Quantum Information

1

Information Encoding

- Information may be Represented in Continuous or Discrete from
- Information can be Represented Directly, Indirectly, or Generatively
  1) Directly is explicit representation.   A set of observed symbols (discrete). There is not direct representation of continuous information because it cannot be explicitly represented.
  2) Indirectly is like a pointer to a data object. A lossless encoding with a known decoding algorithm.
  3) Generatively is based on a generating function that generates a particular sequence of symbols, or a set of continuous functions that can be evaluated at any point to yield the data. Generative continuous data is also usually (always?) indirect.

2

Information Encoding Examples

EXAMPLE: Define a message that conveys the speed of an object in a plane from time t_0 to t_1.

1. Continuous/Direct: A graph drawn on a Cartesian coordinate system, with x-axis labeled in time units from t_0 to t_1 and y-axis labeled in distance/time for speed values.
2. Continuous/Indirect: position of object where x-axis is in units of time and y-axis is in units of distance. Indirect since a transformation must be applied to (in this case first-order time derivative operator is applied) to retrieve the direct information.
3. Continuous/Generative: A Mathematical function defined over the interval t_0 to t_1 for the speed. Can be either direct (like here) or indirect.
4. Discrete versions of above are sampled versions of the continuous information.
5. Some information may be inherently discrete since it is undefined over some intervals.

3

Quantum Information Theory

- BIG Subject – can be Focus of Entire Class
- Covers these Topics
  1) Transmission of Classical (Shannon) Information over a Quantum Channel
     a) noiseless channels
     b) channels with noise
  2) Transmission of Quantum (von Neumann) Information over a Quantum Channel
  3) Tradeoff between Acquisition of Information about a Quantum State and Disturbing the State
  4) Quantifying Quantum Entanglement

4

Von Neumann Information

- Comparison to Shannon Information:
  - Shannon Message: Finite Set of n Symbols chosen from a Finite Set of Symbols
  - Von Neumann Message: Finite Set of n Quantum States chosen from an Ensemble of Quantum States
- Alphabet (the ensemble of quantum states) is represented by a density matrix*, $\boldsymbol{\rho}$, comprised of each possible quantum state $\boldsymbol{\rho}_x$ (in density matrix form) and its associated probability, p_i.

$$\boldsymbol{\rho} = \sum_{i=1}^{n} p_i \boldsymbol{\rho}_i$$

*Density matrices are self-adjoint (Hermitian), positive semi-definite, and of trace 1. Density matrices are diagonalizable, that is, they have a spectral decomposition.

5

Von Neumann Entropy

- Entropy is the Expected Value of the Information Content in a Message
- Comparison to Shannon Entropy:
  - Shannon Message: A set of n symbols $\{x_i\}$

$$H(X) = E\{I(X)\} = -\sum_{i=1}^{n} P[x_i] \log_b (P[x_i])$$

  - Von Neumann Message: A set of quantum states, ρ_i or $|\Psi_i\rangle$

$$S(\rho) = -\mathrm{Tr}(\rho \ln \rho)$$

- Von Neumann entropy requires evaluating the logarithm of the density matrix.

$$\boldsymbol{\rho} = \sum_{i=1}^{n} p_i \boldsymbol{\rho}_i$$

  - if Alphabet is comprised of orthogonal pure states, $|\Psi_i\rangle$:

$$\boldsymbol{\rho} = \sum_{i=1}^{n} p_i |\boldsymbol{\Psi}_i\rangle\langle\boldsymbol{\Psi}_i|$$

