# On the Shannon Capacity Limit

Akhila Raman

*University of California at Berkeley, CA-94720. Email: akhila.raman@berkeley.edu.*

**Abstract**

It is well known that Shannon's Capacity Limit for a communication channel in the presence of white noise is given by $C = W \log(1 + \frac{P}{N})$ where $C$ is the Channel capacity in bits per second, $W$ is the channel bandwidth, $P$ is the average signal power and $N$ is the average noise power. In this paper, we examine the entry conditions for Shannon's Theorem for Capacity Limit and propose a new method which **does not** meet the entry conditions for Shannon's Theorem and hence is **not** bound by the mandates of that theorem and is free to achieve higher channel capacity, if possible. Shannon's Theorem remains safe and sound.

this theorem and if we can get past it.

*Keywords:*

## 1. Introduction

• Scientists believed for a long time that speed of sound cannot be crossed by aircrafts. Supersonic aircrafts have been demonstrated now in civilian and military aircrafts. (Sound Barrier)

• Every theorem in mathematics has entry conditions for that theorem to hold. If entry conditions are **not** satisfied for a specific case, that theorem is **no longer** applicable to that case.

Similarly **Shannon's Capacity Limit theorem** given by $C = W \log_2(1 + \frac{P}{N})$ has **entry conditions** as follows, where $C$ is the channel capacity in bits per second.

1) For n-sample sequence or signal with $n = 2WT$ where $T$ is the duration of the sequence and $W$ is the signal bandwidth and $M$ is the number of symbols per sample, there are $M^n$ possible transmitted sequences in a n-dimensional hyperspace with average power $P$.

2) There is a noise sphere of radius $N$ surrounding each transmitted sequence and as $n \to \infty$, the noise samples will be confined to the surface of that sphere and adjacent spheres do not overlap, so that we can choose the n-sample transmitted sequence which is **closest** to the n-sample received sequence. Shannon said "since overlap of the noise spheres results in confusion as to the message at the receiving point" (Shannon's paper pp452-454). One such criterion is Minimum Mean-Squared Error(**MMSE**) criterion as follows.

$$D_s[m] = \frac{1}{n} \sum_{i=1}^{n} (r[i] - s_m[i])^2, \quad r[i] = s[i] + w[i], \quad m = 1, ...M^n \tag{1}$$

We note that $r(i)$ is the received signal, $s_m[i]$ is the set of all possible transmitted sequences and $w[i]$ is Additive White Gaussian Noise (AWGN). We choose the value of $m$ for which the detection function $D_s[m]$ is minimum, as per **MMSE** criterion, as $n \to \infty$.

• **New method A** is presented in this paper which **does not** meet the entry condition for Shannon's theorem and is not bound by the mandates of that theorem. It **does not** use the idea of non-overlapping noise spheres and instead uses time diversity in two data blocks and a new criterion given by Minimum Mean Multiplication of Errors in Two Data Blocks(**MMMETDB**) and a **new** detection function $D_s^{'}[m]$ in Eq. 2, which has terms involving multiplication and addition of n-sample AWGN noise sequences in the two blocks, when averaged, approach zero, as $n \to \infty$. Noise spheres **can overlap** in this new method and we use the multiplication and addition properties of AWGN sequences to get the desired result.

If Method A results in spectral efficiency (bits per second per hertz) which exceeds the spectral efficiency predicted by Shannon's theorem, it **does not mean** that Shannon's Theorem is violated, simply because Shannon's theorem is **not applicable** to new method A. This is explained in detail in the paragraphs below and in Section 1.4.

