CHAPTER 1

Introduction

We introduce the general problem of optimal information flow in networks, which is the focus of network information theory. We then give a preview of the book with pointers to where the main results can be found.

1.1 NETWORK INFORMATION FLOW PROBLEM

A networked system consists of a set of information sources and communication nodes connected by a network as depicted in Figure 1.1. Each node observes one or more sources and wishes to reconstruct other sources or to compute a function based on all the sources. To perform the required task, the nodes communicate with each other over the network.

- What is the limit on the amount of communication needed?
- How can this limit be achieved?

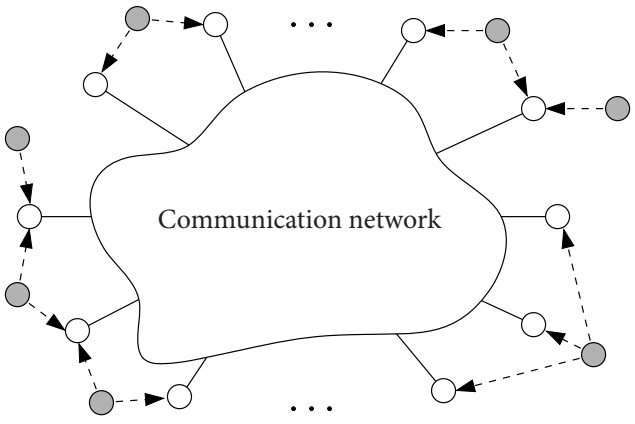

Figure 1.1. Elements of a networked system. The information sources (shaded circles) may be data, video, sensor measurements, or biochemical signals; the nodes (empty circles) may be computers, handsets, sensor nodes, or neurons; and the network may be a wired network, a wireless cellular or ad-hoc network, or a biological network.

These information flow questions have been answered satisfactorily for graphical unicast (single-source single-destination) networks and for point-to-point communication systems.

1.2 MAX-FLOW MIN-CUT THEOREM

Consider a *graphical* (wired) network, such as the Internet or a distributed storage system, modeled by a directed graph $(\mathcal{N}, \mathcal{E})$ with link capacities C_{jk} bits from node j to node k as depicted in Figure 1.2. Assume a unicast communication scenario in which source node 1 wishes to communicate an R-bit message M to destination node N. What is the network capacity C, that is, the maximum number of bits R that can be communicated reliably?

The answer is given by the *max-flow min-cut theorem* due to Ford and Fulkerson (1956) and Elias, Feinstein, and Shannon (1956). They showed that the capacity (maximum flow) is equal to the minimum cut capacity, i.e.,

$$C = \min_{\mathcal{S} \subset \mathcal{N}: 1 \in \mathcal{S},\, N \in \mathcal{S}^c} C(\mathcal{S}),$$

where $C(\mathcal{S}) = \sum_{j \in \mathcal{S},\, k \in \mathcal{S}^c} C_{jk}$ is the capacity of the cut $(\mathcal{S}, \mathcal{S}^c)$. They also showed that the capacity is achieved without errors using simple routing at the intermediate (relay) nodes, that is, the incoming bits at each node are forwarded over its outgoing links. Hence, under this networked system model, information can be treated as a commodity to be shipped over a transportation network or electricity to be delivered over a power grid.

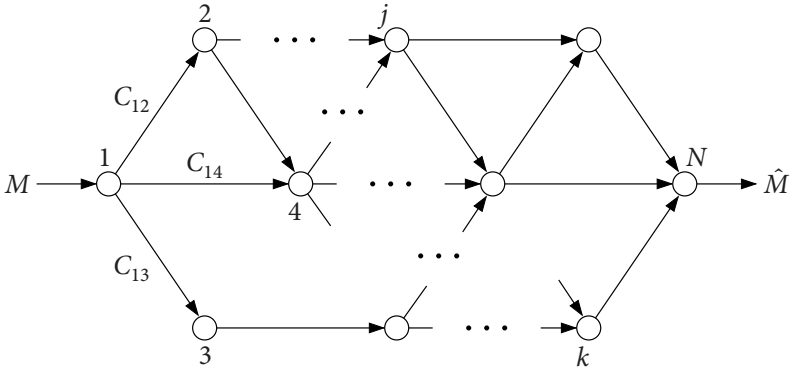

Figure 1.2. Graphical single-source single-destination network.

The max-flow min-cut theorem is discussed in more detail in Chapter 15.

1.3 POINT-TO-POINT INFORMATION THEORY

The graphical unicast network model captures the topology of a point-to-point network with idealized source and communication link models. At the other extreme, Shannon

(1948, 1959) studied communication and compression over a single link with more complex source and link (channel) models. He considered the communication system architecture depicted in Figure 1.3, where a sender wishes to communicate a k-symbol source sequence U^k to a receiver over a noisy channel. To perform this task, Shannon proposed a general *block coding* scheme, where the source sequence is mapped by an encoder into an n-symbol input sequence $X^n(U^k)$ and the received channel output sequence Y^n is mapped by a decoder into an estimate (reconstruction) sequence $\hat{U}^k(Y^n)$. He simplified the analysis of this system by proposing simple discrete memoryless models for the source and the noisy channel, and by using an *asymptotic* approach to characterize the necessary and sufficient condition for reliable communication.

