REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
20/01/2012

2. REPORT TYPE
Final Performance Report

3. DATES COVERED *(From - To)*
1/01/2008-1/01/2011

4. TITLE AND SUBTITLE
Dynamic Wireless Network Based on Open Physical Layer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-08-1-0143

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Lizhong Zheng, EECS, MIT

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
MIT

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR
875 N. Randolph St.

Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0240

12. DISTRIBUTION/AVAILABILITY STATEMENT
Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this project is to develop new interface to the physical layer, which is more suitable for the use of dynamic wireless networks. Our goal is to explore the formulations that gives a new measure of information, both in terms of how difficult it is to transmit such information and in terms of how valuable it is when received. While traditionally, the amount of information is universally measured in number of "bits". Such formulation is particularly suitable when long block codes are used, and the average per unit time rates of information flows are of interests. In dynamic communications, our main observation is that in fact a metric that does not require averaging over asymptotically long period of time is much more meaningful in practice. Moreover, one cannot always think of information as perfectly decoded bits, but rather has to understand how valuable a piece of "soft" information is. That is, when we cannot guarantee reliable decoding of the message, one still has to efficiently process the message, in order that efficiently communication over the entire network is achieved.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Lizhong Zheng
U	U	U	None	6	**19b. TELEPHONE NUMBER** *(Include area code)*

Reset

Final Report

February 18, 2011

Project Title: *Dynamic Wireless Network Based on Open Physical Layer*

Contract No.: FA9550-08-1-0143

Principle Investigator: *Lizhong Zheng, EECS, MIT*

Personnel Supported:

- PI: *Lizhong Zheng*
- PhD Students:
  - *Baris Nakiboglu*
  - *Emmanuel Abbe*
  - *Mina Karzand*

1 Abstract

The goal of this project is to develop new interface to the physical layer, which is more suitable for the use of dynamic wireless networks. Our goal is to explore the formulations that gives a new measure of information, both in terms of how difficult it is to transmit such information and in terms of how valuable it is when received. While traditionally, the amount of information is universally measured in number of "bits". Such formulation is particularly suitable when long block codes are used, and the average per unit time rates of information flows are of interests. In dynamic communications, our main observation is that in fact a metric that does not require averaging over asymptotically long period of time is much more meaningful in practice. Moreover, one cannot always think of information as perfectly decoded bits, but rather has to understand how valuable a piece of "soft" information is. That is, when we cannot guarantee reliable decoding of the message, one still has to efficiently process the message, in order that efficiently communication over the entire network is achieved.

Like "bits", a new information measure can be directly used as a interface to the physical layer. The new interface would be more general than the traditional approach of using communication channels as "reliable" bit pipes, but instead, explicitly model the error and delay of coded transmissions, prioritize the error protections to multiple types of heterogeneous data. With this approach, the higher layer network algorithms would have direct control over the amount of redundancy injected to different data streams, including data various QoS requirements, as well as network control messages. The main challenge of the project is thus to explicitly establish the optimal tradeoff between the reliability and rates of multiple data streams that are encoded and transmitted jointly over a single communication channel. The key tool involved here is information geometry, which is new and powerful analytical tool particularly useful in describing the dynamics of the probability distributions involved in a network communication problem.

One of the most important finding in our project is that when soft information is of concern, there is a difference between how to measure the overall communication efficiency and the communication problem at a particular time instance or a local exchange. That is, the communication scheme based on the optimization of a local or instantaneous metric, might contribute to the overall purpose of communication in different ways, based on how the soft information is processed, and combined with other side information. This finding

implies that a rich collection of metrics on the effectiveness of communication might all be relevant, even if the purpose of the overall communication process or network is to convey information bits. This finding confirms our intuition presented at the beginning of this project, that even in pure data networks, separated from the problems of different types of sources and QoS requirements, information flowing through the network should be intrinsically heterogeneous. Our work thus provides a theoretic framework where such different notions of efficiency can be applied and analyzed, in the context of dynamic communication networks.

