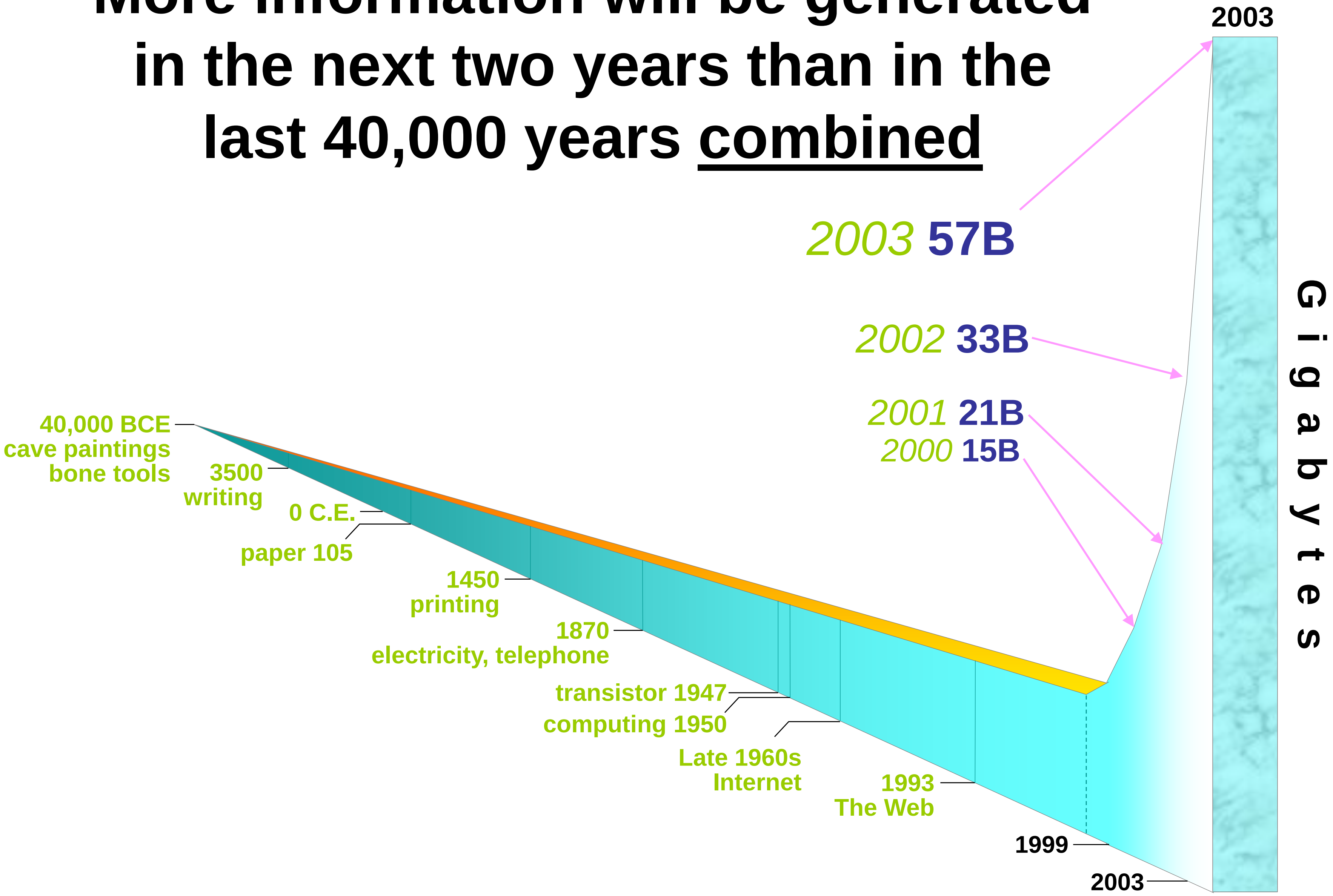
The Power of the Bit: Infinity

Moshe Yanai

IBM Fellow

2010 IEEE Reynold B. Johnson Award

More information will be generated in the next two years than in the last 40,000 years <u>combined</u>



Source: UC Berkeley, School of Information Management and Systems.

Data Storage Growth

- **2006:** 161 Exabytes
- **2010:** 988 Exabytes
- **Every day:** 15 Petabytes of new information generated

…that's 8x more than the information in all U.S. libraries!

