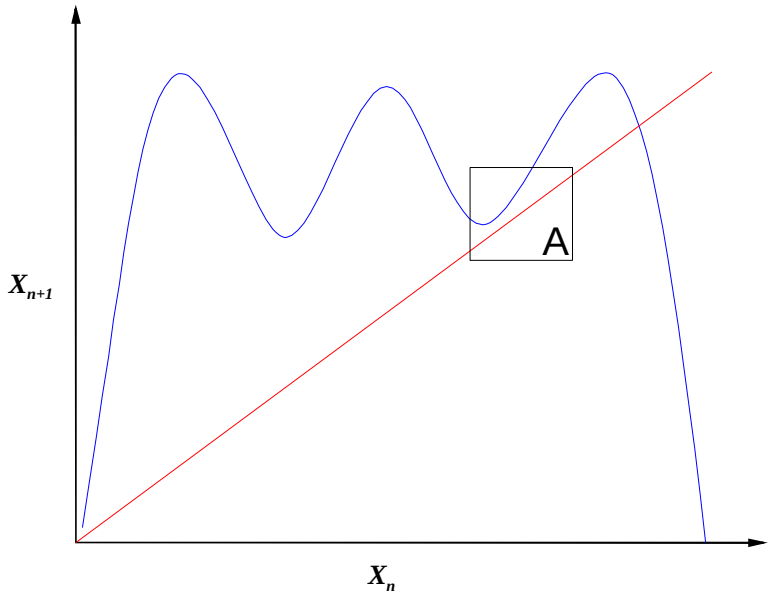

Figure 4.2 Logistic Map: $f_r(x_n) = rx_n(1-x_n)$. Displayed is the $f_r^3(x)$, (third iterate) logistic map showing region of intermittency -A. The logistic map depends on the parameter *r* for its behaviour

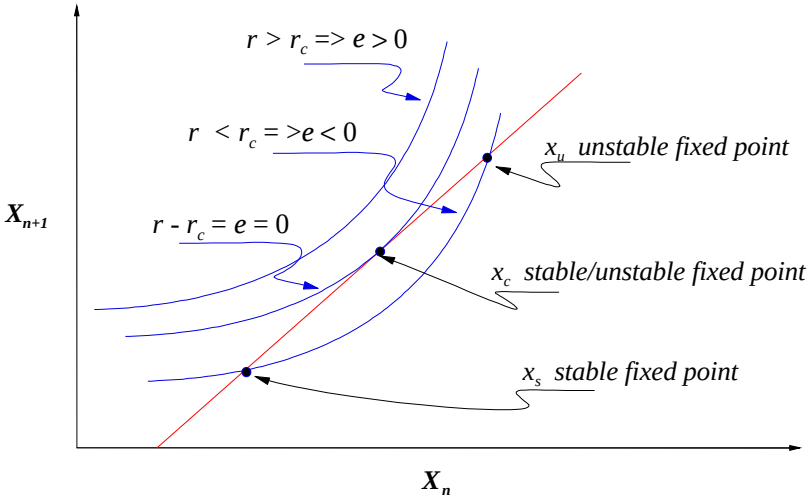

Figure 4.3 Detail of Figure 4.2 around the Region of Intermittency, A, showing: stable fixed pints x_s; unstable fixed points x_u and marginally stable fixed point x_c at the point of criticality. The parameter also shows the relationship between the parameter *e* and the map parameter *r* of the logistic map and its value at criticality, r_c.

In short, any dynamical system that displays inverse tangent bifurcation has Type I intermittency. A map of such a system is the logistic map. The chaotic map models that we will use are constructed from this behaviour. Such models are constructed in one of two ways. Either by:

- approximating the curve at tangency via a Taylor series expansion around the point of tangency, or

- using the structural universality present in the dynamical system at the point of criticality. In this case we have scale invariance in the correlations and use a renormalization group approach to obtain the deterministic equations of interest.

In either event the properties of the system should be able to be derived from either approach, although some properties may be easier to deduce using one method rather than the other.

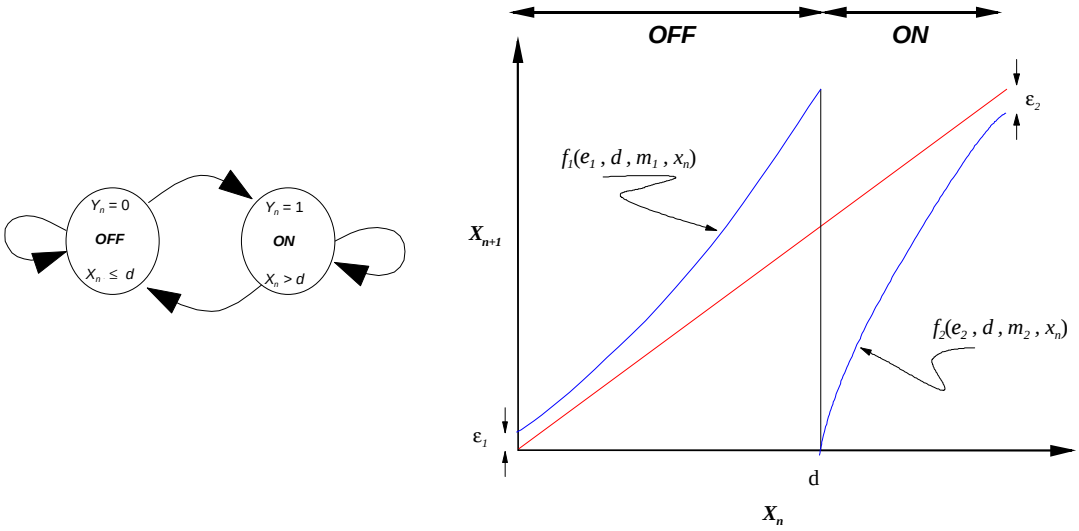

Figure 4.4. *ON/OFF* Source Model interpretation by use of the indicator variable and intermittency map functions

4.4 Chaotic Map Models - Derivation

As we have seen in Section 4.2.2. for the map to be chaotic it must have SIC. In iterated systems SIC arises out of the mixing of the iterates due to some stretching and folding action of the map function. It is the mixing which gives the orbit it apparent randomness. To achieve this mixing we have to have a map function that not only renders intermittent behaviour but one which also incorporates a method for obtaining the mixing as part of the system description. Manneville [MAN80] proposed such a system for maps with Type I intermittency constructed on the unit interval as a method for simplifying the analysis of intermittency. His proposal was to introduce an intermittency map with two functions, $F_1(.)$ and $F_2(.)$, which mutually inject into the critical regions of the map when the iterate value passes some discriminate value, d. $F_1(.)$ is restricted to the interval [0,d] and $F_2(.)$ to the interval (d,1]. Their mutual interaction performs the necessary mixing. Both functions perform stretching: $F_1(.)$ performs a stretching on iterate values in (0,d] to a values in (0,1), while $F_2(.)$ stretches iterates in (d,1] to (0,1). The folding occurs on transitions from $F_1(.) \rightarrow F_2(.)$ and vice versa, i.e. $F_2(.)$ injects into interval (0,d) while $F_1(.)$ injects into interval(d,1). The structure of the map models can be seen in Figure 4.4. From this construction we can see that in order to obtain map functions that yield intermittent behaviour the functions must comply with the following boundary conditions on the interval:

$$\begin{aligned} F_1(0) &= 0, \quad F_1(d) = 1 \\ F_2(1) &= 1, \quad F_2(d) = 0 \end{aligned} . \tag{4.3}$$

4.4.1 Derivation: Taylor's Series Approach

The logistic map is described by the following equation

$$f_r(x) = xr(1-x) \tag{4.4}$$

where *r* is some control parameter. The map can be iterated many times. The behaviour of the map in the region labelled A in Figure 4.2 is the area of interest. This is since the curve of the logistic map and the reflection line form a channel through which the orbits of the system have to pass. It has been shown that this mathematical description is akin to laminar flow in fluids, with ejection from region A being similar to the onset of turbulence [MANN80]. In this context the laminar flow phase of the orbit corresponds to the sojourn time of a traffic source and the emission of the orbit from this region to the arrival of a full cell/packet. The curve of the function near tangency of the logistic map can be approximated by a Taylor series expansion about the point of tangency, x_c. For the third iterate of the logistic map this is:

$$f_r{}^3(x_c - (x - x_c)) \approx f_r{}^3(x_c) + (x - x_c) f_r{}^{3'}(x_c) + \frac{(x - x_c)}{2!} f_r{}^{3''}(x_c) + \cdots + (r - r_c) f_{r_c}{}^{3'}(r_c) + \cdots . \tag{4.5}$$

By letting

$$a_c = \frac{f_r{}^{3'}(x_c)}{2} \quad and \quad b_c = f_{r_c}{}^{3'}(r_c) \tag{4.6}$$

and defining the following

$$y_n \equiv \frac{(x_n - x_c)}{b_c},\ a \equiv a_c b_c > 1,\ e \equiv (r - r_c) \tag{4.7}$$

equ.(4.5) becomes

$$y_{n+1} = a y_n{}^2 + y_n + e\ . \tag{4.8}$$

To determine the value of the constant term, *a*, we apply the boundary conditions, obtaining

$$a = \frac{1 - e - d}{d^2}\ . \tag{4.9}$$

4.4.2 Derivation: Renormalization Group Approach

In this approach we commence with a generalised Taylor's series approximation to the curve at tangency for the unperturbed system $x_{n+1} = x_n + a x_n{}^m$ and in functional form $f(x) = x + ax^m$. For there to be functional self-similarity (structural universality) we need the following condition $f^n(x \to 0) = x + ax^m$. This can be seen in the following example. For the second iterate of the generalised case we have $f^2(x) = (x + ax^m) + a(x + ax^m)^m$, and since $x^m \to 0$ faster than $x \to 0$, we regain, after

appropriate scaling, our original function $f^2(x \to 0) = x + ax^m$. So we can see that there is structural universality present in the generalised function. Furthermore, if repeated operation of the doubling operator *T* leads to a fixed point $f^*(x)$ of *T*, then $f^*(x)$ has the required structure for structural renormalization i.e. $Tf^*(x) = af^*\left[f^*\left(\frac{x}{a}\right)\right] = f^*(x)$, where the term a determines the class of structural universality. Following Hu and Rudnick [HU 82] we write the following recursion relation $f(x) = x'$ in implicit form[10]:

$$G(x') = G(x) - a\,,$$

(4.10)

where *a* is some free parameter. Therefore we can say

$$x'(x) = G^{-1}[G(x) - a] = f(x)\,.$$

(4.11)

For there to be structural universality we require $ax''(x) = x'(ax)$. This follows from $f^*(ax) = af^{*2}(x)$ where the functional self-similarity must also work equally well for forward as well as backward iterates. Applying the function *G(.)* we arrive at

$$G(ax'') = G[x'(ax)] = G(ax) - a\,.$$

(4.12)

However we note the structure at equ.(4.10) then

$$G(x'') = G(x') - a = G(x) - 2a\,.$$

(4.13)

This implies that

$$\frac{1}{2}G(x'') = \frac{1}{2}G(x) - a\,.$$

(4.14)

If we compare equ.(4.14) with equ.(4.12) we note that the solution to the fixed point equation $G^*(x) = G^*(ax)$ must have the property $\frac{1}{2}G^*(x) = G^*(ax)$. Suitable ansatz solutions for $G^*(x)$ that fulfil the boundary conditions for the original function $f(0) = 0,\ f'(0) = 1$ are $1/(x)^{m-1}$ with $2^{\frac{1}{m-}}$ and since $G^{-1}[G^*(x) - a] = f^*(x)$ then

$$f*(x) = \left[\frac{1}{x^{\frac{1}{m-1}} - a}\right]^{\frac{1}{m-1}} = \frac{x}{\left[1 - ax^{(m-1)}\right]^{\frac{1}{m-1}}}\,.$$

(4.15)

[10] Here we note that we are implying the following functional relationship

$$f(x) \xrightarrow{fwd} x'$$
$$x'(x) \xrightarrow{bwd} f(x)\,.$$

We now need to determine the value of the parameter *a*. This can be determined by the map boundary conditions at *d*. With these conditions we obtain

$$a = \frac{1-d^{(m-1)}}{d^{(m-1)}}.$$

(4.16)

4.4.3 Generalised Map Equations

From the derivation of the Taylor and RNG approaches to intermittency we arrive at the following generalised equations for the map family that is used in this study.

Taylor Series Expansion

$$x_{n+1} = F(x_n) = \begin{cases} F_1(x_n) = e_1 + x_n + c_1 x_n^{m_1} & 0 < x_n \le d \\ F_2(x_n) = x_n - e_2 - c_2(1-x_n)^{m_2} & d < x_n < 1 \end{cases},$$

(4.17)

with $c_1 = \frac{(1-e_1-d)}{d^{m_1}}$ and $c_2 = \frac{(d-e_2)}{(1-d)^{m_2}}$.

Renormalization Group

$$x_{n+1} = F(x_n) = \begin{cases} F_1(x_n) = e_1 + \frac{x}{\left[1-c_1 x^{m_1-1}\right]^{\frac{1}{m_1-1}}} & 0 < x_n \le d \\ F_2(x_n) = 1-e_2 - \frac{1-x}{\left[1-c_2(1-x)^{m_2-1}\right]^{\frac{1}{m_2-1}}} & d < x_n < 1 \end{cases},$$

(4.18)

with $c_1 = \frac{(1-e_1)^{m_1-1} - d^{m_1-1}}{(1-e_1)^{m_1-1} d^{m_1-1}}$ and $c_2 = \frac{(1-e_2)^{m_2-1} - (1-d)^{m_2-1}}{(1-e_2)^{m_2-1}(1-d)^{m_2-1}}$.

Indicator variable

In order to simplify things we introduce an indicator variable, y_n, with the following behaviour

$$y_n = \begin{cases} 0 & , 0 < x_n \le d \\ 1 & , d < x_n < 1 \end{cases}.$$

(4.19)

The indicator variable is used to indicate the presence or absence of a packet/full cell. Naturally there is no prohibition to using the reverse definition of the variable. The interpretation of the intermittency map as an *ON/OFF* traffic model is shown in Figure 4.4.

4.5 Transit Time Probability

A property of the map that we require for transit time analysis is the duration that an orbit of the map takes to leave the critical region. This equates to the sojourn time of a source, the length of time spent in a particular state. In terms of iterates of the map, this is the number of iterations required to leave the laminar region. In the simplified model of Manneville

[MANN80] this would be the length of continuous 1's and 0's that would be observed at the indicator variable y_n. By judicious rephrasing of the question what we are really need to find is the probability *P(l)* of having a laminar length of *l*. The determination of this quantity can be more easily approached through the renormalization formulation of the intermittency map.

4.5.1 Derivation: Transit Time Probability

Following Ben-Mizrachi [BEN85], the length of the laminar region is dependent on its starting point x_0. We require the probability of being injected into this region, $P(x_0)$. In our map model the function $f(x_n)$ ceases to affect the orbit past the discriminator value *d*, then the most the iterate value can be is *d*. Therefore we can say that the laminar region is determined by $l = l(x_0, d)$ where *l(.)* is the function determining the length. From the derivation of the RNG method we note that structural universality works for both forward and backward iterates. Using the idea of structural universality we can say that a starting point is $x_0 = x_0(l, d)$ where $x_0(.)$ is the function determining the value of x_0, which we can argue is dependent of the length, *l*, and the end point *d*.

The probability of injection to a start point between x_0 and $x_0 + dx_0$ is denoted by $\hat{P}(x_0)dx_0$. Therefore we can write that the probability of having a laminar length $P(l)$ is

$$\hat{P}(x_0)dx_0 = \hat{P}(x_0(l,d))\left|\frac{dx_0}{dl}\right|dl \equiv P(l)dl \,. \tag{4.20}$$

The function $x_0(l,d)$ can be found via RNG. This is done by noting that the end point *d* must be a fixed point of the doubling operator *T*. Moreover, we are concerned with doubling like behaviour we are therefore interested in lengths $l = 2^n >> 1$. Since forward and backward iterates behave functionally the same then

$$x_0 = X(l,d) \Leftrightarrow d = X(l,x_0) \,. \tag{4.21}$$

We then solve

$$X(l = 2^n, x_0) = \frac{1}{a^n} f^*(a^n x_0) = d$$

$$\frac{1}{a^n} f^*(a^n x_0) = \frac{1}{a^n} \cdot \frac{a^n x_0}{\left[1 - k(a^n x_0)^{m-1}\right]^{\frac{1}{m-1}}}$$

$$d = \frac{x_0}{\left[1 - k(a^n x_0)^{m-1}\right]^{\frac{1}{m-1}}} \tag{4.22}$$

rearranging this to obtain the backward iterate of x_0 i.e. d we obtain the function $x_0(l,d)$

$$x_0 = \frac{d}{\left[1+k\left(a^{\,n}d\right)^{m-1}\right]^{\frac{1}{m-1}}} = x_0\left(l=2^n,d\right).$$

(4.23)

Recall from ansatz solutions to the RNG method we used $a = 2^{\frac{1}{m-1}}$, then as $n \to \infty$ the behaviour of equ.(4.23) is dominated by

$$x_0 \sim \frac{1}{a^{\,n}} = 2^{\frac{n}{1-m}} = l^{\frac{1}{1-m}}.$$

(4.24)

The probability $P(l)$ is then

$$\hat{P}\left(x_0\right)dx_0 = \hat{P}\left(c \times l^{\frac{1}{1-m}}\right) l^{\frac{m}{1-m}} = P(l)dl$$

(4.25)

where c is some constant, and for $l \to \infty$ this becomes

$$P(l) \sim P(0) l^{\frac{m}{1-m}}.$$

(4.26)

The result shows that provided $P(x_0)$ is finite and $x_0 \approx 0$ then we can obtain sojourn times of arbitrary length in the unperturbed system.

Observation: We have seen that because of structural universality, not only is the map function dependent on *m* but also the sojourn time probability is dependent on *m*. It would therefore not be unreasonable to assume that the spectral decay is also solely dependent on this parameter.

4.6 Spectral Decay

We note from Chapter 3 that the spectral decay for stochastic self-similarity has the form

$\lim_{w\to\infty} f(w) = \frac{1}{C_f(H)|w|^{1-2H}}$, where $C_f = \frac{s^{\,2}}{p} C_r \Gamma(2H-1)\sin\left(p(1-H)\right)$ and $s^{\,2} = \mathrm{var}\left(X(t)\right)$.

Since this type of spectral decay refers to *1/f* noise then the intermittency maps must also have a similar spectral decay structure. To obtain the spectral decay of the maps we follow the method used by Ben-Mizrachi [BEN85]:

- We require the correlation function *C(m)* of the indicator variable y_n. We are interested in the behaviour of *C(m)* for $C(m) \xrightarrow{\lim_{m\to\infty}} C(t)$. This requires convolution in the time domain.

- The convolution is easier in the complex frequency $(s = s + iw)$ domain. This requires the Laplace transform.[11]
- We use this result to obtain the spectral density, $f(w)$.
- Unfortunately the Laplace transform is not invertable and therefore we have to take the limit $s \to iw$ of the Laplace transform and its conjugate i.e. $\lim_{s \to iw} \frac{1}{2}\left[C(s) + C^*(s)\right]$.

4.6.1 The Correlation Function

The correlation function is obtained by using Manneville's [MANN80] simplification of the intermittency signal that views the regions of regular laminar flow as periods of constant value indicator variable output. The transitions between laminar and turbulent behaviour make up a binary time series. We analyse this signal through the use of the correlation function *R(m)*

$$R(m) = \sum_{X(0)=0}^{1} \sum_{X(m)=0}^{1} W\left(X(m), X(0)\right) X(m) X(0) \tag{4.27}$$

where $W\left(X(m), X(0)\right)$ is the joint probability of seeing a signal $X(0)$ (0 or 1) at the zeroth iteration and a signal $X(m)$ present at the m^{th} iteration. Using the dependent probability identity[12] we define the following

$$C_m(1|1) \equiv W\left(X(m) = 1 \middle| X(0) = 1\right), \tag{4.28}$$

where $C_m(1|1)$ is the conditional probability of seeing a nonzero signal at the m^{th} iteration. The correlation sum can then be written as

$$\begin{aligned} R(m) &= \sum_{X(0)=0}^{1} \sum_{X(m)=0}^{1} C_m(1|1) W\left(X(0) = 1\right)\left\{X(m) = 1\right\}\left\{X(0) = 1\right\} \\ &= C_m(1|1) W\left(X(0) = 1\right) \end{aligned} \tag{4.29}$$

From this we can see that the low frequency portion of the maps spectrum will be due to the long run (long wavelength) correlations of the indicator variable - long strings of contiguous 1's or contiguous 0's, and that this will be determined by $C_m(1|1)$. If we assume that the probability of a continuous intermission of length *k* is *P(k)*, then we can write $C_m(1|1)$ in an alternative way (dropping the $(1|1)$ notation for convenience)

[11] The Laplace transform is taken in preference to the Fourier transform out of consideration to the boundary conditions.

[12] $P(A|H) = \frac{P(AH)}{P(H)}$.

$$C(m) = \sum_{k=0}^{m} C(m-k)P(k)$$

(4.30)

where $C(m-k)$ is the autocorrelation. We are interested in the behaviour of autocorrelation in the limit as $m \to \infty$. For this reason we can pass into the continuous time domain and write

$$C(t) = \int_0^t C(t-t)P(t)dt + d(t).$$

(4.31)

We need the $d(t)$ term in equ.(4.31) because of the definition of *C(m)*. If we consider *C(*0*)* we obtain the following sum

$$C(0) = \sum_{k=0}^{m=0} C(0-k)P(k).$$

(4.32)

This naturally yields no term. However with a lag of zero *(k = 0)* we should have a correlation of 1 i.e. *C(*0*)* = 1 and *P(*0*)* = 0. Therefore,

$$C(m) = \sum_{k=0}^{m} C(m-k)P(k) + C(0).$$

(4.33)

This is important because in the continuous limit the correlation becomes

$$C(t) = \int_0^t C(t-t)P(t)dt$$

(4.34)

which excludes the autocorrelation with lag zero. For this reason the inclusion of the Dirac delta function is necessary in order to regain this term.

4.6.2 The Spectrum

Normally, to calculate the spectrum we would perform the following cosine transformation

$$f(w) = \int_0^{\infty} C(t)\cos(w)dt.$$

(4.35)

The integral at equ.(4.34) is the convolution integral which can be manipulated better in the complex frequency domain. Taking the Laplace transform of the integral and rearranging it to isolate the correlation term we obtain

$$\int_0^\infty e^{-st}C(t)dt = \int_0^\infty e^{-st}\int_0^t C(t-t')P(t')dt'dt + \int_0^\infty e^{-st}d(t)dt$$

$$C(s) = \frac{1}{1-P(s)}$$

(4.36)

To obtain the spectrum of the correlation we have to solve equ.(4.36) for *C(s)*. Normally this would involve taking the inverse Laplace transform. Unfortunately there is no inverse transform for this equation. A different approach is therefore required to find the spectral decay. This is achieved by taking limits of *C(s)* as the complex frequency, *s*, tends to zero,

$$\lim_{w\to 0} f(w) = \lim_{s\to iw} \frac{1}{2}\left[C(s) + C^*(s)\right].$$

(4.37)

To perform this operation we require the Laplace transform of *P(k)*, where *P(k)* is the probability of a run of *k* 1`s or *k* 0's. This we have already done in section 4.5. The Laplace transform of *P(l)* is

$$P(s) = \frac{\int_1^\infty e^{-sl} l^{\frac{m}{1-m}} dl}{\int_1^\infty l^{\frac{m}{1-m}} dl}.$$

(4.38)

The integration of equ. (4.38) depends on the values that *m* takes. For $m \neq 2, \frac{3}{2}$, *P*(*s*) has the form $P(s) = s^{-z}\Gamma(z,s)$, where Γ(*z*,*s*) is the incomplete gamma function and $z = -1/(m-1)$. This result also has a series interpretation

$$P(s) = s^{-z}\Gamma(z) - \sum_{n=0}^{\infty} \frac{(-1)^n s^n}{n!(z+n)}.$$

(4.39)

For the special cases $m = 2, \frac{3}{2}$, *P*(*s*) has the form $P(s)\big|_{m=2} = \frac{E_2(s)}{m-1}$ and $P(s)\big|_{m=\frac{3}{2}} = \frac{E_3(s)}{m-1}$,

where $E_n(.)$ is the exponential integral $E_n(z) = \int_1^\infty e^{-zt}t^{-n}dt$ for *n*=1,2,…, and *Re*(*z*)>0. In these cases we can use the following equivalence and recursive relationships to again obtain a series representation for *P*(*s*):

$$E_1(z) = -g - \ln(z) - \sum_{n=1}^{\infty} \frac{(-1)^n z^n}{n(n!)} \qquad |\arg z| < p,$$

(4.40)

$$E_{n+1}(z) = \frac{1}{n}\left[e^{-z} - zE_n(z)\right] \qquad n = 1,2,3,\cdots.$$

(4.41)

By substituting the appropriate series representations for $P(s)$ in equ.(4.36) and taking the limits of equ.(4.37) we obtain the following spectral decays[13]

$$\lim_{w\to 0} f(w) \propto \begin{cases} w^{-\frac{1}{m-1}} & 2 < m < 3 \\ \dfrac{1}{w\left|\ln(w)\right|^2} & m = 2 \\ w^{-\frac{2m-3}{m-1}} & \frac{3}{2} < m < 2 \\ \left|\ln(w)\right| & m = \frac{3}{2} \\ const & m < \frac{3}{2} \end{cases} .$$

(4.42)

4.7 Lower and Upper Bounds on the Transit-Time

Since we know the spectral decays dependence on the map parameter *m*, and we know the relationship between spectral decay and the Hurst parameter, *H*, we can now place some limits on the upper and lower bounds on the transit time[14]. The upper and lower bounds can be derived from the Taylor series formulation of the intermittency map. Recall the generalised map function given by

$$x_{n+1} = cx_n^m + x_n + e ,$$

(4.43)

where c is a constant defined in terms of the parameters of the map and e is the width of the critical region[15]. For values of x_n very close to the critical region the distance between successive map iterates is very small. Equation (4.8) can therefore be rearranged to form a difference equation by dividing through by the difference in distance between iterates

$$\frac{x_{n+1} - x_n}{\Delta l} = cx_n^m + e .$$

(4.44)

As $\Delta l \to 0$ then we obtain a differential equation. The solution to this equation yields the time taken to escape the influence of the function, i.e. the transit time. The differential equation corresponding to equ.(4.44) for a given start and stop value of the iterates is given by

$$L = \int_{x_{in}}^{d} \frac{dx}{cx^m + e} ,$$

(4.45)

[13] Both Ben-Mizarachi [BEN85] and Schuster [SCHU95] present extend results on spectral decays, i.e. for $m>3$. The results shown here stop at $m<3$. The reason for this is that for $2<m<3$ we again have a power law decay, admittedly outside the interval of interest (see below). However this observation is included since it may become useful in developing other kinds of traffic model based on this type of chaotic map.

[14] This is so because for positive correlations, i.e. $H\in[½,1]$ m takes values in the range, $m\in[1,2]$.

[15] In actual fact e is the perturbation to the equation.

where L is the number of iterates to leave the left hand (LH) side of the map for an entry point of x_{in} and an exit point of d[†]. Here we note that the upper limit of integration is set to d since (naturally) there can be no higher iterate values than this before escape from the LH half of the map occurs (see Figure 4.4). We know that m is bounded, i.e. $m \in [1,2]$. This yields the following solutions for an upper and lower bound on L.

4.7.1 Case $m = 1$

This case yields the lower bound on the number of iterations required to escape the LH side of the map. For $m = 1$ equ.(4.45) becomes

$$L_{m=1} = \int_{x_{in}}^{d} \frac{dx}{cx+e}.$$

(4.46)

This can be solved quite easily via the transformation $A = cx + e$. The solution to the integral then becomes

$$L_{m=1} = \frac{1}{c}\left\{\ln\left(\frac{cd+e}{cy_{in}+e}\right)\right\}.$$

(4.47)

This yields the lower bound result.

4.7.2 Case $m = 2$

This case yields the upper bound in the number of iterations required to escape the LH side of the map. For $m = 2$ equ.(4.45) becomes

$$L_{m=2} = \int_{x_{in}}^{d} \frac{dx}{cx^2+e}.$$

(4.48)

To solve this equation we apply the transformation $\tan q = y\sqrt{c}/\sqrt{e}$. The integral to this is

$$L_{m=2} = \frac{1}{\sqrt{ce}}\left\{\arctan\left(d\sqrt{\tfrac{c}{e}}\right) - \arctan\left(y_{in}\sqrt{\tfrac{c}{e}}\right)\right\}.$$

(4.49)

This yields the upper bound result.

Comment: In iterated map systems such as the maps that we use, the unit of time at each iteration is "timeless" since the user sets the time unit of the model. This means that each iteration of the map portrays some specific time increment of the system to be modelled. What the transit time actually refers to is the number of iterations required to leave the critical region of the map, not the actual time duration (e.g. seconds, milliseconds etc.) taken to leave the critical region.

[†] Here we can see that the unperturbed map, $e = 0$, poses no analytical problems since

4.8 Invariant Densities - The Statistics of the Dynamics

We have presented the dynamics of the map as a dynamical system in which the iterates of the map follow some pre-described path given by known functions of the LH and RH halves of the map. We now describe the dynamics in a statistical manner. Pruthi [PRU95a] has already investigated the stochastic aspect of chaotic maps. We only summarise the important features here.

The stochastic representation of the orbit is done in terms of the invariant density[16], $r(x)$. This is the probability that an orbit will visit a given interval of the map. For there to be equivalence between the invariant density and a probability, the invariant density must be normalizable, i.e. the sum of all the invariant densities must equal unity. This implies that the integral of the invariant density is finite[17].

The determination of the invariant density stems from solving the Frobenius-Perron equation

$$r_{n+1}(x) = \int r_n(y) d[x - f(y)] dy \tag{4.50}$$

where $f(.)$ is the forward iterate function of the map, $r_n(.)$ is the invariant density at iteration *n*, and the summation is performed over all backward iterates of the map, i.e. $y = f^{-1}(x)$. If the maps are ergodic[18] [19] and chaotic[20] then the $r_n(x)$ becomes independent of time and can be termed invariant. Here we point out that not all maps have invariant densities. The invariant density then becomes the solution to the steady state Frobenius-Perron equation, i.e. $r_{n+1}(x) = r_n(x) = r(x)$, then equ.(4.50) becomes

$$r(x) = \int r(y) d[x - f(y)] dy \,. \tag{4.51}$$

---

equ.(4.45) becomes a simple integral.

[16] Here the term "invariant" is used because the visitations of the map's orbit to points on the interval with arbitrary accuracy containing the map have reached a "steady" and hence "invariant" state.

[17] This does not mean that the density itself be finite - this is in contrast to the interpretation given by Pruthi in [PRU95a] as applied to this particular kind of map i.e. TYPE I intermittency map.

[18] Ergodic in this sense means a unique initial condition that leads to "on average" behaviour.

[19] To prove that the maps are ergodic is very difficult to do theoretically. It essentially boils down to having a positive Lyupanov exponent. If we have a positive Lyapanov exponent then we have a system which will exponentially diverge with a good chance of the orbit of the map visiting all points of the space in which it is described. This spreading is due to the stretching and folding of the map, which makes initial points arbitrarily close to each other spread all over the interval. Then any one starting point will look much the same as any other and the resulting invariant density will be the same. All that we can say about the maps that we use is that they appear to be ergodic since their invariant densities appear to be invariant regardless of the start position and that the averages of the system converge.

[20] See earlier "Required Ingredient for Chaos in Iterated Systems" in this chapter, specifically with regard to item three.

Through the use of the Dirac delta function identities:

$$d[r(t)] = \sum_n \frac{d(t - t_n)}{\left|\frac{dr(t_n)}{dt}\right|}$$

(4.52)

and

$$\int_{-\infty}^{+\infty} f(t)d(t-a)dt = f(a),$$

(4.53)

we can substitute equ.(4.52) into equ.(4.51) and precede the integral by the summation sign without any ill effect.

$$r_{n+1}(x) = \sum_n \int r_n(y) \frac{d(y - y_n)}{\left|df(y_n)/dy\right|} dy .$$

(4.54)

We can now apply the identity given by equ.(4.53) to the numerator of the above in order to obtain the final form that we use in our investigations

$$r_{n+1}(x) = \sum_n \frac{r_n(y_n)}{\left|df(y_n)/dy\right|} .$$

(4.55)

See Appendix B for more information on Dirac impulse function identities.

If we let $Y_i = f_i^{-1}(x)$ be the backward iterate of the $f_i th.$ function and $f_i'(x)$ is the forward differential of the $f_i th.$ function, then for the generalised equations given in section 4.4.3 we can summarise the contributions to the invariant density as

$$r(x) = \begin{cases} \dfrac{r(Y_1)}{\left|f_1(Y_1)\right|} & e_2 - 1 \le x < 1 \\ \dfrac{r(Y_1)}{\left|f_1(Y_1)\right|} + \dfrac{r(Y_2)}{\left|f_2'(Y_2)\right|} & e_1 \le x < e_2 - 1 \\ \dfrac{r(Y_2)}{\left|f_2'(Y_2)\right|} & 0 \le x < e_1 \end{cases} .$$

(4.56)

The dependence of the contributions to the invariant density on the values of e_1 and e_2 can be seen more clearly in Figure 4.5.

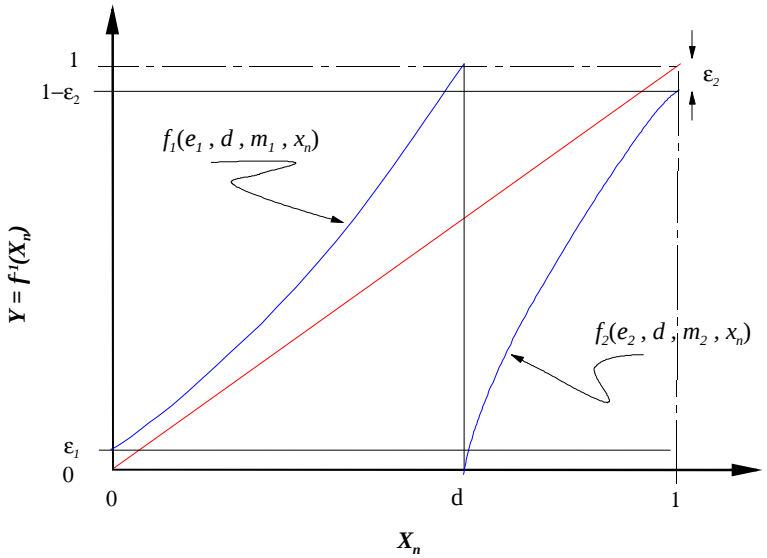

Figure 4.5 Relationship between the backward iterate, y and the zones of contribution to the invariant density

4.9 Invariant Density Approximation

Much effort has been placed in obtaining analytically the invariant density from the map parameters, most notably by Pruthi [PRU95a], who investigated several methods for obtaining the invariant density. More recently, Mondragón [MON98] has made further advances on the analytic approximation of the invariant density via a method which de-couples the LH and RH halves of the map. Both approximations rely on the behaviour around the fixed points to dominate the invariant density. The reasons for concentrating on analytic methods for the approximation of the invariant density are two-fold:

- analytic methods are rapid, and
- they provide a good initial guess for iterative methods (again providing some form of speed up).

4.9.1 The Pruthi Approximation

This method is derived from considerations on a coupled set of the backward iterates of the map functions. If we suppose that the backward iterates of the map functions are

$$a = f_1^{-1}(y) \quad \text{and} \quad b = f_2^{-1}(y),$$

(4.57)

we can then define the probabilities with respect to the invariant density of the map as

$$P(y) \equiv \int_0^y r(x)dx, \quad P(a) \equiv \int_0^a r(x)dx, \quad P(b) \equiv \int_0^b r(x)dx, \quad P(d) \equiv \int_0^d r(x)dx.$$

(4.58)

Under these definitions we can then say

$$P(y) - P(a) = \int_a^y r(x)dx, \qquad P(b) - P(d) = \int_d^b r(x)dx\,.$$

(4.59)

By using the Frobenius-Perron operator and considerations on the contributions of the map functions to the invariant density we obtain

$$\int_0^y r(x)dx = \int_0^a r(x)dx + \int_d^b r(x)dx$$
$$P(y) = P(a) + P(b) - P(d)$$

(4.60)

and for y<<1 we can say

$$r(y) = \frac{P(y) - P(a)}{y - a} \quad \text{and} \quad r(b) = \frac{P(b) - P(d)}{b - d}.$$

(4.61)

Dividing $r(y)$ by $r(b)$ and substituting equ.(4.57) where appropriate we obtain

$$r(y) = \frac{f_2^{-1}(y) - d}{y - f_1^{-1}(y)} r\left(f_2^{-1}(y)\right).$$

(4.62)

By taking the limit y→0 we finally obtain

$$\lim_{y \to 0} r(y) = \frac{f_2^{-1}(y) - d}{y - f_1^{-1}(y)} r(d).$$

(4.63)

By a similar process we obtain

$$\lim_{y \to 1} r(y) = \frac{f_1^{-1}(y) - d}{y - f_2^{-1}(y)} r(d).$$

(4.64)

For more details see [PRU95a pp88-98].

4.9.2 The Mondragón Approximation

This method is applicable when the Taylor's series derivation has been used for the map functions. The assumptions made in this method are:

- the transitions between halves of the map are viewed as random, decoupling the map halves; and
- ε=0.

