Foreword

The first course usually is an appetizer. In the case of Raymond Yeung's *A First Course in Information Theory*, however, another delectable dish get served up in each of the sixteen chapters. Chapters 1 through 7 deal with the basic concepts of entropy and information with applications to lossless source coding. This is the traditional early fare of an information theory text, but Yeung flavors it uniquely. No one since Shannon has had a better appreciation for the mathematical structure of information quantities than Prof. Yeung. In the early chapters this manifests itself in a careful treatment of information measures via both Yeung's analytical theory of I-Measure and his geometrically intuitive information diagrams. (This material, never before presented in a textbook, is rooted in works by G. D. Hu, by H. Dyckman, and by R. Yeung *et al.*) Fundamental interrelations among information measures and Markovianness are developed with precision and unity. New slants are provided on staples like the divergence inequality, the data processing theorem, and Fano's inequality. There is also a clever, Kraft-inequality-free way of proving that the average length of the words in a lossless prefix source code must exceed the source's entropy. An easily digestible treatment of the redundancy of lossless prefix source codes also is served up, an important topic in practice that usually is slighted in textbooks.

The concept of weakly typical sequences is introduced and then used to anchor Yeung's proof of the lossless block source coding theorem. The concept of strongly typical sequences is introduced next. Later extended to joint typicality, this provides a foundation for proving the channel coding theorem in Chapter 8, the lossy source coding (rate-distortion) theorem in Chapter 9, and selected multi-source network coding theorems in Chapter 15. Although the proof of the channel coding theorem follows standard lines, Yeung's tasteful development of the interplay between information quantities and Markovianness readies one's palate for a rigorous proof that feedback around a discrete memoryless channel does not increase its capacity. In most information theory books this basic result of Shannon either does not appear or is relegated to a problem in which the several steps are outlined in order to guide the reader toward the goal. Rate-distortion theory and Shannon's lossy source coding theorem are treated in familiar ways. When proving the latter, one confronts lack of independence of the events $\{(\mathbf{X}, \hat{\mathbf{X}}(i)) \in T^n\}$, where $\mathbf{X}$ is a random source word, $\hat{\mathbf{X}}(i)$ is the ith word in a randomly chosen source code, and T^n is the set of jointly typical vector pairs. In those instances in which this widely unappreciated stumbling block is not overlooked entirely, it usually is addressed via either a non-selfcontained reference or a mammoth problem at the end of the chapter. However, Yeung's thorough earlier development of strong joint typicality concepts allows him to tackle it head-on.

Chapter 10 dishes up a careful treatment of the iterative algorithms for computation of channel capacity and rate-distortion functions pioneered by R. E. Blahut and S. Arimoto, which is generally accepted as today's preferred approach to computational information theory. Moreover, it has the extra advantage that iterative optimization algorithms are finding widespread application to areas as diverse as decoding of turbo and low-density parity-check codes and belief propagation in artificial intelligence and in real and artificial neural nets.

Chapters 11 through 16 are a unique tour de force. In as digestible a fashion as could possibly be expected, Yeung unveils a smorgasbord of topics in modern information theory that heretofore have been available only in research papers generated principally by Yeung and his research collaborators. Chapter 11 is a strong treatment of single-source network coding which develops carefully the relationships between information multicasting and the max-flow min-cut theory. Yeung makes an iron-clad case for how nodes must in general perform coding, not just storing and forwarding. Chapters 12, 13 and 14 on information inequalities of both Shannon and non-Shannon type constitute a definitive presentation of these topics by the master chef himself. Connections with linear programming are exploited, culminating in explication of Information Theory Inequality Prover (ITIP) of R. Yeung and Y.-O. Yan for inequalities of Shannon-type which comes with this book (also WWW-available). This leads, in turn, to the fascinating area of non-Shannon-type information inequalities, pioneered by R. Yeung and Z. Zhang. This material has been found to possess profound implications for the general area of information structures being studied by mathematical logicians and may also contribute to thermodynamics and statistical mechanics wherein the concept of entropy originated and which continue to be heavily concerned with various families of general inequalities. The theory of I-Measure introduced in Chapter 6 provides the essential insight into those of the non-Shannon type inequalities that are discussed here. Multi-source network coding in Chapter 15 is a confounding area in which Yeung and others have made considerable progress but a comprehensive theory remains elusive. Nonetheless, the geometrical framework for information inequalities developed in Chapters 12 and 13 renders a unifying tool for attacking this class of problems. The closing chapter linking entropy to the theory of groups is mouthwateringly provocative, having the potential to become a major contribution of information theory to this renowned branch of mathematics and mathematical physics.

Savor this book; I think you will agree the proof is in the pudding.

Toby Berger
Irwin and Joan Jacobs Professor of Engineering
Cornell University, Ithaca, New York