Figure 1.3. Shannon's model of a point-to-point communication system.

Shannon's ingenious formulation of the point-to-point communication problem led to the following four fundamental theorems.

Channel coding theorem. Suppose that the source is a maximally compressed k-bit message M as in the graphical network case and that the channel is discrete and memoryless with input X, output Y, and conditional probability $p(y|x)$ that specifies the probability of receiving the symbol y when x is transmitted. The decoder wishes to find an estimate $\hat{M}$ of the message such that the probability of decoding error $\mathsf{P}\{\hat{M} \neq M\}$ does not exceed a prescribed value P_e. The general problem is to find the tradeoff between the number of bits k, the block length n, and the probability of error P_e.

This problem is intractable in general. Shannon (1948) realized that the difficulty lies in analyzing the system for any given finite block length n and reformulated the problem as one of finding the *channel capacity* C, which is the maximum communication rate $R = k/n$ in bits per channel transmissions such that the probability of error can be made arbitrarily small when the block length n is sufficiently large. He established a simple and elegant characterization of the channel capacity C in terms of the maximum of the mutual information $I(X; Y)$ between the channel input X and output Y:

$$C = \max_{p(x)} I(X; Y) \quad \text{bits/transmission.}$$

(See Section 2.3 for the definition of mutual information and its properties.) Unlike the graphical network case, however, capacity is achieved only *asymptotically* error-free and using sophisticated coding.

Lossless source coding theorem. As a "dual" to channel coding, consider the following lossless data compression setting. The sender wishes to communicate (store) a source sequence *losslessly* to a receiver over a *noiseless* binary channel (memory) with the minimum number of bits. Suppose that the source U is discrete and memoryless, that is, it

generates an i.i.d. sequence U^k. The sender encodes U^k at rate $R = n/k$ bits per source symbol into an n-bit index $M(U^k)$ and sends it over the channel. Upon receiving the index M, the decoder finds an estimate $\hat{U}^k(M)$ of the source sequence such that the probability of error $\mathsf{P}\{\hat{U}^k \neq U^k\}$ is less than a prescribed value. Shannon again formulated the problem as one of finding the minimum lossless compression rate R^* when the block length is arbitrarily large, and showed that it is characterized by the entropy of U:

$$R^* = H(U) \quad \text{bits/symbol.}$$

(See Section 2.1 for the definition of entropy and its properties.)

Lossy source coding theorem. Now suppose U^k is to be sent over the noiseless binary channel such that the receiver can reconstruct it with some distortion instead of losslessly. Shannon assumed the per-letter distortion $(1/k)\sum_{i=1}^{k} \mathsf{E}(d(U_i, \hat{U}_i))$, where $d(u, \hat{u})$ is a measure of the distortion between the source symbol u and the reconstruction symbol $\hat{u}$. He characterized the *rate–distortion function* $R(D)$, which is the optimal tradeoff between the rate $R = n/k$ and the desired distortion D, as the minimum of the mutual information between U and $\hat{U}$:

$$R(D) = \min_{p(\hat{u}|u):\mathsf{E}(d(U,\hat{U}))\le D} I(U;\hat{U}) \quad \text{bits/symbol.}$$

Source–channel separation theorem. Now we return to the general point-to-point communication system shown in Figure 1.3. Let C be the capacity of the discrete memoryless channel (DMC) and $R(D)$ be the rate–distortion function of the discrete memoryless source (DMS), and assume for simplicity that $k = n$. What is the necessary and sufficient condition for communicating the DMS over the DMC with a prescribed distortion D? Shannon (1959) showed that $R(D) \le C$ is necessary. Since $R(D) < C$ is sufficient by the lossy source coding and channel coding theorems, *separate* source coding and channel coding achieves the fundamental limit. Although this result holds only when the code block length is unbounded, it asserts that using bits as a "universal" interface between sources and channels—the basis for digital communication—is essentially optimal.

We discuss the above results in detail in Chapter 3. Shannon's asymptotic approach to network performance analysis will be adopted throughout the book.

1.4 NETWORK INFORMATION THEORY

The max-flow min-cut theorem and Shannon's point-to-point information theory have had a major impact on communication and networking. However, the simplistic model of a networked information processing system as a single source–destination pair communicating over a noisy channel or a graphical network does not capture many important aspects of real-world networks:

- Networked systems have multiple sources and destinations.
- The task of the network is often to compute a function or to make a decision.

