Information Theory

Master of Logic 2014/15

2nd Block Nov/Dec 2014

Some of these slides are copied from or heavily inspired by the
University of Illinois at Chicago, ECE 534: Elements of Information Theory course
given in Fall 2013 by Natasha Devroye
Thank you very much for the kind permission to re-use them here!

Christian Schaffner

- me
- pure mathematics at ETH Zurich
- PhD from Aarhus, Denmark
- research: quantum cryptography
- c.schaffner@uva.nl

- plays ultimate frisbee

Philip Schulz

- your teaching assistant
- PhD student @ILLC
- working on machine translation
- P.Schulz@uva.nl

Practicalities

- final grade consists of 50-50:
  - 7 weekly homework series, to be graded
  - final exam in week of 31/3/14 - 4/4/14: move to Tuesday, 16 December, but clashes with:
    * Lambda Calculus
    * Foundations of Neural and Cognitive Modelling
- details on course homepage: http://homepages.cwi.nl/~schaffne/courses/inftheory/2014/

Expectations

We expect from you

- be on time
- code of honor (do not cheat)
- focus
- ask questions!

Why multitasking is bad for learning: https://medium.com/@cshirky/why-i-just-asked-my-students-to-put-their-laptops-away-7f5f7c50f368

Expectations

We expect from you

- be on time
- code of honor (do not cheat)
- focus
- ask questions!

You can expect from us

- be on time
- make clear what goals are
- listen to you and respond to email requests
- keep website up to date

Why multitasking is bad for learning: https://medium.com/@cshirky/why-i-just-asked-my-students-to-put-their-laptops-away-7f5f7c50f368

Questions ?

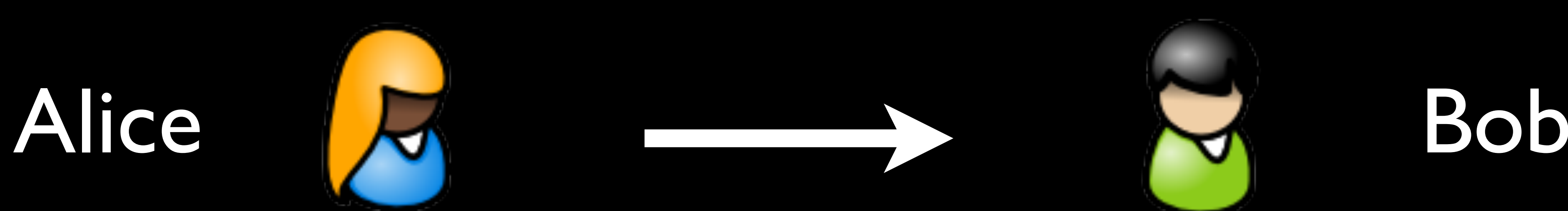
What is communication?

What is communication?

"The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." - C.E. Shannon, 1948

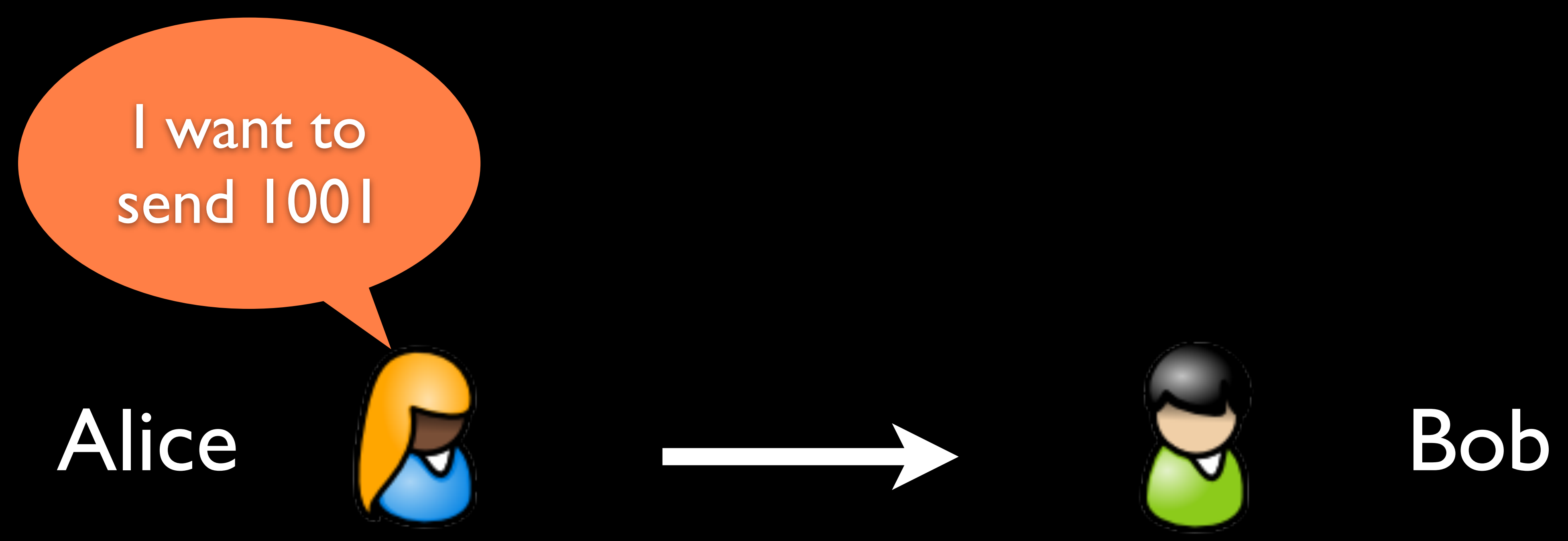
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948



