

Memory: The Dominant Bottleneck in Genomic Workloads



Meysam Taassori, Anirban Nag, Keeton Hodgson, Ali Shafiee† , Rajeev Balasubramonian
University of Utah, Samsung Semiconductor Inc†



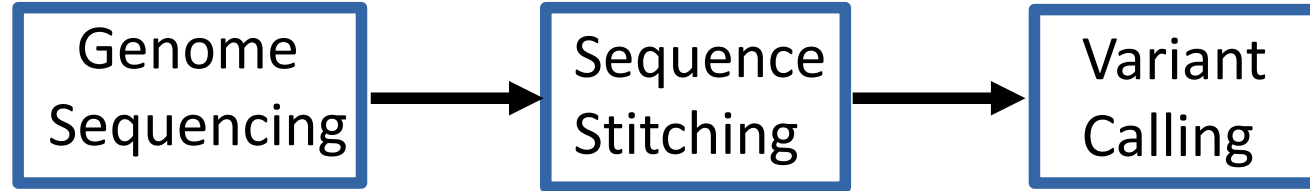
SAMSUNG

February 24, 2018

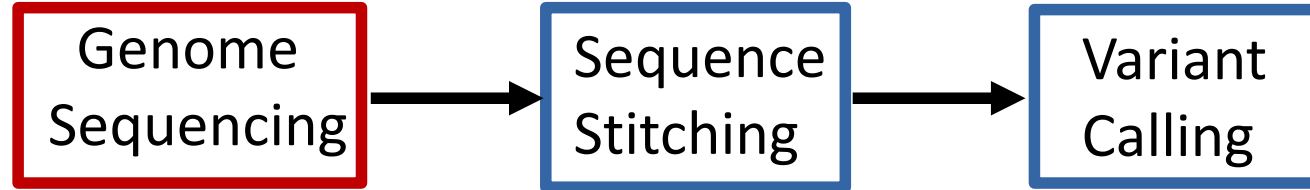
Executive Summary

Generic Genomic Pipeline	Secure Genomic Computation
<div><div>Early Stages</div><div>Dynamic Programming Edit Distance</div></div> <p>Compute Bound</p> <p>The Sequence Alignment Bottleneck</p> <p>Memory Bound</p> <div><div>Early Stages</div><div>Edit Distance</div></div>	<div></div> <p>Security & Privacy for Genomic Workloads</p> <p>Security Properties → Memory Overhead</p> <p>e.g., Intel SGX → Orders of Magnitude</p> <p> Not secure</p> <p>... and even WORSE!</p>

