# A Provable-Security Perspective On Hash Function Design

**Thomas Shrimpton**

Portland State University

February 10, 2010

# A Provable-Security Perspective On Hash Function Design

Blockcipher/Permutation-Based

**Thomas Shrimpton**

Portland State University

February 10, 2010

# Structure of this talk

1. **Basic results** for single-length, one-call, blockcipher-based hash functions

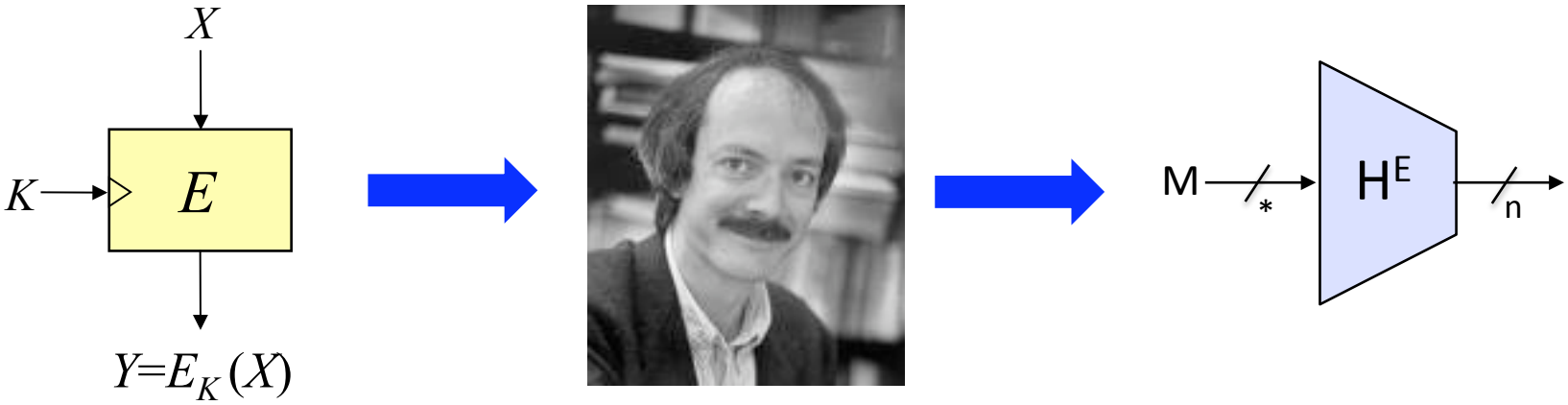
2. Attempts to **maximize speed** lead to questions about fixed-key designs

3. Attempts **increase security** lead to questions about double-length designs

4. Desire for hash functions that **behave like random oracles** leads to new security properties and designs

5. **Skepticism towards idealized models** leads to questions about modeling/assumption

# Building hash function from blockciphers



# Basic results for blockcipher-based schemes

$f(h_{i-1}, m_i) = E_a(b) \oplus c$ $\quad a,b,c \in \{h_{i-1}, m_i, h_{i-1} \oplus m_i, v\}$

[Preneel,Govaerts,Vandewalle'93] analyzed (by attack) 64 blockcipher-based constructions

[Black,Rogaway,S'02] **proved** upper and lowerbounds on collision resistance and preimage resistance

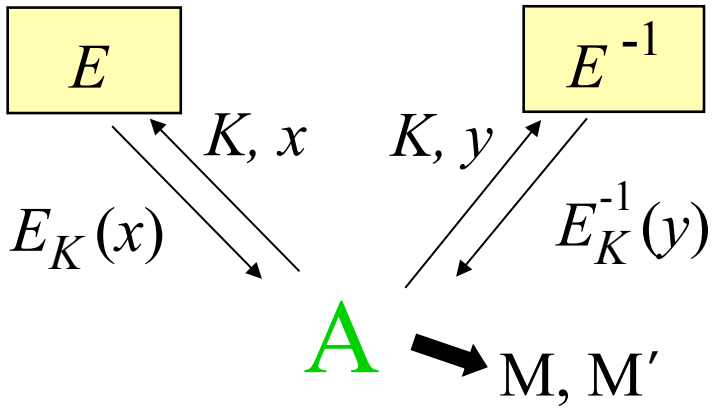
[Stam'09] **generalized the constructions** and reproved bounds

[BRSS'10?] pull it all together

# Collision Resistance in the Ideal Cipher Model

$$\mathbf{Adv}_H^{\mathrm{CR}}(A) = \Pr\left[E \xleftarrow{\$} \mathrm{BC}(k,n);\ (M,M') \xleftarrow{\$} A^{E,E^{-1}} :\ M \neq M' \wedge H_E(M) = H_E(M')\right]$$

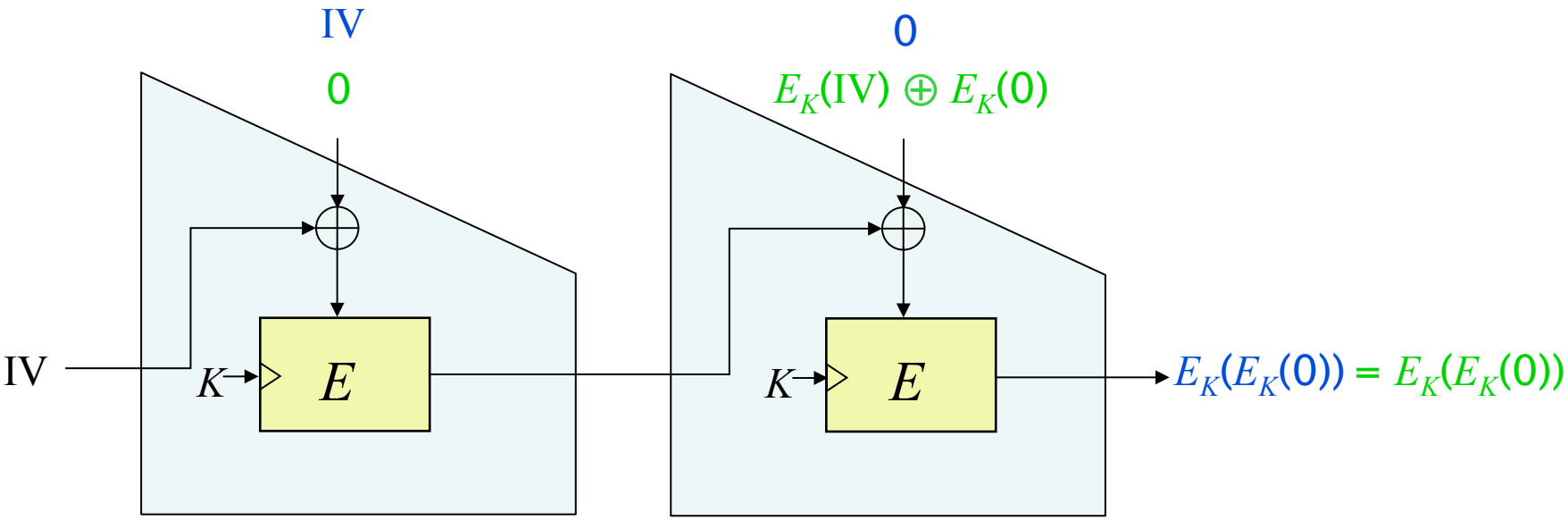
Pick the blockcipher from the set of **all** blockciphers having k-bit keys and n-bit blocksize



# A bad compression function

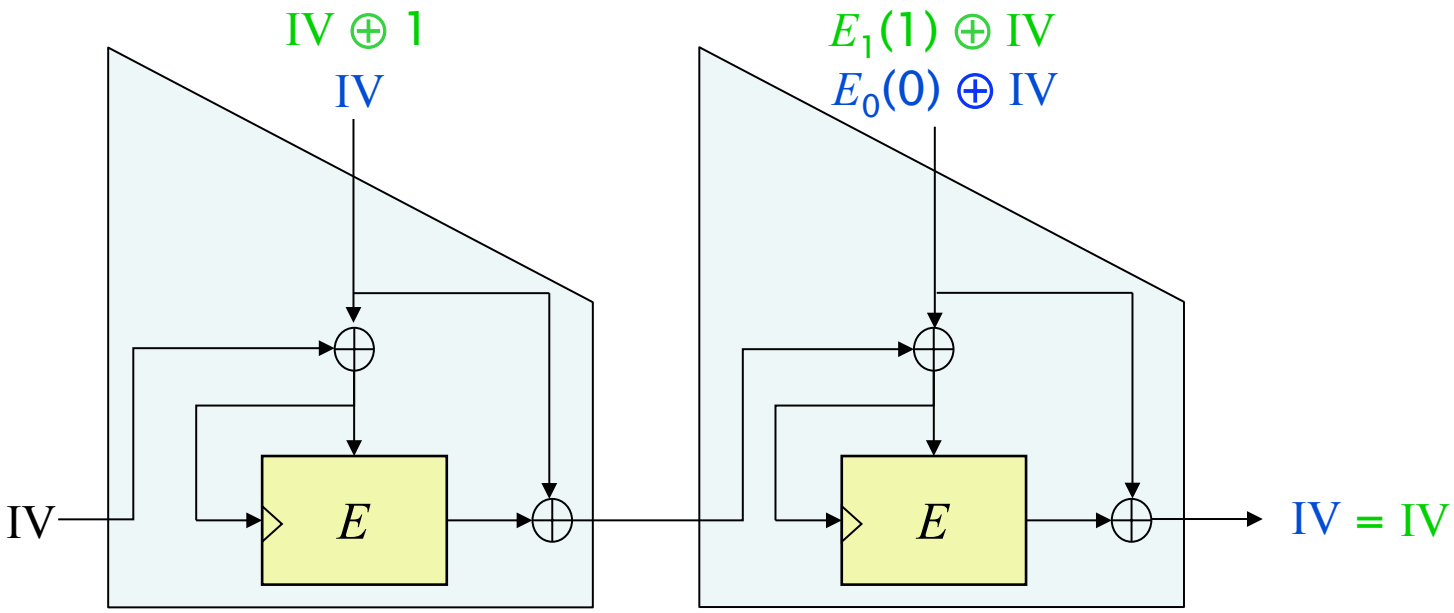
(CBC MAC hash) [Akl'83]

Is this collision-resistant? No.



