CCC 2026 Test of Time Award

Unbalanced Expanders and Randomness Extractors from Parvaresh–Vardy Codes

(appeared in CCC 2007)

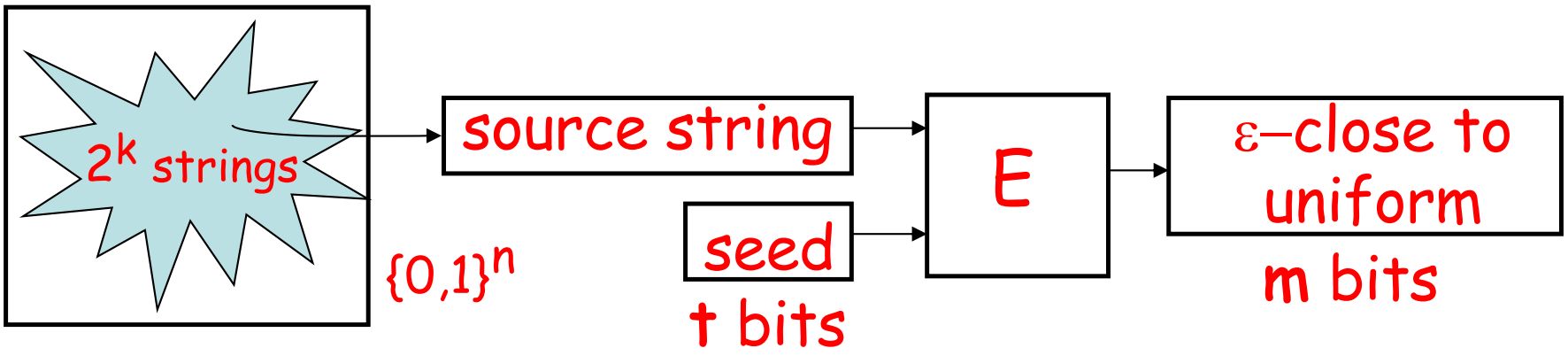
Chris Umans

Caltech

Joint work with: Venkat Guruswami, Salil Vadhan

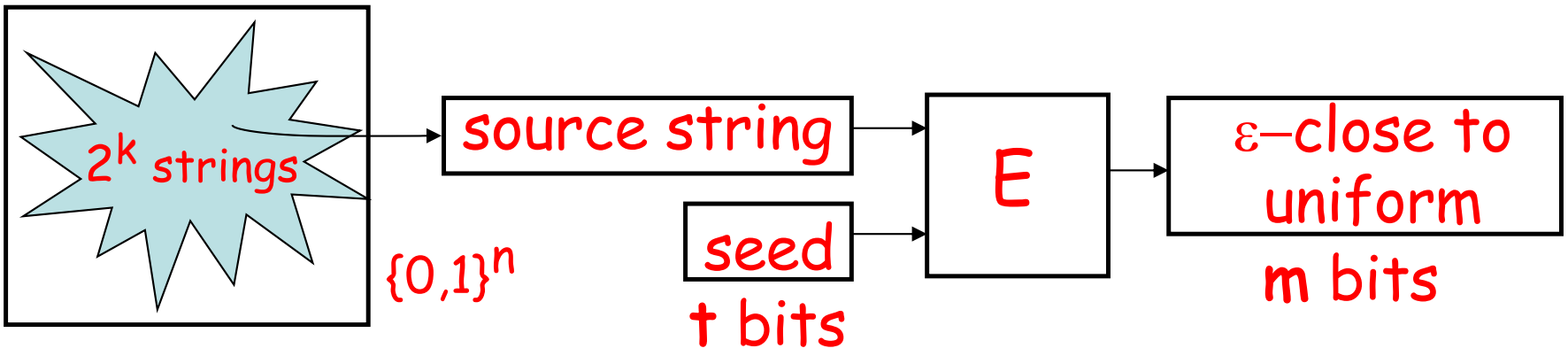
The key objects and the main result

Randomness extractors



- Applications:
  - simulate randomized algorithms using crude source of randomness
  - Many others: expanders, inapproximability results, pseudorandom generators, error-correcting codes, data structures, distributed and network algorithms, embeddings, optimal samplers…

Randomness extractors

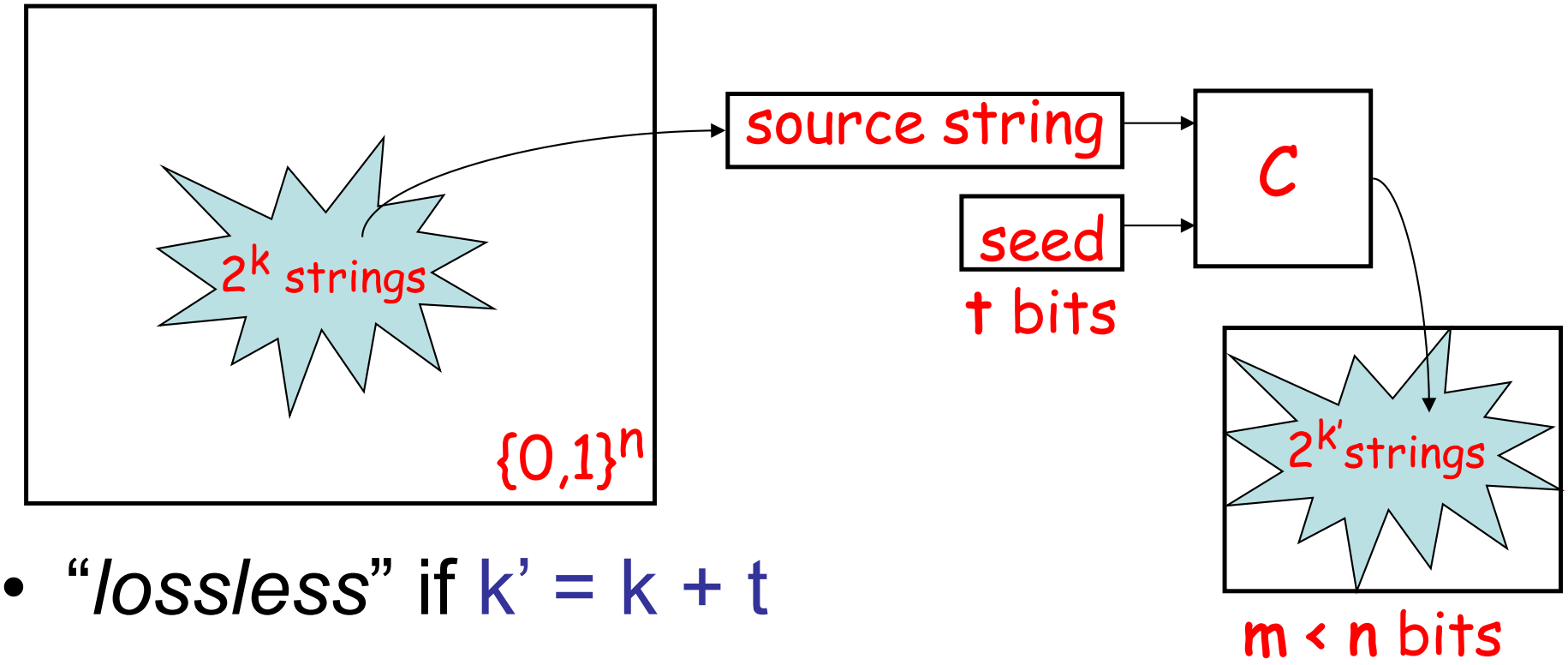


- Dozens of constructions over many years
  - 1990-1999: largely combinatorial
  - 1999: new ingredient -- error-correcting codes
- **Goals**:

	optimal:	[LRVW03]:
short seed	log n + O(1)	O(log n)
long output	m = k + t - O(1)	m = (1 - α)k

