Extending the Conventional Information Theory: Theory and Designs of Layered Partially Decodable Codes

This project is aimed at extending the conventional information theory, which is based on the idea of reliably transmitting all data bits through communication channels, by differentiating different types of data and providing different levels of error protections and thus allowing different parts of the data to be decoded at different nodes in the network. The main challenge addressed in this project is to develop new notions of partial information, and new coding techniques to encode different types of data together, efficiently. Over a 3-year period of research, significant progresses have been made in the following areas: 1) new analysis tools have been developed to establish the fundamental performance limits of layered codes, in terms of the tradeoff between the rates and reliability levels of different parts of data encoded together; 2) new applications of embedded coding in networks, including efficient transmission of network control messages and channel state information, have been investigated, showing significant improvements of the overall performance; 3) new connections to information geometry is observed, which leads to novel approaches to classical network information theory problems, a few instances have been reported.
Project Title: *Extending the Conventional Information Theory: Theory and Designs of Layered Partially Decodable Codes*

Contract No.: FA9550-06-1-0156

Principle Investigator: Lizhong Zheng, EECS, MIT

Personnel Supported:
- PI: Lizhong Zheng
- PhD Students:
  - Shashi Borade
  - Emmanuel Abbe

1 Abstract

This project is aimed at extending the conventional information theory, which is based on the idea of reliably transmitting all data bits through communication channels, by differentiating different types of data and providing different levels of error protections and thus allowing different parts of the data to be decoded at different nodes in the network. The main challenge addressed in this project is to develop new notions of partial information, and new coding techniques to encode different types of data together, efficiently. Over a 3-year period of research, significant progresses have been made in the following areas: 1) new analysis tools have been developed to establish the fundamental performance limits of layered codes, in terms of the tradeoff between the rates and reliability levels of different parts of data encoded together; 2) new applications of embedded coding in networks, including efficient transmission of network control messages and channel state information, have been investigated, showing significant improvements of the overall performance; 3) new connections to information geometry is observed, which leads to novel approaches to classical network information theory problems, a few instances have been reported.

2 Technical Report

During the 3 year period of this project, we have made progresses in two major direction, reported in the PhD theses of two students, Shashi Borade and Emmanuel Abbe, both graduated in the year of 2008. The areas are, respectively, the performance limits and designs of layered codes that allows partial information to be decoded, and the use of geometric tools in the analysis of network information theory problems. At the conceptual level, we have made the following main discoveries.

- The problem formulation of unequal error protections has been extended. We came to realize that there are multiple possible ways to differentiate the error protections of different parts of the message: one can have some subset of bits with higher priority, or could try to avoid some types of errors, misdetection or false alarm for example, to occur when certain subset of messages are sent. Either case has potential practical applications, and both are studied under our framework.
• The main tool we used to study these problems has been much better understood. We found that
the geometric approach we use can not only be applied to error exponent problems, but also to many
network information problems. We developed new theory of local geometric analysis, from where a new
notion of divergence transition is developed. The eigen vectors of such transition can serve as the basis
in the space of probability distributions, and help to bring in geometric insights to these problems.
We use such approaches to study non-Gaussian noise channels as well as some classical multi-terminal
communication problems.

• In applying the results of this project, we have made specific progress on compound channel and channel
with feedbacks. In both cases, strengthen the known results on the performance limits, utilizing the
new techniques developed in this project. We have found new geometrical proof of the entropy power
inequality, which might shed lights to a variety of problems. The specific coding scheme we designed
in this project is now adopted by some industry practice, as candidate coding scheme for some specific
components of commercial wireless networks.

In the following, we give some specific descriptions of the findings we made in this project, over the last
year.

2.1 Unequal Error Protection: Tradeoff between Rates and Reliabilities

The notion of partial decodability at different channel conditions can be easily formulated as unequal error
protection (UEP) problem: as the channel realization becomes atypically poor, one would hope to still decode
a part of the transmitted message, which gives this part of higher priority message a better protection.

Traditionally, unequal error protection and priority coding have been studied mostly by algebraic code
constructions, with little result on the fundamental performance limits, and are restricted to a few special
channels, such as the erasure channel. Moreover, the notion of unequal error protection is mostly limited
to protect some information bits better than the others, where through this project, we realize that in fact
much richer problems should be addressed.

