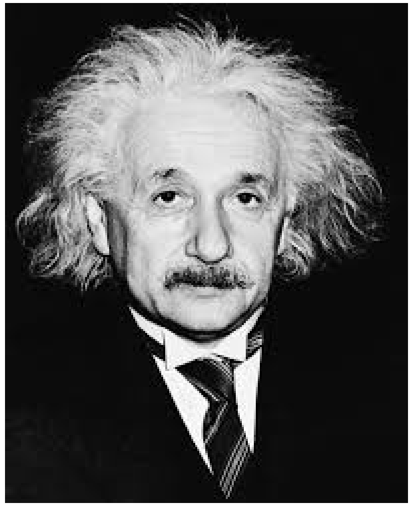
Remembering Shannon

Robert Calderbank, Duke University

Rob Goodman and Jimmy Soni: A Mind at Play: The Life of Claude Shannon and the Birth of the Information Age

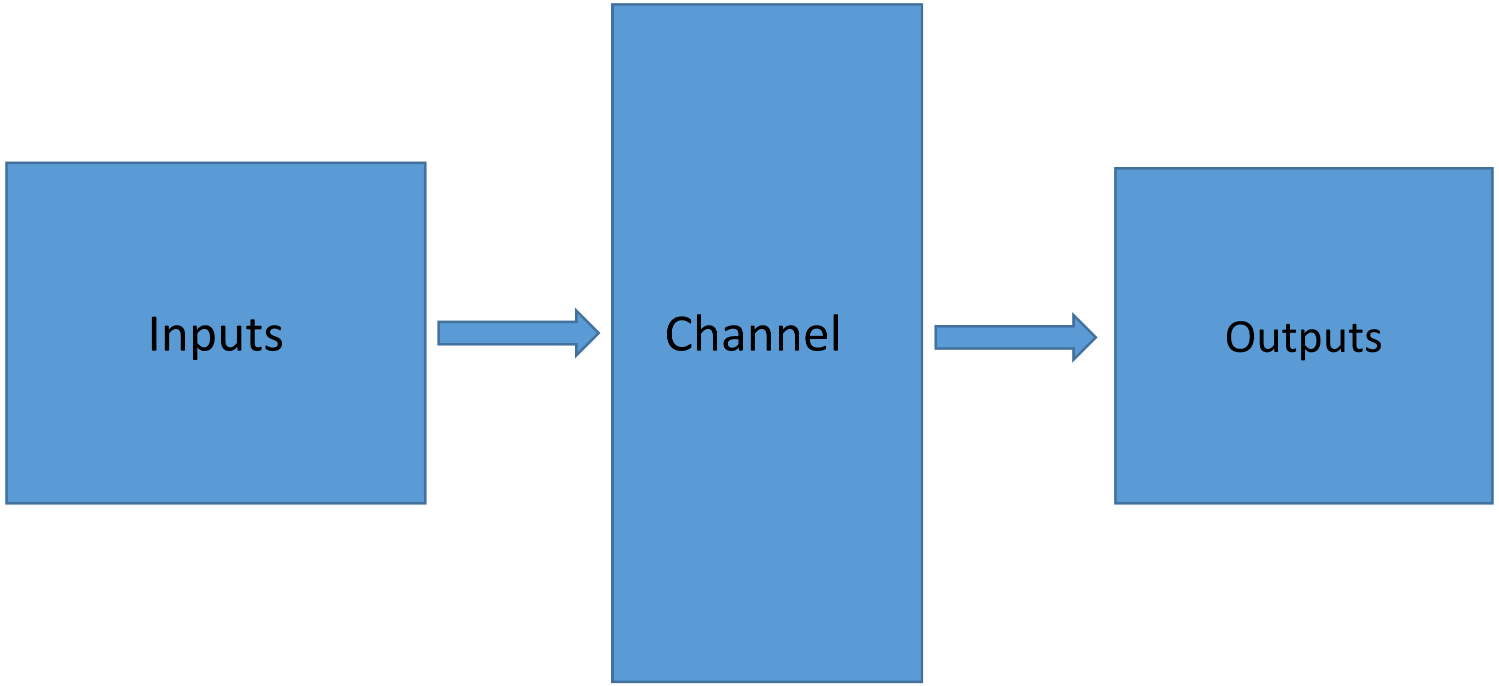
This is the Age of Information

Not the Age of Relativity



So why does Einstein loom so much larger than Shannon ?

The World after Shannon



Inputs

Channel

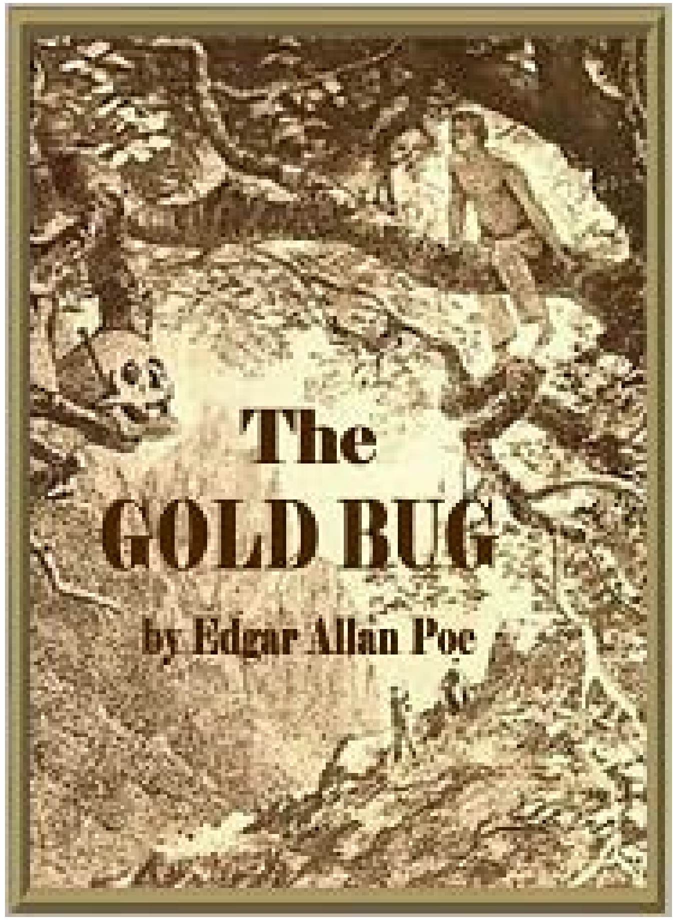
Outputs

Growing up in Gaylord, Michigan

Gaylord
Saginaw
Gaylord, MI
CANADA
Copyright Sperling's BestPlaces

In the 1920s the barbed wire telegraph connected rural communities – as many as 3M farmers

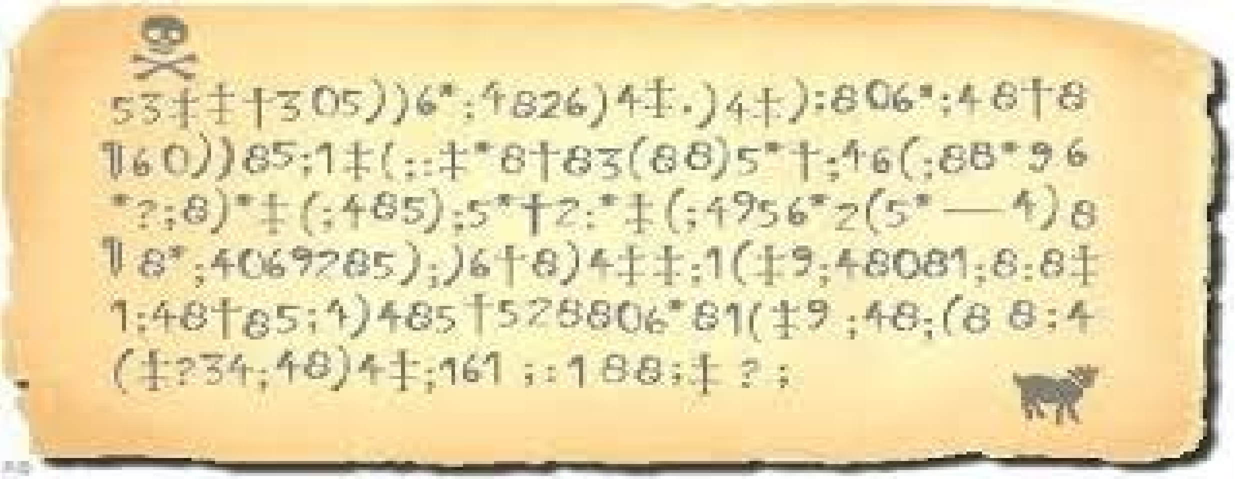
And Claude Shannon built a private network to communicate with a friend who lived half a mile away



Captivated by Secret Writing

Published in 1843, this short story was the most popular and widely read of Poe's works during his lifetime.

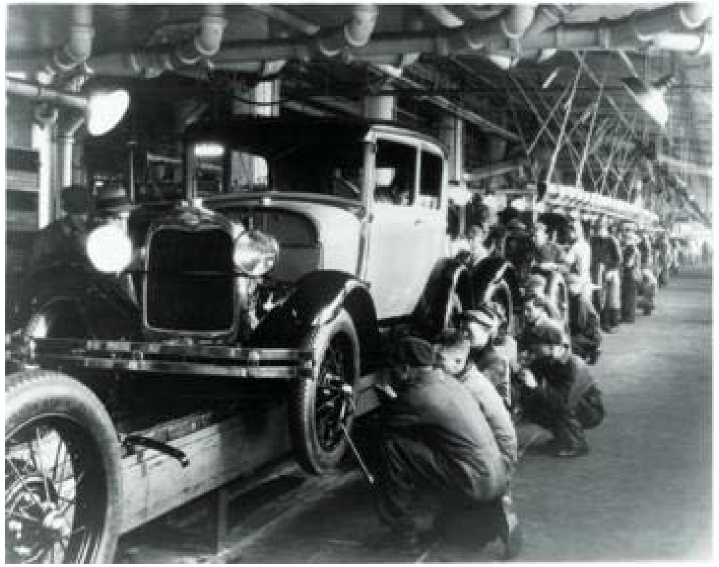
The cipher text provides directions to a pirate hoard



```
53‡‡†305))6*;4826)4‡.)4‡):806*;48†8
¶60))85;1‡(::‡*8†83(88)5*†;46(;88*96
*?;8)*‡(;485);5*†2:*‡(;4956*2(5*—4)8
¶8*;4069285);)6†8)4‡‡;1(‡9;48081;8:8‡
1;48†85;4)485†528806*81(‡9;48;(88:4
(‡?34;48)4‡;161;:188;‡?;
```