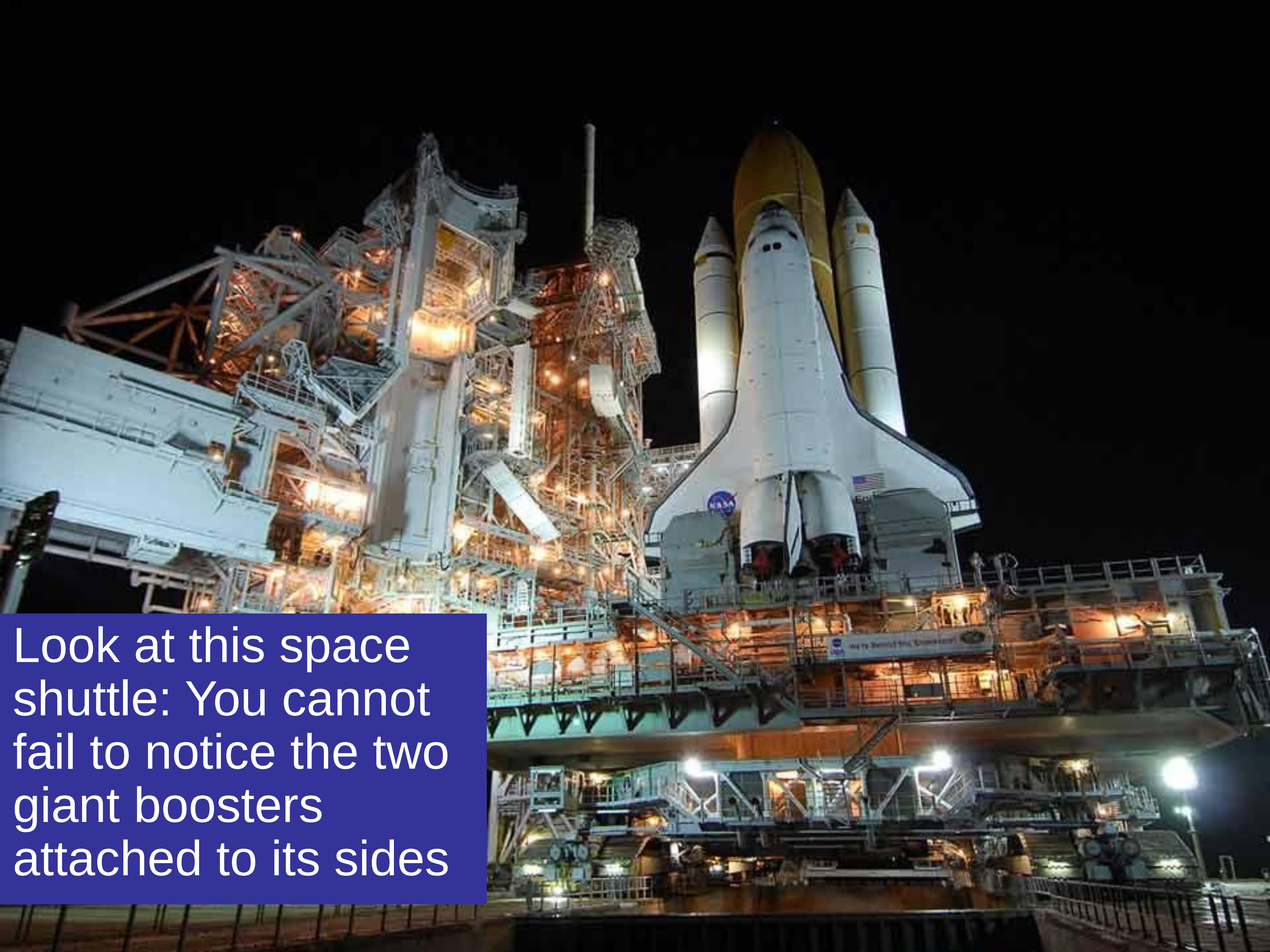
Data Bit has no size…

Here is a brief story about how size limitation impacts technologies other than data storage:

Limitations of the physical world:

