DIMACS Center
Rutgers University

Special Focus on Computational Information Theory and Coding

Final Report

September 2005

Ia. Participants from the program

Participants:

Computational Information Theory and Coding Special Focus Organizers:

Robert Calderbank, Chair, AT&T Labs (now at Princeton University)
Chris Rose, Rutgers University
Amin Shokrollahi, Digital Fountain (now at Ecole Polytechnique Federale De Lausanne)
Emina Soljanin, Bell Labs
Sergio Verdu, Princeton University

Participants of Workshops and Working Groups under the Special Focus:

Workshop on Signal Processing for Wireless Transmission
Dates: October 7 - 9, 2002
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Sergio Verdu, Princeton University
Jerry Foschini, Bell Labs

Workshop on Network Information Theory
Dates: March 17 - 19, 2003
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Piyush Gupta, Bell Laboratories
Gerhard Kramer, Bell Laboratories
Adriaan van Wijngaarden, Bell Labs

Workshop on Complexity and Inference
Dates: June 2 - 5, 2003
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Mark Hansen, Bell Labs
Paul Vitanyi, University of Amsterdam
Bin Yu, University of California, Berkeley

Workshop on Algebraic Coding Theory and Information Theory
Dates: December 15 - 18, 2003
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Alexander Barg, DIMACS/Rutgers University
Alexei Ashikhmin, Bell Labs

Iwan Duursma, University of Illinois

Working Group Meeting: Theoretical Advances in Information Recording
Dates: March 22 - 24, 2004
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Emina Soljanin, Bell Labs
Paul Siegel, University of California
Bane Vasic, University of Arizona
Adriaan J. van Wijngaarden, Bell Labs

Workshop on Theoretical Advances in Information Recording
Dates: March 25 - 26, 2004
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Emina Soljanin, Bell Labs
Paul Siegel, University of California
Bane Vasic, University of Arizona
Adriaan J. van Wijngaarden, Bell Laboratories

Working Group on Network Coding
Dates: January 26 - 27, 2005
Location: DIMACS Center, CoRE Building, Rutgers University

Organizers:
Piyush Gupta, Bell Laboratories
Gerhard Kramer, Bell Laboratories
Emina Soljanin, Bell Laboratories

Special Focus Visitors:

Christina Fragouli, University of Athens, 2/1/03-6/13/03 and 10/12/03-10/19/03
Gregory Kabatianski, Institute for Information Transmission Problems, 5/20/02-6/4/02
Navin Kashyap, Queen's University, 8/16/04-8/20/04
Olgica Milenkovic, University of Colorado at Boulder, 5/27/05-7/27/05
Viacheslav Prelov, Russian Academy of Sciences, 10/1/01-1/31/02
Serap Savari, Bell Labs, 8/1/03-10/31/03
Eran Sharon, Tel Aviv University, 4/5/04-4/15/04
Vitaly Skachek, Technion, 7/7/04-8/23/04
Nedeljko Varnica, Harvard University, 7/1/04-8/31/04
Melda Yuksel, Polytechnic University, 7/1/04-8/31/04
Gilles Zemor, E'Cole National Supérieure des Télécommunications, 8/18/01-8/31/01 and 8/18/02-8/31/02

Special Focus Postdoc:

Xuerong Yong, Hong Kong University of Science and Technology, 2002 - 2003

Ib. Participating Organizations

Telcordia Technologies: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

AT&T Labs - Research: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

NEC Laboratories America: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

Lucent Technologies, Bell Labs: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

Princeton University: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

Rutgers University: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

Avaya Labs: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

HP Labs: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

IBM Research: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

Microsoft Research: Facilities; Personnel Exchanges
Partner organization of DIMACS. Individuals from the organization participated in the program planning and workshops.

Office of Naval Research: Facilities; Personnel Exchanges
ONR partially funded one of the workshops.

The New Jersey Commission on Science and Technology
Provided general support for the Special Focus.

1c. Other Collaborators

The project involved scientists from numerous institutions in numerous countries. There were hundreds of attendees at our workshops, coming from a variety of types of institutions and disciplines. The resulting collaborations also involved individuals from many institutions in many countries.

II. PROJECT ACTIVITIES

Theme of the Program:

The main purpose of coding theory is the reliable transmission or storage of data. It is common to think about information theory as a science concerned with theoretical limits on rates at which reliable transmission or storage of data is possible. During the course of the 50-year life of information theory, these limits (known as channel capacities) have been computed for various important telecommunications channels based on their underlying physics, which was considered given. Computational Information Theory is concerned with techniques (such as channel coding) for achieving channel capacities. Computational information theory has achieved dramatic scientific breakthroughs in recent years, and codes that come close to theoretical limits have been discovered.

This special focus explored the interconnections among coding theory, theoretical computer science, information theory, and related areas of computer science and mathematics, and addressed some of the challenges for the next 50 years - in particular wireless communication, magnetic/optical storage, the role of signal processing in networks, network information theory, and the creation of quantum information theory.

Today, the development of coding and information theory is closely related to the explosion of information technology, with applications to the Internet and the next generation of networks technologies. The rapid development of a myriad of networked devices for computing and telecommunications presents challenging and exciting new issues for coding and information theory.

In 1948 Shannon developed fundamental limits on the efficiency of communication over noisy channels. The coding theorem asserts that there are block codes with code rates arbitrarily close to channel capacity and probabilities of error arbitrarily close to zero. Fifty years later, codes for the Gaussian channel have been discovered that come close to these fundamental limits. Today, theoretical limits on rates at which reliable transmission or storage of data is possible are known for many telecommunications challenges. The same types of challenges face us in today's and tomorrow's new eras of networked, distributed computing and communications.

Coding theory has long used deep mathematical methods. We explored the use of such mathematical ideas from linear systems theory, automata theory, algebraic geometry, etc. in understanding the issues at the interface among coding theory, information theory, and theoretical computer science.

There are many connections between coding theory and theoretical computer science. We explored the connections between coding theory and: upper bounds on the number of random bits used by probabilistic algorithms and the communication complexity of cryptographic protocols; lower bounds on the complexity of approximating numerous combinatorial optimization functions; characterizations of traditional complexity classes such as NP and PSPACE; and the emerging theory of program testing.

Connections between information theory/coding and other parts of science, in particular physics, go back to the need to understand the relations between Shannon's perspective on entropy and the laws of thermodynamics. Today, attention has turned to quantum mechanics and the processing of intact quantum information states, and we explored such connections.

Coding theory is essential in numerous applied areas such as deep space communication, the theory of wireless channels, and optical/magnetic recording. Such applied areas are posing new and challenging problems for coding theorists and were a central focus of this special focus.

The development of the "next generation" of networks technologies leads to many new challenges for coding and information theory. Thinking of the network as a channel should help us to understand these challenges.

This special focus began in summer 2001 and ended in 2005.

Other Support

Activities in this special focus were also supported by NSF grant CCR 00-87022. The two grants together supported a wide variety of Special Focus activities listed herein.

Workshops, Working Groups, Tutorials:

Workshop on Signal Processing for Wireless Transmission
October 7 - 9, 2002
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Sergio Verdu, Princeton University and Jerry Foschini, Bell Labs
Attendance: 138

In contrast to the voiceband telephone channel, the wireless channel suffers from interference from other users and from fading due to destructive addition due to multipath propagation. This workshop explored the ultimate limits that information theory puts on spectral efficiency, as well as the best means of striving toward that efficiency. Multiuser detection and "Dirty Paper Coding" are among the key signal processing countermeasures that promise substantial improvements over existing systems. This workshop investigated such approaches from both a link and network level perspective. The enhancements from multiple antennas were also explored, including means of space-time coding.