Under these assumptions the invariant density can be analytically approximated. The contribution to the invariant density form on map half can then be viewed as

$$\tilde{r}(x) = \frac{\tilde{r}\left(f^{-1}(x)\right)}{f'\left(f^{-1}(x)\right)} + K,$$

(4.65)

where *K* is a constant (this can be viewed as the contribution to the invariant density from the other half of the map), $\tilde{r}(x)$ is the invariant density of the half of the map in question, $f^{-1}(x)$ is

the backward iterate and $f'(x)$ is the derivative of the forward iterate of the map in question. If we assume that equ.(4.65) has the form

$$\tilde{r}(x,d,m,K) \approx -K \frac{\left(1+cmx^{m-1}\right)\left(1-\left(1+cx^{m-1}\right)^{m}\right)}{mc^2x^{2m-2}\left(1+cx^{m-1}\right)^{m-1}} \tag{4.66}$$

where $c = (1-d)/d^m$, the approximation is always valid if x<<1, or in the interval (0,*d*) if *c*<1, or $d^m + d - 1 < 1$. In these cases the invariant density is given by

$$r(x) = \begin{cases} r_2(x,d,m_2,K_2) = \tilde{r}(x,1-d,m_2,1)K_2 & x > d \\ r_1(x,d,m_1,K_1) = \tilde{r}(x,d,m_1,1)K_1 & x < d \end{cases}. \tag{4.67}$$

The *ON* and *OFF* probabilities are then given by

$$\begin{aligned} l &= p_2(x,d,m_2,K_2) = \int_d^1 \tilde{r}(x,1-d,m_2,K_2)dx \\ 1-l &= p_1(x,d,m_1,K_1) = \int_0^d \tilde{r}(x,d,m_1,K_1)dx \end{aligned}. \tag{4.68}$$

The solution for this integral is

$$\begin{aligned} p(x,d,m,K) &= \int_0^x \tilde{r}(y,d,m,K)dy \\ &= \frac{K}{c^2m}\left(\frac{c(1+m)x^{2-m}}{2-m} + c^2mx + \frac{x^{3-2m}}{3-2m}\right) \\ &\quad - KAx^{1-2m}\left[Bx^{2m}\,{}_2F_1\left(m,\frac{1}{m-1};\frac{m}{m-1},z\right)\right. \\ &\quad \left. - x\left\{(m-2)x\,{}_2F_1\left(m,\frac{3-2m}{m-1};\frac{2-m}{m-1},z\right) + Cx^m\,{}_2F_1\left(m,\frac{2-m1}{m-1};\frac{1}{m-1},z\right)\right\}\right] \end{aligned}, \tag{4.69}$$

where $A = \frac{1}{c^2(m-2)(2m-3)m}$, $B = c^2m(2m^2-7m+6)$, $C = c(2m^2-m-3)$, $z = -cx^{m-1}$ and ${}_2F_1(.)$ is the Hypergeometic function. For more detailed description of the method see [MON98].

4.10 Summary

In this chapter we have examined non-linear dynamical maps as teletraffic models. Central to their effectiveness as teletraffic models that can replicate LRD and self-similar traffic is the relationship between extended sojourn times witnessed in real traffic and the maps' orbital behaviour. The part of a map's orbit that produces the correct behaviour occurs near bifurcation in a map with TYPE I intermittency. In order to track the evolution of the map's orbit in this intermittent region we applied boundary conditions to formulations for an

intermittency map derived from Taylor's Series and Renormalization Group approaches. Key properties for this map at the point of bifurcation (the critical point) was obtained through the Renomalization Group approach while behaviour near the critical point was obtained through the Taylor's Series approach. The intermittency map forms the basis for chaotic map models that display LRD and self-similarity and is the subject of the next chapter.

5 Mathematical Analysis of Intermittency Map Models

The previous chapter introduced the basic intermittency map models and the important properties that are relevant to their teletraffic interpretation and application. Pruthi [PRU95a] investigated a variety of chaotic maps and found the fixed point double intermittency map to be of most use. However, his parameterisation of the double intermittency map was limited: he set the parameter e to zero; and conjectured that H depends only on the value of m for one side of the map. Furthermore, his aggregate source models gave poor performance when used for studying queueing behaviour.

A recent result of Willinger *et al* [WIL97] has shown that it is the aggregation of *ON-OFF* sources with LRD in either *ON* or *OFF* states which leads asymptotically to 2nd order self-similarity. This chapter introduces a "family" of intermittency maps, specified by five parameters: e and m in both sides of the map, and d, the discriminator value. This "family" can produce a variety of behaviour when aggregated, which can lead to asymptotic 2nd order self-similarity depending on the map parameters. The following two chapters develop new parameterisation methods that allow these maps to represent aggregate traffic behaviour accurately, focusing on the invariant density.

In this chapter, we present new analysis of the intermittency maps' transit-time (in the general case of perturbed maps), spectral decay, and H (in the specific case for epsilon set to zero). Pruthi and Erramilli [PRU95a, ERR94b] termed e a throwaway parameter. The setting of epsilon to zero was done in part for analytical tractability and also to ensure that there would be bursts over all time scales, i.e. so that the probability of infinite sojourn times in the *ON* or *OFF* period would exist. An infinite sojourn period in the *ON* state would mean that the source was permanently on. Similarly, an infinite sojourn period in the *OFF* state would mean that the source was permanently off. However, Fowler [FOW91] observed that the bursts occurred over all time scales of engineering interest, i.e. over a finite number of time scales. The transit time analysis performed in this chapter shows an important dependence on e. This limits the time scales over which stochastic self-similarity is present supporting the view of Fowler. In terms of chaotic map models this means that e is not such a discardable parameter as was first thought. The Hurst parameter analysis confirms Pruthi's conjecture and shows that, in the limit, the dominant m value determines the value of H.

We also present numerical results for the H behaviour of the maps which illustrate a complex dependence between both portions of the map, and the values of m and d. The H behaviour is predictable, but differs substantially from the limit behaviour suggested by the analysis. We present an empirical fit for H which gives substantially better prediction than analysis for practical cases in a coupled map. In addition, we present a new method for estimating H

based on the correlation structure of the maps, and this is compared with other estimation methods in Chapter 8.

5.1 A Family of Chaotic Maps as Self-similar Traffic Sources

The chaotic map family is described by a common set of equations. Individual members of the family are identified via their parameter values. The common set of equations which describes this map family via the Taylor series formulation is given by

$$x_{n+1} = F(x_n) = \begin{cases} F_1(x_n) = e_1 + x_n + \dfrac{(1-e_1-d)}{d^{m_1}} x_n^{m_1} & 0 < x_n \le d \\ F_2(x_n) = x_n - e_2 - \dfrac{(d-e_2)}{(1-d)^{m_2}} (1-x_n)^{m_2} & d < x_n < 1 \end{cases} , \tag{5.1}$$

where $1 \le m_1, m_2 \le 2$, $0 < d < 1$, $0 \le e_1, e_2 << 1$ are parameters and $0 < x_n < 1$. An *ON-OFF* source model can be constructed from these equations via the use of an indicator variable, y_n, which simulates the presence or absence of teletraffic

$$y_n(x_n) = \begin{cases} 0, & 0 < x_n \le d, \quad \text{passive "}OFF\text{" state} \\ 1, & d < x_n \le 1, \quad \text{active "}ON\text{" state} \end{cases} . \tag{5.2}$$

We refer to the x_n of equ.(5.2) as the underlying dynamics; this corresponds to the realisation of a stochastic process. We refer to the y_n as the overlying dynamics; this corresponds to the visual element of the realisation process.

The nomenclature of the map family is as follows:

- for m_1 and m_2 =1: Bernoulli shift map - this produces SRD in both halves of the map;
- for m_1=1 and m_2>1, or, m_1>1 and m_2=1: single intermittency map - this produces SRD in one half and LRD in the other; and
- for m_1 and m_2>1: double intermittency map; this produces LRD in both halves of the map.

Pruthi [PRU95a] explored many determinist chaotic maps as traffic source models. The model which he found to be of most useful was the Fixed Point Double Intermittency map (FPDI). The reason for this was that the FPDI map rendered LRD in both halves of the map. However, recently Willinger *et al* [WIL97] have shown that it is *ON-OFF* sources having LRD in at least one state (either *ON* or *OFF*) that when aggregated will cause 2nd order self-similarity. Classifying individual maps into a family as described above enables us to view these models as suitable for modelling aggregated data traffic.

5.2 State Transit-time Analysis

The transit time is the time it takes an orbit to leave the *ON* or *OFF* state of a map. Transit-time analysis of chaotic maps is problematic because analytic solutions to the perturbed map can only be found at certain values of the parameter *m* (eg. *m* = 1, 3/2, 2). Further, analytic

forms for the transit time values of the perturbed map for rational $m \in (1,2)$ have not been found for the Taylor formulation of the map family. In this section we develop an approximation for the transit time of the perturbed map for rational *m*. In our development we concentrate only on one half of the map[21].

We recall that the basic equation for the map is given by

$$x_{n+1} = cx_n^m + x_n + e \,,$$

(5.3)

where *c* is a constant defined in terms of the parameters of the map and *e* is the distance between the reflection line and the curve (see section 4.4.3). In actual fact *e* is the perturbation to the equation. Equation (5.3) can be rearranged to yield a difference equation by dividing through by the difference in distance between iterates

$$\frac{x_{n+1} - x_n}{\Delta l} = cx_n^m + e \,.$$

(5.4)

As $\Delta l \to 0$ we approximate a differential equation. The solution to this equation yields the time taken to leave a map state (for example the *ON* or *OFF* state), i.e. the transit time. The differential equation corresponding to equ.(5.4) [22] is

$$L = \int_{x_{in}}^{d} \frac{dx}{cx^m + e} \,,$$

(5.5)

where *L* is the number of iterates to leave a map state for an entry point of x_{in} and an exit point of *d*. We note that the upper limit of integration is set to *d* since higher iterate values mean that the orbit has escaped the map state.

5.2.1 Lower and Upper Bounds on the Transit-time

We can determine the upper and lower bounds on the transit time. We know that *m* is bounded; $m \in [1,2]$, in order to give the correct range of *H*. This yields the following solutions for an upper and lower bound on *L* (see Section 4.7). The case *m* = 1 yields the lower bound

$$L_{m=1} = \frac{1}{c}\left\{\ln\left(\frac{cd+e}{cy_{in}+e}\right)\right\}.$$

(5.6)

The case *m* = 2 yields the upper bound

$$L_{m=2} = \frac{1}{\sqrt{ce}}\left\{\arctan\left(d\sqrt{\tfrac{c}{e}}\right) - \arctan\left(y_{in}\sqrt{\tfrac{c}{e}}\right)\right\}.$$

(5.7)

[21] The other half of the map is identical in structure and can be analysed in the same way.
[22] Here we can see that the unperturbed map, *e* = 0, poses no problems since equ.(5.5) becomes a simple integral.

5.2.2 Transit-time Approximation: Rational powers of $m\hat{I}(1,2)$

We require an expression that describes the transit time for rational powers of m within the domain of m. To achieve this we apply the following transformation

$$A = \frac{cy^m}{e} \qquad dy = \frac{dA}{m\left(\frac{c}{e}\right)^{\frac{1}{m}} A^{\frac{m-1}{m}}} \quad ,$$

to equ.(5.5) which yields the following integral

$$L_{m\in(1,2)} = \frac{e^{-\left(\frac{m-1}{m}\right)}}{mc^{1/m}} \int_{A_{in}}^{A_{out}} \frac{dA}{(A+1)A^{\frac{m-1}{m}}} .$$

(5.8)

We can expand the integrand as a power series (binomially) and form a series of integrals[23]

$$L_{m\in[1,2]} = \frac{e^{-\frac{(m-1)}{m}}}{mc^{\frac{1}{m}}} \sum_{j=0}^{\infty} (-1)^j \int_{A_{in}}^{A_{out}} \frac{A^{\frac{1}{m}}}{A} A^j dA .$$

(5.9)

Integrating this yields

$$L_{m\in[1,2]} = \frac{e^{-\frac{(m-1)}{m}}}{mc^{\frac{1}{m}}} \left\{ \sum_{j=0}^{\infty} (-1)^j \frac{mA_{out}^{\frac{mj+1}{m}}}{mj+1} - \sum_{j=0}^{\infty} (-1)^j \frac{mA_{in}^{\frac{mj+1}{m}}}{mj+1} \right\} .$$

(5.10)

We separate out terms not dependent on j and define the function $k(m)$ as

$$k(m) \equiv \frac{e^{-\left(\frac{m-1}{m}\right)}}{mc^{\frac{1}{m}}} .$$

(5.11)

Using this definition we can define the general transit time through a map state to be the difference in summation terms

$$L_{m\in[1,2]} = mk(m) \left\{ A_{out}^{\frac{1}{m}} \sum_{j=0}^{\infty} \frac{(-A_{out})^j}{mj+1} - A_{in}^{\frac{1}{m}} \sum_{j=0}^{\infty} \frac{(-A_{in})^j}{mj+1} \right\} .$$

(5.12)

As a check we can expand out the summations and recover the upper and lower bounds on the transit time. The expansion of $\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1}$ does not converge easily because of oscillating

[23] We can do this because the reversal of the summation and integration terms, as a result of the series expansion, does not alter the original integration as the additional terms are covered by the summation to infinity preceding the term by term integration.

terms. Moreover since A has a e^{-1} dependence and as *j* becomes large and as $e \to 0$ the oscillations become large exacerbating the convergence of the series.

Transforming the series into a function whose convergence properties are known can ease the problem of convergence. Such a series is the hypergeometric series (see [GRA80] pp1039-1059 and [ABRA65] Ch.13, 15). In particular we are interested in an elementary function of the hypergeometric family known as the Gauss series. The Gauss series is defined as

$$\begin{aligned} F(a,b;c;z) &= {}_2F_1(a,b;c;z) \\ &= \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!} \\ &= \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b+n)}{\Gamma(c+n)} \frac{z^n}{n!} \end{aligned} \quad .$$

(5.13)

The summation terms in equ.(5.12) can be transformed into a hypergeometric function , then the following holds

$$\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1} = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(j+a)\Gamma(j+b)}{\Gamma(j+c)} \frac{Z^j}{j!} = {}_2F_1(a,b;c;z).$$

(5.14)

We now require values for *a*, *b*, *c* and Z. An obvious equivalence is, $-A = Z$ and $a = 1$. The values of *b* and *c* can be found by expanding the Gamma function in the summation

$$\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1} = \frac{\Gamma(c)}{\Gamma(b)} \sum_{j=0}^{\infty} \frac{(j-1+b)(j-2+b)(j-3+b)\cdots(1+b)b\Gamma(b)}{(j-1+c)(j-2+c)(j-3+c)\cdots(1+c)c\Gamma(c)} z^j ,$$

(5.15)

from which we obtain the following values $a = 1, b = \frac{1}{m}, c = \frac{1+m}{m}$ (for a more detailed derivation see Appendix C). Therefore we can say that

$$\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1} = {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{m+1}{m}; z\right).$$

(5.16)

Again as a check we can recover the upper and lower bounds on the transit time. Therefore we can write our original integral given by equ.(5.8) as

$$L_{m\in(1,2)} = mk(m)\left\{A_{out}{}^{\frac{1}{m}}\, {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{1+m}{m}; z_{out}\right) - A_{in}{}^{\frac{1}{m}}\, {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{1+m}{m}; z_{in}\right)\right\}$$

(5.17)

where $A_n = \frac{c}{e} y_n^m$, $Z_n = -A_n$, $k(m) = \frac{e^{-\frac{(m-1)}{m}}}{mc^{\frac{1}{m}}}$, $c = \frac{1-e-d}{d^m}$ and ${}_2F_1(a,b;c;z)$ is a hypergeometric function. A comparison between the theoretical transit times and actual transit times for an intermittency map with the following parameters: $\varepsilon=1*10^{-12}$, d = 0.5. m = 1, 1.5, 1,8, 2. is shown in Figure 5.1. We note that for m = 1.8 there is a slight "wobble" in the theoretical prediction. This is due to the numerical evaluational methods used by Maple™. If another algorithm were used then a better result may be obtained.

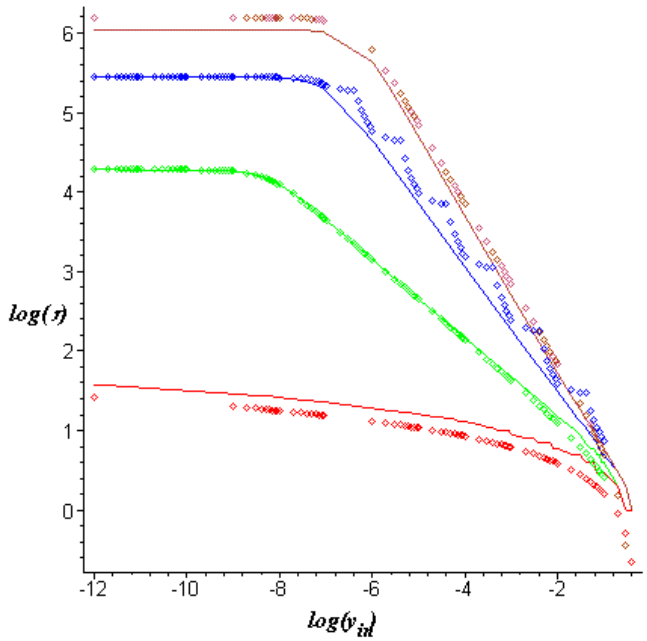

Figure 5.1. Comparisons of map iterations (lines) against theory (crosses) for the case: m=1 (red), m=1.5 (green), m=1.8 (blue), and m=2 (brown)

The hypergeometric function has its convergence defined on the unit circle $|z| = 1$. The convergent behaviour of this series is as follows:

- divergent when $\Re(c-a-b) \le -1$;
- absolutely convergent when $\Re(c-a-b) > 0$;
- conditionally convergent when $-1 < \Re(c-a-b) \le 0$.

Introducing the constant terms we obtain $\Re\left(\frac{1+m}{m} - 1 - \frac{1}{m}\right) = 0$. This implies that the series is conditionally convergent. However, we know that provided *e>0* the transit times are finite and bounded to a maximum value set by *e* and *m=1,2*. Therefore transit-times for *m Î (1,2)* will also be finite and the series converges. We also have the added advantage that these hypergeometric functions can be numerically computed quite rapidly [PRE94].

5.3 The Mean Transit-time

Schuster [SCHU95] has given a general formulation for the mean transit-time

$$\langle l \rangle = \int_0^\infty P(l) l dl \ ,$$

(5.18)

where $P(l)$ is the probability of having a transit time of duration l. This is related to the entry probability at a given point, $P(x_0)$, via the following transformation (see equ.(4.20))

$$P(l) = \hat{P}(x_0) \left| \frac{dx_0}{dl} \right| ,$$

(5.19)

where $\hat{P}(x_0)$ is the estimated probability of having a starting point, x_0.

In the case of chaotic map models, the entry point probability corresponds to the invariant density at the point of entry. This is since the invariant density is the frequency of the orbit's visitation to a particular part of the interval. Hence

$$P(l) = r(x_0) \left| \frac{dx_0}{dl} \right|$$

(5.20)

and the average length is then given by

$$\langle l \rangle = \int_{x \in \{ON\}} r(x) l(x) dx \ .$$

(5.21)

Since $l(x)$ can be expressed as a hypergeometric function, then we can write the mean transit time in terms of this function, which takes a point of entry, x_0, as its argument, i.e. $l(x)$ can be expressed as

$$l(x_0) = mk(m) \left\{ A(d)^{\frac{1}{m}} \, {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{1+m}{m}; z(d)\right) - A(x_0)^{\frac{1}{m}} \, {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{1+m}{m}; z(x_0)\right) \right\}.$$

(5.22)

This leads to the following expression for the mean transit time for a chaotic map source for rational powers of $m \in (1,2)$.

$$\langle l \rangle = mk(m) A(d)^{\frac{1}{m}} \, {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{1+m}{m}; z(d)\right) l - mk(m) \int_{x_0 \in \{ON\}} A(x)^{\frac{1}{m}} \, {}_2F_1\left(1, \tfrac{1}{m}; \tfrac{1+m}{m}; z(x)\right) r(x) dx \ .$$

(5.23)

5.4 Relationship between Transit-time and Correlation

The correlation of the map at the L^{th} iteration, *C(L)*, can be considered to be the probability of all starting points in the interval of interest iterated *L* times and terminating in an emission i.e. $y(x_L)=1$. This can be written as[24]

$$C(L) = l \int_{x\in\{ON\}} r(x)y\left(f^{(L)}(x)\right)dx,$$

(5.24)

where $y\left(f^{(L)}(x)\right)$ is the value of the indicator variable (termination's of 0 or 1), $f^{(L)}(x)$ is the L^{th} forward iterate of the map, $r(x)$ is the invariant density and $x\in\{ON\}$ is the set of all starting points in the *ON* region of the map[25]. The probability *P(l)* of having an iterative sequence of contiguous events (either all ones or zeros) is given by[SCHU95]

$$P(l) = \hat{P}(x_0)\left|\frac{dx_0}{dl}\right|,$$

(5.25)

For the intermittency map family we can approximate the probability of a particular entry point x_0 corresponds to the invariant density at x_0, i.e. $\hat{P}(.)\simeq r(.)$. Therefore we can approximate the invariant density by

$$r(x) = P(l)\frac{dl}{dx}.$$

(5.26)

We can substitute this into equ.(5.22) obtaining

$$C(L) = \frac{1}{l}\int_0^{\infty} P(l)y\left(f^{(L)}(x)\right)dl.$$

(5.27)

The L^{th} forward iterate of the map counts the probability of all combinations of contiguous blocks of length *L*. Hence

$$P(l)y\left(f^{(L)}(x)\right) = \int_0^L dP(l) \equiv P(l<L).$$

(5.28)

Therefore we can say

---

[24] The correlation is defined as $C(L)\equiv E[y(x_n)y(x_{n+m})]$. We are interested in transitions starting and ending in an emission, $C(L)=P[y(x_n)=1, y(x_{n+m})=1]$. Via the probability identity $P(AB)=P(B)P(A|B)$ we obtain $C(L)=P[y(x_n)=1]P[y(x_{n+m})=1|y(x_n)=1]$ which when considering the probability in terms of its invariant density becomes $C(L)=l\int_{x\in\{ON\}} r(x)y\left(f^{L}(x)\right)dx$.

[25] Recall that $l=\int_{x\in\{ON\}} r(x)dx$ .This can be, depending on interpretation, the intervals [0,*d*] or (*d*,1].

$$\lim_{L\to\infty} C(L) = \frac{1}{l}\int_0^L P(l < L)dl \,.$$

(5.29)

If we now employ the following result of Pruthi [PRU95a] concerning the cumulative probability of lengths and derived from the renormalization group approach,

$$P(l < L) = a\, l^{-\frac{1}{m-1}} \,.$$

(5.30)

On substituting the above into equ.(5.29) and by noting that we cannot have correlation lengths less than 1 we obtain

$$\lim_{L\to\infty} C(L) = \frac{a}{l}\int_1^L l^{-\frac{1}{m-1}} dl \,.$$

(5.31)

Evaluating this integral and examining its behaviour as it approaches $L \to \infty$ we arrive at

$$C(L) = \frac{l\,a\,(m-1)}{(m-2)}\left\{ l^{\frac{m-2}{m-1}} - 1 \right\},$$

(5.32)

where α is some constant. As l also plays a part in the transit-time analysis (see equ.(5.22)) then we can also see that there is a connection between the transit-time and the correlation. We use this relationship later to prove a theorem on the Hurst parameter H.

5.5 Hurst Parameter Prediction

Self-similar traffic has a significant effect on the occupancy of buffers in a network: it produces a heavy tailed distribution [NOR93, 95]. This means, from a practical point of view, that providing more buffer space is not a solution to buffer saturation because eventually, the buffer will fill up. The probability of high buffer occupancy has been linked to the LRD of the network traffic [ERR96a]. This means that high values of the Hurst parameter produce a higher than normal probability of buffer occupancy for a given traffic load.

To study and understand packet traffic, a model must be able to reproduce the traffic load and the variability and self-similarity of real traffic. In particular, it is important that the model can generate traffic that has a specific Hurst parameter value. Erramilli [ERR94b, 95a], and Pruthi and Erramilli [PRU95a, 95b], show that these maps are good traffic models and conjectured that the Hurst parameter, H, depends only on one power of the non-linear portion

$$H = \frac{3m_1 - 4}{2(m_1 - 1)} \,.$$

(5.33)

For the case of the intermittency map family we extended Erramilli's conjecture to a theorem using considerations from the correlation function via results from transit-time analysis.

Theorem - Intermittency map Family H dependence on the dominant m. For the intermittency map family, with $m_1, m_2 \in (1,2)$, the Hurst parameter, H, is dependent on the dominant value of m. That is to say

$$H = \begin{cases} \dfrac{3m_1 - 4}{2(m_1 - 1)} & m_1 > m_2 \geq \frac{3}{2} \\ \dfrac{3m_2 - 4}{2(m_2 - 1)} & m_2 > m_1 \geq \frac{3}{2} \\ const. & \frac{3}{2} > m_1, m_2 > 1 \end{cases} . \tag{5.34}$$

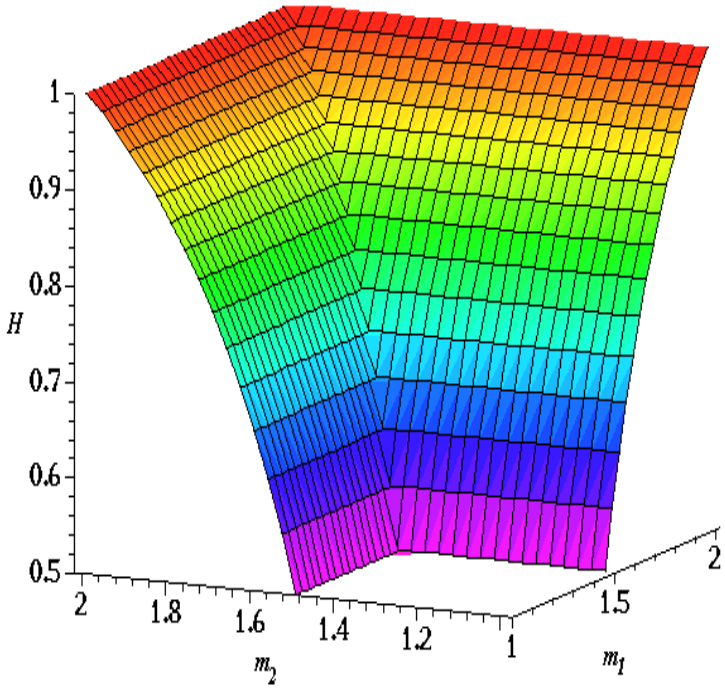

Figure 5.2. Theoretical dependence of H on m_1 and m_2

The theoretical surface of H due to the interaction between m_1 and m_2 is shown in Figure 5.2. The proof for this comes from considerations on the transit time (Sections 5.2.2, 5.3 and 5.4) and its behaviour in the complex plane. If we can show that the transit-time is analytic i.e. is continuous, single valued and has a derivative, then we can show the above theorem to be true.

Definition - Continuity in the complex plane. A function $f(z)$ is said to be continuous at z_0 if for any $\zeta > 0$ $\exists$ $\delta > 0$ such that $\left|f(z) - f(z_0)\right| < z$ whenever $\left|z - z_0\right| < d$.

In the case of the transit-time we need to show that it is continuous, $\forall z, y \in \{ON\} \mapsto A \mapsto z$, $\left|{}_2F_1\left(1, \frac{1}{m}; \frac{m+1}{m}; z\right) - {}_2F_1\left(1, \frac{1}{m}; \frac{m+1}{m}; z_0\right)\right| < z$ whenever $\left|z - z_0\right| < d$. If we write the hypergeometric function as its series expansion

$$\left|\sum_{j=0}^{\infty}\frac{\frac{1}{m}}{j+\frac{1}{m}}z^{j}-\sum_{j=0}^{\infty}\frac{\frac{1}{m}}{j+\frac{1}{m}}{z_0}^{j}\right|<z \Rightarrow \left|\sum_{j=0}^{\infty}\frac{\frac{1}{m}}{j+\frac{1}{m}}\left[\left(z_0+d\right)^{j}-{z_0}^{j}\right]\right|<z$$

then the requirement for continuity can be reduced to the requirement that

$\lim_{j\to\infty}\left(z_0+d\right)^{j}-{z_0}^{j}=k_j>0$ and since the mapping $y\in\{ON\}\mapsto A\mapsto z,$ is a monotonic increasing one then $k_j > 0$ will always be true. Therefore we will always be able to define some $0<k_j<z \quad \forall y\in\{ON\}\mapsto A\mapsto z,\Rightarrow |z-z_0|<d$. Therefore for all z values of interest ${}_2F_1(.;z)$ is continuous.

The derivative of $f(z)$ also exists and is a property of the hypergeometric function

$$\frac{d}{dz}{}_2F_1(a,b;c;z)=\frac{ab}{c}{}_2F_1(a+1,b+1;c+1;z).$$

(5.35)

Since the transit-time stems from a bounded system, and we know that at the upper and lower bounds the function ${}_2F_1(.;z)$ is a monotonic single valued function, and as the function results from a series expansion which will also lead to a single valued monotonic function, then we can say that for values of interest $e>0, y\in\{ON\}$, ${}_2F_1(.;z)$ is analytic. If the function ${}_2F_1(.;z)$ is analytic then

$${}_2F_1(.;z_{m=2})>{}_2F_1(.;z_{1<m<2})>{}_2F_1(.;z_{m=1}).$$

(5.36)

This implies that the lengths L

$$L_{m=2}>L_{1<m<2}>L_{m=1}.$$

(5.37)

This in turn implies that if L is larger then $P(l<L)$ will also be larger and therefore

$$C(L_{m=2})>C(L_{1<m<2})>C(L_{m=1}).$$

(5.38)

If the largest correlation is dependent on the largest m then this will also set the dominant H. This relationship can be seen from earlier work on the interaction between correlation and transit-time. This completes the proof.

Remark: If we apply the relationship of the Hurst parameter to the exponent dependence of the correlation (see Section 5.4) we once again obtain our theoretical dependence on H of the map parameter m

$$\frac{m-2}{m-1}=2H-2\Rightarrow H=\frac{3m-4}{2(m-1)}.$$

(5.39)

We have run numerical experiments using the double intermittency map to determine the dependence of H on m. The results of these experiments are shown in Figure 5.3. What we

can see from these results is that there is a general increase in H with increasing m, which is predicted by the theory. However the plane of H is different from that predicted.

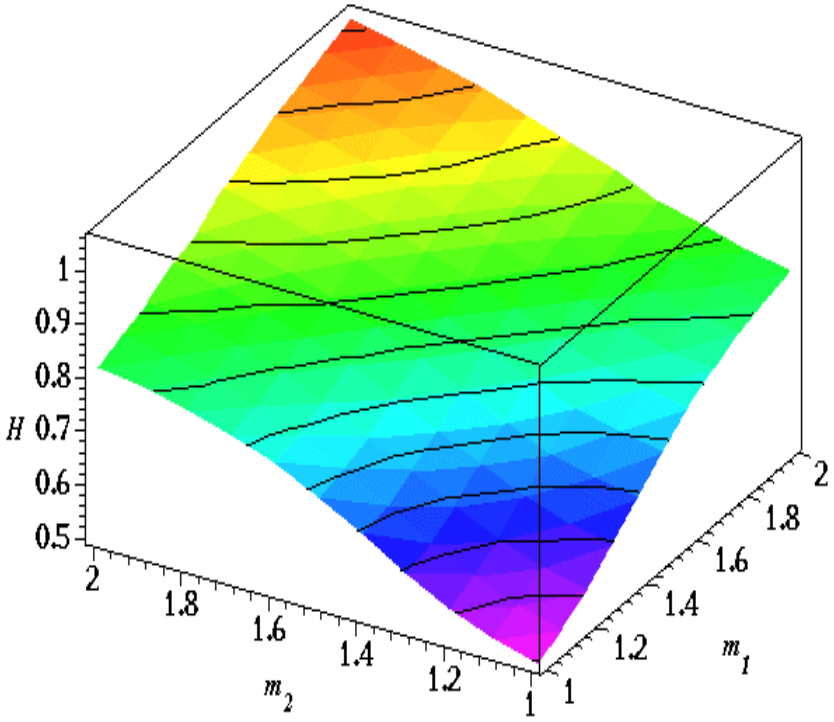

Figure 5.3. H profile for variations in m_1 and m_2 of the double intermittency map: $e_1 = e_2 = 0$ and $d = 0.5$

5.6 Empirical Fit of H

The experimental results show that the dependence of H on m_1 and m_2 is not quite the same as that predicted. In the original derivation of the parameter dependence of H on m the following were assumed:

- a uniform injection probability into the restriction region, and
- that the dynamics of the map once iterated out of the restriction region was independent of the event causing injection into the restriction region.

In our map family the assumption of independence is not valid because the transition between the halves of the map are deterministic. In fact the results suggest that the coupling performs an averaging on H. We can geometrically construct such an averaging behaviour in the following way (see Figure 5.4). We can say that geometrically the radius R is dependent on the values of m_1 and m_2. If we let $m_{\min}=1$ then

$$R = \sqrt{(m_1 - m_{\min})^2 + (m_2 - m_{\min})^2} \ . \tag{5.40}$$

We know tanα from the behaviour of R at $R_{\max}$, the maximum value of $H_{\max}=1$, and H's midrange value $H_{0.5}=1/2$. Hence

$$\tan a = \frac{H_{\max} - H_{0.5}}{\sqrt{2}} = \frac{1}{2\sqrt{2}}.$$

(5.41)

Furthermore we know that $R_{\max}$ occurs when m_1=m_2=2 and $R_{\max}$ has a value of $\sqrt{2}$. The estimate of H is then given by

$$\hat{H} = H_{0.5} + \frac{\sqrt{(m_1 - m_{\min})^2 + (m_2 - m_{\min})^2}}{2\sqrt{2}}.$$

(5.42)

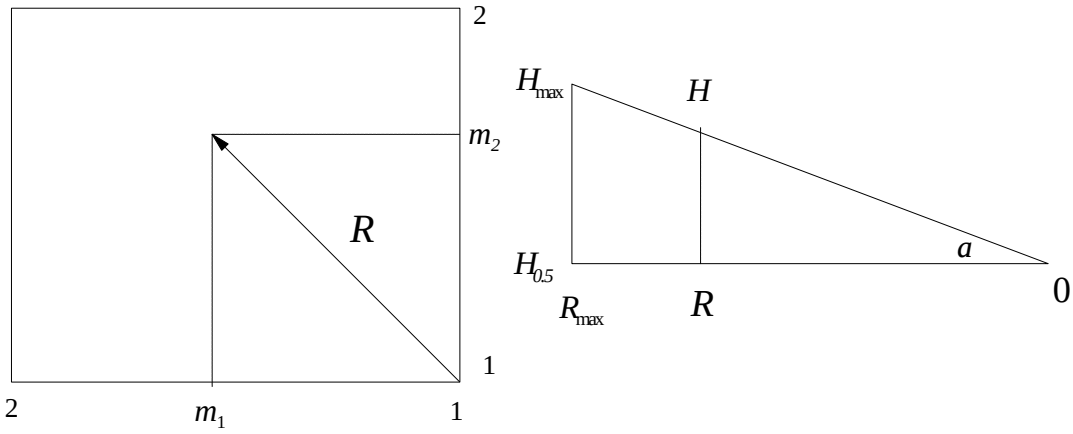

Figure 5.4 Derivation of Empirical Fit

We can see from this construction that coupling imparts dependence of H on contributions from both sides of the map. The H profile for the dependence of H on m_1 and m_2 using this method is shown in Figure 5.5. This has been compared against the actual H profile for m_1 and m_2 and this result is shown in Figure 5.6, where the difference between the empirical fit and the real data is plotted against variations in m_1 and m_2. As can be seen the empirical fit gives a reasonable fit to the data.

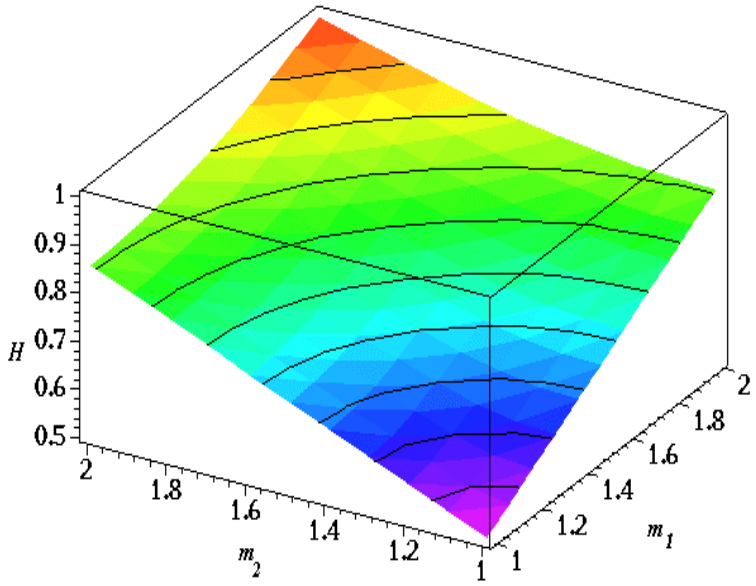

Figure 5.5. Empirical fit on H

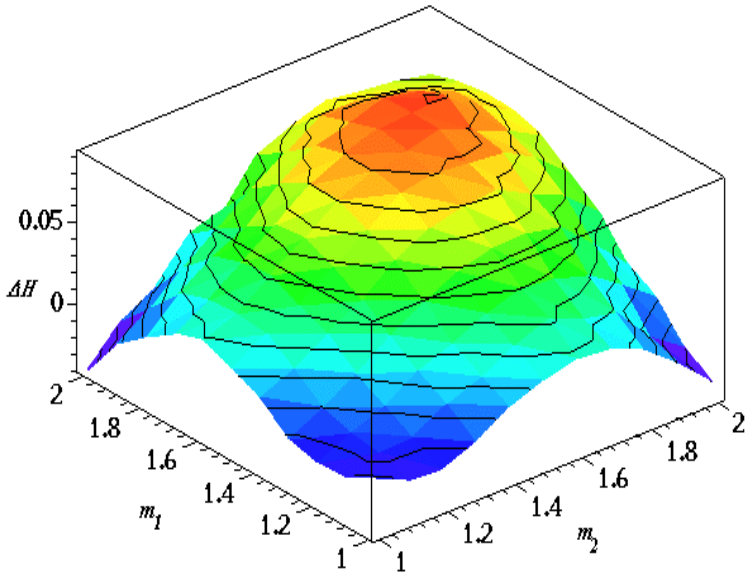

Figure 5.6. Error in H using the Empirical fit

5.7 Dependence of H on other Map Parameters

We have examined the dependence of H on m_1 and m_2 and found that the coupling affects the convergence of H on m. It would then be reasonable to assume that this coupling will also affect H in terms of other map parameters. In order to establish the validity of this assumption, numerical experiments where carried out to establish the dependence of H on:

- e_1 and m_1 while holding m_2 fixed at 1 and d at 0.5
- e_2 and m_2 while holding m_1 fixed at 1 and d at 0.5
- d and m_1 while holding m_2 fixed at 1 and e_1, e_2=0
- d and m_2 while holding m_1 fixed at 1 and e_1, e_2=0

Here we also note that the emitting half of the map contains m_2 and e_2. The Abry-Veitch Wavelet analysis was used to measure H with a sample window size set to 65536 samples. A total of 1000 sample windows were used to establish the value of H for a particular parameter combination. The results of these experiments are shown in Figure 5.7-Figure 5.10.

The results of the effects that the alteration of e and matching m have on H are shown in Figure 5.7 and Figure 5.8. We notice that the only appreciable effect on the alteration of e on H is that any alteration of e away from 0 sets the upper value of H to a constant value less than 1. In the above experiments this was $\overline{H} \approx 0.83$. This was to be expected since e sets the maximum sojourn time and therefore any method which uses a regression technique to a line, as is the case of the Wavelet based method, will be affected by points which alter the general trend of the line. Altering e sets the upper cut off on the transit times (LRD) and hence will

affect the regression fit that can be achieved; however, this does not affect the slope of the line, just where the line is perceived to end.

