Interference Mitigation in Large Random Wireless Networks

Matthew Aldridge

Interference Mitigation in Large Random Wireless Networks

Matthew Aldridge

A dissertation submitted to
the University of Bristol in
accordance with the requirements
for award of the degree of
Doctor of Philosophy in
the Faculty of Science

School of Mathematics

2011

30,000 words

Abstract

A central problem in the operation of large wireless networks is how to deal with interference – the unwanted signals being sent by transmitters that a receiver is not interested in. This thesis looks at ways of combating such interference.

In Chapters 1 and 2, we outline the necessary information and communication theory background. We define the concept of capacity – the highest rate at which information can be sent through a network with arbitrarily low probability of error. We also include an overview of a new set of schemes for dealing with interference known as interference alignment, paying special attention to a channel-state-based strategy called ergodic interference alignment.

In Chapter 3, we consider the operation of large regular and random networks by treating interference as background noise. We consider the local performance of a single node, and the global performance of a very large network.

In Chapter 4, we use ergodic interference alignment to derive the asymptotic sum-capacity of large random dense networks. These networks are derived from a physical model of node placement where signal strength decays over the distance between transmitters and receivers.

In Chapter 5, we look at methods of reducing the long time delays incurred by ergodic interference alignment. We decrease the delay for full performance of the scheme, and analyse the tradeoff between reducing delay and lowering the communication rate.

In Chapter 6, we outline a problem of discovering which users interfere with which; a situation that is equivalent to the problem of pooled group testing for defective items. We then present some new work that uses information theoretic techniques to attack group testing. We introduce for the first

time the concept of the group testing channel, which allows for modelling of a wide range of statistical error models for testing. We derive new results on the number of tests required to accurately detect defective items, including when using sequential 'adaptive' tests.

Chapter 7 concludes and gives pointers for further work.

Acknowledgments

This thesis would not exist without the help of a great many people.

My supervisors are Oliver Johnson and Robert Piechocki – a lot of the work in this thesis is joint work with them. Without Olly and Rob's wisdom, insight, knowledge, help, guidance, encouragement, and gentle prodding, none of this would have been possible. Thank you, Olly; thank you, Rob.

My research for this thesis was funded by Toshiba Research Europe Ltd. Thanks to the Toshiba Telecommunication Research Laboratory in Bristol and its directors. I have also received support from the Engineering and Physical Sciences Research Council, via the University of Bristol 'Bridging the Gaps' cross-disciplinary feasibility account (`EP/H024786/1`). Thanks to Dino Sejdinovic and Olly, who got me involved.

This thesis has benefited, directly or indirectly, from conversations with Olly, Rob, Dino, Henry Arnold, Kara Barwell, Lee Butler, Laura Childs, Justin Coon, Andreas Müller, Clare Raychaudhuri, Magnus Sandell, Andrew Smith, and Will Thompson; conversations that were usually fun as well as useful. It has been checked, in whole or in part, by Olly, Rob, and Laura, and contains considerably fewer errors (mathematical, grammatical, and typographical) than it would have done without their care and attention. My academic reviewers, Ayalvadi Ganesh and Feng Yu, checked three precursors to this document, and had many helpful suggestions. Thanks everyone.

My examiners were Olivier Lévêque and Ganesh. They both read this thesis thoroughly, and made a large number of helpful comments. Their suggestions for improvements and clarifications have made this thesis better, a number of references they located have made it more complete, and fixing the errors they spotted has made it more accurate.

Writing this thesis would not have been possible without free software. Thanks to Donald Knuth, Leslie Lamport, and the LaTeX3 team for LaTeX; to

the AMS, David Carlise, Lars Madsen, and Peter Wilson for various LaTeX macros; to Tino Weinkauf, Sven Wiegand, and the TeXnicCenter team; and to the Inkscape team.

Thanks Mum, thanks Dad, thanks Alice, for many, many things. And thanks Laura (for being awesome like a pigeon).

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others is indicated as such. Any views expressed in the dissertation are those of the author.

Signed:

Date:

Table of contents

List of definitions and theorems

Introduction

A central problem in the operation of large wireless networks is how to deal with *interference* – the unwanted signals being sent by transmitters that a receiver is not interested in. This thesis looks at ways of combating such interference in large random wireless netoworks.

In **Chapter 1: Information**, we briefly summarise information theory in the single user (point-to-point) case.

A *channel* models how signals are corrupted by noise. We pay particular attention to the *Gaussian channel*, which is a good model for real-world wireless communication, and the *finite field channel*, which can be thought of as a discretisation of the Gaussian channel

The *capacity* of a channel tells us how much information we can send through the channel for an arbitrarily low probability of error. *Shannon's channel coding theorem* tells us how to calculate the capacity of a channel. We also demonstrate the capacity of the Gaussian channel under a power constraint.

Fading models how signals can decay and distort when sent over long distances. We investigate three types of fading – *fixed*, *slow*, and *fast* – and show how they affect the channel capacity.

In **Chapter 2: Interference**, we extend our study to multiuser networks.

We look at information theoretic models of wireless networks, concentrating on the *interference network*, where many transmitter–receiver pairs want to communicate through the same medium. This network suffers from the problem of interference.

Weak interference can be ignored and treated as background noise, while strong interference can be decoded and subtracted. The main problem for networks is interference of a similar strength to the desired signal.

We look at *resource division* strategies, which share the channel resources between the users. While such schemes are simple to operate, they perform poorly when the number of users is high.

Of more interest are new *interference alignment* strategies. These work by the following idea: if transmitters plan their signals carefully, then for each receiver the interfering signals can be aligned together, with the desired signal split off separately. Interference alignment techniques offer potentially far higher performance than resource division schemes. We pay particular attention to a channel state-based strategy called *ergodic interference alignment*.

Chapter 3: Regular and Poisson random networks, shows how a simple interference-as-noise technique can be useful when communicating over short hops in well-structured networks.

In a d-dimensional *regular network*, nodes are placed on the grid $\mathbb{Z}^d$. We show that if signals decay like distance$^{-\alpha}$ for $\alpha > d$, then all nodes can communicate at some fixed rate r. We call this *linear growth*, as the sum-rate of communication of a collection of nodes scales linearly with the number of nodes.

We also look at nearest-neighbour communication in *Poisson random networks*, where nodes are placed at random like a Poisson point process. We give bounds on the *outage probability*, the chance that a given link is unable to communicate at some fixed rate. We also show that linear growth occurs with probability tending to 1.

This chapter is joint work with Oliver Johnson and Robert Piechocki.

In **Chapter 4: Sum-capacity of random dense Gaussian interference networks**, we consider *spatially separated IID networks* with *power-law attenuation*, a natural model for wireless networks. We derive the asymptotic sum-capacity of such networks by using ergodic interference alignment to show achievability, and subtle probabilistic and counting techniques to show the converse.

We also give an alternative proof (with an improved rate of convergence) to a recent theorem of Jafar on the sum-capacity of large random networks.

This chapter is joint work with Oliver Johnson and Robert Piechocki. This research has been published in IEEE Transactions on Information Theory *[1], and in the* Proceedings of the 2010 IEEE International Symposium on Information Theory *[2].*

Chapter 5: Delay–rate tradeoff in ergodic interference alignment considers the long blocklengths required to perform ergodic interference alignment. We outline a new scheme called JAP($\mathbf{a}$) and study a *beamforming* extension and derived *child schemes*.

We show how to reduce the time delay for full performance of ergodic interference alignment. We also show how delay can be reduced even further

for the tradeoff of a decrease in communication rate. We analyse the best schemes for small networks, and as the size of the network tends to infinity.

This chapter is joint work with Oliver Johnson and Robert Piechocki. This research has been submiited to IEEE Transactions on Communication Theory *– a preprint is available on the* `arXiv` *[3].*

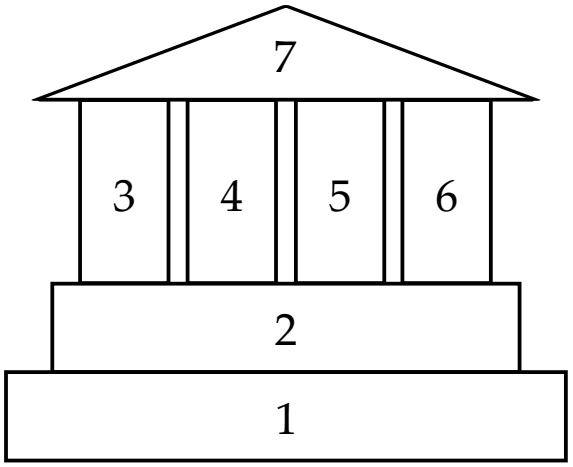
We begin **Chapter 6: Interference, group testing, and channel coding** by considering a problem where receivers must aim to detect which transmitters interfere with them. We show that our formulation of this problem is equivalent to the problem of combinatorial *group testing*.

Recent work by Atia and Saligrama has shown how channel coding techniques can shed light on the problem of group testing. We extend their results, by defining *group testing channels*, and identifying the *only-defects-matter property*, under which an important theorem holds.

We give the first information theoretic analysis of *adaptive group testings* – where test pools can be constructed sequentially based on previous outcomes – by drawing a comparison with the problem of channel coding with feedback.

The thesis finishes with **Chapter 7: Conclusions and further work**.

Schematic representation of chapter dependencies



1

Information

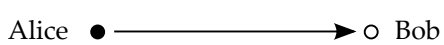
> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.
>
> — Claude E Shannon
>
> *A Mathematical Theory of Communication* [4, page 1]

In this chapter, we examine the subject of information theory, and in particular channel coding, which forms the mathematical basis for studying communication.

We start by giving a brief overview of the subject, and making note of a 'handbook' of useful definitions and facts. We then go through the more formal mathematics of point-to-point communication, concentrating on accurate models for real-life wireless communication. Finally, we study fading, which allows us to model how signals distort as passed through space.

1.1 Information theory: a very short introduction

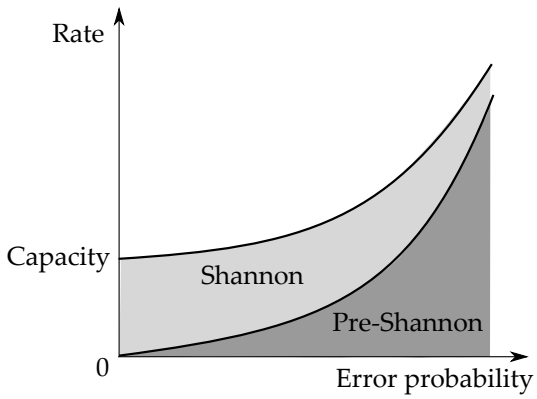
Information theory is the mathematical framework used for studying the transmission of messages in the presence of noise. It was founded by Claude Shannon in his seminal paper of 1948, "A mathematical theory of communication" [4]. Shannon's information theory involves the sending of a message – information that a transmitter (typically called Alice) might wish a receiver (Bob) to know – through a channel, such as a telephone line, an internet connection or a computer cable.



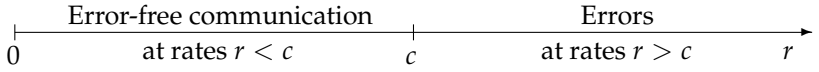
The method, language or standard used to transfer the message is called the *code*. (We use 'code' in the sense of 'Morse code' – there is no intention to keep the transmitted signal secret as well.)

One of the major goals of information theory involves quantifying how much information can be sent down a channel and how quickly. For example, if Alice wants to arrange a meeting with Bob, she might send the message "Hi Bob, meet me at five o'clock on Monday." But if the line is very crackly, Bob might interpret the message incorrectly: "Hi Bob, meet me at nine o'clock on Sunday." How could Alice ensure that this doesn't happen? Perhaps Alice could use a code where she repeats the sentence a number of times, hoping that it would be more likely that Bob could deduce the intended message. But this takes a longer amount of time – we say that her *rate* of communication is very low – and phone calls are expensive, making this undesirable.

Unsurprisingly, there is a trade-off to be made between the rate at which Alice can send the information and the probability that Bob receives it without error. Before Shannon, it was widely assumed that the only way to make the error probability as small as desired was to reduce the rate of communication toward zero too [5, Section 5.1]. However, Shannon discovered that, while this trade-off certainly exists, the error probability can made arbitrarily small while maintaining a communication rate bounded away from zero.



In other words, there is a cutoff rate c such that if we attempt to send information at a rate $r < c$, we can do so with an arbitrarily low risk of error, whereas if we attempt to send information at a rate $r > c$, the probability of error is bounded away from 0. Shannon called this cutoff rate c the *capacity* of the channel.



Shannon managed to calculate the capacity of a number of communication channels in terms of the statistical properties of the noise in a channel. The result is known as *Shannon's channel coding theorem* (and is stated as Theorem 1.11 later).

The aim of this thesis is to find bounds and approximations for capacities of complicated multi-user networks. In particular, we will be interested in networks that model large real-world wireless networks, such as WiFi computer networks or BlueTooth.

The capacity of a channel, such as a wireless link, tells us the maximum rate at which we can send information while being assured the messages are received accurately. Note however that merely knowing the capacity does not give us a method of achieving communication at, or even near, capacity. Nonetheless, the capacity is still a useful benchmark for the quality of a channel. First, it gives us a 'best case' against which we can compare any technologies: if a new code allows us to communicate at a rate near capacity, then this technology is about as good as it's going to get, and there is no need to spend more money on research. Second, studying the mathematical form of the capacity may help us improve the channel itself; for instance, whether extra resources would be best spent increasing power, bandwidth, or the number of antennas.

1.2 Handbook of useful facts

The following basic concepts of information theory will be referred to often in this thesis; we collect them here for reference.

More information is available in any basic information theory textbook – Cover and Thomas's *Elements of Information Theory* [6] is a favourite of mine.

Mass and density functions. For a discrete random variable X, we denote its *probability mass function* by $p(x) := \mathbb{P}(X = x)$. If X is continuous, $p(x)$ denotes its *probability density function*. *Joint* and *conditional mass/density functions* are denoted $p(x, y)$ and $p(y \mid x)$ respectively.

Entropy and related concepts. The *entropy* of a discrete random variable X is

$$\mathbb{H}(X) := \mathbb{E} \log \frac{1}{p(X)} = \sum_x p(x) \log \frac{1}{p(x)}, \tag{HB1}$$

where here, as everywhere in this thesis, $\log := \log_2$ denotes the binary logarithm. When X is continuous, the sum is replaced by an integral. (This last comment holds for all the following definitions.)

The *joint entropy* of the pair (X, Y) is similarly

$$\mathbb{H}(X, Y) := \mathbb{E} \log \frac{1}{p(X, Y)} = \sum_x \sum_y p(x, y) \log \frac{1}{p(x, y)}. \tag{HB2}$$

The *conditional entropy* of Y given X is

$$\mathbb{H}(Y \mid X) := \mathbb{E} \log \frac{1}{p(Y \mid X)} = \sum_x \sum_y p(x, y) \log \frac{1}{p(y \mid x)}. \tag{HB3}$$

If X and Y are independent, then it is easy to show [6, Theorem 2.6.5] that

$$\mathbb{H}(Y \mid X) = \mathbb{H}(Y) \qquad \mathbb{H}(Y + X \mid X) = \mathbb{H}(Y) \tag{HB4}$$

The *relative entropy distance* from one probability function $p(x)$ to another $q(x)$ is

$$\mathbb{D}\big(p(x) \,\|\, q(x)\big) := \mathbb{E}_{p(x)} \log \frac{p(X)}{q(X)} = \sum_x p(x) \log \frac{p(x)}{q(x)} \geq 0. \tag{HB5}$$

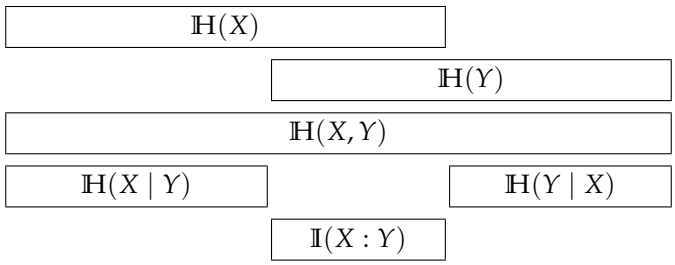
The *mutual information* between X and Y is

$$\begin{aligned} \mathbb{I}(X : Y) &:= \mathbb{D}\big(p(x, y) \,\|\, p(x)p(y)\big) \\ &= \mathbb{E}_{p(x,y)} \log \frac{p(X, Y)}{p(X)p(Y)} \\ &= \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \end{aligned} \tag{HB6}$$

It easy to show [6, Theorem 2.4.1] that

$$\mathbb{I}(X : Y) = \mathbb{H}(Y) - \mathbb{H}(Y \mid X), \tag{HB7}$$

Hu correspondence A useful way to memorise the relationship between these concepts is the Hu correspondence. In the following picture, the area of each rectangle corresponds to the quantity.



The complex Gaussian distribution. A *circularly-symmetric complex Gaussian random variable* $Z \sim \mathbb{C}\mathrm{N}(0, \sigma^2)$ with variance $\sigma^2 = \mathbb{E}\,|Z|^2$ is defined by the probability density function

$$p(z) = \frac{1}{\pi\sigma^2} \mathrm{e}^{-|z|^2/\sigma^2} \qquad z \in \mathbb{C}. \tag{HB8}$$

Maximum entropy. Out of all discrete random variables X on the set $\{0, 1, \ldots, q-1\}$, the maximum entropy is achieved when X is uniform [6, Theorem 2.6.4], giving

$$\max_X \mathbb{H}(X) = \mathbb{H}\big(U(\{0, 1, \ldots, q-1\})\big) = \log q. \tag{HB9}$$

Of all continuous random variables Z on $\mathbb{C}$ with power $\mathbb{E}\,|Z|^2$ at most σ^2, the maximum entropy is achieved when $Z \sim \mathbb{C}\mathrm{N}(0, \sigma^2)$ [6, Example 12.2.1], giving

$$\max_{Z:\mathbb{E}\,|Z|^2 \le \sigma^2} \mathbb{H}(Z) = \mathbb{H}\big(\mathbb{C}\mathrm{N}(0, \sigma^2)\big) = \log(\pi \mathrm{e} \sigma^2). \tag{HB10}$$

Typical set. Let X be a random variable on a countable set $\mathcal{X}$. Given a sequence $\mathbf{X} = (X[1], X[2], \ldots, X[T]) \in \mathcal{X}^T$ of T random draws from X, then $\mathbf{X}$ will very likely take one of only $2^{T\mathbb{H}(X)} \le |\mathcal{X}|^T$ different values, and each of these values is almost equally likely. These values make up the so-called *typical set*, and this property is called the *asymptotic equipartition property*.

We say $(\mathbf{X}, \mathbf{Y})$ is *jointly typical* of (X, Y) if $\mathbf{X}$ is in the typical set of X, $\mathbf{Y}$ is in the typical set of Y, and the pair $(\mathbf{X}, \mathbf{Y})$ is in the typical set of the pair (X, Y).

1.3 Channels, codes, and capacity

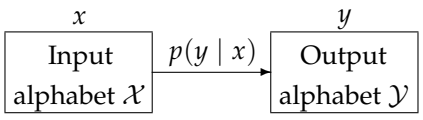
We will now set out mathematically the ideas of channels, codes, and capacity from Section 1.1.

To specify a channel, we need to say what inputs are allowed into the channel, what outputs can be produced, and how the noise randomly corrupts the input.

A common example of a channel is the *binary symmetric channel*. The BSC allows 'bits' – binary digits: 0s and 1s – into the channel. It then either outputs the same bit or, with some fixed probability, flips to the other bit.

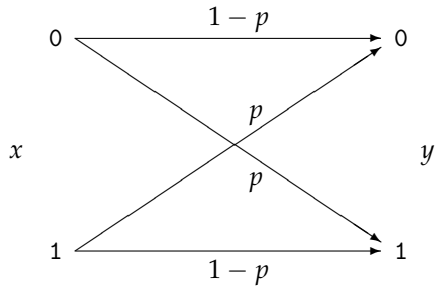
Definition 1.1. A communication *channel* consists of

1. a set $\mathcal{X}$, the *input alphabet*;
2. a set $\mathcal{Y}$, the *output alphabet*;
3. a *probability transition function* $p(y \mid x)$ relating the two.



Definition 1.2. We can now formally define the *binary symmetric channel* with error probability $p < \frac{1}{2}$. This channel is defined by alphabets $\mathcal{X} = \mathcal{Y} = \{0, 1\}$ and transition function

$$p(0 \mid 0) = 1 - p \qquad p(1 \mid 0) = p$$
$$p(0 \mid 1) = p \qquad p(1 \mid 1) = 1 - p.$$



A channel which is an accurate and widely-used model for wireless communication [5, Chapter 5.1] is the *Gaussian channel*. It arises from the sampling of a bandlimited continuous-time channel with white noise [7]. (For convenience, we assume a unit bandwidth.) Gaussian white noise is used for

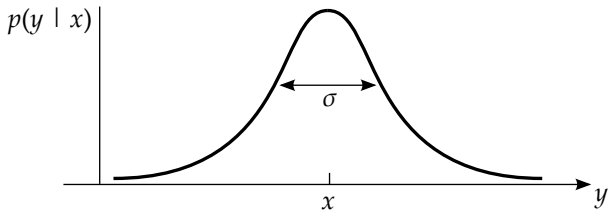
two reasons. First, it seems a reasonable model: the superposition of lots of small pieces of noise ought to (due to the central limit theorem) look roughly Gaussian. Second, it is the simplest case mathematically, and often leads to analytically tractable solutions.

The Gaussian channel takes any number as an input and corrupts it by adding random Gaussian noise. By convention, and for useful modelling reasons, the channel is usually defined in terms of complex numbers [5, Subsection 2.2.4].

Definition 1.3. The *Gaussian channel* with noise power σ^2 has alphabets $\mathcal{X} = \mathcal{Y} = \mathbb{C}$, and probability transition density $p(y \mid x)$ defined implicitly by the relationship $Y = x + Z$, where the $Z \sim \mathbb{C}\mathrm{N}(0, \sigma^2)$ are all independent.

(If the channel is used multiple times, we assume a new random Z is drawn each time.)

Transition density of the Gaussian channel
(real part only)



Another channel we will examine in this thesis is the finite field channel. This channel can be a useful model of a Gaussian channel that has been quantised (or discretised).

Definition 1.4. Let q be prime, and let Z be a random variable defined on the finite field $\mathbb{F}_q = \{0, 1, \ldots, q-1\}$. The *finite field channel* of size q with noise Z has alphabets $\mathcal{X} = \mathcal{Y} = \mathbb{F}_q$, and probability transition density $p(y \mid x)$ defined implicitly by the relationship $Y = x + Z \pmod{q}$. (Again, multiple uses assume independent Zs.)

Note that the BSC is a special case of the finite field channel where $q = 2$ and $Z = 1$ with probability p or $Z = 0$ otherwise.

Now that we have a channel, we can design codes for that channel. A code takes a set of M messages, and encodes each message into a string – called a *codeword* – of length T. After the channel has been used T times to send the codeword, there must be a rule for decoding the received string, to estimate which message was sent.

Definition 1.5. An (M, T)*-code* for the channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$ consists of:

1. a *message set* $\mathcal{M}$ of cardinality M,

2. an *encoding function* $\mathbf{x}\colon \mathcal{M} \to \mathcal{X}^T$,

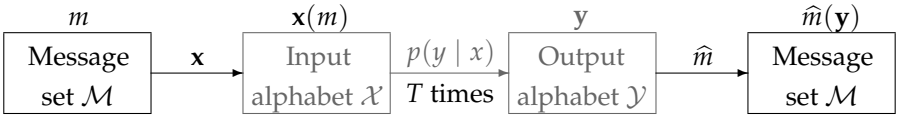
3. a *decoding function* $\widehat{m}\colon \mathcal{Y}^T \to \mathcal{M}$.

The set of codewords $\{\mathbf{x}(m) : m \in \mathcal{M}\}$ is called the *codebook*; the parameter T is called the *block length*.

The problem of channel coding works like this:

- The transmitter (Alice) requires to send a message $m \in \mathcal{M}$.
- Alice encodes this message into a codeword $\mathbf{x}(m) \in \mathcal{X}^T$.
- Alice sends the first letter $x(m)[1]$ of the codeword through the channel to the receiver (Bob). Bob receives a corrupted version $y[1] \in \mathcal{Y}$ of the letter, where the corruption has occurred at random according to $p(y \mid x)$.
- Alice sends the second letter $x(m)[2]$ of the codeword through the channel to Bob. Bob receives a corrupted version $y[2] \in \mathcal{Y}$ of the letter, where the corruption has occurred at random according to $p(y \mid x)$.
- $\vdots$
- Alice sends the final letter $x(m)[T]$ of the codeword through the channel to Bob. Bob receives a corrupted version $y[T] \in \mathcal{Y}$ of the letter, where the corruption has occurred at random according to $p(y \mid x)$.
- Bob now decodes his received word $\mathbf{y}$ to make his estimate $\widehat{m}(\mathbf{y})$ of Alice's original message m. Hopefully, $\widehat{m} = m$, and the message has been communicated successfully.

So the system as a whole looks like this (the channel is in grey):



All channels considered in this thesis will be *memoryless*, in that the channel's current performance is independent of earlier behaviour. In other words, the transmission of codewords follows a product distribution

$$p(\mathbf{y} \mid \mathbf{x}) = \prod_{t=1}^{T} p(y[t] \mid x[t]).$$

So far, we have considered only *static* channels, where the probability transition function remains fixed over time. (Later, we will look at fast-fading

channels, where the transition function is no longer fixed but changes from timeslot to timeslot.)

Note that, from a mathematical point of view, what the messages *are* is unimportant – what matters is how many of them there are. So it often makes sense to choose the message set $\mathcal{M}$ to be something convenient. For example, when dealing with a finite field channel of size q, we often take $\mathcal{M}$ to be $\mathbb{F}_q^S$, which has cardinality $M = q^S$. In particular, when $q = 2$, the message set $\mathcal{M} = \mathbb{F}_2^S = \{0,1\}^S$ is the set of all bit strings of length $S = \log_2 M$.

We define the *rate* of a code to be the number of bits that we can send per channel use. The number of bits is $\log_2 M$, as above, and the number of channel uses is T, so the rate is $(\log_2 M)/T$.

Definition 1.6. The *rate* of an (M,T)-code is defined to be $(\log_2 M)/T$ bits per transmission.

(From now on, all logarithms are to base 2, and we just write log for $\log_2$.)

Also associated with a code we have its *error probability*, the chance that a message is decoded incorrectly. (We take the average error probability across all messages, but if we were to use the maximum error probability, the main results of this chapter would be the same.)

Definition 1.7. The *average error probability* is

$$\begin{aligned} e &:= \frac{1}{M} \sum_{m \in \mathcal{M}} \mathbb{P}\big(\widehat{m}(\mathbf{Y}) \neq m \mid \mathbf{x}(m) \text{ sent}\big) \\ &= \frac{1}{M} \sum_{m \in \mathcal{M}} \sum_{\mathbf{y} \in \mathcal{Y}^T} p\big(\mathbf{y} \mid \mathbf{x}(m)\big)\, \mathbb{1}[\widehat{m}(\mathbf{y}) \neq m]. \end{aligned}$$

We can now give an example of a code for the BSC.

Suppose there are two messages we might wish to send: "No" and "Yes", so $\mathcal{M} = \{\text{No}, \text{Yes}\}$.

A very simple code could assign 0 to be the codeword for "No" and 1 to be the codeword for "Yes". This is a $(2,1)$-code. It clearly has a rate of $(\log 2)/1 = 1$ bit per transmission and error probability p.

To reduce the error probability, we will need a more sophisticated code. One method of coding would be the repetition code where each symbol is repeated T times, so

$$\mathbf{x}(\text{No}) = 00\cdots0 \in \mathcal{X}^T \qquad \mathbf{x}(\text{Yes}) = 11\cdots1 \in \mathcal{X}^T.$$

Definition 1.8. A $(2,T)$-code is called a *T-repetition* code if the codebook consists solely of the all-0 and all-1 codewords of length T.

The obvious method of decoding is the 'majority rule': if $\mathbf{y}$ has more 0s than 1s, decode as "No"; if $\mathbf{y}$ has more 1s than 0s, decode as "Yes". (If there are exactly $T/2$ of each, decoding may be performed arbitrarily.)

Note that the rate of this code is $(\log 2)/T = 1/T$, and the error probability is bounded by $e \geq p^T$, the probability that all T symbols flip.

What does it mean to be able to communicate through a channel at some desired rate r? Well, it means that there must exist a code for the channel with rate at least r and a low probability of error. How low? As low as we desire. If we want to limit the error probability to 5%, then there must be a code with rate at least r and error probability no more than 5%; but if we want the error probability to be as low as 1% or even 0.01%, there has to be a code for that too.

Definition 1.9. Consider a channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$.

A rate r is *achievable* if for any error tolerance $\epsilon > 0$, there exists a code with rate at least r and error probability lower than ϵ.

Otherwise, r is *not achievable*, in that there exists an error threshold ϵ such that there exists no code with rate at least r and error probability lower than ϵ.

The capacity is defined to be the maximum achievable rate.

Definition 1.10. Consider a channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$. Then we define the *capacity* of the channel, c, to be the supremum of all achievable rates:

$$c := \sup\{r : r \text{ is achievable}\}.$$

In other words, all rates r less than c are achievable, but no r above c is achievable.

Shannon calculated the capacity as the maximum mutual information between the input and output of a channel [4, Theorem 11].

The mutual information $\mathbb{I}(X : Y)$ between X and Y can be seen as a measure of 'how independent' X and Y are. If $\mathbb{I}(X : Y)$ is large, then X and Y are highly dependent, so knowledge of the output Y gives us lots of information about the input X; if $\mathbb{I}(X : Y)$ is small, then X and Y are highly independent, so knowledge of the output Y gives us little information about the input X.

Theorem 1.11 (Shannon's channel coding theorem)**.** *Consider a discrete channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$, that is, a channel where both $\mathcal{X}$ and $\mathcal{Y}$ are both countable sets.*

Then the capacity c is given by the formula $c = \max_X \mathbb{I}(X : Y)$, where Y is related to X through $p(y \mid x)$, and the maximum is over all input random variables X defined on $\mathcal{X}$.

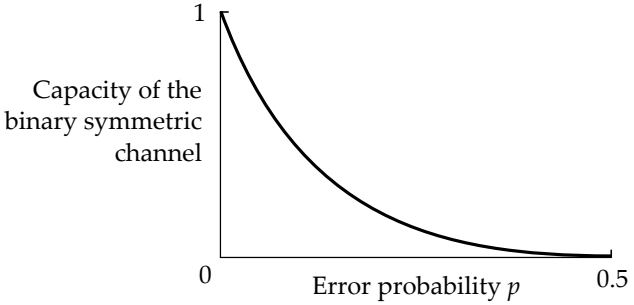
The characterisation of capacity given by Shannon's channel coding theorem (Theorem 1.11) allows us to calculate the capacity of the finite field channel and the BSC.

Theorem 1.12. *The capacity of the finite field channel of size q with noise Z is*

$$c = \log q - \mathbb{H}(Z) =: \mathbb{D}(Z).$$

The capacity of the BSC with error probability p is

$$c = 1 - \left(p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \right).$$



(We use the abbreviation $\mathbb{D}(Z) := \log q - \mathbb{H}(Z)$ since this is equal to the relative entropy distance $\mathbb{D}(p(z) \,\|\, p(u))$ between Z and a uniform random variable U over $\mathbb{F}_q$.)

Proof. Finite field channel. By Shannon's channel coding theorem (Theorem 1.11) we need to calculate the mutual information. This is

$$\begin{aligned} \mathbb{I}(X:Y) &= \mathbb{H}(Y) - \mathbb{H}(Y \mid X) && \text{(HB7)} \\ &= \mathbb{H}(Y) - \mathbb{H}(X+Z \mid X) && (Y = X+Z, \text{ Definition 1.3}) \\ &= \mathbb{H}(Y) - \mathbb{H}(Z). && \text{(HB4)} \end{aligned}$$

This is maximised by choosing X to be uniform on $\mathbb{F}_q$, by (HB9), so that Y is uniform also, giving

$$c = \max_X \mathbb{I}(X:Y) = \log q - \mathbb{H}(Z) = \mathbb{D}(Z),$$

as required.

BSC. This result follows from recalling that the BSC is a special case of the finite field model. □

Later in this thesis, we will often see examples of channels and networks whose capacities are a constant fraction of the finite field channel capacity. If a channel or network has capacity $c = d\,\mathbb{D}(Z)$ for some constant d, we say that the channel has *d degrees of freedom* (also known as the *multiplexing gain* or *pre*-log *term*).

Definition 1.13. Given a discrete channel or network with capacity c, we define the *degrees of freedom* to be $\mathtt{dof} = c/\mathbb{D}(Z)$.

Clearly, the finite field channel itself has a single degree of freedom, that is we have $\mathtt{dof} = 1$.

We have not yet talked about the proof of Shannon's channel coding theorem (Theorem 1.11).

To prove Shannon's channel coding theorem (and related theorems) we must prove two things:

Achievability First, we must show that any rate below capacity $r < c$ is achievable. That is, we must find a sequence of codes all with rates at least $r < c$, but with arbitrarily low error probabilities.

Converse Second, we must show that any rate above capacity $r > c$ is not achievable. That is, we must show that for any sequence of codes all with rates at least $r > c$, the error probabilities must be bounded away from 0.

The converse part is proved using Fano's inequality [8], which bounds the error probability in terms of the conditional entropy across the channel $\mathbb{H}(Y \mid X)$ and the size of the message set M.

Shannon's key insight into the achievable part is the following: instead of trying carefully to design special codes with high rates and low error probabilities, we can instead just pick the code at random. That is, we choose the codeword letters $X(m)[t]$ IID according to some distribution X. If we set $M = \lceil 2^{Tr} \rceil$, then the rate of the code will be at least r. We hope that by choosing T sufficiently large, the error probability will be driven arbitrarily low. (Later, we can optimise over the choice of X.)

This random encoder can be twinned with an effective decoder to show that any rate $r < c$ can be achieved. Two different decoders can be used:

Joint typicality decoder The receiver takes the channel output $\mathbf{y}$ and finds the unique codeword $\mathbf{x}(m)$ such that the pair $(\mathbf{x}(m), \mathbf{y})$ is jointly typical of (X, Y). (See the handbook, Section 1.2, for definitions.) This m is the decoding estimate. (If there is no such $\mathbf{x}(m)$ or it isn't unique, we declare an error.)

Maximum likelihood decoder The receiver takes the channel output $\mathbf{y}$ and decodes to the message most likely to have yielded it. That is, we pick $\widehat{m}$ to maximise

$$p\big(\mathbf{y} \mid \mathbf{x}(m)\big) = \prod_{t=1}^{T} p\big(y[t] \mid x(m)[t]\big)$$

(If the maximum isn't unique, we declare an error.)