Workshop on Network Information Theory
March 17 - 19, 2003
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Piyush Gupta, Gerhard Kramer and Adriaan van Wijngaarden, Bell Labs
Attendance: 132

The DIMACS Workshop on Network Information Theory focused on the area of efficient and reliable communication in multi-terminal settings. This field has recently attracted renewed attention because of key developments that have spawned a rich set of challenging research problems. Applications such as wireless cellular and LAN data services, ad hoc networks and sensor networks should benefit from these developments. The aim of the workshop was to achieve a better understanding of the underlying information theoretic problems and their solutions.

The workshop topics included, but were not limited to, the following areas:

- Large communication networks: analysis, design, asymptotics
- Multi-terminal capacity/coding: relays, multi-access, random access

- Multi-terminal source coding: distributed sources, multiple descriptions
- Network coding: efficiency and reliability

Workshop on Complexity and Inference
June 2 - 5, 2003
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Mark Hansen, Bell Labs; Paul Vitanyi, University of Amsterdam; Bin Yu, University of California, Berkeley
Attendance: 73

The notion of algorithmic complexity was suggested independently by Kolmogorov, Chaitin, and Solomonoff in the 1960's. Both Kolmogorov and Chaitin introduced the concept as a way to formalize notions of entropy and randomness, building on results from theoretical computer science dealing with partial recursive functions. Independently, Solomonoff defined algorithmic complexity in the pursuit of universal priors for statistical inference. In recent years, Rissanen expanded the applicability of these ideas, employing well-established concepts from information theory to frame his principle of Minimum Description Length (MDL) for statistical inference and model selection.

Each of these lines of research has developed methods for describing data (through coding and compression, or by analogy with some formal computing device); and each of these lines has employed some concept of an efficient representation to guide statistical inference. This workshop explored both the foundational aspects of complexity-based inference as well as applications of these ideas to challenging modeling problems. Participants were drawn from the fields of statistics, information and coding theory, machine learning, and complexity theory. Application areas included biology, information technologies, physics and psychology. The following specific topics were covered by the workshop:

- Kolmogorov complexity and inference
- MDL (MML) theory and applications
- Lossy compression and complexity theory
- Complexity and Bayesian methods
- Individual sequence/on-line prediction and predictive complexity
- Compression methods for clustering
- Machine learning and computational complexity
- Complexity and cognitive science
- Applications

Workshop on Algebraic Coding Theory and Information Theory
December 15 - 18, 2003
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Alexander Barg, DIMACS/Rutgers University; Alexei Ashikhmin, Bell Labs; Iwan Duursma, University of Illinois
Attendance: 77

Discoveries made in coding theory in the 1990s brought forward a number of new topics that presently attract the attention of specialists. Present-day developments in coding theory are concentrated around low-complexity code families, algebraic and lattice decoding, code design for multiple access channels, interplay with the theory of random matrices, quantum error correction, quantization, nontraditional applications of coding theory such as alternatives to routing in networks, coding of correlated sources and problems in theoretical computer science. Expanding into these areas enriched coding theory research with new problems and ideas. On the other hand, a number of recent studies in information theory attempt to develop structured code families as an alternative to random coding arguments usually employed to establish performance limits.

This workshop established and strengthened links between algebraic coding and information/communications theory. The workshop featured theoretical contributions in some of the named areas, both new results and tutorial presentations. It had a substantial educational component, exposing coding theorists to a new range of problems and presenting constructive methods to the information theory community. As a result of the workshop, experts in coding theory will be able to identify problems in information theory that can be addressed with coding theory methods, and information theorists will be more familiar with ideas used for code construction.

Working Group Meeting: Theoretical Advances in Information Recording
March 22 - 24, 2004
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Emina Soljanin, Bell Labs; Paul Siegel, Univ. of California, Bane Vasic, University of Arizona; Adriaan J. van Wijngaarden, Bell Labs
Attendance: 22

See write the up for the companion workshop below.

Workshop on Theoretical Advances in Information Recording
March 25 - 26, 2004
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Emina Soljanin, Bell Labs; Paul Siegel, Univ. of California, Bane Vasic, University of Arizona; Adriaan J. van Wijngaarden, Bell Laboratories
Attendance: 25

The last twenty years have witnessed a tremendous increase in recording densities and data rates of man-made recording systems, as well as a steady progression in a variety of methods for information storage and retrieval. The continuing strong demands for more storage capacity and faster access will further drive the development of even more sophisticated storage systems and new materials. Further advances in physics, chemistry and biology may provide new possibilities to further push the limits. Information theory will be playing an important role in identifying principles unique to both natural and artificial storage systems and in building a model and framework for storage in these new materials. These principles are as old as the time "when life divided the labour between two separate activities: chemical work and information storage, metabolism and replication" (as noted by Matt Ridley in "Genome"), and an effort to better understand and mathematically formulate these mechanisms will offer new directions in information and coding theory. It will also motivate a review of existing storage system models and coding and detection methodologies, which, in turn, can provide improvements to new and existing systems as well.

This working group and workshop brought together experts on information storage from a range of different fields. In order to facilitate a synthesis of ideas, the workshop was organized around half-day sessions, each consisting of one or two short presentations followed by active discussion. The six primary topics were:

- recent advances in coding for magnetic storage
- two-dimensional and higher dimensional storage systems
- recent advances in coding for optical recording systems
- sofic systems and constrained coding

- error control coding and iterative decoding for storage systems
- information storage in DNA and combinatorics in computational biology.

Working Group on Network Coding
January 26 - 27, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Piyush Gupta, Bell Laboratories, Gerhard Kramer, Bell Laboratories,
Emina Soljanin, Bell Laboratories
Attendance: 29

Network coding is emerging as a "hot topic" in communications and networking, with many fundamental results appearing rapidly since the publication of the seminal paper by Ahlswede, Cai, Li, and Yeung in the IEEE Transactions on Information Theory in 2000. The celebrated min-cut, max-flow theorem states that a source node can send a commodity through a network to a sink node at the rate determined by the min-cut separating the source and the sink. Surprisingly, by re-encoding at nodes, the min-cut rate can also be achieved when multicasting to several sinks. Perhaps just as importantly, recent results have shown that network coding can improve the security and robustness of networks.

The broad interest in this topic, which has attracted researchers from across the spectrum from coding and networking to computer science, is a result of both the interdisciplinary nature of the problems, and the potential impact that network coding might have on the operation of future networks.

This workshop consisted of talks by leading researchers in network coding. The aim of the workshop was the dissemination of the most recent breakthroughs, and an exchange of ideas to further advance the area.

