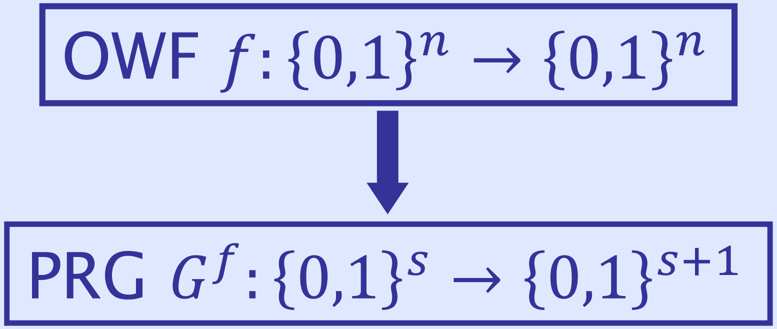
Pseudorandom Generators from One-Way Functions via Computational Entropy

Salil Vadhan
Harvard University

DIMACS Workshop on Complexity of Cryptographic Primitives and Assumptions
June 9, 2017

PRGs from OWFs

Thm [Hastad-Impagliazzo-Levin-Luby `90]:



OWF $f:\{0,1\}^n \to \{0,1\}^n$

PRG $G^f:\{0,1\}^s \to \{0,1\}^{s+1}$

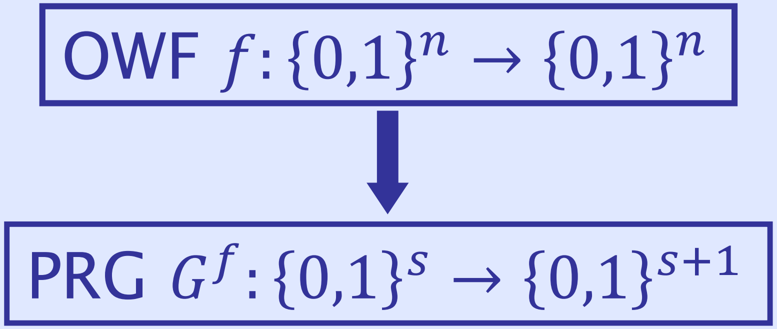
Efficiency measures:

- Seed length: $s = \tilde{O}(n^{10})$ [HILL89], $s = \tilde{O}(n^8)$ [H06].
- # queries to f: $q = \tilde{O}(n^9)$ [HILL89], $s = \tilde{O}(n^7)$ [H06].

[seed = q independent evaluation pts + hash functions]

PRGs from OWFs

Thm [Haitner-Reingold-Vadhan `10, Vadhan-Zheng `11]:



OWF $f\colon \{0,1\}^n \to \{0,1\}^n$

PRG $G^f\colon \{0,1\}^s \to \{0,1\}^{s+1}$

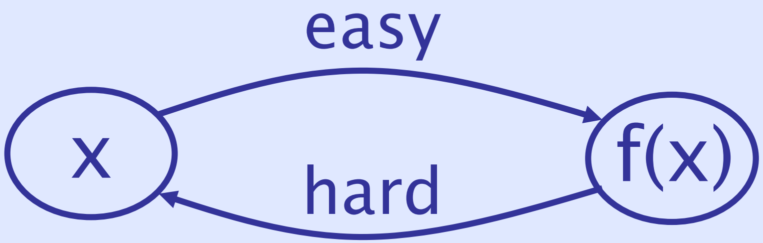
Efficiency measures:

- Seed length: $s = \tilde{O}(n^4)$ [HRV10], $s = \tilde{O}(n^3)$ [VZ11].
- # queries to f: $q = \tilde{O}(n^3)$ [HRV10,VZ11].

Outline

- OWFs & Cryptography
- Notions of pseudoentropy
- OWPs ⇒ PRGs
- OWFs ⇒ PRGs
- Open problems
- Inaccessible Entropy (time permitting)

One-Way Functions [DH76]



