SSD Failures in Datacenters: What? When? And Why?

Iyswarya Narayanan, Di Wang, Myeongjae Jeon, Bikash Sharma, Laura Caulfield,
Anand Sivasubramaniam, Ben Cutler, Jie Liu, Badriddine Khessib, Kushagra Vaid

The 9th ACM Systems And Storage Conference (SYSTOR 2016)

Why SSD Reliability ?

Data reliability

01001100 01001101 11010010 01000000
10011100 10111111 10101111 11000101

SSDs' popularity

46.5% annual growth*

Datacenter decision support

Limited field data

*Source: IDC, Dec 2015

Why SSD Reliability ?

Data reliability

01001100 01001101 11010010 01000000
10011100 10111111 10101111 11000101

SSDs' popularity

46.5% annual growth*

Large scale

Field data

Datacente support

Limited field data

*Source: IDC, Dec 2015

SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.

SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

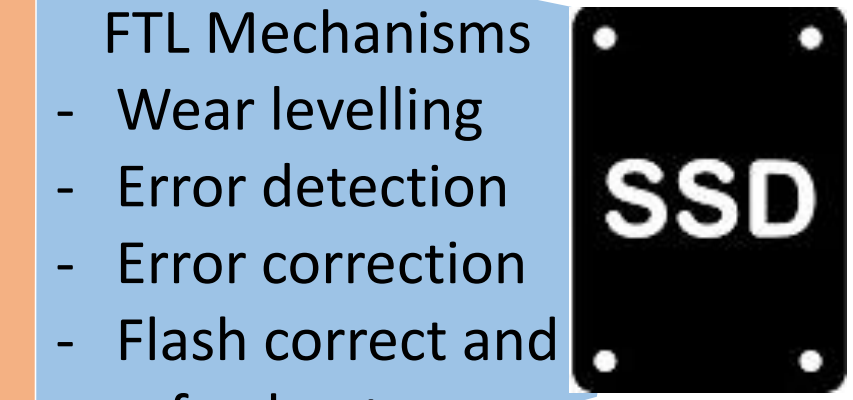
FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.

SSD

SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.



SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.

SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.

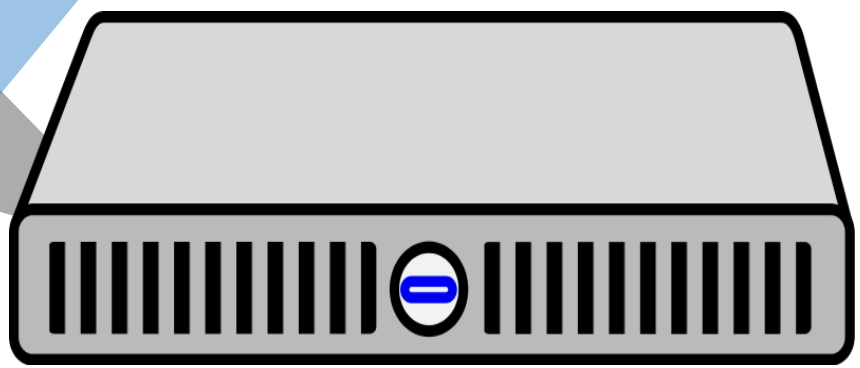
SSD Failures

Flash failures
- Media wear-out
- Data Retention
- Program disturb
- Erase disturb

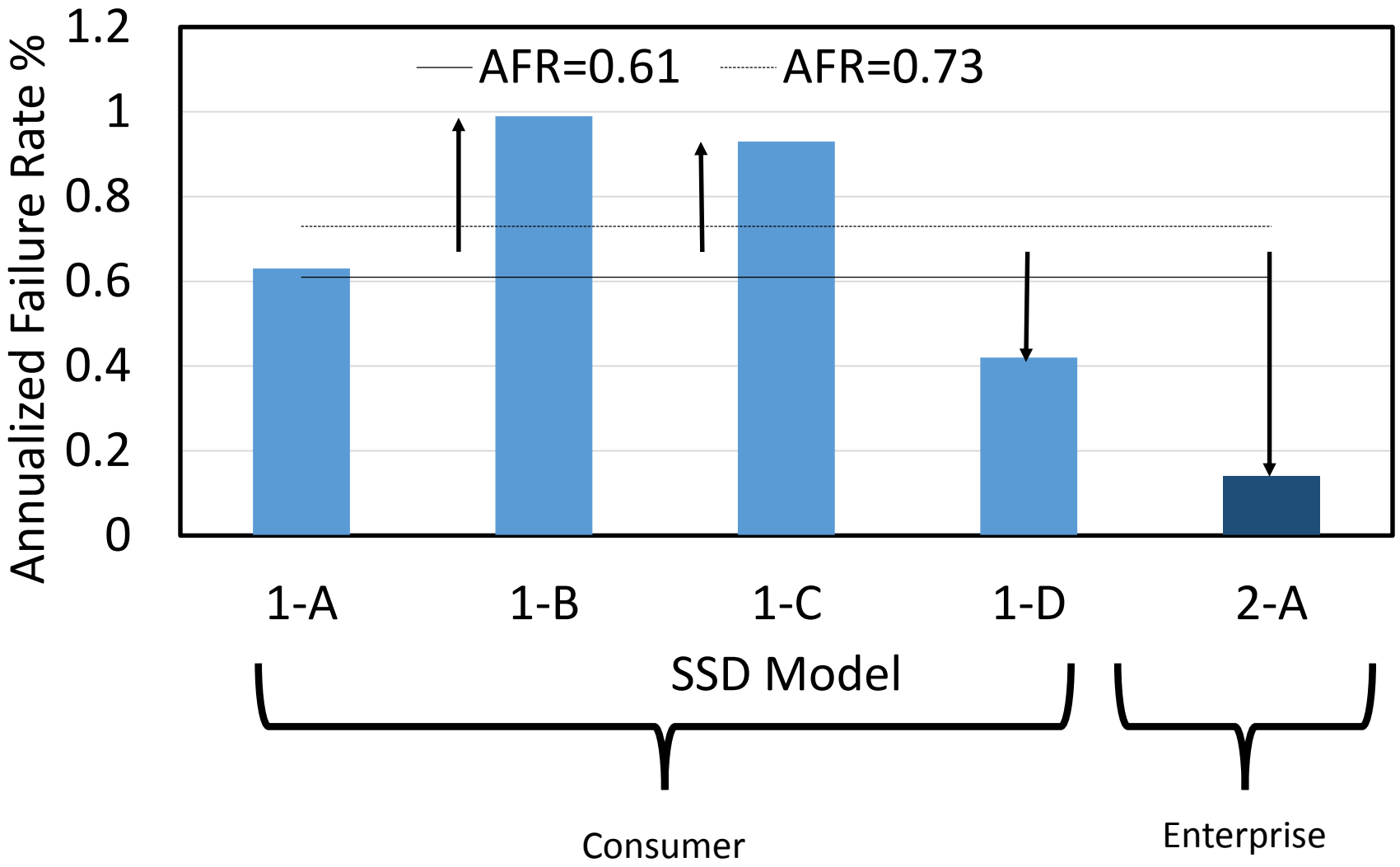
FTL Mechanisms
- Wear levelling
- Error detection
- Error correction
- Flash correct and refresh, etc.

