Wireless Communications: An Information Theoretic Perspective

Ajit Kumar Chaturvedi
Department of EE
IIT Kanpur
akc@iitk.ac.in

Main References

- David Tse and Pramod Viswanath, *Fundamentals of Wireless Communication,* Cambridge, 2005
- Andrea Goldsmith *et al*, *Breaking Spectrum Gridlock With Cognitive Radios: An Information Theoretic Perspective,* Proceedings IEEE, May 2009, 894-914.

2

Outline

- Information Theory
  - Introduction and some key results
- Fading Channel
  - Slow Fading
  - Receive and Transmit Diversity
  - Time and Frequency Diversity
  - Fast Fading
- Multiuser Channel
  - Uplink
  - Capacity Region
- Cognitive Radio

3

Information Theory

- What is information theory ?
  - A way of measuring/weighing/quantifying information in terms of the quantity *entropy*
  - Provides a way of characterizing information flow from source to destination in terms of the quantity *mutual information*
- Why do we need to study information theory ?
  - We need to 'weigh' or 'measure' or 'quantify' the amount of information or rather the information rate produced by a source, say a video camera or a speaker or a file……….
  - We need to find out that given a fidelity criterion how much a bit sequence can be compressed
  - We need to find out the maximum rate at which information can be transmitted in a practical channel which can be modeled as a BSC, erasure, AWGN, fading, multiuser, multi-cell, ……
  - We need to find ways (i.e. codes) to achieve the maximum

4

Key Qualitative Results of Information Theory

- There is a limit beyond which a bit sequence cannot be compressed. Given the distribution of the sequence, this limit can be computed. Compression methods for achieving that limit are known.
- A communication channel can be characterized in terms of a number called capacity which is expressed in bits/s/Hz.
- Even in the presence of noise and fading, coding techniques are known for achieving reliable communication over the channel as long as the rate of information flow does not exceed its capacity.

5

Information Theory

- What are the most significant contributions of information theory ?
  - Lempel-Ziv family of algorithms
  - Turbo and LDPC codes
  - MIMO or multiple transmit and multiple receive antennas
  - Capacity of the AWGN channel

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right) \quad \text{bit/s.}$$

- Which major goals elude (failures) information theory ?
  - Useful insights into networks
  - In contrast to traditional Shannon theory, source is bursty
  - In contrast to traditional Shannon theory, end-to-end delay is finite

6

Some Queries

- We know of Nyquist's result regarding an upper limit on the data rate for zero inter symbol interference in a given finite bandwidth.
- How does Shannon's capacity result reconcile with that ?
- What is the fundamental question that one should ask regarding data rate through a communication channel ?

7

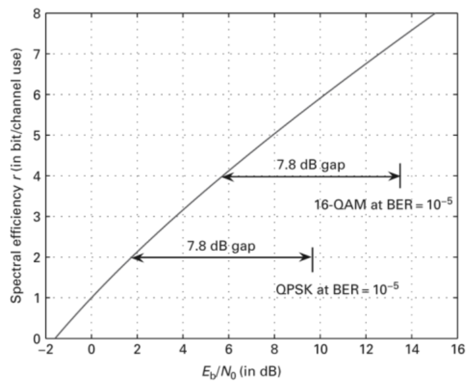
Limit on Spectral Efficiency

- Let R be the information rate, E_b be the energy per information bit and P be the transmitted power
- Then $P=E_bR$
- Define $r=R/W$ as the spectral efficiency
- Using the fact C > R and the relation

$$C = W \log_2 \left(1 + \frac{P}{N_0 W}\right) \text{ bit/s.}$$

- We obtain the following condition for reliable communication $$\frac{E_b}{N_0} > \frac{2^r - 1}{r}.$$

Performance of QAM and QPSK



- As we let spectral efficiency r → 0, we enter a power-limited regime

$$\frac{E_b}{N_0} > \ln 2 \quad (-1.6 \text{ dB})$$

Reliable rate of communication and capacity

- Reliable communication at rate R bits/symbol means that one can design codes at that rate with arbitrarily small error probability.

$$C_{awgn} = \log(1 + \mathsf{SNR}) \text{ bits/s/Hz}$$

- To get reliable communication, one *must code over a long block; this* is to exploit the law of large numbers to average out the randomness of the noise.
- Repetition coding over a long block can achieve reliable communication, but the corresponding data rate goes to zero with increasing block length.

10

Capacity of Fading Channels

- Consider the complex baseband representation of a flat fading channel

$$y[m] = h[m]x[m] + w[m]$$

- Assume that the receiver can perfectly track the fading process i.e. coherent reception
- This can be done by a pilot signal if the channel varies slowly relative to the symbol rate
- Let us first look at the situation when the channel gain is random but remains constant i.e. h[m]=m for all m. This is called the slow fading or the quasi static scenario.

11

Capacity of Fading Channels Contd..

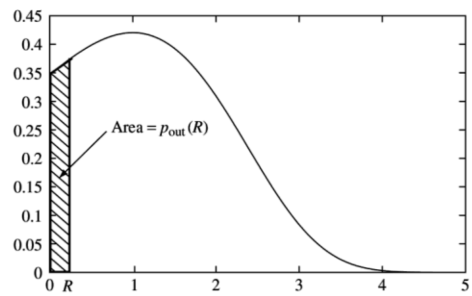
- Conditional on a realization of the channel h, this is an AWGN channel with received SNR as $|h|^2$ SNR
- What is the ultimate performance limit when information can be coded over a sequence of symbols?
- The maximum rate of reliable communication supported by this channel is $\log(1+|h|^2 \text{SNR})$ bits/s/Hz.
- If the channel realization h is such that $\log(1+|h|^2 \text{SNR}) < R$, then whatever the code used by the transmitter, the decoding error probability cannot be made arbitrarily small.
- The system is said to be *in outage, and the outage* probability is

$$p_{\text{out}}(R) := \mathbb{P}\{\log(1+|h|^2 \text{SNR}) < R\}$$

12

Slow Fading Channel Contd..

- Density of log(1+|h|^{2}SNR) , for Rayleigh fading and SNR = 0 dB. For any target rate R, there is a non-zero outage probability.



- the best the transmitter can do is to encode the data assuming that the channel gain is strong enough to support the desired rate R.
- Reliable communication can be achieved whenever that happens, and outage occurs otherwise.

13

Slow Fading Channel Contd..