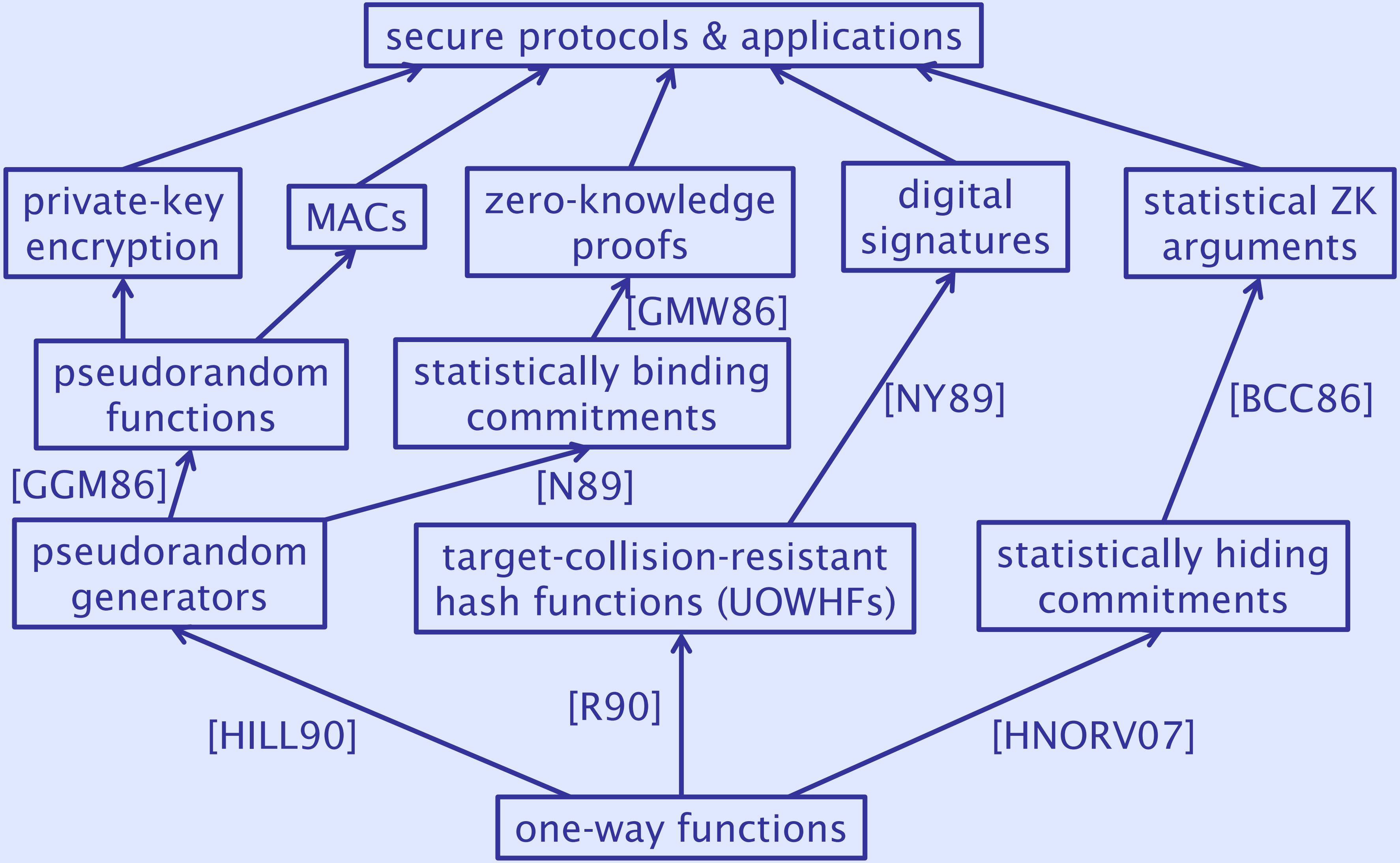
- Candidate: $f(x,y) = x \cdot y$

Formally, a **OWF** is $f : \{0,1\}^n \to \{0,1\}^n$ s.t.

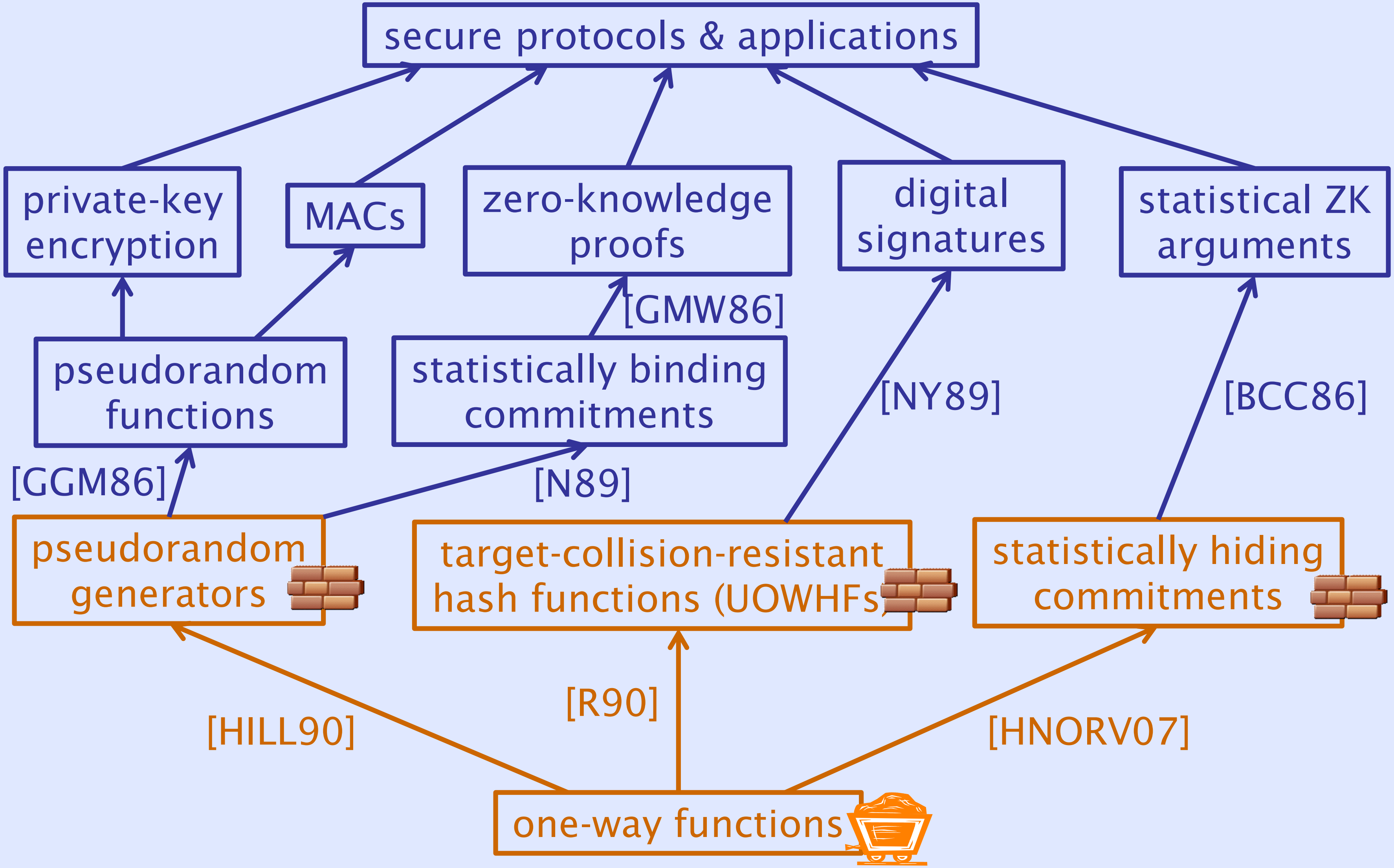
- f poly-time computable
- $\forall$ poly-time A

  $\Pr[A(f(X)) \in f^{-1}(f(X))] = 1/n^{\omega(1)}$ for $X \leftarrow \{0,1\}^n$

OWFs & Cryptography



OWFs & Cryptography



Computational Entropy
[Y82,HILL90,BSW03]

Question: How can we use the “raw hardness” of a OWF to build useful crypto primitives?

Answer [HILL90,R90,HRVW09,…]:

- Every crypto primitive amounts to some form of **“computational entropy”.**
- One-way functions already have a little bit of **“computational entropy”.**

Outline

- OWFs & Cryptography
- Notions of pseudoentropy
- OWPs ⇒ PRGs
- OWFs ⇒ PRGs
- Open problems
- Inaccessible Entropy (time permitting)

Entropy

Def: The **Shannon entropy** of r.v. X is

$$H(X) = E_{x\leftarrow X}[\log(1/\Pr[X=x])]$$

- $H(X)$ = "Bits of randomness in X (on avg)"
- $0 \leq H(X) \leq \log|\mathrm{Supp}(X)|$

X concentrated
on single point

X uniform on
Supp(X)

- **Conditional Entropy:** $H(X|Z) = E_{z\leftarrow Z}[H(X|_{Z=z})]$

(Conditional) Min-Entropy

- **Min-Entropy:**

$$\mathrm{H}_\infty(X) = \min_x \log\left(\frac{1}{\Pr[X{=}x]}\right) = \log\left(\frac{1}{\max_x \Pr[X{=}x]}\right)$$

- **Average Min-Entropy:**

[Dodis-Ostrovsky-Reyzin-Smith `04]

$$\mathrm{H}_\infty(X|Z) = \log\left(\frac{1}{\mathrm{E}_{z\leftarrow Z}\left[\max_x \Pr[X = x|Z = z]\right]}\right)$$

Average Min-Entropy [DORS04]

$$\mathrm{H}_\infty(X|Z) = \log\left(\frac{1}{\mathrm{E}_{z\leftarrow Z}\left[\max_x \Pr[X = x|Z = z]\right]}\right)$$

Properties:

- Equals “guessing entropy”:
  - $\mathrm{H}_\infty(X|Z) = \log\left(\frac{1}{\max_A \Pr[A(Z)=X]}\right)$
- Supports randomness extraction:
  - $(\mathrm{Ext}(X;R), R, Z) \approx_\epsilon (U_m, R, Z)$
  - With m as large as $\mathrm{H}_\infty(X|Z) - 2\log(1/\epsilon) - O(1)$

(HILL) Pseudoentropy

Def [HILL90]: X has **pseudoentropy** $\geq$ k iff there exists a random variable Y s.t.

