

Classification and Capacity of Embedding Mechanisms

In classic communication, the gap between the theoretical Shannon channel capacity and the limitation in practice are filled by systematic and implementable studies on such issues as modulation, coding/decoding, and equalization [5]. Because data hiding problems have close connection with communication, filling the gap between theoretical embedding capacity and practical limitations is important and is the focus of the fundamental issue part (Part-I) of this thesis. More specifically, we follow the classification described in our work [167] to study two major categories of embedding mechanisms that are practically realizable. We shall begin with simplified models, then gradually loosen the assumptions to consider practical situations. The capacity achievable by the two types of schemes are compared, from which the conditions under which each type of schemes is superior to the other one are identified. We consider the following problems and propose solutions:

- *Distortion during and after embedding*: depending on applications, the allowable change introduced by embedding (a.k.a *embedding distortion*) may be

smaller, sometimes, much smaller, than the distortion introduced to the watermarked image (a.k.a *noise*). This is especially the case when the applications preside in a rivalry environment. For example, when watermark is used for ownership protection or access control, the quality of watermarked image/video determines the commercial and artistic value of the digital art works hence the embedding distortion should be well constrained to maintain superb imperceptibility. On the other hand, an adversary who wants to obliterate the watermark may be willing to tolerate some quality degradation. Under this scenario, the noise introduced by an adversary can be significantly larger than the embedding distortion.

- *Actual noise conditions*: an embedding system is generally designed to survive certain noise conditions. This single robustness-capacity tradeoff has limitation in practical applications. First, for applications presiding in a rivalry environment, the actual noise condition may vary dramatically. Second, a watermarked image/video may be compressed or transcoded to different bit rate in order to be delivered through different kinds of communication channels. The desirable amount of information extracted from the image/video could be different depending on the level of compression, as will be explained in Chapter 6. In addition, the information to be embedded usually requires unequal error protection (UEP). Some bits, such as the ownership information and a small amount of control information facilitating the decoding of a larger amount of payload bits, are required to be embedded more robustly than others.
- *Non-stationary property*: due to the non-stationary nature of perceptual sources, the amount of data that can be embedded varies significantly from region to region. This *uneven embedding capacity* adds difficulty to high-rate embedding,

especially in practical systems that need to accommodate diverse multimedia content.

In this chapter, we study the robustness-capacity tradeoff for two major categories of embedding mechanisms. The embedding capacity of simplified channel models for these two kinds of embedding are compared. These studies serves as a guideline for selecting an appropriate embedding algorithm given the design requirements of an application, and as a foundation of multi-level data hiding (Chapter 6), a new embedding framework/algorithm with improved performance. We also discuss in this chapter a number of modulation/multiplexing techniques for embedding multiple bits, quantitatively comparing the advantages and disadvantages of various techniques. The discussion of the uneven embedding capacity problem will be presented in the next chapter.

3.1 Two Types of Data Embedding

The mechanism for embedding one bit in original media is the most basic element in a data hiding system. Many embedding approaches have been proposed in the literature and there are many ways to classify them. For example, some schemes work with the multimedia signal samples while others work with transformed data. We found it beneficial to study the existing embedding approaches under noise-free conditions (i.e., directly passing a watermarked media to a detector) and to examine whether knowledge of the original host media will enhance the detection performance, regardless of whether a detector uses such knowledge or not [167]. Many existing embedding approaches would then fall in one of the following two categories.

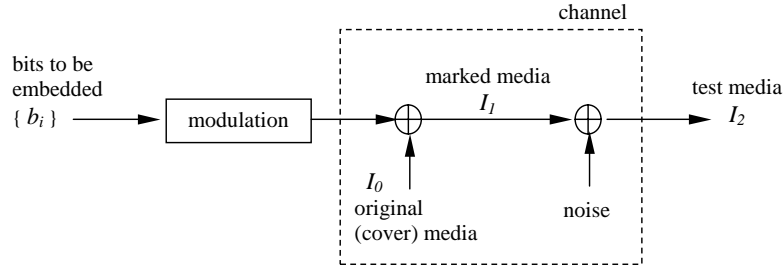


Figure 3.1: Channel model of Type-I embedding.

In the first category (Type-I), the secondary data, possibly encoded, modulated, and/or scaled, is added to the host signal, as illustrated in Fig. 3.1. The addition can be performed in a specific domain or on specific features. Considering the embedding of only one bit, the difference between marked signal I_1 and the original host signal I_0 is a function of b , the bit to be embedded, i.e., $I_1 - I_0 = f(b)$. Although it is possible to detect b directly from I_1 [53], I_0 can be regarded as a major noise source in such detection. Therefore, the knowledge of I_0 will enhance detection performance by eliminating the interference. Additive spread spectrum watermarking is a representative of this category [44, 46].

In the second category (Type-II), the signal space is partitioned into subsets which are mapped by a function $g(\cdot)$ to the set of values taken by the secondary data (e.g., $\{0, 1\}$ for binary hidden data), as illustrated in Fig. 3.2. The marked value I_1 is then chosen from the subset which maps to b , so that the relationship of $b = g(I_1)$ is deterministically enforced. To minimize perceptual distortion, I_1 should be as close to I_0 as possible, where the distance measure is chosen using perceptual models. Unlike the first category, the detector for this type of scheme does not need the knowledge of original value I_0 because the information regarding b is solely carried in I_1 .

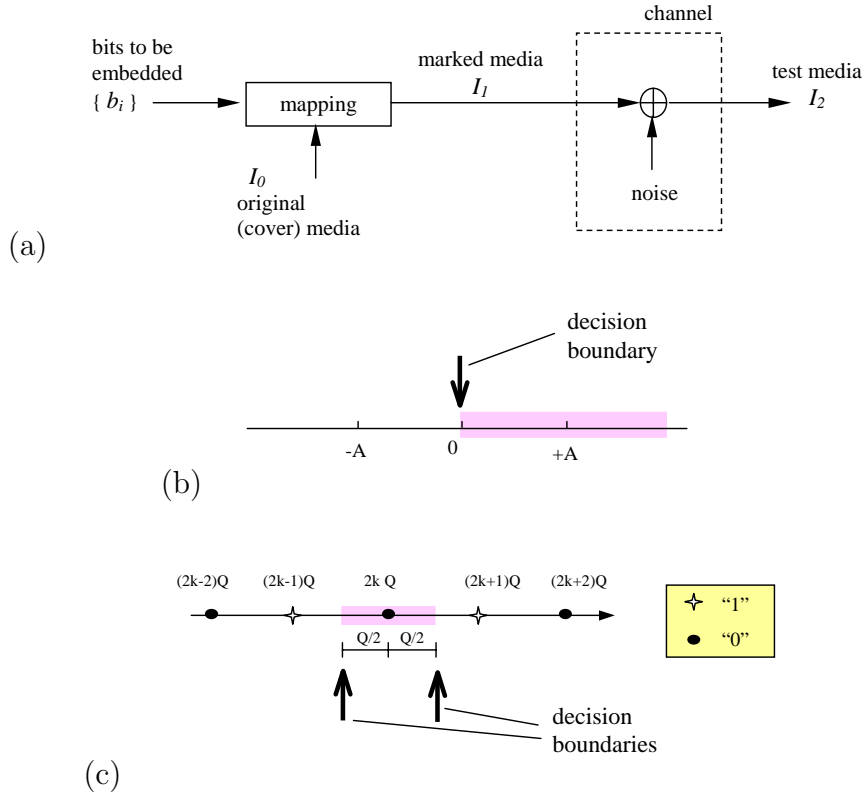


Figure 3.2: Channel model and decision boundaries of Type-II embedding: (a) channel model; (b) single-sided detection decision for sign enforcement with “0” as threshold; (c) detection decision for odd-even enforcement with two boundaries.

