INFORMATION THEORY AND APPLICATIONS IN COMPUTER SCIENCE

Lecture 8 | 2 November 2026

Student notes

8 Communication complexity: equality and fingerprints

Conventions. All alphabets are finite unless stated otherwise. Logarithms are base two, except ln; information is in bits. Terms with probability zero contribute zero to entropy sums, and conditional distributions are used only on events of positive probability. We write $h_2(p) = -p \log p - (1-p)\log(1-p)$ with its continuous endpoint values.

Overview. Communication complexity studies how much information two parties must exchange to compute a function of their separate inputs. Equality illustrates how randomization can reduce communication and streaming memory. We then study transcript factorization and information cost, preparing for lower bounds for set disjointness.

INDEX, tensorization, internal information, and field representation are optional. The proofs use entropy chain rules and Hadamard inner products; the polynomial root bound and binary Fano argument are proved here. Finite-field existence is assumed.

8.1 A protocol is a tree of local decisions

Alice receives $x \in \{0,1\}^n$, Bob receives $y \in \{0,1\}^n$, and

$$\mathrm{EQ}_n(x,y) = \begin{cases} 1, & x = y, \\ 0, & x \neq y. \end{cases}$$

The length $n \geq 1$ is known to both parties. Local computation is free. Neither party initially knows the other's input. A binary protocol is a rooted tree: each internal node names the next speaker, and each outgoing edge is labeled by the bit sent. The speaker's bit can depend on its input, the transcript so far, and its available random coins.

For a fixed public random value r, the speaker at a node and the decision to stop are determined by the public transcript. This prevents uncharged information from being conveyed by who speaks, when a party stops, or message timing. Variable-length messages are parsed bit by bit, so message length carries no free information.

The transcript T is the sequence of communicated bits, ending at a leaf. The worst-case cost c is the maximum length over every input and every coin outcome in the protocol's model. The tree has depth at most c. For proofs we may pad each completed path to length c with fixed dummy bits; this padding carries no additional information and gives at most 2^c possible complete records for each r.

Two output conventions. For the equality algorithms, *Bob-only output* means Bob computes the answer from his input, coins, and the transcript. A leaf need not carry a common answer. *Publicly announced output* means the answer is a function of (R, T) alone. Bob can convert the first convention to the second by sending one final answer bit, charged to the cost. We state which convention is used.

Randomness and error. The public random variable R is known to both parties and independent of the inputs. The private random tapes R_A, R_B are independent of one another and of (X, Y, R). Unless another tolerance is specified, randomized error is *strictly less than* $1/3$ on every fixed input pair. In statements parameterized by ε, the error is at most ε, with probability over all coins. This is worst-case input error, not merely average error under a selected input distribution. Public coins cost no *communication*; they can still be long. Private randomness does not become public unless it is communicated.

Finite coin variables suffice for the constructions. The consistency-event proofs also apply to infinite tapes; the entropy expressions concern finite inputs and transcripts. The lower bounds allow unlimited computation and a separate protocol for each n.

8.2 Deterministic equality: rectangles and a crossing argument

Lemma 8.1 (Rectangle property)**.** *Fix a deterministic protocol and a transcript prefix t. There exist sets A_t of Alice inputs and B_t of Bob inputs such that the input pairs whose executions reach t are exactly $A_t \times B_t$.*

Proof. At the root, every input pair is possible. Suppose the set at a node is $A \times B$. If Alice speaks next, her bit at that node is a function of x and the fixed prefix. For $b \in \{0, 1\}$, let A_b be the Alice inputs that produce b. The child reached by b has input set $A_b \times B$. If Bob speaks, the corresponding set is $A \times B_b$. Induction on the prefix length proves the assertion. The same argument applies to leaves. □

Theorem 8.2 (Exact deterministic equality)**.** *Every deterministic protocol that computes* EQ_n *correctly on every pair with Bob-only output has worst-case cost at least n bits. Sending x achieves n bits. With publicly announced output, the exact optimum is $n + 1$ bits.*

Proof. Consider the 2^n diagonal inputs (z, z). Suppose (x, x) and (y, y), with $x \neq y$, end at the same transcript t. By the rectangle property, the crossed input (x, y) reaches t too. On (y, y) Bob sees input y and transcript t, so he outputs 1. On (x, y) his view is identical; he again outputs 1, a contradiction. Thus all diagonal inputs require distinct leaves. A binary tree of depth c has at most 2^c leaves, so $2^c \geq 2^n$, hence $c \geq n$.