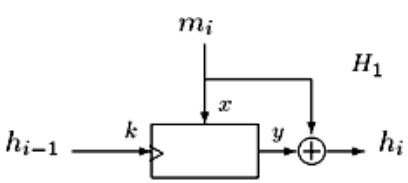
# More complicated, but still bad

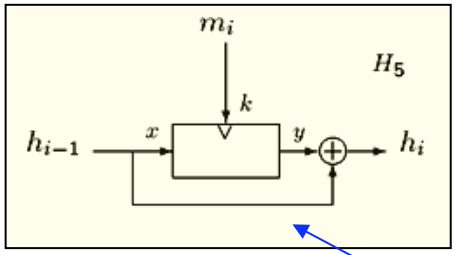
[Preneel,Govaerts,Vandewalle'93]

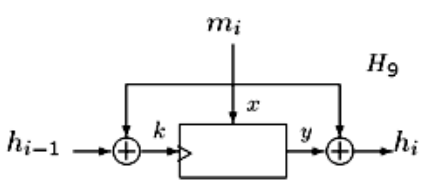


# 12 provably secure compression functions

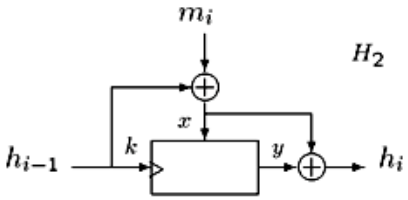
[BRS'02]
[Stam'09]

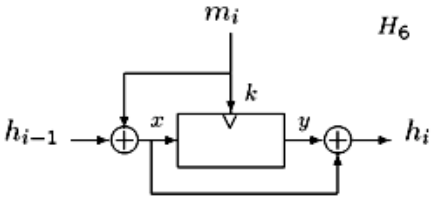


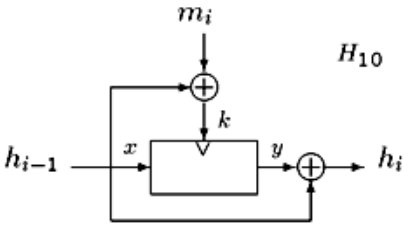



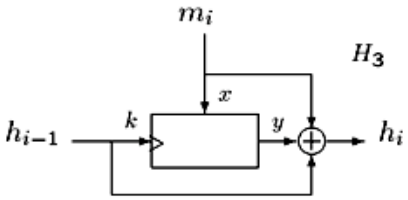



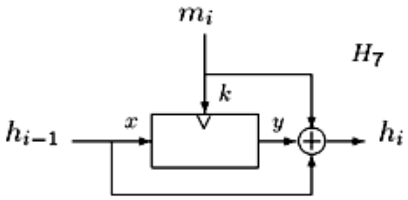

Davies-Meyer

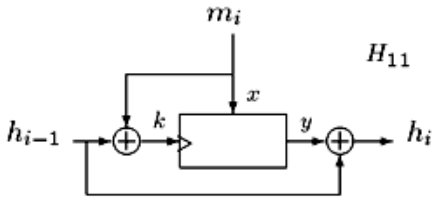


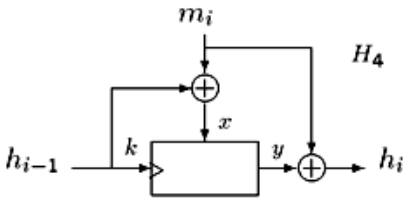



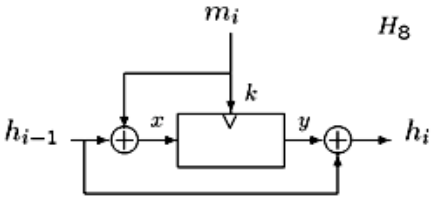



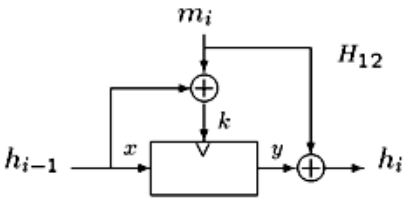



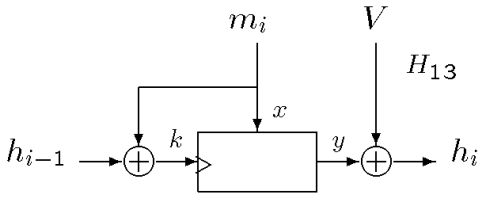



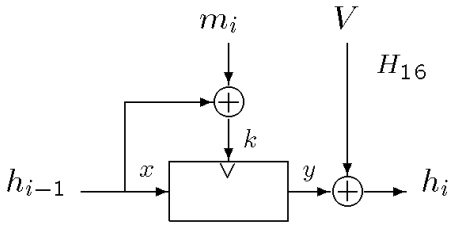



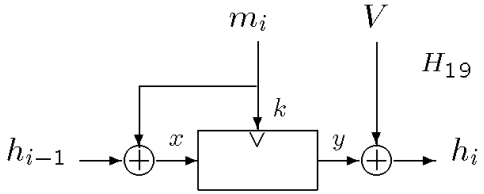



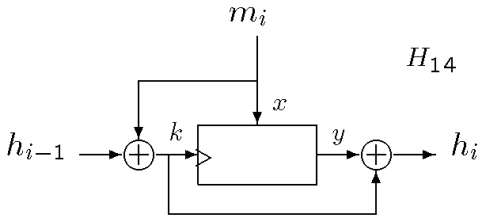
# 8 non-CR Compression functions that MD iterate to CR hashes

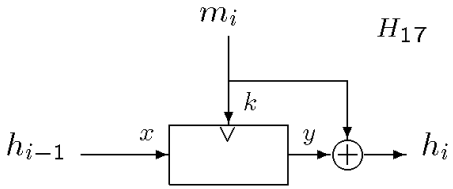
[BRS'02],[Stam'09]

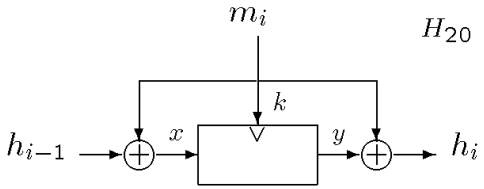


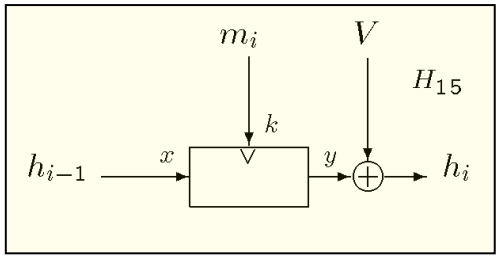



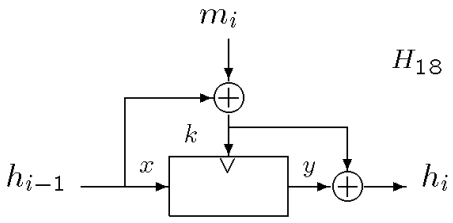




[Rabin'78]




CR: $O(2^{n/2})$ ☺

ePre: $O(2^{n/2})$ ☹

# Structure of this talk

1. **Basic results** for single-length, one-call, blockcipher-based hash functions ✓

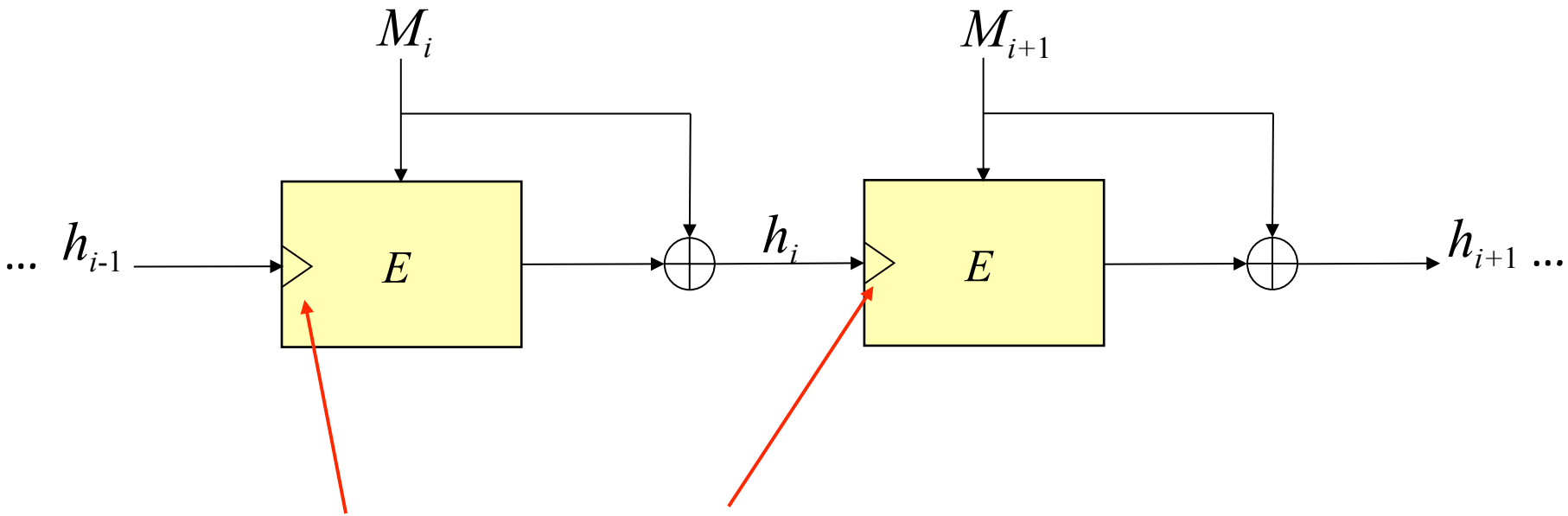
2. Attempts to **maximize speed** lead to questions about fixed-key designs

3. Attempts **increase security** lead to questions about double-length designs

4. Desire for hash functions that **behave like random oracles** leads to new security properties and designs

5. **Skepticism towards idealized models** leads to questions about modeling/assumption

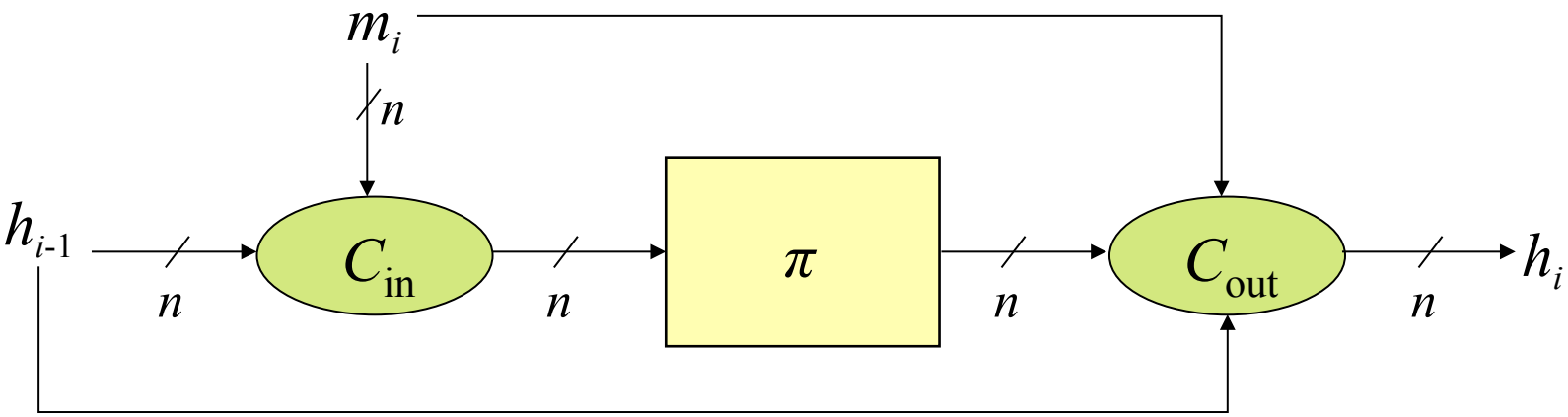
# Do we need to rekey?