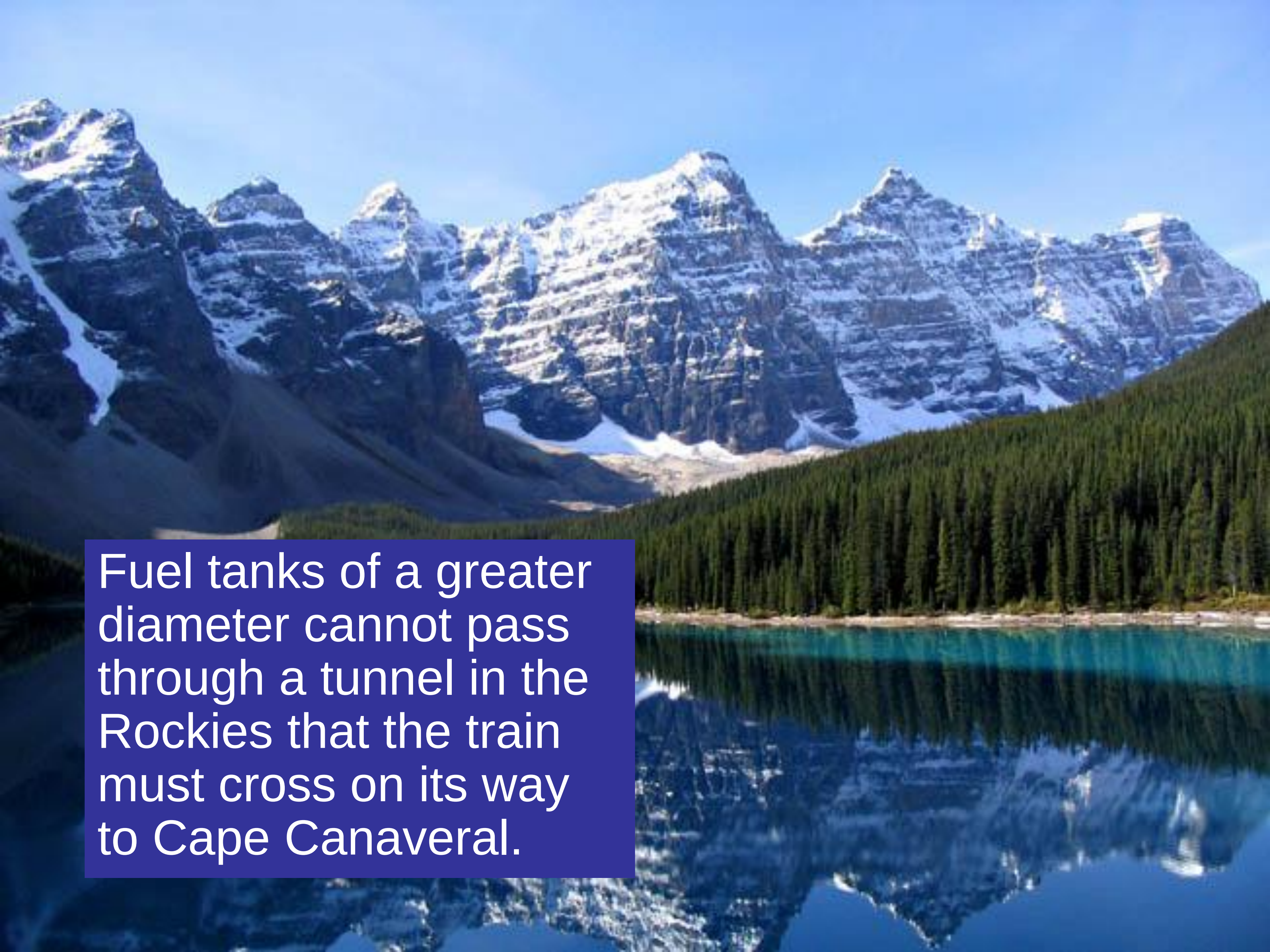
A Horse’s Bottom

You may have heard this, but I want to show you how some of today’s most innovative technological advances are deeply entrenched in history and the past



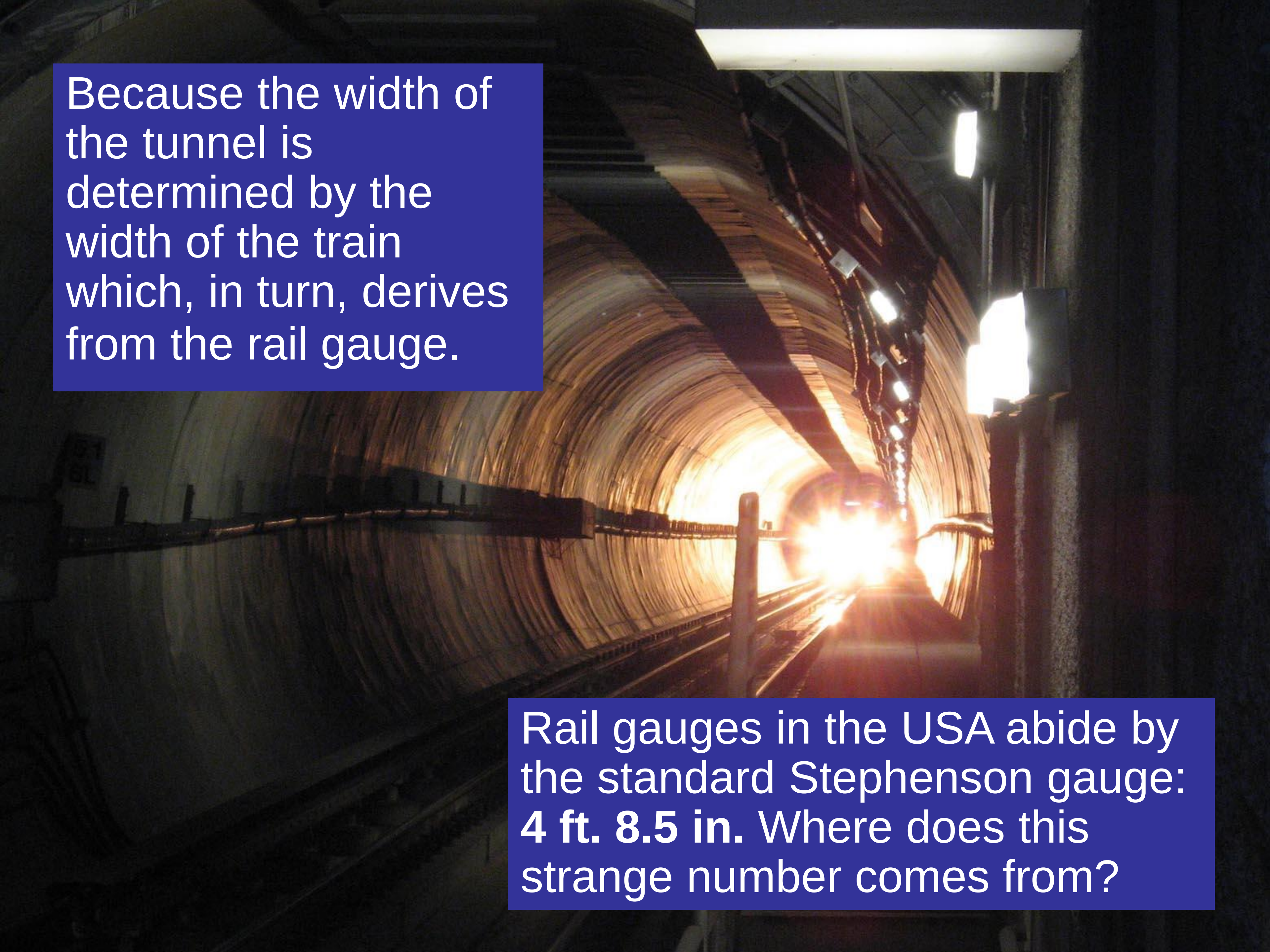
Look at this space shuttle: You cannot fail to notice the two giant boosters attached to its sides

The boosters are manufactured by Thiokol Corp., of Brigham City, Utah. The engineers of this company would rather design fuel tanks of a much larger diameter, but they need to transport them by train from their location to the launching site, and…



Fuel tanks of a greater diameter cannot pass through a tunnel in the Rockies that the train must cross on its way to Cape Canaveral.

