Information

Introduction to Information Theory

Basic concepts of information collecting and processing systems

- Review the theory
- Examine examples

工学系研究科・機械工学専攻
Jean-Jacques Delaunay
jean@mech.t.u-tokyo.ac.jp

Outline

1. What is information?
2. Where does information come from?
3. What is a message?
4. Where does the noise come from?
5. How does noise affect information transmission?

What you need to remember

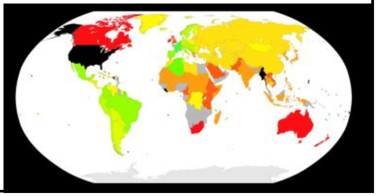
- Definition of information [bit] is universal
- Measurements are always <u>indirect</u> and therefore require <u>calibration</u>
- <u>Noise</u> causes <u>random errors</u>

International system of units (Abbreviated SI)

The seven based units of SI:
meter [m], kilogram [kg], second [s], ampere [A], kelvin [K], candela [cd], mole [mol]

Other units are derived from the based units

SI is not adopted by Burma, Liberia and the US!



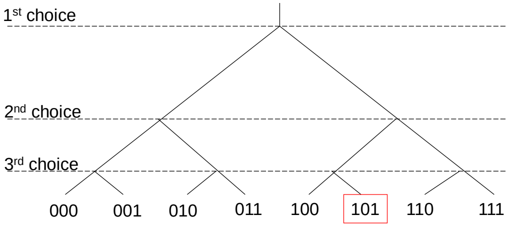
1. What is information?

Information

- Information is a measure of **order**
- Measure of order:
  - Universal measure applicable to any structure, any system (physical laws)
  - Order quantifies the instructions needed to produce a certain organization (binary choices)

Minimum number of binary choices

8 symbols of equally probable states



Three correct consecutive choices
=> message identifying '101' contains 3 bits of information

(bit : binary digit)

Compute the information in any given arrangement from the number of choices we must make to arrive at that particular arrangement.

I am thinking of one person. How many questions do you need to ask to find out?

Information definition

For any number of possibilities *N*, the information *I* for specifying a member

$$I = \log_2\left(\frac{1}{N}\right)$$

- Information is a dimensionless entity
- $I \le 0$, information has to be acquired in order to make the correct choice

Information entropy (Shannon)

1/4

Uncertainty is related to the logarithm of possible symbols (equally likely symbols)

$$H = \log_2(N)$$

Where *N* is the number of equally likely symbols

General formula for uncertainty (non equally likely symbols)

For an infinite string which is formed from an alphabet of M symbols, the averaged uncertainty is

$$H = -\sum_{i=1}^{M} p_i \log_2(p_i) \quad \text{[bits per symbol occurrence]}$$

Symbol $i = 1,\ldots,M$ whose outcome has probability $p_i; \sum_{i=1}^{M} p_i = 1$

Example

When throwing a fair die, the probability of 5 is 1/6.

$$H = -\log_2(P) = -\log_2(1/6) = 2.58 \text{ bits}$$

When 2 dices are thrown independently, the amount of information associated with
thrown dice 1 = 5 AND thrown dice 2 = 4 is

$$P = P_1 \times P_2 = 1/6 \times 1/6 = 1/36$$

$$H = -\log_2(P) = 5.17 \text{ bits}$$

We can also write

$$H = H_1 + H_2 = 5.17 \quad \text{where } H_1 = H_2 = 2.58 \text{ bits}$$

Information entropy: Shannon uncertainty equation

- M the number of available symbols; same frequency of occurrence (equally likely symbols)

The probability of a particular symbol appearing is $P = \frac{1}{M}$

The amount of information per symbol is $I = \log_2\left(\frac{1}{M}\right) = \log_2(P)$

- The total amount of information in a string of length N formed from an alphabet of M symbols is

$$I = N\log_2(P)$$

Information entropy: Shannon uncertainty equation

- Set of M symbols $x_1, x_2, ..., x_M$

Infinite string $N \to \infty$ formed from the M symbols; symbol x_i appears N_i times

