SOLUTION 1. Let $E_1, \ldots, E_k$ be events with $\Pr(E_i) = p_i$. Let $E = \bigcup_i E_i$ be the union of the events. We know, by the union bound, that $\Pr(E) \leq \sum_i p_i$. By noting that the probability of any event is at most 1, we can trivially improve the bound to $\Pr(E) \leq \min\{1, \sum_i p_i\}$. For the rest of this problem, assume that the events $E_1, \ldots, E_k$ are independent.

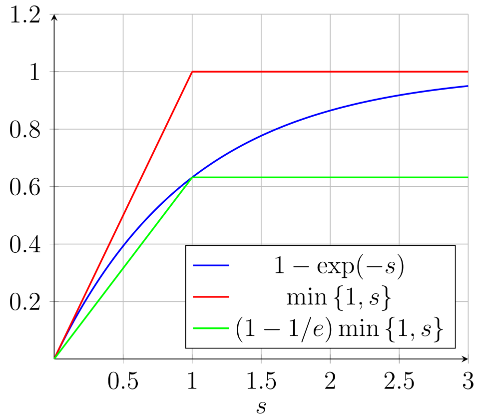
(a) With A^c denoting the complement of an event A, show that $\Pr(E^c) \leq \exp\left(-\sum_i p_i\right)$.
Hint: $1 - x \leq \exp(-x)$.

Solution: Simply following the hint, we have

$$\Pr(E^c) = \Pr\left(\bigcap_i E_i^c\right) = \prod_i \Pr(E_i^c) = \prod_i (1 - p_i) \leq \prod_i \exp(-p_i) = \exp\left(-\sum_i p_i\right).$$

(b) For $s \geq 0$, sketch the functions $1 - \exp(-s)$, and $\min\{1, s\}$. Show that $1 - \exp(-s) \geq (1 - 1/e)\min\{1, s\}$ for $s \geq 0$.
Hint: Consider the two cases (i) $s \in [0, 1]$, and (ii) $s > 1$.

Solution: The curves are sketched below.



(i) When $s \in [0, 1]$, observe that the curves $1 - \exp(-s)$ and $(1 - 1/e)\min\{1, s\}$ have the same endpoints, namely 0 and $1 - 1/e$. Since $1 - \exp(-s)$ is strictly concave in s, and $(1 - 1/e)\min\{1, s\} = (1 - 1/e)s$ is linear, we have $1 - \exp(-s) \geq (1 - 1/e)s$.

(ii) For $s > 1$, observe that $1 - \exp(-s)$ is strictly increasing in s, while $(1 - 1/e)\min\{1, s\} = (1 - 1/e)$ is constant, and they have the same value at $s = 1$. Hence, we have $1 - \exp(-s) \geq (1 - 1/e)$, and we are done.

(c) Combine your results in (a) and (b) to show that, when $E_1, \ldots, E_k$ are independent,

$$\left(1 - \frac{1}{e}\right)\min\{1, s\} \leq \Pr\left(\bigcup_i E_i\right) \leq \min\{1, s\},$$

with $s = \sum_i \Pr(E_i)$.

Solution: The union bound gives us $\Pr\left(\bigcup_i E_i\right) \leq s$, and we also trivially have $\Pr\left(\bigcup_i E_i\right) \leq 1$, which gives us the upper bound. For the lower bound, we have $\Pr\left(\bigcup_i E_i\right) = 1 - \Pr\left(\bigcap_i E_i^c\right) \geq 1 - \exp(-s) \geq (1 - 1/e)\min\{1, s\}$, from (a) and (b).

Moral of the story: For independent events, the trivially improved union bound, $\min\{1, \sum_i \Pr(E_i)\}$ is not only an upper bound to the probability of their union, but also a constant factor approximation to it.

SOLUTION 2. Suppose we design a communication system to send a k-bit message in the following way:

Step 1: We represent a message by a binary sequence $(b_1, \dots, b_k)$, each b_i in $\{0, 1\}$.

Step 2: Pick two vectors v_0 and v_1 in $\mathbb{R}^r$.