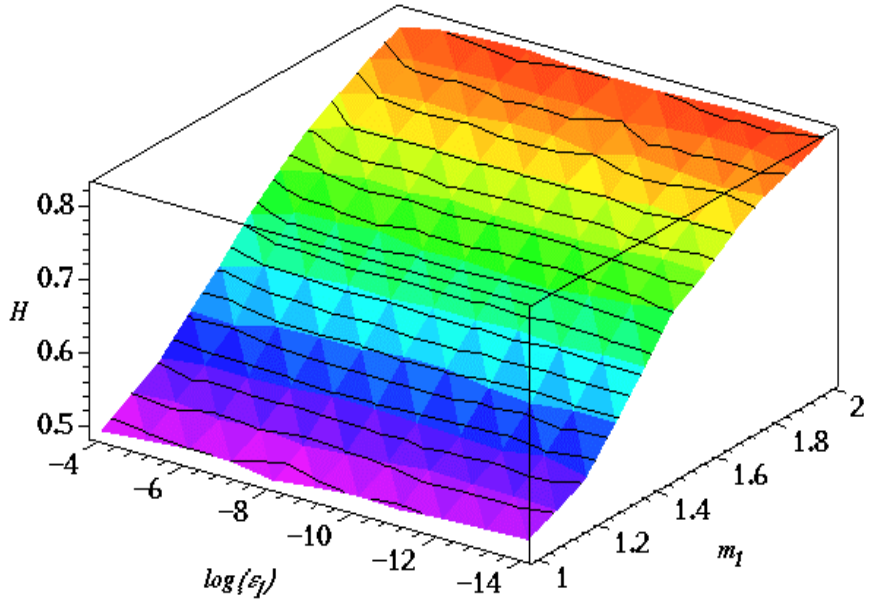

Figure 5.7. *H* profile for variations in m_1 and e_1 of the double intermittency map: m_2 =1, e_2 = 0 and d = 0.5, m_2 is in the *ON* state

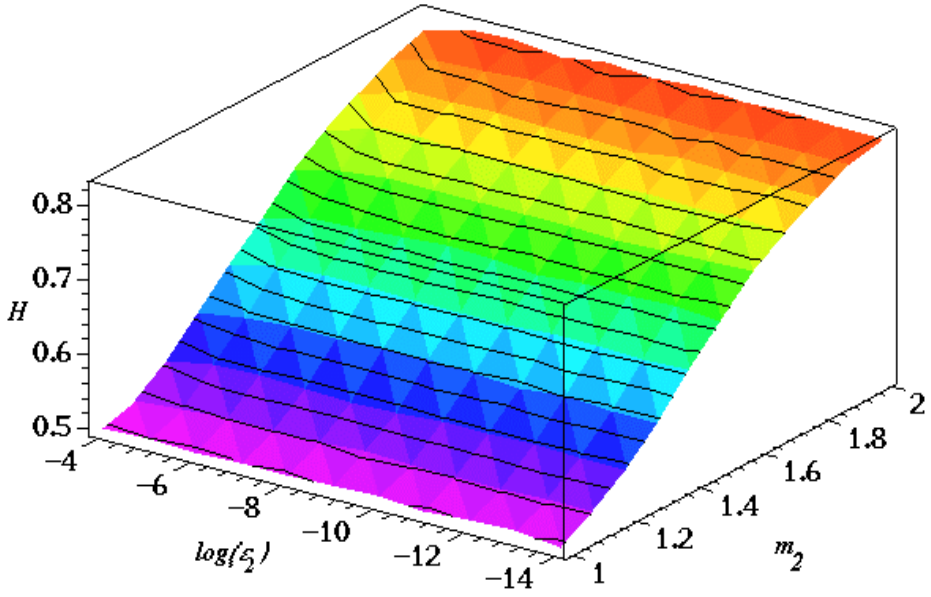

Figure 5.8. *H* profile for variations in m_2 and e_2 of the double intermittency map: m_1 =1, e_1 = 0 and d = 0.5, m_2 is in the *ON* state

The results of the effects of altering d and m on H are shown in Figure 5.9 and Figure 5.10. From these results we noticed that as d alters, the value of H also alters. The value of H diminishes as d moves to lessen the interval in which m exists. Take for example the *OFF* state $x \in (0, d)$, where m_1 describes the order of the polynomial. In this case; as d moves to

lessen the size of the interval, i.e. $d \to 0$, the injections into this interval from the *ON* state are presented with a smaller target area from which long run correlations will result. This has the effect of reducing the number of long run correlation events in an experiment of set length. This in turn affects the convergence of *H*. High *H* values require many long run correlation events to assure convergence on the true value. Less events simply implies that *H* as not yet converged, which is the artefact witnessed in Figure 5.9 and Figure 5.10.

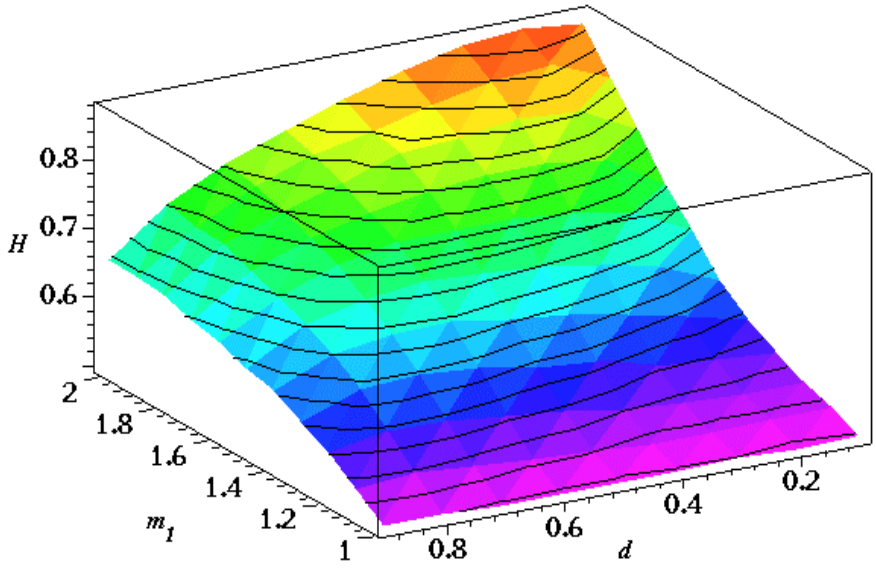

Figure 5.9. *H* profile for variations in m_1 and *d* of the normal map: m_2 =1 and e_1 =e_2 = 0, m_2 is in the *ON* state

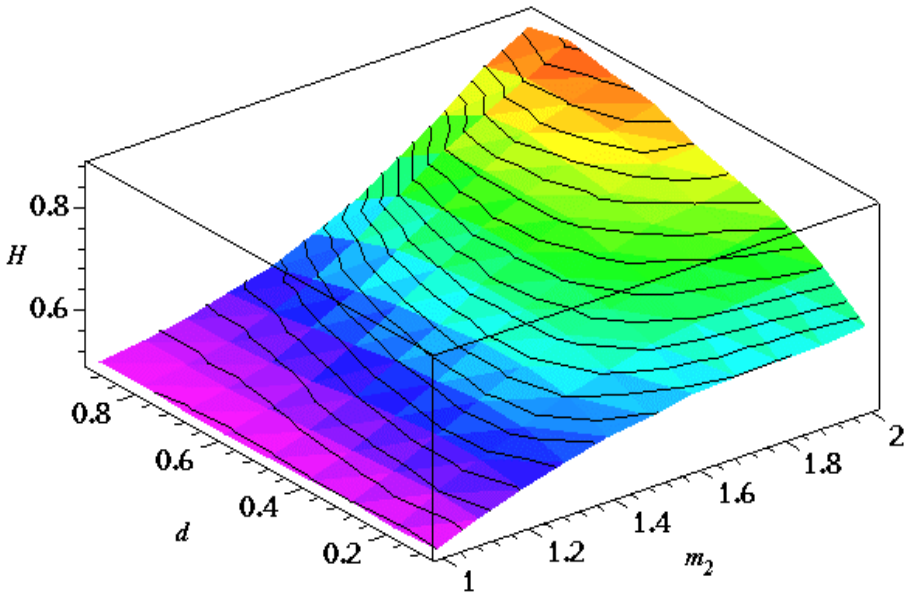

Figure 5.10. *H* profile for variations in m_2 and *d* of the normal map: m_1 =1 and e_1 =e_2 = 0, m_2 is in the *ON* state

We also note from the results, that Figure 5.9 and Figure 5.10 form mirror images of each other. This has the following implications:

- As m_1 and m_2 are varied then the convergence effect on H through the variation of d should be reduced. This is because as d is altered, the LRD from one half of the map reduces while the LRD from the other increases.
- Due to the maps coupling then an average value of H will be observed, hence reducing the effect that d has on the convergence of H.

5.8 Aggregate Behaviour of the Indicator Variable

So far we have only considered the behaviour of the underlying dynamics with regard to H, the correlation and the transit-time. We now direct our attention to the overlying dynamics, i.e. to the behaviour of the indicator variable. In order to consider this behaviour we require a formulation for the measurement of H as observed via the indicator variable. We noted in Chapter 2 that an essential ingredient for self-similarity is aggregation. We also noted that if LRD is present and aggregation is taking place then the variance of the process has a $2H$ dependence. In order to simplify the analysis we follow Ben-Mizrachi [BEN85] and assume that the indicator variable's realisation process is a result of a Bernoulli trial.

We study the maps' self-similar behaviour by observing the cumulative behaviour of the indicator variable after K iterations $z_K = \sum_{i=1}^{K} y_i$, from which the average number of emission will scale as

$$E(z_K) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{K} y_j \right) = KE(y) = K \int_{x\in\{ON\}} r(x)dx = Kl \tag{5.43}$$

and the variance will scale as

$$\begin{aligned} \mathrm{var}(z_K) &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{K} y_j - Kl \right)^2 \\ &= Kl(1-Kl) + 2\sum_{i=1}^{K} \sum_{j>i}^{K} E(y_i y_j) \end{aligned} . \tag{5.44}$$

For derivations of these equations and subsequent ones see Appendix C. If we examine the correlation term in equ.(5.44) we note that the y_i can only take on the value of 0 or 1. Therefore the expectation can only take on non-zero values when y_i and y_j = 1, i.e. $E(y_i y_j) = P\{y_j = 1, y_i = 1\}$. If we consider the expectations in terms of probabilities and use the probability identity $P(A,B) = P(B)P(A|B)$, then the expectation can be written as $E(y_i y_j) = P\{y_j = 1, y_i = 1\} = P\{y_i = 1\}P\{y_j = 1 | y_i = 1\}$. Since $P\{y_i = 1\} = \lambda$, then we have

the relationship $E(y_i y_j) = l \;\; P\{y_j = 1 | y_i = 1\}$. By replacing the expectation with the probability and performing the double summation of equ.(5.44) and defining $C(n = j - i) \equiv P\{y_j = 1 | y_i = 1\}$ we notice that a regular pattern emerges from which we can write

$$\mathrm{var}(z_K) = Kl(1 - Kl) + 2l \sum_{i=1}^{K-1} (K - i) C(i) \,,$$

(5.45)

where *C(i)* is the correlation function. We now have a method for measuring the variance via the correlation function *C*(*i*).

If we recall the relationship between transit-time and correlation (Section 5.4) we can go further and use this relationship to obtain upper bound on the aggregated variance in terms of *H*. A trial fit for the correlation function is $C(i) \approx a\, i^{\,b} + l$ [‡], where *b* contains terms in *m*. We can substitute this ansatz for *C*(*i*)

$$\mathrm{var}(z_K) = Kl(1 - Kl) + 2l \sum_{i=1}^{K-1} \left\{ (K - i)(a i^{\,b} + l) \right\}.$$

(5.46)

We can expand this to obtain

$$\mathrm{var}(z_K) = Kl(1 - Kl) + 2la \sum_{i=1}^{K-1} i^{\,b} (K - i) \,.$$

(5.47)

We can approximate this summation with an integral an obtain an inequality which becomes the upper bound for the aggregate variance

$$\mathrm{var}\left(Z_{K_y} \right) \le Kl(1 - l) + 2la \int_{1}^{K-1} i^{\,b} (K - i) di \quad .$$

(5.48)

On integration and after taking limits as $K \to \infty$ we obtain

$$\mathrm{var}\left(Z_{K_y} \right) \le Kl(1 - l) + 2la \left\{ \frac{K^{\,b+2} - K}{b + 1} - \frac{K^{\,b+2} - 1}{b + 2} \right\} \quad .$$

(5.49)

What we can see here is that as the *K* becomes large the upper bound on the variance has a *b*+2 dependence. When $m \ge \frac{3}{2}$ this relates to a dependence on the dominant map parameter *m* of $(3m - 4)/(m - 1)$. We have tested this method of measuring *H* on the Bernoulli shift map

[‡] Compare with equ.(5.32).

(m_1=m_2=1) where we can recover *H* analytically. For the Bernoulli shift map the invariant density, *r(x)*, is constant. The average load is $l = \int_d^1 r(x)dx = 1-d$ and $C(n=j-i)=l$ for all *n*, hence the variance of a Bernoulli random process is $\mathrm{var}(z_K) = Kl(1-l)$, which has a slope of unity. This equates to a Hurst parameter *H*=½. This has been confirmed by numerical experiments, the results of which are shown in Figure 5.11.

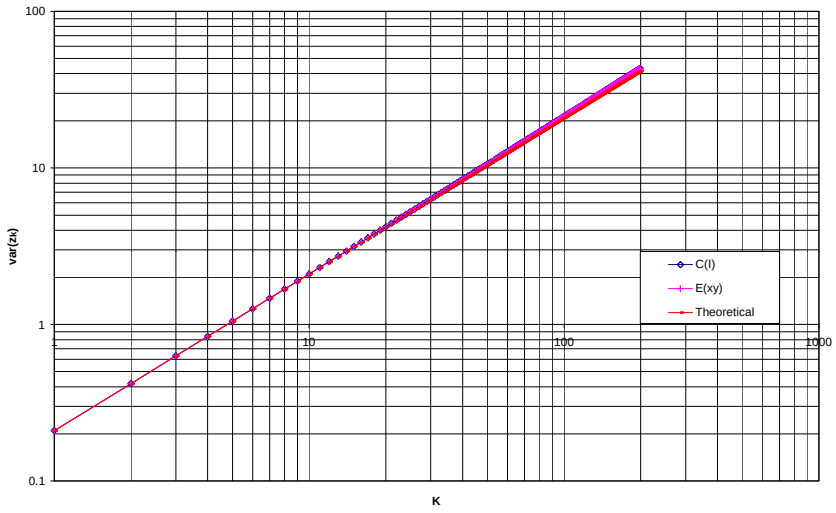

Figure 5.11. Measurement of *H* for the Bernoulli shift map using the for the relationships given by equ.(5.44) and equ.(5.45) against theoretical results.

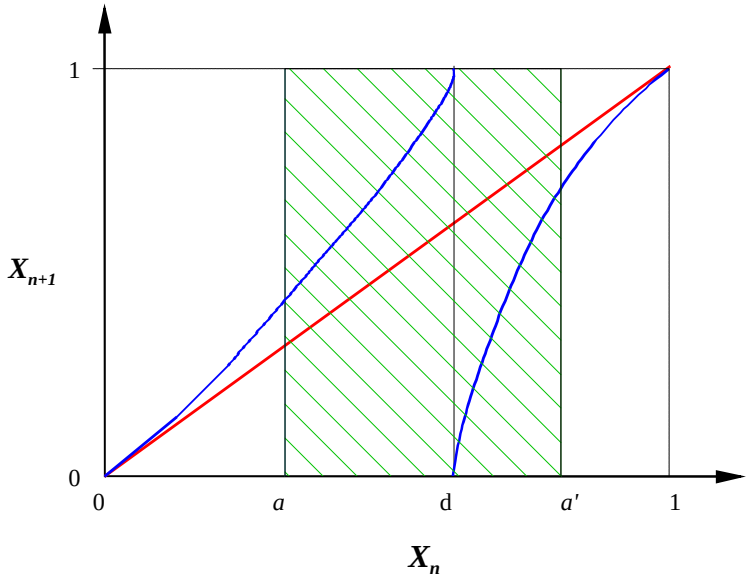

Figure 5.12. Separation of the correlated and decorrelated regions of the map via Ben-Mizrachi's *a*.

5.9 Remarks and Conclusions

Remark: Theoretical Considerations on H and Convergence

The theoretical considerations on *H* stem from Ben-Mizarachi's [BEN85] simplification of an intermittency map model. Their model assumed that the intermittent behaviour was separated by laminar regions caused by random reinjections into the neck region of their map with a

uniform probability distribution. The resulting laminar transients (transients of high correlation) eventually escape past some point *a*, (see left hand side of Figure 5.12).

In the intermittency map family the laminar transitions are coupled because the left-hand (LH) and right-hand (RH) halves of the map are deterministically linked. Therefore, we are dependent on the invariant density of the non-active half of the map for the reinjection probability into the active half of the map. Subsequently the predicted values of *H* become distorted with respect to *d* since this will decide the balance of the reinjection probability. This type of behaviour can be seen in Figure 5.9 and Figure 5.10. We also note that once past $m = 3/2$ the convergence to stable value of *H* becomes problematic. This is since the correlation sum (equ.(5.45)) behaves as a Riemann Zeta function and it is known that the convergence of this function is very slow for exponent values near 1 (see for example Bender and Orzag p397.[BEND84])[26]. This effect is demonstrated in Figure 5.13 where the estimate of *H* from the Abry Veitch wavelet analysis is shown against sample size. We can see that for a value of $m < 3/2$ the convergence is well behaved as the sample size is increased. However the same cannot be said for the case of $m > 3/2$ where the mean value converges slowly and has considerable variance from which confidence in a trend would be hard to justify.

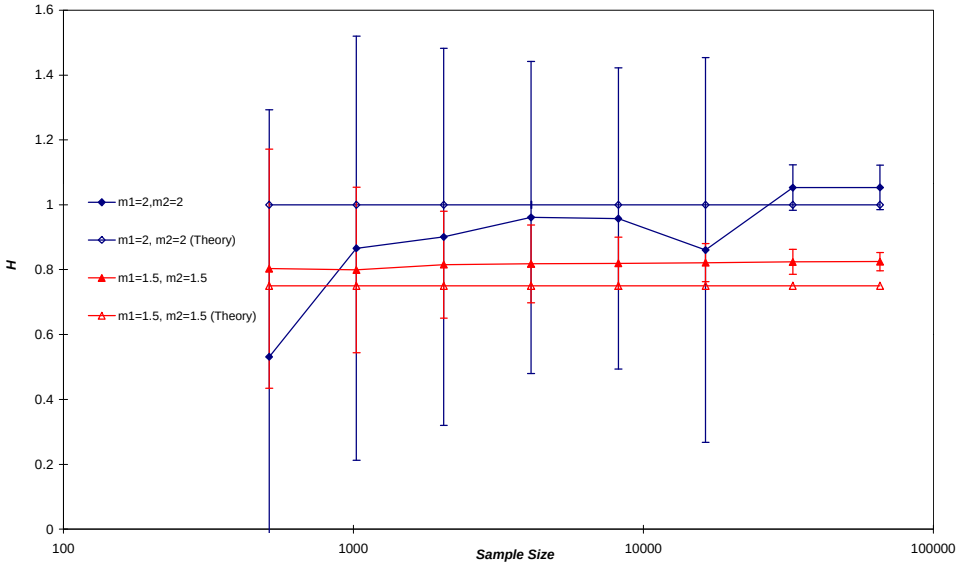

Figure 5.13 Convergence of *H* on sample size for 1000 experiments. *H* calculated using Abry-Veitch wavelet analysis

Remark: The separation of map LRD and SRD behaviour

We note that the theory still gives reasonable estimates of *H*. This is since some decorrelation between the map halves occurs. We can see from Figure 5.12 that past some point *a* in the LH half of the map (and *a`* in the RH half of the map) the long run correlations will end in an apparent random reinjection into the other half of the map. This results in a to-and-fro of

iterations between the LH and RH halves of the map reminiscent of the behaviour observed in the Bernoulli shift map (this is the hatched region in Figure 5.12). We can go further than this and obtain an idea of how big the distance a - a` is. This can be done by comparing when iterations for the Double intermittency and the Bernoulli shift maps have similar transit times for a common starting point, y_{in}[27]. This type of analysis is shown in Figure 5.14, where the differences in transit-time, $\Delta\tau$, between a double intermittency map (with parameters m_1, m_2 = 2 and d = 0.75) and the Bernoulli shift map (with d = 0.75) are compared for y_{in} values in the range, $0 \le y_{in} < d$ and e values in the range $1*10^{-12} \le e \le 1*10^{-4}$. In Figure 5.14 we have also plotted a plane surface (blue) held at a level of $\Delta\tau$ corresponding to when $\Delta\tau$ becomes less than one order of magnitude. Where the (y_{in}, e) plane cuts the plane surface we consider the double intermittency and Bernoulli shift maps to have the same transit time behaviour, i.e. the double intermittency map begins to decorrelate. The y_{in} value at which the intersection between the planes occurs is taken to be the value of a. In this example the point a is considered to be $a \approx 0.12$. Similar results can be obtained for the other half of the map. Therefore we can say that the region of the map that is contained within a-a` is the zone of decorrelation.

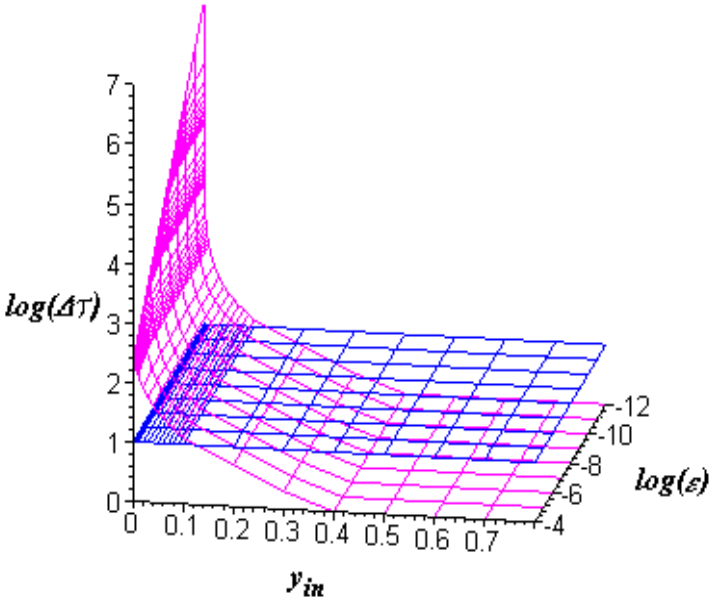

Figure 5.14. *log*(Δτ) against y_{in} and *log* (ε) for case d =0.75

From this last remark we make the following conjecture: Ben Mizrachi's point a exists and acts as the separation between LRD and SRD behaviour. That is to say that when

- $x_n < a$ we have LRD iff ε <<1.
- $x_n > a$ we have SRD.

[26] The Reimann Zeta function $z(b)$ is defined in terms of an infinite sum: $z(b) = \sum_{n=1}^{\infty} n^{-b}$.

This is since for $m \in [3/2,2]$ for $x_n > a$ behaves like $m \in [1,3/2)$. Therefore SRD can be defined as occurring when the order of magnitude, O_n, to leave the state, is the same as an equivalent Bernoulli shift map.

Remark: Decoupling the Map

This last remark leads to the suggestion that if the LH and RH halves of the map could be decoupled, then the theoretical predictions for *H* would improve. This could be done by randomising the transition entry point into the other half of the map. In actual fact work carried out recently by Mondragón [MON98] has investigated this.

Remark: Time Granularity

A more appropriate measure than time is the number of iterates required to leave a map state. This is since the granularity of the map can then be set to reflect the time scale over which the emissions of the map are being considered. In short the iterations are really a timeless quantity that is given meaning by applying some time interval corresponding to each iteration of the map.

Theoretical Behaviour of H

What we have found is that when the map family has both halves of the map coupled, then the behaviour of *H* depends on all the map's parameters. This parameter dependence is not given by a simple formula. Nevertheless, these restrictions on the theory do not constrain their use as traffic models. Due to their evaluational simplicity it is easy to obtain a parameterisation of *H* with the map's parameters, providing a good tool to simulate traffic with a specific self-similarity. It is proposed that since the map family can generate traffic speedily and easily then it would not be difficult to obtain numerical parameterisations of *H* with respect to a given set of map parameters. The differences between the theory and experiment are due to the strong assumption that the *ON* and *OFF* events are independent.

The Contribution

The contribution in this chapter has been:

- A generalised formulation on the transit time which holds for the perturbed map family. The potential that this formulation has is that it can form the basis for input to queueing analysis with heavy tailed *ON/OFF* distributions.
- An improved theory of *H* dependence on the parameter *m* for the map family. The improved theory of *H* dependence is in two parts: a limiting behaviour on the dominant *m*, and an empirical fit for practical cases when $d = 0.5$. Furthermore, we have shown that a numerical parameterisation approach is possible for *H* because the maps produce predictable results. We have also shown that for the coupled map *d* has a major impact on *H* because of the determinism in the transitions that the coupling forces on the map.

[27] Recall that the Bernoulli shift map produces random (decoupled) behaviour.

- A connection between transit-time and the correlation function which has enabled a variance method to be developed for the measurement of *H* from a map's indicator variable.

6 Source Aggregation

The results of Taqqu[TAQ86, 97] and Williger [WIL95, 97] show that *ON-OFF* sources with LRD in at least one state, when aggregated , yield second order self-similar traffic. We know from the previous chapter that non-linear chaotic maps can be parameterised to give LRD in either *ON/OFF* state. Studies carried out by Ben-Mizrachi [BEN85] show that the type of non-linear map used by Pruthi [PRU95a] has the correct type of realisation behaviour that yields a Hurst parameter in the desired range $H \in \left(\frac{1}{2},1\right)$.

Pruthi [PRU95a, 95b] developed a "one-step" and an "*N* one-step" aggregate model. The resulting behaviour of the former was not accurate [I], and the latter sacrificed speed for increased accuracy. In the next chapter we develop a new accelerated "two-step" aggregate model which shows a speed improvement over Pruthi's "*N* one-step" model.

Because non-linear map models can model individual source streams then it is possible to model the aggregated traffic stream behaviour by a single non-linear "equivalent" map. In this chapter we report on aggregation methods that achieve this. The aggregation methods developed here depend on the preservation of the invariant density of the aggregated traffic stream. The motivation for doing this is two-fold, i.e. by preserving the invariant density:

- we preserve the dynamics of the arrivals; and
- we also preserve the LRD structure of the aggregated traffic stream

In this chapter we also develop a new parameterisation method for an equivalent aggregate map iterated *N* times (i.e for a single intermittency map representing *N* sources). We investigate four methods for the parameterisation of the map's discriminator value, *d*, via the invariant density and required load. The map parameter *d* can be viewed as a course adjuster for the bursty behaviour of the map. The parameterisation methods are compared and the most accurate method is then selected. We then use this parameterisation method to parameterise an aggregate map and compare the queueing behaviour of the aggregated map against the queueing behaviour of multiple map sources, showing good performance over a range of parameter values.

6.1 Derivation of the Equivalent Single Map Parameters

The mathematical description of the single intermittency map is given by

$$x_{n+1} = \begin{cases} e + x_n + cx_n^m & ,0 < x_n \le d \\ \dfrac{x_n - d}{1-d} & ,d < x_n < 1 \end{cases} \tag{6.1}$$

where *c = (1-e-d)/d*. The indicator variable $y_n = 1$ indicates when a full packet/cell is generated and is related to the map iterates by

$$y_n = \begin{cases} 0 & , 0 < x_n \le d \\ 1 & , d < x_n < 1 \end{cases} \tag{6.2}$$

The single intermittency map is shown in Figure 6.1. The properties that this map has in common with traffic are:

- it displays *1/f* noise with a power spectrum decay of $\omega^{-(2m-3)/(m-1)}$ where ω is the frequency if $3/2 < m < 2$,
- it produces self-similar traffic with Hurst parameter $H=(3m-4)/(2m-2)$ if $3/2 < m < 2$.

The parameters *e* and *m* are related to the intermittent behaviour of the map. If an iterate of the map is very close to the origin, the orbit slowly moves away from the origin. The average number of iterates to move away from the origin is proportional to $e^{-m/(m-1)}$ and corresponds to a passive period where no traffic is generated. The average time that an orbit of the map spends in the *ON* region can be obtained from the invariant density via the Frobenius-Perron equation (see section 4.8)

$$r(x) = \int_0^1 r(y)d(x - f(y))dy\,. \tag{6.3}$$

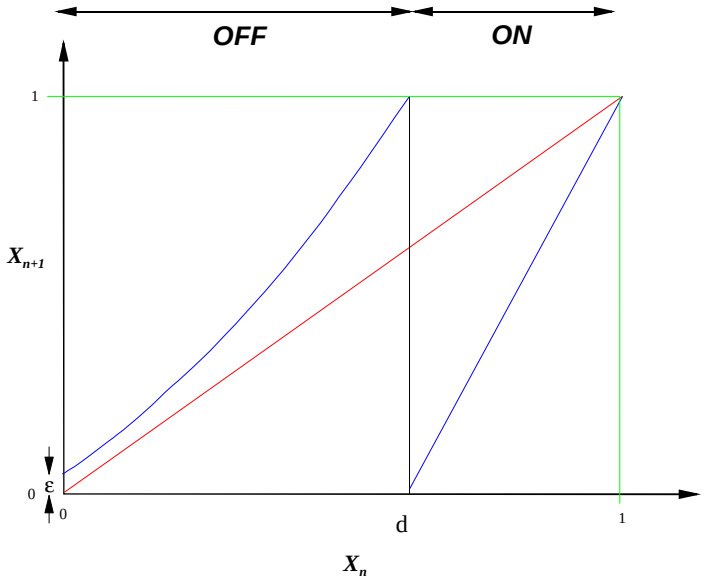

Figure 6.1 The single intermittency map

The invariant density describes how often an orbit of the map visits any region of the interval (0,1). For the case where $m = 2$, the calculation of the invariant density for the intermittency map can be simplified by splitting the invariant density into contributory regions as follows

$$r(x) = \begin{cases} \dfrac{r(y_2)}{|f_2'(y_2)|} & ,0 < x_n \le e \\ \dfrac{r(y_2)}{|f_2'(y_2)|} + \dfrac{r(y_1)}{|f_1'(y_1)|} & ,e < x_n < 1 \end{cases}, \tag{6.4}$$

where $y_i = f_i^{-1}(x)$ the backward iteration of the map and $f_i'(.)$ is the first derivative of the forward iteration of x_n. By performing the backward iteration and taking the derivative, equ.(6.4) becomes

$$r(x) = \begin{cases} (1-d)\,r\left(x(1-d)+d\right) & ,0 < x_n \le e \\ \dfrac{r\left(\frac{\sqrt{4c(x-e)+1}-1}{2c}\right)}{\sqrt{4c(x-e)+1}} + (1-d)\,r\left(x(1-d)+d\right) & ,e < x_n < 1 \end{cases}.$$

(6.5)

As it stands this equation is a bit unpalatable but nevertheless is useful since some approximations can be made for certain points on the invariant density curve

$$\underset{x \in [0,e)}{r(x)} \approx (1-d)\,r(d)$$

(6.6)

$$\lim_{\substack{x \to 1 \\ e \to 0}} r\left(\frac{\sqrt{4c(x-e)}-1}{2c}\right) \to r(d)$$

(6.7)

from which we can make the following approximation for x = 1, ε→0

$$r(1) \approx \frac{r(d)}{(2-d)}.$$

(6.8)

For $e<x<<1$ the bottom of equ.(6.5) can be approximated to give the first differential of the invariant density

$$r'(x) \approx \frac{(1-d)\,p(d) + 2c(x-e)\,r(x)}{e\left(1-2c(x-e)\right)}.$$

(6.9)

With *r'(x) = 0* this yields an implicit equation for peak of the invariant density but more importantly from our point of view we can see that in this region the invariant density falls off as 1/ε.

The relationship between the points given above (equ.(6.6) to equ.(6.8)) and the invariant density is shown in Figure 6.2. The average time that the orbit spends in the *ON* region or equivalently the average number of packets generated by the map after *N* iterations is

$$Nl = N\int_d^1 r(x)dx.$$

(6.10)

Hence, the parameter *d* can be used to adjust the average load of the traffic but is also related to the bursty behaviour of the map.

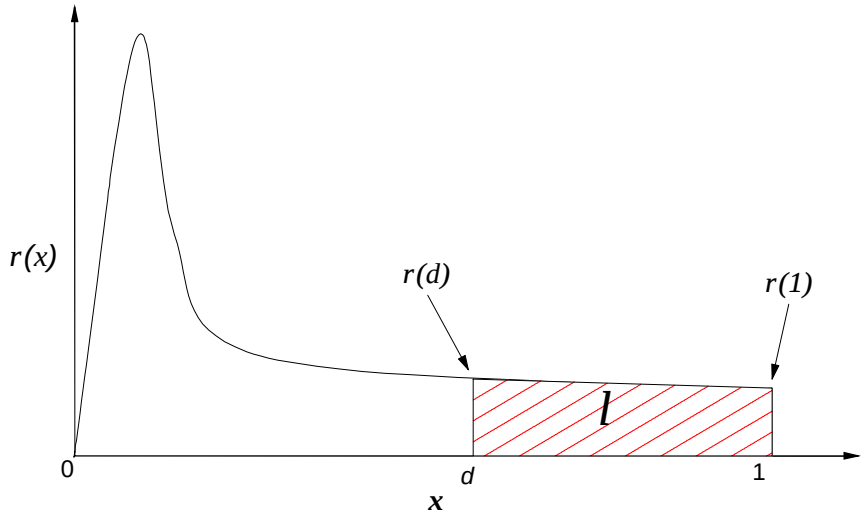

Figure 6.2 The single intermittency map invariant density showing the points *r(d)* and *r(1)*

Under aggregation we require a single map to produce traffic which is similar in output to aggregation of the single maps. To do this we require a method to preserve the tail of the invariant density of the individual map sources[28]. It is the aggregation of many such map sources which gives the LRD and self-similar behaviour [PRU95b]. From equ.(6.8) we know the value of two points on the invariant density curve in terms of the map parameter *d*

$\left(r(x), d\right)$ *and* $\left(r(\mathrm{x}),\ 1\right)$.

(6.11)

These points are used to derive the approximations for the tail in terms of the map parameters.

6.1.1 Exponential Approximation

The tail of the invariant density curve decays as a Pareto law. Hence we use an approximation of the form

$$y = Ax^{-1}.$$

(6.12)

From this we can form a set of simultaneous equations

$$r(x) = Ax^{-1} \quad and \quad r(1) = A,$$

(6.13)

from which we can isolate the constant A

$$r(1) = A \Rightarrow \frac{r(d)}{(2-d)}.$$

(6.14)

Moreover, we now have a general approximation for the tail of the curve in terms of a parameter of the map

[28] Intuitively this is analogous to preserving the events which occur in the tail of the distributions seen in the real traffic measurements [ERR96, PAX95, LEL94].

$$r(x) = \frac{r(d)}{(2-d)} \cdot \frac{1}{x}.$$

(6.15)

The area under the curve between *d* and 1 yields the probability of the source being *ON*, ie λ, and is given by[29]

$$l = \frac{r(d)}{(2-d)} \int_d^1 \frac{1}{x} dx.$$

(6.16)

Integrating the above yields

$$l = \frac{r(d)}{(2-d)} \ln\left(\frac{1}{d}\right)$$

(6.17)

which if rearranged yields

$$d = e^{-\frac{(2-d)}{r(d)} l}.$$

(6.18)

The difficulty with equ (6.18) is that it is a parametric equation. However we can obtain an empirically derived value for *r(1)* and by equ (6.14) we can obtain a value for *r(d)*, and equ (6.18) becomes

$$d = e^{-\frac{l}{r(1)}}.$$

(6.19)

Hence, we can obtain the equivalent map discriminant value and predict the behaviour of the maps under aggregation.

6.1.2 Modified Exponential Approximation

An alternative to the approximation at equ.(6.19) can be made in the following manner. From equ (6.16) we note that on integrating we obtain a logarithm. We can approximate this integral in the following way

$$\frac{1}{x} \cong \frac{1}{x^{1+d}}, 0 < d << \tfrac{1}{2},$$

(6.20)

where $\delta = 1/y$ and ideally $y >> 1$. Subsequently we can obtain an approximation for λ of the form

$$l = \frac{r(d)}{(2-d)} \int_d^1 \frac{1}{x} dx \cong \frac{r(d)}{(2-d)} \int_d^1 \frac{1}{x^{1+d}} dx,$$

(6.21)

which on integration yields

[29] Here we have intentionally used λ for the probability of being on since this ties in with the mean arrival rate into a queueing system.

$$l = \frac{r(d)}{(2-d)}\frac{1}{d}\left[d^{-d} - 1\right].$$

(6.22)

We can then isolate *d* and by substituting for *r(d)* where appropriate we obtain the approximation

$$d = \left(\frac{r(1)}{\frac{l}{y} + r(1)}\right)^{y}.$$

(6.23)

6.1.3 Trapezoidal Approximation

An intuitive approximation for the tail of a distribution that decays very slowly is a trapezoidal approximation. The advantage of this approximation is its simplicity. The equation for the area of a trapezium is

$$A = \frac{1}{2}(b+c)e,$$

(6.24)

where we interpret the above as:

- A is the probability of being on, i.e. λ
- e is the distance *(1-d)*
- b is the point *r(d)* taken from the aggregated density curve, and
- c is the point *r(1)* taken form the aggregated density curve.