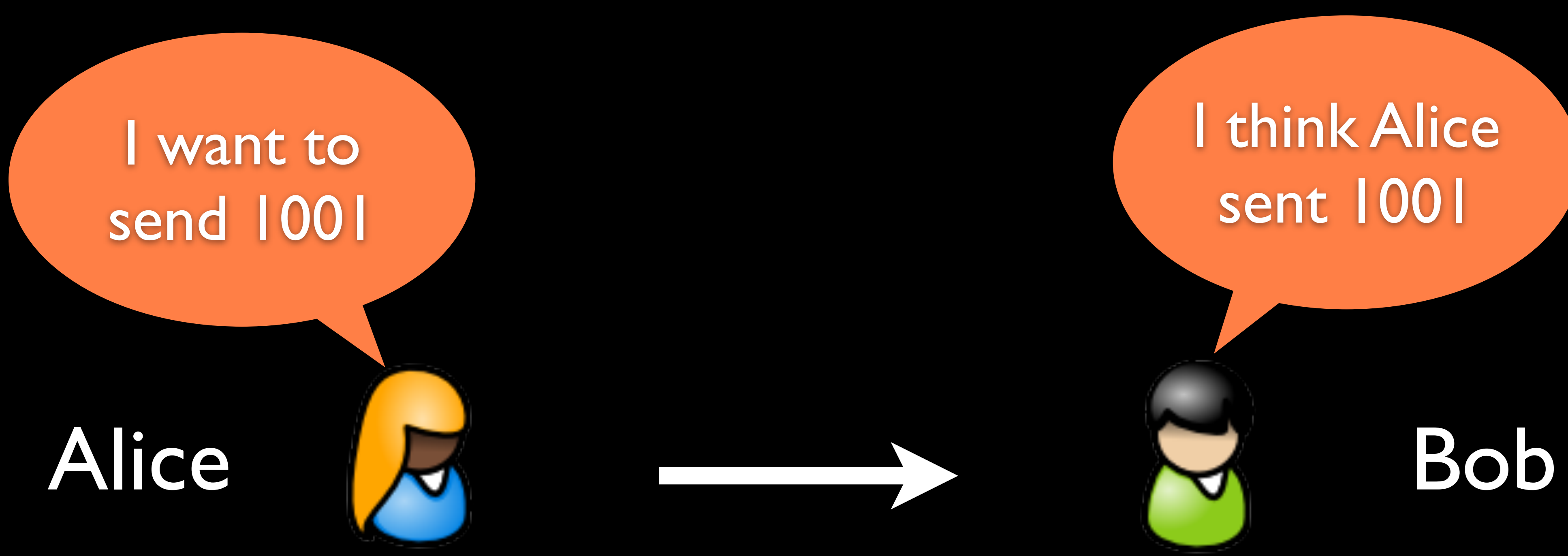
What is communication?