Step 3: The codeword for the message $(b_1, \dots, b_k)$ is then given by the vector $c = (v_{b_1}, \dots, v_{b_k})$ (in $\mathbb{R}^n$ with $n = kr$). For example, let $v_0 = (1, 2, 3)$ and $v_1 = (-1, -3, -2)$ in $\mathbb{R}^3$, then the codeword for the 3-bit message $(0, 0, 1)$ is $(\underbrace{1, 2, 3,}_{0} \underbrace{1, 2, 3,}_{0} \underbrace{-1, -3, -2}_{1})$.

Step 4: The vector c is transmitted and received as $Y = c + Z$ where Z is $\mathcal{N}(0, \sigma^2 I_n)$. Write $Y = (Y_1, \dots, Y_k)$ where each Y_i is in $\mathbb{R}^r$, similarly write $Z = (Z_1, \dots, Z_k)$ where each Z_i is in $\mathbb{R}^r$.

(a) Assuming all 2^k messages are equally likely, show that the procedure: "*for each* $i = 1, \dots, k$, *let* $\hat{b}_i = \arg\min_{b \in \{0,1\}} \|Y_i - v_b\|$ *and estimate the transmitted message as* $(\hat{b}_1, \dots, \hat{b}_k)$" minimizes the probability of error.

Solution: Let H be the random variable associated with the k-bit message, then this rule is exactly the MAP rule obtained by solving $\arg\max_{b^k \in \{0,1\}^k} f_{Y_1,\dots,Y_k|H}(y_1, \dots, y_k \mid b_1, \dots b_k)$.

(b) With $d^2 = \|v_0 - v_1\|^2$, what is $\Pr\left(\hat{b}_i \neq b_i\right)$ (i.e., the probability that the ith bit of the message is received incorrectly)? What is Pr(error) (i.e., the probability that some bit is received incorrectly)? How does Pr(error) compare with $\min\left\{1, kQ(\frac{d}{2\sigma})\right\}$?

Solution: The probability that the ith bit is received incorrectly is $\Pr\left(\hat{b}_i \neq b_i\right) = Q\left(\frac{d}{2\sigma}\right)$. Let E_i be the event that the ith bit is received incorrectly. Then, the probability of error is $\Pr(\text{error}) = \Pr\left(\bigcup_i E_i\right)$. By Problem 1, this is upper bounded by $\sum_i \Pr(E_i) = kQ(\frac{d}{2\sigma})$, and also trivially by 1, hence Pr(error) is upper bounded by $\min\left\{1, kQ(\frac{d}{2\sigma})\right\}$. Further, by Problem 1, we also have that it is lower bounded by $\left(1 - \frac{1}{e}\right)\min\left\{1, kQ(\frac{d}{2\sigma})\right\}$.

(c) With $d^2 = \|v_0 - v_1\|^2$, consider a new system where v_0 and v_1 are replaced by the scalars $d/2$ and $-d/2$. The codewords of the new system $\left(\pm\frac{d}{2}, \dots, \pm\frac{d}{2}\right)$ are now in $\mathbb{R}^k$ instead of $\mathbb{R}^{kr}$. What can you say about the average energy $\mathcal{E}$, average energy per bit $\mathcal{E}_b$, the bit error probabilities, and the message error probability of the new system in terms of the corresponding quantities of the original system?

Solution: Observe that the new system is only a special case of the original system with $r = 1$. Consider a codeword $c = (v_{b_1}, \dots, v_{b_k})$ from the original system, with energy $\|c\|^2 = \sum_{i=1}^k \|v_{b_i}\|^2$. Clearly, this quantity only depends on $\|v_0\|^2$, $\|v_1\|^2$, and the number of 0's and 1's in $b_1, \dots, b_k$, and not explicitly on r. The average energy is thus minimized by a set of codewords that is centered at the origin and minimizes $\|v_0\|^2 + \|v_1\|^2$, which is satisfied by the choice of the scalars $d/2$ and $-d/2$. Thus, the average energy of the new system is at most that of the original system. Since the number of bits is unchanged, the average energy per bit of the new system is also at most that of the original system. The message error probabililty only depends on the geometry of the codewords, which remains unchanged.

