A Journey of Randomness Extractors

Kai-Min Chung

Academia Sinica, Taiwan

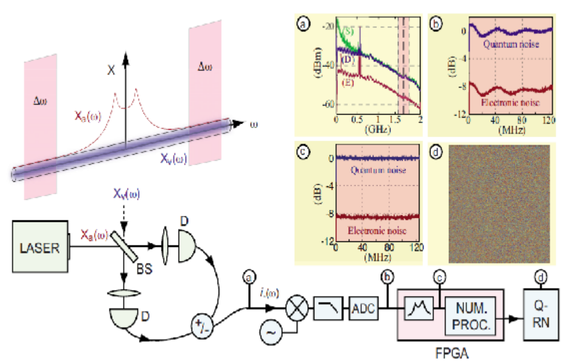
Outline

- Basic concepts
  - Statistical distance
  - Min-entropy
  - Randomness Extractors
- Leftover Hash Lemma:
  - An efficient extractor based on universal hash functions
- Average-case Extractors:
  - Randomness extraction in presence of side information
- (Optional) Quantum-proof Extractors
  - Extraction in presence of quantum side information

Quest for Perfect Randomness

- Randomness is powerful resource
  - Crypto requires truly uniform bits to generate keys
  - Randomized algorithm assumes access to truly uniform bits
- In reality, random sources are not perfect
  - Correlated and biased bits
- Can we turn imperfect source into (almost) uniform bits?

Imperfect random source:



Examples

- IID-Bit source: $X = X_1, X_2, .., X_n \in \{0,1\}$ identical & independent, but biased: for each i, $\Pr[\ X_i = 1\ ] = \delta$ for some unknown δ
  - idea: consider X in pairs,

$$X_i, X_{i+1} = \begin{cases} 01 & \Longrightarrow \text{output } 0 \\ 10 & \Longrightarrow \text{output } 1 \\ 00/11 & \Longrightarrow \text{discard} \end{cases}$$

- Independent-bit source: $X = X_1, X_2, .., X_n \in \{0,1\}$ independent, but with different biased: $\Pr[\ X_i = 1\ ] = \delta_i$ for different δ_i, where $0 < \delta \le \delta_i \le 1 - \delta$ for some constant δ
  - idea: output parity of each t bits

$$|\ \Pr[\bigoplus_{i=1}^{t} X_i = 1] - ½\ | \le 2^{-\Omega(t)}$$

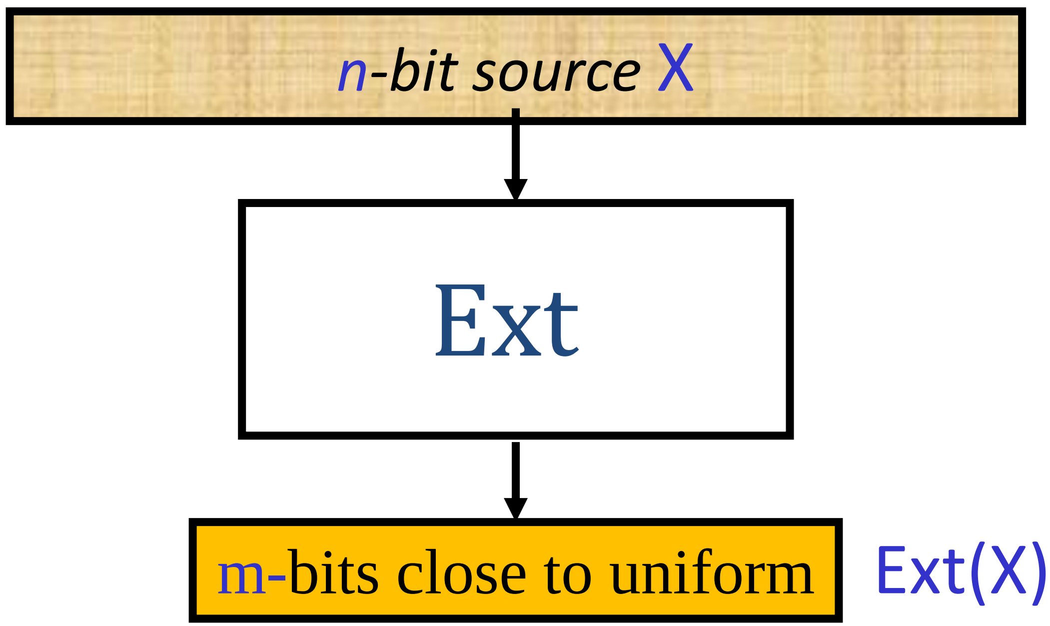
Randomness Extraction

- Source: random variable X over $\{0,1\}^n$ in certain class $\mathcal{C}$
  - IndBits$_{n,\delta}$: $X = X_1, X_2, .., X_n \in \{0,1\}$ independent bits, $\Pr[X_i = 1] = \delta_i$ where $0 < \delta \leq \delta_i \leq 1 - \delta$
  - IndBits$_{n,\delta}$: additionally assume all δ_i are equal

- (Deterministic) extractor: a function $\text{Ext}: \{0,1\}^n \rightarrow \{0,1\}^m$ s.t. $\forall$ source $X \in \mathcal{C}$, Ext(X) is "ε-close" to uniform

Deterministic Extractors

- (Deterministic) extractor: a function $\mathrm{Ext}: \{0,1\}^n \rightarrow \{0,1\}^m$ s.t. $\forall$ source $X \in \mathcal{C}$, $\mathrm{Ext}(X)$ is "ε-close" to uniform



  - single function works for all sources in $\mathcal{C}$
  - only one sample X is available
  - need to define "ε-close" to uniform

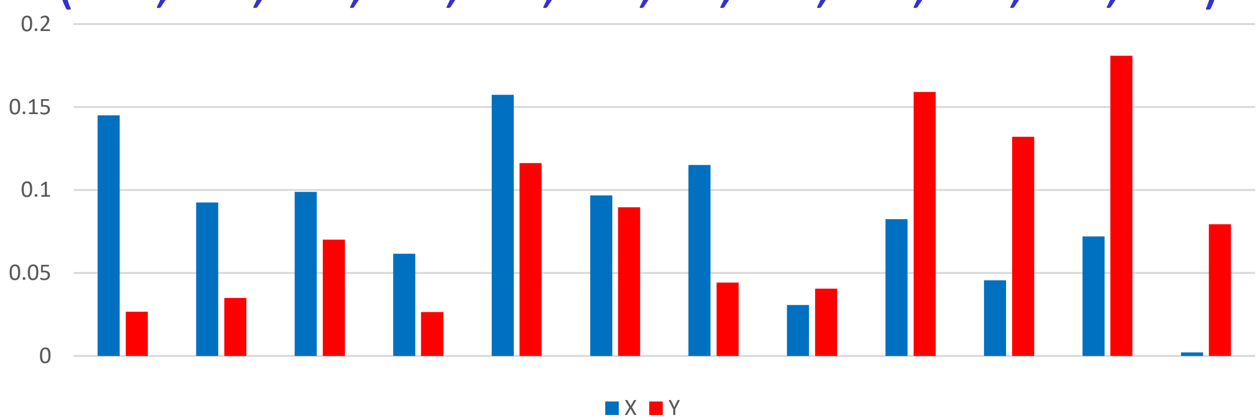
Statistical Distance

- **Def.** Let X, Y be rand. var. over range U, statistical distance between X, Y is defined as

$$\Delta(\mathrm{X},\mathrm{Y}) \stackrel{\text{def}}{=} (1/2)\cdot\sum_{u\in U} |\Pr[X=u]-\Pr[Y=u]|$$

  - View X, Y as vectors over $\mathbb{R}^{|U|}$, it's simply the L1-distance
- **Def.** We say X is ε-close Y if $\Delta(\mathrm{X},\mathrm{Y}) \le \varepsilon$

Example: X = (.15,.09,.10,.06,.16,.09,.11,.03,.08,.04,.078,.002)
Y = (.03,.04,.07,.03,.11,.09,04,.04,.16,.13,.18,.08)