1. $Y \equiv^c X$
2. $H(Y) \geq k$

Interesting when $k > H(X)$, i.e.

Pseudoentropy > Real Entropy,

e.g. X = output of a PRG

(HILL) Pseudoentropy variants

Def [Hsiao-Lu-Reyzin `07]:

X has **pseudoentropy** $\geq$ k given Z iff

$\exists$ a random variable Y s.t.

1. (Y,Z) $\equiv^c$ (X,Z)
2. H(Y|Z) $\geq$ k

Pseudo-min-entropy: require $\mathrm{H}_\infty(Y|Z) \geq k$.

- Supports randomness extraction:

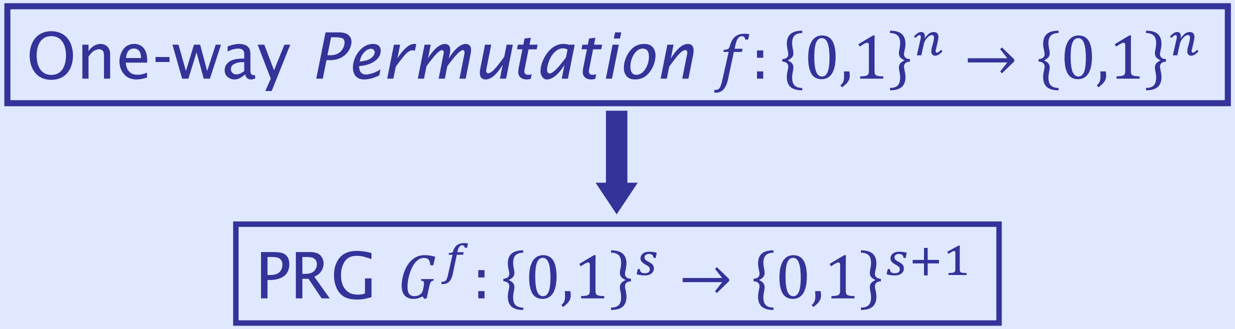
  if Ext is efficiently computable, then
  - $(\mathrm{Ext}(X;R),R,Z) \equiv^c (U_m,R,Z)$
  - With m as large as $k-2\log(1/\epsilon)-O(1)$

Outline

- OWFs & Cryptography
- Notions of pseudoentropy
- OWPs $\Rightarrow$ PRGs
- OWFs $\Rightarrow$ PRGs
- Open problems
- Inaccessible Entropy (time permitting)

OWPs ⇒ PRGs

Thm [Blum-Micali `82, Yao `82, Goldreich-Levin `89]:



One-way *Permutation* $f : \{0,1\}^n \to \{0,1\}^n$

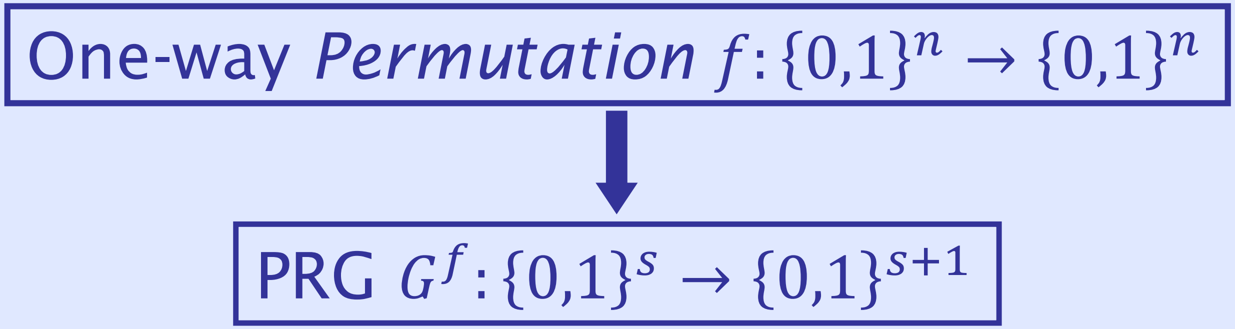
PRG $G^f : \{0,1\}^s \to \{0,1\}^{s+1}$

Efficiency measures:

- Seed length: $s = O(n)$ [GL89]
- # queries to f: $q = 1$ [GL89].

OWPs ⇒ PRGs

Thm [Blum-Micali `82, Yao `82, Goldreich-Levin `89]:



One-way *Permutation* $f:\{0,1\}^n \to \{0,1\}^n$

PRG $G^f:\{0,1\}^s \to \{0,1\}^{s+1}$

Efficiency measures:

- Seed length: $s = O(n)$ [GL89]
- # queries to f: $q = 1$ [GL89].

OWPs ⇒ PRGs

Modern interpretation of proof:

- For $X \leftarrow \{0,1\}^n$, given $f(X)$, X has $\omega(\log n)$ **guessing pseudoentropy** [Hsiao-Lu-Reyzin `07]

  $\forall$ poly-time A, $\Pr[A(f(X))=X] \leq 1/n^{\omega(1)}$

  Note: ordinary pseudoentropy is negligible!

- **Supports randomness extraction:** if Ext is a "reconstructive extractor" then:
  - $(\mathrm{Ext}(X;R), R, Z) \equiv^c (U_m, R, Z)$
  - With m as large as $k - 2\log(1/\epsilon) - O(1)$.

