D I M A C S

Center for Discrete Mathematics and Theoretical Computer Science

DIMACS SPECIAL FOCUS ON DISCRETE PROBABILITY: 1998 ANNUAL REPORT

DIMACS Center • Rutgers University • 96 Frelinghuysen Road • Piscataway, NJ 08854-8018 • USA
TEL: 732-445-5928 • FAX: 732-445-5932 • EMAIL: center@dimacs.rutgers.edu
WWW: http://dimacs.rutgers.edu/ • TELNET: telnet info.rutgers.edu 90

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs - Research, Bell Labs, NEC Research, and Telcordia Technologies (formerly Bellcore)

Overview:

This program began during the 1996 - 1997 academic year and continued into 1997 - 1999 at a much lower level. It is jointly organized and sponsored with the Institute for Advanced Study, which has cohosted visitors and workshops with DIMACS.

Organizers:

- Enrico Bombieri, Institute for Advanced Study
- Jeff Kahn, Rutgers University
- Peter Winkler, Bell Labs

Sponsors:

The National Science Foundation
The National Security Agency
AT&T Labs - Research
Bell Labs
Bellcore
Georgia Tech
The Institute for Advanced Study
NEC Research Institute
National Institute for Standards and Technology (NIST)
Rutgers University
Princeton University

Theme of program:

Probabilistic methods and ideas have by now become central to large parts of discrete mathematics and theoretical computer science. At the same time, recent years have seen a number of major developments in parts of probability and statistical mechanics dealing with topics of an essentially discrete nature (e.g., percolation, random walks and Markov chains, particle systems, branching processes).

A major goal of this program is to foster interaction between the dm/tcs community and the probability and statistics community. This is already happening to some extent, with both the application of increasingly sophisticated tools from probability to dm and tcs, and the solution of problems from probability by combinatorial means. Nonetheless, such interactions, and even awareness of what these communities have to offer each other, are still far more exception than rule. So the program is intended to give a substantial push to a broader and more rapid development of these connections. The program is also reaching out from the mathematical sciences to the physical sciences, especially statistical mechanics.

Workshops Feb. 1998 - Nov. 1998:

- **Miniworkshop on Mathematical Aspects of Ordering and Phase Segregation Dynamics in Lattice and Continuum Systems**
  - Date: May 16, 1998
  - Location: Rutgers University
  - Attendance:
  - Organizers: Joel Leibowitz, Rutgers

This miniworkshop took place just prior to the semi-annual statistical mechanics conference which took place on May 17-18. It was devoted to mathematical aspects of phase segregation dynamics in both lattice and continuum systems, as well as to quasi crystals. These are both topics in which John Cahn, currently at NIST, has been a central figure. The statistical mechanics conference, as well as this meeting, honored his 70th birthday. This meeting focused more on the mathematical issues, while the statistical mechanics meeting focused more on the physical ones.

Workshops Planned for the Rest of 1998-99:

- **Workshop: Joint DIMACS-Georgia Tech Workshop on Combinatorics of Statistical Mechanics Models**
  - Dates: April 29-May 1, 1999
  - Location: Georgia Institute of Technology
  - Organizers: Dana Randall and Prasad Tetali, Georgia Tech

Selected Research Results

New Lower Bounds on the Decoding Error Probability Simon Litsyn (V) derived new upper bounds on the error exponents for maximum likelihood decoding in binary symmetric channels by using an analysis of possible distance distributions of codes. This is an improvement on the best earlier known bounds by Shannon-Gallager-Berlekamp (1967) and McEliece-Omura (1977) for the classical problem of information theory of estimating probabilities of decoding error when a block code is used for information transmission over a binary symmetric channel (BSC), in particular giving exponential bounds on the performance of the best codes. These bounds are of the form $e^{nE(R,p)+o(n)}$ where n is the length of the code, $o(n)/n$ tends to 0 when n grows and $E(R,p)$ is the error exponent depending only on the code rate R and the transition probability p. The function $E(R,p)$ has been thoroughly studied since 1948, when Shannon proved in his seminal paper that the error exponent for maximum likelihood decoding is positive in the interval $0 \leq R \leq C_p$ where C_p is the capacity of the channel. Thus for all rates less than C_P, the error probability can be made as small as one wishes by choosing an appropriate n. The problem of deriving lower and upper bounds on the error exponent has attracted a great deal of attention since then. Thus Litsyn's improvements are of great interest.

Distance Distribution of Codes The problem of estimating possible distributions and deriving upper and lower bounds on individual and integral components is of great importance in coding and information theory. Simon Litsyn (V) derived new lower bounds on interval distributions. The estimates guarantee existence of a distance component in a certain range having essentially (exponentially) large value. As a result of this, in joint work with A. Ashikhmin and A. Barg (PM, BL), he was able to attack two classical problems in coding theory: bounding generalized weights of binary linear codes and estimating the list radius of codes. In particular, a binary list of m radius r code is a subset of binary n-dimensional Hamming space with the property that every point of the space is covered at most m times by spheres of radius r centered at code vectors. If $m = 1$ the list code is just the conventional r-error correcting code. A new asymptotic upper bound on the size of the list codes with fixed m and r growing linearly in n was obtained.

