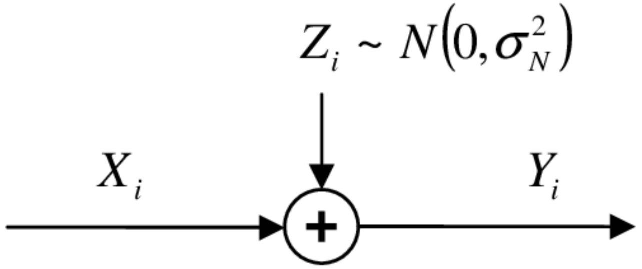
Differential entropy, Gaussian channels

Gaussian Channels

Gaussian channels

- Many communications channels can be modeled as discrete-time continuous-valued channels with additive Gaussian noise.
- The simplest case is i.i.d. noise $\{Z_i\} \sim \mathcal{N}\left(0, \sigma^2\right)$.



Gaussian channels

- If there are no constraints on the input symbols, arbitrarily many (that is, infinitely many) bits could be sent in one use of the channel with any desired small error probability - simply choose input values far enough apart that there will be small probability of uncertainty (tails of the Gaussian density).
- The most common input constraint is power - energy per channel use or per unit time.
- For discrete-time channels, power is defined to be the average input energy per channel use

$$P = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Gaussian channel capacity

The information capacity of a Gaussian channel with power constraint P is:

$$C = \max_{p(x), E[X^2] \leq P} I(X;Y)$$

The maximisation is over all input probability densities with second moment $E\left[X^2\right] \leq P$

- We assume that the noise is independent of the input: $X \perp Z$
- The conditional differential entropy is given by: $f_{Y|X}(y|x) = f_Z(y-x) = f_Z(z) \Rightarrow h(Y|X) = h(Z)$
- $Var\{Y\} = Var\{X+Z\} = Var\{X\} + Var\{Z\} = E\left[X^2\right] + E\left[Z^2\right] = P + \sigma_N^2$
- For a fixed variance, a Gaussian density has the largest differential entropy. Therefore:

$$h(Y) \leq h\left(\mathcal{N}\left(0, P+\sigma_N^2\right)\right) = \frac{1}{2}\log 2\pi e\left(P+\sigma_N^2\right)$$

- The sum of independent Gaussian random variables is also Gaussian, thus we can obtain a Gaussian output density by using an input with Gaussian Density $\mathcal{N}(0, P)$
- hence: $I(X;Y) = h(Y) - h(Y|X) = \frac{1}{2}\log 2\pi e\left(P + \sigma_N^2\right) - \frac{1}{2}\log 2\pi e\sigma_N^2 = \frac{1}{2}\log\frac{P+\sigma_N^2}{\sigma_N^2} = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_N^2}\right)$

Gaussian channel capacity

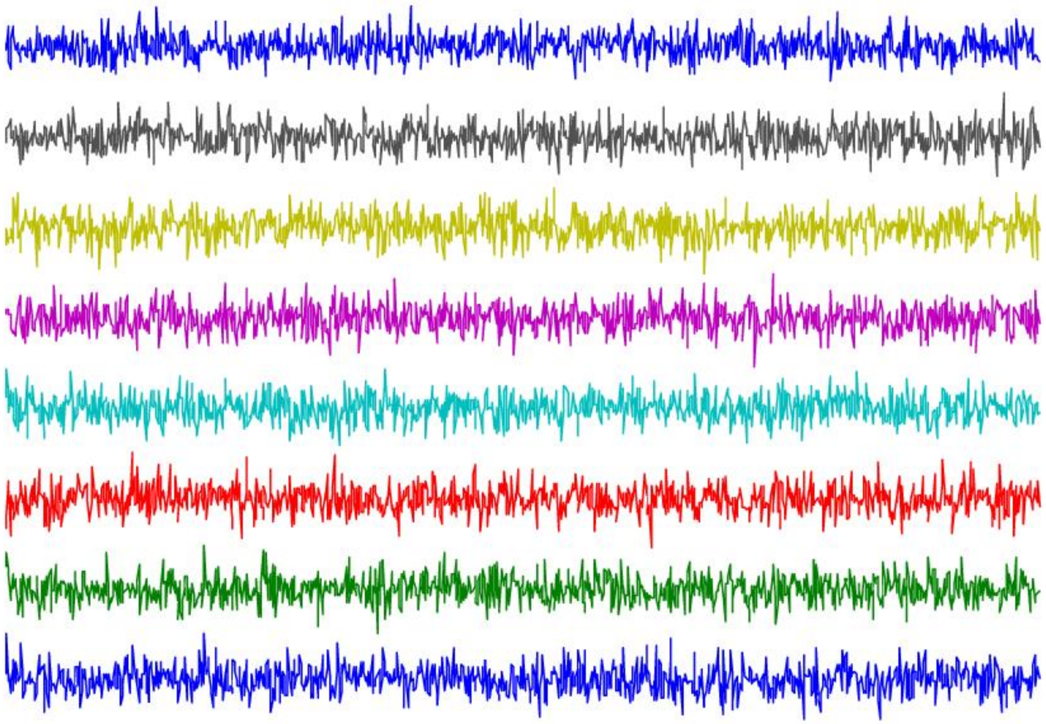
We can summarize these steps as follows: the information capacity of a Gaussian channel with noise power N and power constraint P is

