

# Entropy of Hidden Markov Processes and Connections to Dynamical Systems

## BIRS 5-Day Workshop, 07w5103

Brian Marcus (University of British Columbia)  
Karl Petersen (University of North Carolina-Chapel Hill)  
Tsachy Weissman (Stanford University)

Sept. 30 – Oct. 5, 2007

## 1 Workshop Overview

The focus of this workshop was entropy rate of Hidden Markov Processes (HMP)'s, related information-theoretic quantities and other connections with related subjects. The workshop brought together thirty mathematicians, computer scientists and electrical engineers from institutions in Canada, US, Europe, Latin America and Asia. While some participants were from industrial research organizations, such as Hewlett-Packard Labs and Philips Research, most were from universities, representing all academic ranks, including three postdocs and four graduate students.

The participants came from a wide variety of academic disciplines, including information theory (source, channel and constrained coding), dynamical systems (symbolic dynamics, iterated function systems and Lyapunov exponents), probability theory, and statistical mechanics.

There were eighteen 50-minute lectures from Monday through Friday (see section 4 and the Appendices). This left plenty of time for spirited, yet informal, interaction at breaks, over meals, at a Sunday evening reception and a Wednesday afternoon hike. In addition, there were two problem sessions, one on Tuesday evening to collect and formulate open problems and one on Thursday evening to discuss approaches and possible solutions.

The workshop introduced and improved contacts among groups of experts previously working apart, with different languages and approaches. New joint work is now in progress.

## 2 Overview of the Subject

An HMP is a stochastic process obtained as the noisy observation sequence of a Markov chain. A simple example is a binary Markov chain observed in binary symmetric noise, i.e., each symbol (0 or 1) in a binary state sequence generated by a 2-state Markov chain is flipped with some small probability, independently in time.

Since any stochastic process can be approximated by a Markov process of some order, and since Markov processes are easy to describe, it is not surprising that HMP's are encountered in an enormous variety of applications involving phenomena observed in the presence of noise. These range from speech and optical character recognition, through target tracking, to biomolecular sequence analysis. HMP's are important objects of study, in their own right, in many areas of pure and applied mathematics including information theory, probability theory, and dynamical systems. An excellent survey on HMP's can be found in [10].

A central problem is computation of the Shannon entropy rate of an HMP. The entropy rate of a stationary stochastic process is, by the Shannon-McMillan-Breiman theorem, the asymptotic exponential growth rate of the effective number of different sequences that can be generated by the process. The entropy rate is a measure of randomness of a stochastic process and is the most fundamental notion in information theory. It is closely related to important quantities in statistical mechanics [28].

There is a very simple closed-form formula for the entropy rate of a finite-state Markov chain. But there is no such formula for entropy rate of HMP's, except in very special cases.

The problem of computing entropy rate for HMP's was addressed fifty years ago by Blackwell [6], who discovered an expression for the entropy rate of an HMP as the integral of a very simple integrand with respect to a typically very complicated (and usually singular) measure on a simplex. Shortly afterwards, Birch [5] discovered excellent general upper and lower bounds on the entropy rate. However, until recently, there had been very little progress, with only a few papers on the subject.

Closely related is the problem of computing the capacity of channels with memory. The capacity is defined as the maximum mutual information rate of input and output processes, over all possible input processes. For some channels, this amounts to maximizing the entropy rate of the output process, which would be an HMP if the input process were Markov.

This problem has received a good deal of attention from people working in many different areas, primarily information theory, dynamical systems, statistical mechanics, and probability theory. And in symbolic dynamics, there is the related problem of characterizing which mappings between symbolic dynamical systems always map Markov chains to Markov chains or always lift Markov chains to Markov chains [7]. [34].

### 3 Recent Developments

Recently there has been a rebirth of interest in information-theoretic aspects of HMP's and, in particular, in the computation of entropy rate:

- Application of ideas from filtering, de-noising and state estimation problems to estimation of entropy rate [23].
- New approaches to computing asymptotics of entropy rate as process parameters tend to extreme values [17], [23], [12] [36].
- Application of ideas from statistical mechanics [36], [8].
- Novel variations on earlier techniques [5] giving more efficient upper and lower approximations to entropy rate and bounds on the rates of convergence of the approximations [9], [13].
- Smoothness and analyticity of entropy rate of HMP's [12].
- Connections between Lyapunov exponents for random matrix products and entropy rate of HMP's [16], [17] (earlier, closely related papers on Lyapunov exponents are [3], [4], [25], [26], [29], [30]).
- Monte-Carlo estimation of entropy rate [1], [27]
- Optimization of mutual information rate and application to capacity of noisy channels with memory [16], [19], [33].
- New results on computation of capacity of channels with feedback [35].
- Connections with iterated function systems, a growing area of dynamical systems [2], [31] [32].

## 4 Overview of Lectures

For more content on the lectures, see the list of abstracts in Appendix 2. Most of the lecturers have provided links to their talks, which can be found at:

<http://www.math.unc.edu/Faculty/petersen/>

- Monday:

The first day focused on *tutorials*, aimed to bridge the language and viewpoint barriers among the diverse group of participants. Also included was a short pictorial history, honouring the “masters” who have made fundamental contributions: D. Blackwell, R. Bowen, H. Furstenberg, A. Markov, A. Lyapunov, C. Shannon, and B. Weiss.

