Abstracts for BIRS Workshop on Entropy Rate of Hidden Markov Processes and Connections to Dynamical Systems

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 Autonama 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 nomalized potential of the measure. Even tough 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

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

  where $i_1, .., i_k$ are uniformly random in $[i - R, i + R]$, through an erasure channel. Further you know $i_1, ...i_k$. 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. Slomczynski (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 ϵ 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 Grothendeick 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).