Princeton-Rutgers Seminar Series in Communications and Information Theory 2002 - 2003

Tuesday, October 1, 2002
Speaker: **P. R. Kumar**, University of Illinois
Title: Ad Hoc Wireless Networks: Analysis, Protocols, Architecture, And Towards Convergence
Location: DIMACS Center, Rutgers University, Busch Campus, Piscataway, NJ

Thursday, October 3, 2002
Speaker: **Jacob Ziv**, Professor of Electrical Engineering, Technion--Israel Institute of Technology: President, Israeli Academy of Sciences and Humanities
Title: Classification with Finite Memory Revisited
Location: Princeton University, Friend 101

Thursday, October 3, 2002
Speaker: **Shlomo Shamai (Shitz)**, Technion
Title: On Information Theoretic Aspects of Multi-Cell Wireless Systems
Location: Princeton University, Friend 004

Thursday, November 7, 2002
Speaker: **Michael Luby**, Digital Fountain
Title: LT Codes
Location: Princeton University, Friend 101

Thursday, November 14, 2002
Speaker: **Upamanyu Madhow**, University of California, Santa Barbara

Title: Information-Theoretic Prescriptions for Outdoor Wireless Communication
Location: DIMACS Center, Rutgers University, Busch Campus, Piscataway, NJ

Thursday, November 21, 2002
Speaker: **Tsachy Weissman**, Hewlett-Packard Labs
Title: Universal Discrete Denoising: Known Channel
Location: Princeton University, Friend 101

Monday, March 3, 2003
Speaker: **Michael Honig**, Northwestern University
Title: Asymptotic Methods in Wireless Communications
Location: DIMACS Center, Rutgers University, Busch Campus, Piscataway, NJ

Thursday, March 6, 2003
Speaker: **Babak Hassibi**, California Institute of Technology
Title: Some Asymptotic Results in Wireless Networks
Location: Princeton University, Friend 006

Thursday, March 13, 2003
Speaker: **Robert J. McEliece**, Professor of Electrical Engineering, California Institute of Technology
Title: Belief Propagation on Partially Ordered Sets
Location: Princeton University, Friend 006

Friday, March 14, 2003
Speaker: **Venkat Anantharam**, University of California, Berkeley
Title: Message Passing Algorithms for Marginalization
Location: CoRE Bldg, Rutgers University, Busch Campus, Piscataway, NJ

Monday, April 14, 2003
Speaker: **David N. C. Tse**, University of California, Berkeley
Title: Diversity and Multiplexing: A Tradeoff in Wireless Systems
Location: DIMACS Center, Rutgers University, Busch Campus, Piscataway, NJ

Thursday, April 24, 2003
Speaker: **Andrea Goldsmith**, Stanford University
Title: Capacity Limits of Wireless Channels with Multiple Antennas: Challenges, Insights, and New Mathematical Methods
Time: 4:30-5:30pm
Location: DIMACS Center, Rutgers University, Busch Campus, Piscataway, NJ

Thursday, May 1, 2003
Speaker: **Thomas Richardson**, Flarion Technologies
Title: Principles and Design of Iterative Coding Systems
Time: 4:30-5:30pm
Location: Princeton University, Friend 006

Princeton-Rutgers Seminar Series in Communications and Information Theory 2003 - 2004

Thursday, November 13, 2003
Speaker: **Vaidyanthan Ramaswami**, AT & T Labs

Title: Providing Dial Tone in the Presence of Circuit Congestion due to Long Holding Time Internet Dial-Up Calls - Assuring Emergency Services Access
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, November 20, 2003
Speaker: **Paul Henry**, AT & T Labs
Title: 4G Cellular: Now That's Mobile Data
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, December 4, 2003
Speaker: **Larry Peterson**, Princeton University
Title: A Blueprint for Introducing Disruptive Technology into the Internet
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, December 11, 2003
Speaker: **Nirwan Ansari**, NJIT
Title: IP Traceback by Deterministic Packet Marking
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, February 5, 2004
Speaker: **Sekhar Tatikonda**, Yale University
Title: Markov Channels, Sufficient Statistics, and Capacity
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, February 12, 2004
Speaker: **P. R. Kumar**, University of Illinois
Title: Wireless Networks: From Information Transfer to Sensing and Control
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, February 19, 2004
Speaker: **Sandy Fraser**, Fraser Research
Title: A Packet Switch to Serve One Million Households
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, February 26, 2004
Speaker: **John Tsitsiklis**, MIT
Title: A Game Theoretic View of Efficiency Loss in Network Resource Allocation
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, March 4, 2004
Speaker: **Lang Tong**, Cornell University
Title: On Cross-Layer Design of Wireless Networks
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, April 15, 2004
Speaker: **Steve Sposato**, SBC Network Systems Engineering
Title: Network Design and Optimization for Large Scale Network Deployments
Location: Princeton University, E-Quad B205, Princeton, NJ

Thursday, April 22, 2004
Speaker: **Ness Shroff**, Purdue University

Title: Simplification of Network Dynamics in Large Systems
Location: Princeton University, E-Quad B205, Princeton, NJ

Thurday, May 27, 2004
Speaker: **Venkat Anantharam**, U. C., Berkeley
Title: The Role of Common Randomness in Communications and Control
Location: Princeton University, E-Quad B205, Princeton, NJ

III. PROJECT FINDINGS

In this section, we have indicated affiliations for those members of a collaboration who were participants in the project.

LP Decoding Achieves Capacity. Jon Feldman (Columbia University) and Cliff Stein (Columbia University) developed a linear programming (LP) decoder that achieves the capacity (optimal rate) of a wide range of probabilistic binary communication channels. This is the first such result for LP decoding. More generally, as far as Feldman and Stein are aware, this is the first known polynomial-time capacity-achieving decoder with the maximum-likelihood (ML) certificate property—where output codewords come with a proof of optimality. Additionally, this result extends the capacity-achieving property of expander codes beyond the binary symmetric channel to a larger family of communication channels. Perhaps most importantly, since LP decoding performs well in practice on turbo codes and low-density paritycheck (LDPC) codes (comparable to the popular “belief propagation” algorithm), this result exhibits the power of a new, widely applicable “dual witness” technique for bounding decoder performance. For expander codes over an adversarial channel, Feldman and Stein proved that LP decoding corrects a constant fraction of errors. To show this, they provide a new combinatorial characterization of error events that is of independent interest, and which they expect will lead to further improvements.

Suboptimal Behaviour of Bayes and MDL in Classification Under Misspecification. Peter Grunwald (CWI) and John Langford (IBM) started a succesfull collaboration on one of the main themes of the workshop on Complexity and Inference. They showed - to the great surprise of some of the other workshop participants - that MDL and Bayesian methods for machine learning can be suboptimal in certain problems, even with an infinite amount of data. Overfitting is a central concern of machine learning and statistics. Two frequently used learning methods that in many cases “automatically” protect against overfitting are Bayesian inference and the Minimum Description Length (MDL) Principle. Grunwald and Langford showed that forms of Bayesian and MDL inference that are often applied to classification problems could be inconsistent in the sense that they asymptotically overfit. This means there exist learning problems such that, no matter how much data is available, the generalization errors of the MDL classifier and the Bayes classifier relative to the Bayesian posterior both remain bounded away from the smallest achievable generalization error. There are several plans to continue this line of research.

Best Known Upper Bound on the Capacity of Information Networks. Micah Adler, Nicholas J. A. Harvey (MIT), Kamal Jain (Microsoft), Robert D. Kleinberg, and April Rasala Lehman (MIT) computed the strongest known upper bound on the capacity of general information networks. In an information network, nodes can transmit data by copying bits directly from input to output edges (as in a multicommodity flow) or, more generally, by applying encoding operations to those bits. The notion of capacity studied by Adler, Harvey, Jain, Kleinberg, and Lehman refers to the latter model. Their upper bound on capacity is based on a new technique that combines properties of entropy with a strong information inequality derived from the structure of the network. This blend of information theoretic and graph theoretic arguments generates several interesting results. In particular, they made progress on the undirected k-pairs conjecture, which states that the capacity of an undirected network supporting k point-to-point connections is achievable by multicommodity flows. Their techniques proved the conjecture for a non-trivial class of graphs, and also yielded the first known proof of a gap between the sparsity of an undirected graph and its capacity. They believe that these techniques may be instrumental in resolving the conjecture completely. They demonstrated the importance of the undirected k-pairs conjecture by connecting it with long-standing open questions in Input/Output (I/O) complexity. They also showed that proving the conjecture would provide the strongest known lower bound for computation in the oblivious cell-probe model and give a non-trivial lower bound for two-tape oblivious Turing machines.

