SOLUTION 1. Consider the sequence $M_0, M_1, \ldots$ of matrices constructed recursively as follows:

$$M_0 = [+1] \quad \text{and} \quad M_{r+1} = \begin{bmatrix} +M_r & +M_r \\ +M_r & -M_r \end{bmatrix} \text{ for } r = 0, 1, \ldots$$

For example, we have,

$$M_1 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}, \quad \ldots$$

Note that M_r is a $2^r \times 2^r$ matrix with elements taking values in $\{+1, -1\}$. Each row (and each column) has squared Euclidean norm 2^r.
Now fix $r > 0$ and let $n = 2^r$. Let x be a row of M_r and write $x = [x'\, x'']$ where x' is the left half of x and x'' is the right half (for example, with $r = 2$ and $x = [+1, +1, -1, -1]$ being the 3rd row of M_2, we have $x' = [+1, +1]$ and $x'' = [-1, -1]$). As $\tilde{x} = [\tilde{x}'\, \tilde{x}'']$ steps through all the n rows of M_r, compute the inner products $p' = \langle x', \tilde{x}' \rangle$ and $p'' = \langle x'', \tilde{x}'' \rangle$. (For the example above (p', p'') takes the values $(2, -2)$, $(0, 0)$, $(2, 2)$ and $(0, 0)$ as $\tilde{x}$ steps through $[+1, +1, +1, +1]$, $[+1, -1, +1, -1]$, $[+1, +1, -1, -1]$ and $[+1, -1, -1, +1]$.)

(a) Show that for any row x of M_r, (p', p'') equals $(n/2, n/2)$ once, the value $(n/2, -n/2)$ once, and the value $(0, 0)$ for the remaining $n - 2$ times. Use this to conclude that the rows of M_r are orthogonal to each other.
Hint: One way is by induction on r.

Solution: Direct check by induction. Base case $r = 1$ is trivial. Assume the statement is true for $r \geq 1$. Then check for $r+1$, using the fact that the rows of M_{r+1} are either $[x, x]$ or $[x, -x]$ for some row x of M_r.

Let $B_r = \begin{bmatrix} +M_r \\ -M_r \end{bmatrix}$. The $m = 2n$ rows of B_r consists of the rows of M_r and their negatives. Let $x = [x'\, x'']$ be a row of B_r. As $\tilde{x} = [\tilde{x}'\, \tilde{x}'']$ steps through all the m rows of B_r compute $p' = \langle x', \tilde{x}' \rangle$ and $p'' = \langle x'', \tilde{x}'' \rangle$.

(b) Show that for any row x of B_r, (p', p'') takes the value $(n/2, n/2)$ once, the value $(-n/2, -n/2)$ once, the value $(n/2, -n/2)$ once, the value $(-n/2, n/2)$ once, and the value $(0, 0)$ the remaining $m - 4$ times.
Hint: Use (a).

Solution: Follows almost immediately (a). As $\tilde{x}$ sweeps over $2n$ rows of B_r, it first sweeps over n rows of M_r and then n rows of $-M_r$. Hence it takes all values as in (a) once, and also the negatives of those values once.

(c) Conclude that for any row x of B_r, $p' + p''$ takes the value n once, the value $-n$ once, and the value $(0, 0)$ the remaining $m - 2$ times.
Hint: Use (b).

Solution: Immediate given (b).

SOLUTION 2. Suppose we have a channel whose input $x = [x'\, x'']$ is a real vector of even dimension n (so that x' and x'' are of dimension $n/2$). The channel has two behaviors determined by an internal state $s \in \{1, 2\}$. The output $Y = [Y'\, Y'']$ is given by

$$(Y',\, Y'') = \begin{cases} \left(\sqrt{g}x' + Z',\, x'' + Z''\right) & \text{if } s = 1, \\ \left(x' + Z',\, \sqrt{g}x'' + Z''\right) & \text{if } s = 2, \end{cases}$$

where $g \geq 1$ is a non-negative constant and $Z = [Z'\, Z'']$ has i.i.d. $\mathcal{N}(0, \sigma^2)$ components. In other words, the channel subjects either the first half (if $s = 1$) or the second half (if $s = 2$) of the input vector to an energy gain g, and adds Gaussian noise.
Suppose $c_1, \ldots, c_m$ are the codewords for m equally likely messages for transmission over the channel above. Thus, each c_i is a real vector of (even) dimension n. Write $c_i = [c_i'\, c_i'']$ so that c_i' and c_i'' are real vectors of dimension $n/2$.