Expensive operations;
unnatural way to use a blockcipher

# Permutation-based, generalized compression function

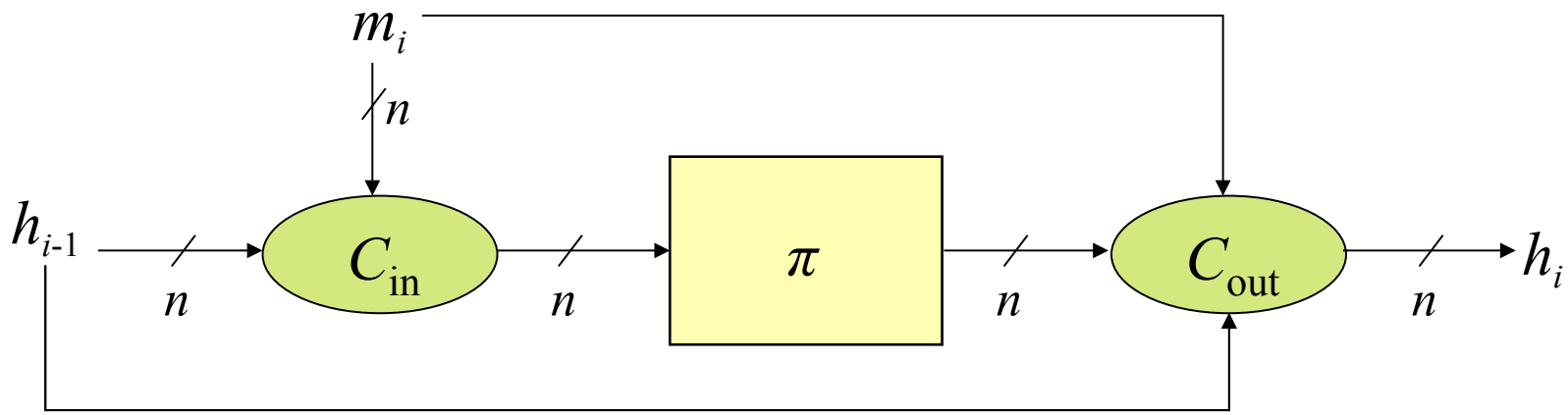
[Black,Cochran,S'05]




Possibly CR, for some $C_{\text{in}}$ and $C_{\text{out}}$?

# CR impossible in the usual model

[BCS'05]




In the ideal cipher model:

compression function — collision after 2 blockcipher calls

If MD iterated — collisions in $\Theta(n + \lg(n))$ calls

# [BCS'05] doesn't say what is (im)possible when…

Computational limits are placed on the adversary

attacks count only queries; time-complexity is still large!

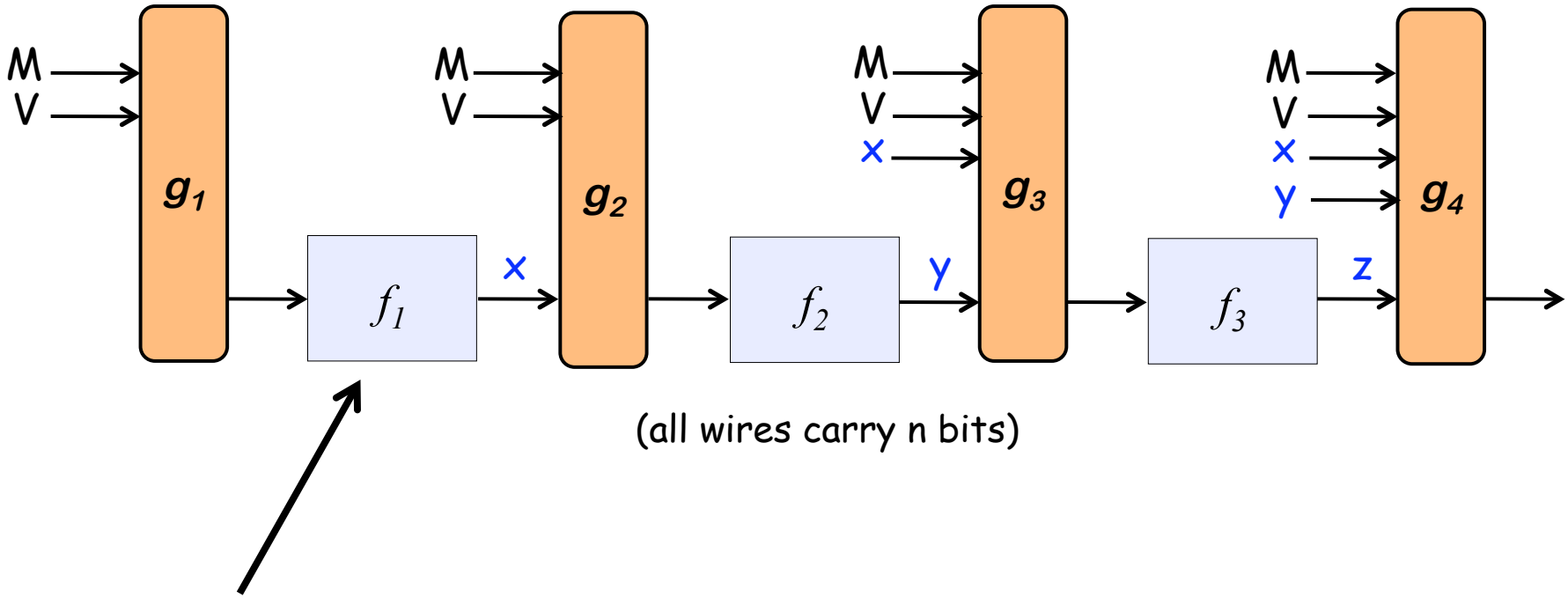
Non-MD constructions are used

what happens if you change the mode?

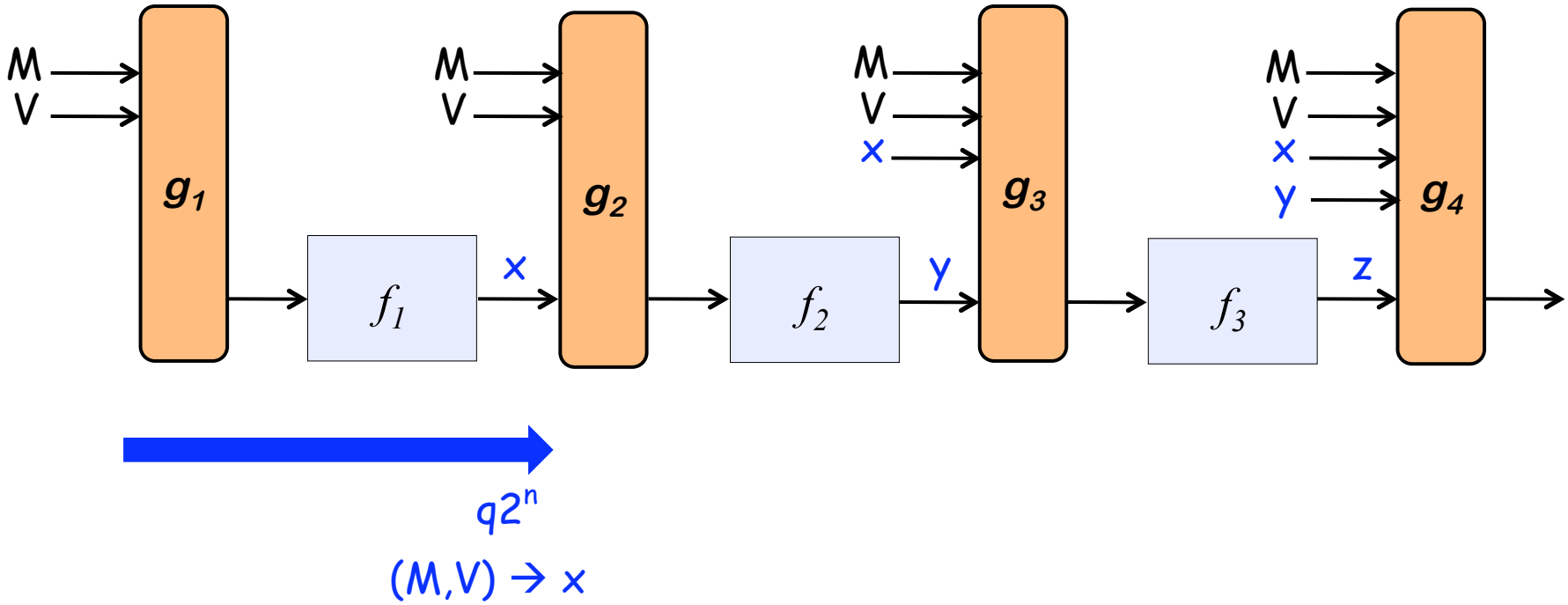
**You use more than one underlying primitive**