Important Properties

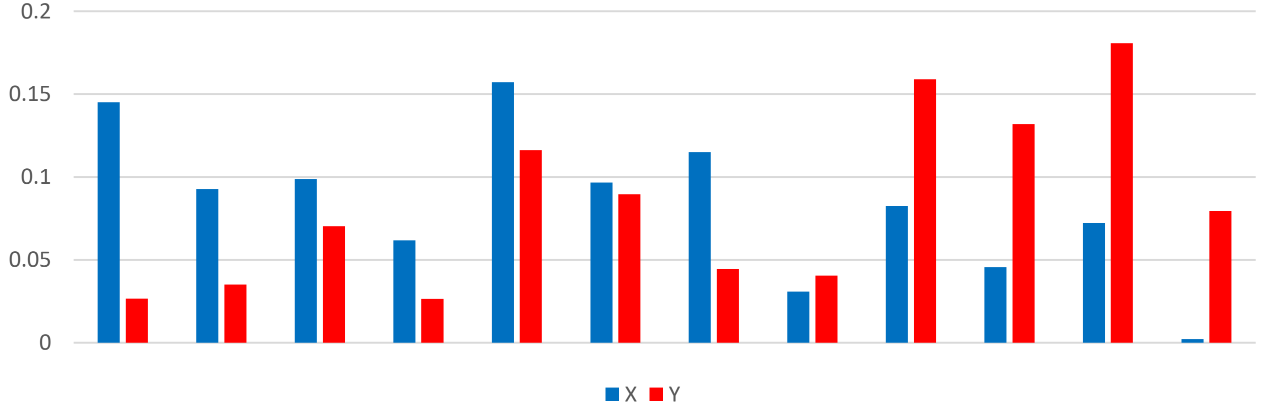
- Operational meaning: max advantage to distinguish X, Y

$$\Delta(\mathrm{X},\mathrm{Y}) = \max_{T \subset U}(\Pr[X \in T] - \Pr[Y \in T])$$

  – In particular, if X is ε-close Y, then for any event T,

$$\Pr[X \in T] \leq \Pr[Y \in T] + \varepsilon$$

Example: X = (.15,.09,.10,.06,.16,.09,.11,.03,.08,.04,.078,.002)
Y = (.03,.04,.07,.03,.11,.09,04,.04,.16,.13,.18,.08)



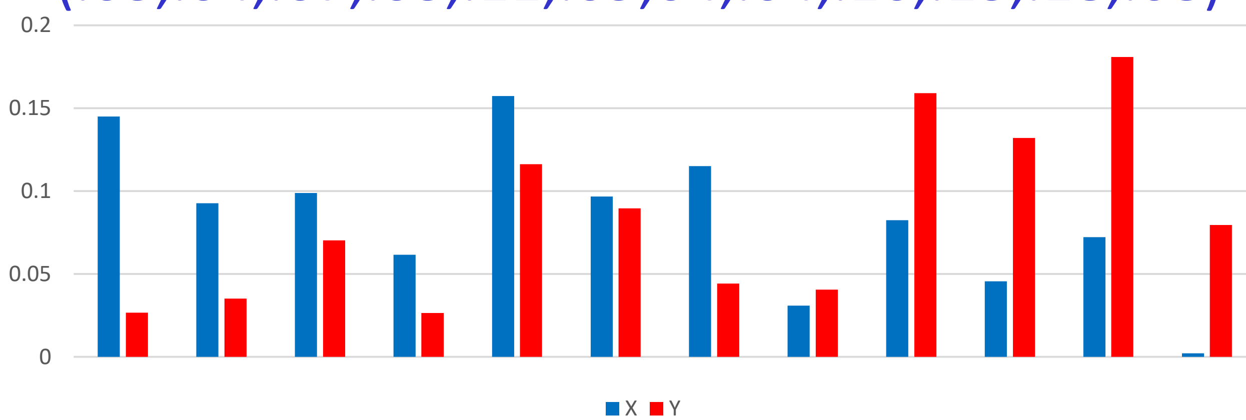
Important Properties

- Post-processing inequality: for any function f,

$$\Delta(f(X), f(Y)) \leq \Delta(X, Y)$$

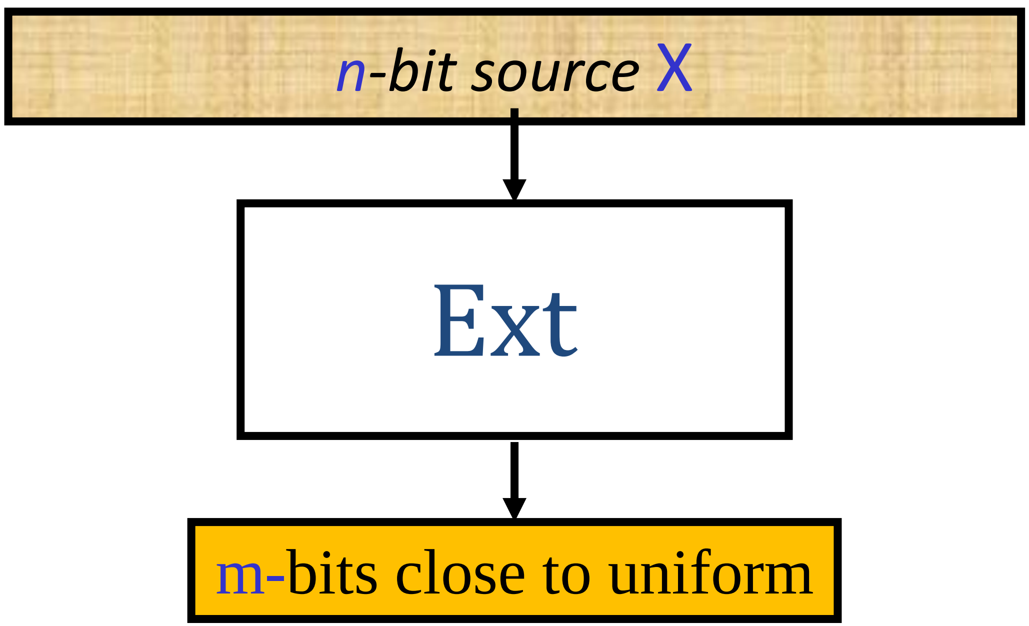
  – I.e., post-processing only decreases statistical distance
  – Equality holds when f is injective

Example: X = (.15,.09,.10,.06,.16,.09,.11,.03,.08,.04,.078,.002)
Y = (.03,.04,.07,.03,.11,.09,04,.04,.16,.13,.18,.08)



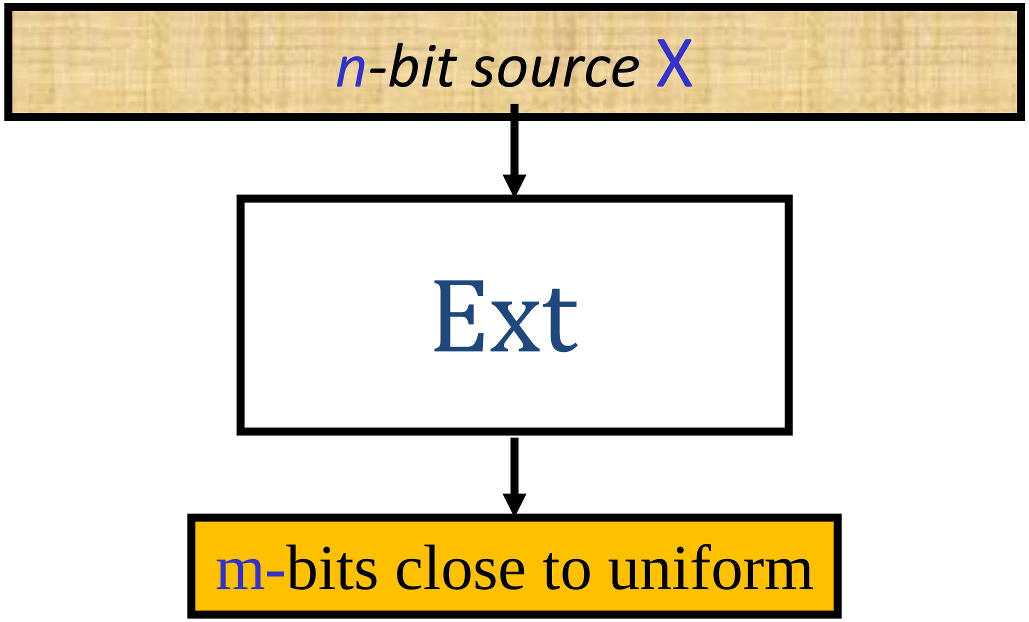
Extractor for $\text{IndBits}_{n,\delta}$

- **Thm.** $\forall$ constant δ, $\forall$ n, m $\in \mathbb{N}$, $\exists$ Ext: $\{0,1\}^n \to \{0,1\}^m$ for $\text{IndBits}_{n,\delta}$ source with error ε = m$\cdot$ $2^{-\Omega(n/m)}$
  - Ext(X) breaks X into m blocks of length $\lfloor n/m \rfloor$ and outputs the parity of each block



Extractor for General Sources?

