This lecture wraps up our discussion of polar codes. Recall from the previous lecture the polar encoding function $E_n$ as defined on strings of length $n = 2^t$. We first recursively define a pre-encoding function $\tilde{E}_n$. Given strings $U$ and $V$ each of length $n/2$, we take

$$\tilde{E}_n(UV) = \tilde{E}_{n/2}(U \oplus V)\tilde{E}_{n/2}(V)$$

and $\tilde{E}_1 = \text{id}_{[2]}$. We visualize the pre-encoding step as a recursion tree of depth $t$, where at each step, we replace $U$ with $U \oplus V$ and $V$ with $V|U \oplus V$. Note that

$$H(U) + H(V) = H(U,V) = H(U \oplus V, V) = H(U \oplus V) + H(V|U \oplus V).$$

Given a string $Z \sim \text{Bernoulli}(p)^n$ of i.i.d. bits, let $W = \tilde{E}_n(Z)$. Because at each step of the recursion $U \oplus V$ has a lower index than $V|U \oplus V$, it is consistent to consider each bit $W_j$ as conditioned upon all previous bits $W_{<j}$. Under this conditioning, each bit in $\tilde{E}_n(Z)$ has either small or large entropy with high probability. More precisely, for every $p \in [0,1]$ and $c > 0$, there exists a $\beta < 1$ such that for all $t \in \mathbb{N}$,

$$\mathbb{P}_{j \in [2^t]}[H(W_j|W_{<j}) \in (c^{-t}, 1 - c^{-t})] \leq O(\beta^t).$$

We call this condition (strong) polarization.

Let $S = \{j \in [2^t] \mid H(W_j|W_{<j}) \geq c^{-t}\}$ be the set of high-entropy bits. The encoding function removes the low-entropy bits from $W$ and outputs $E_n(Z) = W_S$. We exhibit $W_S$ as a member of $\{0,1,\}^n$. Nonetheless, the encoder can transmit the bits in $S$ without having to convey the set $S$ because it only depends on the source Bernoulli$(p)$, of which the encoder and decoder both have knowledge. Metaphorically, $S$ resides in the hardware of the coding scheme.

**Exercise 1.** In terms of the length $n$ of the input string, what is the complexity of encoding? Draw a circuit that computes $E_n$ for the case where $n = 8$ and $p = 1/4$.

In Section 1, we introduce the polar decoding function and show that this function satisfies the relevant desiderata for a coding scheme. In Section 2, we begin a discussion on martingales and sketch a proof of polarization.

### 1 The Polar Decoding Function

Let $D_n$ denote the polar decoding function as defined on strings of length $n = 2^t$. Given $W_S \in \{0,1,\}^n$, we want $D_n(W_S)$ to be the $\hat{Z} \in [2]^n$ that maximizes the probability that $\tilde{E}_n(Z)_S = W_S$ given $Z = \hat{Z}$. We wish to define $D_n$ recursively, but we face a minor obstruction.

Let $Z = UV$ be the concatenation of strings $U$ and $V$ at the first recursive step. The naïve recursion reconstructs $\hat{Z}$ based on the presumption that $Z$ is a string of i.i.d. Bernoulli bits. Let $A = U \oplus V$ and $B = V$. The goal is to take $\hat{Z} = (\hat{A} \oplus \hat{B}, \hat{B})$ upon reconstructing $\hat{A} = (U \oplus V)^\land$ and $\hat{B} = (V | U \oplus V)^\land$. Suppose we successfully decode $\hat{A}$ as to maximize the probability that $\tilde{E}_{n/2}(A)|_S$ agrees with the first half of $W_S$. Indeed, $A$ is a string of i.i.d. Bernoulli bits, and so far there should be no issue.

We now wish to decode $\hat{B}$ as to maximize the probability that $\tilde{E}_{n/2}(B)|_S$ agrees with the last half of $W_S$ given $U \oplus V = \hat{A}$. However, we cannot apply the recursive step here because $B$ is not generally a string of i.i.d. bits. The distribution for $B_i$ is distributed differently from $B_j$ when $A_i$ is not the same as $A_j$.

**Exercise 2.** Let $\hat{A}_i = 0$ and $\hat{A}_j = 1$. Compute the distributions for $B_i$ and $B_j$ given $U \oplus V = \hat{A}$.
Because the inductive step from \( Z \) to \( A \) and \( B \) does not produce identical distributions, we must strengthen the inductive hypothesis to allow \( Z \) to be a string of independent (not necessarily identically distributed) Bernoulli bits. This is equivalent to providing \( D_n \) with additional data \((p_i)_{i=1}^n\) such that \( Z_i \sim \text{Bernoulli}(p_i) \). The true algorithm proceeds as expected.

1. Compute \( a_i = \mathbb{P}[U_i \oplus V_i = 1] \) and take \( \hat{A} = D_{n/2}(W_1, W_2, \ldots, W_{n/2}; a_1, a_2, \ldots, a_{n/2}) \).
2. Compute \( b_i = \mathbb{P}[V_i = 1 | U_i \oplus V_i = \hat{A}_i] \) and take \( \hat{B} = D_{n/2}(W_{n/2+1}, W_{n/2+2}, \ldots, W_n; b_1, b_2, \ldots, b_{n/2}) \).
3. Output \( D_n(W_1, W_2, \ldots, W_n; p_1, p_2, \ldots, p_n) = (\hat{A} \oplus \hat{B}, \hat{B}) \).