$$C = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_N^2}\right)$$

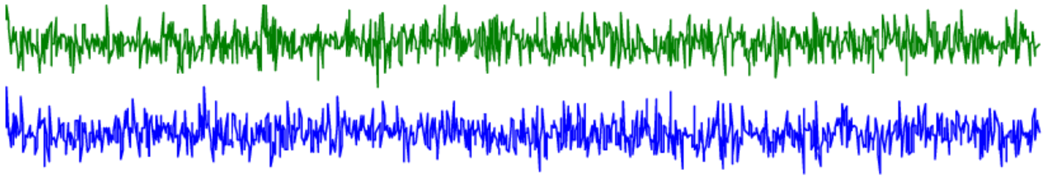
bits per channel use
That is, for any rate $R < C$ and any $\varepsilon > 0$, there exists $\left(2^{Rn}, n\right)$ codes with vanishing probability of error for large enough n.

Gaussian codebooks

Example of a codebook with $n = 1000$ and $k = \log 8$



Geometric interpretation of random coding proof: Sphere Packing



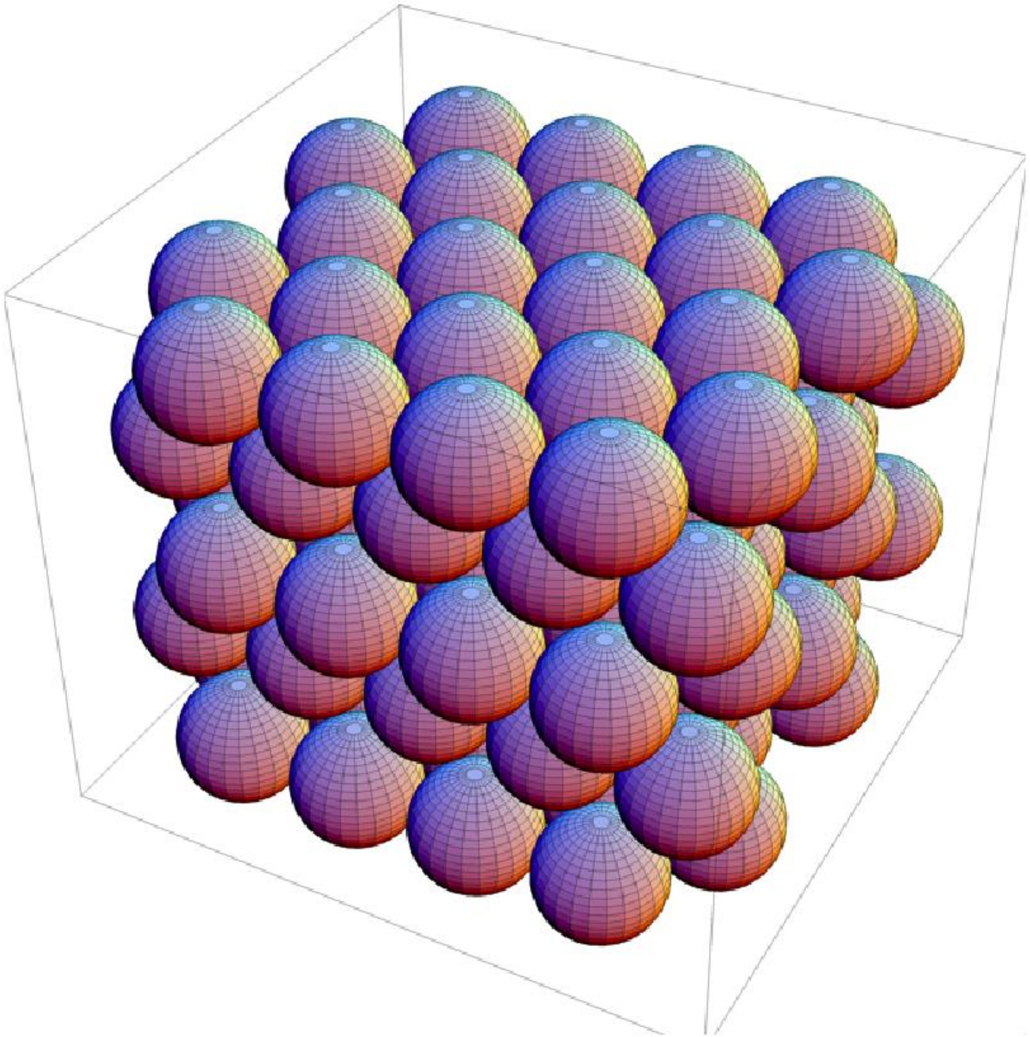
- When noise is i.i.d. Gaussian $\mathcal{N}\left(0, \sigma_N^2\right)$, the output density given a codeword X_n is spherically symmetric and concentrated in a sphere of radius $\sqrt{n\sigma_N^2}$.
- If we transmit codewords whose average power is P, then the received n-tuples will have a Gaussian density with radius $\sqrt{n\left(P+\sigma_N^2\right)}$.
- In order to guarantee reliable decoding, the spheres of radius $\sqrt{n\sigma_N^2}$ surrounding codewords should be (effectively) disjoint.