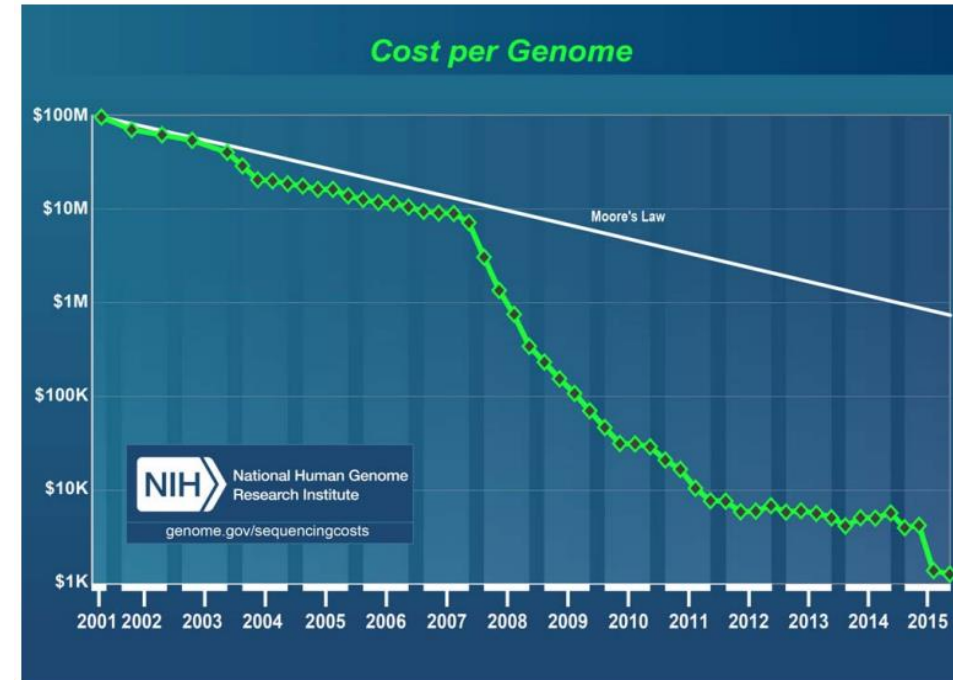
Genome Analysis



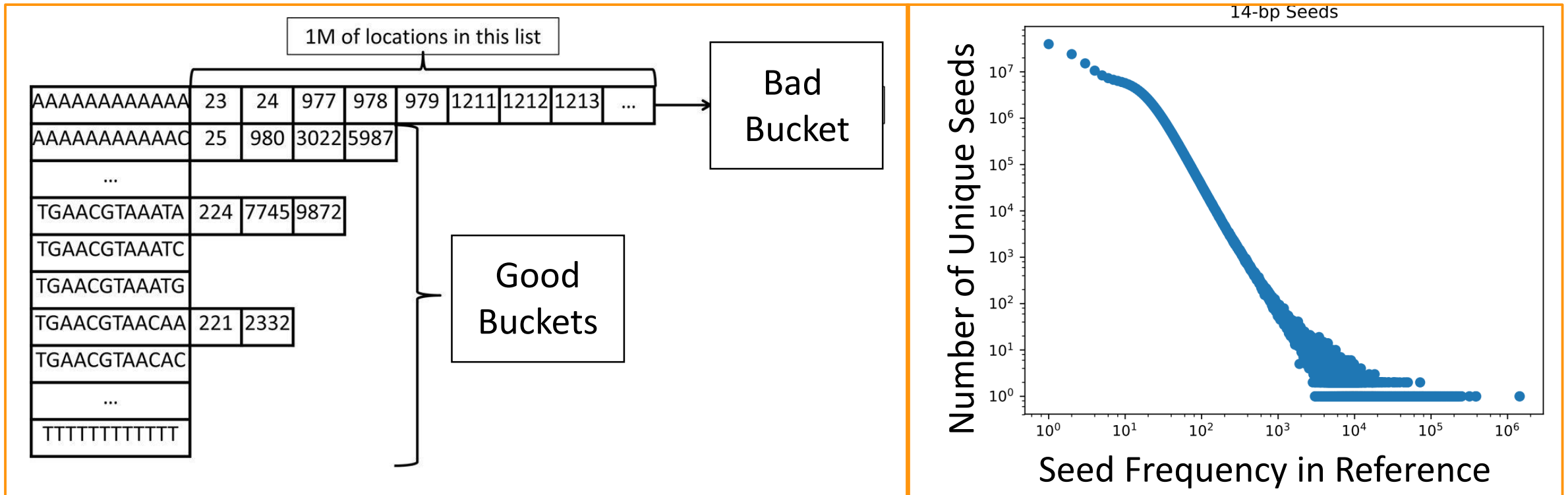
Genome Analysis



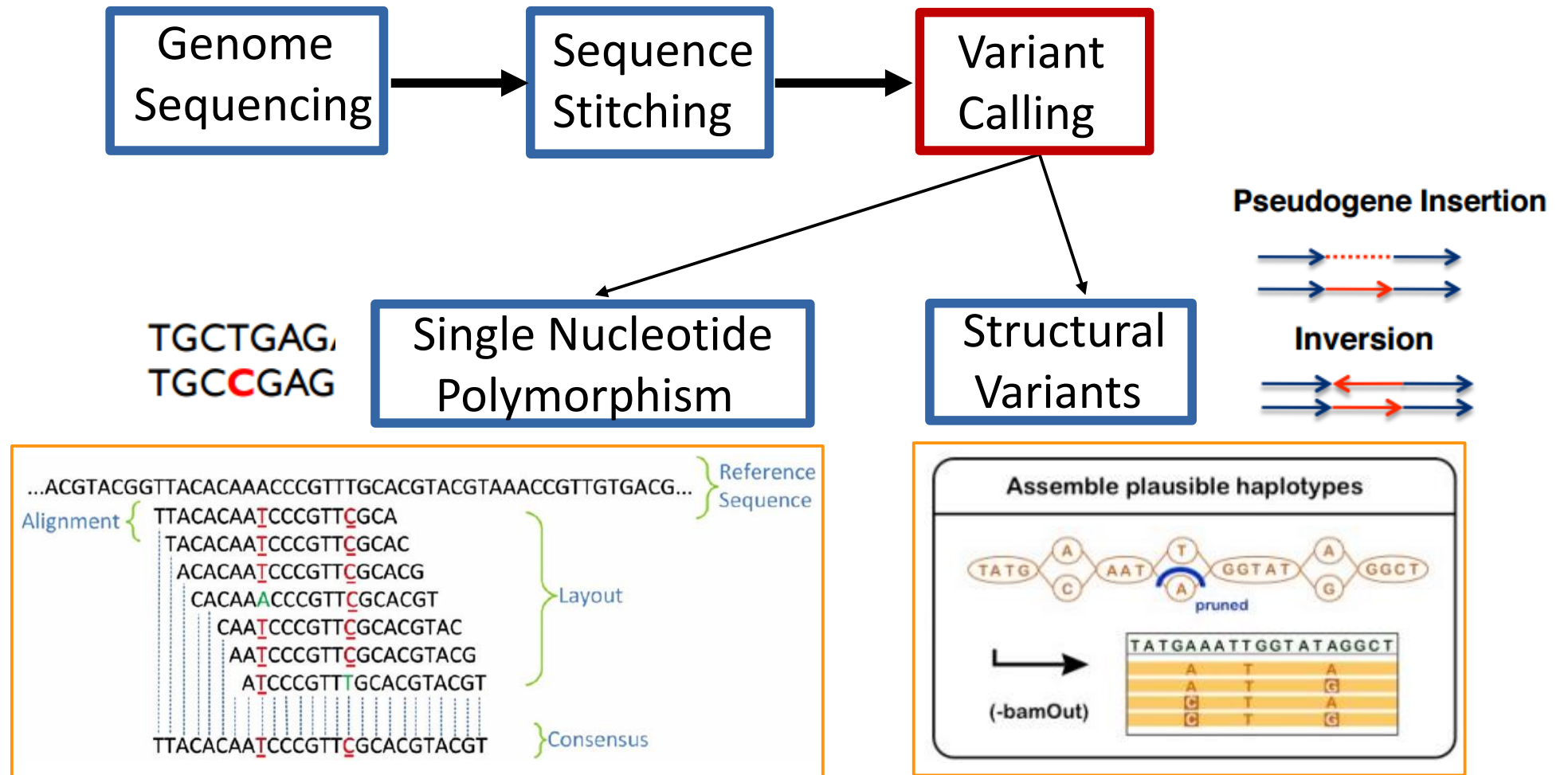
Bandwidth Argument
Wet Lab: 1250 Mbases/min
CPU: 2 Mbases/min



Genome Analysis



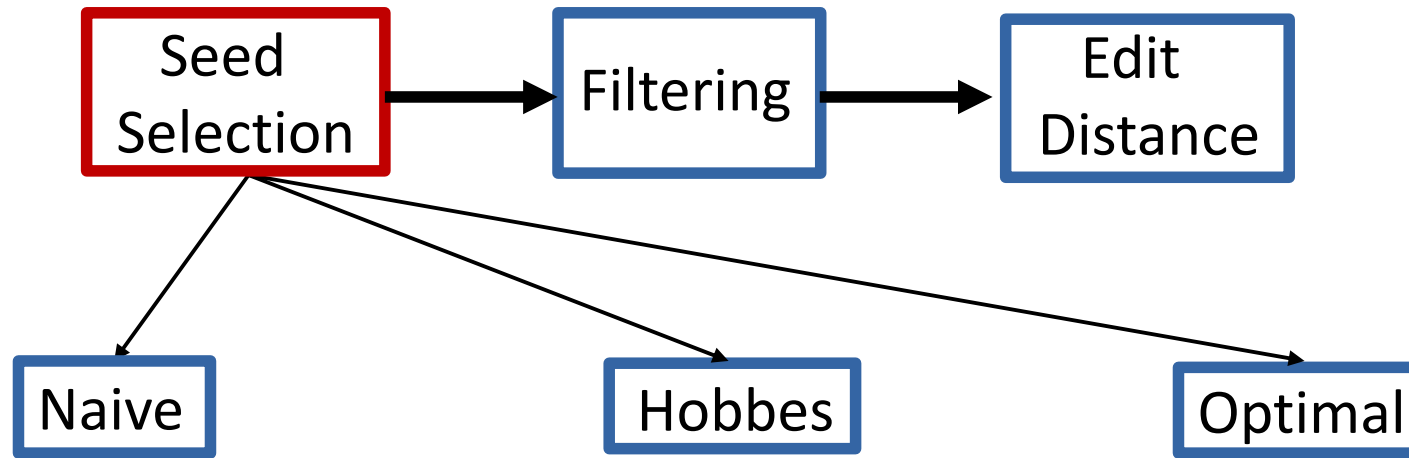
Genome Analysis



Hash Solver Pipeline

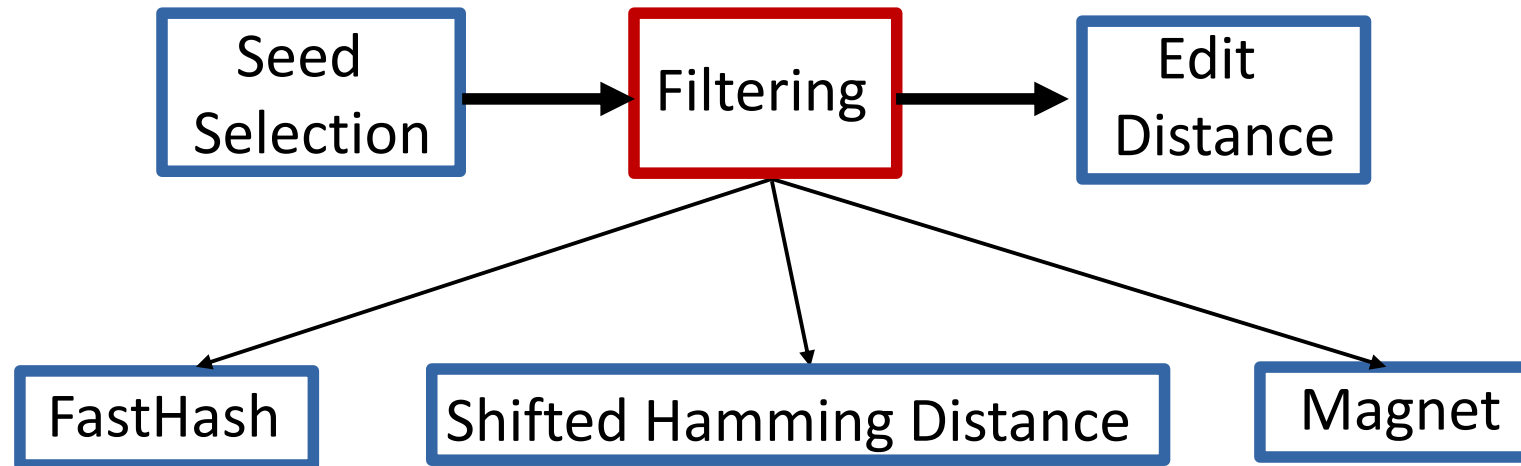


Hash Solver Pipeline



	Naive	Hobbes	Optimal
Memory Fetch	$O(1)$	$O(L)$	$O(L^3)$
Filtering Performance	Baseline	8x less	16x less