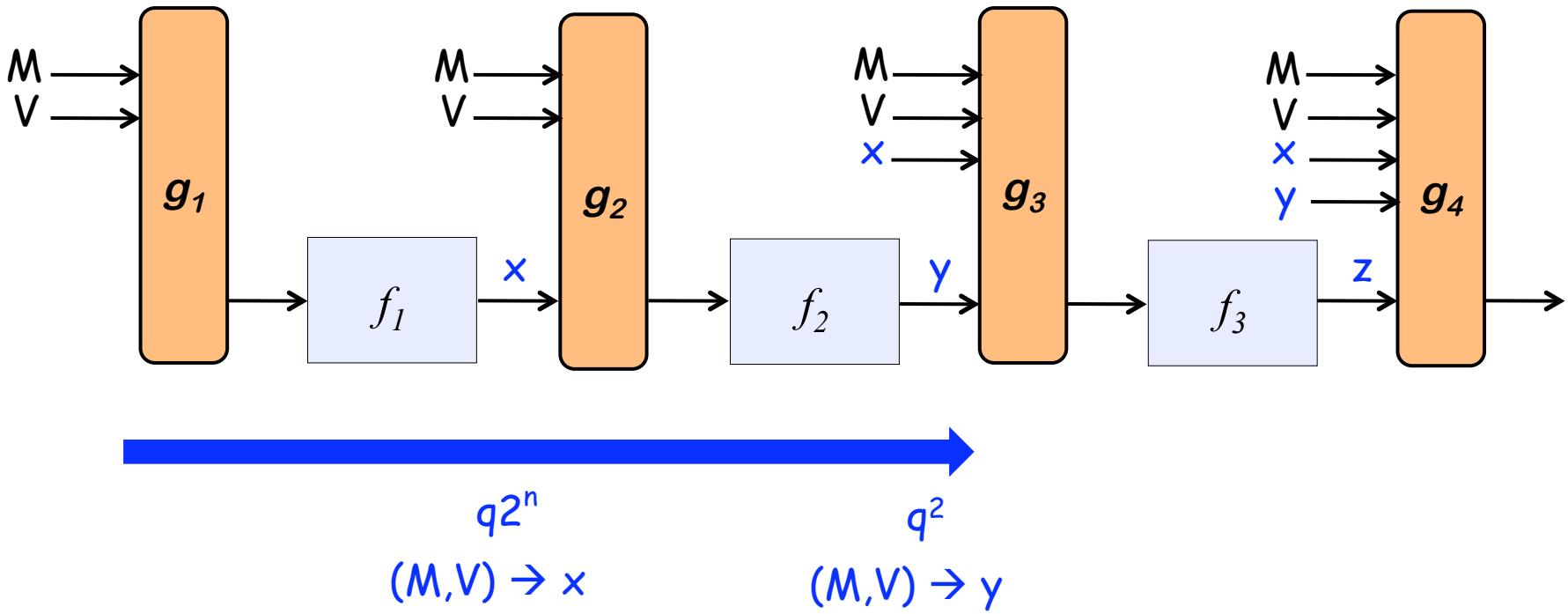
# Yield-based (greedy) attacks

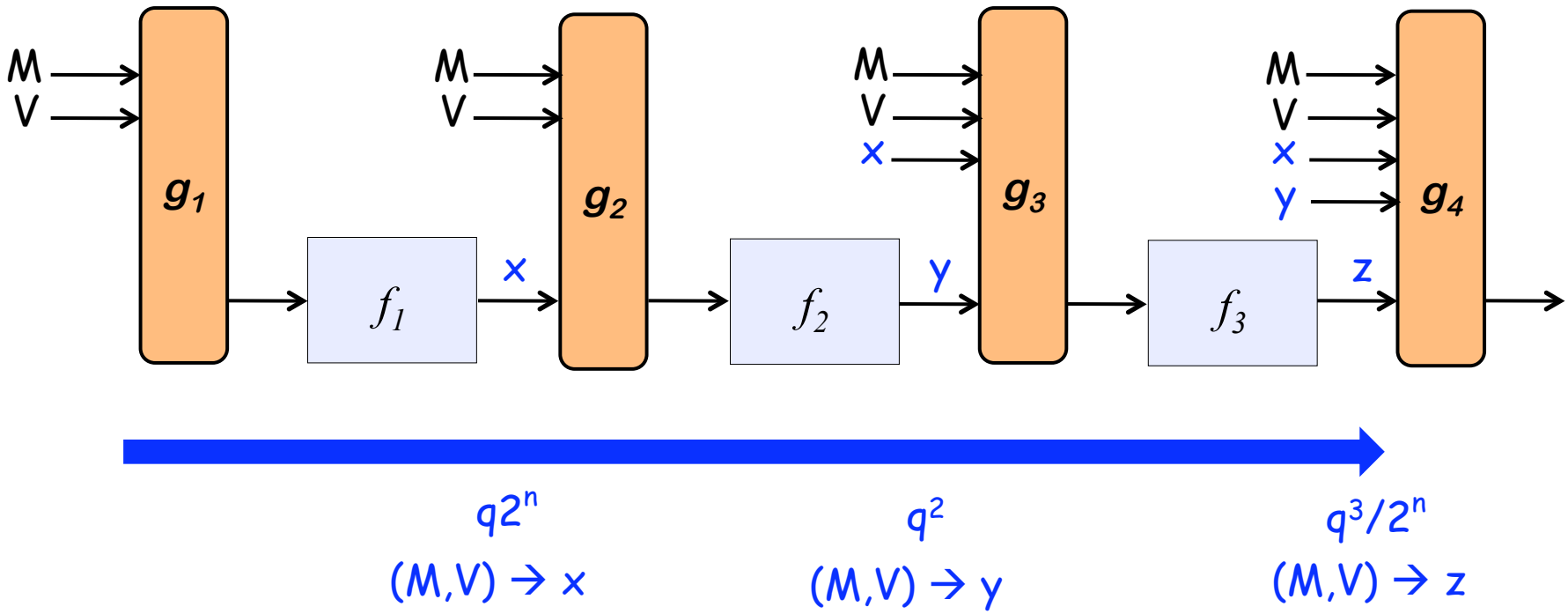
[Rogaway,Steinberger'08],[Stam'08]



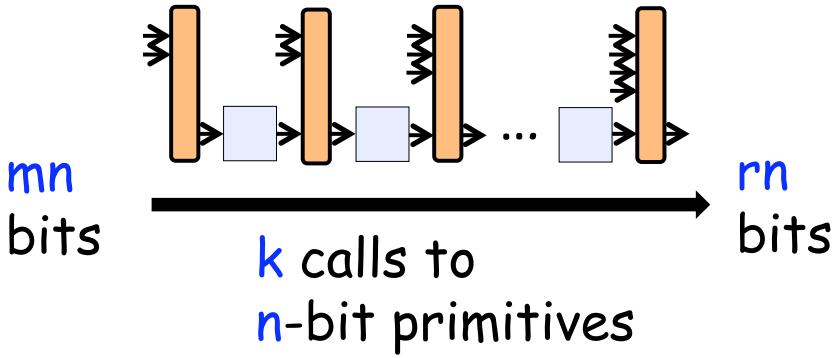

Ask q queries to $f_1$ that maximize the total number of known mappings from (M,V) → x


M
V
$g_1$
$f_1$
x
M
V
$g_2$
$f_2$
M
V
x
y
$g_3$
$f_3$
M
V
x
y
z
$g_4$
$q2^n$
(M,V) → x


M
V
$g_1$
$f_1$
x
M
V
$g_2$
$f_2$
y
M
V
x
$g_3$
$f_3$
z
M
V
x
y
$g_4$
$q2^n$
(M,V) → x
$q^2$
(M,V) → y


M
V
$g_1$
$f_1$
x
M
V
$g_2$
$f_2$
y
M
V
x
$g_3$
$f_3$
z
M
V
x
y
$g_4$
$q2^n$
(M,V) → x
$q^2$
(M,V) → y
$q^3/2^n$
(M,V) → z

# Rogaway-Steinberger result in general




Assuming uniform outputs $q = (2^n)^{1-(m-0.5r)/k}$ queries yield a collision w.h.p.

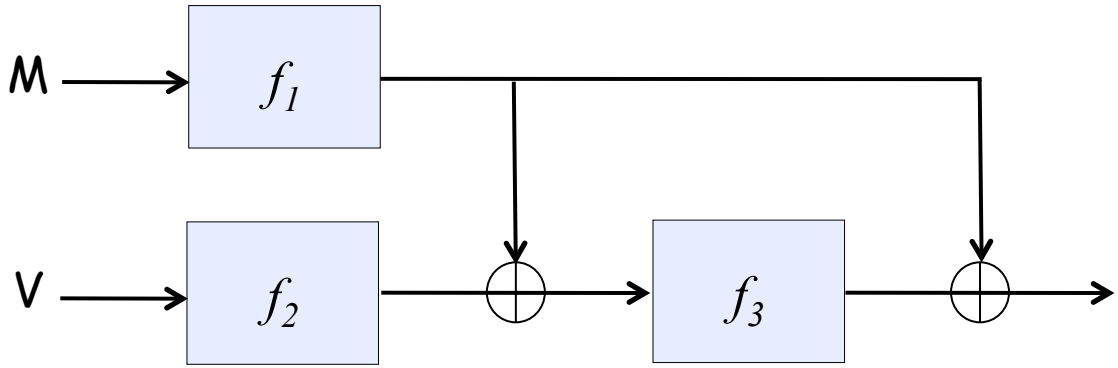
➡ 2n-bit to n-bit compression function (m=2,r=1)

k=2 → $q=2^{n/4}$

k=3 → $q=2^{n/2}$

# (Nearly) optimal compression functions from three non-compressing primitives

[S,Stam'08]
(see also [Rogaway,Steinberger'08])



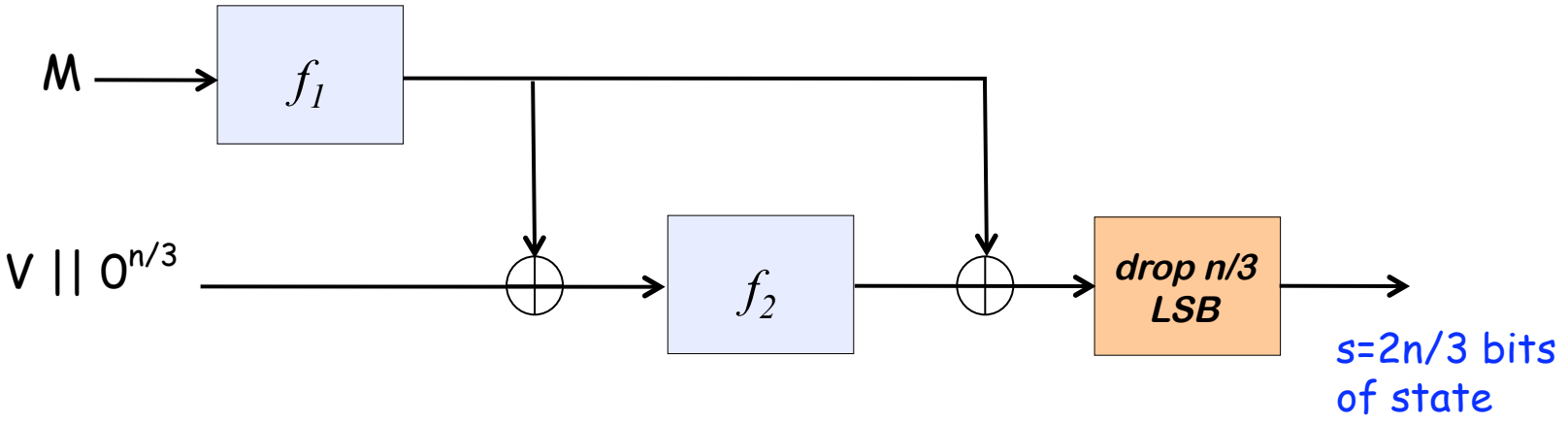

**If** $f_1, f_2, f_3$: $\{0,1\}^n \rightarrow \{0,1\}^n$ are random functions,

(or Davies-Meyer over random permutations)

