# Low-Error Two-Source extractors from efficient non-malleable extractors

DEAN DORON
TEL-AVIV UNIVERSITY

Joint work with AVRAHAM BEN-AROYA ESHAN CHATTOPADHYAY XIN LI AMNON TA-SHMA



# Today's talk

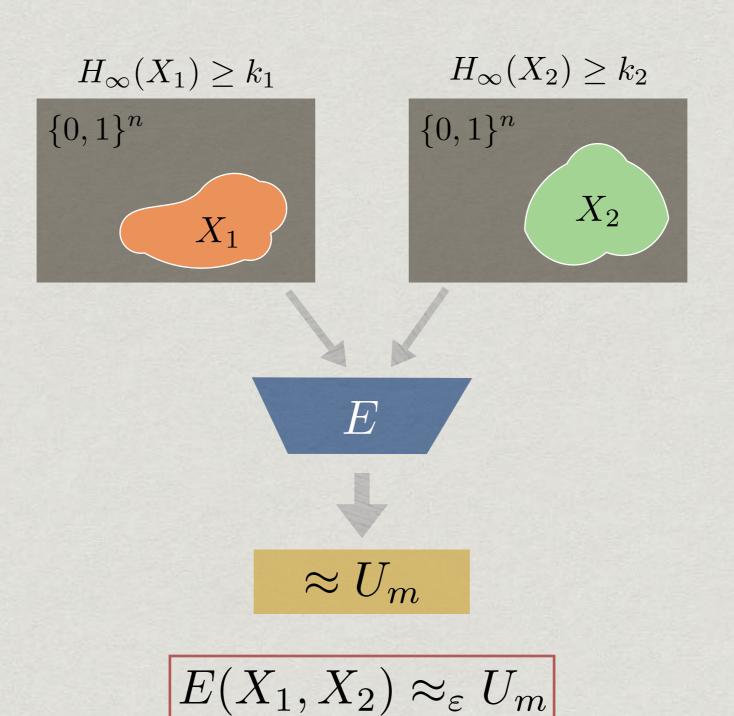
- \* Two-source extractors.
- \* Non-malleable extractors.
- \* Current constructions of two-source extractors via non-malleable extractors and where they fail in achieving small error.
- \* Constructing low-error two-source extractors given "good" non-malleable extractors.

# Today's talk

- \* Two-source extractors.
- \* Non-malleable extractors.
- \* Current constructions of two-source extractors via non-malleable extractors and where they fail in achieving small error.
- \* Constructing low-error two-source extractors given "good" non-malleable extractors.

- \* We say that a source X over  $\{0,1\}^n$  has k minentropy if for every x,  $\Pr[X=x] \le 2^{-k}$ . This is how we model weak sources.
- \* Alternatively, we can think of a weak source X as uniformly distributed over a subset of size  $2^k$ .

- \* We say that a source X over  $\{0,1\}^n$  has k minentropy if for every x,  $\Pr[X=x] \le 2^{-k}$ . This is how we model weak sources.
- \* Alternatively, we can think of a weak source X as uniformly distributed over a subset of size  $2^k$ .
- \* Given two **independent** weak source  $X_1$  and  $X_2$ , we want to extract almost-uniform bits (potentially, almost all the entropy).



- \* Known results for constant error.
- \* Omitted here: many constructions of multi-source extractors.

	min-entropy	
Non-explicit	logn+O(1)	
[CG88]	$(1/2+\delta)n$	
[Raz05]	$(1/2+\delta)n$ , $O(\log n)$	
[Bourgain05]	0.499 <i>n</i>	
[CZ15]	polylog(n)	
[BDT16]	log <sup>1+o(1)</sup> n	
[Cohen16]	logn-poly(loglogn)	
[Li16]	logn·loglogn	

	min-entropy	
Non-explicit	logn+O(1)	
[CG88]	$(1/2 + \delta)n$	
[Raz05]	$(1/2+\delta)n$ , $O(\log n)$	
[Bourgain05]	0.499 <i>n</i>	
[CZ15]	polylog(n)	
[BDT16]	log <sup>1+o(1)</sup> n	
[Cohen16]	logn-poly(loglogn)	
[Li16]	logn·loglogn	

- \* Non-explicitly, we can hope for  $\varepsilon=2^{-\Omega(k)}$ .
- \* Only the constructions of Chor-Goldreich, Raz and Bourgain achieve this.

	min-entropy	
Non-explicit	log n + O(1)	
[CG88]	$(1/2+\delta)n$	
[Raz05]	$(1/2+\delta)n$ , $O(\log n)$	
[Bourgain05]	0.499 <i>n</i>	
[CZ15]	polylog(n)	
[BDT16]	log <sup>1+o(1)</sup> n	
[Cohen16]	logn-poly(loglogn)	
[Li16]	logn·loglogn	

- \* Non-explicitly, we can hope for  $\varepsilon=2^{-\Omega(k)}$ .
- \* Only the constructions of Chor-Goldreich, Raz and Bourgain achieve this.
- \* We will soon see where recent constructions fall short.

	min-entropy	
Non-explicit	logn+O(1)	
[CG88]	$(1/2 + \delta)n$	
[Raz05]	$(1/2+\delta)n$ , $O(\log n)$	
[Bourgain05]	0.499 <i>n</i>	
[CZ15]	polylog(n)	
[BDT16]	log <sup>1+o(1)</sup> n	
[Cohen16]	logn-poly(loglogn)	
[Li16]	logn·loglogn	

- \* Non-explicitly, we can hope for  $\varepsilon=2^{-\Omega(k)}$ .
- \* Only the constructions of Chor-Goldreich, Raz and Bourgain achieve this.
- \* We will soon see where recent constructions fall short.
- \* Viewing it differently: We want the construction to run in time polylog( $1/\varepsilon$ ) instead of poly( $1/\varepsilon$ ).

	min-entropy	
Non-explicit	logn+O(1)	
[CG88]	$(1/2 + \delta)n$	
[Raz05]	$(1/2+\delta)n$ , $O(\log n)$	
[Bourgain05]	0.499 <i>n</i>	
[CZ15]	polylog(n)	
[BDT16]	log <sup>1+o(1)</sup> n	
[Cohen16]	logn-poly(loglogn)	
[Li16]	logn·loglogn	

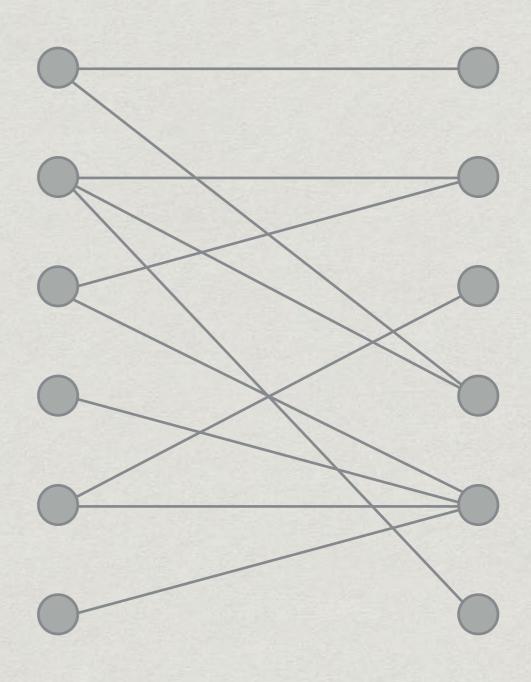
Our goal: Low-error two-source extractors, even for  $\delta n$  min-entropy.

(Preferably outputting many bits as well, but it often goes together...)

- \* The very-high error case is also interesting...
- \* In every N×N bipartite graph there is a ½logN×½logN monochromatic subgraph (a bipartite clique or an independent set).

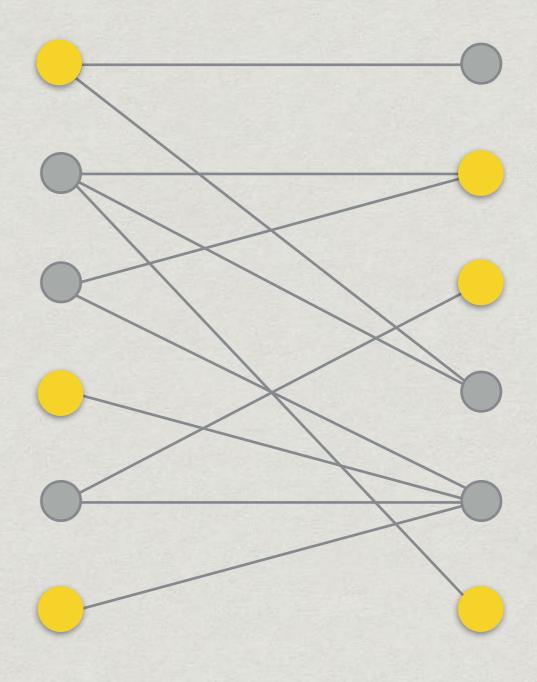
- \* The very-high error case is also interesting...
- \* In every N×N bipartite graph there is a ½logN×½logN monochromatic subgraph (a bipartite clique or an independent set).





- \* The very-high error case is also interesting...
- \* In every N×N bipartite graph there is a ½logN×½logN monochromatic subgraph (a bipartite clique or an independent set).





- \* Erdős (1947) there exists an *N*×*N* bipartite graph with **no** *K*×*K* monochromatic subgraphs, for *K*=2log*N*.
- \* A random graph has this property.
- \* The Erdős \$100 challenge find such an explicit graph, even for  $K=O(\log N)$ .
- \* Still open...

N

- \* We can view every bipartite graph naturally as a function E:[N]×[N]→{0,1}.
- \* The bipartite Ramsey problem: construct explicit matrices with no *K*×*K* constant sub-matrices.