(d) Suppose we need to send k bits (e.g., $k = 100$) using a system as above, and we require the message error probability to be at most α (e.g., $\alpha = 10^{-2}$). Suppose a_1 and a_2 satisfy $Q(a_1) = \frac{\alpha}{k}$ and $(1 - 1/e)\, Q(a_2) = \frac{\alpha}{k}$. Show that if $d/(2\sigma) < a_2$, the error probability requirement cannot be met. What will happen if $d/(2\sigma) \geq a_1$?

Solution: If $a_2 > d/(2\sigma)$, then, from (c),

$$\begin{aligned}
\Pr(\text{error}) &\geq \left(1 - \frac{1}{e}\right) \min\left\{1, kQ\left(\frac{d}{2\sigma}\right)\right\} \\
&= \min\left\{\left(1 - \frac{1}{e}\right), \left(1 - \frac{1}{e}\right) kQ\left(\frac{d}{2\sigma}\right)\right\} \\
&> \min\left\{\left(1 - \frac{1}{e}\right), \left(1 - \frac{1}{e}\right) kQ\left(a_2\right)\right\} \\
&= \min\left\{\left(1 - \frac{1}{e}\right), \alpha\right\}.
\end{aligned}$$

Hence, for sufficiently small α (i.e., smaller than $1 - 1/e \approx 0.63$), we have $\Pr(\text{error}) > \alpha$, i.e., the error probability requirement cannot be met. If $a_1 \leq d/(2\sigma)$, from (c),

$$\begin{aligned}
\Pr(\text{error}) &\leq \min\left\{1, kQ\left(\frac{d}{2\sigma}\right)\right\} \\
&\leq \min\left\{1, kQ\left(a_1\right)\right\} \\
&= \min\left\{1, \alpha\right\} \leq \alpha,
\end{aligned}$$

i.e., the error probability requirement is met.

Moral: (1) If the message is sent 'bit by bit', as in the system described in the beginning of the problem, one may as well use the simpler system in (c). (2) In a system designed as above, the minimal possible value of $(d/2\sigma)^2$ lies between a_2^2 and a_1^2. (Note that $(d/2\sigma)^2$ equals $\mathcal{E}_b/\sigma^2$.)

Solution 3. Consider a communication system with $2n$ equally likely codewords $\pm\sqrt{\mathcal{E}}e_j$, $j = 1, \ldots, n$ where $e_1, \ldots, e_n$ are the unit coordinate vectors in $\mathbb{R}^n$. The receiver receives $Y = c + Z$ where c is one of these codewords and Z is $\mathcal{N}(0, \sigma^2 I_n)$. As the system is sending $k = \log_2(2n)$ bits, the choice $\mathcal{E} = \sigma^2 A \log_2(2n)$ results in an energy per bit $\mathcal{E}_b$ satisfying $\mathcal{E}_b/\sigma^2 = A$.

The MAP rule for this setup is given by the following: find the j for which $|Y_j|$ is largest, and decide that the codeword $\operatorname{sign}(Y_j)\sqrt{\mathcal{E}}e_j$ was transmitted.

Consider the following alternative decoding method. Pick a threshold $t = \alpha\sqrt{\mathcal{E}}$ with $0 \leq \alpha < 1$. If there is exactly one j for which $|Y_j| > t$, decide that the codeword $\operatorname{sign}(Y_j)\sqrt{\mathcal{E}}e_j$ was transmitted. If there is no j for which $|Y_j| > t$ or several j's for which $|Y_j| > t$, then the decoder declares an error. Note that the error probability of the MAP decoder is upper bounded by the error probability of this (suboptimal) decoder, so any upper bound on the error probability of this decoder also upper bounds the probability of error of the MAP rule.