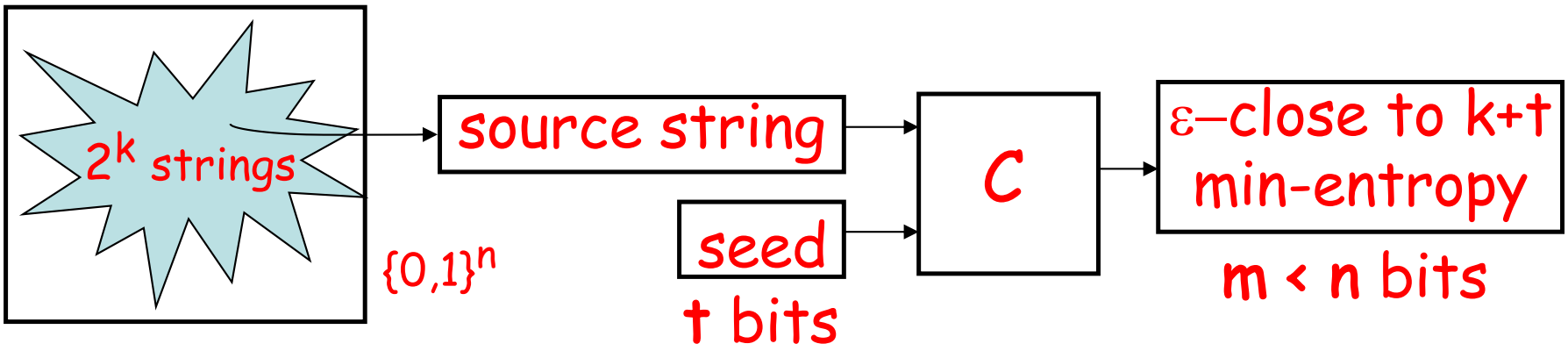
Condensers

- Common intermediate object:



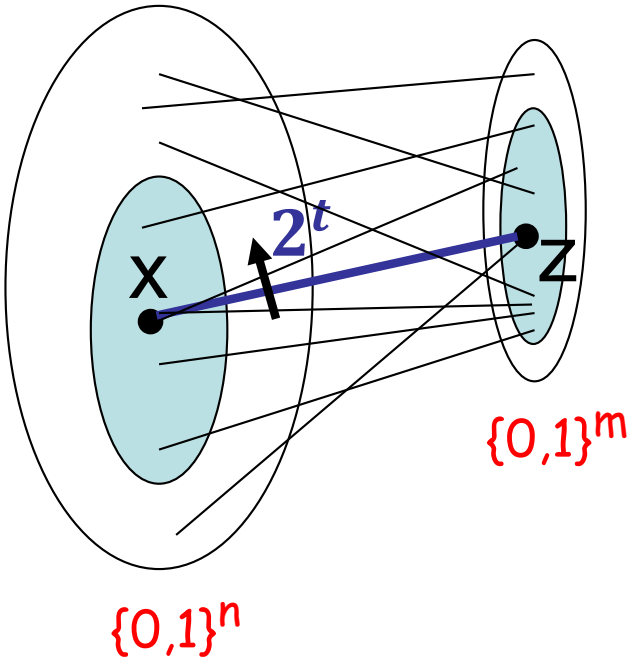
- "*lossless*" if k' = k + t

Lossless condenser ⇔ expander

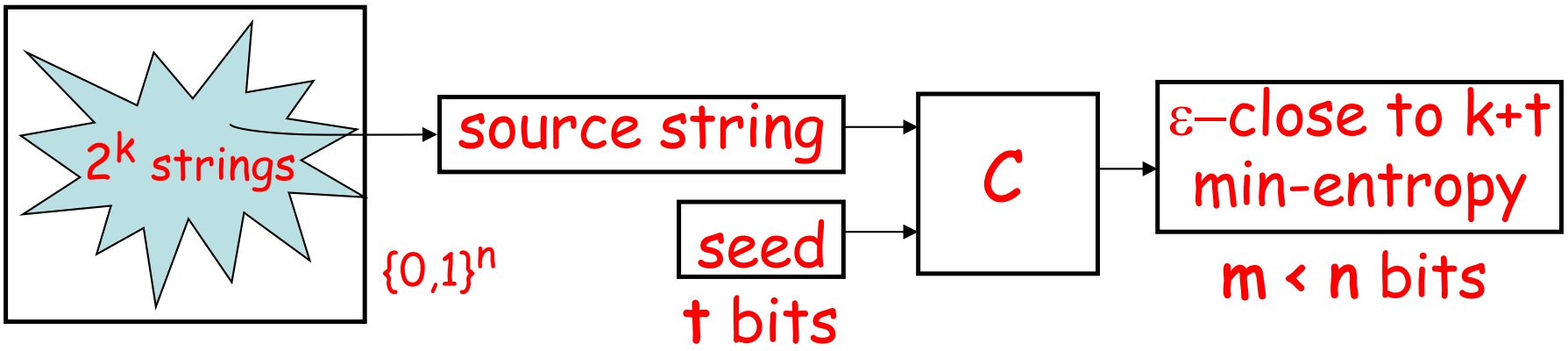


edge (x, z) present iff for some y, C(x, y) = z

C is a (k, ϵ) lossless condenser ⇔ sets of size 2^k expand to $(1-\varepsilon)2^t \cdot 2^k$ [TUZ01]