Sending all n bits of x lets Bob compare with y, proving the upper bound under Bob-only output. For public output, the diagonal inputs still need 2^n different leaves, all labeled 1. Since $n \geq 1$, there is an unequal pair, which must reach an additional leaf labeled 0. Therefore $2^c \geq 2^n + 1$, and the integer c is at least $n + 1$. Sending x and then Bob's answer uses $n + 1$ bits. □

What the rectangle argument actually says. Under public output, every deterministic leaf is monochromatic. Under Bob-only output, a leaf need not be monochromatic across different Bob inputs. The proof above deliberately compares two executions with the *same Bob input*. This distinction prevents an otherwise tempting incorrect proof of the exact n-bit statement.

The equality matrix. For $n = 2$, the equality matrix, with rows and columns in the order $00, 01, 10, 11$, is:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

A rectangle containing two different diagonal entries also contains their off-diagonal corners, which explains the lower bound. This proof uses *product structure*, not merely the fact that the answer has one bit.

Check for understanding. *Why is "the output is only one bit" not an upper bound of one bit on communication?* Neither party knows that output before exchanging information. The difficulty is computing it from split inputs, not storing it.

8.3 A Hadamard coordinate is an equality fingerprint

The Hadamard encoding from the previous lecture is

$$H(x)(r) = x \cdot r = \sum_{i=1}^{n} x_i r_i \pmod 2, \qquad r \in \mathbb{F}_2^n.$$

We do not construct its 2^n-bit truth table. We evaluate one coordinate.

Lemma 8.3 (One random coordinate)**.** *For fixed $x, y \in \{0,1\}^n$ and uniform $r \in \mathbb{F}_2^n$,*

$$\Pr[H(x)(r) = H(y)(r)] = \begin{cases} 1, & x = y, \\ 1/2, & x \neq y. \end{cases}$$

Proof. The first case is immediate. In the second case put $d = x + y \neq 0$ over $\mathbb{F}_2$ and choose j with $d_j = 1$. The masks pair as $\{r, r + e_j\}$. Since $d \cdot (r + e_j) = d \cdot r + 1$, exactly one mask in each pair has $d \cdot r = 0$. Equality of the two fingerprints is precisely this event. Equivalently, distinct Hadamard codewords have relative distance $1/2$. □

Proposition 8.4 (Public-coin equality)**.** *For any integer $k \geq 1$, equality has a one-way public-coin protocol using k communicated bits, nk independent public random bits, and Bob-only output. Equal inputs are always accepted; each unequal input is accepted with probability exactly 2^{-k}. Publicly announcing the output uses $k + 1$ bits.*

Proof. The public tape provides independent uniform masks $r^{(1)}, \ldots, r^{(k)} \in \mathbb{F}_2^n$. Alice sends

$$(x \cdot r^{(1)}, \ldots, x \cdot r^{(k)}).$$

Bob outputs 1 if every sent bit equals his corresponding parity. Equal inputs pass all tests. For a fixed unequal pair, the lemma gives failure probability $1/2$ for each independent test. All tests fail to detect the difference with probability 2^{-k}. Bob's final announcement costs one additional bit. □

For error at most $0 < \delta < 1$, take $k = \lceil \log(1/\delta) \rceil$. In particular, two communicated bits give error $1/4$. The inputs are fixed before the random masks are drawn. A party choosing an input after seeing the masks is a different adversarial model and is not covered by this claim.

Example 8.5 (A concrete parity test)**.** Let $x = 101$ and $y = 111$. Their difference is 010. For the mask $r = 100$, both fingerprints are 1, so the test misses the difference. For $r = 010$, the fingerprints are 0 and 1, so it detects it. Exactly the four masks whose middle bit is 0 miss the difference. Repeating the *same* mask does not reduce the error; independent masks do.

The coding connection. Channel coding uses redundant coordinates to recover a message from a corrupted word. Equality uses distance to *detect disagreement* by sampling a random coordinate.

Check for understanding. *Where did the n bits go?* They became public randomness rather than communication: one test uses a shared n-bit mask.

8.4 Streaming equality: a memory lower bound and a fingerprint

Fix $n \geq 1$, known to the algorithm. The input is a one-pass stream $x_1 \cdots x_n \# y_1 \cdots y_n$, with $x, y \in \{0,1\}^n$ and a delimiter. The algorithm must decide whether $x = y$ after the stream ends; it cannot reread a symbol. Memory counts every stored bit, including any seed that must be reused. No correlated randomness can be recovered from an uncharged external store. The error guarantee is for each fixed stream, over fresh randomness independent of that stream.