**Exercise 3.** The above description of \( D_n \) is incomplete. Express \( a_i \) and \( b_i \) in terms of \( p_i \). Then use the desiderata for \( D_n \) to define the base case \( D_1(W_1; p_1) \).

If we guess \( W \) correctly from \( W_S \) at the maximum recursion depth, the algorithm recovers \( Z \) exactly because \((U, V) \mapsto (U \oplus V, V)\) is idempotent and thus invertible. As a result, the decoding error is union bounded by the sum of the errors at each bit at the maximum recursion depth. With the correct definition of \( D_1 \), the error at bit \( j \) is bounded above by \( c^{-j} \). This gives us a total decoding error of \( nc^{-1} \).

**Exercise 4.** In terms of the length \( n \) of the input string, what is the complexity of successful decoding? Draw a ternary circuit that computes \( D_n \) for the case where \( n = 8 \) and \( p = 1/4 \).

## 2 Polarization and Martingales

We take a random walk on the recursion tree. Let \( Z_i^j \) be node \( j \) at recursion depth \( i \). We begin our walk at \( Z_0^j = Z \). If \( Z_i^j = U_i^j V_i^j \) is the concatenation of strings as before, \( Z_{i+1}^j = U_i^j \oplus V_i^j \) and \( Z_{i+2}^j = V_i^j \). Just as we condition \( V \) on \( U \oplus V \), we condition \( V_i^j \) on \( U_i^j \oplus V_i^j \). Because \( Z_i^j \) appears after all the elements on which it is conditioned, we consider \( Z_i^j \) as conditioned upon \( Z_{<j}^i \).

We define our walk as a sequence \((j_i)_{i=0}^t \) such that \( j_i = 2i - 1 + b_i \) with uniformly random bits \((b_i)_{i=1}^t \). Then there is a correspondence between random walks on the recursion tree and random bits in \( W \). Taking \( Z_i^t = W_j \), gives us a uniform distribution on the bits in \( W \), and taking \((b_i)_{i=1}^t \) to be the binary expansion of a uniformly distributed \( j \in [n] \) gives us a uniform distribution on walks \((j_i)_{i=0}^t \).

Let \( X_t = H(Z_i^t | Z_{<j}^i) \) so that \( X_t = H(W_j | W_{<j}) \). Polarization is equivalent to the condition that for every \( p \in [0, 1] \) and \( c > 0 \), there exists a \( \beta < 1 \) such that for all \( t \in \mathbb{N} \),

\[
\mathbb{P}_{(X_0, \ldots, X_t) \in \mathcal{C}}[X_t \in (c^{-t}, 1 - c^{-t})] \leq O(\beta^t).
\]

We first establish some important properties of \((X_i)_{i=0}^t\).

**Exercise 5.** Show that \( 0 \leq X_i \leq 1 \) for \( 0 \leq i \leq t \). A martingale is a sequence of random variables \( X_i \in L^1(\Omega, \mathcal{F}) \) such that

\[
\mathbb{E}[X_i | X_1, X_2, \ldots, X_{i-1}] = X_{i-1}.
\]

Use the invertibility of \((U, V) \mapsto (U \oplus V, V)\) and the chain rule for entropy to show that \((X_i)_{i=0}^t \) is a Martingale.

These properties surprisingly are not sufficient to show convergence.

**Exercise 6.** Find a bounded Martingale that fails to converge.

To show that the \( X_t \) converges, we want to show that it has variance in the middle and suction at the ends. Variance in the middle, i.e., away from 0 and 1, means that for all \( \tau > 0 \), there exists a \( \sigma > 0 \) such that for all \( i \in [t] \),

\[
X_{i-1} \in (\tau, 1 - \tau) \text{ implies that } \text{Var}[X_i | X_{i-1}] \geq \sigma^2.
\]
Suction at ends means that there exists a $\Theta > 0$ such that for all $c > 0$, there exists a $\tau > 0$ such that

$$
X_{i-1} \leq \tau \text{ implies that } \mathbb{P}\left[X_i < \frac{X_{i-1}}{c}\bigg| X_{i-1}\right] \geq \Theta \text{ and }
$$

$$
X_{i-1} \geq \tau \text{ implies that } \mathbb{P}\left[1 - X_i < \frac{1 - X_{i-1}}{c}\bigg| X_{i-1}\right] \geq \Theta.
$$

Variance in the middle and suction at the ends are together called local polarization.

**Exercise 7.** Write down the distribution for $X_i|X_{i-1}$.

We claim that $X_i$ polarizes locally. We will prove this in the next lecture. It suffices to show that local polarization implies strong polarization.

**Exercise 8.** Here we sketch a proof that local polarization implies strong polarization. Suppose that $X_t$ is locally polarized. Let $\phi_t = \min\{\sqrt{X_t}, \sqrt{1-X_t}\}$. Show that $\mathbb{E}[\phi_{t+1}|\phi_t] \leq \alpha \phi_t$ for some $\alpha < 1$. Deduce via induction and Markov’s inequality that $\mathbb{P}[\phi_t > \alpha^{t/2}] \leq \alpha^{t/2}$. Conclude that $X_t$ is strongly polarized by applying Doob’s martingale inequality to $X_t$ for $t \in [t_0, 2t_0]$. 