SSD

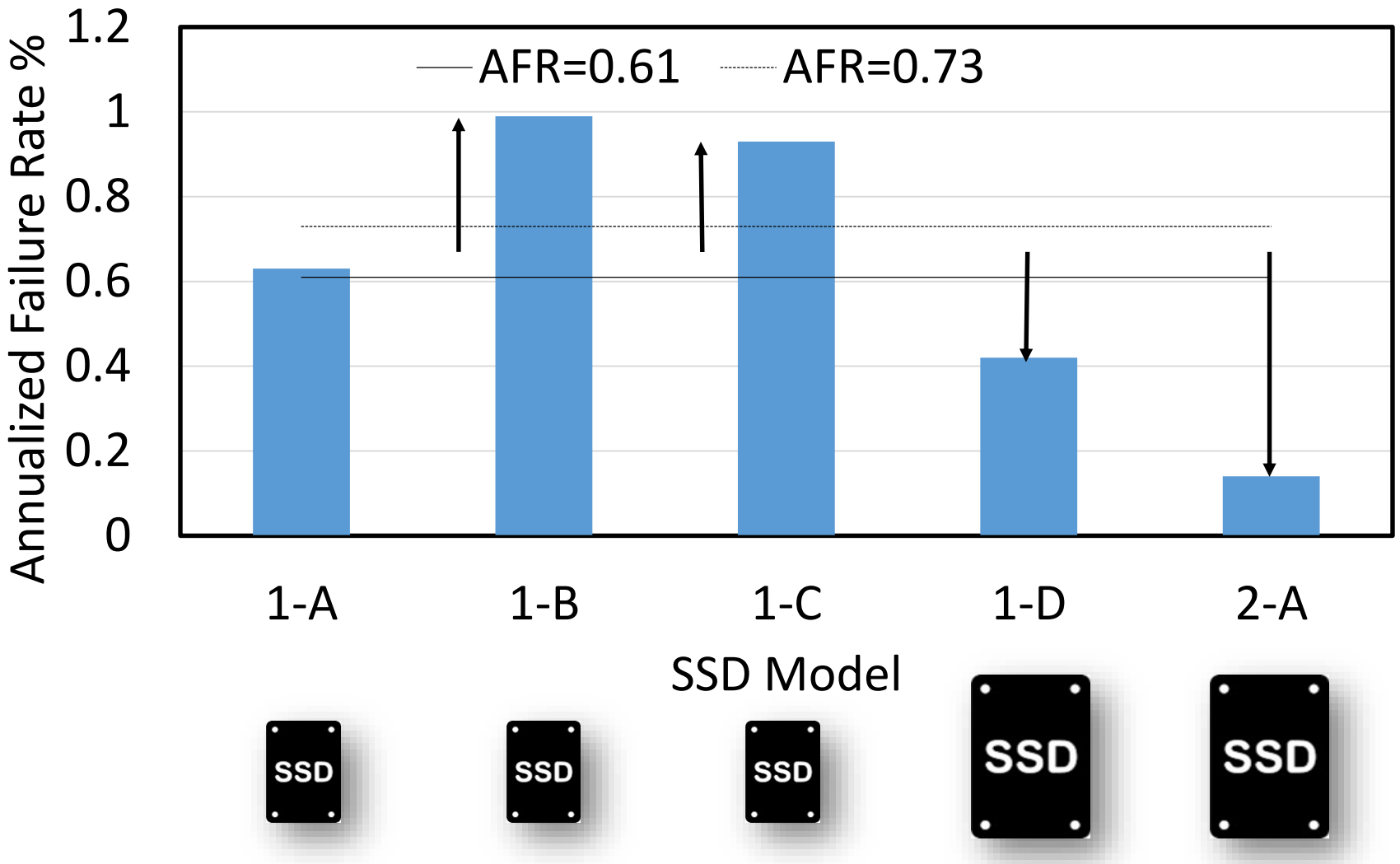
Fail-stop failures



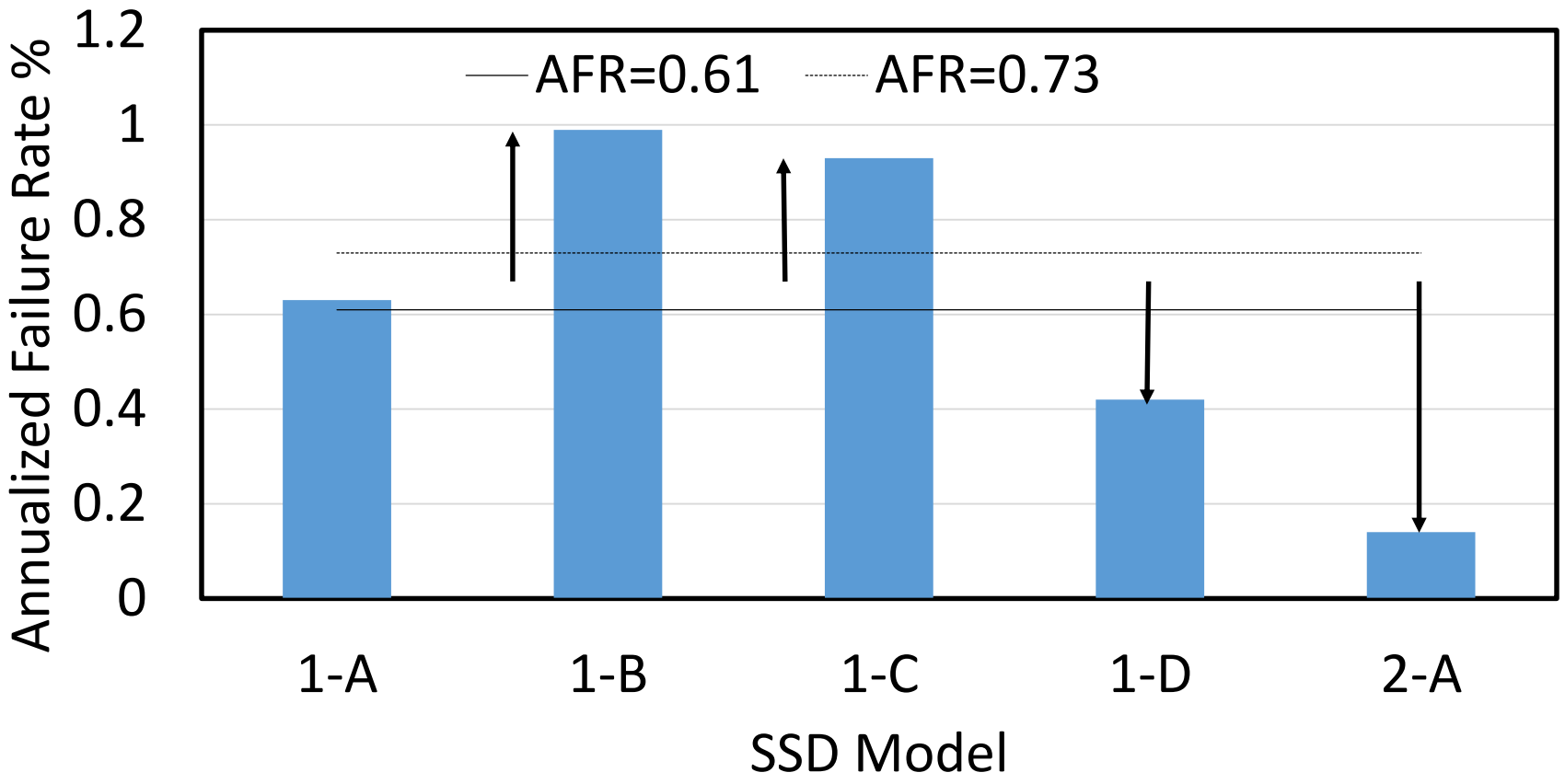
SSD Reliability



SSD Reliability

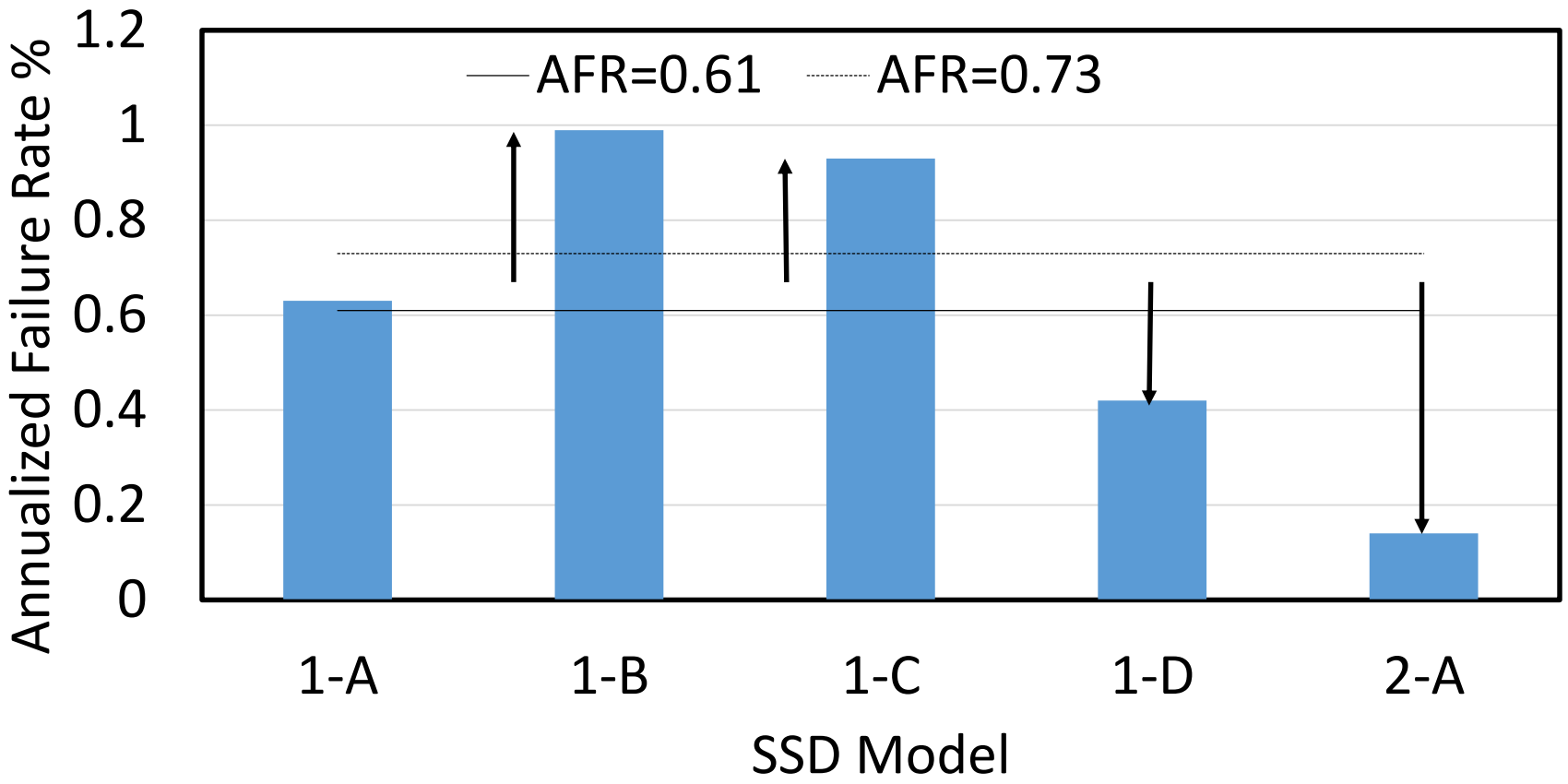


SSD Reliability



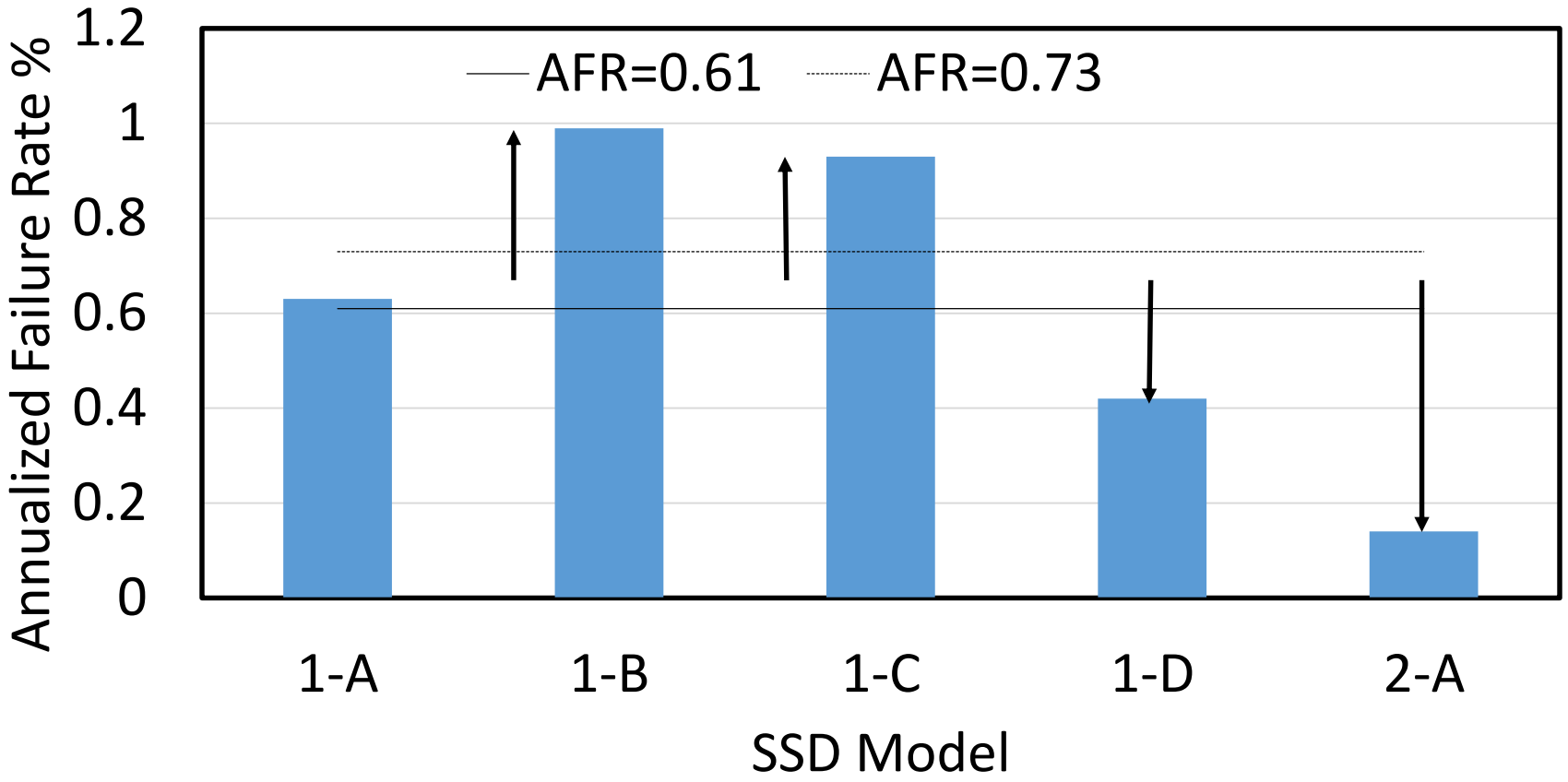
5 large datacenters

SSD Reliability



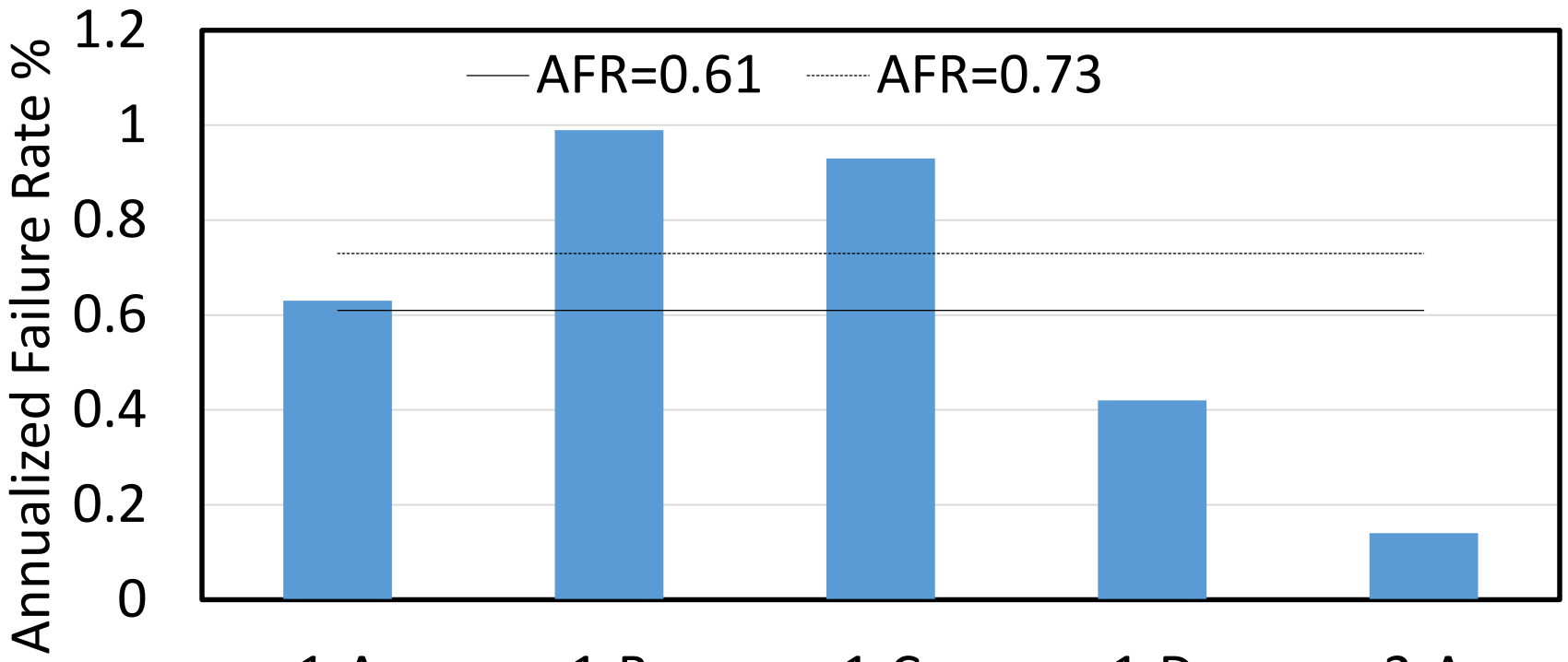
4 major workloads

SSD Reliability



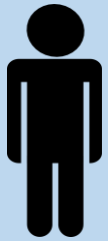
6 different rack SKUs

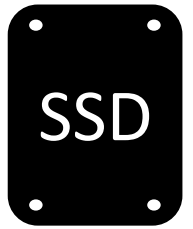
Various factors in production environment could affect SSD failure trends very differently from lab test conditions



Can we understand SSD failures in the presence of various factors ?