A simple example of Type-II is the so called *odd-even embedding*: we choose an even number as I_1 to embed a “0” and an odd number to embed a “1”. Data hiding can also be achieved by enforcing a rather global relationship. For example, one may change the sum of several source components to a nearby even number to encode a “0”, and to an odd number to encode a “1”. This is equivalent to reducing the bits allocated for representing the original vectors and to re-allocate them for conveying side information. By moving from 1-D space to a space of higher dimension, the magnitude of the introduced distortion per dimension is reduced.

Also, there are more choices to select a new signal vector with desired bits embedded in, which allows embedding to be performed in such a way that the human-visual-model-weighted distortion is minimized. On the other hand, the embedding bit rate is reduced, showing a tradeoff between embedding rate and invisibility¹. The odd-even embedding can be viewed as a special case of the table-lookup embedding [78, 163], which uses a lookup table to determine the mapping between the possible values of a media component and the data to be embedded. There are many other possible ways to partition the space and to enforce a desired relationship. One can enforce the ordering of a pair of samples or coefficients v_1 and v_2 . For example, we generate marked coefficients v'_1 and v'_2 close to v_1 and v_2 such that $v'_1 > v'_2$ to embed a “1” and $v'_1 \leq v'_2$ to embed a “0” [63]. One can also enforce signs to embed a “1” or “0”, as used in [64, 65]. Extending the basic ways of enforcement, more sophisticated schemes can be designed and/or analyzed [96]. Many proposed schemes in the literature that claimed to have the ability of non-coherent detection² belong to this category. It is the deterministically enforced relationship on I_1 that removes the need of using original signal I_0 . For the convenience of discussion, we shall refer the collection of image pixels or coefficients on which the relation is enforced as an *embedding unit*. If the enforcement is performed on a quantity derived from the embedding unit (e.g., the sum of a few coefficients, the signs of a coefficient, etc.), we shall refer the quantity as a *feature*.

¹Equivalently, if the embedding distortion per dimension is fixed, the total distortion that can be introduced increases when moving to higher dimensions. This aggregated energy enables embedding more reliably via quantization, as will be discussed in Sec. 3.1.1.

²Non-coherent detection in data hiding refers to being able to detect the embedded data without the use of the original unwatermarked copy. It is also called “blind detection”.

3.1.1 Comparison

The two types of schemes reveal different characteristics in terms of robustness, capacity and distortion (introduced by embedding), as shown in Table 3.1. For Type-I schemes, hypothesis testing is a tool for verifying what hidden data is present in the test media. Spread spectrum embedding, a representative of Type-I, has been demonstrated with excellent robustness and invisibility when the original host media is available in detection [44, 46]. In non-coherent detection, the interference from host signal exists even when there is no subsequent processing or intentional attack ³.

Table 3.1: Comparison of two types of embedding mechanisms

	<i>Type-I (Additive)</i>	<i>Type-II (Relation Enforcement)</i>
Capacity	low (host interference)	high
Robustness	high (rely on long seq.)	low (rely on quantization or tolerance zone)
Example	spread- spectrum embedding	odd-even embedding

We can analyze the detection performance via the following simplified additive model is:

$$\begin{cases} H_0 : Y_i = -S_i + M_i & (i = 1, \dots, n) & \text{if } b = -1 \\ H_1 : Y_i = +S_i + M_i & (i = 1, \dots, n) & \text{if } b = +1 \end{cases} \quad (3.1)$$

where $\{S_i\}$ is a deterministic sequence (sometimes called *watermark*), b is one bit to be embedded and is used to antipodally modulate S_i , M_i is the noise, and n

³Recently, Cox *et al.* modeled the additive embedding as communication with side information and proposed techniques of “informed embedding” to reduce (but not completely eliminate) the negative impact from host interference [54, 55].

is the number of samples/coefficients to carry the hidden information. We further assume b is equally likely to be “-1” and “+1”. In coherent detection where the original source is available, M_i comes from the processing and/or attack applied to the marked copy; in non-coherent detection, M_i consists of noise from the host media and processing/attack. For simplicity, M_i is usually modeled as i.i.d. gaussian distribution $N(0, \sigma_M^2)$, for which the optimal detector is a (normalized) correlator with S_i according to the classic detection theory [2]:

$$T_N = \underline{Y}^T \underline{S} / \sqrt{\sigma_M^2 \cdot \|\underline{S}\|^2} \quad (3.2)$$

where \underline{Y} and \underline{S} are column vectors. This test statistic is gaussian distributed with unit variance and the following mean

$$E(T_N) = b \cdot \sqrt{\|\underline{S}\|^2 / \sigma_M^2} \quad (3.3)$$

$$= b \cdot \sqrt{n \cdot \left(\frac{1}{n} \|\underline{S}\|^2\right) / \sigma_M^2} \quad (3.4)$$

We then compare T_N with zero, and decide H_1 if it is positive and H_0 otherwise. The probability of error is $\mathcal{Q}(E(T_N))$, where $\mathcal{Q}(x)$ is the probability of $P(X > x)$ of a gaussian random variable $X \sim N(0, 1)$. Because $\mathcal{Q}(\cdot)$ is monotonically decreasing, one should raise the ratio of total watermark energy $\|\underline{S}\|^2$ to noise power σ_M^2 to get a lower probability of error. Given the same noise power, this can be achieved by adding watermark to more elements, and/or by raising the watermark power (per element). A watermark with higher power introduces more distortion on the host media. The maximum watermark power is generally determined by perceptual models so that the changes introduced by the watermark are below the *just-noticeable-difference* (JND). Therefore, the remaining way to achieve robustness is to use large n , that is, to collect energy from many weak components of a long signal and to use such a long signal to represent one bit. A longer watermark vector in turn reduces the *capacity* (i.e., the

number of secondary information bits that can be encoded and extracted with very small probability of error).