• **New Method A** uses Gallager's LDPC error correction codes for which $\lim_{n\to\infty} \frac{d_{min}}{n}$ is a constant, where $d_{min}$ is the minimum distance of that code. We use binary signalling with $M = 2$ and $(n, k)$ rate $\frac{1}{2}$ error correction code for which $\lim_{n\to\infty} \frac{d_{min}}{n} = 0.11$. Figure 2.4 in Page 18 in Gallager's book. As $n \to \infty$, the fraction of codewords given by $\frac{2^k}{2^n} = \frac{1}{2^{n-k}} = \frac{1}{2^{\frac{n}{2}}} \to 0$ for rate $\frac{1}{2}$ code and hence error correction codes with $d_{min}$ increasing linearly with $n$ is reasonable.

New Method A uses a **different** detection function as follows.

$$D_s^{'}[m] = \frac{1}{n}\sum_{i=1}^{n}(r[i] - s_m[i])(r^{'}[i] - s_m[i]), \quad r[i] = s[i] + w[i], \quad m = 1, ...2^k$$
$$r^{'}[i] = s[i] + w^{'}[i] \tag{2}$$

We note that $s[i]$ is the transmitted signal, which is one of the set of transmitted sequences $s_m[i]$ for $m = 1, ...2^k$ and $r(i)$ is the received signal and $w[i]$ is Additive White Gaussian Noise (AWGN) in Block 1. We note that $r^{'}(i)$ is the received signal in Block 2 and $w^{'}[i]$ is AWGN in Block 2. We choose the value of $m$ for which the **new detection function** $D_s^{'}[m]$ is minimum, as per the **new** criterion. As $n \to \infty$, $D_s^{'}[m]$ is minimum for the value of $m$ for which $s_m[i] = s[i]$.

• **New method A** is described as follows. In **Block 1**, transmitted n-sample sequence $s[i]$ uses $k = \frac{n}{2}$ bits and $(n, k)$ rate $\frac{1}{2}$ error correction code using Gallager's LDPC codes for which $\lim_{n\to\infty} \frac{d_{min}}{n} = 0.11$ and we use Binary AWGN channel and $w[i]$ is AWGN in Block 1. We get received sequence $r[i] = s[i] + w[i]$.

In **Block 2**, we transmit the same n-sample sequence $s[i]$ and use the same $(n, k)$ rate $\frac{1}{2}$ error correction code and we use an independent Binary AWGN channel and $w^{'}[i]$ is AWGN in Block 2. We get received sequence $r^{'}[i] = s[i] + w^{'}[i]$.

In the receiver, we subtract $s_m[i]$ from the two sequences $r[i]$ and $r^{'}[i]$ and multiply them and compute the mean as in Eq. 2 and get the **new detection function** $D_s^{'}[m]$, for $m = 1, ...2^k$ . We choose the value of $m$ for which the **new detection function** $D_s^{'}[m]$ is minimum, as per the **new** criterion. As $n \to \infty$, $D_s^{'}[m]$ is minimum for the value of $m$ for which $s_m[i] = s[i]$.

We compute the new detection function as follows.

$$D_s^{'}[m] = \frac{1}{n}\sum_{i=1}^{n}(r[i] - s_m[i])(r^{'}[i] - s_m[i]), \quad r[i] = s[i] + w[i], \quad m = 1, ...2^k, \quad r^{'}[i] = s[i] + w^{'}[i] \tag{3}$$

**Case A:** Let is consider the case of the transmitted n-sample sequence $s[i] = s_M[i]$ with average power $P$ derived from Shannon's theorem given by $C = W\log_2(1 + \frac{P}{N})$, for a given spectral efficiency $\frac{C}{W}$ and noise power $N$.

- **Case A1:** In Eq. 3, for $m = M$, we get $r[i] - s_m[i] = w[i]$, $r^{'}[i] - s_m[i] = w^{'}[i]$ and hence $D_s^{'}[M] = \frac{1}{n}\sum_{i=1}^{n} w[i]w^{'}[i]$ which goes to zero as $n \to \infty$ given that $w[i]$ and $w^{'}[i]$ are independent and uncorrelated AWGN sequences.