Shannon himself [4, Section 13] and most subsequent authors (for example [6, Chapter 7], [9, Chapter 10]) use the joint typicality approach, as it gives a fairly simple and short proof.

On the other hand, Gallager [10] used the maximum likelihood approach to prove Shannon's channel coding theorem, by bounding the error probability by $e \leq 2^{-TE(r)}$, where $E(r)$ is called the *error exponent*, and examining the error exponent for different values of r. We shall return to Gallager's maximum likelihood approach in Chapter 6, when proving a similar theorem for group testing.

We now outline the achievability proof using the joint typicality decoder. Basic facts about typical sets are given in the handbook (Section 1.2).

Sketch proof of achievability, Theorem 1.11. As above, we set the number of messages to be $M = \lceil 2^{Tr} \rceil$, choose codeword letters $X(m)[t]$ IID at random according to some distribution X, and decode using a joint typicality decoder. There are two ways we could get an error.

First, the actual codeword $\mathbf{X}$ and the received output $\mathbf{Y}$ could fail to be jointly typical. But the theory of typical sets tell us that this event is very unlikely.

Second, another codeword $\mathbf{X}(\widehat{m})$ could be jointly typical with $\mathbf{Y}$, despite $\mathbf{X}(\widehat{m})$ and $\mathbf{Y}$ actually being independent of each other. Since $\mathbf{X}(\widehat{m})$ and $\mathbf{Y}$ are very likely to (marginally) typical, joint typicality occurs with approximate probability

$$\begin{aligned}\frac{\#\text{ jointly typical } (\mathbf{x}, \mathbf{y})}{\#\text{ typical } \mathbf{x} \times \#\text{ typical } \mathbf{y}} &\approx \frac{2^{T\mathbb{H}(X,Y)}}{2^{T\mathbb{H}(X)}2^{T\mathbb{H}(Y)}} \\ &= 2^{-T(\mathbb{H}(X)+\mathbb{H}(Y)-\mathbb{H}(X,Y))} \\ &= 2^{-T\mathbb{I}(X:Y)},\end{aligned}$$

by standard facts about typical sets (see the handbook, Section 1.2). Hence the probability of error is approximately

$$\begin{aligned}e &\leq \sum_{\widehat{m} \neq m} \Pr(\mathbf{X}(\widehat{m}) \text{ and } \mathbf{Y} \text{ jointly typical}) \\ &\approx \sum_{\widehat{m} \neq m} 2^{-T\mathbb{I}(X:Y)} \\ &= (M-1)2^{-T\mathbb{I}(X:Y)} \\ &= \left(\lceil 2^{Tr} \rceil - 1\right)2^{-T\mathbb{I}(X:Y)} \\ &\leq 2^{Tr}2^{-T\mathbb{I}(X:Y)} \\ &= 2^{-T(\mathbb{I}(X:Y)-r)}.\end{aligned}$$

So provided $r < \mathbb{I}(X:Y)$, then by choosing T large enough, the error probability can be made arbitrarily small.

Choose X to maximise $\mathbb{I}(X : Y)$ to get the result. □

In a sense, this theorem is quite a 'lucky' result: it turns out that the lower bound on capacity given by Shannon's random coding argument and the upper bound given by Fano's inequality coincide, to give us the equality $c = \max_X \mathbb{I}(X : Y)$.

For other networks, we may not be so lucky. However, similar proof strategies can be useful. If we can show that all rates below some r_* are achievable, this gives us a lower bound on capacity: $c \geq r_*$. Conversely, if we can show that no rates above some r^* are achievable, then we have an upper bound: $c \leq r^*$. In the point-to-point case, we have $r_* = r^* = c$. But even if there is a gap between the upper and lower bounds, the result can be useful in giving us an approximation to the capacity. In particular, sometimes there may be a limiting sense in which the upper and lower bounds are asymptotically equal – for example, as signal power or number of users tends to infinity.

It is often useful to think not just of individual codes, but of *families* of codes. One family of codes we have already seen is the repetition code (Definition 1.8).

MacKay [9, Section 11.4] divides families of codes into three separate categories, depending on how effective they are for their channel.

Bad codes In bad families of codes, as we force the error probability to 0, the rate of the codes approaches 0 also.

Good codes In good families of codes, as we force the error probability to 0, the rate of the codes is bounded above 0, but below capacity.

Very good codes In very good families of codes, as we force the error probability to 0, the rate of the codes can be maintained arbitrarily close to capacity.

Earlier we saw that rate of the repetition code is $1/T$, and its error probability is bounded by $e > p^T$. Hence, to force the error probability e to 0, we must send $T \to \infty$, and the rate tends to 0. Hence, the repetition code is a bad code. (MacKay notes, however, that bad codes are not necessarily practically useless [9, p. 183].)

It is sufficient for this thesis to note that Shannon's channel coding theorem (Theorem 1.11) tells us that very good (capacity achieving) codes do exist, and that we can do much better than simple codes like the repetition code. (The design of good and very good practical codes is outside the scope of this thesis.)

In later work in this thesis, rather than finding new codes from scratch, we will instead *adapt* these very good point-to-point channel codes for use in large networks.

A useful class of codes for finite field channels is the class of *linear codes*. If we take $\mathcal{M} = \mathbb{F}_q^S$ for the message set again, then a linear code is a code whose encoding function $\mathbf{x}\colon \mathbb{F}_q^S \to \mathbb{F}_q^T$ is a linear map.

It's often useful to represent this linear map by an $S \times T$ matrix G, so $\mathbf{x}(\mathbf{m}) = \mathsf{G}\mathbf{m}$. We call G the *generator matrix* of the code. The rate of such a code is $(\log M)/T = (S/T)\log q$.

Definition 1.14. Consider a finite field channel of size q. Then a *linear code* is a (q^S, T)-code with message set $\mathcal{M} = \mathbb{F}_q^S$ and encoding function $\mathbf{x}(\mathbf{m}) = \mathsf{G}\mathbf{m}$ for some *generator matrix* $\mathsf{G} \in \mathbb{F}_q^{S\times T}$. Any decoding function may be used.

We call S the *rank* of the code.

For example, the T-repetition code is a linear code with field size $q = 2$, rank $S = 1$ and $1 \times T$ generator matrix $\mathsf{G} = (1\ 1 \cdots\ 1)$.

The important fact about linear codes (at least for finite field channels) is that, when paired with an optimal decoder, very good (capacity achieving) linear codes exist [9, Chapter 14]. Thus, if we restrict our attention only to linear codes, we can still achieve all rates up to the capacity $c = \mathbb{D}(Z)$ of the finite field channel.

Theorem 1.15. *Very good linear codes exist for all finite field channels with nonzero capacity.*

The current state of the art for high-rate practical codes – that is codes with low encoding and decoding complexity and moderate block lengths – is a class of random linear codes called low-density parity-check codes [9, Chapter 47]. (See the textbook of Richardson and Urbanke [11] for more details.)

1.4 Power

We have not yet looked at codes for the Gaussian channel.

The capacity of the Gaussian channel is infinite, as there exist codes with arbitrarily high rates and simultaneously arbitrarily low error probabilities.

To see this, consider the following. Let $\mathcal{M} = \{1, 2, \ldots, M\}$, encode using $x(m) = mN$ for some very large N, and decode to the nearest positive integer to y/N (which should be roughly m). This is an $(M, 1)$-code, with rate $\log M$. By picking N large enough, the error probability can be made arbitrarily small; but by picking M large enough, the rate can be made arbitrarily high.

This is neither mathematically interesting nor physically realistic. Antennas for wireless networks are not capable of transmitting at arbitrarily high powers. Thus, we introduce a *power constraint*: that for all codewords $\mathbf{x}$, the power – the mean square value – is limited by a prescribed value P.

Definition 1.16. The power of a codeword $\mathbf{x}$ is defined to be

$$\overline{|x|^2} := \frac{1}{T}\sum_{t=1}^{T} |x[t]|^2.$$

The power of a code is defined to be the maximum power of any codeword.

So we want to limit our attention to codes whose power is at most the power constraint P.

Definition 1.17. Consider a channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$.

A rate r is *achievable with power P* if for any error tolerance $\epsilon > 0$, there exists a code of power at most P with rate at least r and error probability lower than ϵ.

Otherwise, r is *not achievable with power P*, in that there exists an error threshold ϵ such that there exists no code of power at most P with rate at least r and error probability lower than ϵ.

Definition 1.18. Consider a channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$. Then we define the *capacity of the channel with power P* to be the supremum of all achievable rates:

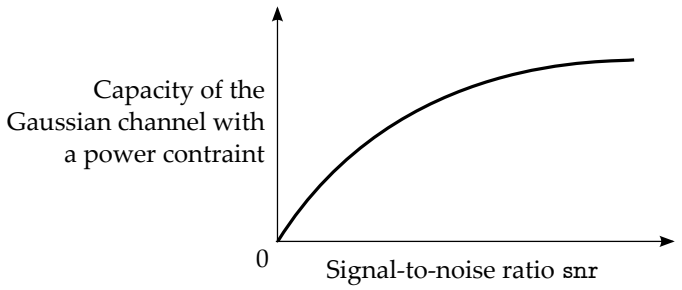
$$c := \sup\{r : r \text{ is achievable with power } P\}.$$

In other words, all rates r less than c are achievable with power P, but no r above c is achievable.

Shannon calculated the capacity of the Gaussian channel with a power constraint in his original paper [4, Theorem 17]

Theorem 1.19. *Consider the Gaussian channel with power constraint P and noise power σ^2.*

Then the capacity is given by the formula $c = \log(1 + \mathtt{snr})$, where we have defined the signal-to-noise ratio $\mathtt{snr} := P/\sigma^2$ *to be the ratio of the signal power to the noise power*



Two useful approximations for the capacity of the Gaussian channel are

$$c = \log(1 + \mathtt{snr}) \approx \begin{cases} \mathtt{snr} \log \mathrm{e} & \text{for small } \mathtt{snr}, \\ \log \mathtt{snr} & \text{for large } \mathtt{snr}. \end{cases} \tag{1.1}$$

So at low $\mathtt{snr}$, capacity grows linearly with $\mathtt{snr}$; whereas at high $\mathtt{snr}$, capacity only grows logarithmically.

Later in this thesis, we will often see examples of channels and networks whose capacities at high-$\mathtt{snr}$ are often a constant fraction of the Gaussian channel capacity $c \approx \log \mathtt{snr}$. If a channel or network has capacity $c = d \log \mathtt{snr} + o(\log \mathtt{snr})$ at $\mathtt{snr} \to \infty$ for some constant d, we say that the channel has *d degrees of freedom*. (This is the Gaussian analogy to the finite field degrees of freedom in Defintion 1.13.)

Definition 1.20. Given a channel with capacity $c(\mathtt{snr})$ and power constraint $P = \mathtt{snr}\,\sigma^2$, we define the *degrees of freedom* to be

$$\mathtt{dof} = \lim_{\mathtt{snr} \to \infty} \frac{c(\mathtt{snr})}{\log \mathtt{snr}}$$

where this limit exists.

From (1.1) the Gaussian channel itself has a single degree of freedom, that is, we have $\mathtt{dof} = 1$.

Sketch proof of Theorem 1.19. In a similar manner to Shannon's channel coding theorem (Theorem 1.11), it can be shown that the capacity of the Gaussian channel with a power constraint is $\max_{X:\mathbb{E}\,|X|^2 \le P} \mathbb{I}(X : Y)$. (This is certainly a believable result: it is the same formula as for discrete channels, with the additional constraint that the expected power satisfies the constraint.)

It remains to calculate the mutual information. This is

$$\begin{aligned} \mathbb{I}(X : Y) &= \mathbb{H}(Y) - \mathbb{H}(Y \mid X) && \text{(HB7)} \\ &= \mathbb{H}(Y) - \mathbb{H}(X + Z \mid X) && (Y = X + Z, \text{ Definition 1.3}) \\ &= \mathbb{H}(Y) - \mathbb{H}(Z) && \text{(HB5)} \\ &= \mathbb{H}(Y) - \log(\pi \mathrm{e} \sigma^2). && (Z \sim \mathbb{C}\mathrm{N}(0, \sigma^2), \text{ Definition 1.3}) \end{aligned}$$

Note that

$$\mathbb{E}\,|Y|^2 = \mathbb{E}\,|X + Z|^2 = \mathbb{E}\,|X|^2 + \mathbb{E}\,|Z|^2 \le P + \sigma^2.$$

Hence the entropy of Y is maximised by choosing X to be complex Gaussian with variance P, by (HB10), so that Y is Gaussian also, with power $P + \sigma^2$,

giving

$$\begin{aligned} c := \max_X \mathbb{I}(X:Y) &= \log\left((\pi e(P+\sigma^2)\right) - \log(\pi e \sigma^2) \\ &= \log\left(\frac{\pi e(P+\sigma^2)}{\pi e \sigma^2}\right) \\ &= \log\left(1+\frac{P}{\sigma^2}\right), \end{aligned}$$

as required. □

Note that the input distribution that achieves capacity is $X \sim \mathbb{C}\mathrm{N}(0, P)$. So the signal is statistically the same – that is, distributed in the same parametric family – as noise, but with a different power. This fact will be useful later.

As we mentioned earlier, Theorem 1.19 tells us that very good (capacity achieving) codes exist for the Gaussian channel. Later, we will adapt these very good point-to-point codes for use in large Gaussian networks.

1.5 Fading

In their standard forms, the finite field and Gaussian channels are represented by the formula $Y[t] = x[t] + Z[t]$. We can interpret this as the signal being transmitted perfectly through the channel, except for the addition of some noise. However, for a more realistic model of wireless networks, we need to account for the way the signal itself transforms as it is sent through the channel. For example, in the Gaussian channel, we might expect the signal power to decay over long distances, and standard physical models suggest that the phase of the signal $\arg x[t]$ will alter as it is transmitted through space [5, Section 2.1].

We can model these concepts by introducing a *fading* (or *channel state*) *coefficient* $H[t]$. Our channels now become $Y[t] = H[t]x[t] + Z[t]$.

We are interested in three cases:

Fixed fading where $H[t] = h$ is a fixed deterministic constant (Subsection 1.4.1);

Fast fading where the $H[t]$ are IID random (Subsection 1.4.2);

Slow fading where $H[t] = H$ is random, but fixed for all time (Subsection 1.4.3).

Before we continue, we have a useful simplification to make. Since the fading Gaussian channel

$$Y[t] = H[t]x[t] + Z[t] \qquad Z[t] \sim \mathbb{C}\mathrm{N}(0, \sigma^2) \qquad \overline{|x|^2} \leq P$$

will be used a lot in this thesis, it makes sense to change our units, so that the noise power and power constraint are both unity. To that end, set

$$\widetilde{Y}[t] := \frac{1}{\sigma} Y[t] \qquad \widetilde{H}[t] := \sqrt{\frac{P}{\sigma^2}} H[t] \qquad \widetilde{x}[t] := \frac{1}{\sqrt{P}} x[t] \qquad \widetilde{Z}[t] := \frac{1}{\sigma} Z[t].$$

Under this change of units we have, after dividing through by σ,

$$\widetilde{Y}[t] = \widetilde{H}[t]\widetilde{x}[t] + \widetilde{Z}[t] \qquad \widetilde{Z}[t] \sim \mathbb{C}\mathrm{N}(0,1) \qquad \overline{|\widetilde{x}|^2} \leq 1. \tag{1.2}$$

From now on, we will solely use this model, so will shall drop the tildes. Note that under this change of units, the signal-to-noise ratio

$$\frac{P|H|^2}{\sigma^2} = \frac{1|\widetilde{H}|^2}{1}$$

remains the same. (Note also that under this change, the Gaussian channel with no fading inherits a fixed fading coefficient $H[t] = h = \sqrt{P/\sigma^2}$.)

1.5.1 Fixed fading

Fixed fading models fading that is constant and predictable, such as the decay in signal power between a non-moving transmitter and a non-moving receiver within a fixed environment.

We model the fading coefficient as a deterministic constant fixed for all time, $H[t] = h$ for all t, giving $Y[t] = hx[t] + Z[t]$.

How does the capacity alter now?

For the finite field channel we have $h \in \mathbb{F}_q$. Note that for nonzero h, the function $x \mapsto hx$ is a bijection, so the channel is equivalent to that without fading ($h = 1$), and still has capacity $\mathbb{D}(Z)$, from Theorem 1.12. On the other hand, if $h = 0$, then Y is always 0, and all signals are indistinguishable, so the capacity is 0. Hence, for the finite field channel, the capacity is

$$c = c(h) = \begin{cases} 0 & \text{if } h = 0 \\ \mathbb{D}(Z) & \text{otherwise.} \end{cases}$$

(For simplicity, we often just assume that h is nonzero, so the capacity is unchanged as $c = \mathbb{D}(Z)$.)

For the Gaussian channel with power constraint $P = 1$ we have $h \in \mathbb{C}$. The power constraint is now $|hx|^2 = |h|^2|x|^2 \leq |h|^2$. So this channel is equivalent to one with no fading, but with the power constraint changed from 1 to $|h|^2$. The capacity is thus

$$c = c(h) = \log(1 + |h|^2) = \log(1 + \mathtt{snr}),$$

with the new convention that $\mathtt{snr}$ denotes the signal-to-noise ratio at the receiver: $\mathtt{snr} = |h|^2 P/\sigma^2 = |h|^2$.

To summarize:

Theorem 1.21. *For $h \neq 0$, the capacity of the finite field channel with fixed fading is $c = \mathbb{D}(Z)$.*

The capacity of the Gaussian channel with fixed fading is $c = \log(1 + \mathtt{snr})$, where $\mathtt{snr} = |h|^2$.

For modelling wireless networks, we will often use a Gaussian channel with fixed fading coefficient h that decays like a power law over distance. That is, we have $h = k\rho^{-\alpha/2}$, where $\rho > 0$ is the distance between a transmitter and receiver and $k > 0$ is a constant.

The parameter $\alpha > 0$ – called the *attenuation* – represents how resistive the environment is to the transmission of radio waves. (Some authors call $\alpha/2$ the attenuation; we do not.) Low α represents an environment with few obstacles to signals; high α implies that a lot of the signal power is absorbed before reaching the receiver. In free space, standard physical considerations imply that $\alpha = 2$ or 3; for built-up areas, values of α of roughly 4 or 5 seem more appropriate [5, Section 2.1]. The capacity of such a channel is $c = \log(1 + k^2\rho^{-\alpha})$, by Theorem 1.21.

(Power-law attenuation fails to be realistic for small distances $\rho \ll 1$. Here the received power would be greater than the transmitted power, which violates the conservation of energy. Some authors therefore prefer alternative models such as $h = \min\{1, k\rho^{-\alpha/2}\}$ or $h = k(\rho + \rho_0)^{-\alpha/2}$ for some fixed constant ρ_0.)

We could also include a fixed phase change in this model by setting $h = k\rho^{-\alpha/2}e^{i\theta}$. In free space, the phase would scale linearly with distance, so $h = k\rho^{-\alpha/2}e^{2\pi i\rho/\lambda}$, where λ is the carrier wavelength [12]. Note that we still have $|h|^2 = k^2\rho^{-\alpha}$, so the capacity is the same.

1.5.2 Fast fading

Fast fading models a situation where the state of the channel is changing rapidly, such as a commuter using a mobile phone on a train. We model this as $H[t]$ being random according to some distribution H but renewing at each channel use; that is, the $H[t]$ are independent and identically distributed like H.

When we deal with fast fading, the performance of a channel will depend on whether the transmitter and receiver know the current value of $H[t]$, or just the general distribution H, and whether the transmitter can use this knowledge to vary their power.

Throughout this thesis, we assume that both the transmitter and the receiver know $H[t]$. This is known as having *perfect channel state information at the transmitter* (CSIT) and *at the receiver* (CSIR). We presume that this knowledge is *causal*, that is, the receiver and transmitter learn $H[t]$ immediately prior

to the transmission of $x[t]$ and reception of $y[t]$ respectively. In other words, they have no prediction of the channel future to use (except, of course, knowing the future channel states will be IID according to H).

When the transmitter has CSIT in a Gaussian channel, we must specify whether or not she can use this information to operate at varying power. So there are two different types of power constraint. Let $x[t](h)$ be the tth codeword letter chosen in channel state h, and let $\mathcal{H}$ be the support of H. (Recall from our simplification (1.2) that we now have $P = 1$.)

Universal A universal power constraint demands that the power constraint is held universally over each channel state realisation. That is, we demand

$$\overline{|x[t](h)|^2} = \frac{1}{T}\sum_{t=1}^{T} |x[t](h)|^2 \leq 1 \qquad \text{for all } h \in \mathcal{H}.$$

Average An average power demands that the power constraint is held when averaged over all channel state realisations. That is, we demand

$$\mathbb{E}_H \overline{|x[t](H)|^2} = \frac{1}{T}\sum_{h\in\mathcal{H}} \mathbb{P}(H = h) \sum_{t=1}^{T} |x[t](H)|^2 \leq 1.$$

(The second term assumes H is discrete; the summation can be replaced by an integral if H is continuous.)

An average power constraint allows a transmitter to use extra power when the channel is at its strongest, and save power when the channel is weak. (For more details, see the textbook of Tse and Viswanath [5, Subsection 5.3.3].)

In this thesis, we always assume a universal power constraint. First, it is mathematically simpler to deal with. Second, it is physically unrealistic for transmitters to operate above their average power for long periods of time.

(Similarly, in a frequency-selective channel, one must specify whether the the power constraint is enforced in each individual subchannel, or an average across all frequencies.)

So, assuming perfect CSIT and CSIR (with a universal power constraint in the Gaussian case) we have the following.

Theorem 1.22. *Consider a fast fading channel, and let $c(h)$ be the capacity of the channel under fixed fading parameter h.*

Then the fast fading capacity is equal to the average fixed fading capacity, in that $c = \mathbb{E}c(H)$.

In the most general case, this theorem is due to Goldsmith and Varaiya [13]. We sketch the achievability proof for the case when H is discrete. (Goldsmith and Varaiya attribute the result for this simpler case to Wolfowitz [14, Theorem 4.6.1].)

Sketch proof. Assume H is discrete, in that it can only take values in some countable set $\mathcal{H}$. Then if we 'collect together' all occasions when $H[t]$ has some particular value h, we can treat that collection of channel uses as being through a fixed fading channel with deterministic fading coefficient h. So at these times, we can achieve rates up to $c(h)$.

Writing $\pi(h, T)$ for the proportion of time periods when $H[t] = h$,

$$\pi(h, T) := \frac{1}{T}\big|\{t \in \{1, 2, \dots, T\} : H[t] = h\}\big| \qquad h \in \mathcal{H},$$

we can achieve the rate

$$r = \lim_{T\to\infty} \sum_{h\in\mathcal{H}} \pi(h, T)c(h).$$

But by the strong law of large numbers, we have the ergodicity property that $\lim_{T\to\infty} \pi(h, T) = \mathbb{P}(H = h)$ (almost surely). Hence (again almost surely),

$$c \geq \sum_{h\in\mathcal{H}} \mathbb{P}(H = h)c(h) = \mathbb{E}c(H).$$

The converse can be proved using Fano's inequality, as with Shannon's channel coding theorem (Theorem 1.11).

The result for continuous H can be derived from this using a quantisation argument [13, Appendix]. □

Since this result relies on the sequence of fading parameters being ergodic, c is sometimes called the *ergodic capacity*. So the above result can be interpreted as 'the ergodic capacity is the average capacity.' (Later, we will see how using interference alignment in networks can allow us to achieve an ergodic capacity that is higher than the average capacity.)

Applying Theorem 1.22 to the finite field channel (Theorem 1.12), we get

$$c = \sum_{h\in\mathcal{H}} \mathbb{P}(H = h)c(h) = \big(1 - \mathbb{P}(H = 0)\big)\mathbb{D}(Z).$$

(Again, we often assume H is never 0, so $c = \mathbb{D}(Z)$ still.)

For the Gaussian channel (Theorem 1.19), we have

$$c = \mathbb{E}c(H) = \mathbb{E}\log(1 + |H|^2) = \mathbb{E}\log(1 + \mathtt{SNR}).$$

(Since $\mathtt{SNR}$ is random here, we capitalise it.)

One type of fast fading for the Gaussian channel could be a rapidly changing phase, $H[t] = ke^{i\Theta[t]}$, where $\Theta[t] \sim \mathrm{U}[0, 2\pi)$ IID over t. This is a good model of wireless communication when there are many paths a signal could take from transmitter to receiver (in a built-up area, for example) [5, Subsection 2.4.2]. Note that here the signal-to-noise ratio is in constant, so the capacity is unchanged.

Another model of wireless communication is *Rayleigh fading* [5, Subsection 2.4.2], where $H[t] \sim \mathbb{C}\mathrm{N}(0, \tau^2)$ for some $\tau > 0$. In this case, $|H[t]|^2$ is exponentially distributed with mean τ^2 [5, (2.53)].

1.5.3 Slow fading

Slow fading models the situation where the state of a channel is varying, but is doing so very slowly, or where the channel state can only be modelled as random, but remains fixed. Here we take $H[t] = H$ as initially random, but remaining fixed for all times $t = 1, 2, \ldots, T$.

Since the channel state is random, so is the capacity $C = c(H)$: if the fading is particularly deep, $H \approx 0$, then the capacity is likely to be very low; if the fading is lighter, then the capacity will be higher. Specifically, under the event that $H = h$ we have $C = c(h)$.

(As with H, when the capacity is a random variable, we capitalise it as C.)

One way to summarise the random variable C would be through its cumulative distribution function $p_{\text{out}}(r) := \mathbb{P}(C \leq r) = \mathbb{P}(c(H) \leq r)$, known as the *outage probability*. We can interpret this as the following: if we are trying to communicate at some fixed rate r, then then $p_{\text{out}}(r)$ is the probability that we are unable to do so – we say the channel is in *outage*.

Definition 1.23. For a slow fading channel with (random) capacity C, the *outage probability* $p_{\text{out}} \colon \mathbb{R}_+ \to [0, 1]$ of the channel is defined by $p_{\text{out}}(r) := \mathbb{P}(C \leq r)$.

The event $\{C \leq r\}$ is called *outage*.

For the Gaussian channel, we have (following Theorem 1.19 and recalling that log denotes $\log_2$)

$$p_{\text{out}}(r) = \mathbb{P}(C \leq r) = \mathbb{P}\big(\log(1 + |H|^2) \leq r\big) = \mathbb{P}(\mathtt{SNR} \leq 2^r - 1),$$

where $\mathtt{SNR} = |H|^2$ is the signal-to-noise ratio.

For the finite field channel, we have (following Theorem 1.12)

$$p_{\text{out}}(r) = \begin{cases} \mathbb{P}(H = 0) & \text{if } 0 \leq r \leq \mathbb{D}(Z) \\ 0 & \text{if } r > \mathbb{D}(Z). \end{cases}$$

In wireless networks, a good model is to position nodes at random and use distance-based attenuation fading. Since distances between nodes are random, the fading is random too. But once the nodes are positioned, the distances remain fixed. Hence, this gives a form of slow fading.

Notes

The section consists of a review of the existing literature; the mathematical content is not claimed to be new.

The basic concepts of information theory as outlined in this chapter are all due to Shannon's original paper [4]. An exception is the concept of relative en-

tropy distance, due to Kullback and Leibler [15]; and the Hu correspondence, due to Hu [16].

The presentation here closely follows the textbook of Cover and Thomas [6, Chapters 2, 7–9, 15]. The textbooks of MacKay [9, Part II] and Tse and Viswanath [5, Chapters 2, 5, 6] were also useful.

Although Shannon [4, Theorem 11] first came up with the channel coding theorem (Theorem 1.11), he provided only a sketch proof; the sketch proof provided here is along the lines of the rigorous proof by Cover [17]. The maximum likelihood approach is due to Gallager [10].

Fading was first studied by Shannon [18]. Our treatment of fading follows closely that of Tse and Viswanath [5, Sections 2.1, 5.4]. The review paper of Biglieri, Proakis, and Shamai (Shitz) [19], and a paper by Caire and Shamai (Shitz) [20] were useful.

2

Interference

In this chapter, we will look at ways of dealing with interference in communications networks.

To start with, we will define information theoretic networks, in a similar manner to our definition of channels in Chapter 1.

We will then look at methods of combating interference – the unwanted signals from other transmitters that a receiver is not interested in.

For the purpose of definiteness, we will consider these in the context of the interference networks and (mostly) the fading Gaussian case. However, the techniques are useful in wider contexts.

We look at some simple schemes – interference as noise, decode and subtract, and resource division – and then look at a family of new schemes known as interference alignment. We pay particular attention to a scheme called ergodic interference alignment, which we will use later in Chapters 4 and 5.

2.1 Wired and wireless networks

In this chapter, we will outline the theory of networks. We will concentrate on accurate models of real-world wireless networks.

Wireless communications are becoming increasingly ubiquitous. From older technologies like radios, to cutting-edge innovations such as WiFi, Bluetooth and ZigBee, the convenience of the untethered nature of wireless is popular on both large and small scales for businesses and consumers alike.

Compared to a wired (or wireline) network, wireless networks provide much greater challenge to engineers and technicians. The main problems are:

Broadcast Each receiver can send only one signal, regardless of how many messages they are trying to send to how many people.

Interference Receivers receive not just the signal corresponding to messages intended for them, but also all of the other transmitted signals as well. These signals are called *interference*.

Superposition Receivers cannot tell which signal corresponds to which message, but rather receive the superposition (that is, the sum) of all such signals.

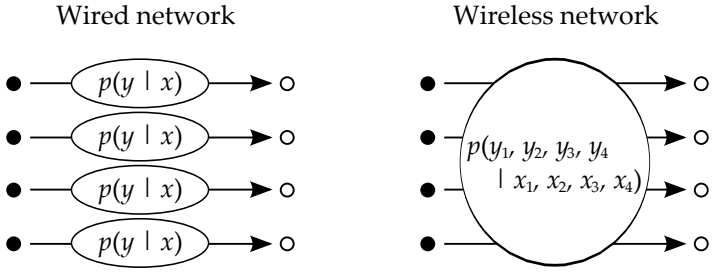
	Wired networks	**Wireless networks**
Transmission	Different signal transmitted down each wire	Same signal broadcast to all receivers
Channel	Interference-free, independent noise along each wire	Interference from other transmitters and background noise
Reception	Different signal received from each wire	Superposition of all signals received
Central difficulty	Scheduling and routing messages around the network	Dealing with interference

In this thesis, we will mainly be looking at networks where each transmitter wishes to send a single message to a single receiver, and each receiver requires a single message from a single transmitter. Thus, the broadcast and superposition problems are less important that that of interference.

We capture this problem mathematically by modelling the network by a probability transition function

$$p(y_1, \ldots, y_n \mid x_1, \ldots x_n)$$

relating all transmitted signals to all received signals.



Later in this chapter, we consider a number of methods for dealing with such interference.

2.2 Networks

Point-to-point links, as we discussed in the previous chapter, are fairly well understood. Networks, however, are much trickier.

By a *network*, we mean a number of transmitters and receivers, all trying to send and receive messages through the same medium.

Definition 2.1. A communications *network* consists of

1. a set $\mathcal{T}$ of transmitters, each with an input alphabet $\mathcal{X}_i$;
2. a set $\mathcal{R}$ of receivers, each with an output alphabet $\mathcal{Y}_j$;
3. a *probability transition function* $p\big((y_j : j \in \mathcal{R}) \mid (x_i : i \in \mathcal{T})\big)$ relating them.

(In general, an agent is allowed to be a *duplex* operator, that is to be both a transmitter and a receiver, which can act as a relay in a network. However, duplex operation will not be used in this thesis, so our definition precludes this.)

Again, we will be interested in the Gaussian and finite field channels with fading. Because there are now many transmitters and receivers, we will let $H_{ji}[t]$ denote the fading coefficient at receiver j from transmitter i.

The Gaussian and finite-field networks work in much the same way as the point-to-point channels, with the change that now receivers experience the superposition (that is, sum) of all signals sent.

Definition 2.2. *Gaussian networks* have $\mathcal{X}_i = \mathcal{Y}_j = \mathbb{C}$ for all i and j. The probability transition measure is implied by the relationship

$$Y_j[t] = \sum_{i \in \mathcal{T}} H_{ji}[t] x_i[t] + Z_j[t] \qquad j \in \mathcal{R},$$

where $Z_j[t] \sim \mathrm{CN}(0, 1)$ independently across j and t.

The *finite field network* of size q with noise Z has $\mathcal{X}_i = \mathcal{Y}_j = \mathbb{F}_q$ for all i and j. The probability transition measure is implied by the relationship

$$Y_j[t] = \sum_{i \in \mathcal{T}} H_{ji}[t] x_i[t] + Z_j[t] \pmod q \qquad j \in \mathcal{R},$$

where $Z_j[t]$ are independently and identically distributed like Z.

When it's convenient, we will write these networks in matrix form, that is

$$\mathbf{Y}[t] = \mathsf{H}[t]\mathbf{x}[t] + \mathbf{Z}[t],$$

where $\mathbf{Y}[t] = \big(Y_j[t] : j \in \mathcal{R}\big)$ is the received vector, $\mathbf{x}[t] = \big(x_i[t] : i \in \mathcal{T}\big)$ is the transmitted vector, $\mathbf{Z}[t] = \big(Z_j[t] : j \in \mathcal{R}\big)$ is the noise vector, and $\mathsf{H}[t] = \big(H_{ji}[t] : i \in \mathcal{T}, j \in \mathcal{R}\big)$ is the channel-state matrix.

To design a code for a network, we need to specify which transmitters are trying to send a message to which receivers; then each transmitter needs an encoding function, and each receiver a decoding function (or more than one, if they are receiving many messages).

Definition 2.3. A *code* for the network

$$\Big(\mathcal{T}, \mathcal{R}, (\mathcal{X}_i : i \in \mathcal{T}), (\mathcal{Y}_j : j \in \mathcal{R}), p\big((y_j : j \in \mathcal{R}) \mid (x_i : i \in \mathcal{T})\big)\Big)$$

consists of

1. a set $\mathcal{L} \subseteq \mathcal{T} \times \mathcal{R}$ of L direct links (we call the other links in $(\mathcal{T} \times \mathcal{R}) \setminus \mathcal{L}$ the *crosslinks*);
2. a *message set* $\mathcal{M}_{ij}$ of cardinality M_{ij} for each link $i \to j \in \mathcal{L}$;
3. an *encoding function* $\mathbf{x}_i \colon \prod_{j:i\to j\in\mathcal{L}} \mathcal{M}_{ij} \to \mathcal{X}_i^T$ for each transmitter $i \in \mathcal{T}$;
4. a *decoding function* $\widehat{m}_{ij} \colon \mathcal{Y}_j^T \to \mathcal{M}_{ij}$ for each link $i \to j \in \mathcal{L}$.