**Then** CR = $O(2^{n/2 - \log(n)})$

# Getting the most out of two calls

[Stam'08]

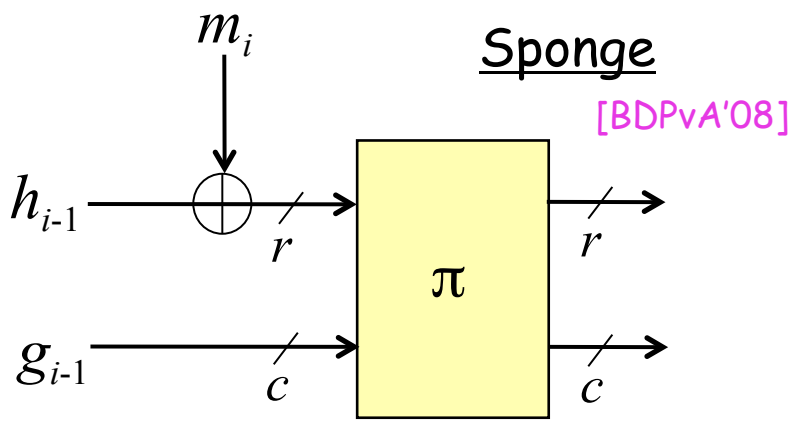


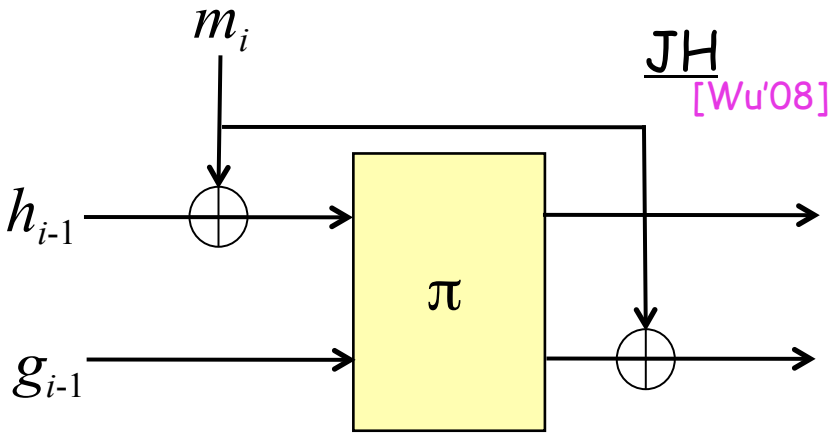

$$CR = O(2^{n/3-\log(n)}) = O(2^{s/2-\log(n)})$$

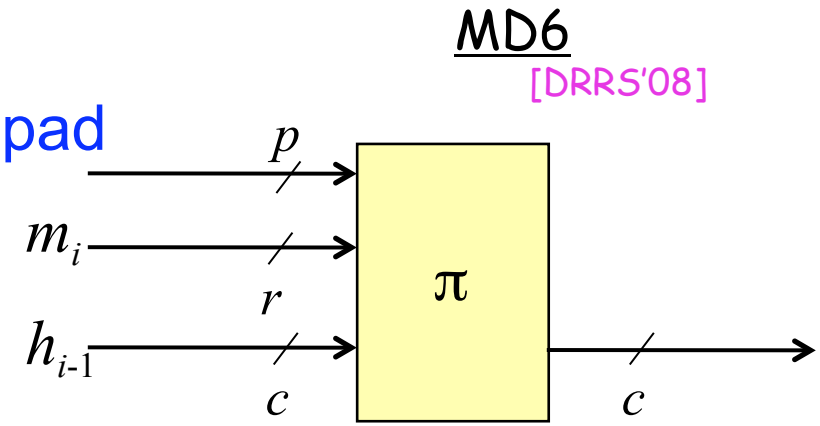
How does this get around the Rogaway-Steinberger $2^{n/4}$ bound?!

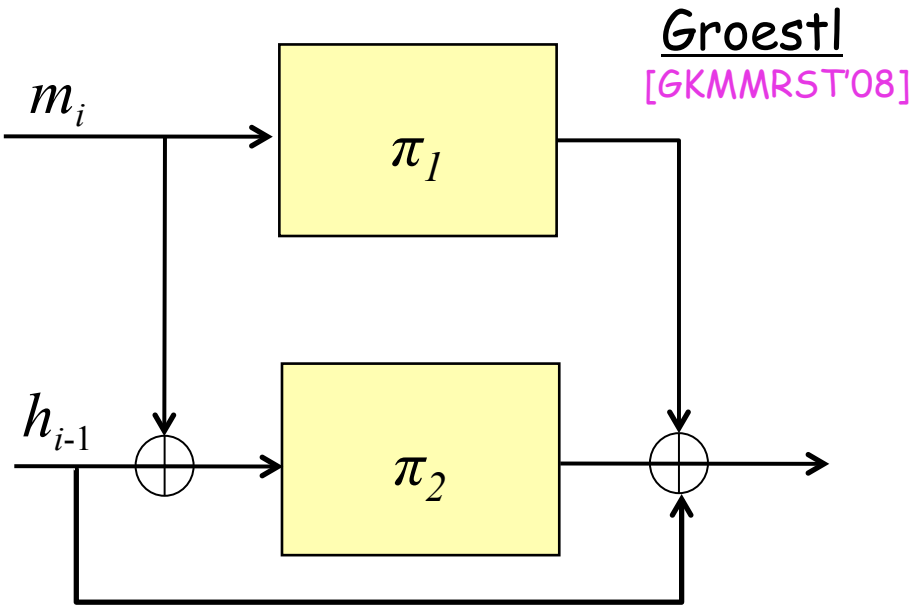
Violates the "uniformity assumption"!

# Other permutation-based examples



# Structure of this talk

1. Basic results for single-length, one-call, blockcipher-based hash functions ✓

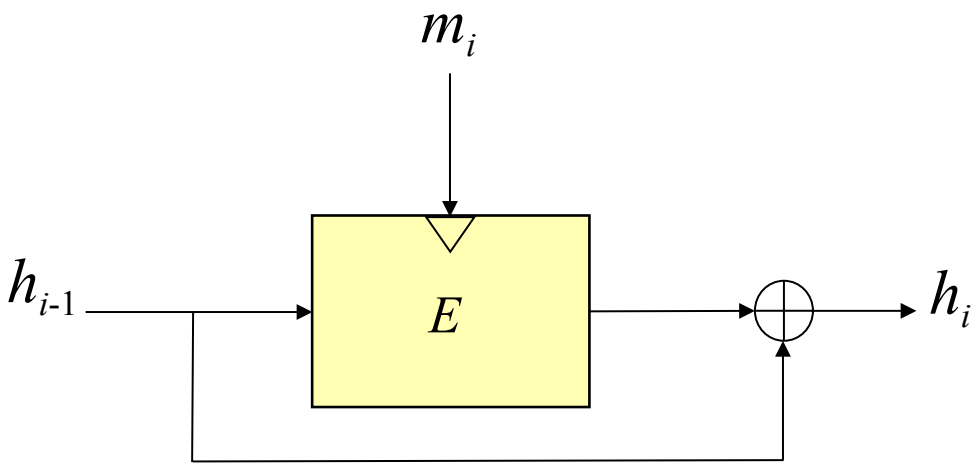
2. Attempts to **maximize speed** lead to questions about fixed-key designs ✓

3. Attempts **increase security** lead to questions about double-length designs

4. Desire for hash functions that **behave like random oracles** leads to new security properties and designs

5. **Skepticism towards idealized models** leads to questions about modeling/assumption

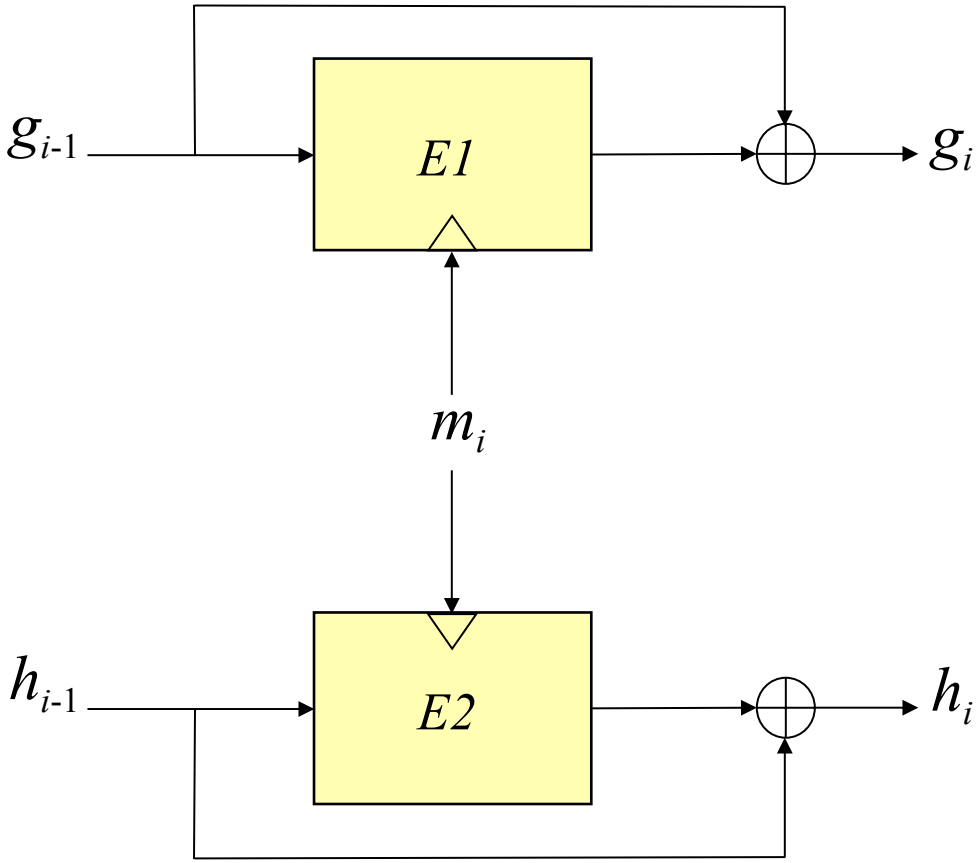
# "With ns like these, who needs enemies"?



