Correlation Extractors and Their Applications

Yuval Ishai
Technion

Based on joint work with
Eyal Kushilevitz
Rafail Ostrovsky
Amit Sahai

What this talk is about

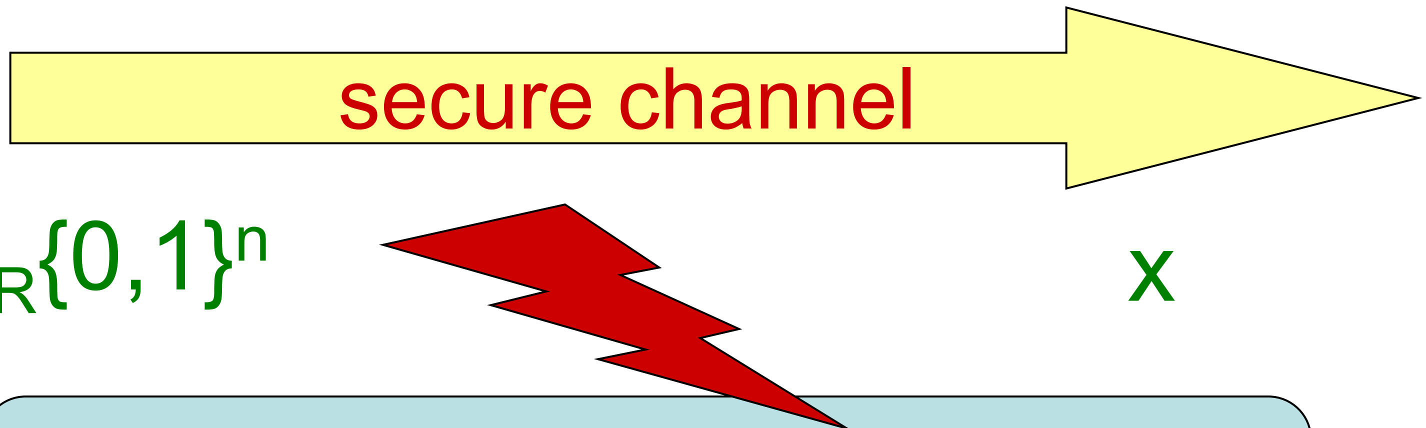
- Extension of randomness extraction and privacy amplification to correlated sources
- Motivated by cryptographic applications

- …but also think about communication channels:
  - Cleaning channels
  - Converting one channel to another
  - Building channels from scratch

Privacy Amplification

[BBR88,BBCM95,...]

Alice Bob



$x \in_R \{0,1\}^n$ x

Eve

What if x is not uniform?

t-dirty: min-entropy n-t

What if x is partially leaked?

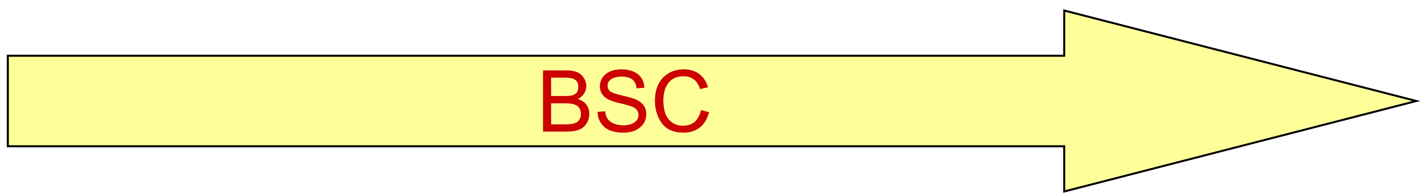
t-leaky: Eve learns f(x), $f:\{0,1\}^n \rightarrow \{0,1\}^t$

- Solved by randomness extractors [NZ96]
  - Alice picks a fresh seed s and sends to Bob over a public channel
  - Both parties output Ext(s,x)

Cleaning other types of channels?

- Noise is useful for crypto! [Wyn75,Csi81,…, CK88,…]

Alice   Bob



$x \in_R \{0,1\}^n$ $y = x \oplus e$

$e_i \sim \mathrm{Bern}(p)$

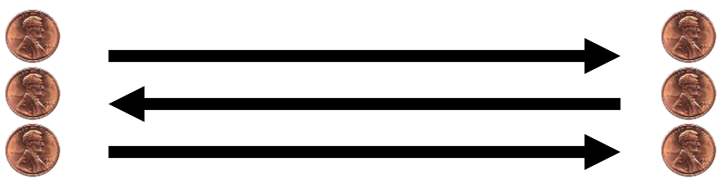
- Noise can be "dirty" or "leaky"
- Can we build a clean BSC from a dirty BSC?
  - Main challenge: protecting against insiders

Correlation Extractors

- Generalize BSC example to any "channel" (X,Y)
- (n,m,t,ε) correlation extractor for (X,Y):

Alice Bob

a' b' (a',b') ← t-dirty $(X,Y)^n$



a b (a,b) ε-close to $(X,Y)^m$

also from point of view of Alice or Bob!