- Wireless communication uses a shared broadcast medium.
- Networked systems involve complex tradeoffs between competition for resources and cooperation for the common good.
- Many networks allow for feedback and interactive communication.
- Source–channel separation does not hold for networks in general.
- Network security is often a primary concern.
- Data from the sources is often bursty and network topology evolves dynamically.

Network information theory aims to answer the aforementioned information flow questions while capturing some of these aspects of real-world networks. In the following, we illustrate some of the achievements of this theory using examples from the book.

1.4.1 Multiple Sources and Destinations

Coding for networks with many sources and destinations requires techniques beyond routing and point-to-point source/channel coding. Consider the following settings.

Graphical multicast network. Suppose we wish to send a movie over the Internet to multiple destinations (multicast). Unlike the unicast case, routing is not optimal in general even if we model the Internet by a graphical network. Instead, we need to use *coding* of incoming packets at the relay nodes.

We illustrate this fact via the famous "butterfly network" shown in Figure 1.4, where source node 1 wishes to send a 2-bit message $(M_1, M_2) \in \{0, 1\}^2$ to destination nodes 6 and 7. Assume link capacities $C_{jk} = 1$ for all edges (j, k). Note that using routing only, both M_1 and M_2 must be sent over the edge $(4, 5)$, and hence the message cannot be communicated to both destination nodes.

However, if we allow the nodes to perform simple modulo-2 sum operations in addition to routing, the 2-bit message can be communicated to both destinations. As illustrated in Figure 1.5, relay nodes 2, 3, and 5 forward multiple copies of their incoming bits,

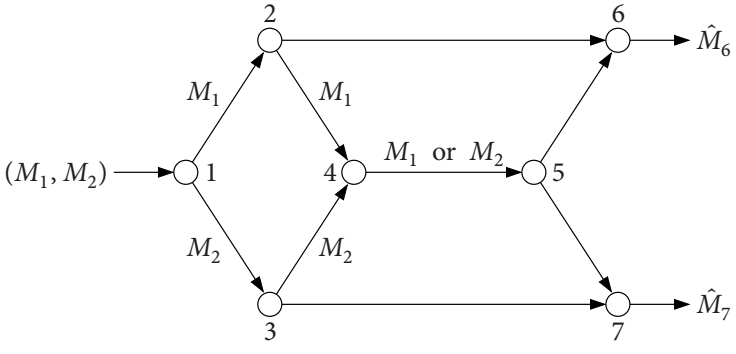

Figure 1.4. Butterfly network. The 2-bit message (M_1, M_2) cannot be sent using routing to both destination nodes 6 and 7.

and relay node 4 sends the modulo-2 sum of M_1 and M_2. Using this simple scheme, both destination nodes 6 and 7 can recover the message error-free.

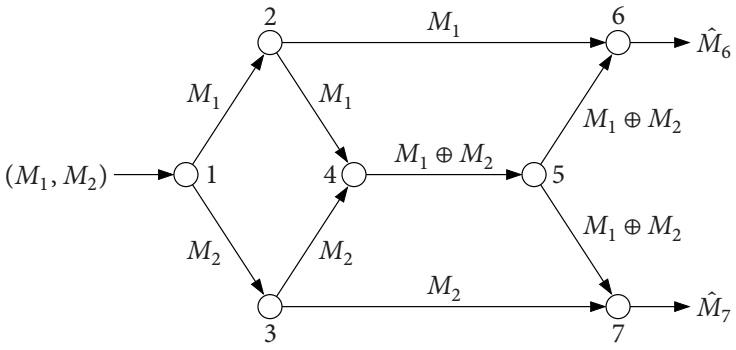

Figure 1.5. The 2-bit message can be sent to destination nodes 6 and 7 using linear network coding.

In Chapter 15, we show that linear network coding, which is a generalization of this simple scheme, achieves the capacity of an arbitrary graphical multicast network. Extensions of this multicast setting to lossy source coding are discussed in Chapters 13 and 20.

Distributed compression. Suppose that a sensor network is used to measure the temperature over a geographical area. The output from each sensor is compressed and sent to a base station. Although compression is performed separately on each sensor output, it turns out that using point-to-point compression is not optimal when the sensor outputs are *correlated*, for example, because the sensors are located close to each other.

Consider the distributed lossless compression system depicted in Figure 1.6. Two sequences X_1^n and X_2^n are drawn from correlated discrete memoryless sources $(X_1, X_2) \sim p(x_1, x_2)$ and compressed separately into an nR_1-bit index M_1 and an nR_2-bit index M_2, respectively. A receiver (base station) wishes to recover the source sequences from the index pair (M_1, M_2). What is the minimum sum-rate R_{sum}^*, that is, the minimum over $R_1 + R_2$ such that both sources can be reconstructed losslessly?