- Can we extract truly uniform bits from any sources?
  - No, if the source is not random, e.g., $X = 0^n$ w.p. 1
- Hope: Ext works whenever X has sufficient "entropy"



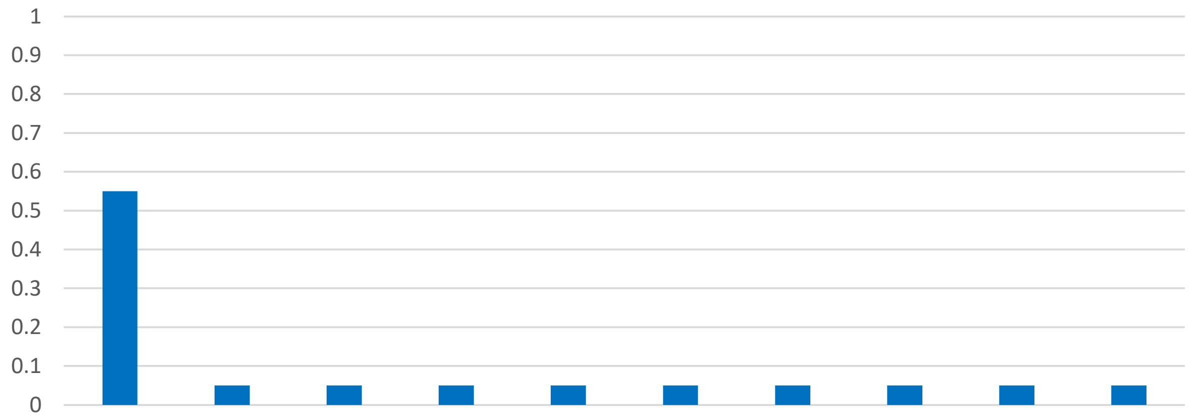
1st Attempt: Shannon Entropy

- **Def.** Shannon entropy $H_{sh}(X)$

  $$H_{sh}(X) \stackrel{\text{def}}{=} \sum_x \Pr[X = x] \log \frac{1}{\Pr[X=x]} = \mathrm{E}_{x \leftarrow X}\left[\log \frac{1}{\Pr[X=x]}\right]$$

  – Not good, consider X defined as follows:
    - w.p. ½, set $X = 0^n$
    - w.p. ½, sample X = uniform on $\{0,1\}^n$

  – $H_{sh}(X) \geq n/2$ but $\Pr[X=0^n] > ½$; can't extract from X

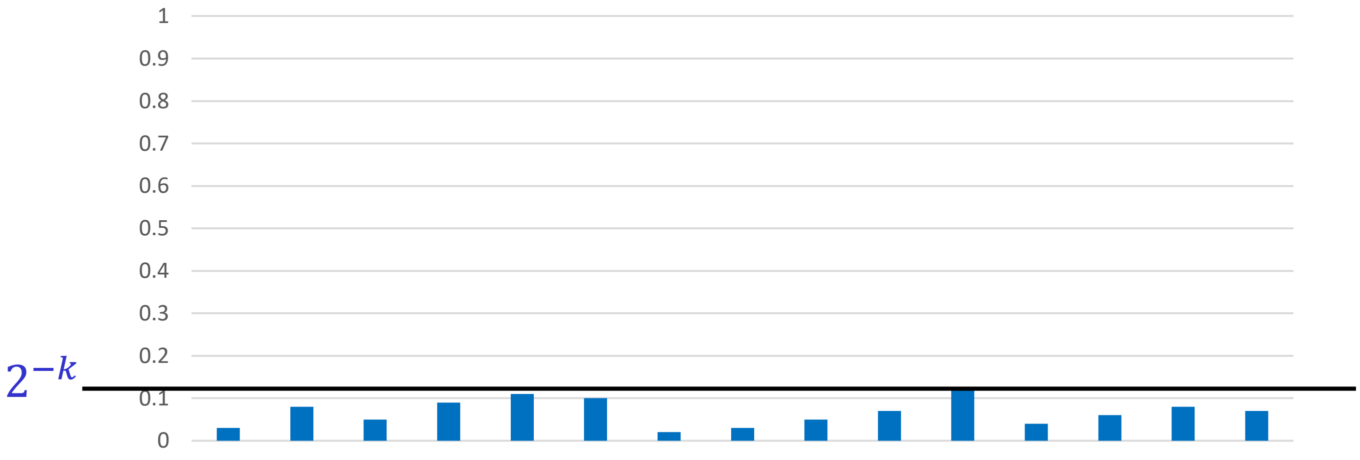


2^{nd} Attempt: Min-Entropy

- **Def.** Min-entropy $H_{min}(X)$

$$H_{min}(X) \stackrel{\text{def}}{=} \max_x \left\{\log \frac{1}{\Pr[X=x]}\right\}$$

  - $H_{min}(X) \geq k$ if for every x, $\Pr[X = x] \leq 2^{-k}$
  - Worst-case notion; possible for extraction
- **Def.** X is k-source if $H_{min}(X) \geq k$
- Extractor for the class of k-sources?



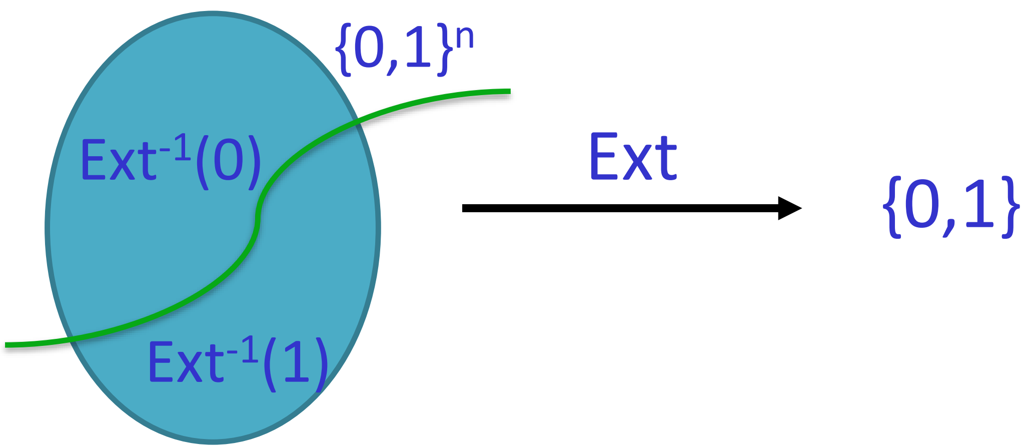
Impossibility of Deterministic Extraction

- **Thm.** For any Ext: $\{0,1\}^n \rightarrow \{0,1\}$, there exists an $(n\text{-}1)$-source X s.t. $\mathrm{Ext}(X)$ = constant

Proof. Consider X_b = uniform on $\mathrm{Ext}^{-1}(b)$

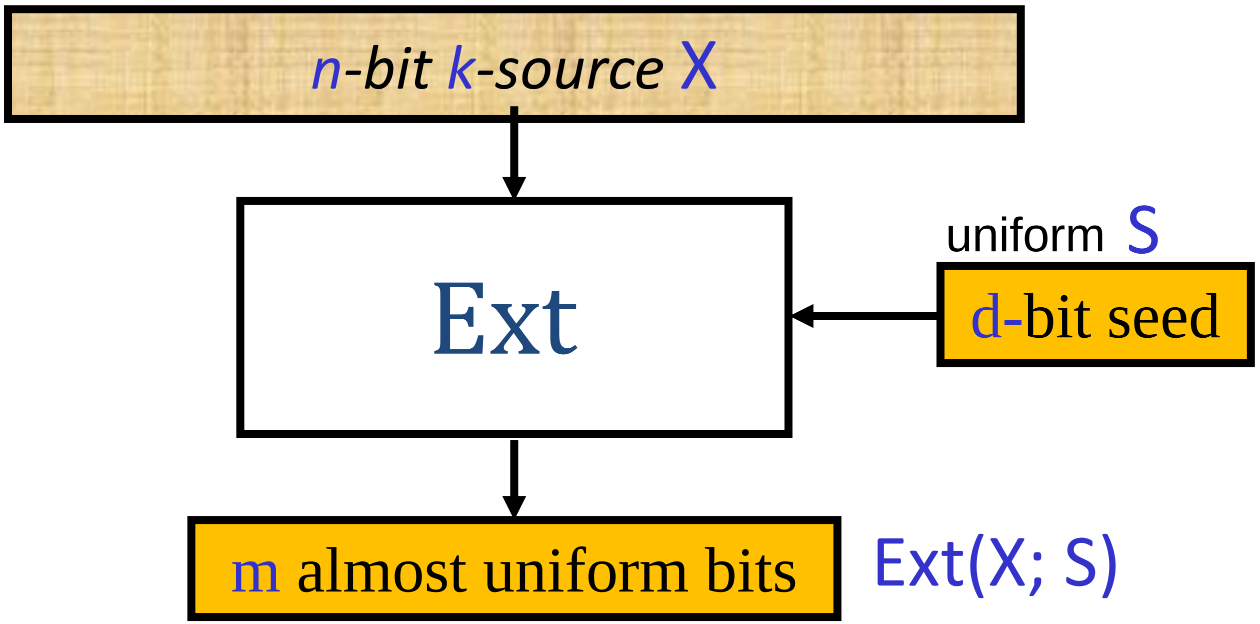
– $\mathrm{Ext}(X_b)$ = constant
– Either $H_{min}(X_0)$ or $H_{min}(X_1) \geq n\text{-}1$