2 Technical Report

Over the duration of this project, we have made good progresses in different aspects. Three graduate students were involved in this project: Emmanue Abbe, Baris Nakiboglu, and Mina Karzand. Emmanuel graduated from MIT and is now a postdoc at EPFL. He continues to work with us as a research affiliate. Baris is graduated in early 2011. Mina Karzand joined MIT in fall 2009, after receiving a MS degree from EPFL. She is currently working on extending our work on feedback channels to network problems. There are three major pieces of work that are related to this project:

2.1 Dynamic Transmission over Feedback Channels

Feedback channels are a good topic to study when we want to understand communication over dynamic environments. Unlike the conventional point-to-point channels, at each time, the transmitter of a feedback channel receives an update of what the receiver has already received so far. As a result, the encoder has to adjust the information content and the way it is transmitted for each single symbol. This is fundamentally different from the conventional approach of pack and coding a long block of data together. Conceptually, if one want s to efficiently use a feedback channel, he needs to have a quantitative way to measure how much information is conveyed at each single symbol time, instead of averaged amount of information over a long block. Consequently, one can hardly define a notion of how many bits that are transmitted and decoded reliably, thus deviating from the conventional wisdom of using number of bits to measure information.

There is a large literature discussing feedback channels, mainly because feedback signals are an important way to enhance the reliability of communication. However, our knowledge on this topic is rather limited. The main ideas in the literature include: 1. using feedback to initiate retransmissions and use variable length codes to improve the reliability, and 2. improved forward error correction coding over a small set of special channels such as binary erasure channel and Gaussian additive noise channels. Furthermore, there is no satisfactory solution for channels with noisy feedbacks. We believe that the reason for this lack of success lies in the use of block average performance metrics for such a highly dynamic communication scenario.

In our work, we focus on designing efficient forward error correction mechanisms that take the advantage of feedback signals, and generalize from the elegant designs over erasure channels or Gaussian channels to general discrete memoryless channels. Our work is built on our previous results reported in this project, on information geometry. We first develop a new geometric measure of the efficiency of dynamic information transmission, which leads to a new formulation of the channel coding problem as a dynamic programming. We solve a simplified version of this dynamic programming problem, and make connection to the classical performance metrics such as the throughputs and the error exponents. We then further fine tune our designs by allowing a dynamically chosen performance metric, according to the transmission time and the history of the received signals. We show that such new insights not only gives nature and simple coding schemes over the dynamic setup, but also improves the performance in terms of the classical metrics such as the error exponents. The advantage of our formulation is several folds:

- First, our work is a proof-of-concept for the new formulation of dynamic communications. It is fundamentally different from the idea of using "bits" as the universal of information measure. The new metrics, as well as the corresponding encoding and decoding schemes, are not based on the assumption of the perfect fidelity between the transmitter and the receiver, and thus can be used to design and evaluate the processing of "soft" information.
- Secondly, while our current work focuses on channels with perfect causal feedbacks, the general framework can be generalized to noisy feedback problems, which is one of the most important open problems in information theory.

- Finally, the new signaling schemes based on the notion of processing soft information is the basic building block for a large number of multi-terminal communication problems, such as cooperative transmission and relay networks.

The key question to answer in this work is how much information is conveyed in a single channel use. While there are many possible definitions given in the literature, a natural measurement is to quantify how far the a postoriori distribution of the message, conditioned on the history of observations, moves after observing the new output of the channel. The answer to this question is naturally connected a geometric view of communication, in that it requires metrics of the lengths and inner products of movements in the space of probability distributions.

One plausible candidate of such metrics is the K-L divergence. Let the posterior distribution of the messages at time t be $\Phi^{(t-1)}$, i.e., conditioned on the observation of Y^{t-1} up to time $t-1$, and that after observing y_t by $\Phi^{(t)}$, one can use

$$D(\Phi^{(t)}||U) - D(\Phi^{(t-1)}||U)$$

as a measure of effectiveness of the communication process at time t. This is equivalent to measuring the reduction of entropy of the unknown message during the t^{th} use of the channel. It can be shown that maximizing this metric is equivalent as maximizing the mutual information of the channel, which result in using the capacity achieving input distribution, P_X^*. The optimal transmission strategy is simply to assign the messages to the input symbols in such a way that the input at time t takes distribution P_X^*.