Extrinsic Information Transfer Functions: Model and Erasure Channel Properties. Extrinsic information transfer (EXIT) charts are a tool for predicting the convergence behavior of iterative decoding of concatenated codes. Workshops in this special focus led to a new collaboration among Alexei Ashikhmin (Lucent Technologies), Gerhard Kramer (Lucent Technologies), and Stephan Ten Brink (Bell Labs). They developed a decoding model that applies to many communication problems, including the iterative decoding of parallel concatenated (turbo) codes, serially concatenated codes, low-density parity-check codes, and repeat-accumulate codes. They defined EXIT functions using the model, and proved several properties of such functions for erasure channels. One property expresses the area under an EXIT function in terms of a conditional entropy. A useful consequence of this result is that the design of capacity-approaching codes reduces to a curve-fitting problem for all the aforementioned codes. A second property relates the EXIT function of a code to its Helleseth-Klove-Levenshtein information functions, and thereby to the support weights of its subcodes. The relation is via a refinement of information functions called split information functions, and via a refinement of support weights called split support weights. Split information functions are used to prove a third property that relates the EXIT function of a linear code to the EXIT function of its dual.

A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication. Theoretical results describing the sum-capacity when using multiple antennas to communicate with multiple users in a known rich scattering environment have challenged researchers to find practical transmission schemes that achieve this capacity. Christian Peel (Bell Labs), Bertrand Hochwald (Lucent Technologies), and A. Lee Swindlehurst (Brigham Young University) introduced a simple encoding algorithm that achieves near-capacity at sum-rates of tens of bits/channel-use. The algorithm is a variation on channel inversion that regularizes the inverse and uses a "sphere encoder" to perturb the data to reduce the energy of the transmitted signal. They showed that the performance difference between channel inversion with and without this perturbation is dramatic. With the perturbation, they achieved linear growth in the sum-rate with the number of users.

Network Information Flow with Correlated Sources. Wireless sensor networks made up of small, cheap, and mostly unreliable devices equipped with limited sensing, processing and transmission capabilities, have recently sparked a fair amount of interest in communications problems involving multiple correlated sources and large-scale wireless networks. An important class of applications for such networks involves a dense deployment of a large number of sensors over a fixed area in which a physical process unfolds. The task of these sensors is then to collect measurements, encode them, and relay them to some data collection point where this data is to be analyzed, and possibly acted upon. Joao Barros (Munich University of Technology) and Sergio Servetto (Cornell University) considered the following network communication setup, originating in a sensor networking application they refer to as the "sensor reachback" problem. In a directed graph G, if there is an edge between node i and node j then node i can send messages to node j over a discrete memoryless channel of capacity Cij. The channels are independent. Each node has a certain probability of observing one of several sources of information. The goal is to solve an incast problem in G: nodes exchange messages with their neighbors, and after a finite number of communication rounds, one of the nodes must have received enough information to reproduce the entire field of observations, with arbitrarily small probability of error. Barros and Servetto found necessary and sufficient conditions for perfect reconstruction. Their main finding is that in this setup a general source/channel separation theorem holds, and that Shannon information behaves as a classical network flow, identical in nature to the flow of water in pipes. This "information as flow" view provides an algorithmic interpretation for their results, among which perhaps the most important one is the optimality of implementing codes using a layered protocol stack.

Efficient Coding Schemes for Commodity Flows Through a Network. The min-cut, max-flow theorem states that a source node can send a commodity through a network to a sink node at the rate determined by the flow of the min-cut separating the source and the sink. It has been shown that by linear re-encoding at nodes in communications networks, the min-cut rate can be also achieved in multicasting to several sinks. Emina Soljanin (Bell Labs), and special focus visitor Christina Fragouli (DIMACS Visitor, EPFL) collaborated on constructing such coding schemes efficiently. Their main contributions were in the design of network codes suitable for applications that require decentralized or local coding strategies. They proposed a subtree decomposition method to group together equivalent network configurations and developed an algorithm that allows local assignment of codes without knowledge of the overall network topology. Based on the subtree decomposition, they derived tight bounds on the network code size, and investigated connections between network coding and convolutional codes. They also started exploring other applications of the proposed subtree decomposition method. This is a new area and there are many open problems and research directions that they plan to explore.

Generalization of LDPC Codes. The research of special focus visitor Eran Sharon, a PhD student at Tel-Aviv University, and Alexei Ashikhmin (Bell Labs) focused on generalization of LDPC codes, combining Multi-edge type LDPC codes and GLD codes. Multi-edge type LDPC codes are a generalization of irregular LDPC codes, which introduce additional structure beyond standard irregularity into the specification of the bipartite graph representing the code. Multi-edge structures have several advantages. In general, with multi-edge structures one can achieve better performance with lower complexity. Multi-edge structures are especially useful for achieving excellent performance under extreme conditions where standard irregular LDPC codes fail. Examples of such conditions include very low or very high code rates, low error floor requirement and short code length. GLD codes are a generalization of LDPC codes such that the vertices in the underlying bipartite graph of the code represent constraints on the codeword symbols by an arbitrary linear block code (not necessarily a parity check code or a repetition code). These codes are more difficult to analyze, and they were mostly studied under a suboptimal hard decision decoding that was shown to correct a constant fraction of errors. The objective of Sharon and Ashikhmin was to develop analysis and design tools based on EXIT charts for multi-edge constructions of GLD codes that will enable further development and improvement of multi-edge constructions especially for short codes, low error floor and fast convergence. They wrote various optimization programs for GLD codes and Multi-Edge constructions (separately) for the Binary Erasure Channel. More specific work was done for Hamming constituent codes as right nodes in GLD codes. They obtained expressions for the EXIT functions of a MAP decoder of a Hamming code over the binary erasure channel, binary symmetric channel and an approximation for the binary input AWGN channel. Future work will include using the tools that were built for designing asymptotically good MultiEdge-GLD codes, optimized for various scenarios, such as fast convergence, low threshold, low error floor, etc. The quality of the designed codes will be determined via finite length simulation. Additional future work may include finite length analysis of GLD codes based on stopping sets and pseudo-codewords analysis.

Properties of Variable Rate Codes for Time Varying Channels. Special focus visitor Nedeljko Varnica (Harvard University) collaborated with Emina Soljanin (Bell Labs) on analyzing constructions and properties of variable rate codes for time varying channels that employ automatic repeat request (ARQ) protocol. They were particularly interested in incremental redundancy coding schemes for Hybrid-ARQ (IR schemes for HARQ) and on scheduling protocols for IR-HARQ. Varnica and Soljanin proposed a robust and new solution to the problem of coding for HARQ. The codes they proposed are based on “rateless” Luby transform (LT) codes and can achieve any level of redundancy. They proposed an algorithm that provides a choice of size of increments in the redundancy. In other words, they devised a transmission protocol that decides how many bits (or symbols) of information are sent in the initial transmission and in each subsequent transmission, if the initial transmissions were unsuccessful. These problems are very important for practical applications when the characteristics of the communication channels are not entirely known or are changing with time. The most obvious and most relevant example of such channels is the standard time-varying wireless channel widely used in modern wireless communications.