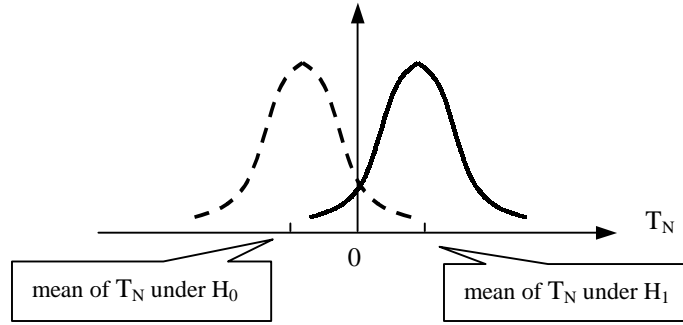


Figure 3.3: Illustration of the distribution of detection statistics (Type-I). Small mean value results in large probability of error.

It is worth mentioning that the hypothesis testing model in Eq. 3.1 is concerned with embedding one bit of information using antipodal modulation on a signal \underline{S} . Another popular model considers the case of a watermark being present versus being absent:

$$\begin{cases} H_0 : Y_i = M_i & (i = 1, \dots, n) & \text{if watermark is absent} \\ H_1 : Y_i = S_i + M_i & (i = 1, \dots, n) & \text{if watermark is present} \end{cases} \quad (3.5)$$

The watermark signal \underline{S} often represents ownership information [44, 46]. The detection statistics of this hypothesis problem is the same as the previous antipodal model. While the threshold can be set according to the Bayesian rule to minimize the overall probability of error as in the previous case, the Neyman-Pearson criterion is often adopted to minimize miss detection probability $P(\text{choose } H_0 | H_1 \text{ is true})$ while keeping the false alarm probability $P(\text{choose } H_1 | H_0 \text{ is true})$ below a bound.

Unlike Type-I, the Type-II schemes are free from the interference from host media and have the ability of coding one bit in only a small number of host components

hence high capacity. Their robustness against processing and attacks generally comes from quantization and/or tolerance zones. For schemes enforcing ordering or sign, instead of making minimal changes to enforce $v'_1 > v'_2$, the embedding mechanism may force $v'_1 > v'_2 + \delta$, where δ is the size of a tolerance zone. As long as distortion is smaller than δ , $v'_1 > v'_2$ can still be retained. For other enforcements, we may apply quantization to obtain robustness [167]⁴. For example, in the odd-even embedding, we pre-quantize the host signal I_0 by step Q , then enforce the quantized value to be an even number to embed a '0' and an odd number to embed a '1'. As shown in Fig. 3.2(c), any further distortion within $(-Q/2, +Q/2)$ will not cause errors in detection. The larger the Q is, the more tolerance we obtain, but also the larger distortion an embedding process may introduce. This is because the mean squared error introduced by embedding, as illustrated in Fig. 3.4, is

$$MSE = \frac{1}{2} \cdot \frac{Q^2}{12} + \frac{1}{2} \cdot \left[\frac{1}{Q} \int_{-Q/2}^0 (x+Q)^2 dx + \frac{1}{Q} \int_0^{Q/2} (x-Q)^2 dx \right] \quad (3.6)$$

$$= \frac{1}{3} Q^2, \quad (3.7)$$

where the host components within $\pm Q$ of kQ are assumed to follow uniform distribution, giving an overall MSE quadratic with respect to Q . The uniform distribution is a good approximation when Q is small. After embedding, noise N is introduced to marked components by processing/attack, then an inverse quantization is performed by a watermark detector. If $|N| < Q/2$, no error will be introduced in detection. If N is uniformly distributed between $-M/2$ and $M/2$ where $M > Q$, the probability of error can be expressed on interval basis:

$$P_e = \begin{cases} 1 - \frac{(2k-1)Q}{M} & M \in [(4k-3)Q, (4k-1)Q] \\ \frac{2kQ}{M} & M \in [(4k-1)Q, (4k+1)Q] \end{cases} \quad (3.8)$$

⁴The enforcement with quantization is formulated in a slightly different way by Chen *et al.* [72] and is referred as *Quantization Index Modulation (QIM)*.

where k is a positive integer. Here P_e fluctuates around $1/2$ and converges to $1/2$ as k goes large. While a more sophisticated set partition may achieve a better tradeoff between the perceptual distortion introduced by embedding and the tolerance against certain processing or attacks, one can see that the tolerance is always limited and is achieved by pre-distortion in the embedding step. By incorporating a proper human visual model, Type-II schemes are suitable for high-rate data hiding applications that do not have to survive severe noise.

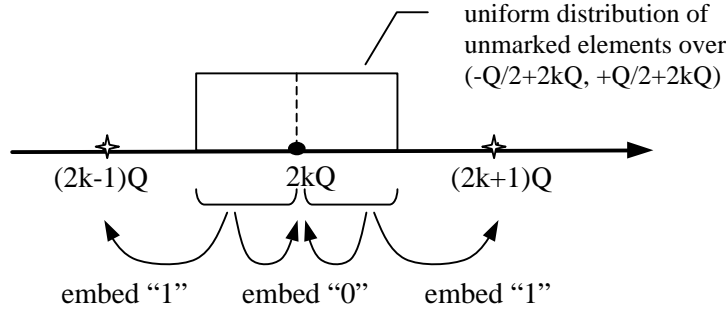


Figure 3.4: Computing MSE distortion by odd-even embedding

For both types of embedding, when the probability of detection error per embedding unit (such as in one or a set of image pixels or coefficients) is non-zero, proper channel coding like error correction codes (ECC) can be applied to achieve reliable embedding under certain noise conditions. Properly constructed codes are ways to approach the embedding capacity to be discussed in Section 3.2.

3.1.2 A Unified View on Two Types

A unified view of the two embedding types can be obtained in terms of set partitioning: both types partition signal space into several subsets, each of which represents

a particular value of hidden data. We have already explained the set partitioning for Type-II. Now considering Type-I, for example, the additive spread spectrum scheme, we can see that the signal space for detection is also partitioned into two parts according to the sign of test statistics T_N which usually takes the form of correlation or a variation of correlation: the positive part represents a ‘1’ and the negative part represents a ‘0’. The difference is that for Type-I, enforcement is not done on the marked media deterministically. The additive embedding alone still leaves non-zero probability for marked components not to be enforced to the desired set in non-coherent detection. Because the host media is a major noise source, we have to rely on a statistical approach (e.g., spreading watermark signal to many components and taking average) to suppress noise and to obtain detection result with small probability of error. This effort is needed even when there is no noise coming from processing/attack.

Most recently, attentions have been paid to Costa’s theoretical work in the early 1980s on the channel capacity under two additive gaussian noise sources with one noise source being known to the sender [97]. Costa showed that the channel capacity is equal to the capacity in the absence of the known noise source and that the optimal transmitter adapts its signal to the state of the known noise source rather than attempting to cancel it. Incorporating Costa’s work into the data hiding problem, as suggested by Moulin *et al.* and Chen *et al.*, provides an alternative unified view on the two embedding types [93, 72].