 0
 1
 0
 1
 1
 1
 0

 0
 0
 0
 1
 1
 0
 0
 1
 0

 0
 0
 1
 0
 1
 1
 1
 0
 1

 0
 1
 0
 1
 0
 1
 0
 1
 0

 1
 1
 0
 1
 0
 0
 0
 1
 0
 1

 0
 0
 1
 0
 0
 0
 1
 0
 1

()

0

N

- \* We can view every bipartite graph naturally as a function E:[N]×[N]→{0,1}.
- \* The bipartite Ramsey problem: construct explicit matrices with no *K*×*K* constant sub-matrices.
- \* The low-error two-source extractors problem: Insist on unbiased sub-matrices, with a very small bias.

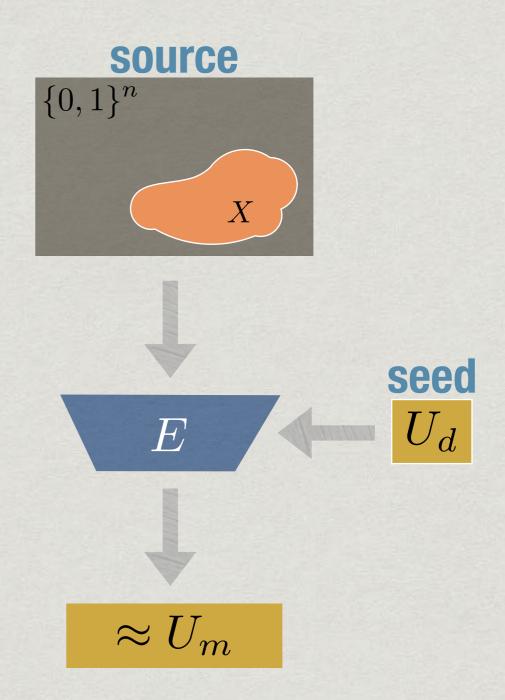
N

N

# Today's talk

- \* Two-source extractors.
- \* Non-malleable extractors.
- \* Current constructions of two-source extractors via non-malleable extractors and where they fail in achieving small error.
- \* Constructing low-error two-source extractors given "good" non-malleable extractors.

- \* A special case of twosource extractors is when one source is completely uniform, the seed.
- \* The seed length can be as small as  $2\log(n/\varepsilon)$ .



\* We say a seeded extractor is **strong** if the output is uniform even given the seed:  $(E(X,Y),Y) \approx_{\varepsilon} (U,Y)$ .

- \* We say a seeded extractor is **strong** if the output is uniform even given the seed:  $(E(X,Y),Y) \approx_{\varepsilon} (U,Y)$ .
- \* Equivalently, for every source X with entropy at least k there exists a set of good seeds of density at least  $1-\varepsilon$  such that for every good seed  $y \in \{0,1\}^d$ ,  $E(X,y) \approx_{\varepsilon} U$ .
- \* We have good strong seeded extractors [LRVW03,GUV07,...].

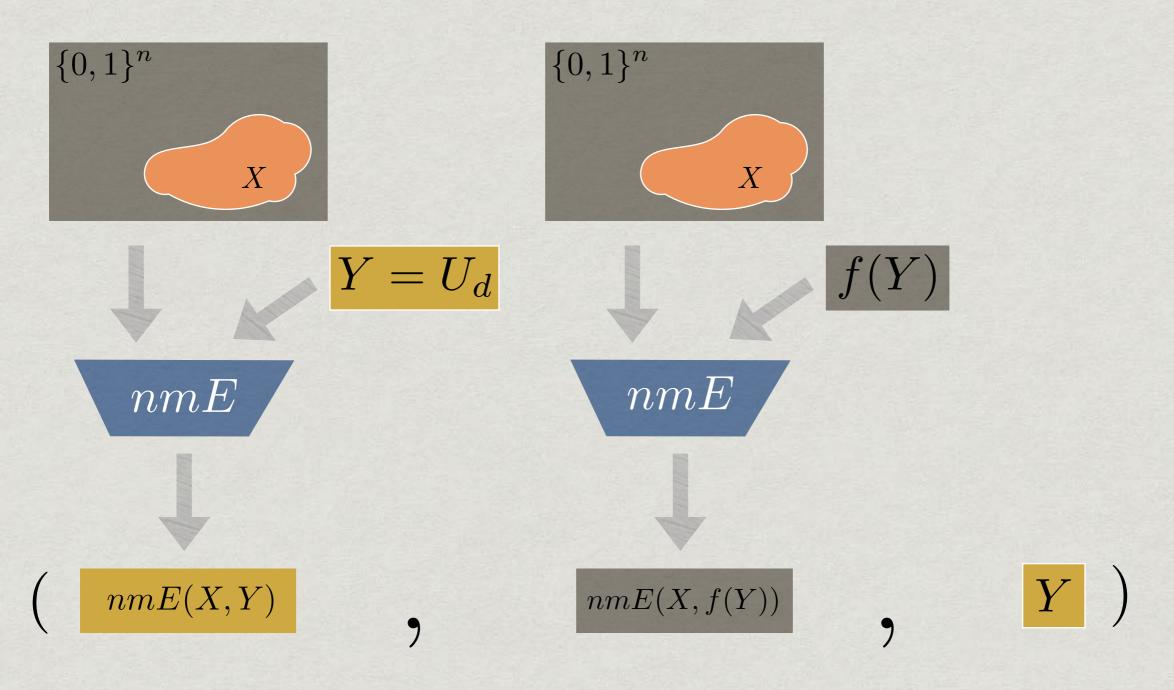
#### Non-malleable extractors [DW09]

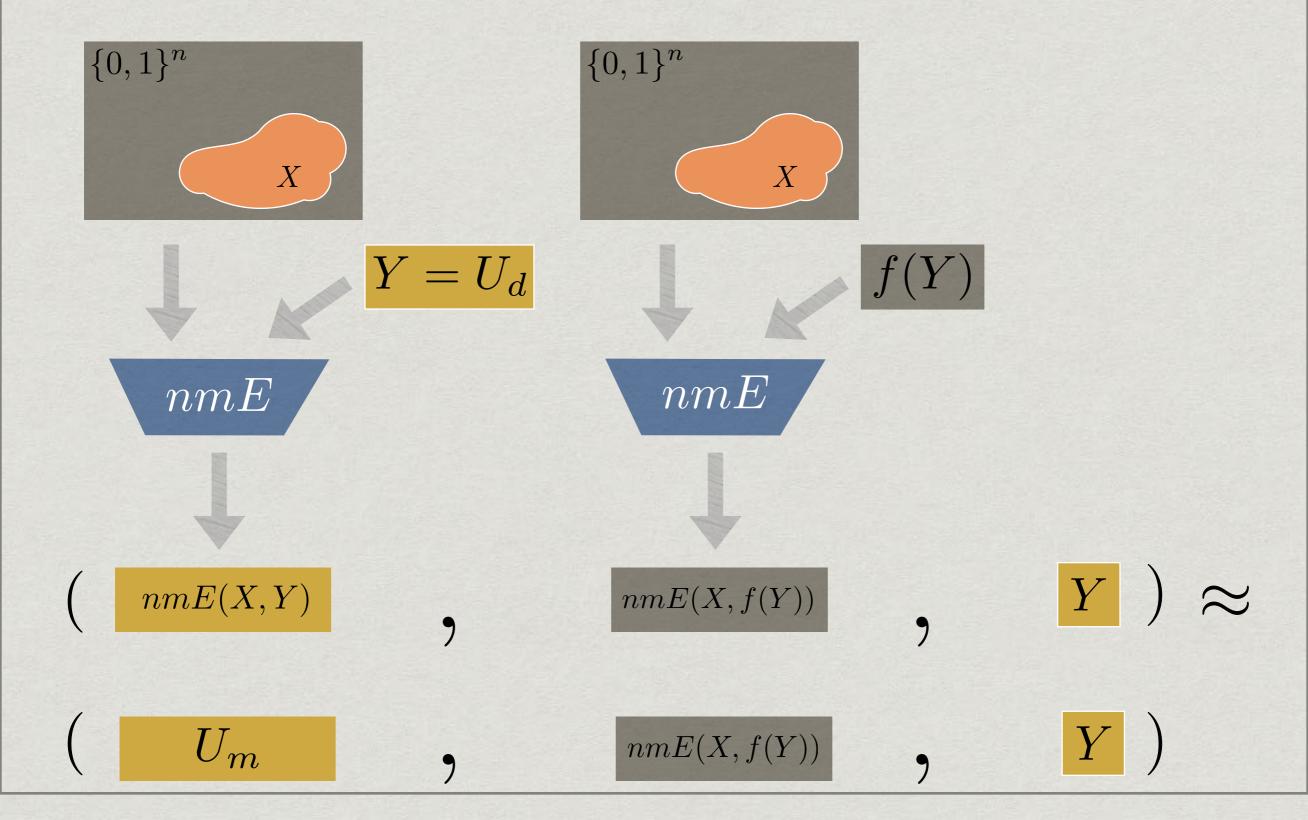
#### Non-malleable extractors [DW09]

- \* A generalization of strong seeded-extractors.
- \* An adversary cannot distinguish between the output nmE(X,Y) and a uniform string, even given the seed Y and the output of nmE on t correlated seeds.

#### Non-malleable extractors [DW09]

- \* A generalization of strong seeded-extractors.
- \* An adversary cannot distinguish between the output nmE(X,Y) and a uniform string, even given the seed Y and the output of nmE on t correlated seeds.
- \*  $(nmE(X,Y),nmE(X,f_1(Y)),...,nmE(X,f_t(Y)),Y)$  is  $\varepsilon$ -close to  $(U,nmE(X,f_1(Y)),...,nmE(X,f_t(Y)),Y)$ .