Lossless condensers

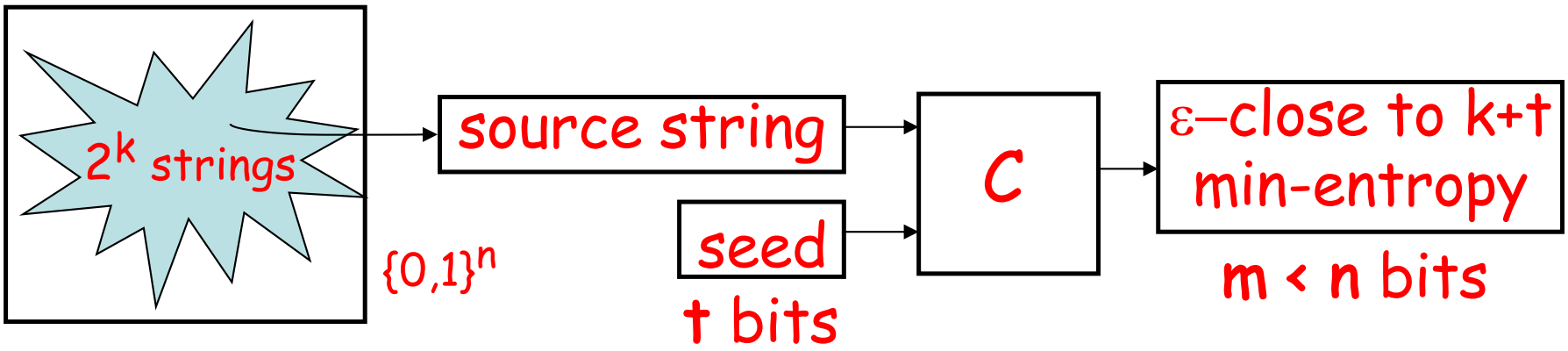


- **Goals**: optimal: [TUZ01]

  short seed log n + O(1) O(log n)

  short output m = k + t + O(1) $m = k^{1+\alpha}$

Lossless condensers



- **Goals**:

	optimal:	[TUZ01]
short seed	log n + O(1)	O(log n) or O($\log^2$ n)
short output	m = k + t + O(1)	m = $k^{1+\alpha}$ or m = O(k)

Our results (extractors)

- **Goals**:

	optimal:	[LRVW03]:
short seed	log n + O(1)	O(log n)
long output	m = k + t - O(1)	m = (1 - α)k

- we matched LRVW03 (and handled small ε)
  - LRVW uses: mergers, condensers, block-wise and somewhere-random sources, locally decodable codes, "win-win" analysis of recursive construction…
  - Our extractor: complete construction+proof in this talk

Our results (condensers)

- **Goals**: optimal: [TUZ01]

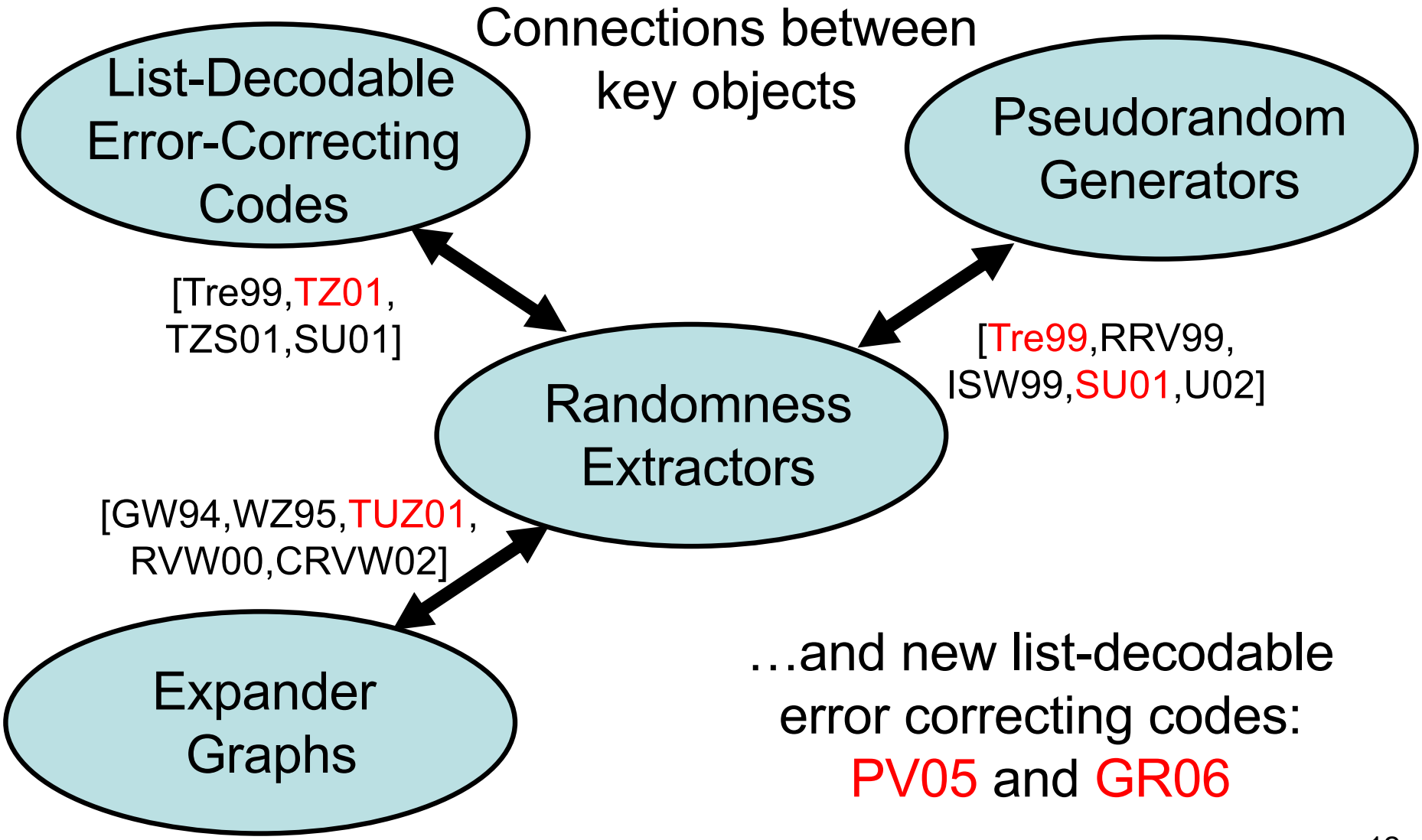
short seed	log n + O(1)	O(log n) or $O(\log^2 n)$
short output	m=k + t + O(1)	$m = k^{1+\alpha}$ or m = O(k)

- we obtained lossless condensers with:
  - seed O(log n)
  - output m = O(k)

Equivalently: bipartite graphs with $(1-\epsilon)$(degree) expansion, and left-degree and RHS size within poly of optimal

What was happening in 2007

What was happening circa 2007



What was happening circa 2007

1999 Luca Trevisan: extractors from error-correcting codes and Nisan-Wigderson PRG

2001 Ta-Shma + Zuckerman: very good list-decodable codes equivalent to extractors

2001 Shaltiel + Umans: extractors from Reed-Muller codes

Developments in coding theory

- RS codes: $f(X) \in \mathrm{F}_q[X]$ of deg. k, $\mathrm{n} \leq q$

f(1)	f(2)	f(3)					f(n)

  – list-decodable from $(kn)^{1/2}$ agreement

- PV 2005: r correlated RS codewords

  E(X) irreducible of deg. k+1, $f_i = f^{h^i} \bmod E$

$f_1(1)$	$f_1(2)$	$f_1(3)$					$f_1(n)$
$f_2(1)$	$f_2(2)$	$f_2(3)$					$f_2(n)$
$f_r(1)$	$f_r(2)$	$f_r(3)$					$f_r(n)$