6

Matrix Logarithm

- General case, given,

$$e^{\mathbf{A}} = \mathbf{B}$$

- Then, $\mathbf{A}$ is the (natural) logarithm of $\mathbf{B}$
- Our matrices of interest have complex components and the exponential of a complex value is not one-to-one, example:

$$e^{i\frac{\pi}{2}} = e^{i\frac{5\pi}{2}} = e^{i\frac{9\pi}{2}}$$

- Thus, some matrices have more than one logarithm
- Power Series Expansions:

$$e^{\mathbf{A}} = \sum_{n=0}^{\infty} \frac{\mathbf{A}^n}{n!} \qquad \ln(\mathbf{B}) = \sum_{n=0}^{\infty} (-1)^{k+1} \frac{(\mathbf{B}-\mathbf{I})^k}{k}$$

- Complex matrices logarithms if, and only if, they are invertible

7

Matrix Logarithm Properties

- A Hermitian matrix, $\mathbf{A}$, is positive-definite if, for every non-zero column vector of $\mathbf{A}$, $|a_i\rangle$, the following scalar is strictly positive (note these scalars are all real since $\mathbf{A}$ is Hermitian).

$$\langle a_i | \mathbf{A} | a_i \rangle > 0$$

- If $\mathbf{A}$ and $\mathbf{B}$ are both positive-definite,

$$\mathrm{Tr}\left[\ln(\mathbf{AB})\right] = \mathrm{Tr}\left[\ln(\mathbf{A})\right] + \mathrm{Tr}\left[\ln(\mathbf{B})\right]$$

$$\ln(\mathbf{AB}) = \ln(\mathbf{A}) + \ln(\mathbf{B})$$

$$\ln(\mathbf{A}^{-1}) = -\ln(\mathbf{A})$$

8

Von Neumann Entropy

- Based on the properties of the Density Matrix and the logarithm of a matrix, Von Neumann Entropy can be shown to be:

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_{i=1}^{n} \lambda_i \log(\lambda_i)$$

- The Von Neumann Entropy of a single pure state is zero. Let: $|\boldsymbol{\Psi}_1\rangle = \boldsymbol{\alpha}|0\rangle + \boldsymbol{\beta}|1\rangle$

$$\rho = \sum_{i=1}^{1} p_i |\Psi_i\rangle\langle\Psi_i| = (1)|\Psi_1\rangle\langle\Psi_1| = |\Psi_1\rangle\langle\Psi_1| = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}\begin{bmatrix} \alpha^* & \beta^* \end{bmatrix} = \begin{bmatrix} \alpha\alpha^* & \alpha\beta^* \\ \alpha^*\beta & \beta\beta^* \end{bmatrix}$$

$$c(\boldsymbol{\lambda}) = |\boldsymbol{\rho} - \boldsymbol{\lambda}\mathbf{I}| = \begin{vmatrix} \boldsymbol{\alpha\alpha}^* - \boldsymbol{\lambda} & \boldsymbol{\alpha\beta}^* \\ \boldsymbol{\alpha}^*\boldsymbol{\beta} & \boldsymbol{\beta\beta}^* - \boldsymbol{\lambda} \end{vmatrix} = (\boldsymbol{\alpha\alpha}^* - \boldsymbol{\lambda})(\boldsymbol{\beta\beta}^* - \boldsymbol{\lambda}) - (\boldsymbol{\alpha\beta}^*)(\boldsymbol{\alpha}^*\boldsymbol{\beta})$$

$$c(\boldsymbol{\lambda}) = (\boldsymbol{\alpha\alpha}^* - \boldsymbol{\lambda})(\boldsymbol{\beta\beta}^* - \boldsymbol{\lambda}) - (\boldsymbol{\alpha\beta}^*)(\boldsymbol{\alpha}^*\boldsymbol{\beta}) = \boldsymbol{\alpha\alpha}^*\boldsymbol{\beta\beta}^* - \boldsymbol{\lambda\alpha\alpha}^* - \boldsymbol{\lambda\beta\beta}^* - \boldsymbol{\alpha\alpha}^*\boldsymbol{\beta\beta}^* + \boldsymbol{\lambda}^2$$

$$= \boldsymbol{\lambda}^2 - \boldsymbol{\lambda}(\boldsymbol{\alpha\alpha}^* + \boldsymbol{\beta\beta}^*) = \boldsymbol{\lambda}^2 - \boldsymbol{\lambda}\left(|\boldsymbol{\alpha}|^2 + |\boldsymbol{\beta}|^2\right) = \boldsymbol{\lambda}^2 - \boldsymbol{\lambda}(1) = \boldsymbol{\lambda}(\boldsymbol{\lambda} - 1) = 0$$

$$\{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\} = \{0, 1\}$$

9

Von Neumann Entropy (cont.)