  [Goldreich-Levin`89, Trevisan`99, Ta-Shma-Zuckerman`01, …]

Guessing pseudoentropy vs. HILL pseudoentropy

Can be very different in general (as we saw), but are equivalent for *short* random variables:

Thm [Impagliazzo `95,..., VZ `12, SGP `15]:
Let (X,Z) $\in$ $\{0,1\}^{O(\log n)} \times \{0,1\}^n$

Guessing pseudoentropy of X given Z
$\geq k$
$\Updownarrow$
Pseudo-min-entropy of X given Z
is $\geq k$

Guessing pseudoentropy vs. HILL pseudoentropy

Can be very different in general (as we saw), but are equivalent for *short* random variables:

Thm [Impagliazzo `95,…, VZ `12, SGP `15]:
Let (X,Z) $\in \{0,1\}^{O(\log n)} \times \{0,1\}^n$

Guessing pseudoentropy of X given Z

$\geq k \pm negl(n)$

$\Updownarrow$

Pseudo-min-entropy of X given Z

is $\geq k$

Guessing pseudoentropy vs. HILL pseudoentropy

Can be very different in general (as we saw), but are equivalent for *short* random variables:

Thm [Impagliazzo `95,…, VZ `12, SGP `15]:
Let (X,Z) $\in \{0,1\}^{O(\log n)} \times \{0,1\}^n$

Guessing pseudoentropy of X given Z

$\geq k$

$\Updownarrow$

Pseudo-min-entropy of X given Z

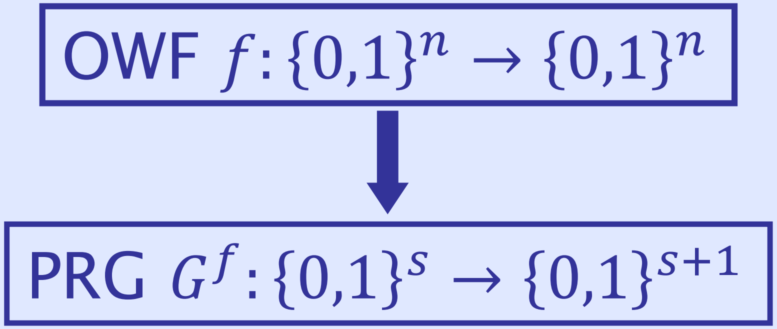
is $\geq k$

Outline

- OWFs & Cryptography
- Notions of pseudoentropy
- OWPs ⇒ PRGs
- OWFs ⇒ PRGs
- Open problems
- Inaccessible Entropy (time permitting)

PRGs from OWFs

Thm [Haitner-Reingold-Vadhan `10, Vadhan-Zheng `11]:



OWF $f: \{0,1\}^n \to \{0,1\}^n$

PRG $G^f: \{0,1\}^s \to \{0,1\}^{s+1}$

Efficiency measures:

- Seed length: $s = \tilde{O}(n^4)$ [HRV10], $s = \tilde{O}(n^3)$ [VZ11].
- # Queries to f: $q = \tilde{O}(n^3)$ [HRV10,VZ11].

Pseudoentropy in a OWF

- **Still true:** For $X \leftarrow \{0,1\}^n$, given $f(X)$, X has $\omega(\log n)$ guessing pseudoentropy:

  $\forall$ poly-time A, Pr[A(f(X))=X] $\leq 1/\mathrm{n}^{\omega(1)}$

- But this may be for trivial information-theoretic reasons, e.g. f(x)=first half of x.
- How to capture *gap* between information-theoretic and computational hardness in X given $f(X)$?

Pseudoentropy in a OWF

Lemma [VZ11]: For $X \leftarrow \{0{,}1\}^n$, given $f(X)$, X has $\omega(\log n)$ **sampling relative entropy**:

for every probabilistic poly-time A
D((f(X),X) || (f(X),A(f(X))) $\geq \omega(\log n)$.

[D = relative entropy/KL Divergence]

cf. distributional one-way functions
[Impagliazzo-Luby `89]: D→ statistical distance

Pseudoentropy in a OWF

Lemma [VZ11]: For $X \leftarrow \{0{,}1\}^n$, given $f(X)$, X has $\omega(\log n)$ **sampling relative entropy**:

for every probabilistic poly-time A
D((f(X),X) || (f(X),A(f(X))) $\geq \omega(\log n)$.

Proof: Applying test T(y,x) = $\begin{cases} 1 \text{ if } y = f(x) \\ 0 \text{ otherwise} \end{cases}$

D((f(X),X) || (f(X),A(f(X)))
$\geq$ D(Bernoulli(1) || Bernoulli($n^{-\omega(1)}$))
$= \log(1/n^{-\omega(1)}) = \omega(\log n)$.

Sampling Relative Entropy vs. Pseudoentropy

Thm [VZ11]: Let $(X,Z) \in \{0,1\}^{O(\log n)} \times \{0,1\}^n$.

X has sampling relative entropy $\geq k$ given Z, i.e. for every probabilistic poly-time A

$$D((Z,X)\|(Z,A(Z))) \geq k$$

$$\Updownarrow$$

The pseudoentropy of X given Z is $\geq H(X|Z)+k$

Problems & solutions:

- Our X is long → break into small pieces
- Can't extract from Shannon entropy → repetitions

Next-bit Pseudoentropy

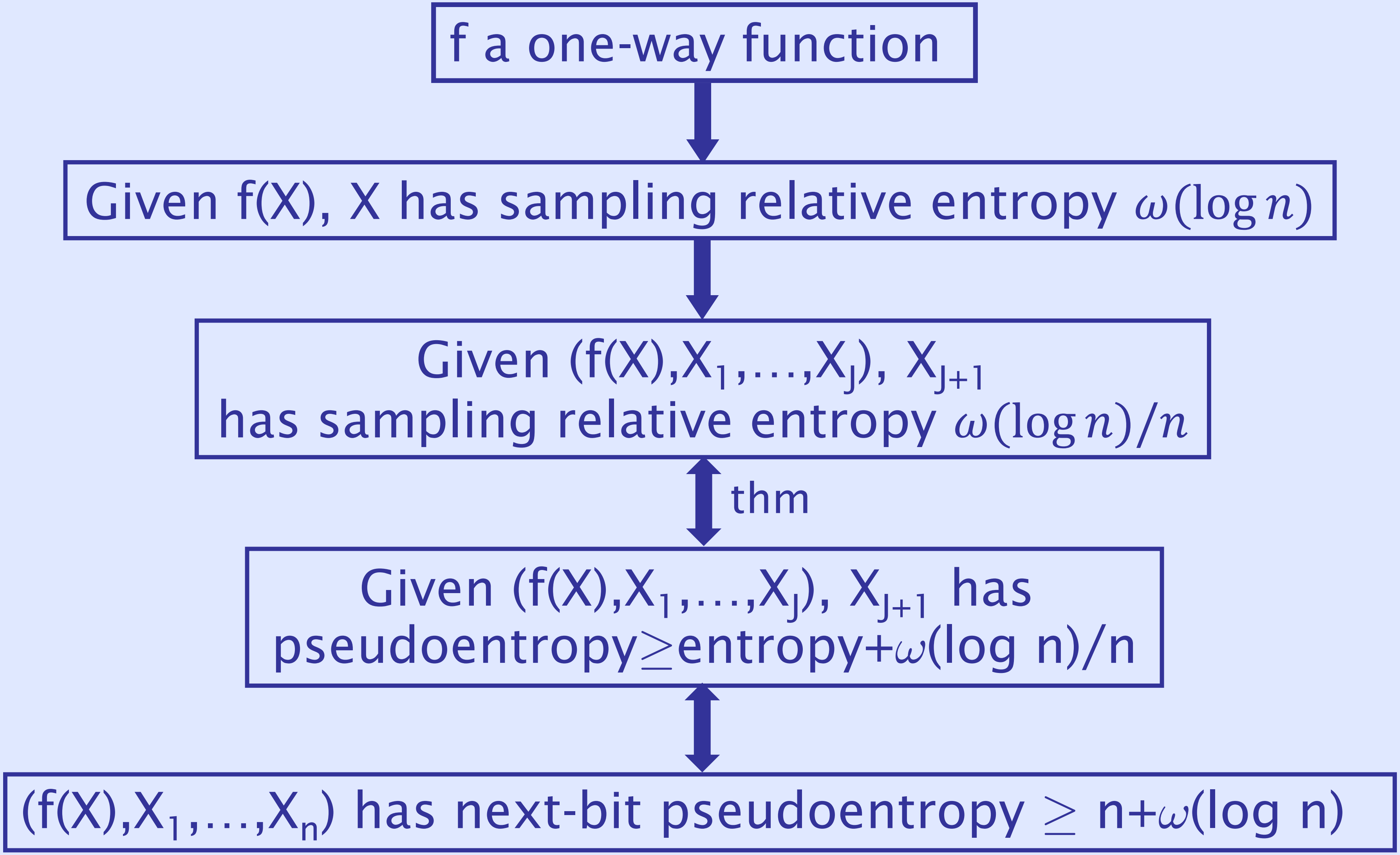
- Thm [HRV10,VZ11]: $(f(X),X_1,\ldots,X_n)$ has **"next-bit pseudoentropy"** $\geq n+\omega(\log n)$.