- Classical case: $X=Y \in_R \{0,1\}$

Main Question

- Are there correlation extractors for arbitrary (X,Y)?
  - If so, how good can they be?

- Question largely unexplored
  - Different from previous extensions of privacy amplification to correlated or “fuzzy” sources [Wyn75,BBR88,Mau91,DRS04,DS05,…]

    Only concerned with secrecy against an external Eve
  - Special cases implicit in literature
    - Special types of correlations, locally imperfect sources
    - No prior study of global imperfections

Main Question

- Are there correlation extractors for arbitrary (X,Y)?
  - If so, how good can they be?

- Question still seems challenging even when
  - allowing non-explicit or heuristic constructions
  - allowing unlimited access to fresh randomness, secure communication

- Source of difficulty:
  Conflict between "structure" and "secrecy"

Randomness extraction meets secure computation

Main Result

[I-Kushilevitz-Ostrovsky-Sahai 2009]

- For any finite (X,Y) there is an efficient, constant-ro… (n,m,t,ε) correlation extractor with:
  - m=Ω…
  - t= Ω… …ge]
  - ε = 2…
  - O(n) communication

constant support size,
rational probabilities

- Assumes semi-honest parties.

Simple Correlations

Very useful for crypto!

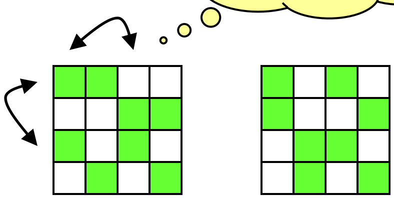
- easy conversion to “chosen input” OTs

  [BG89,Bea95]

- basis for general secure two-party computation
  - requires O(circuit-size) instances of channel

    [GMW87,GV87,GHY87,Kil88,…]

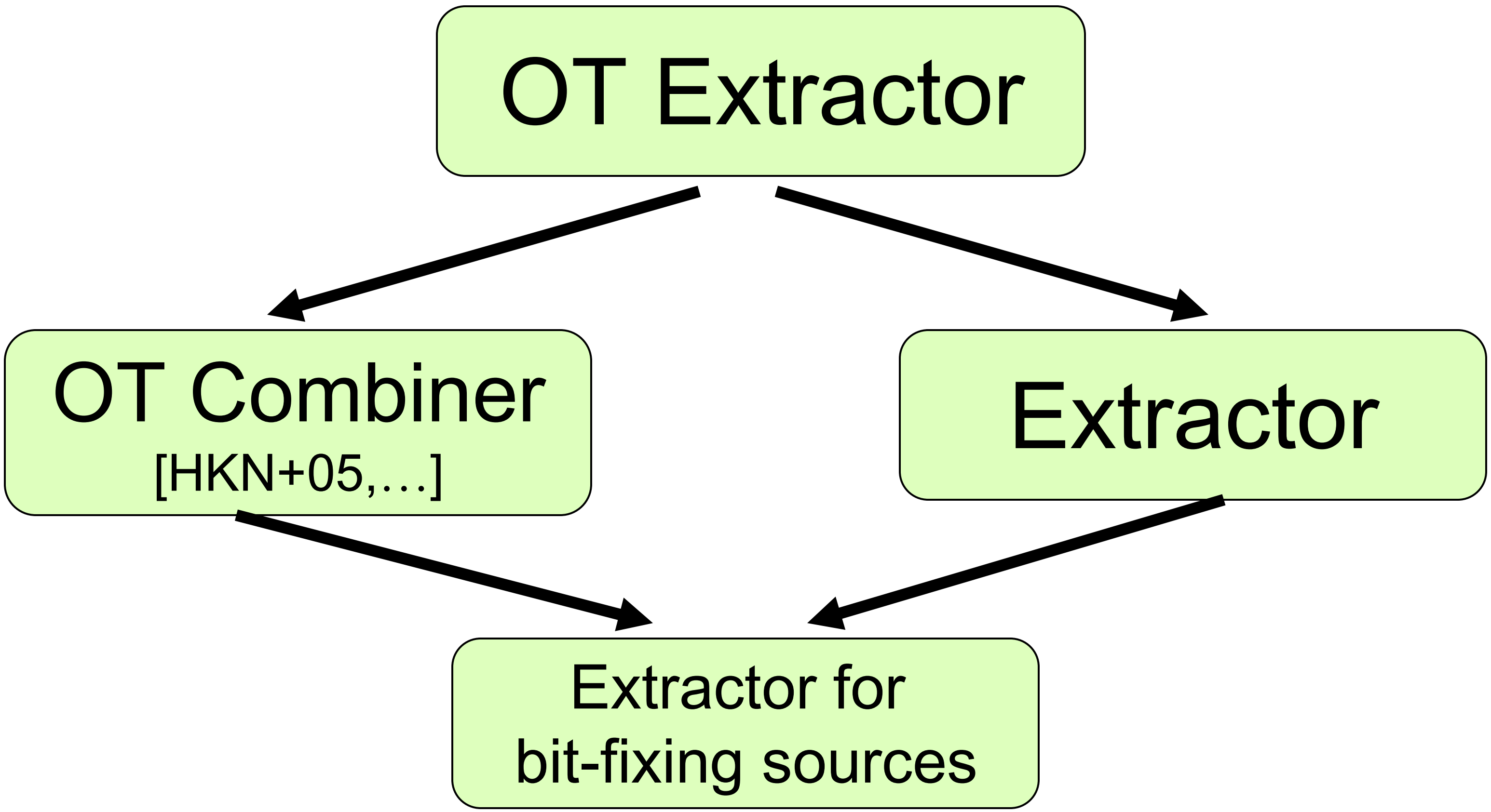
[WW06]