\* Known explicit constructions for t=1 (a partial list). A reduction by [Cohen16] allows us to go to an arbitrary t by roughly paying a factor of t in the entropy and  $t^2$  in the seed-length.

	seed length	min-entropy
[CRS12,DLWZ11]	$\log(n/\varepsilon)$	$(1/2+\delta)n$
[Li12]	$\log(n/\varepsilon)$	0.499 <i>n</i>
[CGL15]	$\log^2(n/\varepsilon)$	$\Omega(d)$
[Cohen16]	$\log(n/\varepsilon)\log(\log(n)/\varepsilon)$	$\Omega(d)$
[CL16]	$\log^{1+o(1)}(n/\varepsilon)$	$\Omega(d)$
[Cohen16]	$log(n) + log(1/\epsilon) poly(loglog(1/\epsilon))$	$\Omega(d)$
[Li16]	$\log(n) + \log(1/\varepsilon) \log\log(1/\varepsilon)$	$\Omega(d)$

- \* We will use an equivalent definition (up to some loss in the error) [CZ15,Cohen16].
- \* nmE is a n.m. extractor if every source induces a set of good seeds of high density such that the output of the extractor on a good seed is close to uniform even conditioned on its output on t other distinct seeds.

- \* We will use an equivalent definition (up to some loss in the error) [CZ15,Cohen16].
- \* nmE is a n.m. extractor if every source induces a set of good seeds of high density such that the output of the extractor on a good seed is close to uniform even conditioned on its output on *t* other distinct seeds.
- \* For every X there exists a set of G of density at least  $1-\varepsilon$  such that for every  $y \in G$  and any  $y_1, ..., y_t \in \{0,1\}^d \setminus \{y\}$  it holds that  $(nmE(X,y),nmE(X,y_1),...,nmE(X,y_t))$  is  $\varepsilon$ -close to  $(U,nmE(X,y_1),...,nmE(X,y_t))$ .

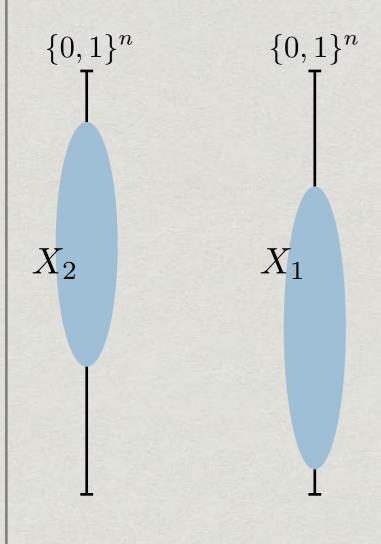
# Today's talk

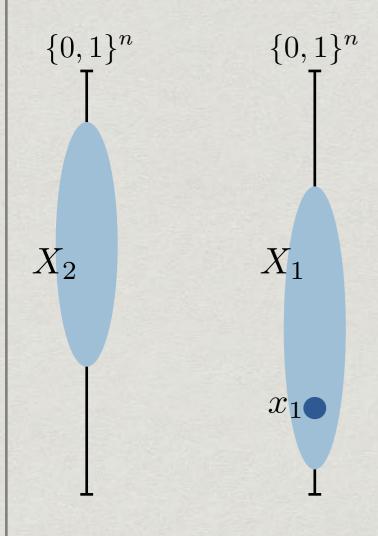
- \* Two-source extractors.
- \* Non-malleable extractors.
- \* Current constructions of two-source extractors via non-malleable extractors and where they fail in achieving small error.
- \* Constructing low-error two-source extractors given "good" non-malleable extractors.

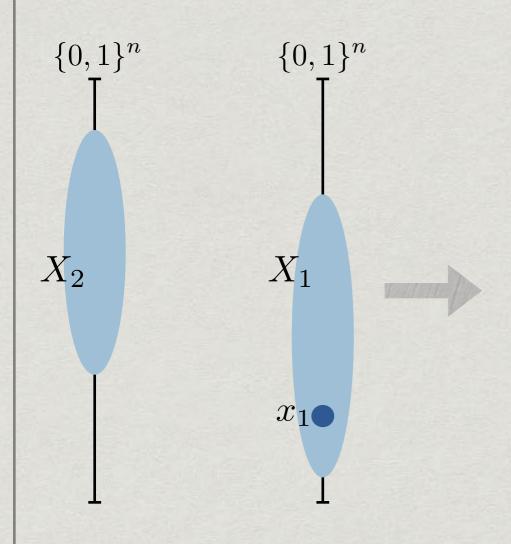
# Current constructions of two-source extractors

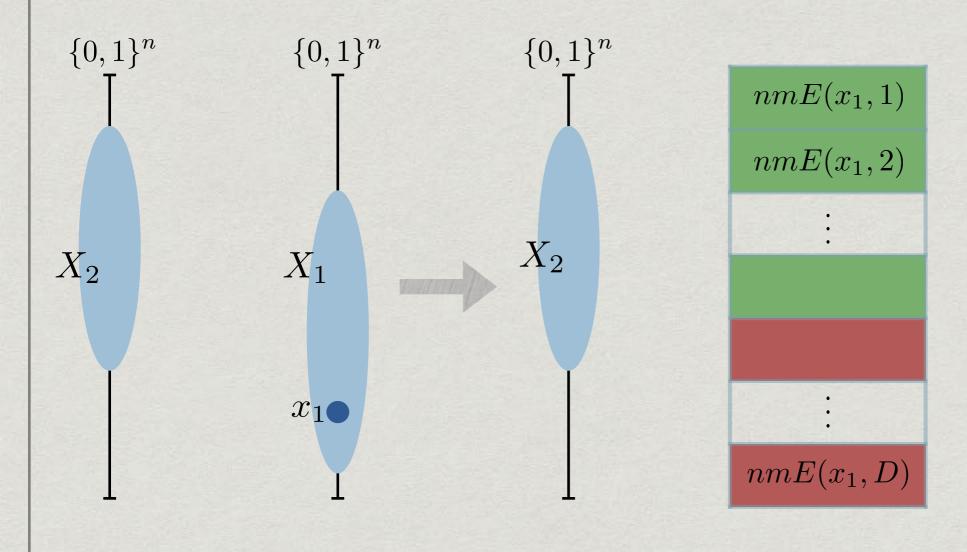
\* All recent constructions of two-source extractors use non-malleable extractors as a central ingredient.

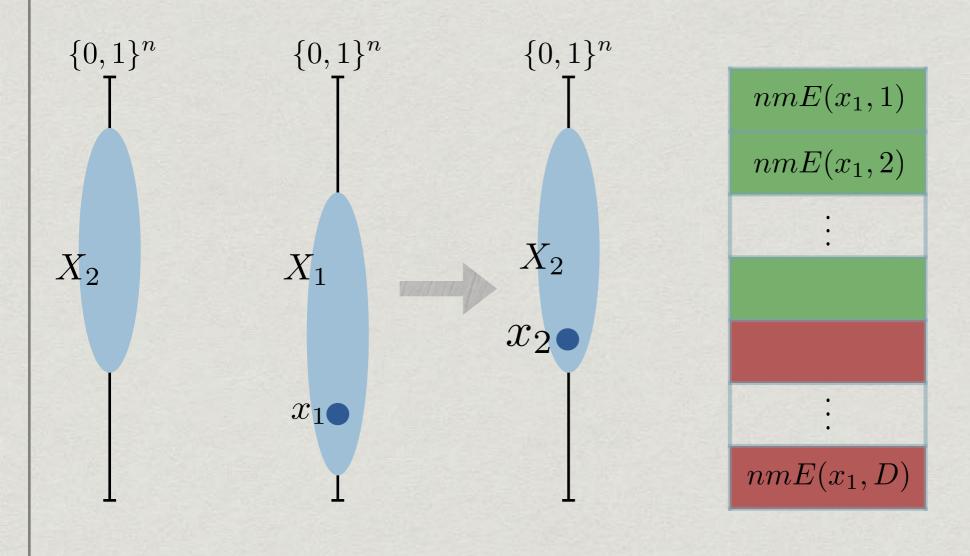
- \* All recent constructions of two-source extractors use non-malleable extractors as a central ingredient.
- \* A bird's-eye view of these constructions: Given two inputs  $x_1$  and  $x_2$ ,
  - \* Generate a table of nmE( $x_1$ ,i) for all seeds  $i \in \{0,1\}^d$ .
  - \* Using  $x_2$ , sample a subset of the rows.
  - \* Apply a *resilient* function on the reduced table.