- Note: $(f(X),X)$ easily distinguishable from every random variable of entropy $> n$.

- Next-bit pseudoentropy: $\exists\ (Y_1,\ldots,Y_n)$ s.t.
  - $(f(X),X_1,\ldots,X_i) \equiv^c (f(X),X_1,\ldots,X_{i-1},Y_i)$
  - $H(f(X))+\sum_i H(Y_i|f(X),X_1,\ldots,X_{i-1}) = n+\omega(\log n)$.

  cf. next-bit unpredictability [Blum-Micali `82]

Next-Bit Pseudoentropy from OWF: Proof Sketch



f a one-way function

Given f(X), X has sampling relative entropy $\omega(\log n)$

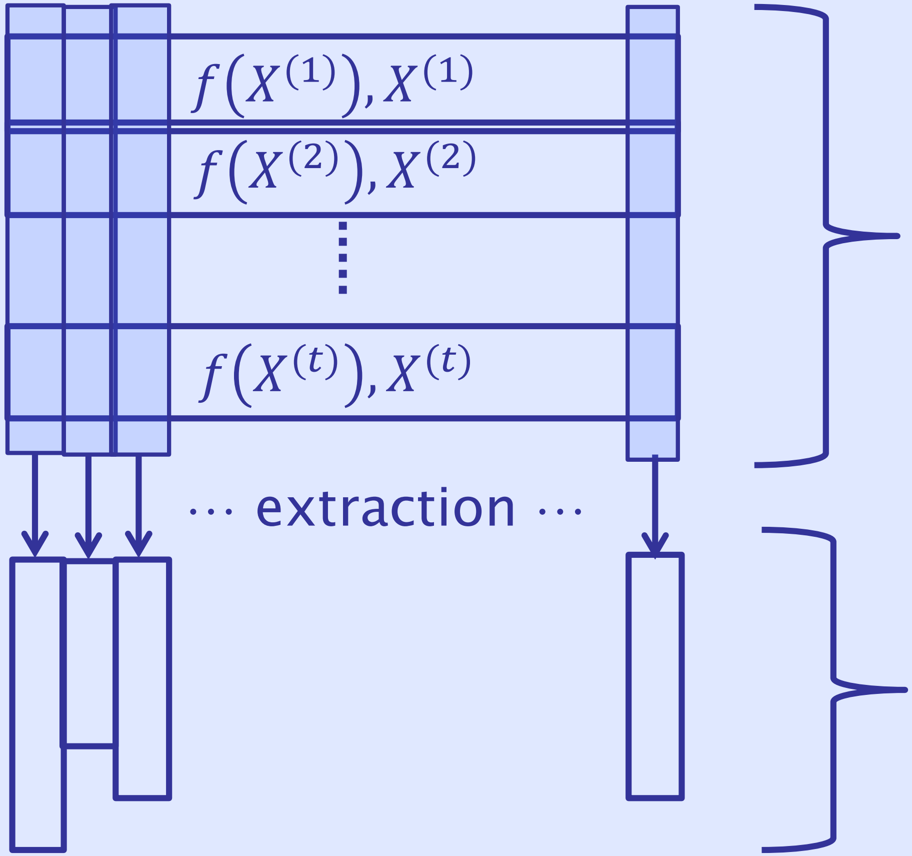
Given $(f(X),X_1,\ldots,X_J)$, X_{J+1} has sampling relative entropy $\omega(\log n)/n$

thm

Given $(f(X),X_1,\ldots,X_J)$, X_{J+1} has pseudoentropy$\geq$entropy+ω(log n)/n

$(f(X),X_1,\ldots,X_n)$ has next-bit pseudoentropy $\geq$ n+ω(log n)

PRGs from OWF: 1st attempt



t=$\tilde{O}$(n^2) repetitions

→ seed length $\tilde{O}$(n^3)

t· (n+log n) – $\tilde{O}(\sqrt{t}\cdot$ n) > t· n pseudorandom bits

Difficulty: how much to extract from each column?

Unknown Entropy Thresholds

- Problem: although we know $H(f(X))+\sum_i H(Y_i|f(X),X_1,\ldots,X_{i-1}) \geq n + \omega(\log n)$, we don't know individual terms.

- Solution: "entropy equalization" [Haitner-Reingold-Vadhan-Wee `09, HRV`10]
  - costs a factor $O(n)$ in # queries to OWF and in seed length.
  - cost in seed length can be eliminated with adaptive queries to OWF [VZ11].

Unknown Entropy Thresholds in Regular OWF

- Problem: Although we know

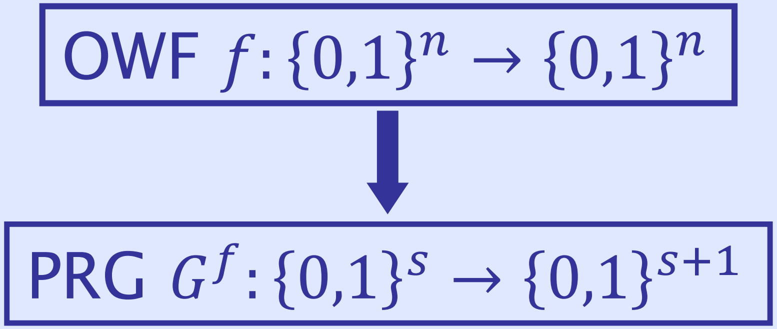
$$H_\infty\big(f(X)\big) + H_\infty(X|f(X)) = n,$$

  we don't know the individual terms.