Proposition 8.6 (Deterministic equality needs n bits of memory)**.** *Every deterministic streaming algorithm that solves this task exactly has at least 2^n memory states immediately after the first half.*

Proof. Suppose distinct first halves $x \neq x'$ led to the same state. Continue both computations with the delimiter and the same second half $y = x$. They now have identical states and future symbols, so they return the same answer. But $x\#x$ must be accepted and $x'\#x$ rejected. Thus every first half must have its own state, requiring at least n stored bits. □

Equivalently, Alice can send a streaming algorithm's boundary state to Bob; a small state would give a short deterministic one-way equality protocol. This is the communication-to-memory reduction in its simplest form.

Why the public random mask needs care in a stream. A Hadamard test uses $\sum_i r_i x_i$ and $\sum_i r_i y_i$. Computing the first parity is easy, but the *same* mask bits are needed again during the second half. Storing a fully random mask takes n bits. Thus the constant-communication public-coin protocol does not by itself establish a constant-memory streaming algorithm. We use a fingerprint with a short, explicitly stored seed.

The polynomial fingerprint. Let $k = \lceil \log(3n) \rceil$ and $q = 2^k$, so $3n \leq q < 6n$. Fix a representation of $\mathbb{F}_q$ by k-bit strings. For a binary string define

$$P_x(z) = x_1 z^{n-1} + x_2 z^{n-2} + \cdots + x_n \in \mathbb{F}_q[z].$$

Choose a uniform $r \in \mathbb{F}_q$ independently of the input and use $P_x(r)$ as its fingerprint. The random seed r and each fingerprint use k bits. All operations below are exact field operations, with the bits embedded as $0, 1 \in \mathbb{F}_q$. No floating-point evaluation is involved.

Lemma 8.7 (Root bound)**.** *A nonzero polynomial of degree d over a field has at most d distinct roots.*

Proof. A nonzero constant has no root. If a is a root, polynomial division gives $P(z) = (z-a)Q(z)$ with $\deg Q = d - 1$. Every other root of P is a root of Q, because its factor $z - a$ is nonzero. Induction bounds the number by $1 + (d-1) = d$; if there is no root, the claim is immediate. □

Theorem 8.8 (Randomized streaming equality)**.** *There is a one-pass algorithm for the task above using $O(\log(n+1))$ bits of memory, which always accepts equal halves and accepts any fixed unequal halves with probability less than $1/3$.*

Proof. Store the uniform seed r and an accumulator $v = 0$. On each successive bit b of the first block use Horner's update

$$v \leftarrow rv + b.$$

After j bits, $v = \sum_{i=1}^{j} x_i r^{j-i}$; multiply the previous sum by r and add x_j to check the invariant. Save the final value $u = P_x(r)$, reset $v = 0$ at the delimiter, and process the second half by the same update. Its final value is $v = P_y(r)$. Accept exactly when $u = v$.

Equal halves always agree. If $x \neq y$, the difference $P_x - P_y$ is a nonzero polynomial of degree at most $n-1$: a differing bit coefficient is nonzero also in characteristic two. The root bound gives

$$\Pr_r[P_x(r) = P_y(r)] \leq \frac{n-1}{q} < \frac{1}{3}.$$

An element of $\mathbb{F}_q$ is a k-bit vector, so the seed is exactly k fresh random bits. Retain only the three field registers r, u, v, the degree-k binary irreducible modulus describing the field, a phase flag, and if needed an $O(\log(n+1))$-bit position counter. Multiplication and reduction use $O(k)$ temporary bits. All stored and working memory, including the reusable seed, is $O(k + \log(n+1)) = O(\log(n+1))$ bits. □

Example 8.9 (A collision is allowed and counted)**.** For $x = 1011$ and $y = 1001$, the difference polynomial is z, so only seed $r = 0$ collides, with probability $1/q$. Excluding zero is no universal fix: for $x = 00, y = 11$, the difference is $z + 1$ and seed $r = 1$ collides instead.

A compact-seed communication protocol as well. With public r, Alice sends the k-bit value $P_x(r)$ and Bob compares it with $P_y(r)$. With only private coins, Alice sends both r and $P_x(r)$, using $2k = O(\log n)$ bits; Bob alone can output. Add one bit if both must know the answer. The Hadamard construction instead uses more shared randomness to obtain constant communication at constant error.