- For Rayleigh fading (i.e., h is $\mathcal{CN}(0\ 1)$), the outage probability is

$$p_{out}(R) = 1 - \exp\left(\frac{-(2^R - 1)}{\mathsf{SNR}}\right)$$

- At high SNR,

$$p_{out}(R) \approx \frac{(2^R - 1)}{\mathsf{SNR}}$$

- Outage probability decays as 1/SNR.
- Coding *cannot significantly improve the* error probability in a slow fading scenario.
- The reason is that while coding can average out the Gaussian white noise, it cannot average out the channel fade, which affects *all the coded symbols.*

14

Slow Fading Channel Contd..

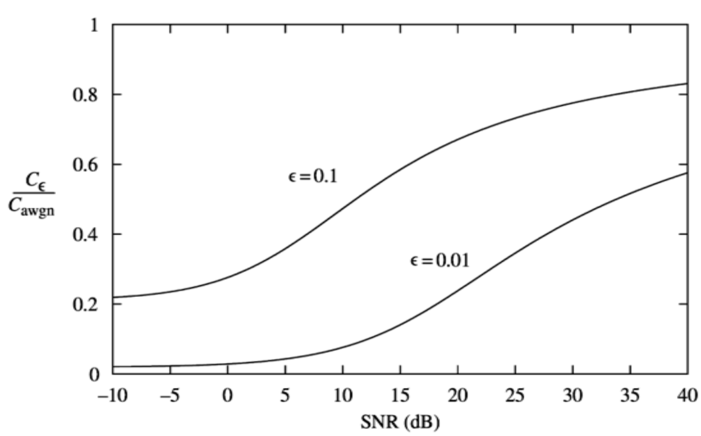
- There is a conceptual difference between the AWGN channel and the slow fading channel.
- In the AWGN channel, one can send data at a positive rate (any rate less than C) while making the error probability as small as desired.
- This cannot be done for the slow fading channel as long as the probability that the channel is in deep fade is non-zero.
- Thus, the capacity of the slow fading channel in the strict sense is zero.
- An alternative performance measure is the ϵ-*outage capacity* C_ϵ.
- This is the largest rate of transmission R such that the outage probability $p_{out}(R)$ is less than ϵ.

$$C_\epsilon = \log(1 + F^{-1}(1-\epsilon)\,\mathsf{SNR})\ \text{bits/s/Hz}$$

where, F is the complementary cumulative distribution function of $|h|^2$ i.e. $F(x)=P[|h|^2 > x]$

15

Slow Fading Channel Contd..



- ϵ-outage capacity as a fraction of AWGN capacity under Rayleigh fading, for ϵ= 0.1 and ϵ= 0.01.

16

Slow Fading Channel Contd..

- It is clear that the impact is much more significant in the low SNR regime.
- At high SNR, $$C_\epsilon \approx \log \mathsf{SNR} + \log(F^{-1}(1-\epsilon))$$ $$\approx C_{\text{awgn}} - \log\left(\frac{1}{F^{-1}(1-\epsilon)}\right)$$
- At low SNR, $$C_\epsilon \approx F^{-1}(1-\epsilon)\mathsf{SNR}\log_2 e$$ $$\approx F^{-1}(1-\epsilon)C_{\text{awgn}}.$$
- Intuitively, the impact of the randomness of the channel is in the received SNR, and the reliable rate supported by the AWGN channel is much more sensitive to the received SNR at low SNR than at high SNR.

17

Receive Diversity

- Let us increase the diversity of the channel by having L receive antennas instead of one.
- Outage occurs whenever:

$$p_{\text{out}}^{\text{rx}}(R) := \mathbb{P}\{\log(1+\|\mathbf{h}\|^2\mathsf{SNR}) < R\}$$

or

$$p_{\text{out}}(R) = \mathbb{P}\left\{\|\mathbf{h}\|^2 < \frac{2^R-1}{\mathsf{SNR}}\right\}$$

- Under independent Rayleigh fading, $||h||^2$ is a sum of the squares of **2L** independent Gaussian random variables and is distributed as Chi-square with 2L degrees of freedom.

$$f(x) = \frac{1}{(L-1)!}x^{L-1}e^{-x}, \qquad x \geq 0$$

18

Receive Diversity Contd..

- Hence at high SNR, the outage probability is given by

$$p_{\text{out}}(R) \approx \frac{(2^R - 1)^L}{L!\,\mathsf{SNR}^L}$$

- We see a diversity gain of L since the outage probability now decays like $1/(\mathsf{SNR})^L$
- This parallels the performance of uncoded transmission, thus coding cannot increase the diversity gain.
- At low SNR and small ϵ

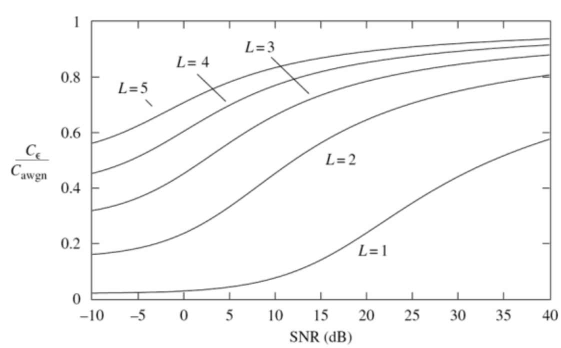
$$C_\epsilon \approx F^{-1}(1-\epsilon)\,\mathsf{SNR}\log_2 e$$
$$\approx (L!)^{\frac{1}{L}}(\epsilon)^{\frac{1}{L}}\,\mathsf{SNR}\log_2 e \text{ bits/s/Hz}$$

- The loss with respect to the AWGN capacity is by a factor of $(\epsilon)^{1/L}$ rather than by ϵ when there is no diversity.

19

Receive Diversity Contd..

- ϵ-outage capacity with L-fold receive diversity, as a fraction of the AWGN capacity log(1+LSNR) for ϵ = 0.01 and different L.



- At ϵ=0.01 and L=2, the outage capacity is increased to 14% of the AWGN capacity.