- Deterministic extractor for k-source is impossible even for extracting 1 bit and even for $k = n\text{-}1$



Seeded Extractors

- Add *short* uniform seed as catalyst for extraction



Ext: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is (k,ε)-*seeded extractor if*

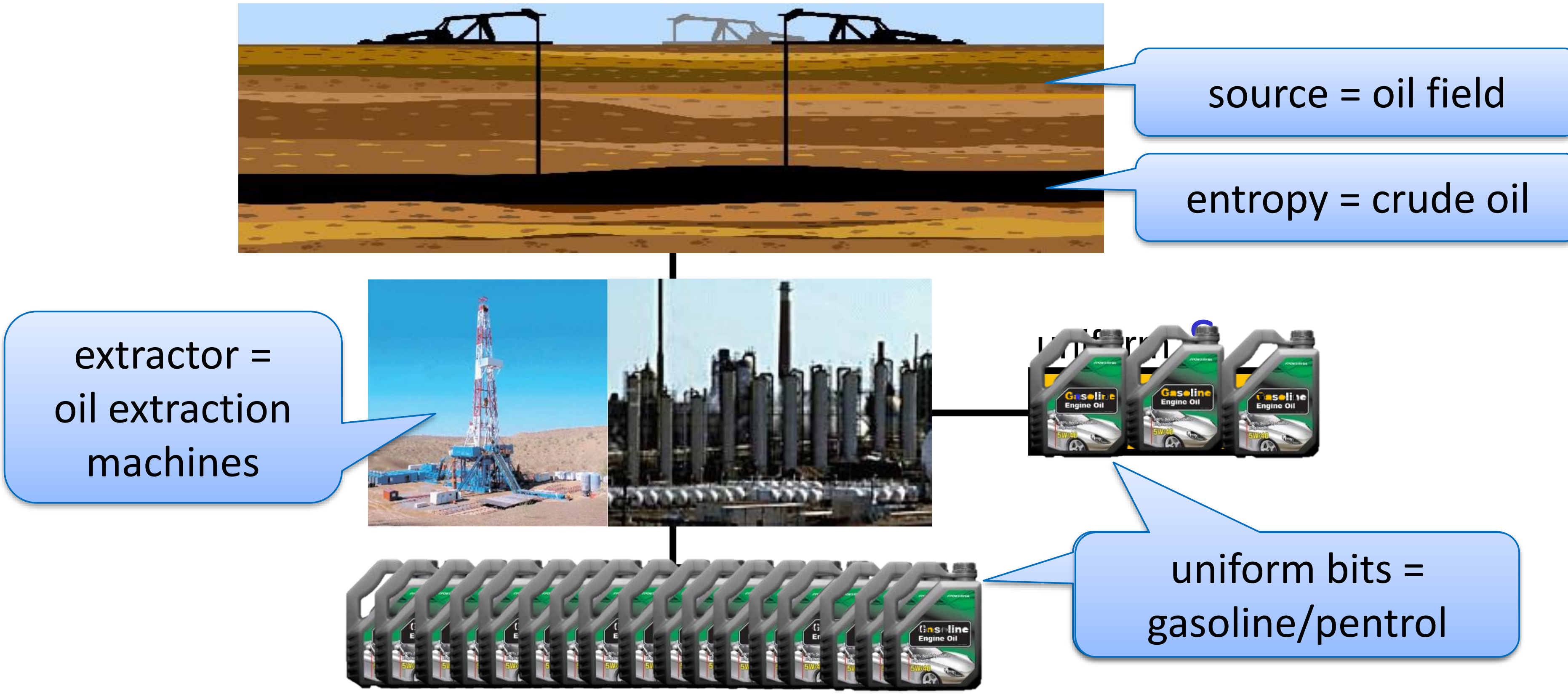
$\forall$ k-source X, Ext(X; S) is ε-close uniform U_m

Pervasive Applications

- Diverse topics in Theoretical Computer Science
  - Cryptography, Derandomization & pseudorandomness [Sis88, NZ93,…], Distributed algorithms [WZ95], Data structures [Ta02], Hardness of Approximation [Zuc93,…]

- Many applications in Cryptography
  - Privacy amplification [BBR88], Bounded-storage model [Lu02,V03], PRG [HILL89], Biometrics [DRS04], Leakage-resilient crypto [DP09]…

An Analogy: Oil Extraction



Ext: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is (k,ε)-*seeded extractor if*

$\forall$ k-source X, $\mathrm{Ext}(X; S)$ is ε-close uniform U_m

Desiderata

- Minimize seed length d
  - Minimize initial gasoline investment
- Maximize output length m, ideally close to min-entropy k
  - Extract and distill all crude oil to gasoline
- Extraction even for small entropy rate k/n
  - i.e., even when oil field has low crude oil content
- Explicit construction: efficient polynomial time extractor
  - Cost-efficiency of oil extraction machines

What We Can Achieve?

- Non-constructively, $\forall$ n, k, ε, $\exists$ (k,ε)-seeded extractor with

  seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$

  output length $m = k + d - 2\log(1/\varepsilon) - O(1)$

  – use logarithmic-length seed
  – extract almost all min-entropy out
  – for any small entropy rate
  – However, not an explicit construction
- Proof: use probability method. See Salil's book.

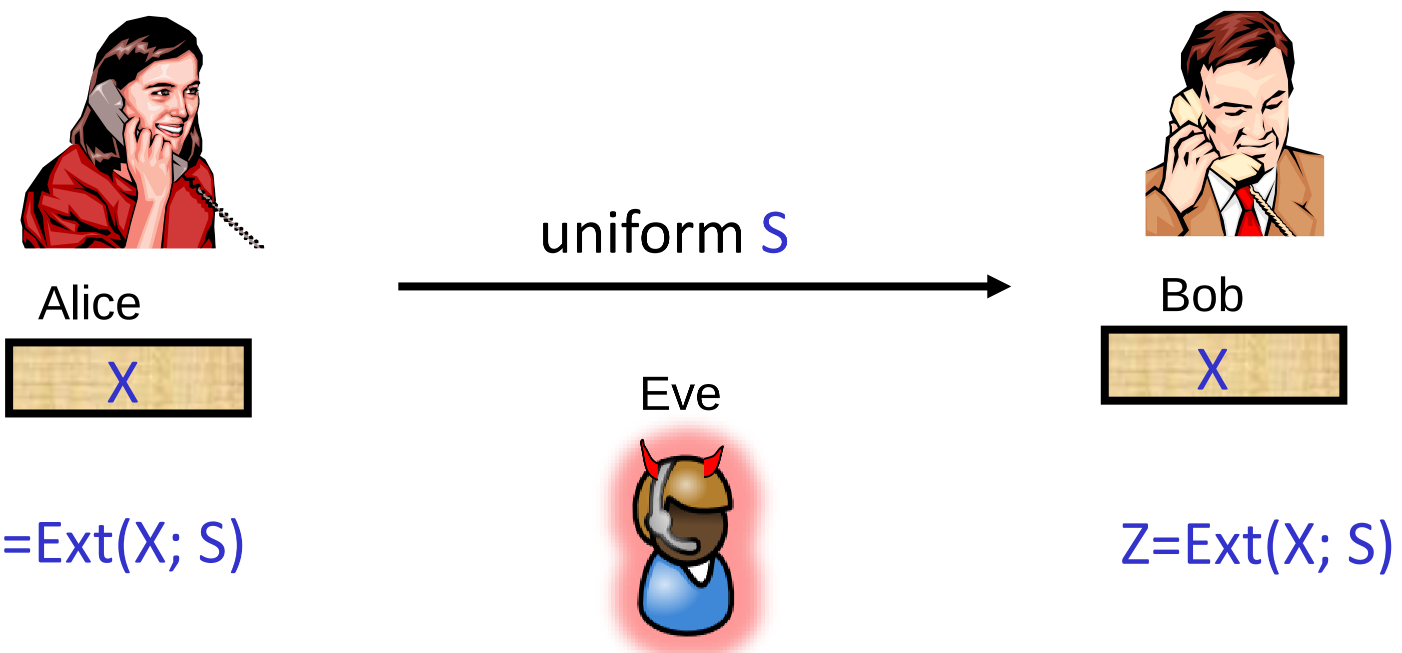
- Research goal: find explicit construction with above parameters