Weissman, an information theorist, begin with an overview of entropy rate of HMP’s and connections to source and channel coding, filtering and de-noising. He prepared so much interesting material, that, by popular demand, he continued his lecture on Monday evening. Boyle, a symbolic dynamicist, followed with a survey of results and problems on mappings from one symbolic dynamical systems to another for which there exists a fully supported Markov chain on the range whose inverse image is fully supported Markov. Verbitsky, who works on the interface of statistical mechanics and symbolic dynamics, surveyed connections between hidden Markov chains and statistical mechanics, including the role of Gibbs states, in one dimension and higher dimensions. Finally, Juang, an electrical engineer, gave a survey of applications of HMP’s to speech recognition.

- Tuesday:

The first three talks focused on work related to *statistical mechanics*. Ugalde presented specific results relating symbolic dynamics, HMP’s and Gibbs states in one dimension; in particular, he gave sufficient conditions for an HMP to be a Gibbs state. Zuk followed by describing relations between HMP’s and the Ising model, as well as various methods in statistical mechanics that can be used to approximate entropy rate of HMP’s, e.g. how to compute higher-order derivatives, with respect to the underlying Markov chain parameters of entropy rate. Then Montanari presented many interesting results on entropy quantities in statistical mechanics.

The last two talks focused on *computing and approximating entropy rates*. Ordentlich described deterministic approaches, as opposed to Monte-Carlo simulation, for computing entropy. He focused on the precision-complexity trade-off for each approach. Cuff followed with results relating a rather specific HMP to a classical channel (Blackwell’s trapdoor channel), showing how results on the channel capacity shed light on entropy of the HMP.

- Wednesday:

Guo presented a framework for computing entropy rate of HMP’s using dynamic programming. Slomczynski followed with a survey on connections among HMP’s, iterated function systems and quantum systems.

- Thursday:

The first two talks focused on *analyticity* issues. The first, by Peres, was set in the more general context of Lyapunov exponents of random matrix products. Peres also described applications to computing Hausdorff dimension of fractal sets which are invariant for certain smooth dynamical systems. The second, by Han, focused on analyticity of entropy rate of HMP’s and explicit computations of Taylor series coefficients, generalizing techniques introduced in Zuk’s talk.

The last three talks focused on *computation of noisy channel capacity*, rather than entropy rate. Pfister led off with approach to computing derivatives of entropy rate of HMP’s and then made the transition to computing derivatives of channel capacity (more precisely derivatives of optimized mutual information rate over inputs of a fixed Markov order) for channels with

memory. Vontobel continued with results and questions regarding optimization of the output entropy rate, or mutual information rate, over all Markov chain input processes to a finite state channel. Finally, Jaquette presented results on entropy rate and noisy channel capacity for channels with hard constrained inputs.

- Friday:

Kavic discussed the general problem of channel capacity, for channels that allow *feedback*, highlighting a recent result showing that Markov chain inputs can achieve channel capacity (in contrast to the situation without feedback).

Finally, Pollicott presented intriguing theoretical results that yield *numerical schemes*, with very fast convergence, for computation of Lyapunov exponents.

## 5 Highlights of Problem Sessions

Most of the lectures included statements of open problems. In the first problem session (Tuesday evening), the open problems stated earlier in the lectures were organized and expanded. Discussion on these problems continued throughout the workshop, with special attention given in the second problem session (Thursday evening). Some partial results and even complete results were given in the second session.

- Calculation of asymptotics of entropy rate of HMP's.

In Zuk's talk, the problem was posed of computing the asymptotics in the "rate transitions" regime, i.e., the case of a binary symmetric Markov chain, in binary symmetric noise, with transition probability  $p$  from one state to the other as  $p$  tends to zero. At the second problem session, Quas presented a solution, with subsequent refinements by Peres; a paper on this result is in preparation.

- Estimates on decay rate of correlations for HMP's.

The decay rate of correlations (as conditioning length increases) is exponential. Estimates on the decay rate should have impact on the domain of analyticity of entropy rate. Verbitsky's lecture contained several estimates and a conjectured improved estimate, which was proven by Peres later in the workshop.

This is intimately connection to computation of best approximations to contraction factor for HMP's,

- Behaviour of measures under continuous factor maps  $\phi$  from one irreducible SFT to another.

Characterize those  $\phi$  that are Markovian, i.e., is the inverse image of some fully supported Markov chain is again fully supported Markov? There is a characterization in terms of functions, known as compensation functions, but it is not clear how to make this effective.

Characterize those  $\phi$  such that every fully supported Markov measure on the domain SFT maps to a Gibbs measure on the range. Chazottes and Ugalde have partial positive results. At the second problem session, Boyle gave a complete characterization in the special case that  $\phi$  is finite-to-one.

- Constructing special Markov chain covers of HMP's:

Any HMP can be realized from many different Markov chains. There does not appear to be a canonical Markov chain for a given HMP. It was proven long ago [21] that there need not be a covering Markov chain of the same entropy as the HMP. But in the second problem session, Peres gave a procedure for constructing covering Markov chains with entropy arbitrarily close to that of the given HMP. This procedure is conceptually simple, but the covering Markov chain is somewhat complicated and impractical. This raises the question of devising less complex alternative schemes.