But why was the tunnel built this wide, and not much wider to begin with?

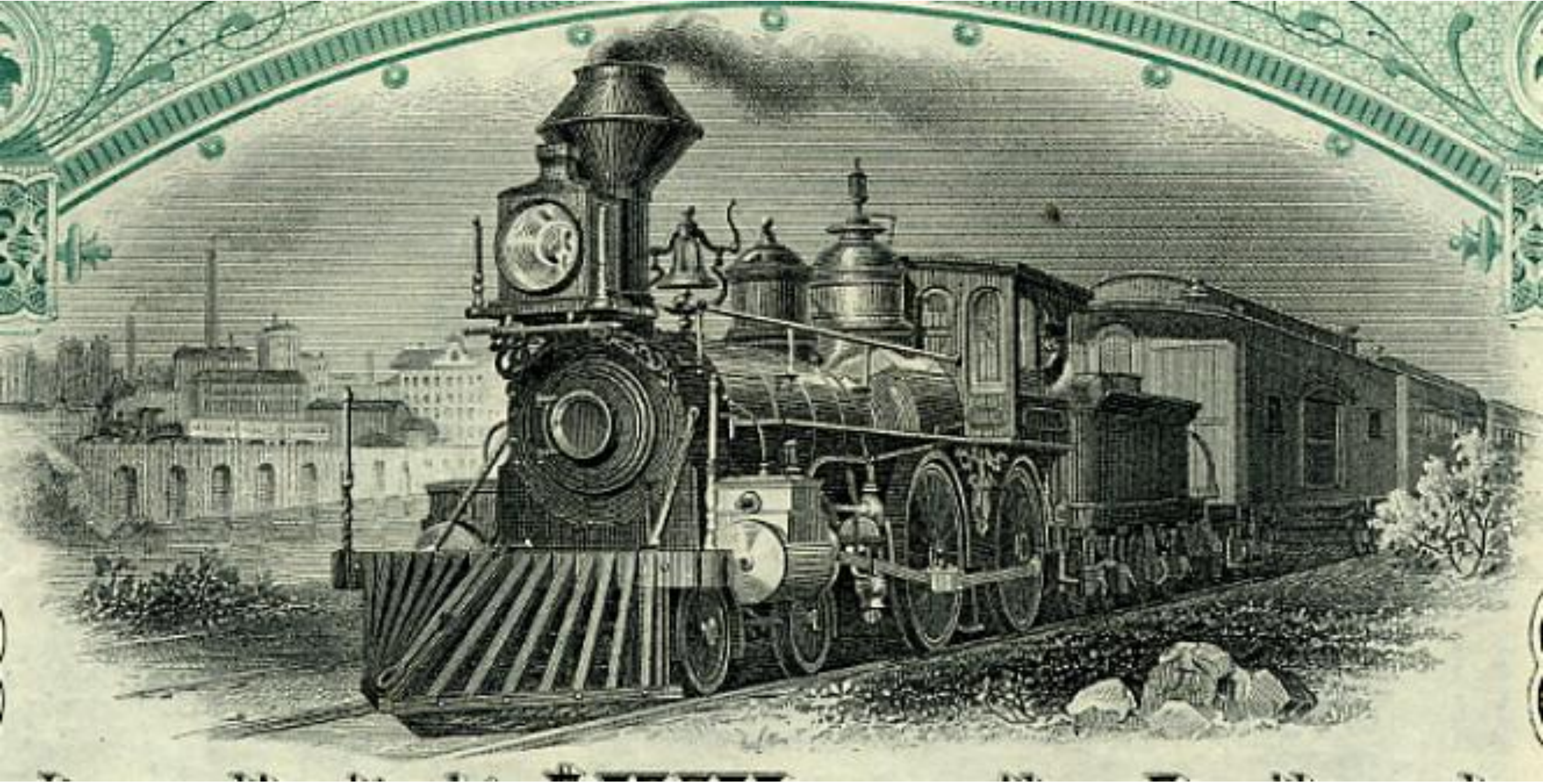


Because the width of the tunnel is determined by the width of the train which, in turn, derives from the rail gauge.