(a) Show that the probability of error (either by declaring an error, or by deciding on a wrong codeword) of this decoder satisfies

$$\begin{aligned}\Pr(\text{error}) &\le Q\left((1-\alpha)\sqrt{\frac{\mathcal{E}}{\sigma^2}}\right) + 2(n-1)Q\left(\alpha\sqrt{\frac{\mathcal{E}}{\sigma^2}}\right)\\ &< Q\left((1-\alpha)\sqrt{\frac{\mathcal{E}}{\sigma^2}}\right) + 2^k Q\left(\alpha\sqrt{\frac{\mathcal{E}}{\sigma^2}}\right).\end{aligned}$$

Solution: Without loss of generality assume that $\sqrt{\mathcal{E}}e_1$ is the transmitted codeword. Then $Y_1 = \sqrt{\mathcal{E}} + Z_1$, and $Y_j = Z_j$ for $j = 2, \ldots, n$. Our decoder will make an error only if (i) $Y_1 \le t$ or (ii) $-Y_1 > t$ or (iii) $|Y_j| > t$ for some $j = 2, \ldots, n$. Note that the event (ii) already included in the event (i), so we only need to consider the union of (i) and (iii). Probability of (i) is

$$\Pr(\sqrt{\mathcal{E}} + Z_1 \le \alpha\sqrt{\mathcal{E}}) = \Pr(-Z_1 \ge (1-\alpha)\sqrt{\mathcal{E}}) = Q((1-\alpha)\sqrt{\mathcal{E}/\sigma^2}).$$

Probability of (iii) is

$$\Pr\Big(\cup_{j=2}^n \{|Z_j| > t\}\Big) \le \sum_{j=2}^n \Pr(|Z_j| > t) = (n-1)2Q(t/\sigma) = 2(n-1)Q(\alpha\sqrt{\mathcal{E}/\sigma^2}).$$

Putting these bounds (i) and (iii) together, we find the desired upper bound to the error probability.

(b) Recall that $\mathcal{E} = kA\sigma^2$. Show that the probability of error is further upper bounded by

$$\tfrac{1}{2}\exp\big(-\tfrac{1}{2}k(1-\alpha)^2 A\big) + \tfrac{1}{2}\exp\big(-\tfrac{1}{2}k\alpha^2 A + k\ln 2\big).$$

Also show that if $A > 2\ln 2$ there is an $0 < \alpha < 1$ for which the probability of error approaches zero as k gets large.

Hint: Use (a) and $Q(x) \le \frac{1}{2}\exp(-x^2/2)$.

Solution: Using (a), the upper bound on $Q(\cdot)$, and noting that $2^k = \exp(k\ln 2)$, the bound follows. Observe that the upper bound consists of the sum of two terms. The first of these approaches zero as k gets large for any $A > 0$ and $\alpha < 1$, so we only need to find an $\alpha < 1$ for which the second term, $\exp(-\delta k)$ (where $\delta = \alpha^2 A/2 - \ln 2$), also approaches zero as k gets large. To that end it suffices to find $\alpha < 1$ for which $\delta > 0$. Observe that $\delta > 0$ is equivalent to $\alpha > \sqrt{(2\ln 2)/A}$. If $A > 2\ln 2$, we have $\sqrt{(2\ln 2)/A} < 1$ and thus there are α's for which $\sqrt{(2\ln 2)/2} < \alpha < 1$.

Moral: If we are given an energy budget in the form energy/bit $= \mathcal{E}_b$, and if this budget satisfies $\mathcal{E}_b/\sigma^2 > 2\ln 2$, then we can, by taking k large enough, meet any desired error probability requirement.

(c) Suppose that $A > 2\ln 2$. Show that

$$\Pr(\text{error}) < \exp\left[-\frac{1}{8}\left(1 - \frac{2\ln 2}{A}\right)^2 Ak\right].$$

Hint: Use (b) and consider the choice $\alpha = \frac{1}{2}\left(1 + \frac{2\ln 2}{A}\right)$. Don't forget to verify that $\alpha < 1$.

