

## **abstract**

COMPUTER SCIENCE/DISCRETE MATH, III  
Topic:

Speaker:

Affiliation:

Date:

Time/Room:

---

Standard information-theoretic results show that data over small, typically binary, alphabets can be compressed to Shannon's entropy limit. Yet most practical sources, such as text, audio, or video, have essentially infinite support. Compressing such sources requires estimating probabilities of unlikely, even unseen, events, a problem considered by Laplace. Of existing estimators, an ingenious if cryptic one derived by Good and Turing while deciphering the Enigma code works best yet not optimally. Hardy and Ramanujan's celebrated results on the number of integer partitions yield an asymptotically optimal estimator that compresses arbitrary-alphabet data patterns to their entropy. The same approach generalizes Fisher's seminal work estimating the number of butterfly species and its extension authenticating a poem purportedly written by The Bard. The talk covers these topics and is self contained.

Joint work with Prasad Santhanam, Krishna Viswanathan, and Junan Zhang.