(a) Suppose the value of s is known to the receiver. What is the decision rule that minimizes the probability of error?

Solution: The MAP rule with knowledge of s is

$$\hat{H}_{\text{MAP}}(Y) = \arg\min_i \begin{cases} \|Y' - \sqrt{g}c_i'\|^2 + \|Y'' - c_i''\|^2 & \text{if } s = 1, \\ \|Y' - c_i'\|^2 + \|Y'' - \sqrt{g}c_i''\|^2 & \text{if } s = 2. \end{cases}$$

(b) Suppose $\|c_1'\| = \cdots = \|c_m'\|$ and $\|c_1''\| = \cdots = \|c_m''\|$. Let

$$\text{score}(i, Y, s) = \begin{cases} \sqrt{g}\langle Y', c_i'\rangle + \langle Y'', c_i''\rangle & s = 1, \\ \langle Y', c_i'\rangle + \sqrt{g}\langle Y'', c_i''\rangle & s = 2. \end{cases}$$

Show that

$$\hat{\imath} = \arg\max_i \text{score}(i, Y, s)$$

is an optimum decision rule for a receiver that observes $Y = [Y'\, Y'']$ and is aware of the value of s.

Solution: It follows by expanding the norm that $\hat{H}_{\text{MAP}}(Y) = \hat{\imath}$ when $\|c_1'\| = \cdots = \|c_m'\|$ and $\|c_1''\| = \cdots = \|c_m''\|$.

Now suppose that the receiver is *not* aware of the value of s. Still supposing that the codewords c_i are as in (b) (i.e., all have equal norms of their first halves, and equal norms of their second halves), consider the following way to assign a score to each message i based on the observation $Y = [Y'\, Y'']$:

$$\text{score}(i, Y) = \max_{s \in \{1,2\}} \text{score}(i, Y, s)$$

and the following two decoding rules. The first rule chooses the message with the highest score, i.e.,

$$\hat{\imath}_1 = \arg\max_i \text{score}(i, Y).$$

The second is based on a threshold t, and chooses the message $\hat{\imath}_2$ if $\hat{\imath}_2$ is the only i for which $\text{score}(i, Y) > t$. If there is no such i, or two or more such i's, it sets $\hat{\imath}_2 = 0$ — note that when $\hat{\imath}_2 = 0$ this decoding rule has certainly made an error.

(c) Argue that when $\hat{\imath}_2 \neq 0$, we have $\hat{\imath}_1 = \hat{\imath}_2$, and thus the probability of error of the second decoding rule is an upper bound to the probability of error of the first decoding rule.

Solution: If $\hat{\imath}_2 = j \neq 0$, then $\text{score}(j, Y) > t$ and $\text{score}(i, Y) \leq t$ for all $i \neq j$. Clearly, we then have that $\text{score}(i, Y)$ is maximized by taking $i = j$, and hence, $\hat{\imath}_1 = j$. Hence, if $\hat{\imath}_2$ is equal to the transmitted message, so is $\hat{\imath}_1$ necessarily, implying that the probability that the second decoding rule is correct is at most that of the first rule, and we are done.

SOLUTION 3. Let $r \geq 0$. Let the codewords for m messages, $c_1, \ldots, c_m$ be the rows of the matrix $\sqrt{\alpha}[B_r\, B_r]$, where B_r is as in Problem 1 above and $\alpha \geq 0$ is chosen to make the energy per bit to equal $\mathcal{E}_b$.

(a) With n denoting the dimension of the c_i's, express n, m and α in terms of r and $\mathcal{E}_b$.