An entropy approach to the hard-core model on bipartite graphs The *hard-core measure* with activity λ on a finite graph G is the measure $hc(\gamma)$ on the set of independent sets of G under which the probability of I is proportional to $\gamma^|I|$. One may take limits of such measures to obtain hard-core measures on infinite graphs, where uniqueness (for given γ) is a central issue. Jeff Kahn (PM, RU) introduced entropy as a tool for studying these measures on regular bipartite graphs of high degree (which includes the lattices $\mathbf{Z}^d$ for large d, the cases of greatest interest for the hard-core model). The entropy approach was used to establish, *inter alia*, asymptotically exact estimates for the expected size of an independent set chosen from $hc(\gamma)$ with $\gamma > 0$ fixed and G as above; for fixed $\gamma > 0$ and G the Hamming n-cube (n large), $hc(\lambda)$ exhibits a "phase transition," meaning that with probability close to 1, almost all vertices of the random independent set have the same parity. An analogue of the phase transition result for large, high-dimensional tori would prove a central conjecture in this area, that the *critical activity* for the hard-core model on $\mathbf{Z}^d$ (the value of γ below which one has a unique hard-core measure) tends to 0 as $d \to \infty$. Indeed, the phase transition result was suggested by J. van den Berg (V) as a test case for this conjecture.

A Fluctuation Theorem for Markov Processes Cohen and Gallavotti noticed that for time-reversible thermostatted dynamics the large deviations for the phase space contraction satisfy a certain symmetry. Their proof is based on the method of Markov partitions and thus assumes the thermostatted dynamics to be sufficiently chaotic. This necessary technical assumption can be proved only in a few cases, such as the case of Chernov, Eyink, Lebowitz, and Sinai. On the Markov partition level the dynamics are essentially described by a Markov chain. Thus it is natural to inquire whether and in what sense the Gallavotti-Cohen fluctuation theorem holds for stochastic Markov processes. Herbert Spohn (V) found that indeed the fluctuation theorem holds with great generality, provided a condition of local detailed balance is satisfied. He used techniques from the theory of large deviations and described systems in thermal equilibrium by reversible Markov processes. For them one has the Einstein relations and the Onsager symmetry. These notions found their natural extension to driven nonequilibrium systems.

A Central Limit Theorem for Convex Chains in the Square Points $P_1, \ldots, P_n$ in the unit square define a convex n-chain if they are below $y = x$ and, together with $P_0 = (0, 0)$ and $P_{n+1} = (1, 1)$, they are in convex position. Under uniform probability, I. Bárány (V), Gunter Rote, Bill Steiger (PM, RU), and Cunhui Zhang (RU) proved an almost sure limit theorem for these chains that used only probabilistic arguments and that strengthened similar limit shape statements established by other authors. An interesting feature is that the limit shape is a direct consequence of the method. The main result was an accompanying central limit theorem for these chains. A weak convergence result implies several other statements concerning the deviations between random convex chains and their limit. This was the first central limit theorem proved for any of the variations of limit shape that have been studied recently by Bárány, Sinai, and Vershik.

The Asymmetric Simple Exclusion Process Pablo A. Ferrari (V) worked on the asymmetric simple exclusion process in one dimension. It is known that for nearest neighbor jumps there are so-called blocking measures. These measures are products, and they approach densities one and zero as the space coordinate goes to plus or minus infinity, respectively. The proof of the existence of these measures uses the fact that they are reversible for the system. It is a striking open problem that the existence of these measures cannot be proven for general jumps. However, for very special values of the jump rates, it is possible to show that the above measures are still reversible for jumps larger than nearest neighbors. Ferrari developed a coupling between two exclusion processes with different rates. The method allowed him to extend the parameter region for which there are invariant blocking measures. The coupling showed that if certain inequalities hold for the jump rates (that essentially say that the first process has more drift than the second), then the existence of blocking measures for the first process implies the existence of an invariant measure for the second one.

DIMACS special foci cross-fertilize each other. The following is one example of a result that arose from the cross-fertilization of two such foci, the one on Discrete Probability (with its emphasis on statistical physics) and the one on Mathematical Support for Molecular Biology.

Statistical Properties of Contact Maps A contact map is a simple representation of the structure of proteins and other chain-like macro-molecules. This representation is quite amenable to numerical studies of folding. Michele Vendruscolo, Balakrishna Subramanian (GS, RU), Ido Kanter, Eytan Domany and Joel Lebowitz (PM, RU) showed that the number of contact maps corresponding to the possible configurations of a polypeptide chain of N amino acids, represented by $(N-1)$ step half avoiding walks on a lattice, grows exponentially with N for all dimensions $D > 1$. They carried out exact enumerations in $D = 2$ on the square and triangular lattices for walks of up to 20 steps and investigated various statistical properties of contact maps corresponding to such walks.