On link $i \to j \in \mathcal{L}$, the rate is $r_{ij} := \log M_{ij}/T$, the rate vector is $\mathbf{r} = (r_{ij} : i \to j \in \mathcal{L})$, and the sum-rate is $r_\Sigma := \sum_{i\to j\in\mathcal{L}} r_{ij}$. The error probability on link $i \to j$ is

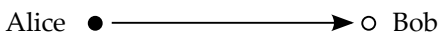
$$e_{ij} := \frac{1}{M_{ij}} \sum_{m \in \mathcal{M}_{ij}} \mathbb{P}\left(\widehat{m}_{ij}(\mathbf{y}_j) \neq m \mid \mathbf{x}_i(m) \text{ sent}\right).$$

When dealing with the Gaussian case, the power of transmitter i is the maximum value of $\overline{|x_i|^2} := 1/T \sum_{t=1}^{T} |x_i[t]|^2$ over all i's codewords $\mathbf{x}_i$.

Some common examples of networks are the following:

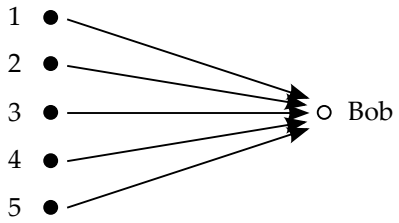
Definition 2.4. The *point-to-point link* is just a special case of a network with

$$\mathcal{T} = \{\text{Alice}\}, \qquad \mathcal{R} = \{\text{Bob}\}, \qquad \mathcal{L} = \mathcal{T} \times \mathcal{R} = \{\text{Alice}\to\text{Bob}\}.$$



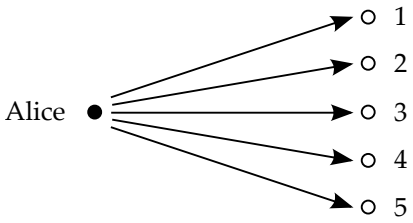
The *multiple-access network* has multiple transmitters sending to one receiver, so

$$\mathcal{T} = \{1, \ldots, n\}, \qquad \mathcal{R} = \{\text{Bob}\}, \qquad \mathcal{L} = \mathcal{T} \times \mathcal{R} = \{1\to\text{Bob}, \ldots, n\to\text{Bob}\}.$$



The *broadcast network* has one transmitter sending to multiple receivers, so

$$\mathcal{T} = \{\text{Alice}\}, \quad \mathcal{R} = \{1, \ldots, n\}, \quad \mathcal{L} = \mathcal{T} \times \mathcal{R} = \{\text{Alice}\to 1, \ldots, \text{Alice}\to n\}.$$



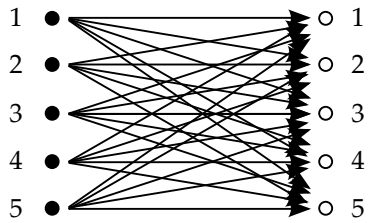
The *interference network* consists of multiple point-to-point links communicating over the same medium, so

$$\mathcal{T} = \{1, \ldots, n\}, \qquad \mathcal{R} = \{1, \ldots, n\}, \qquad \mathcal{L} = \{1{\to}1, 2{\to}2 \ldots, n{\to}n\}.$$

(Note that this differs from n independent point-to-point links, since each receiver is also receiving the signals from the other transmitters, even though they have no use for that signal.)

The *X network* consists of an equal number of transmitters and receivers communicating across all possible links, so

$$\mathcal{T} = \{1, \ldots, n\},\ \mathcal{R} = \{1, \ldots, n\},\ \mathcal{L} = \mathcal{T} \times \mathcal{R} = \{1{\to}1, 1{\to}2, 1{\to}3 \ldots, n{\to}n\}.$$



As with point-to-point links, we are interested in the maximum rate at which we can send information through a network. However, since we now have several competing links in the same network, no one benchmark will describe this. For example, achieving a high rate on one particular link may use up a lot of channel resources, leading to slower communication on another link. (Consider trying to hold a conversation in a room where lots of other people are shouting.)

Instead the set of achievable rate vectors will be a region of L-space; we call this the *capacity region*. (Recall from Definition 2.3 that $L = |\mathcal{L}|$ is the number of links in the network.)

Definition 2.5. Consider a network $\big(\mathcal{T}, \mathcal{R}, (\mathcal{X}_i), (\mathcal{Y}_j), p((y_j) \mid (x_i))\big)$.

A rate vector $\mathbf{r} = (r_{ij} : i \to j \in \mathcal{L})$ is *achievable* for links $\mathcal{L}$ if for any error tolerance $\epsilon > 0$, there exists a code for $\mathcal{L}$ with rate on each link $i \to j$ at least r_{ij} and all error probabilities lower than ϵ.

Otherwise, $\mathbf{r}$ is *not achievable*, in that there exists an error threshold ϵ such that there exists no code for $\mathcal{L}$ with rates at least r_{ij} and all error probabilities lower than ϵ.

Definition 2.6. Consider a network $\big(\mathcal{T}, \mathcal{R}, (\mathcal{X}_i), (\mathcal{Y}_j), p((y_j) \mid (x_i))\big)$. Then we define the *capacity region* of the channel, $\mathcal{C}$, to be the closure of the set of all achievable vectors:

$$\mathcal{C} := \overline{\{\mathbf{r} \in \mathbb{R}_+^L : \mathbf{r} \text{ is achievable}\}}.$$

In other words, all rate vectors $\mathbf{r}$ in the interior of $\mathcal{C}$ are achievable, but no $\mathbf{r}$ outside c is achievable.

Note that the capacity region will always be convex: Suppose the rate vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are both achievable. Then the rate vector $\lambda\mathbf{r}_1 + (1-\lambda)\mathbf{r}_2$, for $\lambda \in [0,1]$ is achievable by operating at $\mathbf{r}_1$ for λT of the time points and at $\mathbf{r}_2$ for the remaining $(1-\lambda)T$ timeslots. This strategy is known as *time sharing*; we discuss this further in Subsection 2.4.1.

Definition 2.7. We define the *sum-capacity* c_Σ to be the maximum achievable sum-rate, so

$$c_\Sigma := \max_{\mathbf{r} \in \mathcal{C}} r_\Sigma = \sup\{r_\Sigma : \mathbf{r} \text{ is achievable}\}.$$

The current knowledge of capacity regions for these networks in the Gaussian and general cases is summarised in the table below.

Network	General case	Gaussian case
Point-to-point	known (Theorem 1.11)	known (Theorem 1.19)
Multiple-access	known [21, 22]	known (Theorem 2.8)
Broadcast	unknown; known for some special cases [23]	known [24]
Interference	unknown; known for some special cases [25]	unknown; known for some special cases [25]; sum-capacity known for most two-user cases [26]
X	unknown	unknown

Later, we will use the capacity region of the multiple-access network. It was discovered independently by Ahlswede [21] and Liao [22] in the 1970s. In the Gaussian case, it simplifies to the following:

Theorem 2.8. *The capacity region of the multiple-access network of n transmitters with fixed fading is the set of* $(r_1, r_2, \ldots, r_n) \in \mathbb{R}_+^n$ *satisfying*

$$\sum_{i \in \mathcal{S}} r_i \leq \log\left(1 + \sum_{i \in \mathcal{S}} \mathtt{snr}_i\right)$$

for all $\mathcal{S} \subseteq \{1, 2, \ldots, n\}$, *where* $\mathtt{snr}_i := |h_i|^2$ *is the signal-to-noise ratio from transmitter i.*

The sum-capacity is

$$c_\Sigma = \log\left(1 + \sum_{i=1}^{n} \mathtt{snr}_i\right).$$

In this thesis, we will mostly be interested in the n-user interference network. We will use the word 'user' to denote a matching transmitter–receiver pair. Hence, an n-user network consists of n transmitters and n receivers. We will mostly be interested in the large n limit.

Recent work by Jafar [27] in the fixed $\mathtt{snr}$, $n \to \infty$ regime has shown much promise. We review Jafar's work in detail later in this chapter and in Chapter 4, and extend it to physical models of wireless networks.

Alternatively, in the fixed n, $\mathtt{snr} \to \infty$ regime, Cadambe and Jafar [28] used interference alignment to deduce the limiting behaviour within $o(\log(\mathtt{snr}))$. These techniques were extended by the same authors [29] to more general models in the presence of feedback and other effects.

For small n, the classical bounds due to Han and Kobayashi [30] as refined by Chong, Montani, Garg and El Gamel [31] for the two-user Gaussian interference network have recently been extended. For example Etkin, Tse, and Wang [32] have produced a characterization of capacity accurate to within one bit. These results were extended by Bresler, Parekh, and Tse [33], using insights based on a deterministic channel which approximates the Gaussian channel with sufficient accuracy, to prove results for many-to-one and one-to-many Gaussian interference channels.

A different approach towards finding the capacity of large communications networks is given by the deterministic approach of Avestimehr, Diggavi, and Tse [34]. They show how capacities can be calculated up to a gap determined by the number of users n, across all values of $\mathtt{snr}$.

More generally, in problems concerning networks with a large number of nodes, the work of Gupta and Kumar [35] uses techniques based on Voronoi tesselations to establish scaling laws. (See also the survey paper of Xue and Kumar [36] for a review of the information theoretical techniques that can be applied to this problem.)

Özgür, Lévêque, and Tse [37] and Özgür and Lévêque [38] use a similar model of dense random network placements, though using the same points as both transmitters and receivers. They describe a hierarchical scheme, where

nodes are successively assembled into groups of increasing size, each group collectively acting as a multiple antenna transmitter or receiver, and restrict to transmissions at a common rate. They show [37, Theorems 3.1, 3.2] that for any $\epsilon > 0$ there exists a constant $k = k(\epsilon)$ depending on ϵ and a fixed constant K such that

$$kn^{1-\epsilon} \leq c_\Sigma \leq Kn \log n. \tag{2.1}$$

These bounds are close to stating that c_Σ grows like n, but without the explicit constant that Jafar [27] and the work in Chapter 4 of this thesis achieve. (Later, we produce a version of the upper bound (2.1) without the logarithmic factor and being explicit about the constant K. Note that this result is proved under a model that differs from that of Özgür, Lévêque, and Tse [37], and the fact that we have a total of $2n$ nodes rather than n – although this is unimportant for asymptotic results. Further, in their work, local collaboration and relaying are both allowed, meaning that the true rate in their scenario could indeed be $n \log n$.)

2.3 Interference as noise

Recall that we mentioned in Section 1.4 that the optimal input distribution to the Gaussian channel is $\mathbb{C}\mathrm{N}(0, P)$, while the noise is $\mathbb{C}\mathrm{N}(0, \sigma^2)$. Thus, interference has the same distribution as noise (after an appropriate scaling). So to receiver j, the received signal

$$Y_j = \sum_{i=1}^{n} h_{ji} x_i[t] + Z_j[t] \qquad Z_j[t] \sim \mathbb{C}\mathrm{N}(0, 1)$$

is statistically indistinguishable from

$$\widetilde{Y}_j = h_{jj} x_j[t] + \widetilde{Z}_j[t] \qquad Z_j[t] \sim \mathbb{C}\mathrm{N}\left(0, 1 + \sum_{i \neq j} |h_{ji}|^2\right).$$

In other words, treating interference as noise allows user j to communicate at rate

$$r_j = \log(1 + \mathtt{sinr}_j) = \log\left(1 + \frac{|h_{jj}|^2}{1 + \sum_{i \neq j} |h_{ji}|^2}\right).$$

Here, $|h_{jj}|^2$ is the received power of the signal, and $\sum_{i \neq j} |h_{ji}|^2$ the received power of the interfering signals from other transmitters. We call

$$\mathtt{sinr}_j := \frac{|h_{jj}|^2}{1 + \sum_{i \neq j} |h_{ji}|^2}$$

the *signal-to-interference-plus-noise ratio* at receiver j.

Theorem 2.9. *Consider an n-user Gaussian interference network. The rates* $r_j = \log(1 + \mathtt{sinr}_j)$ *are simultaneously achievable.*

The convex hull of the set of $\mathbf{r}$ *with* $r_j \leq \log(1 + \mathtt{sinr}_j)$ *for all* j *is an inner bound for the capacity region.*

When the interference is weak, that is when $\sum_{i \neq j} |h_{ji}|^2 \ll 1$, this strategy will be quite effective, as we will have

$$\mathtt{sinr}_j = \frac{|h_{jj}|^2}{1 + \sum_{i \neq j} |h_{ji}|^2} \approx |h_{jj}|^2 = \mathtt{snr}_j.$$

Hence, in this situation, user j can communicate at almost the same rate it could were there no interference at all.

However, if the interference is strong, this strategy will lead to a dramatic decrease in the rate. In this case we will need different strategies.

2.4 Decode and subtract

The tactic of treating interference as noise suggests a method of dealing with strong interference.

Suppose we have just one interfering link 2→1 that is very strong. Then we can treat the interference as signal, and treat the signal itself as noise. This allows us to decode the interfering signal x_2 (with x_1 as noise). Once we have decoded x_2, we know the interference $h_{12}x_2$ (since we have perfect channel state information), allowing us to subtract it. This forms the interference free signal

$$\widetilde{Y}_1 := Y_1 - h_{12}x_2 = (h_{11}x_1 + h_{12}x_2 + Z_1) - h_{12}x_2 = h_{11}x_1 + Z_1.$$

In this case, receiver 1 requires

$$r_1 \leq \log\left(1 + \frac{|h_{12}|^2}{|h_{11}|^2 + 1}\right)$$

to decode the interference, and then $r_1 \leq \log(1 + \mathtt{snr}_1)$ to decode the signal, once the interference has been subtracted.

If the 'interference-to-noise-plus-signal ratio' $|h_{12}|^2/(|h_{11}|^2 + 1)$ is large – that is, if the interference $|h_{12}|^2$ is strong compared to the signal $|h_{11}|^2$– then this will be almost as effective as if no noise were present.

2.5 The problem of mid-level interference

So far, we have two principles:

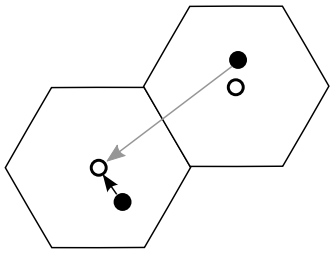
Weak interference should be treated as noise.

Strong interference should first be decoded, and then subtracted.

This leads to a natural question: what about mid-level interference? That is, what is the best way of dealing with interference when the power of the interference is roughly equal to the power of the signal?

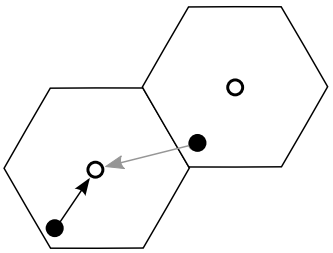
Indeed, there are many plausible real-life situations where the mid-level interference would seem to be the most likely.

For example, in cellular networks (such as mobile phone networks), we have a phenomenon known as the *edge-of-cell effect*. This describes the phenomenon that near the edge of a cell, the strength of a signal is of a very similar level to if it just fell out of the cell. Combating the mid-level interference of these edge-of-cell transmitters is essential to maintaining a high-quality system.



For users near the centre of cells, the signal is much stronger than the interference.

For users near the edge of cells, the signal and interference are of similar strengths.



In fact, dealing with this mid-level interference turns out to be particularly important in multi-user networks; while weak and strong interference are easily dealt with, even rare occurrences of mid-level interference can severely restrict the performance of such a network. The following example, due to Jafar [27], illustrates this.

Consider a two-user Gaussian interference network, as governed by the input–output equations

$$\begin{aligned} Y_1[t] &= h_{11}x_1[t] + h_{12}x_2[t] + Z_1[t] \\ Y_2[t] &= h_{21}x_1[t] + h_{22}x_2[t] + Z_2[t]. \end{aligned}$$

We will use a model with a fast-fading phase. So

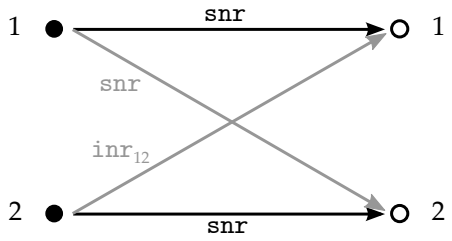
$$\begin{aligned} h_{11}[t] &= \sqrt{\texttt{snr}_1}\exp(\mathrm{i}\Theta_{11}[t]) & h_{12}[t] &= \sqrt{\texttt{inr}_{12}}\exp(\mathrm{i}\Theta_{12}[t]) \\ h_{21}[t] &= \sqrt{\texttt{inr}_{21}}\exp(\mathrm{i}\Theta_{21}[t]) & h_{22}[t] &= \sqrt{\texttt{snr}_2}\exp(\mathrm{i}\Theta_{22}[t]), \end{aligned}$$

where the $\Theta_{ji}[t]$ are IID uniform on $[0, 2\pi)$.

Here $\mathtt{inr}_{ji} = |h_{ji}|^2$ is the *interference-to-noise* ratio at receiver j from transmitter i.

For simplicity, we shall fix the direct links to be of equal strength: $\mathtt{snr}_1 = \mathtt{snr}_2 =: \mathtt{snr}$.

Now suppose just one of the two interfering crosslinks is precisely this mid-level interference: so $\mathtt{inr}_{21} = \mathtt{snr}$ too.



Jafar [27, Lemma 1] then showed the following surprising result:

Theorem 2.10. *The sum-capacity of the above network is* $c_\Sigma = \log(1 + 2\mathtt{snr})$, *regardless of the value of* $\mathtt{inr}_{12}$.

"Regardless of the value of $\mathtt{inr}_{12}$*"!* This is worth emphasising: just one crosslink of this mid-level interference has completely determined the sum-capacity of the whole network.

In Chapter 4, we shall see similar phenomena for networks with many more users, where the study of bottleneck links will be vitally important.

Proof. Direct part. Achievability follows from using ergodic interference alignment (Theorem 2.14), which we consider later in Section 2.6, or by timesharing (Section 2.6.1).

Converse part. Suppose we have a code allowing us to achieve the sum-rate $r_\Sigma = r_1 + r_2$.

Suppose a genie provides receiver 2 with transmitter 1's message (which could only increase the capacity of the network). This allows receiver 2 to cancel the interference due to x_1.

By assumption, receiver 1 can decode his own message, and thus cancel the intended signal x_1 from his received signal y_1.

This leaves the two receivers with statistically equivalent signals. Therefore, if receiver 2 can decode message m_2 – which it can by assumption – then so can receiver 1.

Since receiver 1 is able to decode both messages, the sum rate cannot be more than the sum rate capacity of the multiple access channel seen at receiver 1, which by Theorem 2.8 is $r_\Sigma \leq \log(1 + 2\mathtt{snr})$. □

(We will use a similar proof strategy later to prove Lemma 4.11.)

This shows that dealing well with this mid-level interference will be particularly important. We examine ways of doing so in the next section.

2.6 Resource division

When faced with mid-level interference, a simple way of dealing with it is to share out the channel resources between the transmitters, to stop the interference from getting in the way. (This technique is often known as orthogonalisation.)

This idea is best illustrated by examples, of which we give three below.

A big advantage of resource-division strategies is that they are fairly easy to set up, and require neither detailed ongoing channel knowledge nor high computational complexity. There is also little need for cooperation between users after setup.

In this section, we assume for simplicity that all $\mathtt{snr}$s are equal.

2.6.1 ... by time

In schemes which share out the time resource, each transmitter is given sole use of the channel for some period of time, in return for which they may not transmit the rest of the time.

Consider the case of the finite-field model. At any particular time, only one user has control of the channel, and they can communicate up to their single-user capacity $\mathbb{D}(Z)$. All the other transmitters are silent, so the sum-rate is also $r_\Sigma = \mathbb{D}(Z)$.

Each user can communicate at a rate $r = \mathbb{D}(Z)/n$, for $\mathtt{dof} = 1/n$ degrees of freedom each (recall Definition 1.13), which in particular tends to 0 as the number of users n gets large. Naturally this is undesirable.

For the Gaussian model, transmitters can take advantage of the fact that they will only be transmitting part of the day, yet are operating under an *average* power constraint. Hence, when they are transmitting, transmitters can use power nP instead. Hence, each user can communicate at a rate $r = \frac{1}{n}\log(1+n\mathtt{snr})$.

Note that for low $\mathtt{snr}$ we have

$$\frac{c_\Sigma}{n} = \frac{1}{n}\log(1+n\mathtt{snr}) \approx \frac{1}{n} n\,\mathtt{snr}\log \mathrm{e} = \mathtt{snr}\log \mathrm{e},$$

which is the same as if there was no interference at all. Hence, for low $\mathtt{snr}$, these schemes are optimal.

For high $\mathtt{snr}$ (the more common case), we have

$$\frac{c_\Sigma}{n} = \frac{1}{n}\log(1+n\mathtt{snr}) \approx \frac{1}{n}(\log \mathtt{snr} + \log n) \approx \frac{1}{n}\log \mathtt{snr},$$

a reduction over the single-user case of a factor of $1/n$ again. That is, each user has $\mathtt{dof} = 1/n$ degrees of freedom each, for a total of $\mathtt{dof}_\Sigma = 1$ degrees of freedom all together (recall Definition 1.20).

In summary:

Theorem 2.11. *Consider an n-user finite field interference network. Then the rates* $r_i = \mathbb{D}(Z)/n$ *are simultaneously achievable, for a sum-rate of* $r_\Sigma = \mathbb{D}(Z)$ *and* $\mathtt{dof}_\Sigma = 1$.

Consider an n-user Gaussian interference network. Then the rates $r_i = \frac{1}{n}\log(1 + n\mathtt{snr}_i)$ *are simultaneously achievable. If all* snrs *are equal, this has a sum-rate of* $r_\Sigma = \log(1 + n\mathtt{snr})$, *which has* $\mathtt{dof}_\Sigma = 1$.

Høst-Madsen and Nosratinia showed that $\mathtt{dof} = 1/2$ each, for a total of $\mathtt{dof}_\Sigma = 1$ is optimal for $n = 2$ users, and, more generally, showed that [39, Section IV]

$$1 \leq \mathtt{dof}_\Sigma \leq \frac{n}{2} \qquad \text{for all } n \geq 2. \tag{2.2}$$

They further conjectured that in fact it is the lower bound that is tight, and that $\mathtt{dof}_\Sigma = 1$ is optimal [39, Section IV]. In other words, they conjectured that resource division strategies were also optimal at high snr. We will later see that this is not the case.

Note that this scheme, as with all resource division schemes, requires a small amount of precoordination between users, to decide which user is alloted which timeslot. We do not consider in this thesis the problem of how to conduct this cooperation.

2.6.2 ... by frequency

In schemes that share out the the frequency resource, each user is alloted a section of the frequency spectrum along which they may transmit, while remaining quiet over all other bandwidths.

This leads again to an average per-user capacity of c/n again.

Second-generation (GSM) mobile phone networks use a mixture of resource division by time and by frequency to share a channel of bandwidth 25 MHz between up to 1000 transmitters. First the spectrum is shared, giving 125 channels of bandwidth 200 kHz each. Then within each of these subchannels, the time is divided between up to 8 transmitters in time slots of 577 μs at a time [5, Example 3.1] [40, Example 14.2].

2.6.3 ... by codeword space

Code division multiple access (CDMA) is another way of allowing multiple users to use a shared channel. It works as follows. (For the purpose of simplifying this example, we shall think of a noiseless binary channel.)

Assume each transmitter i wishes to send a symbol $x_i \in \{0, 1\}$. Using CDMA it does this over T channel uses. Each transmitter i is given a vector $\mathbf{v}_i \in \{0, 1\}^T$. If that transmitter wish to send $x_i = 1$, she instead transmits the vector $\mathbf{v}_i$ over T channel uses; if she wishes to send $x_i = 0$, she instead

transmits $\mathbf{0}$, the zero vector of length T. In other words, transmitter i sends $x_i \mathbf{v}_i$.

The receivers then receive the superposition $\mathbf{y} = \sum_{i=1}^{n} x_i \mathbf{v}_i$. If the vectors were linearly independent (for which $T \geq n$ is necessary but not sufficient), then each x_i can be recovered.

Schemes such as this can be thought of as sharing the dimensions of the codeword space $\{0, 1\}^T$ among the n users, and so is another form of resource division.

2.7 Interference alignment

Interference alignment is the name for a new class of schemes for dealing with interference based on the following idea: if transmitters plan their signalling correctly, interference can 'align' at each receiver, with the desired signal split off separately. This allows receivers to share their resources just two ways – half for the signal, and half for all the 'aligned' interference. Thus all users can communicate at (roughly) the rate they could if there were just one interfering link.

In particular, this means that each user can obtain $1/2$ a degree of freedom, leading to $\mathtt{dof} = n/2$ degrees of freedom overall. So in fact, the upper bound (2.2) of Høst-Madsen and Nosratinia [39, Section IV] turns out to be correct, disproving their conjecture that the lower bound $\mathtt{dof} = 1$ was tight. (The conjecture was formally settled in the negative by Cadambe, Jafar, and Wang [41].)

Like resource division, interference alignment can be performed in a number of different ways. We show these by three toy examples, before concentrating on the specific case that will be useful to us later.

These schemes require channel state information at the transmitter (CSIT), in that the signal set in any time slot depends on the channel state coefficients at that time.

2.7.1 ... by codeword space

Note that CDMA (see Section 2.6.3) was, in some sense, wasteful for the interference network, since it allowed every receiver to decode every message, as if it were an X network. Whereas in fact, we only required each receiver to decode its own message.

Interference alignment by codeword choice, due to Cadambe and Jafar [28], develops this idea further. Each receiver receives a superposition of all the transmitters' (faded) signals, but by 'aligning' the interference, the receiver can work out its own signal, at less of a sacrifice than CDMA.

Consider the following 3-user fading interference network:

$$Y_1 = x_1 + \mathrm{i}x_2 + \mathrm{i}x_3 + Z_1$$
$$Y_2 = \mathrm{i}x_1 + x_2 + \mathrm{i}x_3 + Z_2$$
$$Y_3 = \mathrm{i}x_1 + \mathrm{i}x_2 + \mathrm{i}x_3 + Z_3$$

Note for this toy example we have chosen the unusual channel state matrix

$$\mathsf{H} = \begin{bmatrix} 1 & \mathrm{i} & \mathrm{i} \\ \mathrm{i} & 1 & \mathrm{i} \\ \mathrm{i} & \mathrm{i} & 1 \end{bmatrix},$$

which has diagonal entries (corresponding to direct links) all equal to 1, and off-diagonal entries (interfering links) all equal to i.

Suppose now that transmitters send their signal as just real numbers (taking advantage of the power constraint to send at twice the power). The receivers will receive their desired signal in the real subspace, but the interference will be aligned in the (purely) imaginary subspace.

Thus, each user can communicate interference-free at rate $r_i = \frac{1}{2}\log(1 + 2\mathtt{snr})$. Since we saw earlier that this was optimal at high $\mathtt{snr}$ for 2 users, it must certainly be optimal for 3 users too. Hence, this interference network has a sum-capacity of

$$c_\Sigma = \frac{3}{2}\log(1 + 2\mathtt{snr}) > \log(1 + 3\mathtt{snr}),$$

better than the $r_\Sigma = \log(1 + 3\mathtt{snr})$ achievable by resource division.

Cadambe and Jafar [28] managed to develop this idea, to show that it was possible for any values of fading parameters (provided there are no 'unexpected' linear dependencies – this would be avoided almost surely if the fading coefficient were from continuous distributions, for example).

They showed how transmitters can construct their signals so that at each receiver the signal is contained in one subspace of the signal space, and all the interference in another disjoint subspace, with both subspaces using roughly half of the available dimensions.

Specifically, they showed the following [28, Theorem 1]. (Recall the definition of degrees of freedom from Definition 1.20.)

Theorem 2.12. *Consider a Gaussian interference network with fixed fading coefficients h_{ji}. Then the total number of degrees of freedom is $\mathtt{dof}_\Sigma = n/2$.*

That is, for equal $\mathtt{snr}$*s, the sum-capacity is*

$$c_\Sigma = \frac{n}{2}\log \mathtt{snr} + o(\log \mathtt{snr}) \qquad \text{as } \mathtt{snr} \to \infty.$$

Note that the per-user capacity is

$$\frac{c_\Sigma}{n} = \frac{\frac{n}{2}\log(\mathtt{snr}) + o(\log \mathtt{snr})}{n} = \frac{1}{2}\log \mathtt{snr} + o(\log \mathtt{snr}),$$

compared with a single-user rate of

$$c = \log(1 + \mathtt{snr}) = \log \mathtt{snr} + o(\log \mathtt{snr}),$$

so the rate has been roughly halved. This compares well to the reduction to $1/n$ of resource division strategies, at least when $n > 2$.

El Ayach, Peters, and Heath [42] have conducted experiments that show that this interference alignment technique can perform well in real life for $n = 3$ users, with performance close to that predicted by theory.

2.7.2 ... by time

If a network has time delays, we can take advantage of these delays to align interference in the time domain. Interference alignment by time was first considered by Grokop, Tse and Yates [43]. However, due to the computational complexity of such schemes and the lack of physical applicability, it has received little attention since.

Specifically let τ_{ji} be the time delay between transmitter i and receiver j. Thus we have the model

$$Y_j[t] = \sum_{i=1}^{n} h_{ji} x_i[t - \tau_{ji}] + Z_j[t],$$

(with the convention that $x_i[t]$ is 0 for $t \leq 0$).

Consider a toy example with the following time delays:

$$\begin{array}{lll} \tau_{11} = 3 & \tau_{12} = 4 & \tau_{13} = 6 \\ \tau_{21} = 4 & \tau_{22} = 7 & \tau_{23} = 2 \\ \tau_{31} = 2 & \tau_{32} = 8 & \tau_{33} = 1 \end{array}$$

Note that this has been set up so that delays on direct links are odd numbers, while delays on crosslinks are even numbers.

This allows us to use the following strategy. Transmitters only send symbols at the odd numbered times, $t = 1, 3, 5, \ldots$. Then at even-numbered times, receivers will only get their desired signal (since odd + odd = even), and at odd-numbered times, two lots of 'aligned' interference (odd + even = odd).

Hence, users can communicate at half their single-user rate.

Grokop, Tse, and Yates showed a generalisation of this, but it is quite complicated: see their paper [43, Theorem 3.1] for details.

2.7.3 ... by channel state

Consider a fast fading 3-user interference network where the channel state matrix can take either of the following two values

$$\mathsf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \mathsf{H}' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

Note that that this toy example has been set up such that $\mathsf{H} + \mathsf{H}' = \mathsf{I} \pmod 2$.

Nazer, Gastpar, Jafar, and Viswanath [44] discovered a method of interference alignment that can code across the two channel states to recover a single message. Transmitters send the same signal in both states, and receivers combine two estimates to recover the desired message.

Nazer and coauthors named this scheme *ergodic interference alignment*. We investigate this further in the next section.

2.7.4 ... over the rational numbers

Interference alignment by codeword space and by channel state both require a channel which changes over time (or, equivalent, across the frequency spectrum), while interference alignment by time requires the existence of time delays. For some time, this left open the question of whether a form of interference alignment could be performed over a static channel without delays.

The question was answered in the positive Motahari, Gharan, Maddah-Ali, and Khandani, with a scheme they call *real interference alignment* [45].

The strategy works by effectively 'vectorising' the channel. Specifically, we can treat the real numbers $\mathbb{R}$ as a vector space over the rational numbers $\mathbb{Q}$. Then we can treat a real signal $x \in \mathbb{R}$ as a vector $x = \sum_k \lambda_k v_k$, where $\lambda_k \in \mathbb{Q}$ and the $v_k \in \mathbb{R}$ are some basis real numbers.

Using theorems about Diophantine rational approximations to real numbers, Motahari and coauthors deduce that real interference alignment achieves $n/2$ total degrees of freedom for the Gaussian interference channel.

Interested readers are referred to the original paper for further details [45].

2.8 Ergodic interference alignment

It's easiest to analyse ergodic interference alignment by first looking at the finite field channel. For convenience, we will assume that the fast-fading coefficients are IID uniform on $\mathbb{F}_q \setminus \{0\}$.

Recall from Theorem 1.12 that the single-user capacity of the finite field channel with non-zero fading is $\mathbb{D}(Z) := \log q - \mathbb{H}(Z)$.

The main lemma that gets ergodic interference alignment to work is the following [46, Theorem 1 and Corollary 2]. It is based on this observation: although receiver j would normally wish to reconstruct just its own message $\mathbf{m}_j$, it is, in fact, easier to reconstruct the 'pseudomessage' $\widetilde{\mathbf{m}}_j := \sum_{i=1}^n H_{ji}\mathbf{m}_i$.

Lemma 2.13. *Let $\mathcal{M} = \mathbb{F}_q^S$. Consider the finite field interference network. Then each receiver j can decode the linear combination of messages $\widetilde{\mathbf{m}}_j = \sum_{i=1}^n H_{ji}\mathbf{m}_i$ at rate $\mathbb{D}(Z)$.*

Proof. The key here is for all transmitters to use the same linear code. Let the generator matrix of this code be G. Write $\mathbf{g}[t]$ for the tth row of G, so $x_i[t] = \mathbf{g}[t]\mathbf{m}_i$ (since $\mathbf{m}_i$ is a column vector). Then each receiver j sees signal

$$\begin{aligned} Y_j[t] &= \sum_{i=1}^n H_{ji}x_i[t] + Z_j[t] \\ &= \sum_{i=1}^n H_{ji}(\mathbf{g}[t]\mathbf{m}_i) + Z_j[t] \\ &= \mathbf{g}[t]\sum_{i=1}^n H_{ji}\mathbf{m}_i + Z_j[t] \\ &= \mathbf{g}[t]\widetilde{\mathbf{m}}_j + Z_j[t]. \end{aligned}$$

But this is precisely as if the single message $\widetilde{\mathbf{m}}_i$ was sent with the linear code. Since very good linear codes exist (Theorem 1.15), this can be done at rates up to the single-user capacity $c = \mathbb{D}(Z)$. □

The technique proceeds as follows: Match a state H with the *complementary state* $\mathsf{H}' = \mathsf{I} - \mathsf{H}$. Transmitters send the same signal (encoding the same message) in both states. Then after T occurrences of the first state receiver j decodes $\widetilde{m}_j = \sum_{i=1}^n H_{ji}m_i$, at rate $\mathbb{D}(Z)$ and after T occurrences of the complementary state decodes

$$\widetilde{\mathbf{m}}'_j = \sum_{i=1}^n H'_{ji}\mathbf{m}_j = \sum_{i=1}^n (\delta_{ji} - h_{ji})\mathbf{m}_i$$

also at rate $\mathbb{D}(Z)$. Receiver j then computes the message estimate

$$\widehat{\mathbf{m}}_j = \widetilde{\mathbf{m}}_j + \widetilde{\mathbf{m}}'_j = \sum_{i=1}^n (h_{ji} + \delta_{ji} - h_{ji})\mathbf{m}_i = \sum_{i=1}^n \delta_{ji}\mathbf{m}_i = \mathbf{m}_j,$$

as desired. Since decoding the message required twice the blocklength, the rate is half what it would be, $(\log 2^{T\mathbb{D}(Z)})/2T = \mathbb{D}(Z)/2$.