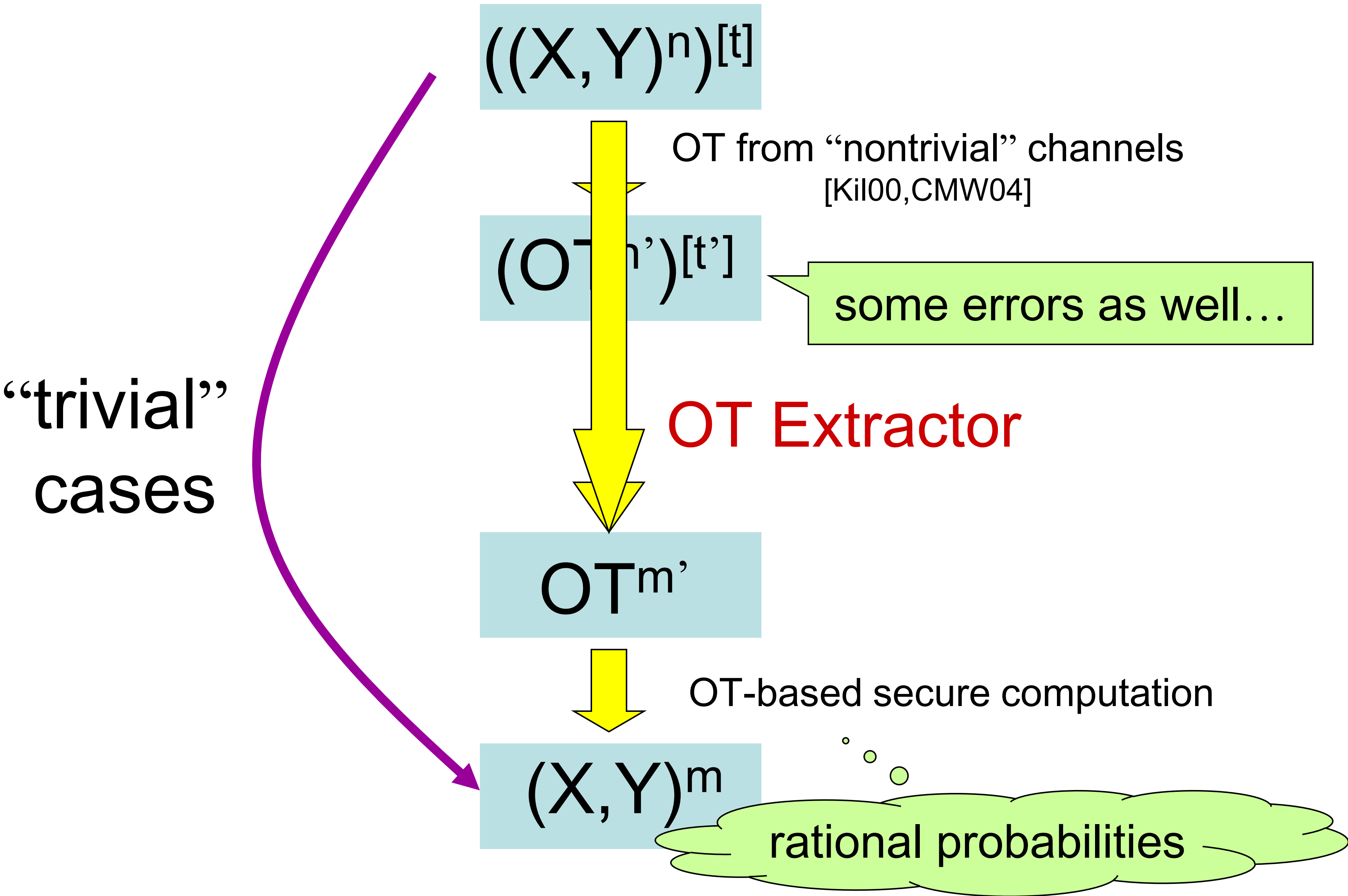
(random) OT channel

OT Extractor

- Building block for general correlation extractors
- Common generalization of previous primitives