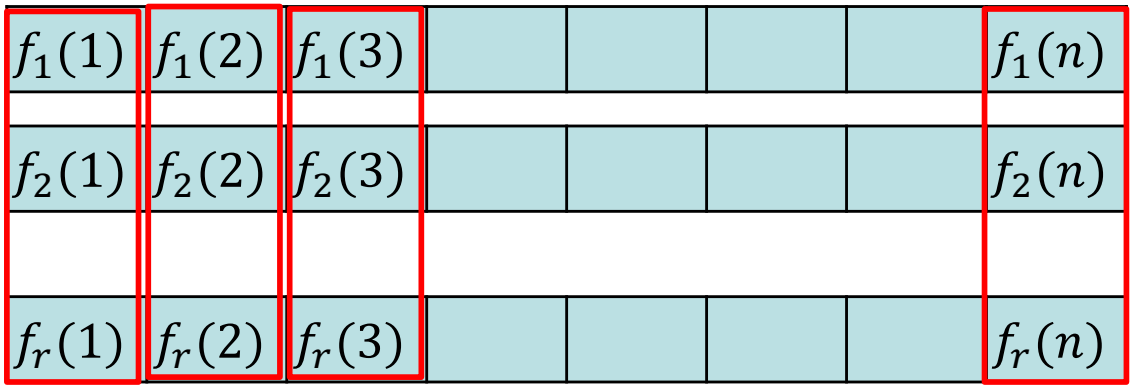
list-dec. from $k^{\frac{r}{r+1}} n^{\frac{1}{r+1}}$ agr.

Developments in coding theory

- GR 2006: r correlated RS codewords are *shifts* of original f

E(X) = $X^q - \alpha$ irreducible, $f_i = f^{q^i}$ mod E

results in $f_i(X) = f(\alpha^i X)$



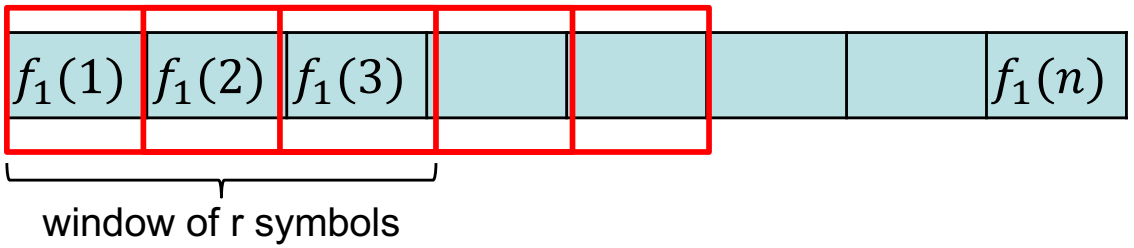
list-dec. from $k^{\frac{r}{r+1}} n^{\frac{1}{r+1}}$ agr.

Developments in coding theory

- GR 2006: r correlated RS codewords are *shifts* of original f

E(X) = $X^q - \alpha$ irreducible, $f_i = f^{q^i}$ mod E

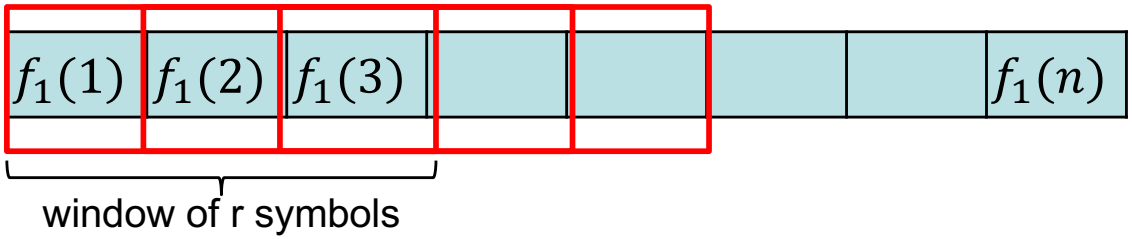
results in $f_i(X) = f(\alpha^i X)$



list-dec. from $k^{\frac{r}{r+1}} n^{\frac{1}{r+1}}$ agr.

What was happening in 2007

- Called “folded Reduced-Solomon” codes



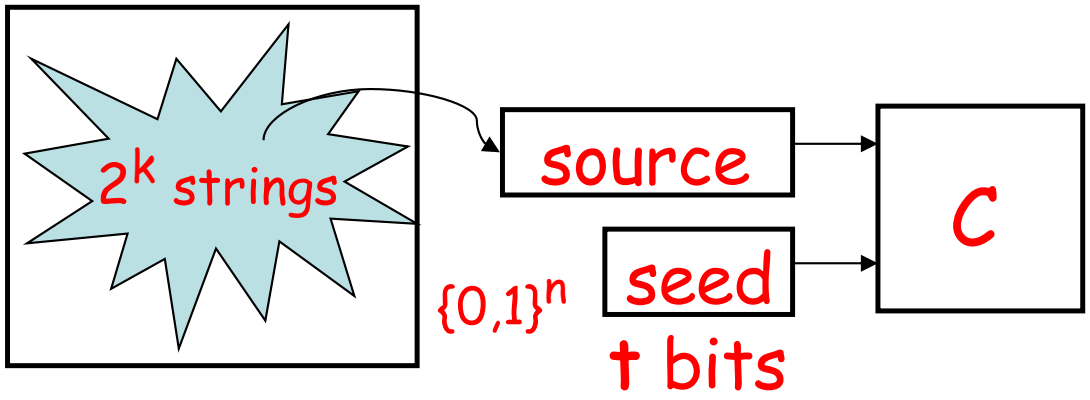
list-dec. from $k^{\frac{r}{r+1}} n^{\frac{1}{r+1}}$ agr.

If f is replaced by RM codeword, this *is* the TZ “extractor code” of the Shaltiel-Umans extractor

…the three GUV authors were skipping a talk at a Banff workshop and talking about this…

Construction
and
proof of main result

The construction



- F_q finite field
- parameter $h \leq q$
- deg. n polynomial $E(Y)$ irreducible over F_q
  - source: degree n-1 univariate polynomial f
  - define $f_i(Y) = f^{h^i}(Y) \bmod E(Y)$

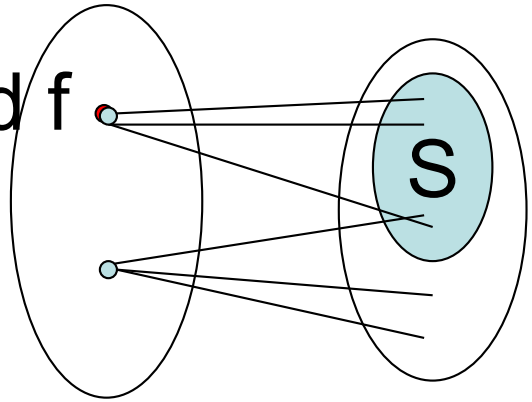
$$C(f, y \in F_q) = (y, f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$$