The most frequent symbol is the most frequent letter: **8** means **E**
The most common sequence is the most common word: **;48** means **the**
and each deciphered symbol leads to the next

Undergraduate Degrees in Mathematics and Engineering

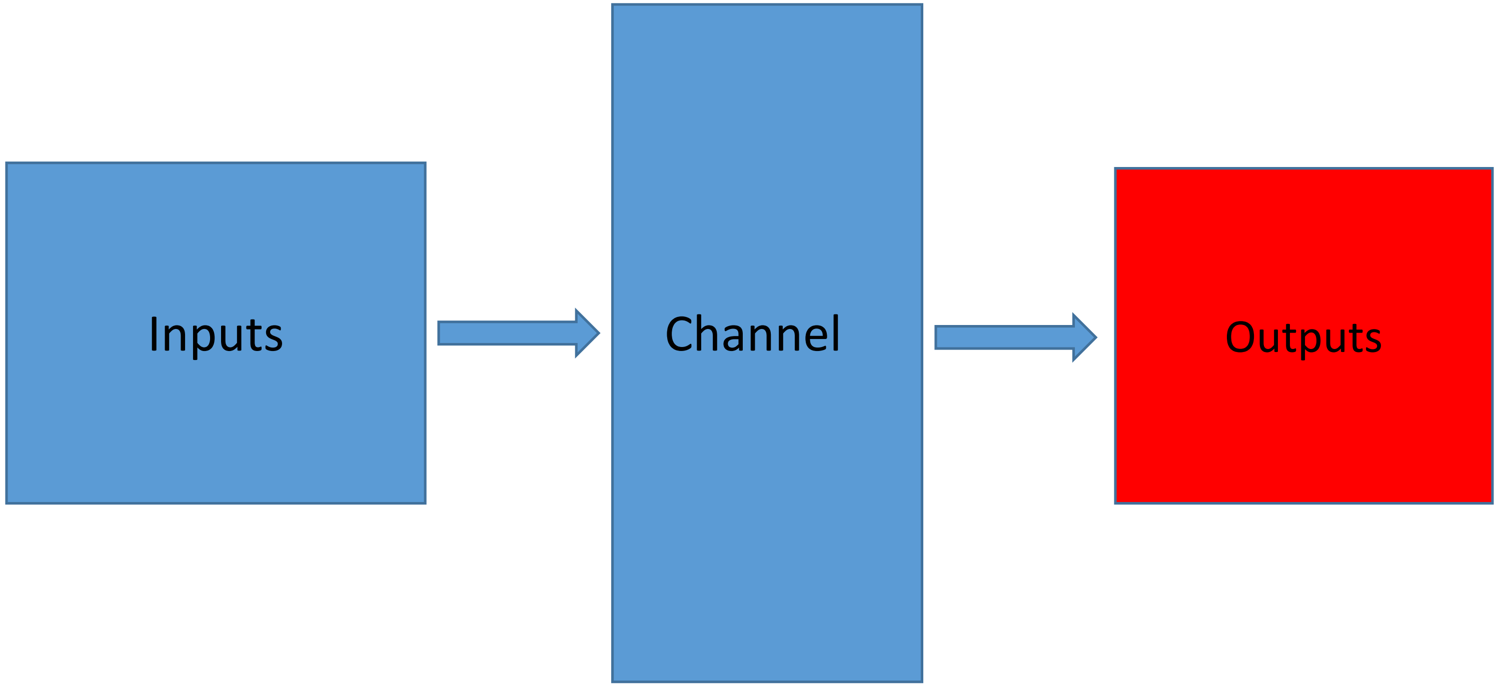


This was the century where the US became the preeminent manufacturing economy through innovations such as the assembly line

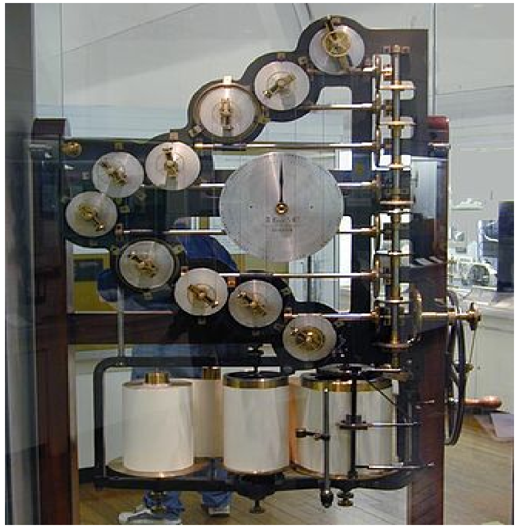
Tinkering was the norm

It was a century where the physical world was expressed in the language of mathematics, and analog computers were developed to scale specific mathematical calculations.

Calculating Outputs



1872 to 1879



Tide-Predicting Machine

A special-purpose mechanical analog computer conceived by William Thomson (later Lord Kelvin) to predict the ebb and flow of sea tides and the irregular variations in their heights.

Thomson had introduced the method of Fourier analysis to tidal patterns in the 1860s.

His machine automates the summation of 20 or more trigonometrical terms and can be seen at the Science Museum, South Kensington, London

Improving Naval Gunnery

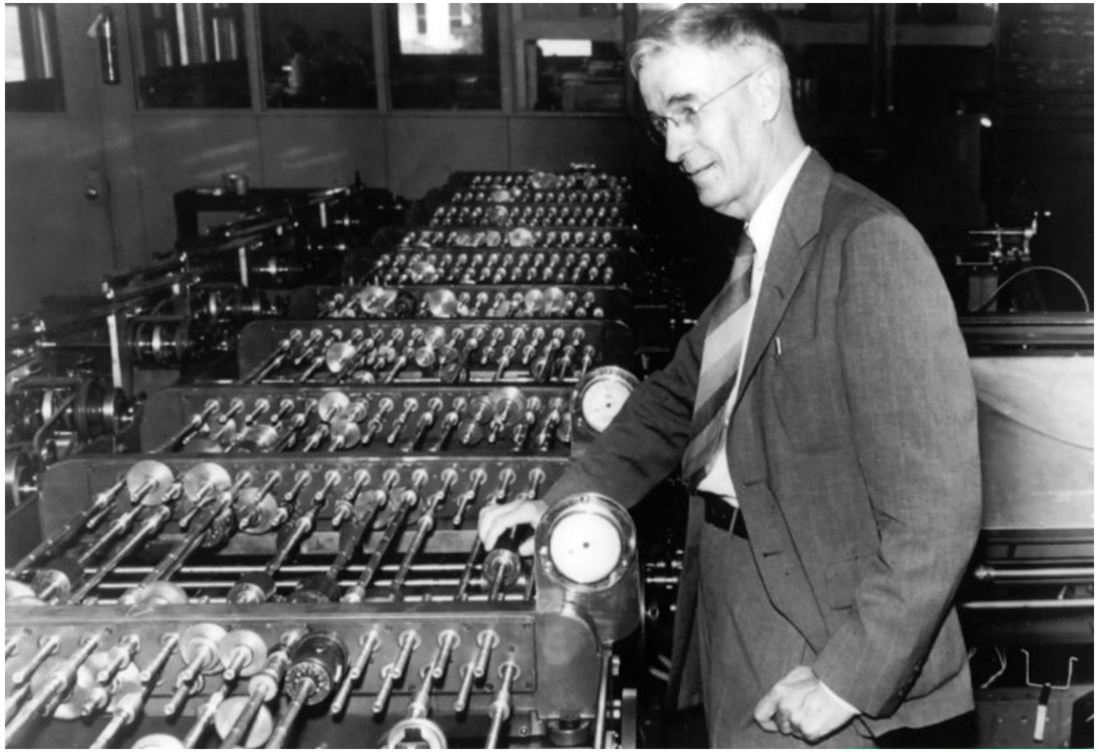
1915 to 1990

Fire-Control Systems enabled remote and automatic targeting with or without the aid of radar or optical sighting.

In 1917, Hannibal Ford, a mechanical engineer from upstate NY developed a machine that automated a key step in the solution of differential equations – this was an analog computer for calculating specific integrals.

Given inputs such as the speed and course of attacking ships, the *Baby Ford* would generate the range to the target, the direction of fire and the angle of the guns.

Vannevar Bush with his Differential Analyzer



This *Differential Analyzer* was an analog computer with some digital components that could solve differential equations with as many as 18 independent variables

Vannevar Bush was an engineer, inventor and science administrator, who during World War II headed the US Office of Scientific Research and Development. He joined the Department of Electrical Engineering at MIT in 1919, and founded the company now known as Raytheon in 1922.

The Computer Before the Digital Revolution

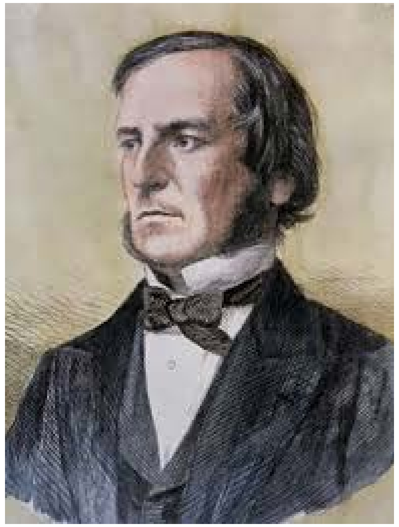
“*A fearsome thing of shafts, gears, strings, and wheels rolling on disks*”

A machine that literally performed the equations in the process of solving them

However with each new equation, the differential analyzer had to be deconstructed and reassembled

Bush dreamed of an analyzer that might be programmable, and he brought Claude Shannon to MIT to develop a control plane – a little box of 100 switches

Shannon not only knew how to design circuits, he also knew Boolean logic – he had learned it in a philosophy class at the University of Michigan



1937 Master’s Thesis

“I think I had more fun doing that than anything else in my life”

Any circuit is represented by a set of equations corresponding to the various relays and switches of the circuit. A calculus is developed for manipulating these equations by simple mathematical processes, most of which are similar to ordinary algebraic algorithms. This calculus is shown to be exactly analogous to the calculus of propositions used in the symbolic study of logic …. The circuit may then be immediately drawn from the equations.