3.2 Quantified Capacity Study

The difference between the two types of schemes in terms of robustness-capacity tradeoff can be quantified using an additive channel model. For simplicity, we assume additive white gaussian noise (AWGN) for both types. Other additive noise conditions such as additive white uniform noise (AWUN) and colored noise can be studied similarly by applying whitening and/or re-computing the capacity based on information theory. It is important to notice that the capacity is tied to a channel model with specific noise distribution and watermark signal constraints. The capacity would be different if the channel is modeled differently, and it is a function of the parameters of the noise distribution and watermark constraints, such as the power of the noise and of the watermark.

Capacity for Type-I Embedding

The channel model of Type-I schemes shown in Fig. 3.1 has continuous input and continuous output (CICO). The additive noise consists of two parts, namely, the interference from the host signal and the noise due to other processing/distortion. For the moment, we assume that (1) the host signal is independent of the processing noise, and (2) both are i.i.d. gaussian distributed. The embedding capacity under the overall AWGN noise is achieved with gaussian distributed input and is equal to

$$C_{CICO} = \frac{1}{2} \log_2 \left(1 + \frac{A^2}{\sigma_I^2 + \sigma^2} \right). \quad (3.9)$$

where σ_I^2 is the power of the original host signal, A^2 is the power of the embedded signal, and σ^2 is the power of additive processing noise. In general, the interference from host signal is much stronger than the processing noise, i.e., $\sigma_I^2 \gg \sigma^2$.

Capacity of Type-II Embedding

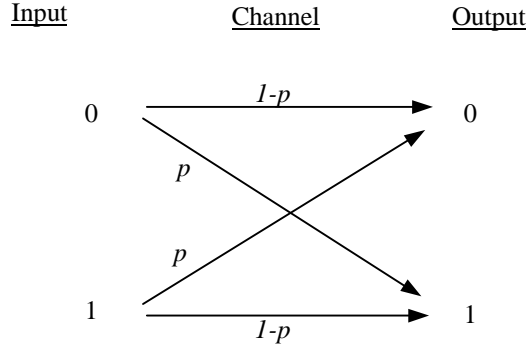


Figure 3.5: A binary symmetric channel (BSC) with a flipping probability of p .

The channel of Type-II schemes has discrete input with the decision boundary being either single sided or double sided, as illustrated in Fig. 3.2. The single-sided case generally corresponds to the embedding relying on sign enforcement, and the double-sided case is common to denser enforcement such as odd-even and lookup table embedding. We shall first study the single sided case, and extend the result to the double-sided case later. Regarding the output of the channel, if a hard decision is used, the channel will be the discrete-input discrete-output (DIDO) binary symmetric channel (BSC) shown in Fig. 3.5. The capacity of this type of channel has been well studied and is given by

$$C_{DIDO} = 1 - h_p \tag{3.10}$$

achieved by equiprobable input, where h_p is the binary entropy

$$h_p = p \cdot \log\left(\frac{1}{p}\right) + (1 - p) \cdot \log\left(\frac{1}{1 - p}\right). \tag{3.11}$$

The probability of bit error p for AWGN and AWUN noise are

$$p_{AWGN} = \mathcal{Q}\left(\frac{A}{\sigma}\right) = \int_{\frac{A}{\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \tag{3.12}$$

$$p_{AWUN} = \begin{cases} \frac{1}{2} - \frac{A}{M} & A < M/2 \\ 0 & A \geq M/2 \end{cases}, \quad \text{with } \sigma = M/\sqrt{12}. \quad (3.13)$$

where the AWUN noise is uniformly distributed between $-M/2$ and $+M/2$ (noise variance $\sigma^2 = M^2/12$), If a soft decision is used, which offers a detector the knowledge of how wrong the decision of a particular element would be, the channel will be discrete-input continuous-output (DICO), having a capacity of

$$C_{AWGN,DICO} = 1 + \frac{A^2}{\sigma^2} \log_2 e - E[\log_2(e^{\frac{2AY}{\sigma^2}} + 1)], \quad (3.14)$$

where $E[\cdot]$ is the expectation with respect to Y , whose probability density function is

$$f(y) = \frac{1}{2\sqrt{2\pi\sigma^2}} e^{-\frac{(y+A)^2}{2\sigma^2}} + \frac{1}{2\sqrt{2\pi\sigma^2}} e^{-\frac{(y-A)^2}{2\sigma^2}}. \quad (3.15)$$

When the noise is AWUN between $-M/2$ and $+M/2$, we can show that:

$$C_{AWUN,DICO} = \begin{cases} \frac{2A}{M} & A < M/2 \\ 1 & A \geq M/2 \end{cases} \Rightarrow C_{AWUN,DICO} = \begin{cases} \frac{A}{\sqrt{3}\sigma} & \frac{A}{\sigma} < \sqrt{3} \\ 1 & \frac{A}{\sigma} \geq \sqrt{3} \end{cases} \quad (3.16)$$

The soft decision allows the *watermark signal-to-noise ratio* A/σ^2 being 2 ~ 5dB lower for the same capacity than the hard decision under AWGN or AWUN noise, as shown in Fig. 3.6. The derivations of $C_{AWGN,DICO}$ and $C_{AWUN,DICO}$ are included as an appendix in Sec. 3.5.

Up to now, we have discussed a simple case with only two signal points $-A$ and $+A$ conveying one bit information. Capacity of different channel models, namely, DICO and DIDO, have been studied and compared. With a few small variations, we can obtain the capacity of several typical Type-II schemes. For practical schemes that enforces signs with a tolerance zone A , the signal are generally enforced to be greater than $+A$ or less than $-A$ to encode one bit, rather than being enforced exactly to

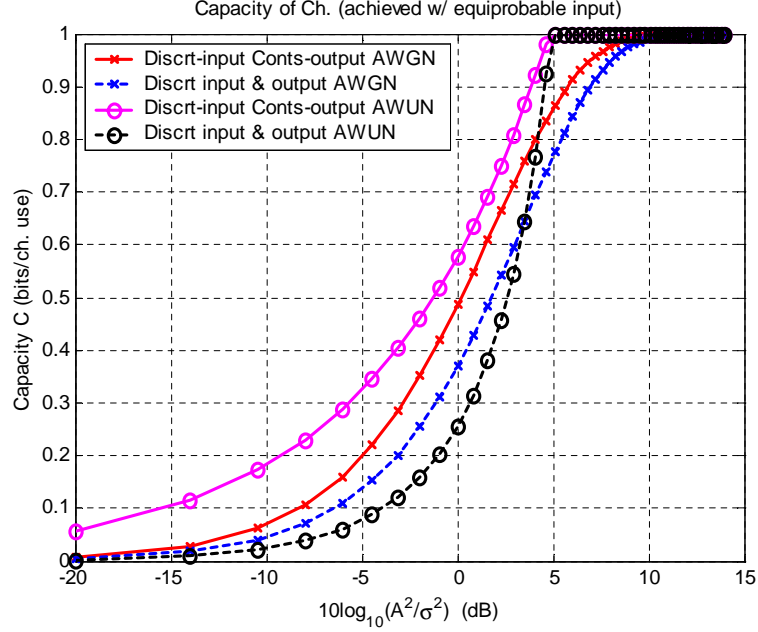


Figure 3.6: Capacity of DICO and DIDO channels under AWGN and AWUN noise.