Hash Solver Pipeline



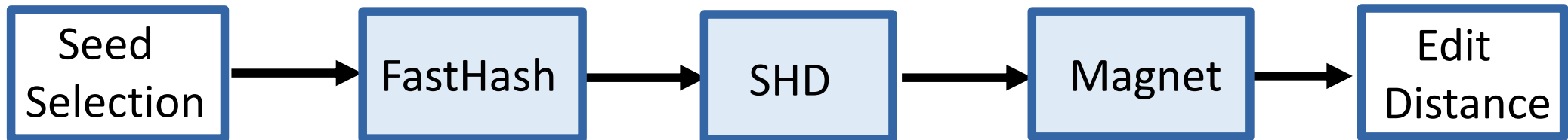
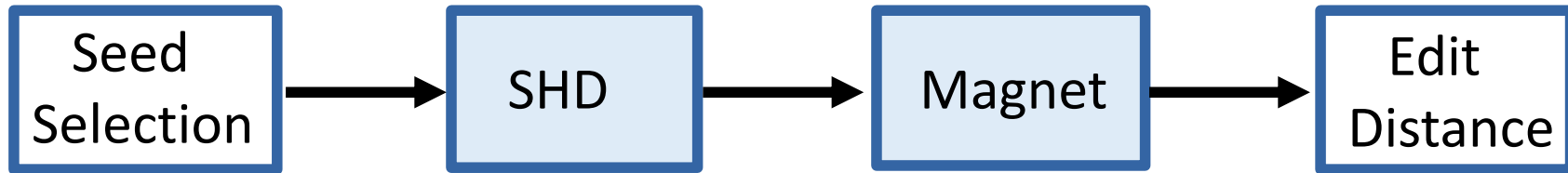
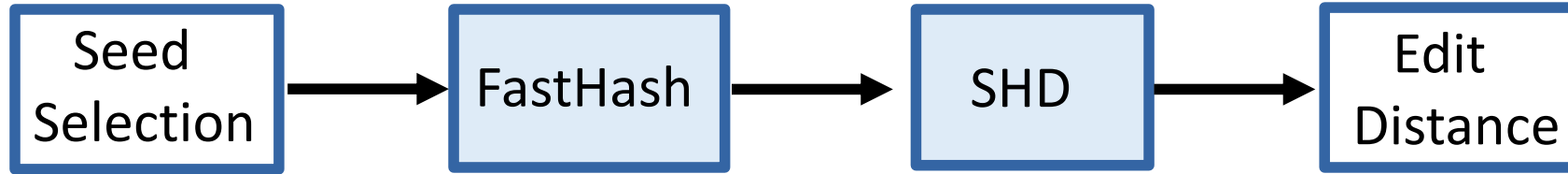
	FastHash	SHD	Magnet
Memory Fetch	$O(1)$	$O(\text{Locations})$	$O(\text{Locations})$
Filtering Performance	Baseline	5x	20x

Hash Solver Pipeline

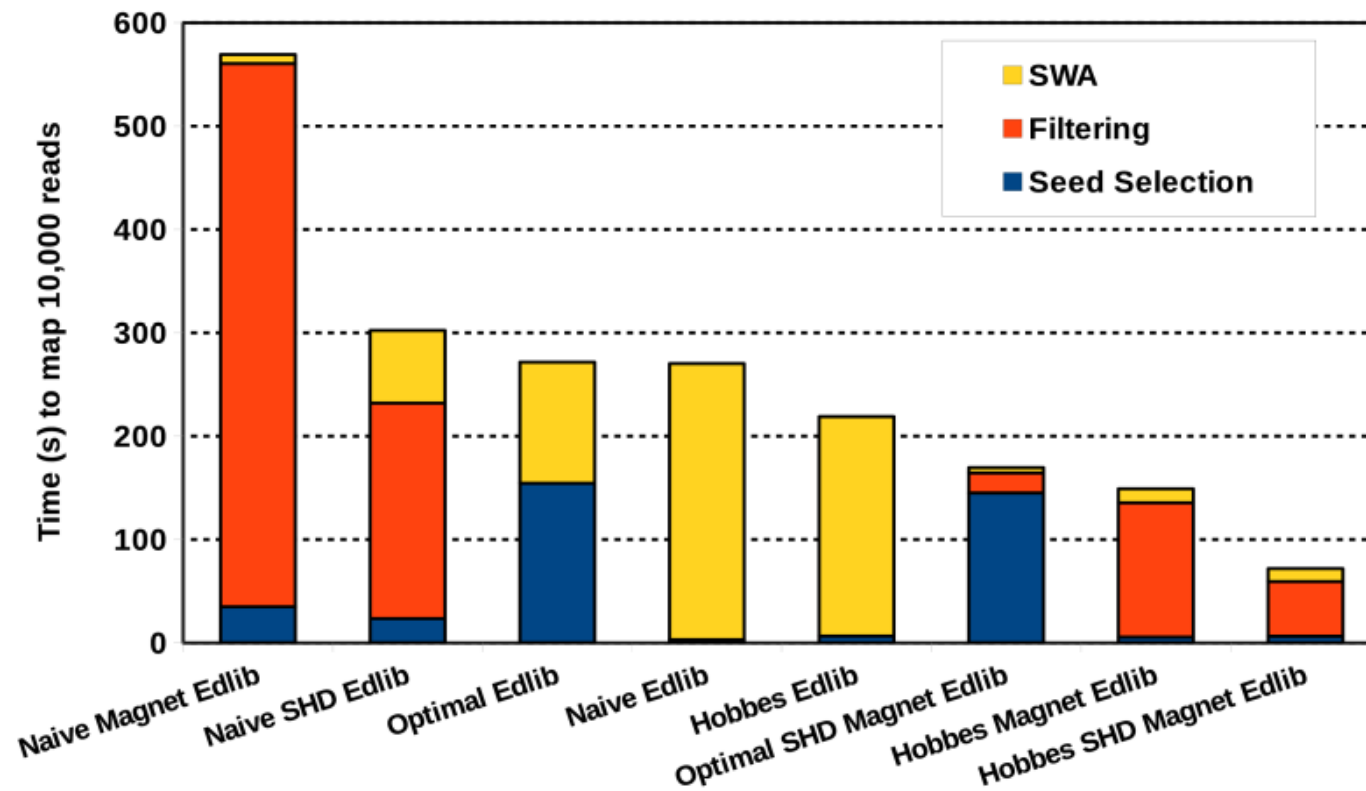


Speedup over scalar
CPU with vector instructions:
16x – 38x

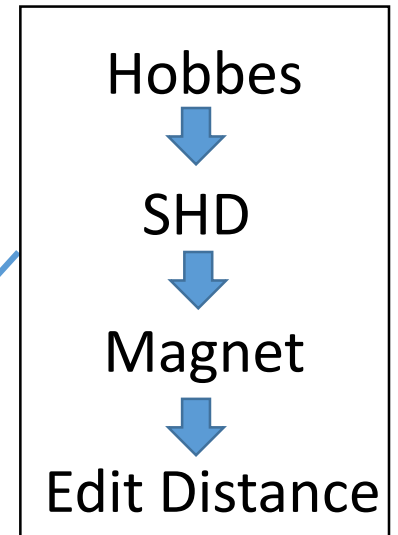
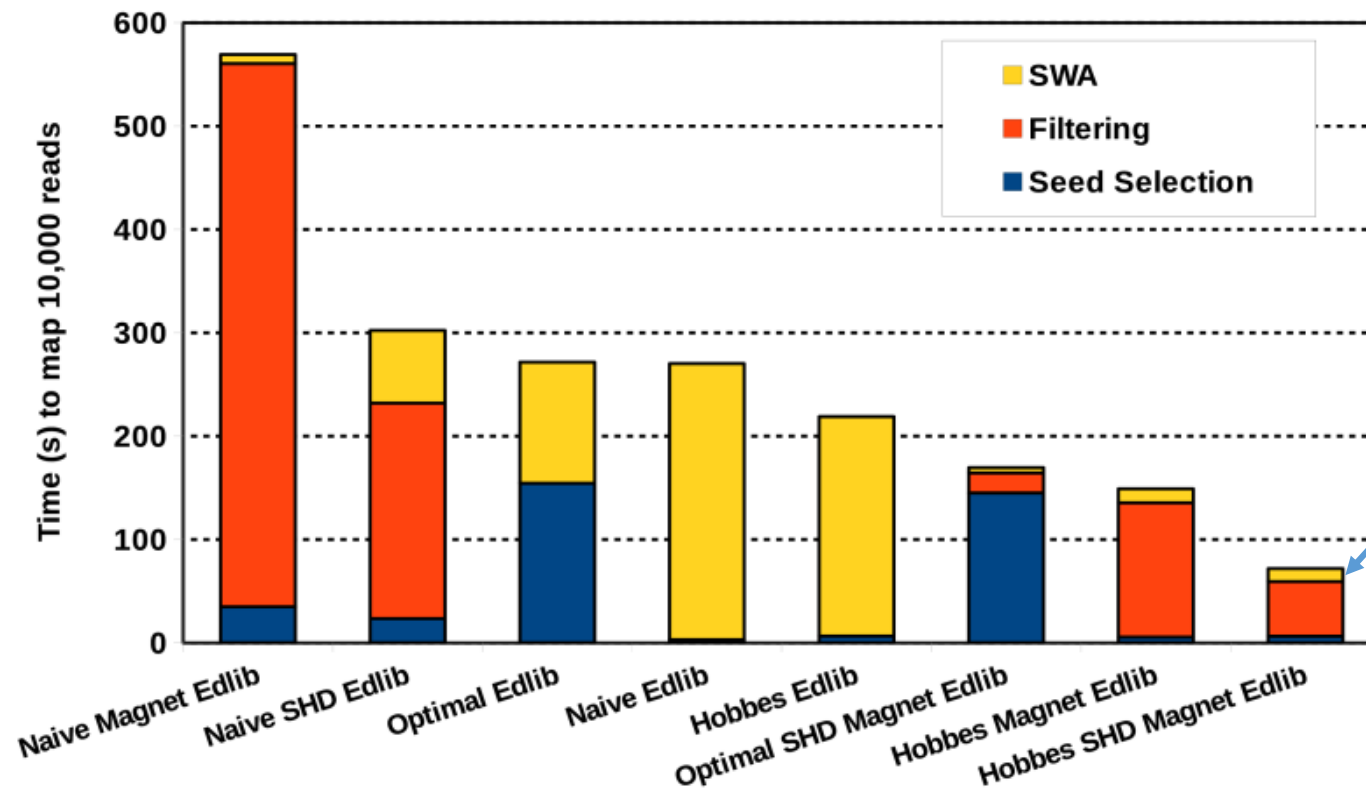
Combining Filtering Algorithms