20

Transmit diversity

- Now suppose there are L transmit antennas but only one receive antenna with a total power constraint of P.
- The transmitter knows the phases and magnitudes of the gains **h** so that it can perform transmit beamforming.
- The capacity of the channel conditioned on the channel gains $\mathbf{h}=[h_1 \ldots\ldots h_L]^t$ is $\log(1+\|\mathbf{h}\|^2 \mathsf{SNR})$
- The outage probability for a fixed rate R is given by

$$p_{\text{out}}^{\text{full-csi}}(R) = \mathbb{P}\{\log(1+\|\mathbf{h}\|^2 \mathsf{SNR}) < R\}$$

- However, this is not true when the transmitter does not know the channel in which case the transmission does not depend on **h**

21

Alamouti Scheme

- The two transmit, single receive channel is written as

$$y[m] = h_1[m]x_1[m] + h_2[m]x_2[m] + w[m]$$

- Let $x_1[1]=u_1$, $x_2[1]=u_2$, $x_1[2]=-u_2^*$, and $x_2[2]=u_1^*$
- For detecting u_1 and u_2, we rewrite this equation as

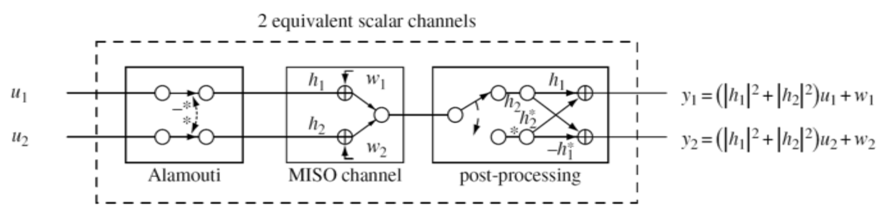
$$\begin{bmatrix} y[1] \\ y[2]^* \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} w[1] \\ w[2]^* \end{bmatrix}$$

- We project **y** onto each of the two columns to obtain the sufficient statistics

$$r_i = \|\mathbf{h}\| u_i + w_i, \qquad i = 1, 2,$$

22

Alamouti Scheme Contd..



•A space-time coding scheme combined with the MISO channel can be viewed as an equivalent scalar channel.
•The outage probability of the scheme is the outage probability of the equivalent channel.

23

Alamouti Scheme Contd..

- Conditioned on h_1, h_2, the capacity of the equivalent scalar channel is
$$\log\left(1 + \|\mathbf{h}\|^2 \frac{\mathsf{SNR}}{2}\right) \text{ bits/s/Hz}$$
- Thus if we now consider successive blocks and use an AWGN capacity achieving code of rate R over each of the streams separately, then the outage probability of each stream is
$$p_{\text{out}}^{\text{Ala}}(R) = \mathbb{P}\left\{\log\left(1 + \|\mathbf{h}\|^2 \frac{\mathsf{SNR}}{2}\right) < R\right\}$$
- Clearly it is worse by 3 dB. The total SNR at the receive antenna at any given time is
$$\left(|h_1|^2 + |h_2|^2\right) \frac{\mathsf{SNR}}{2}$$
- In contrast, when the transmitter knows the channel, the signals add up in phase at the receive antenna and the SNR is
$$\left(|h_1|^2 + |h_2|^2\right) \mathsf{SNR}$$

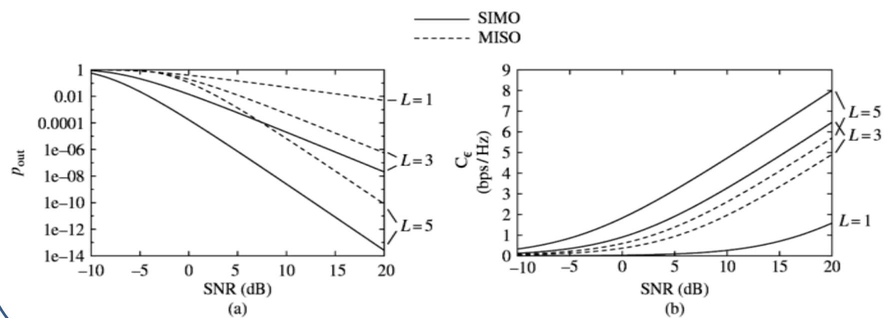
24

Transmit Diversity Contd..

- Comparison of outage performance between SIMO and MISO channels for different L:

  (a) Outage probability as a function of SNR, for fixed R = 1;

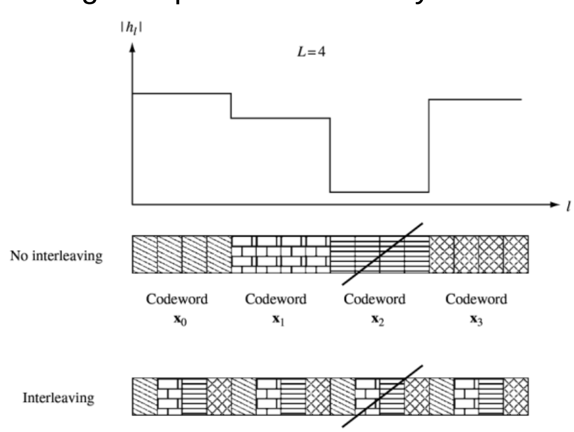
  (b) Outage capacity as a function of SNR, for a fixed outage probability of 10^{-2}.



25

Time and frequency diversity

- Interleaving to exploit time diversity



- In the above example we take one symbol from each coherence period. This can be generalized.

26

Time and Frequency Diversity

- In addition to coding over symbols within one coherence period one can code over symbols from L such periods.
- What is the performance limit in such situations ?
- One can model the situation using the idea of parallel channels. Each of the sub-channels, l=1,…L represents a coherence period of duration T_c symbols:

$$y_\ell[m] = h_\ell x_\ell[m] + w_\ell[m], \qquad m = 1, \ldots, T_c$$

where h_l is the (constant) channel gain during the *l*th coherence period.

- It is assumed that T_c is large enough such that one can code over many symbols in each of the sub-channels.

27

Time and Frequency Diversity

- The maximum rate of reliable communication is

$$\sum_{\ell=1}^{L} \log(1 + |h_\ell|^2 \mathsf{SNR}) \text{ bits/s/Hz}$$

where $SNR=P/N_o$

- If the target rate is R bits/s/Hz per sub-channel, then outage occurs when

$$\sum_{\ell=1}^{L} \log(1 + |h_\ell|^2 \mathsf{SNR}) < LR$$

- Hence, the outage probability of the time diversity channel is precisely

$$p_{\text{out}}(R) = \mathbb{P}\left\{ \frac{1}{L} \sum_{\ell=1}^{L} \log(1 + |h_\ell|^2 \mathsf{SNR}) < R \right\}$$

28

Time and Frequency Diversity Contd.