“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948

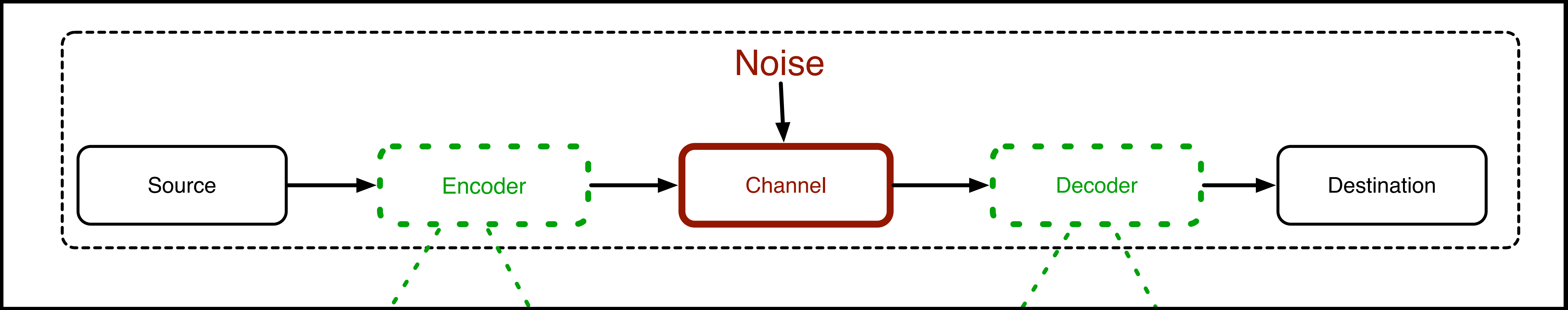


What is communication?

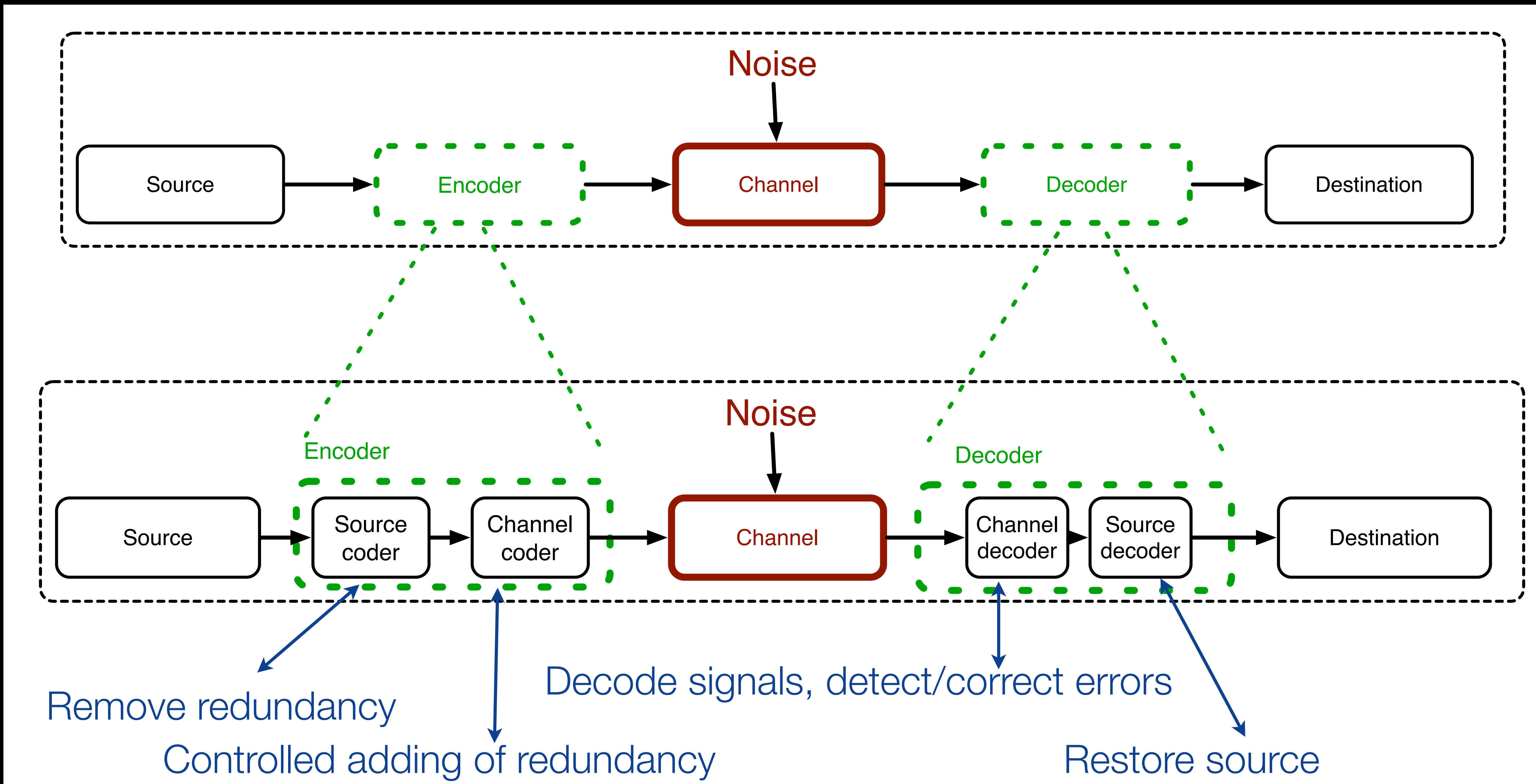
“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.” - C.E. Shannon, 1948