Efficient Codes for Digital Fingerprinting. Alexander Barg (Bell Labs and DIMACS) and Gregory Kabatiansky (Visitor, IPPI RAN, Moscow) studied efficient constructions of codes for digital fingerprinting. Relying on the ideas of their earlier work (DIMACS report 2001-52) they gave a first construction of codes with the identifiable parent property in which identification of a member of the coalition can be accomplished in time logarithmic in the total number of users of the system. Previously known constructions relied basically upon exhaustive search and thus had exponentially slower identification procedures.

Input-Output of Mutual Information of Multidimensional Channels. Slava Prelov (Visitor, Russian Academy of Sciences) and Sergio Verdu (Princeton) obtained a formula for the second-order expansion of the input-output of mutual information of multidimensional channels as the signal-to-noise ratio goes to zero. While the additive noise is assumed to be Gaussian, very general classes of input and channel distributions were considered. As special cases, these channel models include fading channels, channels with random parameters and channels with almost Gaussian noise. When the channel is unknown at the receiver, the second term in the asymptotic expansion depends not only on the covariance matrix of the input signal but also on the fourth mixed moments of its components. The second-order asymptotics of mutual information finds application in the analysis of the bandwidth-power tradeoff achieved by specific (not necessarily optimum) input signaling in the wideband regime.

Performance of Expander Codes on a Binary Symmetric Channel. Alexander Barg (Bell Labs and DIMACS) and Gilles Zemor (Visitor, ENST-Paris) studied performance of expander codes on a binary symmetric channel. This research was initiated during a visit of Zemor to DIMACS in 2001 when he, together with Barg, showed that these codes reach channel capacity under iterative decoding, producing the first example of capacity-attaining linear-time decodable codes. This work was extended during a DIMACS visit of Zemor in August 2002 when they introduced another family of expander codes tailored for transmission at rates in the immediate neighborhood of capacity. For this region they estimated the decrease rate (the error exponent) of error probability of decoding for randomized ensembles codes. The resulting estimate gives a substantial improvement of previous results for expander codes and some other explicit code families.

Digital Fingerprinting Codes. Alexander Barg (Bell Labs and DIMACS), G. R. Blakley, and Gegory Kabatiansky considered a general fingerprinting problem of digital data under which coalitions of users can alter or erase some bits in their copies in order to create an illegal copy. Each user is assigned a fingerprint, which is a word in a fingerprinting code of size M (the total number of users) and length n. They found binary fingerprinting codes secure against size-t coalitions which enable the distributor (decoder) to recover at least one of the users from the coalition with probability of error $exp(-Omega(n))$ for $M=exp(Omega(n))$. This is an improvement over the best known schemes that provide the error probability no better than $exp(-Omega(n^{1/2}))$ and for this probability support at most $exp(O(n^{1/2}))$ users. The construction complexity of codes is polynomial in n. They also found versions of these constructions that afford poly(n)=poly log(M) identification algorithms of a member of the coalition, improving over the best previously known complexity of $Omega(M)$. For the case t=2 they constructed codes of exponential size with even stronger performance, namely, the distributor can either recover both users from the coalition with probability 1-*exp*(*Omega*(*n*)), or identifies one traitor with probability 1.

IV. PROJECT TRAINING/DEVELOPMENT

This project provided support for a postdoctoral researcher at DIMACS. Xuerong Yong's research while at DIMACS is summarized below.

In the study of constrained codes, Yong's research focused on analyzing/computing the channel capacities of two-dimensional constrained codes. These problems are closely related to certain kinds of counting problems in combinatorics and computer science and are motivated by problems in Information Theory arising from the codes for mass data storage/transmission systems. While the study of the Shannon capacity of one-dimensional codes is well developed, the study of two-dimensional codes is still only in its infancy (although extremely active). Yong and his advisor, M. Golin, established new theoretical results on two-dimensional constrained codes and developed new general techniques for bounding channel capacity of constrained codes where the main tools used are matrix theory, graph spectra and combinatorial analysis. As examples, they considered and improved the analyses of the capacity of runlength limited two-dimensional constrained codes (previously considered by N. Calkin and H. Wilf), the channel capacity of read/write isolated memory (previously considered by M. Cohn), and the number of placements of nonattacking kings (a problem due to D. Knuth previously considered by H. Wilf) etc. They were also able to apply the techniques developed to attack other non information-theory related problems.

Yong worked on a number of graph theoretical questions related to reliability of transmission when there are faults in the network. The number of spanning trees of a graph is an important quantity in the study of the reliability of a network in the presence of line faults. Obtaining the exact number of spanning trees for a given graph is a difficult problem, but, for certain special graphs, deriving recurrence formulas is proven to be possible. Boesch et al. and Bedrosian conjectured independently that if G is a square cycle with n nodes, then the number of spanning trees of G is $nF(n)^2$, where F(n) is the nth Fibonacci number. D. Kleitman and B. Golden and F. Boesch and H. Prodinger proved this conjecture independently, and then X. Yong and F. Zhang gave a simpler proof. Since then, M. Golin, Y. Zhang and X. Yong have collaborated on developing techniques for deriving general formulas for the number of spanning trees in different classes of graphs and, in particular, they proved that the number of spanning trees of any circulant graph satisfies a linear recurrence relation. These formulas greatly reduce the amount of time needed to calculate the number of spanning trees.

Yong also worked in the area of graph spectra. The topological structure of a network can be analyzed by characterizing the distribution of its eigenvalues. Yong's deepest result in this area can be described as: In a set of graphs G, each of order n and with negative third largest eigenvalue, there is only one graph that has no eigenvalue equal to -1, all others have the property that there is an index $k \leq n/2$ such that $\lambda_j = -1$, for j=k, k+1, . . . ,n-k+1, where λ_k is the kth largest eigenvalue of G (in some cases the j can run up to n-k+2).

Graduate Students Supported and Their Research Topics

Small research projects by graduate students were also part of the project, aimed at involving them in all of the topics of the Special Focus. In addition, funds were available at each workshop for support of non-local students and non-local students were also invited to be in residence as visitors for periods of a week to several months. Here we list the research of the local students. Research of non-local students is included in the description of research results in the "Findings" section.

Eva Curry, Mathematics, Rutgers University, Summer 2002: Characterization of Low-pass in *n* dimensions.

Navin Goyal, Computer Science, Rutgers University, Summer 2002: Parent Identifying Codes.

V. OUTREACH ACTIVITIES

Special Focus visitors, graduate students, and senior faculty were available to interact with 2- and 4-year college faculty in the DIMACS "Reconnect" program, and with high school teachers in the DIMACS Connect Institute and the DIMACS Bio-Math Connect Institute.

VI. PAPERS/BOOKS/INTERNET

Books and One-Time Publications

Multiantenna Channels: Capacity, Coding and Signal Processing, Editors: Gerard J. Foschini and Sergio Verdú, American Mathematical Society, DIMACS Series, Volume 62, 2003.

Advances in Network Information Theory, Editors: Piyush Gupta, Gerhard Kramer, and Adriaan J. van Wijnggaarden, American Mathematical Society, DIMACS Series, Volume 66, 2004.