Understanding SSD Failures – An analogy

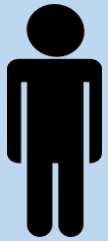




Reactive

Proactive

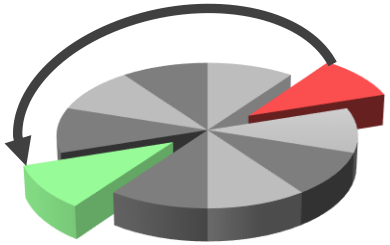
What are the symptoms?



Fever

Unexpected weight loss

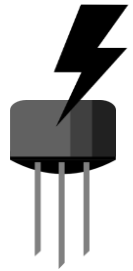
Low blood pressure



Reallocated sectors

011001?00101?

Data errors

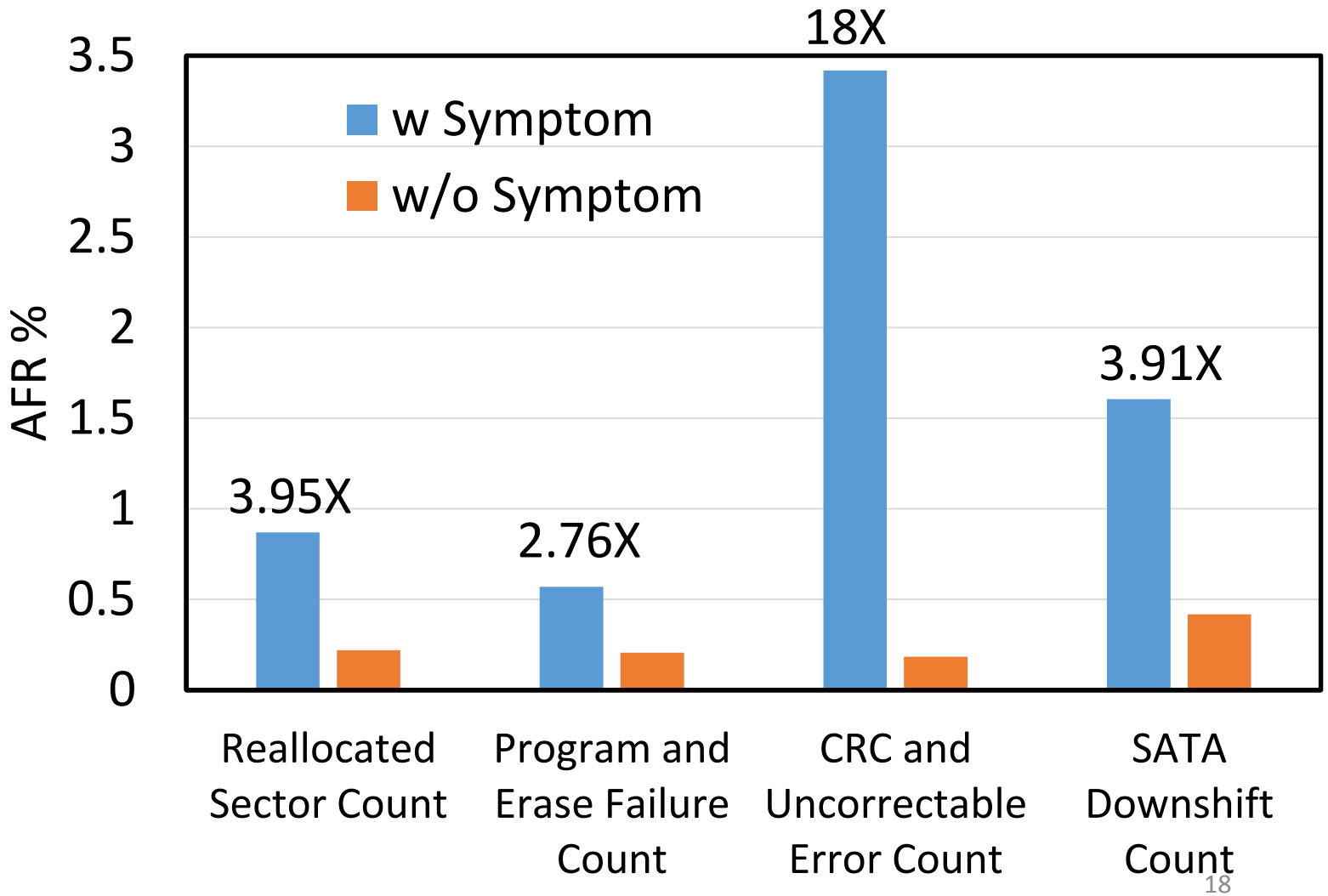


Program and erase failure

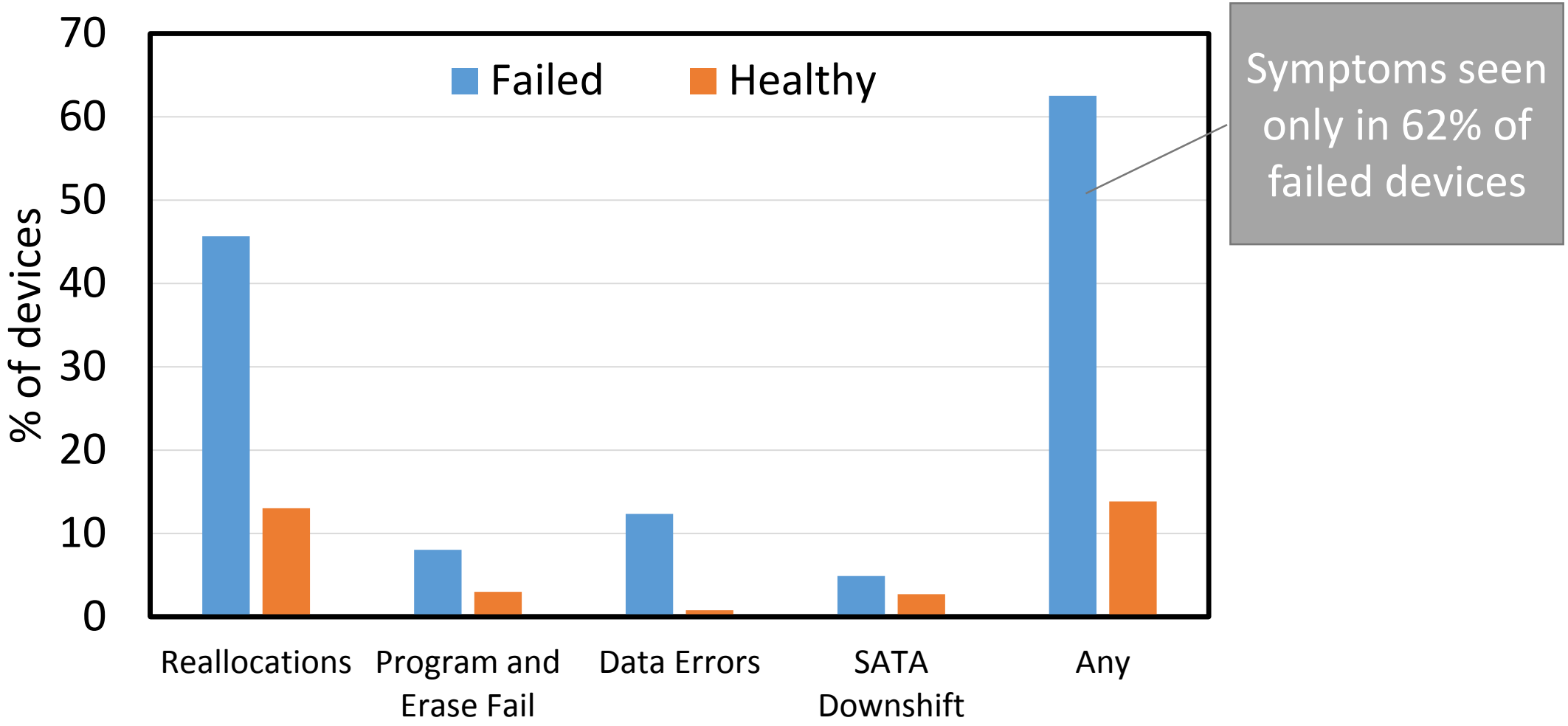
SATA downshift

SSD Failure Symptoms

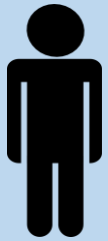
- Reallocated Sector Count
- Program and Erase Fail Count
- CRC and Uncorrectable Error Count
- SATA Downshift Count