$\pm A$ in our simplified model. This implies that the capacity should be higher than that of our model. Regarding the odd-even embedding mentioned earlier, the error regions are two-sided rather than single-sided (Fig. 3.2). We denote the quantization step size as Q and consider the channel input $X = kQ$ (i.e., the enforced values that carry hidden information), error will be incurred when the output Y is in the regions $Y > (k + 1/2)Q$ or $Y < (k - 1/2)Q$. Thus for the model of DICO channel with AWGN noise, the bit error probability p is:

$$\begin{aligned}
 p_{AWGN} &= \min \left\{ 1/2, 2 \cdot \sum_{k=0}^{+\infty} \mathcal{Q}\left(\frac{(4k+1)Q}{2\sigma}\right) - \mathcal{Q}\left(\frac{(4k+3)Q}{2\sigma}\right) \right\} \\
 &= \min \left\{ 1/2, 2 \cdot \sum_{k=0}^{+\infty} \int_{\frac{(4k+1)Q}{2\sigma}}^{\frac{(4k+3)Q}{2\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right\}, \quad (3.17)
 \end{aligned}$$

For high watermark-to-noise ratio ($\frac{Q}{\sigma}$), we may ignore the regions that are far away

from kQ but map to the same bit value as kQ and approximate the probability with an upper bound:

$$p_{AWGN} \approx \min\{1/2, 2 \cdot \mathcal{Q}(\frac{Q}{2\sigma})\} = \min\{1/2, 2 \cdot \int_{\frac{Q}{2\sigma}}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt\}, \quad (3.18)$$

This approximation is based on the fast decay of the tails of gaussian distribution. Plugging the result of p_{AWGN} into Eq. 3.10 yields the channel capacity.

We have discussed *DIDO* and *DICO* channel models that reflect the practically implementable Type-II embedding. The channel model for Type-II embedding can be further improved. Motivated by Costa's techniques in proving the channel capacity [97], Chen *et al.* proposed to incorporate multiplicative scaling into quantization-based enforcement embedding. The enforcement is then linearly combined with the host signal to form a watermarked signal. The scaling factor is a function of watermark-to-noise ratio and has the capability of enhancing the number of bits that can be embedded. Interested readers may refer to [40, 72] for details.

Capacity Comparison for Type-I & Type-II

Fixing the mean squared error introduced by the embedding process as E^2 , we compare the capacity of Type-I and Type-II schemes under AWGN noise with the following simplification. For Type-I, we consider a CICO channel model and assume that the AWGN noise consists of gaussian processing noise (with variance σ^2) and host interference (with standard deviation 10 times as much as that of the watermark signal, i.e., $\sigma_I = 10E$). For Type-II, we consider a DIDO BSC channel with p derived in Eq. 3.17 for odd-even embedding with such quantization step Q that the embedding MSE distortion equals to E^2 , i.e., $Q = \sqrt{3}E$ according to Eq. 3.7. The capacity is thus obtained as:

$$C_I = \frac{1}{2} \log_2\left(1 + \frac{E^2}{(10E)^2 + \sigma^2}\right) \quad (3.19)$$

$$C_{II} = 1 - h_{\min\{1/2, 2 \cdot \sum_{k=0}^{+\infty} \mathcal{Q}(\frac{(4k+1)Q}{2\sigma}) - \mathcal{Q}(\frac{(4k+3)Q}{2\sigma})\}} \quad (3.20)$$

We plot capacity vs. different watermark-to-noise ratio E^2/σ^2 in Fig. 3.7. It shows that the capacity of Type-II is much higher than that of Type-I until the watermark-to-noise ratio (WNR) falls negative, confirming our previous analysis regarding the host interference of Type-I and the pre-distortion nature of Type-II. The comparison suggests that Type-II is useful under low noise condition while Type-I is suitable for severe noise. The capacity of both Type-I and Type-II can be approached via channel coding, such as RS / BCH codes used in [61, 167].

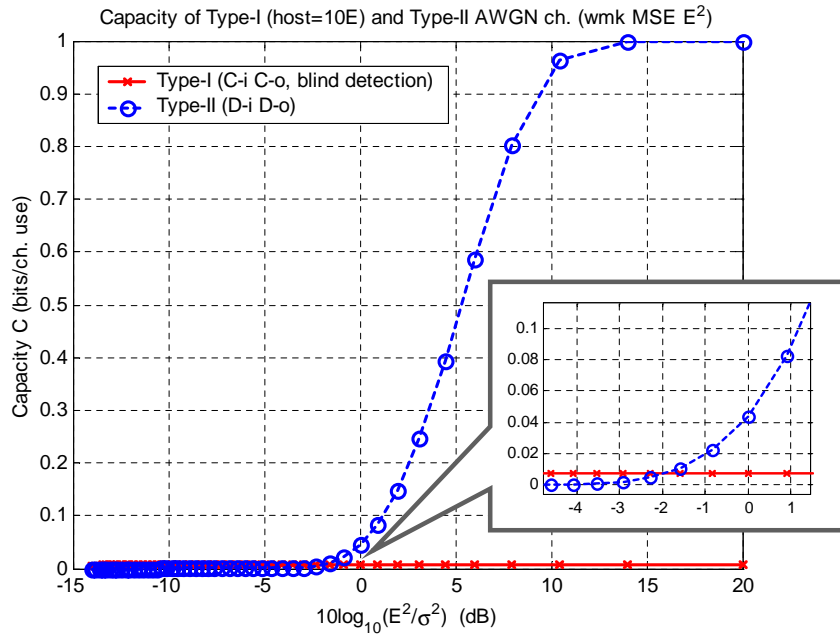


Figure 3.7: Capacity of Type-I (CICO channel) and Type-II (DIDO channel) embedding under AWGN noise.