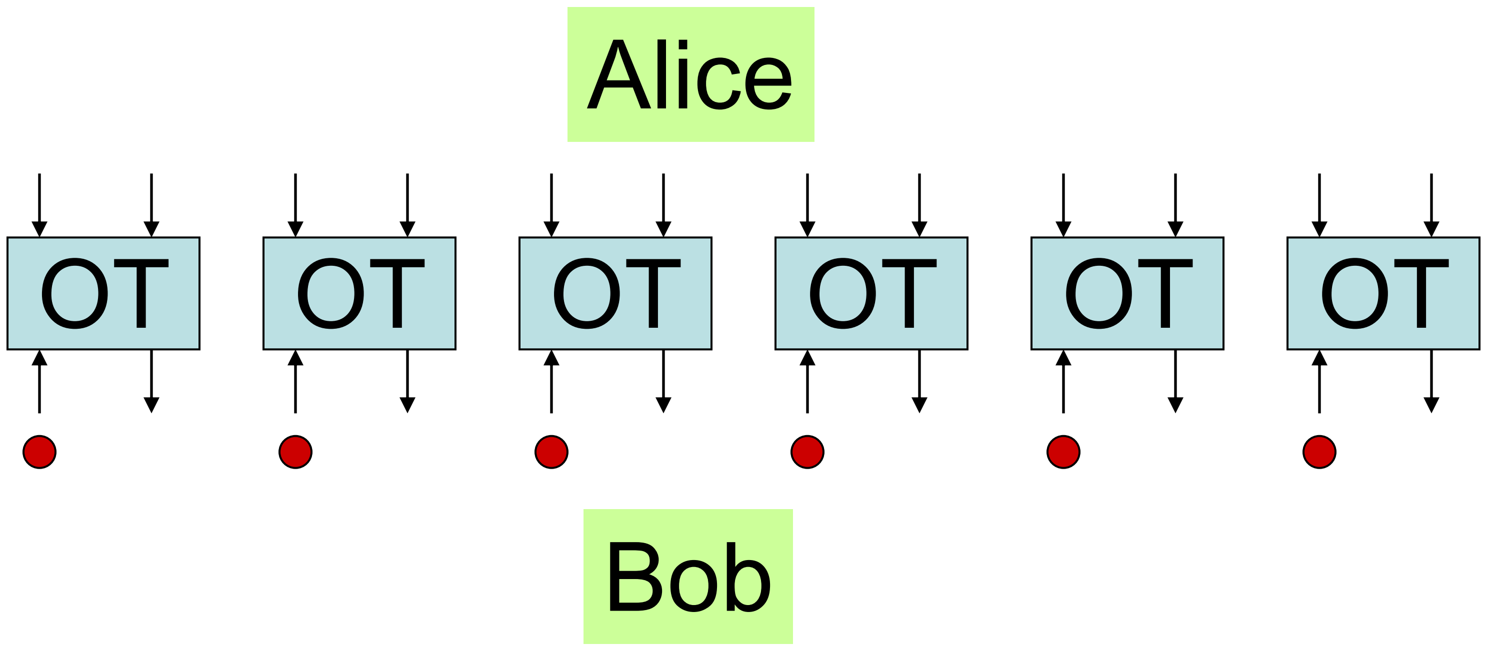
Overview of Construction



Efficient OT Extractors

- Careful combination of secure computation and randomness extraction techniques
  - Simpler with polylog(n,1/ε) loss in m,t
- Idea: Use O(m) “leaky” OTs as a resource for securely computing m fresh OTs.
- Problem: OT-based protocols propagate leakage!
  - Modify computed function to include an extraction step?
  - Leakage still propagates…
- Observation: random OTs are converted into “chosen input” OTs via XORing.