Solution: As $A > 2\ln 2$, the choice of α in the hint satisfies $1/2 \le \alpha < \frac{1}{2}[1+1] = 1$, and is thus a valid choice for the bound in (b). Substituting this value of α in the bound given in (b) we obtain the bound above.

(d) For $A = 4, 6, 8, 10, 12$, what are the values $k(A)$ of k that will make the upper bound to the error probability in (c) less than 10^{-3}?

Solution: All we need to do is to solve for k in $-\frac{1}{8}\left(1 - \frac{2\ln 2}{A}\right)^2 Ak \leq \ln(10^{-3})$, giving $k \geq \frac{24\ln(10)}{A(1-(2\ln 2)/A)^2}$. For $A = 4, 6, 8, 10, 12$ we obtain (after rounding up to the next integer) $k = 33, 16, 11, 8, 6$.

(e) For each of the five values of A in (d), consider a bit-by-bit communication system (à la Problem 2 above) with $\mathcal{E}_b/\sigma^2 = A$ that sends a $k(A)$-bit message. Find the message error probabilities of these systems.

Solution: Recall that $\min\{1, kQ(\sqrt{A})\}$ is a good approximation (within a factor $1 - 1/e$) to the error probability of such systems. The exact value of the error probability is $1 - (1 - Q(\sqrt{A}))^k$. Both these values are given in the table below, rounded to three decimal digits.

A	$k(A)$	$\min\{1, kQ(\sqrt{A})\}$	$1 - (1 - Q(\sqrt{A}))^k$
4	33	0.751	0.532
6	16	0.114	0.109
8	11	0.026	0.025
10	8	0.006	0.006
12	6	0.002	0.002

Note that the error probability of these bit-by-bit methods is much higher than 10^{-3} which is guaranteed by the method described above. The difference is especially evident for low values of A.

SOLUTION 4. Suppose $c_1, \dots, c_m$ are codewords in $\mathbb{R}^n$ and all messages are equally likely. When codeword i is sent, the receiver receives $Y = (Y_1, Y_2)$ in $\mathbb{R}^{2n}$ with either

$$(1)\ Y_1 = c_i + Z,\ Y_2 = \tilde{Z}, \quad \text{or} \quad (2)\ Y_1 = \tilde{Z},\ Y_2 = c_i + Z,$$

with the two cases being equally probable. Here Z and $\tilde{Z}$ are independent, Z is $\mathcal{N}(0, \sigma^2 I_n)$, and $\tilde{Z}$ is $\mathcal{N}(0, \tau^2 I_n)$. If the receiver had "side information" telling it which of (1) and (2) occurred, then it could have decoded the message i based on the part of Y that equals $c_i + Z$. But the receiver does not have such information.
Let $H = (i, b)$ where the binary value b indicates which of (1) and (2) took place.

(a) Consider the following rule to decide the value of H from the observation (y_1, y_2). Find $i_1 = \arg\min \|y_1 - c_i\|$, let $i_2 = \arg\min \|y_2 - c_i\|$. Let $d_1 = \frac{\|y_1 - c_{i_1}\|^2}{\sigma^2} + \frac{\|y_2\|^2}{\tau^2}$ and $d_2 = \frac{\|y_2 - c_{i_2}\|^2}{\sigma^2} + \frac{\|y_1\|^2}{\tau^2}$. Decide

$$\hat{H} = \begin{cases} (i_1, 1) & \text{if } d_1 < d_2, \\ (i_2, 2) & \text{else.} \end{cases}$$

Does this rule minimize $\Pr\left(\hat{H} \neq H\right)$?

Solution: Yes, it is the map rule for H, i.e., it is exactly the rule obtained on computing $\arg\max_h f_{Y_1,Y_2|H}(y_1, y_2 \mid h)$, where $h = (i, b)$.

(b) Let $\hat{\imath}$ be the first component of $\hat{H}$, i.e., $\hat{\imath} = i_1$ if $d_1 < d_2$ and $\hat{\imath} = i_2$ else. Does this rule minimize $\Pr\left(\hat{\imath} \neq i\right)$?