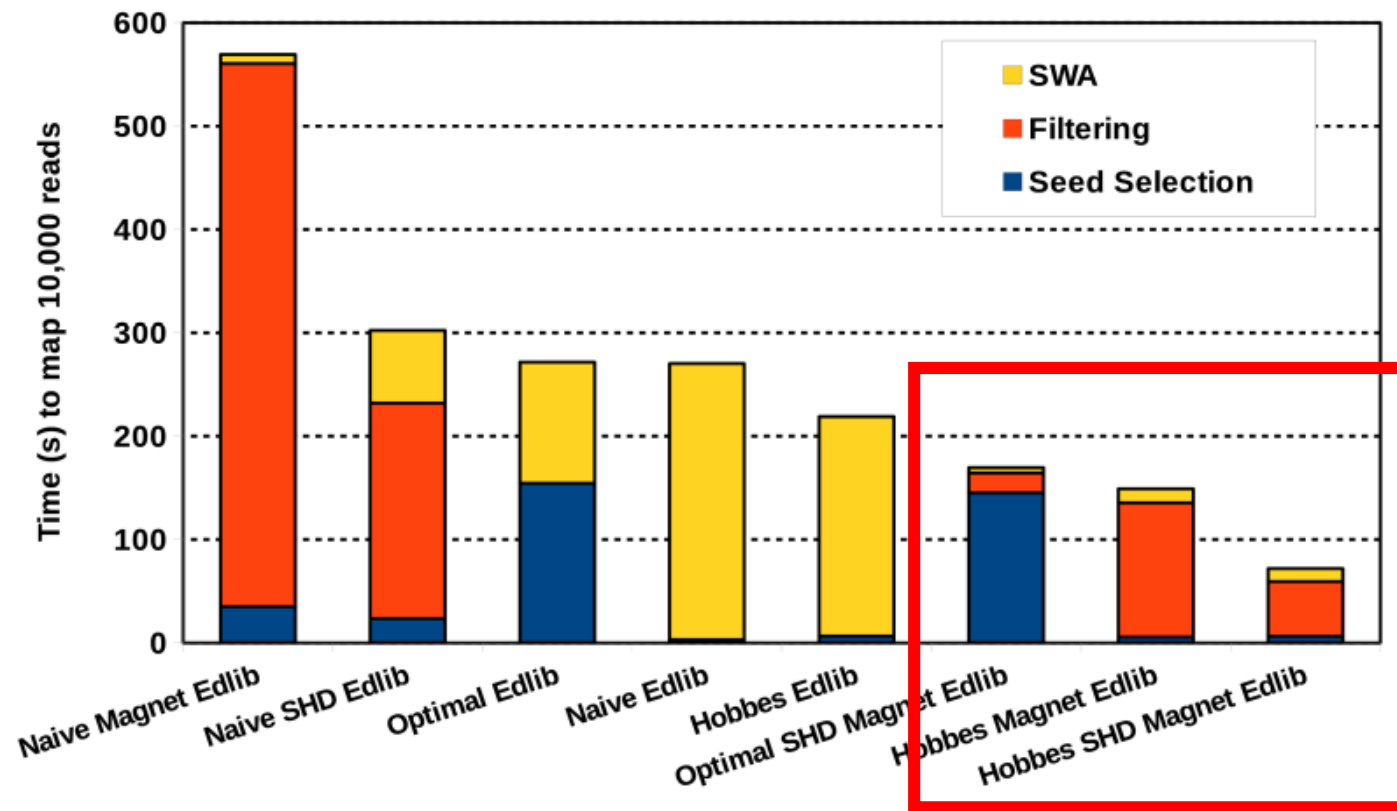
Design Space Exploration



Design Space Exploration



Design Space Exploration




Bottleneck shifts to early stages dominated by memory fetches



Data Breach Alert
COULD AFFECT BLUECROSS BLUESHIELD OF WNY CUSTOMERS

- Newkirk Products, company issuing BCBS plan ID cards
- Server contained names, plan info
- Offer free credit protection

3 STORY
CH LINKED TO BCBS OF WNY
AR-OLD BOY DIES ON KANSAS WATER PAR



UTAH ARCH







Ooops, your files have been encrypted!

What Happened to My Computer?
Your important files are encrypted. Many of your documents, photos, videos, music files and other files are no longer accessible because they have been encrypted. If you are busy looking for a way to recover your files, but do not waste your time. Nobody can recover your files without our decryption service.

Can I Recover My Files?
Sure. We guarantee that you can recover all your files safely and easily. But you have not a lot of time. You must encrypt some of your files. But you want to decrypt them. You only have 3 days to do this. Also, if you don't pay in time, we will have free event.

How Do I Pay?
Payment is accepted in Bitcoin. Please check the current price of Bitcoin. And send the correct amount. After your payment, click on the link to download the decryption tool.

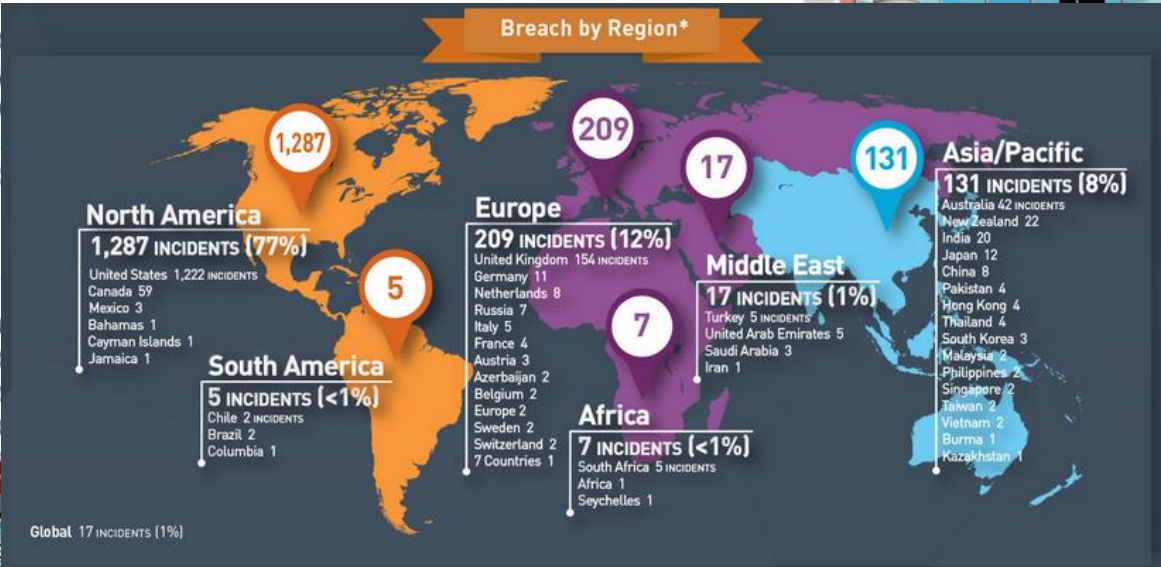
Payment will be raised on
5/16/2017 00:47:55