Why the seed must be random. Why can we not fix a good seed for all unequal pairs? The guarantee fixes x, y first and averages over r; changing that quantifier is invalid. For $n > k$, every fixed seed maps 2^n strings to at most $q = 2^k$ fingerprints, so some pair collides. Why is a checksum modulo two inadequate? It accepts, for example, 00#11 with certainty: those distinct strings have the same parity.

For target error $0 < \delta < 1$, choose $k = \lceil \log(n/\delta) \rceil$. The same proof gives error below δ and space $O(\log(n+1) + \log(1/\delta))$, with exactly k random seed bits.

8.5 Randomized rectangles: probabilities factor

Return to the interactive two-party model, now with arbitrary finite input sets $\mathcal{X}, \mathcal{Y}$. Write

$$q^r_{xy}(t) = \Pr[T = t \mid X = x, Y = y, R = r].$$

The probability is over private coins. We prove a property of the *channel* from the fixed inputs to the transcript; no distribution on inputs is needed for the statement.

Theorem 8.10 (Transcript factorization)**.** *For each public value r and transcript t, there are nonnegative functions $a_x(r,t)$ and $b_y(r,t)$ such that*

$$q^r_{xy}(t) = a_x(r,t) b_y(r,t). \tag{8.1}$$

The same assertion holds when t denotes a reached prefix. If $p(r)$ is the public-coin probability and $W = (R, T)$ is the full public record, then

$$\Pr[W = (r,t) \mid x, y] = p(r) a_x(r,t) b_y(r,t). \tag{8.2}$$

Proof. Fix r, t, x, y. Follow the path described by t. Let $A_x(r,t)$ be the set of Alice private tapes that cause her to send every Alice bit on that path when her input is x and the messages received so far are those specified by t. Define $B_y(r,t)$ by the analogous consistency test for Bob. These tests refer only to the indicated local input and tape, because the incoming messages are already fixed by t.

The execution follows t if and only if both consistency tests succeed. Indeed, induction along the path shows that the next actual bit is the prescribed bit; the converse holds by reading any actual execution. The stopping rule for a complete transcript is already determined by the tree. Since the two private tapes are independent, the probability is the product

$$a_x(r,t) = \Pr[R_A \in A_x(r,t)], \qquad b_y(r,t) = \Pr[R_B \in B_y(r,t)].$$

This proves the prefix and complete-transcript versions. Multiplying by $p(r)$ gives the full-record formula because R is independent of the inputs. □

Why this proof uses entire local tapes. Earlier private random bits can affect later messages. Multiplying unconditional per-message probabilities would ignore these dependencies. The consistency event tests all of one party's choices together, then uses independence only *between the two parties' tapes.*

For a deterministic protocol, these factors are 0–1 indicators and the theorem reduces to the rectangle property. In general the likelihood matrix $(q^r_{xy}(t))_{x,y}$ for a fixed r,t is an outer product. In particular it obeys the pointwise crossing identity

$$q^r_{xy}(t)q^r_{x'y'}(t) = q^r_{xy'}(t)q^r_{x'y}(t). \tag{8.3}$$

Example 8.11 (Forgetting the public coin can destroy factorization)**.** Alice and Bob each have one bit. Let R be a public fair bit. Alice sends $x \oplus R$, then Bob sends $y \oplus R$. If we forget R, the probability of transcript 00, as a matrix indexed by x, y, is

$$\begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

It is not an outer product: its two diagonal entries have product $1/4$ while the off-diagonal entries have product 0. Conditional on R, or with (R,T) retained as the record, the theorem applies exactly.

Likelihood is not the posterior. For a general input distribution μ, the posterior is proportional to $\mu(x,y)a_x(r,t)b_y(r,t)$. A correlated prior need not become independent. The theorem itself is a likelihood statement requiring no independence of X, Y. Its independence assumption concerns the private tapes. The useful product-prior corollary is proved in the optional internal-information section.

8.6 Information carried by a transcript cannot exceed its length

Fix any input distribution μ on (X,Y), possibly correlated. The protocol's external information cost under μ is

$$\mathrm{IC}^{\mathrm{ext}}_\mu(\pi) = I(X,Y;T \mid R).$$

It measures the information a listener to the full public record learns about the input pair. Since R is independent of the inputs, the chain rule gives

$$I(X,Y;R,T) = I(X,Y;T \mid R).$$

This equality does not charge the entropy of the public seed to communication. It says that the seed by itself carries no information about the independently chosen inputs.