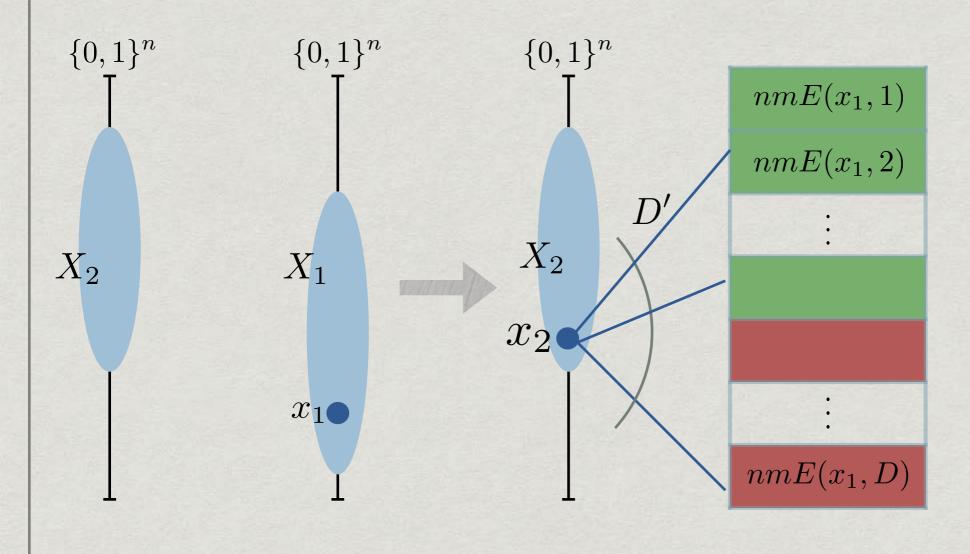


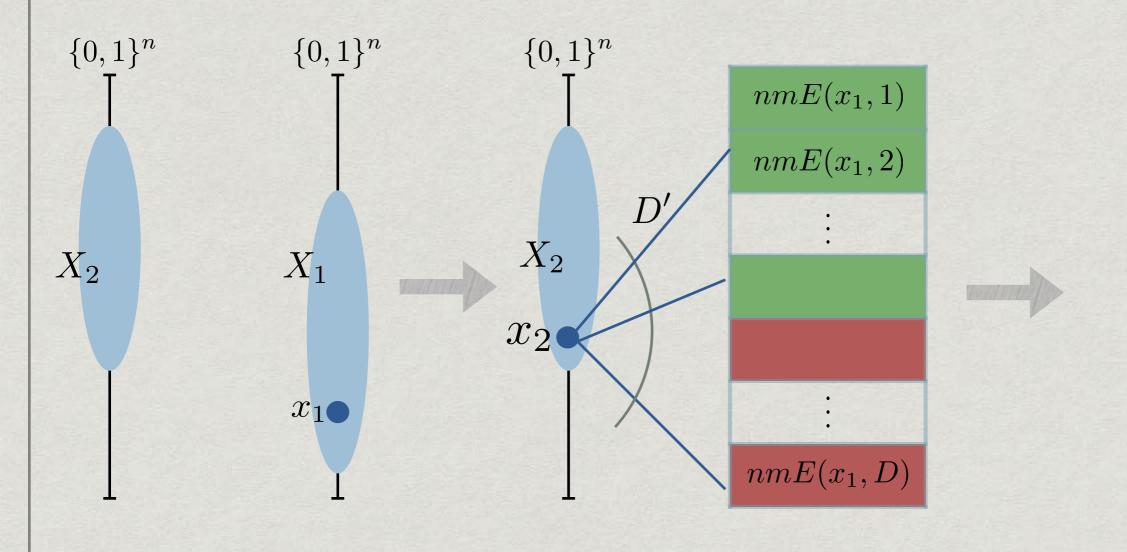


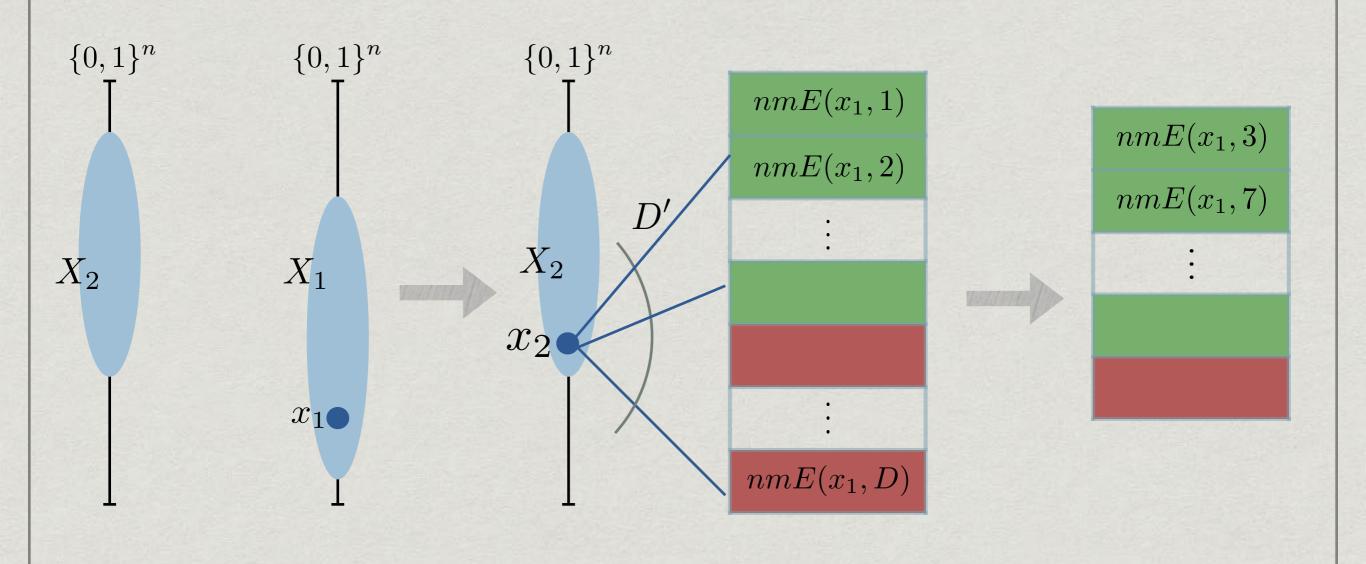


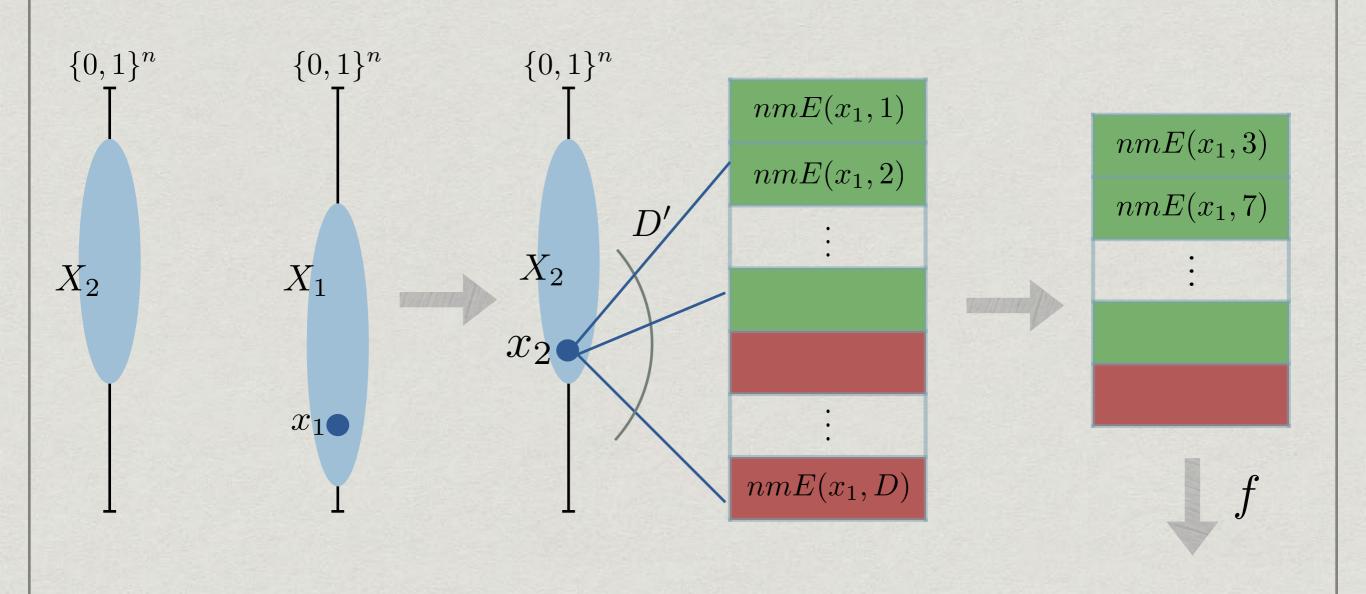


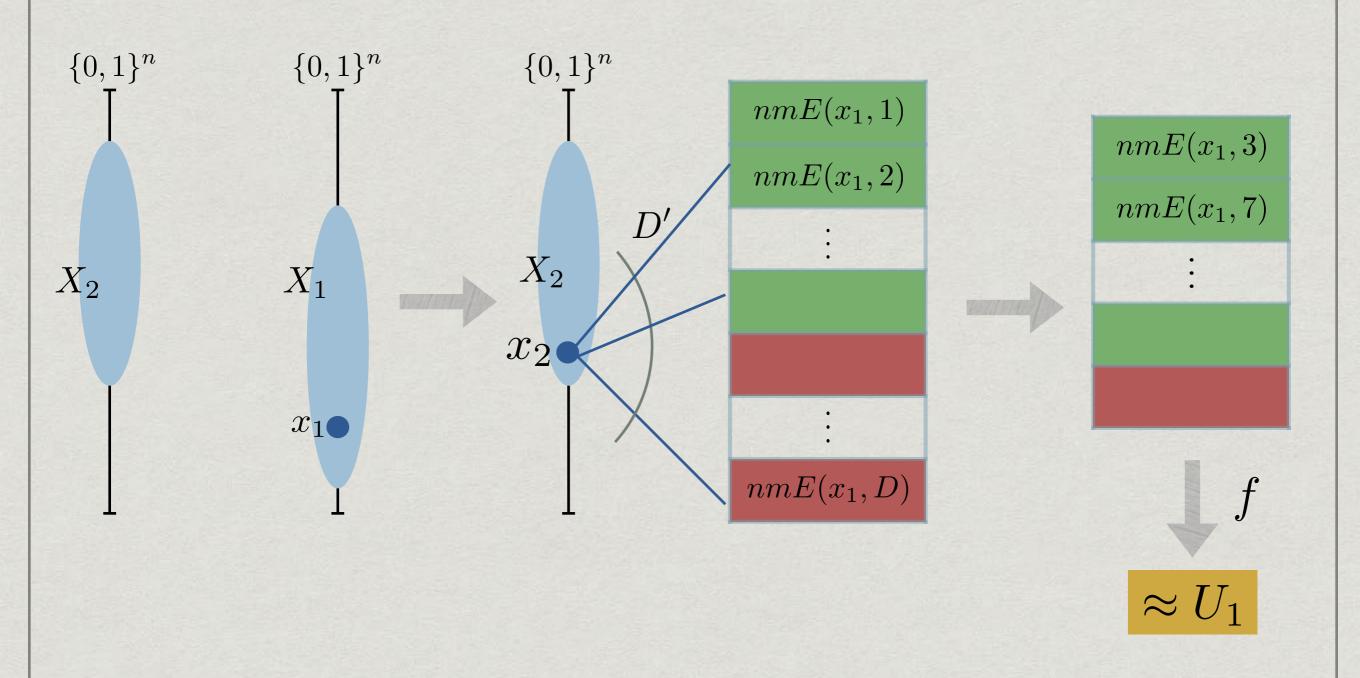












\* The resulting table is close to being uniform and *t*-wise independent in the good rows.