Proof outline

define: $f_i(Y) = f^{h^i}(Y)$ mod irreducible $E(Y)$

$C(f, y \in F_q) = (y, f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$

Expander/lossless cond. equivalence reduces task to proving expansion



– fix $S \subseteq F_q^{m+1}$ of size $(1-\epsilon)q(h^m - 1)$

– "bad" f: for all y, $(y, f_0(y), f_1(y), \ldots, f_{m-1}(y)) \in S$

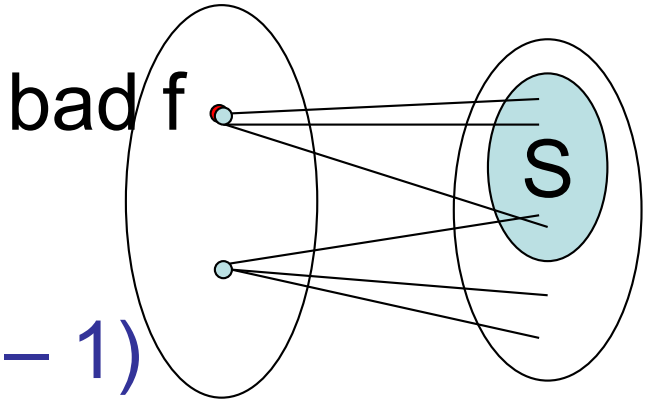
– show number of bad f's is small

– "small": $h^m - 1$ (almost best possible: $|S|/q$)

Proof of expansion

define: $f_i(Y) = f^{h^i}(Y)$ mod irreducible $E(Y)$

$C(f, y \in F_q) = (y, f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$



- Proof:
  - fix $S \subseteq F_q^m$ of size $(1-\epsilon)q(h^m - 1)$
  - non-zero poly.* $Q(Y, Y_0, Y_1, \ldots, Y_{m-1})$ vanishing on S; $\deg(Y) \leq (1-\epsilon)q$, $\deg(Y_i) \leq (h-1)$
  - "bad" f: for all y, $Q(y, f_0(y), \ldots, f_{m-1}(y)) = 0$
  - if $\varepsilon q > nhm$, $R_f(Y) \stackrel{\text{def}}{=} Q(f_0(Y), \ldots, f_{m-1}(Y)) = 0$.

* WLOG not divisible by E(Y)

Proof of expansion

define: $f_i(Y) = f^{h^i}(Y)$ mod irreducible $E(Y)$

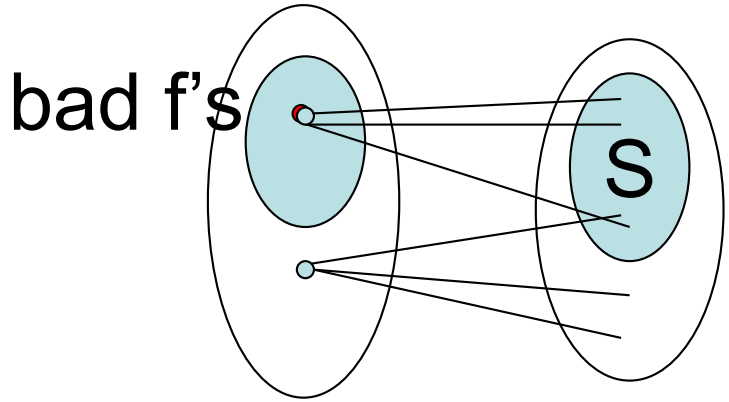
$C(f, y \in F_q) = (y, f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$

- Proof (continued):
  - "bad" f: $R_f(Y) \stackrel{\text{def}}{=} Q(Y, f_0(Y), \ldots, f_{m-1}(Y)) = 0$
  - view f_i as elements of $F_q[Y]/E(Y)$ and Q as a (non-zero) polynomial over extension field
  - for "bad" f: $(f_0, \ldots, f_{m-1})$ is a root of Q
  - f is a root of $Q'(Z) \stackrel{\text{def}}{=} Q(Z, Z^h, Z^{h^2}, \ldots, Z^{h^{m-1}})$

Proof of expansion

define: $f_i(Y) = f^{h^i}(Y)$ mod irreducible $E(Y)$

$C(f, y \in F_q) = (y, f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$



- Proof (continued):
  - for "bad" f:

    f is a root of $Q'(Z) \stackrel{\text{def}}{=} Q(Z, Z^h, Z^{h^2}, \ldots, Z^{h^{m-1}})$
  - number of "bad" f ≤ degree(Q')

    $\leq (h-1)(1 + h + h^2 + \ldots + h^{m-1})$

    $= (h-1)((h^m-1)/(h-1)) = h^m-1$ ■

Condenser parameters

define: $f_i(Y) = f^{h^i}(Y)$ mod irreducible $E(Y)$

$C(f, y \in F_q) = (y, f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$

- proof requires **q** > nm**h**/ε
- output length: (m+1)·log **q**
- entropy: |bad f's| = m·log **h**

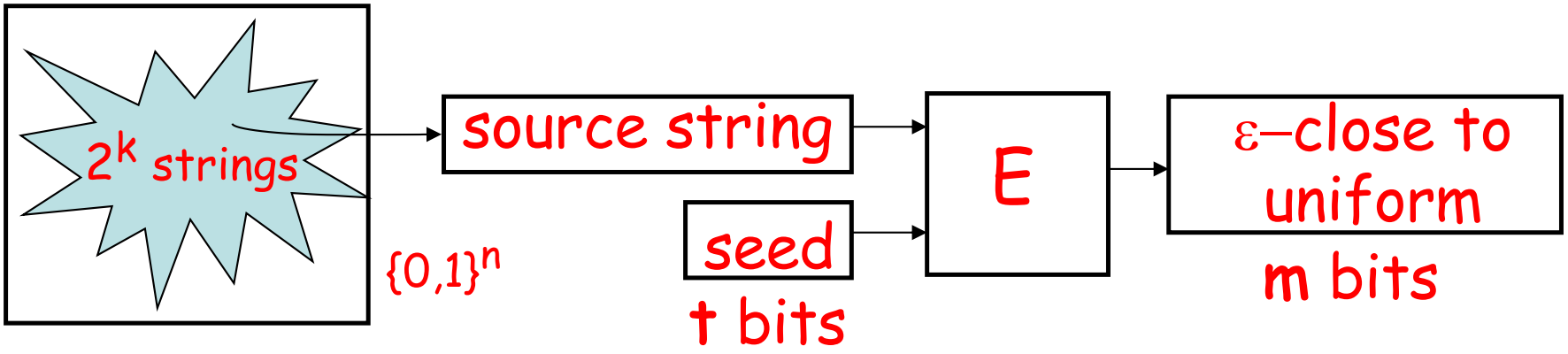
"entropy rate"