Probability of symbol x_i is $p_i = \frac{N_i}{N}$

- Information provided by the x_i symbol is $I_i = N_i \log_2(p_i) = Np_i \log_2(p_i)$
- The total information contains in string N is $I = \sum_{i=1}^{M} Np_i \log_2(p_i)$
- Every time x_i appears, it provides $\frac{I_i}{N_i}$ information per symbol occurence
- The amount of information per symbol occurrence, averaged over all symbols

$$\frac{I}{N} = \sum_{i=1}^{M} p_i \log_2(p_i)$$ [bits per symbol] Shannon uncertainty equation

Information entropy: Rate of information transmission

Amount of information that knowledge about the outcome of an event adds to someone's knowledge

Self information is a decrease in uncertainty *H* before and after communication

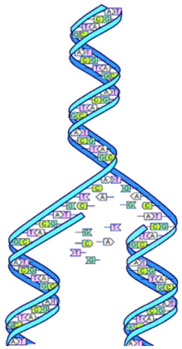
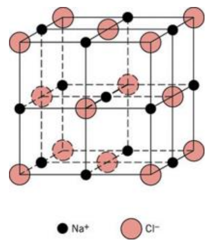
$$R = \underbrace{H_{before}}_{\substack{number\,of\,symbols \\ \equiv \log_2(N)}} - \underbrace{H_{after}}_{transmission\,error}$$

If there is no noise, the information communicated is H_{before}

Information is defined as a decrease in uncertainty

Information in the physical world

Living matter	Nonliving matter
Macromolecules -> complex	Mineral world (e.g. NaCl)
DNA is made of 4 bases (letters) arranged in specific sequences, with a total of 10^9 bases (length) => 4^{10^9} possibilities	Array of chloride and sodium atoms
Less likely to form spontaneously	Salt crystal assembles itself from its atoms in solution

DNA replication	NaCl crystal	New York Times
		
$I = -\log_2(4^{10^9})$	$I = -\log_2(1) = 0$ at $T = 0$ K, no impurities	$I \approx -\log_2\left(2^{81000}\right)$ 1 page ≈ 1000 characters 1 character ≈ 8 bits

Entropy in statistical mechanics

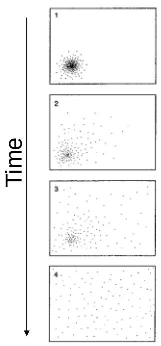
- Measure of the dispersal of energy ("disorder")
- Measure of "disorder"
  - For a system in thermodynamic equilibrium, the number *N* of equivalent ways (microstates) a system can be constructed:

Entropy $S = k_B \ln(N)$

where k_B is the Boltzmann constant, $k_B = 1.3806 \times 10^{-23}$ J/K

- For a perfect crystal at $T = 0$ K, $N = 1$, $S = 0$
- For a liquid phase, the number of equivalent ways for assembly gets very large

Dissipation of order and information



Smoke introduced into a closed chamber
Molecules are initially concentrated near the source
=> significant order and information

Structure being lost. System evolves toward more probable states

Macroscopic structure is lost. Thermodynamic equilibrium.
=> the most probable state, information content zero, entropy maximum

Information content: displacement of probability with respect to the thermod. equilib.

Information and entropy of communication theory

- Entropy $H = \log_2(N)$
- Information $I = \log_2\left(\frac{1}{N}\right)$
- Information is equivalent to entropy

$$\text{Conservation law}: \delta\,\text{information} + \delta\,\text{entropy} = 0$$

An increase in entropy implies a decrease in information, and vice versa.

Summary

Communication theory: measure of order

n binary questions/choices — *N* possibilities/symbols

$$\text{Information [bit]} = -n = \log_2\left(\frac{1}{N}\right) \qquad (2^n = N)$$

$$\text{Entropy [bit]} = \log_2(N)$$

Statistical mechanics: measure of disorder

N number of equivalent ways a system can be constructed

$$\text{Entropy[J/K]} = k_B \ln(N)$$

2. Where does information come from?

Where does information comes from?

Information comes from sensors or transducers:

Sensors/transducers generate responses which can be measured. This measurement creates information.

=>Instrumentation and measurement form the basis of all practical information systems

Transducers

A device that converts one form of energy into another.