A Few Extensions

We have discussed the two possible channel models for Type-II embedding, namely,

DICO and DIDO channels, and have compared them with the CICO channel model for Type-I embedding. Throughout the above discussion, we assumed that the noise is additive and white, and the watermark signal power is the same on all media components. Generalization is possible on two aspects. First, We should consider the so-called *unembeddable* components which have to be left untouched by the embedding mechanism to meet imperceptibility requirement. As we shall see in Chapter 4, these unembeddable components reduce the data hiding capacity. Second, we should consider each media components might incur different host interference, be able to watermarked with different strength, and/or sustain different noise. The channel model can be modified into a parallel channel, with L bands per unit or channel use. The ratio of watermark to interference-plus-noise is different for in each band, but the total noise is assumed to be independent from band to band and is i.i.d. from unit to unit (Fig. 3.8). The channel capacity per unit becomes:

$$C = \max_{P_{X^{(L)}}} I(X^{(L)}; Y^{(L)}), \quad (3.21)$$

where

$$\begin{aligned} I(X^{(L)}; Y^{(L)}) &= I([X_1, \dots, X_L]; [Y_1, \dots, Y_L]) = h(Y^{(L)}) - h(Y^{(L)}|X^{(L)}) \\ &= h(Y^{(L)}) - h(N_1, \dots, N_L) = h(Y_1, \dots, Y_L) - \sum_{i=1}^L h(N_i) \\ &\leq \sum_{i=1}^L h(Y_i) - \sum_{i=1}^L h(N_i) = \sum_{i=1}^L I(X_i; Y_i). \end{aligned} \quad (3.22)$$

If $\{X_i\}$ are independent of each other, the equality of Eq. 3.22 holds, indicating that the total capacity is the sum of the capacity achieved by each individual band. The capacity of an individual band can be determined by the noise power and the just-noticeable-difference of the band, using the approach we discussed in the earlier part of this section. Studying the capacity under correlated noise from unit to unit and/or non-gaussian noise is more involved. It is a direction of future work.

A noteworthy issue regarding the parallel channels is that in classic communication literature, the capacity of L parallel AWGN channels follows a so-called *Water-filling Theorem*, where a constraint on total power is imposed and the power needs to be shared among all channels. The theorem suggests an optimum allocation of the power among L parallel channels, which is analogous to water-filling. Though meaningful in the cases of telecommunication, the constraint on the total power may not be realistic in watermarking problems where the power constraint for each individual channel (possibly in the form of a frequency band or a local region) is determined by perceptual models such as those in [101, 36, 48]. Even though the perceptual constraints may have dependency among several channels, it does not appear to fit the simple constraints on the total power. How to better incorporate the dependency among channels is another direction to be explored.

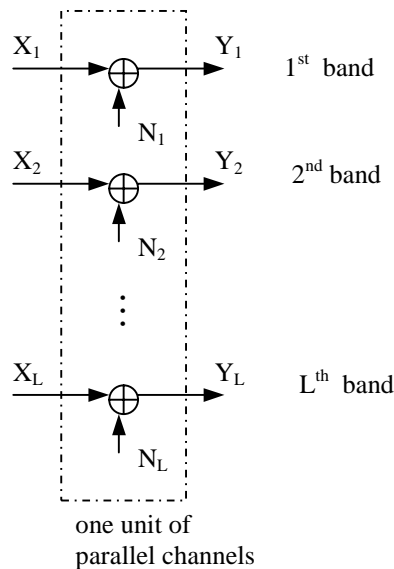


Figure 3.8: Parallel Channel Model With L Bands. Noise in each band is independent but with different variance.

We shall also note from the above analysis that the discrete-input channel model for Type-II has zero probability of detection error if the support of noise distribution is within the decision boundary shown as shaded area in Fig. 3.2. Under the specific model considered there, the channel is able to convey 1 bit per channel use with no error. If we revise the channel model to allow freedom in choosing the input alphabet, the channel capacity is determined by Shannon's zero-error capacity [99, 7], as suggested in [41]. The capacity may exceed 1 bit per channel use for a specific noise distribution and specific power constraint on the watermark.

3.3 Bandwidth via Data Hiding

A specious argument about data hiding is that it could provide *additional* bandwidth to convey secondary information. Actually the bandwidth for conveying secondary information is at an expense of either reducing the bandwidth for conveying host media or increasing the total bandwidth of conveying the watermarked media, depending on whether the quality of host media is reduced or not. Considering the simplest case of embedding one bit in an image (Fig. 3.9), the entire image space \mathcal{S} is always partitioned into two disjoint subsets \mathcal{S}_1 and \mathcal{S}_2 with $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$, regardless of the specific embedding algorithm being used. When a detector sees an image belonging to \mathcal{S}_1 , it will output the embedded bit value as "0"; when it sees an image belonging to \mathcal{S}_2 , it will output the embedded bit value as "1". Assuming the embedded bit takes "0" and "1" equiprobably, the probability of an image falling into the first and the second subset equals to 1/2, respectively. To code any image in the space, one bit is spent on specifying to which subset the image belongs. In other words, one of the bits used in coding an image is actually for conveying the embedded bit.

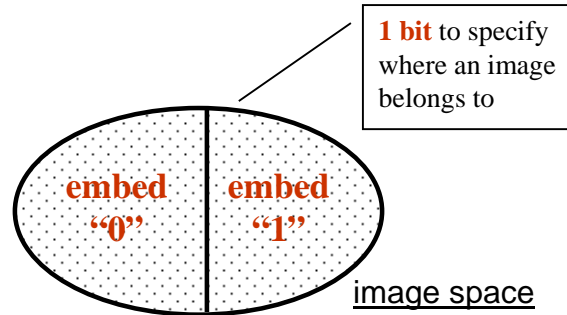


Figure 3.9: Illustration of the bit re-allocation nature of data hiding.

For odd-even embedding in a single image element, the two subsets are obtained by a partition according to the least significant bit (LSB); for an odd-even enforcement applied to the sum of two components, the embedded bit is related to the LSB of both components. For spread spectrum additive embedding, the boundary between the two subsets is commonly determined by a correlator-type detector. If the total number of bits for representing an image do not change during the embedding process, one bit is logically reallocated from representing the image to representing the embedded data, even though there may be more than one bits physically related to the embedded data. In this case, the absolute quality of the image is reduced because of the fewer bits effectively used in image representation (unless there is redundancy in the previous representation). While this indicates that data hiding does not have an advantage in terms of saving bit rate when compared with attaching the secondary data separately to the host media, it does have quite a few advantages, including the ability to associate the secondary data with the host media in a seamless way as well as the standard compliant appearance and the low computation complexity in practical applications. We will discuss this in Chapter 8.