Time Left
02:23:57:37

Your files will be lost on
5/20/2017 00:47:55

Time Left
06:23:57:37

Over 200,000 victims



Data Breach Alert

COULD AFFECT BLUECROSS BLUES

Newkirk Products, com
issuing BCBS plan ID c

Server contained names,

plan info

offer free

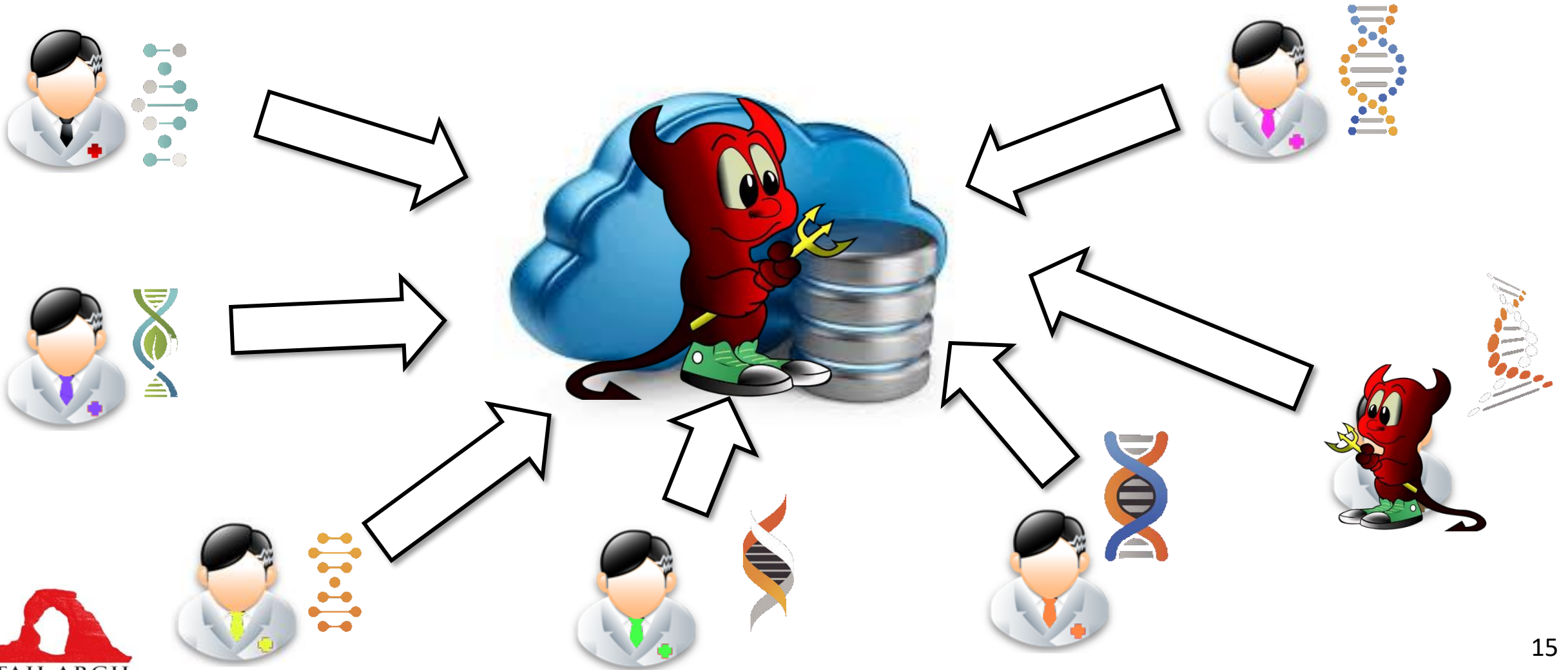
left protection

3 STORY

CH LINKED TO BCBS OF WNY

AR-OLD BOY DIES ON KANSAS WATER PAR

Security & Privacy for Genomic Data

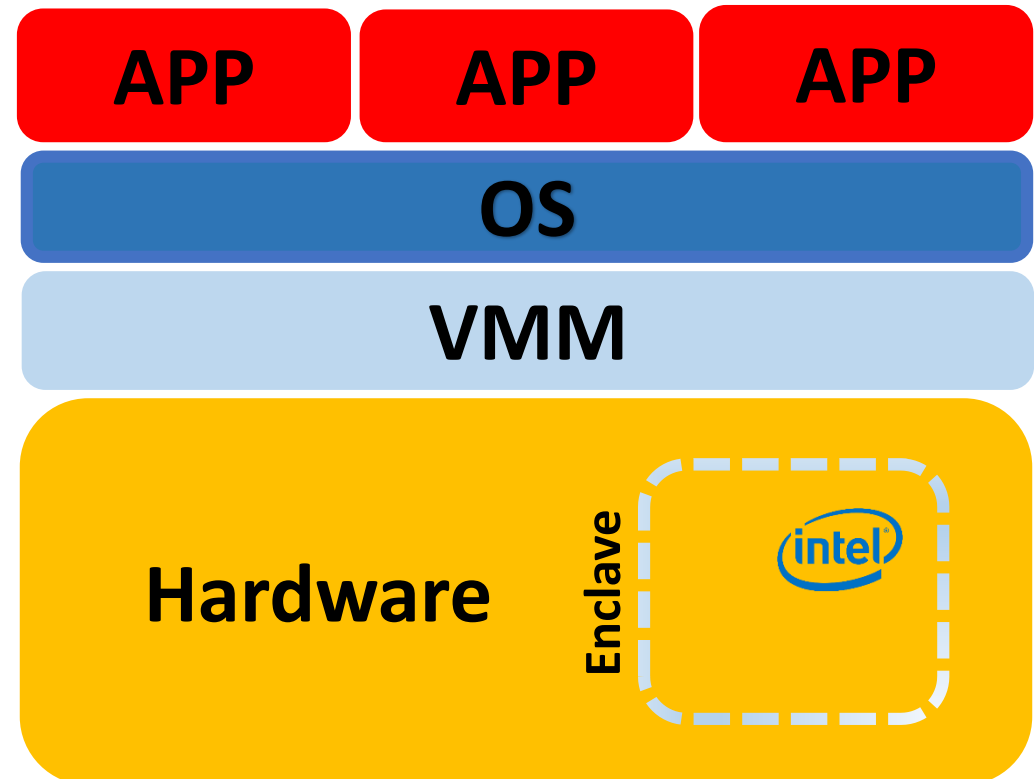


Secure Genome Computation

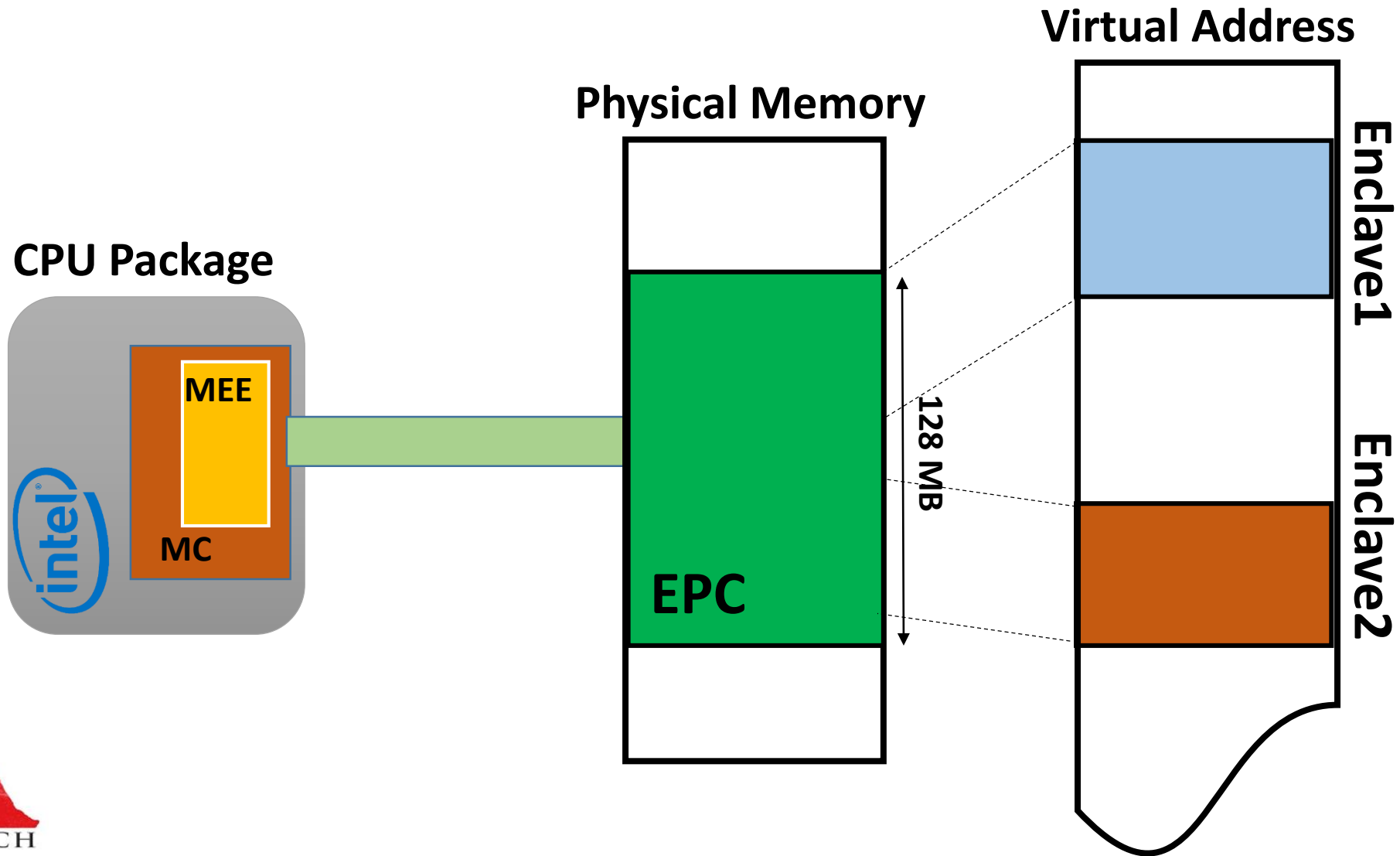
Intel® Software Guard eXtensions
(Intel® SGX)