  seed length $d = O(\log n) + O(\log(1/\varepsilon))$

  output length $m = 0.99k$

Privacy Amplification

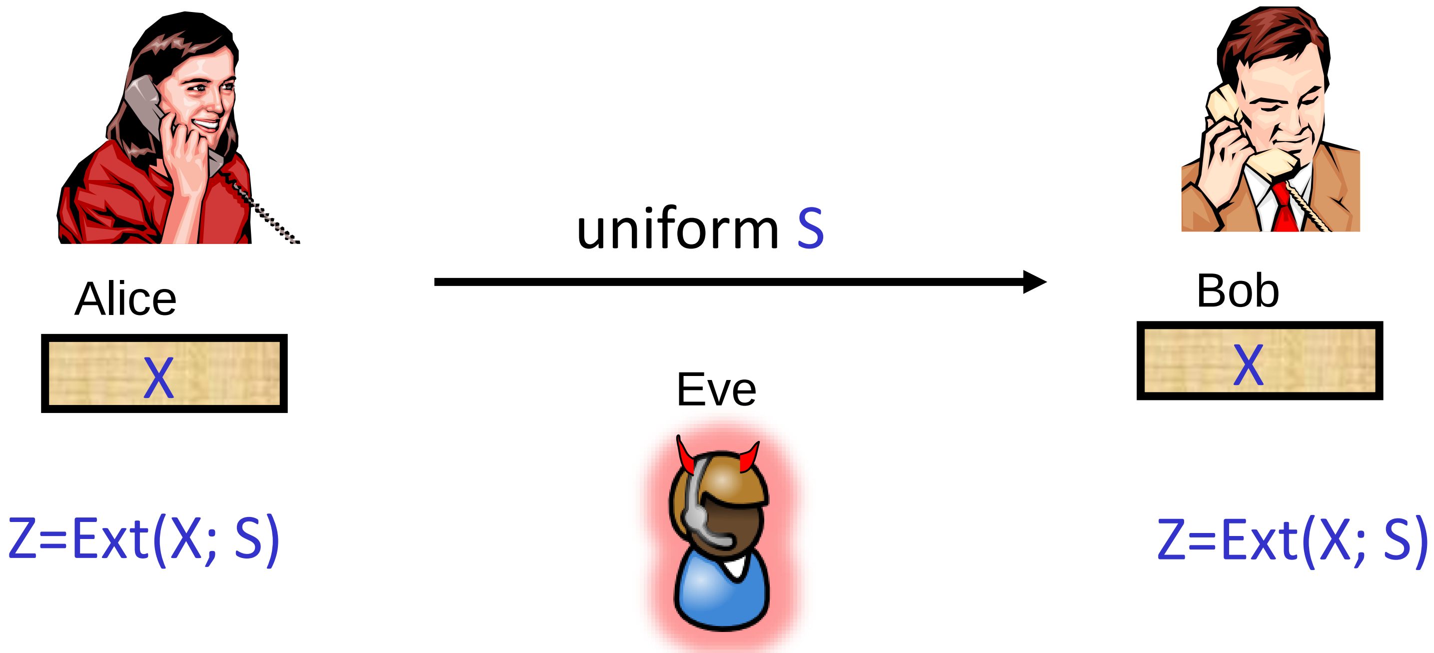
- Alice & Bob share secret weak random source X
- Goal: extract uniform key Z against eavesdropper Eve using public authenticated channel
- Issue: Eve learns seed S, may leak info about Ext(X; S)



$$\mathrm{Ext}(X; S) \approx_{\varepsilon} U_m$$

Privacy Amplification

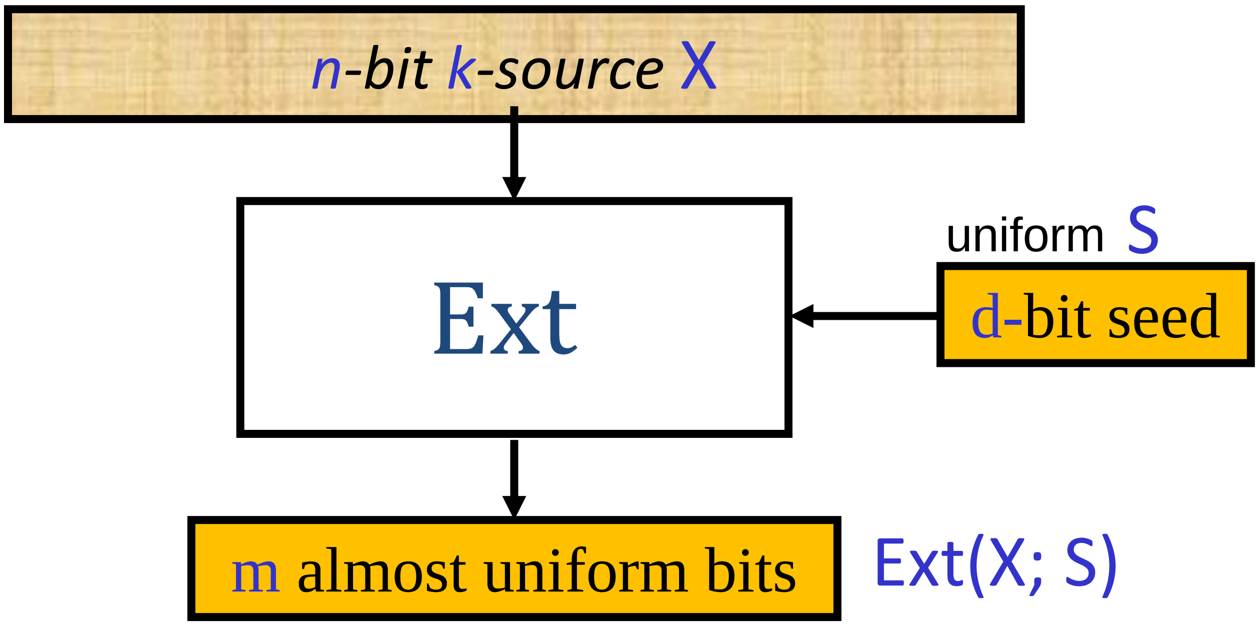
- Alice & Bob share secret weak random source X
- Goal: extract uniform key Z against eavesdropper Eve using public authenticated channel
- Issue: Eve learns seed S, may leak info about Ext(X; S)



Need: $(\mathrm{Ext}(X; S), S) \approx_{\varepsilon} (U_m, S)$

Strong Seeded Extractors

- Require Ext(X; S) close to uniform even given the seed S



Ext: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is (k,ε)-*strong seeded extractor if*

$\forall$ k-source X, $(\mathrm{Ext}(X; S), S) \approx_\varepsilon (U_m, S)$

Privacy Amplification

- Alice & Bob share secret weak random source X
- Goal: extract uniform key Z against eavesdropper Eve using public authenticated channel

Alice

X

uniform S

$\longrightarrow$

Bob

X

Eve

Z=Ext(X; S)

Z=Ext(X; S)

$$(\mathrm{Ext}(X; S), S) \approx_{\varepsilon} (U_m, S)$$

Parameters for Strong Extractors

- Non-constructively, $\forall$ n, k, ε, $\exists$ (k,ε)-seeded extractor with
  seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$
  output length $m = k + d - 2\log(1/\varepsilon) - O(1)$
  - use logarithmic-length seed
  - extract almost all min-entropy out
  - for any small entropy rate
  - However, not an explicit construction
- Proof: use probability method. See Salil's book.

- Research goal: find explicit construction with above parameters
  seed length $d = O(\log n) + O(\log(1/\varepsilon))$
  output length $m = 0.99k$

Parameters for Strong Extractors

- Non-constructively, $\forall$ n, k, ε, $\exists$ (k,ε)-strong seeded extractor with

  seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$

  output length $m = k \cancel{+ d} - 2\log(1/\varepsilon) - O(1)$
  - use logarithmic-length seed
  - extract almost all min-entropy out
  - for any small entropy rate
  - However, not an explicit construction
- Proof: use probability method. See Salil's book.

- Research goal: find explicit construction with above parameters

  seed length $d = O(\log n) + O(\log(1/\varepsilon))$

  output length $m = 0.99k$
- Strong property is usually important in crypto

An Explicit Strong Extractor --- Leftover Hash Lemma

Leftover Hash Lemma

- **Thm.** $\forall$ n, k, ε, $\exists$ efficient (k,ε)-strong seeded extractor with
  seed length $d = n$
  output length $m = k - 2\log(1/\varepsilon)$
  - use linear-length seed 
  - extract almost all min-entropy out
  - for any small entropy rate
  - explicit construction 

- Extremely useful in cryptography!