3.4 Modulation and Multiplexing Techniques for Embedding Multiple Bits

An embedding scheme generally specifies a particular way to hide one-bit information in multimedia source. Modulation/multiplexing techniques can be applied on top of it to embed multiple bits. Evolving from classic communication [5], the following strategies are commonly used:

- *Amplitude modulo modulation* (for Type-II embedding ⁵)

In general, B bits can be embedded in each embedding unit by enforcing a feature derived from this unit into one of M subsets, where $B = \log_2 M$. A straightforward example extended from odd-even embedding is to enforce the relation via modulo- M operation to hide B bit per element. That is,

$$I_1 = \arg \min_{I \text{ s.t. } I=kQ, k \in \mathbf{Z}, \text{mod}(k, M)=m} |I - I_0| \quad (3.23)$$

where $[\cdot]$ represents the rounding operation, $m \in \{0, 1, \dots, M-1\}$ represents the B -bit information to be embedded, I_0 is the original image feature, I_1 is the watermarked feature, and Q is the quantization step size for obtaining robustness. Assuming I_0 follows uniform distribution in each quantization interval $(kQ - \frac{Q}{2}, kQ + \frac{Q}{2})$ where k is an integer, we can show that the MSE distortion introduced by embedding is $Q^2 M^2 / 12$. This indicates that with the minimal separation Q between the M subsets being fixed, larger embedding distortion will be introduced by a larger M ; with fixed MSE embedding distortion, the

⁵The Type-I additive embedding formulated in Eq. 3.1 (the antipodal modulation) and Eq. 3.5 (the on-off modulation) can be viewed as amplitude modulation. For blind detection of additive embedding that is subject to host interference, using amplitude modulation to convey more than two constellation points are rare in practice. We therefore focus on the amplitude modulo modulation that is applicable to Type-II embedding.

enforced relation with a larger M has smaller separation hence tolerates less distortion. The idea is easily extensible to table lookup embedding or other enforcement scheme and the analysis can be done similarly.

- *Orthogonal & Biorthogonal modulation*

Mainly used with Type-I additive embedding, the *orthogonal modulation* uses M orthogonal signals to represent $B = \log_2 M$ bits by embedding one of the M signals into the host media. A detector computes the correlation with respect to all M signals. The signal that gives the largest correlation and exceeds some threshold will be decided as the signal embedded by the sender and the corresponding B -bit value will be determined accordingly. A variation, so-called *biorthogonal modulation*, encodes $\log_2 2M = (B + 1)$ bits by adding or subtracting one of M signals. The computational complexity of detection is exponential with respect to the number of bits being conveyed, therefore is inefficient except for small M .

- *“TDMA” type modulation*

For data hiding in images, this type of modulation partitions an image into non-overlapped regions and hides one or several bits in each region. For audio, it means to partition an audio into time segments and to hide one or several bits in each segment. A video can be partitioned into regions within each frame and into time segments across frames. “TDMA” type modulation is a simple way to realize orthogonal embedding for both Type-I and Type-II, i.e., the bits embedded in different regions or segments do not interfere with each other. However, it could suffer from the problem of uneven embedding capacity, which will be explained in Section 4.

- “CDMA” type modulation

For Type-I additive embedding, encoding B bits to a watermark signal \underline{w} in the following way provides more efficient detection than orthogonal modulation in terms of the computational complexity:

$$\underline{w} = \sum_{k=1}^B b_k \cdot \underline{u}_k, \quad (3.24)$$

where $b_k = \pm 1$. In general, the vectors $\{\underline{u}_k\}$ are chosen to be orthogonal to each other. The orthogonality implies that the total signal energy is the sum of the energy allocated for each bit. If a fixed amount of energy is uniformly allocated to each bit, the energy per bit will be reduced as B increases, implying a decrease in detection reliability and more generally, a limit on the total number of bits that can be hidden for low error rate extraction. “TDMA” is a special case with the supports of \underline{u}_k being non-overlapped with each other in the sample domain (i.e., the pixel domain for image and the time domain for audio). Alternatively, we can choose orthogonal but overlapped $\{u_k\}$, similar to CDMA in communication [6]. Uneven embedding capacity is no longer a concern as we can choose $\{u_k\}$ such that each bit is spreaded all over the media. But B orthogonal sequences have to be generated and shared with a detector, which may be non-trivial for large B . The TDMA and CDMA approach can be combined to encode multiple bits.

For Type-II embedding, multiple bits can be embedded by enforcing relations deterministically along multiple directions that are orthogonal to each other. Swanson *et al.* and Alghoniemy *et al.* proposed to embed multiple bits in an image block by enforcing relations on the projections of a feature vector along several orthogonal directions [66, 68]. The total modification introduced by

embedding is the sum of the change along each direction, implying a tradeoff among capacity, robustness, and imperceptibility.

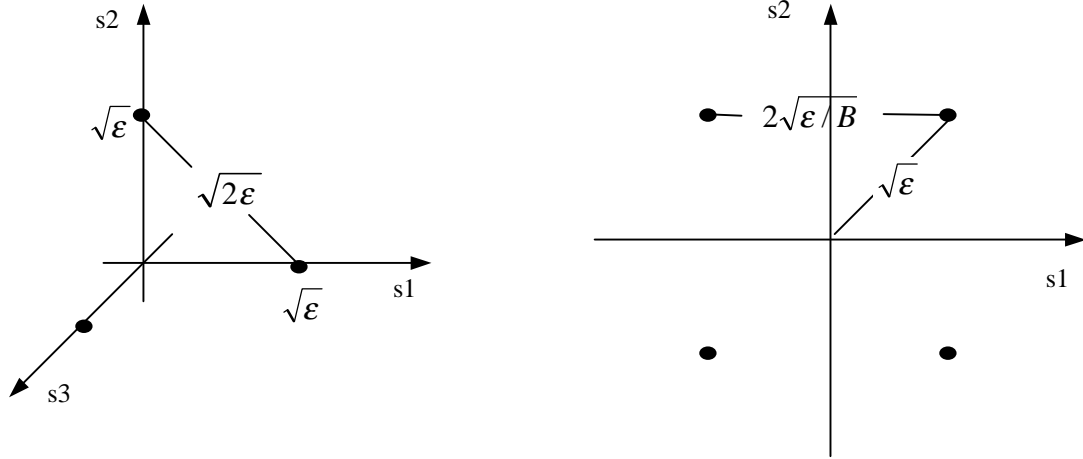


Figure 3.10: Comparison of distance between signal constellation points for orthogonal modulation (left) vs. TDMA/CDMA-type modulation (right) with total signal energy being fixed at \mathcal{E} .

The orthogonal modulation and TDMA/CDMA-type modulation can be compared by studying the distance between signal constellation points that represent the secondary data (Fig. 3.10). Considering the case of conveying B bits using total energy \mathcal{E} . The minimum distance between signal points is $\sqrt{2\mathcal{E}}$ for orthogonal modulation, and is $2\sqrt{\mathcal{E}/B}$ for TDMA/CDMA. When $B > 2$, orthogonal modulation gives smaller probability of error at a cost of detection complexity.

Expanding a bit more, we summarize the quantified comparison among the modulation/multiplexing techniques discussed above ⁶ in Table 3.2. The modulation/multiplexing is applied to one embedding unit of S elements. The quantity

⁶The modulo- M modulation extended from odd-even embedding is taken as a representative of amplitude modulation.