Algebraic Coding Theory and Information Theory, Editors: Alexei Ashikhmin, Alexander Barg, and Iwan Duursma, American Mathematical Society, DIMACS Series, 2005, to appear.

Theoretical Advances in Information Recording, Editors: Adriaan van Wijngaarden, Emina Soljanin, Paul Siegel, and Bane Vasic, American Mathematical Society, DIMACS Series, 2007, to appear.

Journal Articles

A. Alloum, J.J. Boutros, G. Shamir and L. Wang, "Non-systematic LDPC codes via scrambling and splitting," *Proceedings of the Forty-Third Annual Allerton Conference on Communication, Control, and Computing*, University of Illinois, September 30, 2005

A. Ashikhmin, G. Kramer and S. ten Brink, "Extrinsic information transfer functions: Model and erasure channel properties," *IEEE Trans. Inform. Theory*, **50** (2004), 2657-2673.

A. Barg and G. Kabatiansky, "A class of i.p.p. codes with efficient identification," *Journal of Complexity*, **20** (2004), 137-147.

A. Barg and G. Zémor, "Error exponents of expander codes under linear-time decoding," *SIAM Journal of Discrete Mathematics*, **17** (2004), 426-445.

A. Barg and G. Zemor, "Multilevel expander codes," in preparation.

A. Barg and G. Zemor, "Distance properties of expander codes," *IEEE Transactions on Information Theory*, to appear. Preprint at http://www.arxiv.org/abs/cs.IT/0409010.

J. Barros and S. D. Servetto, "Network information flow with correlated sources," *IEEE Transactions on Information Theory*, submitted, 2003 (original title: *The Sensor Reachback Problem*); Revised, February 2005.

J. Barros and S. D. Servetto, "Coding theorems for network information flow with correlated sources," in the *Proceedings of the IEEE International Symposium on Information Theory* (ISIT), Adelaide, Australia, September 2005.

J. Barros, C. Peraki and S. D. Servetto, "Efficient network architectures for sensor reachback," in the *Proceedings of the IEEE International Zuerich Seminar on Communications*, Zuerich, Switzerland, February 2004.

J. Barros and S. D. Servetto, "Coding theorems for the sensor reachback problem with partially cooperating nodes," in *Advances in Network Information Theory*, Piyush Gupta, Gerhard Kramer, and Adriaan J. van Wijnggaarden (eds.), American Mathematical Society (AMS), DIMACS Series, Volume **66**, 2004.

S.F. Bush and J. Miller, "A homology-driven boolean network as a framework for analyzing RNAi performance," in preparation.

R. Cilibrasi and P. Vitanyi, "Clustering by compression," *IEEE Trans Inform Th*., **51** (2005), 1523-1545.

R. Cilibrasi, P. Vitanyi and R. de Wolf, "Algorithmic clustering of music," *Computer Music Journal*, **28** (2004), 49-67.

R. Cilibrasi, P. Vitanyi and R. de Wolf, "Algorithmic clustering of music," *Proc. IEEE Conf on Web Delivery of Music, 2004,* to appear.

C. Fragouli and E. Soljanin, "Information flow decomposition for network coding," *IEEE Trans. Inform. Theory,* June 2004, submitted.

C. Fragouli and E. Soljanin, "Decentralized network coding based on subtree decomposition," *InfoCom 2004*, submitted.

C. Fragouli, "Connections between convolutional codes and network coding," *ICC 2004*, in preparation.

J. Feldman, R. Koetter and P.O. Vontobel, "The benefit of thresholding in LP decoding of LDPC codes," in *Proc. 2005 IEEE International Symposium on Information Theory*, Adelaide, Australia, September 4-9, 2005, to appear.

J. Feldman and C. Stein, "LP decoding achieves capacity," in *2005 ACM-SIAM Symposium on Discrete Algorithms* (SODA), January 2005.
(http://www.columbia.edu/~jf2189/pubs/lpcap_soda_final.pdf)

M. Golin, X. Yong, Y. Zhang and L. Sheng, "New upper and lower bounds on the channel capacity of read/write isolated memory," *Discrete Applied Mathematics*, **140** (2004), 35-48. Preliminary version, in *2000 IEEE International Symposium on Information Theory* (ISIT 2000), Italy.

P.D. Grünwald and J. Langford, "Suboptimal behaviour of Bayes and MDL in classification under misspecification," in *Proceedings of the Seventeenth Annual Conference on Computational Learning Theory* (COLT 2004).

P.D. Grünwald and J. Langford, “Suboptimal behaviour of Bayes and MDL in classification under misspecification,” *Machine Learning Journal*, submitted.

M.K Gupta, “The quest for error-correction in Biology,” *IEEE EMB (Engineering and Medicine in Biology) Magazine Special Issue on Communication Theory, Coding Theory and Molecular Biology*, December 2004), to appear.

N. Harvey, R. Kleinberg and A. Lehman, “On the capacity of information networks,” *IEEE Transactions on Information Theory & IEEE/ACM Transactions on Networking*, March 14, 2005, submitted.

Y. Hanatani, T. Horiyama and K. Iwama, “Hajos calculus on planar graphs,” *Proc. 4th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications*, 2005.

B. Hochwald, C. Peel and A. Swindlehurst, “A vector-perturbation technique for near-capacity multi-antenna multi-user communication Part II: Perturbation,” *IEEE Transactions on Communications*, February 2005.

K. Iwama and S. Tamaki, “Improved upper Bbounds for 3-SAT,” *Proc. SODA 2004*, (2004), 321-322.

G. Kramer, M. Gastpar and P. Gupta, “Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inform. Theory*, 2004, submitted.

G. Kramer and S.A. Savari, “Cut sets and information flow in networks of two-way channels,” in *Proceedings of the 2004 IEEE International Symposium on Information Theory*, Chicago, llinois, June 2004.

G. Kramer and S.A. Savari, “On networks of two-way channels,” in *Algebraic Coding Theory and Information Theory*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, A. Ashikhmin and A. Barg (eds.), **68** (2005), 133-143.

J.N. Laneman and G. Kramer, “Window decoding for the multiaccess channel with generalized feedback,” *IEEE Int. Symp. Inform. Theory*, Chicago, IL, (2004), 281.

M. Li, X. Chen, X. Li, B. Ma and P. Vitanyi, “The similarity metric,” *Proc. 14th ACM-SIAM Symp. Discrete Algorithms*, 2003.

M. Li, X. Chen, X. Li, B. Ma and P. Vitanyi, “The similarity metric,” *IEEE Trans Inform. Th.,*. to appear.

O. Milenkovic, “Asymptotic enumeration of combinatorial structures in regular LDPC code ensembles,” 2005, submitted.

O. Milenkovic, “Asymptotic spectra of trapping sets in regular and irregular LDPC code ensembles,” 2005, submitted.

T.J. Oechtering and H. Boche, “Capacity of a Gaussian FIR linear relay network,” in *Proc. WirelessCom 2005*, Maui, Hawaii, 2005.

T.J. Oechtering and A. Sezgin, “A new cooperative transmission scheme using the space-time delay code,” in *Proc. ITG Workshop on Smart Antennas*, Munich, Germany, 2004.

C.B. Peel, B.M. Hochwald and A.L. Swindlehurst, “A vector-perturbation technique for near-capacity multi-antenna multi-user communication - Part I and II,” available from http://mars.bell-labs.com/.

C. Peel, B. Hochwald and A. Swindlehurst, “A vector-perturbation technique for near-capacity multi-antenna multi-user communication Part I: Channel inversion and regularization,” *IEEE Transactions on Communications*, January 2005.