Solution: $n = m = 2^{r+1}$ by noting that M_r is a $2^r \times 2^r$ matrix. $\alpha = (r+1)\mathcal{E}_b/2^{r+1}$ by noting that the energy of any c_i is $n = 2^{r+1}$ and each c_i conveys $\log_2 m = r + 1$ bits.

Suppose the codewords above are used to communicate over the channel in Problem 2. Let s denote the channel state and $\bar{s}$ denote the 'other' state (i.e., $\bar{s} = 3 - s$).

(b) Fix a threshold t. Show that, conditional on i being the transmitted message, the probability of error of the second decoding rule that uses the threshold t is upper bounded by

$$\Pr\big(\text{score}(i, Y, s) \leq t\big) + \sum_{i' \neq i} \sum_{s' \in \{1,2\}} \Pr\big(\text{score}(i', Y, s') > t\big).$$

Solution: Let i be the transmitted message. There are only two ways that the second decoding rule can make an error: (1) $\text{score}(i', Y) > t$ for some $i' \neq i$, or (2) $\text{score}(i, Y) \leq t$. By the union bound, the error probability of the second decoding rule is upper bounded by the sums of the probabilities of these events. As $\text{score}(i', Y) = \max_{s' \in \{1,2\}} \text{score}(i', Y, s')$, we have $\text{score}(i', Y) > t$ only if $\text{score}(i', Y, s') > t$ for some s'. Using another union bound to upper bound this probability gives us the required terms.

(c) Let $\beta = \frac{1}{2}\alpha(g+1)n$. Show that, conditional on i being the transmitted message, with s being the channel state,

$$\text{score}(i, Y, s) \sim \mathcal{N}(\beta, \sigma^2\beta).$$

Solution: Suppose $s = 1$ and i is the transmitted message. Then $\text{score}(i, Y, 1) = \sqrt{g}\langle Y', c_i'\rangle + \langle Y'', c_i''\rangle$ with $Y' = \sqrt{g}c_i' + Z'$ and $Y'' = c_i'' + Z''$. A direct computation gives $\text{score}(i, Y, 1) = g\langle c_i', c_i'\rangle + \langle c_i'', c_i''\rangle + \sqrt{g}\langle c_i', Z'\rangle + \langle c_i'', Z''\rangle$. Since Z' and Z'' are zero mean Gaussian vectors, $\text{score}(i, Y, 1)$ is Gaussian and has mean $g\langle c_i', c_i'\rangle + \langle c_i'', c_i''\rangle = \alpha\frac{n}{2}(g+1) = \beta$. The variance of $\text{score}(i, Y, 1)$ is $(g\|c_i'\|^2 + \|c_i''\|^2)\sigma^2 = \beta\sigma^2$. An identical computation results for $s = 2$ as well.

(d) Conditional on i being the transmitted message, with s being the channel state, show that

(i) there is one value of $i' \neq i$ for which

$$\text{score}(i', Y, s) \sim \mathcal{N}(-\beta, \sigma^2\beta) \quad \text{and} \quad \text{score}(i', Y, \bar{s}) \sim \mathcal{N}(-\alpha\sqrt{g}n, \sigma^2\beta);$$

(ii) for the remaining $m-2$ values of $i' \neq i$,

$$\text{score}(i', Y, s) \sim \mathcal{N}(0, \sigma^2\beta) \quad \text{and} \quad \text{score}(i', Y, \bar{s}) \sim \mathcal{N}(0, \sigma^2\beta).$$

Hint: Use 1(c).

Solution: We can compute the mean and variance of each score by repeating simikar computations as in part (c) and using 1(c).

(e) For a threshold $t \geq 0$, show that the error probability of the second decoding rule is upper bounded by

$$Q\left(\frac{\beta - t}{\sqrt{\sigma^2\beta}}\right) + (2m-2)Q\left(\frac{t}{\sqrt{\sigma^2\beta}}\right).$$

Hint: Use (b), (c) and (d) and the fact that $Q(\cdot)$ is a decreasing function.