Intel® SGX Provides

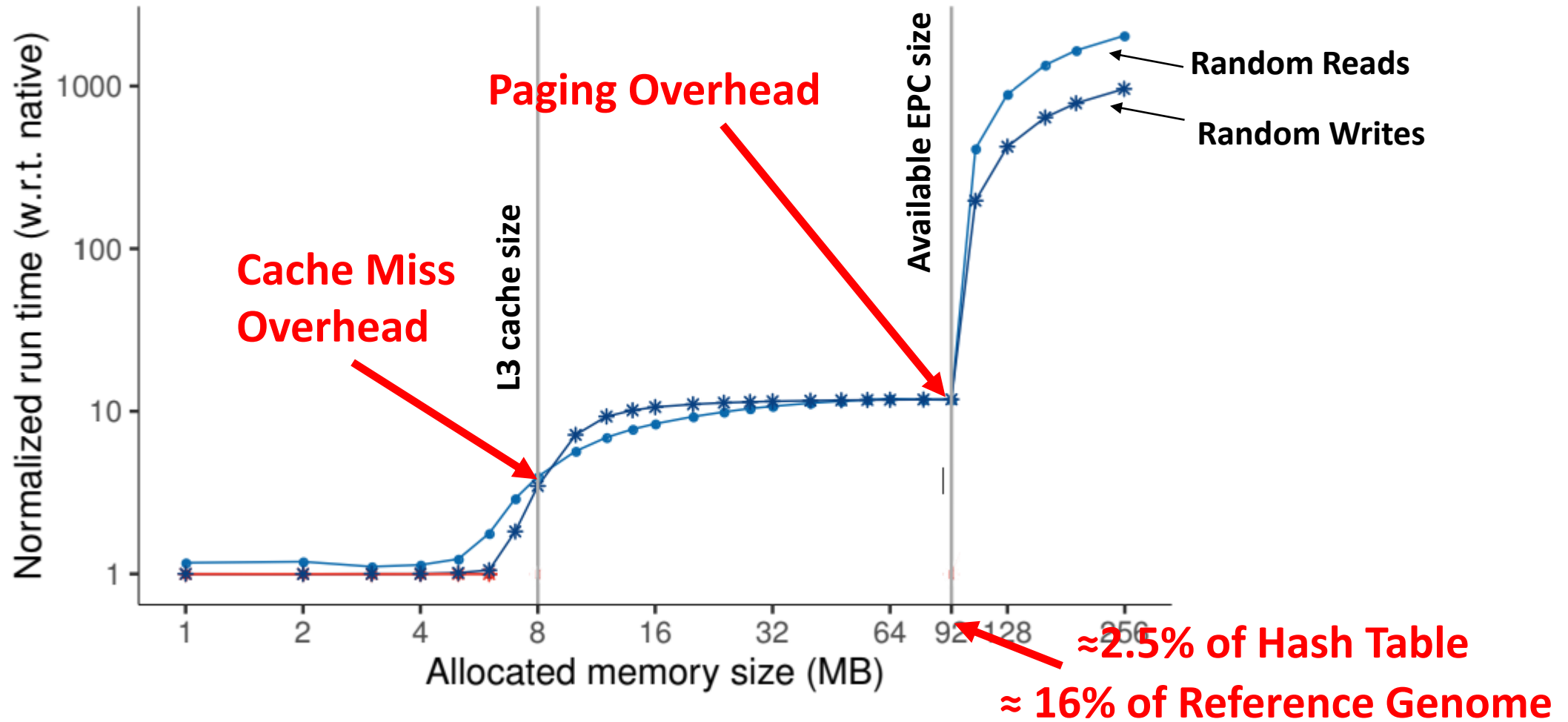
- ☐ Confidentiality
- ☐ Integrity Verification



Intel® SGX Overhead



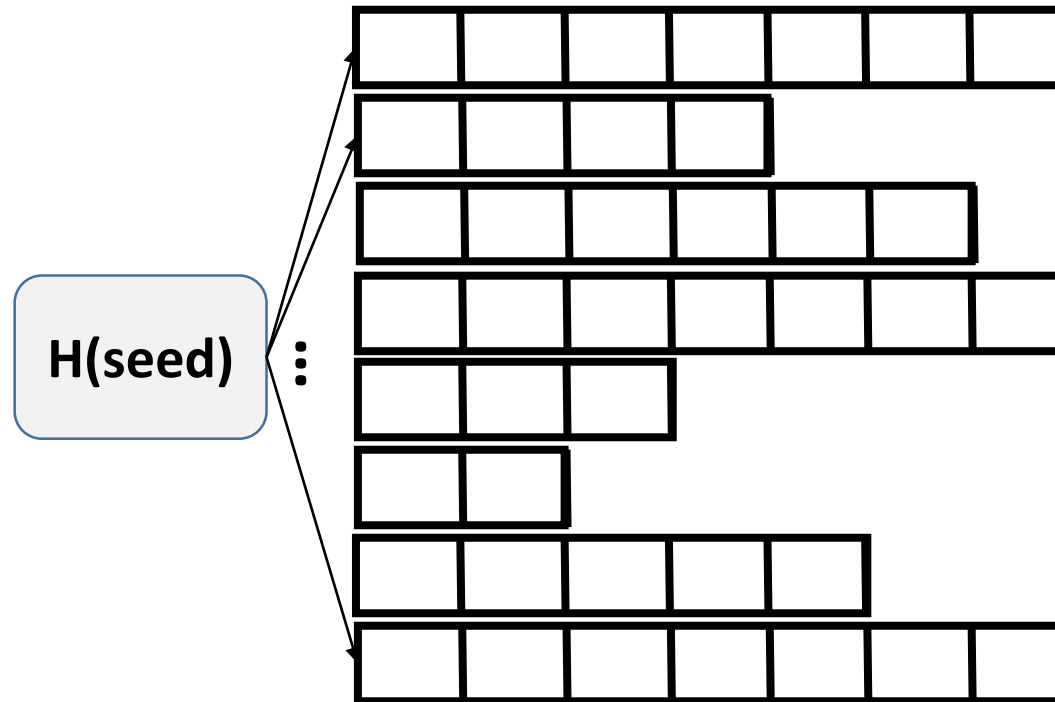
Intel[®] SGX Overhead (Cont.)



Intel[®] SGX is not Safe Enough!

Information Leakages

- Page Level Access pattern
 - Controlled-Channel Attack



Intel[®] SGX is not Safe Enough!