- Even though this outage performance can be achieved with or without transmitter knowledge of the channel, the coding strategy is vastly different.
- With transmitter knowledge, dynamic rate allocation and separate coding for each sub-channel suffices.
- Without transmitter knowledge, coding across different coherent periods will be necessary.
- The parallel channel can model frequency diversity as well.
- Thus using OFDM, a slow frequency-selective fading channel can be converted into a set of parallel sub-channels, one for each sub-carrier.

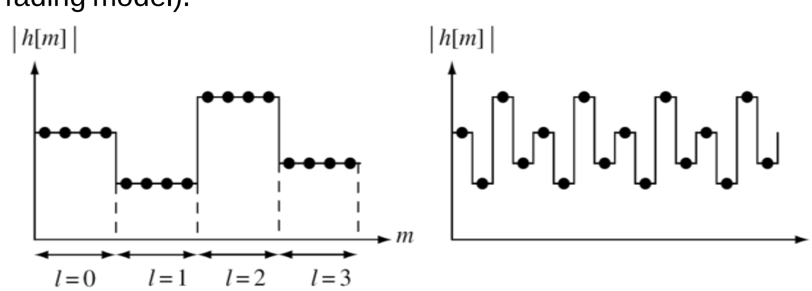
29

Fast Fading Channel Contd..

- If the codeword length spans many coherence periods, it is referred to as a the fast fading regime
- Let us consider a simple model of a fast fading channel:

$$y[m] = h[m]x[m] + w[m]$$

where $h[m]=h_l$ remains constant over the lth coherence period of T_c symbols and is iid across different coherence periods (block fading model).



(a) (b)

30

Fast Fading Channel Contd.

- Suppose coding is done over L such coherence periods. If T_c >>1, we can model this as L parallel sub-channels that fade independently. The outage probability is

$$p_{\text{out}}(R) = \mathbb{P}\left\{ \frac{1}{L}\sum_{\ell=1}^{L} \log(1+|h_\ell|^2\mathsf{SNR}) < R \right\}$$

- For finite L, the quantity

$$\frac{1}{L}\sum_{\ell=1}^{L} \log(1+|h_\ell|^2\mathsf{SNR})$$

is random and there is a non-zero probability that it will drop below any target rate R. Hence we have to resort to the notion of outage.

31

Fast Fading Channel Contd.

- However as L → ∞, the law of large numbers says that

$$\frac{1}{L}\sum_{\ell=1}^{L} \log(1+|h_\ell|^2\mathsf{SNR}) \to \mathbb{E}[\log(1+|h|^2\mathsf{SNR})]$$

- Now we can average over many independent fades of the channel by coding over a large number of coherence time intervals and a reliable rate of communication can indeed be achieved.
- In this situation it is now meaningful to assign a positive capacity to the fast fading channel

$$C = \mathbb{E}[\log(1+|h|^2\mathsf{SNR})] \text{ bits/s/Hz}$$

32

Discussion

- We have seen that time varying fading channels can also have a well defined capacity just like time invariant AWGN channels.
- However, the two situations are quite different from the operational viewpoint.
- In the AWGN channel information flows at a constant rate of log (1+SNR) and reliable communication can take place as long as the coding block length is large enough to average out the white Gaussian noise.
- The resulting delay is smaller and may not be a big concern.

33

Discussion

- In the fading channel information flows at a variable rate of log $(1+|h[m]|^2\text{SNR})$.
- The coding block length now needs to be large enough to average out both the Gaussian noise and the fluctuations of the channel.
- To average out the latter, the coded symbols must span many coherence periods and the coding/decoding delay can be quite significant.
- Interleaving reduces the coding block length but not the delay.

34

Performance Comparison

- The capacity of the fading channel is always less than that of the AWGN channel with the same SNR.
- This follows from Jensen's inequality which says that E[f(x)] ≤ f(E[x]) if f is a strictly concave function.
- Intuitively, the gain from the times when the channel strength is above the average cannot compensate for the loss from the times when the channel strength is below average.
- This follows from the law of diminishing marginal return on capacity from increasing the receiving power.

35

Performance Comparison

- At low SNR, the capacity of the fading channel is

$$C = \mathbb{E}[\log(1 + |h|^2 \mathsf{SNR})] \approx \mathbb{E}[|h|^2 \mathsf{SNR}] \log_2 e = \mathsf{SNR} \log_2 e \approx C_{\text{awgn}}$$

- At high SNR,

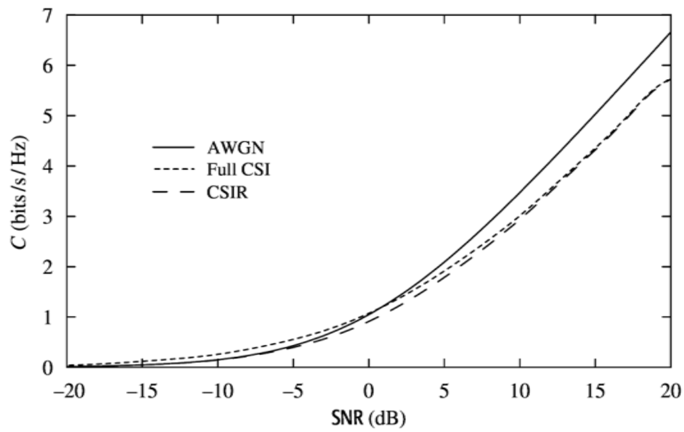
$$C \approx \mathbb{E}[\log(|h|^2 \mathsf{SNR})] = \log \mathsf{SNR} + \mathbb{E}[\log |h|^2] \approx C_{\text{awgn}} + \mathbb{E}[\log |h|^2]$$

  i.e. a constant difference with the AWGN capacity at high SNR. The difference is -0.83 bits/s/Hz for the Rayleigh fading channel
- Equivalently, 2.5 dB more power is needed in the fading case to achieve the same capacity as in the AWGN case

36

Capacity Comparison

- Comparison of the Rayleigh fading channel capacity with the AWGN capacity as a function of SNR



Transmitter Side Information

- So far we have assumed that only the receiver can track the channel
- There are ways in which channel information can be obtained at the transmitter in TDD as well as FDD systems. For eg. power control in the CDMA system also conveys this information.

Slow Fading: Channel Inversion

- In the slow fading case with no channel knowledge at the transmitter, outage occurs whenever the channel cannot support the target data rate R.
- With channel knowledge one option is to control the transmit power such that R can be delivered no matter what the fading state is.

Channel Inversion

- This is known as the channel inversion strategy: the received SNR is kept constant irrespective of the channel state.
- This means that huge amount of paper is required when the channel is bad.
- Practical systems are peak-power constrained and this will not be possible beyond a threshold.
- Systems like IS-95 use a combination of channel inversion and diversity to achieve a target rate with reasonable power consumption.

