WORKSHOP REPORT:
NSF Sponsored Workshop
on Joint Source-Channel Coding:

October 1999

DRAFT 3

Organizers:

Michelle Effros, California Institute of Technology, Co-Chair

Kenneth Zeger, University of California, San Diego, Co-Chair

Participants:

Toby Berger, Cornell University

Philip Chou, Microsoft

Nariman Farvardin, University of Maryland

Thomas Fischer, Washington State University

William R. Gardner, Qualcomm

Robert M. Gray, Stanford University

Nikil S. Jayant, Georgia Institute of Technology

Rajiv Laroia, Lucent Technologies

Upamanyu Madhow, University of Illinois at Urbana-Champaign

Michael W. Marcellin, University of Arizona

James W. Modestino, Rensselaer Polytechnic Institute

David L. Neuhoff, University of Michigan, Ann Arbor

Alon Orlitsky, University of California, San Diego

Kris Popat, Xerox PARC

Kannan Ramchandran, University of California, Berkeley

James A. Storer, Brandeis University

Vinay Vaishampayan, AT&T Labs Research

Zhen Zhang, University of Southern California

Program Directors:

John H. Cozzens, National Science Foundation

Rodger Ziemer, National Science Foundation

The views expressed here are those of the participants of this workshop and do not necessarily reflect the views of any organization public or private.

I Executive Summary

As described in the recent report of the President's Information Technology Advisory Committee (PITAC), improvements in information technology are transforming the way we work, learn, communicate, access information, practice health care, buy and sell goods, conduct research, understand our environment, build things, and run government agencies. Behind many of these developments and the developments needed for future information technologies are past and future advances in communications.

For the past 50 years, the field of communications has developed under the assumption that the two basic components of the communications problem can be performed independently with no performance degradations relative to joint design. Briefly, this idea, known as the "separation principle," states that "source coding" – or efficient information representation – and "channel coding" – or protection of information from corruption by noise – can and should be performed independently. By dividing a single difficult problem into two simpler problems, the separation principle has led to 50 years of enormous advances in communications theory and technology. Examples of technologies made possible by communications advances include CD ROMs, fast modems, wireless communication systems, and the Internet.

Ironically, many of the modern information technology systems made possible by advances achieved through the use of the separation principle actually violate the conditions upon which the optimality of that principle relies. For such systems, performance improvements may be achieved by moving from separate design and operation of source and channel codes to "joint source-channel code" design and operation.

The following report describes discussions and findings of the NSF-Sponsored Workshop on Joint Source-Channel Coding. The workshop was held in San Diego, California on October 12 and 13, 1999. Twenty-two experts, coming from a variety of different disciplines and points of view, participated in the workshop. The goals of the workshop were to: provide a contextual basis for understanding the array of applications in which source-channel coding can play a central role; assess the potential impact of source-channel coding technology in those applications; identify and characterize the major open research problems associated with the design, development, and deployment of source-channel coding technology; and develop a conceptual framework for future cross-cutting research activities in this inherently interdisciplinary area.

The committee's results include a careful description of system properties that lead to the violation of the separation principle. These properties include existence of critical resource constraints (such as power constraints for hand-held wireless devices), networks (such as the Internet) in which users must share channel resources, applications (such as multimedia communications) characterized by source or channel heterogeneity, and systems (such as mobile communications devices) with unknown or time-varying source and channel properties. The committee's assessment is that these properties pervade an enormous array of information systems, including many key network technologies. Wireless communications systems and the Internet are two prime examples considered in this report. Given the pervasive nature of these properties, the committee recommends focused research in this area in order to: (1) advance the theory needed for characterizing the performance and fundamental limitations of joint source-channel coding systems, (2) develop methods for designing joint source-channel coding systems, and (3) apply the theory and design techniques to address practical problems. The committee also puts forward recommendations for encouraging productive research interaction between industry and academia in this area of enormous interest and importance to both.

II Introduction and Motivation

Advances in information technology are transforming the way we live our lives. The impact is pervasive, affecting everything from communications to health care, commerce, and education. Since much of information technology is, at its core, communications technology, developments in communications represent key enabling advances for the information technology revolution currently underway.