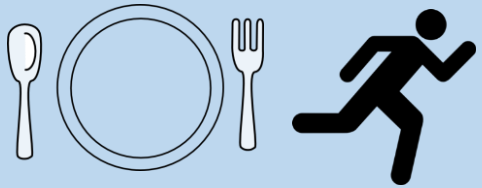


Insufficiency of symptom only diagnosis



What are the factors?

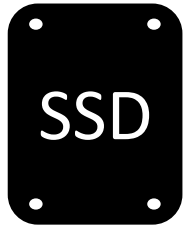




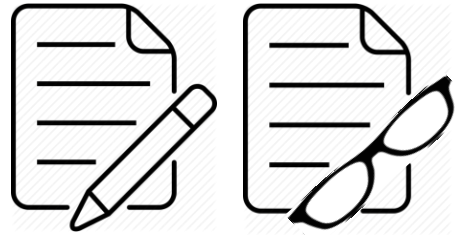
Lifestyle

Environmental agents

Genetics

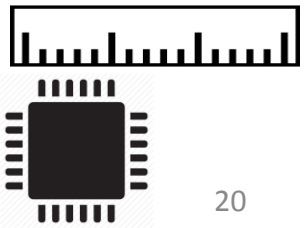


Workload



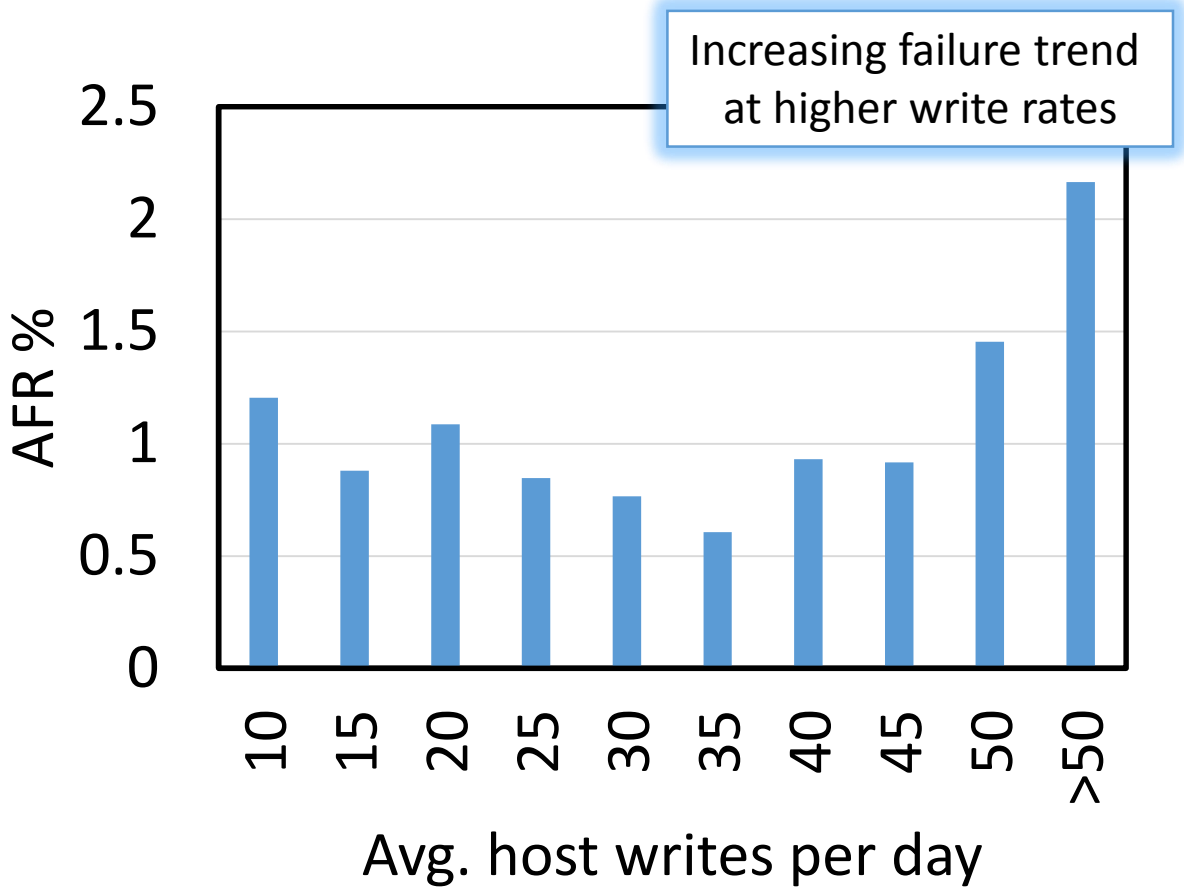
Production environment

Design decisions



Device level correlating factors

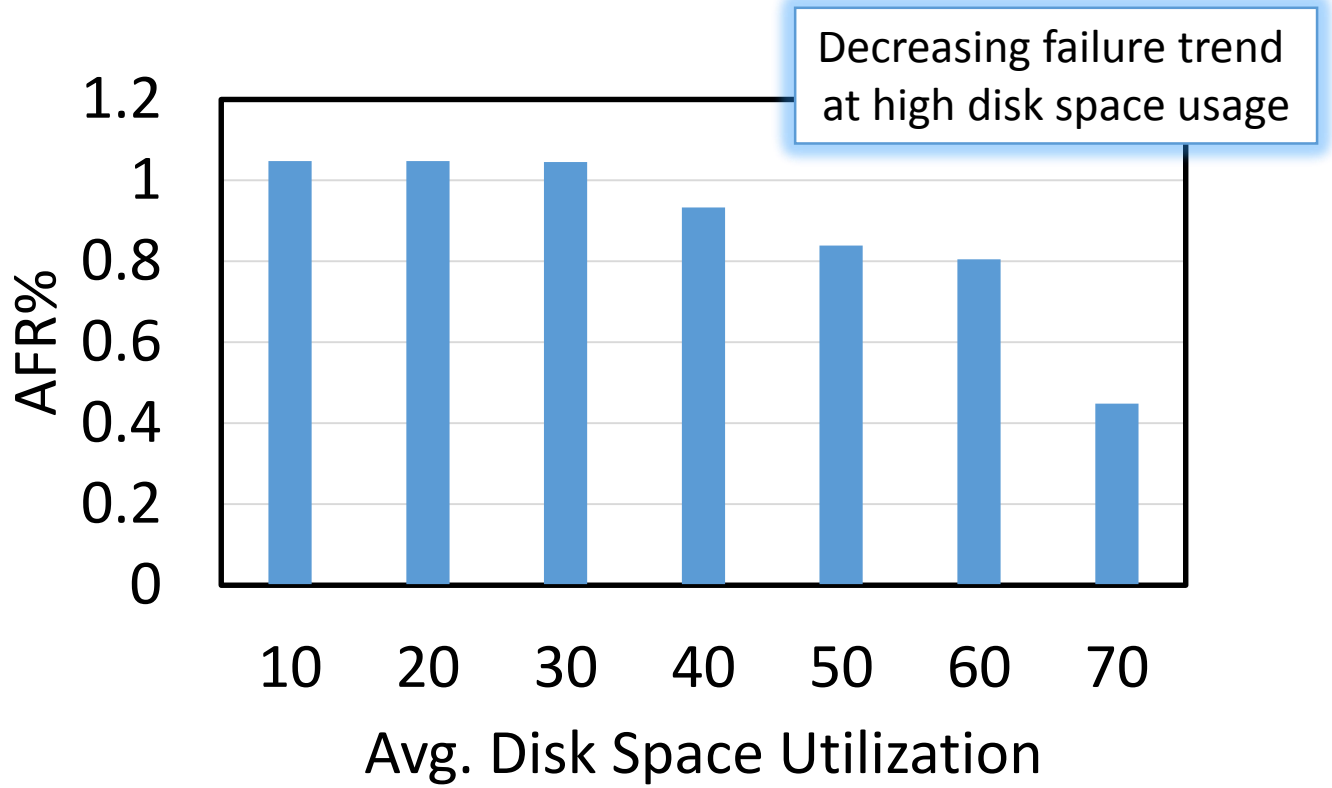
- Average write rate of a device
- Average read rate of a device
- Total read and/or write usage
- Write Amplification
- Read Write Ratio



More results in the paper

Server level correlating factors

- SSD space utilization
- Disk space utilization
- Memory utilization
- Processor utilization