ε-biased secure computation



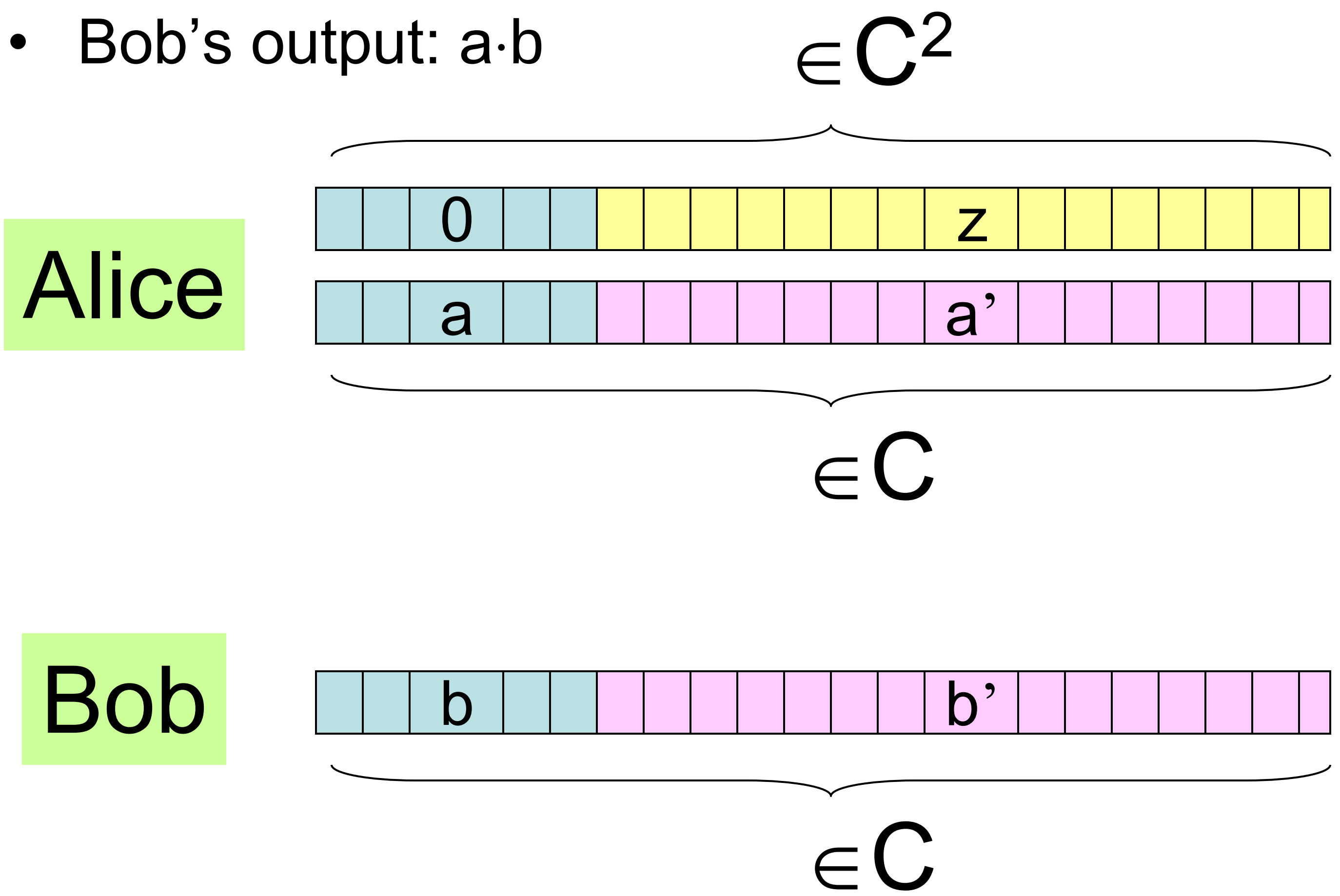
- Goal: Generate m "fresh" OTs using O(m) calls to an OT oracle while making Bob's oracle inputs ε-biased
- Masking with outputs of leaky oracle will keep Bob's fresh OT selections private [AR94,GW97]
- Need to reverse & repeat the process for protecting Alice.

Building Block

- Explicit family of linear codes $C_n : F^{k(n)} \to F^n$ such that
  - F has characteristic 2
  - The dual distance of C_n is $\Omega(n)$
  - The linear code C^2_n spanned by pointwise products of $c_i, c_j \in C_n$ has minimal distance $\Omega(n)$
- Examples:
  - RS codes (non-constant F) [BGW88,…]
  - AG codes (constant F) [CC06, CCX11]
- Can't use random codes (even non-explicitly)
  - last requirement implies efficient decoding [CDG+05]