Theorem 8.12 (External and conditional information budgets)**.** *If a protocol has worst-case communication at most c, then for every input distribution*

$$I(X,Y;T \mid R) \le c.$$

More generally, let D be any auxiliary random variable jointly distributed with (X,Y), with all protocol coins independent of the joint triple (X,Y,D). The protocol need not be given D. Then

$$I(X,Y;T \mid D,R) \le c, \qquad I(X,Y;R,T \mid D) = I(X,Y;T \mid D,R). \tag{8.4}$$

Proof. For fixed r, at most 2^c leaves can occur. Further conditioning on $D = d$ changes their probabilities but cannot add possible leaves. Entropy is at most the logarithm of the support size, so

$$I(X,Y;T \mid D,R) = H(T \mid D,R) - H(T \mid X,Y,D,R) \le H(T \mid D,R) \le c.$$

Taking D constant proves the first assertion. For the displayed equality, expand by the chain rule and use $I(X,Y;R \mid D) = 0$, which follows from independence of R from the *joint* triple (X,Y,D). □

Conditioning and side information. First, D can be an analyst's hidden variable; conditioning on D does not mean that Alice and Bob receive it. Second, conditional mutual information can increase under extra conditioning. We did not infer $I(X,Y;T \mid D,R) \le I(X,Y;T \mid R)$; that inequality is generally false. We bounded the conditional transcript support directly.

Example 8.13 (An independent seed can still be essential to interpreting a message)**.** Let X and R be independent fair bits, and let Alice send $T = X \oplus R$. Then $I(X;T) = 0$, but $I(X;T \mid R) = 1$. A listener who knows the public seed recovers X. Thus forgetting public randomness can undercount information as well as destroy factorization.

8.7 The bridge to disjointness: the answer need not explain the conversation

Define $\mathrm{DISJ}_n(x,y) = 1$ when no coordinate has $x_i = y_i = 1$. Sending x and one answer bit costs $n+1$ bits. Our next target is to prove that randomized protocols with any fixed error $\varepsilon < 1/2$ still require $\Omega(n)$ bits, even with arbitrary interaction. Equality shows why a deterministic lower bound alone cannot establish that conclusion.

For one coordinate, the complementary output is $\mathrm{AND}(a,b)$. Imagine a statistical channel that reveals only that answer: its transcript is 0 on $(0,0),(0,1),(1,0)$ and 1 on $(1,1)$. It is not the transcript of a legal protocol. For transcript 0, factorization would require

$$\underbrace{P_{00}(0)P_{11}(0)}_{1\cdot 0=0} = \underbrace{P_{01}(0)P_{10}(0)}_{1\cdot 1=1},$$

a contradiction. Equivalently, the three zero-inputs are not a rectangle. A distributed conversation must reveal something beyond the answer.

A worked protocol, not yet a lower bound. Alice sends her bit a. If it is zero, stop and output zero; if it is one, Bob sends b and both output b. The possible transcripts are $0, 10, 11$. Under the uniform distribution on the three zero-inputs the answer is constant, but the transcript is 0 with probability $2/3$ and 10 with probability $1/3$. Because the protocol is deterministic, its external information cost is

$$I(A,B;T) = H(T) = h_2(1/3) > 0.$$

This only computes the cost of *one* protocol. The goal is to prove a positive lower bound for *every* bounded-error AND protocol that is correct on all four inputs, measuring leakage on zero-inputs, and then accumulate that cost across coordinates. The randomized factorization is what lets correctness on $(1,1)$ constrain transcripts on the other inputs.

Why output entropy is not the right lower bound. On disjoint inputs the DISJ answer is always one, so its entropy is zero. A constant-answer protocol is correct on that support but fails on intersecting inputs. Worst-case correctness outside the support, together with protocol structure, is the source of the forthcoming lower bound. The quantitative lower bound is developed in Lecture 9.

8.8 Summary and review questions

1. *What changed between deterministic equality and fingerprinting?* The error allowance and the quantifier over fresh random coins. For fixed unequal inputs, a random test is unlikely to collide; no one short fixed fingerprint must separate every pair.
2. *Why is streaming memory larger than public-coin communication in these constructions?* A reusable random seed must be stored in the streaming state. The polynomial construction counts it explicitly.
3. *Does a randomized transcript still correspond to a monochromatic rectangle?* For fixed public coins its likelihood factorizes. Monochromatic leaves are guaranteed for exact deterministic protocols with public output. Bob-only output and bounded-error protocols need not have that property.
4. *Can a constant answer still force information leakage?* Yes, when correctness is also required on inputs outside the analysis support. Proving the quantitative version is the next task.