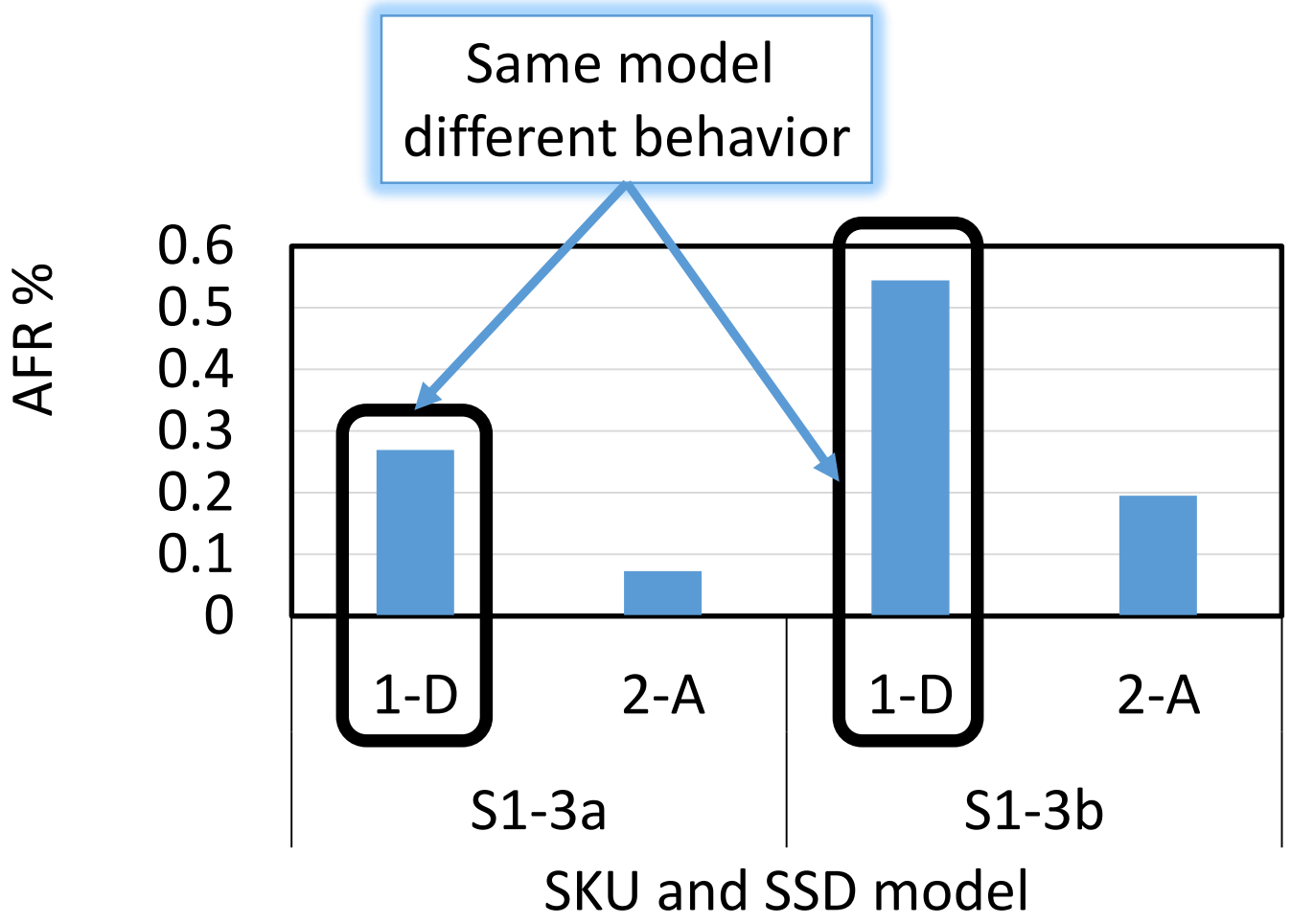


More results in the paper

Datacenter factors

Rack SKU

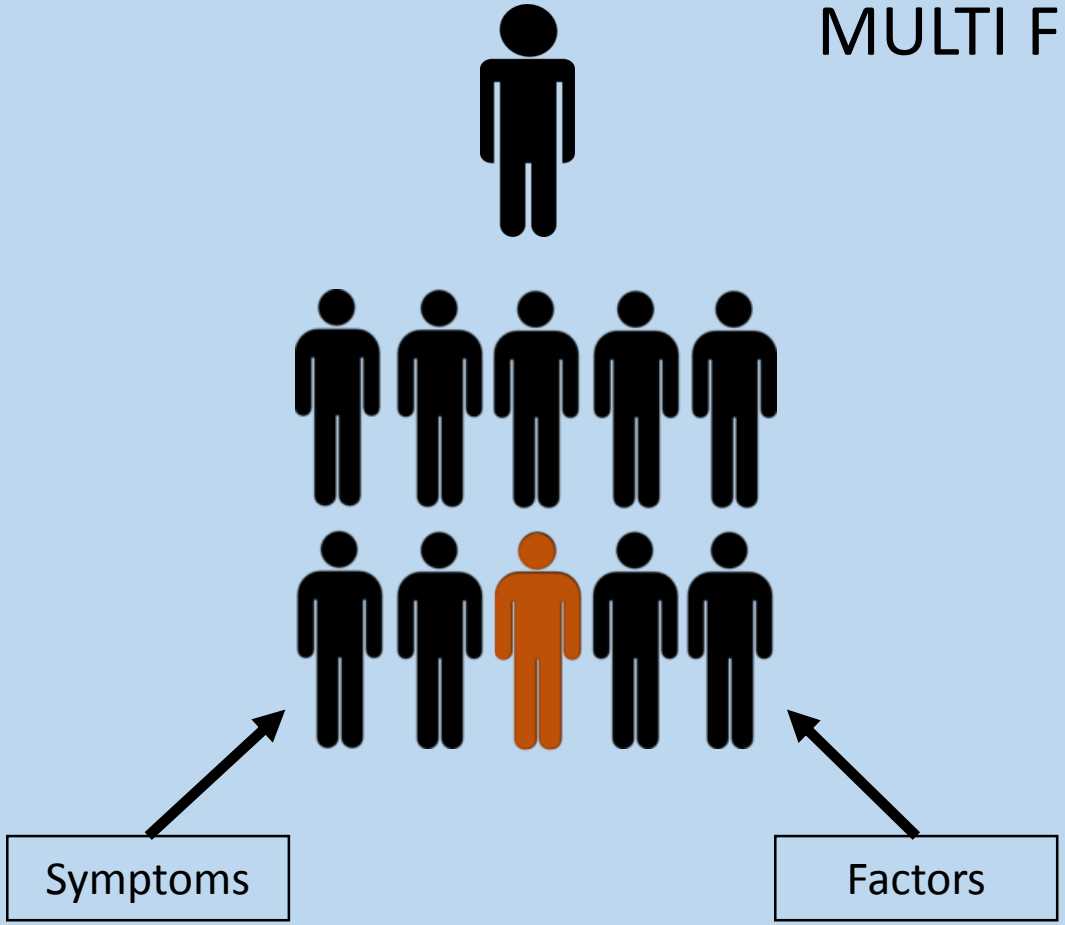
Datacenter Facility

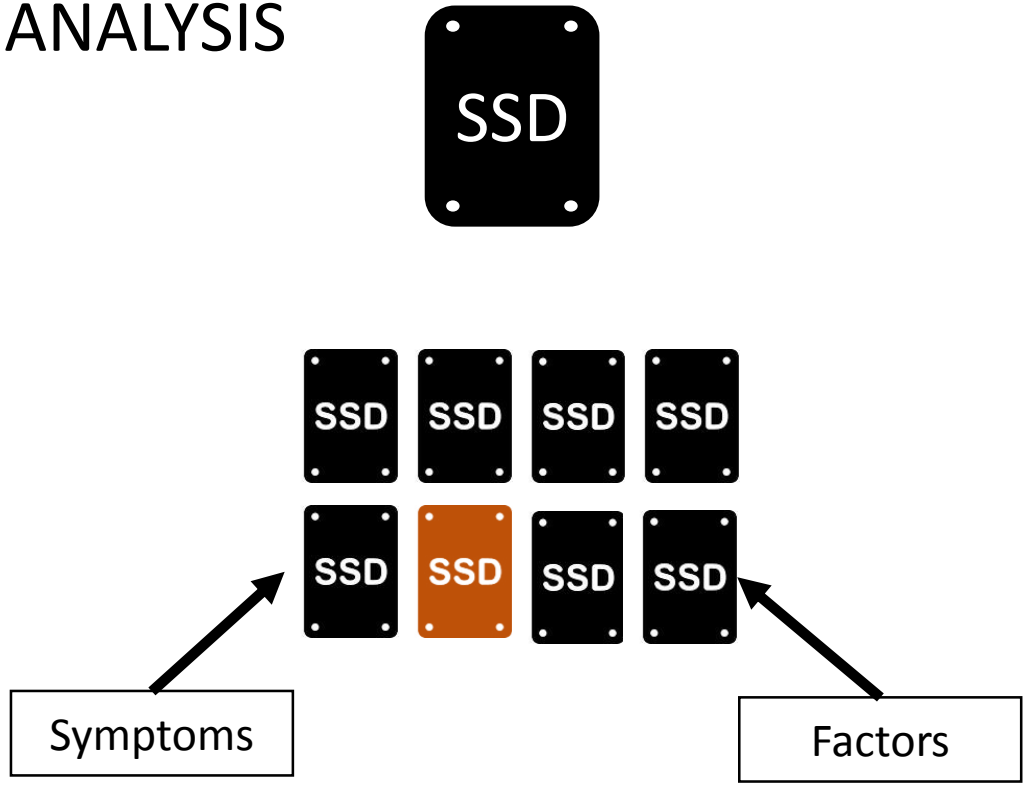


More results in the paper

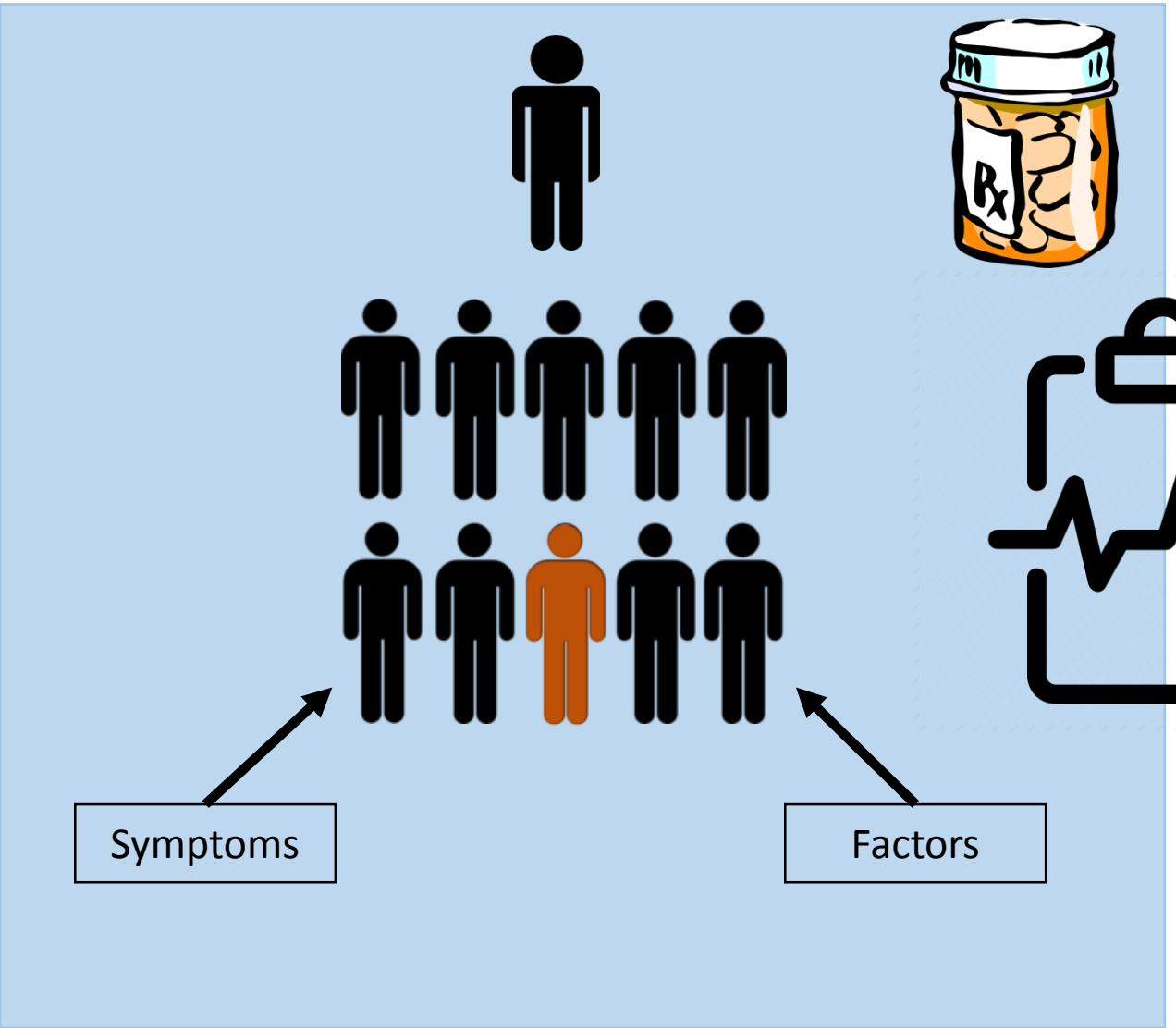