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_{i=1}^{n} \lambda_i \log(\lambda_i)$$

- The Von Neumann Entropy of a single pure state is zero. Let:

$$|\boldsymbol{\Psi}_1\rangle = \boldsymbol{\alpha}|0\rangle + \boldsymbol{\beta}|1\rangle \qquad \rho = \begin{bmatrix} \alpha\alpha^* & \alpha\beta^* \\ \alpha^*\beta & \beta\beta^* \end{bmatrix} \qquad \{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2\} = \{0, 1\}$$

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_{i=1}^{n} \lambda_i \log(\lambda_i) = -(0)\ln(0) - (1)\ln(1) = 0 \ \ \text{nats}$$

10

Von Neumann & Shannon Entropy

- Definition of Von Neumann entropy: $S(\rho) = -\mathrm{Tr}(\rho \ln \rho)$
- Because the Density matrix is Hermitian, it is Diagonalizable (can be expressed as a spectral decomposition) for some orthonormal basis set $\{|a_i\rangle\}$. $\rho = \sum_{i=1}^{n} \lambda_i |a_i\rangle\langle a_i|$
- Consider the ensemble as the alphabet, A: $A = \{\lambda_i, |a_i\rangle\}$
- Then, the Von Neumann entropy reduces to the Shannon entropy: $S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_{i=1}^{n} \lambda_i \log(\lambda_i) = H(A)$
- Thus, if the alphabet consists of a set of orthogonal pure states with probability, p_i, $A=\{p_i, |\Psi_i\rangle\}$ then the Von Neumann and Shannon entropy are identical ($p_i=\lambda_i$):

$$S(\rho) = -\mathrm{Tr}(\rho \log \rho) = -\sum_{i=1}^{n} p_i \log(p_i) = H(A)$$

11

Von Neumann & Shannon Entropy

- Shannon and Von Neumann entropy provide the incompressible information content of a message consisting of a set of symbols from an alphabet
- When the Von Neumann alphabet consists of a set of pure and orthonormal states, the two entropy values are identical and we can quantify information content in "bits" not "qubits"!
- Quantum Information Theory is concerned with the interpretation and use of Von Neumann entropy
- Case where alphabet contains quantum states, $\boldsymbol{\rho}_i$, that are not orthogonal is interesting and opens a new area not covered by classical Shannon Information theory

12

Von Neumann Entropy Properties

- Purity: Von Neumann entropy of a single pure state is zero (previously proven).
- Invariance: Entropy is unchanged when the density matrix undergoes a unitary change of basis defined by $\mathbf{U}$.

  Since $S(\boldsymbol{\rho})$ depends only on its eigenvalues, this is obvious. $S(\rho) = S(\mathbf{U}\rho\mathbf{U}^{-1})$
- Maximum: When $\boldsymbol{\rho}$ has non-vanishing eigenvalues, the diagonal matrix $\mathbf{D}$ in its spectral decomposition obeys:

  $$S(\rho) \le \log(\mathbf{D})$$

  thus, information content or entropy can be shown to be maximized when all non-zero eigenvalues are equal, which means each quantum state in a message is chosen randomly (equally likely, uniformly distributed).

13

Von Neumann Entropy Properties (cont.)