By substituting appropriate values we arrive at the following expression for λ

$$l = \frac{1}{2}\left(r(d) + r(1)\right)(1-d).$$

(6.25)

From equ (6.14) we substitute *r(1)* for the *r(d)* term and we then obtain an expression for *d* in terms of the invariant densities of the map

$$d = \frac{4 \pm \sqrt{16 - 4\left(\frac{3r(1)-2l}{r(1)}\right)}}{2}.$$

(6.26)

6.1.4 Improved Exponential Approximation

We can also add a further constant term to equ (6.12) and obtain a hyperbolic equation of the form

$$y = Ax^{-1} + B.$$

(6.27)

We can obtain simultaneous equations in terms of known points of the invariant density. Solving these equations we arrive at expressions for A and B

$$\begin{aligned} A &= d\,r(1) \\ B &= (1-d)\,r(1) \end{aligned}.$$

(6.28)

The approximation for the invariant density is then

$$r(x) = r(1)\left[\frac{d}{x} + (1-d)\right].$$

(6.29)

From this we can obtain λ,

$$l = r(1)\int_{d}^{1}\left[\frac{d}{x} + (1-d)\right]dx.$$

(6.30)

Integrating equ.(6.16) yields

$$l = r(1)\left[(1-d)^2 - d\ln(d)\right].$$

(6.31)

What we now have is an implicit equation in terms of *d*. This is impossible to solve analytically. However we can still solve this numerically in the following way. We define a new function F(*d*,*h*), where *h* is defined as $\lambda / r(1)$. F(*d*,*h*) is then

$$F(d,h) = (1-d)^2 - d\ln(d) - h\,.$$

(6.32)

We can solve F(*d*,*h*) = *0* numerically using Newton-Raphson.

$$d_{n+1} = d_n - \frac{F(d_n,h)}{dF(d_n,h)}.$$

(6.33)

The numerical solution for *d* against η is

$$d = 1.00037 - 1.00005h + 0.426561h^2 - 0.258587h^3.$$

(6.34)

An interesting point from this result is that knowing that the emissions (full/empty cells/packets) from the maps are totally independent of each other, i.e. the events are mutually exclusive, then we can use the same technique to predict the required map *d* parameter from combination (aggregation) of more than two maps. This is since under normalisation we have the following behaviour

$$l_{NL} = \frac{l_1 + l_2 + \cdots + l_{N1}}{N}$$
$$r(1)_{NL} = \frac{r(1)_1 + r(1)_2 + \cdots + r(1)_N}{N},$$

(6.35)

where λ_{NL} and $r(1)_{NL}$ refer to the single equivalent map values, and N is the total number of sources being aggregated. Since the definition of h_T is in effect $\lambda_{NL}/r(1)_{NL}$ then cancellation of N occurs. We can therefore say that

$$h_T = \frac{\sum_{i=1}^{N} l_i}{\sum_{i=1}^{N} r(1)_i}.$$

(6.36)

6.2 Limitations on the Methods

As $d \rightarrow 1$ the map becomes increasingly linear in its behaviour and as a result tends to a white noise generator (random number generator in the interval (0,1)). Similarly if $d \rightarrow 0$ then again the map tends to a white noise generator. The difference between the two is in the interpretation of the indicator variable. For example, if iterates of the map falling to the left of d are interpreted as a packet generation/full cell emission then as $d \rightarrow 1$ the source becomes permanently off. The extreme effects of d on the invariant density can be seen in Figure 6.3. What we gather from this observation is that for the intermittency map to be a good imitator of bursty traffic then d has to be placed in-between the extremes.

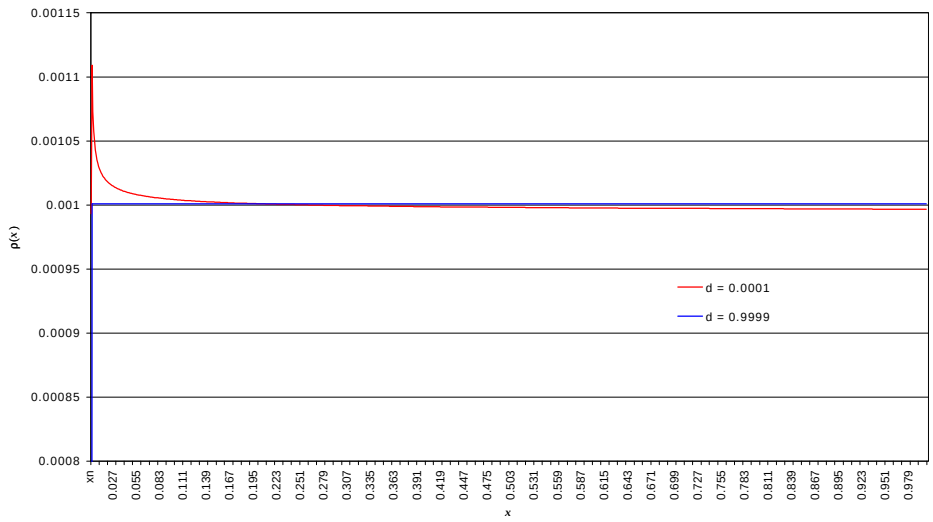

Figure 6.3 Effects of extreme d on invariant density of the single intermittency map

As we have seen the usefulness of the single intermittency map as a model of a bursty traffic source breaks down as the map discriminant, d, approaches the limits of the map interval. For

this reason some realistic limits must be assumed for the traffic behaviour under aggregation that we wish to model, i.e. over what range of utilisation values do we wish to predict the effects of aggregation using the intermittency map. Studies of connectionless traffic [FOW91, SHO80] suggested external daily link utilisation values of around 1% and external peak minute utilisation values of around 17% as being realistic; Fowler [FOW91] also mentions that the peak internal minute utilisation is occasionally in excess of 50%. Under these guidelines we have assumed the overall aggregated load for the simulators to be in the range

$$0.01 < r_T \leq 0.5, \tag{6.37}$$

where r_T is the overall utilisation for the traffic entering the simulators.

6.3 Aggregation Scenario

In order to test out the ideas of source aggregation using chaotic maps the following scenario was used. This scenario assumes that utilisation of the network will remain at levels equal to or less than 50%.

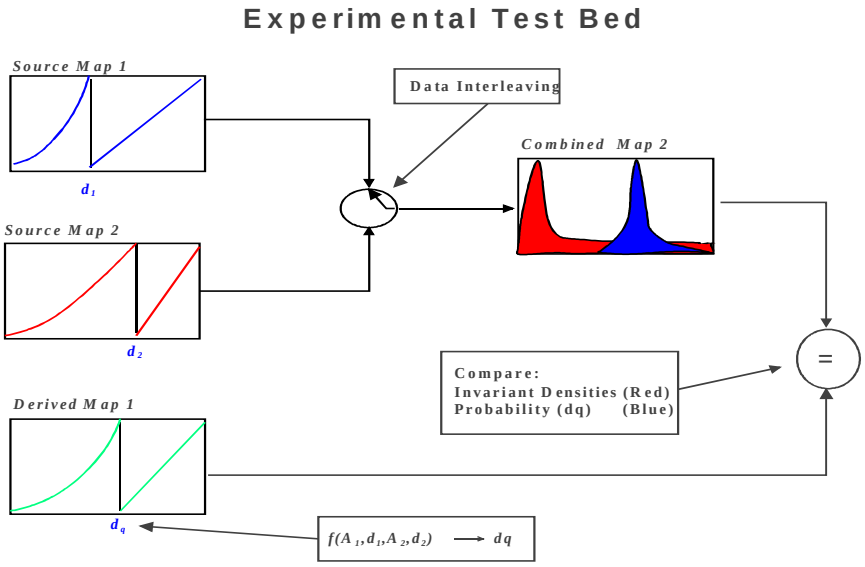

Figure 6.4 Map aggregation test-bed

The test-bed models traffic on a link between nodes, i.e. we are examining the traffic profiles before they enter any kind of queueing system. A schematic of the test-bed is shown in Figure 6.4. We aggregate single intermittency maps with different parameters - initially two such sources are aggregated. We then recreate the aggregated traffic behaviour through a single "equivalent" map that has parameters derived from the parent maps. After aggregation we compare the aggregated traffic stream statistics with the derived single map statistics. The error in behaviour between the single map and the aggregated traffic is then compared.

The four prediction methods given in section 6.1 have been compared over the utilisation ranges given in section 6.2. For convenience in these evaluations we assume that the total *ON* probability of the map (area under the invariant density curves between *d* and the interval limit

1) is equivalent to the mean inter-arrival time, λ, of the sources entering a deterministic server queueing system, such that the combined λ of the maps is equal to the utilisation of the system.

In order to obtain the utilisation values required one of the maps' *d* parameters is held fixed at a value near the interval extrema (d_f = 0.1 or 0.9). The other maps *d* parameter is permitted to vary in the interval [0.1, 0.9]. The ε value for the maps is held fixed at ε = 0.0001 throughout these sets of experiments.

The results are obtained in the following manner:

- The λ for the individual maps are calculated. - the map densities are obtained by using the Frobenius-Perron operator.
- The maps' λ are then combined and λ_{NL} calculated.
- From the individual map densities the $r(1)$ are extracted and combined to give $r(1)_{NL}$.
- The λ_{NL} and $r(1)_{NL}$ are then used to predict a value of d_n for a new single map which will yield the same λ_{NL} .
- This new map invariant density is then obtained using Frobenius-Perron.
- The map λ_n is calculated and compared against λ_{NL}. and this comparison is recorded. This is done through obtaining an error value, i.e.

$$E = \frac{l_n - l_{NL}}{l_{NL}} .$$

(6.38)

- The difference between the map *d* values is also recorded and is termed the Δ value, ie

$$\Delta = \left| d_f - d \right| .$$

(6.39)

The results are displayed as a plots of the error in the single map λ against the discriminant Δ value. Table 6.1 summarises the plot scenarios.

Fig.	Map_1 *d* (fixed)	Map_2 *d* (range)
Figure 6.5	0.1	0.1 to 0.9
Figure 6.6	0.9	0.9 to 0.1
Figure 6.8	0.25	0.25 to 0.9
Figure 6.9	0.9	0.9 to 0.25

Table 6.1 Summary of Plot Scenarios

6.4 Results: Prediction of d

From the results show in Figure 6.5and Figure 6.6 we noticed that:

- the Improved Exponential Approximation (IEA) method performed the best with an error range of +15% to -5 % error,

- the next best approximation is the Trapezoidal (TRAP) with an error range of -18% to -11%.
- the remaining prediction techniques do not perform as well. But it is interesting to see that the Modified Exponential Approximation (MEA) follows closely the Exponential Approximation (EA) (this may yet turn out to be a useful approximation technique).

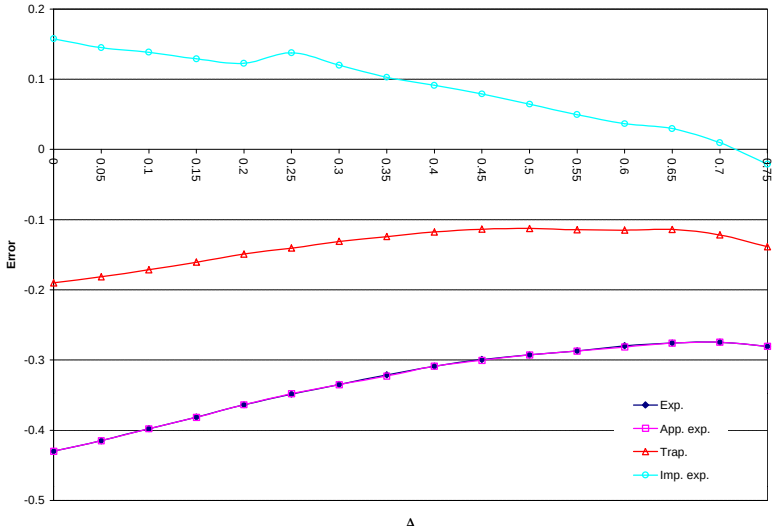

Figure 6.5 Comparison for various λ prediction methods for a two map scenario: Map_1 d value fixed at 0.1. Map_2 variable d $\hat{I}(0.1, 0.9)$

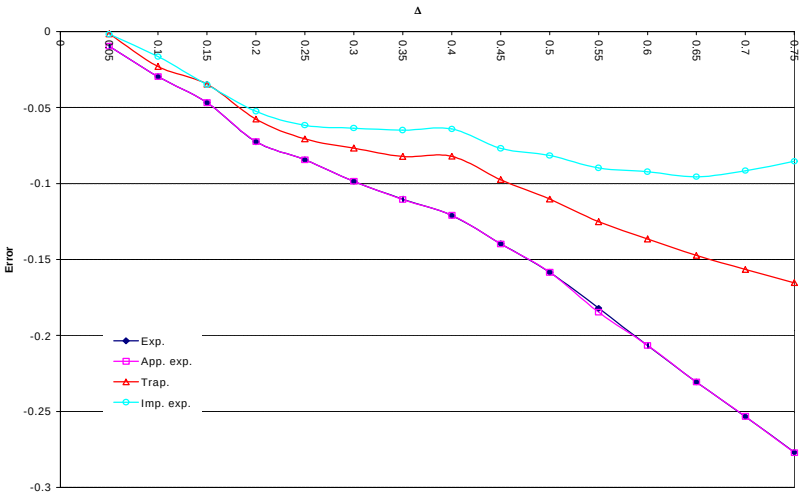

Figure 6.6 Comparison for various λ prediction methods for a two map scenario: Map_1 d value fixed at 0.9. Map_2 variable d $\hat{I}(0.1, 0.9)$

Although values of d can be obtained for $h > 1$ we know from the function $F(d,h)$ that as $h > 1$, d goes out of bounds and it can no longer be relied upon. A plot of the function $F(d,h)$ is shown in Figure 6.7. For this reason a two part algorithm is used to determine d. For $h > 1$ we use the polynomial approximation given in equ.(6.34) and for $h \leq 1$ we use equ.(6.33) to find d.

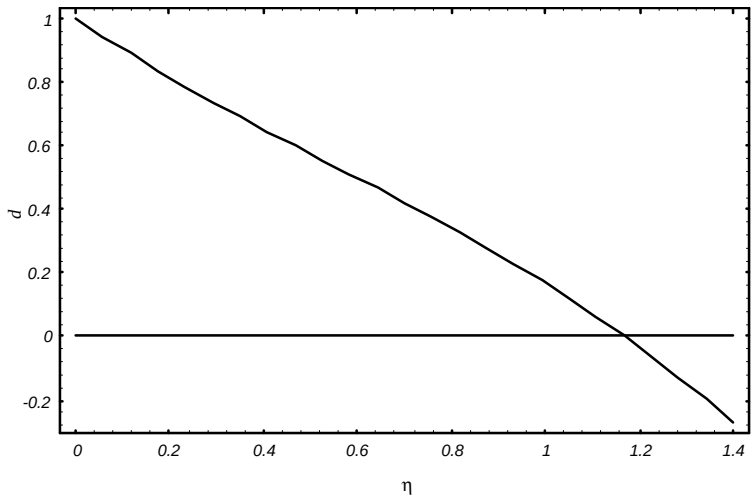

Figure 6.7 Curve fit of function $d = \mathrm{F}(h)$

We can see from this that the approximation technique can only be used with certainty in the range $0 < h \leq 1$. For this reason the experiments were re-run with d in a reduced range (0.25, 0.9) corresponding to the permissible range for h. These results are shown in Figure 6.8 and Figure 6.9. From these plots we can see that by restricting the range over which d varies the accuracy of the prediction for the IEA improves. This is since less of the peak of the invariant density is included in the desired λ value. In the best case the error is around ±5%. Moreover the range of utilisation values covered still remains within the experimental objectives outlined earlier, and this can be seen from Figure 6.10.

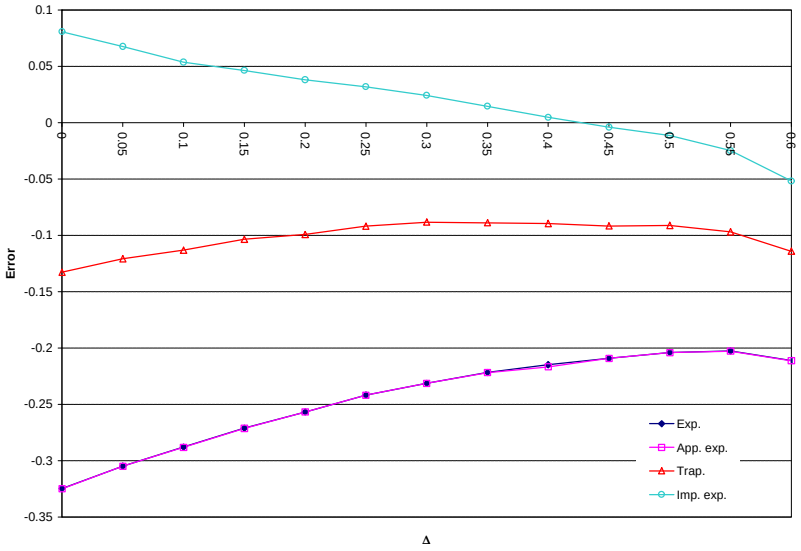

Figure 6.8 Comparison for various λ prediction methods for a two map scenario: Map_1 d value fixed at 0.25. Map_2 variable d $\hat{I}(0.25, 0.9)$

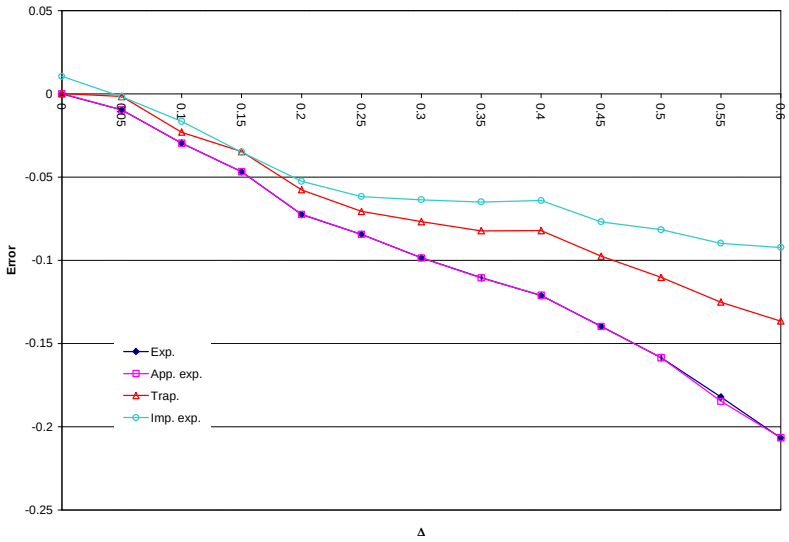

Figure 6.9 Comparison for various λ prediction methods for a two map scenario: Map_1 d value fixed at 0.9. Map_2 variable d $\hat{I}(0.25, 0.9)$

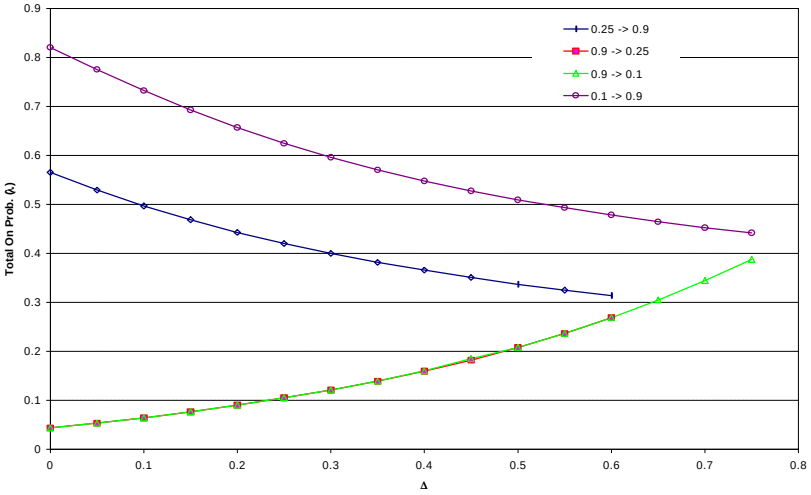

Figure 6.10 The range of λ values against Δ used in the prediction experiments

6.5 Sensitivity of the IEA Method

All the estimation methods presented for d were derived for the case $m = 2$. Since these estimation methods rely on the estimation of two points, $r(d)$ and $r(1)$, of the invariant density (which is dependent on m) then the methods for estimating d presented in this chapter should perform reasonably well for $m \in [1,2]$. Of the methods investigated, the IEA method performed the best. However we note that this method is sensitive to values of $r(1)$; this is because $r(1) << 1$ and is in the denominator. In order for the method to function properly $r(1)$ must have a stable value. In this subsection we explore the use of regression techniques to stabilise the effects of $r(1)$ on d and we check to see if the prediction method is still applicable for $\frac{3}{2} \leq m < 2$.

The instability in the IEA method can be seen in Figure 6.11 and Figure 6.12 where the %error in d and l_{NL} are shown respectively for different values of m and e. The most striking feature in these figures is that the variability in the result increases as e decreases. This can be explained by the behaviour of the map. As $e \to 0$ more of the orbits time will be spent in the restriction region (near the origin of Figure 6.1). This then affects the estimate of the invariant density at $r(1)$ for an experiment of the same run length, i.e. for the same number of iterations with a decreasing value of e less events will be recorded for the event $r(1)$. Hence the estimate of d will be adversely affected.

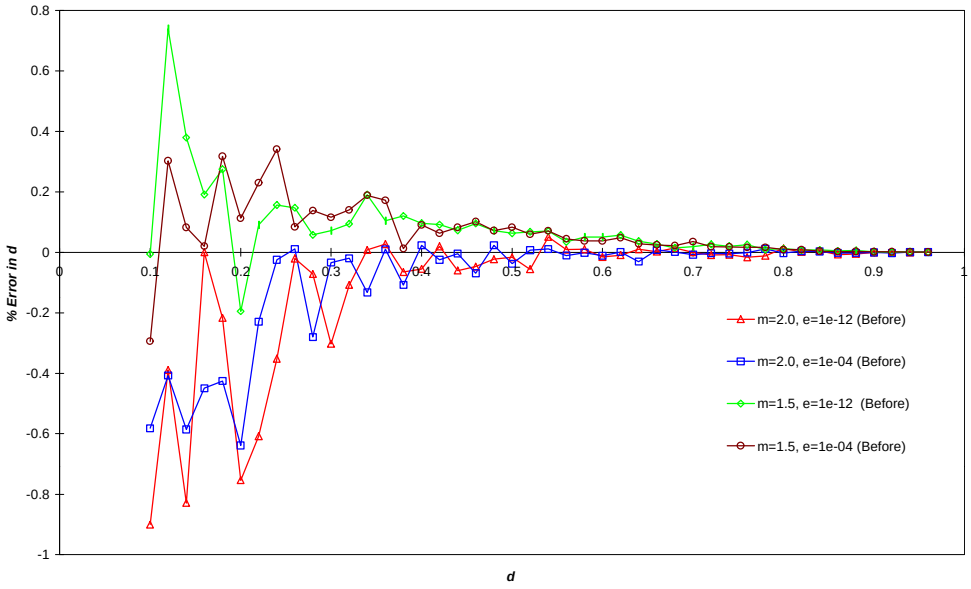

Figure 6.11 Plot of d vs *% Error in* d before corrections for the case $e = 1*10^{-4}, 1*10^{-12}$

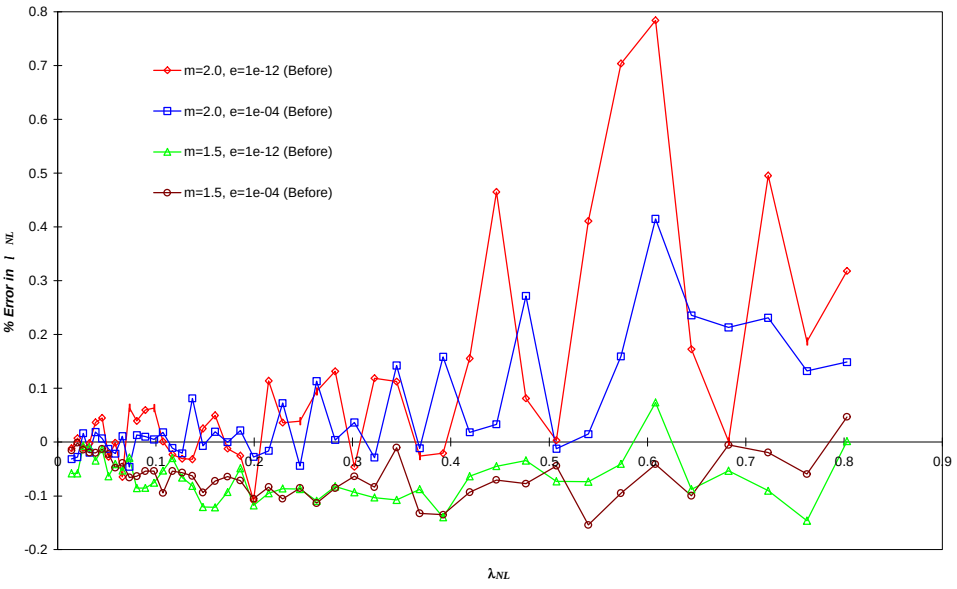

Figure 6.12 Plot of l_{NL} vs *% Error in* l_{NL} before corrections for the case $e = 1*10^{-4}, 1*10^{-12}$

Applying standard least mean square regression techniques to l_{NL} and d can reduce this problem. This is achieved in the following manner:

- We know that l_{NL} is a smoothed estimate of the cumulative density in the *ON* region of the map, i.e. all the individual effects witnessed in the individual bins of the invariant density are averaged out, and hence smoothed.
- A regressed curve fit can be obtained for various values of d against l_{NL} (see Figure 6.13).
- A regressed curve fit can be obtained for various values of $r(1)$ against d (see Figure 6.14).
- Then for a required value of l_{NL} we can obtain d. From d we can obtain $r(1)$. Knowing l_{NL} and $r(1)$ we can estimate d via the preferred method.

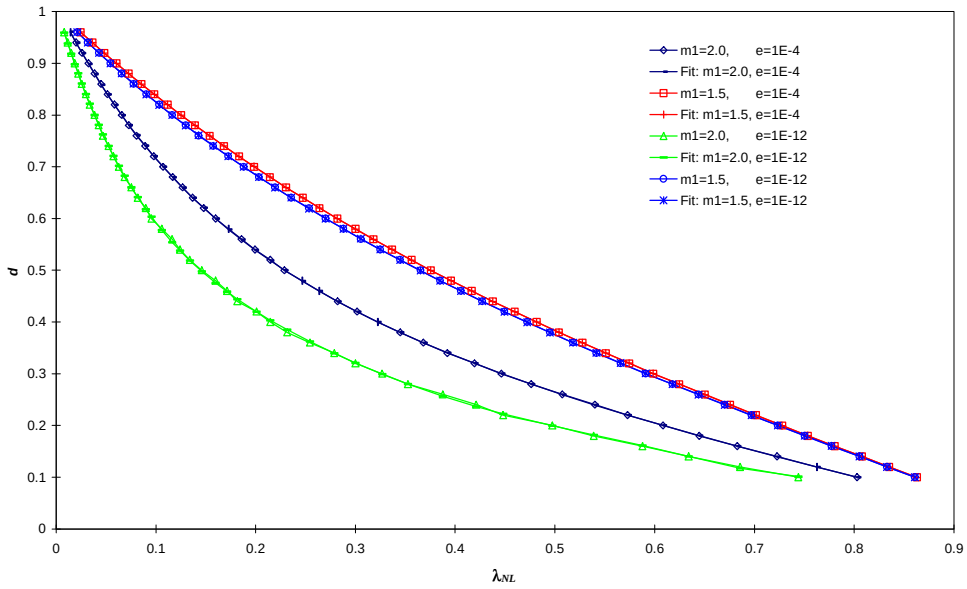

Figure 6.13 Plot of l_{NL} vs d for the cases $e = 1*10^{-4}, 1*10^{-12}$

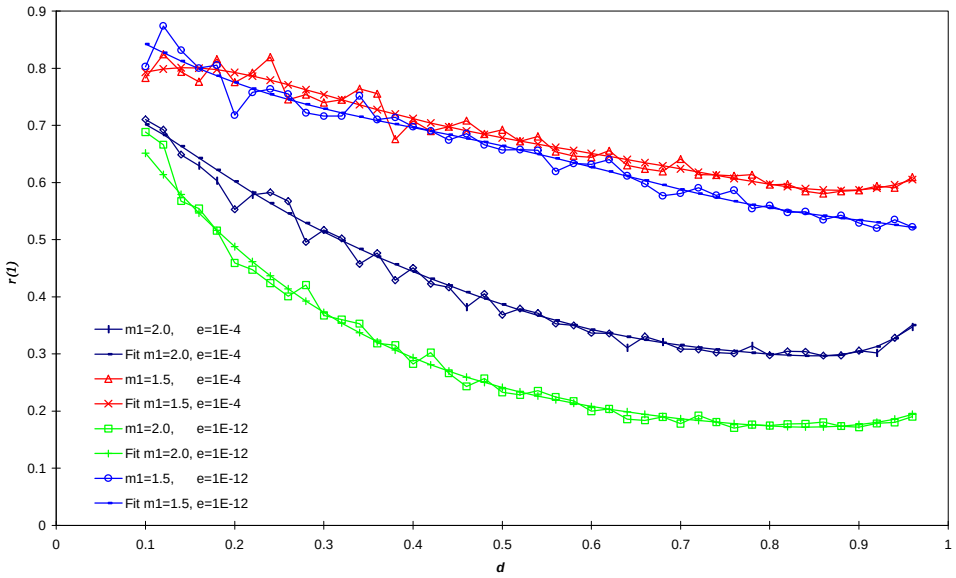

Figure 6.14 Plot of d vs $r(1)$ for the cases $e = 1*10^{-4}, 1*10^{-12}$

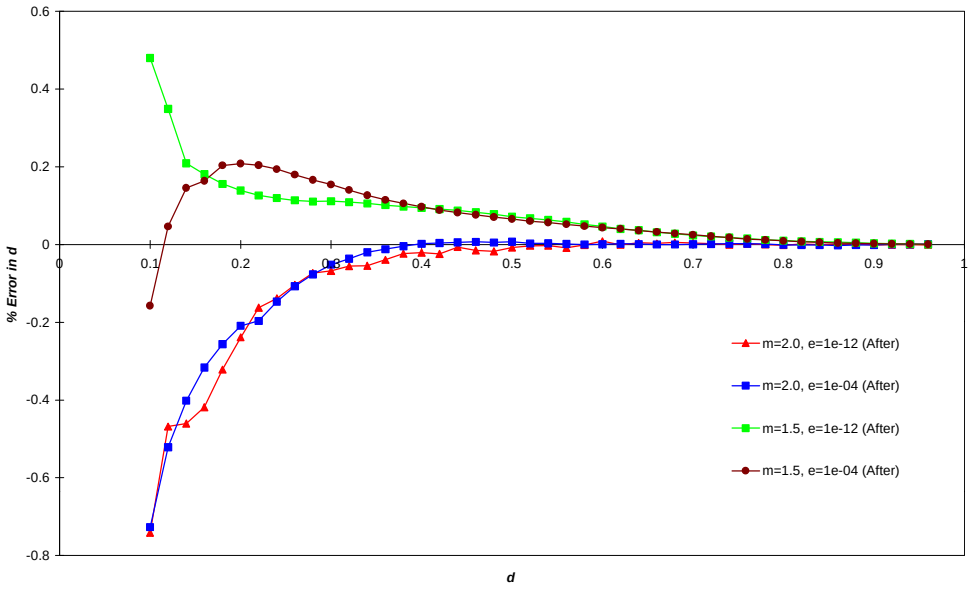

Figure 6.15 Plot of *d* vs *% Error in d* after corrections for the case $e = 1*10^{-4}, 1*10^{-12}$

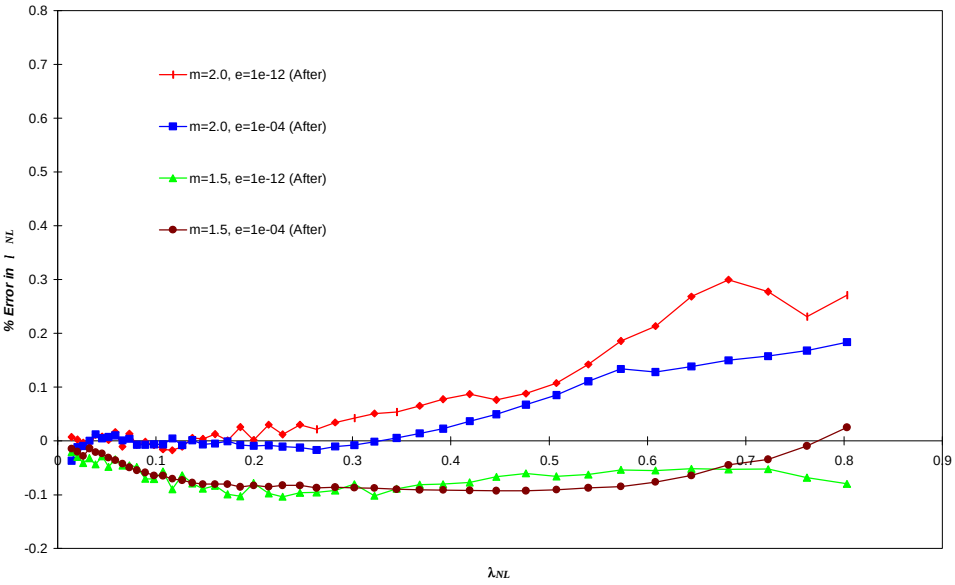

Figure 6.16 Plot of l_{NL} vs *% Error in* l_{NL} after corrections for the case $e = 1*10^{-4}, 1*10^{-12}$

A natural improvement for this method would be to have an accurate analytic approximation of l_{NL} for a given set of map parameters. This is an area recommended for further work. We can see by comparing Figure 6.11 with Figure 6.15 that this method improves the reliability of the estimated value of *d*. The curves in Figure 6.15 are much smoother than those of Figure 6.11 and, more importantly, the method gives reasonable approximations for *d* (error of 10% or less) from around *d*=0.3. The same improvement in l_{NL} can be seen when comparing Figure 6.12 with Figure 6.16. The improvement in stability allows l_{NL} loads of up to 50% with around 10% error. The other feature that we note is that as *m* decreases the error remains fairly constant over the whole interval of *d* regardless of the value of *e* used. The reason for this is that as $m \rightarrow 1$ the map behaviour tends towards that of a Bernoulli shift map in which all

the regions of the map are visited with equal probability and hence $r(1)$ will be visited with (relatively) more frequency and its variability will be reduced.

6.6 Queueing Behaviour of the IEA Method

The single equivalent map's queueing behaviour in terms of cell loss probability (CLP) has been compared against the original individual queueing effects of the independent sources feeding an M/D/1 queue. As a point of reference the Schormans *et al* M/D/1 [SCHO96] approximation is included in the result. The approximation for the M/D/1 loss probability for cell scale queueing is

$$P(k > K) = (1 - l)\left(e^{l} - 1 - \frac{2l}{3}\right)^{k}, \tag{6.40}$$

where l is the arrival rate and k is the queue state. To obtain the queueing statistics, the queues were run for a total of 30 experiments, each experiment consisting of $30*10^6$ cell slots (map iterations). Two scenarios of the single equivalent map at the extremes of behaviour were compared:

- The modelling of an aggregate total of N=5, 50 sources with $e = 1*10^{-4}$. The individual sources were chosen such that the mean d value was known (d=0.6) *a priori*. The service rate was then set in order to obtain a system load of 0.8 (the individual overall system loads determined to be 0.795 (N=5) and 0.888 (N=50)). The results of this comparison are shown in Figure 6.17.
- The modelling of an aggregate total of N=5, 44 sources with $\varepsilon = 1*10^{-12}$. The mean value of d for the aggregate sources was chosen to be 0.6 a priori. The results for this scenario are given in Figure 6.18.

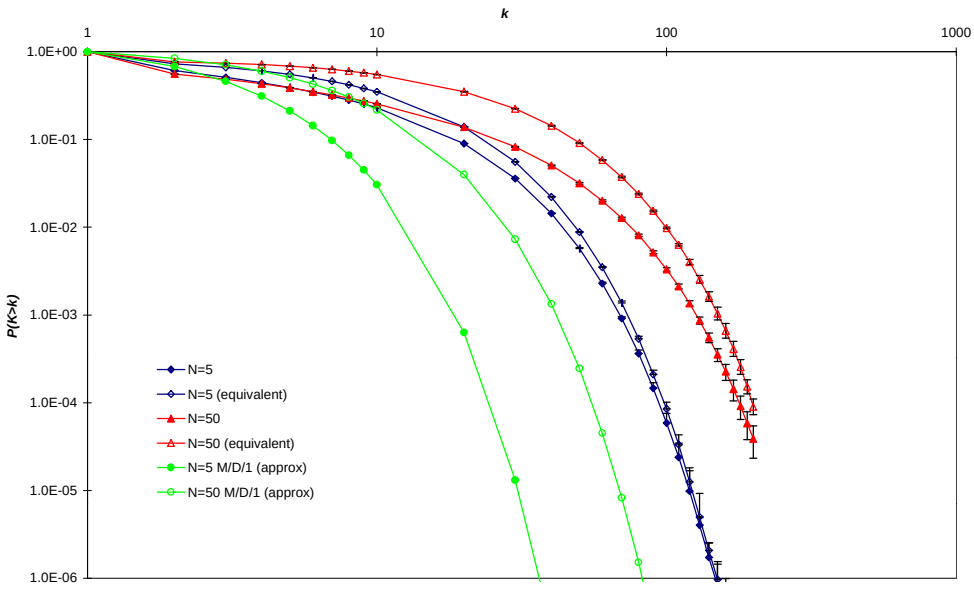

Figure 6.17 CLP with its 95% confidence interval for single intermittency map and equivalent map for the case N=5 and N=50 Sources. Load for N=5 - 0. 0.795. Load for N=50 0.888. Map $e = 1*10^{-4}$. Mean d= 0.6 equivalent d =0.599862, m_1=2.0 m_2 =1

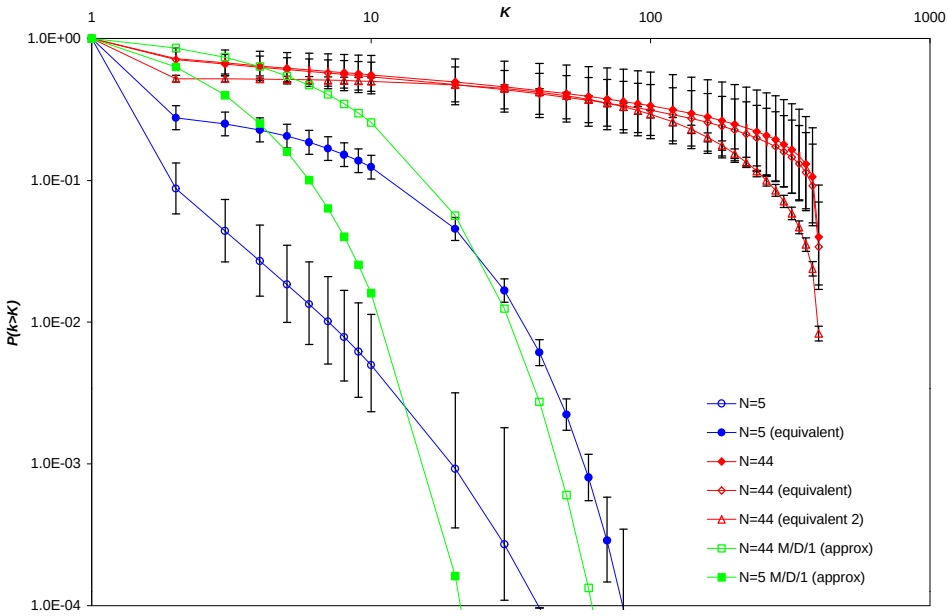

Figure 6.18 CLP with its 95% confidence interval for single intermittency map and equivalent map: for the case N=5 and 44 sources. Load 0. 76 (N=5) and 0.9 (N=44). Map parameters $e = 1*10^{-12}$. Mean d= 0.6 equivalent d=0.599862 (N=5) and 0.602584 (N=44), m_1=2.0 m_2 =1

We can see from Figure 6.17 that the single equivalent map and the original traffic queue behaviour converges in the tail. This result leads to the following comment on the sampling method used with regard to the equivalent map orbit and the total number of iterations in each experiment.