Solution: No, this is not the MAP rule for decoding i. Let I be the random variable corresponding to the message i and B correspond to the "state" of the channel b, then the MAP rule for decoding I is given by

$$\begin{aligned}
&\arg\max_i f_{Y_1,Y_2|I}(y_1, y_2 \mid i) \\
&= \arg\max_i \frac{1}{2}\left[f_{Y_1,Y_2|I,B}(y_1, y_2 \mid i, (1)) + f_{Y_1,Y_2|I,B}(y_1, y_2 \mid i, (2))\right] \\
&= \arg\max_i \left[f_Z(y_1 - c_i) f_{\tilde{Z}}(y_2) + f_{\tilde{Z}}(y_1) f_Z(y_2 - c_i)\right].
\end{aligned}$$

Since Z and $\tilde{Z}$ are both normally distributed, this rule is the sum of two exponential quantities, which does not give the same rule as (a). Hence, taking the first component of $\hat{H}$ from (a) does not minimize $\Pr\left(\hat{\imath} \neq i\right)$.

Let $\hat{\imath}_o(y_1, y_2, b)$ be the MAP estimator of a receiver that somehow has access to the side information as mentioned above, i.e., it is the decision made from the observation (y_1, y_2, b).

(c) Let $\hat{b}$ be the second component of $\hat{H}$ as above, i.e., $\hat{b} = 1$ if $d_1 < d_2$ and $\hat{b} = 2$ else. Justify the following inequalities:

$$\begin{aligned}
\Pr\left(\hat{\imath}_o \neq i\right) &\overset{(c_0)}{\leq} \Pr\left(\hat{\imath} \neq i\right) \\
&\overset{(c_1)}{\leq} \Pr\left(\hat{H} \neq H\right) \\
&\overset{(c_2)}{=} \Pr\left(\hat{b} \neq b\right) + \Pr\left(\hat{b} = b \text{ and } \hat{\imath} \neq i\right) \\
&\overset{(c_3)}{=} \Pr\left(\hat{b} \neq b\right) + \Pr\left(\hat{b} = b \text{ and } \hat{\imath}_o \neq i\right) \\
&\overset{(c_4)}{\leq} \Pr\left(\hat{b} \neq b\right) + \Pr\left(\hat{\imath}_o \neq i\right).
\end{aligned}$$

Solution: (c_0) follows from (b), since $\hat{\imath}_o$, being the MAP estimator which also has access to the channel state b, cannot do any worse than the rule in (a).
(c_1) follows from the fact that if $\hat{\imath} \neq i$, then it must be true that $\hat{H} \neq H$, since $H = (i, b)$.
(c_2) follows by considering the event $A = \{\hat{b} \neq b\}$. Clearly, $\Pr\left(\hat{H} \neq H\right) = \Pr\left(\hat{H} \neq H \text{ and } \hat{b} \neq b\right) + \Pr\left(\hat{H} \neq H \text{ and } \hat{b} = b\right)$. The first term is at most $\Pr\left(\hat{b} \neq b\right)$ and the second term is $\Pr\left(\hat{b} = b \text{ and } \hat{\imath} \neq i\right)$, since if $\hat{b} = b$, we have $\hat{H} \neq H$ if and only if $\hat{\imath} \neq i$.

Moral: The message error probabilities of the receiver with and without side information differ at most by $\Pr\left(\hat{b} \neq b\right)$. If $\Pr\left(\hat{b} \neq b\right)$ is small, then not much is lost by not having the side information about the channel state.

(d) Suppose that $H = (i, 1)$. Show that $\hat{b} \neq 1$ only if there exists $i' \in \{1, \ldots, m\}$ with

$$\frac{\|Z\|^2}{\sigma^2} + \frac{\|\tilde{Z}\|^2}{\tau^2} > \frac{\|c_i + Z\|^2}{\tau^2} + \frac{\|\tilde{Z} - c_{i'}\|^2}{\sigma^2}.$$

Hint: How does the left-hand side compare to d_1?