- Concavity: Consider the non-zero eigenvalues of $\boldsymbol{\rho}$,

  $$\lambda_1, \lambda_2, \cdots, \lambda_n \ge 0, \quad \lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$$

  then,

  $$S(\lambda_1\rho_1 + \lambda_2\rho_2 + \cdots + \lambda_n\rho_n) \ge \lambda_1 S(\rho_1) + \lambda_2 S(\rho_2) + \cdots + \lambda_n S(\rho_n)$$

  This means that the message entropy is LARGER if less is known about how the state was prepared! This is due to the convexity of the logarithm function.

Several other interesting and useful theoretical properties, but let us turn to an application in Quantum Information Theory, that of Data Compression.

14

Classical Data Compression

15

Classical and Quantum Messages

- Consider a finite alphabet of N distinct and discrete symbols with probability, p_i: $A_C = \{(x_i, p_i) \mid i = 1, N\}$
- A message, M, of length n is a specific sequence n symbols chosen from the alphabet
- Shannon Information content of a n-length message is:

$$I(M) = \sum_{i=1}^{k} \frac{1}{\log(P[s_i])} = -\sum_{i=1}^{k} \log(P[s_i])$$

- The average or expected amount of information over all messages of length n is the entropy
- Entropy differs depending on how the symbols are physically represented, or encoded
  - Classical encoding: Shannon entropy, $H(M)$
  - Quantum encoding: Von Neumann entropy, $S(\boldsymbol{\rho}_M)$

16

Classical Message Encoding

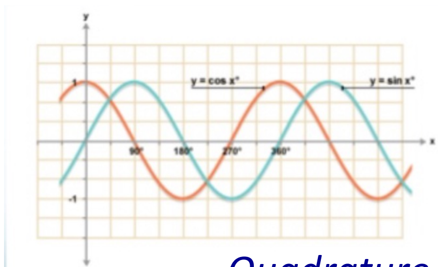
- Information is content is a sequence of symbols Each symbol encodes a fixed amount of classical information in units such as nats or bits
- Each message is a Sequence of Symbols
  - when symbols represent groups of bits, a Symbol is often called a "baud" in honor of Emile' Baudot (early telegraphy engineer)
  - the rate of information transfer is Symbols per second, or "baud" rate when in units of bits per second*
- Symbols are represented, or encoded, by an observable of a physical quantity that is unique for each symbol
  - measurement is modeled as a bijective function
  - example physical quantity is a voltage sine wave with some specific amplitude and relative phase such as 8-valued Quadrature Amplitude Modulation (*8QAM*)

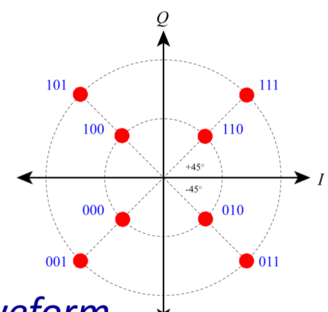
*Bits per second is often (somewhat erroneously) reported in Hertz and called the "Bandwidth." More precisely it should be referred to as the "data rate."