The field of communications describes the theory, methodologies, and technologies for transmitting and storing information. The information considered includes sources such as text, speech, images, audio, video, and scientific data. Traditionally, the notion of a communications system divides into two separate functions or subsystems. The first, called a "source code" or "compression algorithm," aims to provide an efficient representation of the information; efficient data representations speed information transmission and minimize information storage-space requirements. The second, called a "channel code" or "error control code," works towards correcting or concealing errors arising from flaws in the storage or transmission system; channel codes protect information from corruption due to effects like electronic noise in a circuit, damage to a storage device (such as a scratch on a compact disk), loss of information packets in a network, and failure of a communications link.

The separation principle, which divides a single communications system into separate source coding and channel coding subsystems that are then designed and run independently, is both extremely powerful and extremely limiting to the advancement of communications technology. The power of the separation approach is that it divides a single complex problem into two simpler problems. This divide and conquer strategy has led to a wide variety of advances in both the theory and the practical design of source coding and channel coding systems. Theoretical advances include greater understanding of optimal source coding and channel coding methodologies and their performance. Practical advances include good, fast algorithms for source coding and channel coding. For example, modern compression systems typically halve the file size or transmission time needed for text and reduce by a factor of 40 or more those required for good image reproduction. Similarly, advances in channel coding have led to channel coding methods achieving reliable information communications and storage; in fact, the performance of these techniques is nearing the theoretical limits for simple common channels (e.g., single-user Gaussian noise on linear channels). These advances have enabled a wide array of information technologies, including CD ROMs, fast modems, wireless communications, the Internet, and digital libraries.

The power of the separation principle is balanced by its limitations. The use of separation is theoretically motivated; the "separation theorem" demonstrates that under idealized assumptions, there is no performance penalty associated with designing independent source and channel codes. Yet, despite this result, the separation approach does not guarantee the best possible performance for all systems. In particular, proof of the optimality of separation treats only issues of file size and reliability – ignoring such practical design criteria as computational complexity and delay. As a result, communications systems built from independently designed source codes and channel codes may require greater computational resources and cause higher delay (or "latency") than jointly designed systems. Further the idealized assumptions require a single-transmitter, single-receiver ("node-to-node") system and known source and channel distributions at code design time. Since these conditions are often violated in modern communications systems (e.g., multiuser networks like the internet fail to meet the node-to-node condition while the time-varying channels of mobile communications systems violate the known-channel assumption), the separation theorem fails to apply in a wide array of practical applications. Finally, the separation principle ignores the imperfections observed in real communications systems. In particular, source codes are typically designed assuming that the channel code will correct *all* errors introduced by the channel. Unfortunately, this assumption of a perfect channel code leads to extremely fragile source codes in which a single bit error can potentially yield catastrophic source code failure. This fact is unfortunate in real communications systems where uncorrected errors are inevitable (although often hidden from the end user). Similarly, channel codes are typically designed assuming that all bits created by the source code are equally important. This approach, which results from the assumption of a perfect source code, leads to the design of channel codes that protect all bits equally. This fact is unfortunate since many practical source codes output bit streams in which the bits are not of equal importance. As a result of all of these factors, reliance on the separation principle may lead to performance degradation relative to what could be achieved through joint design.

A communication system in which the source code and channel code are designed or operated in a dependent fashion is called a "joint source-channel code". Unlike separation-based techniques, joint source-channel code design techniques rely on the joint or cooperative optimization of communication system components. Further, the joint source-channel coding approach allows for strategies where the choice of source coding parameters varies over time or across users in a manner that in some way depends on the channel or network characteristics. Likewise, joint source-channel coding techniques allow for systems in which the choice of channel code, modulation, or network parameters varies with the source characteristics. In short, joint source-channel coding applies to any system in which the conditions of the separation theorem are violated and thus dependence of the source and channel code arises.