Understanding SSD Failures – An analogy

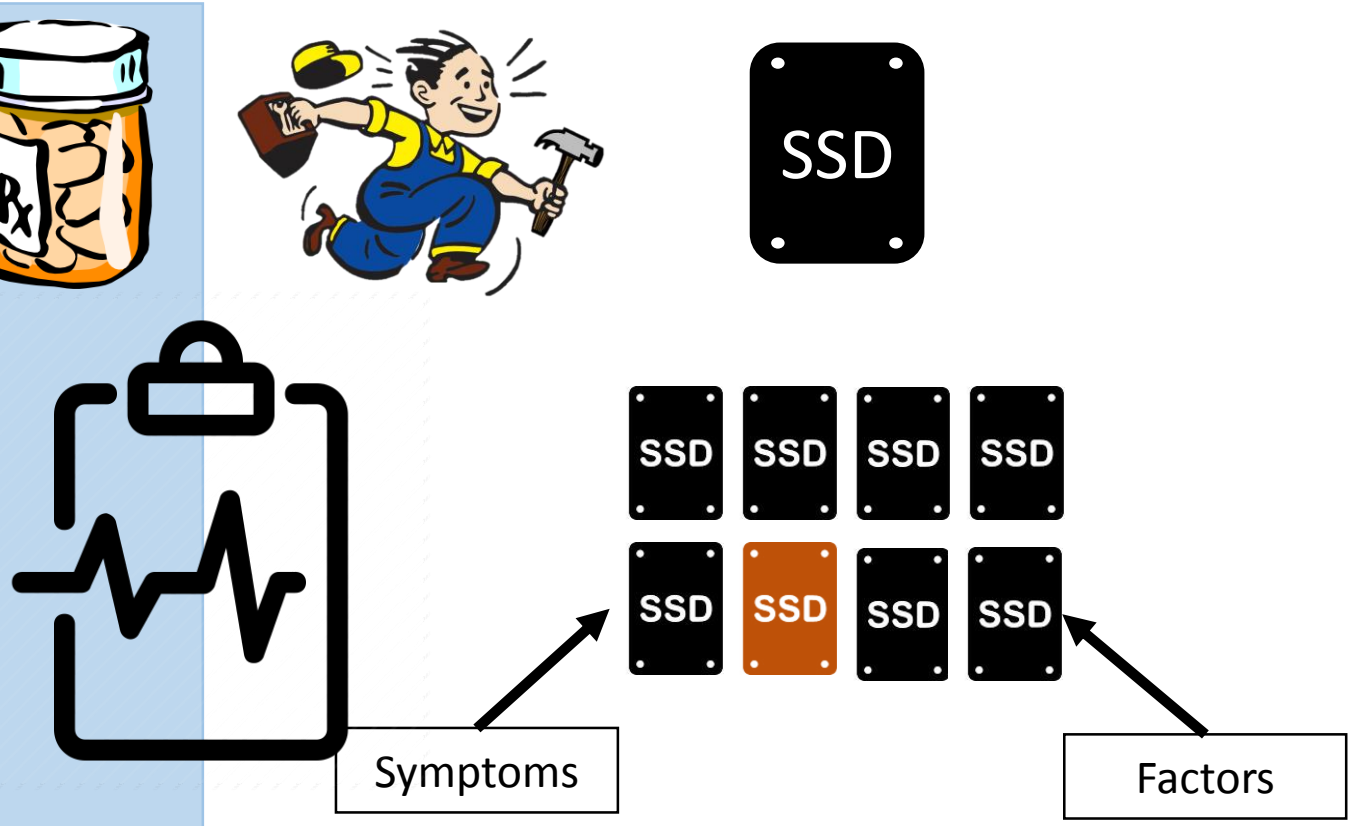
MULTI FEATURE ANALYSIS





Understanding SSD Failures – An analogy





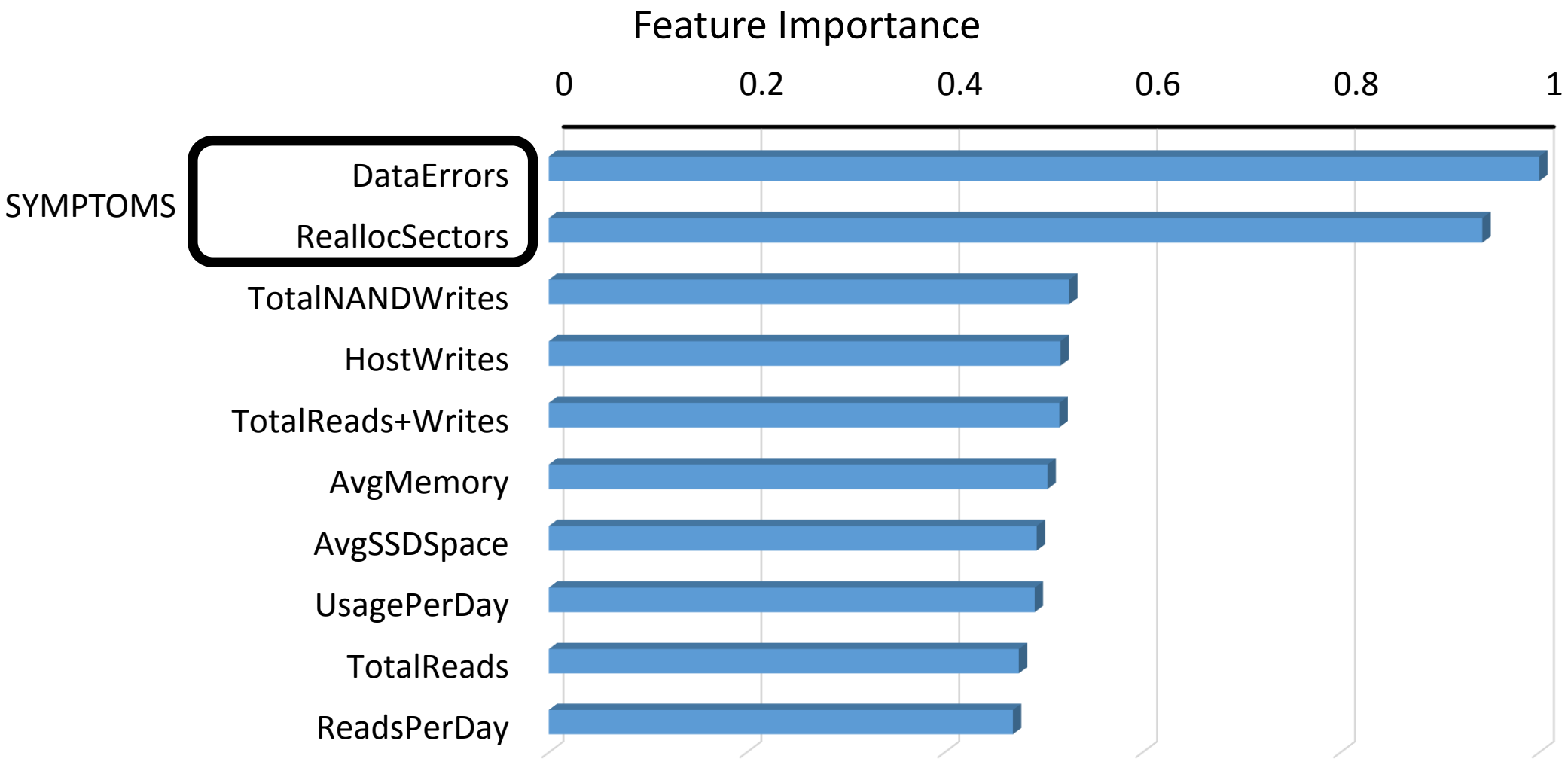
Random forest based binary classification
Permutation feature ranking

Understanding *What* ?