17

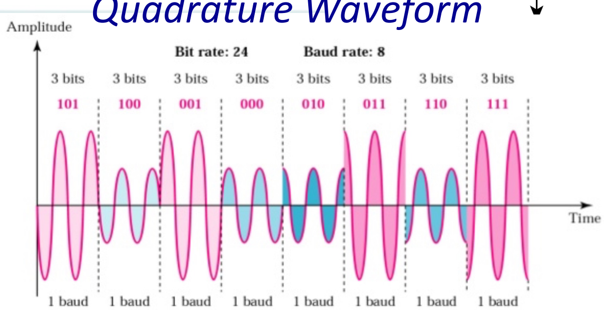
8QAM Example Waveform

Quadrature = Sine Wave + Cosine Wave



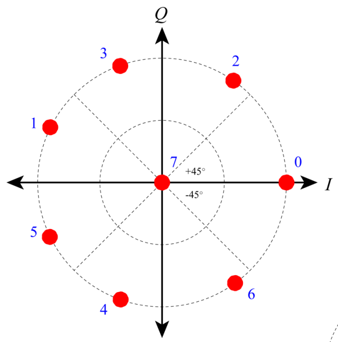


Quadrature Waveform

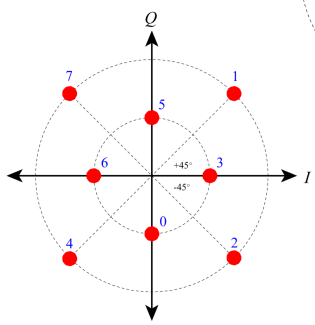


18

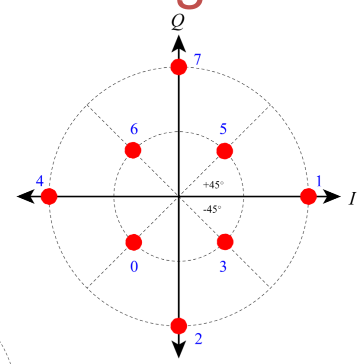
Other 8QAM Encodings



Circular Spacing



Star 1 Spacing

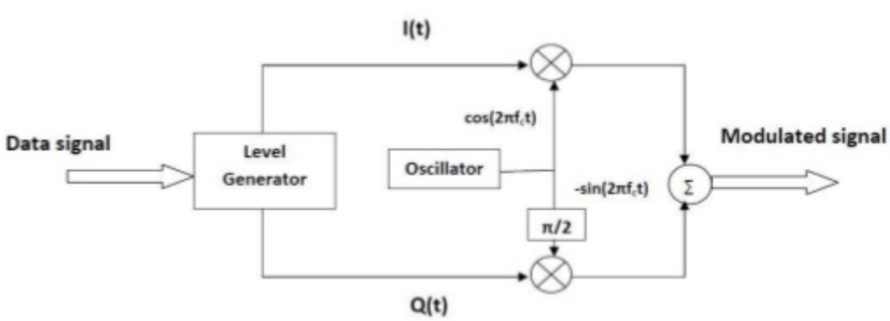


Star 2 Spacing

19

8QAM Example Encoder

8QAM Modulator (transmitter frontend)

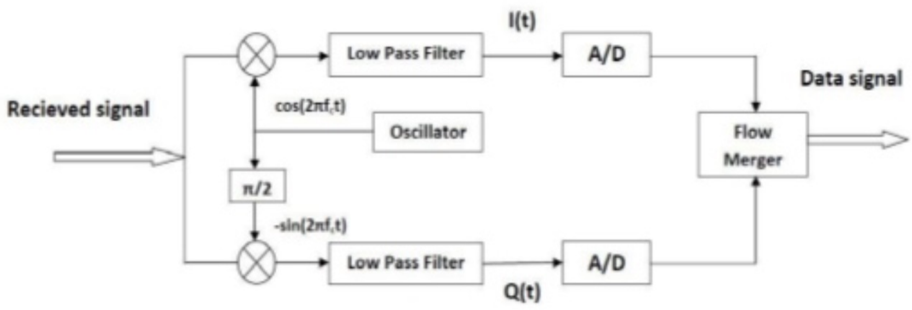


Quadrature = Sine Wave + Cosine Wave

20

8QAM Example Decoder

8QAM Demodulator (receiver frontend)



21

Classical Compression

- Objective of Data Compression is to represent Information with as few symbols from the alphabet as possible
- Shannon entropy is Maximized when Symbols in a message are Equally likely
  - a purely random stream contains the most information since the probability of occurrence for symbol is as small as possible (*i.e.*, a uniform distribution)
- For an n-symbol, classical message, we can calculate the maximum possible information content, I_{MAX}, as:

$$I_{MAX}(n) = \max\left\{\frac{1}{\log(P[x_1])} + \frac{1}{\log(P[x_2])} + \cdots + \frac{1}{\log(P[x_n])}\right\}$$

$$= \left\{\frac{1}{\log(\min\{P[x_1]\})} + \frac{1}{\log(\min\{P[x_1]\})} + \cdots + \frac{1}{\log(\min\{P[x_1]\})}\right\}$$