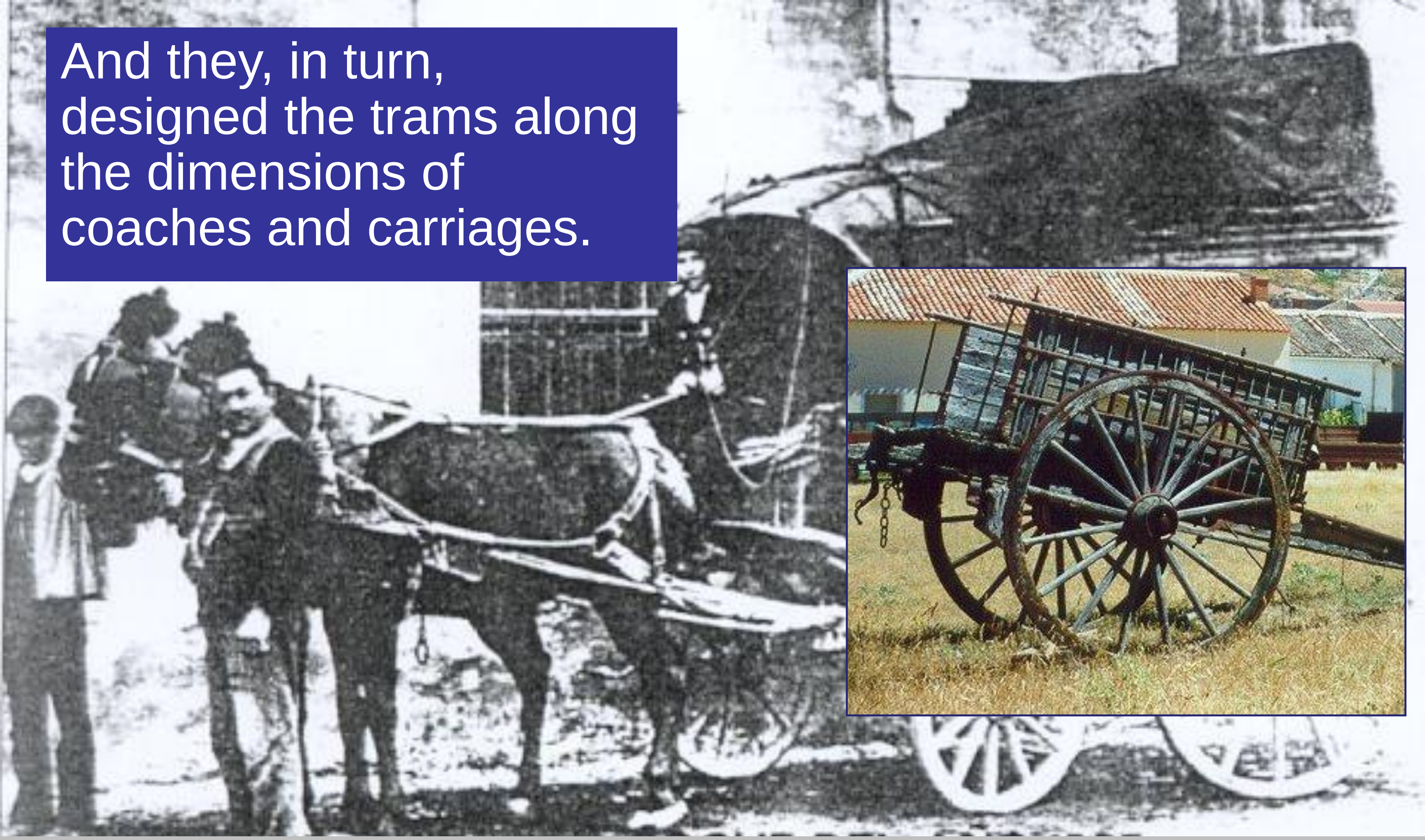
Rail gauges in the USA abide by the standard Stephenson gauge: **4 ft. 8.5 in.** Where does this strange number comes from?

This rail gauge was used in Britain and later adopted in the USA because the early trains were purchased there.

Fair enough.
But why did the British choose this gauge?



Because trains were built by engineers who had built city trams that used this gauge.

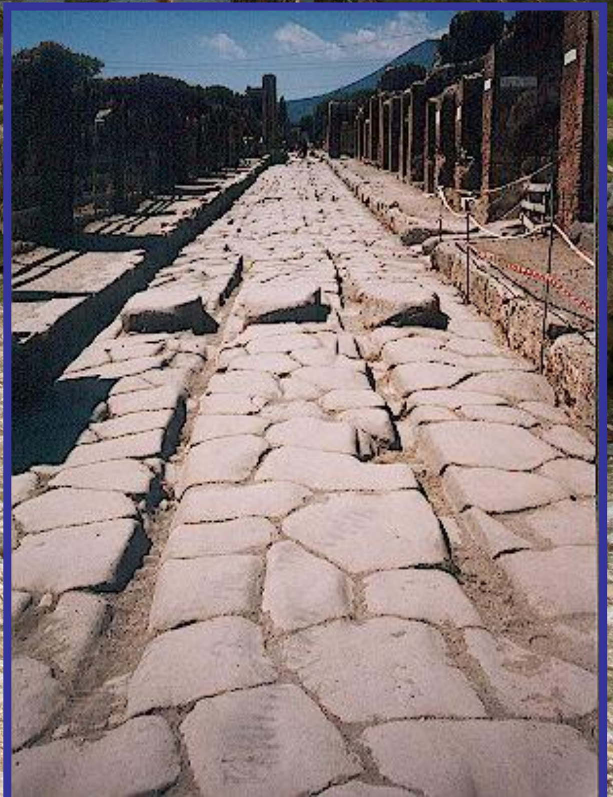


And they, in turn, designed the trams along the dimensions of coaches and carriages.

But why this size?