$\mathcal{W} = \frac{\mathcal{Y}}{\mathcal{X} \cdot \mathcal{Z}^2}$ measures the energy efficiency of embedding, where \mathcal{Y} is the MSE distortion per element introduced by embedding, and \mathcal{Z} is the minimum separation between the enforced constellation points hence reflects the robustness against noise. Because \mathcal{W} describes the MSE embedding distortion per bit per unit squared separation distance, a smaller value is more preferable. We can see that except for very small S and B , biorthogonal techniques has the smallest \mathcal{W} values, while the amplitude modulo technique gives large \mathcal{W} values as M goes larger – it equals to $\frac{1}{3}$ for $M = 2$, and to $\frac{2}{3}$ for $M = 4$. TDMA and CDMA modulation, being applicable to both Type-I and Type-II embedding under blind detection and having a constant \mathcal{W} value of $\frac{1}{4}$, show a good balance between energy efficiency and detection complexity, therefore are suitable for many applications. Further, TDMA or CDMA can be combined with orthogonal or biorthogonal modulation to enhance the embedding rate while balancing the detection complexity.

3.5 Appendix - Derivations of Type-II Embedding Capacity

In this appendix section, we derive the capacity under DICO channel model for Type-II embedding. We shall consider AWGN and AWUN noises, and show the capacity under these channels follow Eq. 3.14 and Eq. 3.16, respectively.

According to information theory [7], the channel capacity is

$$C = \max_{p(x)} I(X; Y) \quad (3.25)$$

where $I(X; Y)$ is the mutual information between two random variables X and Y .

Table 3.2: Comparison of Modulation/Multiplexing Techniques

(S elements per embedding unit, $B \leq S$)

	<i>Amplitude Modulo</i>	<i>TDMA / CDMA</i>	Orthogonal	Biorthogonal
Type-I embed.		v	v	v
Type-II embed.	v	v		
\mathcal{X} # embedded bits per element	$\frac{\log_2 M}{S}$	$\frac{B}{S}$	$\frac{\log_2 B}{S}$	$\frac{\log_2 2B}{S}$
\mathcal{Y} MSE distortion per element	$\frac{Q^2 M^2}{12S}$	$\frac{\mathcal{E}}{S}$	$\frac{\mathcal{E}}{S}$	$\frac{\mathcal{E}}{S}$
\mathcal{Z} minimum separation	Q	$2\sqrt{\frac{\mathcal{E}}{B}}$	$\sqrt{2\mathcal{E}}$	$\sqrt{2\mathcal{E}}$
$\mathcal{W} = \frac{\mathcal{Y}}{\mathcal{X} \cdot \mathcal{Z}^2}$	$\frac{M^2}{12 \log_2 M}$	$\frac{1}{4}$	$\frac{1}{2 \log_2 B}$ $\left(\geq \frac{1}{2 \log_2 S}\right)$	$\frac{1}{2(1+\log_2 B)}$ $\left(\geq \frac{1}{2(1+\log_2 S)}\right)$

For a channel with continuous outputs, we have

$$I(X; Y) = h(Y) - h(Y|X) = h(Y) - h(X + Z|X) = h(Y) - h(Z) \quad (3.26)$$

where $h(\cdot)$ is the differential entropy of a continuous random variable, $h(\cdot|\cdot)$ is the conditional differential entropy, and Z is additive noise that is independent of the channel input. Consider first the case of AWGN noise $N(0, \sigma^2)$, whose differential entropy is known as $\frac{1}{2} \log(2\pi e \sigma^2)$. We have

$$I(X; Y) = E[-\log f_Y] - \frac{1}{2} \log(2\pi e \sigma^2) \quad (3.27)$$

where the expectation $E[\cdot]$ is with respect to the random variable Y whose probability

density function (p.d.f.) f_Y is a bimodal gaussian (?), i.e.,

$$f_Y(y) = P(X = -A) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+A)^2}{2\sigma^2}} + P(X = +A) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-A)^2}{2\sigma^2}} \quad (3.28)$$

By symmetry, the capacity is achieved by equiprobably input, i.e., $P(X = -A) = P(X = +A) = 1/2$. We now have

$$\log f_Y(y) = -\log 2 - \frac{1}{2} \log(2\pi\sigma^2) - \frac{A^2}{2\sigma^2} \log e - \frac{y^2}{2\sigma^2} \log e + \log(e^{-B} + e^B) \quad (3.29)$$

where $B = yA/\sigma^2$. The term $\log(e^{-B} + e^B)$ can be simplified as

$$\log(e^{-B} + e^B) = \log \frac{e^{2B} + 1}{e^B} = \log e^{\frac{2yA}{\sigma^2}} + 1 - \frac{yA}{\sigma^2} \log e \quad (3.30)$$

We take expectation with respect to Y on every term in $-\log f_Y(y)$ and obtain

$$h(Y) = \log 2 + \frac{1}{2} \log(2\pi\sigma^2) + \frac{A^2}{2\sigma^2} \log e + \frac{\log e}{2\sigma^2} E(Y^2) - E[\log e^{\frac{2AY}{\sigma^2}} + 1] \quad (3.31)$$

where the term $E[\frac{YA}{\sigma^2} \log e]$ vanishes because Y has zero mean. With $E(Y^2) = \sigma^2$ and some more rearrangement, we arrive at

$$C_{AWGN,DICO} = \log 2 + \frac{A^2}{\sigma^2} \log e - E[\log(e^{\frac{2AY}{\sigma^2}} + 1)]. \quad (3.32)$$

Therefore, the channel capacity in unit of bit per channel use under AWGN noise is

$$\boxed{C_{AWGN,DICO} = 1 + \frac{A^2}{\sigma^2} \log_2 e - E[\log_2(e^{\frac{2AY}{\sigma^2}} + 1)]} \quad (3.33)$$

For AWUN noise between $-M/2$ and $+M/2$ (noise variance $\sigma^2 = M^2/12$), the differential entropy of noise is

$$h(Z) = \int_{-M/2}^{M/2} \frac{1}{M} \log M dz = \log M \quad (3.34)$$

The shape of the output Y 's distribution depends on the relations between M and A . We can show that

$$h(Y) = \begin{cases} \frac{2A}{M} h(p) + \log M & A < M/2 \\ h(p) + \log M & A \geq M/2 \end{cases} \quad (3.35)$$

where p is the probability of the channel input $p = P(X = -A)$, and $h(p)$ is the binary entropy defined as $h(p) = -p \cdot \log p - (1 - p) \cdot \log(1 - p)$. By noticing $h(p)$ assumes its maximum at $p = 1/2$, we have

$$C_{AWUN,DICO} = \begin{cases} \frac{2A}{M} & A < M/2 \\ 1 & A \geq M/2 \end{cases} \quad (3.36)$$

where the capacity is achieved by equiprobable inputs.