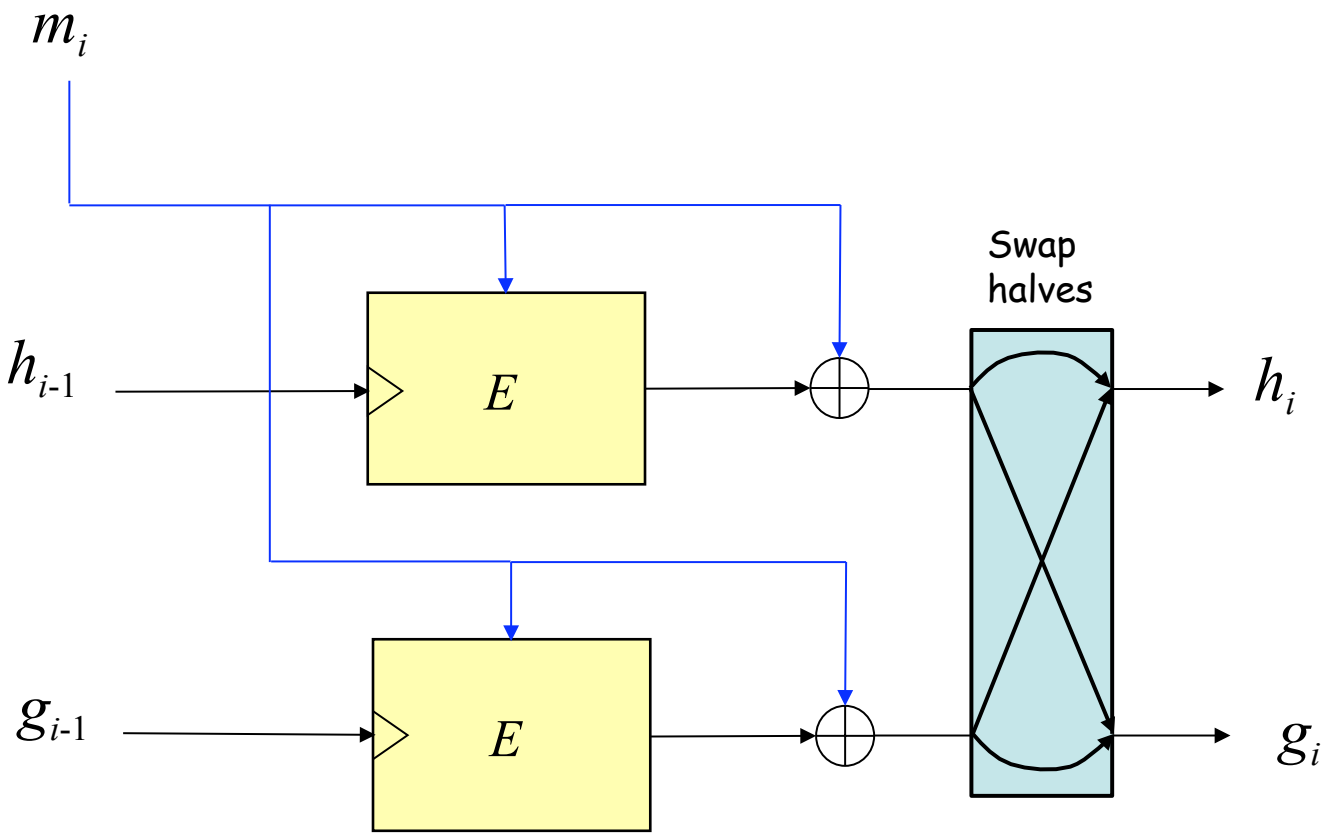

Davies-Meyer is provably CR up to $2^{n/2}$ queries

DES: $2^{32}$
AES: $2^{64}$

# "Parallel DM": CR to $2^n$ ? No... $2^{n/2}$

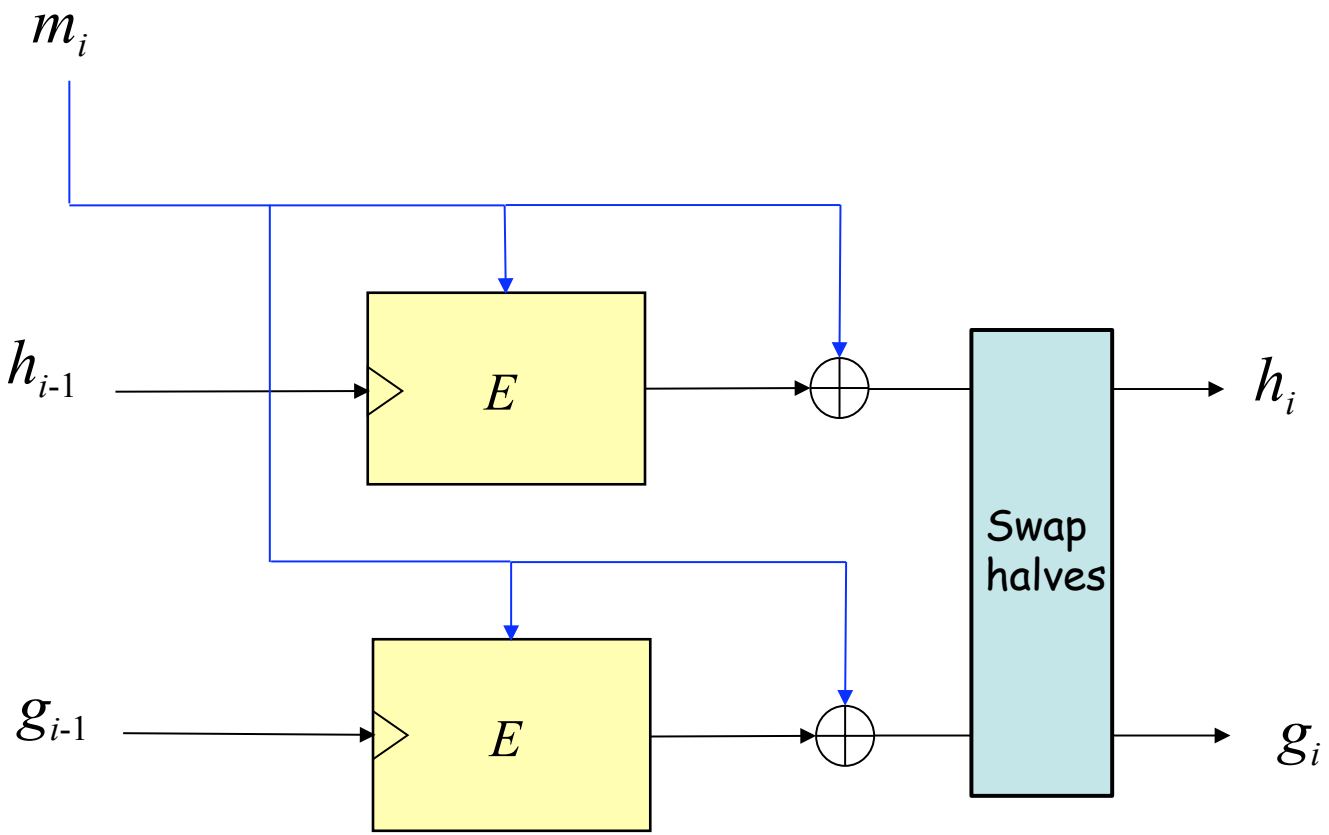


# The MDC-2 compression function (~ "parallel MMO")




Trivial CR bound in the iteration is $2^{n/2}$

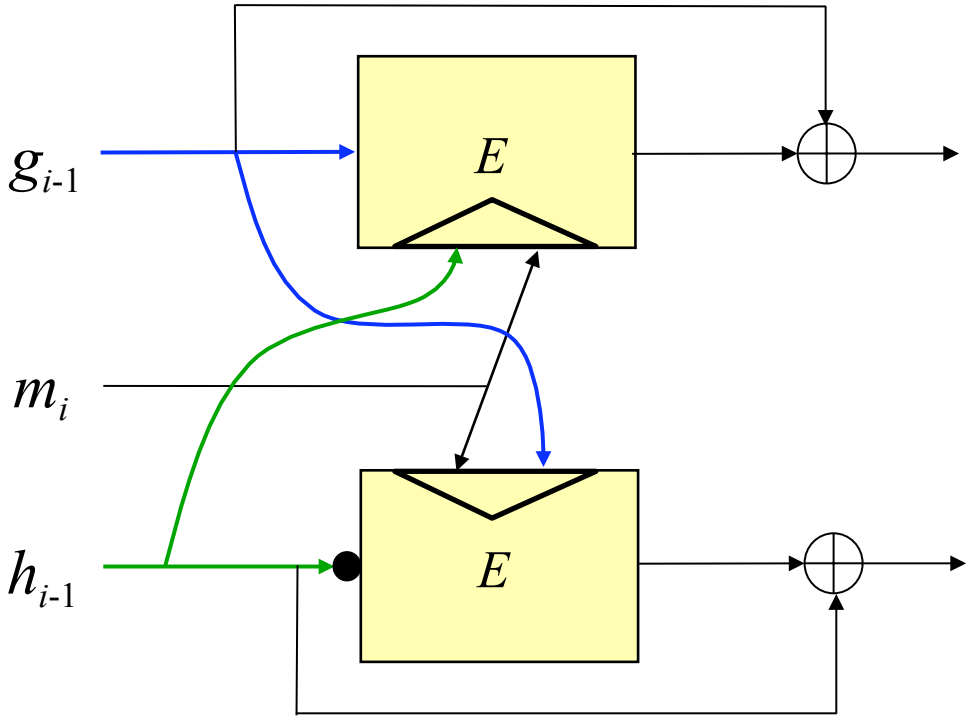
# The MDC-2 compression function




Steinberger recently showed that the iteration of MDC-2 has collision resistance of $2^{3n/5}$ in the ideal cipher model (concretely, $2^{74.9}$ for 256 bits of output)

[Steinberger'07]