- \* The resulting table is close to being uniform and *t*-wise independent in the good rows.
- \* We need f to be resilient:
  - \* Say we have D' players.  $\varepsilon$ -fraction of them are malicious, and the rest are t-wise independent.

- \* The resulting table is close to being uniform and *t*-wise independent in the good rows.
- \* We need f to be resilient:
  - \* Say we have D' players.  $\varepsilon$ -fraction of them are malicious, and the rest are t-wise independent.
  - \* The honest players draw their random bit and later the malicious players draw as they wish.

- \* The resulting table is close to being uniform and *t*-wise independent in the good rows.
- \* We need f to be resilient:
  - \* Say we have D' players.  $\varepsilon$ -fraction of them are malicious, and the rest are t-wise independent.
  - \* The honest players draw their random bit and later the malicious players draw as they wish.
  - \* With high probability, the outcome is not biased the malicious players cannot substantially bias the outcome.

\* A corollary of [KKL88] — even one malicious player can bias the output with probability at least logD'/D'.

- \* A corollary of [KKL88] even one malicious player can bias the output with probability at least logD'/D'.
- \* We cannot hope for an error smaller than 1/D', and D' is the size of our table.

- \* A corollary of [KKL88] even one malicious player can bias the output with probability at least logD'/D'.
- \* We cannot hope for an error smaller than 1/D', and D' is the size of our table.
- \* Thus, the running time is at least  $1/\varepsilon$ .

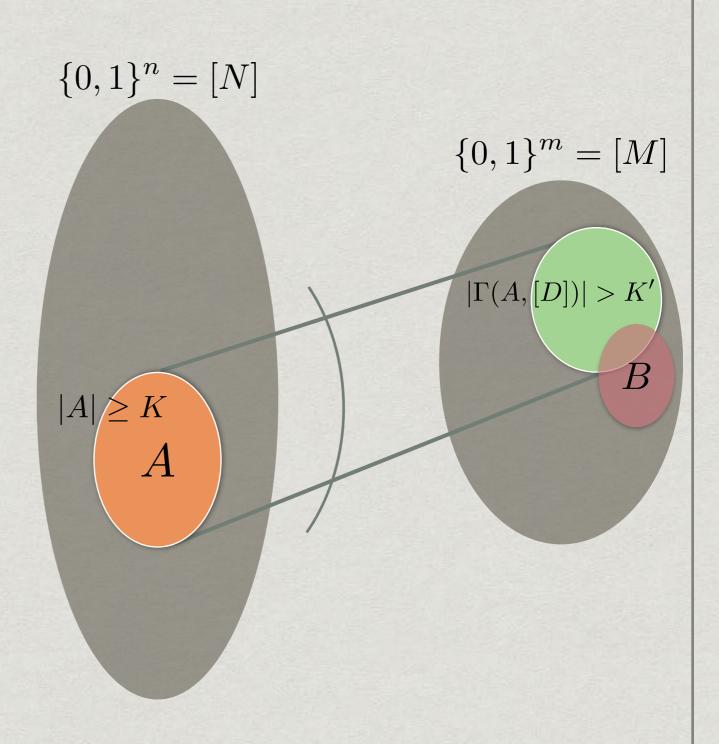
## Today's talk

- \* Two-source extractors.
- \* Non-malleable extractors.
- \* Current constructions of two-source extractors via non-malleable extractors and where they fail in achieving small error.
- \* Constructing low-error two-source extractors given "good" non-malleable extractors.

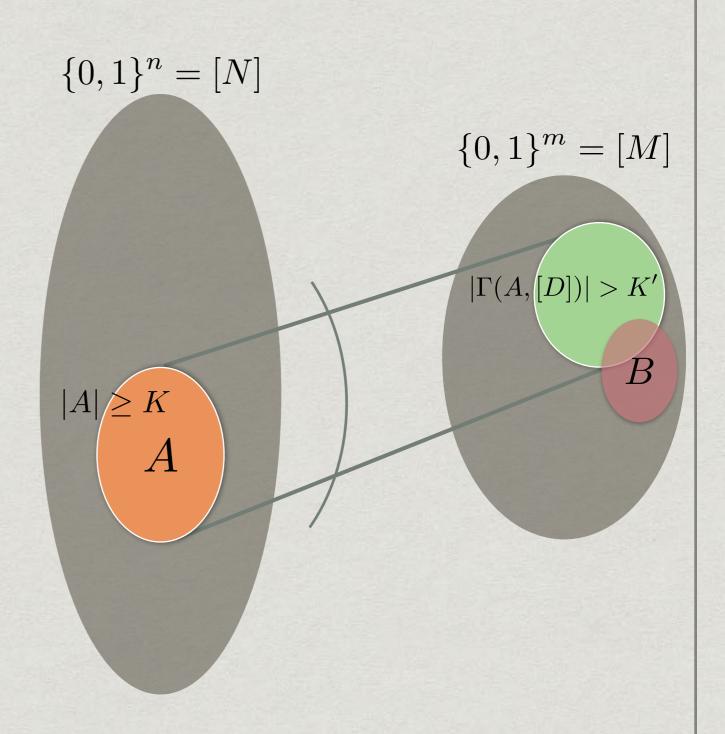
\* We should abandon resilient functions if we want to get a small error.

- \* We should abandon resilient functions if we want to get a small error.
- \* Instead of trying to sample and then employ *t*-wise in the good rows, let's just try and **hit** a good row.

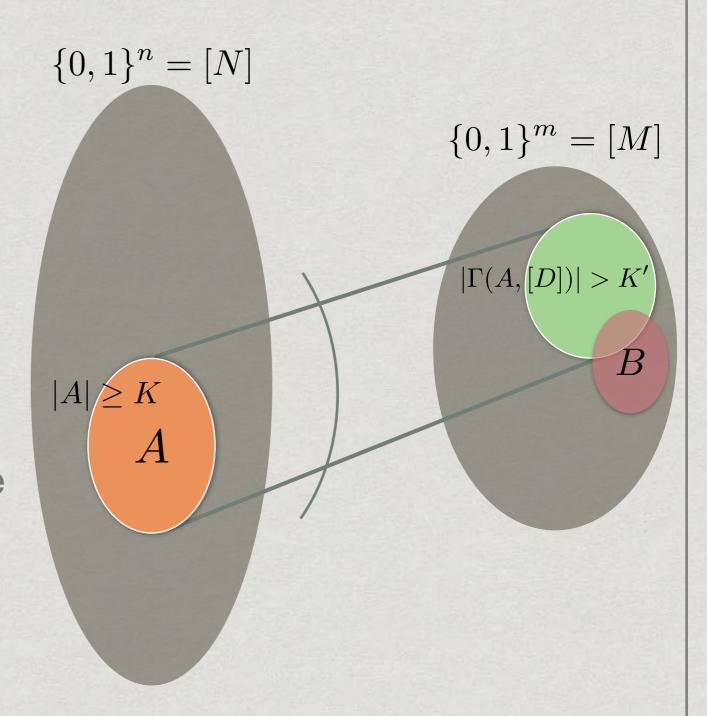
- \* We should abandon resilient functions if we want to get a small error.
- \* Instead of trying to sample and then employ *t*-wise in the good rows, let's just try and **hit** a good row.
- \* As usual, we hit with a disperser...



\*  $\Gamma:\{0,1\}^n \times [D] \to \{0,1\}^m$  is a (K,K')-disperser if for every set A of cardinality at least K,  $\Gamma$  maps A to a set of cardinality greater than K'.

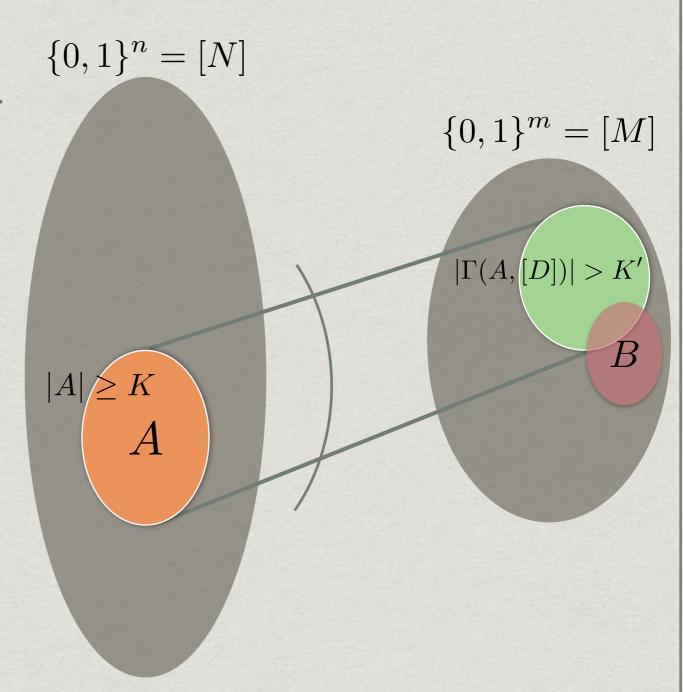


- \*  $\Gamma:\{0,1\}^n \times [D] \to \{0,1\}^m$  is a (K,K')-disperser if for every set A of cardinality at least K,  $\Gamma$  maps A to a set of cardinality greater than K'.
- \* We are interested in the case where K' is small compared to  $2^m$ . That is, we want to avoid **small** bad sets.