Universal Hash Functions

- Let $\mathcal{H} = \{ h: \{0,1\}^n \to \{0,1\}^m \}$ be a family of hash functions.
  - Let H denote a random hash function from $\mathcal{H}$
- **Def.** We say $\mathcal{H}$ is *universal* if for every $x \neq x' \in \{0,1\}^n$,

$$\Pr[H(x) = H(x')] \leq 2^{-m}$$

  - i.e., prob. of hash collision on x and x' is small *for every* $x \neq x'$

- Example: $\mathcal{H} = \{ h_s : s \in GF(2^n) \}$, where $h_s(x)$ = first m bits of $s \cdot x$
  - Note that $h_s(x) = h_s(x')$ implies $s \cdot (x-x') = 0^m z$ for some $z \in \{0,1\}^{n-m}$.
  - Each z determines $s = (0^m z)/(x-x')$, so at most 2^{n-m} out of 2^n h_s.
  - So $\Pr[H(x) = H(x')] \leq 2^{n-m}/2^n = 2^{-m}$.

Extractor Construction

- Let $\mathcal{H} = \{ h: \{0,1\}^n \rightarrow \{0,1\}^m \}$ be a family of hash functions.
  - Let H denote a random hash function from $\mathcal{H}$
- **Def.** We say $\mathcal{H}$ is universal if for every $x \neq x' \in \{0,1\}^n$,

$$\Pr[\, H(x) = H(x')] \leq 2^{-m}$$

  - i.e., prob. of hash collision on x and x' is small *for every* $x \neq x'$

- Define $\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \rightarrow \{0,1\}^m$ by $\mathrm{Ext}(x, h) = h(x)$
  - i.e., use seed h to select a hash function to hash x
  - need seed length $d = n$

Why Does It Work?

- Define Ext: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ by Ext(x, h) = h(x), where h is from universal hash family $\mathcal{H} = \{\, h\colon \{0,1\}^n \to \{0,1\}^m \,\}$

$$\Pr[\, H(x) = H(x')] \leq 2^{-m} \text{ for every } x \neq x' \in \{0,1\}^n$$

- Want to show $(\mathrm{Ext}(X; H), H) \approx_\varepsilon (U_m, H)$, or $(H, H(X)) \approx_\varepsilon (H, U_m)$

- Analyze via "collision probability"
  - Step 1. Z has small "collision probability" $\Longrightarrow$ Z is close to uniform
  - Step 2. Show (H, H(X)) has small "collision probability"

Collision Probability

- **Def.** Let Z be a rand. var. over $[M]$. Define *collision probability* of Z as $CP(Z) \stackrel{\text{def}}{=} \Pr[\, Z = Z' \,]$, where Z' is an independent copy of Z.
  - E.g., for uniform distribution $U_{[M]}$, $CP(U_{[M]}) = 1/M$
- View Z as vector $v \in \mathbb{R}^M$, i.e., $v_i = \Pr[\, Z = i \,]$, then $CP(Z)$ is the square of $L2$-norm of v.
  - $CP(Z) = \Pr[\, Z = Z' \,] = \sum_i \Pr[\, Z = Z' = i \,] = \sum_i v_i^2 = \|v\|_2^2$
- Intuition: uniform distribution minimize collision probability. If $CP(Z) \approx CP(U_{[M]})$, then Z is close to $U_{[M]}$
- **Lemma.** $CP(Z) \leq (1+\varepsilon)/M \Longrightarrow \Delta(Z, U_{[M]}) \leq \sqrt{\varepsilon}/2$

Small CP ⇒ Close to Uniform

Lemma. $CP(Z) \leq (1+\varepsilon)/M \Longrightarrow \Delta(Z, U_{[M]}) \leq \sqrt{\varepsilon}/2$

Proof. Define $w \in \mathbb{R}^M$ by $w_i = (v_i - 1/M)$.

- Note $\Delta(Z, U_{[M]}) = \frac{1}{2} \cdot \|w\|_1$
- Let's compute $\|w\|_2^2 = \sum_i (v_i - 1/M)^2$

  $= \sum_i v_i^2 - \sum_i (2v_i/M) + \sum_i (1/M)^2$

  $= CP(Z) - 1/M$
- Thus, $\|w\|_2^2 \leq \varepsilon/M$, or $\|w\|_2 \leq \sqrt{\varepsilon/M}$
- By relation between L1 and L2-norm $\|w\|_1 \leq \sqrt{M} \cdot \|w\|_2 \leq \sqrt{\varepsilon}$
- So $\Delta(Z, U_{[M]}) \leq \sqrt{\varepsilon}/2$

CP(H, H(X)) is Small

Lemma. $CP(H,H(X)) \leq (1/D)\cdot((1/M)+(1/K))$

- Notation: $D = 2^d$, $M = 2^m$, $K = 2^k$

Proof. $CP(H,H(X)) = \Pr[\, (H,H(X)) = (H', H'(X'))]$

$$= \Pr[H = H']\cdot\Pr[H(X) = H(X') \,|\, H = H']$$

$$= (1/D)\cdot(\, \Pr[X=X']\cdot\Pr[H(X) = H(X') \,|\, H = H' \wedge X=X'] +$$

$$\Pr[X\neq X']\cdot\Pr[H(X) = H(X') \,|\, H = H' \wedge X\neq X']\,)$$

$$\leq (1/D)\cdot(\, CP(X) + (1/M))$$

- $CP(X) = \sum_x \Pr[X = x]^2 \leq (\max_x \Pr[X = x])\,(\sum_x \Pr[X = x]) \leq 1/K$

Put Things Together

- **Lemma.** $CP(Z) \leq (1+\varepsilon)/M \Longrightarrow \Delta(Z, U_{[M]}) \leq \sqrt{\varepsilon}/2$
- **Lemma.** $CP(H,H(X)) \leq (1/D)\cdot((1/M)+(1/K))$
- Recall we set $m = k - 2\log(1/\varepsilon)$, so $(1/K) = (\varepsilon^2/M)$
- So $\Delta((H,H(X)), (H,U_m)) \leq \varepsilon/2$

Thm. $\forall$ n, k, ε, $\exists$ efficient (k,ε)-strong seeded extractor with

seed length $d = n$

output length $m = k - 2\log(1/\varepsilon)$

Average-case Extractors

Privacy Amplification

- Alice & Bob share secret weak random source X
  - Since Eve may learn some leakage information E about X
- Goal: extract uniform key Z against eavesdropper Eve using public authenticated channel

uniform S

Alice

X

Bob

X

Eve

E

Z=Ext(X; S)

Z=Ext(X; S)

Need: (Ext(X; S), S, E) $\approx_{\varepsilon}$ (U_m, S, E)

Conditional Min-Entropy

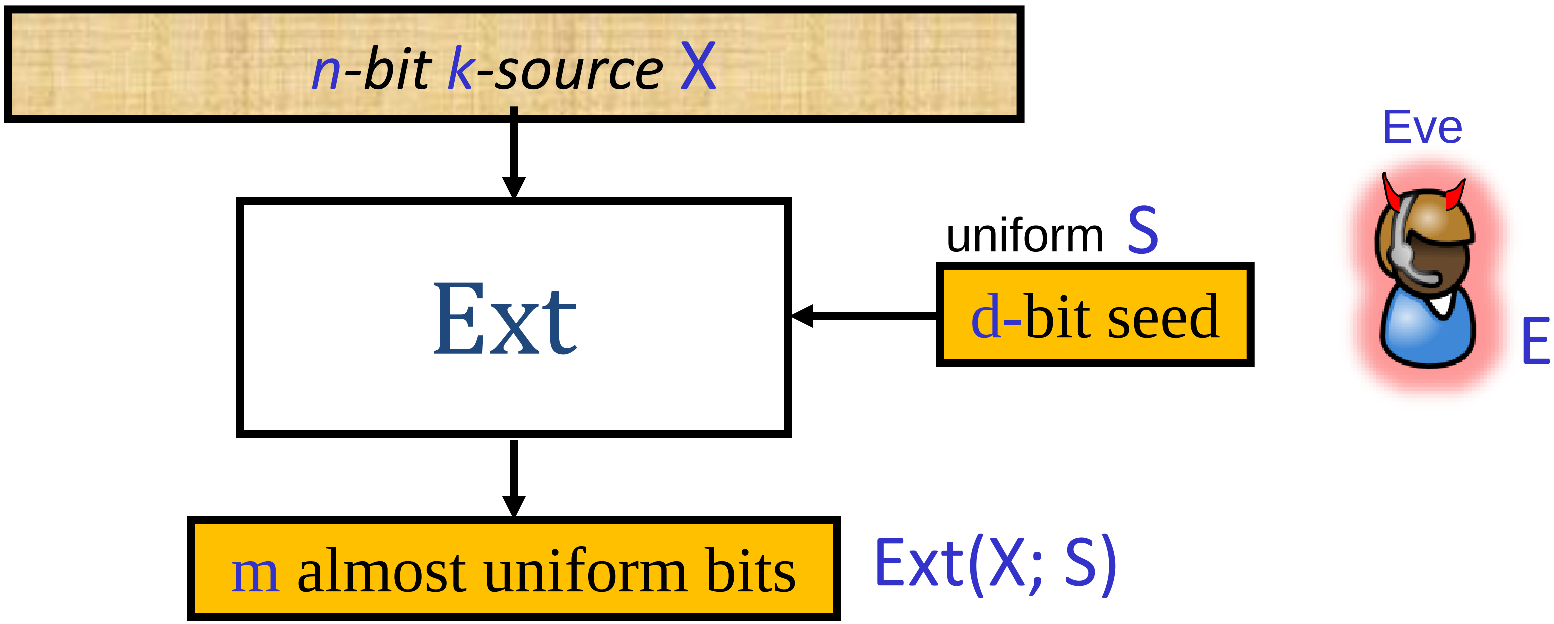
- How to measure min-entropy of X given side information E?
- Guessing Probability: $P_{guess}(X|E)$

$$P_{guess}(X|E) \stackrel{\text{def}}{=} \max \Pr[\text{ guess } X \text{ correctly given } E\ ]$$