If each sender uses a point-to-point code, then by Shannon's lossless source coding theorem, the minimum lossless compression rates for the individual sources are $R_1^* = H(X_1)$ and $R_2^* = H(X_2)$, respectively; hence the resulting sum-rate is $H(X_1) + H(X_2)$. If instead the two sources are jointly encoded, then again by the lossless source coding theorem, the minimum lossless compression sum-rate is $H(X_1, X_2)$, which can be much smaller than the sum of the individual entropies. For example, let X_1 and X_2 be binary-valued sources with $p_{X_1,X_2}(0,0) = 0.495$, $p_{X_1,X_2}(0,1) = 0.005$, $p_{X_1,X_2}(1,0) = 0.005$, and $p_{X_1,X_2}(1,1) = 0.495$; hence, the sources have the same outcome 0.99 of the time. From the joint pmf, we see that X_1 and X_2 are both Bern(1/2) sources with entropy $H(X_1) = H(X_2) = 1$ bit per symbol. By comparison, their joint entropy $H(X_1, X_2) = 1.0808 \ll 2$ bits per symbol pair.

Slepian and Wolf (1973a) showed that $R_{\text{sum}}^* = H(X_1, X_2)$ and hence that the minimum

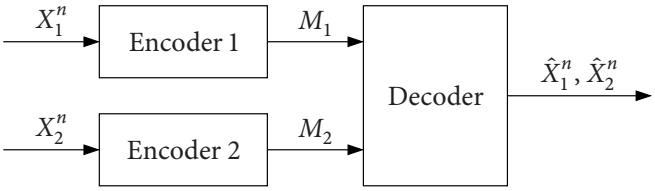

Figure 1.6. Distributed lossless compression system. Each source sequence X_j^n, $j = 1, 2$, is encoded into an index $M_j(X_j^n) \in [1 : 2^{nR_j})$, and the decoder wishes to reconstruct the sequences losslessly from (M_1, M_2).

sum-rate for distributed compression is asymptotically the same as for centralized compression! This result is discussed in Chapter 10. Generalizations to distributed lossy compression are discussed in Chapters 11 and 12.

Communication for computing. Now suppose that the base station in the temperature sensor network wishes to compute the *average* temperature over the geographical area instead of the individual temperature values. What is the amount of communication needed?

While in some cases the rate requirement for computing a function of the sources is the same as that for recovering the sources themselves, it is sometimes significantly smaller. As an example, consider an n-round online game, where in each round Alice and Bob each select one card without replacement from a virtual hat with three cards labeled 1, 2, and 3. The one with the larger number wins. Let X^n and Y^n be the sequences of numbers on Alice and Bob's cards over the n rounds, respectively. Alice encodes her sequence X^n into an index $M \in [1 : 2^{nR}]$ and sends it to Bob so that he can find out who won in each round, that is, find an estimate $\hat{Z}^n$ of the sequence $Z_i = \max\{X_i, Y_i\}$ for $i \in [1 : n]$, as shown in Figure 1.7. What is the minimum communication rate R needed?

By the aforementioned Slepian–Wolf result, the minimum rate needed for Bob to reconstruct X is the conditional entropy $H(X|Y) = H(X, Y) - H(Y) = 2/3$ bit per round. By exploiting the structure of the function $Z = \max\{X, Y\}$, however, it can be shown that only 0.5409 bit per round is needed.

This card game example as well as general results on communication for computing are discussed in Chapter 21.

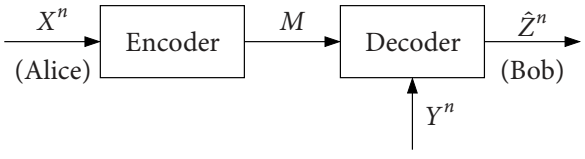

Figure 1.7. Online game setup. Alice has the card number sequence X^n and Bob has the card number sequence Y^n. Alice encodes her card number sequence into an index $M \in [1 : 2^{nR}]$ and sends it to Bob, who wishes to losslessly reconstruct the winner sequence Z^n.

1.4.2 Wireless Networks

Perhaps the most important practical motivation for studying network information theory is to deal with the special nature of wireless channels. We study models for wireless communication throughout the book.

The first and simplest wireless channel model we consider is the point-to-point Gaussian channel $Y = gX + Z$ depicted in Figure 1.8, where $Z \sim \mathrm{N}(0, N_0/2)$ is the receiver noise and g is the channel gain. Shannon showed that the capacity of this channel under a prescribed average transmission power constraint P on X, i.e., $\sum_{i=1}^{n} X_i^2 \le nP$ for each codeword X^n, has the simple characterization

$$C = \frac{1}{2}\log(1 + S) = \mathsf{C}(S),$$

where $S = 2g^2P/N_0$ is the received signal-to-noise ratio (SNR).

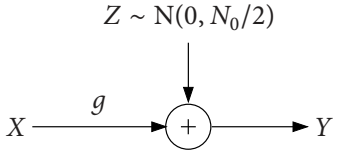

Figure 1.8. Gaussian point-to-point channel.