- In Figure 6.17 the results were obtained by modelling the N number of sources iterated a total of K iterations. This leads to the single equivalent map being iterated a total of KN times with the equivalent map being reseeded at the end of the KN iterations. This means that only one sample orbit was taken for each experiment and the sample length of the orbit was KN iterations.
- However there is an alternative sampling strategy. We can run the experiment with the single equivalent map being given a total of N seeds and each seed point is iterated K times i.e. N sample orbits of length K were taken from the orbit of the single equivalent map.

A comparison of the two strategies is shown in Figure 6.18. The results of the first sample strategy are shown as N=44 (equivalent 2) in Figure 6.18. and that of the second as N=44 (equivalent) in Figure 6.18. What we can see from this that taking more samples (of shorter length) from the orbit improves the convergence of the model.

A final observation that we can draw from these results is that the sojourn time is reflected in the queueing results. After a region affected by the sojourn lengths of the map (the true LRD affected region of the queue statistics) the queue decay will again become cell scale in nature.

This can be reasoned in the following way: the burst scale queueing is due to the excess arrivals building up in the queue, the arrivals in this region will have burst scale queueing decays. After the mean burst length has been accommodated, any subsequent arrival will be due to individual arrivals entering the queue after the burst. These arrivals will have cell scale decays.

6.7 Summary

In this chapter we have developed an aggregate map model that is iterated *N* times to represent the *N* aggregate sources. This model preserves the invariant density of the aggregate traffic stream and as such retains the bursty nature of the original traffic. This development represents an improvement over the Pruthi "one step" aggregation method reported in [PRU95a, 95b]. The aggregated map model uses the IEA parameterisation method for determining the aggregate map *d* value. We have seen that the IEA method for determining *d* works well with an error span of +15% to -5 %. The stability in the IEA method can be improved by the use of regression techniques. We have also shown that the method works reasonably well for range $m \in [1,2]$; this is because the invariant density takes into account the effects of *m*. The range of *m* covers the LRD behaviour of the map and because the choice of *ON* or *OFF* portion of the map is purely arbitrary then the LRD behaviour can be chosen to be in either state. The single equivalent map derived has queueing behaviour that shows convergence in the tail when compared to the queueing behaviour generated by the original traffic stream. There are limits to the IEA method which limit the arrival rate of the map when fixed to one time resolution. However if traffic is being modelled in scenarios where the individual traffic source arrival rates, λ, are in a spread which is relatively tight[30], then the IEA technique will predict parameters for an equivalent single map to within ±5% of the actual λ_{NL} figure.

[30] An example of such a tight spread would be where the maximum span ($\lambda_{max} - \lambda_{min}$) = 0.48 and studies [FOW91, SHO80] show this to be a reasonable figure.

7 A Fast Source Aggregation Model

Pruthi [PRU95a] produced two aggregate traffic models: a "one-step" and an "*N* one-step". The "step" refers to the number of stages that are used in determining the number of packets being emitted at each iteration of the map. One-step means that there is a single step in obtaining the number of emissions. *N* one-step means that the single step is repeated *N* times to determine the number of emissions.

The "one-step" model of Pruthi produced *Lk* packets at each iteration if in the *ON* portion of the map, where *L* is the number of sources being aggregated and *k* is the packet length. This approach gave inaccurate queueing behaviour. In the "*N* one-step" model of Pruthi there are *N* individual IID. maps, each map emitting *k* packets in length. At the end of the iteration cycle there will have been *kj* packets emitted with probability *P*(*j*) where *j* is taken from the binomial distribution $P(j) = {}^{N}C_{j}\, l^{j}(1-l)^{N-j}$ and λ is the total average *ON* probability. This approach sacrifices speed for greater accuracy.

The single equivalent map of the previous chapter iterates an equivalent map *N* times, to produce any number from 0 up to *N* in a time slot. This produces better queueing behaviour, much closer to that obtained by multiplexing *N* individual maps. However, the invariant density parameterisation method is limited to a single intermittency map.

The new "Bulk Property" map introduced in this chapter is a "two-step" approach which uses the iterate value in the *ON* portion of the map to index into a specific batch size varying between 1 and *N*. This approach can be applied to either single or double intermittency maps. This chapter describes how the Bulk Property batch size is parameterised from the invariant density, presents results for the *H* dependence on the map parameters, and illustrates the speed up over other methods.

7.1 Current Thinking: ON/OFF Sources

Current thinking on the traffic modelling of high speed networks suggests that individual sources can be described by *ON/OFF* models in which the model is either emitting traffic at a maximum rate or is completely idle [WILL94]. An example of this interpretation is the use of the indicator variable of the intermittency map family. There have been recent studies on intermittency maps which employ this *ON/OFF* interpretation by emitting a fixed number of packets/full ATM cells on each iteration in the *ON* state of the map [PRU95b]. Queueing behaviour produced from the use of such intermittency maps produces extremes of behaviour, i.e. either;

- when the tail of the invariant density corresponds to the *ON* portion of the map the state probabilities decay exponentially; or

- when the head of the invariant density corresponds to the *ON* portion of the map we obtain a power law decay of the state probabilities.

In both instances it has been shown that when intermittency maps with the *ON* behaviour described above are aggregated then the queueing behaviour (state occupancy) tends to a stretched exponential [PRU95b]. The speed with which the stretched exponential distribution is achieved depends on the *ON* interpretation being used. Faster convergence is achieved when using the second interpretation.

If we are to model an aggregated traffic stream as an *ON/OFF* model then the maximum rate behaviour described above would be incorrect. For example, suppose there are 50 sources feeding into a single link. If an *ON/OFF* model of the type described above were used to model this behaviour, then there would be 50 sources emitting full ATM cells when *ON*, and no full cells being emitted when *OFF*. Clearly this cannot be true. There must be a compromise between the extremes which properly describes the actual behaviour. To resolve this problem we propose a continuum of behaviour between the extremes.

7.2 The Bulk Property

In the single intermittency maps used so far we have interpreted the *ON* region as the partition of the interval corresponding to m_2=1 with e_2=0. What we have noticed is that a single emission event (cell or packet) is coincident with a near uniform invariant density in the emitting region, i.e. the uniform invariant density gives rise to a single emission. We could then view a non-uniform invariant density as corresponding to the probability of more than one emission. This notion leads to an interesting interpretation of the intermittency map family and its invariant density that can be used to

- depict network behaviour; and
- provide a method of giving a continuum of *ON* behaviour.

Furthermore it was noticed in experiments conducted with the equivalent single intermittency map of Chapter 6 that it yielded the same mean emission rate whether the map was emitting from its linear region (m_2=1) or from its non-linear region (m_1>1). The map remains the same, suggesting that the map is depicting the underlying behaviour of the network traffic and not just the sources.

What these observations lead to is the “Bulk Property” interpretation for the intermittency map family. This interpretation comes from what the “bulk” of the network traffic is doing, i.e. are we examining a single source or are we examining a collection of sources. The bulk property is used to relate the map dynamics at each iteration to the aggregate number of emitting sources. The manner in which this is achieved is by aligning the mode of the invariant density to the mode emission rate of the sources. In this way the invariant density

not only characterises the sojourn times for *ON* and *OFF* states but also gives a measure of the "*ON*-ness" of the aggregated sources. This idea is illustrated in Figure 7.19.

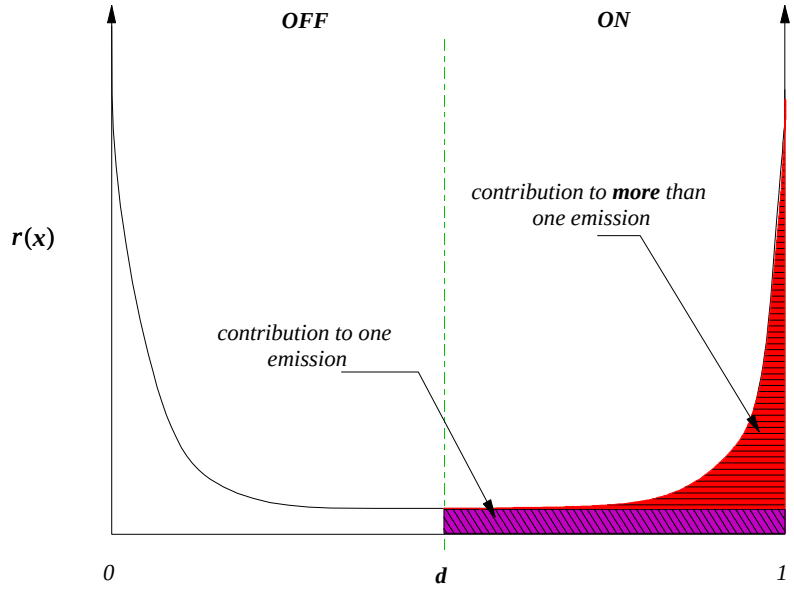

Figure 7.1 Aggregate map: Contribution of $r(x)$ to the number of emissions

To simplify this approach it is easier to work with the cumulative distribution functions of the map and the probability of the number of sources being *ON* arranged in back-to-back fashion. This is illustrated in Figure 7.20.

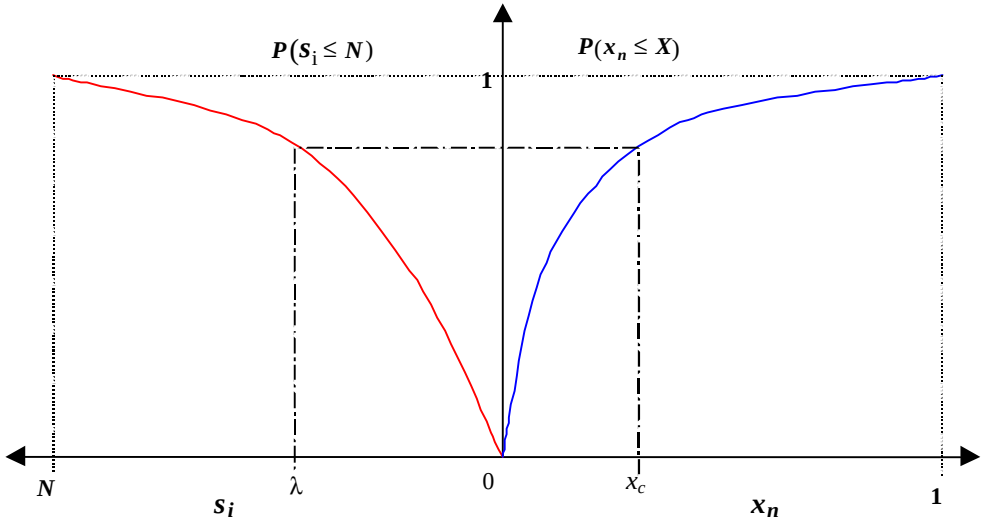

Figure 7.2 Single equivalent map cumulate invariant density $P(X \geq x_n)$ iterate x_c mapped to the corresponding source emission cumulate probability $P(N \geq s_i)$ l of emitting sources

We can see from Figure 7.20 that in terms of the cumulate probabilities a map iterate value, x_n, has a corresponding number of sources, s_i, that are *ON* associated with it.

7.3 The Bulk Property Map

The Bulk Property map is a *single map* source that is the *aggregated equivalent* of the *N* map sources. This aggregated map is iterated once and the value of the iterate yields information in the *ON*-ness of the emitting sources. The benefit of this approach is two-fold:

- it permits a speed-up in the simulation of on-line traffic generation; and
- it acts as an enabling step towards the application of chaotic control to networks via Coupled Map Lattice techniques (see Chapter 9).

To develop the Bulk Property map, we start by considering aggregation as taking place in the aggregation plane, where the types of aggregation are given by the axes Sources and Time (see Figure 7.21).

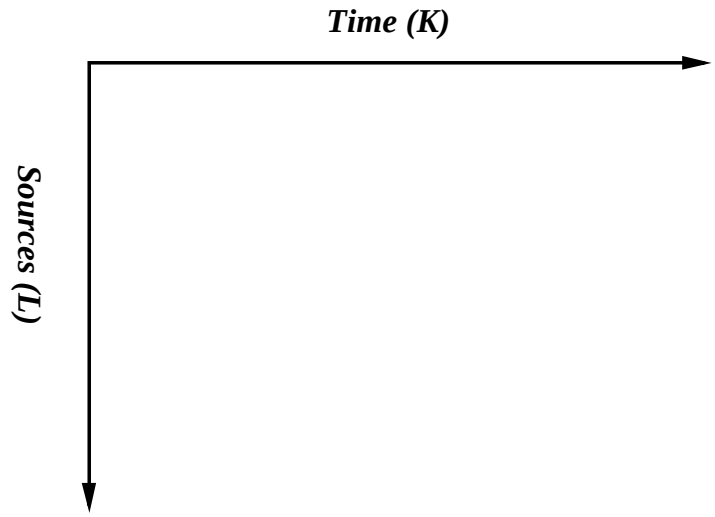

Figure 7.3. Conceptual view of the aggregation plane

In the Bulk Property map the evolution of the map's iterates leads to aggregation in time, and the superposition of the source probabilities on top of the iterates leads to aggregation in sources.

7.3.1 Aggregation in Time

The aggregation in time comes from considering the probability of a single emission from a total of L sources. If we assume that the sources are IID and that members of the map family can model the traffic streams then the probability of emission from an individual source is

$$P_j(1) = l_j = \int_{x \in \{ON\}} r_j(x)dx ,$$

(7.41)

and the probability of no emission is given by $P_j(0) = 1-\lambda_j$. If we now consider two cell streams that are aggregated under the following rule: *zeros are only preserved in the output stream when there are two coincident zeros in the input streams*, then the probability of an emission (at least one) in the output stream would be

$$P_{out}^{(2)}(1) = P_1(1)P_2(1) + P_1(1)P_2(0) + P_1(0)P_2(1) .$$

(7.42)

The probability of no emission in the output stream would then be

$$P_{out}^{(2)}(0) = P_1(0)P_2(0) .$$

(7.43)

It is easier to write the probability of an emission in the output stream in terms of the probability of no emission

$$\begin{aligned} P_{out}^{(2)}(1) &= 1 - P_1(0)P_2(0) \\ &= 1 - P_{out}(0) \end{aligned}.$$

(7.44)

This can be readily extended to *L* sources

$$P_{out}^{(L)} = 1 - \prod_{j=1}^{L} P_j(0) \equiv \Lambda_L ,$$

(7.45)

where Λ_L is the aggregate map *ON* probability i.e. $\Lambda_L = \int_{d_{agg}}^{1} r_{agg}(x)dx$ and d_{agg} and r_{agg} are the aggregate map discriminate and invariant densities respectively.

7.3.2 Aggregation in Sources

The aggregation of sources arises out of considering the number of emitting sources at each iteration of the map. If we suppose that the sources are IID then the mean arrival rate at each iteration interval is

$$\bar{l}_T = \frac{1}{L}\sum_{i=1}^{L} l_i .$$

(7.46)

The probability of *M* emitting sources in a total number of *L* sources is given by

$$f_L(M) = \binom{L}{M} \bar{l}_T{}^M \left(1 - \bar{l}_T\right)^{L-M} .$$

(7.47)

The cumulate probability distribution that we superimpose onto the invariant density is then

$$\Phi_L(m \le M) = \sum_{j=0}^{M} f_L(j).$$

(7.48)

The mean of such a distribution is $L\bar{l}_T$.

7.3.3 Effect of Aggregation on the Mean and Variance of the Bulk Property Map

The results of equ.(7.45) and equ.(7.47) are used to show the effect of aggregation on the cumulate traffic produced by the Bulk Property map. The Bulk Property map aggregates both in time, *K*, and sources *L*. We have formulations for the variance and the mean of a single source aggregated in time (see Chapter 5). Recall that we had defined the cumulate traffic from a single source as

$$z_K \equiv \sum_{i=1}^{K} y_i \ ,$$

(7.49)

where y_i is the value of the indicator variable at discrete time i. Also recall that the expectation and variance of the cumulate traffic from a single source was given by

$$E(z_K) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{K} y_j \right) = KE(y) = K \int_{x\in\{ON\}} r(x)dx = Kl$$

(7.50)

$$\mathrm{var}(z_K) = Kl(1-Kl) + 2l \sum_{i=1}^{K-1} (K-i)C(i)$$

(7.51)

where $C(i)$ is the correlation function and $r(x)$ is the map's invariant density. What we require is a similar development of the aggregated map. There are two reasons for this:

- we would like to have a method which can check on the Hurst performance of the aggregated map; and
- we would like to develop a method of measuring the Hurst parameter of the aggregated traffic on-line.

In order to do this we reassess the function of the indicator variable. We define the indicator variable for the aggregated map as

$$g_n(x_n) = \begin{cases} 0 & x_n \in \{OFF\} \\ 1 & x_n \in \{ON\} \end{cases} .$$

(7.52)

For the aggregated map the indicator variable acts as a step function for the probability of the number of sources being on i.e. if $g_n = 1$ then we will have the probability $f_L(M)$ of having M sources on at discrete time n, while if $g_n = 0$ then there are no sources on. This is illustrated in Figure 7.22.

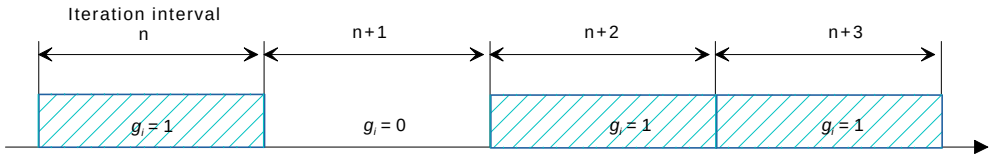

Figure 7.4 Interpretation of g_i as a step function: If g_i =1 then M_i is the number of sources on. If g_i =0 then M_i=0

We can see that there are two things to consider:

- the probability of the aggregated map being *ON* in the time *K*. If we define $V_K \equiv \sum_{i=1}^{K} g_i$ then the required probability is given by $E(V_K) = \lim_{L\to\infty} \frac{1}{L} \sum_{n=1}^{L} \left(\sum_{i=1}^{K} g_i \right) = K\Lambda_L$ [†];
- the mean number of emissions at each iteration from the aggregate of *L* sources, which is $E(f_L(M)) = L\bar{l}_T$ by virtue of the source distribution being Binomial.

In light of the behaviour of the indicator variable we can therefore say that in the aggregation of *L* sources over time *K*, the average number of emissions, S_K, will be

$$E(S_K) = E(V_K)E(f_L(M)) = K\Lambda_L L\bar{l}_T .$$

(7.53)

Here we note that there is an alternative formulation for the above result that may be useful later. We can define the expectation of S_K in the following way

$$E(S_K) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{i=1}^{K} g_i M_i \right),$$

(7.54)

where g_i is the indicator variable and M_i is the number of *ON* sources at discrete time *i*. Since the g_i is independent of the M_i, i.e. the M_i and the g_i can be generated separately without affecting the end result, then we can say

$$E(S_K) = \sum_{k=1}^{K} \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_n M_n \right)$$

(7.55)

and because of the independence we can write

$$\begin{aligned} E(S_K) &= \sum_{j=1}^{K} \left(\left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_n \right) \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} M_n \right) \right), \\ &= K\Lambda_L L\bar{l}_T \end{aligned}$$

(7.56)

which is the same result as equ.(7.53).

We are now in a position to obtain the variance of the aggregate process var(S_K). For this we use the definition of the variance

[†] Where $\Delta = \int_{x\in\{ON\}} r_{agg}(x)dx$ and $r_{agg}(x)$ is the invariant density of the aggregated map.

$$\begin{aligned}\mathrm{var}(S_K) &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{j=1}^{K} g_i M_i - K\Lambda_L L\, l_T\right)^2 \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{j=1}^{K} g_i M_i \sum_{j=1}^{K} g_i M_i - 2K\Lambda_L L\, l_T \sum_{j=1}^{K} g_i M_i - (K\Lambda_L L\, l_T)^2\right)\end{aligned} \tag{7.57}$$

$$\mathrm{var}(S_K) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{j=1}^{K}(g_i M_i)^2 + 2\sum_{i=1}^{K}\sum_{j>i}^{K}(g_i g_j M_i M_j) - 2K\Lambda_L L\bar{l}_T \sum_{j=1}^{K} g_i M_i - (K\Lambda_L L\bar{l}_T)^2\right) \tag{7.58}$$

By applying similar techniques to those used in Chapter 5 for the variance of a single map model we can obtain the variance of the aggregate map model. After simplification in which $C(i)$ is the correlation term for the map and $C(i) = C(n = j - i)$, we can similarly define the correlation terms for the *M* sources as $\mathbb{C}(i) = \mathbb{C}(n = j - i)$. Therefore we can write

$$\mathrm{var}(S_K) = K\Lambda_L L\bar{l}_T\left[f_L(0) + L\bar{l}_T(1 - K\Lambda_L)\right] + 2L\Lambda_L\bar{l}_T\sum_{i=1}^{K-1}(K-i)C(i)\mathbb{C}(i). \tag{7.59}$$

Equation(7.59) tells us that the correlation term is dependent on the correlation of the map which possesses LRD and on the correlation of the number of sources emitting, which is SRD. The net result is that the LRD dominates since it has a slower decay and therefore we can say that the underlying dynamics of the Bulk Property map will permeate through to the overlying dynamics of the indicator variable (the actual number of sources emitting). This means that no matter what distribution we impose on top of the underlying dynamics the LRD and hence the *H* will still be there.

7.3.4 Consideration of the Invariant Density

The Bulk Property map associates the invariant density to the probability of a number of emissions. In order to perform this association we introduce the following concept.

Equivalent uniform invariant density. The equivalent uniform invariant density, $r_1(x)$, is the contribution of the invariant density in the *ON* state which corresponds to only one source being active. In the intermittency map this can be thought of as the invariant density at the point *d* being constant over the whole of the *ON* interval i.e. $r_1(x) = r(d) \quad \forall x \in \{ON\}$. Under this definition we can say that any $r(x) > r_1(x)$ contributes to the *ON*-ness of other emitting sources. This concept is illustrated in Figure 7.19.

Definition - Modified cumulate invariant density. This is defined as

$$\mathrm{P}_M\left(x_n \le X\right) \equiv \left| \frac{\int_d^X r(x)dx}{\int_{x_n \in \{ON\}} r(x)dx} \right| \tag{7.60}$$

where $r(x)$ is the invariant density of the aggregated map and $X \in \{ON\}$.

We use the concepts of equivalent uniform and modified cumulate invariant densities in the following way. The distribution of $r(x) > r_1(x)$ is not uniform. This implies that the aggregated map has favoured intervals that the orbits of the map prefer. This preference in the orbits amounts to a preference in the number of emitting sources. If x_n is the iterate at discrete time *n* then this corresponds to some number of emitting sources M_n, i.e. $x_n \Rightarrow M_n$.

Theorem 1: *Existence of* $E(x_n)$.

$\exists x_n$, $x_n \in \{ON\}$ such that $E(x_n)$ has a finite mean i.e. $E\left(x_n\right) = \int_{x_n \in \{ON\}} x_n r(x)dx < \infty$.

Proof. The normality of the aggregate map is assured by exclusion of the fixed points from the interval in which the map exists, i.e.

$$\forall x_n \in (0,1) \quad \mathrm{P}(1) = \int_0^1 r(x)dx < \infty. \tag{7.61}$$

Then any subset of $x \in (0,1)$ will also have a summable finite invariant density, i.e.

$$\forall x_n \in \{ON\} \subset (0,1) \;\; \exists x_n \text{ such that } E\left(x_n\right) = \int_{x_n \in \{ON\}} x_n r(x)dx < \infty.$$

Corollary: If the invariant density is normalisable for all subsets of $x \in (0,1)$, then we can say that there exits a modified cumulate invariant density that is finite and bounded.

Remark*:* We can now state the following correspondence in probability between the modified cumulate invariant density and the probability distribution of the number of emitting sources as a direct consequence of the equivalent uniform invariant density. There exits a modified cumulate invariant density $\mathrm{P}_M\left(x_n \le X\right)$ such that

$$\mathrm{P}_M\left(x_n \le X\right) \Leftrightarrow \Phi_L\left(m_n \le M\right). \tag{7.62}$$

Consequently we can say

$$\frac{d\mathrm{P}(x_n \le X)}{dx} \approx \left[\frac{\Phi_L(m_n + \Delta m \le M) - \Phi_L(m_n \le M)}{\Delta m}\right], \tag{7.63}$$

where $\approx [.]$ signifies the closest discrete probability to the modified invariant density.

Theorem 2: *Existence of mappings of the aggregate map iterate to the number of emitting sources.*

$$\begin{array}{ll} \exists g & \forall x_n \in (d, E(x_n)) \overset{g}{\mapsto} f_L(k) \in (1, L\bar{I}_T) \\ \exists h & \forall x_n \in [E(x_n), 1] \overset{h}{\mapsto} f_L(k) \in [L\bar{I}_T, L] \end{array},$$

where $L\bar{I}_T$ is the expected number of sources that are on.

Proof: The proof for these mappings arises out of the proof for the existence of $E(x_n)$ and knowledge of the invariant density. That is to say if x_n has a $r(x_n)$ then

$\forall x_n \in \{ON\} \exists r(x_n) \Leftrightarrow P_M(x_n \le X)$,

and since by correspondence in probability

$$\frac{\Phi_L(m_n + \Delta m < M) - \Phi_L(m_n < M)}{\Delta m} = f_L(M),$$

where $f_L(M)$ is the probability density of M sources emitting out of a total of L sources, then $\forall x_n \in \{ON\} \exists x_n \Leftrightarrow m_n$ and this implication is achieved by the mapping g and h, since $E(x_n)$ and $L\bar{I}_T$ are finite and are contained within their respective intervals. Therefore $E(x_n)$ and $L\bar{I}_T$ act as natural partition values for those intervals in which the mapping exists.

Both g and h are 1:1 mappings that are continuous and single valued functions. This is illustrated in Figure 7.23. The functions g and h are hard to obtain. However the proof outlined above indicates a method for circumventing the mapping which is outlined below:

1. Obtain x_n by iterating the aggregate map.
2. Obtain $\mathrm{P}_M(x_n \le X)$.
3. Obtain equivalent $\Phi_L(m_n \le M)$.
4. Find solutions for $[\Phi_L(k_n + \Delta k \le K) - \Phi_L(k_n \le K)]/\Delta k$.
5. If more than one solution: perform Bernoulli trial for the selection of the solution.
6. Return m_n.

Steps 2 -5 can be simplified and sped up through the use of look-up tables, since $\mathrm{P}_M(x_n \le X)$, $\Phi_L(k_n \le K)$ and $[\Phi_L(k_n + \Delta k \le K) - \Phi_L(k_n \le K)]/\Delta k$ can be tabulated.

The advantage of this method is that even though the original modified invariant density of the aggregate map may not be multi-valued it will still give an indication of intervals that yield highly probable emissions from those which yield low probability emissions. For example

with $e = 0$ the modified invariant density is single valued. However the mapping to the source probability density yields multi-valued results.

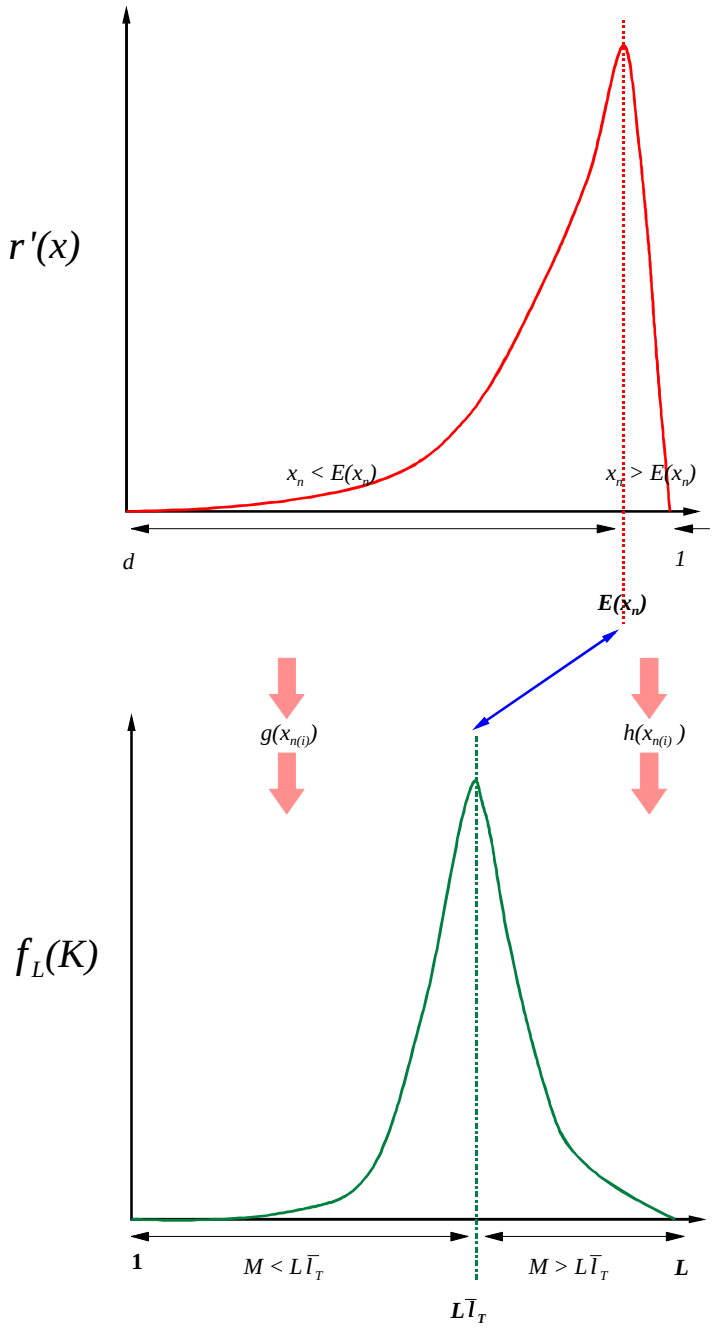

Figure 7.5 Mapping of invariant density to source *pdf.* for the double intermittency map

Strictly speaking the aggregate map maps a specific iterate to an emission sequence. To accomplish this coupling between an iteration and a number of sources emitting would be difficult. A way around this problem of coupling is to view the iteration in terms of its cumulative invariant density and to relate this to the cumulate density of source emissions as outlined above. Under this interpretation we view the cumulate invariant density as the probability of all orbits (x_n) having been visited up to X_n. This then equates to a probability of

emitting all sources *k* up to source *K*. This is illustrated in Figure 7.24. In this way there is connectivity between the iterate value and the number of sources emitting.

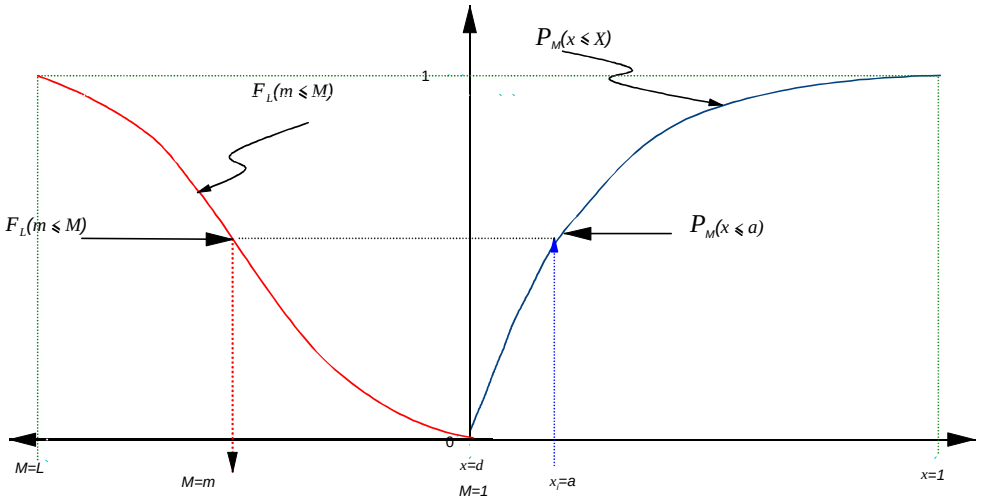

Figure 7.6 Equivalence relations for the cumulative distributes of the sources and invariant density

Observation: *"d-walk-back" and emergent self-similar and LRD behaviour*

The *ON* probability of the aggregated map is set by

$$l_T = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} l_n \ .$$

(7.64)

If we now suppose that the number of sources increases, this would imply that the network utilisation increases. This is usually the case in real networks. The implication of this in the aggregated map is that as $l_T \to 1$ (increase in utilisation) $d \to 0$. The movement of *d* towards 0 is termed "*d-walk-back*". The effect of *d-walk-back* on the traffic generated by the aggregated map is an apparent loss of independence of the sources. The justification of this is in the following argument: a self-similar source is characterised as much by its silent sojourn periods as by its active sojourn periods[31]. Therefore as the aggregation level increases the probability of a silent period being retained in the output aggregate traffic diminishes. This is the equivalent of a loss in information. The net effect is that the independent nature of the sources begins to appear dependent and hence correlations appear in the aggregated traffic which lead to self-similar and LRD behaviour.

Remark: Aggregation Order

Here we note that the aggregation performed by the Bulk Property map is aggregating in the correct order since the source terms are included within the summations to *K*, i.e. we are aggregating in the limits described by Taqqu *et al*.[TAQ97].

[31] For example, consider a self-similar source modelled by a chaotic map that has the dominant *m* in the *OFF* state.

7.4 The Source Aggregation of Erramilli and Pruthi

Willinger *et al* [WILL97] used an aggregated process $W^{*}_{M,b}(j)$ generated by:

- aggregating 0/1 sequence output of *M ON-OFF* sources; and then,
- aggregating the resulting sequence in time over non-overlapping blocks of size *b*.

If each source has a heavy tail sojourn time in *ON* and *OFF* (see Chapter 3 for more details) and the probability that given a period of continuous zeros (0) and ones (1) = ½ the aggregated process $G_{H,s}(t), t \geq 0$ is FGN, then the following limit in distribution applies

$$\lim_{b\to\infty}{}_{d}\lim_{M\to\infty}{}_{d}\frac{\sqrt{M}}{b^{H}}\left(W^{*}_{M,b}(j)-\frac{bM}{2}\right)=G_{H,s}(j).$$

(7.65)

The implication of this is that the map family will tend to FGN when aggregated over *M* sources and then *b* blocks of output data.

7.4.1 Pruthi's Aggregation Model

Pruthi [PRU95a] describes mathematically the output behaviour of his "*N* one-step" aggregation model in the following way. He considered the aggregate behaviour of a number of sources $\{i{:}1..N\}$ over a discreet time index *n* = 1,2,…

$$x^{(i)}_{n+1}=\begin{cases} f_1\left(x^{(i)}_n\right) & 0\leq x^{(i)}_n < d \\ f_2\left(x^{(i)}_n\right) & d\leq x^{(i)}_n \leq 1 \end{cases}$$

(7.66)

with the packet cell generation for the source *i* being given by

$$y^{(i)}_n=\begin{cases} 0 & 0\leq x^{(i)}_n < d \\ 1 & d\leq x^{(i)}_n \leq 1 \end{cases}.$$

(7.67)

The number of packets generated $k, k\in[1,\infty)$ is the number of packets generated by each source at every iteration when in the *ON* state (fixed length). The output of the traffic stream is then

$$Y_n = k\sum_{i=1}^{N} y^{(i)}_n .$$

(7.68)

For the characterisation of the aggregated traffic we assume IID. therefore

$$l = E\left[y^{(i)}_n\right]=\int_{ON} r(x)dx .$$

(7.69)

The mean traffic rate is then

$$r = E[Y_n] = Nkl \ .$$

(7.70)

Because the *N* sources are independent the probability of *j* data sources is given by

$$P(j) = \binom{N}{j} l^j (1-l)^{N-j},$$

(7.71)

for $j = 0,1,2,\cdots,N$. Therefore at each iteration a batch of *kj* packets is generated with the probability given above. The variance of the output process is then given by

$$\text{var}(Y_n) = E\left[\left(Y_n - r\right)^2\right] = Nk^2 l(1-l).$$

(7.72)

The peakedness, more formally the index of dispersion for counts (IDC), is given by

$$a = \frac{\text{var}\left(Y_n\right)}{E\left(Y_n\right)} = k(1-l).$$

(7.73)

7.4.2 Scaling the fluctuations

For the double intermittency map the problem exists of finding the mean. To get around this Pruthi constructed a zero mean process from the aggregate of the individual maps,

$$Y_n^z = Y_n - \frac{kN}{2}$$

(7.74)

i.e. half the mean value from the cumulative total. A fluctuation term is then constructed $\frac{2}{k}\sqrt{\frac{\text{var}(Y_n)}{N}}$ to which the mean, $E\left[Y_n\right]$ is added. Therefore the aggregated output of the maps can then be written as

$$Y_n^t = \frac{2}{k}\sqrt{\frac{\text{var}(Y_n)}{N}} Y_n^z + E\left[Y_n\right]$$
$$= \frac{2}{k}\sqrt{\frac{\text{var}(Y_n)}{N}} Yn - \sqrt{N\,\text{var}(Y_n)} + E\left[Y_n\right].$$

(7.75)

7.5 Bulk-property Map H-profiles

We have run numerical experiments on the bulk-property map to examine it's dependence of *H* on its parameter values. The experiments were conducted in a manner identical to that outlined in Chapter 5. The source emission distribution for the bulk-property map was chosen to be 100 IID. sources with a mean emission rate of 0.014190 cells/second. This corresponds to 100 bursty Ethernet sources at 2.2 Mb/s feeding into a 155Mb ATM backbone link.