39

Capacity of Fast Fading with Transmitter Channel Knowledge

- Let us consider the block fading model

$$y[m] = h[m]x[m] + w[m]$$

where $h[m]=h_l$ remains constant over the l^{th} coherence period of T_c (>>1) symbols and is iid across different coherence periods.

- The channel over L such coherence periods can be modeled as a parallel channel with L sub-channels that fade independently.
- For a given realization of the channel gains the capacity of this parallel channel is

$$\max_{P_1,\ldots,P_L} \frac{1}{L}\sum_{\ell=1}^{L} \log\left(1+\frac{P_\ell |h_\ell|^2}{N_0}\right)$$

subject to $\frac{1}{L}\sum_{\ell=1}^{L} P_\ell = P$

40

Transmitter side information

- The optimal power allocation is waterfilling:

$$P_\ell^* = \left(\frac{1}{\lambda} - \frac{N_0}{|h_\ell|^2}\right)^+$$

where λ satisfies

$$\frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{1}{\lambda} - \frac{N_0}{|h_\ell|^2}\right)^+ = P \qquad \text{(X)}$$

- In the context of the frequency selective channel, waterfilling is done over OFDM sub-carriers but here waterfilling is done over time.
- In both cases, the basic problem is power allocation over a parallel channel.

Waterfilling

- The optimal power P_l depends on the channel gain and λ which in turn depends on all other channel gains through (X)
- Hence implementing this scheme would require knowledge of the future channel states. Fortunately, as $L \to \infty$, this non-causality requirement goes away and by the law of large numbers (X) converges to

$$\mathbb{E}\left[\left(\frac{1}{\lambda} - \frac{N_0}{|h|^2}\right)^+\right] = P$$

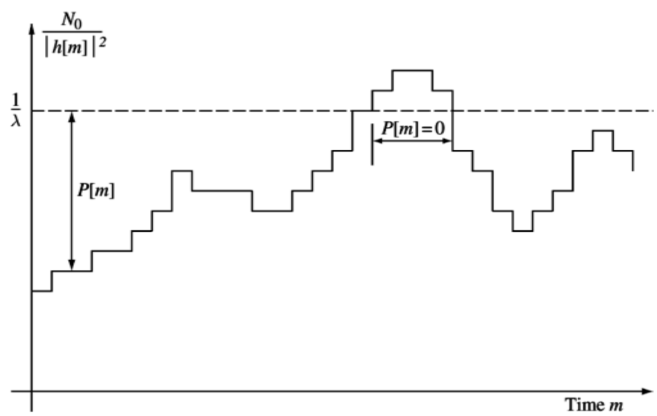
for almost all realizations of the fading process {h[m]}.
- λ depends only on the channel statistics but not on the specific realization of the fading process.
- Hence, the optimal power at any time depends only on the channel gain h at that time

$$P^*(h) = \left(\frac{1}{\lambda} - \frac{N_0}{|h|^2}\right)^+$$

Pictorial Representation of Waterfilling

- The capacity of the fast fading channel with transmitter channel knowledge is

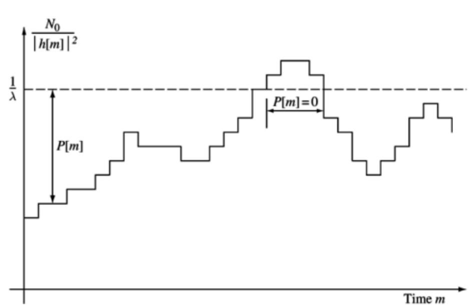
$$C = \mathbb{E}\left[\log\left(1 + \frac{P^*(h)|h|^2}{N_0}\right)\right] \text{ bits/s/Hz}$$



43

Discussion

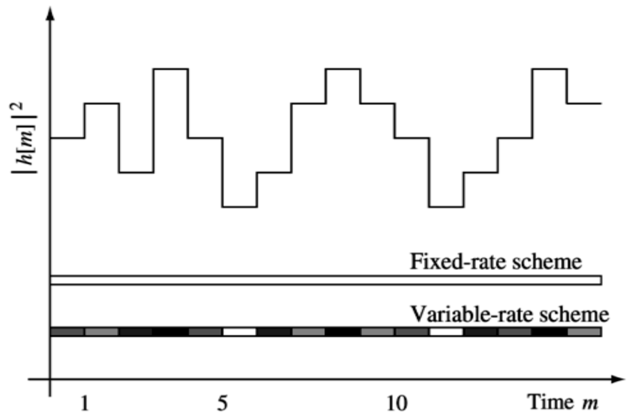
- In general, the transmitter allocates more power when the channel is good and less or even no power when the channel is poor.
- This is precisely the opposite of the channel inversion strategy.
- Only the magnitude of the channel gain is needed to implement the waterfilling scheme.



44

Fixed and Variable Rate Coding

- Variable-rate coding scheme consists of a set of codes of different rates, one for each channel state h.
- When the channel is in state h, the code for that state is used. This can be done since both the transmitter and receiver can track the channel.



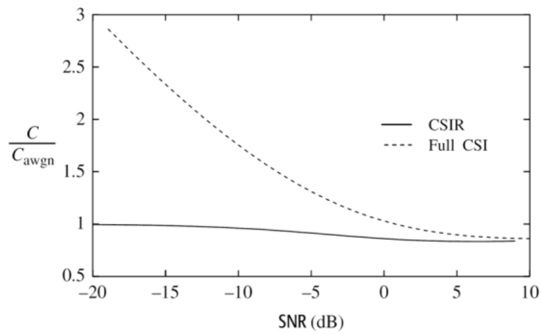
45

Discussion

- A transmit power of $P^*(h)$ is used when the channel gain is h.
- The rate of that code is $\log(1+ P^*(h)|h|^2/N_o)$ bits/s/Hz.
- Coding across channel states is not necessary
- This is in contrast to the case without transmitter channel knowledge where a single fixed-rate code with the coded symbols spanning across different coherence periods is needed.
- Thus knowledge of the channel state at the transmitter not only allows dynamic power allocation but simplifies the code design problem.