- Solution: "the randomized iterate" [Goldreich-Krawczyk-Luby `88, Haitner-Harnik-Reingold `07]:
  - Costs factor of $O(n)$ in adaptive queries to OWF
  - Costs a factor of $O(\log n)$ in seed length
  - Cost in #queries is *necessary* for black-box reductions [Holenstein-Sinha `12]

PRGs from OWFs

Thm [Haitner-Reingold-Vadhan `10, Vadhan-Zheng `11]:



OWF $f: \{0,1\}^n \rightarrow \{0,1\}^n$

PRG $G^f: \{0,1\}^s \rightarrow \{0,1\}^{s+1}$

Efficiency measures:

- Seed length: $s = \tilde{O}(n^4)$ [HRV10], $s = \tilde{O}(n^3)$ [VZ11].
- # queries to f: $q = \tilde{O}(n^3)$ [HRV10,VZ11].

Outline

- OWFs & Cryptography
- Notions of pseudoentropy
- OWPs ⇒ PRGs
- OWFs ⇒ PRGs
- Open problems
- Inaccessible Entropy (time permitting)

PRGs from OWFs

- # queries to f: $q = \tilde{O}(n^2) \times O(n)$ [HRV10,VZ11].

Shannon entropy to min-entropy

Unknown entropy thresholds (necessary by [HS12])

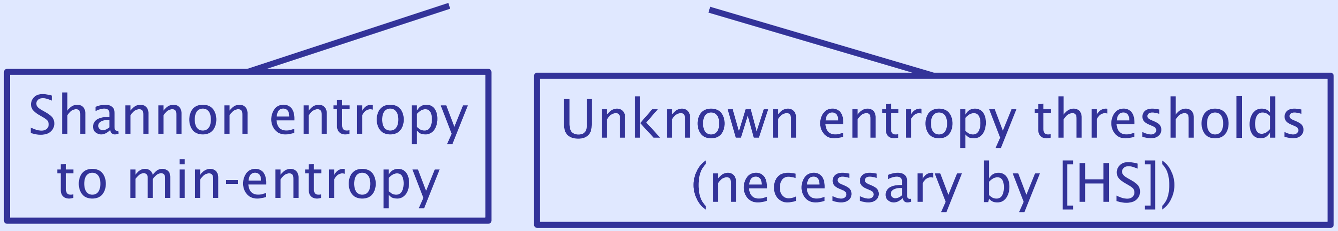
- Seed length: $s = O(q \cdot n)$ [HRV10], $s = \tilde{O}\left(n^2\right) \cdot n$ [VZ11].

Non-adaptive queries

Adaptive queries

PRGs from OWFs

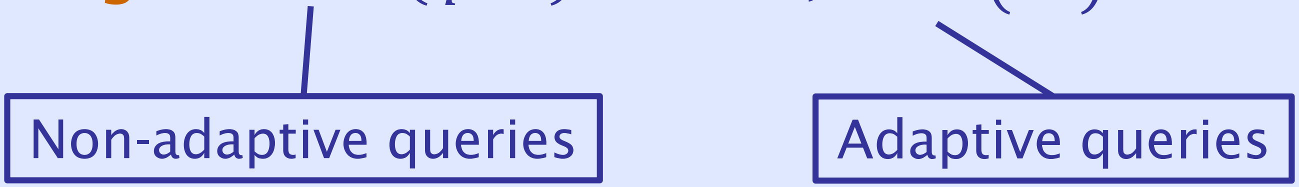
- # queries to f: $q = \tilde{O}(n^2) \times O(n)$ [HRV10,VZ11].



Shannon entropy to min-entropy

Unknown entropy thresholds (necessary by [HS])

- Seed length: $s = O(q \cdot n)$ [HRV10], $s = \tilde{O}\left(n^2\right) \cdot n$ [VZ11].



Non-adaptive queries

Adaptive queries

Open Problems:

- Find a better construction or better black-box lower bounds.
- There could be a construction with $O(n)$ seed length and #queries.

PRGs from OWFs

- # queries to f: $\boldsymbol{q = \widetilde{O}\left(n^2\right)} \times O(n)$ [HRV10,VZ11].

Shannon entropy to min-entropy

Unknown entropy thresholds (necessary by [HS12])

- Seed length: $\boldsymbol{s = O(q \cdot n)}$ [HRV10], $s = \widetilde{O}\left(n^2\right) \cdot n$ [VZ11].

Non-adaptive queries

Adaptive queries

Why do we obtain Shannon entropy?

- Separating pseudoentropy of f(X) and X.
- Breaking X into blocks.