Example: an electric motor is a transducer that converts mechanical energy into a voltage (or vice versa).

Transducers are important components in many types of sensor => convert the physical quantity to be measured into a voltage.

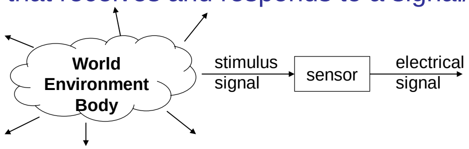
Transducers include sensors and actuators

-Sensors: Devices which transform a stimulus into an electrical signal (thermometer)

-Actuators: Devices which transform an input signal into motion (electrical motor)

Sensor

A device that receives and responds to a signal/stimulus.



Biological sensor	Man-made sensor
Electrochemical character	Physical, chemical, biochemical characters
Ion transport in nerve fibers	Information is processed and transported in electrical form

Instrument

An instrument senses some form of input and responds by producing an appropriate output.

(Examples of instrument: balance, voltmeter, oscilloscope, PC, TV set…)

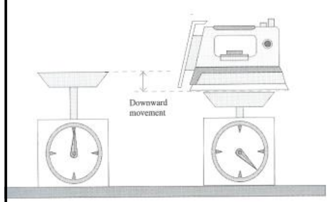
Important properties:

-Calibration
-Response time
-Range
-Noise

=> Measurement errors

=> Amount of information is limited

Measurement



Compression force from the squeezed spring balances the force from the increased weight.

We observe:
-pointer rotates (120 degrees)
-pan moves down (2 cm)

We do not measure weight directly. (indirect indication)
=> we have to calibrate the balance using known weights.
=> we compare the weight of the iron with a set of standard weights.

All measurements are comparisons with some defined standard.

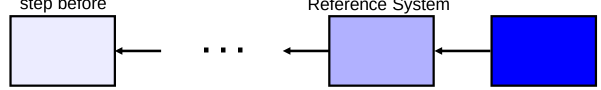
Calibration process

Where did the known weights come from?

End User — System was calibrated in the step before

Makers — Calibrate a system against the Primary Reference System

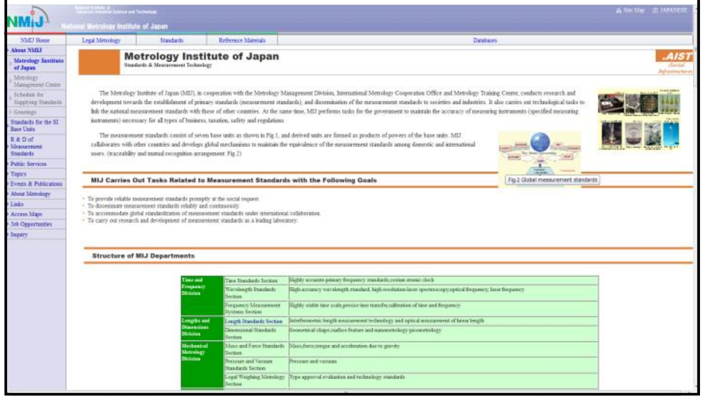
National standard Laboratory — Primary Reference System



Define what we mean by “1 kg”, “1 sec”.

Traceability: The ability to trace the calibration of a device through a chain of calibrations to a primary standard. Traceability does not guarantee accuracy…

The Metrology Institute of Japan



Calibration of measurement devices always necessary?

Calibration standards are always necessary, however special techniques may be applied to use un-calibrated devices. For example:

- double weighting measurements (un-calibrated balance, Borda XVIII century)

- reciprocity calibration for microphones (piezoelectric)

Double weighting measurements

M (unknown mass) is measured two times (left and right side). We write the two equilibrium conditions (torque balance):

$$\begin{cases} Mgr_2 = M_1 g r_1 \\ Mgr_1 = M_2 g r_2 \end{cases}$$

$$M^2 g^2 r_2 r_1 = M_1 M_2 g^2 r_1 r_2$$

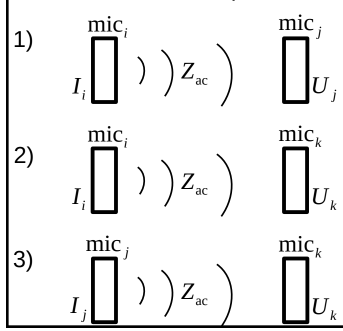
$$M = \sqrt{M_1 M_2}$$

g gravitational acceleration

In this measurement, the instrument is not calibrated. M_1 and M_2 are used as the weight standards.