- **Case A2:** For index $m = M_1 \neq M$ for which the sequences $s_M[i]$ and $s_{M_1}[i]$ differ by $d_{min}$ bits, we get $r[i] - s_m[i] = w[i] + d_m[i]$, $r^{'}[i] - s_m[i] = w^{'}[i] + d_m[i]$ where $d_m[i] = \pm 2$ at the bit positions where $s_M[i]$ and $s_{M_1}[i]$ differ and $d_m[i] = 0$ at other bit positions.

Hence $D_s^{'}[M_1] = \frac{1}{n}\sum_{i=1}^{n}[w[i]w^{'}[i] + d_m^2[i] + d_m[i](w[i] + w^{'}[i])]$ which approaches $\frac{4d_{min}}{n} = 0.44$ as $n \to \infty$ given that

.

a) $D_1 = \frac{1}{n}\sum_{i=1}^{n} d_m^2[i] = \frac{4d_{min}}{n} = 4 * 0.11 = 0.44$ as $n \to \infty$.

.

b) $D_2 = \frac{1}{n}\sum_{i=1}^{n} d_m[i](w[i] + w^{'}[i]) = 0$ as $n \to \infty$ given that $w[i] + w^{'}[i]$ is an AWGN sequence and multiplying it by $d_m[i] = \pm 2$ gives a random variable(RV) with a certain probability density function(PDF), which when summed over $i = 1, ...n$ gives a RV with Gaussian density function, as $n \to \infty$ , using **Ergodicity** of random process which uses the fact that ensemble average equals the time average. Using **Central Limit Theorem**, we see that addition of RV's with an arbitrary density function, we get a Gaussian RV. (Details in Section 1.5)

.

c) $D_3 = \frac{1}{n}\sum_{i=1}^{n} w[i]w^{'}[i] = 0$ given that $w[i]w^{'}[i]$ is a random variable(RV) with a certain probability density function(PDF), which when summed over $i = 1, ...n$ gives a RV with Gaussian density function, as $n \to \infty$ , using **Ergodicity** of random process which uses the fact that ensemble average equals the time average. Using **Central Limit Theorem**, we see that addition of RV's with an arbitrary density function, we get a Gaussian RV. (Details in Section 1.5)

- **Case A3:** For index $m = M_2 \neq M$ for which the sequences $s_M[i]$ and $s_{M_2}[i]$ differ by **more than** $d_{min}$ bits, we get $r[i] - s_m[i] = w[i] + d_m^{'}[i]$, $r^{'}[i] - s_m[i] = w^{'}[i] + d_m^{'}[i]$ where $d_m^{'}[i] = \pm 2$ at the bit positions where $s_M[i]$ and $s_{M_2}[i]$ differ and $d_m^{'}[i] = 0$ at other bit positions.

Hence $D_s^{'}[M_2] = \frac{1}{n}\sum_{i=1}^{n}[w[i]w^{'}[i] + (d_m^{'})^2[i] + d_m^{'}[i](w[i] + w^{'}[i])]$ which approaches a value $> \frac{4d_{min}}{n}$ as $n \to \infty$ using arguments in Case 2.

We choose the value of $m$ for which the **new detection function** $D_s^{'}[m]$ is minimum, as per the **new** criterion, as $n \to \infty$. We set a **detection threshold** of $\frac{0.44}{2} = 0.22$. Hence we are **guaranteed a minimum** $D_s^{'}[M_1] = \frac{4d_{min}}{n} = 0.44$ as $n \to \infty$, if we choose the wrong transmitted sequence. Hence, as $n \to \infty$, we choose the correct transmitted sequence with 100 percent probability and bit error rate goes to zero.

• **Case B:** Let is consider the case of the transmitted n-sample sequence $s[i] = s_M[i]$ with average power $P$ **lower than** that derived from Shannon's theorem, for a given spectral efficiency $\frac{C}{W}$ and noise power $N$.