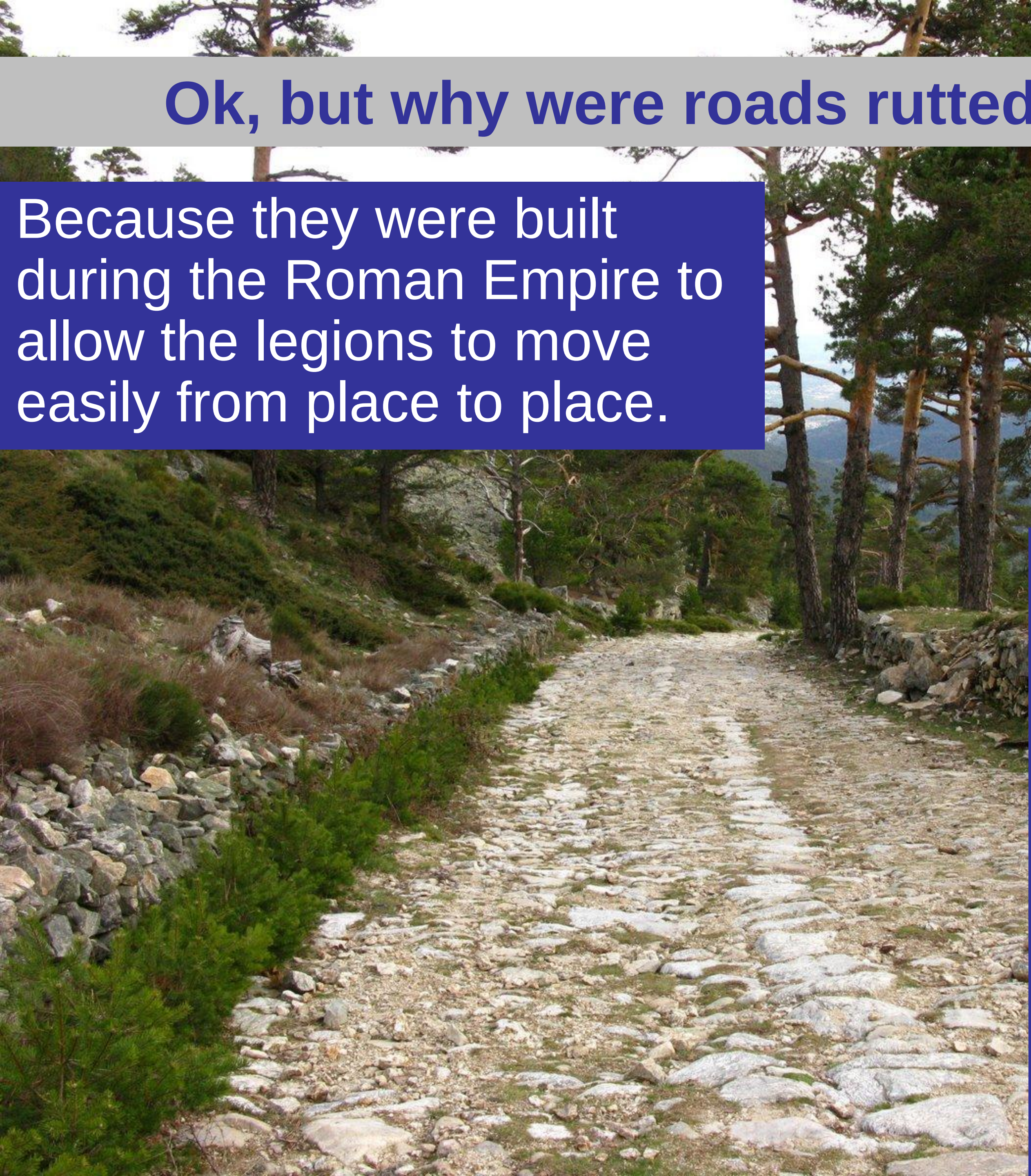
And why was this standard chosen for coaches?

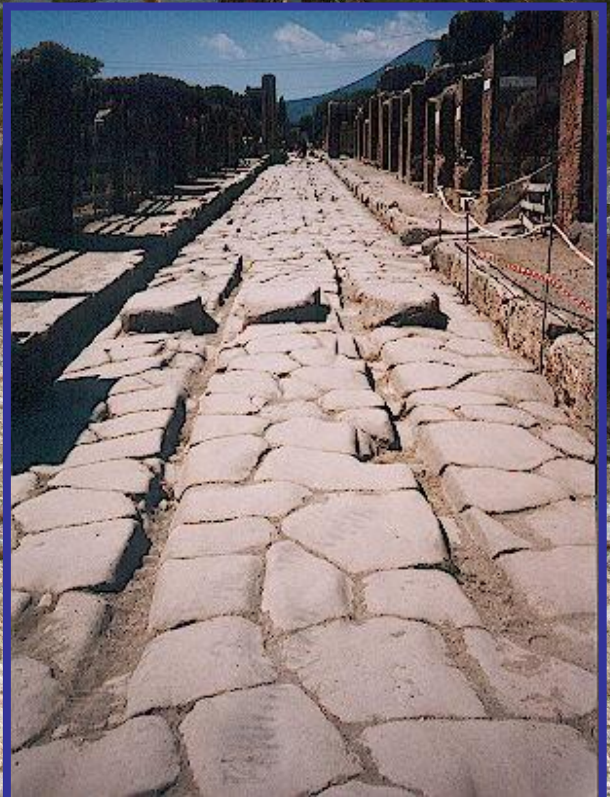
Because all over Europe roads were already rutted to this distance from earlier times, and placing the coach wheels at different widths would cause them to vibrate and eventually break down.