Reciprocity calibration (1/2)

The technique exploits the reciprocal nature of transduction mechanisms such as the piezoelectric effect (microphones). Uncalibrated microphones *i*, *j*, and *k* are used. The microphones are placed facing each other with a well known acoustic transfer impedance between their diaphragms.



acoustic transfer impedance Z_{ac}

current measurement I
$\rightarrow$ source of sound

voltage measurement U
$\rightarrow$ response to pressure wave

elecrical impedance

$$Z_{i,j} = \frac{U_j}{I_i} = M_i Z_{ac} M_j$$

Reciprocity calibration (2/2)

$$\begin{cases} Z_{i,j} = \dfrac{U_j}{I_i} = M_i Z_{ac} M_j & i \to j \\ Z_{i,k} = \dfrac{U_k}{I_i} = M_i Z_{ac} M_k & i \to k \\ Z_{j,k} = \dfrac{U_k}{I_j} = M_j Z_{ac} M_k & j \to k \end{cases}$$

$$\begin{cases} Z_{i,j} Z_{i,k} = M_i^{\,2} Z_{ac}^{\,2} M_j M_k \\ Z_{j,k} = M_j Z_{ac} M_k \end{cases}$$

$$\begin{cases} M_i = \sqrt{Z_{i,j} Z_{i,k}} \Big/ Z_{ac} \sqrt{M_j M_k} \\ Z_{j,k} = M_j Z_{ac} M_k \end{cases}$$

$$M_i = \sqrt{\frac{Z_{i,j} Z_{i,k}}{Z_{j,k} Z_{ac}}} = \sqrt{\frac{1}{Z_{ac}} \frac{Z_{i,j} Z_{i,k}}{Z_{j,k}}}$$

The electrical transfer impedance is determined by measuring currents and voltages only.

Accuracy and precision

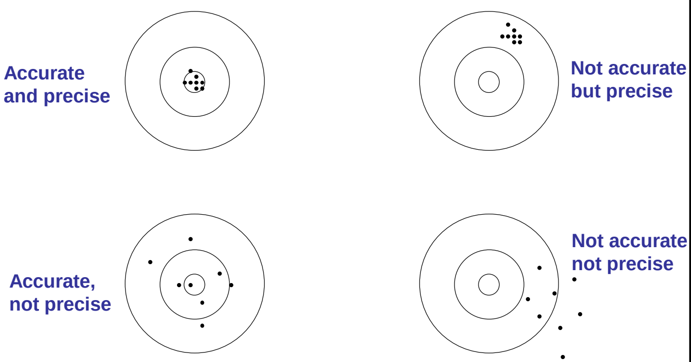
We can never make perfect measurements with absolute accuracy or precision:

1) Noise effect (thermal noise, shot noise…)

2) Calibration process

=> The amount of information we can collect is always finite.

Accuracy and precision

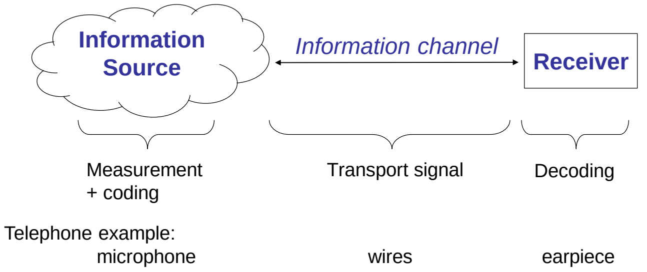


Summary

- Information is collected by instruments which perform some kind of measurement.
- All measurements are comparisons with a set of standards.
- The amount of information we can collect is always finite, limited by the effects of noise, saturation and response time.

3. What is a message?

Signal and message



Information is in the form of a signal which carries a message.