# $2^n$ CR is possible… with a 2n-bit key



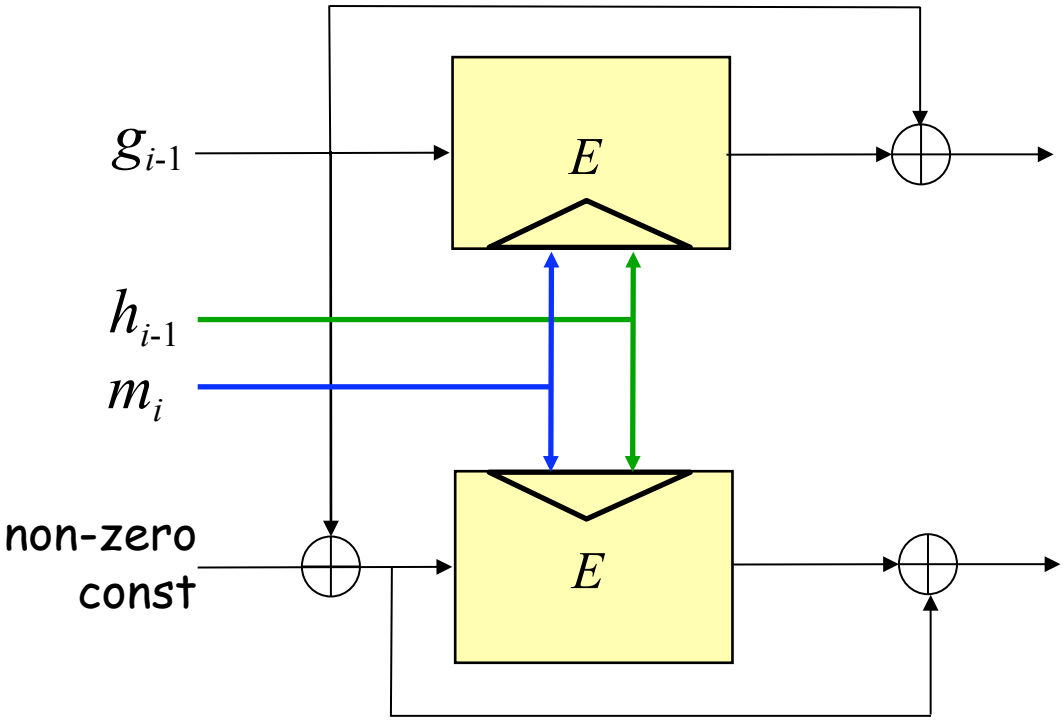

Abreast Davies-Meyer recently proved secure to ~ $2^n$

[Fleishman,Gorski,Lucks'09]
[Lee,Kwon'09]

Proof must deal with cycles of query "reuse"; for Abreast DM the cycle length is 6.

# A nice DBL construction with one key scheduling

[Hirose'06]




A recent paper by Özen and Stam gives a framework for proving CR/ePre security of class of DBL constructions [Özen,Stam'09]

# Structure of this talk

1. **Basic results** for single-length, one-call, blockcipher-based hash functions ✓

2. Attempts to **maximize speed** lead to questions about fixed-key designs ✓

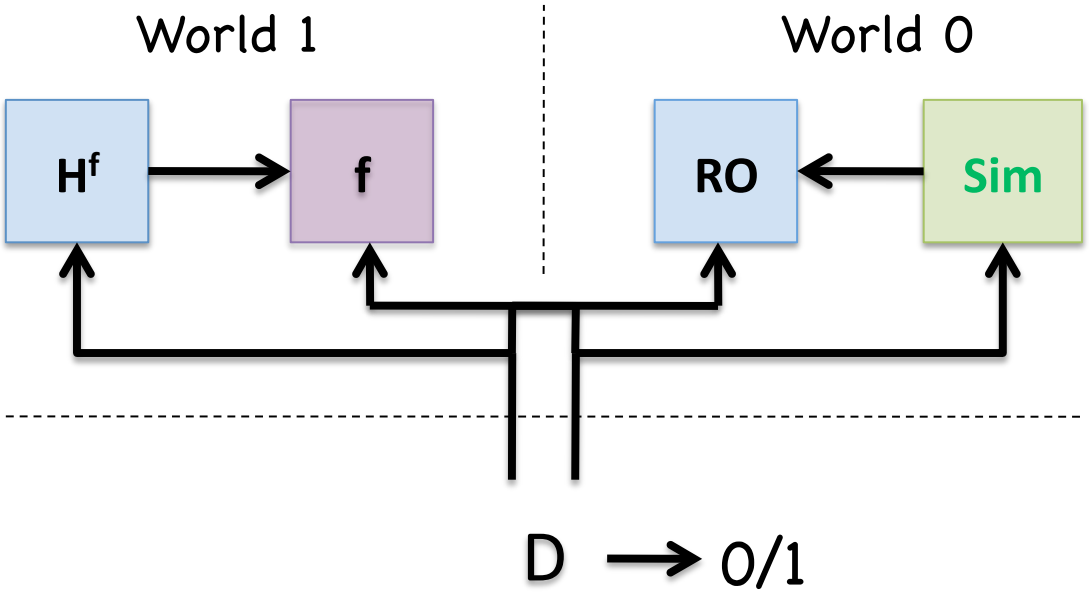
3. Attempts **increase security** lead to questions about double-length designs ✓

4. Desire for hash functions that **behave like random oracles** leads to new security properties and designs

5. **Skepticism towards idealized models** leads to questions about modeling/assumption

# Indifferentiability from a RO

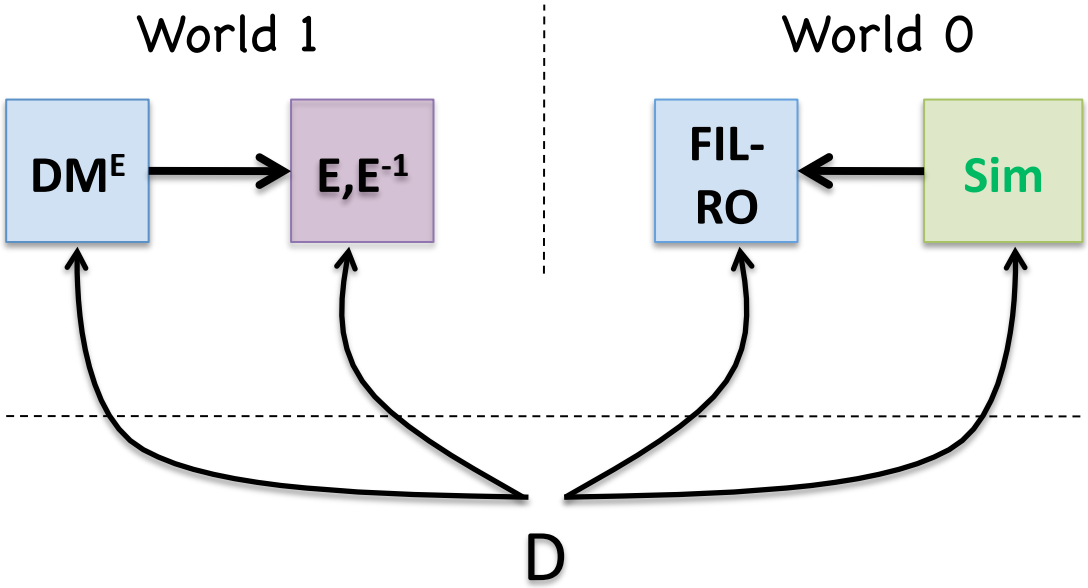
[Maurer,Renner,Holenstein'04], [CDMP'05]



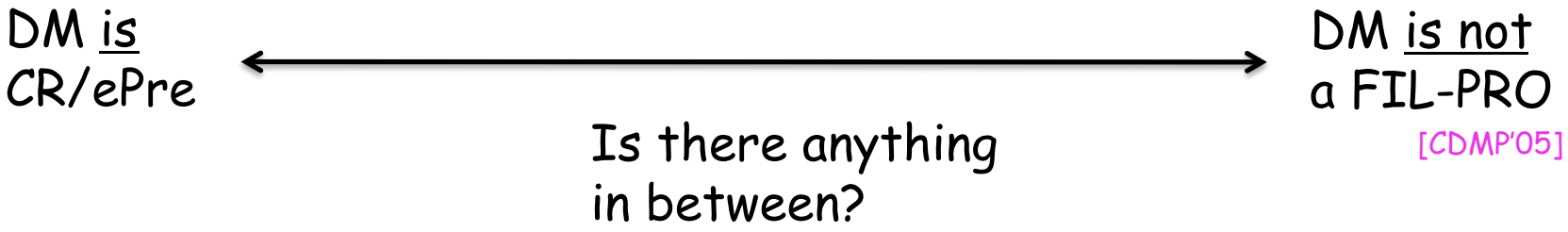

Sim simulates f, trying to make World 0 indistinguishable from World 1.

If ∃ Sim ∀ D the distinguishing advantage is "small" we call $H^f$ a pseudo random-oracle (PRO)

# Indifferentiability from a FIL-RO



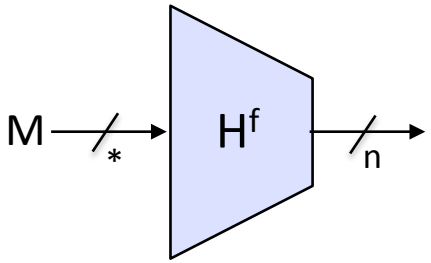

When **E** is an ideal cipher, is **DM**$^{\mathbf{E}}$ an FIL-PRO?

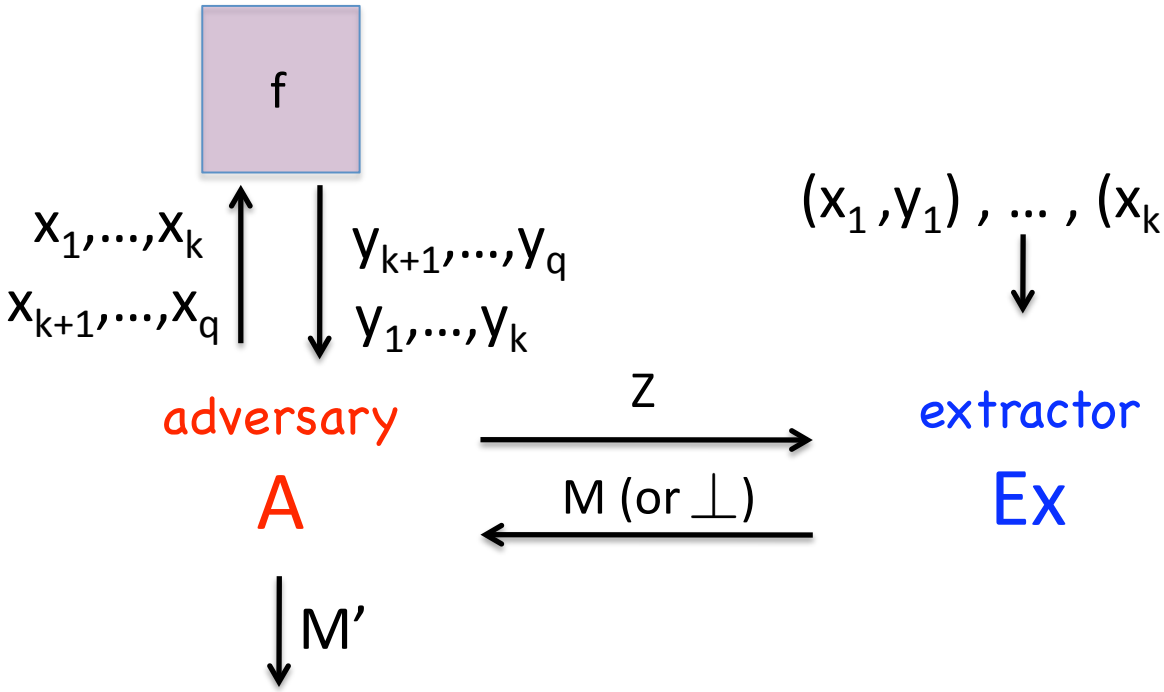



Yes: Preimage-awareness

# Preimage Awareness (PrA)

[Dodis,Ristenpart,S'09]