(Observe that the receiver needs to perform two separate estimates, and cannot simply add the channel outputs together. To do this would lead to a channel of the form $Y = x_j + Z * Z$, where the convolution $Z * Z$ means the sum of two IID copies of Z. The overall rate in this case is $\mathbb{D}(Z * Z) \leq$

$\mathbb{D}(Z)/2$, unless Z is deterministic, with strict inequality unless $\mathbb{D}(Z) = 0$ [47]. In general, the K-fold convolution has relative entropy from the uniform $\mathbb{D}(Z * \cdots * Z)$ that usually decreases exponentially in K [47].)

Nazer and coauthors use a typical set argument to show that sufficiently many channel states can be matched up in this way, showing that with high probability each matrix and its complement show up almost the same number of times. They use this to prove the following theorem [44, Lemma 3 and Theorem 2]:

Theorem 2.14. *For the model as outlined above, the rates $r_i = \mathbb{D}(Z)/2$ are simultaneously achievable.*

Using a quantisation argument, they show the following [44, Theorem 3]:

Theorem 2.15. *For the fast fading Gaussian interference channel with symmetric fading (that is, H and $-\mathsf{H}$ have the same distribution), the rates $r_i = \frac{1}{2}\mathbb{E}\log(1 + 2\mathtt{SNR})$ are simultaneously achievable.*

We will use this result more in Chapter 4.

Notes

The section consists of a review of the existing literature; the mathematical contents is not claimed to be new.

On networks, the textbook of Cover and Thomas [6, Chapter 15] and the review papers of El Gamal and Cover [48] and Kramer [49] were useful.

The first detailed study of the point-to-point link was by Shannon [4], of the multiple access network was by Ahlswede [21] and Liao [22], of the broadcast network was by Cover [24], and of the interference network was by Carleial [50].

The strategies of interference as noise, decode-and-subtract, and resource division are old and well-known, making tracking down details of their discovery difficult. The textbooks by Tse and Viswanath [5] and by Goldsmith [40] give good background on this material.

The concept of interference alignment – first discovered in the alignment by codeword choice paradigm – is due to Cadambe and Jafar [28], and also to Maddah-Ali, Motahari, and Khandani [51], who independently discovered a similar method (published in the same issue of the same journal).

The toy example of interference alignment by codeword choice is due to Jafar [52]. The example of interference alignment by time is after Grokop, Tse, and Yates [43]. Ergodic interference alignment was discovered by Nazer and coauthors [44].

A tutorial by Jafar [52] was useful for the material on interference alignment.

3

Regular and Poisson random networks

In this chapter, we look at two networks based on models of how nodes are positioned in space. In the *regular network*, nodes are positioned at regular spacings, as in a grid; in the *Poisson random network*, nodes are positioned at random according to a Poisson point process.

We examine how large networks can operate using simple 'interference as noise' techniques. In particular, we show the important relationship between the attenuation α of the signals (which describes how quickly signals die off over distance) and the dimension d of the network.

3.1 Model

We use the model of a Gaussian network with slow fading based on power-law attenuation. That is we have a countable set of points (*nodes*) placed in d-dimensional Euclidean space $\mathbb{R}^d$ – each node $i \in \mathbb{Z}_+$ is positioned at the point $\mathbf{T}_i \in \mathbb{R}^d$.

On the tth channel use, the signal received by node j is

$$Y_j[t] = \sum_{i \neq j} h\rho(i,j)^{-\alpha/2} x_i[t] + Z_j[t].$$

Here, $\rho(i,j) = \|\mathbf{T}_j - \mathbf{T}_i\|$ is the Euclidean distance between nodes i and j. We call h the fixed fading parameter. Large h corresponds to signals being much more powerful than noise; small h corresponds to signals being much less powerful than noise. To concentrate on the interference-limited regime, we will sometimes consider the limit $h \to \infty$, which is equivalent to the noiseless network with $h = 1$,

$$Y_j[t] = \sum_{i \neq j} \rho(i,j)^{-\alpha/2} x_i[t].$$

The Euclidean norm in $\mathbb{R}^d$ will be denoted $\| \ \|$, where d is the dimension of the network. It will be useful later to define

$$v(d) := \frac{\pi^{d/2}}{\Gamma(1 + d/2)},$$

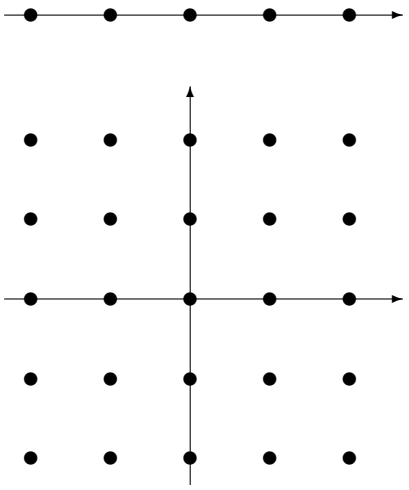
which is the volume of the Euclidean unit ball in $\mathbb{R}^d$.

The $d = 2$ case is the most commonly studied, as this obviously has real-world applications. The case $d = 3$ has applications in, for example, tall office buildings; and the $d = 1$ is attracting more attention for 'car-to-car' protocols (see for example recent work by the US Department of Transportation [53]), where a long road can be modelled as a one-dimensional line.

3.2 Regular networks

In a d-dimensional *regular network*, (see for example Xie and Kumar [54]), nodes are placed at points of $\mathbb{Z}^d$ in d-dimensional space $\mathbb{R}^d$. In particular, any two nodes are a distance at least 1 from each other.

Below we show regular networks in one and two dimensions respectively.



Each node $\mathbf{t} \in \mathbb{Z}^d$ is a transmitter, transmitting to one of its $2d$ nearest neighbours, chosen arbitrarily. We write $\mathbf{t} \to \mathbf{r}$ to indicate that node $\mathbf{t}$ transmits to node $\mathbf{r}$.

All nodes will use standard Gaussian codebooks of power $P = 1$, generated independently of each other. This means that – power aside – signals are statistically indistinguishable from a) each other, and b) the background noise. We use the principle that interference should be treated as noise (as in Section 2.3).

Following Gupta and Kumar [35] the interference at any node $\mathbf{r} \in \mathbb{Z}^d$

$$I = \sum_{\substack{\mathbf{t} \neq \mathbf{r} \\ \mathbf{t} \not\to \mathbf{r}}} h \|\mathbf{r} - \mathbf{t}\|^{-\alpha}, \tag{3.1}$$

and using interference-as-noise the communication rate

$$r := \log(1 + \mathtt{sinr}) = \log\left(1 + \frac{h}{I+1}\right)$$

is achievable for each link $\mathbf{r} \to \mathbf{t}$.

That is, if the interference I is finite, then every node can transmit at this fixed rate. This is known as *linear growth* for the following reason: if we have a sequence of sets of nodes $(S_n : n \in \mathbb{N})$, where $S_n \subset \mathbb{Z}$ is of cardinality n, then there exists an achievable rate n-tuple $(r_i : i \in S_n)$ such that

$$\sum_{\mathbf{j} \in S_n} r_j = nr = O(n).$$

The following theorem generalises the work of Xie and Kumar [54], who proved it for the case $d = 2$.

Theorem 3.1. *The d-dimensional linear network supports linear growth, provided the ratio of the attenuation to the dimension of the network is sufficiently large, specifically if $\alpha > d$.*

Proof. We proceed by induction on the dimension d, showing that the interference $I = I(\alpha, d)$ is finite for $\alpha > d$. Without loss of generality, receiver $\mathbf{0}$ is receiving a message from transmitter $\mathbf{t}^* := (1, 0, \ldots, 0)$.

First, the base case, $d = 1$. The interference is

$$I(\alpha, 1) = h \sum_{t \neq 0,1} |t|^{-\alpha} = h\left(2\sum_{t=1}^{\infty} t^{-\alpha} - 1\right),$$

which is finite for $\alpha > 1$, as desired.

The inductive hypothesis is that $I(\alpha, d-1)$ is finite.

Now the inductive step. Again, the interference is

$$I(\alpha, d) = h \sum_{\mathbf{t} \neq \mathbf{0}, \mathbf{t}^*} \|\mathbf{t}\|^{-\alpha} \leq h \sum_{\mathbf{t} \neq \mathbf{0}} \|\mathbf{t}\|^{-\alpha}.$$

We now split $\mathbb{Z}^d$ into the d different $(d-1)$-dimensional coordinate spaces (where at least one coordinate is 0), and the 2^d open orthants (where all coordinates are nonzero). This gives

$$I(\alpha, d) \leq dI(\alpha, d-1) + h2^d \sum_{\mathbf{t} \in \mathbb{N}^d} \|\mathbf{t}\|^{-\alpha}. \tag{3.2}$$

The first term in (3.2) is finite by the inductive hypothesis; we concentrate on the second term. By treating the point $\mathbf{1} = (1, 1, \ldots, 1)$ separately, we have

$$h2^d \sum_{\mathbf{t}\in\mathbb{N}^d} \|\mathbf{t}\|^{-\alpha} = h2^d d^{-\alpha/2} + h2^d \sum_{\mathbf{t}\in\mathbb{N}^d\setminus\{\mathbf{1}\}} \|\mathbf{t}\|^{-\alpha}. \tag{3.3}$$

The second term in (3.3) can be approximated by an integral, since for $\mathbf{t} \in \mathbb{N}^d$,

$$\|\mathbf{t}\|^{-\alpha} \le \int_{t_1-1}^{t_1} \cdots \int_{t_d-1}^{t_d} \|\mathbf{t}\|^{-\alpha} \, \mathrm{d}t_1 \cdots \mathrm{d}t_d.$$

Hence,

$$\begin{aligned} h2^d \sum_{\mathbf{t}\in\mathbb{N}^d\setminus\{\mathbf{1}\}} \|\mathbf{t}\|^{-\alpha} &\le h2^d \int_1^\infty \cdots \int_1^\infty \|\mathbf{t}\|^{-\alpha} \, \mathrm{d}t_1 \cdots \mathrm{d}t_d \\ &= h \int_{\mathbb{R}^d\setminus[-1,1]^d} \|\mathbf{t}\|^{-\alpha} \, \mathrm{d}\mathbf{t} \\ &\le h \int_{\mathbb{R}^d\setminus B(\mathbf{0},1)} \|\mathbf{t}\|^{-\alpha} \, \mathrm{d}\mathbf{t}, \end{aligned}$$

where $B(\mathbf{0}, 1)$ is the d-dimensional unit ball centered at the origin. We now use a change of coordinates to $\rho = \|\mathbf{t}\|$, $\mathbf{s} = \mathbf{t}/\rho$, so that $\mathrm{d}\mathbf{t} = \rho^{d-1} \, \mathrm{d}\rho \, \mathrm{d}\mathbf{s}$. This gives

$$\begin{aligned} h2^d \sum_{\mathbf{t}\in\mathbb{N}^d\setminus\{\mathbf{1}\}} \|\mathbf{t}\|^{-\alpha} &\le h \int_{\|\mathbf{s}\|=1} \int_{\rho=1}^\infty \rho^{-\alpha} \rho^{d-1} \, \mathrm{d}\rho \, \mathrm{d}\mathbf{s} \\ &= hdv(d) \int_{\rho=1}^\infty \rho^{-(\alpha-d)-1} \, \mathrm{d}\rho \\ &= hdv(d) \left[-\frac{1}{\alpha-d} \rho^{-(\alpha-d)} \right]_1^\infty \\ &= h\frac{dv(d)}{\alpha-d}, \end{aligned} \tag{3.4}$$

which is finite. Putting together (3.2), (3.3), and (3.4), we get

$$I(\alpha, d) \le I(\alpha, d-1) + h2^d d^{-\alpha/2} + h\frac{dv(d)}{\alpha-d} < \infty.$$

The inductive step is complete and the theorem is proven. □

Note that this result is the best possible. (By 'best possible', we mean that Theorem 3.1 is not true for $\alpha \le d$. We do not claim our bounds on $I(\alpha, d)$ are as tight as possible.) If $\alpha \le d$, then the interference is

$$I(\alpha, d) \ge h \sum_{t=1}^\infty t^d t^{-\alpha} \ge h \sum_{t=1}^\infty t^d t^{-d} h \sum_{t=1}^\infty 1 = \infty,$$

and the signal-to-interference-plus-noise ratio is 0.

It is worth noting that a simple bound for $I(\alpha, d)$ is

$$I(\alpha, d) \leq h\frac{\alpha}{\alpha - d}2^{d-1}(d+1)!.$$

This can be proved inductively, using $v(d) \leq 2^d$ (that is, the volume of the unit sphere is less than that of the surrounding cube).

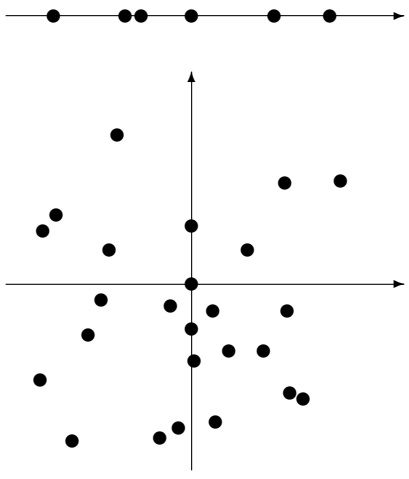
3.3 Poisson random networks

In this section, we define the Poisson random network, where nodes are distributed like a Poisson process. We give a local result – bounds on the outage probability of a single transmission – and a global result – showing that linear growth occurs in the network with high probability.

3.3.1 Node positioning model

In a d-dimensional Poisson random network (studied extensively by Haenggi [55, 56], and Dhillon, Ganti, and Andrews [57], among others), the set of nodes $\{\mathbf{T}_i : i \in \mathbb{Z}_+\}$ are placed in $\mathbb{R}^d$ as a Poisson point process of density 1 (without loss of generality). For simplicity, we will translate the points such that $\mathbf{T}_0$ is at the origin $\mathbf{0}$, and relabel the nodes by order of distance from the origin. So we have $0 = \|\mathbf{T}_0\| \leq \|\mathbf{T}_1\| \leq \cdots$ (and strict inequalities almost surely).

The figures below shows Poisson random networks in one and two dimensions.



We want to model the scenario of a multihop network; that is, where messages are sent to distant nodes by being successively passed over short distances by a number of intermediate nodes. For this model, we shall assume

that each node broadcasts its intended signal, whilst picking up the signal from its nearest neighbour. We will concentrate on the communication over just these short hops – the large-scale strategy has been studied by many others, particularly the multihop strategy of Gupta and Kumar [35] and the work on hierarchical cooperation by Özgür, Lévêque, and Tse [37].

If node i's nearest neighbour is node j, we again write $j \to i$. In particular, the nearest neighbour to node 0, at the origin, is node 1 at $\mathbf{T}_1$, so $1 \to 0$.

All nodes will use standard Gaussian codebooks of power $P = 1$, generated independently of each other. Nodes treat signals other than they are picking up as Gaussian noise, as discussed in Section 2.3.

3.3.2 Outage probability

Given the link $j \to i$ and the positions of nodes i and j, the position of other nodes in the network is random. In particular, for any given rate $r > 0$, we cannot guarantee that j can communicate to i at rate r. This is because the other nodes might (with non-zero probability) crowd round i, drowning out the intended signal from j.

Hence, we need to study the *outage probability* of the network (as discussed previously in Subsection 1.5.3).

Definition 3.2. We define the outage probability as

$$p_{\text{out}}(r) = \mathbb{P}\left(r > \log\left(1 + \texttt{SINR}_{ji}\right)\right).$$

Here (and throughout) $\mathbb{P}$ denotes probability over the Poisson point process.

The signal-to-interference-plus-noise ratio at node j, $\texttt{SINR}_{ji}$, has marginal distribution function

$$F_{\texttt{SINR}}(s) = \mathbb{P}\left(\frac{h\|\mathbf{T}_j - \mathbf{T}_i\|^{-\alpha}}{I_{ji} + 1} \leq s\right),$$

independent of the link $i \to j$ where the interference is

$$I_{ji} = \sum_{k \neq i,j} h\|\mathbf{T}_j - \mathbf{T}_k\|^{-\alpha}.$$

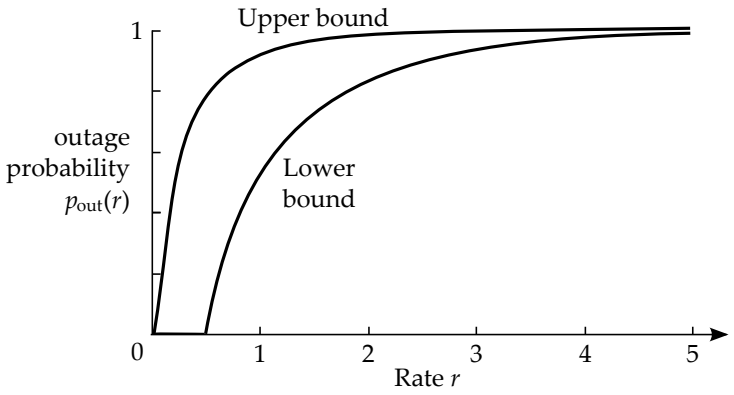
From now on, we deal with the link $1 \to 0$, without loss of generality. We will suppress subscripts that are no longer necessary.

An important special case is the high power regime $h \to \infty$, which is simpler to deal with mathematically and is important for studying networks that are interference limited rather than noise limited. Note that since

$$\begin{aligned}\lim_{h\to\infty} \texttt{SINR} &= \lim_{h\to\infty} \frac{h\|\mathbf{T}_j - \mathbf{T}_i\|^{-\alpha}}{\sum_{k\neq i,j} h\|\mathbf{T}_j - \mathbf{T}_k\|^{-\alpha} + 1} \\ &= \frac{\|\mathbf{T}_j - \mathbf{T}_i\|^{-\alpha}}{\sum_{k\neq i,j} \|\mathbf{T}_j - \mathbf{T}_k\|^{-\alpha}},\end{aligned}$$

this is equivalent to taking $h = 1$ and ignoring the noise term.

The next two theorems give upper (Theorem 3.3) and lower (Theorem 3.4) bounds on the outage probability $p_{\text{out}}(r)$. The figure below shows these bounds for the common case $d = 2$, $\alpha = 3$ in the high power $h \to \infty$ regime.



Theorem 3.3. *Consider a d-dimensional Poisson random network with fixed fading parameter h, and suppose $\alpha > d$. Then the outage probability is upper-bounded by*

$$p_{out}(r) \leq \frac{d(2^r-1)}{\alpha - d}\left(1 + 2\frac{2^r-1}{h}v(d)^{-\alpha/d}\,\Gamma\left(2+\frac{\alpha}{d}\right)\right) + \exp\left(-v(d)\left(2\frac{2^r-1}{h}\right)^{-d/\alpha}\right).$$

In the high power $h \to \infty$ regime, we have

$$p_{out}(r) \leq \frac{d(2^r-1)}{\alpha - d}\left(1 - \exp\left(-\frac{\alpha - d}{d(2^r-1)}\right)\right).$$

Note that since

$$\frac{d}{\alpha - d} = \frac{1}{\alpha/d - 1},$$

the attentuation α and dimension d only enter this theorem through their ratio.

Proof. First, note that the outage probability can be rewritten as

$$p_{\text{out}}(R) = \mathbb{P}\left(r > \log\left(1 + \texttt{SINR}\right)\right) = F_{\texttt{SINR}}(s),$$

where we have defined $s := 2^r - 1$. So now we need only bound $F_{\texttt{SINR}}(s)$, the probability that SINR is large.

We will bound $F_{\texttt{SINR}}(s)$ by conditioning on the position of the nearest neighbour. So

$$\begin{aligned} F_{\texttt{SINR}}(s) &= \mathbb{P}(\texttt{SINR} \leq s) \\ &= \int_0^\infty f_{\|\mathbf{T}_1\|}(t)\mathbb{P}\left(\texttt{SNR} \leq s \,\middle|\, \|\mathbf{T}_1\| = t\right) \mathrm{d}t \\ &= \int_0^\infty f_{\|\mathbf{T}_1\|}(t)\mathbb{P}\left(I \geq \frac{1}{s}ht^{-\alpha} - 1 \,\middle|\, \|\mathbf{T}_1\| = t\right) \mathrm{d}t. \end{aligned} \tag{3.5}$$

It is known [58, Theorem 1] (and can be easily shown) that $\|\mathbf{T}_1\|^d$ has an exponential distribution with parameter $v(d)$,

$$\|\mathbf{T}_1\|^d \sim \mathrm{Exp}\big(v(d)\big),$$

giving

$$f_{\|\mathbf{T}_1\|}(t) = dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}. \tag{3.6}$$

We will also need to bound the probability that the interference is large, which we do using the conditional Markov inequality. Specifically, when $\frac{1}{s}ht^{-\alpha} - 1 > 0$, we have

$$\mathbb{P}\left(I \geq \frac{1}{s}ht^{-\alpha} - 1 \,\middle|\, \|\mathbf{T}_1\| = t\right) \leq \min\left\{\frac{1}{\frac{1}{s}ht^{-\alpha} - 1}\,\mathbb{E}\left(I \,\middle|\, \|\mathbf{T}_1\| = t\right), 1\right\}, \tag{3.7}$$

and when $\frac{1}{s}ht^{-\alpha} - 1 \leq 0$, we take the trivial bound 1 instead.

We concentrate on the simpler high power scenario first. Note that in the limit $h \to \infty$, we always have $\frac{1}{s}ht^{-\alpha} - 1 > 0$.

For the moment, we concentrate on the first argument. We can write

$$\mathbb{E}\left(I \,\middle|\, \|\mathbf{T}_1\| = t\right) = \mathbb{E}\sum_{\mathbf{T}\in\mathcal{P}(t)} h\|\mathbf{T}\|^{-\alpha},$$

where $\mathcal{P}(t)$ is the set of points of the Poisson process outside the ball of radius t about the origin. Using Campbell's theorem (see for example the monograph of Haenggi and Ganti [55, Theorem A.2]), we then have

$$\begin{aligned}
\mathbb{E}\sum_{\mathbf{T}\in\mathcal{P}(t)} h\|\mathbf{T}\|^{-\alpha} &= \int_{\|\mathbf{u}\|\geq t} h\|\mathbf{u}\|^{-\alpha}\,\mathrm{d}\mathbf{u} \\
&= hdv(d)\left[-\frac{1}{\alpha - d}\rho^{-(\alpha-d)}\right]_{\rho=t}^{\infty} \\
&= h\frac{dv(d)}{\alpha - d}t^{-(\alpha-d)}.
\end{aligned} \tag{3.8}$$

Substituting (3.8) and (3.6) this back into equation (3.5) gives

$$\begin{aligned}
F_{\mathtt{SINR}}(s) &\leq \lim_{h\to\infty}\int_0^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}\min\left\{\frac{1}{\frac{1}{s}ht^{-\alpha} - 1}h\frac{dv(d)}{\alpha - d}t^{d-\alpha}, 1\right\}\mathrm{d}t \\
&= \int_0^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}\min\left\{\frac{1}{\frac{1}{s}t^{-\alpha}}\frac{dv(d)}{\alpha - d}t^{d-\alpha}, 1\right\}\mathrm{d}t \\
&= \int_0^{t^*} dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}st^\alpha\frac{dv(d)}{\alpha - d}t^{d-\alpha}\,\mathrm{d}t \\
&\qquad + \int_{t^*}^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}st^\alpha\,\mathrm{d}t,
\end{aligned}$$

where

$$t^* = \left(s\frac{dv(d)}{\alpha - d}\right)^{-1/d}$$

is the point where we cross over from one argument of the 'min' to the other.

Calculating these integrals using a substitution $y = v(d)x^d$ gives

$$F_{\mathtt{SINR}}(s) \leq \frac{ds}{\alpha - d}\left(1 - \mathrm{e}^{-(\alpha-d)/ds}\right).$$

This proves the theorem for the high power regime.

We now move back to the general case. Returning to (3.7), we now have

$$\begin{aligned}\int_0^\infty f_{\|\mathbf{T}_1\|}(t)\mathbb{P}\left(I \geq \frac{1}{s}ht^{-\alpha} - 1 \,\middle|\, \|\mathbf{T}_1\| = t\right)\mathrm{d}t \\ \leq \int_0^{t^{**}} dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}\min\left\{\frac{1}{\frac{1}{s}ht^{-\alpha}-1}h\frac{dv(d)}{\alpha-d}t^{d-\alpha}, 1\right\}\mathrm{d}t \\ + \int_{t^{**}}^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}\,\mathrm{d}t,\end{aligned}$$

where $t^{**} = (s/h)^{-1/\alpha}$ is the point at which $\frac{1}{s}ht^{-\alpha} - 1 = 0$.

We could evaluate this integral numerically, or express it as a complicated sum of Gamma functions. Instead, we will show a simple bound.

When $t \leq 2^{-1/\alpha}t^{**}$, we have $\frac{s}{h}t^\alpha \leq \frac{1}{2}$, and hence

$$\frac{1}{\frac{1}{s}ht^{-\alpha}-1}h = \frac{st^\alpha}{1-\frac{s}{h}t^\alpha} \leq st^\alpha\left(1 + 2\frac{s}{h}t^\alpha\right).$$

This allows us to bound the integral by

$$\begin{aligned}\int_0^{2^{-1/\alpha}t^{**}} dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}s\left(1+2\frac{s}{h}t^\alpha\right)\frac{dv(d)}{\alpha-d}t^d\,\mathrm{d}t \\ + \int_{2^{-1/\alpha}t^{**}}^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}\,\mathrm{d}t \\ \leq \frac{ds}{\alpha-d}\left(1 + 2\frac{s}{h}v(d)^{-\alpha/d}\Gamma\left(2+\frac{\alpha}{d}\right)\right) + \exp\left(-v(d)\left(2\frac{s}{h}\right)^{-d/\alpha}\right),\end{aligned}$$

as required, where the first term follows upon replacing the upper limit of the integral by ∞. □

Theorem 3.4. *Under the same conditions as Theorem 3.3, the outage probability is bounded below by*

$$p_{out}(r) \geq 1 - \frac{1}{(2^r-1)^{d/\alpha}}.$$

Proof. The key is to observe that the interference I is at least as large as the contribution coming from the second-nearest neighbour $\mathbf{T}_2$. That is,

$$\begin{aligned}\mathbb{P}\left(I \geq \frac{1}{s}ht^{-\alpha} + 1 \,\middle|\, \|\mathbf{T}_1\| = t\right) &\geq \mathbb{P}\left(I \geq \frac{1}{s}ht^{-\alpha} \,\middle|\, \|\mathbf{T}_1\| = t\right) \\ &\geq \mathbb{P}\left(h\|\mathbf{T}_2\|^{-\alpha} \geq \frac{1}{s}ht^{-\alpha} \,\middle|\, \|\mathbf{T}_1\| = t\right).\end{aligned}$$

Rearranging this, we get

$$\begin{aligned}\mathbb{P}\Big(h\|\mathbf{T}_2\|^{-\alpha} \geq \frac{1}{s}ht^{-\alpha} \,\Big|\, \|\mathbf{T}_1\| = t\Big) &= \mathbb{P}\Big(\|\mathbf{T}_2\|^{-\alpha} \geq \frac{1}{s}t^{-\alpha} \,\Big|\, \|\mathbf{T}_1\| = t\Big)\\ &= 1 - \mathbb{P}\Big(\|\mathbf{T}_2\| > ts^{1/\alpha} \,\Big|\, \|\mathbf{T}_1\| = t\Big)\\ &= 1 - \exp\Big(-v(d)\big((ts^{1/\alpha})^d - t^d\big)\Big),\end{aligned}$$

where the final result follows on considering the probability that the annulus $\{\mathbf{u} \in \mathbb{R}^d : t < \|\mathbf{u}\| \leq ts^{1/\alpha}\}$ is empty.

Again, combining this with Equations (3.5) and (3.6), we obtain a lower bound on the outage probability of the form

$$\begin{aligned}&\int_0^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d}\left(1 - \mathrm{e}^{-v(d)t^d(s^{d/\alpha}-1)}\right)\mathrm{d}t\\ &\qquad = \int_0^\infty dv(d)x^{d-1}\mathrm{e}^{-v(d)t^d}dt - \int_0^\infty dv(d)t^{d-1}\mathrm{e}^{-v(d)t^d s^{d/\alpha}}\mathrm{d}t,\end{aligned}$$

and the result follows on making the change of variables $y = v(d)t^d$. □

3.3.3 Linear growth

Recall that we say that *linear growth* occurs when the sum-rate of n users scales linearly with n. That is, if we have a sequence of sets of nodes $\big(S_n : n \in \mathbb{N}\big)$, where $S_n \subset \mathbb{Z}_+$ is of cardinality n, then there exists an achievable rate n-tuple $(r_{i\to j} : i \in S_n)$ such that $\sum_{\mathbf{i}\in S_n} r_{i\to j} = O(n)$.

In particular if a 'proportion' p of links were to support a given rate r, then we would have

$$\sum_{\mathbf{i}\in S_n} r_{i\to j} \approx pnr = O(n),$$

which would be sufficient to show linear growth.

In particular, if communication on distinct links were independent, this would be true with $p = 1 - p_{\text{out}}(r)$. Although the links are not independent, links 'far enough away' are. Thus by splitting the network into 'close' and 'distant' nodes we can prove the theorem.

Theorem 3.5. *Consider a d-dimensional Poisson random network with attenuation α in the interference-limited $h \to \infty$ regime. Suppose $\alpha > d$. Then we have linear growth with probability tending to 1 as $n \to \infty$ at rate $O(n^{-(1-d/\alpha)})$.*

The following Chernoff-type bound on Poisson random variables will be useful later.

Lemma 3.6. *Let $X \sim Po(\lambda)$. Then $\mathbb{P}(X \geq \mathrm{e}\lambda) \leq \mathrm{e}^{-\lambda}$.*

Proof. We have, using Markov's inequality,

$$\mathbb{P}(X \geq \mathrm{e}\lambda) = \mathbb{P}(\mathrm{e}^X \geq \mathrm{e}^{\mathrm{e}\lambda}) \leq \frac{\mathbb{E}\,\mathrm{e}^X}{\mathrm{e}^{\mathrm{e}\lambda}} = \frac{\mathrm{e}^{(\mathrm{e}-1)\lambda}}{\mathrm{e}^{\mathrm{e}\lambda}} = \mathrm{e}^{-\lambda}. \qquad \square$$

We are now in the position to prove Theorem 3.5.

Proof of Theorem 3.5. First recall that $h \to \infty$ is equivalent to taking $h = 1$ an deleteing the noise term.

Let n be some fixed integer. (Later, we will consider the sum rate of n nodes, and let $n \to \infty$.) Fix the nearest-neighbour link $i \to j$.

Given a node j situated at the point $\mathbf{T}_j$, we divide the other nodes into those *close* to j

$$C(j) := \{k \neq i, j : \|\mathbf{T}_j - \mathbf{T}_k\| \leq n^a\}$$

and those *distant* from j

$$D(j) := \{k \neq i : \|\mathbf{T}_j - \mathbf{T}_k\| > n^a\},$$

where a is some parameter to be chosen later.

Outage occurs when the signal-to-interference-plus-noise ratio is insufficient to support some given rate $r = \log(1+s)$. We will consider whether an outage event is caused primarily by close or distant nodes. Specifically, we define the events

$$\mathrm{Out}_C(j) := \left\{ \frac{\|\mathbf{T}_j - \mathbf{T}_i\|^{-\alpha}}{\sum_{k \in C(j)} \|\mathbf{T}_j - \mathbf{T}_k\|^{-\alpha}} \leq \frac{s}{2} \right\},$$
$$\mathrm{Out}_D(j) := \left\{ \frac{\|\mathbf{T}_j - \mathbf{T}_i\|^{-\alpha}}{\sum_{k \in D(j)} \|\mathbf{T}_j - \mathbf{T}_k\|^{-\alpha}} \leq \frac{s}{2} \right\}.$$

Note that if a link $i \to j$ is in outage, at least one of $\mathrm{Out}_C(j)$ and $\mathrm{Out}_D(j)$ must be occurring. Note also that the (marginal) distributions of $\mathrm{Out}_C(j)$ and $\mathrm{Out}_D(j)$ are independent of the node j. When the node index is irrelevant, we suppress it.

Suppose n nodes all try to communicate at the rate $r = \log(1+s)$. Then for $\epsilon \in (0, \frac{1}{2})$,

$$\begin{aligned}
&\mathbb{P}(\text{total rate} \leq (1-2\epsilon)nr) \\
&\quad \leq \mathbb{P}(\text{number of outages} \geq 2\epsilon n) \\
&\quad = \mathbb{P}(\text{number of distant outages} \geq \epsilon n \cup \text{number of close outages} \geq \epsilon n) \\
&\quad \leq \mathbb{P}(\text{number of distant outages} \geq \epsilon n) + \mathbb{P}(\text{number of close outages} \geq \epsilon n) \\
&\quad = \mathbb{P}\left(\sum_{i=0}^{n-1} \mathbb{1}[\mathrm{Out}_D(j)] \geq \epsilon n\right) + \mathbb{P}\left(\sum_{i=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \geq \epsilon n\right). \qquad (3.9)
\end{aligned}$$

We bound the two terms above separately.

For the first term in (3.9), by Markov's inequality, we have

$$\mathbb{P}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_D(j)] \geq \epsilon n\right) \leq \frac{\mathbb{E}\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_D(j)]}{\epsilon n} = \frac{1}{\epsilon}\mathbb{P}\big(\mathrm{Out}_D\big).$$

Using the same ideas as in the proof of Theorem 3.3, we can show that this term tends to zero for $a > 0$ at rate $O(n^{-a(\alpha-d)})$. (The key is to change the lower limit in the integral in (3.8) to n^a.)

We bound the second term in (3.9) using the idea that for most j and k, $C(j)$ and $C(k)$ are disjoint, so the corresponding contributions will be independent.

Let N_j denote the number of nodes whose 'close' regions overlap with j's 'close' region; that is

$$N_j := \#\{k \neq j : C(j) \cup C(k) \neq \varnothing\}.$$

Note that N_j is the number of nodes in a ball of radius $2n^a$, so is Poisson with mean $2^d v(d) n^{ad}$.

We write $\mathcal{N} = \{\max_{0\leq j\leq n-1} N_j \geq 2^d v(d) n^{ad}\mathrm{e}\}$ for the event that one of the N_j is particularly large. We will argue that the event $\mathcal{N}$ can be ruled out with high probability, allowing good control of the growth of the variance.