Communication requires a code

Before a signal can be used to communicate some specific information in the form of a message, the sender and the receiver must have agreed on the details of how the actual signals are to be used (code).
=> Distinguish one code symbol from another

Communicate information requires:
Addresser
Addressee
Channel
Code
Context
Message

Code includes Error Detection -Parity bit-

We use redundancy to check for transmission errors

Odd parity bit: the parity bit of each word can be set to 1 when it contains an even number of 1.

Parity = 0: odd numbers of 1
Parity = 1: even numbers of 1

Before Transmission: Original word 8 bit word	Before Transmission: Parity bit added 9 bit word	Parity	After Transmission: After transmission 9 bit word	Parity
10100100	010100100	0	010100100	0
01010101	101010101	1	101010101	1
11101111	011101111	0	011101111	0
11101111	011101111	0	011111111	1

Parity bit (Redundant information)

Transmission error

Note that there exists two types of parity bits: even parity bit and odd parity bit.

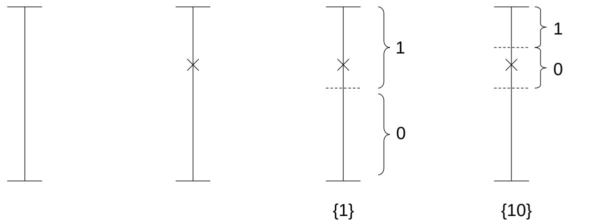
Conversion into binary numbers

Define a signal range (wide enough to ensure that the signal is inside the chosen range)

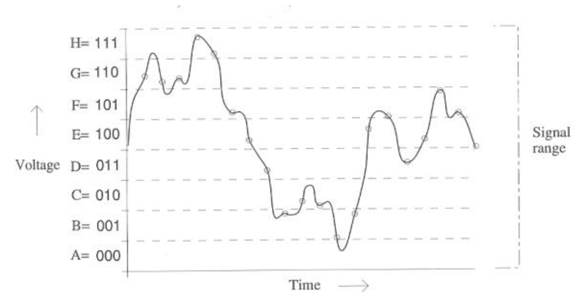
Sample

Conversion to a 1-digit binary

Conversion to a 2-digit binary



Sampling



Value indicated by the nearest available binary number.

Signal level moves out of the initial range: signal is said to have been clipped

Examples of analog-to-digital conversions

Image sensors

Oscilloscope (DAQ card)

-Signal amplitude is coded using 12 bits for each pixel
-Acquisition rate is 10 frames per second

-Signal amplitude is coded using 8 bits
-Sample rate is 2 Giga points per second
-bandwidth 100 MHz

How much information in a message?

Bit:

Yes/no question; minimum possible amount of information.
Ask *n* questions => *n* bits of information (coded by 2^n distinct symbols)
Note: ask an extra question double the number of symbols, but it doesn't provide twice as much information.

2 bits ($2^2 = 4$ symbols) -> 3 bits ($2^3 = 8$ symbols)
=> 3/2 as much information

Total amount of information contained in a message:

Information = Number of samples × number of bits

Number of samples = measurement duration / sampling time

Number of bits = $\log_2$(number of symbols)

Capacity of an information channel

The amount of information that can be communicated through a channel is limited by the properties of that channel:

- Range
- Noise
- Response time (bandwidth)

Effect of noise

The amount of information an analog channel can convey is limited by the noise level.

- Channel maximum range 1 V; noise level 1 mV

Measurement with an Analog to Digital Convertor (ADC):

Set voltage range to 1 V

8-bit ADC => 1 V / 2^8 = 1/256 = 0.0039 V

10-bit ADC => 1 V / 2^{10} = 1/1024 = 0.00097 V ≈ noise

> 10-bit ADC => obtain NO extra information

Effect of noise: Dynamic range

Definition of the Dynamic range: the ratio of maximum signal amplitude to the minimum amplitude detectable

Signal has a finite dynamic range
=> Amount of information limited by the noise level
(the effect seen in the previous slide is a consequence of the signal having a finite dynamic range)

$$\text{Dynamic Range} = 10\log\left(\frac{P_{max}}{P_{noise}}\right) \quad [\text{decibels}]$$

where P_{max} is the maximum power of the signal

P_{noise} is the mean noise power.