$$\mathrm{r} = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

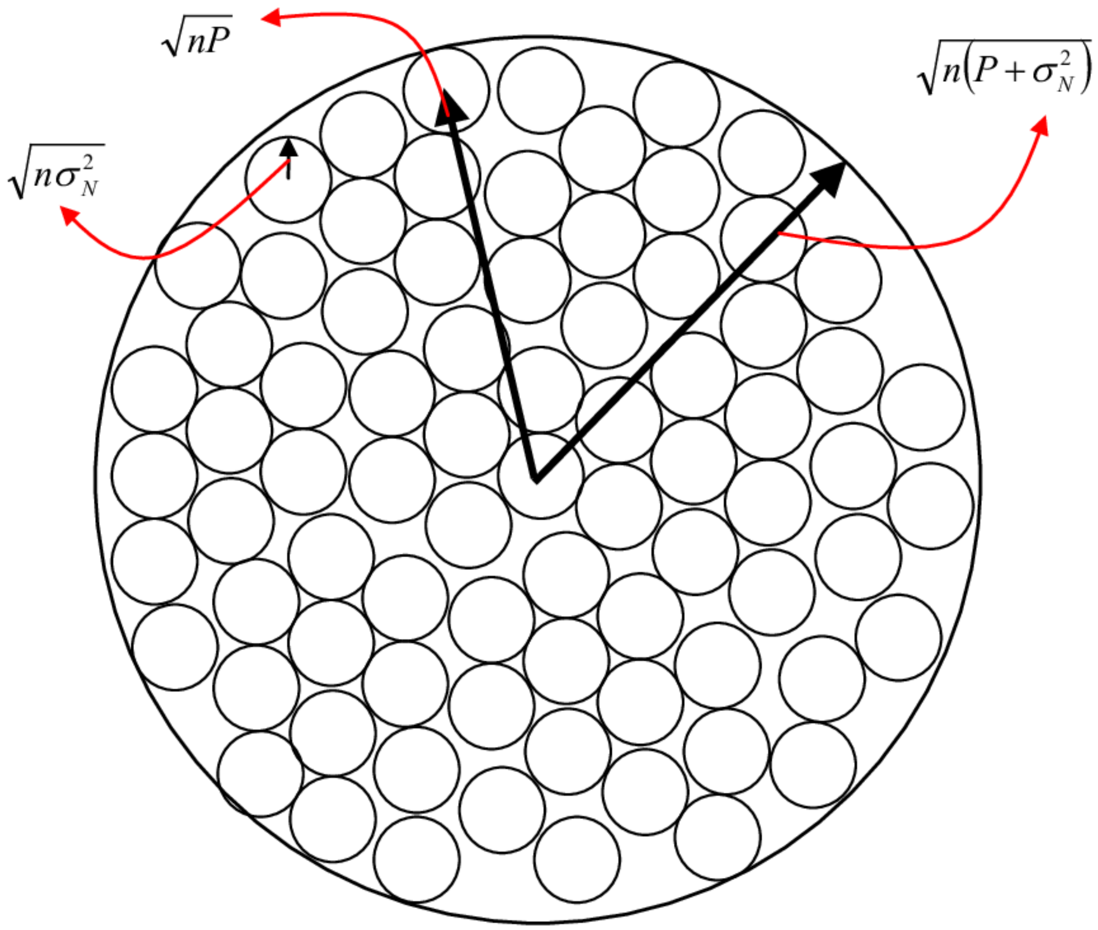
$$\mathrm{r} = \sqrt{n\sigma_x^2}$$

Sphere Packing



The Sphere packing problem

Sphere Packing



Geometric interpretation

Denote by $V_n(r) = A_n r^n$ volume of a sphere with radius r (A_n is dimension only dependent constant).
The total number of codewords is bounded above by:

$$M \leq \frac{V_n\left(\left(n\left(P+\sigma_N^2\right)\right)^{\frac{n}{2}}\right)}{V_n\left(\left(n\sigma_N^2\right)^{\frac{n}{2}}\right)} = \frac{\left(n\left(P+\sigma_N^2\right)\right)^{\frac{n}{2}}}{\left(n\sigma_N^2\right)^{\frac{n}{2}}} = \left(1+\frac{P}{\sigma_N^2}\right)^{\frac{n}{2}}$$

Thus the rate of the code is

$$R = \frac{1}{n}\log_2 M \leq \frac{1}{2}\log\left(1+\frac{P}{\sigma_N^2}\right)$$