We exploit the fact that, conditioned on the event $\mathcal{N}^c$, for any j there are at most $1 + 2^d v(d) n^{ad}\mathrm{e}$ indices k such that $\mathrm{Cov}\left(\mathbb{1}[\mathrm{Out}_C(j)], \mathbb{1}[\mathrm{Out}_C(k)] \mid \mathcal{N}^c\right)$ is non-zero, and each such covariance is no greater than 1. Hence we can control the growth of the variance of the sum as

$$\begin{aligned}\mathrm{Var}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \,\middle|\, \mathcal{N}^c\right) &\leq \sum_{j=0}^{n-1} (1 + 2^d v(d) n^{ad}\mathrm{e})\, \mathrm{Var}\left(\mathbb{1}[\mathrm{Out}_C(j)] \mid \mathcal{N}^c\right) \\ &\leq (1 + 2^d v(d) n^{ad}\mathrm{e}) n. \end{aligned} \tag{3.10}$$

By the union bound and Lemma 3.6, we have

$$\mathbb{P}(\mathcal{N}) \leq \sum_{j=0}^{n-1} \mathbb{P}(N_j \geq 2^d v(d) n^{ad}) \leq n\mathrm{e}^{-2^d v(d) n^{ad}}. \tag{3.11}$$

Using the law of total probability, and substituting (3.10) and (3.11) into Chebyshev's inequality gives

$$\begin{aligned}\mathbb{P}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \geq \epsilon n\right) &= \mathbb{P}(\mathcal{N})\mathbb{P}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \geq \epsilon n \,\middle|\, \mathcal{N}\right) \\ &\quad + \mathbb{P}(\mathcal{N}^c)\mathbb{P}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \geq \epsilon n \,\middle|\, \mathcal{N}^c\right) \\ &\leq n\mathrm{e}^{-2^d v(d) n^{ad}} + \mathbb{P}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \geq \epsilon n \,\middle|\, \mathcal{N}^c\right)\end{aligned}$$

$$\leq n\mathrm{e}^{-2^d v(d) n^{ad}} + \frac{\mathrm{Var}\left(\sum_{j=0}^{n-1} \mathbb{1}[\mathrm{Out}_C(j)] \mid \mathcal{N}^c\right)}{n^2\left(\epsilon - \mathbb{E}(\mathbb{1}[\mathrm{Out}_C] \mid \mathcal{N}^c)\right)^2}$$
$$\leq n\mathrm{e}^{-2^d v(d) n^{ad}} + \frac{(1 + 2^d v(d) n^{ad} \mathrm{e}) n}{n^2\left(\epsilon - \mathbb{P}(\mathrm{Out}_C \mid \mathcal{N}^c)\right)^2}.$$

As we send $n \to \infty$, this tends to 0 at rate $O(n^{ad-1})$.

Setting $a = 1/\alpha$ means the two terms both tend to 0 at rate

$$ad - 1 = -a(\alpha - d) = -\left(1 - \frac{d}{\alpha}\right),$$

as desired.

□

3.4 Further work

The strength of a Poisson model is the extensibility and mathematical flexibility of the Poisson point process.

For just one example, in a spatially inhomogeneous Poisson process, there is a density function $\lambda\colon \mathbb{R}^d \to \mathbb{R}_+$, such that the number of nodes in some region A is $\mathrm{Po}\left(\int_A \lambda(\mathbf{x})\,\mathrm{d}\mathbf{x}\right)$. Under what conditions on λ is linear growth still possible?

Many other open questions remain.

Notes

The new work in this chapter is joint work with Oliver Johnson and Robert Piechocki. It has not previously been published, although the research was conducted prior to some since-published work [59, 55]

The monographs of Baccelli and Błaszczyszyn [59] and Haenggi and Ganti [55] were useful background on the SINR approach to stochastic networks, as was the work of Haenggi [56], Xie and Kumar [54], and Gupta and Kumar [35].

The textbook of Kingman [60] was useful background material for Poisson processes.

4

Sum-capacity of random dense Gaussian interference networks

In this chapter, we will find approximations to the sum-capacity of interference networks with many users. Our interference networks will be motivated by physical models of wireless networks.

When we say many users, we will be allowing the number of users n to tend to infinity, and looking at asymptotic behaviour. We will be doing this without expanding the area in which the nodes reside, so the nodes will get packed closer and closer together – for this reason, such networks are called *dense* networks.

We will model transmitters and receivers as being placed randomly in space. This means that signal- and interference-to-noise ratios will be random too, but fixed for the duration of communication – and hence a form of slow fading.

The crucial insight to prove these results is that the performance in these networks is tightly constrained by the performance on a few so-called bottleneck links. Ideas from interference alignment (specifically ergodic interference alignment) are crucial in performing well on these links, and hence in the whole network.

The main result (Theorem 4.3) will be to show, for the model we will consider, that the sum-capacity C_Σ is roughly $\frac{n}{2}\mathbb{E}\log(1+2\mathtt{SNR})$. Specifically, we show that C_Σ/n converges in probability to $\frac{1}{2}\mathbb{E}\log(1+2\mathtt{SNR})$ as $n\to\infty$.

In this chapter, random positions will give a form of slow fading, and we will also add fast fading, to make a realistic model of real-world networks.

All results in this chapter hold provided that this fast fading is symmetrical (in the sense that H_{ji} and $-H_{ji}$ are identically distributed, for all i and j).

For simplicity and concreteness, we will choose to give all our results in the context of random phase fading, so

$$H_{ji}[t] = |H_{ji}| \exp(\mathrm{i}\Theta_{ji}[t]) = \begin{cases} \sqrt{\mathtt{SNR}_i} \exp(\mathrm{i}\Theta_{ji}[t]) & \text{for } i = j, \\ \sqrt{\mathtt{INR}_{ji}} \exp(\mathrm{i}\Theta_{ji}[t]) & \text{for } i \neq j. \end{cases}$$

In Section 4.6, we briefly discuss how to apply other fading models such as Rayleigh fading.

4.1 Introduction

Recently, progress has been made on many-user approximations to the sum-capacity C_Σ of random Gaussian interference networks.

In particular, in a 2009 paper, Jafar [27, Theorem 5] proved a result on the asymptotic sum-capacity of a particular random Gaussian interference network:

Theorem 4.1. *Suppose direct* SNR*s are fixed and identical, so* $\mathtt{SNR}_i = \mathtt{snr}$ *for all* i*, and suppose that all* INR*s are IID random and supported on some neighbourhood of* snr*. Then the average per-user capacity* C_Σ / n *tends in probability to* $\frac{1}{2}\log(1 + 2\mathtt{snr})$ *as* $n \to \infty$*.*

We examine Jafar's result in detail later.

(Here and elsewhere, we use C_Σ to denote the sum-capacity of the network, and interpret C_Σ / n as the average per-user capacity.)

A subsequent result by Johnson, Aldridge and Piechocki [1, Theorem 4.1] concerned a more physically realistic model, the standard dense network:

Theorem 4.2. *Suppose receivers and transmitters are placed IID uniformly at random on the unit square* $[0, 1]^2$*, and suppose that signal power attenuates like a polynomial in* $1/\mathrm{distance}$*. Then the average per-user capacity* C_Σ / n *tends in probability to* $\frac{1}{2}\mathbb{E}\log(1 + 2\mathtt{SNR})$ *as* $n \to \infty$*.*

In this chapter, we prove a similar, but more general, result to Theorem 4.2, with a neater proof, using ideas from Jafar's proof of Theorem 4.1. We assume transmitters and receivers are situated independently at random in space (not necessarily uniformly), and that the power of a signal depends in a natural way on the distance it travels.

Specifically our result is the following (full definitions of non-italicised technical terms are in Section 4.2):

Theorem 4.3. *Consider a Gaussian interference network formed by n pairs of nodes placed in a* spatially separated IID network *with* power-law attenuation. *Then the average per-user capacity C_Σ/n converges in probability to $\frac{1}{2}\mathbb{E}\log(1+2\mathtt{SNR})$, in that for all $\epsilon > 0$*

$$\mathbb{P}\left(\left|\frac{C_\Sigma}{n} - \frac{1}{2}\mathbb{E}\log(1+2\mathtt{SNR})\right| > \epsilon\right) \to 0 \quad \textit{as } n \to \infty.$$

The direct part of the proof uses interference alignment; specifically, we take advantage of ergodic interference alignment (see Section 2.6).

The converse part of the proof uses the idea of 'bottleneck links' developed by Jafar [27]. An information theoretic argument gives a capacity bound on such bottleneck links, and probabilistic counting arguments show there are sufficiently many such links to tightly bound the sum-capacity of the whole network.

Before going any further, we should mention two similar results in the same area using different interference alignment techniques.

A paper by Özgür and Tse [61] proves linear scaling in interference networks by showing that for any value of $\mathtt{snr}$ and n, the sum-rate is bounded below by $k_1 n \log(1 + k_2 \mathtt{snr})$, where $k_1, k_2 > 0$ are universal constants.

Secondly, a paper by Niesen [62] bounds the capacity region (rather than just the sum-capacity) of arbitrary dense networks, albeit with a factor of $O(\log n)$ separating the inner and outer bounds.

4.2 Model

We outline the model we will use. We will model separately how messages are transmitted, and how nodes are positioned.

These ideas were introduced in an earlier paper [1], but were not fully exploited, due to that paper's concentration on the standard dense network.

4.2.1 Communication model

We will use the n-user Gaussian interference network as our main model. That is, we have

$$Y_j[t] = \sum_{i=1}^{n} H_{ji}[t]x_i[t] + Z_j[t].$$

The fading coefficients $H_{ji}[t]$ will be made up of a fast fading part, representing the moment-to-moment changes in the channel, and a slow fading part, the power attenuation due to node placing.

The results require the fast fading to be symmetric, in that H_{ji} and $-H_{ji}$ are identically distributed – this is ensured by the random phase we use (and

is also satisfied by Rayleigh fading). For ease of notation and to simplify exposition, will assume the fast fading takes the form of a (uniformly) random phase change (although we briefly discuss more general models in Section 4.6). Therefore, we can write the fading coefficients in modulus–argument form as

$$H_{ii} = \exp(\mathrm{i}\Theta_{ii}[t])\sqrt{\mathtt{SNR}_i}, \qquad H_{ji} = \exp(\mathrm{i}\Theta_{ji}[t])\sqrt{\mathtt{INR}_{ji}} \quad j \neq i,$$

where the $\Theta_{ji}[t]$ are IID uniform on $[0, 2\pi)$, and $\mathtt{SNR}_i = |H_{ii}|^2$ and $\mathtt{INR}_{ji} = |H_{ji}|^2$ are the squared moduli (which are constant over time).

Our results are in the context of so-called 'line of sight' communication models, without multipath interference. That is, we consider a model where signal strengths attenuate deterministically with distance according to some function a.

Definition 4.4. Fix transmitter node positions $\{\mathbf{T}_1, \ldots, \mathbf{T}_n\} \in \mathbb{R}^d$ and receiver node positions $\{\mathbf{R}_1, \ldots, \mathbf{R}_n\} \in \mathbb{R}^d$, and consider Euclidean distance $\| \quad \|$ and an *attenuation function* $a \colon \mathbb{R}_+ \to \mathbb{R}_+$.

We define $\mathtt{SNR}_i = a(\|\mathbf{R}_i - \mathbf{T}_i\|)$, and for all pairs with $i \neq j$, define $\mathtt{INR}_{ji} = a(\|\mathbf{R}_j - \mathbf{T}_i\|)$.

We consider the n-user Gaussian interference network. So transmitter i sends a message encoded as a codeword $\mathbf{x}_i = (x_i[1], \ldots, x_i[T])$ to receiver i, under a power constraint $\frac{1}{T}\sum_{t=1}^{T} |x_i[t]|^2 \leq 1$ for each i. The tth symbol received at receiver j is given as

$$Y_j[t] = \exp(\mathrm{i}\Theta_{jj}[t])\sqrt{\mathtt{SNR}_j}x_j[t] + \sum_{i \neq j} \exp(\mathrm{i}\Theta_{ji}[t])\sqrt{\mathtt{INR}_{ji}}x_i[t] + Z_j[t],$$

where the noise terms $Z_j[t]$ are independent standard complex Gaussian random variables, and the phases $\Theta_{ji}[t]$ are independent $U[0, 2\pi)$ random variables independent of all other terms. The $\mathtt{INR}_{ji}$ and $\mathtt{SNR}_i$ remain fixed over time, since the node positions themselves are fixed, but the phases are fast-fading, in that they are renewed for each t.

Definition 4.5. We say an attenuation function a has *power-law attenuation* if there exist constants α and k_{att} such that for all ρ, we have $a(\rho) \leq k_{\mathrm{att}}\rho^{-\alpha}$.

In particular, standard power-law decay of the form $a(\rho) = h\rho^{-\alpha/2}$ clearly satisfies this with $k_{\mathrm{att}} = h$ and α set to $\alpha/2$. Other models we discussed in Section 1.4 such as $a(\rho) = h\max\{1, \rho^{-\alpha/2}\}$ and $a(\rho) = h(\rho + \rho_0)^{-\alpha/2}$ also satisfy this.

For brevity, we write S_{ji} for the random variables $\frac{1}{2}\log(1 + 2\mathtt{INR}_{ji})$ (when $i \neq j$), and S_{ii} for $\frac{1}{2}\log(1 + 2\mathtt{SNR}_i)$ which are functions of the distance between the transmitters and receivers. In particular, since the nodes are positioned

independently, under this model the random variables S_{ji} are identically distributed, and S_{ji} and S_{lk} are IID when $\{i, j\}$ and $\{k, l\}$ are disjoint.

We will also write $E = \mathbb{E}S_{ii} = \frac{1}{2}\mathbb{E}\log(1 + 2\mathtt{SNR})$, noting that this is independent of i. (It is also true that $E = \mathbb{E}S_{ji}$ for all i and j.) Lemma 4.12 later ensures that E is indeed finite.

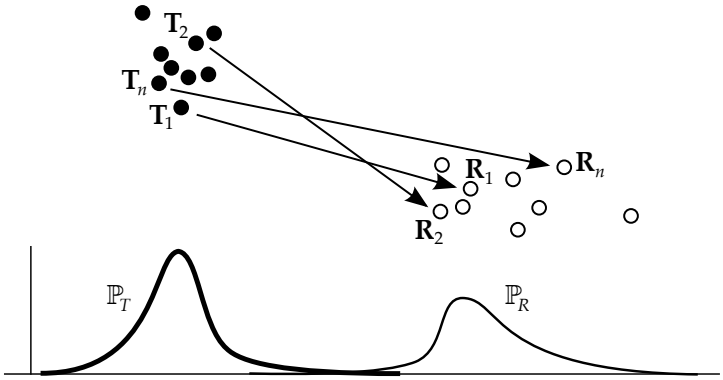
4.2.2 Node position model

We believe that our techniques should work in a variety of models for the node positions. We outline one very natural scenario here.

Definition 4.6. Consider two probability distributions $\mathbb{P}_T$ and $\mathbb{P}_R$ defined on d-dimensional space $\mathbb{R}^d$. Given an integer n, we sample n transmitter node positions $\mathbf{T}_1, \ldots, \mathbf{T}_n$ independently from the distribution $\mathbb{P}_T$. Similarly, we sample n receiver node positions $\mathbf{R}_1, \ldots, \mathbf{R}_n$ independently from distribution $\mathbb{P}_R$. We refer to such a model of node placement as an *IID network*.

(Equivalently, we could state that transmitter and receiver positions are distributed like two independent non-homogeneous Poisson processes, conditioned such that there are n points of each type.)

We pair the transmitter and receiver nodes up so that transmitter i at $\mathbf{T}_i$ wishes to communicate with receiver i at $\mathbf{R}_i$ for each i.



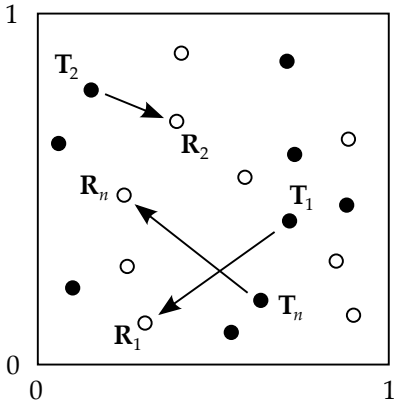
Transmitters and receivers that are very close to each other will lead to very strong interference or signals. These extreme occurences could prevent the network from operating as we would like. Also, as we remarked in Section 1.4, our attenuation models lose physical accuracy at very small distances. For this reason, we will demand that our node positioning model ensures that this is rare, a property we call *spatial separation*.

Definition 4.7. Let $\mathbf{T} \sim \mathbb{P}_T$ and $\mathbf{R} \sim \mathbb{P}_R$ be placed independently in $\mathbb{R}^d$. We say the IID network is *spatially separated* if there exists constants $\beta > 0$ and k_{sep} such that for all ρ

$$\mathbb{P}(\|\mathbf{R} - \mathbf{T}\| \leq \rho) \leq k_{\text{sep}}\rho^{\beta}.$$

In particular, we can show that the standard dense network is spatially separated.

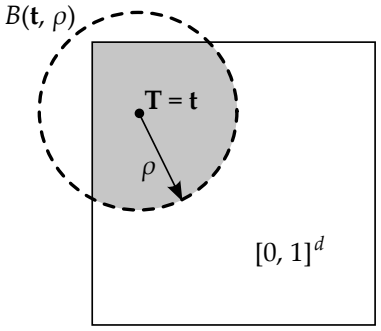
Definition 4.8. The *d-dimensional standard dense network* is an IID network defined by $\mathbb{P}_T$ and $\mathbb{P}_R$ being independent uniform measures on $[0,1]^d$.



Lemma 4.9. *The standard dense network is spatially separated.*

Proof. We need to show that Definition 4.7 is fulfilled. By conditioning on $\mathbf{T} = \mathbf{t}$, we get

$$\begin{aligned}
&\mathbb{P}(\|\mathbf{R} - \mathbf{T}\| \le \rho) \\
&\quad = \int_{[0,1]^d} \mathbb{P}_R\big(B(\mathbf{t},\rho) \cap [0,1]^d\big)\, \mathrm{d}\mathbf{t} \\
&\quad \le \int_{[0,1]^d} \mathbb{P}_R\big(B(\mathbf{t},\rho)\big)\, \mathrm{d}\mathbf{t} \\
&\quad \le \int_{[0,1]^d} v(d)\rho^d\, \mathrm{d}\mathbf{t} \\
&\quad = v(d)\rho^d,
\end{aligned}$$



where $v(d)$ is the volume of the d-dimensional unit ball $B(\mathbf{0}, 1)$. Taking $k_{\text{sep}} = v(d)$, $\beta = d$ gives the result. □

The standard dense network has been the subject of much research (see the review paper of Xue and Kumar [36] and references therein). However, we emphasise that our result holds for a wider range of models.

4.3 Jafar network

We now review in more detail Jafar's important result. We call a network with fixed equal `snr`s and IID `INR`s a *Jafar network*. (Note that the Jafar network cannot be written as an IID network with power-law attenuation.)

Theorem 4.1 restated. *Suppose direct* SNR*s are fixed and identical, so* $\mathtt{SNR}_i = \mathtt{snr}$ *for all* i*, and suppose that all* INR*s are IID random and supported on some neighbourhood of* snr*. Then the average per-user capacity* C_Σ/n *tends in probability to* $\frac{1}{2}\log(1+2\mathtt{snr})$ *as* $n \to \infty$.

Proving the direct part of this result is simple: the central result of ergodic interference alignment (Theorem 2.15) tells us that the rates $\frac{1}{2}\log(1+2\mathtt{snr})$ are simultaneously achievable by all users.

For the converse part, Jafar defines [27, proof of Theorem 5] the crosslink $i \to j$ as being a *ϵ-bottleneck link* if the sum-capacity of the two-user network with links $i \to i$ and $j \to j$ with crosslink $i \to j$ has sum-rate bounded by

$$r_i + r_j \leq \log(1+2\mathtt{snr}) + \epsilon. \tag{4.1}$$

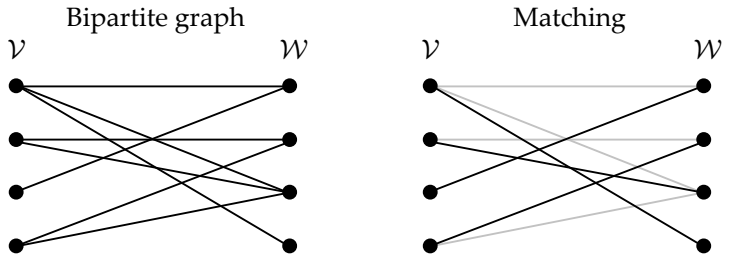
for some fixed $\epsilon > 0$. Analysis of these bottleneck links then gives the converse result.

We will use a similar method to this to prove our main result (Theorem 4.3). We will need to alter the definition of bottleneck links slightly to fit our needs. Also, the entire converse part is made more complicated: while the Jafar network has lots of convenient independences between the snr and INRs, we are not so lucky. Therefore extra care must be taken.

We also give here an alternative proof of Theorem 4.1. This proof uses techniques from graph theory, and gives a faster rate of convergence than Jafar's own proof – exponential, rather than $O(n^{-2})$.

Alternative proof of Theorem 4.1. We need to show that every user is involved in a bottleneck link. We will first review some facts from random graph theory.

Form a random bipartite graph by taking two sets $\mathcal{V}$, $\mathcal{W}$ of vertices each of size K, and making each edge from $\mathcal{V}$ to $\mathcal{W}$ present independently with probability δ (there are no edges within either $\mathcal{V}$ or $\mathcal{W}$). A *matching* is a set of k of the edges such that every vertex in the graph is adjacent to one edge – so each vertex $v \in \mathcal{V}$ is matched to a vertex $w \in \mathcal{W}$ by an edge vw.



A classical result due to Erdös and Rényi [63, Theorem 2] (originally stated in terms of random matrices) states that the probability that a matching fails to exist tends to 0 for any $\delta = \delta(K) = (\log K + \omega(1))/K$, where $\omega(1)$ is, using Bachmann–Landau asymptotic notation, a term that tends to ∞.

We recall the argument where δ is a fixed constant, so that we can be precise about the bounds, rather than just working asymptotically. Following Walkup [64, Definition 1], we say that a subset $\overline{\mathcal{V}} \subseteq \mathcal{V}$ of size k and a subset $\overline{\mathcal{W}} \subseteq \mathcal{W}$ of size $K-k+1$ form a *blocking pair* of size k if no edge of the graph connects $\overline{\mathcal{V}}$ to $\overline{\mathcal{W}}$. Walkup [64, Section 3] uses König's theorem (equivalently one can use Hall's marriage theorem) to deduce that

$$\begin{aligned}\mathbb{P}(\text{no matching from } \mathcal{V} \text{ to } \mathcal{W}) &\leq \sum_{k=1}^{K} \sum_{\substack{|\overline{\mathcal{V}}|=k \\ |\overline{\mathcal{W}}|=K-k+1}} \mathbb{P}\left((\overline{\mathcal{V}}, \overline{\mathcal{W}}) \text{ a blocking pair}\right) \\ &= \sum_{k=1}^{K} \binom{K}{k}\binom{K}{K-k+1}(1-\delta)^{k(K-k+1)} \\ &= 2\sum_{k=1}^{(K+1)/2} \binom{K}{k}\binom{K}{k-1}(1-\delta)^{k(K-k+1)}.\end{aligned}$$

We split this sum into the terms where $k \leq \sqrt{K}$ and those where $k > \sqrt{K}$. Using the bounds

$$(1-\delta)^{k(K-k+1)} \leq \begin{cases} \exp\left(-\delta\,\dfrac{K+1}{2}\right) & \text{for } k \leq \sqrt{K}, \\ \exp\left(-\delta\,\dfrac{K^{3/2}}{2}\right) & \text{for } k > \sqrt{K}, \end{cases}$$

we get the bound

$$\mathbb{P}(\text{no matching}) \leq 2\sqrt{K}K^{2\sqrt{K}}\exp\left(-\delta\,\frac{K+1}{2}\right) + 2^{2K}\exp\left(-\delta\,\frac{K^{3/2}}{2}\right).$$

We deduce that the probability of a complete matching failing to exist decays at an exponential rate in K.

We can now prove Theorem 4.1.

Construct a random bipartite graph by dividing the receiver-transmitter links into two groups $\mathcal{V}$ and $\mathcal{W}$ of size $K = n/2$. Choose $\epsilon > 0$ and, for $i \in \mathcal{V}$ and $j \in \mathcal{W}$ include the edge ij if either of the crosslinks $i \to j$ or $j \to i$ is an ϵ-bottleneck link, which in the Jafar network occurs independently with some probability δ.

We seek a matching on this graph. For each pair (i, j) that is successfully matched up, the corresponding two-user channel is an ϵ-bottleneck, and hence has the bound $r_i + r_j \leq \log(1 + 2\mathtt{snr}) + \epsilon$ by (4.1). for any achievable rates. If every edge is matched this way, we have

$$r_\Sigma = \sum_{i=1}^{n} r_i = \sum_{\text{matches } (i,j)} r_i + r_j \leq \frac{n}{2}\left(\log(1+2\mathtt{snr}) + \epsilon\right)$$

The high probability of a matching existing implies exponential decay of

$$\mathbb{P}\left(\left|\frac{c_\Sigma}{n} - \frac{1}{2}\log(1+2\mathtt{snr})\right| > \epsilon\right) \to 0$$

proving the theorem with exponential convergence. □

For our IID networks, extra dependencies between links make the picture much more difficult, so we have been unable to find a proof that extends this random bipartite graph method. Whether or not exponential convergence holds for IID networks with power-law attenuation is an open problem.

4.4 Proof: achievability

We can now prove our main theorem, Theorem 4.3, by breaking the probability into two terms, which we deal with separately. So

$$\mathbb{P}\left(\left|\frac{C_\Sigma}{n} - E\right| > \epsilon\right) = \mathbb{P}\left(\frac{C_\Sigma}{n} - E < -\epsilon\right) + \mathbb{P}\left(\frac{C_\Sigma}{n} - E > \epsilon\right). \tag{4.2}$$

Bounding the first term of (4.2) corresponds to the achievability part of the proof. Bounding the second term of (4.2) corresponds to the converse part, and represents our major contribution.

We prove the direct part using ergodic interference alignment.

Proof. The first term of (4.2) can be bounded relatively simply, using ergodic interference alignment. A theorem of Nazer, Gastpar, Jafar, and Vishwanath (Theorem 2.15 of this thesis) implies that the rates

$$R_i = \frac{1}{2}\log(1+2\mathtt{SNR}_i) = S_{ii}$$

are simultaneously achievable.

This implies that $C_\Sigma \geq \sum_{i=1}^n R_i = \sum_{i=1}^n S_{ii}$. This allows us to bound the first term in (4.2) as

$$\mathbb{P}\left(\frac{C_\Sigma}{n} - E < -\epsilon\right) \leq \mathbb{P}\left(\frac{\sum_{i=1}^n S_{ii}}{n} < E - \epsilon\right).$$

But $E = \mathbb{E}S_{ii}$, so this probability tends to 0 by the weak law of large numbers. □

4.5 Proof: converse

We now need to show that the second term of (4.2) tends to 0 too. Specifically, we must prove the following: for all $\epsilon > 0$

$$\mathbb{P}\left(\frac{C_\Sigma}{n} \geq E + \epsilon\right) \to 0 \qquad \text{as } n \to \infty. \tag{4.3}$$

The proof of the converse part is the major new part of this chapter. First, bottleneck links are introduced, and we prove a tight information-theoretic bound on the capacity of such links. Second, a probabilistic counting argument ensures there are (with high probability) sufficiently many bottleneck links to bound the sum-capacity of the entire network.

4.5.1 Bottleneck links

The important concept is that of the bottleneck link, an idea first used by Jafar [27] and later adapted by Johnson, Aldridge, and Piechocki [1] in the following form:

Definition 4.10. We say the link $i \to j$, $i \neq j$, is an *ϵ-bottleneck link*, if the following three conditions hold:

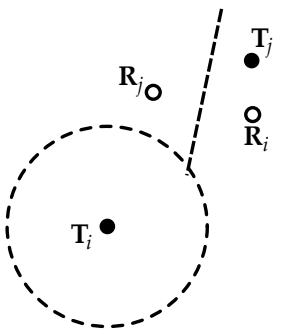
B1 $S_{ii} \leq E + \epsilon/2$,

B2 $S_{ji} \leq E + \epsilon/2$,

B3 $S_{jj} \leq S_{uj}$.

We let B_{ij} be the indicator function that the crosslink $i \to j$ is a ϵ-bottleneck link. We also define the *bottleneck probability* $\beta := \mathbb{E}\, B_{ij}$ to be the probability that a given link is an ϵ-bottleneck which is independent of i and j for an IID network. (We suppress the ϵ dependence for simplicity.)

This definition does have a physical interpretation (although the only motivation for it is that it allows us to prove the bound in Lemma 4.11).



The physical interpretation is this: fix the position $\mathbf{T}_i$ of transmitter i. Condition B1 requires that receiver i is sufficiently far away from $\mathbf{T}_i$, and condition B2 requires that receiver j is sufficiently far away too. Condition B3 requires that transmitter j is closer to receiver j than to receiver i.

The crucial point about bottleneck links is the constraints they place on achievable rates in a network.

Lemma 4.11. *Consider a crosslink $i \to j$ in a n-user Gaussian interference network. If $i \to j$ is a ϵ-bottleneck link, then the sum of their achievable rates is bounded by $r_i + r_j \leq 2E + \epsilon$.*

Proof. First, note that we make things no worse by considering the two-user subnetwork:

$$Y_i = \exp(\mathrm{i}\Theta_{ii})\sqrt{\mathtt{SNR}_i}X_i + \exp(\mathrm{i}\Theta_{ij})\sqrt{\mathtt{INR}_{ij}}X_j + Z_i$$
$$Y_j = \exp(\mathrm{i}\Theta_{ij})\sqrt{\mathtt{INR}_{ji}}X_i + \exp(\mathrm{i}\Theta_{ji})\sqrt{\mathtt{SNR}_j}X_j + Z_j$$

where receiver i needs to determine message m_i, and receiver j message m_j. (The time index is omitted for clarity.)

From bottleneck conditions B1 and B2 we have

$$1 + 2\mathtt{SNR}_i \leq \exp(2E + \epsilon), \quad 1 + 2\mathtt{INR}_{ij} \leq \exp(2E + \epsilon).$$

Summing and taking logs gives

$$\log(1 + \mathtt{SNR}_i + \mathtt{INR}_{ij}) \leq 2E + \epsilon. \tag{4.4}$$

We combine this with the argument given by Jafar [27], which we discussed earlier (Section 2.3). Let r_i and r_j be jointly achievable rates for the subnetwork. In particular, receiver i can determine message m_i with an arbitrarily low probability of error.

We certainly do no worse if a genie presents message m_i to receiver j, so we assume receiver j can indeed recover m_i. But condition B3 ensures that it is easier for receiver i to determine m_j than it is for receiver j (since the weighting is larger in the first case). So since receiver j can recover m_j (as r_j is achievable), receiver i can recover m_j also.

Because receiver i can determine both m_i and m_j, these two signals must have been transmitted at a sum-rate no higher than the sum-capacity of the Gaussian multiple-access channel (Theorem 2.8). Hence,

$$r_i + r_j \leq \log(1 + \mathtt{SNR}_i + \mathtt{INR}_{ij}) \leq 2E + \epsilon,$$

where the second inequality comes from (4.4). □

4.5.2 Three technical lemmas

A few technical lemmas are required in order to prove (4.3).

First, we need to ensure that very high SNRs are very rare (Lemma 4.12). Second, we need to show that bottleneck links will actually occur (Lemma 4.13). Last, we must show that the number of bottleneck links cannot vary too much (Lemma 4.14).

Under any network model where these three lemmas are true, our theorem will hold. We emphasise that our model of IID networks with power-law attenuation is one such model; we believe the result holds more widely.

Lemma 4.12. *Consider a spatially separated IID network, with power-law attenuation. Then for any $\eta > 0$,*

$$\mathbb{P}\left(\max_{1 \le i \le n} S_{ii} > n^{\eta/2} \right) = O(n^{-1}) \qquad \text{as } n \to \infty.$$

In fact, in our case convergence to 0 is considerably quicker than $O(n^{-1})$, but this is sufficient.

Proof. First, we have by the union bound

$$\mathbb{P}(\max S_{ii} > n^{\eta/2}) \le n\mathbb{P}(S_{11} > n^{\eta/2}). \tag{4.5}$$

Now we apply the definition of $S_{11} := \frac{1}{2}\log(1 + 2\mathtt{SNR})$ and recall that $\mathtt{SNR}_1 = a(\|\mathbf{R}_1 - \mathbf{T}_1\|)$ (Definition 4.4) to get

$$\begin{aligned}\mathbb{P}(S_{11} > n^{\eta/2}) &= \mathbb{P}\left(\mathtt{SNR}_1 > \frac{1}{2}(2^{2n^{\eta/2}} - 1)\right) \\ &= \mathbb{P}\left(a(\|\mathbf{R}_1 - \mathbf{T}_1\|) > \frac{1}{2}(2^{2n^{\eta/2}} - 1)\right).\end{aligned}$$

Since a is a power-law attenuation function, we have

$$\begin{aligned}\mathbb{P}(S_{11} > n^{\eta/2}) &\le \mathbb{P}\left(k_{\text{att}}\|\mathbf{R}_1 - \mathbf{T}_1\|^{-\alpha} > \frac{1}{2}(2^{2n^{\eta/2}} - 1)\right) \\ &= \mathbb{P}\left(\|\mathbf{R}_1 - \mathbf{T}_1\| < \left(\frac{1}{2k_{\text{att}}}(2^{2n^{\eta/2}} - 1)\right)^{-1/\alpha}\right);\end{aligned}$$

and since the network is spatially separated, we have

$$\mathbb{P}(S_{11} > n^{\eta/2}) \le k_{\text{sep}}\left(\frac{1}{2k_{\text{att}}}(2^{2n^{\eta/2}} - 1)\right)^{-\beta/\alpha} = O(n^{-2})$$

(and obviously much tighter than $O(n^{-2})$). Together with (4.5), this gives the result. □

(It is worth noting that this fast decay in the tails of S_{ii} ensures that the expectation $E = \mathbb{E}S_{ii}$ does indeed exist and is finite.)

We will often condition off this event; that is, condition on the complementary event $\{\max S_{ii} \le n^{\eta/2}\}$. We use $\mathbb{P}_n$, $\mathbb{E}_n$ and Var_n to denote such conditionality, and write $\beta_n = \mathbb{E}_n B_{ij}$ for the conditional bottleneck probability.