Effect of noise: Signal to noise ratio

Signal to Noise Ratio (SNR) in decibels

$$\text{SNR} = 10\log\left(\frac{P_{actual\,signal}}{P_{noise}}\right)$$

Note: SNR is different from the dynamic range ($P_{actual\ signal} \le P_{max}$)

$$\text{SNR} = 10\log\left(\frac{V_{actual\,signal}^2}{V_{noise}^2}\right) = 20\log\left(\frac{V_{actual\,signal}}{V_{noise}}\right)$$

$$\text{Dynamic Range} = 10\log\left(\frac{V_{max}^2}{V_{noise}^2}\right) = 20\log\left(\frac{V_{max}}{V_{noise}}\right)$$

where

$V_{actual\,signal}$ actual signal rms

V_{max} maximum signal rms

V_{noise} noise rms

Effect of noise: Decrease in transmission rate

4/4

Two equally likely symbols transmitted every second :

1) Case without noise (no transmission error)

$H_{before} = 1, H_{after} = 0 \Rightarrow R = 1$ bit/sec

2) Case with noise

probability that a 1 is correctly transmitted is 0.99

probability that a 0 is correctly transmitted is 0.99

$H_{before} = 1$

$H_{after} = -(0.99\log_2(0.99) + 0.01\log_2(0.01)) = 0.081$

$\Rightarrow R = 1 - 0.081 = 0.919$ bit/sec $<$ 1 bit/sec

Before transmission	After transmission
1 or 0 H = log2(2) = 1 bit	**Without noise** -We receive a 1 P(1) = 1 => H=log2(1)=0 P(0) = 0 -We receive a 0 P(1) = 0 P(0) = 1 => H=log2(1)=0 **With noise** -We receive a 1 P(1) = 0.99 P(0) = 0.01 -We receive a 0 P(1) = 0.01 P(0) = 0.99

Effect of response time

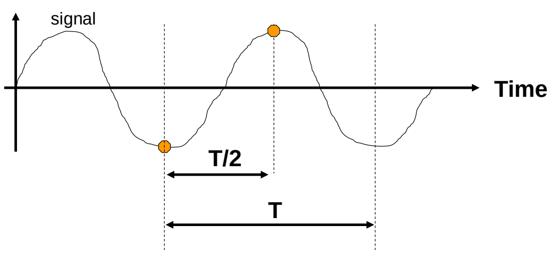
1/2

Limitation on how quickly the voltage can be changed (finite response time of any system).

Example: wires take 1 μs to respond to a change.
Sampling at a frequency > 1 MHz do NOT provide extra information.

Maximum signal frequency the wires can carry?
One cycle, one up and down: 1 μs + 1 μs = 2 μs
=>maximum frequency = 0.5 MHz

Sampling rate = 2 × maximum frequency



Time:
Signal period: T
Sampling: T/2

Frequency:
Signal frequency 1/T
Sampling frequency 2/T

Sampling frequency = 2 × maximum signal frequency

2/2

Bandwidth

Bandwidth: the range of frequencies a channel can carry.

Usually bandwidth = maximum frequency (range of frequencies extends to dc)

Sampling rate = 2 × Bandwidth

Summary

- Information system consists of an information source connected to a receiver through an information channel.
- A set of information is a message which is sent as a signal using a code made up of symbols.
- The amount of information a channel can carry is limited by its range, the level of noise and its response time.

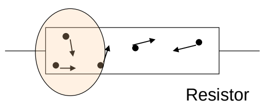
4. Where does noise come from?

Noise sources

Noise type	Origin	System	Noise spectrum
Thermal (Johnson)	Thermal agitation	Electronic systems	Flat (normally distributed)
Shot	Quantization of electrical charge	Electronic systems	Flat (Poisson-distributed)
1/f	System dynamics	Chaotic systems	1/f

Thermal noise

Mobile electrons move in a random fashion:



=> at any particular instant there may be more e- near one end of the conductor than the other.

Thermal noise occurs in all systems which are not at absolute zero: cannot be avoided.