What is more interesting is that there are a variety of different metrics exist in the communication literature. For example, other than achieving the capacity, one might ask what kind of input would maximize the error exponent of a finite length communication session. Or alternatively, what kind of metrics one should use in the space of probability distributions so that the resulting optimization problem would yield the error-exponent optimal solutions.

We solved this problem, and the detailed works are reported in [**?**]. It turns out that when Renyi entropy is used for the metrics on the space of probability distributions, the resulting optimization yield a simple coding strategy, which we call titled posteriori matching. Following this coding strategy, we can demonstrated the best known error performance for the feedback problem.

Exploring problems like this has potentially deep impacts. As we pointed out in the earlier work of this project, the currently widely used performance metrics and formulations of communication problems are only suitable for point-to-point communications, and in order to address dynamic network problems, new formulations are needed. We believe that a geometric view is the key of solving the new dynamic communication problems. However, it is important that such new studies can be consistent and compatible with the existing 60 years of works in information theory. The impact of our work can be realized if we can show that the new geometric formulations are natural generalizations to the commonly accepted formulations. To that end our work on the feedback channel serves precisely that purpose.

A different aspect of the feedback problem is that it suggests a new symbol-by-symbol transmission scheme based on soft information. In network communication problems, it is often the case that a node, such as a relay, face the puzzle of what information from his observation should be extracted and forwarded. This puzzle comes from the fact that relays often has to process "soft information", i.e., from his observation, it cannot extract a number of reliable information bits that the destination wants, but only noisy "hints" of these bits. It is often hard to quantify what information is needed by the destination, and in what how to transmit signals that contribute to that need. At a higher level, the encoder of a feedback channel also face the same problem. In the process of transmission, the receiver has received only some "soft" information about the message, and the question is how to transmit the next single symbol, which does not quite allow decoding, but is efficient in contributing to a decision at the end of the communication session.

We solved the problem in the feedback setting. At each time, the encoder would compute the posterior distribution of the message, and match that to the optimal input distribution over the channel. One way to see this is that the encoder "answers" the question about the message from the receiver in the best way he can. Thus, the transmission is all about what is not known at the receiver at the time, and hence is the most efficient. We proved that such signaling indeed achieves optimality according to some information theoretic measure. Following this result, we are currently working on relay channels, and we hope to report more progress on that at the end of this project.

3 Global Geometry of Non-Gaussian Distributions

Another main results we obtained during the last year is an extension of our work on finding the geometric structure of non-Gaussian distributions.

In our previous works, we have defined a notion of divergence transition map (DTM). For a given discrete memoryless channel with a particular pair of input and output distributions, we can view the channel as a linear map that maps the neighborhood of the input distributions to that of the output distribution. This way, we can exactly quantify how the changes in the input distributions affects the outputs; describe which information is lost through the channel; and based on these develop a new way to optimize mutual information as well as network capacity.

The key observation is that when describing the divergence transition map, we have a linear map acting on a neighborhood of distributions, which can be viewed as a neighborhood on the manifold, and hence a linear space itself. Thus, to describe the divergence transition, it is generally of interests to find out the eigenvalues and the eigenvectors of the DTM. It turns out that for Gaussian additive channel with Gaussian inputs, the eigenvectors of the DTM have a particularly elegant form, namely, the Hermite polynomials. Using this observation, we came up with new ways to derive local optimality of several important results, including the monotonicity of central limit theorem, a new derivation of entropy power inequality, and some new results on Gaussian interference channels. The key advantage of this eigen analysis is that an input distribution is a perturbation from the Gaussian distribution

$$f = g_v(1 + \delta H^{[k]})$$

along the k^{th} order Hermite polynomial $H^{[k]}$, then the output of an additive Gaussian noise channel is

$$f * g_{\sigma^2} = g_{v+\sigma^2}(1 + \delta \mu_k H^{[k]})$$

which is still a perturbation from the corresponding Gaussian distribution along the k^{th} order Hermite polynomial, with a scaling factor μ_k, i.e., the corresponding singular value.