The next two lemmas concern showing that conditional probabilities are nonzero. However, we have for any event A,

$$\begin{aligned}\mathbb{P}(A) = \mathbb{P}(A \mid \max S_{ii} \le n^{\eta/2})\mathbb{P}(\max S_{ii} \le n^{\eta/2}) \\ + \mathbb{P}(A \mid \max S_{ii} > n^{\eta/2})\mathbb{P}(\max S_{ii} > n^{\eta/2}).\end{aligned}$$

and hence by Lemma 4.12 we have the two bounds

$$\mathbb{P}(A) \leq \mathbb{P}(A \mid \max S_{ii} \leq n^{\eta/2}) + \mathbb{P}(\max S_{ii} > n^{\eta/2}) = \mathbb{P}_n(A) + O(n^{-1})$$
$$\mathbb{P}(A) \geq \mathbb{P}(A \mid \max S_{ii} \leq n^{\eta/2})\mathbb{P}(\max S_{ii} \leq n^{\eta/2}) = \mathbb{P}_n(A)\big(1 - O(n^{-1})\big),$$

and so

$$\mathbb{P}(A) = \mathbb{P}_n(A) + O(n^{-1}). \tag{4.6}$$

This will be useful in the next two proofs.

Lemma 4.13. *Consider a spatially separated IID network, with power-law attenuation. Then the conditional bottleneck probability β_n is bounded away from 0 for all n sufficiently large.*

Proof. First note that by (4.6), we need only show that the unconditional bottleneck probability β is nonzero.

Second, note that by the exchangeability of $\mathbf{R}_i$ and $\mathbf{R}_j$, we have

$$\mathbb{P}(\text{B1 and B2 and B3}) \geq \frac{1}{2}\mathbb{P}(\text{B1 and B2}).$$

It is left to show that $\mathbb{P}(\text{B1 and B2})$ is non-zero.

Note that B1 requires S_{ii} to be less than its expectation plus ϵ. So $\mathbf{R}_i$ must be situated such that this has nonzero probability. So $\mathbf{T}_i$ has a nonzero probability of being positioned such that B1 occurs. But $\mathbf{T}_i$ and $\mathbf{T}_j$ are also exchangeable, so we are done. □

Lemma 4.14. *Consider a spatially separated IID network with power-law attenuation. Then, conditional on $\{\max_i S_{ii} < n^{\eta/2}\}$,*

$$\mathrm{Var}_n(\#\text{ bottleneck links}) = \mathrm{Var}_n\left(\sum_{i \neq j} B_{ij}\right) = O(n^3),$$

where the sum is over all crosslink pairs (i, j), $i \neq j$.

In general, one might assume that $\mathrm{Var}_n(\#\text{ bottleneck links})$ would be proportional to the square of the total number of links, and thus be $O(n^4)$. However, because of the independences in the IID network, the variance is in fact much lower.

Proof. First consider the unconditional version. We have

$$\mathrm{Var}\left(\sum_{i \neq j} B_{ij}\right) = \sum_{i \neq j}\sum_{k \neq l} \mathrm{Cov}(B_{ij}, B_{kl}).$$

The important observation is that for i, j, k, l all distinct, B_{ij} and B_{kl} are independent giving $\mathrm{Cov}(B_{ij}, B_{kl}) = 0$. (This is because they depend only on

the position of distinct and independently-positioned nodes.) Hence there are only $O(n^3)$ non-zero terms in the sum. Each non-zero covariance term in the sum is bounded by variance of the indicator function, so

$$\operatorname{Cov}(B_{ij}, B_{kl}) \leq \operatorname{Var} B_{ij} = \beta(1-\beta) \leq \frac{1}{4}.$$

But by (4.6), if $\operatorname{Cov}(B_{ij}, B_{kl}) = 0$, then the conditional covariance is very small $\operatorname{Cov}_n(B_{ij}, B_{kl}) = O(n^{-1})$. Hence,

$$\operatorname{Var}_n\left(\sum_{i\neq j} B_{ij}\right) \leq O(n^3)\frac{1}{4} + O(n^4)O(n^{-1}) = O(n^3),$$

as desired. □

4.5.3 Completing the proof

We are now in a position to prove (4.3), and hence prove Theorem 4.3.

Proof. We need to show

$$\forall \epsilon > 0 \quad \forall \delta > 0 \quad \exists N \quad \forall n \geq N \quad \mathbb{P}\left(\frac{C_\Sigma}{n} \geq E + \epsilon\right) \leq \delta.$$

So choose $\epsilon > 0$, $\delta > 0$, fix $n \geq N$ (where N will be determined later), and pick a fixed rate vector $\mathbf{r} \in \mathbb{R}_+^n$ with sum-rate

$$\frac{r_\Sigma}{n} > E + \epsilon; \tag{4.7}$$

we need to show that $\mathbb{P}(\mathbf{r} \text{ is achievable}) \leq \delta$. (Here, we are writing $r_\Sigma := \sum_{i=1}^n r_i$ for the sum-rate.)

We divide into two cases: when there is a very high SNR, which is unlikely to happen; and when there is not, in which case $\mathbf{r}$ is unlikely to be achievable. Formally,

$$\begin{aligned}
&\mathbb{P}(\mathbf{r} \text{ achievable}) \\
&\quad = \mathbb{P}(\mathbf{r} \text{ achievable} \mid \max S_{ii} \leq n^{\eta/2})\mathbb{P}(\max S_{ii} \leq n^{\eta/2}) \\
&\qquad\qquad + \mathbb{P}(\mathbf{r} \text{ achievable} \mid \max S_{ii} > n^{\eta/2})\mathbb{P}(\max S_{ii} > n^{\eta/2}) \\
&\quad \leq \mathbb{P}(\mathbf{r} \text{ achievable} \mid \max S_{ii} \leq n^{\eta/2}) + \mathbb{P}(\max S_{ii} > n^{\eta/2}) \\
&\quad \leq \mathbb{P}_n(\mathbf{r} \text{ achievable}) + \frac{\delta}{2},
\end{aligned} \tag{4.8}$$

for n sufficiently large, by Lemma 4.12. We need to bound the first term in (4.8).

First, note that our assumption on $\max_i S_{ii}$ means that if $r_i > 2n^{\eta/2}$, then we break the single-user capacity bound, since we would have

$$r_i > 2n^{\eta/2} \geq 2\max_j S_{jj} \geq 2S_{ii} = \log(1 + 2\mathtt{SNR}_i) > \log(1 + \mathtt{SNR}_i)$$

meaning $\mathbf{r}$ is not achievable, and we are done. Thus we assume this does not hold; that

$$r_i \leq 2n^{\eta/2} \quad \text{for all } i. \tag{4.9}$$

(The rest of our argument closely follows Jafar [27, proof of Theorem 5].)

Now, if $\mathbf{r}$ is achievable, it must at least satisfy the constraints on the ϵ-bottleneck links $i \to j$ from Lemma 4.11, and hence also the sum of those constraints. So

$$\begin{aligned}
\mathbb{P}_n(\mathbf{r} \text{ achievable}) &\leq \mathbb{P}_n(r_i + r_j \leq 2E + \epsilon \text{ on bottleneck links } i \to j) \\
&\leq \mathbb{P}_n\left(\sum_{i \neq j} B_{ij}(r_i + r_j) \leq \left(\sum_{i \neq j} B_{ij}\right)(2E + \epsilon)\right) \\
&= \mathbb{P}_n(U \leq V),
\end{aligned} \tag{4.10}$$

where we have defined

$$\begin{aligned}
U &:= \frac{1}{n(n-1)} \sum_{i \neq j} B_{ij}(r_i + r_j), \\
V &:= \frac{1}{n(n-1)} \left(\sum_{i \neq j} B_{ij}\right)(2E + \epsilon).
\end{aligned}$$

The conditional expectations of U and V are

$$\mathbb{E}_n U = 2\beta_n \frac{r_\Sigma}{n}, \quad \mathbb{E}_n V = \beta_n(2E + \epsilon) = 2\beta_n \left(E + \frac{\epsilon}{2}\right).$$

Note that since $\beta_n > 0$ by Lemma 6, we can rewrite (4.7) as

$$\mathbb{E}_n U > \mathbb{E}_n V + \beta_n \epsilon,$$

or equivalently,

$$\mathbb{E}_n U - \frac{\beta_n \epsilon}{2} > \mathbb{E}_n V + \frac{\beta_n \epsilon}{2}.$$

The proof is completed by formalising the following idea: since the expectations are ordered $\mathbb{E}_n U > \mathbb{E}_n V$, we can only rarely have the opposite ordering $U < V$. Hence the expression in (4.10) is small.

Formally, by (the conditional version of) Chebyshev's inequality and the union bound, we have

$$\begin{aligned}
\mathbb{P}_n(U \leq V) &\leq \mathbb{P}_n\left(U \leq \mathbb{E}_n U - \frac{\beta_n \epsilon}{2} \text{ or } V \geq \mathbb{E}_n V + \frac{\beta_n \epsilon}{2}\right) \\
&\leq \mathbb{P}_n\left(|U - \mathbb{E}_n U| \geq \frac{\beta_n \epsilon}{2}\right) + \mathbb{P}_n\left(|V - \mathbb{E}_n V| \geq \frac{\beta_n \epsilon}{2}\right) \\
&\leq \left(\frac{2}{\beta_n \epsilon}\right)^2 \operatorname{Var}_n U + \left(\frac{2}{\beta_n \epsilon}\right)^2 \operatorname{Var}_n V \\
&= \frac{4}{\beta_n^2 \epsilon^2}(\operatorname{Var}_n U + \operatorname{Var}_n V).
\end{aligned} \tag{4.11}$$

Using Lemma 4.14 we can bound these variances as

$$\begin{aligned}\operatorname{Var}_n U &= \frac{1}{n^2(n-1)^2} \operatorname{Var}_n \left(\sum_{i \neq j} B_{ij}(r_i + r_j) \right) \\ &\leq \frac{1}{n^2(n-1)^2} O(n^3) 16 n^{\eta} \\ &= O(n^{-(1-\eta)}). \\ \operatorname{Var}_n V &= \frac{1}{n^2(n-1)^2} O(n^3)(2E+\epsilon)^2 = O(n^{-1}),\end{aligned}$$

where we used (4.9) to bound $\operatorname{Var}_n U$. Choosing η to be less than 1, we can ensure N is sufficiently large that for all $n \geq N$,

$$\operatorname{Var}_n U + \operatorname{Var}_n V \leq \frac{\beta_n^2 \delta \epsilon^2}{8}.$$

This makes (4.11) into $\mathbb{P}_n(U \leq V) < \delta/2$. Together with (4.10) and (4.8), this yields the result. □

4.6 Conclusion

In this chapter we have defined IID interference networks with power-law attenuation. We have shown that this setup fulfills necessary properties for the average per-user capacity C_Σ / n to tend in probability to $\frac{1}{2}\mathbb{E}\log(1 + 2\mathtt{SNR})$. We have also noted that this result is not unique to our setup. We briefly mention one more example.

Suppose, for example, that Rayleigh fading is added to our model. That is, now let $\mathtt{SNR}_i := |H_{ii}|^2 a(\|\mathbf{T}_i - \mathbf{R}_i\|)$ and $\mathtt{INR}_{ji} := |H_{ji}|^2 a(\|\mathbf{T}_i - \mathbf{R}_j\|)$, where the H_{ji} are IID standard complex Gaussian random variables.

Because ergodic interference alignment still works with Rayleigh fading [44, Section IV], the direct part of the theorem still holds. But also, because the fading coefficients are IID, the independence structure from the non-fading case remains, ensuring Lemmas 4.12–4.14 hold. Hence, the theorem is still true.

Characterising *all* networks for which such a limit for average per-user capacity exists is an open problem.

At the moment, Theorem 4.3 should perhaps be regarded as being of theoretical interest. That is, our major contribution is to provide a sharp upper bound on the performance of interference networks. However, the lower bound relies on an ergodic interference alignment which, while rigorously proved, may not be feasible to implement in practice for large number of users. Examination of the proof of the effectiveness of ergodic interference alignment [44, Theorem 1] shows that, even for a model with alphabet size q, the channel needs to be used $O((q-1)^{n^2})$ times. Even for $n \sim 10$, this is a prohibitive requirement. We approach this problem in the next chapter.

Notes

The new work in this chapter – specifically Theorem 4.3 and its proof, and the definition of the IID network – is joint work with Oliver Johnson and Robert Piechocki. This chapter is based on two papers we wrote [1, 2]. This work benefited from the advice of anonymous reviewers for from *IEEE Transactions on Information Theory* and the 2010 IEEE International Symposium on Information Theory.

The ideas in this chapter were first studied by Jafar [27] – in particular, the concept (although not the exact definition) of bottleneck links comes from that paper.

The earlier of our two papers [1] (which actually appeared in publication later) considered only the standard dense network, but defined many of the important concepts in this chapter. The full general proof was first published in the later of our two papers [2] (which appeared earlier).

With the exception of our new proof of Theorem 4.1 from our earlier paper [1], the material on Jafar networks (Section 4.3) is due to Jafar.

5

Delay–rate tradeoff for ergodic interference alignment

5.1 Introduction

Earlier, in Section 2.8, we discussed the ergodic interference alignment scheme of Nazer, Gastpar, Jafar, and Viswanath (which we hereafter to refer to as the NGJV scheme).

Recall that we considered a model of communication over a finite field $\mathbb{F}_q$ of size q. Since the NGJV scheme (see Section 2.8) requires a particular $n \times n$ channel matrix with entries in $\mathbb{F}_q \setminus \{0\}$ to occur, the expected delay for a particular message is $(q-1)^{n^2}$ (which is roughly q^{n^2} for large q). It is clear that even for n and q relatively small, this is not a practical delay. (For $n = 6$ and $q = 3$, for example, the delay is $2^{36} \approx 7 \times 10^{10}$.)

There are five questions we would like to try to answer:

1. Can we find a scheme that, like NGJV, achieves half the single-user rate, but at a lower time delay?

2. Can we find schemes that have lower time delays than NGJV, even at some cost to the rate achieved?

3. Specifically, which schemes from Question 2 perform well for situations where we have few users (n small)?

4. Specifically, which schemes from Question 2 perform well for situations where we have many users ($n \to \infty$)?

5. What is a lower bound on the best time delay possible for any scheme achieving a given rate for a given number of users?

In Sections 5.3 and 5.4, we define a new set of schemes, called JAP (Subsection 5.3.1), a beamforming extension JAP-B (Subsection 5.3.4), and child schemes derived from them (Section 5.4) that have lower time delays than the NGJV scheme, for a variety of different rates, answering Question 2. As a special case, examined in Subsection 5.3.5, the JAP-B($[n]$) schemes achieve half the single-user rate, like NGJV, while reducing the time delay from q^{n^2} to $q^{(n-1)(n-2)}$, answering Question 1. In Section 5.5, we answer Questions 3 and 4, by finding and analysing the JAP schemes that perform the best for small and large n; the table on page 109 and the graphs on page 110 illustrate the best schemes for small n, and Theorems 5.6 and 5.7 give the asymptotic behaviour of the schemes. Question 5 remains an open problem (although we do give a lower bound on the delay achievable for the schemes listed above).

Koo, Wu, and Gill [65] have previously attempted to answer Questions 2 and 3. We briefly outline their work at the end of Section 5.2.

5.2 Model

We give our results in the context of the finite-field interference network with fast-fading.

Recall that the single-user capacity of this channel is $\log q - \mathbb{H}(Z) =: \mathbb{D}(Z)$. Extending our previous definition of degrees of freedom (Definition 1.13) to multiple users, we have the following:

Definition 5.1. Given an achievable symmetric rate point $(r, r, \ldots, r)$, we define the *symmetric per-user degrees of freedom* to be $\mathtt{dof} = r/\mathbb{D}(Z)$.

In particular, it's clear that a single user can achieve 1 degree of freedom.

We define the expected time delay for the NGJV scheme to be the average number of time slots we must wait after seeing a channel matrix H until we see the corresponding matrix $\mathsf{I} - \mathsf{H}$. The time delay is geometrically distributed with parameter p, where p is the probability that the random channel matrix takes the value $\mathsf{I} - \mathsf{H}$. The mean of this random variable is $1/p$; hence the problem of finding the average time delay is reduced to a problem of finding the probability that a desired matrix appears in the next time slot. Since a channel matrix has n^2 entries, each of which needs to take the correct one value of $q - 1$ possible values, the average time delay is

$$D = \frac{1}{\left(\frac{1}{q-1}\right)^{n^2}} = (q-1)^{n^2} \sim q^{n^2}. \tag{5.1}$$

(Here and elsewhere, we write $f(q) \sim g(q)$ if $f(q)/g(q) \to 1$ as $q \to \infty$.)

As we mentioned before, this expected delay will be quite large even for modest values of q and n. For this reason, we will concentrate on the delay exponent.

Definition 5.2. An interference alignment scheme with expected delay $D \sim kq^T$ for some k and T has *delay exponent* T and *delay coefficient* k. More specifically, we have $T := \lim_{q\to\infty} \log D / \log q$.

We regard reduction of the delay exponent as the key aim, with the delay coefficient playing a secondary role. In particular, the finite field model is in some sense an abstraction of the model where channel coefficients are Gaussians quantized into a set of size q, where q is chosen large enough to reduce quantization error. When q is large, the delay exponent T dominates the delay coefficient k in determining size of the expected delay D.

From Theorem 2.14, we know that the NGJV schemes achieve $\mathtt{dof} = 1/2$, and we have just shown in (5.1) it requires a delay exponent of n^2.

We also mention some new schemes outlined in a recent paper by Koo, Wu, and Gill [65]. They attempted to answer our Questions 2 and 3, by finding schemes – we call them KWG schemes – with lower delay than the NGJV scheme. The KWG schemes suggest matching a larger class of matrices than simply H and $\mathsf{I} - \mathsf{H}$. By analysing the hitting probability of an associated Markov chain, they were able to reduce the expected delay, at the cost of a reduction in rate (and hence degrees of freedom). However, their schemes only affect the delay by a constant multiple, with the shortest-delay scheme only reducing the delay to $0.64(q-1)^{n^2} \sim 0.64q^{n^2}$ with a sum-rate of $0.791\mathbb{D}(Z)$ [65, page 5]. That is, the KWG schemes only reduce the delay coefficient k, leaving the delay exponent as $T = n^2$. For modest q and n (say $q = 3$, $n = 6$, again), we regard this delay as still impractical. Since the KWG schemes achieve a lower rate than the NGJV scheme for the same delay exponent, we shall only compare our results with the NGJV scheme.

5.3 New alignment schemes: JAP and JAP-B

5.3.1 Three important observations

In the NGJV scheme, all receivers were able to decode their message by summing their two pseudomessages

$$\sum_{i=1}^{n} h_{ji}[t_0]\mathbf{m}_i + \sum_{i=1}^{n} h_{ji}[t_1]\mathbf{m}_i = \mathbf{m}_j \qquad \text{for } j = 1, \ldots, n.$$

In other words, the NGJV scheme relies on the linear dependence

$$\mathsf{H}[t_0] + \mathsf{H}[t_1] = \mathsf{I}.$$

This scheme has a large delay, because, given $\mathsf{H}[t_0]$, there is only one matrix, $\mathsf{H}[t_1] = \mathsf{I} - \mathsf{H}[t_0]$, that can complete the linear dependence. If there were a large collection of matrices that could complete the dependence, then the delay would be lower.

We make three observations to this end.

First, while NGJV matches two channel states $\mathsf{H}[t_0]$ and $\mathsf{H}[t_1]$ to form this linear dependence, we could use more than two. That is, if we have $K+1$ channel matrices $\mathsf{H}[t_0], \mathsf{H}[t_1], \ldots, \mathsf{H}[t_K]$ such that

$$\mathsf{H}[t_0] + \mathsf{H}[t_1] + \cdots + \mathsf{H}[t_K] = \mathsf{I},$$

than receivers j can sum the $K+1$ pseudomessages to recover their message,

$$\sum_{i=1}^{n} h_{ji}[t_0]\mathbf{m}_i + \sum_{i=1}^{n} h_{ji}[t_1]\mathbf{m}_i + \cdots + \sum_{i=1}^{n} h_{ji}[t_K]\mathbf{m}_i = \mathbf{m}_j.$$

Note that the transmission of a single message is now split among $K+1$ channel states, rather than 2 as in NGJV. This means that the degrees of freedom of this scheme is reduced to $1/(K+1)$ from NGJV's $1/2$.

Second, any linear combination of channel state matrices that sums to I is sufficient. That is, if there exist scalars $\lambda_0, \lambda_1 \in \mathbb{F}_q$ such that

$$\lambda_0 \mathsf{H}[t_0] + \lambda_1 \mathsf{H}[t_1] = \mathsf{I},$$

then all receivers can recover their message by forming the linear combination of pseudocodewords

$$\lambda_0 \sum_{i=1}^{n} h_{ji}[t_0]\mathbf{m}_i + \lambda_1 \sum_{i=1}^{n} h_{ji}[t_1]\mathbf{m}_i = \mathbf{m}_j \qquad \text{for } j = 1, \ldots, n.$$

Third, NGJV requires all receivers to be able to decode their messages at the same time. However, receiver j can decode its message if

$$\sum_{i=1}^{n} h_{ji}[t_0]\mathbf{m}_i + \sum_{i=1}^{n} h_{ji}[t_1]\mathbf{m}_i = \mathbf{m}_j$$

regardless of whether this equality holds for other receivers as well. In other words, receiver j can decode its message if

$$\begin{aligned} h_{jj}[t_0] + h_{jj}[t_1] &= 1 \\ h_{ji}[t_0] + h_{ji}[t_1] &= 0 \qquad \text{for } i \neq j. \end{aligned}$$

Putting these three observations together, we make the following conclusion: Let $\mathsf{H}[t_0], \mathsf{H}[t_1], \ldots, \mathsf{H}[t_K]$ be a sequence of $K+1$ channel state matrices. If there exist scalars $\lambda_0, \lambda_1, \ldots, \lambda_K$ such that for some j

$$\lambda_0 h_{jj}[t_0] + \lambda_1 h_{jj}[t_1] + \cdots + \lambda_K h_{jj}[t_K] = 1 \tag{5.2}$$

$$\lambda_0 h_{ji}[t_0] + \lambda_1 h_{ji}[t_1] + \cdots + \lambda_K h_{ji}[t_K] = 0 \qquad i \neq j, \tag{5.3}$$

then receiver j can recover its message by forming the linear combination of pseudocodewords

$$\lambda_0 \sum_{i=1}^{n} h_{ji}[t_0]\mathbf{m}_i + \cdots + \lambda_K \sum_{i=1}^{n} h_{ji}[t_K]\mathbf{m}_i = \mathbf{m}_j.$$

In fact, we only require

$$\lambda_0 h_{jj}[t_0] + \lambda_1 h_{jj}[t_1] + \cdots + \lambda_K h_{jj}[t_K] \neq 0$$
$$\lambda_0 h_{ji}[t_0] + \lambda_1 h_{ji}[t_1] + \cdots + \lambda_K h_{ji}[t_K] = 0 \qquad \text{for } i \neq j.$$

since the coefficients λ_k can be rescaled to make the top equation (5.2) equal to 1 without breaking the bottom equation (5.3).

Or, writing $\mathbf{h}_j^{\text{int}}$ for the interference vector

$$\mathbf{h}_j^{\text{int}} = (h_{j1}, \ldots, h_{jj-1}, h_{jj+1}, \ldots, h_{jn}),$$

we can again rewrite the requirement as

$$\lambda_0 h_{jj}[t_0] + \lambda_1 h_{jj}[t_1] + \cdots + \lambda_K h_{jj}[t_K] \neq 0 \tag{5.4}$$
$$\lambda_0 \mathbf{h}_j^{\text{int}}[t_0] + \lambda_1 \mathbf{h}_j^{\text{int}}[t_1] + \cdots + \lambda_K \mathbf{h}_j^{\text{int}}[t_K] = \mathbf{0}. \tag{5.5}$$

If n equalities like the (5.4) and (5.5) above hold, we say that "receiver j can recover its message from $\mathsf{H}[t_0]$, $\mathsf{H}[t_1], \ldots, \mathsf{H}[t_K]$."

The time delay of this scheme is $t_K - t_0$.

Recall that the average delay is the reciprocal of the probability that a random matrix allows a receiver to recover its message. Thus it will be useful to note the following lemma.

Lemma 5.3. *Conditional on the interference vectors $\mathbf{H}_j^{int}[t_0], \ldots, \mathbf{H}_j^{int}[t_K]$ being linearly dependent, the probability that receiver j can recover its message is $1 - O(q^{-1})$.*

Note that we only use a matrix when we know for certain that it fulfills our desired criteria – we merely need to know what the probability is that the next matrix will do, in order to calculate the expected delay.

Proof. Since the interference vectors are linearly dependent, there exists a linear combination

$$\lambda_0 \mathbf{H}_j^{\text{int}}[t_0] + \lambda_1 \mathbf{H}_j^{\text{int}}[t_1] + \cdots + \lambda_K \mathbf{H}_j^{\text{int}}[t_K] = \mathbf{0}$$

where $L > 0$ of the λ_k are nonzero. Thus, receiver j can recover its message provided that the corresponding linear combination

$$\lambda_0 H_{jj}[t_0] + \lambda_1 H_{jj}[t_1] + \cdots + \lambda_K H_{jj}[t_K] \tag{5.6}$$

is nonzero; call the probability that this happens p.

When $\lambda_k \neq 0$, then $\lambda_k H_{jj}[t_k] =: V_k$ is uniform on $\mathbb{F}_q \setminus \{0\}$, and when $\lambda_k = 0$, then $\lambda_k H_{jj}[t_k] = 0$ too. So (5.6) is the sum of L random variables V_k IID uniform on $\mathbb{F}_q \setminus \{0\}$. We can write the mass function of each V_k as $(1+\rho)U - \rho\delta_0$, where U is uniform on $\mathbb{F}_q$, δ_0 is a point mass on 0, and $\rho = 1/(q-1)$. Then the mass function of the L-fold convolution is

$$(1 - (-\rho)^L)U + (-\rho)^L \delta_0.$$

Hence, the probability that (5.6) is zero is

$$1 - p = (1 - (-\rho)^L)\frac{1}{q} + (-\rho)^L = \frac{1}{q} + \frac{1}{q(q-1)^{L-1}} = O(q^{-1}).$$

The result follows. □

5.3.2 The scheme JAP(**a**)

We now present our new scheme.

The idea behind the scheme is as follows: We start by seeing some channel state $\mathsf{H}[t_0]$. We then set t_1 to be the first time slot that allows receivers 1 to a_1 to recover their message (where a_1 is decided on in advance). Next, we set t_2 to be the first time slot that allows receivers the next a_2 receivers to recover their message. And so on, until all n receivers have recovered their message.

Specifically, fix $K \leq n$ and a sequence $[a_1, a_2, \dots, a_K] =: \mathbf{a}$ of length K and weight n; that is, in the set

$$\mathcal{A}(n, K) := \left\{ \mathbf{a} \in \mathbb{Z}_+^K : \sum_{k=1}^{K} a_k = n \right\}.$$

We write A_k for the partial sums $A_k := a_1 + a_2 + \cdots + a_k$ (so in particular $A_1 = a_1$ and $A_K = n$).

Then we define the scheme JAP($\mathbf{a}$) as consisting of the following $K + 1$ steps:

Step 0: Start with a matrix $\mathsf{H}[t_0]$.

Step 1: Set t_1 to be the first timeslot that allows the first a_1 receivers $1, 2, \dots, A_1$ to recover their message from $\mathsf{H}[t_0], \mathsf{H}[t_1]$.

⋮

Step *k*: Set t_k to be the first timeslot that allows the next a_k receivers $A_{k-1} + 1, A_{k-1} + 2, \dots, A_k$ to recover their message from $\mathsf{H}[t_0], \mathsf{H}[t_1], \dots, \mathsf{H}[t_k]$.

⋮

Step *K*: Set t_K to be the first timeslot that allows the final a_K receivers $A_{K-1} + 1, A_{K-1} + 2, \dots, A_K$ to recover their message from $\mathsf{H}[t_0], \mathsf{H}[t_1], \dots, \mathsf{H}[t_K]$.

By the end of this process, all $n = A_K$ receivers have recovered their message.

Since the message was split over $K + 1$ time slots, the common rate of communication is $\mathbb{D}(Z)/(K + 1)$, which corresponds to $\mathtt{dof} = 1/(K + 1)$.

5.3.3 Delay exponent of JAP schemes

We now examine the delay exponent for our new schemes.

Theorem 5.4. *Consider the n-user finite field interference network. Fix K and* $\mathbf{a} \in \mathcal{A}(n, K)$. *We use the scheme JAP*$(\mathbf{a})$ *as outlined above. Then*

1. *the expected time for the kth round to take place is* $D \sim q^{T_k(\mathbf{a})}$, *where* $T_k(\mathbf{a}) = a_k(n-k-1)$;
2. *the delay exponent for the whole scheme is*

$$T(\mathbf{a}) := \max_{1 \le k \le K} T_k(\mathbf{a}) = \max_{1 \le k \le K} a_k(n-k-1).$$

Proof. Recall that the expected delay is the reciprocal of the probability the desired match can be made.

Suppose we are about to begin stage k of a scheme JAP$(\mathbf{a})$. By Lemma 5.3, the probability we can complete the stage is $1 - O(q^{-1})$ multiplied by the probability that the interference vectors for the next a_k receivers

$$\mathbf{H}_j^{\text{int}}[t_0], \mathbf{H}_j^{\text{int}}[t_1], \dots, \mathbf{H}_j^{\text{int}}[t_k]$$

are linearly dependent.

If the first $k-1$ interference vectors are already linearly dependent, then we are done (with high probability, by Lemma 5.3). Assume they are not.

Write $\mathcal{S}$ for the span of the first $k-1$ interference vectors for one of the desired a_k receivers j, so

$$\mathcal{S} := \operatorname{span}\{\mathbf{H}_j^{\text{int}}[t_0], \dots, \mathbf{H}_j^{\text{int}}[t_{k-1}]\}.$$

Since all possible interference vectors in $(\mathbb{F}_q \setminus \{0\})^{n-1}$ are equally likely, the probability that the next matrix completes a linear dependence is

$$\frac{|\mathcal{S} \cap (\mathbb{F}_q \setminus \{0\})^k|}{|(\mathbb{F}_q \setminus \{0\})^{n-1}|} = \frac{q^k s}{(q-1)^{n-1}},$$

where s is the proportion of vectors in $\mathcal{S}$ with no zero entries. By counting the possible coefficients in $\mathbb{F}_q$ used in the span, the inclusion–exclusion formula gives us

$$s = 1 - (K-1)\frac{1}{q} + O\left(\frac{1}{q^2}\right) = 1 - O(q^{-1}).$$

Hence, the desired probability is

$$\frac{q^k s}{(q-1)^{n-1}}\left(1 - O(q^{-1})\right) = \frac{q^k\left(1 - O(q^{-1})\right)}{(q-1)^{n-1}}\left(1 - O(q^{-1})\right) \sim q^{-(n-k-1)}$$

(where the $1 - O(q^{-1})$ term comes from Lemma 5.3).

This property that a linear dependence is completed must hold for all a_k receivers, which happens with probability $(q^{-(n-k-1)})^{a_k} = q^{-a_k(n-k-1)}$, hence the first result.

For the second result, note that, as $q \to \infty$, the delay is dominated by the delay for the slowest round. □

5.3.4 Improving delay with beamforming: JAP-B

Beamforming slightly improves the performance of JAP($\mathbf{a}$) schemes, combining ideas from the original Cadambe–Jafar interference alignment [28] with the JAP scheme.

In round k we can guarantee that the interference matches up for receiver $l := A_{k-1} + 1$. Each transmitter i, instead of repeating their message w_i, rather encodes $(h_{li}[t_k]])^{-1} h_{li}[t_0]\mathbf{m}_i$. (Since the coefficient h_{li} cannot be 0 and q is prime, the inverse term certainly exists.) The total received interferences at receiver l at times t_0 and t_k are both equal to $\sum_{i \neq l} h_{li}[t_0]\mathbf{m}_i$, so can be estimated and cancelled.

We refer to such schemes that take advantage of beamforming as JAP-B($\mathbf{a}$) schemes.

Theorem 5.5. *The delay exponent of a JAP-B($\mathbf{a}$) scheme indexed by sequence $\mathbf{a}$ is*

$$T_B(\mathbf{a}) := \max_{1 \leq k \leq K} (a_k - 1)(n - k - 1).$$

Proof. At each round, receiver $l = A_{k-1} + 1$ will automatically be able to recover its message, leaving the JAP scheme to align interference for the other $a_k - 1$ users. (Independence of the coefficients h_{ji} ensures that the scheme still has the same problem to solve.) □

In particular, the JAP-B scheme will always outperform the JAP scheme with the same sequence $\mathbf{a}$.

5.3.5 An interesting special case: JAP-B($[n]$)

An interesting special case of the JAP-B schemes is the case when $K = 1$ and $a_1 = n$; we call this scheme JAP-B($[n]$).

In this case we have $1/(K+1) = 1/2$ degrees of freedom for a rate of $\mathbb{D}(Z)/2$. From Theorem 5, we see that the delay exponent is

$$T_B([n]) = (a_1 - 1)(n - 1 - 1) = (n-1)(n-2).$$

Effectively, the JAP-B($[n]$) scheme works by using beamforming to automatically cancel transmitter 1's interference, then for users $2, 3, \ldots, n$ requiring the existence of diagonal matrices $\mathsf{D}_0, \mathsf{D}_1$ such that $\mathsf{D}_0\mathsf{H}[t_0] + \mathsf{D}_1\mathsf{H}[t_1] = \mathsf{I}$.

Note that this is the same rate as is achieved by the original NGJV scheme, but that the delay exponent has been reduced from NGJV's n^2 to

$$(n-1)(n-2) = n^2 - (3n-2).$$

For small n in particular, this is a worthwhile improvement (see figure, p. 110).

5.4 Child schemes: using time-sharing

Another way to generate new alignment schemes is by time-sharing schemes designed for a smaller number of users.

Call the NGJV, KWG, JAP and JAP-B schemes 'parent schemes'. Given a parent scheme for the m-user network, we can modify for the any n-user network with $n > m$, giving what we call a 'child scheme'.

Specifically, we use resource division by time (see Subsection 2.6.1) to split the network into $\binom{n}{m}$ subnetworks, each of which contains a unique collection of just $m < n$ of the users. Within each of these m-user subnetworks, a parent scheme is used, while the other $n - m$ transmitters remain silent.

Resource division by time is often known as *time-division multiple access* or TDMA – we use that abbreviation in the rest of this chapter.

Such a child scheme clearly has the same delay exponent as the parent scheme, with the rate – and thus the degrees of freedom – reduced by a factor of m/n. So an m-user JAP-B scheme shared between n users gives $\texttt{dof} = m/n(K+1)$.