Thermal noise

The thermal noise power is $P = 4k_B TB$ [W]

=> the rms of the noise voltage accross a resistor is

$$\sqrt{\langle v_{noise}{}^2 \rangle} = \sqrt{4\, k_B\, T\, B\, R} \quad [V]$$

where k_B is the Boltzman constant (1.38×10^{-23} J/K)

T is the resistor's temperature [K]

B is the observed bandwidth [Hz]

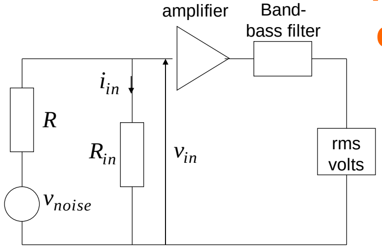
R is the resistor's resistance [Ω]

Example : thermal noise accross a 1 kΩ resistor at 300 K

produces fluctuations : $\sqrt{\langle v_{noise}{}^2 \rangle / B} = \sqrt{4 \times 1.38 \times 10^{-23} \times 300 \times 1000}$

$= 4 \times 10^{-9}$ V/$\sqrt{\text{Hz}}$

Noise power observation



$$i_{in} = \frac{v_{noise}}{R + R_{in}}; \quad v_{in} = R_{in} \times i_{in}$$

$$\text{mean noise power} = \langle v_{in} \times i_{in} \rangle = \frac{R_{in}}{(R + R_{in})^2} \langle v_{noise}{}^2 \rangle$$

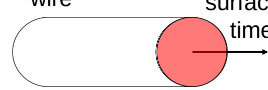
maximum power is obtained for $R_{in} = R$:

$$\langle v_{in} \times i_{in} \rangle_{\max} = \frac{1}{4R} \langle v_{noise}{}^2 \rangle = k \times T \times B$$

Shot noise

Shot noise arises because of the quantization of the electrical charge ($e = 1.6 \times 10^{-19}$ C).

Repeat the same measurement $i_{k=1,m}$



$$i_k = n_k \frac{e}{\Delta t},$$

$$I = \langle n_k \rangle \frac{e}{\Delta t} \quad \text{where } \langle n_k \rangle = 1/m \sum_{i=1}^{m} n_k$$

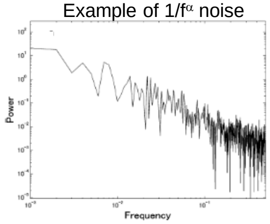
$i_k \neq I$ Each carrier has its own velocity and separation from its neighbors

Shot noise cannot be avoided.

1/f noise

Thermal noise and shot noise are ‘white’ noises (have flat power spectra)

Example of $1/f^{\alpha}$ noise



1/f noise :

noise power spectrum $1/f^{\alpha}$

where $0.5 \le \alpha \le 2$

1/f power spectrum observed in:

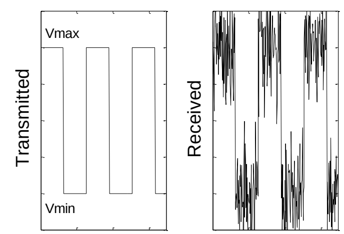
- Some chaotic systems, such as in the Bernoulli map (deterministic chaos).
- Brownian motion (red noise).

Summary

- Thermal noise and shot noise arises from the quantization of electrical charge. These types of noise are unavoidable.
- Thermal noise and Shot noise have a uniform noise power spectral density (white noise)

5. How does noise affect information transmission?

Digital transmission over a noisy channel



Message sent as a stream of binary digits

$$\text{Decision level} = V_{Decision} = \frac{V_{\max} + V_{\min}}{2}$$

Transmission code rules:

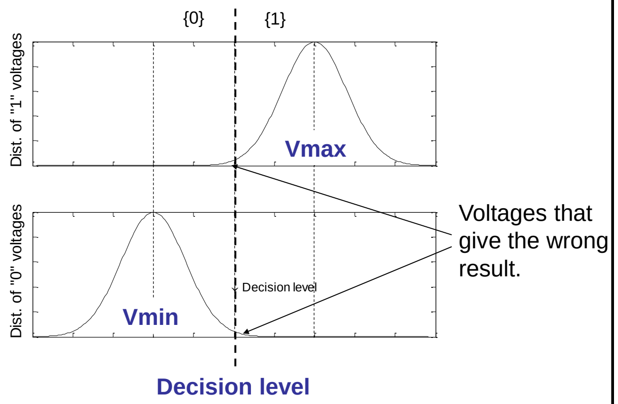
$\text{Voltage} \ge \text{Decision level} \Rightarrow \text{received a } 1$