Generic communication block diagram



Generic communication block diagram



History of (wireless) communication

- Smoke signals

History of (wireless) communication

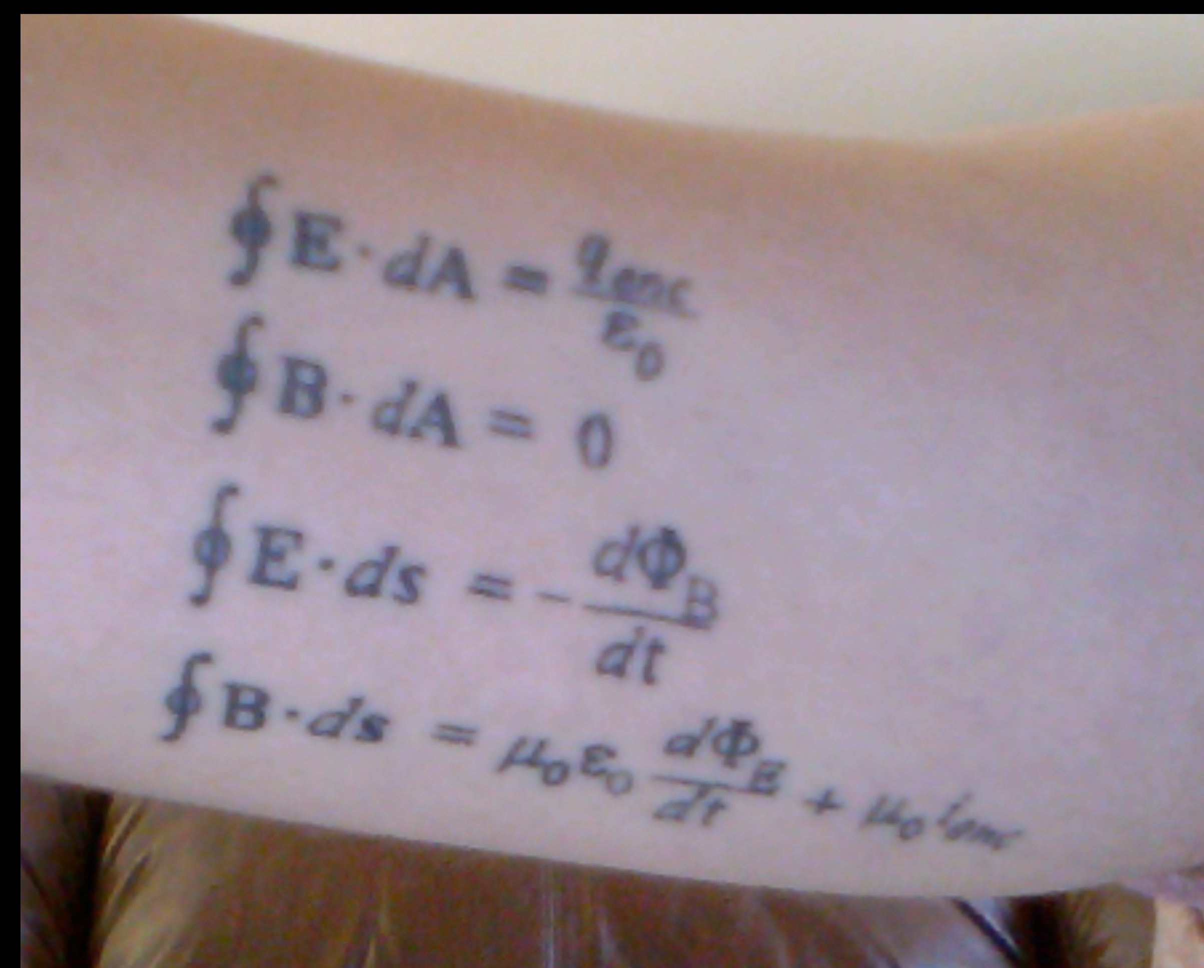
- Smoke signals
- 1861: Maxwell's equations

$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{q_{enc}}{\varepsilon_0}$$

$$\oint \mathbf{B} \cdot d\mathbf{A} = 0$$

$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$

$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 \varepsilon_0 \frac{d\Phi_E}{dt} + \mu_0 i_{enc}$$