- Complexity-Precision trade-offs for computing entropy rate of an HMP:

This topic was the major focus of Ordentlich's talk. The main problems are to determine the fundamentally optimum precision-complexity tradeoff as well as effective bounds on the approximation error as a function of the number of operations needed to construct the approximation.

Ordentlich described various approaches for deterministic approximations of the entropy rate. One approach is based on the classical Birch approximations and their more recent variable length improvements due to Egner et. al. [9]. For the former, bounds on the exponential rate of convergence of the approximations are known, but the exact rate is not known. For the latter, only empirical evidence was known.

Another approach is to approximate, via quantization, the stationary distribution of a simplex-valued Markov process which is closely related to Blackwell's measure and, hence, is closely linked to the entropy rate. This approach leads to an approximation procedure, described by Ordentlich, yielding the best known precision-complexity tradeoff.

- Biological models:

As described above, an HMP is the result of a Markov chain observed at the receiving end of a noisy channel. In this case, the channel is described by conditional probabilities of changing one symbol into another. In a biological setting, these changes could be viewed as "mutations." In addition to mutations, one could also allow insertions and deletions. The analysis of this type of model is important in biology, where, for instance, one might want to decide if two given genomes were derived from a common ancestor and, if so, what was the sequence of changes (mutations (M), insertions (I) and deletions (D)) that led to the two outcomes. More precisely, given two strings  $Y$  and  $Y'$  one could ask to find the most likely string of M's, I's and D's, which would transform a common ancestor of  $Y$  and  $Y'$  into  $Y$  and  $Y'$ . It was observed that even noisy channels which allow only deletions is difficult to analyze.

- Characterization of those stationary stochastic processes that have only countably many conditional measures.

Kitchens and Tuncel [20] characterized those stationary stochastic processes that have only finitely many conditional measures. The analogous countable case is of interest since if the process were an HMP, with countable support of the Blackwell measure, entropy rate can be expressed as a series expansion.

- Capacity of finite-state channels

Continuing on the thread of Pfister's and Vontobel's talks, the longstanding problem of computing the capacity of a noisy finite-state channel was discussed. One approach would, to some extent, depend on the question of whether a particular information rate function is concave as a function of the input distribution [33]. Computation of capacity for channels which allow feedback is better understood [35], but there are still open problems in that area.

- Capacity of noisy-constrained channels

Low-order asymptotics have been developed for capacity of binary symmetric channels whose inputs are constrained by a finite-type constraint [18], [14]. There was discussion of how to extend this work to higher order asymptotics and more complex channels.

- Ising models on trees and graphs: There was considerable discussion on the problem of extending results on entropy quantities to these kinds of models.

## 6 Appendix 1: Schedule

Sunday, Sept. 30:

4 PM: check In at Banff Centre

7:30 - 9:30 PM: Informal reception, Corbett Hall, 2nd floor lounge

Monday, Oct 1:

9-9:15: Introductions

9:15-10:15: T. Weissman (EE, Stanford): Overview of entropy rate of HMP's.

10:15 - 10:45: Coffee

10:45 - 11:45: M. Boyle (Math, Maryland): Overview of Markovian maps

Lunch

1:00 - 2:00 (optional) tour of Banff Centre

2:00 - 2:15 Group photo

2:15 - 2:30: B. Marcus (Math, UBC), K. Petersen (Math, UNC): a very brief history (in photos)

2:30 - 3:30: E. Verbitsky (Philips-Eindhoven): Thermodynamics of Hidden Processes

3:30 - 4:00: Coffee

4:00 - 5:00: B. H. Juang (ECE, Georgia Tech): Hidden Markov Model and its Application in Speech Recognition – A Tutorial.

Tuesday, Oct 2:

9:00 - 10:00: E. Ugalde (Math, Universida Autonoma de San Luis Potosi): On Gibbs measures and lumped Markov chains

10:00 - 10:30: Coffee

10:30 - 11:30: O. Zuk (Physics of Complex Systems, Weizmann Institute): HMP's Entropy Rate - Statistical Mechanics and Taylor Series Expansions

Lunch

1:30 - 2:30: A. Montanari (EE, Stanford): The rank of random band diagonal matrices in the Kac limit

2:30 - 3:30: E. Ordentlich (HP Labs - Palo Alto): Deterministic algorithms for computing/approximating the HMP entropy rate.

3:30 - 4:00: Coffee

4:00 - 5:00 P. Cuff (EE, Stanford): Entropy Rates of Hidden Markov Processes emerge from Blackwell's Trapdoor Channel

7:30 - 9:00 PM 1st Problem Session

Wednesday, Oct 3:

9:00 - 10:00: D. Guo (EECS, Northwestern): On The Entropy and Filtering of Hidden Markov Processes Observed Via Arbitrary Channels

10:00 - 10:30: Coffee

10:30 - 11:30: W. Slomczynski (Jagiellonian University): Entropy integral formula: from hidden Markov processes to quantum systems.