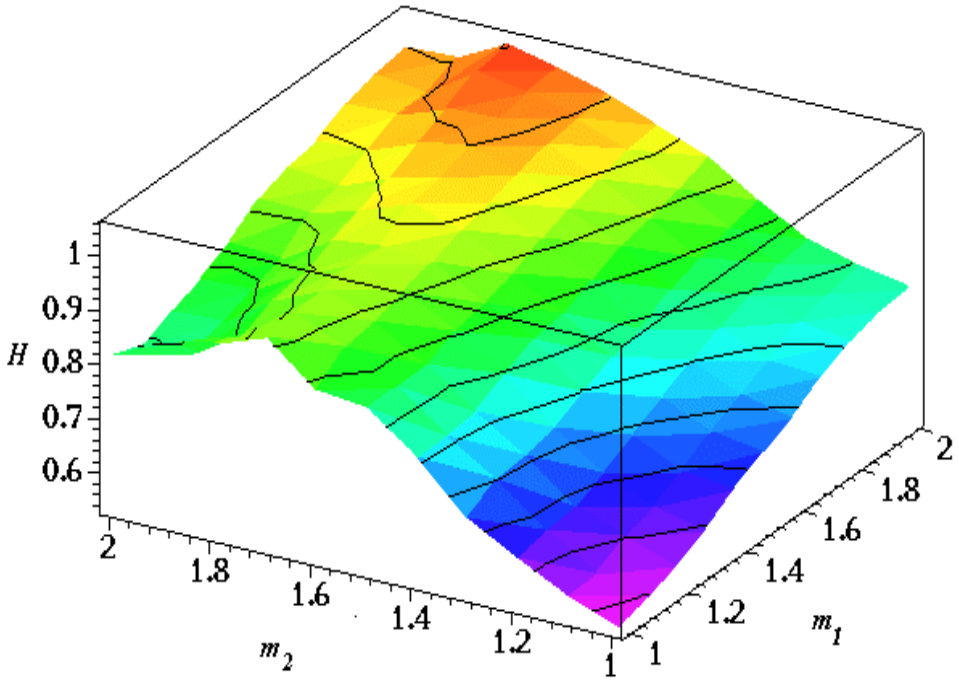

Figure 7.7 *H* profile for variations in m_1 and m_2 of the Bulk Property map: $e_1 = e_2 = 0$ and $d = 0.5$. Modelling 100 Sources, Mean Arrival Rate per source 0.014190 (2.2Mb Ethernet link on a 155Mb ATM link)

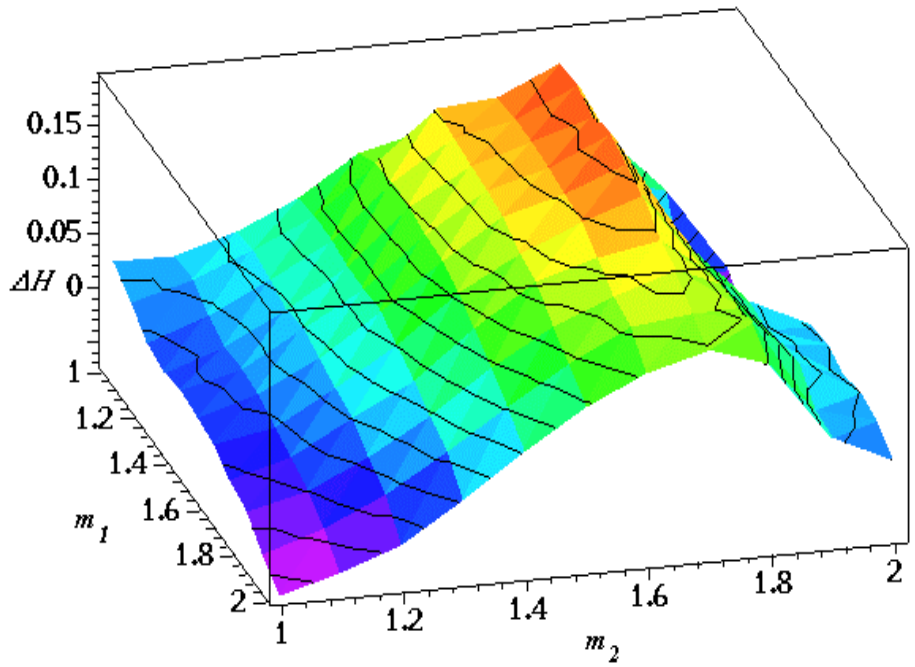

Figure 7.8. Error in *H* using the Empirical fit for variations in m_1 and m_2 of the Bulk Property map: $e_1 = e_2 = 0$ and $d = 0.5$. Modelling 100 Sources, Mean Arrival Rate per source 0.014190 (2.2Mb Ethernet link on a 155Mb ATM link)

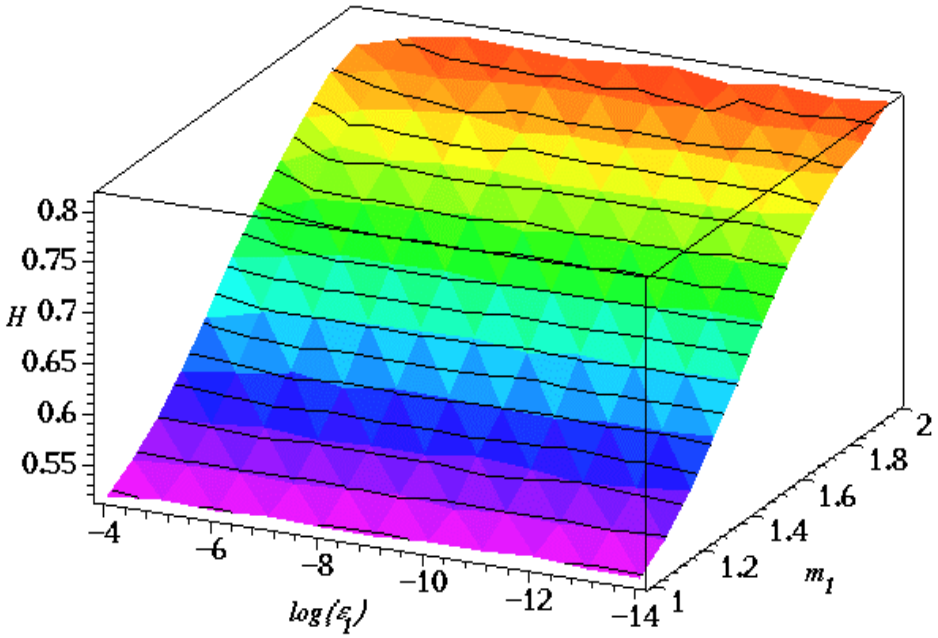

Figure 7.9. *H* profile for variations in m_1 and e_1 of the Bulk Property map: m_2 =1, e_2 = 0 and d = 0.5, m_2 is in the *ON* State, Modelling 100 Sources, Mean Arrival Rate per source 0.014190 (2.2Mb Ethernet link on a 155Mb ATM link)

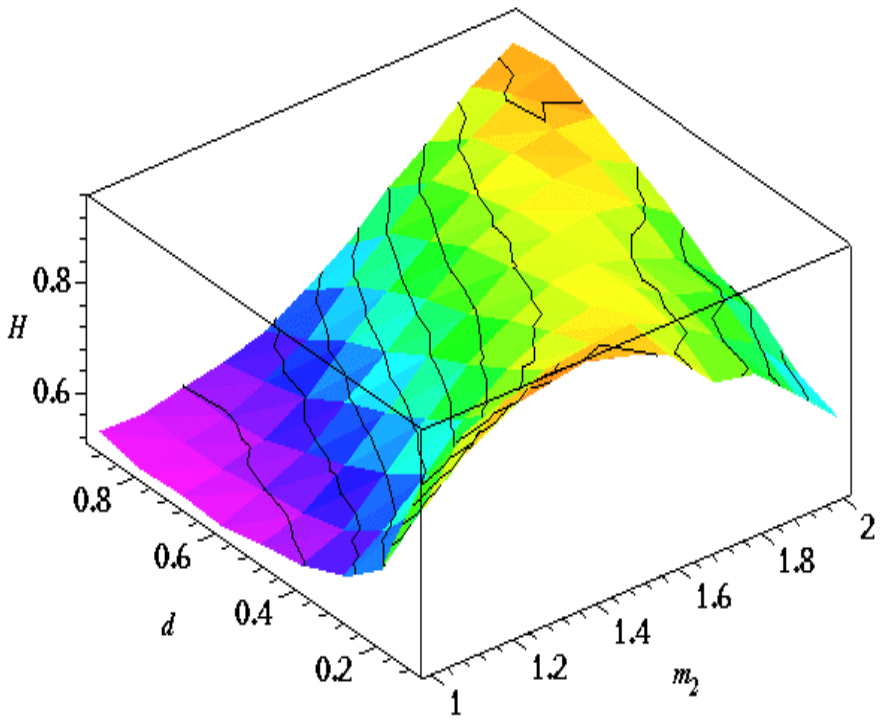

Figure 7.10. *H* profile for variations in m_2 and e_2 of the Bulk Property map: m_1 =1, e_1 = 0 and d = 0.5, m_2 is in the *ON* State, Modelling 100 Sources, Mean Arrival Rate per source 0.014190 (2.2Mb Ethernet link on a 155Mb ATM link)

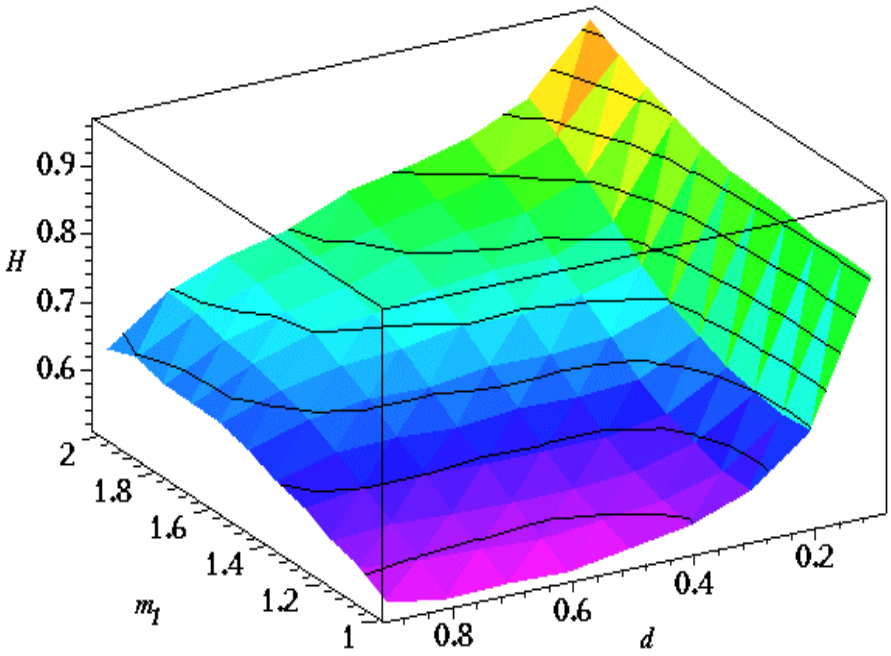

Figure 7.11. *H* profile for variations in m_1 and *d* of the Bulk Property map: m_2 =1 and e_1 =e_2 = 0, m_2 is in the *ON* State, Modelling 100 Sources, Mean Arrival Rate per source 0.014190 (2.2Mb Ethernet link on a 155Mb ATM link)

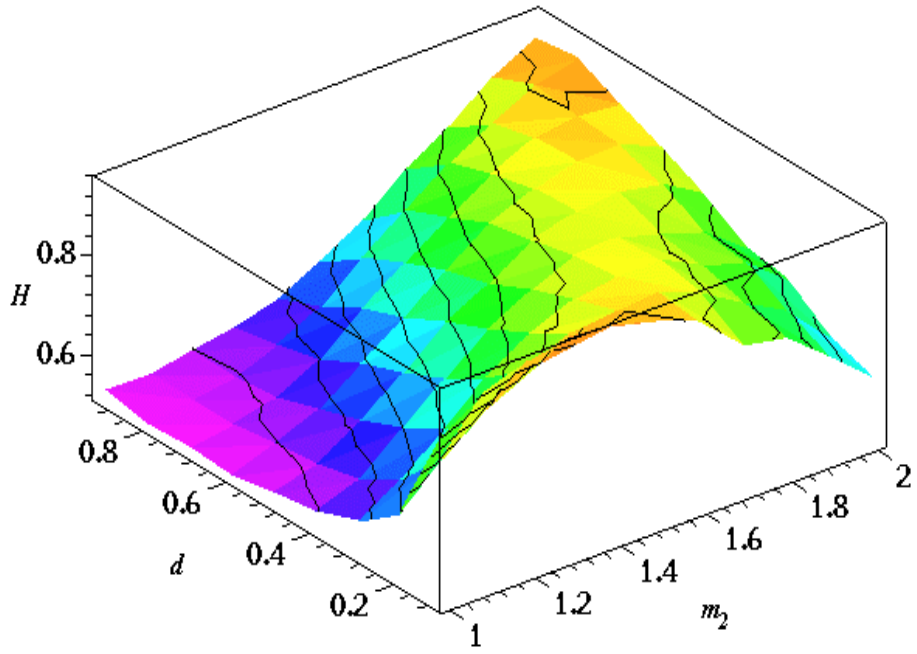

Figure 7.12. *H* profile for variations in m_2 and *d* of the Bulk Property map: m_1 =1 and e_1 =e_2 = 0, m_2 is in the *ON* State, Modelling 100 Sources, Mean Arrival Rate per source 0.014190 (2.2Mb Ethernet link on a 155Mb ATM link)

The results of Figure 7.25-Figure 7.30 show that the bulk-property map has its *H* profile governed by the underlying dynamics. Generally the bulk-property map follows the same

profiles obtained for the intermittency maps of Chapter 5. However it is noticeable that there are artefacts present in these results which are due to the algorithm that imposes the source distribution onto the underlying dynamics. This is most noticeably evidenced in the results where variations of parameters (m_2 and e_2) in the emitting half of the map are made (see Figure 7.28 and Figure 7.30).

7.6 Bulk Property Map - Speed up

The Bulk Property map has been compared for speed of sample generation against the results reported in Chen *et al* [CHEH96]. Although the methods reported in [CHEH96] are for off-line generation of self-similar samples we can compare the time it takes to generate a given number of samples. Under the test conditions stipulated by Chen – Sun SPARC 5 running at 70 MHz requiring the generation of 16384 samples – the best performing algorithm was the Maximum Likelihood Estimator (MLE). This generated the required number of samples in 0.6 seconds. The bulk properties map generated the same number of samples in 0.03 seconds.

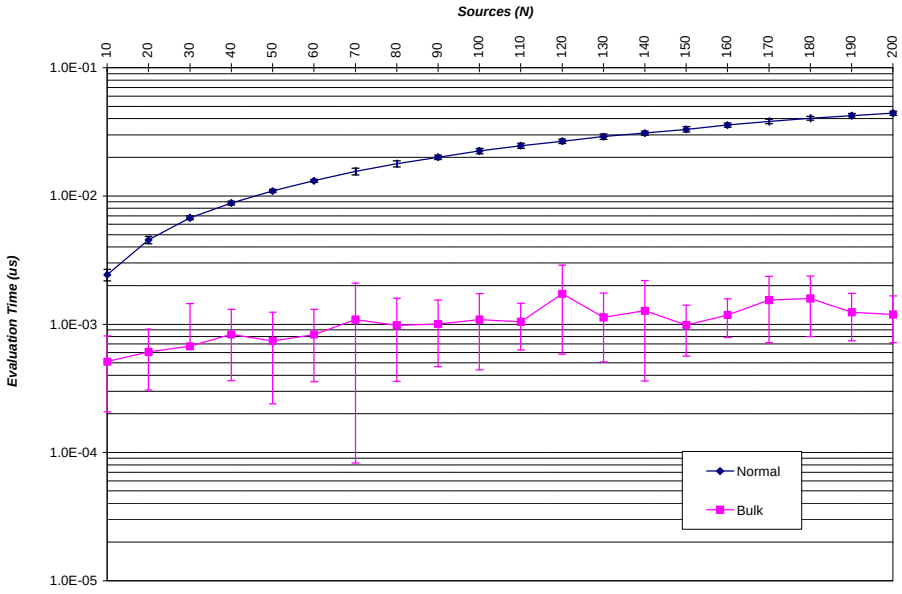

Figure 7.13 Comparative evaluation times (in μs) for the Normal – Pruthi sample generation (standard map) and the bulk properties map for increasing number of sources, N, together with their 95%tile error bars

The second comparison performed was against the sample generation method proposed by Pruthi and outlined earlier. In this comparison the number of aggregated sources modelled are increased form 10 to 200. The average time to generate a sample is then calculated along with the 95% confidence interval around the generation time. The results of this comparison are shown in Figure 7.31. What we can see from this result is that for the Pruthi method, the sample generation time increases as the number of modelled aggregated sources increases, whereas the Bulk Property map has a generation time which is more or less constant as the number of modelled aggregated sources increases. However it should be pointed out that the variability of the generation time is much higher for the Bulk Property map than for the Pruthi method. This variability is due to the search algorithm employed by the map.

7.7 Summary

In this chapter we have investigated a new map – the Bulk Property map. This map models aggregate self-similar traffic on-line. We have seen that the underlying dynamics of the map parameters permeates through to the output variable of the map enabling H to be set in the same way as the double intermittency map. Moreover we have seen that imposing a source onto the underlying dynamics permits the Bulk Property map to produce sample output between one and two orders of magnitude faster than MLE and Pruthi's method. However, the Bulk Property map is still in its early stages and artefacts in the H profiles need to be explained. These artefacts could be sourced from a number of places; for example the search algorithm used to find the batch size could be imparting an effect on H since the index is derived from the value of the map iterate. Furthermore, because map is still coupled it could be suffering from similar problems at the extremes of d. Or indeed the artefacts could be combinations of the above. However, progress can be made in counteracting these artefacts since their effects are predictable, so a numerical parameterisation would be reliable. This map is nonetheless promising and should be developed further.

8 On-line Measurement of H

There are many estimators for *H*. There are those which are heuristic in nature, for example *R/S* statistics, correlograms, and there are those which are more statistically robust, such as Whittle's estimator [BER94] and more recently Abry and Veitch's wavelet based estimator [ABR98]. The more rigorous the estimator the more complex in nature it becomes. For an on-line estimator of *H* the practical requirements alter slightly from the theoretical ones. These alterations really reduce to one of a compromise between speed and accuracy. The best on-line estimator would be one that:

- can produce a reasonably stable estimate of *H* in the fastest possible time (stability in convergence);
- is reasonably stable to short term fluctuations in the arrival rate;
- gives an idea of the error around the measurement in *H*;
- in terms of hardware would require as little memory as possible and, could classify the readings as they are taken.

In this chapter we asses the Indicator Variable Variance (IVV) method developed for measuring *H* in Chapter 5 as an on-line measurement method of *H* from a single traffic source. We compare this method for accuracy against measurements taken by the Abry and Veitch's wavelet based estimator. The interest in comparing these methods for accuracy lies in the fact that the IVV estimation method is derived from a correlation structure where as the Abry and Veitch method is wavelet based. From a teletraffic point of view a correlational method may be more significant in terms of understanding the queueing behaviour produced by self-similar traffic.

8.1 Assessment of the IVV Method

The IVV method calculates *H* from the variance of a single source as the lag increases over which the variance is calculated. *H* is measured from the cumulative arrival process of the source via the asymptotic relationship, $\text{var}(Z_K) \propto K^{2H}$, for $K >> 1$ (see Chapter 3). Self-similarity in a traffic stream is embodied in the cumulative arrival process, $z_K = \sum_{i=1}^{K} y_i$, where *K* indicates the lag and y_i indicates the presence of a full cell/packet in the interval *i*. The suitability of the IVV methods as on-line indicators of *H* arises out of noting that the mean arrival rate for a given lag *K*, l_K, is the average of the cumulate arrival process, i.e. $l_K = \langle z_K \rangle / K$ and by noting that the cumulate arrivals generalise irrespective of the source used. Therefore the IVV method should be applicable to any traffic stream.

The aim of this section is to assess the IVV method against similar computational methods for stability of *H*, convergence onto a stable value of *H* and computational speed. The IVV

method was compared by calculating the cumulative variance of a single map source. In a map source this amounts to measuring the variance of the indicator variable y_n. Three methods were compared in the numerical experiments for their convergence, i.e. all the methods give rise to the same value, however their convergence behaviour is different. These methods ranged from the most simplistic to those rather more involved. The equations for the methods are

$$\mathrm{var}\left(Z_{K_y}\right) = Kl\left(1-Kl\right)+2l\sum_{i=1}^{K-1}\left(K-i\right)C\left(i\right) \qquad \text{IVV,} \tag{8.1}$$

$$\mathrm{var}\left(Z_{K_y}\right) = Kl\left(1-Kl\right)+2\sum_{i=1}^{K}\sum_{j>i}^{K}E\left(y_i y_j\right) \qquad \text{Method 1,} \tag{8.2}$$

$$\mathrm{var}\left(Z_{K_y}\right) = Kl\left(1-Kl\right)+2l\sum_{i=1}^{K}\sum_{j>i}^{K}P\left(y_j=1\middle|y_i=1\right) \qquad \text{Method 2,} \tag{8.3}$$

where λ is the mean arrival rate, *E(.)* is the expectation operator, *P(.)* is the dependent probability and *C(.)* is the correlation. For the derivation of these equations see Chapter 5. We compared the IVV method, Method 1 and Method 2 against two separate evaluational methods, Comp A and Comp B.

$$\mathrm{var}\left(Z_{K_y}\right) = E\left(Z_{K_y}\right)^2 - \left(E\left(Z_{K_y}\right)\right)^2 \qquad \text{Comp A.} \tag{8.4}$$

$$\mathrm{var}\left(Z_k\right) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{j=1}^{K}y_j - Kl\right)^2 \qquad \text{Comp B.} \tag{8.5}$$

Comp A and Comp B are there to check the validity of IVV and methods 1 and 2. This is because these methods do not involve any kind of measurement of the correlation structure possessed by the map source.

A number of numerical experiments were undertaken on IVV and methods 1 and 2. Each numerical experiment consisted of 100 sub-experiments of 80,000 full cells. For each sub-experiment the variance was calculated using a selected method. The gradient was then calculated from the variance. At the end of the experiment an overall gradient was calculated (cumulated reading over all of the sub-experiments) and the batch mean value of the gradient was calculated from the empirical distribution of the gradients of the sub-experiments. The result of this was plotted on a doubly logarithmic plot. A double intermittency map with the parameters $e_1 = e_2 = 0$, $m_1=1.6$, $m_2=1.8$ and $d=0.7$ was used throughout the experiments. The maximum lag that was examined was $K=400$.

8.2 Assessment Results

The stability of *H* for variations in arrival rate, λ are shown in Table 8.1 and Figure 8.1. Table 8.1 shows that the mean value of *H* for each of Comp A and B falls within the 90% confidence intervals for *H* for all the methods used. Moreover, Table 1 also shows that IVV performs best since it is closer to the target value of *H*=0.8.

	Mean *H*	*H*: 5%tile	*H*: 95%tile
Method 1	0.742334	0.902709	0.58196
Method 2	0.794418	0.887013	0.701822
IVV	0.794295	0.88011	0.708481
Comp A	0.842139		
Comp B	0.841648		

Table 8.1 Relative stability of *H* for variations in λ

Figure 8.1 shows that the numerical evaluation of *H* is sensitive to the estimation of λ over the 90% confidence interval for λ. Method 2 and IVV are more stable to variations in λ than Method 1.

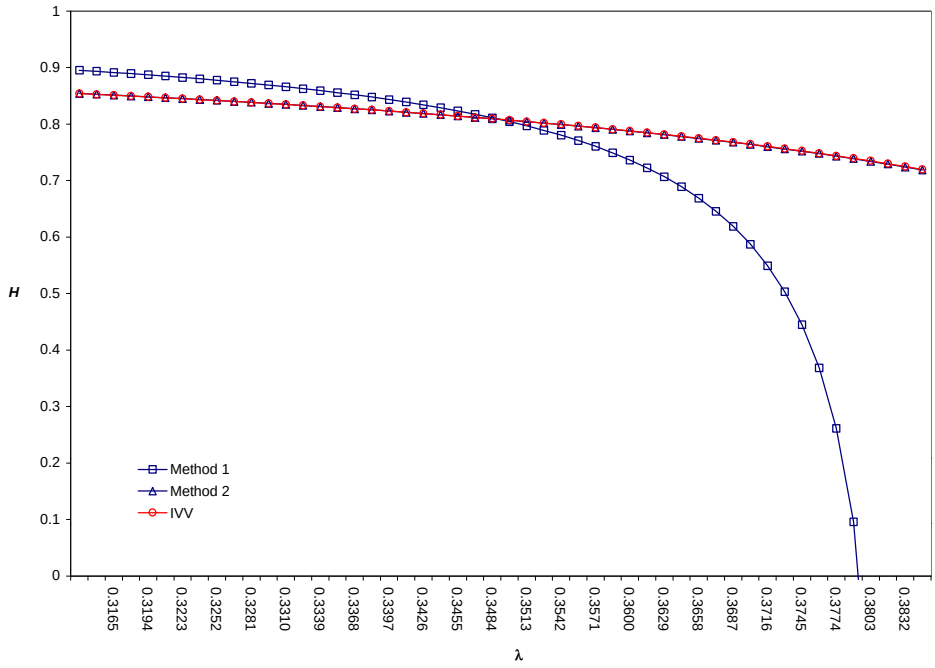

Figure 8.1 Relative stability of Hurst Parameter (*H*) for Method 1, Method 2 and IVV vs. Arrival Rate(λ).

Figure 8.2 displays the relative convergence of the measurement methods onto a stable value of *H*. The leftmost Y axis gives the measure of the convergence of the calculational methods. This is achieved by calculating the running average of the slope value after the 10^{th} experiment. As more experiments are performed, the chosen method should start to converge on a stable value for the slope (and hence *H*) this is since the law of large numbers starts to affect the result. The rightmost Y axis gives a measure of the variation in convergence. This can be viewed as the stability of the method. If the variance of the running average is converging on a fixed value then the method is stable. Its relative stability can be assessed by how quickly the variance converges on to a fixed value. What we observe from these results is that if the mean slope value converges and the variance of the mean slope also converges

(i.e. the error around the mean becomes stable) then the method is stable. With this in mind we can see that that Comp A is stable, Method 1 is stable but is converging very slowly, and Method 2 and the IVV method display the best stability since they appear to converge the fastest. Another possible way of assessing the relative performance of the methods lies in the number of missing readings (gaps in the data traces in Figure 8.2). These missing readings arise in the following manner. Since the slope is plotted on a doubly logarithmic plot then this implies that all the readings are positive (otherwise no logarithmic value can be obtained). A lack of positiveness in the reading can indicate that the reading has not yet stabilised, hence it is an indicator of stability. However why this should be so is a matter of further investigation.

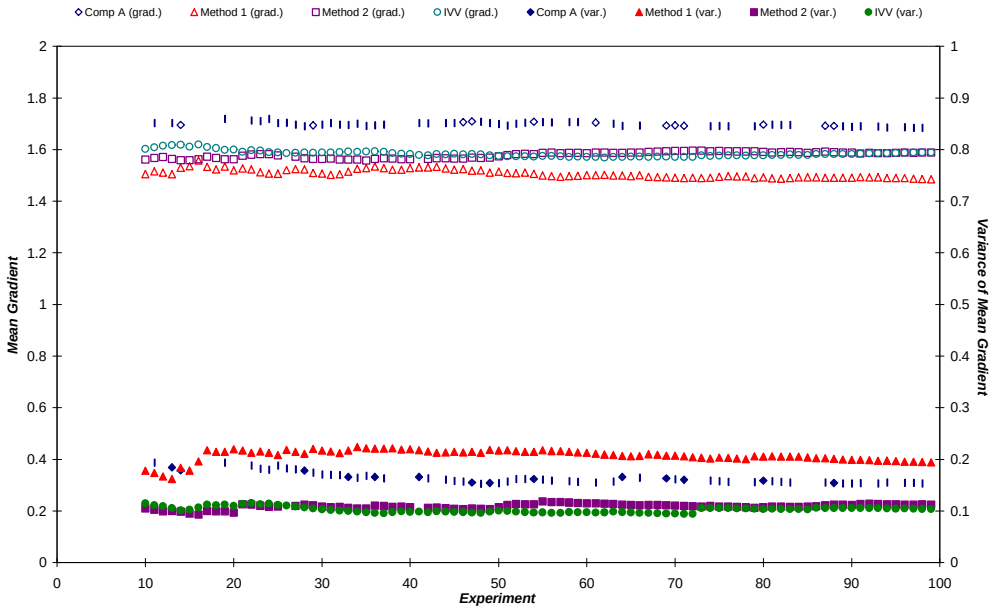

Figure 8.2 Relative convergence on a stable value of *H* for Comp A, Method 1, Method 2 and IVV

Figure 8.3 shows the speed of the IVV method, Methods 1 and Method 2. Two different processors were used, one running at 70 MHz and the other running at 143 MHz. It is clear from the figure that IVV outperforms the other evaluation methods.

Figure 8.4 and Figure 8.5 show the overall slopes for the variance of Z_k against K on doubly logarithmic plots for the numerical experiments conducted. Figure 8.4 shows the effects of the worst case influence of the arrival rate on the slopes of methods 1,2 and IVV. Comp A and B methods are included in the traces for comparison. Figure 8.5 shows the effect of the best case influence of the arrival rate on the slopes of methods 1,2 and IVV. Once again Comp A and B methods' traces are included for comparison.

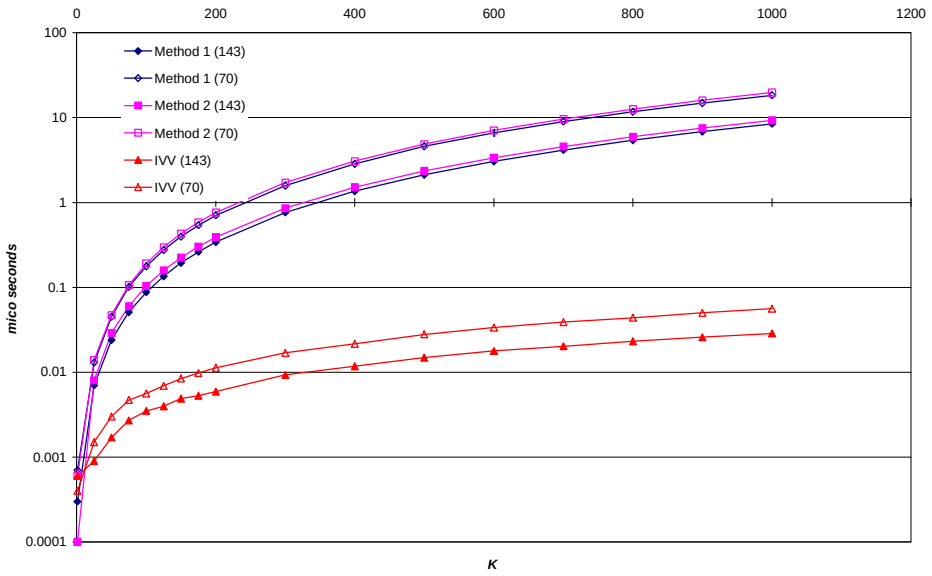

Figure 8.3 CPU system time (µS) vs. Lag (*K*). Two processor speeds shown: (143) indicates 143MHz Ultra Sparc, (70) indicates 70MHz µ Sparc 2.

What these plots show is the variation and hence the convergence of Methods 1,2 and IVV are quite good compared with Comp A and B. This is observed in the linearity (lack of fluctuations) in the traces for Methods 1,2 and IVV. Method 2 and IVV are far more stable to variations in the arrival rate since they both follow the traces of Comp A and B far more closely than Method 1.

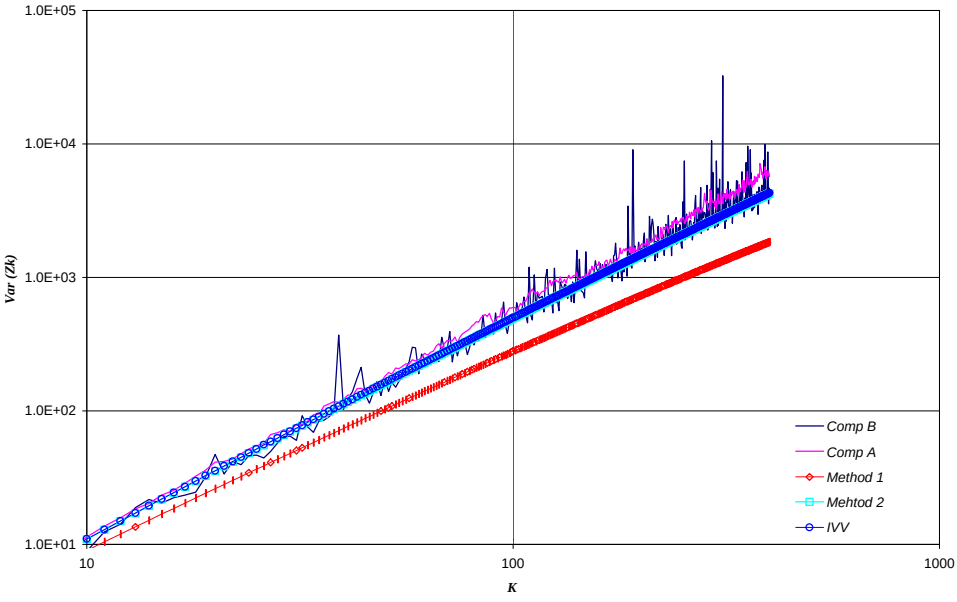

Figure 8.4 Variance Slopes: Qualitative Assessment (Worst Case)

The other notable point that the results presented in Figure 8.4 and Figure 8.5 show is that for large *K* the variance has long range dependence which exhibits itself as a variance which has both a linear dependence with *K* on a doubly logarithmic plot and a gradient bounded between 1 and 2. This equates to $H \hat{I} (½,1)$.

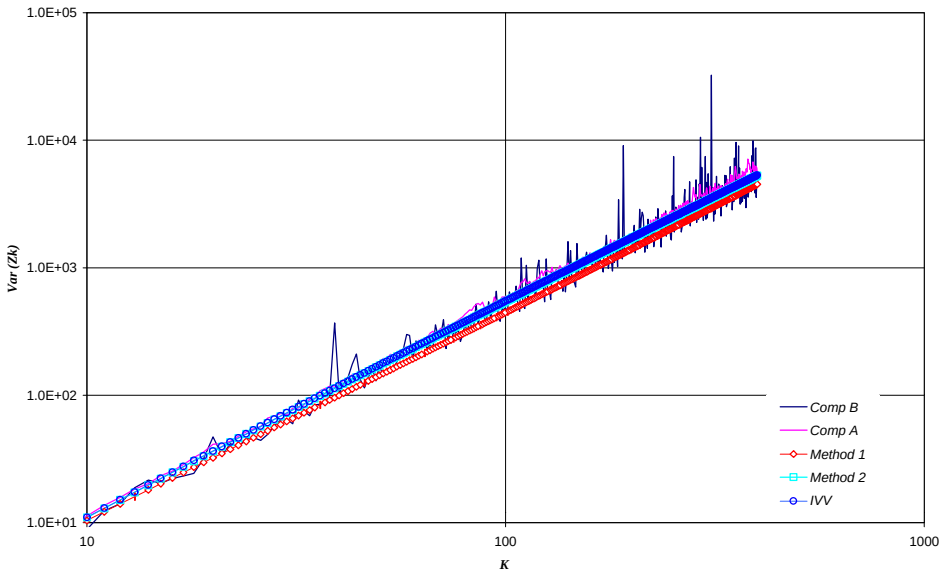

Figure 8.5 Variance Slopes: Qualitative Assessment (Best Case)

The conclusion that we can draw from the assessment of the IVV method is that it can produce relatively stable values of *H* on-line once the correlation statistics have been amassed. We can see from Figure 8.2 that the experiments required around 800000 sample points for stability. However once this number of samples had been reached an estimate of *H* could be obtained every subsequent sample. It should also be pointed out that for higher values of *H*, a greater number of samples will be required in order to obtain the stable estimates of *H*.

8.3 Accuracy Comparison of IVV and Abry-Veitch H estimation methods

The IVV and Abry-Veitch methods are very different in their approaches to obtaining an estimate of *H*. Essentially the IVV method measures the cumulate arrivals to obtain the scaling and hence *H*, whereas the Abry-Veitch method uses a window sampling technique and wavelet binary decimation in order to determine the scaling and hence *H*. In both instances the final result depends on some form of linear regression to a gradient line which determines the estimate of *H*. Notwithstanding these differences one would expect that given an identical set of experimental conditions both would render the same estimated value of *H* within some given margin for error. Two sets of experiments were conducted to see if this assumption was true. These experiments revealed interesting results that require further investigation.

In the first set of experiments a given combination of parameters m_1 and m_2 (see Table 8.1) with e_1, e_2 =0 were fed into an intermittency map. IVV, Abry-Veitch and R/S methods measured the resulting time series for *H*. The R/S method was included as an additional reference marker. The window size for the Abry-Veitch method was set to 2^{15} samples (this was following guidance obtained from results in Chapter 5). The experiment was then repeated with the value of m_1 and m_2 reversed. The results of these experiments are tabulated in Table 8.2 and displayed in Figure 8.6.

The second set of experiments consisted of repeating the experiments conducted in Chapter 5 required to produce the results seen in Figure 5.7 of that chapter. In these experiments a double intermittency map with the following parameters: m_2=1, e_2=0, and d=0.5 with m_1 and e_1 variable was used to generate a time series. The time series produced by this map was measured for H using the IVV method. The results of these experiments are shown in Figure 8.7.

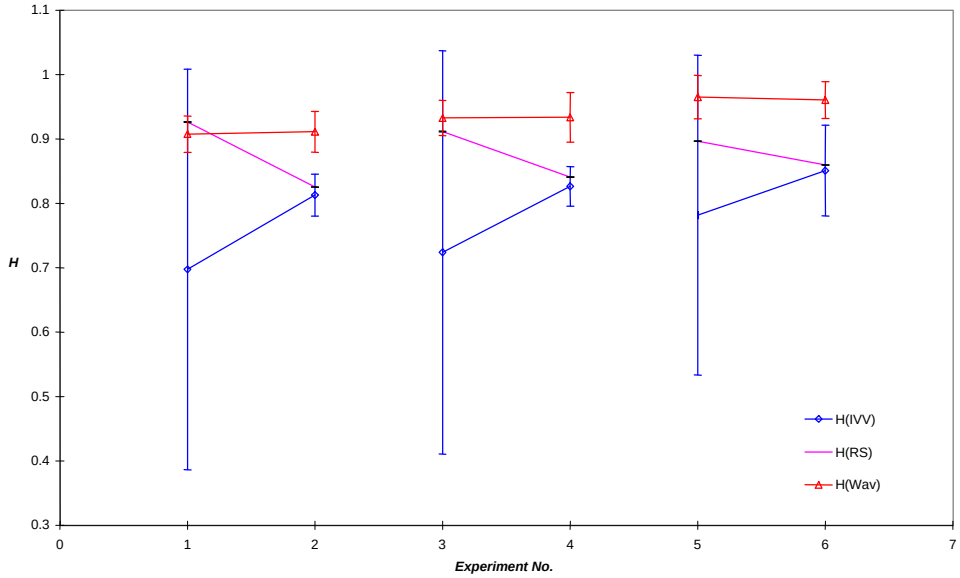

Figure 8.6 Behaviour of H on dominant m**.**

Expnt	m_1	m_2	$\hat{H}$ (IVV)	$\hat{H}$ (R/S)	$\hat{H}$ (Wave)	±95%Conf (IVV)	±95%Conf (RS)	±95%Conf (Wave)
1	1.5	1.8	0.697486	0.926343	0.907626	0.311187	0.000553	0.028213
2	1.8	1.5	0.812903	0.825452	0.911258	0.032674	0.000259	0.031741
3	1.6	1.8	0.723913	0.91165	0.932778	0.313278	0.000533	0.027295
4	1.8	1.6	0.826458	0.841017	0.933758	0.030684	0.000445	0.038404
5	1.7	1.8	0.781727	0.896696	0.965118	0.248511	0.000428	0.03363
6	1.8	1.7	0.850897	0.859873	0.960563	0.070523	0.000385	0.028605

Table 8.2 Table of Experiment Number, Double intermittency map m parameter values, the resulting estimate on H for map variance, R/S statistic and wavelet analysis methods together with their 95% confidence values

The striking feature that the results of Figure 8.6 show is that while Abry-Veitch retains the same value of H for the given parameter set which is predicted by the theory, the IVV method shows entirely different behaviour, most notably:

- when the dominant m is in the emitting *ON* part of the map, the variability of H is enormous. This is in contrast to the dominant m being in the *OFF* part of the map.
- A lower mean value of H is obtained when the dominant m is in the emitting half of the map.