- $h = (nm/\varepsilon)^c$, $q = h^{(c+1)/c}$ so (log h)/(log q) = c/(c+1)
- we chose h = poly(n/ϵ), so q = poly(n/ϵ),
  - seed length t = log q = O(log n + log(1/ε))

Getting an extractor

Zuckerman 97: extract $(1-\alpha)k$ bits from k-minentropy for any constant entropy rate with seed $t = O(\text{optimal})$

- First apply condenser, then extract $(1-\alpha)k$ bits from k-minentropy with $t = O(\text{optimal})$



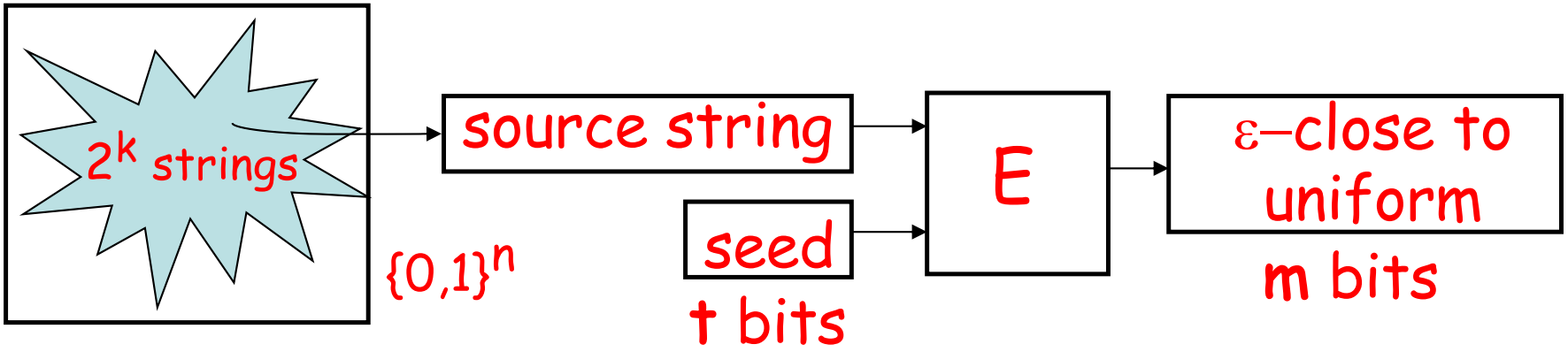
Our results (recap)

Extractor:	optimal:	[GUV]:
short seed	log n + O(1)	O(log n)
long output	m = k + t - O(1)	m = (1 - α)k
	best to date [DKSS09, TU11]	$\alpha = 1/\mathrm{polylog}(n)$

Lossless Condenser:	optimal:	[GUV]
short seed	log n + O(1)	O(log n)
short output	m = k + t + O(1)	m = (1 + α)k

Why was this paper influential?

Why was this paper influential?



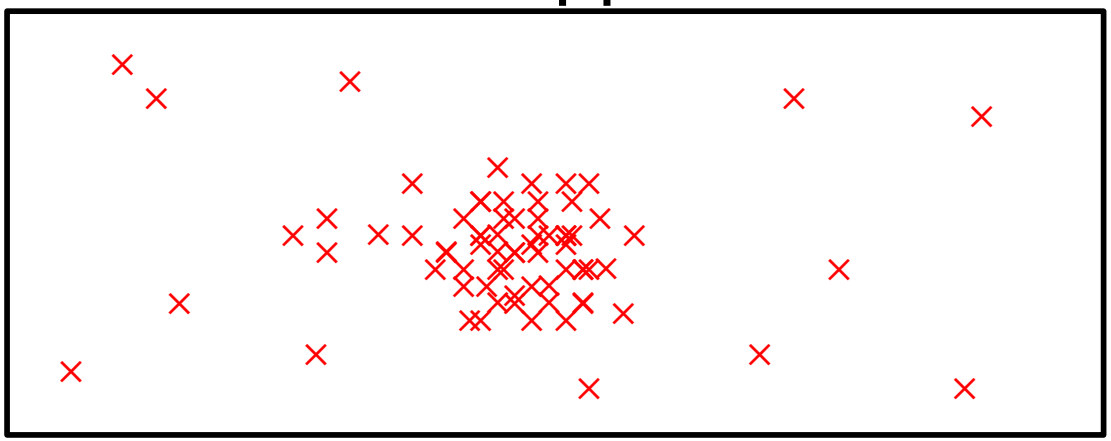
Extractor applications

Seed length:

$\mathrm{poly}(\log n)$

$O(\log n)$

$\log n$



Output length →

k^α αk $(1-\alpha)k$ k $k+t$

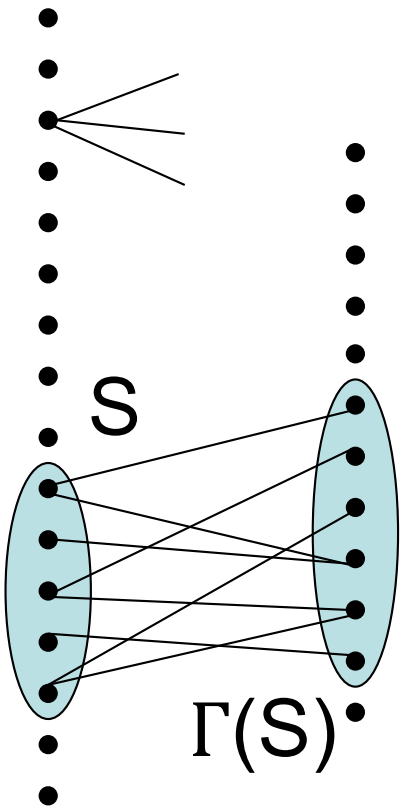
Why was this paper influential?

- The construction and proof are
  - easy!
  - self-contained
  - algebraically “nice”
  - adaptable / modifiable for other applications

Why was this paper influential?

- the expanders have near-optimal expansion which gives the unique-neighbor property

(1 - α) fraction of nodes have a unique neighbor



- Analysis gives efficient list-recovery of PV codes from very large input lists

Some later works that used GUV in interesting ways

Later works that used GUV

- Cheraghchi + Indyk 2016:
  - "Nearly Optimal Deterministic Algorithm for Sparse Walsh–Hadamard Transform"
  - careful choice of parameters makes the function C(f, y) linear for each fixed seed y
  - more generally: compressed sensing from linear lossless condensers, GUV with modification has near-optimal parameters and fast reconstruction/decoding.

Later works that used GUV

- Kane, Nelson + Woodruff 2009:
  - "Revisiting Norm Estimation in Data Streams"
  - re-engineers GUV construction to make it space-efficient
  - uses in Armoni PRG for bounded-space branching programs to obtain improvements:
    - derandomizing ℓ_p estimators
    - Improved streaming dimensionality reduction.