Afternoon: free, e.g., hike to Sulphur mountain

Thursday, Oct 4:

9:00 - 10:00: Y. Peres (Microsoft): Analyticity of Lyapunov exponents

10:00 - 10:30: Coffee

10:30 - 11:30: G. Han (Math, Hong Kong U.): Analyticity and Derivatives of entropy rate for HMP's

Lunch

1:30 - 2:30 H. Pfister (ECE, Texas A& M): The Derivatives of Entropy Rate and Capacity for Finite-State Channels

2:30 - 3:30 P. Vontobel (HP Labs - Palo Alto): Optimizing Information Rate Bounds for Channels with Memory

3:30 - 4:00 Coffee

4:00 - 5:00 P. Jacquet (INRIA): Entropy of HMP and asymptotics of noisy input-constrained channel capacity

7:30 - 9:00 PM 2nd Problem Session

Friday, Oct 5:

9:00 - 10:00: A. Kavcic (ECE, Hawaii): Markov and hidden Markov Processes in communication channels used with feedback

10:00 - 10:15: Coffee

10:15 - 11:15: M. Pollicott (Math, Warwick): Computing integrals, Lyapunov exponents and entropy using cycle expansions

12:00 Check Out of Banff Centre

## 7 Appendix 2: Abstracts

MONDAY:

- *T. Weissman (EE, Stanford):*

*Title: Overview of entropy rate of HMP's.*

Abstract: I will present an overview on the entropy rate of hidden Markov processes (HMP's), information and coding-theoretic motivation for its study, and some of its connections to dynamical systems, to non-linear filtering, and to statistical physics. Particular attention will be given to:

- Alternative representations: via the Blackwell measure, as a Lyapunov exponent, and as a partition function in statistical physics.
- Bounds and approximations (stochastic and deterministic), and their complexity-precision tradeoffs.
- Asymptotic regimes and analyticity.

- *M. Boyle (Math, Maryland):*

*Title: Overview of Markovian maps*

Abstract: A topological Markov shift is the support of a Markov chain (measure); that is, it is the set of infinite sequences all of whose finite subwords have strictly positive probability (measure). A topological Markov shift can support many different Markov chains, including higher-order chains (on which the past and future become independent after conditioning on finitely many steps in the past).

Now let  $f$  be a sliding block code from a topological Markov shift  $S$  onto another topological Markov shift  $T$ . We assume  $S$  is irreducible (it is the support of an irreducible/ergodic Markov chain).

Then there is a dichotomy: either every Markov measure on  $T$  lifts (via  $f$ ) to a Markov measure on  $S$ , or every Markov measure on  $T$  does not lift to a Markov measure on  $S$ . In the former case, the map  $f$  is called Markovian. The Markovian condition is a thermodynamic phenomenon and is the first of a range of conditions on the regularity of the map  $f$ . I will try to explain this condition, the related conditions, and related work due to myself, Petersen, Quas, Shin, Tuncel, Walters and others.

- *E. Verbitsky (Philips-Eindhoven):*

*Title: Thermodynamics of Hidden Processes*

Abstract: Hidden Markov processes have a number of very strong properties. I will argue that some of these properties can be explained using the language of Statistical Mechanics and Thermodynamic Formalism. For similarly defined hidden Markov random fields ( $d > 1$ ) the picture is much more complex. I will illustrate it with a number of examples and open questions.

- *B.H. Juang (ECE, Georgia Institute of Technology):*

*Title: Hidden Markov Model and its Application in Speech Recognition – A Tutorial.*

Abstract: Speech signals are produced everyday and are considered one physical-behavioral phenomenon that is most unique and intriguing. A speech signal carries a code that can be understood (decoded) by the listener but the same code may be realized as acoustic signals with vastly different physical properties. Variations across pitch, power level, prosodic manifests, and talker (including gender, the articulatory apparatus, etc.) are observed to be wide and broad. The hidden Markov model/process has been successfully developed as an effective modeling tool for this rather complex signal in several applications, most notably automatic speech recognition. In this talk, we present justifications for use of hidden Markov process for speech modeling, elaborate the mathematical development of such a tool over the past two decades, and discuss applications of this mathematical formalism in practical systems that are in use in our daily life.

## TUESDAY:

- *E. Ugalde (Math, Universida Autonoma de San Luis Potosi):*

*Title: On Gibbs measures and lumped Markov chains*

Abstract: Gibbs measures are fully determined by continuous functions or potentials, and admit a nice thermodynamic characterization. In the symbolic case, for finite type or sofic subshifts, they may arise as measures induced from Markovian ones, after amalgamation of symbols in the original alphabet. We have found sufficient conditions for the induced measure to be Gibbsian, and under those conditions we are able to determine the normalized potential of the measure. Even though in general the measure induced by amalgamation of symbols is not Gibbsian, an induced potential can be defined almost everywhere. The induced measure admits therefore a thermodynamic characterization (joint work with J.-R. Chazottes)

- *O. Zuk (Physics of Complex Systems, Weizmann Institute):*

*Title: HMP's Entropy Rate - Statistical Mechanics and Taylor Series Expansions*

Abstract: Hidden Markov Models are very similar to models encountered in statistical physics - specifically the Ising model in a random field. In this talk I will discuss the similarities and differences between the two models.