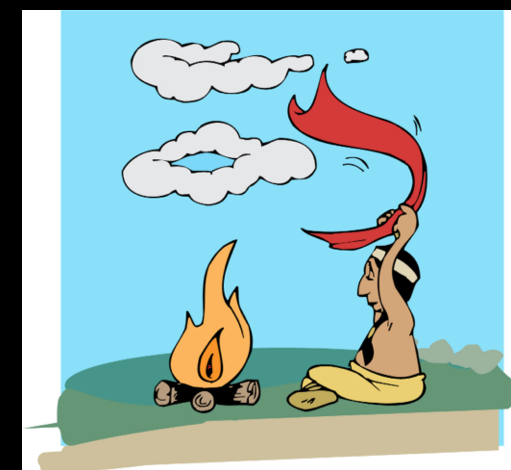


History of (wireless) communication

- Smoke signals
- 1861: Maxwell's equations
- 1900: Guglielmo Marconi demonstrates wireless telegraph

$$\oint \mathbf{E}\cdot d\mathbf{A} = \frac{q_{enc}}{\varepsilon_0}$$
$$\oint \mathbf{B}\cdot d\mathbf{A} = 0$$
$$\oint \mathbf{E}\cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$
$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0\varepsilon_0\frac{d\Phi_E}{dt} + \mu_0 i_{enc}$$

History of (wireless) communication



$$\oint \mathbf{E}\cdot d\mathbf{A} = \frac{q_{enc}}{\varepsilon_0}$$

$$\oint \mathbf{B}\cdot d\mathbf{A} = 0$$

$$\oint \mathbf{E}\cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$

$$\oint \mathbf{B}\cdot d\mathbf{s} = \mu_0\varepsilon_0\frac{d\Phi_E}{dt} + \mu_0 i_{enc}$$

- Smoke signals
- 1861: Maxwell's equations
- 1900: Marconi demonstrates wireless telegraph
- 1920s: Edwin Howard Armstrong demonstrates FM radio

Big Open Questions

- mostly analog
- ad-hoc engineering, tailored to each application
- is there a general methodology for designing communication systems?
- can we communicate reliably in noise?
- how fast can we communicate?

Claude Elwood Shannon

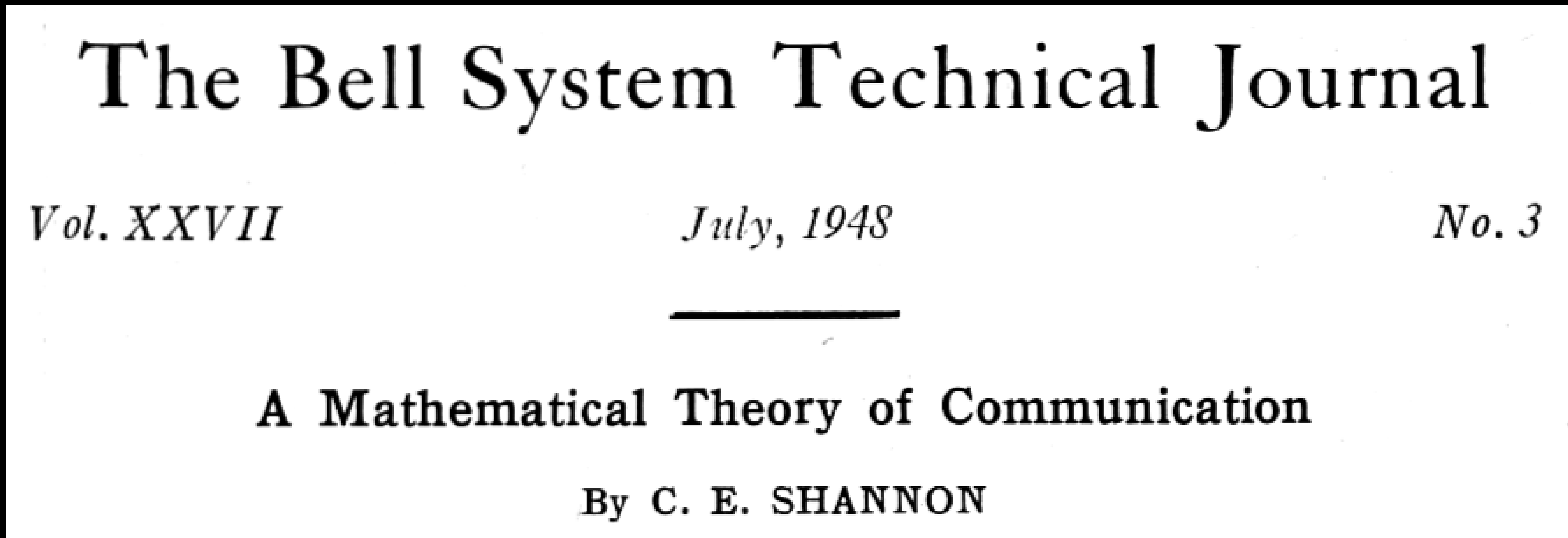
1916 - 2001

- Father of Information Theory
- Graduate of MIT 1940:
  "An Algebra for Theoretical Genetics"
- 1941-1972: Scientist at Bell Labs
- 1958: Professor at MIT:

  When he returned to MIT in 1958, he continued to threaten corridor-walkers on his unicycle, sometimes augmenting the hazard by juggling. No one was ever sure whether these activities were part of some new breakthrough or whether he just found them amusing. He worked, for example, on a motorized pogo-stick, which he claimed would mean he could abandon the unicycle so feared by his colleagues ...
- juggling, unicycling, chess
- ultimate machine

History of (wireless) communication

- BITS !
- arguably, first to really define and use "bits"
- *"He's one of the great men of the century. Without him, none of the things we know today would exist. The whole digital revolution started with him."* -Neil Sloane, AT&T Fellow



The Bell System Technical Journal

Vol. XXVII — July, 1948 — No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

- Introduced a new field: Information Theory

What is communication?