Given the above observations about the conditions under which the assumptions for separation fail, we next list four examples of conditions under which we expect significant gains from joint source-channel coding. First, gains are expected for systems with critical resource constraints. Examples of potential resource constraints include rate or bandwidth for data transmission, complexity for low-cost systems, power for hand-held devices, and delay for real-time systems. Joint source-channel coding yields performance improvements in such scenarios by allowing for optimal allocation of a single user's resources between the source code, channel code, and modulation scheme of his system. The goal here is to choose the allocation that yields the best end-to-end system performance. For example, in a system in which power is a critically limited resource, the power required to compress a signal should be no larger than the power savings achieved by reducing the source description length. Likewise in "real-time" systems, where delay limitations represent a critical constraint, the delay introduced by the source code must be justified by a greater reduction in the time required to transmit the data using the more efficient data representation. By encouraging careful tradeoffs between the resources consumed by the source code and those consumed by the channel code and modulation schemes, critical resource constraints create fertile environments for the application of joint source-channel coding techniques.

A second example of a scenario where the potential performance benefits of joint source-channel coding are high is multiuser systems with shared channels. Wired data networks, where packets of information travel through a sequence of shared links on their way from their point of origin to their point of destination, are one example of this type of shared-resource environment. Wireless communications systems, where users attempting to send information through the system compete for the limited system bandwidth, are a second example. In systems in which users share a common channel, the signals transmitted by different users can interfere with each other. Thus one user's "source" contributes to the noise in another user's "channel." The competition for network resources and the dependence of the channel on the sources feeding information into the channel make shared-channel environments prime candidates for application of joint source-channel coding solutions. In particular, while joint design of source and channel codes across users in a shared-channel environment is typically not possible, some form of cooperation can limit interference between users and thereby lead to better overall system performance. System protocols establish a framework in which these types of cooperation take place. For example, in packet-based networks, system protocols describe methods for breaking a single user's information into packets and routing those information packets, along with the packets from other users, through the network. In some sense, the protocol becomes the "channel" through which the data description passes. Here joint source-channel coding techniques lend insight into how to achieve better system performance. In particular, we can engineer the source to match the channel by modifying source rates to send more information when network congestion is low and less when network congestion is high. Likewise, we can engineer the channel – or the network protocols – to match the source statistics. For example, in packet networks we can use the bits in the headers of data packets to instruct the routers on which packets have low priority and can be dropped if necessary to avoid overflow. This approach, which is analogous to existing methods for using bits in the data packet body to describe to the source and channel decoder how to decode a particular transmitted message, allows for control of the channel based on the importance of the information in the source. These joint source-channel coding approaches for engineering sources that yield better channels and channels that behave in ways better suited to the source are a natural match for shared-channel environments.

A third basic scenario in which significant gains are expected from the application of joint source-channel coding methodologies is in multiuser systems with heterogeneous sources, channels, topologies, or users. Source heterogeneity arises from the fact that different data types traversing a single communications system have different sensitivities to loss, corruption, and delay. Channel heterogeneity arises from the fact that different channels within a single network may have different noise characteristics, available rates, and, in packet networks, delay and jitter.[1] User heterogeneity refers to the differences between the needs and quality-of-service requirements of different users, the different computational capabilities of their systems, and the different bandwidth access available to them. Heterogeneity of topology refers to the fact that the same network can be simultaneously used as a point-to-point communication system (with or without feedback), a broadcast system, a multiple access system, and so on. A variety of joint source-channel coding techniques may be employed in networks exhibiting these types of heterogeneity. For systems with heterogeneous sources, unequal error protection can be used to protect more error- or

[1] Jitter refers to the differences in delay from packet to packet.

delay-sensitive data from corruption or the need for retransmission. In sources with heterogeneous channels or users, layered or multiresolution source coding allows transmission at multiple rates (for multiple channels) or makes information simultaneously accessible with different reproduction qualities to multiple users with varying needs and system capabilities; the result is a "democratization" of information access, where information can be made available to all users despite the potentially very different bandwidths and computational capabilities of their access devices. Finally, network topology can affect the choice of source and channel codes. For example, separation is known to fail for some network topologies even when the sources and channel are known and the system delay and complexity are unconstrained. Further, feedback in a network system topology can affect source and channel coding choices. For example, using feedback to request repeated transmission of missing or corrupted data packets can obviate traditional channel coding; prioritization of packets by relative importance of the included source content can then lead to prioritized repeat requests for scenarios where delay constraints are tight. The matching of source coding parameters to heterogeneous channels and networks and channel coding parameters, modulation techniques, and network protocols to source characteristics in these heterogeneous environments are inherently joint source-channel coding approaches.