I will also describe our asymptotic results for the entropy rate and other quantities of interest in both models in various regimes, including an algorithm for calculating the Taylor series coefficients of the entropy rate.

I will end with numerical results and conjectures for the radius of convergence of the Taylor expansion. (based on joint work with E. Domany, I. Kanter and M. Aizenman).

- *A. Montanari (EE, Stanford):*

*Title: The rank of random band diagonal matrices in the Kac limit*

Abstract: Consider a stream of iid Bernoulli(1/2)  $x_t$ . At time  $t$ , you observe the mod 2 sum

$$y_t = x_{i_1}(t) + \dots + x_{i_k}(t)$$

where  $i_1, \dots, i_k$  are uniformly random in  $[i - R, i + R]$ , through an erasure channel. Further you know  $i_1, \dots, i_k$ . I provide several estimates of  $H(X^t|Y^t)/t$  in the large  $R$  limit. In physics this is known as the Kac limit after the seminal work of Marc Kac.

- *E. Ordentlich (HP-Labs, Palo Alto):*

*Title: Deterministic algorithms for computing/approximating the HMP entropy rate.*

Abstract: We survey known deterministic algorithms for approximating the entropy rate of hidden Markov models. We will consider the well known approach based on truncated conditional entropies as well as less studied approaches based on quantized likelihood processes. The various approaches will be compared on a complexity versus accuracy basis, to the extent that this tradeoff is known.

- *P. Cuff (EE, Stanford):*

*Title: Entropy Rates of Hidden Markov Processes emerge from Blackwell's Trapdoor Channel*

Abstract: Blackwell's trapdoor channel is a simple channel with memory that has been widely investigated during the past four decades. The non-feedback channel capacity has not been solved analytically, but we find the feedback capacity to be the logarithm of the golden ratio.

During the investigation of the trapdoor channel we find that, with the assistance of feedback, the channel can be transformed into an equivalent memoryless channel with a constrained input. A Markov source as input satisfies the constraints, but calculating the resulting mutual information requires finding the entropy rate of a Hidden Markov process.

In hindsight, after finding the capacity of the trapdoor channel with feedback, we recognize that we can express the entropy rate of a particular class of Hidden Markov processes in closed form. This is the class of Hidden Markov processes for which the conditional entropy rate of the states given the observations is zero, so the entropy rate is that of the joint states and observations, which is a Markov process. We will comment on relations between the HMP transition probabilities and satisfiability of said conditional entropy rate condition.

### WEDNESDAY:

- *D. Guo (EECS, Northwestern):*

*Title: On The Entropy and Filtering of Hidden Markov Processes Observed Via Arbitrary Channels*

Abstract: We study the entropy and filtering of hidden Markov processes (HMPs) which are discrete-time binary homogeneous Markov chains observed through an arbitrary memoryless channel. A fixed-point functional equation is derived for the stationary distribution of an input symbol conditioned on all past observations. While the existence of a solution to this equation is guaranteed by martingale theory, its uniqueness follows from contraction mapping property. In order to compute this distribution, the fixed-point functional equation is firstly converted to a linear system through quantization and then solved numerically using quadratic optimization. The entropy or differential entropy rate of the HMP can be computed in two ways: one by exploiting the average entropy of each input symbol conditioned on past observations, and the other by applying a relationship between the input-output mutual information and the stationary distribution obtained via filtering.

- *W. Slomeczynski (Jagiellonian University):*

*Title: Entropy integral formula: from hidden Markov processes to quantum systems.*

Abstract: We investigate the notion of dynamical entropy (or entropy rate) in the context of the statistical (or operational) approach to dynamical systems. In this approach we can

distinguish the kinematical and dynamical parts. In kinematics we define states of the system and observables, and in dynamics, the evolution operators describe changes of the state. Moreover, we describe mathematical objects depicting measurement procedures, called measurement instruments. This approach makes it possible to describe both classical and quantum phenomena by a single mathematical formalism. States are defined as the positive elements of the unit sphere in a certain ordered vector space and evolution operators as Markov (stochastic) operators in this space. By an equilibrium we mean a fixed point of a given Markov operator.

We present a method of computing entropy rate based on an integral formula. This method enables us to generalise some old formulae for dynamical entropy and to prove new ones, to work out numerical methods for computing entropy, and to investigate the basic properties of dynamical systems.

The reasoning leading to the proof of the integral formula is based on: attributing an iterated function system to each dynamical system and measurement instrument, investigating the properties of the iterated function system guaranteeing the existence and the uniqueness of an invariant measure, and justifying the integral formula using the properties of the iterated function system.

Integral formulae for entropy rate were previously shown in particular cases, where the state space was finite-dimensional, by David Blackwell for functions of a Markov chain and by Mark Fannes, Bruno Nachtergaele and Leo Slegers for the so-called algebraic measures. Here we present a unified approach to the problem and show general results utilising two techniques: the first uses the compactness of subsets of the state space in certain weak topologies, the second is based on employing the projective metric in the state space. Applying these methods, we obtain results concerning iterated function systems on the state space and dynamical entropy for many concrete state space types.