Synthesis of a target circuit:

$x'y'z + x'yz + xy'z + xyz' + xyz$

becomes

$yz(x+x') + y'z(x+x') +xyz'$

becomes

$yz + y'z + xyz'$

becomes

$z(y+y') + xyz'$

becomes

$z + xyz'$

becomes

$z +xy$

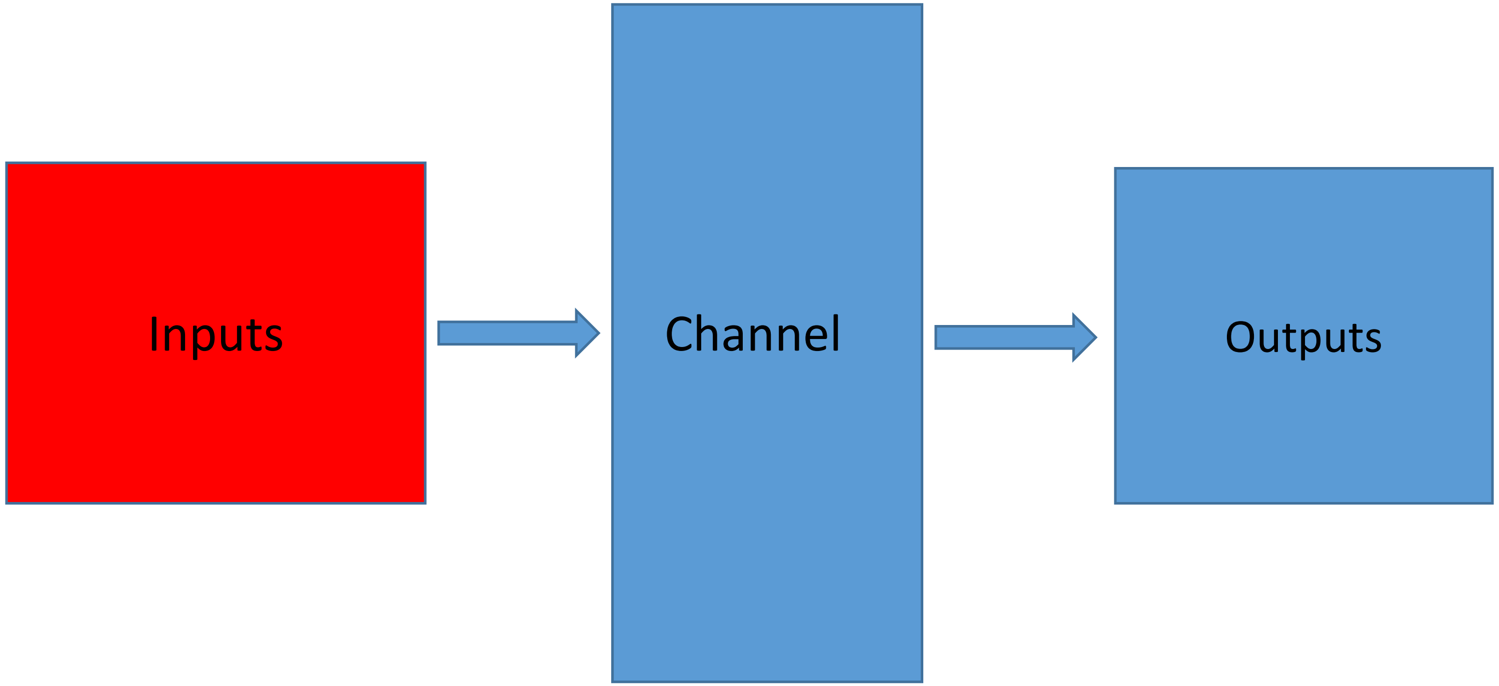
The Dawn of Digital Computing

A decade later the differential analyzer is effectively obsolete, replaced by digital computers where vacuum tubes implemented thousands of logic gates

1937 is also the year that Alan Turing shows that any solvable mathematical problem could in principle be solved by machine.

And in World War II he builds a machine that breaks codes

The Art of Measurement

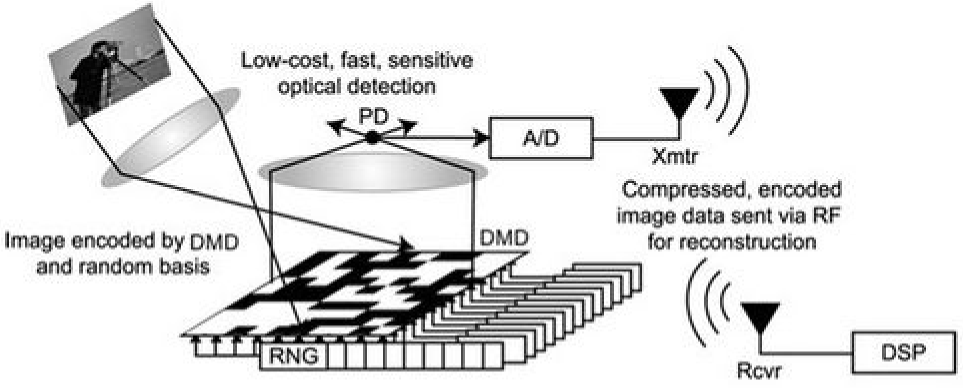


Measurement: Ancient and Modern

1935 Development of weighing designs by Yates

1949 Development of the Multi-Slit Spectrometer by Golay

2008



Compressed Sensing: The Rice Single Pixel Camera reconstructs images by measuring pixels in combination

Golay and Multi-Slit Spectrometry

Far Infrared Spectroscopy identifies molecules by detecting the characteristic absorption frequencies of specific chemical bonds

1949 Spectrometer with spinning disks and slits encoding Walsh functions

1951 Spectrometer with fixed slits encoding Golay complementary pairs

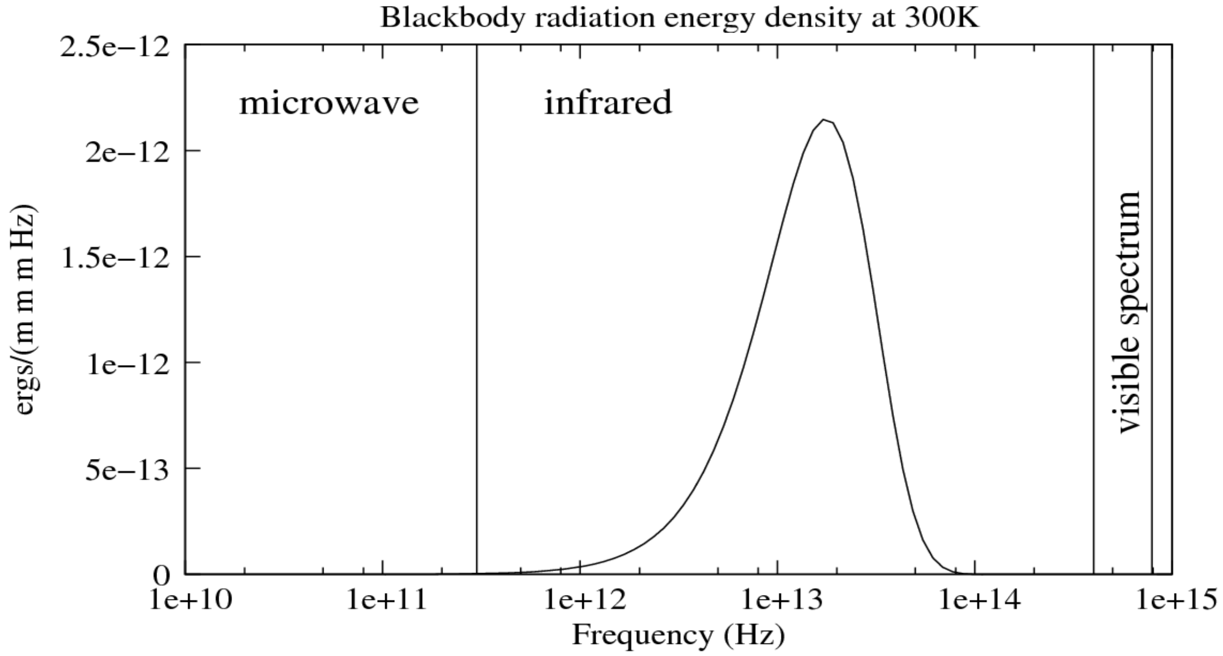
1952 Pro. IRE Survey Paper *Bridges Across the Infrared Radio Gap*

Obstacles to Infrared Spectroscopy

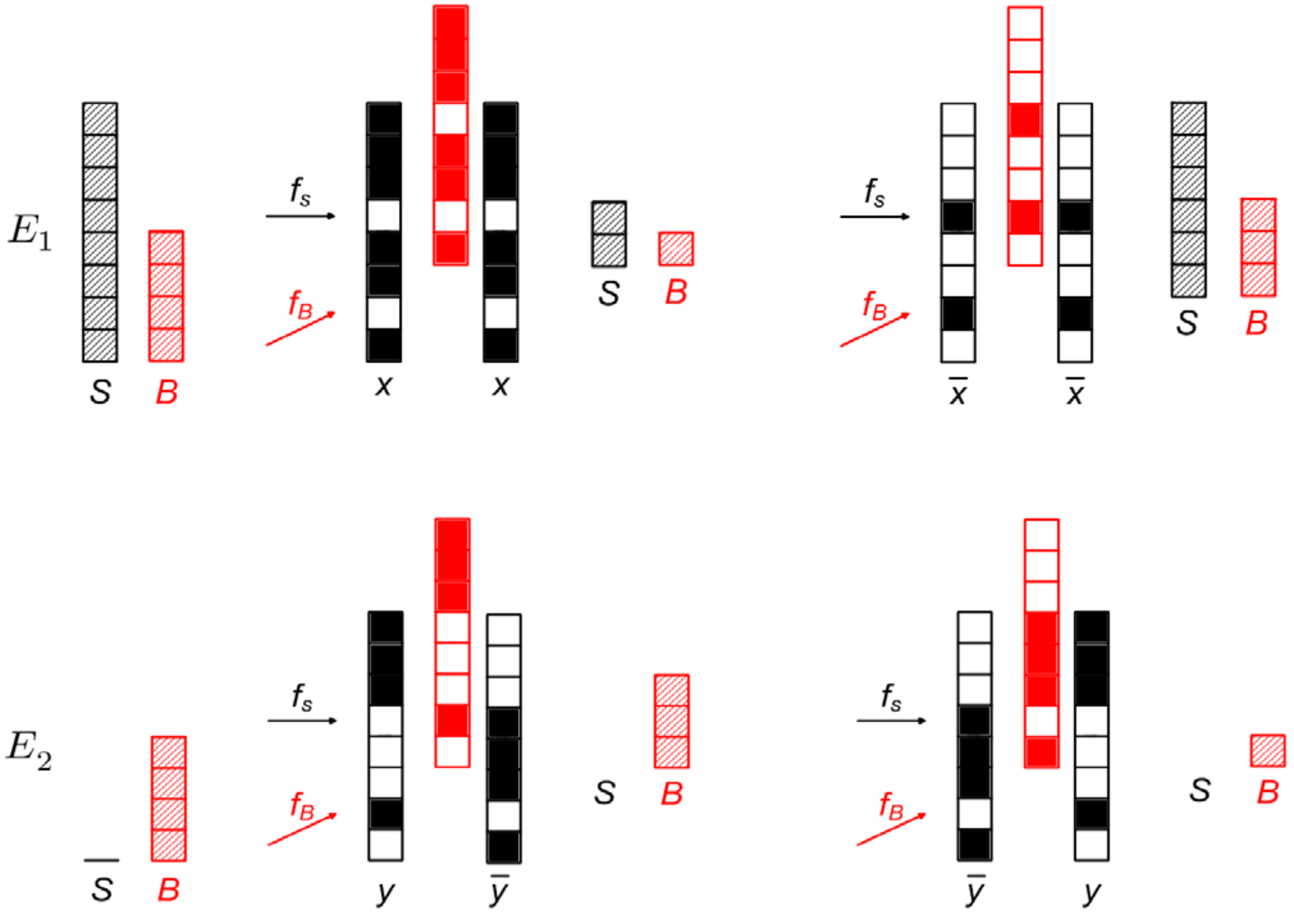
Sources of interest are typically small, thus emit and absorb weakly