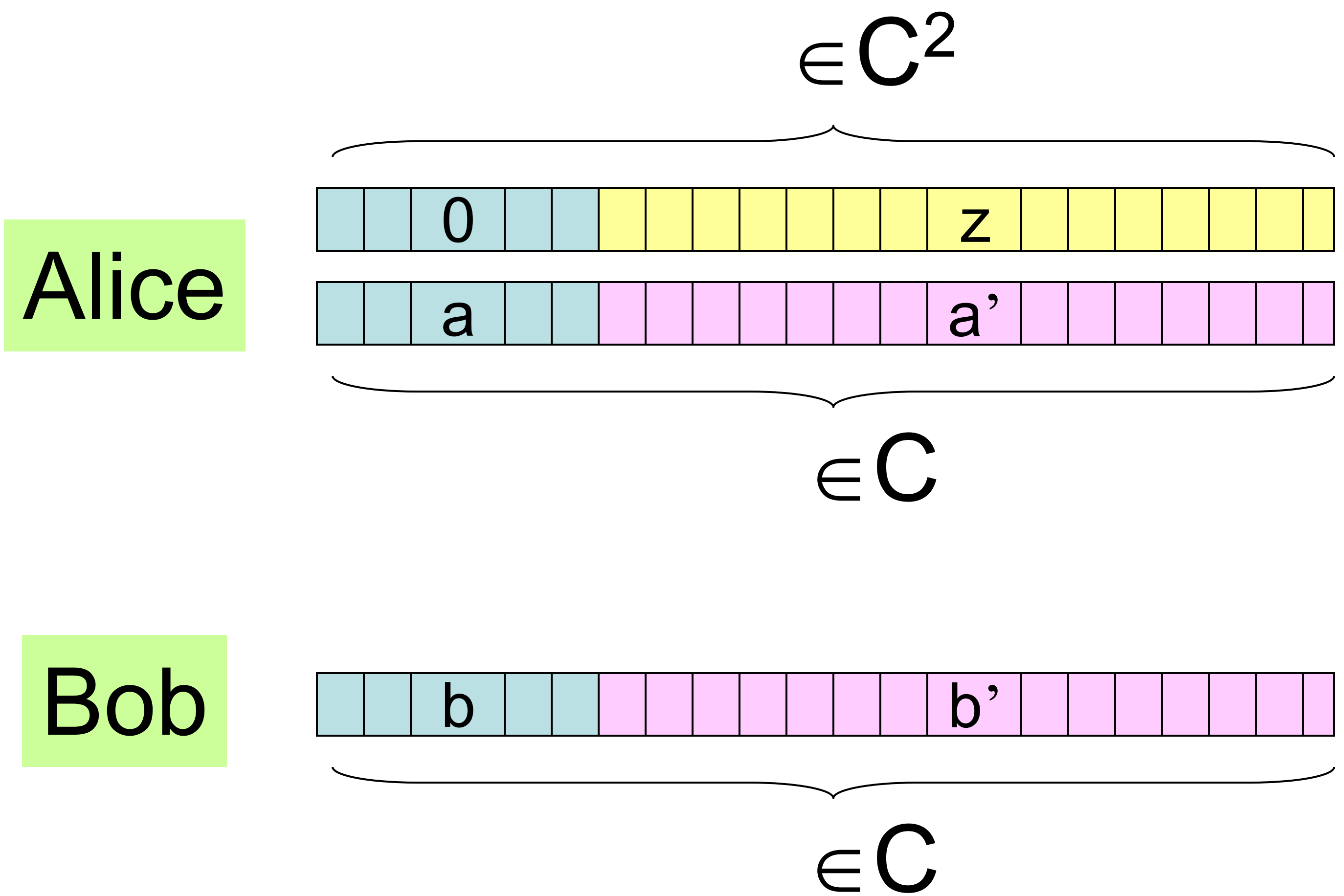
ε-biased protocol for AND^m

- Alice's input: $a \in \{0,1\}^m$
- Bob's input: $b \in \{0,1\}^m$
- Bob's output: $a \cdot b$