The fourth and final category of problems in which joint source-channel coding is expected to achieve significant gains is applications characterized by unknown or time-varying sources, channels, or networks. The issues arising here are the same as those considered for systems with heterogeneous sources, channels, networks, or users, and the techniques that arise to combat these problems are likewise similar. The difference here is that the potential variation occurs across time for a *single* source, channel, or network rather than varying across space or system use. The goal for time-varying systems is to achieve graceful performance fluctuations as the system evolves and changes with time.

Examples of existing technologies and applications in which the above-described characteristics appear are bountiful. For example, second generation (G2) voice-oriented digital cellular phones combine critical resource constraints, a shared-channel environment, and extreme channel variations over time and space. Critical resource constraints include power constraints for hand-held devices and delay constraints for real-time voice communications. The very noisy shared channel environment arises in part from extreme competition for bandwidth, a problem that increases as cellular systems increase in popularity and use. Further, wireless channels experience extreme variations over time and user location due to changes in such characteristics as distance from base stations, interference from other users, shadowing effects, and fading characteristics. The combined effect of all of these characteristics makes digital cellular telephony a prime candidate for the application of joint source-channel coding techniques.

Another existing system exhibiting many of the above characteristics is the Internet. Due to its shared-channel environment and heterogeneous components, the Internet is characterized by extreme channel variations over space and time. Variations over space describe the variations between the channels observed by different sender-receiver pairs in the network. These variations can be extreme, with bandwidths ranging from 10s of kbps (kilobits per second) to 10s of Mbps (megabits per second), packet loss probabilities varying from fractions of a percent to 25% or more, and delay and jitter varying from milliseconds to seconds. In addition to these variations over space, are variations over time. A single link's bandwidth, packet loss probability, delay, and jitter can vary enormously over time as competing processes begin and end. In distance learning, for example, downloading a web page during a real-time video transmission can cause severe performance degradations. User mobility adds further heterogeneity to the system, bringing with it all of the time and space variations described above for cellular telephony systems. Finally, network topology further complicates the issue since the Internet is simultaneously used for point-to-point, broadcast, multiple access, and feedback communications applications.

Other examples of current systems and applications characterized by the above properties of critical resource constraints, shared channels, system heterogeneities, and time variance are numerous. The tight delay constraints of video telephony, video conferencing, and video-on-demand applications create a resource-constrained environment. In video broadcast applications, differences between the systems and links of different users create a system characterized by user heterogeneity. The desire for ubiquitous access, fast browsing, and easy downloading in a large data bases such as a digital library creates another heterogeneous user and channel environment. By failing to meet the separation theorem's conditions for optimality, each of these applications becomes a candidate for the application of joint source-channel coding techniques.

With advances in information technology, the properties that cause the separation theorem to fail are becoming more common. For example, G2.5 and G3 digital cellular systems will add simultaneous data transmission to their voice services, allow variable data rates (with up to 2 Mbps for stationary conditions and up to 144 kbps for high mobility), run the Internet protocol on top of the data link, and offer multimedia (image, video, and audio) services on top of the Internet protocol. As a result, these future systems will be characterized by even greater source, channel, and user heterogeneity and more time variance than G2 systems. Similarly, next generation Internet will add even greater channel and source heterogeneities to an already diverse environment. New Internet channel heterogeneities will arise from both the increase, by several orders of magnitude, of allowed transmission capacity for high-end users and the increase in use by low-data-rate mobile users. Increased use of multimedia applications over the Internet will likewise lead to greater source heterogeneities. Other future applications in which joint source-channel coding may play a significant role include last mile technologies, distributed sensor arrays, network transport of digital video and images, packet-switched video conferencing, multimedia delivery on digital subscriber loops, and digital audio delivery systems.