46

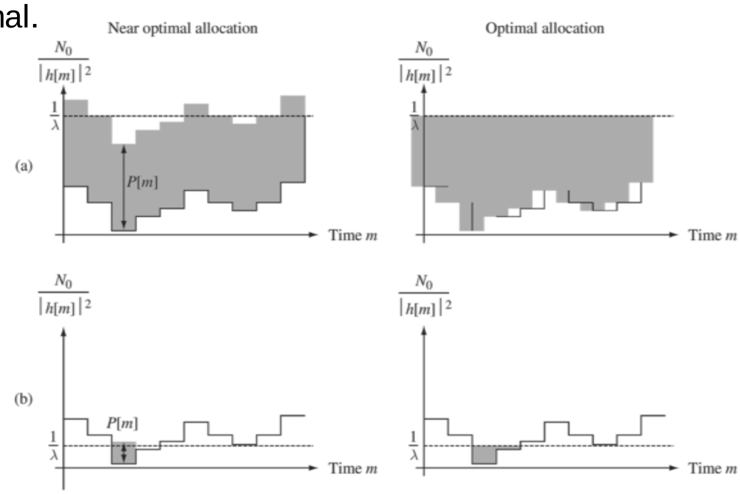
Performance as a Fraction of AWGN Capacity



- At low SNR, the capacity with full CSI is significantly larger than the CSIR capacity.
- Recall that at low SNR, the CSIR capacity is same as the AWGN capacity.
- This means that the capacity of the fading channel can be much larger than when there is no fading.

Transmitter Side Information

- High SNR: allocating equal powers at all times is almost optimal.



- Low SNR: allocating all the power when the channel is strongest is almost optimal.

Discussion

- In a fading channel when SNR is low, with CSIT the transmitter opportunistically transmits only when the channel is near it peak.
- In contrast, in a non-fading AWGN channel the channel stays constant and there are no peaks to take advantage of.
- Overall the performance gain from full CSI is not that large compared to CSIR unless the SNR is very low.
- On the other hand, full CSI potentially simplifies the code design problem as coding across channel states is not necessary. In contrast, with CSIR one has to interleave and code across many channel states.
- Despite the power inefficiency of channel inversion as compared to waterfilling, it offers a constant rate of flow of information and so the associated delay is independent of channel variations.

IS-95 and IS-856 (CDMA 2000 1x EV-DO)

- The contrast between power control in IS-95 and rate control in IS-856 is roughly analogous to that between channel inversion and waterfilling.
- In IS-95 power is allocated dynamically to a user to maintain a constant target rate at all times.
- In IS-856 rate is adapted to transmit more information when the channel is strong which is suitable for data since it does not have a stringent delay requirement.
- However, unlike waterfilling there is no dynamic power adaptation in IS-856, only rate adaptation.

Rate Adaptation in IS-856

Requested rate (kbits/s)	SINR threshold (dB)	Modulation	Number of slots
38.4	−11.5	QPSK	16
76.8	−9.2	QPSK	8
153.6	−6.5	QPSK	4
307.2	−3.5	QPSK	2 or 4
614.4	−0.5	QPSK	1 or 2
921.6	2.2	8-PSK	2
1228.8	3.9	QPSK or 16-QAM	1 or 2
1843.2	8.0	8-PSK	1
2457.6	10.3	16-QAM	1

Impact of Prediction Uncertainty

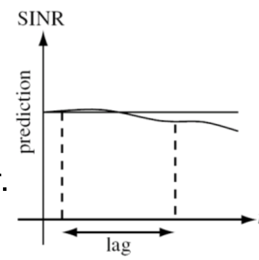
- At 1.9 GHz coherence time varies from 25 ms (3km/h speed) to 2.5 ms (30 km/h speed) to less than 1 ms for speeds approaching 100 km/h
- At one extreme the channel can be quite accurately predicted while on the other extreme this is not possible but there is significant time diversity over the duration of the packet.
- Recall that the fast fading capacity is given by

  $$C = \mathbb{E}\left[\log\left(1+|h|^2\mathsf{SNR}\right)\right] \approx \mathbb{E}[|h|^2]\mathsf{SNR}\log_2 e \text{ bits/s/Hz}$$

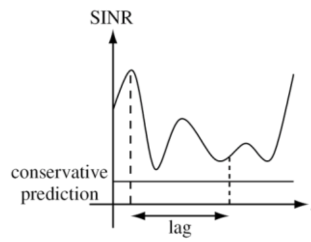
  in the low SNR regime.
- Thus to determine an appropriate rate it suffices to predict the average SINR which is quite easy to predict.

Impact of Prediction Uncertainty Contd..

- Coherence time is long compared to the prediction time lag; predicted SINR is accurate. Near perfect CSI at transmitter.



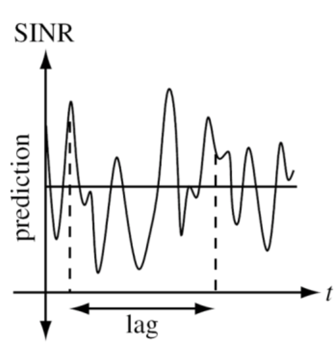
- Coherence time is comparable to the prediction time lag, predicted SINR has to be conservative to meet an outage criterion.



53

Impact of Prediction Uncertainty Contd..

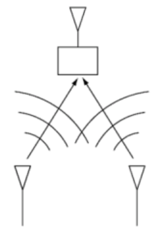
- Coherence time is short compared to the prediction time lag; prediction of average SINR suffices. No CSI at the transmitter.



54

Multiuser Uplink Channel

Capacity via successive interference cancellation



- Received signal with two users

$$y[m] = x_1[m] + x_2[m] + w[m],$$

- User k has an average power constraint P_k

- Capacity region is the set of all possible pairs (R_1 , R_2) such that simultaneously user 1 and user 2 can reliably communicate at rate R_1 and R_2 respectively
- Since the two users share the same bandwidth there is naturally a tradeoff between the reliable communication rates of the users.

55

Multiuser Uplink Channel

- From the capacity region, one can derive performance measures of interest. For eg:
- The symmetric capacity:

$$C_{\text{sym}} := \max_{(R,R)\in\mathcal{C}} R$$

- The sum capacity:

$$C_{\text{sum}} := \max_{(R_1,R_2)\in\mathcal{C}} R_1 + R_2$$

- In general, the capacity region is the set of all rates satisfying the three constraints

$$R_1 < \log\left(1 + \frac{P_1}{N_0}\right)$$

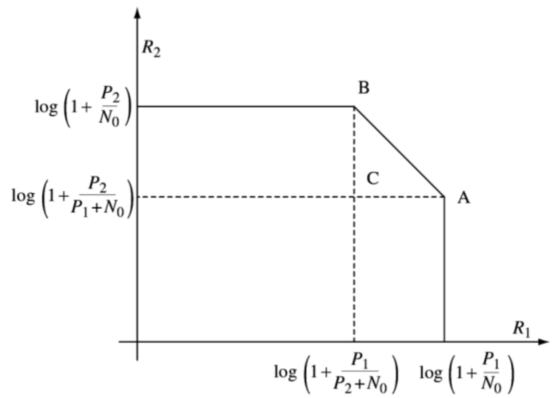
$$R_2 < \log\left(1 + \frac{P_2}{N_0}\right)$$

$$R_1 + R_2 < \log\left(1 + \frac{P_1 + P_2}{N_0}\right)$$

56

Capacity Region

- Two-user uplink channel.