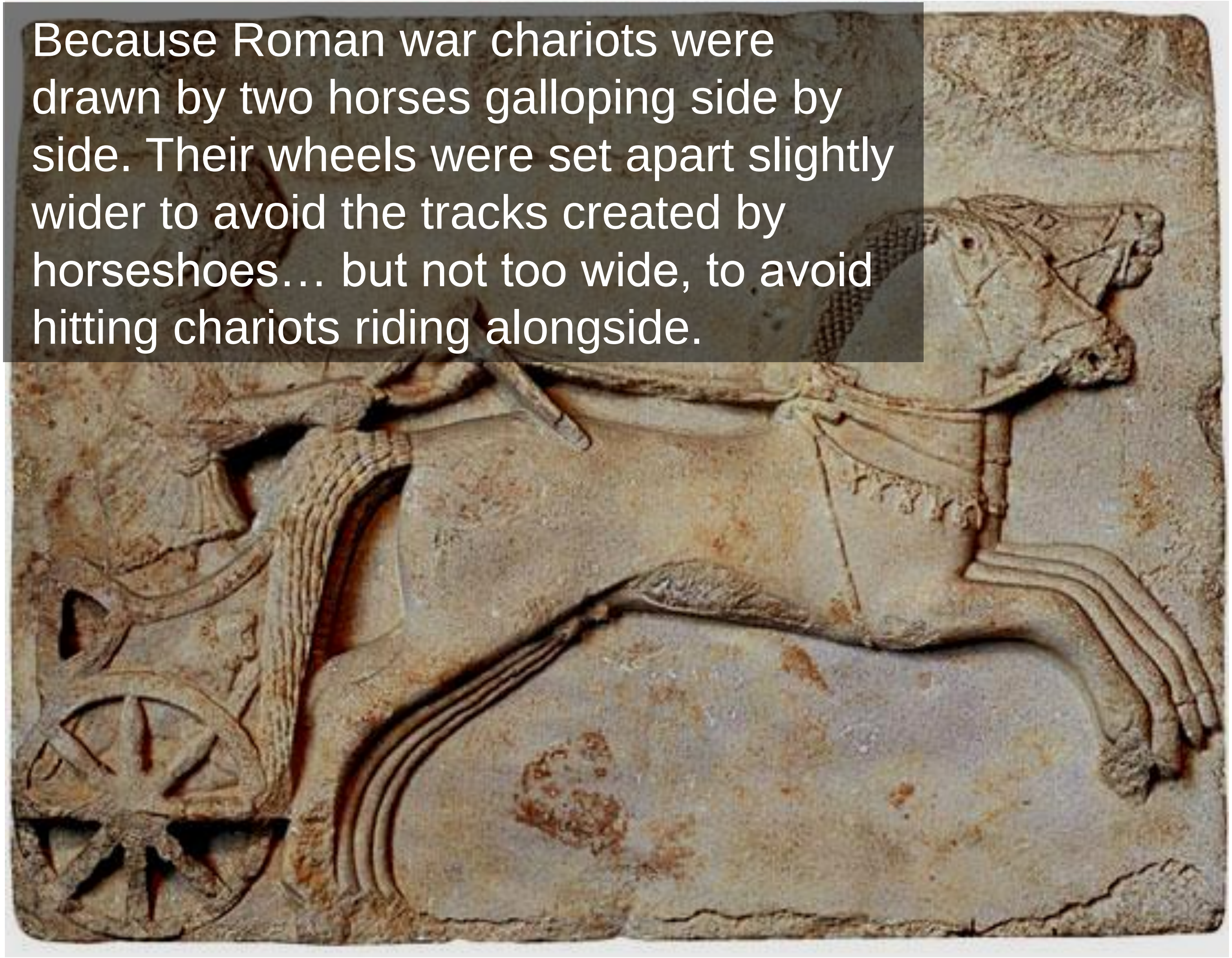
Ok, but why were roads rutted in this way?

Because they were built during the Roman Empire to allow the legions to move easily from place to place.





And why did the Romans choose this width?



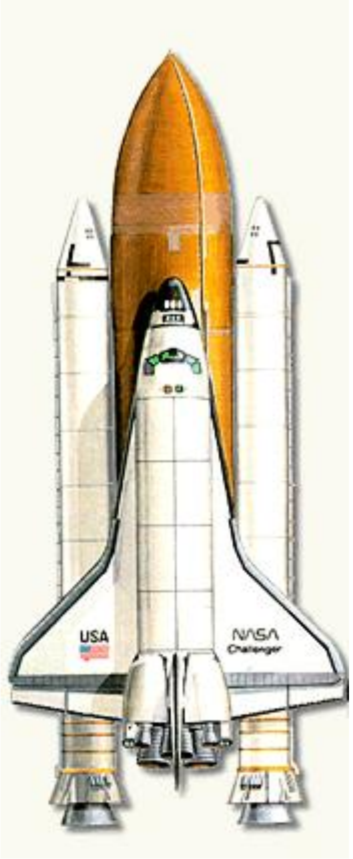
Because Roman war chariots were drawn by two horses galloping side by side. Their wheels were set apart slightly wider to avoid the tracks created by horseshoes… but not too wide, to avoid hitting chariots riding alongside.