Later works that used GUV

- Cheraghchi 2009
  - "Capacity Achieving Codes from Randomness Condensers"
- Bauwens + Zimand 2020
  - "Universal Codes in the Shared-Randomness Model for Channels with General Distortion Capabilities"
- both need linear lossless condensers with near optimal parameters, and decoding

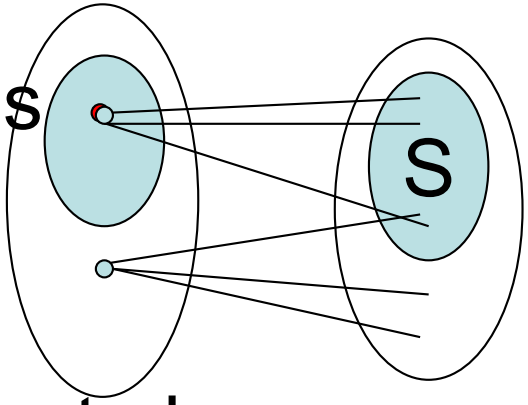
Later works that used GUV

- Kalev + Ta-Shma 2022
  - "Unbalanced Expanders from Multiplicity Codes"
  - replace correlated PV codewords $f_1, f_2, \ldots, f_m$ with multiplicity code codewords $f', f'', \ldots, f^{(m)}$
  - entirely different and beautiful proof using differential equations, ideas from list-decoding algorithms for multiplicity codes
  - similar parameters
  - Chattopadhyay, Gurumukhani, Ringach, Zhao 2025: two-sided expander

An open problem
and a
step toward a solution

Recall condenser parameters

- proof requires **q** > nm**h**/ε
- |S| = h^m
- right-hand-side size = q^m
- inequality forces constant entropy rate because



RHS size = $q^m \geq (h^m)^{1+\alpha}$ for const. $\alpha > 0$

|S| ≈ |RHS|/poly(n) "logarithmic entropy rate deficiency" yields optimal output length extractors with O(log n) seed length. Can this be achieved?

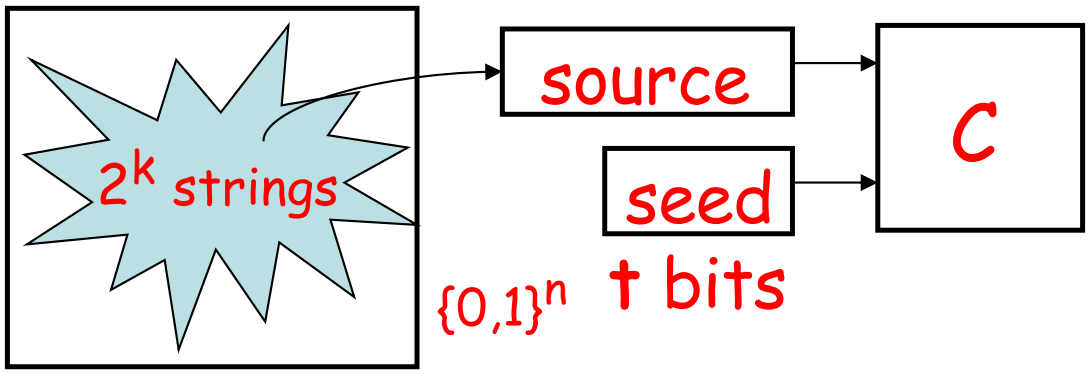
Improving the entropy rate to 1-o(1)

- Ta-Shma + Umans 2011, new ideas:
  - lossy condenser analysis (almost the same)
  - total-degree bound on Q, vanishing on set S with high multiplicity
  - generalization of the "PV curve"

$$f \mapsto (f, f^h, f^{h^2}, \ldots, f^{h^{m-1}})$$

  - elements of $F_{q=h^2}$ are linear polys over F_h
  - two layers of "GUV analysis"

First modification



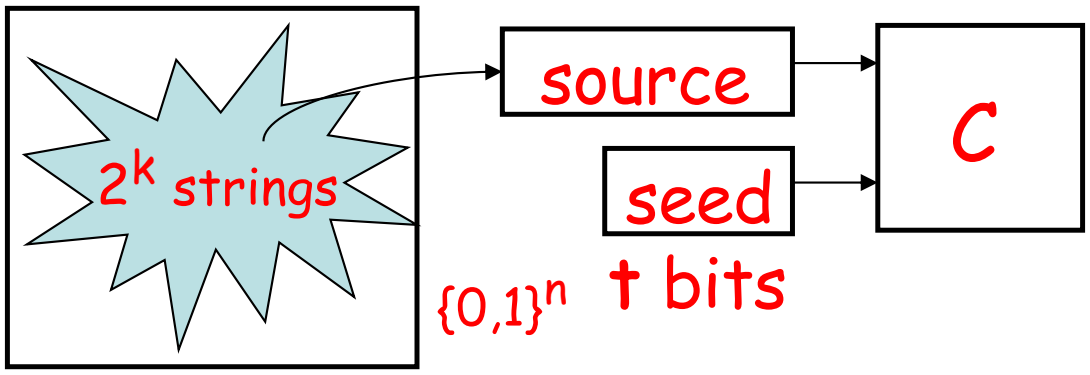
- F_q finite field
- deg. n polynomial E(Y) irreducible over F_q
  - source: degree n-1 univariate polynomial f
  - $f_i(Y) = f^i(Y) \bmod E(Y)$, $G_i(f)(Y)$ for $G_i: F_{q^n} \to F_{q^n}$

$$C(f, y \in F_q) = (f_0(y), f_1(y), f_2(y), \cdots, f_{m-1}(y))$$

(deg(G_i) will be h^{m-1} – same as before)

Second modification

degree 2 extension



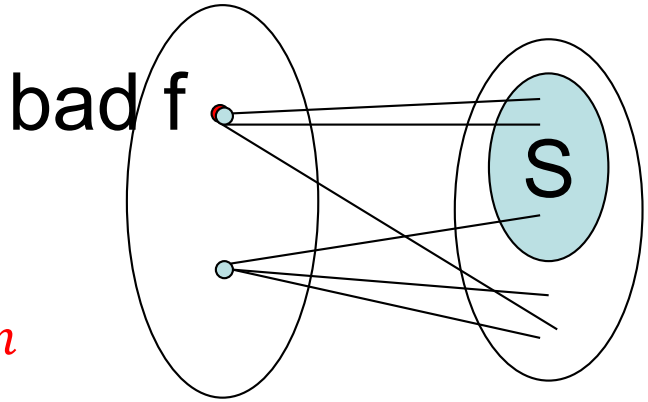
- $F_q = F_h[Z]/D(Z)$
- deg. n polynomial $E(Y)$ irreducible over F_q
  - source: degree n-1 univariate polynomial f
  - $f_i(Y) = G_i(f)(Y)$ for $G_i: F_{q^n} \rightarrow F_{q^n}$

$$C(f;\ y \in F_q,\ w \in F_h) = (f_0(y)(w),\ f_1(y)(w),\ f_2(y)(w),\ \cdots,\ f_{m-1}(y)(w))$$

now C maps into F_h^m

Proof of expansion

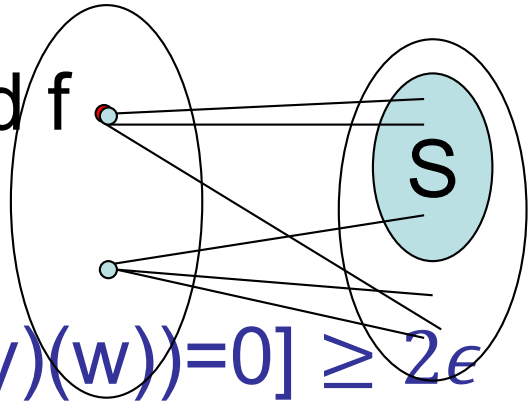
define: $f_i(Y) = G_i(f)(Y)$ mod irreducible $E(Y)$
$C(f; y \in F_q, w \in F_h) = (f_0(y)(w), f_1(y)(w), \cdots, f_{m-1}(y)(w))$