57

Successive Interference Cancellation

- Something surprising happens: user 1 can achieve its single-user bound while at the same time user 2 can get a non-zero rate; in fact as high as its rate at point A, i.e.,

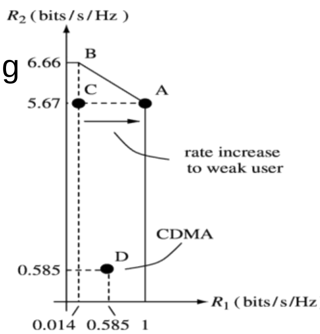
$$R_2^* = \log\left(1+\frac{P_1+P_2}{N_0}\right) - \log\left(1+\frac{P_1}{N_0}\right) = \log\left(1+\frac{P_2}{P_1+N_0}\right)$$

- Each user encodes its data using a capacity-achieving AWGN channel code.
- The receiver decodes the information in two stages
- It first decodes the data of user 2 treating the signal from user 1 as Gaussian interference and then can achieve the above rate.
- Then it subtracts user 2's signal from the aggregate signal and can decode user 1 which is left only with background noise.

58

Comparison with CDMA

- In the case when the received powers of the users are very disparate, successive cancellation (point A) can provide a significant advantage to the weaker user compared to conventional CDMA decoding (point C).
- The conventional CDMA solution is to control the received power of the strong user to equal that of the weak user (point D), but then the rates of both users are much lower. Here, P1/N0 = 0 dB, P2/N0 = 20 dB.



59

Information Theoretic View of CR

- In the terminology of information theory, a cognitive radio is defined by the availability and utilization of network side information.
- Formally speaking, a cognitive radio is a wireless communication system that intelligently utilizes any available side information about the
  - Activity
  - Channel conditions
  - Codebooks
  - messages of other nodes sharing the spectrum.

Some Possible CR Scenarios

- Based on the type of the side information available, along with the regulatory constraints, cognitive users seek to
  - Underlay
  - Overlay

    Or
  - Interweave

  their signals with those of existing users without significantly impacting their communication
- Information Theoretic perspective seeks the capacity regions of these scenarios.

Capacity and Degrees of Freedom

- Obtaining the capacity region of a wireless network is an open problem for most networks of interest.
- However, useful insights can be obtained by approximate/ asymptotic characterizations of the capacity, for example, the network degrees of freedom.
- Intuitively, it represents the number of accessible interference-free dimensions.
- The concept of degrees of freedom was first introduced in the context of single-user multiple-input multiple-output (MIMO)systems by Zheng *et. al.,*2003[1].

[1]L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels ", IEEE Trans. Inf. Theory, vol. 49, May 2003.

Degrees of Freedom

- The degrees of freedom of a channel is the dimension of the received signal space.
- In a channel with two transmit and a single receive antenna, the received signal is a scalar and hence it has only one degree of freedom.
- But in a channel with one transmit and L receive antenna, the received signal lies in an L-dimensional vector space.

$$\mathbf{y} = \mathbf{h}x + \mathbf{w} \quad \text{where} \quad \mathbf{y} := [y_1, \ldots, y_L]^t$$

$$\mathbf{h} = [h_1, \ldots, h_L]^t$$

$$\mathbf{w} = [w_1, \ldots, w_L]^t$$

63

Examples

- However the received signal is only one dimensional (it does not span the L dimensional space)

$$\mathbf{y} = \mathbf{h}x + \mathbf{w}$$

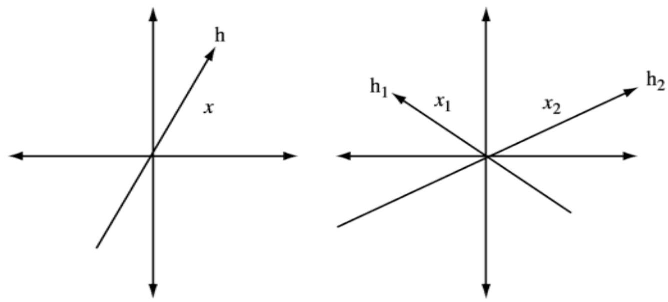
- Hence the degree of freedom is only one
- However, in a 2 2 channel there are potentially two degrees of freedom.

$$\mathbf{y} = \mathbf{h}_1 x_1 + \mathbf{h}_2 x_2 + \mathbf{w}$$

64

Examples Contd..

- In the 1 2 channel, the signal space is one-dimensional, spanned by **h**.
- In the 2 2 channel, the signal space is two-dimensional, spanned by $\mathbf{h}_1$ and $\mathbf{h}_2$.



65

Degrees of Freedom

- A single-user MIMO channel with M transmit and N receive antennas and a full rank channel matrix has min(M; N) degrees of freedom.
- In a network with M transmitters and N receivers, some dimensions cannot be resolved because of the distributed nature of the signal processing.
- At high SNR, the number of degrees of freedom is the principal determinant of the network capacity.
- The sum capacity is given by

$$C_{\Sigma}(\mathrm{SNR}) = d\log(\mathrm{SNR}) + o(\log(\mathrm{SNR}))$$

where d is the network degrees of freedom and o(log(SNR)) represents the approximate error term which becomes negligible compared to log(SNR) as SNR increases

Cognitive Radio Paradigms

- Underlay, Overlay and Interweave

Underlay

- Allows cognitive users to operate if the interference caused to noncognitive users is below a given threshold.
- Assumes cognitive users have knowledge of the interference caused by their transmission to the receivers of noncognitive users.
- In this scenario, a cognitive user is often called a secondary user which cannot significantly interfere with the communication of primary users.

Underlay Contd..