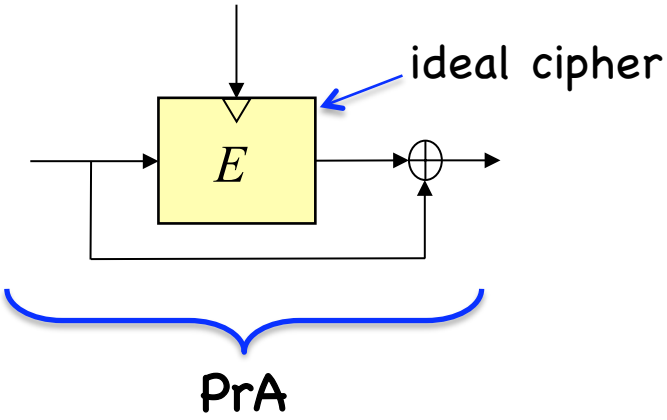

**A** wins if:

1) $H^f(M')=Z$ and
2) $M' \neq$ value previously returned by **Ex** on Z

if ∃ **Ex** such that ∀ **A** the winning probability is "small", then we say that H is **preimage-aware**.

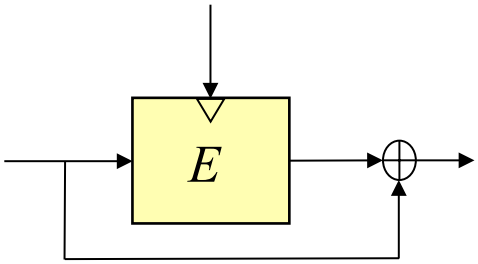
# Davies-Meyers is PrA

(also other optimally CR blockcipher-based compression functions!)



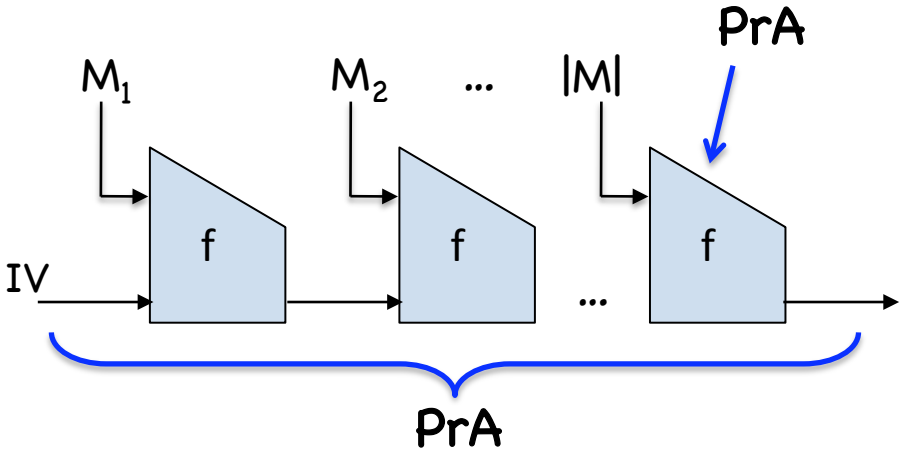

[DRS'09]

Davies-Meyers is **PrA**




+

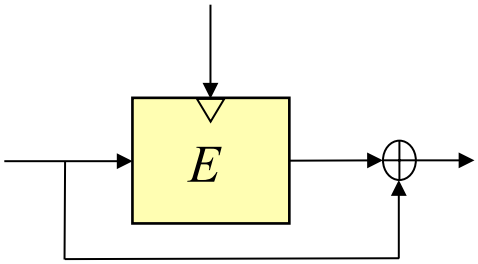
MD is **PrA**-preserving




Note: MD is not PRO-preserving
(length extension…)

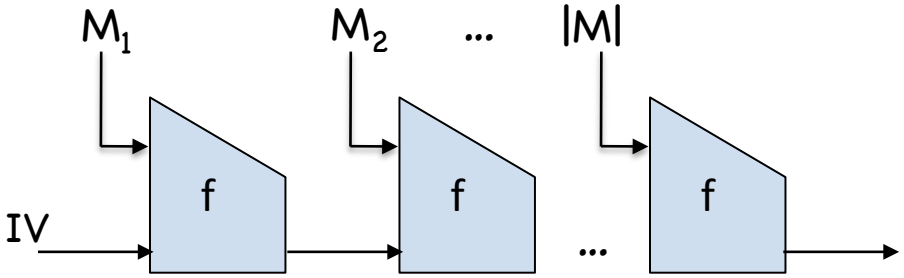
[DRS'09]

Davies-Meyers is **PrA**



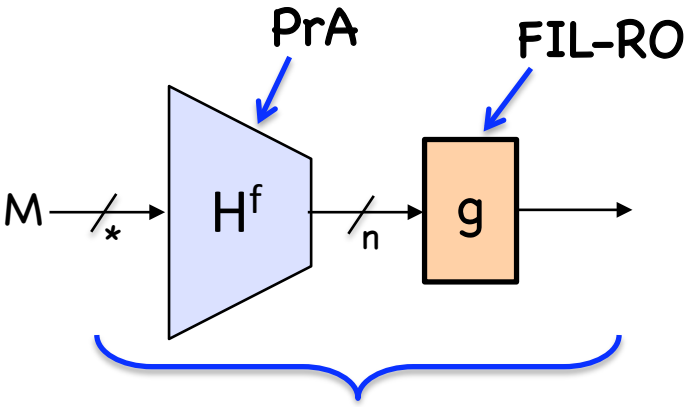

+

MD is **PrA**-preserving




+

"VIL-PrA + FIL-RO = VIL-PRO"




Indifferentiable from VIL-RO

[DRS'09]

# Structure of this talk

1. **Basic results** for single-length, one-call, blockcipher-based hash functions ✓

2. Attempts to **maximize speed** lead to questions about fixed-key designs ✓

3. Attempts **increase security** lead to questions about double-length designs ✓

4. Desire for hash functions that **behave like random oracles** leads to new security properties and designs ✓

5. **Skepticism towards idealized models** leads to questions about modeling/assumption

# Why the Ideal Cipher Model?
## (Why not PRP?)

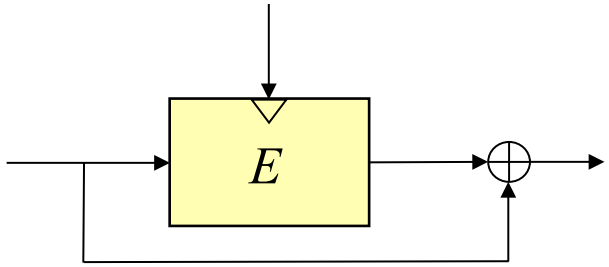
$$\mathbf{Adv}^{\mathrm{prp}}_{E}(A) \;=\; \Pr\left[K \xleftarrow{\$} \mathcal{K} :\; A^{E_K(\cdot)} \Rightarrow 1\right] - \Pr\left[\pi \xleftarrow{\$} \mathrm{Perm}(n) :\; A^{\pi(\cdot)} \Rightarrow 1\right]$$

A good PRP is computationally indistinguishable from a truly random permutation if the key is secret

Also, [Hirose'02] and Hopwood and Wagner (sci.crypt'02) exhibit PRPs that break the good PGVs

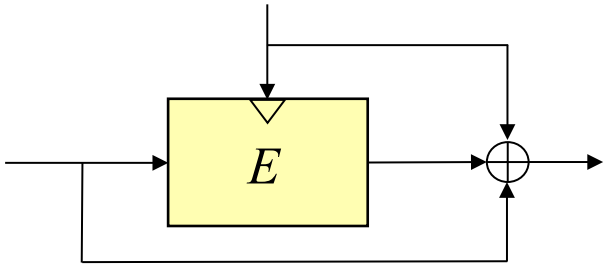
# "Human Ignorance" could save us!

[Rogaway'06]




6 out of 12 ICM CR functions
have this basic "Davies-Meyer shape"...

$$\mathbf{Adv}_E^{DM}(A) = \Pr\left[(K,X),(K',X') \xleftarrow{\$} A : \; E_K(X) \oplus X = E_{K'}(X') \oplus X'\right]$$




...the other 6 have the
"Miyaguchi-Preneel shape"

$$\mathbf{Adv}_E^{MP}(A) = \Pr\left[(K,X),(K',X') \xleftarrow{\$} A : \; E_K(X) \oplus X \oplus K = E_{K'}(X') \oplus X' \oplus K'\right]$$

# Revisiting the ICM

Algorithm for building an ideal n-bit cipher E:

for all $K \in \{0,1\}^k$

Pick permutation $\pi$ uniformly over $\{0,1\}^n$

Assign $E_K = \pi$

end

Fix a distribution $D_\pi$ over n-bit permutations

for all $K \in \{0,1\}^k$

    Pick permutation $\pi$ over $\{0,1\}^n$ according to $D_\pi$

    Assign $E_K = \pi$

end

## What are interesting distributions $D_\pi$? Up to you!

$D_\pi$ : a distribution with statistical distance $\leq \varepsilon$ from uniform

$D_\pi$ : a distribution with min-entropy $\geq \delta$

Can you build secure comp. functions? Iterations?

$D_\pi$ : pick uniformly from permutations such that $f(x) = \pi(x) \oplus x$ is itself a permutation.

"Davies-Meyer cipher"

$D_\pi$: pick uniformly from permutations such that $f(x) = \pi(x) \oplus x$ has a bias away from some particular value V

Possibly useful for Domain separation a la NMAC?

# Generalizing one step further…

Fix a sequence of distributions $\{D_\pi^K\}_{K \in \{0,1\}^k}$

for all $K \in \{0,1\}^k$

Pick permutation $\pi$ according to $D_\pi^K$

Assign $E_K = \pi$

end

Recent Shabal analysis is (kind of) like this…

Fix relation R(K,X,Y), and build E so that for all (K,X) we have $R(K,X,E_K(X))=1$

[BCCCFGIMNPPRTV'09]

# We've learned a lot, but still things to do!

(Im)possibility results for computationally bounded adversaries

Closing gaps between query- and time-complexity of attacks

Is there anything interesting between PrA and CR?

Proofs in weaker idealized models

Proofs using strong (?) standard model assumptions

감사합니다

Thank you!