Blackbody radiation from the environment and the equipment itself is strongly concentrated in the infrared spectrum and overlaps the signal of interest

Detectors were temperature sensors that could not by themselves distinguish between different frequencies of infrared radiation, but merely integrated thermal energy received.



Measurement with Golay Complementary Pairs



$$E_S = E_1 - E_2$$

since

$$R_x(k) + R_y(k) = 2L\delta_{k,0}$$

The Concept of Entropy

Look at a compact disc under a microscope and you will see music represented as a sequence of pits, or in mathematical terms as a sequence of 0s and 1s, commonly referred to as bits.

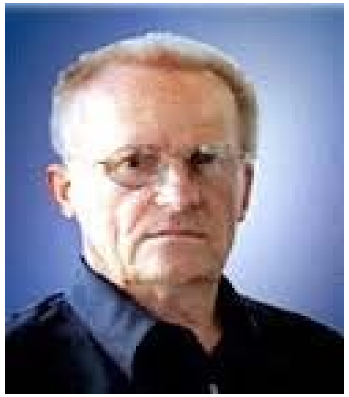
The foundation of our Information Age is this transformation of speech, audio, images and video into what we now call digital content.

Nyquist and Hartley had defined information as choice from a set of symbols

Shannon thought in terms of measuring the uncertainty that is overcome

When we communicate in English there are rules that restrict our freedom to choose the next letter

Shannon used the frequencies of letters to randomly generate text, then introduced memory – n-gram models for letters and words – and recorded the passage from gibberish to passable text



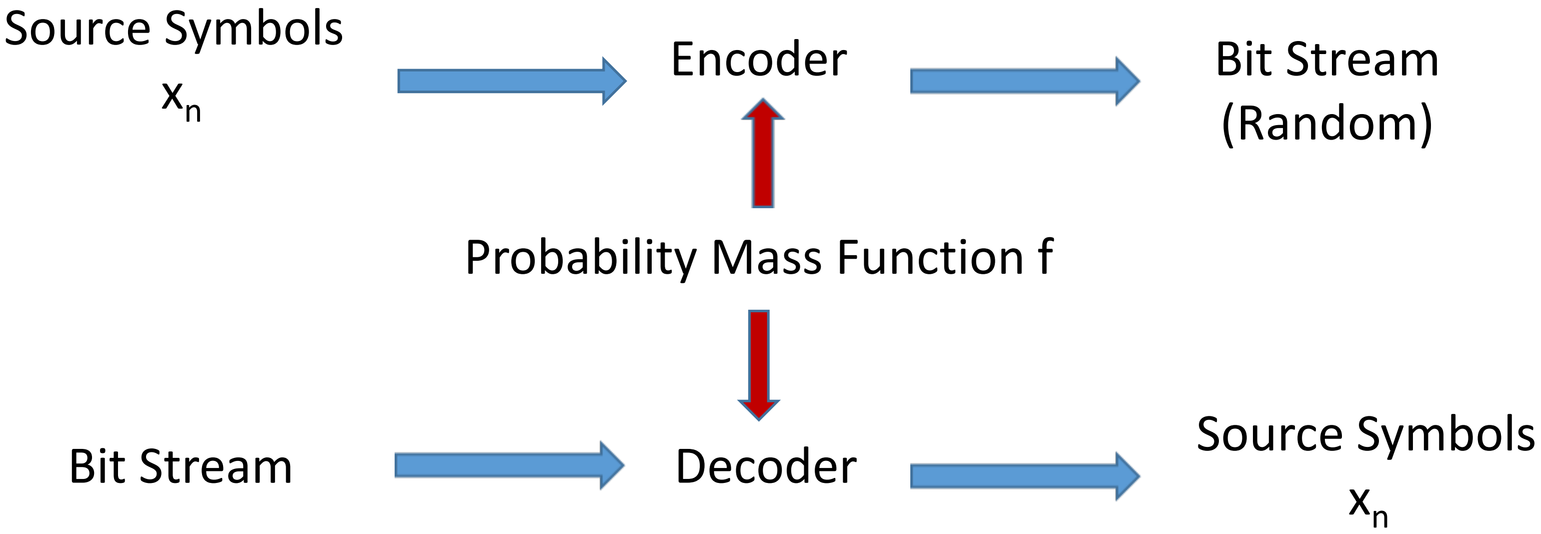
Arithmetic Coding

Anticipated by Elias, invented by Rissanen, rendered practical by Witten

Arithmetic codes are widely used in modern image and video compression

- JBIG, JPEG, JPEG-2000, H.263, …..

Encoder and decoder share a probabilistic model for the source



Source Symbols x_n

Encoder

Bit Stream (Random)

Probability Mass Function f

Bit Stream

Decoder

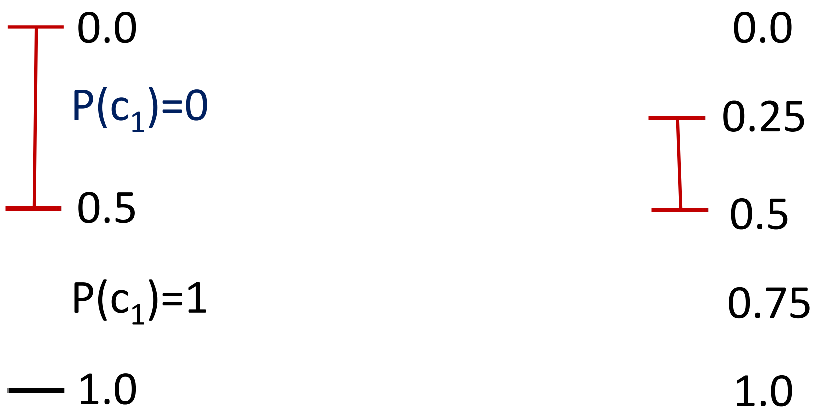
Source Symbols x_n

Arithmetic Coding in Action

Example: Compressing a random codeword from C = < (0011), (1100) >

The arithmetic encoder represents the codeword as an interval within the real interval [0, 1)
- the length of the interval is the probability of the string, given the model

(0011)



0.0

$P(c_1)=0$

0.5

$P(c_1)=1$

1.0

0.0

0.25

0.5

0.75

1.0

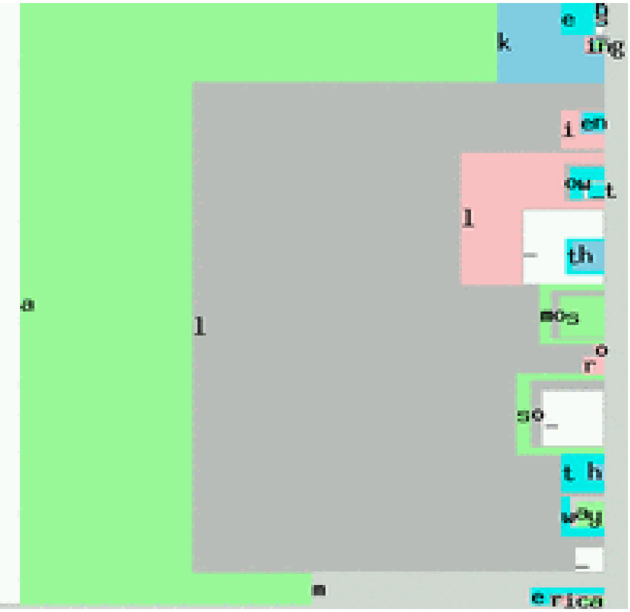
[0.25, 0,5)

$c_1 = 0$ $c_1 = c_2$ $c_3 = 1$ $c_4 = c_3$

Source symbols are encoded as a dyadic interval, which is represented by a binary string

Dasher

Dasher is a text entry system developed by David MacKay that enables an individual who might have only one muscle to type 30 words per minute

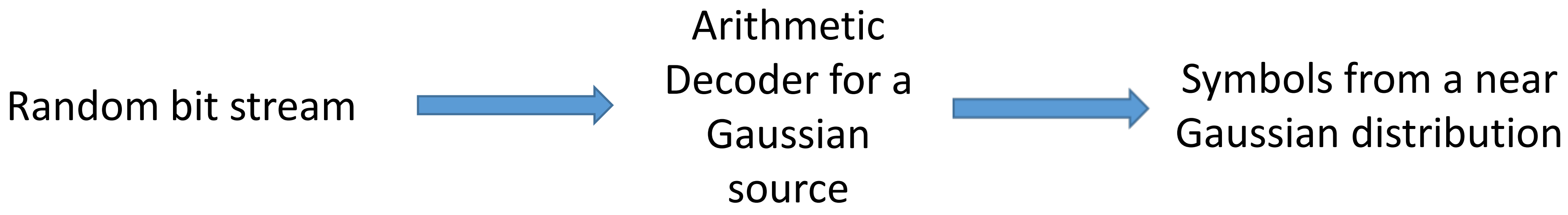


https://www.youtube.com/watch?v=nr3s4613DX8

Arithmetic Coding in Reverse

Shannon proved that Gaussian inputs are optimal for communication over a Gaussian channel