Solution: Since $b = 1$, we have that $Y_1 = c_i + Z$ and $Y_2 = \tilde{Z}$. We have $\hat{b} \neq 1$ exactly when $d_1 \geq d_2$. Suppose we have $d_1 \geq d_2$. Then,

$$\begin{aligned}
\frac{\|Z\|^2}{\sigma^2} + \frac{\|\tilde{Z}\|^2}{\tau^2} &= \frac{\|Y_1 - c_i\|^2}{\sigma^2} + \frac{\|Y_2\|^2}{\tau^2} \\
&\geq \arg\min_i \frac{\|Y_1 - c_i\|^2}{\sigma^2} + \frac{\|Y_2\|^2}{\tau^2} \\
&= \frac{\|Y_1 - c_{i_1}\|^2}{\sigma^2} + \frac{\|Y_2\|^2}{\tau^2} = d_1 \\
&\geq d_2 = \arg\min_i \frac{\|Y_2 - c_i\|^2}{\sigma^2} + \frac{\|Y_1\|^2}{\tau^2} \\
&= \arg\min_i \frac{\|\tilde{Z} - c_i\|^2}{\sigma^2} + \frac{\|c_i + Z\|^2}{\tau^2},
\end{aligned}$$

i.e., there must exist some $i' \in \{1, \ldots, m\}$ such that $\frac{\|Z\|^2}{\sigma^2} + \frac{\|\tilde{Z}\|^2}{\tau^2} > \frac{\|c_i + Z\|^2}{\tau^2} + \frac{\|\tilde{Z} - c_{i'}\|^2}{\sigma^2}$.

(e) From now on, suppose $\sigma = \tau$. Use the union bound to upper bound $\Pr\left(\hat{b} \neq 1 \mid H = (i, 1)\right)$ by $\sum_{i'=1}^{m} Q\left(\sqrt{\frac{\|c_i\|^2 + \|c_{i'}\|^2}{4\sigma^2}}\right)$.

Solution: From part (d), taking $\sigma = \tau$, given $H = (i, 1)$, we have $\hat{b} \neq 1$ only if there exists an $i' \in \{1, \ldots, m\}$ such that $\|Z\|^2 + \|\tilde{Z}\|^2 > \|c_i + Z\|^2 + \|\tilde{Z} - c_{i'}\|^2$. Hence, by the union bound, we have

$$\begin{aligned}
\Pr\left(\hat{b} \neq 1 \mid H = (i, 1)\right) &= \Pr\left(\bigcup_{i'=1}^{m} \|Z\|^2 + \|\tilde{Z}\|^2 > \|c_i + Z\|^2 + \|\tilde{Z} - c_{i'}\|^2\right) \\
&\leq \sum_{i'=1}^{m} \Pr\left(\|Z\|^2 + \|\tilde{Z}\|^2 > \|c_i + Z\|^2 + \|\tilde{Z} - c_{i'}\|^2\right).
\end{aligned}$$

To obtain the desired bound, it suffices to show that

$$\Pr\left(\|Z\|^2 + \|\tilde{Z}\|^2 > \|c_i + Z\|^2 + \|\tilde{Z} - c_{i'}\|^2\right) \leq Q\left(\sqrt{\frac{\|c_i\|^2 + \|c_{i'}\|^2}{4\sigma^2}}\right).$$

This follows immediately by simplifying the event on the left-hand side, as

$$\begin{aligned}
& \|Z\|^2 + \|\tilde{Z}\|^2 > \|c_i + Z\|^2 + \|\tilde{Z} - c_{i'}\|^2 \\
\iff & \|Z\|^2 + \|\tilde{Z}\|^2 > \|c_i\|^2 + \|Z\|^2 + 2\langle c_i, Z\rangle + \|\tilde{Z}\|^2 + \|c_{i'}\|^2 - 2\langle c_{i'}, \tilde{Z}\rangle \\
\iff & 2\langle c_{i'}, \tilde{Z}\rangle - 2\langle c_i, Z\rangle > \|c_i\|^2 + \|c_{i'}\|^2.
\end{aligned}$$

Observing that $2\langle c_{i'}, \tilde{Z}\rangle - 2\langle c_i, Z\rangle$ is a normal random variable with mean 0 and variance $4\sigma^2(\|c_i\|^2 + \|c_{i'}\|^2)$, we are done.