A wireless network can be turned into a set of separate point-to-point Gaussian channels via time or frequency division. This traditional approach to wireless communication, however, does not take full advantage of the broadcast nature of the wireless medium as illustrated in the following example.

Gaussian broadcast channel. The downlink of a wireless system is modeled by the Gaussian broadcast channel

$$Y_1 = g_1X + Z_1,$$
$$Y_2 = g_2X + Z_2,$$

as depicted in Figure 1.9. Here $Z_1 \sim \mathrm{N}(0, N_0/2)$ and $Z_2 \sim \mathrm{N}(0, N_0/2)$ are the receiver noise components, and $g_1^2 > g_2^2$, that is, the channel to receiver 1 is stronger than the channel to receiver 2. Define the SNRs for receiver $j = 1, 2$ as $S_j = 2g_j^2P/N_0$. Assume average power constraint P on X.

The sender wishes to communicate a message M_j at rate R_j to receiver j for $j = 1, 2$. What is the *capacity region* $\mathscr{C}$ of this channel, namely, the set of rate pairs (R_1, R_2) such that the probability of decoding error at both receivers can be made arbitrarily small as the code block length becomes large?

If we send the messages M_1 and M_2 in different time intervals or frequency bands, then we can reliably communicate at rate pairs in the "time-division region" $\mathscr{R}$ shown in Figure 1.9. Cover (1972) showed that higher rates can be achieved by adding the codewords for the two messages and sending this sum over the entire transmission block. The

stronger receiver 1 decodes for both codewords, while the weaker receiver 2 treats the other codeword as noise and decodes only for its own codeword. Using this *superposition coding* scheme, the sender can reliably communicate the messages at any rate pair in the capacity region $\mathscr{C}$ shown in Figure 1.9b, which is strictly larger than the time-division region $\mathscr{R}$.

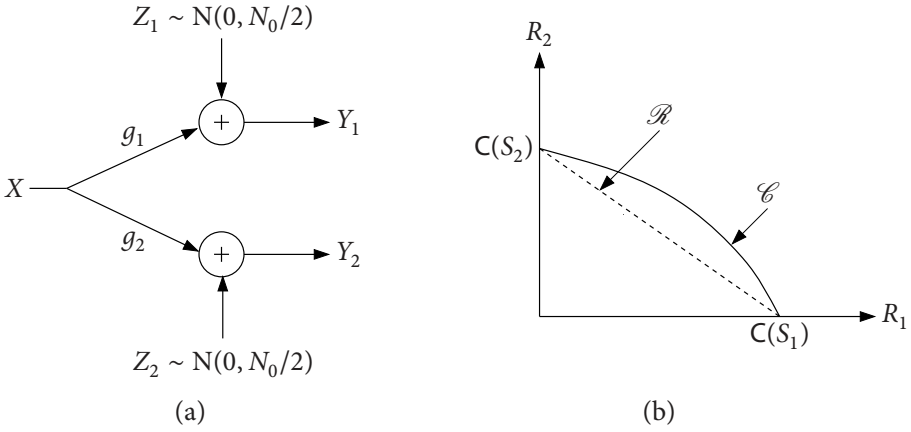

Figure 1.9. (a) Gaussian broadcast channel with SNRs $S_1 = g_1^2 P > g_2^2 P = S_2$. (b) The time-division inner bound $\mathscr{R}$ and the capacity region $\mathscr{C}$.

This superposition scheme and related results are detailed in Chapter 5. Similar improvements in rates can be achieved for the uplink (multiple access channel) and the intercell interference (interference channel), as discussed in Chapters 4 and 6, respectively.

Gaussian vector broadcast channel. Multiple transmitter and receiver antennas are commonly used to enhance the performance of wireless communication systems. Coding for these multiple-input multiple-output (MIMO) channels, however, requires techniques beyond single-antenna (scalar) channels. For example, consider the downlink of a MIMO wireless system modeled by the Gaussian vector broadcast channel

$$\mathbf{Y}_1 = G_1\mathbf{X} + \mathbf{Z}_1,$$
$$\mathbf{Y}_2 = G_2\mathbf{X} + \mathbf{Z}_2,$$

where G_1, G_2 are r-by-t channel gain matrices and $\mathbf{Z}_1 \sim \mathrm{N}(0, I_r)$ and $\mathbf{Z}_2 \sim \mathrm{N}(0, I_r)$ are noise components. Assume average power constraint P on $\mathbf{X}$. Note that unlike the single-antenna broadcast channel shown in Figure 1.9, in the vector case neither receiver is necessarily stronger than the other. The optimum coding scheme is based on the following *writing on dirty paper* result. Suppose we wish to communicate a message over a Gaussian vector channel,

$$\mathbf{Y} = G\mathbf{X} + \mathbf{S} + \mathbf{Z}$$

where $\mathbf{S} \sim \mathrm{N}(0, K_{\mathbf{S}})$ is an *interference* signal, which is independent of the Gaussian noise

$\mathbf{Z} \sim \mathrm{N}(0, I_r)$. Assume average power constraint P on $\mathbf{X}$. When the interference sequence $\mathbf{S}^n$ is available at the receiver, it can be simply subtracted from the received sequence and hence the channel capacity is the same as when there is no interference. Now suppose that the interference sequence is available only at the sender. Because of the power constraint, it is not always possible to presubtract the interference from the transmitted codeword. It turns out, however, that the *effect* of interference can still be completely canceled via judicious precoding and hence the capacity is again the same as that with no interference!