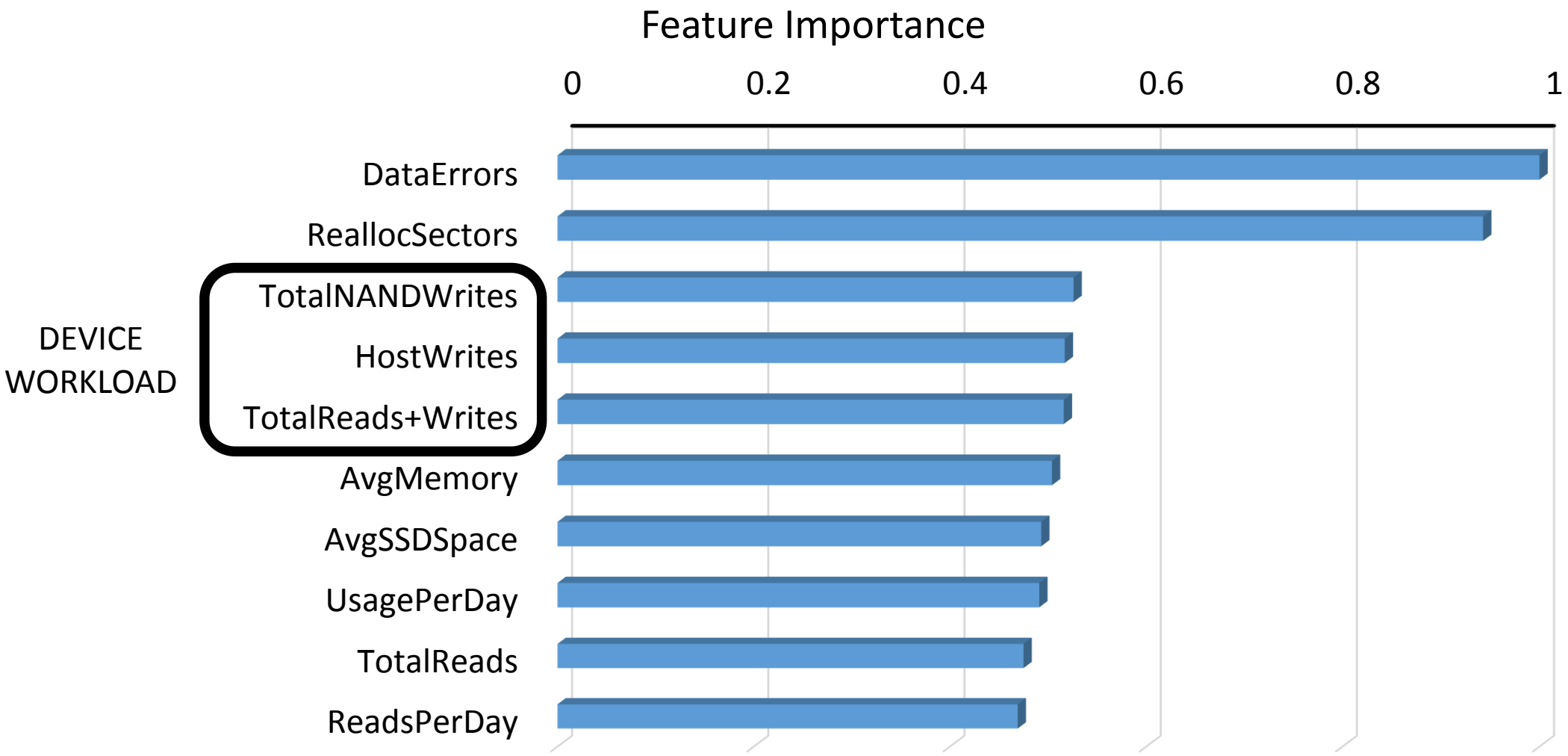
What

are the important factors ?
is their order of importance ?
are the important combinations?

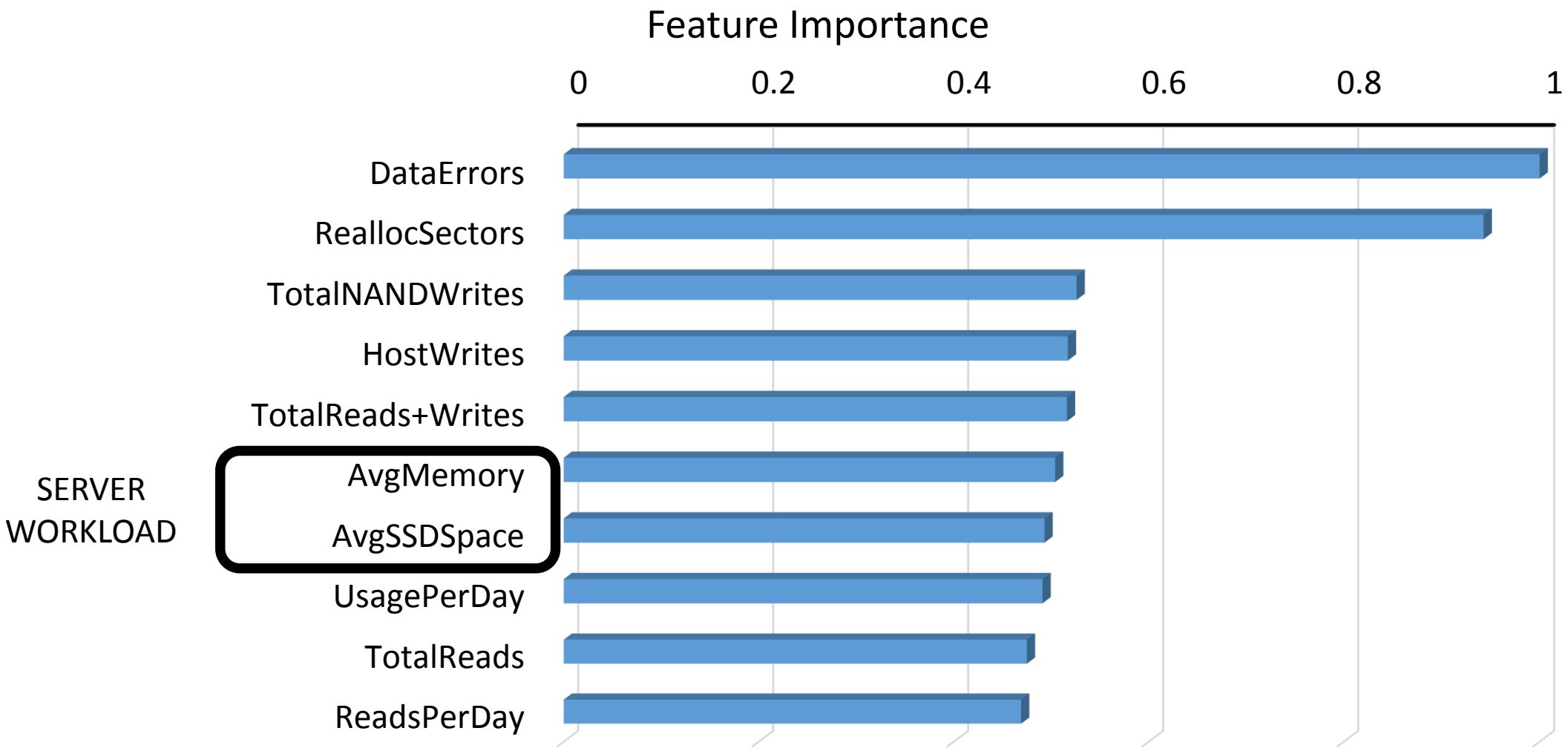
Understanding *What* ?



Understanding *What* ?



Understanding *What* ?



Understanding *What* ?

Combinations of top 8 important features

Frequent Combinations

SYMPTOMS

Condition	Class
Data Errors <=1 & Reallocated Sectors<=5	H
Data Errors<=1& WAF<=1	H
Media Wear-out=100 & WAF<=1	H
Avg. SSD space >=10	F

Understanding *What* ?

Combinations of top 8 important features

Frequent Combinations

SYMPTOMS + WORKLOAD

Condition	Class
Data Errors <=1 & Reallocated Sectors<=5	H
Data Errors<=1& WAF<=1	H
Media Wear-out=100 & WAF<=1	H
Avg. SSD space >=10	F

Understanding *What* ?

Combinations of top 8 important features

Frequent Combinations

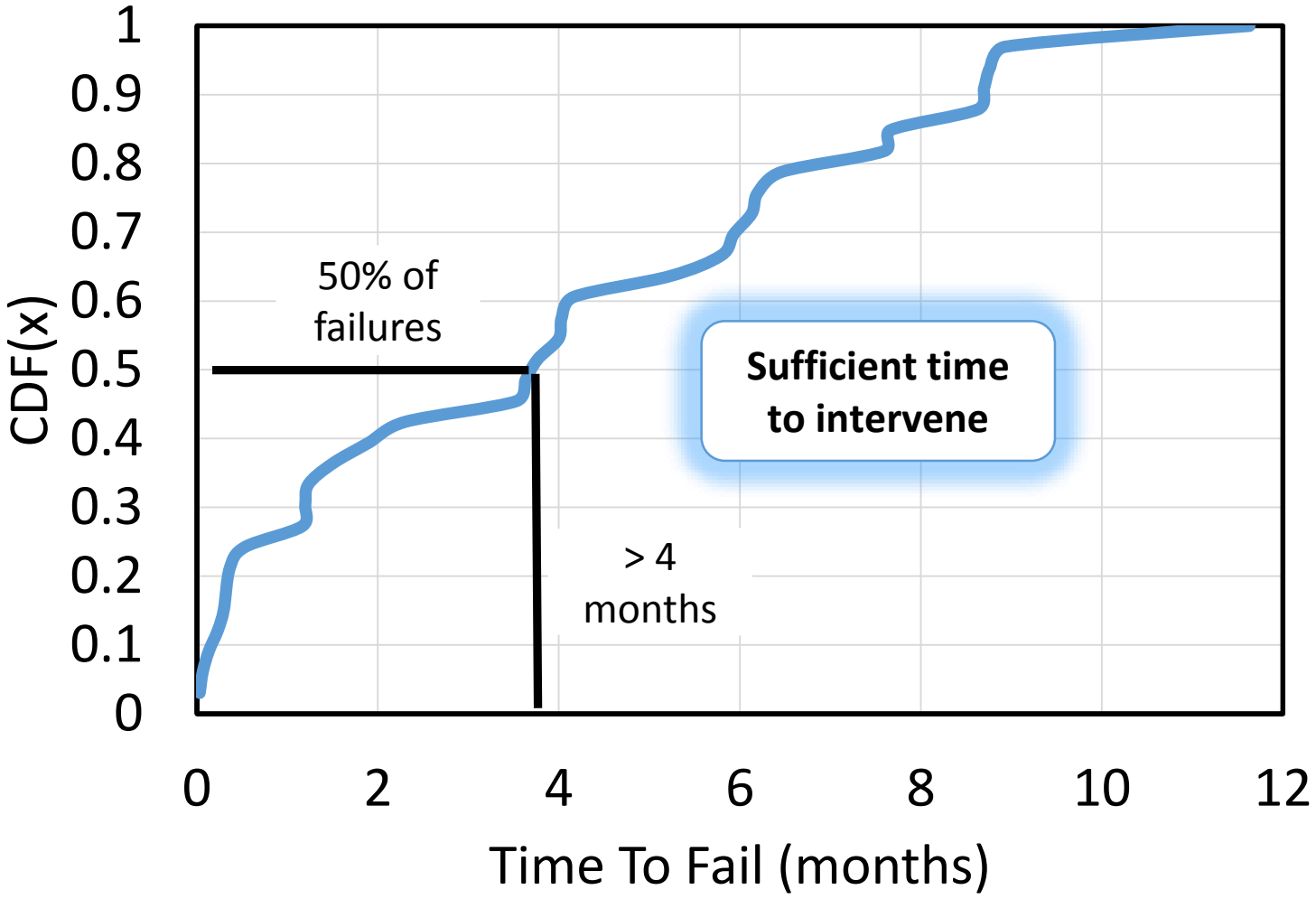
Condition	Class
Data Errors <=1 & Reallocated Sectors<=5	H
Data Errors<=1& WAF<=1	H
Media Wear-out=100 & WAF<=1	H
Avg. SSD space >=10	F