V. V. Prelov and S. Verd, “Second-order asymptotics of mutual information,” *IEEE Trans. Information Theory*, **50** (2004), 1567-1580.

C. Rose and G. Wright, “Inscribed matter as an energy efficient means of communication with an extraterrestrial civilization,” *Nature*, **431** (2004).

G. I. Shamir and J. J. Boutros, “Non-systematic low-density parity-check codes for nonuniform sources,” in *Proceedings of the IEEE International Symposium on Information Theory*, ISIT-2005, Adelaide, Australia, (2005), 1898-1902.

A. Talip, X. Yong and H. Inaba, “Further analysis of the number of spanning trees in circulant graphs,” in *Proceedings of 2004 International Conference on Software, Telecommunications and Computer Networks* (SoftComm2004), Italy

P.A. van der Helm, “Transparallel processing by hyperstrings,” in *Proceedings of the National Academy of Science of the United States* (PNAS), **101** (2004), 10862–10867.

S. Vishwanath, G. Kramer, S. Shamai (Shitz), S. Jafar and A. Goldsmith, “Capacity bounds for Gaussian vector broadcast channels,” in *Multiantenna Channels: Capacity, Coding and Signal Processing*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, G.J. Foschini and S. Verdu (eds.), **62** (2003), 107-122.

X. Yong, “Analyzing the codes that avoid specified differences by binary tree,” in *Proceedings of 2003 International Symposium on Information and Communication Technologies* (ISICT2003), Ireland.

Y. Zhang, X. Yong and M. Goldin, “Chebyshev polynomials and spanning tree formulas for circulant and related graphs,” *Discrete Mathematics*, to appear.

Talks

A. Alloum, J.J. Boutros, G. Shamir and L. Wang, “Non-systematic LDPC codes via scrambling and splitting,” P*roceedings of the Forty-Third Annual Allerton Conference on Communication, Control, and Computing*, University of Illinois, September 30, 2005.

M. Adler, N.J.A. Harvey, K. Jain, R.D. Kleinberg and A.R. Lehman, “On the capacity of information networks,” *ACM-SIAM Symposium on Discrete Algorithms* (SODA 06), Miami, Florida, January 2006.

A. Barg and G. Zémor, “Concatenated codes: Serial and parallel,” *International Symposium on Information Theory*, Yokohama, Japan, June 2003.

A. Barg and G. Zémor, “Distance properties of expander codes,” *International Symposium on Information Theory*, Chicago IL June 28, 2004.

S.F. Bush, panel session at *IEEE ICCCN 2004*. Includes J. Miller as a panelist.

J. Feldman, R. Koetter and P.O. Vontobel, "The benefit of thresholding in LP decoding of LDPC codes," in *IEEE International Symposium on Information Theory*, Adelaide, Australia, September 4-9, 2005.

J. Feldman and C. Stein, "LP decoding achieves capacity," *ACM-SIAM Symposium on Discrete Algorithms* (SODA), January 2005.
(http://www.columbia.edu/~jf2189/pubs/lpcap_soda_final.pdf)

C. Fragouli, "Decentralized network coding," Bell Labs, Murray Hill, June 12 2003.

C. Fragouli, "Decentralized network coding," AT&T Research Labs, Florham Park, June 18 2003.

C. Fragouli, "Decentralized network coding," DIMACS, June 13, 2003.

P.D. Grünwald and J. Langford, "Suboptimal behaviour of Bayes and MDL in classification under misspecification," *Seventeenth Annual Conference on Computational Learning Theory* (COLT 2004).

P. Grunwald, "Suboptimal behaviour of Bayes and MDL in classification under misspecification," *2005 Graybill Conference*, Fort Collins, Colorado.

Y. Hanatani, T. Horiyama and K. Iwama, "Hajos calculus on planar graphs," *4th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications*, 2005.

K. Iwama and S. Tamaki, "Improved upper bounds for 3-SAT," *SODA 2004*, 2004.

G. Kramer and S.A. Savari, "Cut sets and information flow in networks of two-way channels," *2004 IEEE Int. Symp. Inform. Theory*, Chicago, IL, June 27-July 2, 2004.

J.N. Laneman and G. Kramer, "Window decoding for the multiaccess channel with generalized feedback," *2004 IEEE Int. Symp. Inform. Theory*, Chicago, IL, June 27-July 2, 2004.

G.I. Shamir and J.J. Boutros, "Non-systematic low-density parity-check codes for nonuniform sources," *IEEE International Symposium on Information Theory*, ISIT-2005, Adelaide, Australia, Sept. 4-9, 2005.

E. Soljanin, "Network coding: From graph theory to algebraic geometry," Columbia University, February 2004.

E. Soljanin, "Network coding based on subtree decomposition," Information Sciences Seminar, CalTech, September 10, 2004.

Websites

Website for Computational Information Theory and Coding
http://dimacs.rutgers.edu/SpecialYears/2001_COD/

VIII. CONTRIBUTIONS WITHIN DISCIPLINE

This project has been inherently interdisciplinary. The project involves both fundamental theoretical issues and major outreach components to other communities and other parts of computer science and mathematics. This diversity of programs, feeding off of each other, led to many of the synergies that make DIMACS successful. Computational Information Theory and Coding explored the interface among coding theory, information theory, and parts of computer science and mathematics.

The interdisciplinary emphasis has led to a variety of collaborations, often among people with quite different interests who had never before collaborated or even known of each other. Here are a few selected examples of the collaborations that have been fostered among participants.

"During a meeting on coding and information theory a new research topic has been identified in a discussion between G. Kabatiansky (IPPI, Moscow, Russia) P. Narayan (U. Maryland), and myself (A. Barg, U. Maryland). The topic concerns a link between fingerprinting digital data and a class of problems for a multiple access channel. First capacity-type results for this problem were obtained since."
Alexander Barg
University of Maryland

"I attended the Working Group Meeting and Workshop on Theoretical Advances In Information Recording as having written papers asserting that error-correcting codes were essential to the faithful communication of genetic information, in order to counteract the cumulative effect on DNA of errors due to radiations. It was a pleasure for me to discuss these ideas in such a competent and friendly working group. Meeting Ms. Olgica Milenkovic (University of Colorado), a young and talented researcher, was especially gratifying for me, since she based some of her own research on my earlier works. We had some interesting talks at the meeting and continued to correspond by e-mail. More generally, the meeting and workshop have been the opportunity for informal talks with many of the participants, which I found fruitful and pleasant... I am personally convinced that mixing people and disciplines is the best way to promote the advancement of science, so this kind of meeting is extremely useful."
Gerard Battail
E.N.S.T. Paris (retired)

"Interaction with Jonathan Miller (Baylor College of Medicine) who also attended DIMACS has resulted indirectly in an RNAi related cross-discipline publication (currently being written) and a panel session at IEEE ICCCN 2004 led by Stephen F. Bush (GE Global Research) and including Jonathan Miller as a panelist. A paper is in-progress on a homology-driven boolean network as a framework for analyzing RNAi performance."
Stephen F. Bush
GE Global Research

"There certainly have been a number of new collaborations and research ideas. In my personal case, I started successful collaboration with John Langford (IBM) on one of the main themes of the workshop. We showed - to the great surprise of some of the other workshop participants - that MDL and Bayesian methods for machine learning can be suboptimal in certain problems, even with an infinite amount of data. There are several plans to continue this line of research. The program has certainly had an influence on my career, giving me a chance to get to know people from the prestigious Toyota Technological Institute and the Wharton School of the University of Pennsylvania and Duke University. I am now in regular contact with all of these people, and I am sure that new collaborations will follow with some of them."
Peter Grunwald
CWI