22

Efficiency

- This means the ***maximum possible information*** in a message* of length n is when it is composed of ***equally likely*** symbols (p_i=$1/N$) that are independent and identically distributed (uniform distribution)
- Figure of merit, M, for ***source encoding efficiency*** is the actual amount of information represented in a message of length n divided by the theoretical maximum amount of information in a message of length n

$$M = \frac{H_{MSG}\left(X\right)}{H_{RAND}\left(X\right)} \times 100\%$$

*Shannon Information is maximized in a symbol string that appears to be purely random since such strings are comprised of symbols with minimum probability of occurrence.

23

8QAM Message Information

- Assume the alphabet is a collection of N pairs comprising a unique symbol and its probability of occurrence:

$$A_C = \left\{\left(x_i, p_i\right) | i = 1, N\right\}$$

- Assume symbols are equally likely to be used: $p_i = \frac{1}{N}$
- Assume n=7, $8QAM$ alphabet (N=8), and Message* is Y=2344250
- Message Symbol Distribution (histogram of symbols):

NUMBER OF OCCURRENCES:	1	0	2	1	2	1	0	0
SYMBOL:	0	1	2	3	4	5	6	7

- Message Symbol Percentages: $\hat{p}_i = \frac{n_i}{n} \times 100$ $\left\{\frac{1}{7}, \frac{0}{7}, \frac{2}{7}, \frac{1}{7}, \frac{2}{7}, \frac{1}{7}, \frac{0}{7}, \frac{0}{7}\right\}$
- Message Entropy, assuming that each message of length n is independent of past or future messages:

$$H_{MSG}\left(Y\right) = \sum_{i=1}^{7} H_{MSG}\left(X_i \to x_i\right) = -\sum_{i=1}^{7} \hat{p}_i \log_2\left(\hat{p}_i\right) = -\left[3\left(\frac{1}{7}\right)\log_2\left(\frac{1}{7}\right) + 2\left(\frac{2}{7}\right)\log_2\left(\frac{2}{7}\right)\right]$$

$$= 1.2032 + 0.5164 = 1.7196 \text{ bits}$$

*www.random.org/integers was used to generate the example message string

24

8QAM Message Information (cont.)

- Assume n=7, $8QAM$ alphabet, and Message* is Y=2344250_8

$$H_{MSG}(Y) = \sum_{i=1}^{7} H_{MSG}(X_i \to x_i) = -\sum_{i=1}^{7} \hat{p}_i \log_2(\hat{p}_i) = 1.7196 \text{ bits}$$

- Theoretical Maximum Information:

$$H_{RAND}(X) = -\sum_{i=1}^{7} P[X = x_i] \log_2(P[X = x_i]) = -\sum_{i=1}^{7}\left(\frac{1}{8}\right)\log_2\left(\frac{1}{8}\right) = -7\left(\frac{1}{8}\right)\log_2\left(\frac{1}{8}\right) = 2.625 \text{ bits}$$

- Message Information Efficiency:

$$M = \frac{H_{MSG}(X)}{H_{RAND}(X)} \times 100\% = \frac{H_{MSG}(X)}{H_{RAND}(X)} \times 100\% = \frac{1.7196}{2.625} \times 100\% = 66\%$$

- Efficiency indicates that the Message has redundancy and that lossless data compression may be possible.
- If an Ideal Lossless Compression Method could be found, the same amount of information should theoretically be possible to represent in n_{comp} Symbols:

$$n_{comp} = \lceil M \times n \rceil = \lceil (0.66)(7) \rceil = 5$$

*www.random.org/integers

25