WORKLOAD

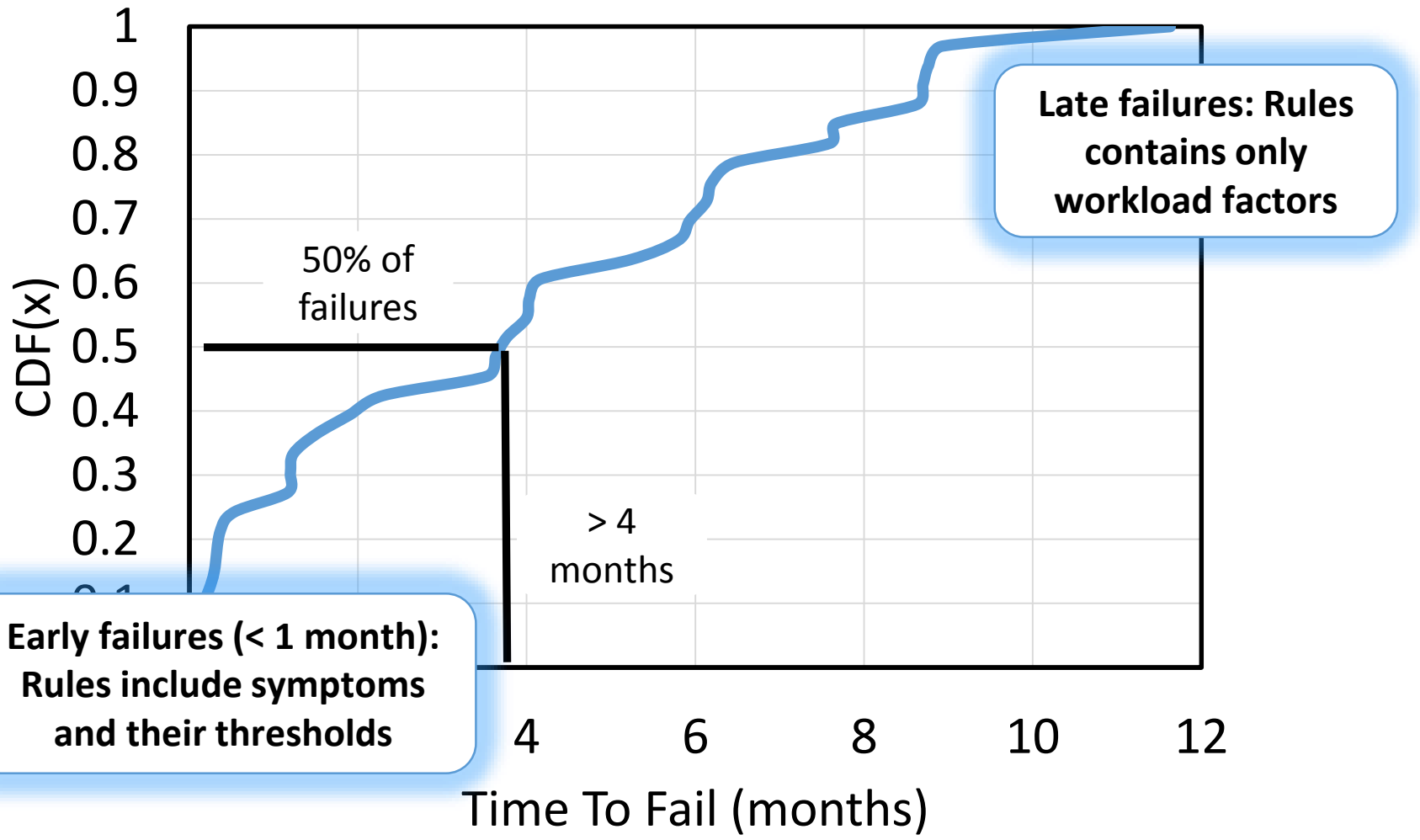
Understanding *When* ?

What is the duration between detection and failure?
signatures characterize SSD survivability?

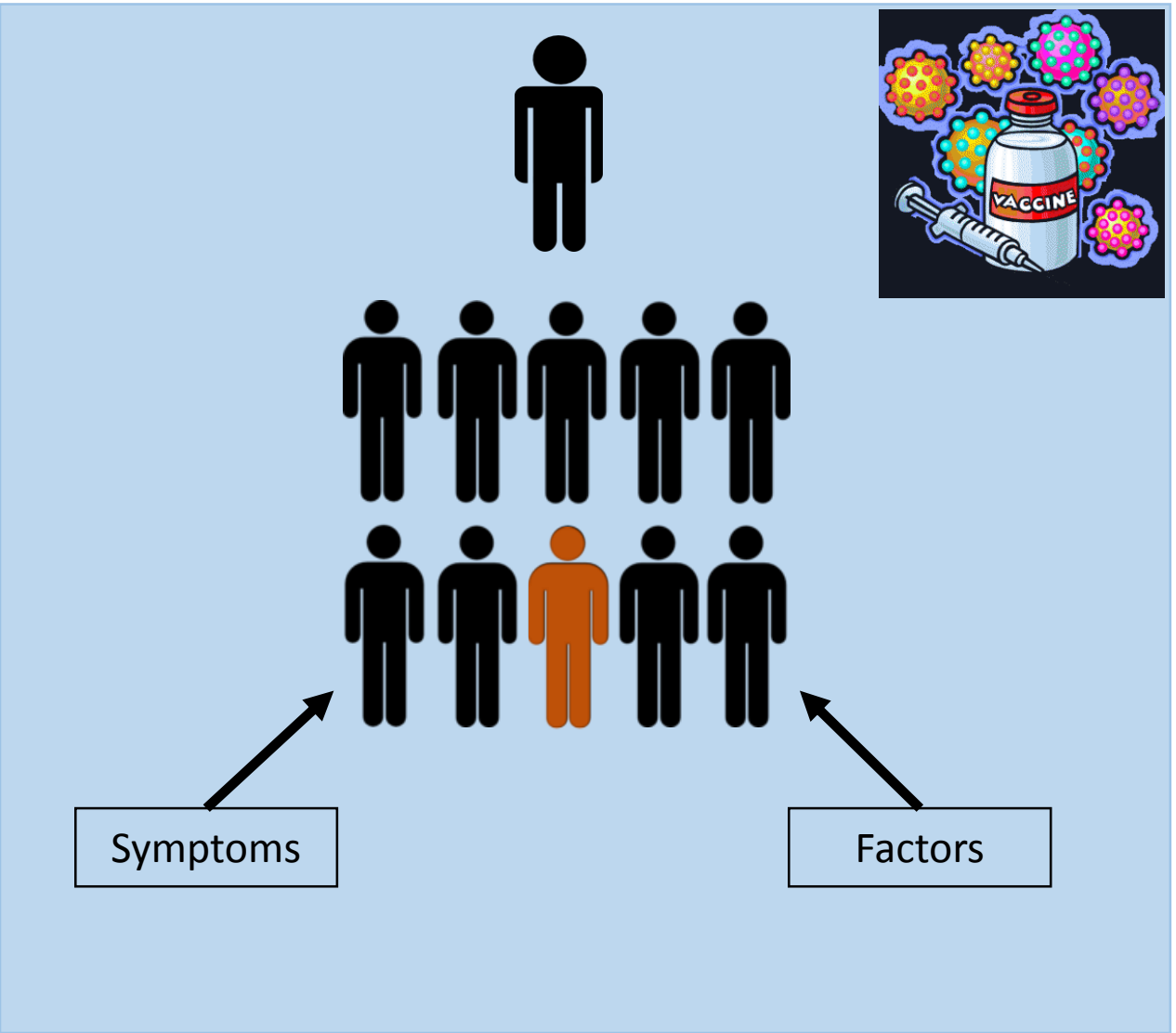
Understanding *When* ?

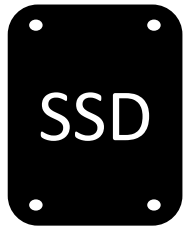


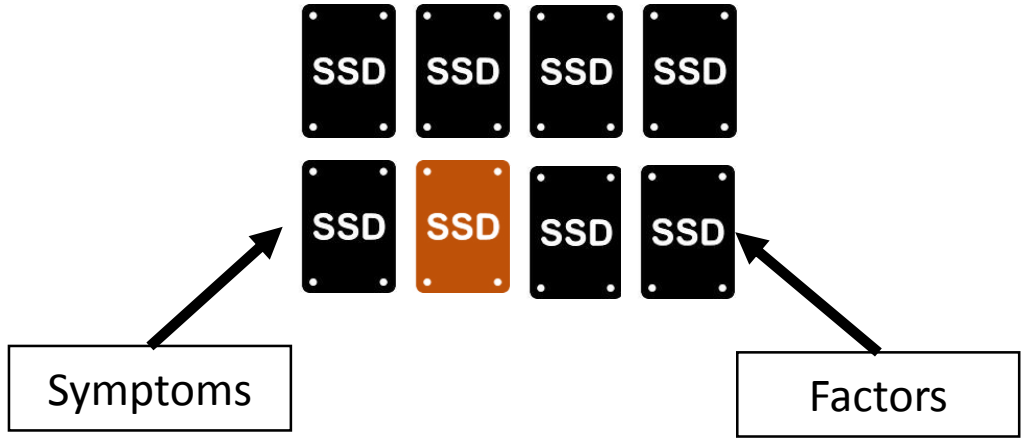
Understanding *When* ?



Understanding SSD Failures – An analogy





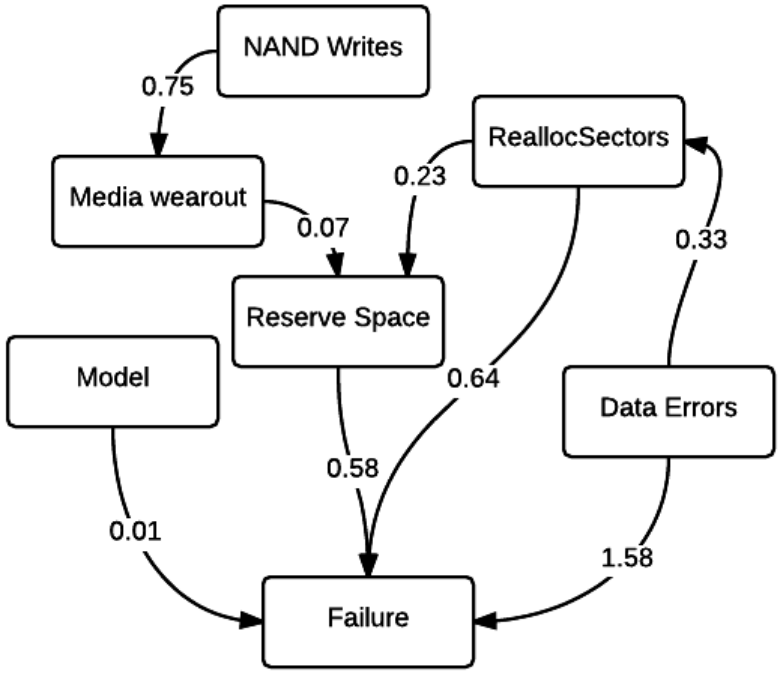


Observation based causal estimate
Probabilistic causal models and Pearl's do-calculus

Understanding *Why* ?

What factors impact SSD reliability?
is their magnitude of impact?

Understanding *Why* ?



SSD model and symptoms have direct impact

Workload impacts failures through media wearout

Concluding Remarks

- SSD Failures in the field
- Factors -> Symptoms -> Failures
- Important Symptoms: Data Errors and Reallocated Sectors
  - High intensity and rapid progression fails early
- Important factors: NAND Writes, Total Reads and Writes, etc.
- Direct impact: SSD Model and Symptoms
- Indirect impact: Workload through wear-out
- Future direction: prediction and control