- \* Used to reduce error in onesided probabilistic algorithms.
- \* [RT]: When K' is not too large, say  $K'=\varepsilon M$ , the lower bound on the degree is

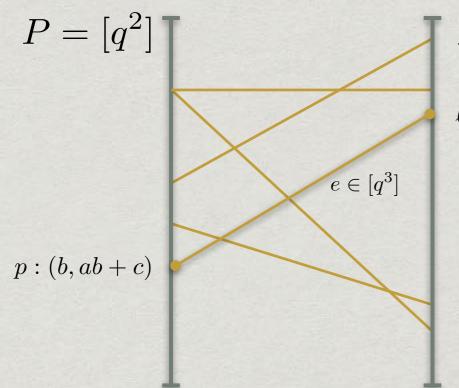
$$D = \Omega\left(\frac{\log\frac{N}{K}}{\log\frac{1}{\varepsilon}}\right)$$



\* Quite amazingly, when  $K=N^{\delta}$  for a constant  $\delta$ , there exist explicit constructions that achieve this bound [BKSSW05,Raz05,Zuck06].

- \* Quite amazingly, when  $K=N^{\delta}$  for a constant  $\delta$ , there exist explicit constructions that achieve this bound [BKSSW05,Raz05,Zuck06].
- \* The key ingredient in Zuckerman's construction: A points-lines incidence graph.

The input source is distributed, over  $[q]^3$ , among the edges of the graph.



$$L = [q^2]$$

$$\ell: y = ax + c$$

$$\Gamma: \mathbb{F}_q^3 \times [2] \to \mathbb{F}_q^2$$



\* This gives a degree-2 disperser, and we can recurse.

# Zuckerman's disperser

- \* This gives a degree-2 disperser, and we can recurse.
- \* For  $K=N^{\delta}$ , where  $\delta$  is arbitrary, the dependence is

$$D = (1/\delta)^{O(1)} \frac{n}{\log \frac{1}{\varepsilon}}$$

# Zuckerman's disperser

- \* This gives a degree-2 disperser, and we can recurse.
- \* For  $K=N^{\delta}$ , where  $\delta$  is arbitrary, the dependence is

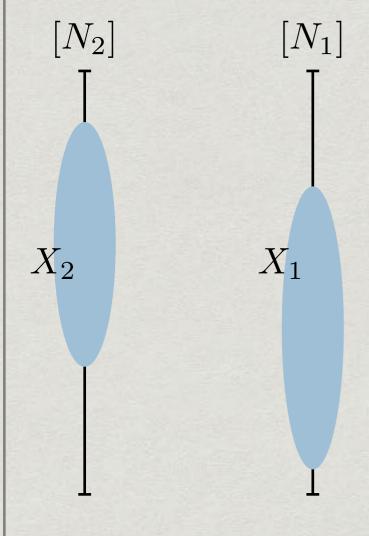
$$D = (1/\delta)^{O(1)} \frac{n}{\log \frac{1}{\varepsilon}}$$

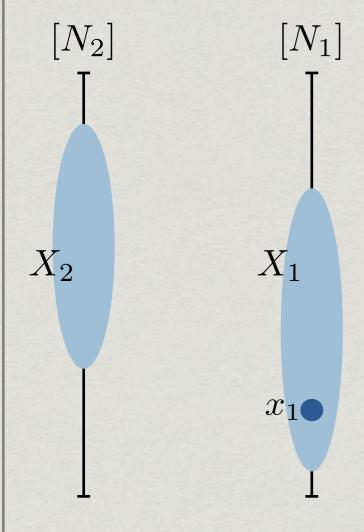
\* Also, the output length is determined by the number of recursion steps, and we have  $m = \delta^{O(1)}n$ .

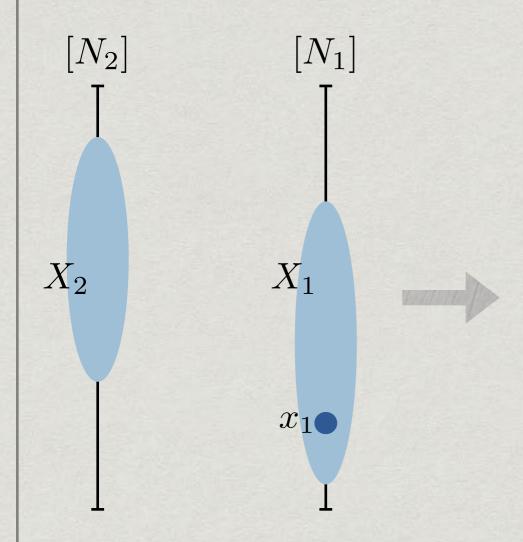
\* We are given a source  $X_1$  over  $[N_1]$  with entropy  $k_1$  and a source  $X_2$  over  $[N_2]$  with min-entropy  $k_2$ .

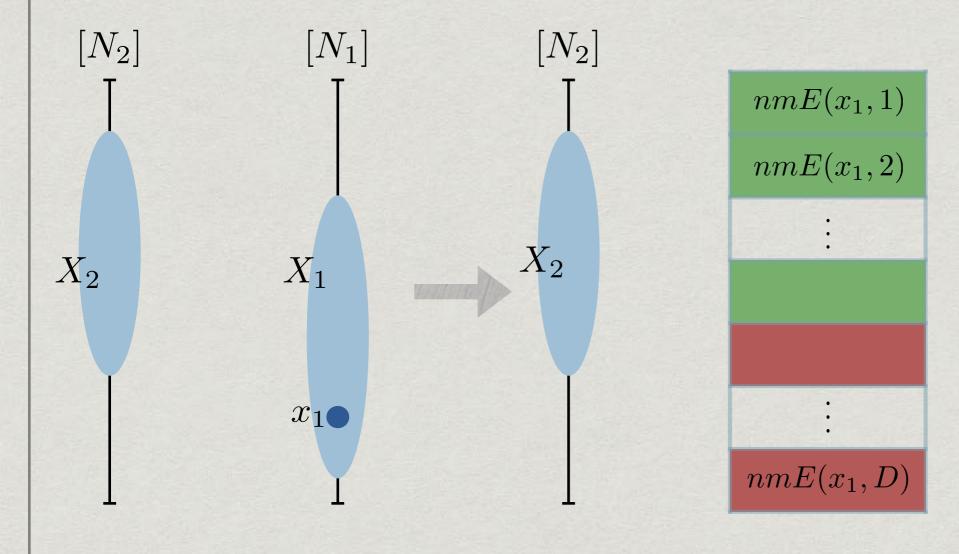
- \* We are given a source  $X_1$  over  $[N_1]$  with entropy  $k_1$  and a source  $X_2$  over  $[N_2]$  with min-entropy  $k_2$ .
- \* Ingredients:
  - \* nmE:  $[N_1] \times [D] \rightarrow \{0,1\}^m$ , a strong **t**-n.m. extractor with error  $\varepsilon$ .
  - \*  $\Gamma: [N_2] \times [t+1] \rightarrow [D]$ , a  $(\varepsilon K_2, \varepsilon D)$ -disperser.

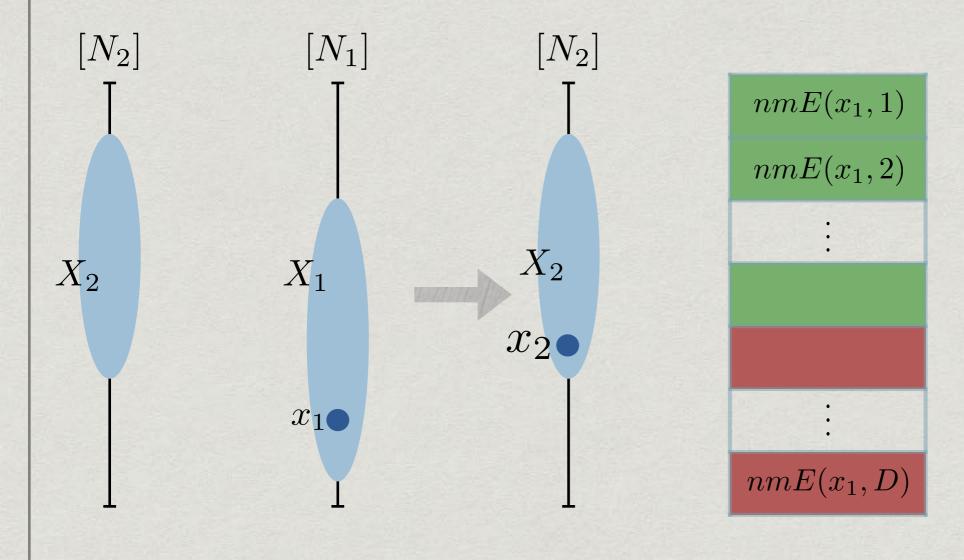
- \* We are given a source  $X_1$  over  $[N_1]$  with entropy  $k_1$  and a source  $X_2$  over  $[N_2]$  with min-entropy  $k_2$ .
- \* Ingredients:
  - \* nmE:  $[N_1] \times [D] \rightarrow \{0,1\}^m$ , a strong **t**-n.m. extractor with error  $\varepsilon$ .
  - \*  $\Gamma: [N_2] \times [t+1] \rightarrow [D]$ , a  $(\varepsilon K_2, \varepsilon D)$ -disperser.
- \* On input  $x_1, x_2$ , output  $\bigoplus_{i \in [t+1]} nmE(x_1, \Gamma(x_2, i))$ .