Applications of the integral formula include hidden Markov processes, kernel operators, Frobenius-Perron operators, and quantum systems (Srinivas-Pechukas-Beck-Graudenz entropy, quantum jumps, coherent states entropy).

#### THURSDAY:

- *Y. Peres (Microsoft):*

*Title: Analyticity of Lyapunov exponents*

Abstract: I will describe the relevance of the entropy of HMM and of Lyapunov exponents to determining dimension of slices and projections of fractals, then survey the analyticity of Lyapunov exponents via the polynomial approximation approach, the significance of obtaining explicit domains of analyticity and the Hilbert metric. Finally, I will work out a simple but illuminating example of integration with respect to coin tossing measures and determine a domain of analyticity there.

- *G. Han (Math, Hong Kong U.):*

*Title: Analyticity and Derivatives of entropy rate for HMP's*

Abstract: We prove that under mild positivity assumptions the entropy rate of a hidden Markov chain varies analytically as a function of the underlying Markov chain parameters. A general principle to determine the domain of analyticity is stated. We also show that under the positivity assumptions the hidden Markov chain itself varies analytically, in a strong sense, as a function of the underlying Markov chain parameters. For a natural class of hidden Markov chains called "Black Hole", we show that one can exactly compute any derivatives of entropy rate.

- *H. Pfister (ECE, Texas A& M):*

*Title: The Derivatives of Entropy Rate and Capacity for Finite-State Channels*

Abstract: This talk discusses a number of topics related to the entropy rate and capacity of finite-state channels. A simple formula is given for the derivative of the entropy rate and it is used to compute closed-form expansions for the channel capacity in the high noise regime. The relationship between this formula and previous results is discussed.

The derivative formula is then extended to the Lyapunov exponent of a sequence of random matrices. In particular, we discuss i.i.d., Markov, and hidden Markov matrix sequences. The last case is closely related to the derivative of the divergence between two hidden Markov processes.

The talk concludes with a short discussion of ergodic properties and mixing conditions of the forward Baum-Welch (a.k.a. BCJR) algorithm.

- *P. Vontobel (HP-Labs, Palo Alto):*

*Title: Optimizing Information Rate Bounds for Channels with Memory*

Abstract: We consider the problem of optimizing information rate upper and lower bounds for communication channels with (possibly large) memory. A recently proposed auxiliary-channel-based technique allows one to efficiently compute upper and lower bounds on the information rate of such channels.

Towards tightening these bounds, we propose iterative expectation-maximization (EM) type algorithms to optimize the parameters of the auxiliary finite-state machine channel (FSMC). We provide explicit solutions for optimizing the upper bound and the difference between the upper and the lower bound and a method for the optimization of the lower bound for data-controllable channels with memory. We discuss examples of channels with memory, for which application of the developed theory results in noticeably tighter information rate bounds.

Interestingly, from a channel coding perspective, optimizing the lower bound is related to increasing the achievable mismatched information rate, i.e. the information rate of a communication system where the maximum-likelihood decoder at the receiver is matched to the auxiliary channel and not to the true channel.

(This talk is based on joint work with Parastoo Sadeghi (ANU) and Ramtin Shams (ANU).)

- *P. Jacquet (INRIA):*

*Title: Entropy of HMP and asymptotics of noisy input-constrained channel capacity*

Abstract: In this talk, we consider the classical problem of noisy constrained capacity in the case of the binary symmetric channel (BSC), namely, the capacity of a BSC whose input is a sequence from a constrained set. We derive an asymptotic formula (when the noise parameter is small) for the entropy rate of a hidden Markov chain, observed when a Markov chain passes through a binary symmetric channel. Using this result we establish an asymptotic formula for the capacity of a binary symmetric channel with input process supported on an irreducible finite type constraint, as the noise parameter  $\epsilon$  tends to zero. For the  $(d, k)$ -Run Length Limited (RLL) constraint, we show that when  $k \leq 2d$ , the difference between the noisy capacity and noiseless capacity is  $O(\epsilon)$  and when  $k > 2d$ , it is  $O(\epsilon \log \epsilon)$  with explicitly computable constants (joint work with G. Han, B. Marcus, G. Seroussi, and W. Szpankowski).

#### FRIDAY:

- *A. Kavcic (ECE, Hawaii):*

*Title: Markov and hidden Markov Processes in communication channels used with feedback*

In this talk, we consider finite memory communications channels (finite-state channels, or state-space representable channels). Such channels are reasonably good models for magnetic and optical data storage, wireless communications in multipath environments, and communications through band-limited media. The channel capacity is typically obtained by optimizing the channel input process to maximize the entropy of the channel output. If the channel input is a Markov process, then the channel output is a hidden Markov process, and the problem is equivalently stated as the maximization of the entropy of a hidden Markov process. It is well known that even if the channel has finite memory, the channel capacity is generally not attained by a finite-memory channel input process, so generally, finite-memory Markov processes do not achieve the capacities of finite-memory channels. However, if feedback from the receiver to the transmitter is utilized, then a certain class of finite-memory conditionally Markov sources do achieve the feedback capacities of finite-memory channels. We establish some basic results for this case: 1) Finite-memory conditionally Markov sources achieve the capacities of finite-memory channels, 2) The optimal processor of the feedback is the forward recursion of the sum-product algorithm (i.e, the forward recursion of the Baum-Welch algorithm, or the Kalman-Bucy filter, depending on the application), 3) This generalizes Shannon's well-known result that memoryless sources achieve the (feedback) capacities of memoryless channels, i.e., we now have that finite-memory conditionally Markov sources achieve the capacities of finite-memory channels. An interesting consequence is that decoders for codes that achieve feedback capacities need not utilize long buffer memories, but rather the decoders can be implemented using extremely simple detection/estimation techniques already available in the statistical signal processing literature. We give several examples of how this applies to some well-known single-input-single-output channels. Further, we consider the open problem of establishing the capacity (or capacity bounds) for the relay channel, and show that similar results apply for relay channels with either deterministic or randomly fading finite intersymbol interference memory.

