



# The Science of Information: From Communication to DNA Sequencing

## Lecture Abstract:

Information theory is the science behind modern day communication engineering. Before information theory, the design of communication systems was ad hoc and tied to the specific source and specific physical medium of communication. By focusing instead on an abstract but quantifiable notion of information, the theory provides a unified basis for the design of all communication systems and introduces new ways of communicating that upends decades of engineering intuition. Can this way of thinking benefit other fields?

In this talk we explore one such possibility: DNA sequencing, the basic workhorse of modern day biology and medicine. The dominant technique is shotgun sequencing where many randomly located short reads are extracted from the DNA sequence and assembled to reconstruct the original sequence. Today there are multiple sequencing technologies and many assembly algorithms designed for them.

A basic but yet open information theoretic question is: given a sequencing technology and the statistics of the DNA sequence, what is the minimum read length and the minimum number of reads required for reliable reconstruction, and what is an optimal assembly algorithm that achieves this minimum? We provide an exact answer to this question for DNA sequences modeled by short-range correlated models, and an approximate answer for sequences with repeat statistics extracted from actual genomic data. These results form the basis of a systematic data-driven approach to designing optimal assembly algorithms.

This is joint work with Guy Bresler, Ma'ayan Bresler and Abolfazl Motahari.

**Date: Dec 14, 2012**  
**Time: 2:30 - 3:30pm**

**Venue: T.Y. Wong Hall, 5/F**  
**Ho Sin Hang Engineering Bldg.**

**Speaker: Prof. David Tse**  
**University of California at Berkeley**

David Tse was born in Hong Kong. He received the B.A.Sc. degree in systems design engineering from University of Waterloo in 1989, and the M.S. and Ph.D. degrees in electrical engineering from Massachusetts Institute of Technology in 1991 and 1994 respectively. From 1994 to 1995, he was a postdoctoral member of technical staff at A.T. & T. Bell Laboratories. Since 1995, he has been at the EECS Department in the University of California at Berkeley, where he is currently a Professor. He received a NSERC four year graduate fellowship from the government of Canada in 1989, and a NSF CAREER award in 1998, the Best Paper Awards at the Infocom 1998 and Infocom 2001 conferences, the Erlang Prize in 2000 from the INFORMS Applied Probability Society, the IEEE Communications and Information Theory Society Joint Paper Award in 2001, the Information Theory Society Paper Award in 2003, the 2009 Frederick Emmons Terman Award from the American Society for Engineering Education, and a Gilbreth Lectureship from the National Academy of Engineering in 2012. He was the Technical Program co-chair of the International Symposium on Information Theory in 2004, and was an Associate Editor of the IEEE Transactions on Information Theory from 2001 to 2003. He is a coauthor, with Pramod Viswanath, of the text "Fundamentals of Wireless Communication", which has been used in over 60 institutions around the world.

## Event Sponsors:

**Institute of Theoretical Computer Science and Communications**



IEEE Information Theory Society

$$C = W \log(1 + \text{SNR})$$



網絡編碼研究所  
Institute of Network Coding