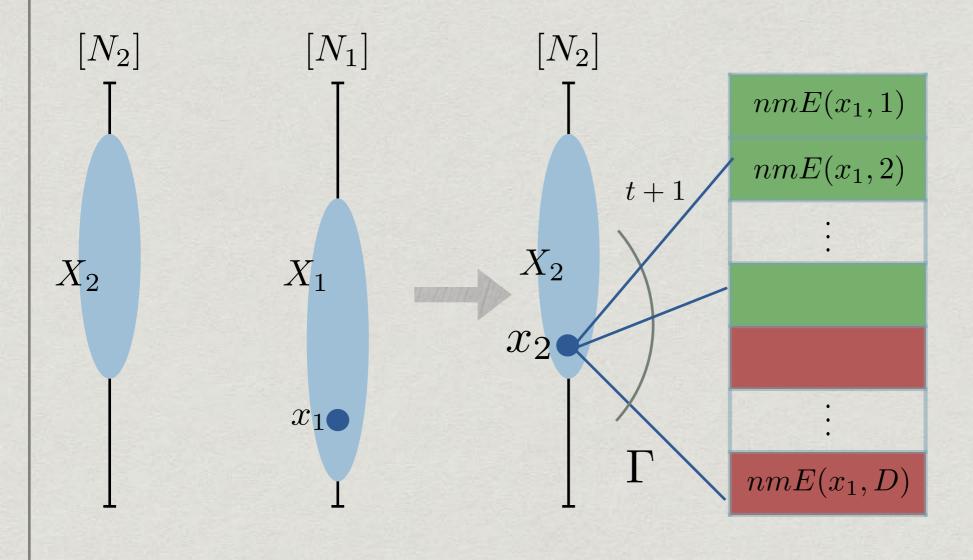


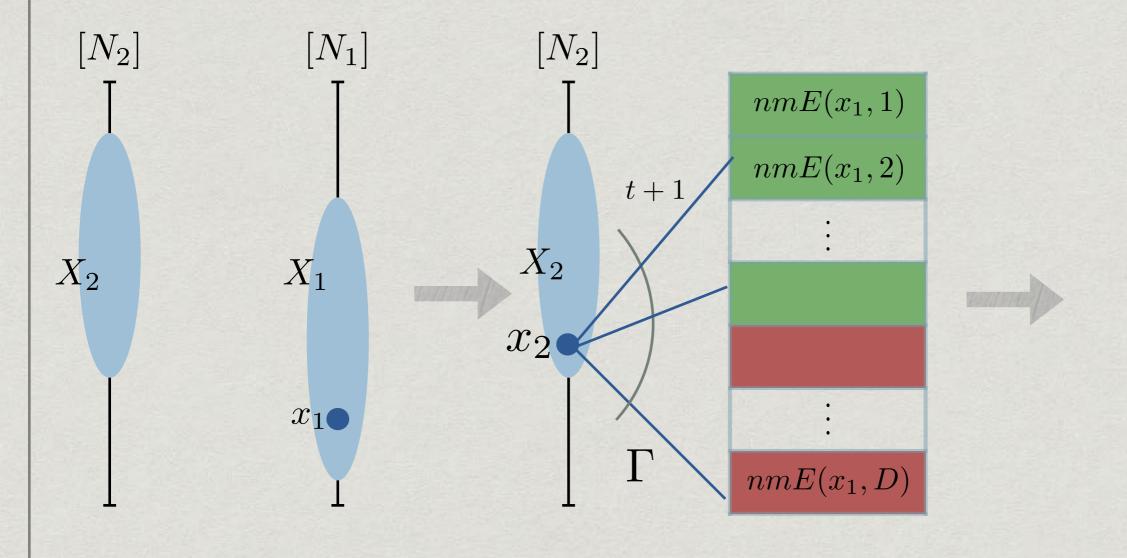


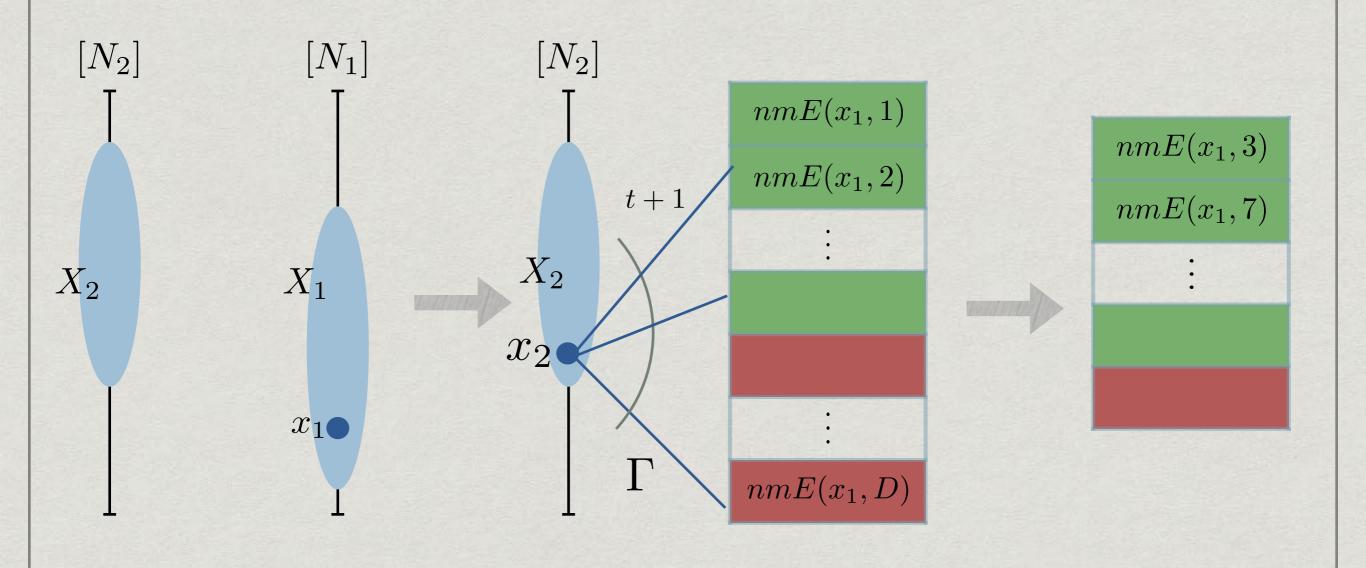


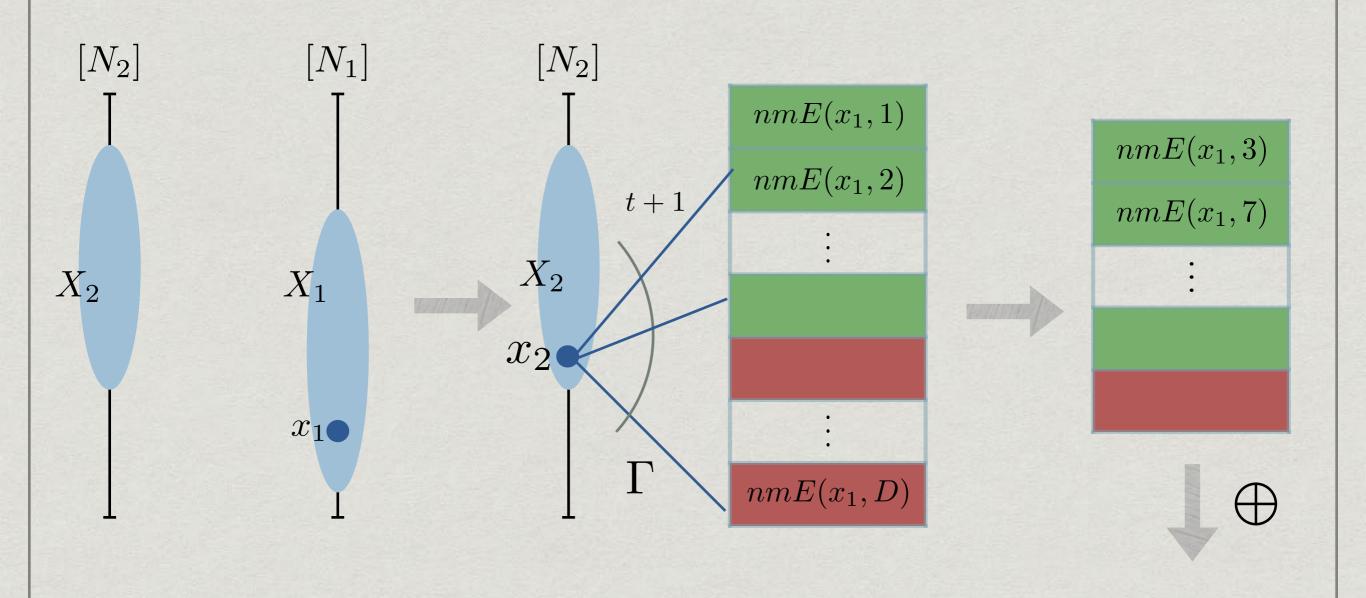


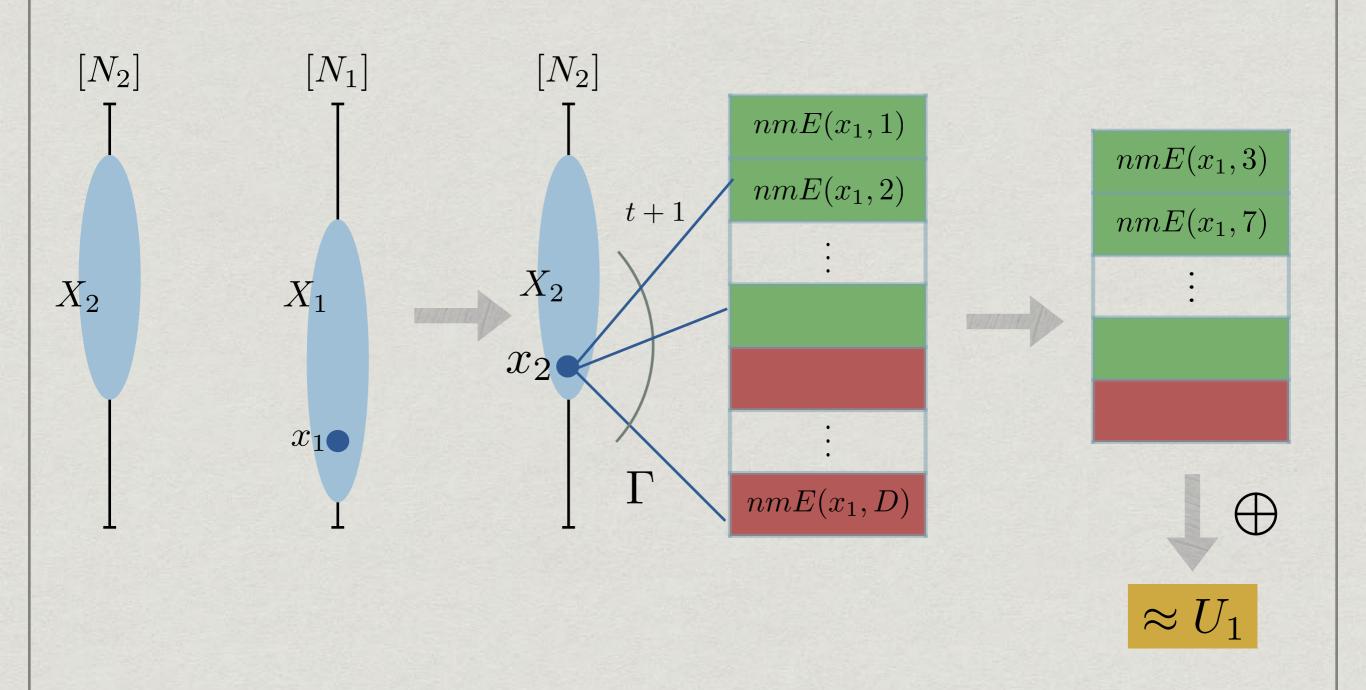


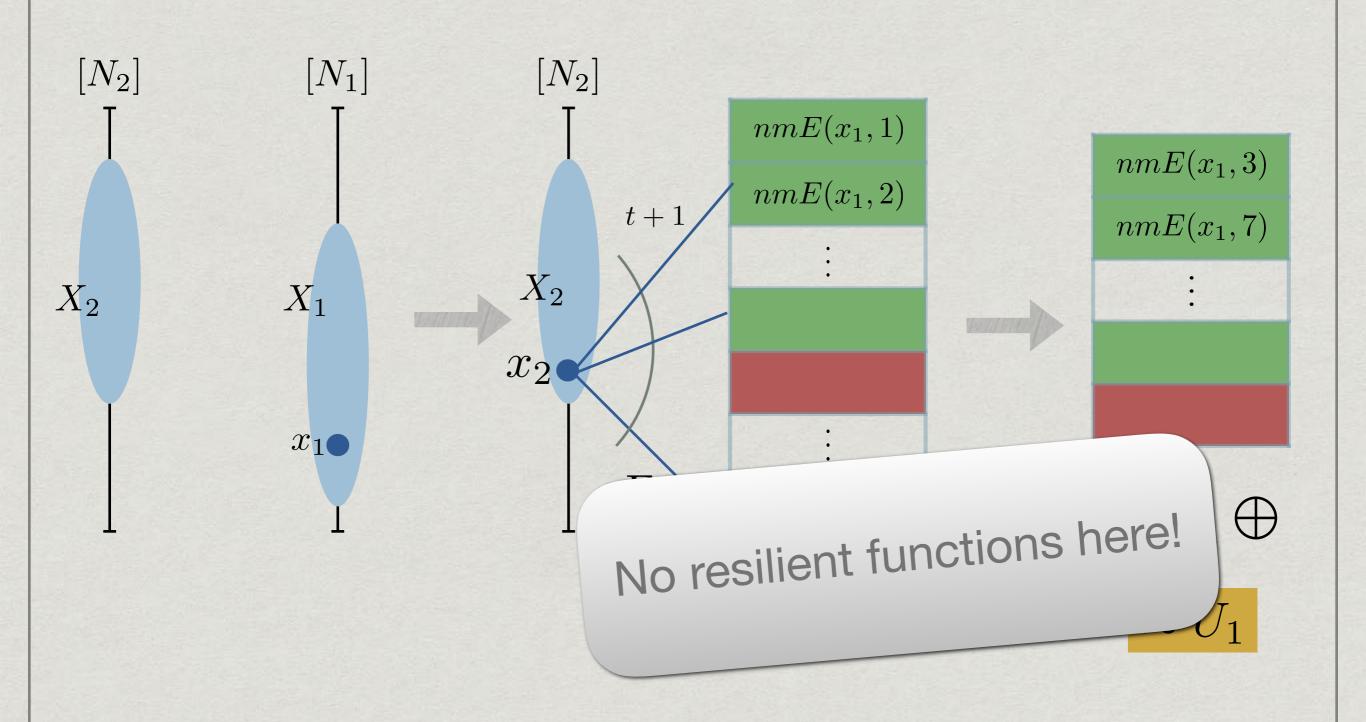












\* The source  $X_1$  defines a set of good and bad seeds. Let G be the set of good seeds, of density at least  $1-\varepsilon$ .

- \* The source  $X_1$  defines a set of good and bad seeds. Let G be the set of good seeds, of density at least  $1-\varepsilon$ .
- \* By the properties of  $\Gamma$ , the number of elements  $x_2$  for which  $\Gamma(x_2,[t+1])$  contains only bad seeds is at most  $\varepsilon K_2$ .

- \* The source  $X_1$  defines a set of good and bad seeds. Let G be the set of good seeds, of density at least  $1-\varepsilon$ .
- \* By the properties of  $\Gamma$ , the number of elements  $x_2$  for which  $\Gamma(x_2,[t+1])$  contains only bad seeds is at most  $\varepsilon K_2$ .
- \* Thus, with probability at least  $1-\varepsilon K_2/K_2=1-\varepsilon$ , the input  $x_2$  samples t+1 seeds of nmE, one of which, y, is good.

\* In such a case, nmE(X,y) is  $\varepsilon$ -close to uniform, even condition on t arbitrary outputs! This is since:

- \* In such a case, nmE(X,y) is  $\varepsilon$ -close to uniform, even condition on t arbitrary outputs! This is since:
- \* For every  $y \in G$  and any  $y_1, ..., y_t \in \{0,1\}^o \setminus \{y\}$  it holds that  $(nmE(X,y),nmE(X,y_1),...,nmE(X,y_t))$  is  $\varepsilon$ -close to  $(U,nmE(X,y_1),...,nmE(X,y_t))$ .

- \* In such a case, nmE(X,y) is  $\varepsilon$ -close to uniform, even condition on t arbitrary outputs! This is since:
- \* For every  $y \in G$  and any  $y_1, ..., y_t \in \{0,1\}^d \setminus \{y\}$  it holds that  $(nmE(X,y),nmE(X,y_1),...,nmE(X,y_t))$  is  $\varepsilon$ -close to  $(U,nmE(X,y_1),...,nmE(X,y_t))$ .
- \* Hence, the parity of these random variables is also close to uniform, and the overall error is  $2\varepsilon$ .

\* So, if the n.m. extractor can support small error (and existing constructions can), we get a construction with a small error.

\* The parity is not resilient... What happened here?

- \* The parity is not resilient... What happened here?
  - \* Instead of sampling (with a good sampler) D' rows from the table and applying a resilient function, we pick a drastically smaller sample set — of size t+1.

- \* The parity is not resilient... What happened here?
  - \* Instead of sampling (with a good sampler) D' rows from the table and applying a resilient function, we pick a drastically smaller sample set — of size t+1.
  - \* Instead of requiring that the number of malicious players is small, we have the weaker requirement that not *all* of the players in our sample set are malicious.

\* Or, when does it work? We have no option but to look closer into the parameters.

- \* Or, when does it work? We have no option but to look closer into the parameters.