Some very rudimentary joint source-channel coding strategies are already in use in existing technologies and applications. For example, G2 voice-oriented digital cellular phones such as the North American and European TDMA digital telephony standards use unequal error protection to maximize decoded speech quality while meeting the required bit constraints. Error concealment, which takes advantage of speech redundancy by using interpolation to approximate missing or corrupted symbols, is also employed. It can be argued that the use of unequal error protection yields a 20 – 40% savings in aggregate data rates in these systems. As a result, the use of joint source-channel coding techniques like unequal error protection played a first-order role in the difficult but finally successful transition from analog cellular to digital cellular systems.

There are many other examples of the successful use of very simple joint source-channel coding techniques in existing systems. For example, error concealment is used in video telephony and conferencing to hide the effects of uncorrected channel errors. Video-on-demand systems use prioritized automatic repeat requests so that higher priority data such as audio may be reliably received at the expense of losing lower priority data such as video when high loss conditions are experienced. Finally, video broadcast applications very commonly use layered coding in order to simultaneously provide lower quality data reproductions to users with lower available data-rates and higher quality data reproductions to users with higher available data-rates using a single embedded data description.

The properties given above as examples of appropriate scenarios for the use of joint source-channel coding and the joint source-channel coding techniques briefly described there barely scratch the surface of the problem and its solution. A more detailed description of a single system gives a better picture of the system characteristics of which joint source-channel coding may eventually take advantage.

Example: The Internet

As described earlier, the Internet is an example of a network in which there are tremendous channel variations over time and space. Not noted earlier, however, is the fact that the channel variation is also correlated in time and space. The correlation in time arises from channel memory. Much of the packet loss experienced in packet-based channels arises from buffer overflow. Since buffers, once full, take time to drain, packet losses in a network often occur in bursts. The correlation in space arises from the shared-channel environment. For example, two neighboring paths through a network may share one or more nodes or links. Overflow in a shared node generally results in lost packets for both channels. Correlations may also exist for two paths not sharing any nodes since both may be affected by high transmission rates on a third path that crosses both of them.

These characteristics create many challenges and opportunities for joint source-channel coding in networked environments. Even in the simplest case, with only a single sender and a single receiver, the channel between the sender and receiver is not known a priori, is in fact time varying, and the memory of the channel is usually large compared to the delay constraint for the application. If there is a single sender broadcasting (actually, multicasting) to multiple receivers, then the sender must deal with coding for a heterogeneous mixture of unknown but correlated channels. This is the problem of joint source-channel coding for the broadcast channel. One way to deal with this problem is with feedback from the receivers to the sender. Each receiver can report, for example, its average transmission rate, loss rate, delay, and jitter. Such receiver reports are highly correlated because the network paths between adjacent receivers have many links in common. Although it is known that feedback from a single receiver to a single sender does not increase the capacity of a memoryless channel, the broadcast feedback problem involves multiple receivers and a channel with memory. In this environment, there is good reason to take advantage of feedback, which in a highly connected network is readily available (although over unreliable channels). Feedback from the multiple receivers to the single sender can be considered a problem in joint source-channel coding for the so-called multiple access channel. It is known that source coding for correlated sources plus channel coding for the multiple access channel is inferior, in general, to joint source-channel coding. In other words, in multiuser communication networks, it is known that the separation theorem does not hold, even when the sources and channels are known and the delay is unconstrained.