Arithmetic Coding provides an efficient way of generating an approximately Gaussian distribution



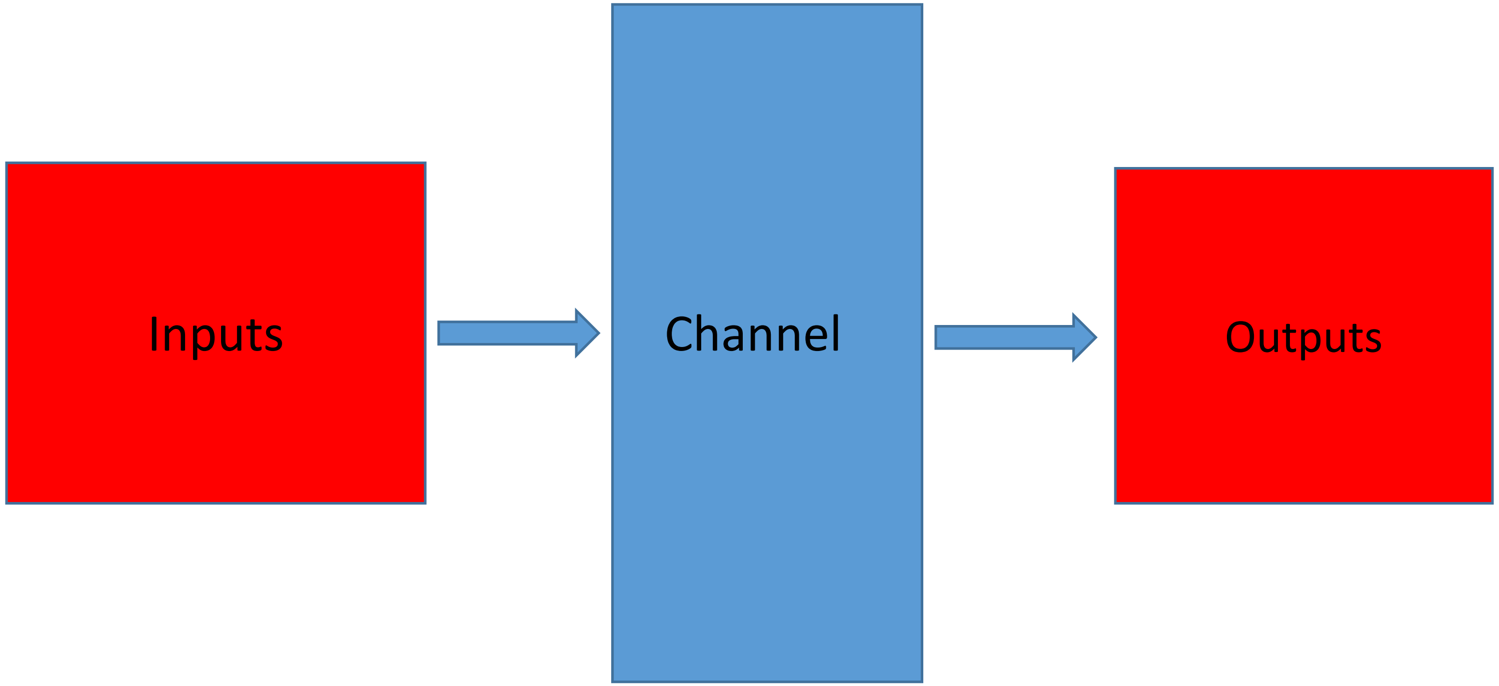
Random bit stream

Arithmetic Decoder for a Gaussian source

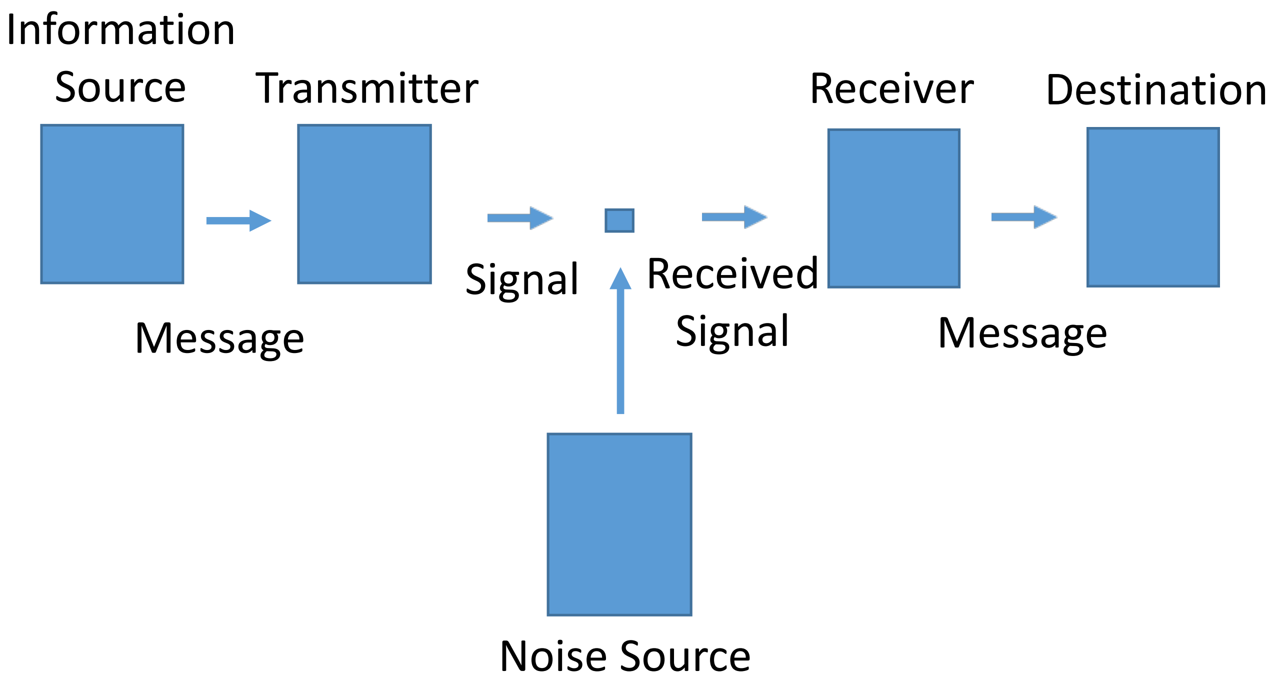
Symbols from a near Gaussian distribution

This is an example that illustrates the duality of compression and communication

The Art of Signaling



A Mathematical Theory of Communication



The *information source* produces the message

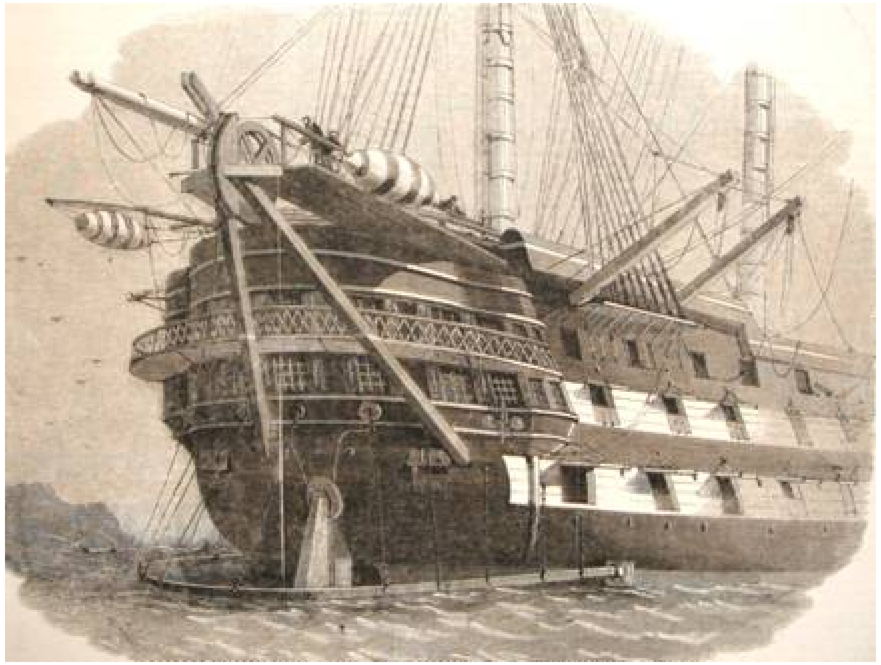
The *transmitter* encodes the message in a form capable of being sent as a signal

The *channel* is the medium through which the signal passes

The *noise source* represents the errors that distort or corrupt the signal on the way to its destination

The *receiver* decodes the message, reversing the action of the transmitter

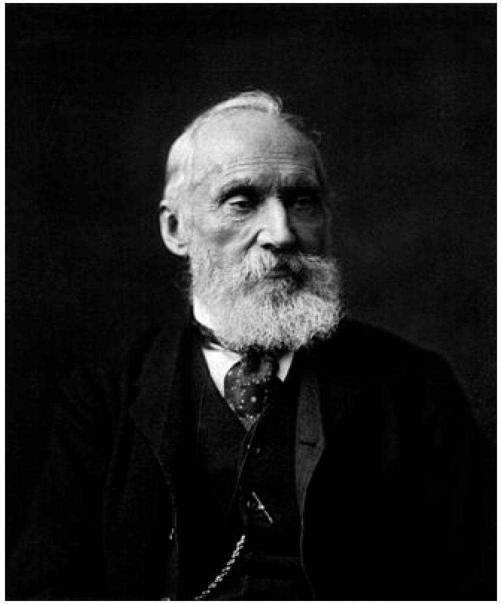
The *destination* is the recipient of the message



Speaking Louder

Before Shannon it was commonly believed that the only way of achieving arbitrarily small probability of error in a communication channel , such as a telephone line, was to increase signal energy or to reduce the transmission rate to zero.