For $n \geq 1$ and target error $0 < \delta < 1/2$, put $k = \lceil \log(1/\delta) \rceil$. The main bounds are

$$\begin{aligned} \text{deterministic EQ, Bob outputs} &: n \text{ bits}, \\ \text{public-coin EQ, Bob outputs} &: k \text{ bits}, \\ \text{streaming EQ, exact deterministic} &: s \geq n, \\ \text{streaming EQ, error below } 1/3 &: s = O(\log(n+1)), \\ \text{every protocol and analysis distribution} &: I(X, Y; T \mid D, R) \leq c. \end{aligned}$$

Further reading. The sampled-coordinate protocol uses the Hadamard encoder from Lecture 7. For background on communication and information cost, see Mark Braverman's overview of information complexity.

8.9 Optional: adding information across coordinates

The following inequality lets us account for information about independent coordinates separately, including when independence holds only after conditioning on a hidden variable.

Lemma 8.14 (Conditional tensorization)**.** *Let $Z_1, \ldots, Z_n$ be independent conditional on D, and let R be independent of $(Z_1, \ldots, Z_n, D)$. For every transcript random variable T,*

$$I(Z_1, \ldots, Z_n; T \mid D, R) \geq \sum_{i=1}^{n} I(Z_i; T \mid D, R). \tag{8.5}$$

If the Z_i collectively describe the inputs to a c-bit protocol, the left side is at most c.

Proof. Conditional independence gives

$$H(Z_1, \ldots, Z_n \mid D, R) = \sum_i H(Z_i \mid D, R).$$

Conditional subadditivity gives

$$H(Z_1, \ldots, Z_n \mid T, D, R) \leq \sum_i H(Z_i \mid T, D, R).$$

For completeness, this subadditivity follows from the entropy chain rule and the fact that removing $Z_1, \ldots, Z_{i-1}$ from the conditioning can only increase the ith conditional entropy. Subtract the second displayed inequality from the first equality to get (8.5). The upper bound follows directly from the transcript support: for fixed public coins a depth-c tree has at most 2^c leaves, so $I(Z_1, \ldots, Z_n; T \mid D, R) \leq H(T \mid D, R) \leq c$. □

This is not a statement that a transcript splits into one message per coordinate. A single bit may concern many coordinates. The inequality allocates *information*, using independence in the input distribution.

For a single coordinate, choose an independent fair selector $D \in \{A, B\}$ and a fair bit U, and set

$$D = A: \ (X, Y) = (U, 0), \qquad D = B: \ (X, Y) = (0, U).$$

The selector belongs to the analyst; it is not given to the protocol. The inputs are independent given D, but correlated without it. The pair 11 never occurs. Draw n independent copies, and write $Z_i = (X_i, Y_i)$ and $D = (D_1, \ldots, D_n)$. The preceding proved tools give

$$c \ \geq \ I(X, Y; T \mid D, R) \geq \ \sum_{i=1}^{n} I(X_i, Y_i; T \mid D, R).$$

Every sampled pair is disjoint, so the answer has entropy zero under this distribution. Consequently a proof based only on the entropy of the answer cannot establish the desired lower bound.

Completing the lower bound requires showing that correctness on *all* inputs, including the absent one-coordinate input 11, forces a positive information contribution. It uses square-root probability distances and *cut-and-paste*, whose algebraic starting point is (8.3), and then an explicit legal coordinate embedding. Both steps are developed in Lecture 9.

8.10 Additional example (optional): INDEX and streaming

This example applies coordinate recovery to one-way communication and then converts the lower bound into a streaming memory lower bound. It uses entropy and binary Fano; transcript factorization is not needed.

8.10.1 The task and why direction matters

Alice receives $x \in \{0,1\}^n$, Bob receives $i \in [n]$, and Bob must output

$$\mathrm{INDEX}_n(x,i) = x_i.$$

Only Alice may send a message. Public randomness R and private randomness R_A, R_B are independent of the inputs and each other. Worst-case error is at most $0 \le \varepsilon < 1/2$. Alice's message has worst-case cost c bits under the same fixed-length or prefix-free tree convention as above; Bob's local output is not charged. Arbitrary local computation is allowed.

If interaction were permitted, Bob could send i in $\lceil \log n \rceil$ bits, and Alice could answer with one bit. The linear lower bound below is therefore specifically *one-way Alice-to-Bob*.

Theorem 8.15 (One-way INDEX)**.** *Every such protocol satisfies*

$$c \ge n\big(1 - h_2(\varepsilon)\big). \tag{8.6}$$

The same bound holds if error at most ε is required only on average for a uniform independent pair (X, I).