This scheme is applied to the Gaussian vector broadcast channel as follows.

- To communicate the message M_2 to receiver 2, consider the channel $\mathbf{Y}_2 = G_2\mathbf{X}_2 + G_2\mathbf{X}_1 + \mathbf{Z}_2$ with input $\mathbf{X}_2$, Gaussian interference $G_2\mathbf{X}_1$, and additive Gaussian noise $\mathbf{Z}_2$. Receiver 2 recovers M_2 while treating the interference signal $G_2\mathbf{X}_1$ as part of the noise.
- To communicate the message M_1 to receiver 1, consider the channel $\mathbf{Y}_1 = G_1\mathbf{X}_1 + G_1\mathbf{X}_2 + \mathbf{Z}_1$, with input $\mathbf{X}_1$, Gaussian interference $G_1\mathbf{X}_2$, and additive Gaussian noise $\mathbf{Z}_1$, where the interference sequence $G_1\mathbf{X}_2^n(M_2)$ is available at the sender. By the writing on dirty paper result, the transmission rate of M_1 can be as high as that for the channel $\mathbf{Y}_1' = G_1\mathbf{X}_1 + \mathbf{Z}_1$ without interference.

The writing on dirty paper result is discussed in detail in Chapter 7. The optimality of this scheme for the Gaussian vector broadcast channel is established in Chapter 9.

Gaussian relay channel. An ad-hoc or a mesh wireless network is modeled by a Gaussian *multihop* network in which nodes can act as relays to help other nodes communicate their messages. Again reducing such a network to a set of links using time or frequency division does not take full advantage of the shared wireless medium, and the rate can be greatly increased via node cooperation.

As a canonical example, consider the 3-node relay channel depicted in Figure 1.10a. Here node 2 is located on the line between nodes 1 and 3 as shown in Figure 1.10b. We assume that the channel gain from node k to node j is $g_{jk} = r_{jk}^{-3/2}$, where r_{jk} is the distance between nodes k and j. Hence $g_{31} = r_{31}^{-3/2}$, $g_{21} = r_{21}^{-3/2}$, and $g_{32} = (r_{31} - r_{21})^{-3/2}$. Assume average power constraint P on each of X_1 and X_2.

Suppose that sender node 1 wishes to communicate a message M to receiver node 3 with the help of relay node 2. On the one extreme, the sender and the receiver can communicate *directly* without help from the relay. On the other extreme, we can use a *multihop* scheme where the relay plays a pivotal role in the communication. In this commonly used scheme, the sender transmits the message to the relay in the first hop and the relay recovers the message and transmits it to the receiver concurrently in the second hop, causing interference to the first-hop communication. If the receiver is far away from the sender, that is, the distance r_{31} is large, this scheme performs well because the interference due to the concurrent transmission is weak. However, when r_{31} is not large, the interference can adversely affect the communication of the message.

In Chapter 16, we present several coding schemes that outperform both direct transmission and multihop.

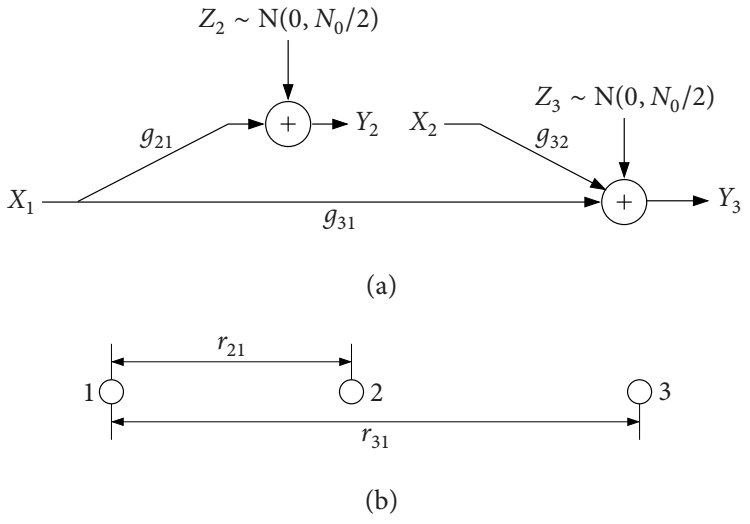

Figure 1.10. (a) Gaussian relay channel. (b) Node placements: relay node 2 is placed along the lines between sender node 1 and receiver node 2.