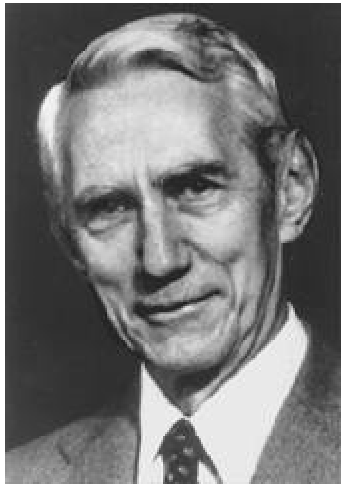
William Thomson
(Lord Kelvin)



In 1858 Kelvin disagreed with Wildman Whitehouse, the Electrician of The Atlantic Telegraph Company, about how to operate the new transatlantic telegraph cable.

Kelvin advocated more sensitive detection

Whitehouse thought that electricity needed to be given a bigger kick – *and fried the cable*



Speaking More Redundantly

Shannon characterized a channel by a single parameter, the *channel capacity*, and showed that it was possible to transmit encoded information at any rate below capacity with arbitrarily small probability of error.

His method of proof was to show the existence of a single good code by averaging over all possible codes

He established fundamental limits on the efficiency of communication through noisy channels without producing an explicit example of a good code that would achieve the highest capacity.

It took more than 50 years for coding theorists to discover families of codes that come close to these fundamental limits

In the Beginning

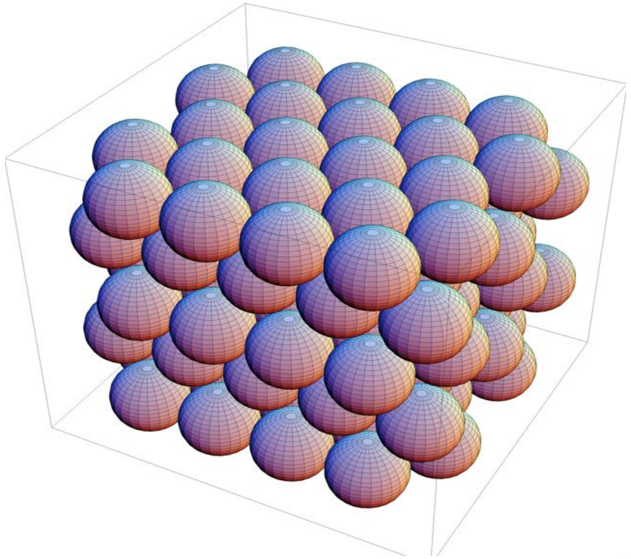
Fisher, Hamming and Golay

1954 Reed Muller Codes

1960 Reed Solomon Codes

Shannon established fundamental limits on what was possible, setting up the challenge of finding specific families of codes that achieve capacity.

And coding theorists turned to discrete geometry



Symmetry

How many spheres can touch a given sphere in N-dimensional Euclidean space?

$\tau(2) = 6$: spheres centered at hexagonal lattice points

1694

$\tau(3) = 12$: Newton

2017

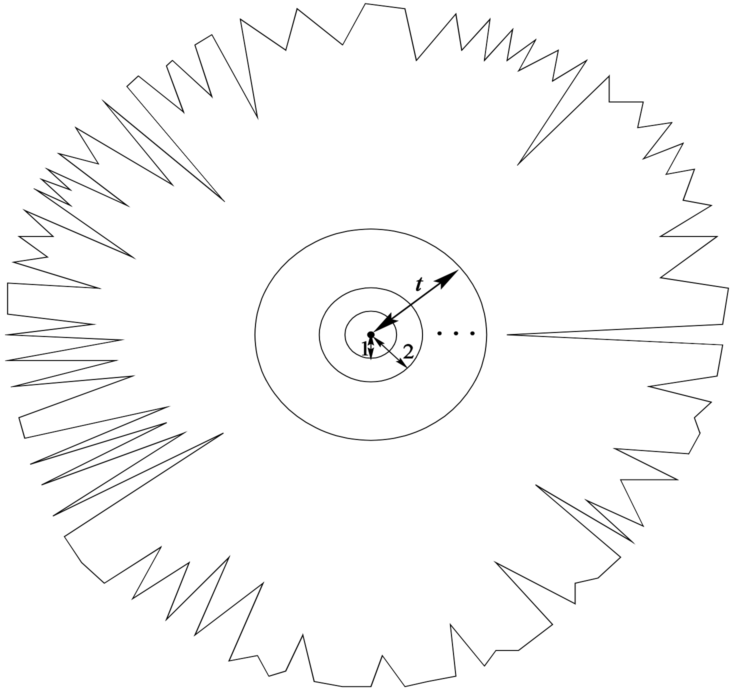
$\tau(8) = 240$: Maryna Viazovska
Spheres are centered at E_8 lattice points

David McKay

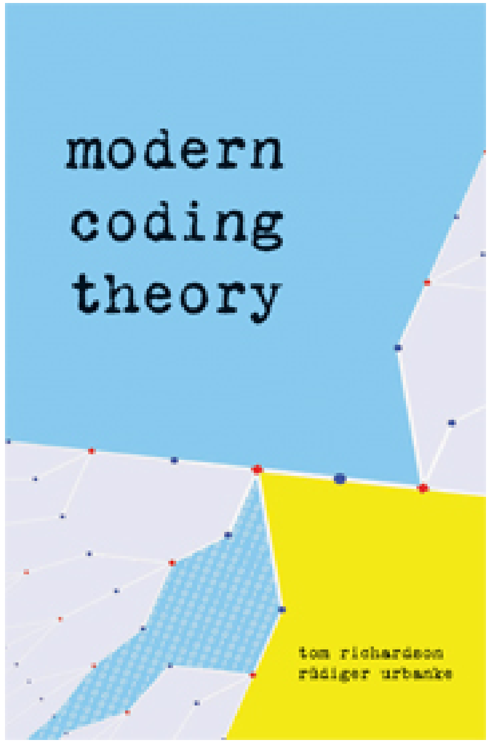
While sphere-packing is a fascinating mathematical subject, the aim of maximizing the distance between codewords in a code has only a tenuous relationship to Shannon’s aim of reliable communication

Berlekamp’s Bats: Hamming space in the vicinity of a codeword typically looks like this – the t-sphere around the codeword takes up a small fraction of the space.

The implication is that we are able to tolerate a small number of low weight codewords



Iterative Decoding



Gallager (1960 PhD Thesis) So far ahead of its time that it had to be rediscovered

As computation becomes more possible, coding theory changes character & symmetry fades from consciousness

1993 **Turbo Codes:** Discovery by Berrou, Glavieux and Thitimajshima

1995 **MacKay and Neal:** Good codes based on very sparse matrices

1998 **Luby, Mitzenmacher, Shokrollahi, and Spielman:** Analysis of low-density codes and improved designs using irregular graphs

Symmetry & Analysis of Algorithms

Felix Klein: His 1872 Erlangen Program classifies geometries by means of the underlying symmetry groups

Proof by sleight of hand

EXIT Functions: If a sequence of codes achieves capacity on erasure channels under bit-MAP decoding then the EXIT functions must exhibit a sharp transition from 0 to 1 at the erasure value 1-r.

Talagrand (1994): Monotone Boolean Functions invariant under a transitive permutation group exhibit a sharp 0-1 transition

Reed Muller Codes achieve capacity on erasure channels under symbol-MAP decoding: EXIT functions of linear codes with doubly transitive permutation groups can be characterized in terms of symmetric monotone Boolean functions

Beyond Double Transitivity

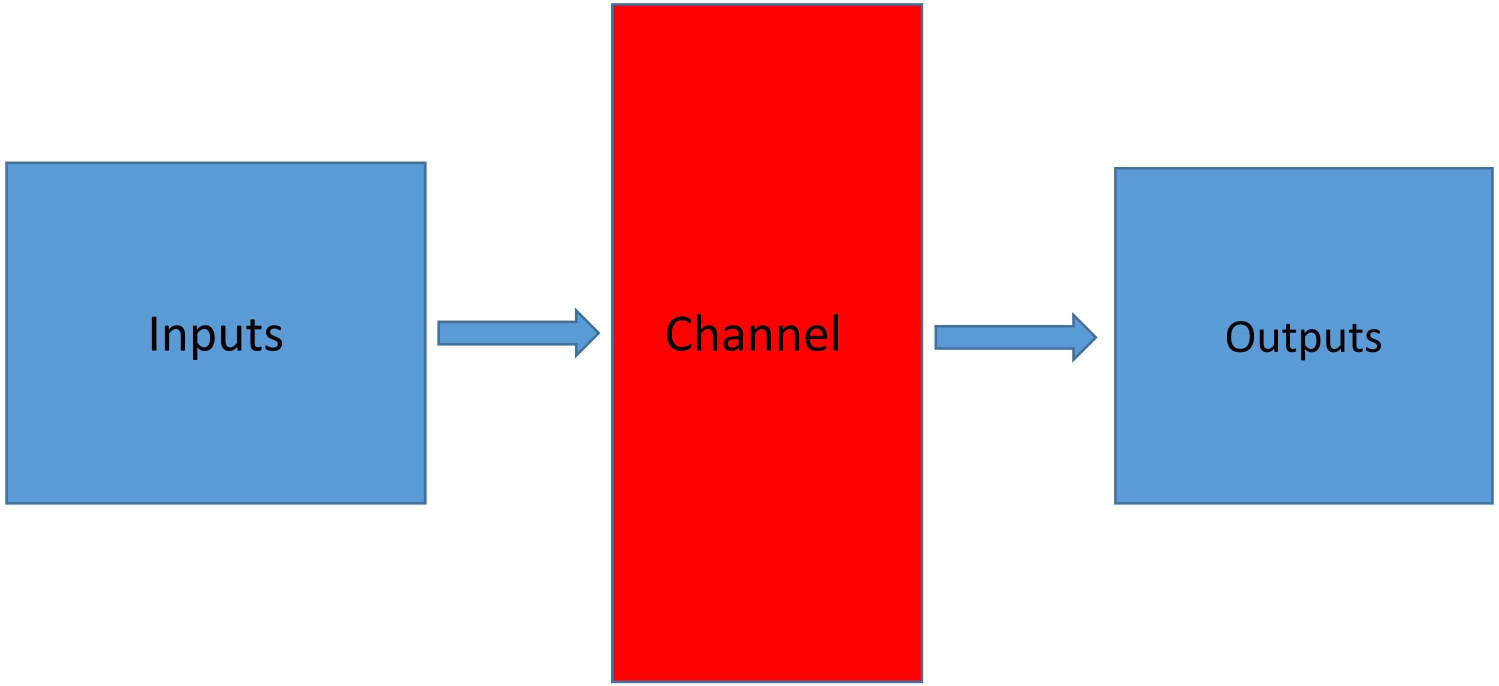
ITW 2016: Santhosh Kumar, RC, and Henry Pfister, Beyond Double Transitivity: Capacity-Achieving Cyclic Codes on Erasure Channels.