What is information?

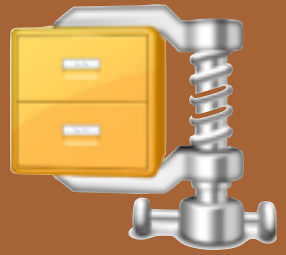
How much can we compress information?

How fast can we communicate?

Main Contributions of Inf Theory

Source coding

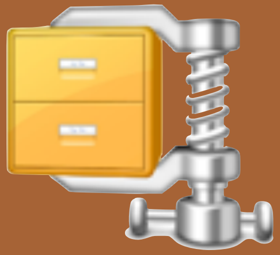
- source = random variable
- ultimate data compression limit is the source's **entropy H**



Main Contributions of Inf Theory

Source coding

- source = random variable
- ultimate data compression limit is the source's **entropy H**



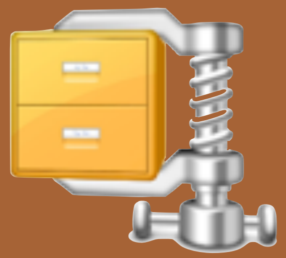
Channel coding

- channel = conditional distributions
- ultimate transmission rate is the **channel capacity C**

Main Contributions of Inf Theory

Source coding

- source = random variable
- ultimate data compression limit is the source's **entropy H**



Channel coding

- channel = conditional distributions
- ultimate transmission rate is the **channel capacity C**

Reliable communication possible $\Leftrightarrow$ H < C

Reactions to This Theory

- Engineers in disbelief
- stuck in analogue world

How to approach the predicted limits?

Shannon says: can transmit at rates up to say 4Mbps over a certain channel without error. How to do it?

It Took 50 Years To Do It

How to approach the predicted limits?

review article by [Costello Forney 2006]

It Took 50 Years To Do It

- 50's: algebraic codes

How to approach the predicted limits?

review article by [Costello Forney 2006]

It Took 50 Years To Do It

- 50‘s: algebraic codes
- 60‘s 70‘s: convolutional codes

How to approach the predicted limits?

review article by [Costello Forney 2006]

It Took 50 Years To Do It

- 50's: algebraic codes
- 60's 70's: convolutional codes
- 80's: iterative codes (LDPC, turbo codes)

How to approach the predicted limits?

review article by [Costello Forney 2006]

It Took 50 Years To Do It

- 50's: algebraic codes
- 60's 70's: convolutional codes
- 80's: iterative codes (LDPC, turbo codes)
- 2009: polar codes

How to approach the predicted limits?

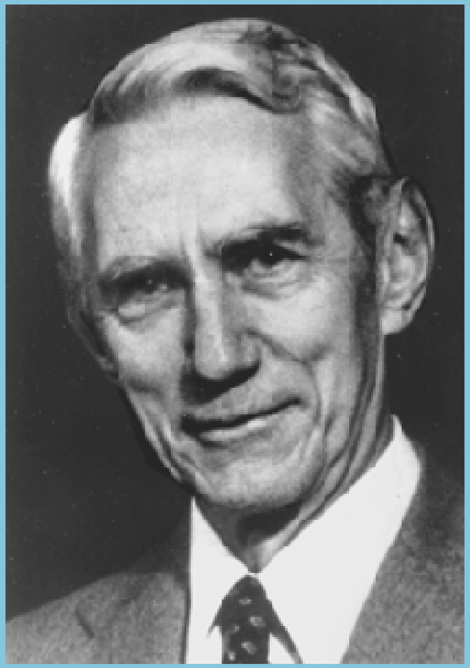
review article by [Costello Forney 2006]

It Took 50 Years To Do It

- 50‘s: algebraic codes
- 60‘s 70‘s: convolutional codes
- 80‘s: iterative codes (LDPC, turbo codes)
- 2009: polar codes

How to approach the predicted limits?

review article by [Costello Forney 2006]



Claude Shannon — Born on the planet Earth (Sol III) in the year 1916 A.D. Generally regarded as the father of the Information Age, he formulated the notion of channel capacity in 1948 A.D. Within **several decades**, mathematicians and engineers had devised practical ways to communicate reliably at data rates within 1% of the Shannon limit . . .