- *Decode–forward.* The direct transmission and multihop schemes are combined and further enhanced via coherent transmission by the sender and the relay. The receiver decodes for the signals from both hops instead of treating the transmission from the first hop as interference. Decode–forward performs well when the relay is closer to the sender, i.e., $r_{21} < (1/2)r_{31}$.
- *Compress–forward.* As an alternative to the "digital-to-digital" relay interface used in multihop and decode–forward, the compress–forward scheme uses an "analog-to-digital" interface in which the relay compresses the received signal and sends the compression index to the receiver. Compress–forward performs well when the relay is closer to the receiver.
- *Amplify–forward.* Decode–forward and compress–forward require sophisticated operations at the nodes. The amplify–forward scheme provides a much simpler "analog-to-analog" interface in which the relay scales the incoming signal and transmits it to the receiver. In spite of its simplicity, amplify–forward can outperform decode–forward when the relay is closer to the receiver.

The performance of the above relaying schemes are compared in Figure 1.11. In general, it can be shown that both decode–forward and compress–forward achieve rates within 1/2 bit of the capacity, while amplify–forward achieves rates within 1 bit of the capacity.

We extend the above coding schemes to general multihop networks in Chapters 18 and 19. In particular, we show that extending compress–forward leads to a noisy network coding scheme that includes network coding for graphical multicast networks as a special case. When applied to Gaussian multihop multicast networks, this noisy network coding scheme achieves within a constant gap of the capacity independent of network topology,

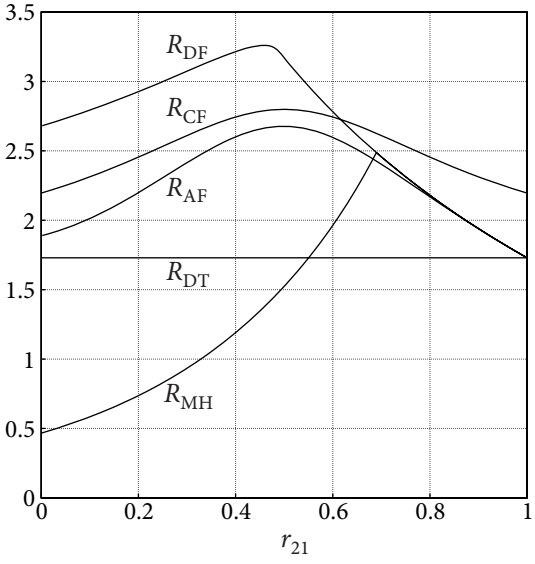

Figure 1.11. Comparison of the achievable rates for the Gaussian relay channel using direct transmission (R_{DT}), multihop (R_{MH}), decode–forward (R_{DF}), compress–forward (R_{CF}), and amplify–forward (R_{AF}) for $N_0/2 = 1$, $r_{31} = 1$ and $P = 10$.

channel parameters, and power constraints, while extensions of the other schemes do not yield such performance guarantees.

To study the effect of interference and path loss in large wireless networks, in Chapter 19 we also investigate how capacity scales with the network size. We show that relaying and spatial reuse of frequency/time can greatly increase the rates over naive direct transmission with time division.

Wireless fading channels. Wireless channels are *time varying* due to scattering of signals over multiple paths and user mobility. In Chapter 23, we study fading channel models that capture these effects by allowing the gains in the Gaussian channels to vary randomly with time. In some settings, channel capacity in the Shannon sense is not well defined. We introduce different coding approaches and corresponding performance metrics that are useful in practice.

1.4.3 Interactive Communication

Real-world networks allow for feedback and node interactions. Shannon (1956) showed that feedback does not increase the capacity of a memoryless channel. Feedback, however, can help simplify coding and improve reliability. This is illustrated in the following example.

Binary erasure channel with feedback. The binary erasure channel is a DMC with binary input $X \in \{0, 1\}$ and ternary output $Y \in \{0, 1, \mathrm{e}\}$. Each transmitted bit (0 or 1) is erased

$(Y = \mathrm{e})$ with probability p. The capacity of this channel is $1 - p$ and achieving it requires sophisticated block coding. Now suppose that noiseless causal feedback from the receiver to the sender is present, that is, the sender at each time i has access to all previous received symbols Y^{i-1}. Then we can achieve the capacity simply by retransmitting each erased bit. Using this simple feedback scheme, on average $n = k/(1 - p)$ transmissions suffice to reliably communicate k bits of information.

Unlike point-to-point communication, feedback can achieve higher rates in networks with multiple senders/receivers.