- As the values of *m* approach each other, the variability in the map's output due to the dominant *m* in the emitting half of the map diminishes.

In addition to these observations, the R/S method (used as an additional marker) more or less gives the same results as the Abry-Veitch method when the dominant *m* is in the emitting half of the map while it gives the same results as the IVV method when the subdominant *m* is in the emitting half.

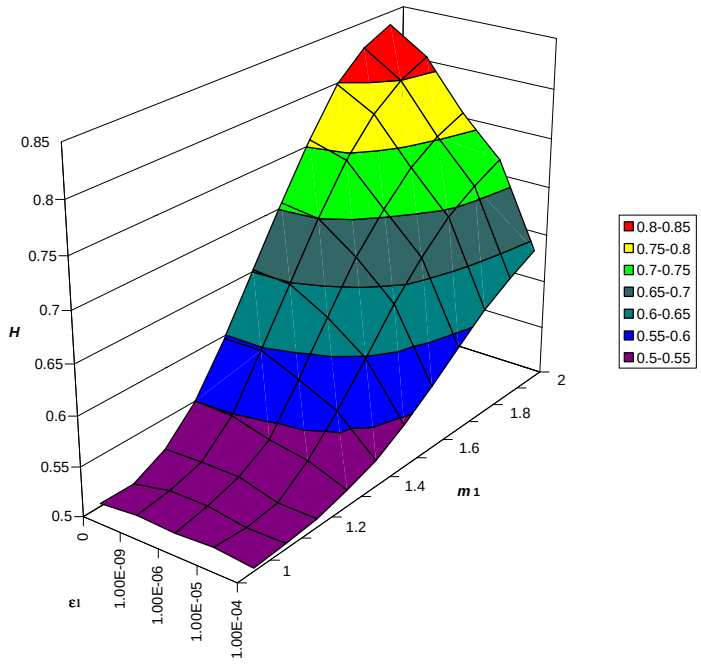

Figure 8.7 Single intermittence map variations in e_1 and m_1 against H using IVV

Recently Naranyan [NAR98] has formulated queueing behaviour dependent on H for input traffic that is self-similar. However if one combines the observations given above with recent reports that self-similar traffic with the same H produces vastly queueing behaviour [VAT98], then one can put forward the conjecture that the variability of H and the LRD of the emitting half of the source would give a better indication of the self-similar behaviour of the queue. The results shown in Figure 8.6 support this view and oppose the formulation proposed by Naranyan.

The results of Figure 8.7 show that *e* has an effect on H. It sets the upper value of H. This is in contrast to the results of Figure 5.7. The theory states that H (asymptotically) is not dependent on *e*. This is borne out by the results of Chapter 5 Figures 5.7 and 5.8. However we know that *e* sets the upper limit to the scaling of the burst of the source, i.e. the maximum sojourn time, and therefore sets the LRD which does affect H. This is evidenced in the results of Figure 8.7. This contradiction shows that to be able to capture the effect of self-similarity in measurements an appreciation of the way in which H is measured is required. Although both methods regress to a line, the Abry-Veitch method determines only the slope of the low

frequency component that requires the same regression of the linear points. IVV, on the other hand, is affected by the capping that e has on the slope. This last observation suggests that reliance on a single method for determining H on-line may prove unreliable. Comparing the results of Figure 8.7 and Figure 5.7 show this.

8.4 Conclusion

What we have seen in this chapter is that the IVV method is computationally tractable after a cumulation time required amassing the statistics. We have also seen that it converges at a reasonable rate, is stable and that this method is promising for the on-line measurement of H. However, we note from the accuracy comparisons with the Abry-Veitch method that H by itself as a parameter for modelling self-similar traffic in a "parsimonious" manner may not be enough. From these results it appears that there is a unique H associated with a given set of parameters. However it also appears that H does not describe adequately the instantaneous variability of the traffic, which may ultimately affect the behaviour of the queue, especially when the dominant LRD in a source may not be due to the active phase of a source. It is also worth noting that reliance on a single method for determining H on-line may prove unwise. The IVV method is derived from the correlation structure of the map (an LRD source). The implication of this is that it will yield results for H that more closely match the effect that H has on queueing behaviour i.e. the manner in which the arrivals into a queue appear. Furthermore, the IVV method is responsive to the effect of limiting the transit time (effects of e) and to whether the LRD is in the *ON* or *OFF* phase of a source. Both of these effects are missed by the Abry-Veitch method and for this reason the IVV method establishes its usefulness as a complementary method to the Abry-Veitch method. This confirms the usefulness of the map as an accurate and flexible model.

One final comment. Asymptotic queue behaviour dependent on H may have no practical meaning. Based on this we can further conjecture that what may be of more use is a dynamical formulation of the queue behaviour, as this can yield instantaneous predictions on "short term" buffer occupancy derived from dynamical formulations of long term correlation and scaling at the point of criticality i.e. the point at which we obtain bursts over all time scales of engineering interest.

9 Discussion and Further Work

9.1 Parameter Effects on H

In Chapter 5 we noted that the transit time through a map state depends on all the parameters of the map. That is to say that the transit time depends on e and m of the map state under examination, together with d which is common to all map states, and $r(x_0)$, which is the injection probability into the point x_0 in the state being examined. $r(.)$ is influenced by the parameters of the other map state. We also saw that the length of contiguous events (such as emissions of 1's or 0's) defines the LRD produced by the map state and hence the H that the model produces. We noted that e has the effect of setting the upper limit to the correlation lengths. We also noted that m sets the value of H. This is a limiting relationship as the frequency $\omega \to 0$. H, in actual fact, relates the gradient of curves such as the variance and correlation against lag. However we also noted that as d approaches the boundaries of the map (0,1) it affects the value of H that is witnessed. The reason for this is that the coupling between map states is governed by a set of equations. Due to this coupling the injections into the critical region of the map (the region that controls the LRD) is affected by the behaviour of the other map state. As d moves to its boundary values we require iterate values in the other map state which are very close to d. With d near its limits the invariant density is near uniform for a large part of the map interval. The effect of this is that the injection probability becomes very small. This affects the convergence rate on H. Fewer injections into the critical region mean that H takes longer to measure (as long-run contiguous chains of 1's or 0's become less frequent). However, it is thought that convergence will improve by decoupling the map states, thereby enabling true random injections into the map states on a transition.

9.2 Deficiencies – Invariant Density Approximation

A deficiency in the chaotic map models that we have used relates to their ability to obtain the invariant density analytically. Moreover, if a non-linear traffic source were to be characterised then this would require at least a good approximation of the sources' invariant density under all parameter conditions from which we can iterate rapidly to a final solution for the invariant density. At present we can approximate the invariant density of the chaotic maps under limited conditions, principally with $e = 0$. What is required is parameterisation that accounts for $e \neq 0$ and m_1, m_2 in their full range $m_i \in [1,2]$. The suggestion given above, that of decoupling the map states, may be a reasonable first step in this direction. The lack of adequate methods for obtaining individual densities limits the utility of these methods. This is an area recommended for further work.

9.3 Limitations of the Models

In Chapter 6 and 7 we explored two methods for aggregating traffic streams. In Chapter 6 we focused on obtaining the aggregate invariant density of the equivalent map. This was done by averaging the invariant densities of the *N* individual maps over a number of cardinal points of the invariant density (e.g. $r(1)$ and $r(d)$) in order to produce the equivalent density. Then the map was iterated *N* times per emission interval in order to produce the traffic. The resulting traffic stream was queued and compared against the original traffic trace. We also compared the behaviour of the equivalent map against Pruthi's interpretation in which the aggregation of *N* sources equates to a single map emitting *N* emissions per iteration. What we observed was that for a small number of sources the equivalent map produced queuing behaviour closer to the original trace than the method advocated by Pruthi. However we also noted that as the aggregation level increased, the value of *d* "walked back" to the origin. We have seen that this type of behaviour adversely affects the output behaviour of the map (see Chapter 5).

In order to overcome this obstacle we devised another method of aggregation. This is the method presented in Chapter 7. This method links the invariant density of the equivalent map to the probability density of the number of emitting sources. The type of linkage described here actually implies a correspondence (coupling) between a particular iterate value and a particular number of emitting sources. In other words, there is an x_n iterate value for which there is a unique m_n sources being emitted. Unfortunately this type of coupling between overlying and underlying dynamics is difficult to achieve practically. We therefore relaxed the condition on the coupling to a probabilistic one via considerations on the cumulate invariant density and the cumulate source emission density. In other words, for an $x_n<X$ we have the probability that the orbit has visited all points up to X; this leads to an $m_n<M$, which is the probability of m_n sources up to a maximum of M sources, being emitted. In this way we presume some coupling between underlying and overlying dynamics. Using this method we achieved an order of magnitude speed-up over iterating *M* sources independently.

This limitation of the model due to *d* "walk back" can be overcome in a practical way by viewing the time resolution of the map iterates. The low value of *d* is due to the fact that we wish to model particular load values. If these load values are viewed from a single maximum link rate then it may well be that the individual source rate is very low compared to the link rate, which leads to a low source load value. However, when this is viewed at the tributary link rate, the load may be reasonable in the sense that *d* is no longer near its extremes. In practical terms this means that rather than iterate the map at the maximum link rate (say every iteration) we iterate the map at a reduced rate (for example every 6^{th} iteration) which corresponds to the tributary link rate.

9.4 Superimposition of Source Distributions

A major assumption in Chapter 6 and Chapter 7 is that the sources being aggregated are IID. In practice this assumption may be an oversimplification of the reality that may exist in future

networks. This is because feed-back control methods are currently being proposed and implemented for data traffic in ATM and IP6. The use of feed-back methods implies that the behaviour of data traffic arriving at a buffer will be dependent on all the other data sources entering the same buffer. The same is true if priority buffering is used at the buffers. The stochastic behaviour of the individual sources then becomes dependent rather than independent. However, because the source distribution is simply a superimposition on the underlying dynamics, then provided that the network still sees aggregate traffic with high *H* values, a new source distribution which reflects reality would still work because we know that the underlying dynamics permeates through to the overlying dynamics.

In fairness, we should also point out that superimposition of the source distribution as used by the bulk properties map does distort the anticipated value of *H* from the model. This distortion needs to be removed in order to make the model more effective. This is an area for further work.

9.5 Effectiveness of H

From the investigations into chaotic maps as traffic models we make the following remarks on the effectiveness of *H* as a parameter of measured traffic and as a parameter to be modelled in traffic simulations.

We noted in Chapter 8 from the comparisons of the IVV against Abry-Veitch methods of measuring *H*, that *H* by itself as a parameter for modelling self-similar traffic in a "parsimonious" manner may not be enough. This is because, while it appears that there is a unique *H* associated with a given set of parameters, it also appears that, while *H* may describe the long term scaling, it cannot adequately describe the "instantaneous" short-term variability of the traffic. It is this "instantaneous" behaviour which may ultimately affect the behaviour of the queue, especially when the dominant LRD in a source may not be due to the active (*ON*) phase of a source. Because of these observations we submit that reliance on a single method for determining *H* on-line may prove unwise and that, consequently, a combination of dissimilar measurement methods as behavioural indicators of *H* on-line may prove more useful. This submission is supported by comparing the results of Figure 8.6, Figure 8.7 and Figure 5.7.

We have also examined, in a limited way, how *H* affects queue behaviour (Chapter 6). We noted that it is possible to obtain *H* values for shorter measurement periods but that the variability around the mean value of *H* is high (Chapter 8). We also noted that the LRD prevalent in traffic arriving at a queue may not necessarily be the LRD value which dictates the unique *H* value of the aggregate traffic (Chapter8). We also noted from experiments and observations for high values of *H* that *H* converges very slowly (Chapter 5 and Chapter 8). From this evidence we submit that *H* may have no practical meaning for queues other than for assuring buffer overflow.

With regard to modelling we note that, for true high value *H* traffic measured over realistic measurement periods, evaluating queue behaviour (i.e. running experiments and gathering statistics until some state evaluation of the queue behaviour is obtained) may become limited, if not problematic. This is because high *H* affects convergence and relatively small measurement periods result in high variability in the value of *H*; this makes the collection of stable queue statistics problematic when modelling networks which have high *H* traffic. In this light a different approach to examining queues has to be thought of.

These observations lead us to the conclusion that while *H* is useful for determining the long term scaling of the traffic it only forms part of the picture. The act of buffer overflow is a dynamical process. This is in contrast to the accepted stochastic view of accounting for enough buffer space such that, on average, the buffer overflow becomes acceptable. High values of *H* imply that almost surely, regardless of the buffer size, the buffer will overflow. We therefore further conjecture that what may be of more use is a dynamical formulation of the queue behaviour which can yield instantaneous predictions on "short term" buffer occupancy derived from:

- dynamical formulations of long term correlation, and
- dynamical scaling at the point of criticality i.e. the point at which traffic becomes bursty over all time scales of engineering interest.

9.6 Remarks on Traffic Modelling and Stationarity

High *H* values cast some doubt on ideas of stationarity, at least in its practical sense. It takes *H* a long time to converge on stable results (longer than one can feasibly measure). This indicates, in a practical sense, that the time series behaves in a non-stationary way over the measurement period. Under this light perhaps it is better to view the problem of self-similar traffic and its modelling in a different way. The realism in modelling comes from modelling what is present in the traffic streams. The benefit of modelling comes from observing the effect of the model on network elements (buffers and the like). If stationarity is an issue due to high variability then analysis via other means such as dynamics may again prove useful, for example the map family has *H* in built into the map, however the map family is flexible enough to have its parameters adjusted on line. In this manner the non-stationary nature of the traffic could be captured. The determination of such an approach is an area that is recommended for further work.

9.7 Remarks on Chaotic Maps and ON-OFF Self-similar Traffic Modelling

Taqqu's theorem, mentioned in Chapter 3 [TAQ97], is really an extension/justification of the self-similar traffic model first presented by Norros in [NOR93]. The main difference with Norros' model is that Taqqu employs a validity condition that states that in order to obtain

self-similar traffic the individual *ON-OFF* sources must have an element of LRD either in *ON* or *OFF* states. We summarise the results of Taqqu in the following manner:

- the *ON* state must have LRD, AND/*OR* the *OFF* state must have LRD,
- the *ON* and *OFF* states are independent of each other,
- the *ON* and *OFF* states need not have the same distribution.

This is interesting from a chaotic map modelling standpoint because the conditions stated above correspond to a chaotic map model:

- that belongs to the intermittency map family with a map parameter combination of $e_j \to 0$ and $m_j \in (1,2)$, where e_j has to be small enough to have burst behaviour over the time scales of interest;
- where the aggregated distributions of *ON* periods can be assumed to be Gaussian distributed at any instance, *t*, of inspection for a sufficiently large aggregation level.

This aggregate behaviour can be modelled by a bulk properties map with random injection into the *ON* and *OFF* regions of the map, i.e. a decoupled bulk properties map.

9.8 Chaotic Maps - Potential Application Areas

The importance of *H* lies in the fact that queue length distributions are sensitive to its value [NOR93]. The higher the value of *H* the higher the probability of high queue state occupancy. The implication of controlling *H* (and its derivative) is that congestion within the network can be controlled. An important first step in this direction is through the on-line modelling of network elements such as switches. As we have seen the bulk properties map technique enables the adjustment of *H* via the alteration of parameters ε and *m*. We therefore submit that the bulk properties map technique is a contribution in this direction.

The bulk properties map can be used to generate self-similar traffic on-line with $H \in (1/2,1)$. This is in contrast with other approaches such as FBM and FGN which generate samples off-line. We have also seen that manipulation of the map parameter *e* away from zero increasingly destroys the self-similarity produced by the chaotic map. These results (on-line generation and on-line parameter manipulation) are promising for the development of network control techniques based on chaos theory and the coupling of chaotic maps (for the description of such a technique see [MON97b]). These techniques rely solely on local information present at a network element, and as such prove attractive as areas of future research because:

- they are inherently scaleable; and
- they would reduce global control signalling across networks.

On this theme we again comment on the theorem put forward by Taqqu [TAQ97] with regard to the modelling of aggregate heterogeneous *ON-OFF* traffic sources. The heterogeneous traffic equation

$$W_M^*(Tt) = Tt\left(\sum_{r=1}^{R} M^{(r)} \frac{m_1^{(r)}}{m_1^{(r)} + m_2^{(r)}}\right) + \sum_{r=1}^{R} T^{H^{(r)}} s_{\lim}^{(r)} \sqrt{L^{(r)}(T) M^{(r)}} . B_{H^{(r)}}(t),$$

(9.6)

has implications for the control of networks from a "fairness" allocation point of view for the following reasons:

- the result can be viewed as a model description of traffic entering an aggregation point, such as a switch;
- the controlled traffic mean level is dependent on the mean value of aggregation and this is given by the term

$$\widehat{W_M^*}(Tt) \approx Tt\left(\sum_{r=1}^{R} M^{(r)} \frac{m_1^{(r)}}{m_1^{(r)} + m_2^{(r)}}\right).$$

(9.7)

This relies only on local knowledge of the *R* types of sources arriving at the node, and,

- the allocated bandwidth is a proportion to the ratio of the mean sojourn lengths i.e. the ratio

$$m_1^{(r)} \Big/ \left(m_1^{(r)} + m_2^{(r)}\right).$$

(9.8)

Network control is affected by altering the variability term that is represented by the second summative term, that is

$$\mathrm{var}\left(W_M^*(Tt)\right) \approx \sum_{r=1}^{R} T^{H^{(r)}} s_{\lim}^{(r)} \sqrt{L^{(r)}(T) M^{(r)}} . B_{H^{(r)}}(t).$$

(9.9)

More precisely, network control is affected by altering the variance in the traffic stream produced by traffic of type *r*, where *r* is the traffic from the most dominant LRD traffic source. Therefore by modelling the source traffic on-line we can identify the dominant LRD traffic and take controlling actions accordingly.

Further insight into network control may be gained from employing some ideas from the dynamics of Lévy flights. Formally Lévy flights have been studied as jump models. These are models in which the particle moves instantaneously between periods of halt. The use of jump models is not unknown in teletraffic analysis because they are the base behaviour behind that of Markov chains (for an explanation on the relationship between jump models and Markov chains see [KLE75, GIL92]). The interesting thing about Lévy flights is that they can also be modelled by chaotic maps which exhibit intermittency [ZUM93, GEI92, KLA93]. In these models, constant velocity Lévy flights are interpreted as particles moving with constant velocity between halts, with the intermittency region of the map providing the constant velocity episodes. When applying these models to teletraffic we may interpret the constant

velocity episodes as a cell or packet transiting through a network before appearing as an emission into a queue i.e. a halt.

The reason for pursuing dynamical formulations for network behaviour, for example the one outlined above, rather than the stochastic ideas commonly employed, is that the network congestion problems itemised by Fowler [FOW91] can be approached using non-linear chaotic control ideas developed by Ott, Grebogi and Yorke[OTT90] and extended to lattice networks by Mondragón and Arrowsmith [MON97b]. For this reason further investigation of non-linear dynamical map models, with emphasis placed on developing such models for chaotic control of networks, is encouraged.

9.9 Non-linear Control - A New Network Control Paradigm

In this subsection we propose a new network control methodology based on the results of the research reported in this thesis. The aim of our approach is to consider the control of a network containing self-similar traffic which we can model using chaotic (intermittency) maps. Each chaotic map describes the traffic generated by an element of the network (computer data, voice data, etc.). These maps are coupled due to the aggregation, switching, and routing of the traffic in the network. A network of such maps is similar in form to a lattice structure. Such lattice structures can be controlled using chaotic control methods. The method of controlling "chaos" developed by Ott, Grebogi and Yorke (OGY) [OTT90] has recently been extended to the control of non-linear networks such as lattices. There have been several successful approaches to controlling a network of chaotic elements (Mondragón and Arrowsmith [MON97b], Oketani *et al*, [OKE95] Sepulchre and Babloyants [SEP95], Youssefmir and Huberman [YOU95]). We begin by outlining current network control methodology by citing the example of network control applied to ATM.

9.9.1 Traditional Control

Traditional control in network traffic is based on the following ethos. There are three levels of immediate control which look at the time scales of call-by-call duration. These levels of control can be summarised as[CHET95]:

instantaneous - this level of control relates to the selective discard of cells, cell scheduling, traffic shaping and User Parameter Control (UPC);

end-to-end propagation delay - this level of control relates to the tagging of cells, explicit forward congestion indication and fast resource protocol; and

end-to-end round trip delay - this relates to routing and Connection and Admission Control (CAC) functions.

Above these immediate levels of control lies the management plane which traditionally looks at time scales longer than that of a call-by-call duration, i.e. it looks at the medium to long term control of the network such as provisioning and restructuring of the network in order to cope with changing demands on the system.

In order to achieve the above levels of control, the following kinds of control are applied:

1. ***Call level control*** - Permits traffic onto the network provided that there are enough resources on the requested path which permit the incoming traffic to propagate across the network without causing congestion at the cell level.
2. ***Cell level control*** - Allocates resources in the network (such as buffer space) in order to accommodate the call in terms of cells or allow some traffic loss according to some pre-agreed cell loss rate.

Both (1) and (2) have been traditionally handled on entry onto the network via CAC algorithms. They have taken various guises from:

- purely mathematical/theoretical in their approach to the CAC problem (examples of this type can be found in Gibbens *et al* [GIB95] and Borst and Mitra [BOR96]) to,
- those based on heuristics/intelligent learning approaches (examples of these can be found in Ramalho [RAM96] and Hiramatsu [HIRA90]).

As admission into the network is carried out on a link by link basis then the decisions made are usually conservative because the bandwidth for the call is governed by the link rate of the slowest link (bottle neck) on the call path. Additionally (ii) can be approached via the design /dimensioning of switches and cell level control methods such as UPC.

Chaotic control can be applied to the three levels of control mentioned above. In chaotic control, ***call level control*** is termed "order". This is the selection of a call or burst based on a weighted decision derived from the dynamics of the system (network). ***Cell level control*** is termed "procession" and is the effect of control on the dynamical system which permits the transfer of data between source and sink.

9.9.2 Dynamical Systems Approach to Teletraffic

As a first stage in developing a chaotic control method for telecommunication networks we have studied how H changes with the map parameters. A decision to alter H based on the adjustment of these parameters can be interpreted as the controlling action. For example, an alteration in the value of e imposes an upper cut-off on the correlation, an alteration in d changes the mean traffic load and changes in m_1 and m_2 change the sojourn time of the *ON* and *OFF* states i.e. the LRD of the traffic and its H value. These effects have already been studied in Chapters 5,6, and 7.

Chapter 5 shows how an individual source can be modelled by a member of the map family. Chapter 6 and 7 show how aggregate sources can be modelled by an equivalent map which preserves the traffic load and H. The parameterisation of this "equivalent" map is reported in Samuel *et al* [I, II], and Chapter 8 suggests how to measure H on-line.

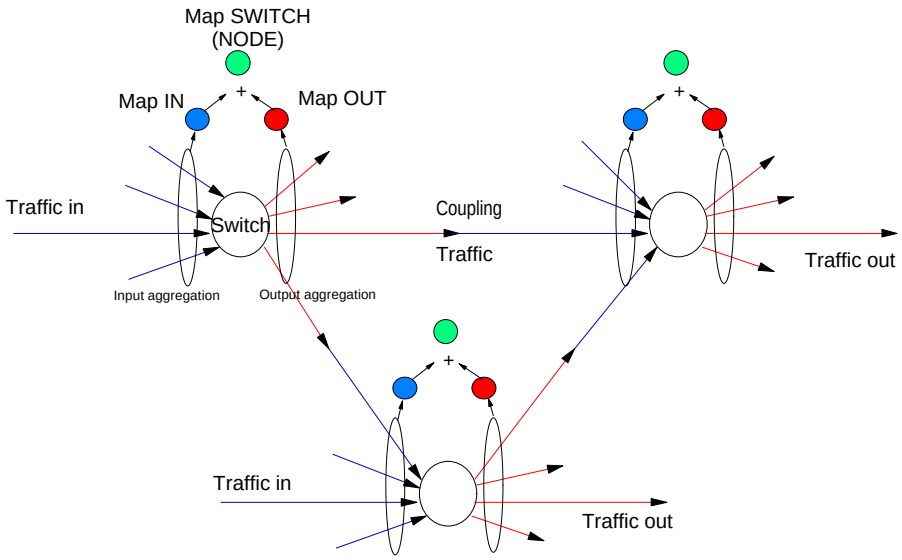

Figure 9.1 Lattice interpretation of a telecommunications network

9.9.3 The MAPS Control Paradigm

We have mentioned how chaotic maps can be used to provide aggregate models for self-similar traffic. We considered the case of aggregation at a single node [I]. A natural extension of nodal models is to couple them together to form networks and then attempt to model and control the characteristics of these networks. In such a model the nodes would be the switching sites and the couplings between the nodes would represent the links between the switching sites. The simplest mathematical models which resemble such a construction are Coupled Map Lattices (CML) where a dynamical system at each node will produce the local traffic input. Coupling will be provided by external input from neighbouring nodes due to the queueing (aggregation) and switching. This is shown schematically in Figure 9.1.

Investigations have been made into regular lattice structures where coupling exists between nodal sites. An important property has been the discovery that global control across all nodes can be obtained via local control at each node [MON97a, 97b, ARR96, OTT90].

In these investigations the dynamical behaviour of each nodal site is modelled by a chaotic map and the coupling to the neighbouring nodes imparts perturbations into the orbits of the dynamical system containing the node. Since the chaotic map's orbit possesses the property that any orbit will approach arbitrarily closely every point of the plane described by the chaotic map, then at some point the orbit must take it near to a desired control state. A small feedback control applied at the target point in the orbit places the dynamics of the node into a required state, since the same structure occurs at all nodes. Experimental evidence shows that if a desired control state is prescribed for all nodes then eventually the lattice becomes controllable. However, it is possible for neighbouring dynamical behaviour to kick a node out of equilibrium via the coupling and so "occasional feedback control" is introduced where the feedback control is activated within the control region around the desired equilibrium for only part of the allowable time [MON97a].

The outline given above (see also [I]) describes how each node site of the lattice structure can be modelled by a controlling map. The individual traffic streams entering a node can be modelled by chaotic maps and in certain models the maps can be aggregated into a single map which describes the behaviour of the traffic at the node. The next step is to provide on-line information on *H*. This is necessary to enable the construction of an active dynamic control environment.

We propose a control scheme that actively manipulates the value of *H*. Our intended approach is based in manipulating *H* via the mean, peakedness and LRD of the traffic stream characteristics by means of a local control strategy. The "philosophy" is not to destroy the chaotic behaviour of the traffic but instead to use its variability as a method of control. Chaotic systems are everywhere unstable and thus a small change in the system at any instant produces a large change at later times (this is SIC or, more colloquially, "the butterfly effect"). This gives the controller "agility" to changes in the traffic over many timescales. Moreover, as was noted earlier, successful control of coupled chaotic systems can be instigated by the local control of each system. The implication of this to networks should be significant, since it suggests that for relatively small control actions applied locally the congestion/buffer occupancy on local and remote switches (relative to the control site) should be reduced. It is thought that in transport systems such as ATM mechanisms like ABR via a chaotically initiated control sequence would provide a mechanism for the reduction of *H* and subsequent control of congestion.

We are proposing two different mechanisms to control traffic that is already in the network. The first mechanism seeks to reduce the variability of the traffic in a specific channel. This can be done by "careful" introduction of empty cells. This control mechanism would modify but not destroy the highly variable behaviour of the traffic. We conjecture that, individually, each of these controlled channels would change very little but that these changes would have a larger effect when the traffic is aggregated in the queue.

The second mechanism is based in a random selector of calls in a node. The random selector would choose which call to admit by weighting dynamically the statistics of the traffic variability. The selector, modelled by a chaotic map, would assign larger probabilities to some channels than others but all the channels would have positive probabilities to be served.

The first control mechanism is termed "Procession" and the second "Order". Conceptual views of the proposed chaotic control regime can be found in Figure 9.2.

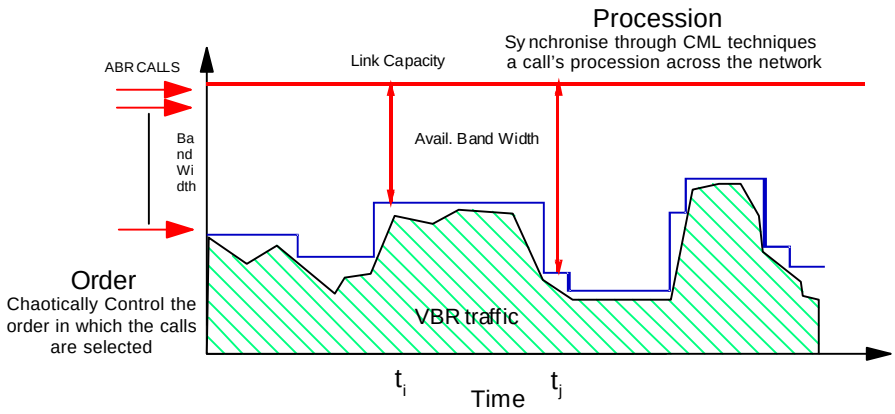

Figure 9.2 Conceptual view of chaotic network control as applied to ABR

These two mechanisms of control can be developed using chaotic maps as models of self-similar traffic because it is known that they have the correct characteristics and, moreover, that they can model high traffic rates efficiently [I, II].

10 Conclusions

In this thesis, we have applied techniques from non-linear dynamics to the teletraffic modelling of modern packetised telecommunications networks. Measurements from these networks have revealed self-similar behaviour, i.e. burstiness ranging over all time-scales of engineering interest. This burstiness can have a significant impact on network resource occupancy. We have developed a novel teletraffic framework for the use of non-linear dynamics models in the form of intermittency maps which consist of parsimonious, parameterisable and predictable models for the accelerated simulation of both self-similar behaviour of individual sources and aggregate traffic in such networks.

The teletraffic framework consists of a family of maps based on the intermittency map with single *ON-OFF* interpretations for individual sources, and either an *N* times iterated equivalent single maps, or a two-step bulk property interpretation for aggregate traffic representation. These aggregate traffic models are novel and provide up to two orders of magnitude speed-up over other methods (FBM/FGN and Pruthi's *N* one-step method). Further, we have significantly extended the characterisation of individual source models previously carried out by Pruthi and Erramilli. The extension accounts for the impact of all five parameters (*e* and *m* for both states, and *d*) on *H*, the parameterisation for load via the invariant density, and the parameterisation for heavy tailed sojourn times in the *ON* and *OFF* states via the transit-time.

The theoretical analysis of the maps with respect to *H* confirms the conjecture put forward by Pruthi that asymptotically *H* is only dependent on the dominant value of *m*. Numerical results show that convergence is slow and that for the coupled map *H* differs substantially from the theory. However the deviation from the theoretical is predictable and this has lead to an empirical fit for the asymptotic dependence of *H* on *m*. An important feature that these results also show is that the underlying dynamics of the map persist in all of the map interpretations. The numerical results also show limitations of parameter ranges on *H*, particularly for *d* (*d* must lie in the range 0.1 to 0.9). However, this limit can be overcome practically to some degree by manipulating the time resolution of the iterates. A further parameter limitation stems from *e;* any value of $e > 0$ effectively limits range of time-scales over which LRD occurs. These numerical limitations apply to all map interpretations.

We have developed a method of measuring *H* via the map's variance (IVV). We have shown that this method is promising for measuring *H* on-line. We have also found by comparing the IVV and Abry-Veitch methods for measuring *H*, that *H* by itself as a parameter for modelling self-similar traffic in a "parsimonious" manner may not be enough. This conclusion is drawn from observing different queueing behaviour with input traffic having the same *H*. This conclusion also leads to the practical suggestion that reliance on a single method for

determining *H* on-line may prove unwise. Additionally, the results from the comparison of IVV and Abry-Veitch show the flexibility that these map models have in specifying key LRD behaviour that determines the impact on queueing. This flexibility is derived from the intuitive relationship that these map models have to their underlying physical *ON-OFF* process of cell/packet transfer in networks.

A further asset that increases the flexibility of these maps is that they can depict non-stationary traffic. This is because the maps can have their parameters altered dynamically. Furthermore manipulation of the maps' parameters may point to the development of a control schemes for telecoms networks based on non-linear dynamics.

Finally, analysing and understanding the statistical behaviour of these non-linear dynamics models is a significant step towards developing the theoretical framework necessary for statistical-chaotic control policies to address congestion avoidance in telecoms networks.

References

[I] Samuel L.G., Pitts J.M., and Mondragón R.J., **Towards the Control of Communication Networks by Chaotic Maps: Source Aggregation,** in Vol 2b Proc. ITC15, (1997)"Teletraffic Contributions for the Information Age" (eds. V Ramaswami and P.E. Wirth), Elseveier, pp1369-1378.

[II] Samuel L.G., Pitts J.M., and Mondragón R.J., **Fast Self-similar Traffic Generation**, IEE. 14th UK Teletraffic Symposium, pp. 8/1 - 8/4, March 1997.

[III] Samuel L.G, Pitts J.M., Mondragón R. J., and Arrowsmith D.K., **The MAPS control paradigm: using chaotic maps to control telecoms networks**, *in Proc. IFIP BC'98, (1998) "Broadband Communications The future of telecommunications"* (eds. P.J. Kühn and R.Ulrich), Chapman & Hall, pp371-382.

[IV] Samuel L.G., Pitts J.M and Mondragón R.J. **Applications of Non-linear Dynamics to Network Modelling,** IEE. 15th UK Teletraffic Symposium, pp. 19/1 - 19/5, March 1998.

[V] Ho L.T.H., Pitts J.M., Samuel L.G., and Schormans J.A., **Buffering packetised data: an examination of the crossover effect,** Electronics Letters, Vol. 34, No. 17, pp1637 -1638.

[VI] Samuel LG. Pitts JM. Mondragón RJ. Arrowsmith DK. **The maps paradigm: chaotic control of networks**. Sixth IEE Conference on Telecommunications (Conf. Publ. No.451). IEE. 1998, pp.57-61. London, UK.

[ABRA65] Abramowitz M and Stegun I.A.(Eds), **Handbook of Mathematical Functions**, Dover Publications, New York, 1965.

[ABR95] Abry P., Gonçalvès P. and Falndrin P., **Wavelets, spectrum analysis and 1/f processes**, pp 15-29 in: Lecture Notes in Statistics, No 103, Springer-Verlag, Berlin, 1995.

[ABR98] Abry P and Veitch D., **Wavelet Analysis of Long-Range-Dependent Traffic**, IEEE Trans. on Information Theory, Vol. 44, No. 1, 1998, pp 2-15.

[ADD95] Addie R G, Zukerman M, Neame T. **Performance of a Single Server Queue with Self-Similar Input**, IEEE ICC95 Seattle, Washington, pp461-465.

[ARR96] Arrowsmith, D.K., Lansbury, A.N., and Mondragon, R.J. **Control of Arnold circle map**. Int. Jour of Bif and Chaos, 1996, 6, pp437-453.

[BEND84] C. M. Bender and S. A. Orszag., **Mathematical Methods for Scientists and Engineers**, McGraw Hill, 1984

[BEN85] Ben-Mizrachi A., Procaccia I., Rosenberg N., Schmidt A. and Schuster H. G., **Real and apparent divergencies in low-frequency spectra of non-linear dynamical systems**, Physical Review A, Vol 31, No. 3 , March 1985. pp1830-1840.

[BER94] Beran J., **Statistics for Long-Memory Processes**, Chapman & Hall, New York, 1994.

[BER95] Beran J., Sherman R., Taqqu M .S., Willinger W. **Long Range Dependence in Variable-Bit Rate Video Traffic.** IEEE Trans on Communications, Vol 43, No 2/3/4, Feb/March/April 1995. pp. 1566-1579.

[BOR96] Borst S and Mitra D., **Optimal connection admission control in ATM networks**, Performance Evaluation Vols 27 & 28 (1996) pp 437-451.

[CHEH96] Chen H. M., Mellor J., and Mars P., **Comparisons of Simulation Algorithms for Self-Similar Traffic Models**. IEE 13th UKTS, 1996, pp8/1-8/8.

[CHET95] Chen, T. and Liu, S. **ATM Switching Systems**, Artech House, Boston, 1995.

[COL80] Collet, P. and Eckmann, J. P., **Iterated Maps of the Interval as Dynamical Systems**, Birkhäuser, Boston, 1980.

[CRO-A] Crosby S., Leslie I., Huggard M., Lewis J T, Toomey F. and Walsh C., **Bypassing Modelling: Further Investigations of Entropy as a Traffic Descriptor in the Fairisle ATM network**, Proc. of 1st Workshop on ATM Traffic Management, Paris 1995.

[CRO-B] Crosby S., Leslie I., Huggard M., Lewis J T, McGurk B., and Russell R **Predicting bandwidth requirements of ATM and Ethernet traffic,** IEE 13th UK Teletraffic Symposium, Glasgow, March 1996.

[PRY91] de Prycker, M. **Asynchronous Transfer Mode: Solution for Broadband ISDN**. Ellis Horwood, London., 1991.

[DEV94] Devaney R.L., **An introduction to Chaotic Dynamical Systems 2nd. Ed.** Addison-Wesley, Reading Massachusetts, 1994.

[DUF94a] Duffield N G, Lewis J T, O'Connell N, Russell R, Toomey F. **Statistical issues raised by the Bellcore data** Eleventh UK Teletraffic Symposium, Cambridge, 23-25 March 1994.

[DUF94b] Duffield N G and O'Connell N. **Large deviations for arrivals, departures, and overflow in some queues of interacting traffic** Eleventh UK Teletraffic Symposium, Cambridge, 23-25 March 1994.

[DUF95a] Duffield, N.G., Lewis, J.T., O'Connell, N., Russell, R. and Toomey, F. **Entropy of ATM Traffic Streams: A Tool for Estimating QoS parameters**, IEEE Jour. on Sel. Areas in Comm., Vol. 13, No 6, 1995, pp980-990.