The transport of multimedia data across networks like the Internet poses even more problems. Existing reliable transport protocols (TCP/IP) are matched to data applications requiring guaranteed robust performance. To guarantee robust performance, TCP/IP requests retransmission of all lost or corrupted packets, thereby achieving the desired protection from noise at the expense of an imposed delay. Unfortunately, the resulting delay makes these protocols a poor choice for real-time data traffic, which is characterized by high delay sensitivity and variable tolerances to noise. In the networking field, two approaches to this problem have emerged: integrated services and differentiated services. The integrated services approach stresses guaranteed qualities of service for different service classes, while the differentiated services approach stresses different, though not guaranteed, qualities of service for different service classes. Guaranteed quality of service is more difficult to implement and consumes more network resources, but provides stable channels for traditional separation-based source coding algorithms such as MPEG video. On the other hand, differentiated services, in conjunction with joint source-channel coding, could prove to be significantly more resource-efficient than integrated services. As an example, a simple differentiated services protocol that calls for lower priority packets to be dropped during congestion before higher priority packets would obviate the need for explicit channel coding and unequal error protection. This could result in a lower transmission rate, higher quality, lower delay, and lower network resource usage compared to a separation-based solution. In general, information placed in packet headers (such as priority tags, protocol parameters, or even programmatic scripts destined for active routers) can be used by joint source-channel coders to change the way different packets are handled by the network, potentially as a function of buffer delay, packet loss, changing transmission bandwidth, and so forth, either measured by the routers directly, or received as feedback from downstream routers.

The Internet provides a rich variety of research problems in joint source-channel coding. The solutions to these problems would likely have a major impact on networks of the future.

III Research Goals and Challenges

Three pressing research goals for joint source and channel coding are (1) develop appropriate theory for characterizing the performance and fundamental limitations of joint source-channel coding systems, (2) determine methods for designing joint source-channel coding systems, and (3) apply the theory and design techniques to address practical problems. A variety of more specific research objectives addressing these problems is listed below. These objectives represent many of the significant research goals in joint source-channel coding. The order in which these items are listed reflects no prioritization over the described projects. Much of this research is expected to be multidisciplinary in nature, combining expertise from such areas as signal communications and information theory, networking and queuing theory, as well as input from areas including economics and psychology.

Performance and Fundamental Limits

Many of the major challenges in developing appropriate theory for joint source-channel coding are related to understanding the tradeoffs between the quality, delay, complexity, channel usage, and power inherent in code design. Quality refers to the end-to-end distortion between the original data and its reproduction. Delay refers to the difference between the times at which data enters the encoder and exits the decoder. Complexity refers to the requirements of the encoding and decoding algorithms, including such things as number of arithmetic operations per data sample and memory use, which may be heavily influenced by implementation considerations. (Complexity is also a primary factor in determining how much power a particular joint source-channel coding implementation consumes.) Channel usage refers to issues such as the number of channel uses per sample, channel bandwidth, and channel signal-to-noise ratio (which includes both channel noise and transmit power); both implementation and transmit power use become important factors in energy-limited applications (e.g., battery-powered devices). It is also important to limit transmit power when one user's transmitted power can significantly impact the performance of others.

While development of a comprehensive theoretical framework for joint source-channel coding is a suitable long-term goal, it is evident from prior work that completely characterizing the relevant tradeoffs between quality, delay, complexity, channel usage, and power is extremely difficult, and that achieving progress in this area is very challenging.

Potential projects in this area include:

- The development of a theoretical framework for joint source-channel coding, suitable for both point-to-point communications and network communications, that includes the relevant tradeoffs between quality, delay, complexity, channel usage, and power. Such a framework is needed for both lossless and lossy data requirements, and for situations where synchronization is an issue.

- The development of performance measures appropriate for joint source-channel code design and evaluation, possibly including such things as performance bounds or different quality metrics for lossless and lossy coding applications.

Models

Modern communication systems are complex. Analytical research is typically based on the study of relatively simple models (abstracted from physical or idealized systems). There is great need for the development of tractable source and channel models for use in joint source-channel coding research. The models developed should be rich enough to capture the features and constraints of the problems at hand yet simple enough to provide the insight necessary for algorithm development. Such models could be developed progressively from consideration of simple point-to-point communication systems to more complex multiuser networks and should be driven by corresponding developments in theory.