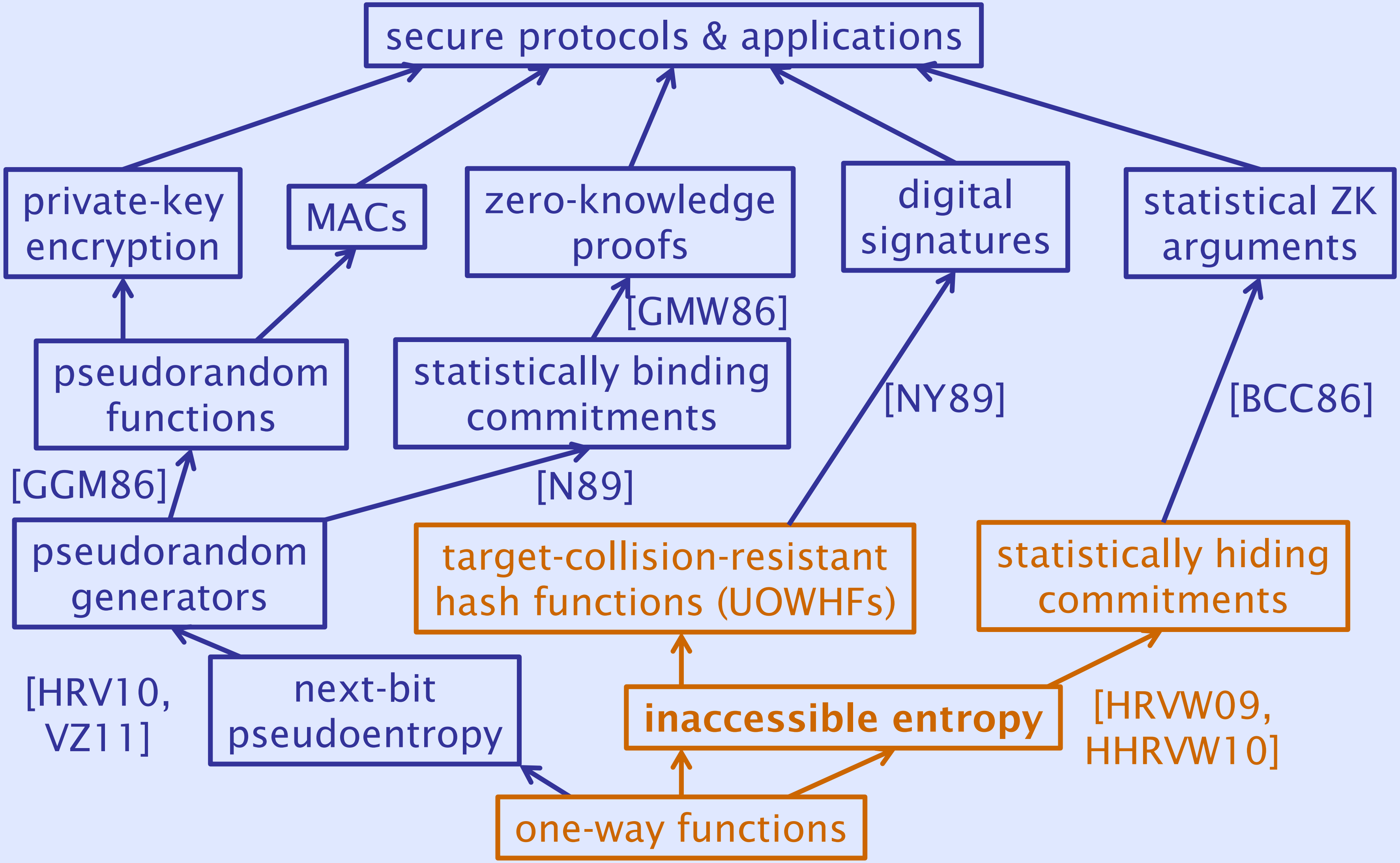
Converting Shannon Entropy to Min-Entropy

Thm [Goldreich-Sahai-Vadhan `99]: There is an oracle algorithm $A^{(\cdot)}\colon \{0,1\}^s \to \{0,1\}^m$ making $q = O(n^2)$ (independent) queries to an input oracle $X : \{0,1\}^n \to \{0,1\}^n$ such that:

1. $H(X(U_n)) \geq \frac{n}{2} + 1 \Rightarrow A^X(U_s)$ negl(n)−close to U_m
2. $H(X(U_n)) \leq \frac{n}{2} \Rightarrow |\mathrm{Support}(A^X(U_s))| \leq \mathrm{negl}(n) \cdot 2^m$.

Q: superlinear lower bounds on q or s?

OWFs & Cryptography



secure protocols & applications

private-key encryption

MACs

zero-knowledge proofs

digital signatures

statistical ZK arguments

[GMW86]

pseudorandom functions

statistically binding commitments

[NY89]

[BCC86]

[GGM86]

[N89]

pseudorandom generators

target-collision-resistant hash functions (UOWHFs)

statistically hiding commitments

[HRV10, VZ11]

next-bit pseudoentropy

inaccessible entropy

[HRVW09, HHRVW10]

one-way functions

Outline

- OWFs & Cryptography
- Notions of pseudoentropy
- OWPs ⇒ PRGs
- OWFs ⇒ PRGs
- Open problems
- Inaccessible Entropy (time permitting)

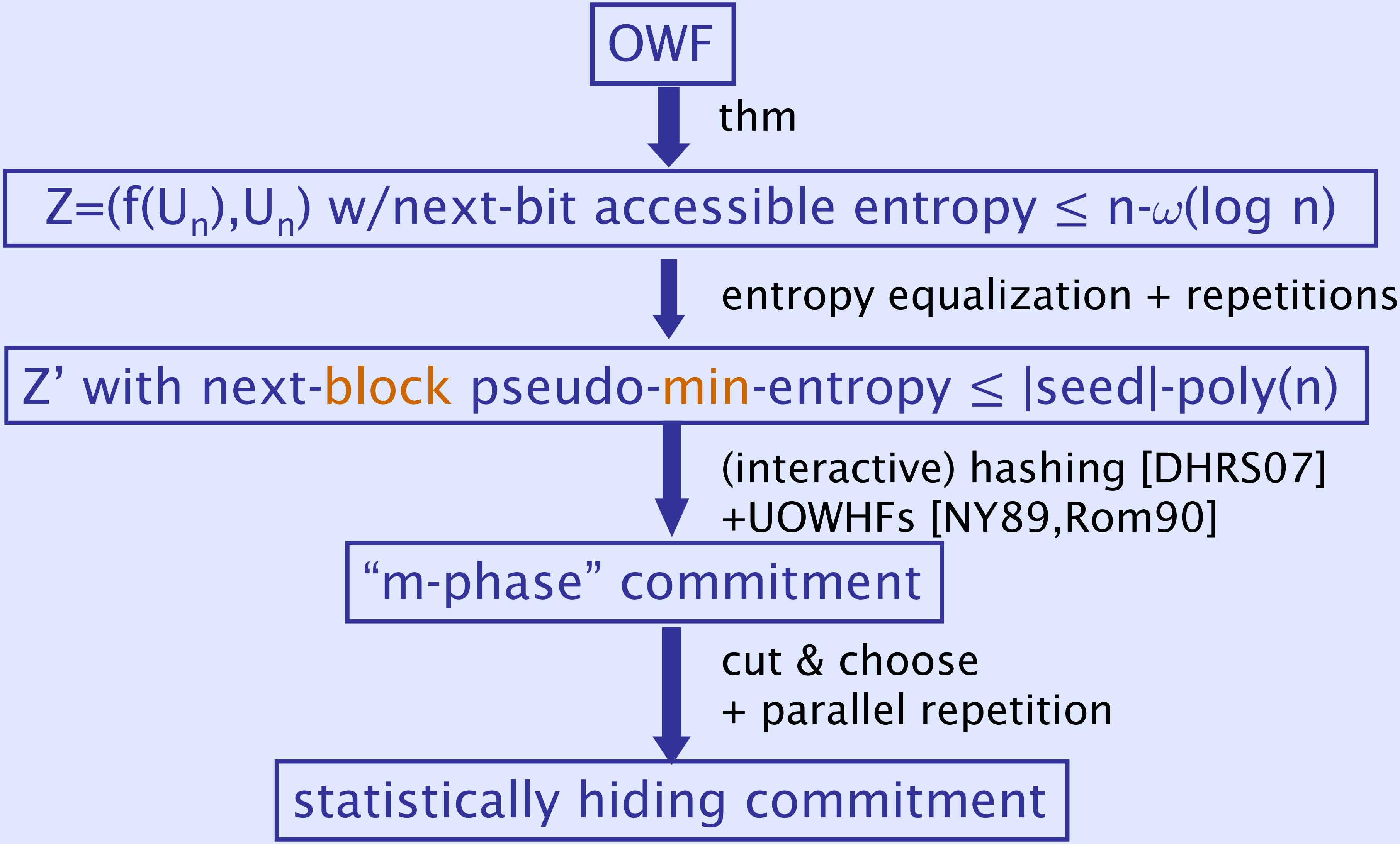
Inaccessible Entropy [HRVW09,HHRVW10]