(f) Assume that $\|c_i\| = \sqrt{\mathcal{E}}$ for all $i \in \{1, \ldots, m\}$ and $\frac{\mathcal{E}_b}{\sigma^2} > 4 \ln 2$ where $\mathcal{E}_b$ is the energy per bit. Use (c) and (e) to show that $\Pr\left(\hat{\imath} \neq i\right) - \Pr\left(\hat{\imath}_{\mathrm{o}} \neq i\right)$ approaches 0 as m grows.

Hint: What happens to $\Pr\left(\hat{b} \neq b\right)$ as m grows?

Solution: By (c), we have that $\Pr\left(\hat{\imath} \neq i\right) - \Pr\left(\hat{\imath}_{\mathrm{o}} \neq i\right)$ is at most $\Pr\left(\hat{b} \neq b\right)$, which, by symmetry, is equal to $\Pr\left(\hat{b} \neq 1 \mid H = (i,1)\right)$ Hence, by (e), it is enough to show that $\sum_{i'=1}^{m} Q\left(\sqrt{\frac{\|c_i\|^2+\|c_{i'}\|^2}{4\sigma^2}}\right)$ goes to 0 as m goes to infinity. This follows immediately, using the fact that $\|c_i\| = \sqrt{\mathcal{E}} = \sqrt{\mathcal{E}_b \log_2 m}$ for all i, as

$$\begin{aligned}\sum_{i'=1}^{m} Q\left(\sqrt{\frac{\|c_i\|^2+\|c_{i'}\|^2}{4\sigma^2}}\right) &= mQ\left(\sqrt{\frac{\mathcal{E}_b \log_2 m}{2\sigma^2}}\right) \\ &\leq \frac{1}{2}\exp\left(\ln m - \frac{\mathcal{E}_b \log_2 m}{4\sigma^2}\right) \\ &= \frac{1}{2}\exp\left[\log_2 m\left(\ln 2 - \frac{\mathcal{E}_b}{4\sigma^2}\right)\right].\end{aligned}$$

Takeaways from theory for the implementation.

1. The goal of Problem 1 is to show that the union bound, while it is an upper bound, is in fact a good approximation, since it is only a constant factor of $1 - 1/e$ away from the actual error probability. This gives us a mathematical tool to effectively analyze the systems that we design.

2. The goal of Problems 2 and 3 is to provide two possible coding schemes for converting bits into codewords for transmission over the channel. Problem 2 gives a bit-by-bit approach, and Problem 3 describes a *biorthgonal code.* Our analysis shows that using a bit-by-bit communication system may not be the best idea when the energy is required to be small, as seen in Problem 3(e) — we need significantly more energy to obtain the same error probability when doing so as compared to a more clever design such as the biorthogonal code.

3. Note that so far, we assumed that our channel was an AWGN channel. The channel over which the communication is to take place in the implementation is not simple AWGN, but it can be thought of as two parallel AWGN channels, where only one of them sees the actual input; the receiver receives both outputs and does not know which of the parallel channels actually had the correct input. Problem 4 deals with this channel specifically. The MAP rule for decoding over this channel involves computing sums of exponentials of various terms, and cannot be obtained as an extension of a decoding scheme for the simple AWGN channel. The scheme given in Problem 4(a) describes how the decoding can be done with minimal additional effort over a simple AWGN channel (compute the MAP estimates for each of the parallel channels, compute d_1 and d_2, which are simply sums of distances, and compare them). Though this is not the optimal rule which minimizes the message error probability, it is considerably simpler than the MAP rule to decode the message, and has an error probability that is asympotically equal to the optimal rule which also has access to the side information. Hence, implementing this biorthogonal code with the decoding rule mentioned in Problem 4(a) is expected to give a working implementation (that also meets the channel constraints on the energy and number of samples transmitted, but this is to be checked).