Solution: We use the bound in part (b). The first term is $\Pr\left(\text{score}(i, Y, s) \leq t\right) = Q\Big(\frac{\beta - t}{\sqrt{\sigma^2\beta}}\Big)$, as $\text{score}(i, Y, s) \sim \mathcal{N}(\beta, \sigma^2\beta)$. The second term is the sum over $i' \neq i$ of terms $f(i') \coloneqq \sum_{s'=s,\bar{s}} \Pr\left(\text{score}(i', Y, s') > t\right)$. There is one value of $i' \neq i$ for which $\text{score}(i', Y, s) \sim \mathcal{N}(-\beta, \sigma^2\beta)$ and $\text{score}(i', Y, \bar{s}) \sim \mathcal{N}(-\alpha\sqrt{g}n, \sigma^2\beta)$. For this i', the term $f(i') = Q\Big(\frac{\beta+t}{\sqrt{\sigma^2\beta}}\Big) + Q\Big(\frac{\alpha\sqrt{g}n+t}{\sqrt{\sigma^2\beta}}\Big)$, which is upper bounded by $2Q\Big(\frac{t}{\sqrt{\sigma^2\beta}}\Big)$ as $Q(\cdot)$ is decreasing. For the remaining $m-2$ values of $i' \neq i$, we have $f(i') = 2Q\Big(\frac{t}{\sqrt{\sigma^2\beta}}\Big)$. Hence, the bound in part (b) is further upper bounded by the given expression.

(f) Fix $0 < \epsilon < 1$ and set $t = (1-\epsilon)\beta$. Show that the error probability of the second decoding rule is upper bounded by

$$Q\left(\epsilon\sqrt{\beta/\sigma^2}\right) + (2m-2)Q\left((1-\epsilon)\sqrt{\beta/\sigma^2}\right).$$

Solution: Follows immediately by substituting $t = (1-\epsilon)\beta$ in the bound of part (e).

(g) Show that as r grows, the first term above approaches zero, and if $(1+g)(1-\epsilon)^2\mathcal{E}_b > \sigma^2 4\ln 2$, the second term approaches zero too.
Hint: For the last claim, use the fact that for $x \geq 0$, $Q(x) \leq \frac{1}{2}\exp(-x^2/2)$.

Solution: The first term is $Q\left(\epsilon\sqrt{\beta/\sigma^2}\right) \leq \frac{1}{2}\exp\left(-\frac{\epsilon\beta}{\sigma^2}\right) = \frac{1}{2}\exp\left(-\frac{\epsilon\alpha(g+1)2^{r+1}}{\sigma^2}\right)$. We are done if we show that the upper bound goes to zero, as it is also lower bounded by zero (the argument of Q is positive). As $r \to \infty$, the argument of exp goes to $-\infty$, and hence, the first term goes to zero. The second term can be upper bounded as follows:

$$\begin{aligned}
&(2m-2)Q\left((1-\epsilon)\sqrt{\beta/\sigma^2}\right) \\
&\leq 2^{r+1}\exp\left(-\frac{(1-\epsilon)^2\beta}{2\sigma^2}\right) = 2^{r+1}\exp\left(-\frac{(1-\epsilon)^2(r+1)\mathcal{E}_b(g+1)}{4\sigma^2}\right) \\
&= \exp\left[-(r+1)\left(\frac{(1-\epsilon)^2\mathcal{E}_b(g+1)}{4\sigma^2} - \ln 2\right)\right], \qquad (*)
\end{aligned}$$

which goes to zero if $(1+g)(1-\epsilon)^2\mathcal{E}_b > \sigma^2 4\ln 2$.

(h) Conclude that if $\mathcal{E}_b/\sigma^2 > (4\ln 2)/(1+g)$, the error probability of the first decoding rule approaches zero as r grows.