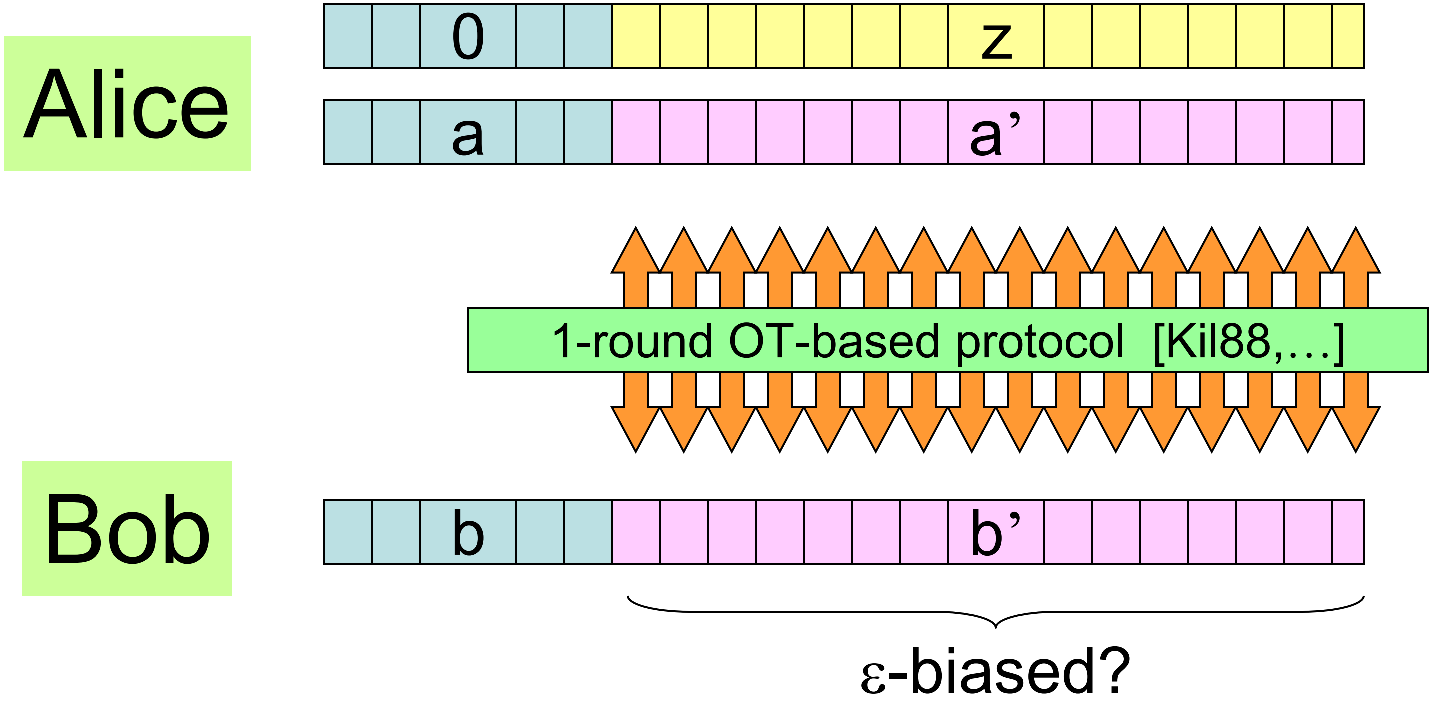
ε-biased protocol for AND^m

- $a'\cdot b'+z$ is the suffix of a random codeword from C^2 which starts with $a\cdot b$ ➔ reveals no info beyond $a\cdot b$