Models needed include:

- Models for speech, video, images, audio, and text that can be used to devise better source codes;

- Models for the output of source encoders for use in the design of channel codes (e.g., for unequal error protection);

- Models for wireless channels that capture the effects of mobility and multiple access;

- Models for Internet traffic that can be used when devising end-to-end protocols.

Design

Systematic design methodologies are required to broaden the applicability of joint source-channel coding techniques. Potential projects include the development of design methodologies that addressing the following problems:

- Channel-optimized source coding, including error concealment and recovery methods;
- Source-optimized channel coding, including unequal error protection methods;
- Synchronization;
- Feedback in source-channel coding;
- Coding for time-varying sources or channels;
- Multiple description coding;
- Robustness to model variations;
- Flexibility to different data types.

Applications

Joint source-channel coding theory and design techniques should be applied to address practical problems, including the design of joint source-channel codes for multiuser or network systems. One challenge to the accomplishment of this goal is the fact that network protocols attempt to isolate link and applications layers – a separation-style approach. In contrast, joint source-channel coding inherently involves interaction between the link and applications layers. A challenge of joint source-channel coding is to find ways for the application layer to crack through the network stack all the way to the link layer. Another challenge is to use joint source-channel coding research to inform protocol design.

Specific projects include:

- Evaluation of the effects of network protocols, including the need for networks to handle a variety of data sources;
- Evaluation of the effects of packetization and routing;
- Re-examination of the typical network separation of channel coding into the physical layer and source coding into the application layer;
- Application-specific needs for asymmetry in encoding and decoding complexity;
- Power allocation in shared-resource environments.

IV Roles of Academia and Industry

Little is currently understood theoretically about joint source-channel coding, and advances will require contributions from both industry and academia. It is therefore of interest to determine how best to encourage productive interaction between industry and academia in the context of joint source-channel coding. Several characteristics distinguish this topic, and may have bearing on interactions between industry and academia. Foremost, the problem area is of great immediate interest to both academicians and members of industry. In addition, joint source-channel coding requires systems-level design and optimization, which industry is perhaps more inclined towards than academia. Further, it is an area in which applications are expected to strongly motivate and inform the theory, and vice-versa. Finally, it is an area where fast-paced development outstrips theoretical understanding.

The emphasis and role of research in this area is likely to be different for industry and academia. The pace of development in the wireless and Internet industries, as well as the competitive climate in these areas, has largely necessitated the allocation of research resources on the part of industry to shorter-term goals. However, there are important, fundamental, longer-term research topics that need to be pursued if the technology is to advance to its potential. As academics take up more of the fundamental research agenda, it is important that they work closely with industry for a number of reasons. First, close ties between industry and academia allow for fundamental research based on practical insights into where the most important research problems lie. Second, these ties allow for design choices informed by an understanding of their theoretical implications.

Areas of interaction

The time from algorithm development to implementation is rapidly decreasing due to advances in processor speeds and memory as well as feature size reduction in VLSI. It is therefore becoming possible to devise sophisticated algorithms for end-to-end optimization, whether at the level of a single wireless link, or of an end-to-end Internet protocol. As discussed earlier, the development of good source and channel models is critical to such end-to-end optimizations. Given the wealth of real-world experience available in industry and the traditional academic strengths in abstract reasoning, model development is a prime area for industry-academic collaboration.

Another broad area for industry-academic cooperation is in the formulation of design criteria. Since the objective of joint source-channel coding is to perform end-to-end optimization, an understanding of the constraints facing the engineers who will ultimately build the system is important even in the earliest phases of algorithm development. For example, the need for low power consumption in small hand-held devices implies certain limitations on the complexity of transmitter and receiver algorithms in such devices. On the other hand, advances in VLSI may temper such limitations in the long term. Such rapid changes in technology make close interaction between academia and industry a must in formulating meaningful problems. Another example pointed out at the workshop was that, from an overall systems perspective, it is often preferable to provide a flexible framework that allows for system-wide optimization, rather than over-optimizing for a specific source or channel model. Multimedia transmission over heterogeneous networks is an example of a broad class of applications that could benefit from such a flexible framework. Industry-academic interactions allows such tradeoffs between flexibility and optimality to be made in an informed fashion.