- Conditional Min-Entropy: $H_{min}(X|E) \stackrel{\text{def}}{=} \log 1/P_{guess}(X|E)$
- Sanity check: $P_{guess}(X) = \max_x \Pr[X=x]$, so $H_{min}(X) = \log 1/P_{guess}(X)$
- In general: $P_{guess}(X|E) = \mathrm{E}_{e \leftarrow E}[\ \max_x \Pr[X=x|E=e]]$
- Conditional min-entropy ≈ unpredictability of the source given E

Average-Case Strong Extractors

- Extract conditional min-entropy from X given E



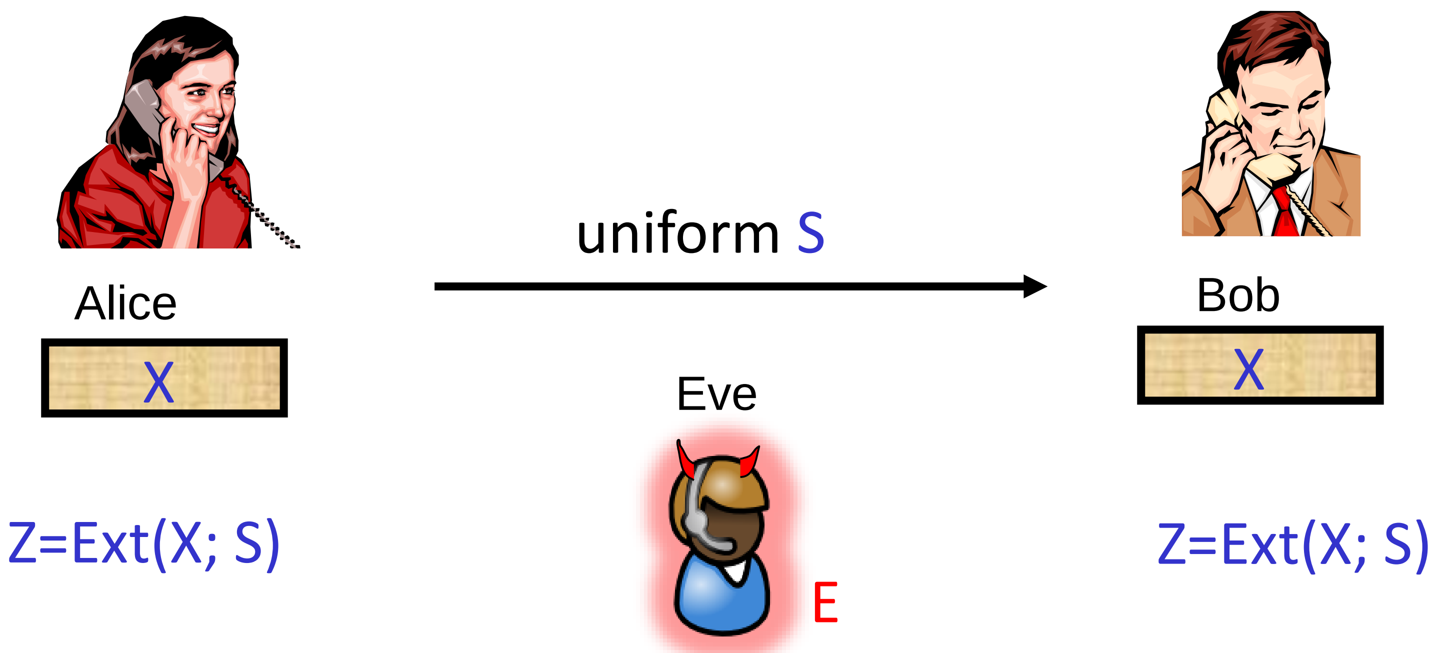
Ext: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is (k,ε)-*average-case strong extractor*

if $\forall$ (X, E) with $H_{min}(X|E) \geq k$,

$$(\mathrm{Ext}(X; S), S, E) \approx_{\varepsilon} (U_m, S, E)$$

Privacy Amplification

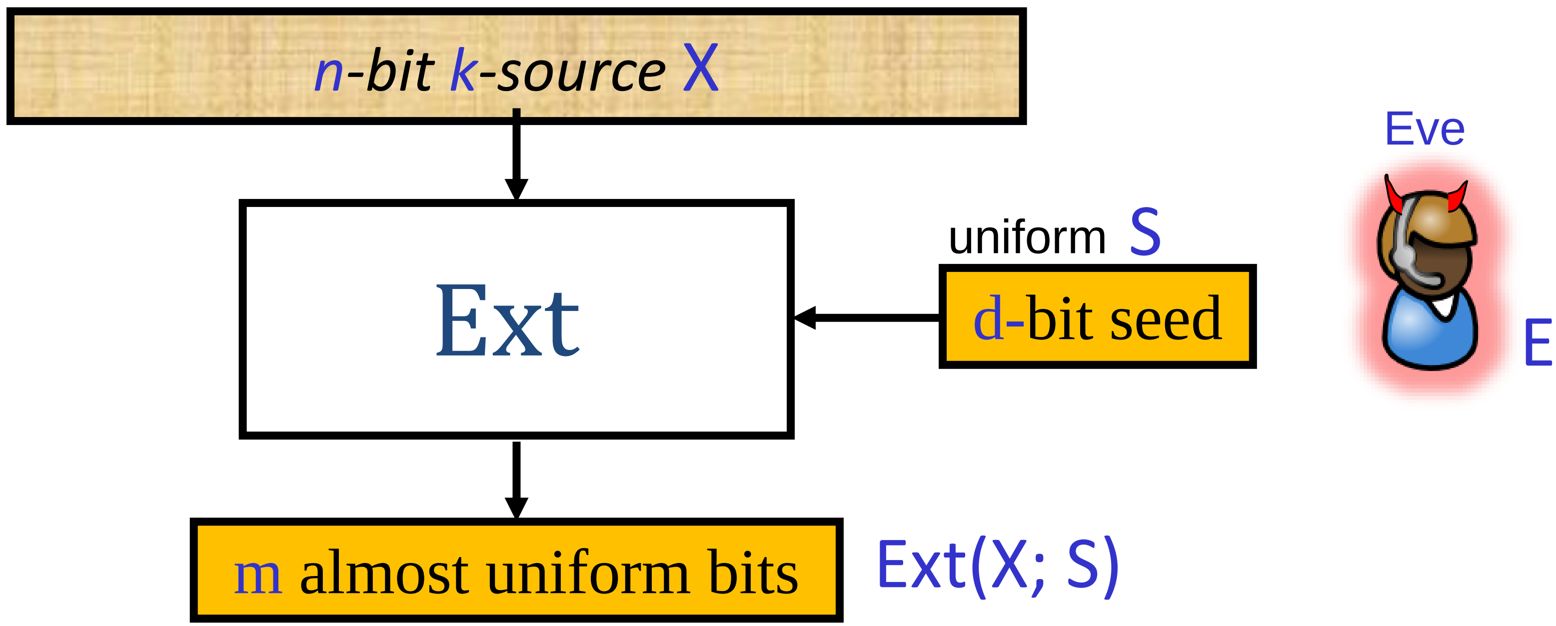
- Alice & Bob share secret weak random source X
  - Since Eve may learn some leakage information E about X
- Goal: extract uniform key Z against eavesdropper Eve using public authenticated channel



$$(\mathrm{Ext}(X; S), S, E) \approx_{\varepsilon} (U_m, S, E)$$

Interpretation

- Conditional min-entropy ≈ unpredictability
- Statistical distance ≈ distinguishing advantage



- Extractor: distill unpredictability to indistinguishability
  - Can't predict source ⟹ can't distinguish output from uniform

Every Extractor is Average-Case Ext.