Using this approach, we can define a new coordinate system around the Gaussian distribution, and parameterize all the non-Gaussian distributions in this neighborhood. With this new tool, we can then quantitatively answer many questions, such as "how does the non-Gaussianness evolve when passed over a channel", and "how to cancel the non-Gaussianness when the noise is not Gaussian". These questions are directly related to central limit theorem, entropy power inequality, Gaussian boradcasting channels, interference and relay channels, and give new insights to the optimization of capacities for these problems.

The main limitation of the above approach is that they are limited to local statements. That is, we could only consider distributions that are "close" to Gaussian. To generalize this work to arbitrary distributions, we have made the following progresses in the past year.

- First, we made the following observation. if $f_1 = g(1 + \delta_1 H^{[k_1]})$, $f_2 = g(1 + \delta_2 H^{[k_2]})$, then the convolution of these two distributions $f_1 * f_2$, when written as Gaussian perturbations, has perturbation on $H^{[k_1]}, H^{[k_2]}$, as by the local approximation. In addition, the approximation error is simply on the direction of $H^{[k_1+k_2]}$. The amount of these perturbations can all be precisely calculated. In this way, we can indeed compute the convolution between non-Gaussian distributions in general, in terms of perturbation from Gaussian, and without any approximation. Following this observation, we can re-derive CLT and some other results in the most general form, albeit the algebra, which involves some combinatorics, is rather complicated. The main drawback of such brute force approach is that for the global picture, Hermite polynomials are no longer the eigen vectors of the DTM, and hence it is not surprising that the calculation based on the Hermite basis is a bit complicated.
- We can also derive the eigen structure of the DTM by Gaussian noise channel, at a neighborhood around an arbitrary non-Gaussian distribution. This involves the Christoffel symbols that describes the connection over the distribution manifold. The resulting eigen structure can then be used to analyze the non-Gaussian distributions as before. The drawback of this approach is that the eigen structure no longer has a simple form. Such complications are in some sense not surprising, as it appears that analyzing the global geometry of probability distributions do require the description of the global structure of the distribution manifolds.

We are currently working on using such global analytical tools to extend our earlier results on interference channels, some partial results are reported in [**?**]

4 Dynamic Communication and Instantaneous Efficiency

The most general way to describe a communication process is through *belief evolution*. We refer to the distribution of the messages, conditioned on the receiver's knowledge as the belief at the receiver. As the receiver accumulates observations, either over time or over different paths through the network, this belief would move from a uniform distribution, corresponds to no knowledge at the receiver at all, to a deterministic one, where the receiver can make a decision of which message is transmitted. The advantage of this approach is it is very general, without any assumptions of block code, long term behavior, or any sense of reliability. The disadvantage is also obvious: belief vectors live in a high dimensional space, and characterizing their movements requires strong geometric tools.

Designing communication protocols according to belief evolution has several difficulties. First, the control one has to maneuver the belief vector is through the channel input, which only controls the belief vector is a stochastic way. Thus the effectiveness of a control protocol can be measure only by averaging over the randomness of the channel. Secondly, the dimensionality of the belief space is often very large, much larger than the size of the input alphabet . Thus, one can hardly hope to drive the belief vector towards the "right" direction, pointing at the desired corner of the belief space, but rather bundle many messages together, and hope to move the belief vector towards one face that contains the correct message. As a result, one needs to frequently readjust the control to make the belief vector to zig-zag towards a particular direction. Conventional error correction coding can be viewed as one way to do that. Furthermore, the coding protocol cannot depend on the correct message being sent, thus it has to be simultaneously efficient for all possible messages.

While all of the above difficulties can, to some extent, be addressed with a good dynamic programing solution, the most difficult part of this approach is that the notion of efficiency is not unique. This can be seen from a simple example of $M = 3$. Starting from 3 equally likely messages, if one wishes to make a decision among the messages after the communication session, then it is desirable that the probability of a particular message "stands out". For example, $P_1 = [0.6, 0.2, 0.2]$ gives a probability of error of 0.4, which is better than $P_2 = [0.5, 0.5, 0]$. On the other hand, if there is side information available or if the communication session continues before a final decision is made, one can easily check that P_2 can be more desirable, as one of the messages is completely ruled out, and in information theoretic terms, the entropy/uncertainty of P_2 is smaller. This example says that the notion efficiency of an intermediate step of communication highly depends on how the communication results are to be used, or what time this step is, during the communication session.