Encyclopedia Galactica, 166th ed.

Robert J. McEliece, Shannon Lecture 2004

Applications

- Communication Theory
- Computer Science (e.g. in cryptography)
- Physics (thermodynamics)
- Philosophy of Science (Occam's Razor)
- Economics (investments)
- Biology (genetics, bio-informatics)

Topics Overview

- Entropy and Mutual Information
- Entropy Diagrams
- Perfectly Secure Encryption
- Data Compression
- Coding Theory
- Channel-Coding Theorem
- Zero-Error Information Theory
- Noisy-Channel Theorem
- Application to Machine Learning

Questions ?

Example: Letter Frequencies

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

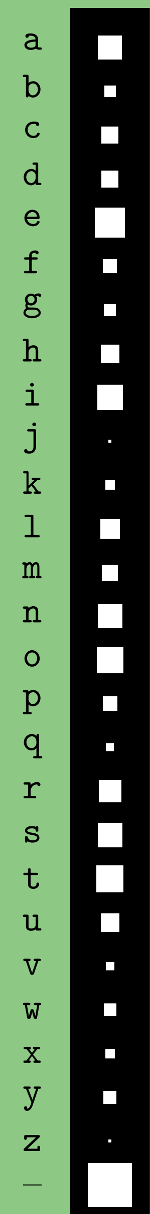


Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from *The Frequently Asked Questions Manual for Linux*). The picture shows the probabilities by the areas of white squares.

Book by David MacKay

Example: Letter Frequencies

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

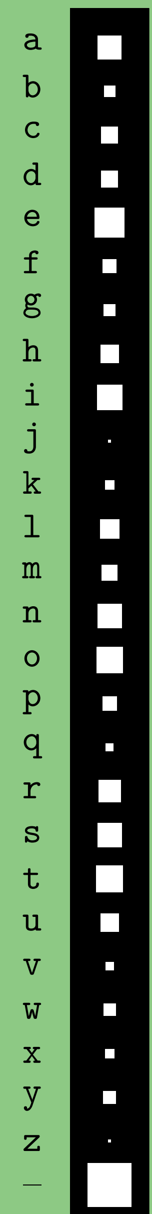


Figure 2.1. Probability distribution over the 27 outcomes for a randomly selected letter in an English language document (estimated from *The Frequently Asked Questions Manual for Linux*). The picture shows the probabilities by the areas of white squares.