- *M. Pollicott (Math, Warwick):*

*Title: Computing integrals, Lyapunov exponents and entropy using cycle expansions*

Abstract: I will describe an approach which is based upon the study of certain analytic functions (called dynamical determinants) and studied by Ruelle. In certain cases, some of the above quantities can be "read off" from these functions. Using some classical ideas on determinants (originating with Grothendieck in the 1950s) one can rapidly approximate these analytic functions by polynomials. "Cycle expansions" refers to the explicit method, used by Cvitanovic et al, for computing these polynomials (and thus computing numerically the associated quantities).

## References

- [1] D. Arnold, H-A Loeliger, On the information rate of binary-input channels with memory, Proceedings of IEEE ICC, 2001, v. 9, pp.2692-2695.
- [2] L. Arnold, Random Dynamical Systems, Springer Verlag, Berlin, Heidelberg, 1998.
- [3] L. Arnold, V. M. Gundlach, and L. Demetrius, Evolutionary formalism for products of positive random matrices, Annals of Applied Probability, 4:859-901, 1994.
- [4] R. Atar and O. Zeitouni, Lyapunov Exponents for Finite State Nonlinear Filtering, SIAM Journal on Control and Optimization, 35(1):36-55, 1997.
- [5] J. J. Birch, Approximations for the entropy for functions of Markov chains, Ann. Math. Statist., 33:930-938, 1962.
- [6] D. Blackwell, The entropy of functions of finite-state markov chains, Trans. First Prague Conf. Information Thoery, Statistical Decision Functions, Random Processes, 1957, 13-20.

- [7] M. Boyle and S. Tuncel, Infinite-to-one codes and Markov measures, *AMS Transactions*, 285 (1984), 657 – 684.
- [8] J.-R. Chazottes and E. Ugalde, Projection of Markov measures may be Gibbsian, *Journal of Statistical Physics*, 111 (2003), 1245-1272.
- [9] S. Egner, V. Balakirsky, L. Tolhuizen, S. Baggen, and H. Hollmann, On the entropy rate of a hidden markov model, In *Proceedings of the 2004 IEEE International Symposium on Information Theory*, page 12, Chicago, U.S.A., 2004.
- [10] S. Y. Ephraim and N. Merhav, Hidden Markov Processes, *IEEE Transactions on Information Theory*, 48, 2002, 1518–1569.
- [11] R. Gharavi and V. Anantharam, An upper bound for the largest lyapunov exponent of a markovian product of nonnegative matrices, *Theoretical Computer Science*, Vol. 332, Nos. 1-3, pp. 543-557, February 2005.
- [12] G. Han and B. Marcus, Analyticity of entropy rate of a hidden Markov chain, *IEEE Transactions on Information Theory*, 52(2006), 5251-5266.
- [13] G. Han and B. Marcus. Derivatives of Entropy Rate in Special Families of Hidden Markov Chains. *IEEE Transactions on Information Theory*, Volume 53 (2007), 2642 - 2652.
- [14] G. Han and B. Marcus. Asymptotics of noisy constrained capacity, *Proc. ISIT 2007, Nice*, 991-995.
- [15] B. Hochwald and P. Jelenkovic, State Learning and mixing in entropy of Hiddern Markov Processess and the Gilbert-Elliot channel, *IEEE Transactions on Information Theory*, 45 (1999), 128–138.
- [16] T. Holliday, A. Goldsmith, and P.Glynn, On Entropy and Lyapunov Exponents for Finite State Channels, *IEEE Transactions on Information Theory*. 52 (2006).
- [17] P. Jacquet, G. Seroussi, and W.Szpankowski, On the entropy of a hidden Markov process, *Data Compression Conference 2004: 362-371* (full paper to appear in *Theoretical Computer Science*).
- [18] P. Jacquet, G. Seroussi, and W. Szpankowski. Noisy Constrained Capacity. *ISIT 2007, Nice*, 986-990.
- [19] A. Kavcic, On the capacity of Markov sources over noisy channels, *Proceedings of IEEE Globecom*, 2001, v. 5, pp. 2997-3001.
- [20] B. Kitchens and S. Tuncel. Finitary measures for subshifts of finite type and sofic systems, *AMS Memoirs*, 58(338), 1985.
- [21] B. Marcus, K. Petersen and S. Williams, Transmission rates and factors of markov chains, *Contemporary Mathematics*, 26:279–294, 1984.
- [22] C. Nair, E. Ordentlich and T.Weissman, On asymptotic filtering and entropy rate for a hidden Markov process in the rare transitions regime, *Proceedings of ISIT 2005, Adelaide*.
- [23] E. Ordentlich and T. Weissman, On the optimality of symbol by symbol filtering and denoising, *IEEE Transactions on Information Theory*, 52 (2006), 19-40.
- [24] E. Ordentlich and T. Weissman, New bounds on the entropy rate of hidden Markov processes, *San Antonio Information Theory workshop*, 2004.
- [25] Y. Peres, Analytic dependence of Lyapunov exponents on transition probabilities, volume 1486 of *Lecture Notes in Mathematics*, Lyapunov’s exponents, *Proceedings of a Workshop*, Springer Verlag, 1990.