In particular, time-sharing the NGJV schemes for smaller networks gives a collection of schemes with a lower delay exponent $m^2 < n^2$ than the main NGJV scheme for a given number of users, reducing the degrees of freedom from $1/2$ to $m/2n$.

(We are not aware that the idea of time-sharing NGJV schemes has previously appeared in the literature. However, the idea seems simple enough that we regard this as the 'current benchmark' against which we should compare our new schemes.)

Interestingly, it seems that child schemes derived from time-sharing an NGJV-like JAP-B($[n]$) parent scheme are particularly effective, and very often performs better than other JAP-B schemes. We discuss this point further in the next section.

5.5 Best schemes

5.5.1 General case

Given a number of users n and a desired number of degrees of freedom $\texttt{dof} = 1/(K+1)$, we wish to find a scheme with the lowest delay exponent.

For $K = n-1$ or n, when $\mathtt{dof} = 1/n$ or $1/(n+1)$, the best JAP-B schemes have delay exponent $T_B([1,\dots,1,2]) = T_B([1,\dots,1,1,1]) = 0$. This is the same delay exponent as TDMA, which has $\mathtt{dof} = 1/n$ also. Thus we need not consider schemes with $K = n-1$ or n.

For $K \le n-2$ the best parent scheme will be a JAP-B scheme with parameter vector $\mathbf{a} \in \mathcal{A}(n,K)$. We write $T(n,K)$ for this best delay exponent, that is

$$T(n,K) := \min_{\mathbf{a}\in\mathcal{A}(n,K)} T_B(\mathbf{a}) = \min_{\mathbf{a}\in\mathcal{A}(n,K)} \max_{1\le k\le K} (a_k - 1)(n-k-1).$$

We can bound $T(n,K)$ as follows.

Theorem 5.6. *Fix n and $K \le n-2$. For $T(n,K)$ as defined above, we have the following bounds:*

$$\frac{n}{K}(n-2) - (2n-K-2) \le T(n,K) \le \frac{n}{K}(n-2)$$

The gap between the upper and lower bounds grows linearly with n.

The following lemma on partial harmonic sums will be useful.

Lemma 5.7. *Let $S(n,K)$ be the partial harmonic sum*

$$S(n,K) := \sum_{k=1}^{K} \frac{1}{n-k-1} = \frac{1}{n-2} + \cdots + \frac{1}{n-K-1}.$$

Then we have the bounds

$$\frac{K}{n-2} \le S(n,K) \le \frac{K}{n-K-2}.$$

Of course, tighter bounds are available by comparing $\sum 1/k$ to $\int 1/x \,\mathrm{d}x$, but this suffices for our needs.

Proof. There are K terms in the sum, the largest of which is $1/(n-K-2)$ and the smallest of which is $1/(n-2)$. □

We can now prove Theorem 5.6.

Proof of Theorem 5.6. The value of $T(n,K)$ is lower-bounded by the value of the same minimisation problem relaxed to allow the a_k to be real. That is,

$$\begin{aligned} T(n,K) &= \min_{\mathbf{a}\in\mathbb{Z}_+^K : \sum_k a_k = n} \max_{1\le k\le K} (a_k - 1)(n-k-1) \\ &\ge \min_{\mathbf{a}\in\mathbb{R}_+^K : \sum_k a_k = n} \max_{1\le k\le K} (a_k - 1)(n-k-1). \end{aligned}$$

The relaxed problem is solved by waterfilling, setting $a_k - 1 = c/(n-k-1)$. Requiring the weight of $\mathbf{a}$ to be n forces

$$c = \frac{n-K}{S(n,K)} \ge \frac{(n-K)(n-K-2)}{K},$$

where we have used Lemma 5.7. Rearrangement gives the lower bound.

An upper bound is obtained by using the same c and taking

$$a_k - 1 = \left\lceil \frac{c}{n-k-1} \right\rceil \leq \frac{c}{n-k-1} + 1.$$

This gives

$$\begin{aligned} T_B(\mathbf{a}) &\leq c + \max_k (n-k-1) \\ &= \frac{n-K}{S(n,K)} + (n-2) \\ &\leq \frac{(n-K)(n-2)}{K} + (n-2), \end{aligned}$$

where we have used Lemma 5.7. Rearrangement gives the upper bound. □

5.5.2 Few users: small n

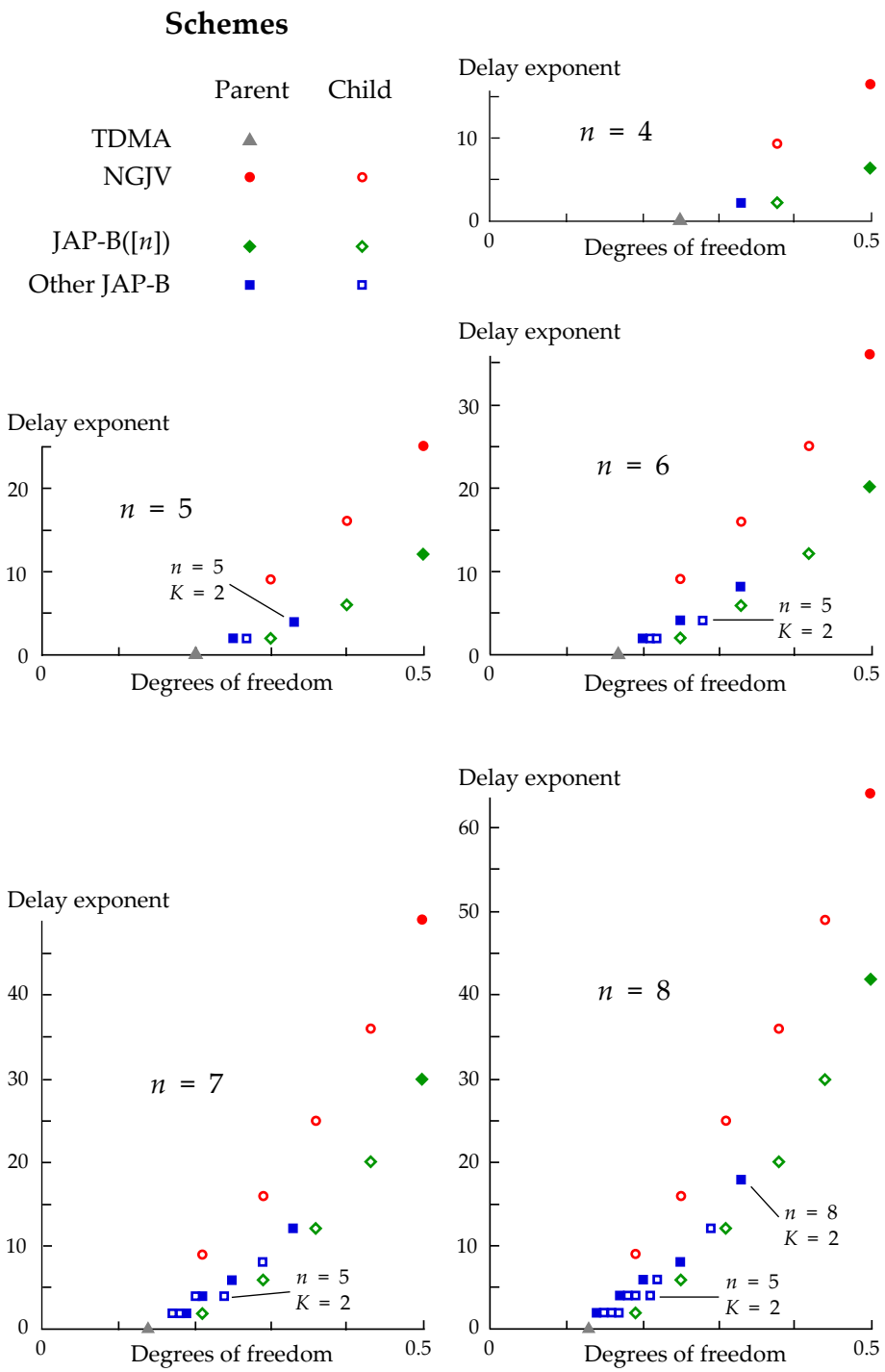
For small values of n, we can find the best parent JAP-B schemes by hand. (The task is simplified by noting that the optimal a_k will be nonzero and increasing in k.) The table below gives the delay exponents of the best JAP-B schemes for $n = 3, \ldots, 8$ and $K \leq n-2$.

Best JAP-B($\mathbf{a}$) schemes for small values of n and K, and their delay exponents.

	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
$K=1$	2	6	12	20	30	42
dof $=1/2$	[3]	[4]	[5]	[6]	[7]	[8]
$K=2$	0	2	4	8	12	18
dof $=1/3$	TDMA	[1,3]	[2,3]	[3,3]	[3,4]	[4,4]
$K=3$		0	2	4	6	8
dof $=1/4$		TDMA	[1,1,3]*	[1,2,3]*	[2,2,3]	[2,3,3]
$K=4$			0	2	4	6
dof $=1/6$			TDMA	[1,1,1,3]*	[1,1,2,3]*	[1,2,2,3]*
$K=5$				0	2	4
dof $=1/6$				TDMA	[1,1,1,1,3]*	[1,1,1,2,3]*
$K=6$					0	2
dof $=1/7$					TDMA	[1,1,1,1,1,3]*
$K=7$						0
dof $=1/8$						TDMA

*Asterisks mean that the choice of $\mathbf{a}$ achieving this delay exponent is non-unique.

Delay-rate tradeoff for TDMA, NGJV and JAP-B schemes for n-user interference networks



We can also consider child schemes based on parent JAP-B schemes. The figure on page 110 plots the performance of NGJV and all JAP-B schemes, as well as child schemes derived from them, for $n = 3, \ldots, 7$. Note that for many values of n and $\mathtt{dof}$, the scheme with the lowest delay exponent is JAP-B$([n])$ or one of the child schemes derived from it. (Note however, that the the parent schemes with $n = 5, K = 2$ and $n = 8, K = 2$, as well as child schemes derived from them, outperform JAP-B$([n])$ for some degrees of freedom.)

5.5.3 Many users: $n \to \infty$

We now consider the performance of schemes in the many-user limit $n \to \infty$.

In particular, we are interested in two limiting regimes, specifying how the degrees of freedom $\mathtt{dof}(n)$ should scale with the number of users n. In regime I the per-user rate is held constant; in regime II the sum-rate is kept constant, so each user's individual rate falls like $1/n$.

- **Regime I**, where we hold the degrees of freedom constant as $n \to \infty$. That is, we want to communicate at fixed fraction of the single-user rate, as in the NGJV scheme. In this regime I, we take $\mathtt{dof}(n) = \alpha$ for some $\alpha \in (0, 1/2]$. (The NGJV scheme corresponds to $\alpha = 1/2$.)
- **Regime II**, where we allow the degrees of freedom to fall as the number of users increases, scaling like $1/n$. That is, we want to communicate at a fixed multiple of the rate allowed by resource division schemes like TDMA. In regime II, we take $\mathtt{dof}(n) = \beta/n$ for some $\beta \geq 1$. (TDMA corresponds to $\beta = 1$.)

First, we consider how parent JAP-B schemes perform in the many-user limit.

Theorem 5.8. *For regimes I and II, as above, and as $n \to \infty$, we have the following results for the delay exponent $T(n)$ of parent JAP-B schemes:*

- ***Regime I:*** *Fix $\alpha \in (0, 1/2]$. Then the delay exponent for $\mathtt{dof}(n) = \alpha$ scales quadratically like*
$$T(n) \sim \frac{1}{\lfloor 1/\alpha \rfloor - 1} n^2.$$
- ***Regime II:*** *Fix $\beta > 1$. Then the delay exponent for $\mathtt{dof}(n) = \beta/n$ scales linearly, in that $T(n) = O(n)$, or more specifically,*
$$\left(\beta + \frac{1}{\beta} - 2\right) n - o(n) \leq T(n) \leq \beta n + o(n).$$

Proof. For regime I, we need $\mathtt{dof} = 1/(K+1) \geq \alpha$, so we take
$$K = \left\lfloor \frac{1}{\alpha} - 1 \right\rfloor = \left\lfloor \frac{1}{\alpha} \right\rfloor - 1.$$

But the general bounds on delay exponents from Theorem 5.6 tell us that for fixed K we have $T(n,K) \sim \frac{1}{K}n^2$. The result follows.

For regime II, $1/(K+1) = \mathtt{dof} = \beta/n$, so we need

$$K = \left\lfloor \frac{n}{\beta} - 1 \right\rfloor = \frac{n}{\beta} + O(1).$$

Hence

$$\frac{n}{K} = \frac{n}{n/\beta + O(1)} = \beta + o(1).$$

Putting this into the bounds from Theorem 5.6 gives

$$(\beta + o(1))(n-2) - \left(2n - \frac{n}{\beta} + O(1)\right) \leq T(n) \leq (\beta + o(1))(n-2).$$

Rearranging gives the result. □

Note that in regime I with $\alpha = 1/2$, we get $T(n) \sim n^2$, the same as NGJV.

We noted previously that child schemes produced by sharing the parent scheme JAP-B($\lfloor m \rfloor$) were particularly effective. The following theorem shows this.

Theorem 5.9. *For regimes I and II, as above, and as $n \to \infty$, we have the following results for the delay exponent $T(n)$ of child schemes based on JAP-B($\lfloor m \rfloor$) parent schemes:*

- ***Regime I:*** *Fix $\alpha \in (0, 1/2]$. Then the delay exponent for $\mathtt{dof}(n) = \alpha$ scales quadratically, in that*

$$T(n) = 4\alpha^2 n^2 - 6\alpha n + O(1) \sim 4\alpha^2 n^2.$$

- ***Regime II:*** *Fix $\beta > 1$. Then the delay exponent for $\mathtt{dof}(n) = \beta/n$ is constant, in that*

$$T(n) = (\lfloor 2\beta \rfloor - 1)(\lfloor 2\beta \rfloor - 2).$$

Proof. Recall from Section 5.4 that sharing the scheme JAP-B($\lfloor m \rfloor$) amongst n users gives $\mathtt{dof} = m/2n$ for delay exponent $T = (m-1)(m-2)$.

For regime I, note that $m/2n = \mathtt{dof}(n) = \alpha$, so we need to take $m = \lfloor 2\alpha n \rfloor$, giving $T(n) = (\lfloor 2\alpha n \rfloor - 1)(\lfloor 2\alpha n \rfloor - 2)$. The result follows.

For regime II, note that $m/2n = \mathtt{dof}(n) = \beta/n$, so we need to take $m = \lfloor 2\beta \rfloor$, giving $T(n) = (\lfloor 2\beta \rfloor - 1)(\lfloor 2\beta \rfloor - 2)$. □

Note that asymptotically, this means that in both regimes child schemes from JAP-B($\lfloor m \rfloor$) parent schemes are asymptotically more effective than any other parent scheme. This is because

$$4\alpha^2 n^2 \leq \frac{1}{\lfloor 1/\alpha \rfloor - 1} n^2$$

(with inequality unless $\alpha = 1/2$, when no child scheme will achieve the desired degrees of freedom) and any constant is less than $(\beta - 2)n$ for n sufficiently large.

Note also that by the same argument as the above proof, sharing the NGJV parent scheme gives $T(n) = 4\alpha^2 n^2$ in regime I, which is less good than sharing JAP-B($[m]$), but the same to first-order terms.

5.6 Conclusion

In the Secion 5.1, the questions we attempted to answer were:

1. Can we find a scheme that, like NGJV, achieves half the single-user rate, but at a lower time delay?
2. Can we find schemes that have lower time delays than NGJV, even at some cost to the rate achieved?
3. Specifically, which schemes from Question 2 perform well for situations where we have few users (n small)?
4. Specifically, which schemes from Question 2 perform well for situations where we have many users ($n \to \infty$)?
5. What is a lower bound on the best time delay possible for any scheme achieving a given rate for a given number of users?

In answer to question 2, we defined the new sets of parent schemes JAP and the even more effective JAP-B, and also derived child schemes from them. We noted that these had lower time delays – and sometimes significantly lower – at the costs of some loss in rate (or equivalently degrees of freedom). We saw that the child schemes from JAP-B($[n]$) schemes were often particularly effective.

In answer to question 1, we noted that the JAP-B($[n]$) schemes keep the degrees of freedom to $1/2$ while reducing the delay exponent from n^2 to $(n-1)(n-2) = n^2 - (3n-2)$.

In answer to Questions 3 and 4, we explicitly found the best schemes JAP-B schemes for $n \leq 8$, and analysed the asymptotic behaviour of our schemes as $n \to \infty$.

Question 5 remains an open problem.

Notes

This chapter is joint work with Oliver Johnson and Robert Piechocki, and is based on our paper [3].

The NJGV scheme is due to Nazer, Gastpar, Jafar, and Vishwanath [44].

The delay–rate tradeoff problem was first studied by Koo, Wu, and Gill [65].

6

Interference, group testing, and channel coding

6.1 Building the interference graph

In Section 2.6, we looked at resource division schemes such as resource division by time (Subsection 2.6.1). We noted that for a user to communicate without interference, it was necessary for all of the other users to stand idle. However, if not every receiver gets interference for every transmitter, than it might be possible for more than one user to communicate through the network at once.

For concreteness, consider the N-user finite field interference network with fixed fading, so

$$Y_{jt} = \sum_{i=1}^{N} h_{ji} x_{it} + Z_t \pmod{q}.$$

For the moment, so that we can concentrate on the interference, we will assume that the noise Z is 0 with probability 1, so

$$Y_{jt} = \sum_{i=1}^{N} h_{ji} x_{it} \pmod{q}.$$

If we have for some $i \neq j$ that $h_{ji} = h_{ij} = 0$, then both users i and j can communicate simultaneously and interference-free. (Recall that we use the word "user" to mean a transmitter–receiver pair. So we mean that if $h_{ji} = h_{ij} = 0$, then transmitter i can communicate to receiver i and simultaneously transmitter j can communicate to receiver j, both links without interference.)

More generally, we can build the *interference graph* to show which users interfere with which.

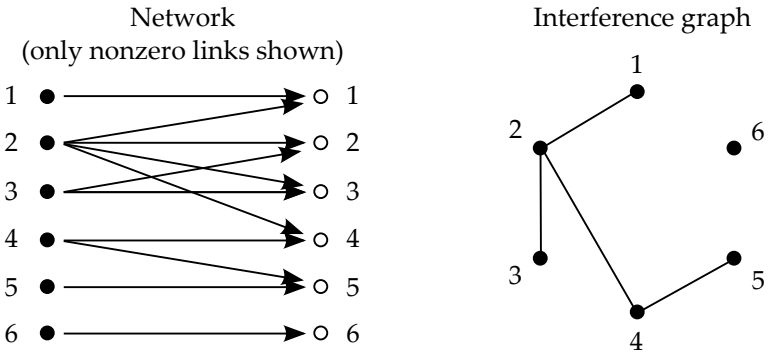
Definition 6.1. The *interference graph* of an N-user interference network with fixed fading has vertex set

$$\mathcal{V} := \{1, 2, \ldots, N\}$$

and edge set

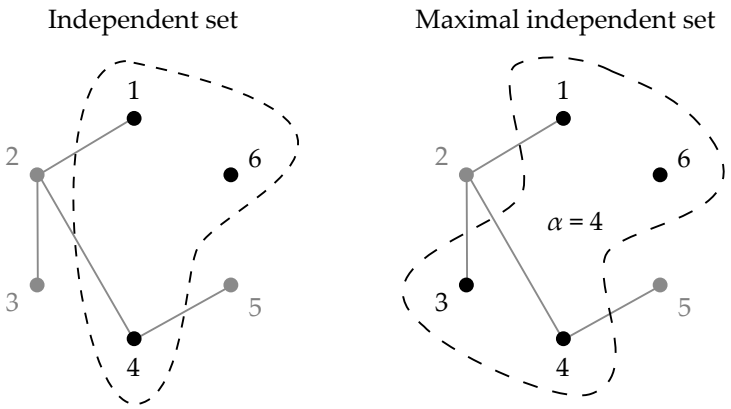
$$\mathcal{E} := \{ij : h_{ji} \neq 0 \text{ or } h_{ij} \neq 0\}.$$

The figure below shows the nonzero links in a network (transmitters on the left; receivers on the right), and the interference graph derived from it.



So two users i and j can communicate simultaneously and interference-free if they are not joined by an edge in the interference graph. More generally, any independent set of the interference graph can communicate at the same time. (Recall that a set of vertices $\mathcal{U} \subseteq \mathcal{V}$ is called *independent* if no edge in the graph joins one vertex in $\mathcal{U}$ to another.)

Optimal use of such a resource division strategy requires operating using only maximal independent sets. In particular, the maximum degrees of freedom achievable by a resource division scheme is $\mathtt{dof} = \alpha$, where the *interference number* α of the graph is the size of the largest independent set.



Given such a network, the users need to find out which other users they interfere with. Here is one method they could use to do so: Let each transmitter i choose a random nonzero message $m_i \in \mathbb{F}_q \setminus \{0\}$. Then, for each of T timeslots, each transmitter i either sends m_i or just sends the empty message 0. In other words, let $x_{it} = 1$ denote that transmitter i communicates in timeslot t, and $x_{it} = 0$ that she does not. Then each receiver j receives the signal

$$y_{jt} = \sum_{i=1}^{N} h_{ji} x_{it} m_i.$$

For large q, the probability that $y_{jt} = 0$ when at least one transmitter has $x_{it} = 1$ is small – for the moment we neglect this. Then if $y_{jt} = 0$, receiver j knows that $h_{ji} = 0$ for all i with $x_{it} = 1$; conversely, if $y_{jt} \neq 0$, receiver j knows that $h_{ji} \neq 0$ for at least one i with $x_{it} = 1$.

Receiver j wants to discover which h_{ji} are nonzero (but doesn't need to know their actual values) in as few tests T as possible – this T will depend on the number of users N, the number K of transmitters i for which $h_{ji} \neq 0$, and the acceptable error probability ϵ. This is equivalent to the problem of *group testing*, which we will outline more fully in the next section.

Our model would be more accurate if we included a noise term Z that wasn't always 0 and if we did not neglect the possibility that signals cancel each other out. Thus we would like our group testing protocols to be robust to this noise and to still have a small probability of error ϵ. In this chapter, we investigate how a channel coding approach to group testing can help with this.

Similar problems in multiuser networks have also been studied from a group testing perspective. Berger and coauthors [66] and Capetanakis [67] studied this problem using a model where more than one interfering message results in a collision where all messages are lost (rather than our model where signals are superposed at the receiver). Zhang, Luo, and Guo [68] studied the Gaussian network, where low-interference links $h_{ji} \approx 0$ are assumed to be zero and those signals are treated as noise.

In the rest of this chapter, we outline the problem of group testing, and explore a new approach to it using techniques from channel coding (as outlined in Chapter 1).

We define for the first time *group testing channels*, which operate much like communications channels, and identify an important property where 'only defects matter' that allows us to prove a theorem on the number of tests needed for accurate group testing to be possible. We also give the first information theoretic bound on adaptive group testing, by drawing an analogy to channel coding with feedback.

6.2 Group testing: a very short introduction

The problem of *group testing* concerns detecting the defective members of a set of items through the means of pooled tests. (In our previous example, for receiver j, think of the 'defective items' as the interfering transmitters with $h_{ji} \neq 0$.) Group testing as a subject dates back to the work of Dorfman [69] in 1940s studying practical ways of testing soldiers' blood for syphilis, and has received much attention from combinatorialists and probabilists since.

The setup is as follows: Suppose we have of a set N items, of which a subset $\mathcal{K}$ of size K is defective. To identify $\mathcal{K}$, we could test each of the N items individually for defectiveness. However, when K is small compared to N, most of the tests will give negative results. A less wasteful method is to test *pools* of numerous items together at the same time. After a number T of such pooled tests, it should be possible to deduce which items were defective.

Let $x_{it} = 1$ denote that item i is included in test t. In the so-called *deterministic case*, a test of $n_t := |\{i : x_{it} = 1\}|$ items of which $k_t := |\{i \in \mathcal{K} : x_{it} = 1\}|$ are defective gives a negative result $y_t = 0$ if no defects are tested ($k_t = 0$) and a positive result if at least one defect is pooled into the test ($k_t \geq 1$).

After T tests, we make an estimate $\widehat{\mathcal{K}}$ of the defective set, with some average probability of error ϵ. We want to choose our tests in such a way that ϵ is small, while keeping T as low as possible.

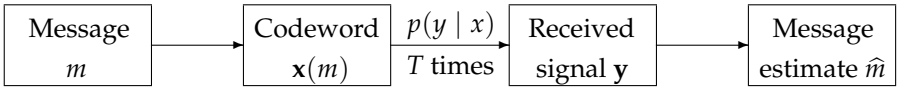
Traditionally, this has been seen as a combinatorial problem: given N and K, one aims to find an $N \times T$ *testing matrix* $\mathsf{X} = (x_{it})$ such that all $\binom{N}{K}$ possible defective sets $\mathcal{K}$ give a different sequence of test results $y_1, \ldots, y_T$. This gives a zero error probability $\epsilon = 0$, and one is interested in how small T can be made. (See, for example, the textbook of Du and Hwang [70] for more details on the combinatorial approach to group testing.)

However, an alternative approach is to use random pools. That is, we set X_{it} to be random 0s or 1s – typically, $X_{it} = 1$ with some probability p IID across i and t, where p may depend on K and N. One then investigates how big T must be compared to N and K in order to keep the average error probability ϵ low.

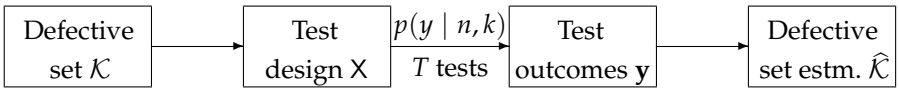
Recent progress has been made on this channel coding approach by Atia and Saligrama [71], by comparing the problem to the classical problem of channel coding, first studied by Shannon [4]. (See Chapter 1 for more details on channel coding.)

The two figures below show an interesting similarity between the two problems. Note that our goals are slightly different, though – in channel coding we wish to maximise the number of messages M for large blocklengths T; whereas in group testing we wish to minimise the number of tests T for a large number of items N.

Channel coding:



Group testing:



In single-user point-to-point channel coding, M messages are encoded into codewords $(x_{m1}, \dots, x_{mT})$ of length T. After being sent through some channel, the codewords are received as $(y_1, \dots, y_T)$, and the original message is estimated as $\widehat{m}$, with the hope that the average error probability ϵ will be small.

Shannon's celebrated channel coding theorem [4, Theorem 11] (see Theorem 1.11 of this thesis) tells us how large we can make M compared to T, while still being sure that the error probability stays small. Shannon's breakthrough was to study a random coding scheme, were the X_{mt} are all IID according to some distribution X. One way to phrase the achievability part of Shannon's theorem – Shannon offered a similar phrasing as an alternative in his original paper [4, Theorem 12] – is the following:

Theorem 6.2 (Shannon's channel coding theorem)**.** *Consider a communications channel $(\mathcal{X}, \mathcal{Y}, p(y \mid x))$. Let $M^* = M^*(T, \epsilon)$ be the maximum number of messages that can be sent through the channel with blocklength T and error probability at most $\epsilon \in (0, 1)$. Then*

$$M^* \geq 2^{T \max_X \mathbb{I}(X:Y) + o(T)} \qquad \text{as } T \to \infty.$$

Atia and Saligrama [71, Theorem III.1] adapted Gallager's proof [10] of the achievability part of Shannon's channel coding theorem to give a similar result. This time, we're interested in how many tests T are required to keep the error probability arbitrarily low.

Theorem 6.3. *Consider group testing in the deterministic case.*

Let $T^* = T^*(N, K, \epsilon)$ be the minimum number of tests necessary to identify K defects among N items with error probability at most $\epsilon \in (0, 1)$. Then $T^* \leq \overline{T} + o(\log N)$ as $N \to \infty$, where

$$\overline{T} = \min_p \max_{\mathcal{L} \subset \mathcal{K}} \frac{\log_2 \binom{N-K}{|\mathcal{L}|} \binom{K}{|\mathcal{L}|}}{\mathbb{I}(\mathbf{X}_{\mathcal{K} \setminus \mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}.$$

(It's worth noting that the term inside the maximisation depends only on the cardinality $|\mathcal{L}|$ of $\mathcal{L}$, not on $\mathcal{L}$ itself.)

In channel coding, the main interest is in finding the value of M^* for different channels – the limit

$$\lim_{T\to\infty} \frac{\log M^*}{T} = \max_X \mathbb{I}(X:Y) =: c$$

exists and is called the channel capacity (see Definitions 1.6 and 1.10).

In group testing, for the deterministic case, Atia and Saligrama [71, Theorem V.1] showed that we have

$$\overline{T} = O(K \log N) \qquad \text{as } K \to \infty \text{ and } N \to \infty.$$

a bound that Sejdinovic and Johnson [72, Theorem 2] improved to

$$\overline{T} \sim \mathrm{e}K \frac{\log(K(N-K))}{\log K} = \mathrm{e}K \left(1 + \frac{\log(N-K)}{\log K}\right) \quad \text{as } K \to \infty \text{ and } N \to \infty.$$

(Here, as elsewhere, we use $f(x) \sim g(x)$ to mean that $f(x)/g(x) \to 1$.)

Here, we will attempt to find to what range of *group testing channels* the result of Atia and Saligrama can be extended, and some bounds on $\overline{T}$ – and hence T^* – for those channels. We also investigate further insights that channel coding can give to group testing.

6.3 Channels

In channel coding, many different types of communication can be modelled by using different channels. Recall from Definition 1.1 that a communication channel is defined by stating what inputs $x \in \mathcal{X}$ and outputs $y \in \mathcal{Y}$ the channel can have, and what the probability $p(y \mid x)$of each output is given each input.

We want to do the same for group testing. The input is already constrained: there are N items each of which can be in ($x_i = 1$) or not in ($x_i = 0$) the pool, so the input alphabet is $\{0,1\}^N$. So we must define the output alphabet and the probability function.

We will assume that in a testing pool there is no 'order' to the items, nor will any elements not placed in the pool affect the outcome of the test, nor can we distinguish between the items other than whether or not they are defective. Hence, the outcome can only depend on two things, the number of items in the test pool n, and the number of those items that are defective k.

Definition 6.4. A *group testing channel* consists of

- an output alphabet $\mathcal{Y}$,
- a probability transition function $p(y \mid n,k)$ relating the number of items n and defectives k in a testing pool to the test outcome y.

The problem of group testing is then to come up with test designs that have a low error probability for as few tests as possible.

Definition 6.5. A *testing pool* for N items consists of a vector $\mathbf{x} = (x_i) \in \{0,1\}^N$, where $x_i = 1$ denotes that item i is included in the pool and $x_i = 0$ denotes that item i is not included in the pool. We define $n := |\{i : x_i = 1\}|$ to be the total number of items in the pool and $k := |\{i \in \mathcal{K} : x_i = 1\}|$ to be the number of defective items in the pool.

A *test design* of T tests for N items consists of

- a sequence $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ of T *testing pools* (which can be summarized by the *testing matrix* $\mathsf{X} = (x_{it}) \in \{0,1\}^{N \times T}$);
- a *defective set detection function* $\widehat{\mathcal{K}} \colon \mathcal{Y}^T \to [N]^{(K)}$, where $[N]^{(K)}$ is the collection of subsets of $\{1, 2, \ldots, N\}$ of size K.

We can now describe the deterministic case discussed earlier as an example of a group testing channel under Defintion 6.4.

Definition 6.6. The *deterministic channel* has output alphabet $\mathcal{Y} = \{0,1\}$ and probability transition function

$$p(1 \mid n,k) = \begin{cases} 0 & \text{if } k = 0, \\ 1 & \text{if } k \geq 1, \end{cases} \qquad p(0 \mid n,k) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{if } k \geq 1. \end{cases}$$

Atia and Saligrama [71, Subsection II-C] also studied two ways in which error could be introduced into group testing, which they called the additive and dilution models. They showed that their main result (Theorem 6.3) also holds true for these two channels.

In the additive model, a negative pool can actually return a false positive result, with some fixed probability q. This could happen in our interference graph example from Section 6.1 if we included a noise term Z. This model can be recast as an example of a group testing channel.

Definition 6.7. The *addition channel* with addition probability $q > 0$ has output alphabet $\mathcal{Y} = \{0,1\}$ and probability transition function

$$p(1 \mid n,k) = \begin{cases} q & \text{if } k = 0, \\ 1 & \text{if } k \geq 1, \end{cases} \qquad p(0 \mid n,k) = \begin{cases} 1-q & \text{if } k = 0, \\ 0 & \text{if } k \geq 1. \end{cases}$$

Atia and Saligrama [71, Table 1] calculated that for the addition channel, as $K \to \infty$ and $N \to \infty$, we have

$$\overline{T} = O\left(\frac{K \log N}{1-q}\right),$$

Using the work of Sejdinovic and Johnson [72, Theorem 6] we can improve this to

$$\overline{T} \sim \; eK\frac{\log(K(N-K))}{\log\frac{1}{q}},$$

by setting $u = 0$ and optimising $\alpha = 1$ in their equation (22) and rearranging. (Interestingly, this is discontinuous with the deterministic channel at $q = 0$.)

The dilution model describes the case where a very small number of defective items in a testing pool might be 'drowned out' by the nondefective items. Specifically, any defective items in the test may each evade the test independently with some probability u, with the potential to cause false negative results. This could happen in our interference graph example from Section 6.1 if we did not neglect the possibility that superposed interfering signals can cancel each other out.

Definition 6.8. The *dilution channel* with dilution probabilty $u \geq 0$ has output alphabet $\mathcal{Y} = \{0, 1\}$ and probability transition function

$$p(1 \mid n, k) = \begin{cases} 0 & \text{if } k = 0, \\ 1 - u^k & \text{if } k \geq 1, \end{cases} \qquad p(0 \mid n, k) = \begin{cases} 1 & \text{if } k = 0, \\ u^k & \text{if } k \geq 1. \end{cases}$$

Atia and Saligrama [71, Table 1] calculated that for the addition channel we have

$$\overline{T} = O\left(\frac{K \log N}{(1-u)^2}\right),$$

Again, using the work of Sejdinovic and Johnson [72, Theorem 6] we can improve this to

$$\overline{T} \sim \mathrm{e}^{1+u+f(u)} K \left(1 + \frac{\log(N-K)}{\log K}\right)$$

where

$$0 \leq f(u) \leq \frac{u^2}{1-u} = u^2 + u^3 + \cdots;$$

we do this by optimising $\alpha = 1/(1-u)$ in their equation (24) and rearranging. In other words, the addition channel requires about $e^u \approx 1 + u$ times as many tests as the deterministic channel, for small u.

Sejdinovic and Johnson also considered a channel that combines the additive and dilutive noise models.