The arguments in Case A continue to hold for this case and we choose the value of $m$ for which the **new detection function** $D_s^{'}[m]$ is minimum, as per the **new** criterion, as $n \to \infty$. We set a **detection threshold** of $\frac{0.44}{2} = 0.22$. Hence we are **guaranteed a minimum** $D_s^{'}[M_1] = \frac{4d_{min}}{n} = 0.44$ as $n \to \infty$, if we choose the wrong transmitted sequence. Hence, as $n \to \infty$, we choose the correct transmitted sequence with 100 percent probability and bit error rate goes to zero. Case B requires larger number of samples than Case A to get the stated results. Numerical simulations are shown in Section 1.1.

Thus new method gives us spectral efficiency which **exceeds** the value given by Shannon's capacity theorem. It **does not mean** that Shannon's Theorem is violated, simply because Shannon's theorem is **not applicable** to new method A, because it **does not** meet the entry conditions requried by Shannon's theorem and instead uses a different detection function. This is explained in detail in Section 1.4.

### *1.1.* ***Numerical Simulations.***

We can use Matlab to simulate the detection function in the new method A and verify that the results in Section 1 hold. As an example, Matlab program shows that the new detection function $D_s^{'}[m] = \frac{1}{n}\sum_{i=1}^{n}[w[i]w^{'}[i] + d_m^2[i] + d_m[i](w[i] + w^{'}[i])]$ which approaches $\frac{4d_{min}}{n} = 0.44$ and all the terms involving AWGN go to zero as $n \to \infty$.

The above equation is simulated **at and below** the signal power required for channel capacity specified by Shannon's theorem, by choosing larger and larger values of noise power, for the case of binary signalling and spectral efficiency $\frac{C}{W} = 0.25$, $\frac{C}{W} = 0.5$ and $\frac{C}{W} = 1$. For this simulation, we choose each transmitted sample equals $\pm 1$ with the two symbols separated by a distance of $\pm 2$ and average signal power of 1 and choose suitable average noise power to verify if new method can go beyond channel capacity stated by Shannon's Theorem.

Numerical simulations of new method A is shown in Matlab program, for $\frac{C}{W} = 0.5$, $\frac{C}{W} = 1$ and $\frac{C}{W} = 2$, with plots for system operating at Shannon's Limit (plot) and 3 dB below it (plot and plot). It is noted that $\frac{C}{W}$ is **reduced** by a factor of 4 given that we use a rate $\frac{1}{2}$ error correction code and we repeat the data bits in the second block. We see that the above plots confirm the results for new method A in previous sections. We set a **detection threshold** of $\frac{0.44}{2} = 0.22$ and we see that noise component terms $D_2$ and $D_3$ reduce much below this thereshold, as number of samples $n$ increases.

We can do the simulations for higher $\frac{C}{W}$ with larger number of samples and longer computation times. As we reduce the signal power further below Shannon's Limit or equivalently increase noise power above Shannon's Limit, we see that we require larger number of samples and longer computation times. This is shown in Section 1.2.

### *1.2.* ***Higher channel capacity***

To achieve higher channel capacity and spectral efficiency, we can use M-ary signalling for each sample and the number of bits per sample is given by $b = \log_2 M$ and symbols separated by a distance of $\pm 2$ with average signal power larger than 1 and we choose suitable average noise power to verify if new method can go beyond channel capacity stated by Shannon's Theorem.

It is noted that channel spectral efficiency $\frac{C}{W}$ is **reduced** by a factor of 4 given that we use a rate $\frac{1}{2}$ error correction code and we repeat the data bits in the second block. We can improve channel spectral efficiency as follows.

- **Method 1:** We can use $(n, k)$ code with code rate $R = \frac{k}{n} > 0.5$ with $0.5 < R < 1$ for which $\lim_{n\to\infty} \frac{d_{min}}{n} < 0.11$ which is constant. Figure 2.4 in Page 18 in Gallager's book. As $n \to \infty$, the fraction of codewords given by $\frac{2^k}{2^n} = \frac{1}{2^{n-k}} = \frac{1}{2^{(1-R)n}} \to 0$ and hence error correction codes with $d_{min}$ increasing linearly with $n$ is reasonable. For example, if we choose $k = 0.99n$, we get a much smaller $\lim_{n\to\infty} \frac{d_{min}}{n} = 0.01$ and hence we need to consider larger number of samples and higher computational complexity to produce desired results.