A natural set of metrics of interests are the family of Renyi divergences. With a parameter of α that can be tuned, Renyi entropy of a given distribution corresponds to the Shannon entropy, at $\alpha = 1$, to the probability of detection error, at $\alpha = \infty$. This gives a continuum from of metrics of information transmissions that is of "general purpose" and agnostic to how the information is used, to that reflects the information used for decision making only. We developed a new approach to design communication based on instantaneous optimization of Renyi divergence, based on the intuition of choosing the parameter α according to the time and the current knowledge at the receiver. We show that such instantaneous communication schemes can out-perform those based on capacity optimization, using the toy example of quantum detections.

We are interested in quantum detection problems for several reasons.

- First, quantum processing are by nature lossy and instantaneous. A quantum state cannot by duplicated and hardly stored. Thus, when a measurement is made to a quantum state, the state itself gets destroyed. In designing such measurement, it is often the case that the measurement results from other related quantum states are available. Thus, a measurement here can often be lossy, and cannot capture the sufficient statistic for inference. It has to be designed based on the temporary and limited information. Thus, the concept of lossy processing is necessary in such designs.
- The geometry of quantum detections is particularly interesting. The embedding of quantum states and probability amplitudes are different from that for that classical probability distributions. The classical notion of K-L divergence and the Renyi family are replaced by Von Neumann entropy and its generalizations. Thus, we are facing a fundamentally different way to characterize the belief space, with the promises of much stronger processing tools leading to quantum channel and quantum computer designs. Clearly, further understanding of this geometry and its impacts to information processing are of great theoretic values.

- Quantum algorithm designs is a wide open area. In particular, while promising of improving the computational efficiency by orders of magnitude, the exact characterization of the information flows in such algorithms is not well understood. We hope that understanding quantum algorithms from the viewpoint of information geometry can offer fruitful insights and improvements.

In our recent work [**?**], we studied the coherent quantum detection problem, and the resulting capacity, in terms of the amount of classical information that can be conveyed through a quantum channel. We show that the best known binary quantum detector, the Dolinar receiver, can indeed be derived by instantaneous optimization. This result helps us to generalize Dolinar receiver from binary to general hypothesis testing problems, as well as the cases with coded transmissions, particularly in the regime of high photon efficiency. Our approach is also used to develop a new way to prove converse results, which not only improved our theoretic understanding of such problems, but also made great impacts on the practical designs of quantum receivers.

References

[1] Shashibhushan Borade, Baris Nakiboglu, and Lizhong Zheng, *Unequal Error Protection: Some fundamental limits and optimal strategies*, IEEE Transactions on Information Theory, Vol 55, pp 5511-5539, December, 2009.

[2] Emmanuel Abbe and Lizhong Zheng, *Coding along Hermite Polynomials for Interference Channels*, IEEE Information Theory Workshop, 2009

[3] Emmanuel Abbe, and Lizhong Zheng, *Coding along Hermite polynomials for Gaussian noise channels*, International Symposium on Information Theory(ISIT) 2009

[4] Emmanuel Abbe, and Lizhong Zheng, *Linear universal decoding for compound memoryless channels: a local to global geometric approach*, IEEE Transactions, on Information Theory, Vol. 56, December 2010

[5] Emmanuel Abbe, and Lizhong Zheng, *A coordinate system for Gaussian networks*, submitted to IEEE Transactions on Information Theory, August, 2011

[6] Siva K. Gorantla, Baris Nakiboglu, Todd P. Coleman, and Lizhong Zheng, *Multi-layer Bit-wise Unequal Error Protection for Variable Length Blockcodes with Feedback*, submitted to IEEE Transactions on Information Theory, January, 2011, (arXive: 1101.1934)

[7] Baris Nakiboglu, L. Zheng, *Error-and-Erasure Decoding for Block Codes with Feedback*, submitted to IEEE Transactions on Information Theory, March, 2009

[8] Hye-Won Chung, Saikat Guha, and Lizhong Zheng, *On Capacity of Optical Channels with Coherent Detection*, submitted to International Symposium on Information Theory(ISIT), 2011