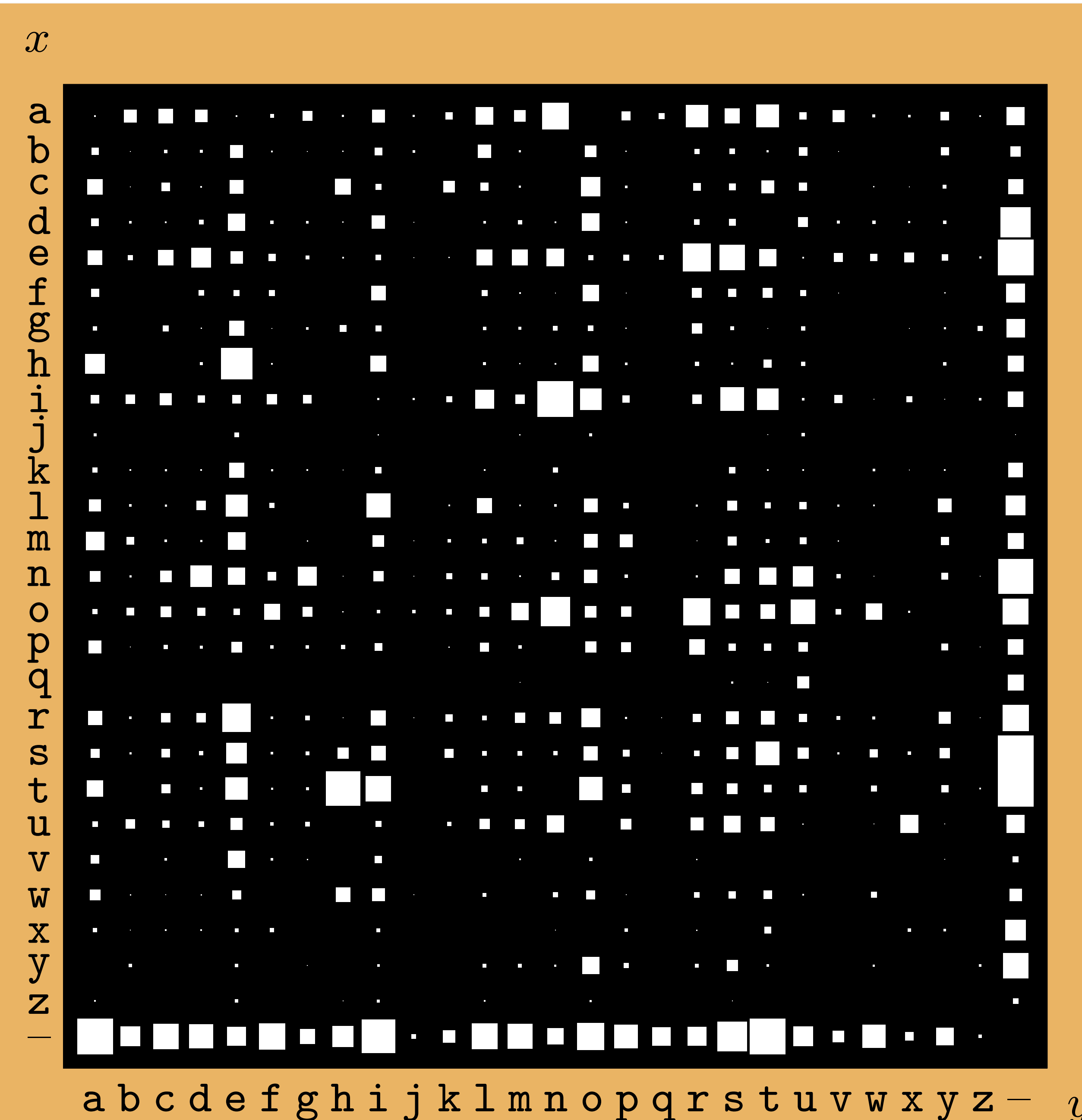


Figure 2.2. The probability distribution over the 27×27 possible bigrams xy in an English language document, *The Frequently Asked Questions Manual for Linux.*

Book by David MacKay

Example: Surprisal Values

from http://www.umsl.edu/~fraundorfp/egsurpri.html

situation	probability $p = 1/2^{\#bits}$	surprisal #bits = $\ln_2[1/p]$
one equals one	1	0 bits
wrong guess on a 4-choice question	3/4	$\ln_2[4/3]$ ~0.415 bits
correct guess on true-false question	1/2	$\ln_2[2]$ =1 bit
correct guess on a 4-choice question	1/4	$\ln_2[4]$ =2 bits
seven on a pair of dice	$6/6^2$ =1/6	$\ln_2[6]$ ~2.58 bits
snake-eyes on a pair of dice	$1/6^2$ =1/36	$\ln_2[36]$ ~5.17 bits
random character from the 8-bit ASCII set	1/256	$\ln_2[2^8]$ =8 bits =1 byte
N heads on a toss of N coins	$1/2^N$	$\ln_2[2^N]$ =N bits
harm from a smallpox vaccination	~1/1,000,000	~$\ln_2[10^6]$ ~19.9 bits
win the UK Jackpot lottery	1/13,983,816	~23.6 bits
RGB monitor choice of one pixel's color	$1/256^3$ ~5.9×10^{-8}	$\ln_2[2^{8*3}]$ =24 bits
gamma ray burst mass extinction event TODAY!	<1/(10^9*365) ~2.7×10^{-12}	hopefully >38 bits
availability to reset 1 gigabyte of random access memory	$1/2^{8E9}$ ~$10^{-2.4E9}$	8×10^9 bits ~7.6×10^{-14} J/K
choices for 6×10^{23} Argon atoms in a 24.2L box at 295K	~$1/2^{1.61E25}$ ~$10^{-4.8E24}$	~1.61×10^{25} bits ~155 J/K
one equals two	0	∞ bits

i	a_i	p_i	$h(p_i)$
1	a	.0575	4.1
2	b	.0128	6.3
3	c	.0263	5.2
4	d	.0285	5.1
5	e	.0913	3.5
6	f	.0173	5.9
7	g	.0133	6.2
8	h	.0313	5.0
9	i	.0599	4.1
10	j	.0006	10.7
11	k	.0084	6.9
12	l	.0335	4.9
13	m	.0235	5.4
14	n	.0596	4.1
15	o	.0689	3.9
16	p	.0192	5.7
17	q	.0008	10.3
18	r	.0508	4.3
19	s	.0567	4.1
20	t	.0706	3.8
21	u	.0334	4.9
22	v	.0069	7.2
23	w	.0119	6.4
24	x	.0073	7.1
25	y	.0164	5.9
26	z	.0007	10.4
27	-	.1928	2.4
	$\sum_i p_i \log_2 \frac{1}{p_i}$		4.1

Table 2.9. Shannon information contents of the outcomes a–z.

Book by David MacKay