Finally, while the development of fundamental theory and algorithms is, of necessity, based on somewhat idealized models, simulation and testing in a realistic environment is an area in which industry has substantially more resources than are typically available in university laboratories. Such performance evaluation is therefore a major area of potential collaboration between industry and academia, especially because it is the first step to potential technology transfer. An example of this is the implementation and evaluation of joint source-channel codes in a wireless mobile environment, which would not be possible to recreate at a large scale in a university environment.

Modes of Interaction

There are many fruitful ways in which academia and industry can interact. Perhaps the most obvious is through joint research projects. These can take the form of true collaborations between engineers from both organizations, but more often consist of academia performing research and reporting results to industry. This approach is often a result of the need for industry to remain silent about research approaches that they believe will be important to short term competitiveness. Indeed, even the knowledge of which problems need attention in the short term can represent a competitive advantage.

Perhaps the most serious impediment to joint research projects between academia and industry is the intellectual property (IP) issue. The usual opening stance of both organizations is sole ownership of all IP, with no rights for the other party without additional licensing fees. Much attention has been focused on large efforts involving centers, with relatively less attention for smaller (e.g., single investigator) projects. Some progress has been made on both fronts, and IP agreements have been reached in many cases. Nevertheless, much work remains to resolve this thorny issue.

Other modes of interaction between academia and industry exist and can be quite fruitful. Many of these involve exchange or visits of personnel. Internships for students or faculty in industry can provide perspective, direction, and research ideas to be followed upon return to academia. Likewise, visits by industrial researchers to academic labs can benefit the university lab as well as the industrial visitor. While relatively uncommon in U.S. industry, Japanese and Korean engineers often visit university research labs. While similar to internships (although usually more focused and shorter term), faculty consulting can provide similar benefits to both academia and industry. On the other side of the equation, industrial practitioners frequently teach courses at universities, and can co-adviser for theses and dissertations. One final and highly successful mode of academic/university interaction is through entrepreneurship and start-up companies.

There are a variety of ways in which the NSF fosters cooperation between academia and industry. Perhaps the most direct is via matching funds that leverage industrial contributions. Currently, most of the matching is done via large NSF supported centers (such as the ERC and UICRC). We encourage funding for smaller scale programs (e.g., individual PI) such as the GOALI and CAREER programs. Programs that sponsor industrial implementation of research (e.g., SBIR) help spur innovation and move ideas from academia to industry and should be encouraged as well.

On a more intellectual side, the NSF in cooperation with joint industrial and academic panels can identify areas of mutual interest to both communities and promote them via calls for proposals. Finally, to ameliorate the intellectual property impediments mentioned earlier, the NSF can help establish general guidelines for IPR agreements. This could be done either by publicizing existing standards or conventions, or by convening industrial and academic representatives to devise a general framework, or even model agreements, for the sharing of intellectual property rights.

V Concluding Remarks

This report describes the findings of the 1999 NSF Sponsored Workshop on Joint Source-Channel Coding Research. The report and its recommendations address three main goals. First, the report provides a view of the conditions and scenarios in which joint source-channel coding can potentially achieve performance improvements over independent source and channel code designs. These conditions include resource-constraints, multiuser interactions, heterogeneity, and time-variation or uncertainty. Second the report describes key research goals and challenges in the field of joint source-channel coding. This summary includes specific suggestions for research leading to advancement of theory, design, and application of joint source-channel codes. Third, the report comments on the roles of academia and industry in meeting the research challenges discussed. In particular, the report suggests specific research problems for which collaboration could be particularly beneficial and suggests methods for encouraging productive interactions between companies and universities in tackling this problem of great interest to both.

The open research problems described in this report present formidable challenges for academic and industrial research. The motivation for tackling these challenges is both the advancement of basic understanding and the technological advances expected from these advances. The need to address these challenges is made urgent by both the speed at which communication technology development is taking place in this field and the enormous societal impact of the resulting information technology.