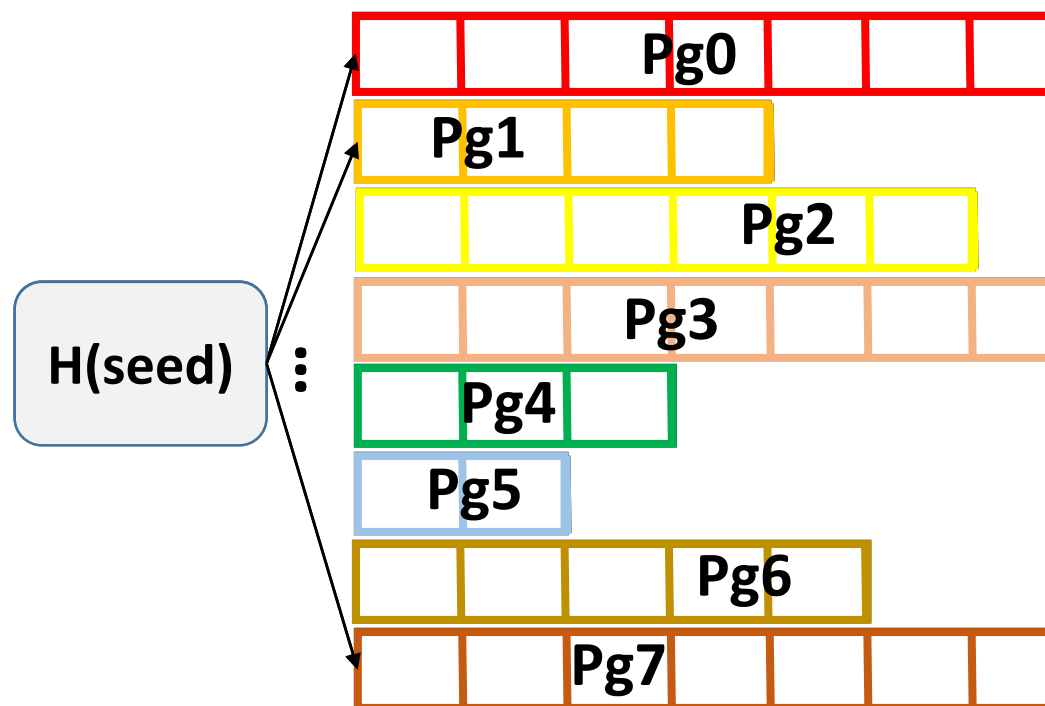
Information Leakages

- Page Level Access pattern
 - Controlled-Channel Attack

Access → Seed **0**, **7**, **5**, **7**, **4**, **4**

↓ ↓ ↓ ↓ ↓ ↓

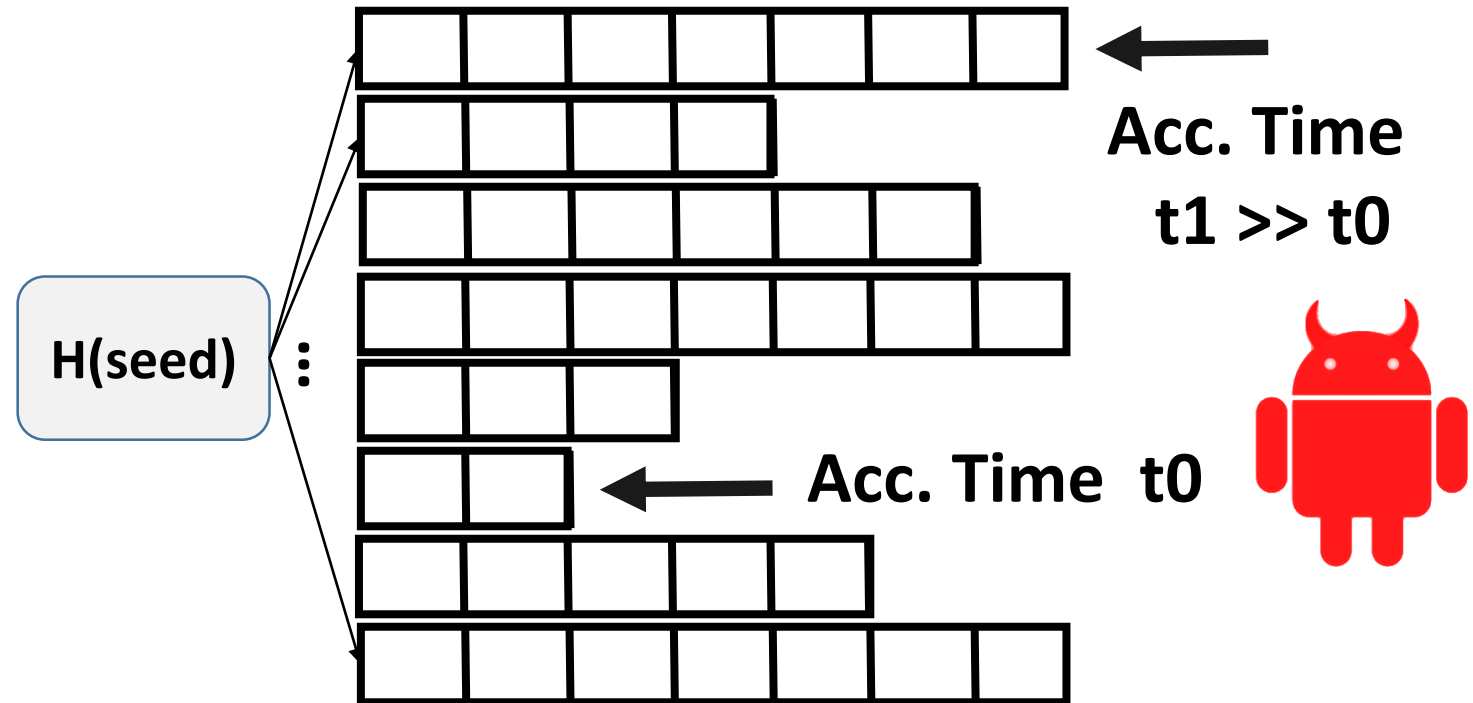
Page fault → page 0, 7, 5, 7, 4, 4



Intel® SGX is not Safe Enough!

Information Leakages

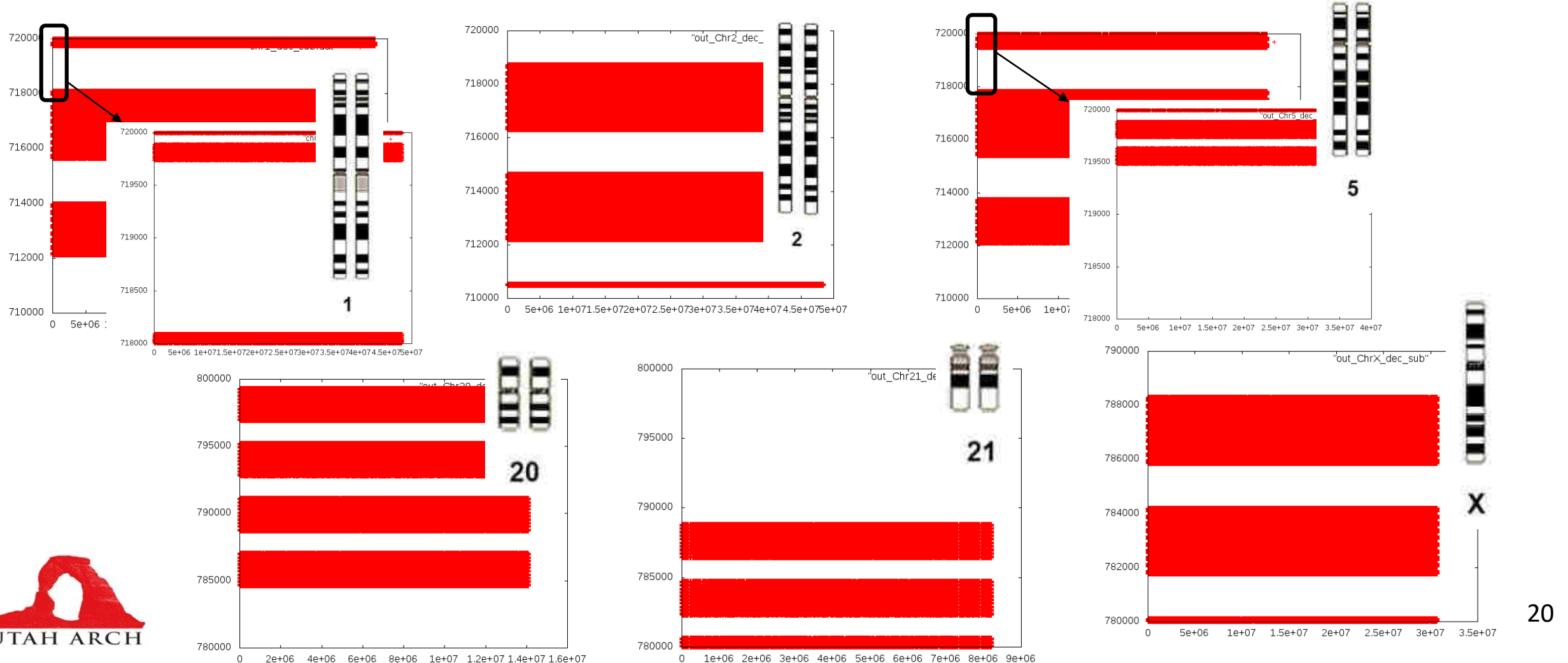
- Page Level Access pattern
- Controlled-Channel Attack
- Timing Side Channel
 - Memory Intensity