- Proof:
  - fix $S \subseteq F_h^m$ of size $\approx (\epsilon h/2)^m$
  - non-zero poly. $Q(Z_0, Z_1, \ldots, Z_{m-1})$ vanishing on S with multiplicity 2r, total degree ϵhr -1
  - "bad" f: $\Pr_{y,w}[Q(f_0(y)(w),\ldots,f_{m-1}(y)(w))=0] \geq 2\epsilon$

Proof of expansion

define: $f_i(Y) = G_i(f)(Y)$ mod irreducible $E(Y)$
$C(f;\ y \in F_q,\ w \in F_h) = (f_0(y)(w),\ f_1(y)(w),\ \cdots,\ f_{m-1}(y)(w))$



- Proof (continued):
  - "bad" f: $\Pr_{y,w}[Q(f_0(y)(w),\ldots,f_{m-1}(y)(w))=0] \geq 2\epsilon$
  - averaging: for ϵ fraction of y's

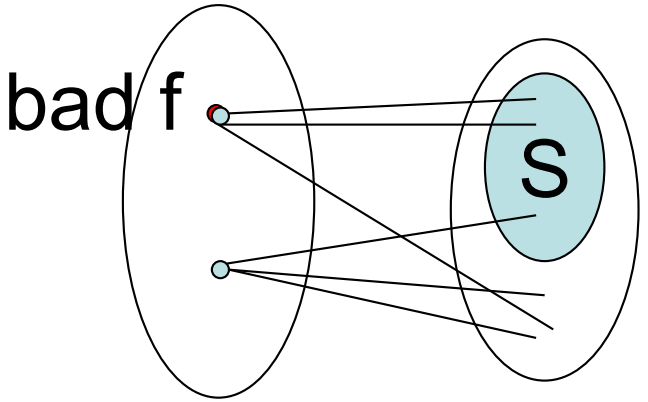
$$\Pr_w[Q(f_0(y)(w),\ \ldots,\ f_{m-1}(y)(w)) = 0] \geq \epsilon$$

  - because $\epsilon hr > \deg(Q)$, for these y's we have

$$P_{f,y}(W) \stackrel{\text{def}}{=} Q(f_0(y)(W),\ \ldots,\ f_{m-1}(y)(W)) = 0$$

Proof of expansion

define: $f_i(Y) = G_i(f)(Y)$ mod irreducible $E(Y)$
$C(f;\ y \in F_q,\ w \in F_h) = (f_0(y)(w), f_1(y)(w), \cdots, f_{m-1}(y)(w))$



- Proof (continued):
  - for the ϵq y's for which

$$P_{f,y}(W) \stackrel{\text{def}}{=} Q(f_0(y)(W), \ldots, f_{m-1}(y)(W)) = 0$$

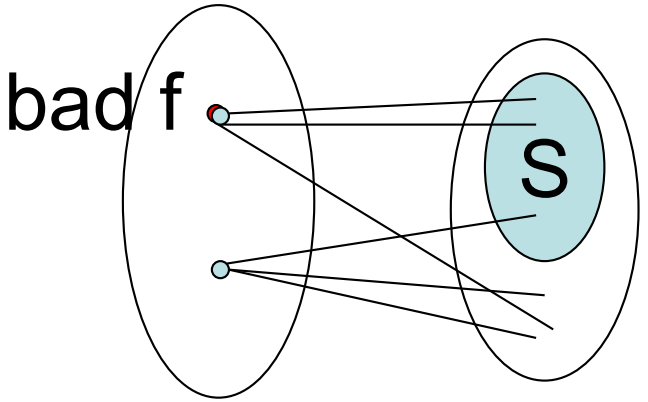
we have $Q(f_0(y), \ldots, f_{m-1}(y)) = 0$

  - because $\epsilon q > n \cdot \deg(Q)$ (pick h large enough)

$$R_f(Y) \stackrel{\text{def}}{=} Q(f_0(Y), \ldots, f_{m-1}(Y)) = 0$$

Proof of expansion

define: $f_i(Y) = G_i(f)(Y)$ mod irreducible $E(Y)$
$C(f; y \in F_q, w \in F_h) = (f_0(y)(w), f_1(y)(w), \cdots, f_{m-1}(y)(w))$



- Proof (continued):
  - for the f's for which

$$R_f(Y) \stackrel{\text{def}}{=} Q(f_0(Y), \ldots, f_{m-1}(Y)) = 0$$

we have $Q(f_0, \ldots, f_{m-1}) = 0$

  - f a root of $Q'(Z) \stackrel{\text{def}}{=} Q(G_1(Z), G_2(Z), \ldots, G_m(Z))$
  - number of "bad" f ≤ degree(Q') ≤ ϵrhm

Improving the entropy rate to 1-o(1)

- Sets of size $\approx h^m$ expand to sets size

$$\approx \left(\frac{h}{const}\right)^m.$$

- Gives $m \log h \rightarrow m(\log h - O(1))$ cond.
- Entropy rate: (log h – O(1))/(log h)

- Condense+extract: seed length t=O(log n), output length m = (1 – 1/polylog(n))k

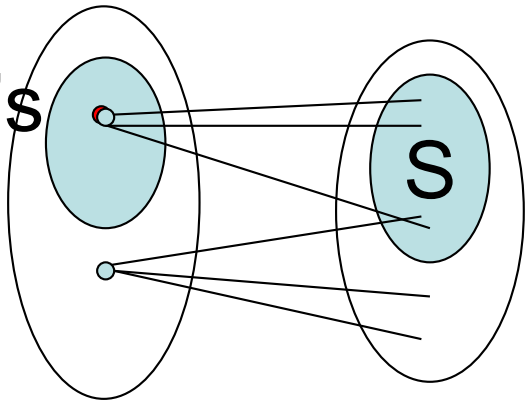
Open problem again

- Handle sets S of size |RHS|/poly(n)

  "logarithmic entropy rate deficiency"

- # bad f's at most poly(n) ·|S|

  "logarithmic entropy loss"



- Would yield optimal output length extractors, with O(log n) seed length.

Thank you!