Proof. Let X be uniform on $\{0,1\}^n$, and let I be uniform on $[n]$, independent of X and all coins. Let

$$M = f(X, R, R_A), \qquad \widehat{X}_I = g(M, R, I, R_B), \qquad e = \Pr[\widehat{X}_I \ne X_I] \le \varepsilon.$$

Alice does not know I, so I is independent of (X, M, R). This independence is the exact point where the communication direction enters.

We first prove the binary Fano step explicitly. Put $V = (M, R, I, R_B)$ and $E = \mathbf{1}\{\widehat{X}_I \ne X_I\}$. For fixed V, the estimate is known and $X_I = \widehat{X}_I \oplus E$. Thus E and X_I determine each other given V, and

$$H(X_I \mid V) = H(E \mid V) \le H(E) = h_2(e) \le h_2(\varepsilon).$$

The last inequality uses monotonicity of binary entropy on $[0, 1/2]$. Since Bob's private tape is independent of (X, M, R, I), removing it from the conditioning does not change the first entropy. Hence

$$H(X_I \mid M, R, I) \le h_2(\varepsilon). \tag{8.7}$$

Independence and uniformity of I imply

$$H(X_I \mid M, R, I) = \frac{1}{n} \sum_{j=1}^{n} H(X_j \mid M, R).$$

By the entropy chain rule and conditioning reducing entropy,

$$H(X \mid M, R) \le \sum_{j=1}^{n} H(X_j \mid M, R) \le n h_2(\varepsilon).$$

The input bits are independent and uniform and R is independent of X, so $H(X \mid R) = n$. It follows that

$$I(X; M \mid R) = H(X \mid R) - H(X \mid M, R) \ge n(1 - h_2(\varepsilon)).$$

Finally, for each r the message has at most 2^c possible values, so $I(X; M \mid R) \le H(M \mid R) \le c$. Combining the inequalities gives (8.6). Only the average error $e \le \varepsilon$ was used in the proof. □

The governing idea. The message is chosen before Alice learns which coordinate Bob needs. Good average recovery of a uniform random coordinate bounds the sum of the coordinate uncertainties. Those bounds force the message to contain linear information about the entire string. Bob need not recover all coordinates at once; recovering a uniform random one suffices.

Example 8.16 (Interpreting the constant)**.** For error $1/3$, $1 - h_2(1/3) \approx 0.0817$, a positive constant. The proof therefore yields a linear lower bound, although it does not optimize its constant. For zero error it gives $c \geq n$, attained by sending x.

8.10.2 A one-pass streaming lower bound with all state counted

Consider the promised stream consisting of a bit string $x_1 \cdots x_n$, a delimiter, and a binary encoding of $i \in [n]$ of known length $\lceil \log n \rceil$. The task is to output x_i after reading the suffix. For $n = 1$ the index is implicit. The index is not known during the first block. Use the same read-once fresh-randomness model as for streaming equality: any persistent seed, generator state, tape pointer needed for reuse, or saved random bits belong to the workspace.

Corollary 8.17 (Streaming query lower bound)**.** *If a one-pass randomized algorithm solves this task with worst-case error at most $\varepsilon < 1/2$ and at most s bits of mutable state at the delimiter, then*

$$s \geq n(1 - h_2(\varepsilon)).$$

In particular its maximum workspace is $\Omega(n)$ for fixed error below $1/2$.

Proof. Turn the algorithm into a one-way INDEX protocol. Alice simulates its processing of her block x and the delimiter, using private fresh coins, then sends the complete memory state to Bob in s bits. Bob continues from this state on his index suffix, using independent fresh coins, and outputs the algorithm's answer.

All persistent memory, including any reusable seed, is in the transmitted s bits. Fresh coins after the split have their original distribution, so the output distribution on each (x, i) is unchanged. The INDEX theorem therefore applies. □

Equality versus INDEX. Streaming equality asks whether two whole blocks agree, and a short randomized polynomial fingerprint suffices. A later arbitrary coordinate query demands a summary that supports many possible questions, forcing linear memory. The difficulty depends on the query, not just the input length or the one-bit output size.

Relation to disjointness. If Bob's set is the singleton $\{i\}$, then $\mathrm{DISJ}_n(x, e_i) = 1 - x_i$. Thus INDEX immediately gives a linear bound for *one-way Alice-to-Bob* disjointness. It does not give a linear bound for interactive disjointness: on singleton inputs Bob can send i first. The later lecture's interactive lower bound needs its separate factorization, distance, and embedding argument.