Binary erasure multiple access channel with feedback. Consider the multiple access channel (MAC) with feedback depicted in Figure 1.12, where the channel inputs X_1 and X_2 are binary and the channel output $Y = X_1 + X_2$ is ternary, i.e., $Y \in \{0, 1, 2\}$. Suppose that senders 1 and 2 wish to communicate independent messages M_1 and M_2, respectively, to the receiver at the same rate R. Without feedback, the *symmetric capacity*, which is the maximum rate R, is $\max_{p(x_1)p(x_2)} H(Y) = 3/4$ bits/transmission.

Noiseless causal feedback allows the senders to *cooperate* in communicating their messages and hence to achieve higher symmetric rates than with no feedback. To illustrate such cooperation, suppose that each sender first transmits its k-bit message uncoded. On average $k/2$ bits are "erased" (that is, $Y = 0 + 1 = 1 + 0 = 1$ is received). Since the senders know through feedback the exact locations of the erasures as well as the corresponding message bits from both messages, they can cooperate to send the erased bits from the first message (which is sufficient to recover both messages). This cooperative retransmission requires $k/(2 \log 3)$ transmissions. Hence we can increase the symmetric rate to $R = k/(k + k/(2 \log 3)) = 0.7602$. This rate can be further increased to 0.7911 by using a more sophisticated coding scheme that sends new messages simultaneously with cooperative retransmissions.

In Chapter 17, we discuss the *iterative refinement* approach illustrated in the binary erasure channel example; the cooperative feedback approach for multiuser channels illustrated in the binary erasure MAC example; and the two-way channel. In Chapters 20

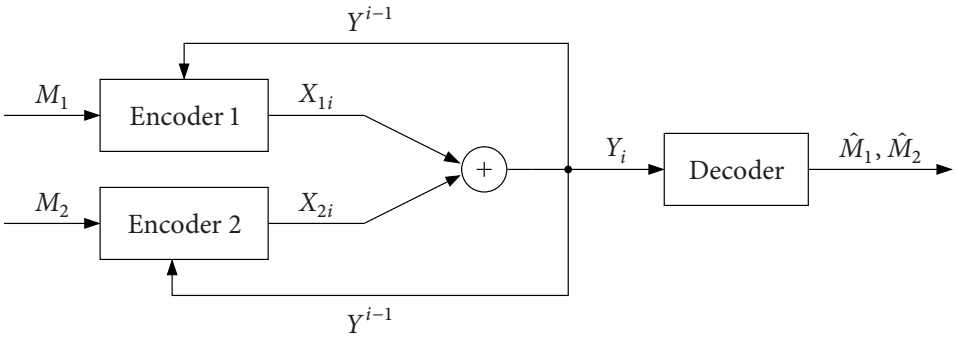

Figure 1.12. Feedback communication over a binary erasure MAC. The channel inputs X_{1i} and X_{2i} at time $i \in [1 : n]$ are functions of (M_1, Y^{i-1}) and (M_2, Y^{i-1}), respectively.

through 22, we show that interaction can also help in distributed compression, distributed computing, and secret communication.

1.4.4 Joint Source–Channel Coding

As we mentioned earlier, Shannon showed that separate source and channel coding is asymptotically optimal for point-to-point communication. It turns out that such separation does not hold in general for sending correlated sources over multiuser networks. In Chapter 14, we demonstrate this breakdown of separation for lossless communication of correlated sources over multiple access and broadcast channels. This discussion yields natural definitions of various notions of *common information* between two sources.

1.4.5 Secure Communication

Confidentiality of information is a crucial requirement in networking applications such as e-commerce. In Chapter 22, we discuss several coding schemes that allow a legitimate sender (Alice) to communicate a message reliably to a receiver (Bob) while keeping it secret (in a strong sense) from an eavesdropper (Eve). When the channel from Alice to Bob is stronger than that to Eve, a confidential message with a positive rate can be communicated reliably without a shared secret key between Alice and Bob. By contrast, when the channel from Alice to Bob is weaker than that to Eve, no confidential message can be communicated reliably. We show, however, that Alice and Bob can still agree on a secret key through interactive communication over a public (nonsecure) channel that Eve has complete access to. This key can then be used to communicate a confidential message at a nonzero rate.

1.4.6 Network Information Theory and Networking

Many aspects of real-world networks such as bursty data arrivals, random access, asynchrony, and delay constraints are not captured by the standard models of network information theory. In Chapter 24, we present several examples for which such networking issues have been successfully incorporated into the theory. We present a simple model for random medium access control (used for example in the ALOHA network) and show that a higher throughput can be achieved using a broadcasting approach instead of encoding the packets at a fixed rate. In another example, we establish the capacity region of the asynchronous multiple access channel.

1.4.7 Toward a Unified Network Information Theory

The above ideas and results illustrate some of the key ingredients of network information theory. The book studies this fascinating subject in a systematic manner, with the ultimate goal of developing a unified theory. We begin our journey with a review of Shannon's point-to-point information theory in the next two chapters.