The main results we obtained within the past year is what we believe a complete formulation of the
problem. We have two different ways to differentiating the error protection to different sub-messages. First,
the conventional definition, where we have two sub-messages, each with a certain rate, and their error
protections are measured by the corresponding error exponents; in the second formulation, we would consider
the data message as one choice among many candidates, and the error protection can be a requirement to
avoid a particular type of error when a particular message is chosen to be transmitted. An example of this
is that one can transmit a normal data stream, with an extra "fire alarm" message, where we would try to
have small misdetection probability for this special message. Note that this special protection is not towards
any bit, and thus is more general and fundamental. We refer to the first formulation as "bit-wise" UEP, and
the second "message-wise" UEP.

Along both approaches, we have obtained some nice new results. In terms of bit-wise UEP, we studied
the problem to encode two independent bit streams, at rate $R_1, R_2$, and to obtain the best tradeoff between
the decoding error probability in terms of error exponents $E_1, E_2$. We characterized the 4-dimensional
tradeoff between $(R_1, R_2, E_1, E_2)$ in general, and found that the analytical solution surprisingly simple and
intuitive for very noisy channels, i.e. channels with low capacity. In the later case, several concrete coding
algorithms following the theoretic results are developed, and some have adopted by industrial practice almost
instantly. For message-wise UEP, we considered the problem of transmitting one of $2^{nR}$ possible messages,
corresponding to $nR$ bits, where we require that there are $2^{nR}$ messages need to be protected better than
the others, i.e., conditioned on any of these special messages being sent, the probability of error should be
much smaller. We derived the outer bound of performance in this case, and found a universally optimal
procedure to apply conventional codes for this new problem.

2.2 Geometric Approach, Divergence Transition MAP

One important lesson we learnt from the research of error exponent is the use of geometric approaches in
information theory problems. The conventional approaches based on the combination of a menu of different
bounding techniques often hide the insights from an information theoretic solution, where the information
geometry often give a clear view of the structure of the problems. A particularly simple case is when the
distributions of concern are all close to each other in some sense, in which case information geometry is reduced into much simpler Euclidean geometry. For this reason, most known results on error exponents are limited to the very noisy channel, including our result on UEP mentioned earlier in this report.

We took the view that one can indeed use such simplification to address a much larger variety of problems, including some more complicated network information theory problems. Often in these problems, the question is whether a single letter characterization of the solution is optimal; or in other words, to find the structure of the optimal solution for the multi-letter K-L divergence optimization problems. While K-L divergence optimization quickly becomes intractable as the dimensionality increases, the local approximation of K-L divergence, approximating divergence by the Euclidean distance between distributions, is often much easier to handle.

Using this insight, we studied a few classical multi-terminal information theory problems, including source coding with single helper, degraded broadcasting channel, and general broadcasting channel with degraded message sets. In each cases, we re-derived the known capacity result from a very different approach. Applying the similar approach to some new problems, we solve the open problem of degraded message sets over broadcasting channel for more than 2 users, under the very noisy assumption. We further identify the key feature of a probabilistic transition that affects the efficiency of shifting information from one target to another, and is directly gives the slope of the capacity region. This observation connects our new results well too the vision of layered coding in this project, as the performance of each layer of sub-code is characterized.

2.3 Applications of Layered Codes

As described in the project proposal, the application of layered coding techniques can have significant impacts to network operations. Most importantly, as different types of data are encoded together, a new interface to the physical layer is defined, where the higher layer protocols can control not only the data rates, but also the tradeoffs between the reliability and delays of different streams of data. This approach allows the control and protocol information in networks to be treated in the same way as the data traffic, efficiently utilizing the channel resources, and allowing fast adaptation. Specific applications include embedding feedback and retransmission requests in two way data communications, and the use of geometric approach to design universal decoders for compound channels.

We have derived new outer bounds for channel with feedbacks and erasure. The idea is that the receiver can claim erasure to avoid making wrong decisions, when the confidence level of decoding is low. Our result gives a new approach to find the optimal tradeoff between the probability of erasure and probability of error, in terms of erasure and error exponents. Our result features a clear construction of a two-phase coding technique, and a new bounding technique, which we expect to be used in more general problems.

We studied the compound channel, where the precise channel realization is not known, but restricted in some set. Such abstraction can be applied to most wireless fading channels and channels with interference. We derived the achievable rate under the assumption that a additive decoding metric is used, which is a good practical constraint but hardly addressed in information theory literature. Inspired by our geometric approach, we derived the optimal decoder in this class, and completely characterized the tradeoff between decoding complexity and the loss from capacity.

3 Publications