8.11 Optional: internal information

External information concerns a listener who knows the public record. Internal information instead measures what the two participants learn about each other's inputs.

Define the conventional internal information cost by

$$\mathrm{IC}_\mu^{\mathrm{int}}(\pi) = I(X;T \mid Y,R) + I(Y;T \mid X,R).$$

It measures information about Alice's input given Bob's input, plus the reverse quantity. This definition does not explicitly condition on the private tapes; its relation to the complete private views is explained below.

Proposition 8.18 (Internal information is also bounded by communication)**.** *For every joint input distribution and every c-bit protocol in our model,*

$$I(X;T \mid Y,R) + I(Y;T \mid X,R) \le c.$$

Proof. Pad transcripts to c bits, writing $T = (T_1, \dots, T_c)$ and assigning a speaker to each padded node. Consider a fixed public value and reached prefix u. If Alice speaks next, the prefix factorization gives, whenever the denominator is positive,

$$\Pr[T_j = b \mid x,y,r,T_{<j} = u] = \frac{a_x(r,ub)b_y(r,u)}{a_x(r,u)b_y(r,u)} = \frac{a_x(r,ub)}{a_x(r,u)}.$$

The Bob factor is unchanged because the new bit imposes a constraint only on Alice's tape. The ratio does not depend on y. Consequently this Alice bit contributes zero to $I(Y;T_j \mid X,R,T_{<j})$ and at most one bit to $I(X;T_j \mid Y,R,T_{<j})$. At a Bob node the roles reverse. A dummy bit contributes zero to both.

If the speaker varies with the prefix, apply these observations separately to each prefix and average: the first contribution is at most the probability of an Alice node, and the second at most that of a Bob node. Their sum is at most one. Sum over j and use the mutual-information chain rule for each direction. The result is at most c. □

Corollary 8.19 (Product priors remain product after a transcript)**.** *Suppose X and Y are independent given a random variable D, and the public and private coins are independent of (X,Y,D). Then X and Y are independent given (D,R,T).*

Proof. For fixed d,r,t of positive probability, the prior has form $u_d(x)v_d(y)$. Bayes' formula and (8.1) make the posterior proportional to

$$[u_d(x)a_x(r,t)]\,[v_d(y)b_y(r,t)].$$

Its normalizing sum factors into the product of the two separate sums, so the normalized posterior is a product as well. □

A useful equality under a product prior. For arbitrary random variables and a conditioning variable V, expanding the entropies gives

$$\begin{aligned} &I(X;T \mid Y,V) + I(Y;T \mid X,V) - I(X,Y;T \mid V) \\ &\quad = I(X;Y \mid T,V) - I(X;Y \mid V). \end{aligned}$$

If X,Y are independent given D, take $V = (D,R)$. The prior independence and the posterior-product corollary to Theorem 8.10 make both terms on the right zero. Thus for this situation,

$$I(X,Y;T \mid D,R) = I(X;T \mid Y,D,R) + I(Y;T \mid X,D,R).$$

This is a conditional statement using the specified product structure. It is not an identification of unconditional external and internal information under every correlated input distribution.

Private views and the conventional formula. One can define what Bob learns using his entire view (Y, R, R_B, T). Prefix factorization also shows X and R_B are conditionally independent given (Y, R, T): after fixing Y, R, T, Bob's private tape has its local consistency distribution, independent of X. As $I(X; R_B \mid Y, R) = 0$ initially, the chain rule in two orders gives

$$I(X; T \mid Y, R, R_B) = I(X; T \mid Y, R).$$

The symmetric identity holds for Alice. Thus the stated formula agrees with information learned through the complete private views in this standard independent-coins protocol model.

8.12 Optional: representing finite fields

Field existence is the usual finite-field algebra fact. A representation can be fixed from n before reading the stream. It can also be found uniformly in $O(k)$ workspace: enumerate monic degree-k binary polynomials and reject a candidate if a *nonconstant* monic polynomial of a degree from 1 through $\lfloor k/2 \rfloor$ divides it. A reducible degree-k polynomial has a nonconstant factor in that range; an irreducible polynomial has none. The enumeration and polynomial divisions use $O(k)$ bits and terminate at an irreducible one.

Elements are binary polynomials of degree less than k, reduced modulo the selected degree-k polynomial. The modulus occupies $k + 1$ bits; elementary multiplication and polynomial long division need $O(k)$ working bits.