- [26] Y. Peres, Domains of analytic continuation for the top Lyapunov exponent. *Ann. Inst. H. Poincaré Probab. Statist.*, 28(1):131–148, 1992.
- [27] H. D. Pfister, J. B. Soriaga and P. H. Siegel, On the achievable information rates of finite state ISI channels, *Proceedings of IEEE Globecom*, 2001, v. 5, 2992-2996.
- [28] D. Ruelle, *Thermodynamic formalism : the mathematical structures of classical equilibrium statistical mechanics*, Addison-Wesley Pub. Co., Advanced Book Program, Reading, Mass, 1978.
- [29] D. Ruelle, Analyticity properties of the characteristic exponents of random matrix products, *Adv. Math.*, 32:68–80, 1979.
- [30] D. Ruelle, Differentiation of SRB states, *Comm. Math. Phys.*, 187(1):227–241, 1997.
- [31] W. Ślomyński, *Dynamical Entropy, Markov Operators and Iterated Function Systems*, Wydawnictwo Uniwersytetu Jagiellońskiego, Krakow, 2003.
- [32] O. Stenflo, Uniqueness of invariant measures for place-dependent random iterations of functions, volume 132 of *IMA Vol. in Math. and its Appl.*, pages 13–32, Springer Verlag, 2002.
- [33] P.O. Vontobel, A. Kavcic, D. Arnold, and H.-A. Loeliger A generalization of the Blahut-Arimoto algorithm to finite-state channels, *IEEE Transactions on Information Theory*, to appear.
- [34] P. Walters, Relative pressure, relative equilibrium states, compensation functions and many-to-one mappings between subshifts, *AMS Transactions*, 296 (1986), 1-31.
- [35] S. Yang, A. Kavcic, S. Tatikonda, The Feedback Capacity of Finite-State Machine Channels, *IEEE Transactions on Information Theory*, 51 (2005), 799-810
- [36] O. Zuk, I. Kanter, and E. Domany, Asymptotics of the entropy rate for a hidden markov process, *DCC 2005*, pp. 173-182. Journal version: The Entropy of a Binary Hidden Markov Process *J. Stat. Phys.*, 121(2005), 343-360.
- [37] O. Zuk, E. Domany, I. Kanter and M. Aizenman, Taylor Series Expansions for the Entropy Rate of Hidden Markov Processes, *ICC 2006, Istanbul*; Journal version: From Finite-System Entropy to Entropy Rate for a Hidden Markov Process *IEEE Sig. Proc. Letters*, 13(2006), 517 – 520.

In a dynamical system, a rule maps states  $x \in X$  forward in time. Familiar examples include discrete-time maps and differential equations in which the states are elements of  $\mathbb{R}^n$ . At time  $t$ , the current state  $x(t)$  of a dynamical system provides all the information necessary to calculate future states, and information about past states is redundant. Thus the sequence of states in a dynamical system satisfies the Markov property, which I will more formally define in (1.7). The notion is that the underlying Markov process chugs along unaware of the observations and that, when the process arrives at each successive state  $s(t)$ , an observation  $y(t)$  is produced in a fashion that depends only on the state  $s(t)$ . e. f. d. Hidden Markov processes in the context of symbolic dynamics Mike Boyle, University of Maryland, College Park Karl Petersen, University of North Carolina, Chapel Hill On the preservation of Gibbsianness under symbol amalgamation Jean-Ren  Chazottes, Polytechnique, Palaiseau Edgardo Ugalde, Universidad Aut noma de San Luis Potos - A note on a complex Hilbert metric with application to domain of analyticity for entropy rate of. Computing entropy rates for hidden Markov processes Mark Pollicott, University of Warwick. 223. 8. Hidden Markov measures provide a natural way to model systems governed by chance 1. Hidden Markov processes in the context of symbolic dynamics Mark Boyle and Karl Petersen 2. On the preservation of Gibbsianess under symbol amalgamation Jean-Ren  Chazottes and E. Ugalde 3. A note on a complex Hilbert metric with application to domain of analyticity for entropy rate of hidden Markov processes Guangyue Han, Brian Marcus and Yuval Peres 4. Bounds on the entropy. rate of binary hidden Markov processes Erik Ordentlich and Tsachy Weissman 5. Entropy rate for hidden Markov chains with rare transitions Yuval Peres and Anthony Quas 6. The capacity of finite-state channels in the hi