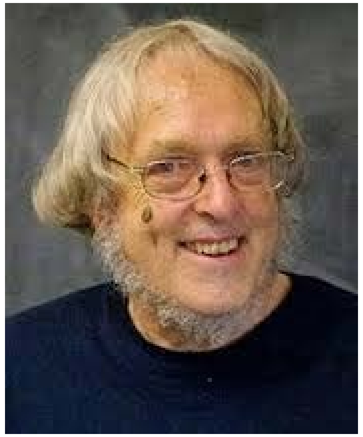
Proof: For a sequence of codes to achieve capacity on erasure channels under bit-MAP decoding, the EXIT functions must exhibit a sharp transition from 0 to 1 at the erasure value 1-r.

Apply Talagrand's lower bound on influences for monotone Boolean functions to show that a monotone Boolean function whose symmetry group does not have small orbits must exhibit a sharp transition.

Consider cyclic codes over GF(q) of length N(n) dividing $q^{t(n)} – 1$ where the smallest prime factor of t(n) increases without limit.

Identifying the Channel





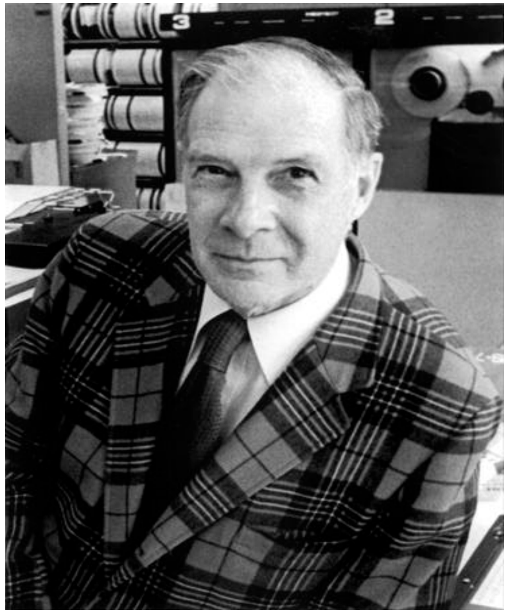
P.J. Cameron, Some Bridges between Codes and Designs

Discovery of the Hamming Codes

1942

R.A. Fisher discovers the simplex or first-order Reed Muller code in the course of his work on design of experiments

“Using blocks of 2^m plots it is possible to test all combinations of so many as N factors in such a way that all interactions confounded (2^{N-m} in number) shall involve not less than three factors each”



1950

R.W. Hamming discovers his family of perfect single error correcting codes with block length $N=2^m – 1$ and dimension $k = N-m$

Fractional Factorial Designs

Fisher investigated how N factors might affect the yield of some process, assuming that the yield is only affected significantly by interactions of small numbers of factors

Suppose for simplicity that the ith factor can take one of two levels, say 0 or 1

Fisher tested treatment combinations lying in a subgroup B of Z_2 x ….. x Z_2

The subgroup C, dual to B, represents combinations of treatments that are confounded in the experiment – if $e_i + e_j + e_k$ is in C, then the main effect of factor I cannot be distinguished from the interaction of factors j and k.

Fisher worked at Rothamsted Experimental Station between 1919 and 1933

Comparing Codes and Designs

In a **Factorial Design** we want

Large Weight in C so that potentially significant combinations of factors are not confounded

Few Trials so B should be small
- Note however that $|C| = 2^m / |B|$

In an **Error Correcting Code** we want

Large Weight in C so that codewords can be distinguished at the output of a Binary Symmetric Channel

Large Number of Codewords in C so that more information can be communicated

Hamming Codes Satisfy Both Specifications

Duality between Packing and Covering

Philippe Delsarte explored the geometry of error-correcting codes in his 1973 PhD thesis

An Algebraic Approach to the Association Schemes of Coding Theory

His thesis transformed coding theory

It provided the mathematical foundation for the linear programming bounds derived by McEliece, Rodemich, Rumsey and Welch, also those derived by Levenshtein.

The Future

Is Hard to Predict

But the way that Shannon worked will continue to produce breakthroughs:

Start with an engineering challenge of operational significance, derive a simple model that captures essential features, identify mathematical quantities that govern the influence of inputs on outcomes, and finally derive fundamental limits on what is possible.

Looking Back at Source Coding: Stochastic models are used to describe the dynamics of an information source, and it is the entropy rate of the stochastic process that determines the minimum number of codewords that can represent the source to within a target fidelity.

Looking Back at Channel Coding: The mutual information between channel input and channel output that determines the maximum rate at which communication is possible.

Shannon Joined Bell Laboratories in 1941

1876 Alexander Graham Bell invents the Telephone

1919 Rotary Dial

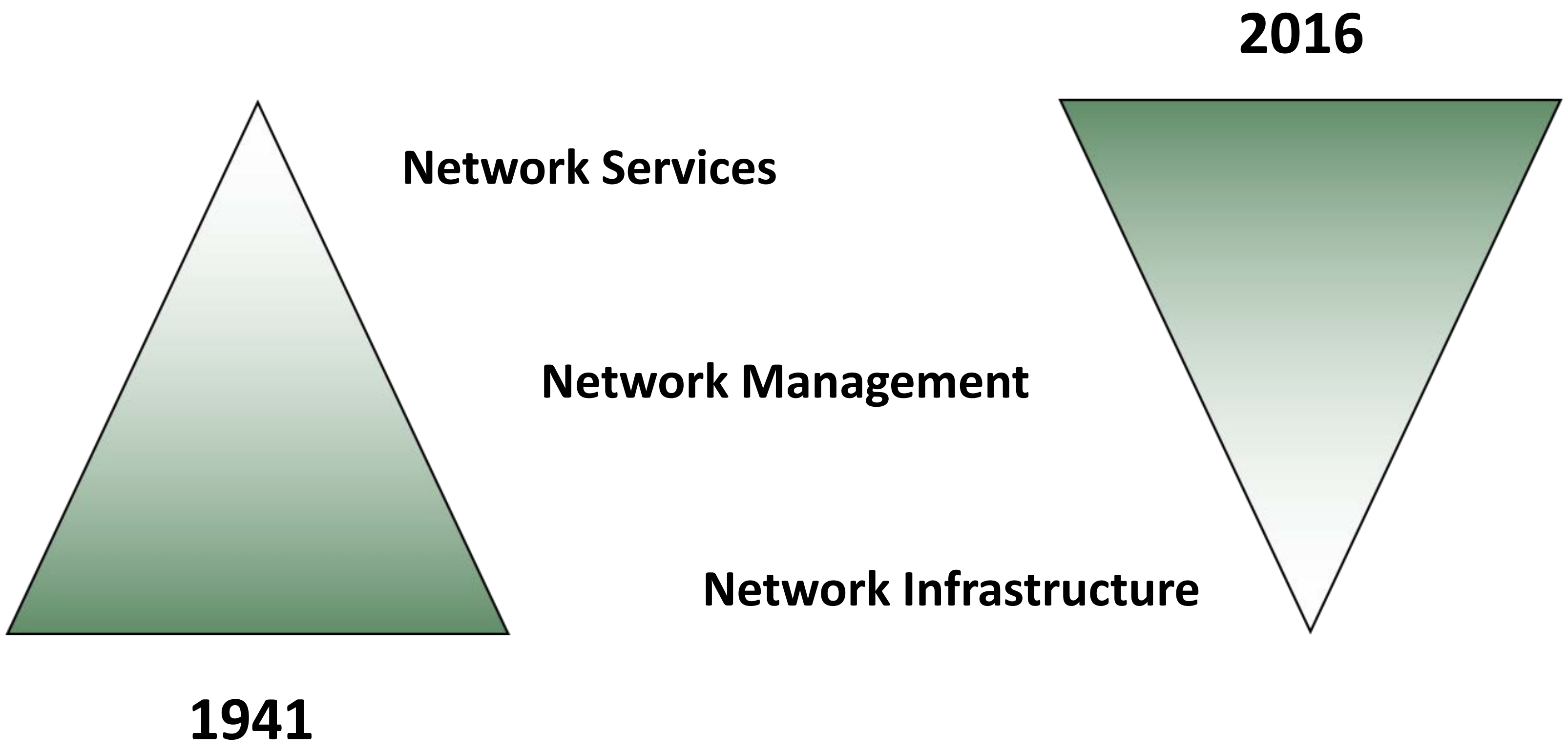
1925 Bell Laboratories Founded

1927 Transatlantic Service via Radio

1951 Direct Dial

It was a unique research environment sustained by vertical integration of service infrastructure, network/service/customer management and network infrastructure within a monopoly business model