Solution: If $\mathcal{E}_b/\sigma^2 > (4\ln 2)/(1+g)$, there exists an $\epsilon > 0$ sufficiently close to zero such that $(1+g)(1-\epsilon)^2\mathcal{E}_b > \sigma^2 4\ln 2$. Hence, by part (g), the error probability of the second decoding rule goes to zero as r grows. But by Problem 2(c), this error probability is an upper bound to that of the first decoding rule, while, consequently, must also go to zero as r grows.

Takeaways from theory for the implementation.

1. The goal of Problem 1 is to obtain a characterization of the possible inner products of the rows the matrix M_r (called a Hadamard matrix). The reason for doing so is that we require these inner products to compute the error probability of a code that uses these rows as codewords, as we do in Problem 3.

2. Problem 2 introduces two decoding rules $\hat{\imath}_1$ and $\hat{\imath}_2$. It is easy to see that the second is suboptimal (Problem 2(c)). The motivation for the second decoding rule is that it is easy to analyze its performance, unlike the first decoding rule. However, as the second decoding rule is necessarily worse (in terms of error probability) than the first, we get an upper bound to the error probability of the first rule for free.

3. We see that using such a "repeated biorthogonal" (each codeword is obtained by repeating a row of the Hadamard matrix or its negative) code allows us to have an error probability that decays exponentially with the number of bits, if $\mathcal{E}_b$ is at least $4\sigma^2\ln 2/(1+g)$. In the practical part of the project (refer to the project description file):

   (a) The channel is *nearly* the same, in the sense that half of the samples are boosted by a power gain g. However, the samples experiencing this gain are the samples at either odd indices or at even indices. In the theory part above, the samples experiencing the gain were either the first half or the second half of the codeword. Hence, to get the same performance as suggested by theory, we must use as codeword an "interleaved" version of the rows of $\alpha[B_r\, B_r]$, i.e., if $c_i = [c_i'\, c_i']$ is a row of the matrix (with c_i' having dimension $n/2$), the codeword sent should be $[c_{i,1}'\, c_{i,1}'\, c_{i,2}'\, c_{i,2}' \ldots c_{i,n/2}'\, c_{i,n/2}']$. The decoding should also be modified appropriately — Y' should be the odd indices and Y'' should be the even indices of the received codeword (instead of the first half and second half).

   (b) We have $g = 10, \sigma^2 = 10$, so as long as $\mathcal{E}_b > 2.52$, we can make the error probability as small as desired by choosing a sufficiently large r.

   (c) We can send at most 10^6 samples to send 240 bits, hence if we use the code in Problem 3, we must send $240/(r+1)$ symbols, each of codelength 2^{r+1}. Hence, we require r to satisfy $240 \cdot 2^{r+1}/(r+1) \leq 10^6$, which gives $r \leq 15$.

   (d) We have an energy constraint of 2,000 to send 240 bits, so we are allowed to use $\mathcal{E}_b \leq \frac{2000}{240} \approx 8.3$. This is sufficiently larger than 2.52, so we expect to get a sufficiently small error probability even at $r = 15$. Computing the expression (*) in Problem 3(f) for $r = 15$ and $\mathcal{E}_b = 8.3$, and optimizing over the choice of ϵ, we get that the error probability of the second decoding rule is upper

bounded by 0.017. Hence, to send 240 bits, the error probability of the second rule is upper bounded by $0.017 \cdot \frac{240}{r+1} \approx 0.26$. The probability of getting an error in both attempts is thus at most 0.07. A tighter bound can be obtained by using the fact that $Q(x) \leq \min\{\frac{1}{2}\exp(-x^2/2), \frac{1}{\sqrt{2\pi}x}\exp(-x^2/2)\}$ which gives us that the error probability of the second rule is at most 0.0039, implying that the probability of getting both attempts wrong while sending 240 bits is at most $\approx \left(\frac{240}{r+1} \cdot 0.039\right)^2 = 3.4 \times 10^{-3}$, or a total success probability of at least 99.6%. This gives us enough confidence that the code in Problem 3 should work for the channel in question (up to a rearrangement, as in (a) above), at least using $r = 15$ and an energy of 2000.