- **Method 2:** We can **replace** time diversity method which repeats second data block, with receiver antenna diversity with a second antenna and we get 3 dB gain in SNR, given AWGN is independent in the two paths. To compare new method with two antenna diversity with Shannon's system, we reduce transmitted signal power by 3 dB.

Using Method 1 and 2, we get channel spectral efficiency $\frac{C}{W} \approx 2$ for binary signalling system, with 3 dB reduced transmitted signal power. Numerical simulations of new method A using Method 1 and 2 is shown in Matlab program, for $\frac{C}{W} = 1.98$, $\frac{C}{W} = 3.96$ and $\frac{C}{W} = 4.92$, with plots for system operating 1 dB below Shannon's Limit (plot and plot), using larger number of samples and higher computational complexity and we see that this system can operate below Shannon's Limit.

### *1.3.* ***Computational Complexity***

New method A needs to compare received signal with $M^n$ possible transmitted sequences resulting in very high computational complexity. We can reduce this complexity as follows.

Given the rate $\frac{1}{2}$, $(n, k)$ Gallager's LDPC error correction code used in new method A, for which $\lim_{n\to\infty} \frac{d_{min}}{n} = 0.11$, we can decode received sequence and quantize it to the transmitted codeword $T$ or one of the nearby codewords, depending on average noise power. As noise power gets larger, received sequence gets quantized into one of the nearby codewords given by $T'$.

Hence we can compute the detection function $D_s^{'}[m]$ with $s_m[i]$ set to the codeword sequence $T$ or one of the nearby codewords $T'$ and use the algorithm below.

1) If $D_s^{'}[m]$ goes to zero, as $n \to \infty$, then we have decoded the correct transmitted codeword and we exit.

2) If $D_s^{'}[m] = \frac{4d_{min}}{n}$, as $n \to \infty$, then we have decoded a codeword near the correct transmitted codeword and we compute $D_s^{'}[m]$ for one of the nearest neighbour codewords, till $D_s^{'}[m]$ goes to zero, as $n \to \infty$ and then we exit.

3) If $D_s^{'}[m] > \frac{4d_{min}}{n}$, as $n \to \infty$, then we have decoded a codeword slightly farther than the nearest neighbours to the correct transmitted codeword and we compute $D_s^{'}[m]$ for one of the nearest neighbour codewords, till $D_s^{'}[m]$ goes to zero, as $n \to \infty$ and then we exit.

This algorithm requires lesser computations than a brute force computation which computes $D_s^{'}[m]$ for all possible $M^n$ codewords.

### *1.4.* ***Differences between entry conditions in Shannon's Theorem and New Method A***

- Shannon's Theorem uses white-noise like n-sample transmitted sequences, as $n \to \infty$. New Method A uses **binary signalling** with each sample taking values of $\pm 1$ with **uniform** probability density function and hence the n-sample sequence is **not** like white noise with **gaussian** density function.

- Shannon's Theorem **does not** use $(n, k)$ error correction codes and channel capacity stated in the theorem equals source capacity. New method A **uses** $k = \frac{n}{2}$ bits and $(n, k)$ rate $\frac{1}{2}$ error correction code using Gallager's LDPC codes for which $\lim_{n\to\infty} \frac{d_{min}}{n} = 0.11$. If we do not use error correction codes and $d_{min} = 1$, then new method A does not work, given that $\frac{1}{n}\sum_{i=1}^{n} d_m^2[i] = \frac{4d_{min}}{n} = 0$ as $n \to \infty$ in Case A2 and Case A3 in Eq. 3.