$\text{Voltage} < \text{Decision level} \Rightarrow \text{received a } 0$

Noise produces random errors

=> Information received is not always correct

Transmission errors



Probability that a "1" will be correctly received :

$$P(v \geq V_{Decision}) = \int_{V_{Decision}}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{v-V_{\max}}{\sigma}\right)^2} dv$$

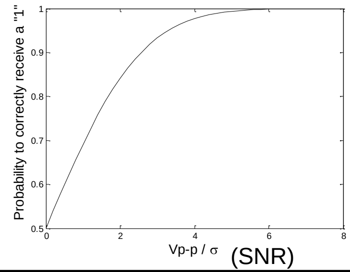
$$= \int_{-V_{p-p}/2}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2} du \quad \text{(change of variables } u = v - V_{\max}\text{)}$$

where $V_{p-p} = V_{\max} - V_{\min}$

Note that for $V_{p-p} = 0$

$\Rightarrow P(v \geq V_{Decision}) = 50\%$

only two possibilities: "0" or "1"!



Noise can be helpful!

1/3

Dithering

Digital transmission system

-Quantization interval: range/2^bit

-Signal details smaller than the quantization interval are lost

Dithering

-Superimpose random noise on the signal (noise fluctuation should be a little larger than the quantization interval)

-Quantize/digitize

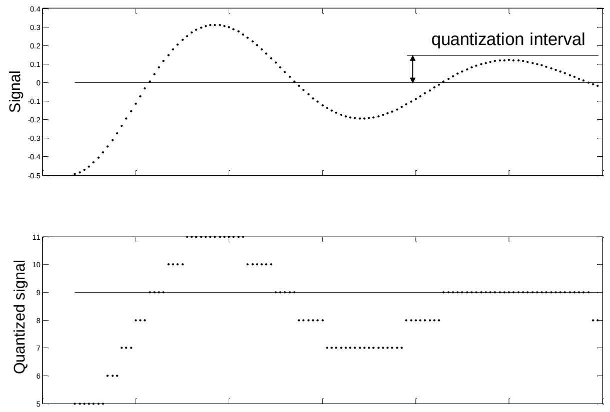
-Transmit

-Filter (low-pass filter), time information

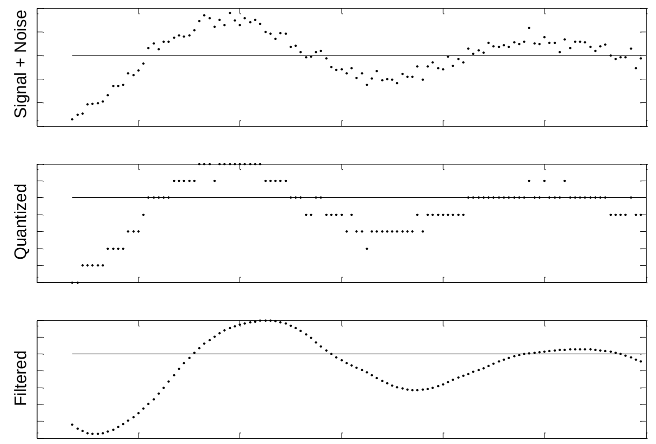
=> Signal details smaller than the quantization interval can be transmitted (overcome the distortion caused by quantization)

Image dithering refers to a different technique: diffuse color information to nearest pixels.

2/3



3/3



Signal averaging

- Voltmeter reading in the presence/absence of noise:
  - 3-digit voltmeter
  - Reading in the absence of noise : 5.50 V
  - If noise > quantization interval 0.01 V

    => readings : 5.51, 5.50, 5.50, 5.51, 5.49, etc.

    average of the readings 5.503 V => better precision!!!