- Example: if $h : \{0,1\}^n \rightarrow \{0,1\}^{n-k}$ is collision-resistant and $X \leftarrow \{0,1\}^n$, then
  - $H(X|h(X)) \geq k$, but
  - To an efficient algorithm, once it produces $h(X)$, X is determined $\Rightarrow$ "accessible entropy" 0.
  - Accessible entropy $\ll$ Real Entropy!

- Thm [HRVW09]: f a OWF $\Rightarrow$ $(f(X)_1,\ldots,f(X)_n,X)$ has "next-bit accessible entropy" $n-\omega(\log n)$.
  - cf. $(f(X),X_1,\ldots,X_n)$ next-bit pseudoentropy $n+\omega(\log n)$.

OWF $\Rightarrow$ Statistically Hiding Commitments

[Haitner-Reingold-Vadhan-Wee `09]



OWF

thm

$Z=(f(U_n),U_n)$ w/next-bit accessible entropy $\leq n-\omega(\log n)$

entropy equalization + repetitions

Z' with next-block pseudo-min-entropy ≤ |seed|-poly(n)

(interactive) hashing [DHRS07]
+UOWHFs [NY89,Rom90]

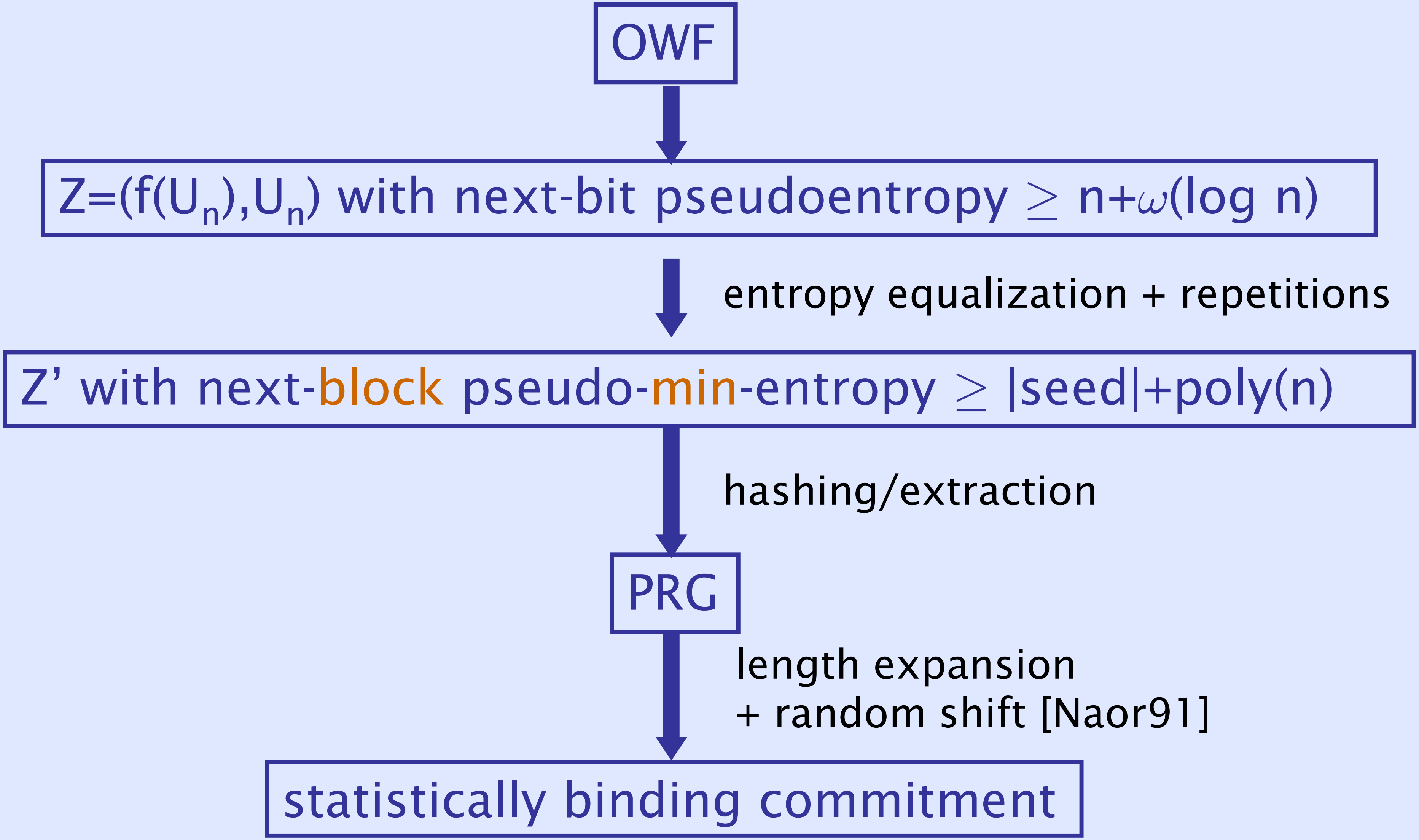
"m-phase" commitment

cut & choose
+ parallel repetition

statistically hiding commitment

OWF ⇒ Pseudorandom Generators

[Haitner-Reingold-Vadhan `10]



OWF

$Z=(f(U_n),U_n)$ with next-bit pseudoentropy $\geq n+\omega(\log n)$

entropy equalization + repetitions

Z’ with next-block pseudo-min-entropy $\geq$ |seed|+poly(n)

hashing/extraction

PRG

length expansion
+ random shift [Naor91]

statistically binding commitment

Conclusion

Complexity-based cryptography is possible because of gaps between real & computational entropy.

“Secrecy”

pseudoentropy > real entropy

“Unforgeability”

accessible entropy < real entropy

Research Directions

- *Formally* unify inaccessible entropy and pseudoentropy.
- From OWF on n bits, can we construct:
  - PRGs with $O(n)$ seed and/or # queries to f?
  - Statistically hiding commitments with $O(n)$ communication and/or # queries to f? (n.b. $\tilde{\Theta}(n)$ optimal for round complexity [Haitner-Harnik-Reingold-Segev `07, HRVW `09])
- More applications of inaccessible entropy in crypto or complexity (or mathematics?)