- Shannon's Theorem uses the idea of non-overlapping noise spheres of radius $N$ surrounding each transmitted sequence in n-dimensional space and as $n \to \infty$, the noise samples will be confined to the surface of that sphere and adjacent spheres do not overlap, so that we can decode the transmitted sequence by using Minimum Mean-Squared Error(**MMSE**) criterion using detection function $D_s[m]$ in Eq. 1.(Shannon's paper pp452-454)

New method A **does not** use the idea of noise spheres. Instead it uses time diversity in two data blocks and uses a new Minimum Mean Multiplication of Errors in Two Data Blocks(**MMMETDB**) criterion and a **new** detection function $D_s^{'}[m]$ in Eq. 3, which has terms $D_3$ and $D_2$ involving multiplication and addition of n-sample noise sequences in the two blocks, when averaged, approach zero, as $n \to \infty$. Noise spheres **can overlap** in this new method and we use the multiplication and addition properties of AWGN sequences to get the desired result.

Hence the **entry conditions** specified in Shannon's Channel Capacity Theorem are **not** satisfied by the new method A and hence Shannon's Theorem is **not** applicable to new method A and hence new method A is **not** bound by channel capacity limits stated by Shannon's Theorem and is free to achieve **higher** channel capacity.

### *1.5.* ***Multiplication and addition of AWGN sequences***

b) We want to show that $D_2 = \frac{1}{n}\sum_{i=1}^{n} d_m[i](w[i] + w^{'}[i]) = 0$, as $n \to \infty$. We can divide the sequence $a_e[i] = d_m[i](w[i] + w^{'}[i])$ into $M$ blocks of $n_b$ samples each and write $w_n[m] =$

$\frac{1}{n_b} \sum_{i=(m-1)*n_b+1}^{(m-1)*n_b+n_b} d_m[i](w[i] + w^{'}[i])$ and compute $D_2 = \frac{1}{M}\sum_{m=1}^{M} w_n[m]$.

We see that $w[i] + w^{'}[i]$ is an AWGN sequence and multiplying it by $d_m[i] = \pm 2$ gives a random variable(RV) with a given probability density function(PDF), which when summed over $i = 1, ...n_b$ samples in a given block, gives a RV $w_n[m]$ with Gaussian density function, as $n_b \to \infty$ , using **Ergodicity** of random process which uses the fact that ensemble average equals the time average and using **Central Limit Theorem**, we see that addition of RV's with an arbitrary density function, we get a Gaussian RV $w_n[m]$. Then we compute the mean $D_3 = \frac{1}{M}\sum_{m=1}^{M} w_n[m]$ which approaches zero, as $n \to \infty$.

.

c) We want to show that $D_3 = \frac{1}{n}\sum_{i=1}^{n} w[i]w^{'}[i] = 0$, as $n \to \infty$. We can divide the sequence $b[i] = w[i]w^{'}[i]$ into $M$ blocks of $n_b$ samples each and write $w_n[m] = \frac{1}{n_b} \sum_{i=(m-1)*n_b+1}^{(m-1)*n_b+n_b} w[i]w^{'}[i]$ and compute $D_3 = \frac{1}{M}\sum_{m=1}^{M} w_n[m]$.

We see that $w[i]w^{'}[i]$ is a random variable(RV) with a given probability density function(PDF), which when summed over $i = 1, ...n_b$ samples in a given block, gives a RV $w_n[m]$ with Gaussian density function, as $n_b \to \infty$ , using **Ergodicity** of random process which uses the fact that ensemble average equals the time average and using **Central Limit Theorem**, we see that addition of RV's with an arbitrary density function, we get a Gaussian RV $w_n[m]$. Then we compute the mean $D_3 = \frac{1}{M}\sum_{m=1}^{M} w_n[m]$ which approaches zero, as $n \to \infty$.

Simulations in **awgn ergodic tests.m and shannon noise dmin ofdm.m**

## 2. Conclusion

In this paper, we have examined the implications of Shannon's Capacity theorem and examined if we can get past it.