Thm. If $\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is (k,ε)-strong extractor, then Ext is $(k+\log(1/\varepsilon), 2\varepsilon)$-average-case strong extractor.

Lemma. If in (X, E), X has conditional min-entropy k conditioned on E, then w.p. $1-\varepsilon$ over $e \leftarrow E$, $X|_{E=e}$ is a $(k-\log(1/\varepsilon))$-source.

Lemma $\Rightarrow$ **Thm**:

- Ext works for good $X|_{E=e}$ with error ε
- Ext may fail on bad $X|_{E=e}$ but $X|_{E=e}$ bad w.p. at most ε

$\Rightarrow$ Ext works for (X, E) with error $\leq 2\varepsilon$

Lemma. If in (X, E), X has conditional min-entropy k conditioned on E, then w.p. $1-\varepsilon$ over $e \leftarrow E$, $X|_{E=e}$ is a $(k\text{-}\log(1/\varepsilon))$-source.

Proof. Suppose not, i.e.,

w.p. $> \varepsilon$ over $e \leftarrow E$,

$$H_{min}(X|_{E=e}) \leq k - \log(1/\varepsilon)$$

$$\Rightarrow P_{guess}(X|_{E=e}) \geq 2^k/\varepsilon$$

$\Rightarrow P_{guess}(X|E) > \varepsilon \cdot 2^k/\varepsilon > 2^k$

$\Rightarrow H_{min}(X|E) < k$, a contradiction.

In Fact, Can Do Better!

Leftover hash lemma:

Thm. $\forall$ n, k, ε, $\exists$ efficient (k,ε)-average-case strong extractor with

seed length $d = n$

output length $m = k - 2\log(1/\varepsilon)$

– Use "conditional collision probability" in analogous way

In general:

Thm. Any (k,ε)-strong extractor is a $(k,3\varepsilon)$-average-case strong extractor

Summary

- Conditional min-entropy ≈ unpredictability
- Statistical distance ≈ distinguishing advantage
- Extractor: distill unpredictability to indistinguishability
  - Oil extraction analogy
  - Features: strong, average-case, “quantum-proof”, “non-malleable”

- Non-constructively, $\forall$ n, k, ε, $\exists$ (k,ε)-strong extractor with
  seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$
  output length $m = k - 2\log(1/\varepsilon) - O(1)$

Summary

- Leftover hash lemma: $\forall$ n, k, ε, $\exists$ explicit (k,ε)-extractor with
  seed length $d = n$
  output length $m = k - 2 \log(1/\varepsilon)$
  - Collision prob.: useful way to bounds distance to uniform

- Best-known explicit construction
  seed length $d = O(\log n) + O(\log(1/\varepsilon))$
  output length $m = 0.99k$

Quantum-Proof Extractors

Privacy Amplification

- Alice & Bob share secret weak random source X
- Goal: extract uniform key Z against eavesdropper Eve using public authenticated channel
- What if the side information E is quantum?

Alice

X

uniform S

Bob

X

Eve

E

Z=Ext(X; S)

Z=Ext(X; S)

Need: (Ext(X; S), S, E) $\approx_{\varepsilon}$ (U_m, S, E) for quantum E

How to Think about Quantum?

- Some physical resource generalizes classical world and is sometime more powerful
  - Quantum information: generalize classical information and sometimes more useful
  - Quantum computation: generalize classical computation and sometimes much more powerful! (e.g., Shor’s algorithm)

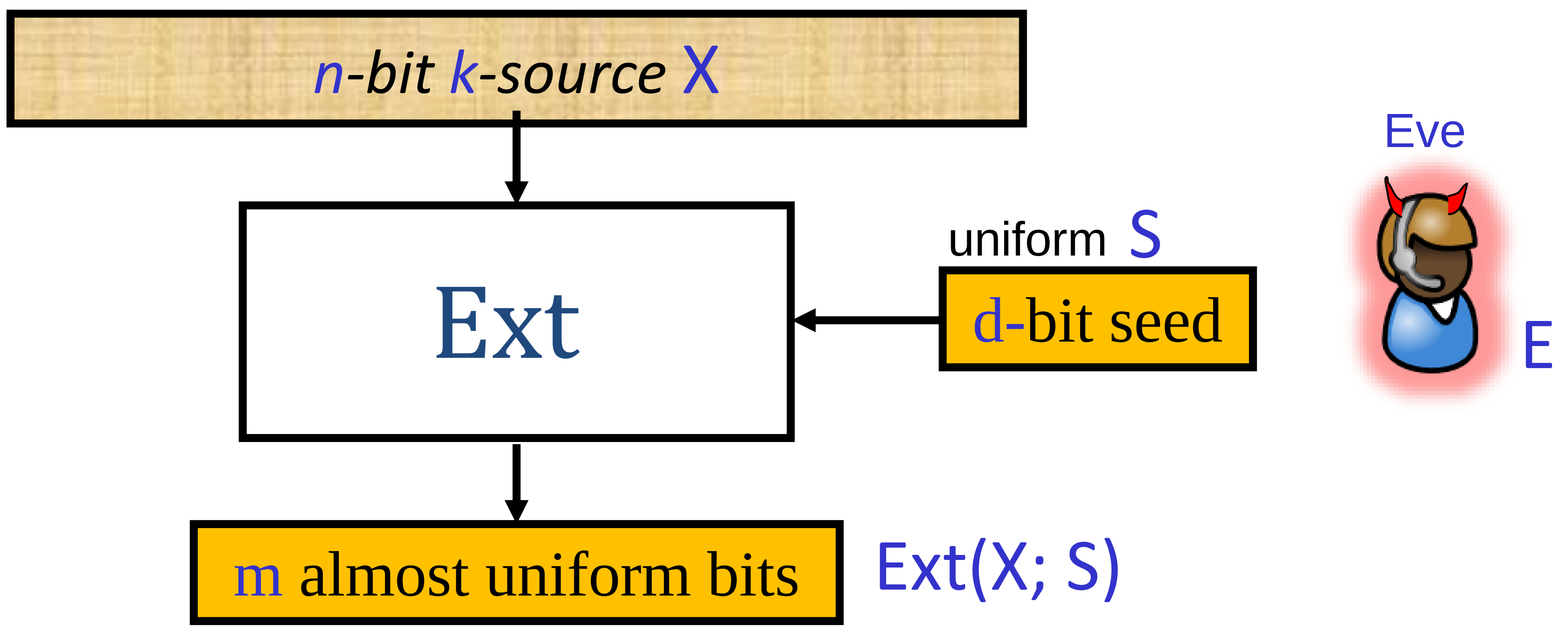
- Randomness extraction in presence of quantum side information
  - Harder task since Eve holds more useful information
  - Operational definition generalizes

Operational Definitions Generalize

- Entropy measure: conditional min-entropy
  - Cond. Min-entropy: $H_{min}(X|E) = \log 1/P_{guess}(X|E)$, where
  - Guessing Probability:

    $P_{guess}(X|E) \stackrel{\text{def}}{=} \max \Pr[$ guess X correctly given $E\ ]$
  - Min-entropy ≈ unpredictability
- Distance measure: trace distance
  - Trace distance ≈ max distinguishing advantage
- Extractor: distill unpredictability to indistinguishability
  - Can't guess source ⟹ can't distinguish output from uniform
  - $(\mathrm{Ext}(X; S), S, E) \approx_\varepsilon (U_m, S, E)$ for quantum E

Quantum-Proof Strong Extractors



Ext: $\{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is (k,ε)-*quantum-proof strong extractor*

if $\forall$ (X, E) with $H_{min}(X|E) \geq k$,

$$(Ext(X; S), S, E) \approx_{\varepsilon} (U_m, S, E)$$ for quantum E

What Remains True in Quantum?

- Conditional min-entropy ≈ unpredictability
- Statistical distance ≈ distinguishing advantage

- Extractor: distill unpredictability to indistinguishability
  - Oil extraction analogy

- Non-constructively, $\forall$ n, k, ε, $\exists$ (k,ε)-strong extractor with
  seed length $d = \log(n-k) + 2\log(1/\varepsilon) + O(1)$
  output length $m = k - 2\log(1/\varepsilon) - O(1)$

What Remains True in Quantum?

- Leftover hash lemma: $\forall$ n, k, ε, $\exists$ explicit (k,ε)-extractor with
  seed length $d = n$
  output length $m = k - 2 \log(1/\varepsilon)$
  - Collision prob.: useful way to bounds distance to uniform

- Best-known explicit construction
  seed length $d = O(\log n) + O(\log(1/\varepsilon))$
  output length $m = 0.99k$