Definition 6.9. The *addition/dilution channel* with addition probability $q > 0$ and dilution probabilty $u \geq 0$ has output alphabet $\mathcal{Y} = \{0, 1\}$ and probability transition function

$$p(1 \mid n, k) = \begin{cases} q & \text{if } k = 0, \\ 1 - u^k & \text{if } k \geq 1, \end{cases} \qquad p(0 \mid n, k) = \begin{cases} 1 - q & \text{if } k = 0, \\ u^k & \text{if } k \geq 1. \end{cases}$$

Working on the unproven assumption that Atia and Saligrama's result (Theorem 6.3) also holds for the addition/dilution channel, Sejdinovic and Johnson [72, Theorem 3] prove a similar result for the addition/dilution channel. (Although note that Sejdinovic and Johnson [72, reference 1] were working from an earlier preprint of Atia and Saligrama's paper [71, version 2]. This had a different, and less rigorous, proof based on typical set decoding, rather than Gallager's maximum likelihood approach as in the most recent version [71, version 4].)

Atia and Saligrama and Sejdinovic and Johnson defined their channels in terms of complicated Boolean sums and products of random vectors and matrices with the testing matrix X. But our definition in terms of probability transition functions makes the behaviour of such channels clearer, and should allow the proof of more universal theorems. It also makes it much easier to define new channels to model testing behaviour – and many other existing models can be reformulated as group testing channels.

Definition 6.10. The *erasure channel* is a model that works like the deterministic channel, but fails to produce a result with some fixed erasure probability ϵ. That is, $\mathcal{Y} = \{\mathtt{0}, \mathtt{?}, \mathtt{1}\}$ and

$$p(\mathtt{1} \mid n,k) = \begin{cases} 0 & \text{if } k = 0, \\ 1-\epsilon & \text{if } k \geq 1, \end{cases} \quad p(\mathtt{?} \mid n,k) = \epsilon, \; p(\mathtt{0} \mid n,k) = \begin{cases} 1-\epsilon & \text{if } k = 0, \\ 0 & \text{if } k \geq 1. \end{cases}$$

The *dilution threshold* only gives a positive result if a sufficient proportion of the tested items are defective, above some threshold $\theta \in (0,1)$. That is, $\mathcal{Y} = \{\mathtt{0}, \mathtt{1}\}$ and

$$p(\mathtt{1} \mid n,k) = \begin{cases} 0 & \text{if } k/n < \theta, \\ 1 & \text{if } k/n \geq \theta, \end{cases} \qquad p(\mathtt{0} \mid n,k) = \begin{cases} 1 & \text{if } k/n < \theta, \\ 0 & \text{if } k/n \geq \theta. \end{cases}$$

The *counting channel* gives as the output the number of defective items in the set. That is, the probability transition function $p(y \mid n,k)$ defined implicitly be the relation

$$Y = k.$$

It's worth noting that group testing under the counting channel model is equivalent to 0–1 compressed sensing with sparsity exactly k. (A model equivalent to the count channel has previously been studied by Shapiro and Fine [73], Erdős and Rényi [74], and others – for more details see the textbook of Du and Hwang [70, Section 11.2].)

The *overflow channel* gives as a result exactly how many defective items were in the test, up to some limit l. That is, $\mathcal{Y} = \{0, 1, 2, \ldots, l\}$ and probability function definied implicitly by the relation

$$Y = \max\{k, l\}.$$

(When $l = n$, this is equivalent to the counting channel; when $l = 1$ this is equivalent to the deterministic channel; when $l = 2$ this is the message collision model of Capetanakis [67].)

The *symmetric channel* gives a negative result if all items are nondefective, a positive result if all items are defective, and an uncertain result if there is a mixture of defective and nondefective items. That is, $\mathcal{Y} = \{0, ?, 1\}$ and relation

$$Y = \begin{cases} 0 & \text{if } k = 0 \\ ? & \text{if } 0 < k < n \\ 1 & \text{if } k = n. \end{cases}$$

(A model equivalent to the symmetric channel has previously been studied by Sobel, Kumar, and Blumenthal [75] and Hwang [76].)

These are just a few examples – many more realistic error models for group testing can be formulated as channels, and perhaps wider use can be made of this new concept.

6.4 When only defects matter

Note that for many of the group testing channels we have mentioned – including all those studied by Atia and Saligrama and by Sejdinovic and Johnson – the output depends only on k, the number of defects in the test, and not on n, the total number of items in the test. In other words, 'only defects matter', and the number of nondefects in the test is irrelevant.

Definition 6.11. A channel $(\mathcal{Y}, p)$ whose probability function $p(y \mid n, k) = p(y \mid k)$ is dependent only on k and not on n is said to have the *only-defects-matter* property.

In the examples from Defintion 6.11 and earlier, the deterministic, addition, dilution, addition/dilution, erasure, counting, and overflow channels have the only-defects-matter property. The dilution threshold and symmetric channels do not have the only-defects-matter property

Another way to state the only-defects-matter property is that for a channel where only defects matter, we can make the simplification

$$\mathbb{P}(Y \mid \mathbf{X}) = \mathbb{P}(Y \mid \mathbf{X}_{\mathcal{K}}).$$

Making this simplification is crucial to the proof of Atia and Saligrama [71, for example Section III.A]. Indeed, this is the only specific point about the deterministic, additive, and dilution channels that Atia and Saligrama use in their proof. Hence, we have the following:

Theorem 6.12. *Consider a group testing channel where only defects matter. Let $T^* = T^*(N, K, \epsilon)$ be the minimum number of tests necessary to identify K defects among N items with error probability at most $\epsilon \in (0, 1)$. Then $T^* \leq \overline{T} + o(\log N)$ as $N \to \infty$, where*

$$\overline{T} = \min_{p} \max_{\mathcal{L} \subset \mathcal{K}} \frac{\log_2 \binom{N-K}{|\mathcal{L}|}\binom{K}{|\mathcal{L}|}}{\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}.$$

The proof is the same as that of Atia and Saligrama's theorem [71, Theorem III.1]. We briefly outline the proof here.

Sketch proof. We need to analyse the error probability of a test design. To do this, we will analyse the probability our estimated defective set $\widehat{\mathcal{K}}$ of cardinality K overlaps with the true defective set $\mathcal{K}$ on a set $\mathcal{L}$.

Given a set $\mathcal{L}$ There are $\binom{N-K}{K-|\mathcal{L}|}$ such sets, and $\binom{K}{|\mathcal{L}|}$ sets $\mathcal{L}$ of each possible cardinality. Using a technique similar to Gallager's proof of Shannon's coding theorem, we can bound the probability that we make an error in $|\mathcal{L}|$ places as

$$\epsilon \leq \binom{N-K}{K-|\mathcal{L}|}\binom{K}{|\mathcal{L}|} 2^{-T\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}}:\mathbf{X}_{\mathcal{L}},Y)}.$$

Hence we require

$$T > \frac{\log \binom{N-K}{K-|\mathcal{L}|}\binom{K}{|\mathcal{L}|}}{\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}.$$

We need this to be true for every $\mathcal{L} \subset \mathcal{K}$, and can optimise the result over the test design parameter p. □

Identifying the only-defects-matter property as the crucial factor for proving Theorem 6.12 means that Sejdinovic and Johnson's bound on $\overline{T}$ for the addition/dilution channel is now rigorously proven.

Whether Theorem 6.12 – or a similar theorem – holds for channels without the only-defects-matter property is an open problem.

Calculating, or finding good bounds for, the value of $\overline{T}$ or T^* for the erasure, counting, and overflow channels is also an open problem.

6.5 Converse part and adaptive testing

Atia and Saligrama [71, Theorem IV.1] also provide a lower bound on the number of tests needed in group testing. The proof is along the lines of Shannon's converse to the channel coding theorem, and uses Fano's inequality. As before, the Atia–Saligrama proof in fact applies to all channels where only defects matter.

Theorem 6.13. *Consider a group testing channel where only defects matter. Let $T^* = T^*(N, K, \epsilon)$ be the minimum number of tests necessary to identify K defects*

among N items with error probability at most $\epsilon \in (0,1)$. Then $T^* \geq \underline{T} - o(\log N)$ as $N \to \infty$, where

$$\underline{T} = \min_p \max_{\mathcal{L} \subset \mathcal{K}} \frac{\log_2 \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}.$$

Note that, unlike for channel coding, the bounds $\overline{T}$ and $\underline{T}$ do not coincide. Therefore, the exact number of tests needed for group testing to work is not known.

We now present a proof of the converse that is slightly simpler than Atia and Saligrama's. Our proof is based on theirs [71, Theorem IV.1], with some simplifications based on the standard proof of the converse of Shannon's coding theorem, as exposited by Cover and Thomas [6, Section 7.9].

Proof. Suppose a genie reveals to us some subset $\mathcal{L} \subset \mathcal{K}$ of the defective set, leaving us to work out the remaining $K - |\mathcal{L}|$ defective items. Given $\mathcal{L}$, let $\mathcal{K}$ be the random defective set chosen uniformly among the possible $\binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}$ sets of size K of which $\mathcal{L}$ is a subset.

Then we have the following:

$$\begin{aligned}
\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} &= \mathbb{H}(\mathcal{K} \mid \mathcal{L}) && \text{(definition of entropy)} \\
&= \mathbb{H}(\mathcal{K} \mid \widehat{\mathcal{K}}, \mathcal{L}) + \mathbb{I}(\mathcal{K} : \widehat{\mathcal{K}} \mid \mathcal{L}) && \text{(HB7)} \\
&\leq 1 + \epsilon \log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} + \mathbb{I}(\mathcal{K} : \widehat{\mathcal{K}} \mid \mathcal{L}) && \text{(Fano's inequality)} \\
&\leq 1 + \epsilon \log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} + \mathbb{I}(\mathsf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{Y} \mid \mathsf{X}_{\mathcal{L}}), && (6.1)
\end{aligned}$$

where the final step is uses the data-processing inequality and the fact that only defects matter.

We can bound the mutual information term in (6.1) as

$$\begin{aligned}
&\mathbb{I}(\mathsf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{Y} \mid \mathsf{X}_{\mathcal{L}}) \\
&\quad = \mathbb{H}(\mathbf{Y} \mid \mathsf{X}_{\mathcal{L}}) - \mathbb{H}(\mathbf{Y} \mid \mathsf{X}_{\mathcal{K}}) && \text{(HB7)} \\
&\quad = \sum_{t=1}^{T} \left(\mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathsf{X}_{\mathcal{L}}) - \mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathsf{X}_{\mathcal{K}})\right) && \text{(chain rule)} \\
&\quad = \sum_{t=1}^{T} \left(\mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathsf{X}_{\mathcal{L}}) - \mathbb{H}(Y_t \mid \mathbf{X}_{\mathcal{K}t})\right) && \text{(memorylessness)} \\
&\quad \leq \sum_{t=1}^{T} \left(\mathbb{H}(Y_t \mid \mathbf{X}_{\mathcal{L}t}) - \mathbb{H}(Y_t \mid \mathbf{X}_{\mathcal{K}t})\right) && \text{(conditioning reduces entropy)} \\
&\quad = \sum_{t=1}^{T} \mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}\,t} : Y_t \mid \mathbf{X}_{\mathcal{L}t}) && \text{(HB7)} \\
&\quad = T\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : Y \mid \mathbf{X}_{\mathcal{L}}). && (6.2)
\end{aligned}$$

(Here, we write $\mathbf{X}_{\mathcal{K}t} := (X_{it} : i \in \mathcal{K})$ for fixed t.

But we can rewrite this mutual information as

$$\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : Y \mid \mathbf{X}_{\mathcal{L}}) = \mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y) - \mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}) = \mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y), \tag{6.3}$$

since $\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}}$ and $\mathbf{X}_{\mathcal{L}}$ are independent.

Putting together (6.3), (6.2), and (6.1), we get

$$\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} \leq 1 + \epsilon \log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} + T\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y).$$

We can rearrange this to get

$$\epsilon \geq 1 - T\frac{\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}} - \frac{1}{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}.$$

Sending $N \to \infty$, it is clear that we require

$$T \geq \frac{\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|}}{\mathbb{I}(\mathbf{X}_{\mathcal{K}\setminus\mathcal{L}} : \mathbf{X}_{\mathcal{L}}, Y)}$$

to get the error probability arbitrarily low.

This has to be true for all $\mathcal{L} \subset \mathcal{K}$, and we can optimise over the test inclusion parameter p. This gives the result. □

So far, we have been looking at *nonadaptive group testing*, where all the test pools are decided on ahead of time.

Instead, we could consider *adaptive group testing* where tests are performed sequentially, and the makeup of a testing pool can depend on the results of previous tests. That is, the x_{it} are functions of the previous test outcomes $(y_1, y_2, \ldots, y_{t-1})$.

Definition 6.14. An *adaptive test design* of T tests for N items consists of a sequence $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ of T testing pools, where the testing pool for the tth test $\mathbf{x}_t = \mathbf{x}_t(y_1, y_2, \ldots, y_{t-1})$ can depend on earlier test outcomes.

This is analogous to channel coding with feedback, where the encoding function x_t at time t can depend on previous channel outputs $y_1, y_2, \ldots, y_{t-1}$.

As Shannon [77, page 15] showed, for discrete memoryless channels, feedback does not increase the channel capacity. (Although it can help in simplifying encoding and decoding [6, Section 7.12].)

Due to the non-tightness of the bounds on testing in the nonadaptive case, we will not be able to show that adaptive group testing requires the *same* number of tests as nonadaptive testing, but we will be able to show that it obeys the same lower bound and requires no more tests than the nonadaptive case.

This is (as far as we are aware) the first application of information theoretic techniques to adaptive group testing.

Theorem 6.15. *Let T^*_{NA} and T^*_A (dependent on N,K, and ϵ) be the minimum number of tests necessary to identify K defects among N items with error probability at most $\epsilon \in (0,1)$ for nonadaptive and adaptive group testing respectively.*

Then, as $N \to \infty$, we have the inequalities

$$\underline{T} - o(\log N) \leq T^*_A \leq T^*_{NA} \leq \overline{T} + o(\log N)$$

where $\underline{T}$ and $\overline{T}$ are as in Theorems 6.3 and 6.13.

Proof. The third inequality was proven in Theorem 6.3. The second inequality is trivial, as nonadaptive group testing is merely a special case of adaptive group testing where the tester chooses to ignore the information of previous test results.

To prove the first inequality, we adapt the proof of Theorem 6.13, and Shannon's proof that feedback fails to improve channel capacity [77, Theorem 6], as exposited by Cover and Thomas [6, Theorem 7.12.1].

We begin exactly the same way as the proof of Theorem 6.13, to get

$$\log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} \leq 1 + \epsilon \log \binom{N-|\mathcal{L}|}{K-|\mathcal{L}|} + \mathbb{I}(\mathcal{K} : \widehat{\mathcal{K}} \mid \mathcal{L}). \tag{6.4}$$

We again use the data processing inequality (but in a slightly different way) and the only-defects-matter property on the mutual information term in (6.4), to write

$$\begin{aligned}
&\mathbb{I}(\mathcal{K} : \widehat{\mathcal{K}} \mid \mathcal{L}) \\
&\quad \leq \mathbb{I}(\mathcal{K} \setminus \mathcal{L} : \mathbf{Y} \mid \mathcal{L}) && \text{(data processing and only-defects-matter)} \\
&\quad = \mathbb{H}(\mathbf{Y} \mid \mathcal{L}) - \mathbb{H}(\mathbf{Y} \mid \mathcal{K}) && \text{(HB7)} \\
&\quad = \sum_{t=1}^{T} \big(\mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{K}) - \mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{L})\big) && \text{(chain rule)} \\
&\quad = \sum_{t=1}^{T} \big(\mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{L}, \mathbf{X}_{\mathcal{L}t}) - \mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{K}, \mathbf{X}_{\mathcal{K}t})\big) \\
&\qquad && (\mathbf{X}_{\mathcal{J}t} \text{ a function of } Y_1, \ldots, Y_{t-1}, \mathcal{J}) \\
&\quad \leq \sum_{t=1}^{T} \big(\mathbb{H}(Y_t \mid \mathbf{X}_{\mathcal{L}t}) - \mathbb{H}(Y_t \mid Y_1, \ldots, Y_{t-1}, \mathcal{K}, \mathbf{X}_{\mathcal{K}t})\big) && \text{(cond. reduces entr.)} \\
&\quad = \sum_{t=1}^{T} \big(\mathbb{H}(Y_t \mid \mathbf{X}_{\mathcal{L}t} - \mathbb{H}(Y_t \mid \mathbf{X}_{\mathcal{K}t})\big), && (6.5)
\end{aligned}$$

where (6.5) is because, conditional on $\mathbf{X}_{\mathcal{K}t}$, we know that Y_t is independent of previous Ys and the defective set $\mathcal{K}$.

We can now pick back up with the proof of Theorem 6.13, two lines above (6.2), to complete the proof. □

It is tempting to wonder if in fact non-adaptive and adaptive group testing require exactly the same number of tests $T_A^* = T_{NA}^*$. This would be in contrast to results for the deterministic model in the zero-error case, where it is known that adaptive group testing can be performed in strictly fewer tests than nonadaptive (in the $K \to \infty, N \to \infty$ regime) [70, page 139].

Similarly, Shannon showed that feedback for channel coding does help in the zero-error case, but not the arbitrarily-small-error case [77].

6.6 Further work

The investigation of applying information theoretic ideas to group testing is at a very early stage and there are many open questions.

Can we find an analogue of Theorem 6.12 where nondefects also matter? That is, can we drop the only-defects-matter property. The proof of Theorem 6.12 relies on the fact that, given $\mathcal{K} \cap \widehat{\mathcal{K}} =: \mathcal{L}$, the random variables $\mathbb{P}(Y \mid \mathbf{X}, \mathcal{K} \text{ defective})$ and $\mathbb{P}(Y \mid \mathbf{X}, \widehat{\mathcal{K}} \text{ defective})$ are conditionally independent. While this is no longer true when nondefects matter too, they are still conditionally independent given $\mathcal{L}$ and the total number of items n in the test. Perhaps this promises a way forwards.

What is T^* for different channels? Is there a reliable method to calculate, exactly or approximately, T^* for different channels. Is there an easy way to do so? What are the minimising choices of p?

Can we close the gap between $\overline{T}$ and $\underline{T}$ to tighten our bounds?

Does nonadaptive group testing require more tests than adaptive? That is, does $T_A^* = T_{NA}^*$ asymptotically.

The wider view: What other information-theoretic techniques can be useful? Sejdinovic and Johnson [72] have tried using message-passing decoding algorithms for computationally feasible defective set detection. Cheraghchi and coauthors [78] have used information theoretic techniques to study group testing on graphs. What else is there to try?

That's just a few brief suggestions.

Notes

This chapter has benefited from discussions with Oliver Johnson, Dino Sejdinovic, and Robert Piechocki.

Group testing was first studied by Dorfman [69] in the context of testing soldiers' blood for syphilis.

The information theoretic approach to group testing is due to Atia and Saligrama [71]. A recent paper of Sejdinovic and Johnson was also useful [72].

The textbook of Du and Hwang [70] was useful for material on group testing. The textbook of Cover and Thomas [6] was useful for material on channel coding.

7

Conclusions and further work

In this thesis, we've examined a number of different ways of combating interference in wireless networks.

- In Chapter 3 we saw how the simple interference-as-noise technique can be effective over short hops in well-structured networks.
- In Chapter 4, we saw how interference alignment can give tight bounds on the performance of large random networks.
- In Chapter 5, we examined the tradeoff between delay and communication rate when using ergodic interference alignment.
- In Chapter 6, we examined how group testing can help with network performance, and how channel coding techniques can help with group testing.

As we've gone through, we have left a number of pointers to open questions and further work (in particular, see Sections 3.4, 4.6, 5.6, and 6.6). We now have the opportunity to take a brief look at the wider picture.

- While interference alignment has been an important theoretical breakthrough, it has had few practical benefits to date. The work in Chapter 5 of this thesis on delay–rate tradeoff is a step in the right direction, but more work is needed. How can we reduce the complexity of schemes? Can we reduce the amount of channel state information required? Can we reduce blocklengths further? Are the schemes robust to errors in channel estimation?

- There is no Shannon's channel coding theorem for networks. That is, there is no general result that tells one how to calculate the capacity region (or even just the sum-capacity) of a network. (The best current result is the cutset bound of Cover and Thomas [6, Theorem 15.10.1].) Even a result for the Gaussian case seems far off – Jafar refers to a result in the Gaussian case for only interference networks as "the holy grail of network information theory" [27, page 1].

- Most capacity theorems for networks – including those in this thesis – are nonconstructive. That is, no explicit codes are given. Are some coding schemes particularly well suited to interference alignment schemes, or can standard codes (like low-density parity-check codes) be adapted to this new area?

- Theorems about sum-capacity (such as ours in Chapter 4) do not give everything we want to know about a network. After all, a network operating solely at its optimal sum-rate may provide very poor performance for some unfortunate users. Concepts of fairness and cooperation also need to be taken into account.

- As wireless devices become ever more ubiquitous, many engineering problems in this area become ever more severe. Networks must be set up 'ad hoc' with little prior knowledge (the example in Section 6.1 shows one way of approaching this); also, interference will become a much bigger problem, and 'green' technologies with low power consumption will be important.

- Work on the connection between group testing and channel coding is at a very early stage – in Section 6.6 we listed a number of future directions for research. Are other mathematical problems that involve inference about unknown quantities approachable from an information theoretic point of view? How far can Shannon's work at a 1940s telephone company take us?

References

[1] Oliver JOHNSON, Matthew ALDRIDGE, and Robert PIECHOCKI. Interference alignment-based sum capacity bounds for random dense Gaussian interference networks. *IEEE Transactions on Information Theory*, **57**:1, 282–290, 2011.

[2] Matthew ALDRIDGE, Oliver JOHNSON, and Robert PIECHOCKI. Asymptotic sum-capacity of random Gaussian interference networks using interference alignment. *Proceedings of the 2010 IEEE International Symposium on Information Theory*, 410–414, 2010.

[3] Oliver JOHNSON, Matthew ALDRIDGE, and Robert PIECHOCKI. Delay–rate tradeoff for ergodic interference alignment, version 2. `arXiv:1004.0208v2 [cs:IT]`, 2011.

[4] Claude E SHANNON. A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423 and 623–656, 1948.

[5] David TSE and Pramod VISWANATH. *Fundamentals of Wireless Communication*. Cambridge University Press, 2006.

[6] Thomas M COVER and Joy A THOMAS. *Elements of Information Theory*, second edition. Wiley–Interscience, 2006.

[7] Claude E SHANNON. Communication in the presence of noise. *Proceedings of the IRE*, **37**:1, 10-21, 1949.

[8] Robert M FANO. *Transmission of Information: A statistical theory of communications*. MIT Press, 1961.

[9] David JC MACKAY. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[10] Robert G GALLAGER. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, **11**:1, 3–18, 1965.

[11] Tom RICHARDSON and Rüdiger URBANKE. *Modern Coding Theory*. Cambridge University Press, 2008.

[12] Ayfer ÖZGÜR, Olivier LÉVÊQUE, and David TSE. Linear capacity scaling in wireless networks: beyond physical limits? *2010 Information Theory and Applications Workshop*, 2010.

[13] Andrea J GOLDSMITH and Pravin P VARAIYA. Capacity of fading channels with channel side information. *IEEE Transactions on Information Theory*, **43**:6, 1986–1992, 1997.

[14] Jacob WOLFOWITZ. *Coding Theorems of Information Theory*, second edition. Springer–Verlag, 1964.

[15] Solomon KULLBACK and Richard A LEIBLER. On information and sufficiency. *Annals of Mathematical Statistics*, **22**:1, 79–86, 1951.

[16] HU Kuo Ting (translated by D LIEBERMAN). On the amount of information. *Theory of Probability and Its Applications*, **7**:4, 439–447, 1962.

[17] Thomas M COVER. An achievable rate region for the broadcast channel. *IEEE Transactions on Information Theory*, **21**:4, 226–228, 1975.

[18] Claude E SHANNON. Channels with side information at the transmitter. *IBM Journal of Research and Development*, **2**:4, 289–293, 1958.

[19] Ezio BIGLIERI, John PROAKIS, and Shlomo SHAMAI (SHITZ). Fading channels: information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, **44**:6, 2619–2692, 1998.

[20] Giuseppe CAIRE and Shlomo SHAMAI (SHITZ). On the capacity of some channels with channel state information. *IEEE Transactions on Information Theory*, **45**:6, 2007–2019, 1999.

[21] Rudolf AHLSWEDE. Multi-way communication channels. *Proceedings of the 2nd International Symposium on Information Theory*, 23–51, 1971.

[22] Henry H-J LIAO. Multiple access channels. PhD thesis, University of Hawaii, Honolulu, 1972.

[23] Thomas M COVER. Comments on broadcast channels. *IEEE Transactions on Information Theory*, **44**:6, 2524–2530, 1998.

[24] Thomas M COVER. Broadcast channels. *IEEE Transactions on Information Theory*, **18**:1, 2–14, 1972.

[25] Gerhard KRAMER. Review of rate regions for interference channels. *International Zurich Seminar on Communications*, 162–165, 2006.

[26] Xiaohu SHANG, Gerhard KRAMER, and Biao CHEN. A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels. *IEEE Transactions on Information Theory*, **55**:2, 689–699, 2009.

[27] Syed A JAFAR. The ergodic capacity of interference networks. `arXiv: 0902.0838 [cs:IT]`, 2009.

[28] Viveck CADAMBE and Syed A JAFAR. Interference alignment and degrees of freedom of the *K*-user interference channel. *IEEE Transactions on Information Theory*, **54**:8, 3425–3441, 2008.

[29] Viveck CADAMBE and Syed Ali JAFAR. Degrees of freedom of wireless networks with relays, feedback, cooperation, and full duplex operation. *IEEE Transactions on Information Theory*, **55**:5, 2334–2344, 2009.

[30] Te Sun HAN and Kingo KOBAYASHI. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, **27**:1, 49–60, 1981.

[31] Han-Fah CHONG, Mehul MOTANI, Hari Krishna GARG, and Hesham EL GAMEL. On the Han–Kobayashi region for the interference channel. *IEEE Transactions on Information Theory*, **54**:7, 3188–3195, 2008.

[32] Raul H ETKIN, David NC TSE, and Hua WANG. Gaussian interference channel capacity to within one bit. *IEEE Transactions on Information Theory*, **54**:12, 5534–5562, 2008.

[33] Guy BRESLER, Abhay PAREKH, and David NC TSE. The approximate capacity of the many-to-one and one-to-many Gaussian interference channels. *IEEE Transactions on Information Theory*, **56**:8, 4566–4592, 2010.

[34] A Salman AVESTIMEHR, Suhas N DIGGAVI, and David NC TSE. Wireless information flow: a deterministic approach. *IEEE Transactions on Information Theory*, **57**:4, 1872–1905, 2011.

[35] Piyush GUPTA and PR KUMAR. The capacity of wireless networks. *IEEE Transactions on Information Theory*, **46**:2, 388–404, 2000.

[36] Feng XUE and PR KUMAR. Scaling laws for ad-hoc wireless networks: an information theoretic approach. *Foundation and Trends in Networking*, **1**:2, 145–270, 2006.

[37] Ayfer ÖZGÜR, Olivier LÉVÊQUE, and David NC TSE. Hierarchical cooperation achieves optimal capacity scaling in *ad hoc* networks. *IEEE Transactions on Information Theory*, **53**:10, 3549–3572, 2007.

[38] Ayfer ÖZGÜR and Olivier LÉVÊQUE. Throughput-delay trade-off for hierarchical cooperation in ad hoc wireless networks. *2008 International Conference on Telecommunicatons*, 2008.

[39] Anders HØST-MADSEN and Aria NOSRATINIA. The multiplexing gain of wireless networks. *Proceedings of the 2005 IEEE International Symposium on Information Theory*, 2065–2069, 2005.

[40] Andrea GOLDSMITH. *Wireless Communications*. Cambridge University Press, 2005.

[41] Viveck R CADAMBE, Syed Ali JAFAR, and Chenwei WANG. Interference alignment with asymmetric complex signalling—settling the Høst-Madsen–Nosratinia conjecture. *IEEE Transactions on Information Theory*, **56**:9, 4552–4565, 2010.

[42] Omar EL AYACH, Stephen W PETERS, and Robert W HEATH Jr. The feasibility of interference alignment over measured MIMO–OFDM channels. *IEEE Transactions on Vehicular Technology*, **59**:9, 4309 –4321 , 2010.

[43] Leonard GROKOP, David NC TSE, and Roy D YATES. Interference alignment for line-of-sight channels. `arXiv:0809.3035 [cs:IT]`, 2009.

[44] Bobak NAZER, Michael GASTPAR, Syed A JAFAR, and Sriram VISHWANATH. Ergodic interference alignment. *Proceedings of the 2009 IEEE International Symposium on Information Theory*, 1769–1773, 2009.

[45] Abolfazl Seyed MOTAHARI, Shahab Oveis-GHARAN, Mohammad-Ali MADDAH-ALI, and Amir Keyvan KHANDANI. Real interference alignment: exploiting the potential of single antenna systems, version 2. `arXiv:0908.2282v2 [cs:IT]`, 2009.

[46] Bobak NAZER and Michael GASTPAR. Computation over multiple-access channels. *IEEE Transactions on Information Theory*, **53**:10, 3498–3516, 2007.

[47] Oliver JOHNSON and Yurii SUHOV. Entropy and convergence on compact groups. *Journal of Theoretical Probability*, **13**:3, 843–857, 2000.

[48] Abbas EL GAMAL and Thomas M COVER. Multiple user information theory. *Proceddings of the IEEE*, **68**:12, 1466–1483, 1980.

[49] Gerhard KRAMER. Topics in multi-user information theory. *Foundation and Trends in Communications and Information Theory*, **4**:4–5, 265-444, 2008.

[50] Aydano B CARLEIAL. Interference channels. *IEEE Transactions on Information Theory*, **24**:1, 60–70, 1978.

[51] Mohammad Ali MADDAH-ALI, Abolfazl Seyed MOTAHARI, and Amir Keyvan KHANDANI. Communication over MIMO X channels: interference alignment, decomposition, and performance analysis. *IEEE Transactions on Information Theory*, **54**:8, 3457–3470, 2008.

[52] Syed A JAFAR. Interference alignment. Tutorial at *2010 IEEE International Symposium on Information Theory*, 2010.

[53] *ITS Strategic Research Plan, 2010–2014: Transforming transport through connectivity*. U.S. Department of Transportation, Research and Innovative Technology Administration, 2010.

[54] Liang-Liang XIE and PR KUMAR. A network information theory for wireless communication: scaling laws and optimal operation. *IEEE Transactions on Information Theory*, **50**:5, 748–767, 2004.

[55] Martin HAENGGI and Radha Krishna GANTI. Interference in large wireless networks. *Foundations and Trends in Networking*, **3**:2, 127-248, 2009.

[56] Martin HAENGGI. A geometric interpretation of fading in wireless networks: theory and applications. *IEEE Transactions on Information Theory*, **54**:12, 5500-5510, 2008.

[57] Harpreet S DHILLON, Radha Krishna GANTI and Jeffrey G ANDREWS. A tractable framework for coverage and outage in heterogeneous cellular networks. *Proceedings of the Information Theory and Applications Workshop, 2011*, 2011.

[58] Martin HAENGGI. On distances in uniformly random networks. *IEEE Transactions on Information Theory*, **51**:10, 3584–3586, 2005.

[59] Francois BACCELLI and Bartłomiej BŁASZCZYSZYN. Stochastic geometry and wireless networks – volume 1: theory. *Foundations and Trends in Networking*, **3**:3–4, 249–449, 2009.

[60] John FC KINGMAN. *Poisson Processes*. Oxford Studies in Probability, **3**, Clarendon Press, 1992.

[61] Ayfer ÖZGÜR and David TSE. Achieving linear scaling with interference alignment. *Proceedings of the 2009 IEEE International Symposium on Information Theory*, 1754–1758, 2009.

[62] Urs NIESEN. Interference alignment in dense wireless networks. *IEEE Transactions on Information Theory*, **57**:5, 2889–2901, 2011.

[63] Paul ERDŐS and Alfréd RÉNYI. On random matrices. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, **8**, 455–460, 1963.

[64] David W WALKUP. Matchings in random regular bipartite digraphs. *Discrete Mathematics*, **31**:1, 59–64, 1980.

[65] Joseph C KOO, William WU, and John T GILL III. Delay–rate tradeoff for ergodic interference alignment in the Gaussian case. *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, 1069–1075, 2010.

[66] Toby BERGER, Nader MEHRAVARI, Don TOWSLEY, and Jack WOLF. Random multiple-access communication and group testing. *IEEE Transactions on Communications* **32**:7, 769–779, 1984.

[67] John I CAPETANAKIS. Generalized TDMA: the multi-accessing tree protocol. *IEEE Transactions on Communications*, **27**:10, 1476–1484, 1979.

[68] Lei ZHANG, Jun LUO, and Dongning GUO. Compressed neighbor discovery for wireless networks, version 2. `arXiv:1012.1007v2 [cs:NI]`, 2010.

[69] Robert DORFMAN. The detection of defective members of large populations. *Annals of Mathematical Statistics*, **14**:4, 436–440, 1943.

[70] Ding-Zhu DU and Frank K HWANG. *Combinatorial Group Testing and Its Applications*, second edition. Series on Applied Mathematics, **12**, World Scientific Publishing, 2000.

[71] George ATIA and Venkatesh SALIGRAMA. Boolean compressed sensing and noisy group testing, version 4. `arXiv:0907.1061v4 [cs:IT]`, 2010.

[72] Dino SEJDINOVIC and Oliver JOHNSON. Note on noisy group testing: asymptotic bounds and belief propagation reconstruction. *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computing*, 998–1003, 2010.

[73] HS SHARPIRO (proposed by) and NJ FINE (solution by). Counterfeit coins. *The American Mathematical Monthly*, **67**:7, 697–698, 1960.

[74] Paul ERDŐS and Alfréd RÉNYI. On two problems of information theory. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, **8**, 229–243, 1963.

[75] Milton SOBEL, Satindar KUMAR, and Saul BLUMENTHAL. Symmetric binomial group-testing with three outcomes. *Statistical Decision Theory*

and Related Topics: Proceedings of a symposium at Perdue University, 1970, 119–160, 1971.

[76] Frank K HWANG. Three versions of a group testing game. *SIAM Journal on Algebraic and Discrete Methods*, **5**:2, 145–153, 1984.

[77] Claude E SHANNON. The zero-error capacity of a noisy channel. *IRE Transactions on Information Theory*, **2**:3, 8–19, 1956.

[78] Mahdi CHERAGHCHI, Armin KARBASI, Soheil MOHAJER, and Venkatesh SALIGRAMA. Graph-constrained group testing. *Proceedings of the 2010 IEEE International Symposium on Information Theory*, 1913–1917, 2010.