[DUF95b] Duffield N G, Lewis J T, O'Connell N, Russell R, Toomey F. **Predicting Quality of Service for Traffic with Long-Range Fluctuations**, IEEE ICC95 Seattle, Washington, 1995, pp473-477.

[ERR94a] Erramilli A., Singh R. P., and Pruthi P., **Chaotic Maps as Models of Packet Traffic.** ITC 14 pp 329-338 , June 1994.

[ERR94b] Erramilli A, Pruthi P, Willinger W. **Modelling Packet Traffic with Chaotic Maps**, ISRN KTH/IT/R-94/18—SE, Stockholm-Kista, Sweden, Aug. 94.

[ERR94c] Erramilli A, Gordon J and Willinger W**. Applications of Fractals in Engineering for Realistic Traffic Processes**. ITC14 pp 35-44 1994

[ERR95a] Erramilli A., Singh R. P., and Pruthi P., **An application of deterministic chaotic maps to model packet traffic**, Queueing Systems 20 (1995) pp 171-206.

[ERR95b] Erramilli A, Pruthi P, Willinger W. **Recent Developments in Fractal Traffic Modelling.** Proc ITC International Teletraffic Seminar. St. Petersberg, 26 June - 2July, 1995.

[ERR95c] Erramilli A, Pruthi P, Willinger W. **Self-Similarity in High-Speed Network Traffic Measurements: Fact or Artifact?.** Proc. 12th Nordic Teletraffic Seminar NTS12, Espoo, Finland, 22-24 Aug., 1995.

[ERR96] Erramilli A., Naranyan O. and Willinger W., **Experimental Queueing Analysis with Long-range Dependent Packet Traffic.** IEEE/ACM Trans on Networking, Vol 4, No 2, April 1996, pp209-223.

[FAN97] Fan, Z. and Mars, P., **The Impact of the Hurst Parameter and its Crossover effect on Long Range Dependent Traffic Engineering**, IEE 14th UKTS, 10/1-10/8.,1997.

[FOW91] Fowler H.J, Leland W. E. **Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management**. IEEE Jour. on Sel. Areas in Comm., Vol 9, No 7, Sept. 1991, pp1139-1149.

[GAR95] Garrod C., **Statistical Mechanics and Thermodynamics**, Oxford University Press, Oxford, 1995.

[GEI85] Geisel T, Nierwetberg J and Zacherl A**, Accelerated Diffusion in Josephson Junctions and Related Chaotic Systems**, Phyical Review Letters, Vol 54, No 7 , Feb. 1985, pp 616-619.

[GIB95] Gibbens R.J., Kelly F.P. and Key P.B., **A Decision-Theoretic Approach to Call Admission Control in ATM Networks**, IEEE Jour. on Sel. Areas in Comm., Vol 13, No 6, Aug. 1995, pp1101-1114.

[GIL92] Gillespie, D.T., **Markov Processes an Introduction for Physical Scientists**, Academic Press, Boston, 1992.

[GRA80] Gradshteyn I. S and Ryzhik I. M., **Table of Integrals, Series and Products**, Academic Press, New York, 1980.

[GUL92] Gulick, D., **Encounters with chaos**, McGraw-Hill, NewYork,1992.

[HEA94] Heagy, J.F., Platt, N. and Hammel, S.M., **Characterization of on-off intermittency**, Physical Review E, Vol. 49, No. 2, pp1140-1150, Feb. 94.

[HIRA90] Hiramatsu A, **ATM communication Network Control by Neural Networks**, IEEE Trans. on Neural Networks, Vol 1, No 1 March 1990, pp12-30.

[HIR82a] Hirsch, J.E., Huberman, B. A. and Scalpino, D.J., **Theory of Intermittency**, Physical Review A, Vol 35, No. 1. Pp519-532. Jan. 82.

[HIR82b] Hirsch, J.E., Nauenberg, M. and Scalpino, D.J., **Intermittency in the presence of Noise: A renormalization group formulation**. Physics Letters, Vol 87A No. 8. pp391-393. Feb. 82.

[HUA95] Huang C, Devetsikiotis M, Lambadaris I, Kaye R.,. **Fast Simulation for Self-Similar Traffic in ATM Networks**, IEEE ICC95 Seattle, Washington, 1995, pp438-444.

[KAU94] Kaufmann Z., Lustfield H. and Bene J., **Eigenvalue spectrum of the Frobenius-Perron operator near intermittency**, Physical Review E, Vol 53, No 2, pp 1416-1421, Feb 94.

[KLA93] Klafter J , Zumofen G, and Shlesinger M.F**, Lévy walks in Dynamical Systems**, Physica A, Vol 200, 1993, pp 222-230

[KLE75] Kleinrock L. **"Queueing Systems Volume 1: Theory",** John Wiley and Sons, New York, 1975.

[KRI96] Krishnan, K.R., **A new class of performance results for a fractional Brownian traffic model**, Queueing systems 22, 1996, pp277-285.

[LEW96] Lewis J. T. and Russel R., **An Introduction to Large Deviations**, Tutorial Proceedings, Performance 1996, Lausanne.

[LAU95] Lau W, Erramilli A, Wang J L, Willinger W. **Self-Similar Traffic Generation: The Random Midpoint Displacement Algorithm and its Properties**, IEEE ICC95 Seattle, Washington, 1995, pp473-477.

[LAW91] Law, A.M and Kelton, W. D. **Simulation Modelling and Analysis 2nd. Ed.,** McGraw-Hill, New York, 1991.

[LEL93] Leland W.E**.,** Taqqu M. S., Willinger W., and Wilson D**., On the Self-Similar Nature of Ethernet Traffic.** Proc. ACM SIGComm '93. San Fransisico, CA, Sept.1993.

[LEL94] Leleand W.E., Taqqu, M. S., Willinger W., and Wilson D**. On the Self-Similar Nature of Ethernet Traffic (Extended Version).** IEEE/ACM Trans on Networking, Vol 2, No 1, Feb 1994., pp1-15

[LUS96] Lustfield H., Bene J. and Kaufmann Z., **The Correlation Functions Near Intermittency in a One-Dimensional Piecewise Parabolic Map**, Journal of Statistical Physics, Vol. 83, No 5/6, pp 1199-1210, 1996.

[MAN63] Mandelbrot B., and Berger J.M., **A New Model for Error Clustering in Telephone Circuits**, IBM Journal , July 1963.

[MAN65] Mandelbrot B., **Self-Similar Error Clusters in Communication Systems and the Concept of Conditional Stationarity**, IEEE Trans. On Communication Technology March 1965, pp71-90.

[MAN68a] Mandelbrot B., and Van Ness J.W., **Fractional Brownian Motions, Fractional Noises and Applications**, SIAM Review Vol. 10, No 4, October 1968, pp 422-437.

[MAN68b] Mandelbrot B., and Wallis J. R., **Computer Experiments with Fractional Gaussian Noise** Water Resources Research, Vol. 5, No 1, February 1968, pp 228 - 267.

[MAN68c] Mandelbrot B., and Wallis J. R., **Computer Experiments with Fractional Gaussian Noise Part2 , Rescaled Ranges and Spectra**. Water Resources Research, Vol. 5, No 1, February 1968, pp 242- 259.

[MANN80] Manneville. P., **Intermittency, self-similarity and 1/f spectrum in dissipative dynamical systems**. Le Journal de Physique, Vol 41, No 11, November 1980, pp 1235-1243.

[MCC95] McCauley J.L. **Chaos, Dynamics and Fractals and Algorithmic Approach to Deterministic Chaos**, Cambridge University Press, Cambridge, 1995.

[MEI91] Meier-Hellstern, K., Wirth, P., Yan, Y-L., and Hoeflin, D., **Traffic models for ISDN data users: office automation application**, Proc. 13th International Teletraffic Congress, Copenhagen 1991.

[MOL97] Molnár, S. and Vidács**,** A. **On Modelling and Shaping Self-similar ATM Traffic,** in Vol 2bProc. ITC15, "Teletraffic Contributions for the information Age" (eds. V Ramaswami and P.E. Wirth), Elseveier, 1997, pp1409-1420.

[MON97a] Mondragón, R.J., and Arrowsmith, D.K. **Tracking unstable fixed points in parametrically dynamic systems**. Phys. Lett. A, Vol.229, No.2. 1997.

[MON97b] Mondragón, R.J., and Arrowsmith, D.K., **On Control of Coupled Map Lattices: Using local dynamics to predict controllability.** Int. Jour of Bif and Chaos. 7:No 2,1997, 383-399.

[MON98] Mondragón, R.J., **A New Model of Packet Traffic Using Chaotic Maps**, Preprint Dept of Elec Eng. QMW, Aug. 98.

[MUL95] Mullin, T (Ed.) **The Nature of Chaos**, Clarendon Press, Oxford, 1995.

[NAR98] Naranyan O., **Exact Asymptotic Queue Length Distribution for Fractional Browninan Traffic.** Advances in Performance Analysis, Vol. No.1, 1998, pp39-63.

[NOR93] Norros I., **Studies on a model for connectionless traffic, based on fractional Brownian motion.** Conf. On Applied Probability in Engineering, Computer and Communication Sciences, Paris, June 16-18 1993.

[NOR95] Norros, I. **The Management of Large Flows of Connectionless Traffic on the Basis of Self-Similar Modelling**, IEEE ICC95 Seattle, Washington, 1995, pp451-455.

[OCO] O'Connell N. **Large deviations in queueing networks,** DIAS Technical Report, DIAS-APE-94-13.

[OKE95] Oketani, N., Ushio, T.and Hirai, K., **Decentralised control of chaos in nonlinear networks**, Phys. Lett. A, 198, 1995, pp 327-332

[OTT90] Ott, E. Grebogi, C. Yorke, A.J. **Controlling Chaos"** Physical Review Letters, Vol. 64, No. 11, March 1990, pp1196-1199.

[OTT94] Ott E., **Chaos in Dynamical Systems.** Cambridge University Press, Cambridge, 1994.

[PAX95] Paxson V, Floyd S. **Wide Area Traffic: The failure of Poisson Modelling.** IEEE/ACM Trans on Networking, Vol 3, No 3, June 1995, pp226-244.

[PER94] Percival I., and Richards |D., **Introduction to Dynamics.** Cambridge University Press, Cambridge, 1994.

[PITT96] Pitts J.M, and Schormans J A, **Introduction to ATM Design and Performance,** John Wiley and Sons, Chichester, 1996.

[POM80] Pomeau, Y. and Manneville, P. **Intermittent Transition to Turbulence in Dissipative Dynamical Systems**, Commun. Math. Phys. Vol 74, 1980, pp189-197.

[PRE94] Press W.H., Teukolsky S.A., Vetterling W. T. and Falnnery B. P., **Numerical Recipes in C**, Cambridge University Press, Cambridge, 1994.

[PRI94] Priestley M.B., **Spectral Analysis and Time Series.** Academic Press, London, 1994.

[PRO83] Procaccia, I. And Schuster, H., **Functional renomalization-group theory of universal 1/f noise in dynamical systems**. Physics Review A, Vol 28, No.2., Aug. 83, pp1210-1212.

[PRU95a] Pruthi, P. **An Application of Chaotic Maps to Packet Traffic Modeling**, PhD Dissertation, Royal Institute of Technology, Sweden, ISRN KTH/IT/R--95/19--SE, 1995.

[PRU95b] Pruthi. P, Erramilli A. **Heavy-Tailed ON/OFF Source Behaviour and Self-Similar Traffic**. Proc ICC 95 June 1995.

[RAM96] Ramalho M., **Application of an Automatically Designed Fuzzy Logic Decision Support System to Connection Admission Control in ATM Networks**, PhD Thesis, Queen Mary and Westfield College, University of London, December 1996.

[ROM] Romerias, F.J. Grebogi, C. Ott, E.. Dayawansa , W.P.. **Controlling Chaotic Dynamical Systems**, Physica D, Vol 58, pp1165-192.

[ROS87] Ross. S.M., **Introduction to Probability Models**, 4th. Ed., Academic Press, Boston, 1987.

[SAI94] Saito, H., **Teletraffic Technologies in ATM Networks**. Artech House, Boston.,1994.

[SLI95] Slimane B, Le-Ngoc T. **A Doubly Stochastic Poisson Model for Self-Similar Traffic**, IEEE ICC95 Seattle, Washington, pp456-460.

[SEP93] Sepulchre, J. and Babloyantz, A. , **Controlling chaos in a network of oscillators,** Physical Review E, 48 (2), 1993, 945-950

[SHO80] Shoch J. F, and Hupp J. A. **Measured performance of an Ethernet local network**. Cammun. ACM, vol 23 No. 12, pp711-721 Dec. 1980.

[SCHO93] Schormans, J.A., Scharf, E.M. and Pitts, J.M. **Waiting time probabilities in a statistical multiplexer with priorities**, IEE Proc. I, vol.140, No. 4, 1993, pp301-307.

[SCHO96] Schormans, J.A., Pitts, J.M., Clements B.R. and Scharf, E.M. **Approximation to M/D/1 for ATM CAC, buffer dimensioning and cell loss performance**, Electronics Letters, Vol. 32, No. 3, pp164 -165.

[SCHO97] Schormans, J., Azmoodeh, M., Gordhan, S. and Davison, R., **Buffer Partitioning Formula for Different Service Classes of ATM Traffic**, submitted to IEE Proc. Comms, 1997.

[SCHR91] Schroeder, M., **Fractals, Chaos, Power Laws, Minutes form an Infinite Paradise**, W.H. Freeman, New York, 1991.

[SCHU95] Schuster. H.G., **Deterministic Chaos an Introduction 3rd. Ed**. VCH Verlagsgesellschaft, Weinheim Germany, 1995.

[TAQ86] Taqqu M. S. and Levy J.B. **Using Renewal Processes to Generate Long Range Dependence and High Variability**, Dependence in Probability and Statistics, Progress in Prob. and Vol 11, Birkhauser, Boston, 1986, pp 73-89.

[TAQ97] Taqqu M.S., Willinger W. and Sherman R., **Proof of a Fundamental Result in Self-similar Traffic Modelling**, ACM Computer Communications Review, Vol 1997 pp5-23.

[VAT98] Vaton S, Moulines E., **A locally stationary Semi-Markovian Representation for Ethernet LAN Traffic Data**, in Proc. IFIP BC'98, (1998) "Broadband Communications The future of telecommunications" (eds. P.J. Kühn and R.Ulrich), Chapman & Hall, pp525-535.

[VEI92] Veitch D., **Novel Models of Broadband Traffic**, 7th Australian Teletraffic Research Seminar, Murray River, Australia, 1992.

[WIL94] Willinger W., **"Traffic Modelling of High-Speed Networks: Theory and Practice",** Stochastic Networks, Kelly and Williams (Eds.), Springer-Verlag, New York, 1994.

[WIL95] Willinger W., Taqqu M.S., Sherman R. and Wilson D.V., **Self-similarity through high variablility: Statistical analysis of ethernet LAN traffic at the source level**. Proc.ACM Sigcomm '95. Cambridge MA 1995, pp 100-113.

[WIL97] Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V. **Self-Similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level**. EEE/ACM Transactions on networking, Vol. 5, No. 1, 1997, pp71-86.

[WILS74] Wilson K.G. and Kogut J.B., **The Renormalization Group and the ε Expansion**, Phys Rep. Vol 12C No 2, August 1974, pp 77-199.

[WILS75] Wilson K.G., **The renormalization group: Critical phenomena and the Kondo problem**, Rev. Mod. Phys. Vol 47, No. 4, October 1975, pp 773-839.

[YAN94] Yang H.L. and Ding E.J., **On-off intermittency in random map lattices**, Physical Review E, Vol. 50, No.5. ppR3295-R3298, Nov. 94.

[YOU95] Youssefmir, M. and Huberman, B., **Local control for large assemblies of nonlinear elements,** xyz.lan.gov, prepreint , 1995.

[ZUM93] Zumofen G, Klafter J and Blumen A**, Lévy walks and propagators in intermittent chaotic systems**, Phyical Review E, Vol 47, No 3 , March 1993, pp 2183-2186.

Appendices

Appendix A - Covariance Structure of Self-similar Processes

In order to get from equ.(3.28) to equ.(3.35) in more detail we perform the following. Because of stationarity, the lag is the important criterion therefore we can make the following simplification:

$$g(k) = \mathrm{cov}\left(X_t, X_{t+k}\right) = \mathrm{cov}\left(X_1, X_{1+k}\right) = E\left[X_t X_{t+k}\right].$$

Now if Y_t process has a zero mean then we can also say that its incremental process X_t will also have a zero mean. This helps greatly, since we can write the product term with the equivalent expansion

$$\begin{aligned} X_t X_{t+k} &= X_1 X_{1+k} \\ &= \left(\sum_{j=1}^{k+1} X_j\right)^2 + \left(\sum_{j=2}^{k} X_j\right)^2 - \left(\sum_{j=1}^{k} X_j\right)^2 - \left(\sum_{j=2}^{k+1} X_j\right)^2 \end{aligned} \tag{A.1}$$

We can now write the covariance using this expansion

$$\begin{aligned} \mathrm{cov}\left(X_1, X_{1+k}\right) &= \frac{1}{2} E\left\{\left(\sum_{j=1}^{k+1} X_j\right)^2 + \left(\sum_{j=2}^{k} X_j\right)^2 - \left(\sum_{j=1}^{k} X_j\right)^2 - \left(\sum_{j=2}^{k+1} X_j\right)^2\right\}. \\ &= E\left[X_1 X_{1+k}\right] \end{aligned} \tag{A.2}$$

If we expand out the expansion term by term we see that we are left with the following equivalent terms

$$\left(\sum_{j=1}^{k+1} X_j\right)^2 = Y_{k+1} - Y_0$$

$$\left(\sum_{j=2}^{k} X_j\right)^2 = Y_{k-1} - Y_0$$

$$\left(\sum_{j=1}^{k} X_j\right)^2 = \left(\sum_{j=2}^{k+1} X_j\right)^2 = Y_k - Y_0$$

The covariance can now be written with the inclusion of the equivalent terms

$$\mathrm{cov}\left(X_t, X_{t+k}\right) = \frac{1}{2}\left\{E\left[\left(Y_{k+1} - Y_0\right)^2\right] + E\left[\left(Y_{k-1} - Y_0\right)^2\right] - E\left[\left(Y_k - Y_0\right)^2\right] - E\left[\left(Y_k - Y_0\right)^2\right]\right\}. \tag{A.3}$$

This is the covariance for incremental process X_t. We now apply the following result

$E\left[Y_{t-s}^2\right] = (t-s)^{2H} E\left[Y_1^2\right] = (t-s)^{2H} s^2$ and obtain

$$\mathrm{cov}\left(X_t, X_{t+k}\right) = \frac{s}{2}\left\{(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right\} = g(k).$$

(A.4)

The correlation *r(k)* comes from the standard result $r(k) = g(k)/g(0)$ and since $g(0) = s$ then

$$r(k) = \frac{1}{2}\left\{(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right\}.$$

(A.5)

What remains now is to determine the asymptotic behaviour of *r(k)* as $k \to \infty$. To be able to do this we apply Taylor's expansion to the function *r(k)*. To do this effectively we rewrite equ (A.5) in the following way.

$$r(k) = \frac{k^{2H}}{2}\left\{\left(1+\tfrac{1}{k}\right)^{2H} - 2 + \left(1-\tfrac{1}{k}\right)^{2H}\right\}.$$

(A.6)

We then define the function *g(.)* as

$$g(x) \equiv (1+x)^{2H} - 2 + (1-x)^{2H}.$$

(A.7)

The correlation can then be written in terms of the function *g(.)*

$$r(k) = \frac{k^{2H}}{2} g\left(\frac{1}{k}\right).$$

(A.8)

Using Taylor's expansion we find the dominant terms and determine the functions asymptotic behaviour[32]. The correlation can be written then as

$$\begin{aligned} r(k) = \frac{k^{2H}}{2}\Big\{&\Big[(1+x)^{2H} - 2 + (1-x)^{2H}\Big] + \\ &x\Big[2H\left\{(1+x)^{2H-1} - (1-x)^{2H-1}\right\}\Big] + \\ &x^2\Big[2H(2H-1)\left\{(1+x)^{2H-2} - (1-x)^{2H-2}\right\}\Big] + \\ &x^2\Big[2H(2H-1)(2H-2)\left\{(1+x)^{2H-3} - (1-x)^{2H-3}\right\}\Big] + \cdots\Big\} \end{aligned}$$

(A.9)

Evaluating the differential terms at *x*=0 yields $r(k) = \frac{k^{2H}}{2}\left\{x^2 2H(2H-1) + H.O.T\right\}$, which is the first non vanishing term plus higher order terms. Since *x = 1/k* then these higher order terms vanish rapidly as $k \to \infty$ and the asymptotic behaviour of the correlation is then

[32] Recall that Taylor's expansion is :

$$g(x) = g(0) + xg'(0) + \frac{x2g''(0)}{2!} + \frac{x^3 g'''(0)}{3!} + \cdots + \frac{x^n g^n(0)}{n!} + \cdots$$

and the first three derivatives are given as

$$g'(x) = 2H(1+x)^{2H-1} - 2H(1-x)^{2H-1}$$

$$g''(x) = 2H(2H-1)(1+x)^{2H-2} - 2H(2H-1)(1-x)^{2H-2}$$

$$g'''(x) = 2H(2H-1)(2H-2)\left\{(1+x)^{2H-3} - (1-x)^{2H-3}\right\}.$$

$$r(k) \approx H(2H-1)k^{2H}\left(\frac{1}{k}\right)^2 .$$

(A.10)

Appendix B - Dirac Impulse Function

The Dirac Impulse function is defined as

$$d(t) = \begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases}.$$

(A.11)

For convenience the *d(t)* is normalised such that

$$\int_{-\infty}^{+\infty} d(t)dt = 1 .$$

We now introduce two further identities that use this basic definition:

$$\int_{-\infty}^{+\infty} f(t)d(t-a)dt = f(a)$$

(A.12)

and

$$d[r(t)] = \sum_n \frac{d(t-t_n)}{\left|\frac{dr(t_n)}{dt}\right|} .$$

(A.13)

What equ.(A.12) states is that the function *f(t)* is only switched on at the pulse position given by *d(t-a)* and can only take on a value at the time of activation i.e. when $t = a$. We can view equ.(A.13) in the following manner. If *d(t)* is an impulse at the origin and $d(t-t_n)$ describes a pulse shifted from the origin, or alternatively a pulse at t_n, then *d[r(t)]* describes a series of pulses in which the position of the pulse is being given by the roots of the function *r(t)*. To illustrate this consider the following example. Suppose we have a function *r(t)*which has the following form $r(t) = t^2 - 1$. This has roots at $(t-1)(t+1) = 0$. equ.(A.13) then becomes:

$$d[t^2 - 1] = \sum_{n=1}^{N=2} \frac{d(t-t_n)}{\left|\frac{dr(t_n)}{dt}\right|} .$$

(A.14)

We note that *N*=2 is since there are two roots which will yield an impulse. The first differential of *r(t)* is simply $\frac{dr(t)}{dt} = 2t$. Therefore the full expansion of equ.(A.14) becomes

$$d[t^2 - 1] = \frac{d(t-t_1)}{|2t_1|} + \frac{d(t-t_2)}{|2t_2|} = \frac{1}{2}(d(t+1) + d(t-1)) .$$

(A.15)

Appendix C - Transit-Time Analysis

Trial constant values for *a*, *b* and *c*.

Trial value for constant a.

At this point it may be useful to recall the following Gamma Function identities

$\Gamma(n+z) = (n-1+z)(n-2+z)\cdots(1+z)z\Gamma(z)$,

$\Gamma(1+z) = z\Gamma(z) = z!$,

$\Gamma(1) = 1$.

There is an obvious choice for the constant a. Since there is a factorial term in the RHS denominator of equ.(5.14) which is absent from the LHS then a natural choice for a is a=1, since use of the Gamma identity $\Gamma(1+z)$ with a=1 will cancel out the factorial term occurs.

Trial values for constants b and c.

To obtain trial values for b and c we note that

$$\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1} = \frac{\Gamma(c)}{\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(b+j)}{\Gamma(c+j)} z^j \,. \tag{A.16}$$

By using the Gamma function identities we expand out the $\Gamma(1+j)$ terms

$$\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1} = \frac{\Gamma(c)}{\Gamma(b)} \sum_{j=0}^{\infty} \frac{(j-1+b)(j-2+b)(j-3+b)\cdots(1+b)b\Gamma(b)}{(j-1+c)(j-2+c)(j-3+c)\cdots(1+c)c\Gamma(c)} z^j \,. \tag{A.17}$$

The $\Gamma(b)$ and $\Gamma(c)$ are common for all summation terms in the RHS and can therefore be cancelled with the $\Gamma(.)$ outside the summation and we can therefore write

$$\sum_{j=0}^{\infty} \frac{(-A)^j}{mj+1} = \sum_{j=0}^{\infty} \frac{(j-1+b)(j-2+b)(j-3+b)\cdots(1+b)b}{(j-1+c)(j-2+c)(j-3+c)\cdots(1+c)c} z^j \,. \tag{A.18}$$

Furthermore, if we assume some favourable cancellation occurs in the numerator and denominator of the RHS and by rewriting the numerator and denominator of the LHS then,

$$\sum_{j=0}^{\infty} \frac{\left(\frac{1}{m}\right)}{\left(\frac{mj+1}{m}\right)} (-A)^j = \sum_{j=0}^{\infty} \frac{b}{(j-1+c)} z^j \,. \tag{A.19}$$

Equating numerator and denominator of both sides yields the following result

$$b = \frac{1}{m}$$
$$c = \frac{1+m}{m} \,. \tag{A.20}$$

Therefore we can say that

$$\sum_{j=0}^{\infty}\frac{(-A)^j}{mj+1}={}_2F_1\left(1,\tfrac{1}{m};\tfrac{m+1}{m};z\right).$$

(A.21)

We know this to be true since substitution of our trial values in to our equations recovers the original series representation. Therefore we can write our original integral given by equ.(5.10) as

$$L_{m\in(1,2)}=mk(m)\left\{A_{out}{}^{\frac{1}{m}}\,{}_2F_1\left(1,\tfrac{1}{m};\tfrac{1+m}{m};z_{out}\right)-A_{in}{}^{\frac{1}{m}}\,{}_2F_1\left(1,\tfrac{1}{m};\tfrac{1+m}{m};z_{in}\right)\right\}$$

(A.22)

where $A_n=\frac{c}{e}y_n^m$, $Z_n=-A_n$, $k(m)=\frac{e^{-\frac{(m-1)}{m}}}{mc^{\frac{1}{m}}}$, $c=\frac{1-e-d}{d^m}$ and ${}_2F_1(a,b;c;z)$ is a hypergeometric function

Recovery of Upper and Lower bounds from the Series Formulation

As a check we introduce our trial values for a, b, and c into our original equation and see if we recover our original series

$$\begin{aligned}{}_2F_1\left(1,\tfrac{1}{m};\tfrac{1+m}{m};z\right)&=\frac{\Gamma\left(\frac{1+m}{m}\right)}{\Gamma(1)\Gamma\left(\frac{1}{m}\right)}\sum_{n=0}^{\infty}\frac{\Gamma(1+n)\Gamma\left(\frac{1}{m}+n\right)}{\Gamma\left(\frac{1+m}{m}+n\right)}\frac{z^n}{n!}\\&=\frac{\Gamma\left(\frac{1+m}{m}\right)}{\Gamma\left(\frac{1}{m}\right)}\sum_{n=0}^{\infty}\frac{\left(n-1+\frac{1}{m}\right)\left(n-2+\frac{1}{m}\right)\cdots\left(1+\frac{1}{m}\right)\frac{1}{m}\Gamma\left(\frac{1}{m}\right)}{\left(n-1+\frac{1+m}{m}\right)\left(n-2+\frac{1+m}{m}\right)\cdots\left(1+\frac{1+m}{m}\right)\frac{1+m}{m}\Gamma\left(\frac{1+m}{m}\right)}z^n\\&=\sum_{n=0}^{\infty}\frac{1}{(mn+1)}z^n\end{aligned}$$

(A.23)

Which is the original series. As a further set of checks we can insert appropriate values of *m* and recover the following hypergeometirc relations:

for *m*=1 we obtain ${}_2F_1(1,1;2;-z)=\frac{1}{z}\ln(1-z)$,

(A.24)

and for *m*=2 ${}_2F_1\left(1,\tfrac{1}{2};\tfrac{3}{2};-z^2\right)=\frac{1}{z}\arctan(z)$.

(A.25)

Intermittency Map Variance Structures

This section relates to the formulation of the variance structure for the intermittency family of maps. Ranging from the Bernoulli shift map to the Double intermittency map. What is required to show is that the decay of the correlation structure and hence *H* can be determined in terms of the map behaviours, i.e. through the statistical analysis of the maps in particular its variance. There are three equations that are linked to the variance that we will show derivations for: equ.(5.44), equ.(5.45) and equ.(5.49). Additionally we need to show that equ.(5.44) and equ.(5.45) are functionally the same.

Derivation of equ.(5.44)

We first define Z_{K_y} as the total number of cells generated after *k* iterations using the indicator variable interpretation y, i.e.

$$y_n(x_n) = \begin{cases} 0, & 0 < x_n \le d \\ 1, & d < x_n \le 1 \end{cases}$$

$$Z_{K_y} \equiv \sum_{i=1}^{K} y_i \ .$$

(A.26)

The average number of packets is then given by

$$E\left(Z_{K_y}\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} Z_{nK_y}$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{i=1}^{K} y_i \right) .$$

(A.27)

The *N* trials of can be viewed as independent and therefore the order of summation can be reversed

$$E\left(Z_{K_y}\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{K} \left(\sum_{n=1}^{N} y_n \right).$$

(A.28)

This limit is related to the invariant density of the map via the following relationship

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} y_n = l = \int_{d}^{1} r(x)dx$$

(A.29)

which is the average number of emissions of the map (mean number of cells generated). We can therefore say that

$$E\left(Z_{K_y}\right) = Kl \ .$$

(A.30)

The variance can be written in the traditional manner

$$\mathrm{var}\left(Z_k\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{j=1}^{K} y_j - Kl \right)^2 .$$

(A.31)

Expanding out the bracketed terms we arrive at

$$\mathrm{var}\left(Z_{K_y}\right) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{K} y_j \sum_{j=1}^{K} y_j - 2Kl \sum_{j=1}^{K} y_j + (Kl)^2 \right).$$

(A.32)

We now examine the terms with in the brackets of equ.(A.32)

$$\sum_{j=1}^{K} y_j \sum_{j=1}^{K} y_j = \left(y_1 + y_2 + \cdots + y_K\right)\left(y_1 + y_2 + \cdots + y_K\right) \\ = \left(y_1^2 + y_2^2 + \cdots + y_K^2\right) + 2\left(y_1 y_2 + y_1 y_3 + \cdots + y_{K-1} y_K\right)$$

(A.33)

Since by the definition of y then $y_i y_j = y_j y_i$ and the above reduces to

$$\sum_{j=1}^{K} y_j \sum_{j=1}^{K} y_j = \sum_{i=1}^{K} {y_i}^2 + 2\sum_{i=1}^{K}\sum_{j>i}^{K} y_i y_j \,.$$

(A.34)

We can now introduce the limits to this term

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{j=1}^{K} y_j \sum_{j=1}^{K} y_j \right) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{i=1}^{K} {y_i}^2 + 2\sum_{i=1}^{K}\sum_{j>i}^{K} y_i y_j \right) \\ = Kl \; + 2\sum_{i=1}^{K}\sum_{j>i}^{K} E\left(y_i y_j\right)$$

(A.35)

Examining the second term and introducing the limits

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \left(2Kl \sum_{j=1}^{K} y_j \right) = \sum_{j=1}^{K} \left(2Kl \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} y_n \right). \\ = 2\left(Kl\right)^2$$

(A.36)

The final term after introducing limits is simply

$$\lim \frac{1}{N} \sum_{n=1}^{N} \left(Kl\right)^2 = \left(Kl\right)^2 .$$

(A.37)

The variance in equ.(A.32) can then be written as

$$\mathrm{var}\left(Z_{K_y}\right) = Kl\left(1 - Kl\right) + 2\sum_{i=1}^{K}\sum_{j>i}^{K} E\left(y_i y_j\right).$$

(A.38)

This is the same as equ.(5.44).

Derivation of equ.(5.45)

To show that equ.(5.44) and equ.(5.45) are the same we proceed as follows. We begin by examining the correlation term of equ.(5.44) i.e. the $E\left(y_i y_j\right)$ term. Since the y_i can only take on values of 1 or 0 (see definition of y_i) then the expectation can only take on non-zero values when y_i and y_j =1, i.e. $E\left(y_i y_j\right) = P\left\{y_j = 1, y_i = 1\right\}$. Therefore if we consider the expectations in terms of probabilities and use the probability identity $P\left(A,B\right) = P\left(B\right)P\left(A|B\right)$.

The expectation ca be written as $E(y_i y_j) = P\{y_j = 1, y_i = 1\} = P\{y_i = 1\}P\{y_j = 1|y_i = 1\}$ and since $P\{y_i = 1\}$ is just the probability of being on, λ. Then we have the following relationship $E(y_i y_j) = l \; P\{y_j = 1|y_i = 1\}$. We can now write equ.(13) in terms of this relationship

$$\text{var}(Z_{K_y}) = Kl(1-l) + 2l\sum_{i=1}^{K}\sum_{j>i}^{K} P\{y_j = 1|y_i = 1\}$$

(A.39)

If the summation term is expanded out we obtain the following pattern (the case of *K*=4 is shown)

$$\sum_{i=1}^{K=4}\sum_{j>i}^{K=4} P\{y_j = 1|y_i = 1\} = P\{y_2 = 1|y_1 = 1\} + P\{y_3 = 1|y_2 = 1\} + P\{y_4 = 1|y_3 = 1\} +$$
$$P\{y_3 = 1|y_1 = 1\} + P\{y_4 = 1|y_2 = 1\} + \quad .$$
$$P\{y_4 = 1|y_1 = 1\}$$
$$\text{i=1} \qquad \text{i=2} \qquad \text{i=3}$$

(A.40)

The above summation has been arranged in a matrix formation so that the probabilities for a given i are arranged in columns and that probabilities where the difference between i and j are the same form the rows . If we now define the following:

$$C(n = j - i) \equiv P\{y_j = 1|y_i = 1\}.$$

(A.41)

We can therefore write the summation given above as

$$\sum_{i=1}^{K=4}\sum_{j>i}^{K=4} P\{y_j = 1|y_i = 1\} = C(1) + C(1) + C(1) +$$
$$C(2) + C(2) + \quad .$$
$$C(3)$$
$$\text{i=1} \quad \text{i=2} \quad \text{i=3}$$

(A.42)

This RHS forms a series in terms of $C(i)$ and there for we have an equivalent summation

$$\sum_{i=1}^{K=4}\sum_{j>i}^{K=4} P\{y_j = 1|y_i = 1\} = \sum_{i=1}^{K-1}(K-i)C(i).$$

(A.43)

The variance can therefore be written

$$\text{var}(Z_{K_y}) = Kl(1-Kl) + 2l\sum_{i=1}^{K-1}(K-i)C(i).$$

(A.44)

This is the same as equ.(5.45).

Derivation of equ.(5.49)

To derive euq.(5.49) we have to use the ansatz $C(i) \approx a\, i^{\,b} + l$. We can write equ.(A.44). in terms of the ansatz

$$\mathrm{var}\left(Z_{K_y}\right) \approx Kl\left(1-Kl\right) + 2l\sum_{i=1}^{K-1}\left\{(K-i)\left(a\, i^{\,b} + l\right)\right\}. \tag{A.45}$$

Expanding out the bracketed term and splitting the summations we obtain

$$\mathrm{var}\left(Z_{K_y}\right) \approx Kl\left(1-Kl\right) + 2lKa\sum_{i=1}^{K-1} i^{\,b} + 2l^2K\sum_{i=1}^{K-1}1 - 2l\,a\sum_{i=1}^{K-1} i^{\,b+1} - 2l^2\sum_{i=1}^{K-1} i\,. \tag{A.46}$$

In order to simplify equ.(A.46) we use the following series identities

$$\sum_{i=1}^{K} i = \frac{K(K+1)}{2};\quad \sum_{i=1}^{K} Kl = K^2 l\,. \tag{A.47}$$

We apply the above to each the terms in equ. .(A.46)

$$\begin{aligned} 2lKa\sum_{i=1}^{K-1} i^{\,b} - 2l\,a\sum_{i=1}^{K-1} i^{\,b+1} &= 2Ka\left(1^b + 2^b + \cdots + (K-1)^b\right) - 2l\,a\left(1.1b + 2.2b + \cdots + (K-1)(K-1)^b\right) \\ &= 2l\,a\left(1^b(K-1) + 2^b(K-2) + \cdots + (K-1)^b\left(K-(K-1)\right)\right) \\ &= 2l\,a\sum_{i=1}^{K-1} i^b(K-i) \end{aligned} \tag{A.48}$$

$$\begin{aligned} 2l^{\,2}K\sum_{i=1}^{K-1}1 - 2l^2\sum_{i=1}^{K-1} i &= 2l^2K(K-1) - l^2K(K-1) \\ &= l^2K(K-1) \end{aligned} \tag{A.49}$$

The variance can then be written as

$$\begin{aligned} \mathrm{var}\left(Z_{K_y}\right) &\approx Kl\left(1-Kl\right) + Kl^2\left(K-1\right) + 2l\,a\sum_{i=1}^{K-1} i^b\left(K-i\right) \\ &\approx Kl\left(1-l\right) + 2l\,a\sum_{i=1}^{K-1} i^b\left(K-i\right) \end{aligned} \tag{A.50}$$

Equ.(A.50) can be analysed in terms of an inequality relationship. The summation term can be replaced by a continuous integration, and since they are equal only in the limit as $\Delta i \to 0$ then we can form the following inequality

$$\mathrm{var}\left(Z_{K_y}\right) \approx Kl\left(1-l\right) + 2l\,a\sum_{i=1}^{K-1} i^b\left(K-i\right) \le Kl\left(1-l\right) + 2l\,a\int_{1}^{K-1} i^b\left(K-i\right)di \quad . \tag{A.51}$$

Carrying out the integration we obtain

$$\mathrm{var}\left(Z_{K_y}\right) \approx Kl(1-l) + 2la\sum_{i=1}^{K-1} i^b (K-i) \le Kl(1-l) + 2la\left\{\frac{K(K-1)^{b+1} - K}{b+1} - \frac{(K-1)^{b+2} - 1}{b+2}\right\} \quad .$$

(A.52)

If we let K$\rightarrow \infty$ we can see that the RHS becomes

$$\mathrm{var}\left(Z_{K_y}\right) \le Kl(1-l) + 2la\left\{\frac{K^{b+2} - K}{b+1} - \frac{K^{b+2} - 1}{b+2}\right\} \quad .$$

(A.53)

This is the same as equ.(5.49).