"The DIMACS meeting on network coding lead to several new ideas and collaborations for me. After conversations with other meeting participants, we discovered a new research result showing that network coding can lead to a linear improvement in bandwidth for k-pairs communication problems in directed graphs... The discussions at DIMACS also motivated our development of Informational Dominance… Our discussions at the workshop also lead to a collaboration with Kamal Jain (Microsoft), resulting in a joint submission to the ACM/SIAM Symposium on Discrete Algorithms, 2006."
Nicholas Harvey
PhD Student in Theoretical Computer Science
Massachusetts Institute of Technology

Emina Soljanin (Bell Labs) and one of the Computational Information Theory and Coding Special Focus visitors Christina Fragouli (DIMACS Visitor, EPFL) started working on (at that time very new) area known as Network coding. Subsequently, Emina won a 5-year NSF medium ITR grant for Network Coding with G. Kramer and P. Gupta (Bell Labs), R. Koetter (Univ. Ill., Urbana-Champaign), M. Effros (Caltech), M. Medard (MIT) and D. Karger (MIT)). The DIMACS visitor Christina Fragouli won an FNS (Swiss counterpart to NSF) award to do research on network coding for a year.

IX. CONTRIBUTIONS -- OTHER DISCIPLINES

As noted in the section on Contribuitons Within Discipline, this project was inherently interdisciplinary. Among the most important new directions of work stressed were connections of information theory and coding to statistical physics.

The Workshop on Theoretical Advances in Information Recording served as a platform for work by Christopher Rose and Gregory Wright on extraterrestrial communications, which later became the cover story in the September, 2004 issue of *Nature*.

X. CONTRIBUTIONS -- HUMAN RESOURCE DEVELOPMENT

Many of the comments in the section on Contributions within Discipline illustrate the human resource development contributions of this project. A major contribution is the impact on the research programs and careers of the participants. This project fostered new collaborations both within disciplines and among disciplines, both within academe and between academe and industry. These collaborations often involved students and helped shape the direction of their research careers. Here are some examples.

"The two workshops I attended (one on coding theory and one on network info theory) were very useful in giving me a good overview of the current activity and interesting issues in these areas. The workshops were very well arranged and the choice of speakers was excellent. These workshops put forth many open questions before us and have helped me in formulating a research problem for my PhD. Meeting other students and researchers in these areas was also very beneficial."
Shashi Borade
MIT

"I was one of the speakers for DIMACS Workshop on Network Information Theory. I would like to mention the impact it has had on my students. Three of my students (Deniz Gunduz, Zinan Lin and Melda Yuksel) attended the workshop. It was a great source of inspiration for all of them. They not only were able to meet some of the prominent researchers in the area, but they also learned a lot about the field. This was especially very good for them, because they were at the beginning of their PhD studies."
Elza Erkip
Department of Electrical Engineering
Polytechnic University

“I (and my students) attended 3 of the workshops: Signal Processing for Wireless Transmission, Workshop on Algebraic Coding Theory and Information Theory and Working Group Meeting on Network Coding. Although all 3 were inspiring and assisted me in my research and in terms of getting exposed to new ideas, the first and third of the above were instrumental in defining the theses for two of my students: Radhika Gowaikar and Amir F. Dana.”
Babak Hassibi
Caltech

“The program was wonderful and one of the best conferences I have ever attended. For me, it has led to new collaborations, new research ideas, and better understanding of current research results. I am finishing my thesis on "Distributed Detection and Coding in Information Networks." I will be finished this term.”
Shan-Yuan Ho
MIT

“I participated to the Workshop on Complexity and Inference and was financially supported by DIMACS. Although the workshop did not contribute directly to my PhD, it did help me a lot as a junior researcher. I got acquainted with theoretical information theory, which was a lack in my education. Moreover, the contact with the senior researchers in the field inspired me a lot. I will use the theory to found the 'inference' part of my PhD (which I'll finish next year) on a philosophical & theoretical way.”
Jan Lemeire
PhD-student VUB, Brussels, Belgium

“I am a graduate student working with Prof. Ralf Koetter at UIUC. I had the opportunity to attend the DIMACS workshop on Algebraic Coding Theory and the workgroup meeting on Network Coding. I would like to begin by thanking the organizers (for giving me) an opportunity for me to attend the events, where I was exposed to many (new) and different ideas. I met Dr. Gerhard Kramer in the workshop on Algebraic coding theory and later went on to do an internship at Bell Labs, Lucent Technologies with him. This is a very fruitful collaoboration, which is ongoing.”
Niranjan Nayak Ratnakar
University of Illinois

“First of all I would like to thank you for the excellent workshop (I joined the network information theory). The workshop was the first conference I attended as a Phd student and therefore it showed me a lot of new problem areas and research topics. I also learned some new methods and concepts. Afterwards I started to study some more publications of some presenters. Today I'm working on a concept for relay networks and I mention that I was inspired by the algebraic network coding approach from Koetter and Medard, which I have seen first at your network information theory workshop…In the future, I still work on the linear relay coding concept, so that the workshop had substantial impact on my work today.”
Tobias J. Oechtering
Technische Universität Berlin

“Both Prof. Joseph Boutros from ENST, Paris, and myself (from the University of Utah) participated (in) the DIMACS Workshop on Algebraic Coding Theory and Information Theory on December 2003. Discussions that we had had during that workshop started up a very fruitful contact between the two of us that within the following two years led to a visit that I had in ENST, which led to a joint research project on design of channel codes for coding redundant sequences. This joint work already led to two conference publications, one in the International Symposium of Information Theory - ISIT-2005, and the other in the Allerton Conference - 2005. We are constantly working together now to advance this research, where the interaction now also includes work by graduate students from both departments.”

Gil Shamir
Assistant Professor
University of Utah

"Although there has not been any direct research output resulting from participating (in) the workshop (network coding), I find the workshop extremely valuable to me as well as other participants. It provided me with a unique opportunity to explain some difficult part of my work in great detail, which cannot possibly be done in a regular conference where new research results are expected. I believe the same applies to other presenters, and the audience benefit (among them are students) a lot."
Raymond Yeung
The Chinese University of Hong Kong

"My stay at DIMACS gave me the opportunity to look at new research directions, and investigate interdisciplinary problems, that involve networking, coding theory, graph theory, algebra and discrete mathematics. This would not have been possible without the support of DIMACS."
Christina Fragouli
DIMACS visitor, EPFL

"I continued working with Gerhard Kramer and we interacted with Bruce Shepherd and Chandra Chekuri. Gerhard visited me at Michigan and we continued our collaboration… I am a new professor and didn't have outside funding during my two visits (to DIMACS.) I now have my first NSF grant, which builds off the work that Gerhard and I did in 2003."
Serap Savari
Lucent Technologies

XI. Contributions to Resources for Research and Education

Gerhard Kramer developed and taught a course on "Topics in Multi-terminal Information Theory." This topic was directly related to the subject matter of the Workshop on Network Information Theory that he co-organized as part of the Special Focus. Gerhard taught the course internally (for Bell Labs researchers, summer students, and visitors) and externally. He then developed a university course on this topic.

XII. Contributions Beyond Science and Engineering