Variant Calling Footprint

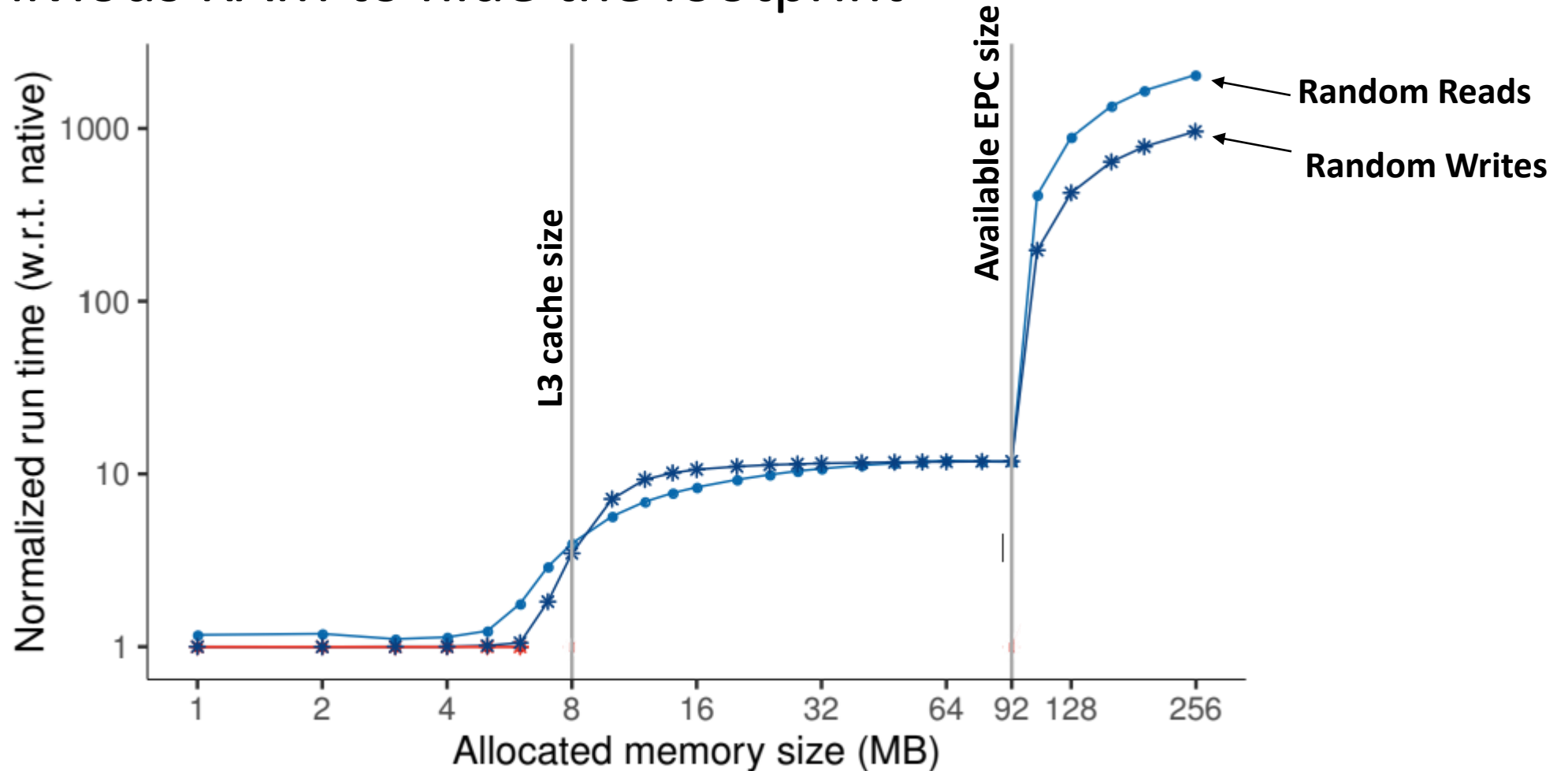
SAMTools → Variant Calling for different Chromosomes

Intel PIN → Track virtual addresses



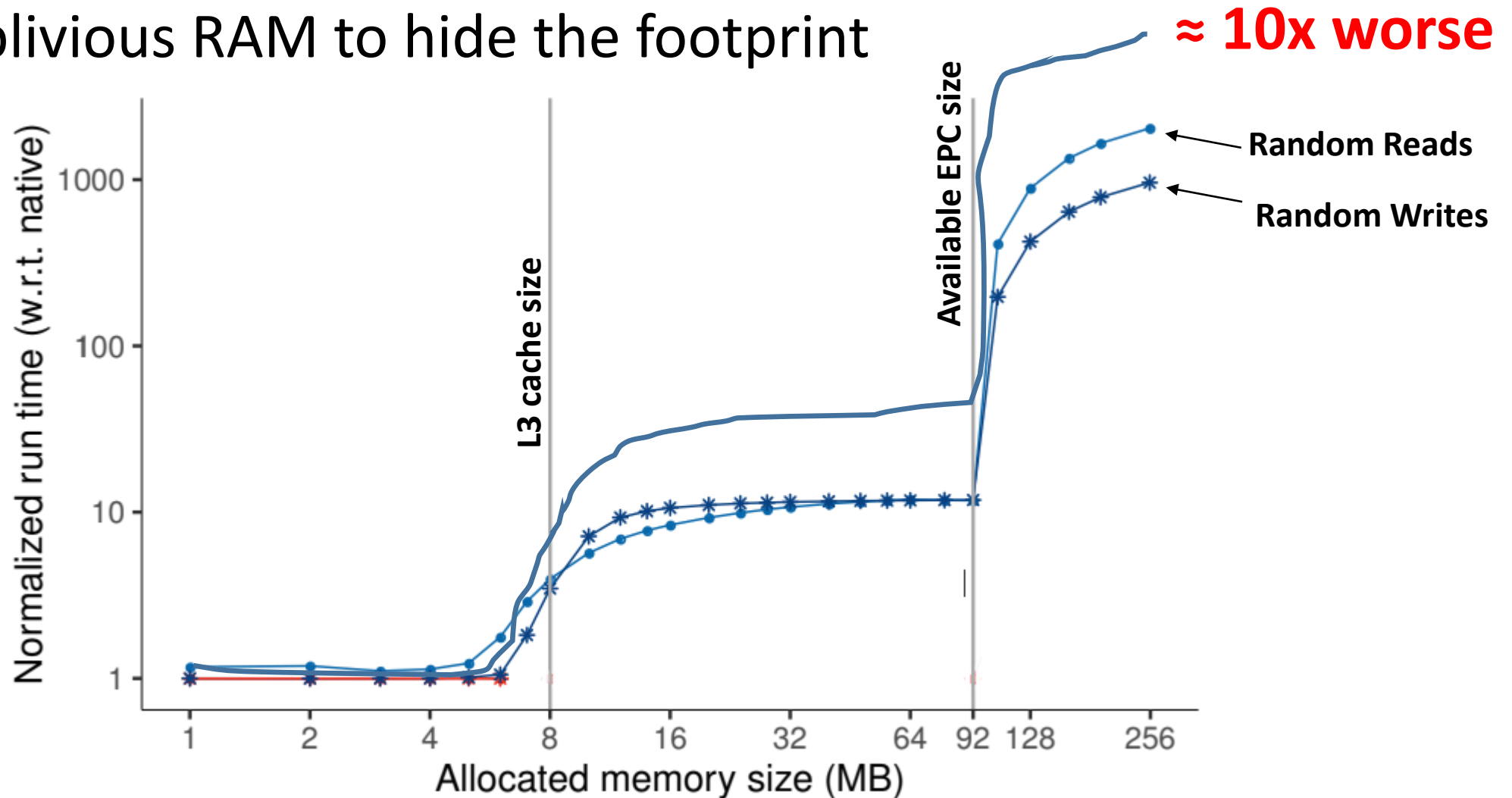
Even More Performance Reduction?

Using Oblivious RAM to hide the footprint



Even More Performance Reduction?

Using Oblivious RAM to hide the footprint



Conclusion

- Sequence alignment is memory bound when edit distance is replaced with simpler computations.
- Information leakages may threaten privacy in genomic workloads
- Security requirements intensify the memory bandwidth dramatically