Telecommunications has Changed Dramatically



2016

Network Services

Network Management

Network Infrastructure

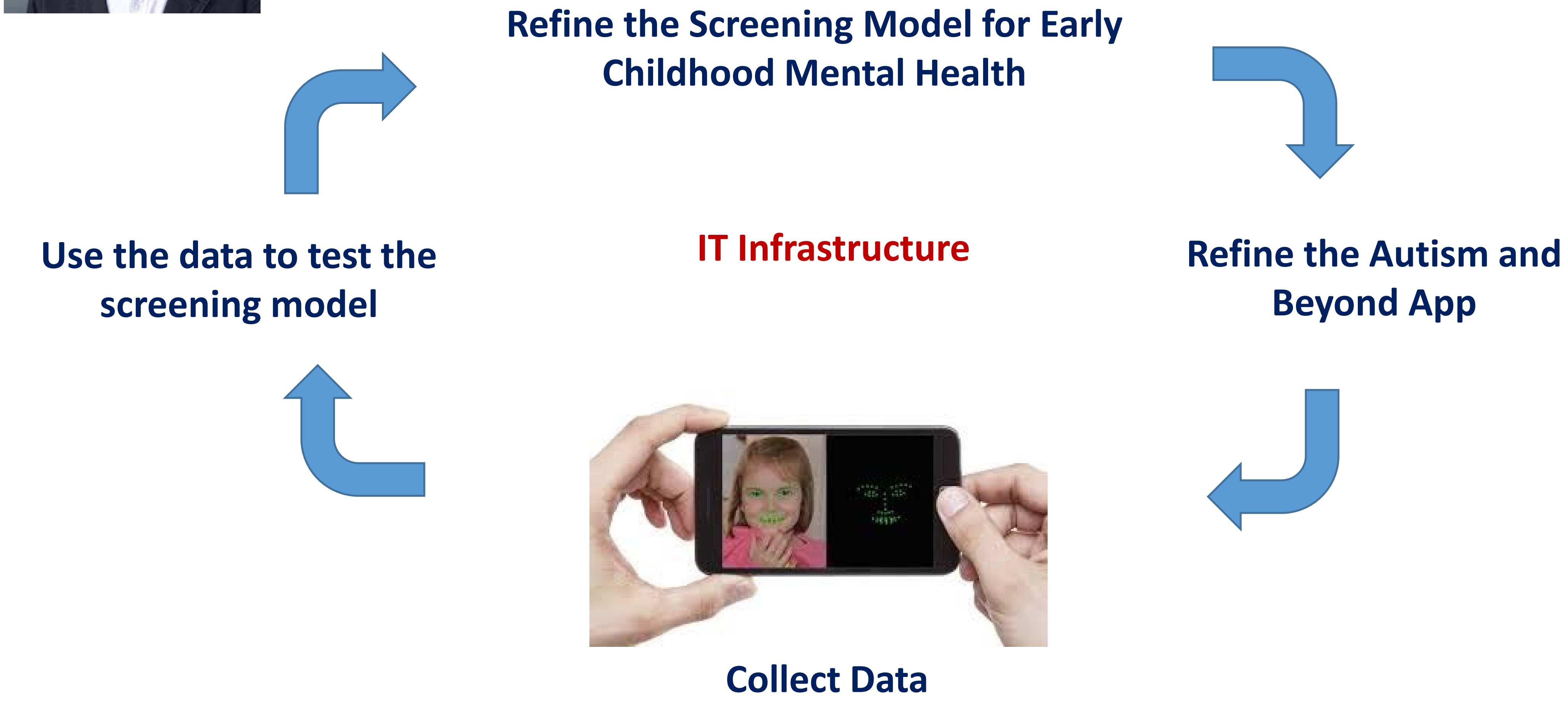
1941

Today the most important research component is software.

Followed by the science of operations at massive scale

Traditional communications research has become less important

Modern Network Services Assume Connectivity



Refine the Screening Model for Early Childhood Mental Health

Refine the Autism and Beyond App

IT Infrastructure

Collect Data

Use the data to test the screening model

And Interference is more important than Noise

Communication over wireless channels that are limited by interference rather than noise

New Network Models: For example, the deterministic model proposed by Avestimehr, Diggavi and Tse (2007)

New Communication Techniques: For example, methods of interference alignment proposed by Cadambe and Jafar (2008)

Compression of massive tables when there are dependencies among the fields

The Lempel-Ziv Algorithm (1978) provides an *asymptotically optimal* solution to the problem of universal data compression and it finds application in the Unix file compression utility *compress*

How then to configure compression to take maximal advantage of these dependencies in the non-asymptotic regime?

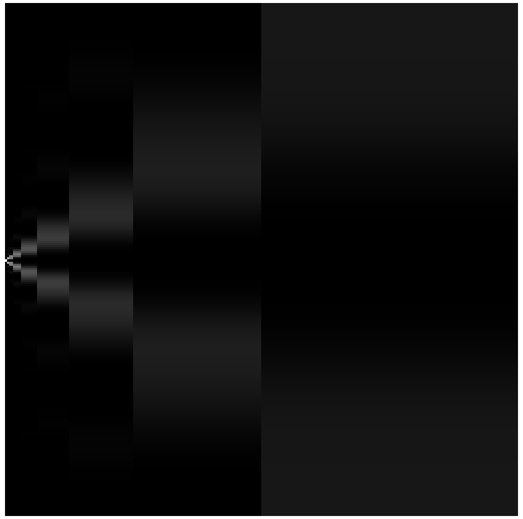
Measurement is being Transformed by Sparse Models

And perhaps it was image compression and the role played by wavelet representations that encouraged interest in sparse models

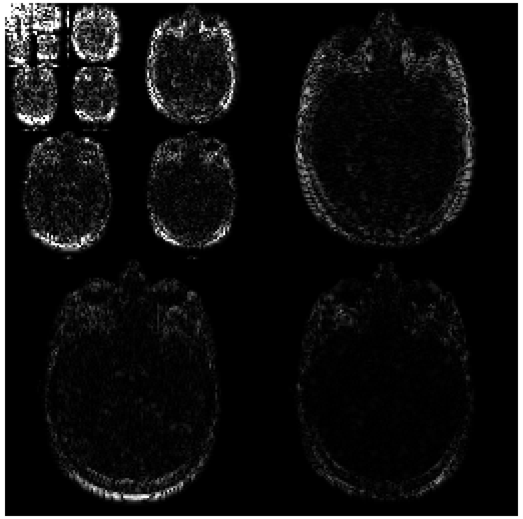
MRI scan performed at Cambridge University Hospitals (GE 1.5T)

that Direct Undersampled Signal Recovery

Optimizing resolution in MRI by being more discriminating about where samples are taken

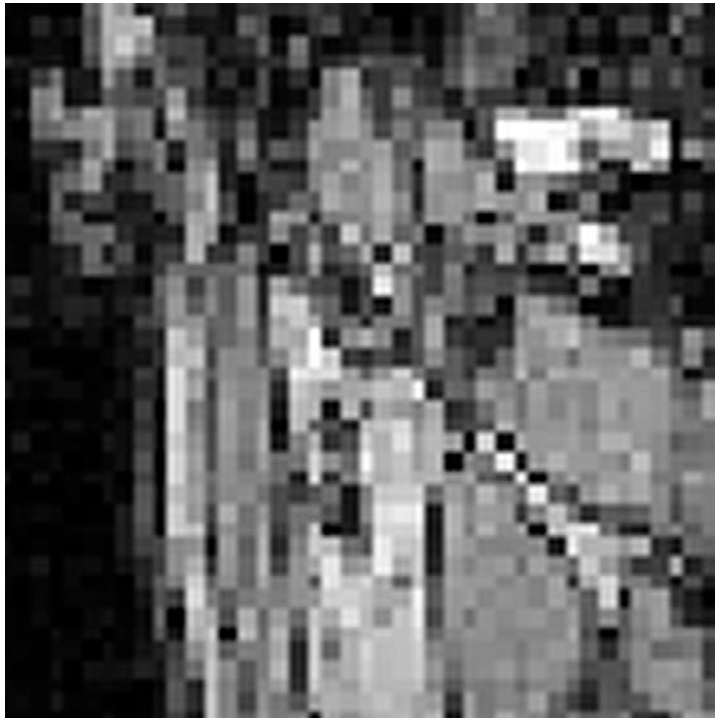


Inner products between measurement basis (Fourier) and reconstruction basis (db4)

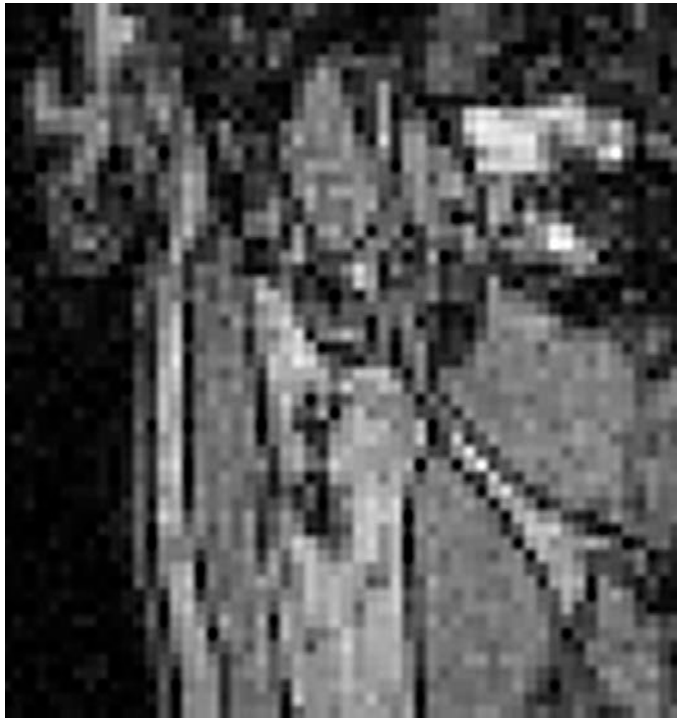


Db4 wavelet scales and coefficients

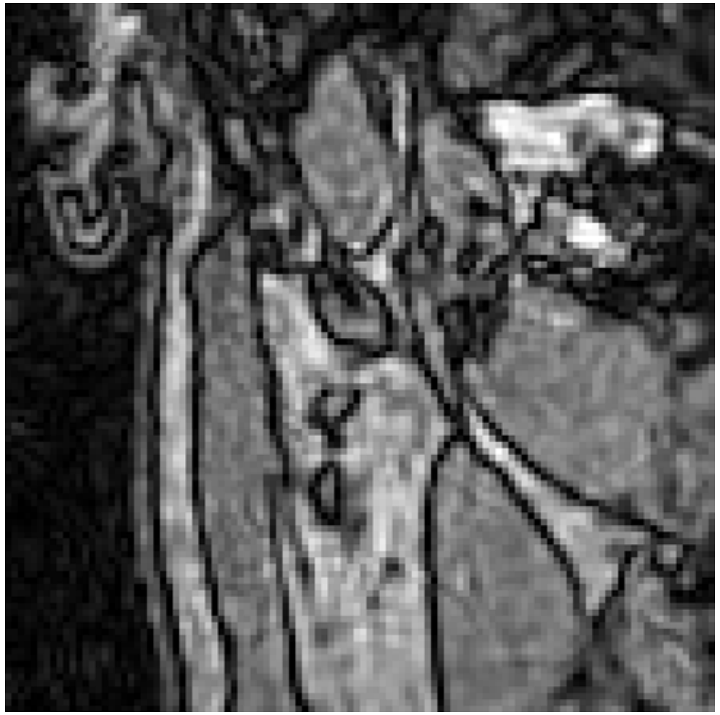
and Enhance Resolution in Undersampled Physical Imaging



Acquisition Time **T**
Fully Sampled: **1.2 x 1.2 x 1.2 mm**



Acquisition Time: **3T/2**
Fully Sampled: **0.8 x 0.8 x 0.8 mm**



Acquisition Time **T**
Subsampled: **0.4 x 0.4 x 0.4 mm**

Bogdan Roman and Martin Graves: University of Cambridge, UK

Thank You for Your Attention