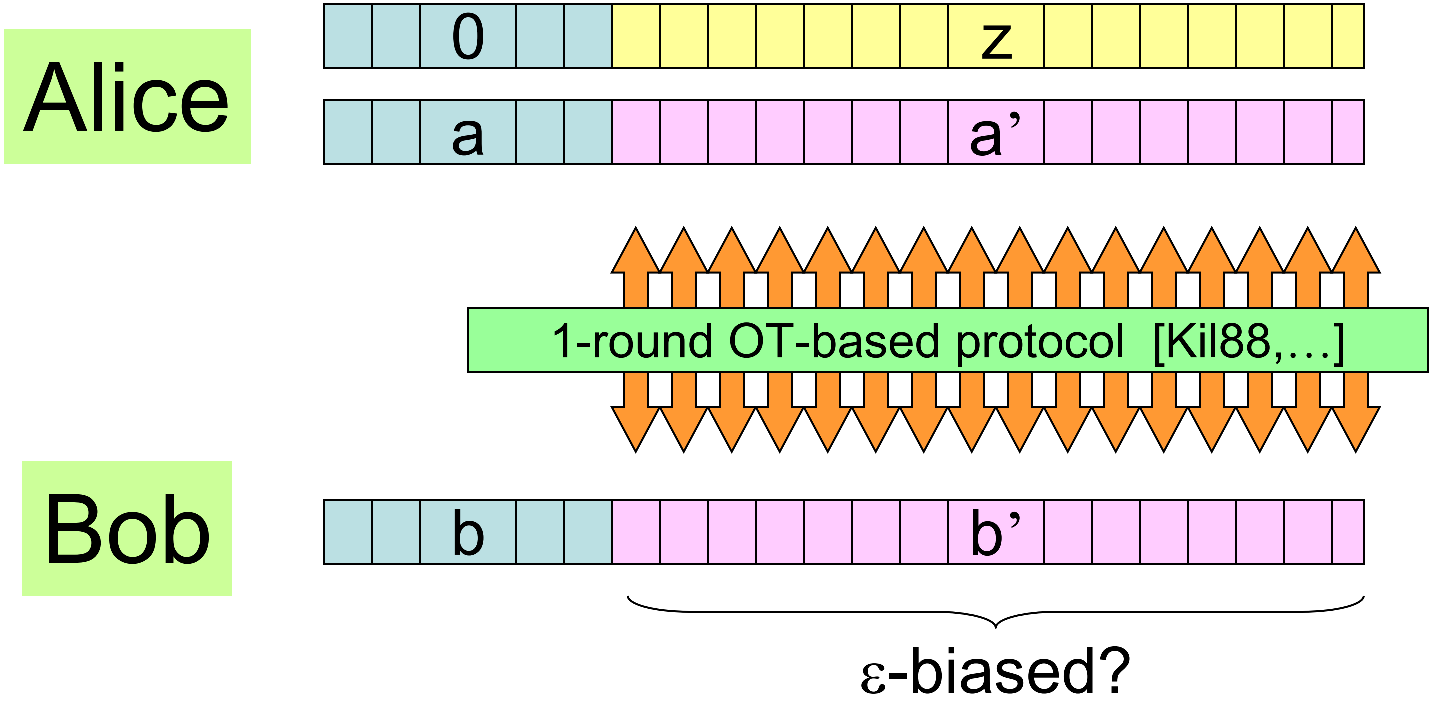
ε-biased protocol for AND^m

- $a' \cdot b' + z$ is the suffix of a random codeword from C^2 which starts with $a \cdot b$ ➔ reveals no info beyond $a \cdot b$
  - Good distance of C^2 guarantees that $a \cdot b$ can be recovered



ε-biased protocol for AND^m

- Good dual distance of C, $|F|=2^c$ ➔ b' is $\Omega(m)$-wise independent
  - But not ε-biased!
- Apply random 3-bit majority encoding to each bit of b'
  - Makes b' ε-biased with $\varepsilon=2^{-\Omega(m)}$
  - Incorporate decoding into OT-based secure computation protocol



Applications

- Protecting protocols against leakage
- Efficient reductions between channels
- Communication-efficient secure computation

protecting against leakage

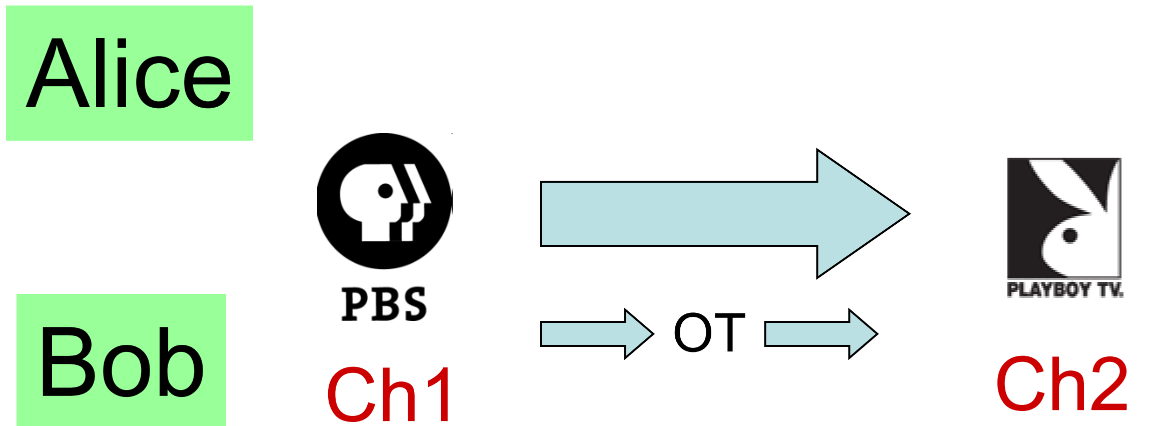
OT generation process

leaky storage

OT Extractor

OT-based protocol

efficient reductions between channels



- Much work on OT from noisy channels
  - BSC, “unfair” channels, Gaussian channels, …
  - poly(k) invocations of Ch1 per OT instance, even in semi-honest model
- OT extractors ➔ constant-rate OTs from any nontrivial channel
  - Bonus feature: leakage-resilience

communication-efficient secure computation

- Secure two-party computation, standard model
- Communication of typical protocols: poly(k) per gate
- [Gentry09]: poly(k) (|input|+|output|) overall!
- But… sometimes life is a sequence of finite tasks
  - circuit of size O(|output|)
  - even [Gentry09] requires poly(k) communication per gate
- Application of OT extractors
  - Constant-rate OT protocol under Θ-Hiding Assumption [CMS99,GR05]

    → general circuit evaluation with O(1) bits per gate

    → constant-rate realization of any discrete channel!
  - Previously known under a nonstandard assumption [IKOS08]

Conclusions

- Defined correlation extractors
- Constructed (n,m,t,ε) extractor for every finite (X,Y)
  - $m=\Omega(n)$
  - $t=\Omega(n)$
  - $\varepsilon = 2^{-\Omega(n)}$
  - $O(n)$ communication
- Several applications, all with “constant rate”
  - Cleaning channels
  - Reducing channels to each other
  - Building channels from scratch!
    - Computationally, under Θ-hiding assumption

Further Research

- Better parameters
  - Maximize leakage resilience and rate
  - Minimize round complexity
  - Better dependence on domain size?
- Malicious parties
- Multi-party setting
- Computational setting
  - Protecting computationally-secure two-party protocols against side-channel attacks