- The interference constraint for the noncognitive users may be met by using
  - Multiple antennas to direct the cognitive signals away from the noncognitive receivers
  - A wide bandwidth over which the cognitive signal can be spread below the noise floor, then despread at the cognitive receiver (e.g. UWB and spread spectrum)
- Since the interference constraints are quite restrictive, cognitive users are limited to short range communication.
- Underlay paradigm is most common in the licensed spectrum (e.g., UWB underlays many licensed spectral bands)

Overlay Scenario

- A cognitive user has knowledge of the noncognitive users' codebooks and its messages as well.
- It assumes the noncognitive message is known at the cognitive transmitter when the noncognitive user begins its transmission.
- This is impractical for an initial transmission
- However, it may hold for a message retransmission or if the cognitive user is close by and the noncognitive user send its message to it prior to transmission
- Knowledge of a noncognitive user's message and/or codebook can be exploited in a variety of ways to either cancel or mitigate the interference seen at the cognitive and noncognitive receivers.

Overlay Contd..

- This information can be used to cancel the interference at the cognitive receiver, by DPC.
- The cognitive users can use this knowledge and assign part of their power for their own communication and the remainder to relay the noncognitive transmissions.
- The increase in the noncognitive user's SNR can be offset by the decrease in it's SNR due to the interference by the remainder of the cognitive user's transmit power.
- This guarantees that the noncognitive user's rate remains unchanged while the cognitive user allocates part of its power for its own transmissions.

Overlay Contd..

- Overlay paradigm can be applied to either licensed or unlicensed band communications.
- In licensed bands, cognitive users would be allowed to share the band with the licensed users since they would not interfere with, and might even improve, their communications.
- In unlicensed bands, cognitive users would enable a higher spectral efficiency by exploiting message and codebook knowledge to reduce interference.

Interweave Scenario

- Based on the idea of opportunistic communication and was the original motivation for cognitive radio
- There exist temporary space–time–frequency voids, referred to as spectrum holes, that are not in constant use in both the licensed and unlicensed bands.
- The interweave technique requires knowledge of the activity information of the noncognitive (licensed or unlicensed) users in the spectrum.
- All users can be considered as cognitive with existing users as primary and new users as secondary that cannot interfere with communications already taking place

Comparative View

Underlay	Overlay	Interweave
Channel Side Information: Cognitive (secondary) transmitter knows the channel strength to noncognitive (primary) receiver(s).	Codebook Side Information: Cognitive nodes know channel gains, codebooks and the messages of the noncognitive users.	Activity Side Information: Cognitive user knows the spectral holes in space, time or frequency when the noncognitive user is not using these holes.
Cognitive user can transmit simultaneously with noncognitive user as long as interference caused is below an acceptable limit.	Cognitive user can transmit simultaneously with noncognitive user, the interference to noncognitive user can be offset by using part of the cognitive user 's power to relay the noncognitive user's message.	Cognitive user can transmit simultaneously a with noncognitive user only in the event of a false spectral hole detection.
Cognitive user's transmit power is limited by the interference constraint.	Cognitive user can transmit at any power , the interference to noncognitive user can be offset by relaying the noncognitive user's message.	Cognitive user's transmit power is limited by the range of its spectral hole sensing.

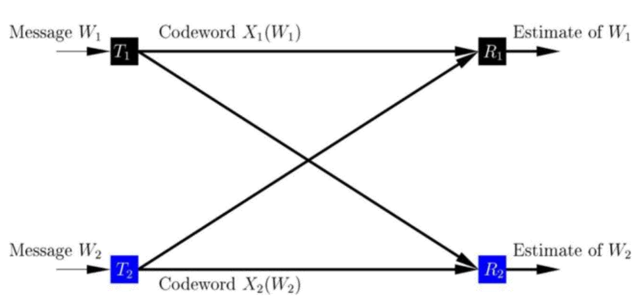
Interference Channels

- Interference channel provides a fundamental building block to the capacity as well as encoding and decoding strategies for cognitive networks
- Multiple pairs wish to communicate simultaneously in the presence of mutual interference.
- The users are not assumed to be cognitive – they do not monitor the activity or decode messages of other users.
- However, it is usually assumed that all terminals know the channel gains and the codebooks.
- The communication problem is to determine the highest rates that can simultaneously be achieved – capacity region.

Interference Channels Contd..

- This performance can serve as a benchmark to evaluate the gains of cognition.
- Even for the smallest interference network consisting of two transmitter–receiver pairs, this problem has remained unsolved for more than three decades.
- Clearly we still have a long way to go before we understand the science of mitigating and exploiting interference.

Two Transmitter Two Receiver Scenario



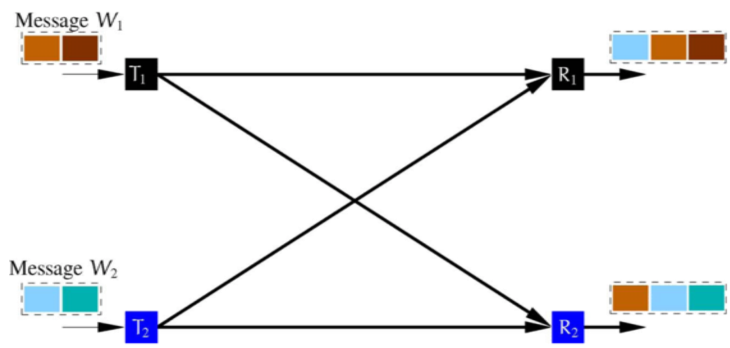
Capacity Region

- Information-theoretic analysis towards obtaining the capacity region is typically performed in two steps.
  - Propose a specific encoding and decoding scheme and evaluate its achievable rate region.
  - Determine the an outer bound to the rate region that cannot be exceeded by any encoding scheme.
- If the two bounds meet, the capacity region is knownand the proposed encoding scheme is capacity achieving.
- The capacity region is known when the interference is *strong.* The interfering signal can be jointly decoded with the desired signal.

Rate-Splitting

- In general, the interference is not strong enough to allow for decoding of the unwanted message without reducing its rate.
- In this case, rate-splitting can be used at the encoders to allow the receivers to decode a part of the unwanted message. Each encoder divides its message into two sub-messages and encodes them separately.
- A receiver decodes one sub-message of the other user and cancels a part of this interference.
- This will increase his rate but will lower the rate for the other pair. Thus there is a tradeoff between sending a message only to the desired receiver and allowing partial decoding at the other.

Rate – Splitting Contd..



- User 1 and 2 sub-messages are represented with shades of brown and blue, respectively.
- Rate-splitting achieves the best rates known today