- \* First, note that the disperser dictates  $n_2$ , the length of the second source, and typically it is smaller than  $n_1$ .

- \* Or, when does it work? We have no option but to look closer into the parameters.
- \* First, note that the disperser dictates  $n_2$ , the length of the second source, and typically it is smaller than  $n_1$ .
- \* A potential circular hazard the degree of  $\Gamma$  should be at least t+1, but the degree of  $\Gamma$  also depends on the seed length of the n.m. extractor, which in turn depends on t...

\* Let's check this circularity on the board...

#### Our result

- \* We see that the seed length of the n.m. extractor plays a crucial role. Say there exists an explicit n.m. extractor with seed length *d* and supports entropy *k*<sub>1</sub>. Our results:
  - \* If  $d=ct\log(n_1/\varepsilon)$  for a small enough constant c, there exists an explicit two-source extractor with small error for entropies  $k_1$  and  $k_2=an_2$  (for every constant a).

#### Our result

- \* We see that the seed length of the n.m. extractor plays a crucial role. Say there exists an explicit n.m. extractor with seed length *d* and supports entropy *k*<sub>1</sub>. Our results:
  - \* If  $d=t^{\gamma}\log(n_1/\varepsilon)$  for a small enough constant  $\gamma$ , there exists an explicit two-source extractor with small error for entropies  $k_1$  and  $k_2=n_2^{\beta}$  for some constant  $\beta$ .

#### Our result

- \* We see that the seed length of the n.m. extractor plays a crucial role. Say there exists an explicit n.m. extractor with seed length *d* and supports entropy *k*<sub>1</sub>. Our results:
  - \* If  $d=\log(n_1/\varepsilon)+O(t)$ , there exists an explicit two-source extractor with small error for entropies  $k_1$  and  $k_2=n_2^\beta$  for every constant  $\beta$ .

\* Non-explicitly, our constraints on d are easily satisfied. The seed length of a probabilistic construction is  $d=2\log(n/\varepsilon)+O(\log t)$ .

- \* Non-explicitly, our constraints on d are easily satisfied. The seed length of a probabilistic construction is  $d=2\log(n/\varepsilon)+O(\log t)$ .
- \* Taking a closer look on recent constructions of non-malleable extractors, we see that  $d=\Omega(k)$  and  $k=\tilde{O}(t^2\log(n/\varepsilon))$ .
- \* Very roughly, this coupling between *d* and *k* is inherent when you do alternating extraction.

\* To summarize...

- \* To summarize...
- \* Due to [CZ15,BDT16] we know that n.m. extractors with short seed length supporting small entropies give rise to good two-source extractors with constant error.

- \* To summarize...
- \* Due to [CZ15,BDT16] we know that n.m. extractors with short seed length supporting small entropies give rise to good two-source extractors with constant error.
- \* This work: N.m. extractors also give rise to twosource extractors with small error, as long as the seed-length's dependency on *t* is good.

\* The moral: Keep constructing non-malleable extractors, with techniques that go beyond alternating extraction.