(e) It turns out this bound is still rather loose and we can do much better in practice, see Figure 1. Using $r = 15$ and an energy of 2000 to send 240 bits, we have an error with probability $1/2000 = 5 \times 10^{-4}$, hence the actual error rate in the demo (two attempts) is $(5 \times 10^{-4})^2 = 2.5 \times 10^{-7}$ (compare against the theoretical bound of 3.4×10^{-3}). Hence, we can use a much lower energy in practice: Suppose our target success rate in the demo is 99.9%, i.e., an error probability at most 0.001 to send 240 bits using two attempts. Then, the required error probability for our code should be $\sqrt{0.001} \approx 3.2 \times 10^{-2}$, which we can achieve using an energy under 1450 with $r = 15$ or even 1700 with $r = 11$ (which is mapping two characters per codeword).

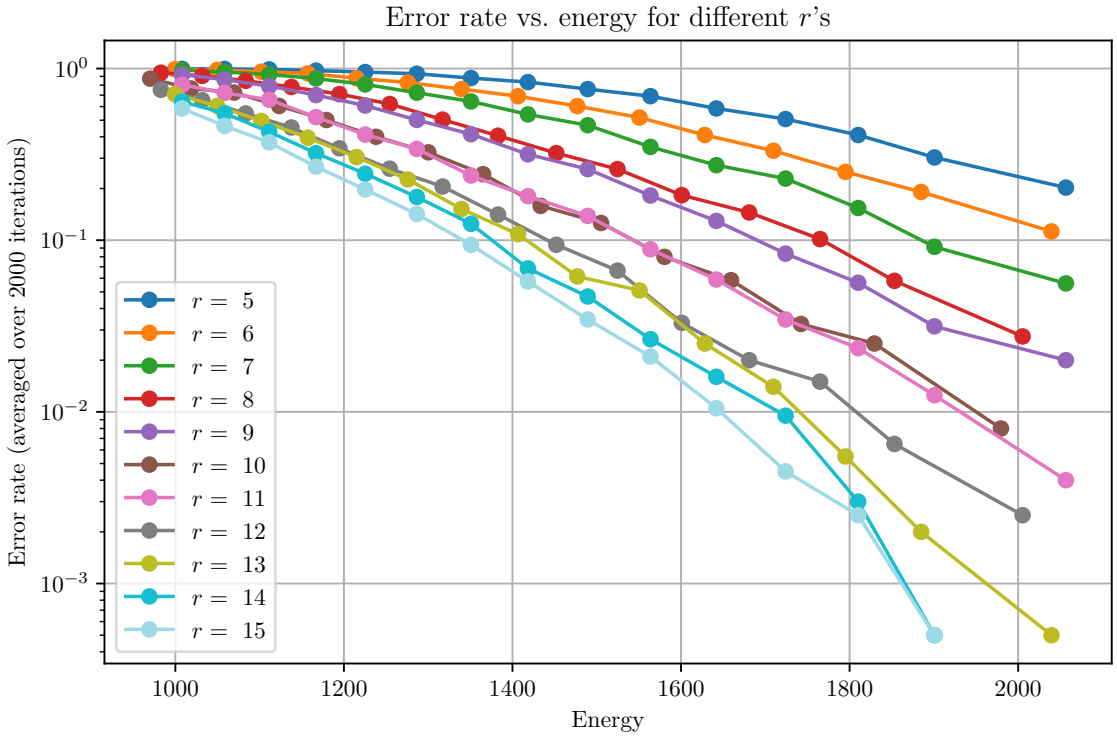

Figure 1: Empirical error probability over 2000 iterations versus average signal energy for different values of r. We consider the frame error, i.e., it is considered to be an error if at least one of the 40 characters is decoded incorrectly. At $r = 14, 15$, we get zero errors out of 2000 iterations for the highest value of energy, so the average error probability is likely to be smaller than 1/2000. To get r's that do not divide the 240 bits exactly, simply pad the original 240 bit message with some additional bits so that it is a multiple of $r + 1$.