We are now close to answering our original question:

The standard rail gauge in the USA (4 ft. 8.5 in.) was chosen because more than 2,200 years ago, on a distant continent, Roman war chariots were built to the dimensions of…

a horse's bottom!!

Summary



The design constraints of the space shuttle, the world’s most advanced transportation system, derive from…

the dimensions of a horse’s bottom!

Physical limitations:

A few more examples

1908
Ford T

2010 Audi
A5 Cabriolet

American Airlines
DC4, 1938

Boeing 787
New Dreamliner

Titanic - 1912

Newest Carnival
Cruise

On the other hand…

What has been happening with information?

Today:

- **Billions** of people are authoring information, which then flows from…
- **A trillion** intelligent devices, sensors, and all manner of instrumented objects

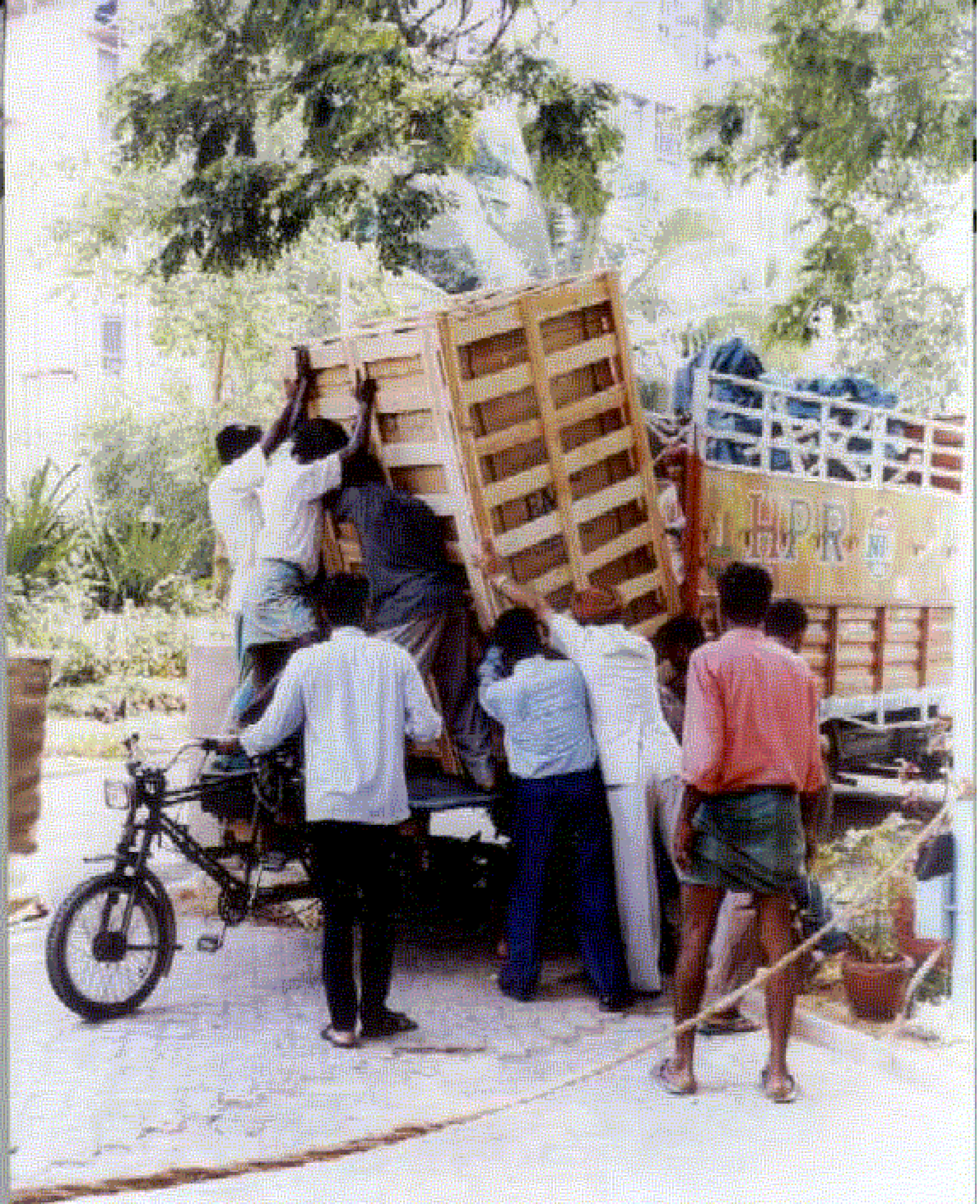
One billion camera phones were sold in 2007, up from 450 million in 2006. Annual growth rate of 3G devices: **30%**

Data buildup: Some examples

More than **147 million U.S. Internet users:**

- ❖ watched an average of 101 videos per viewer in January 2009
- ❖ downloaded **93,536** terabytes of data in one month
- ❖ **3.12 petabytes/**a day

By 2011, the Web will be used by an estimated 2 billion people… and connected to by a trillion objects: cars, appliances, cameras, roadways, pipelines… comprising the "Internet of Things."

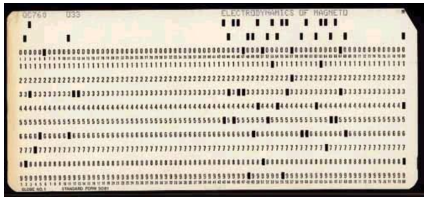


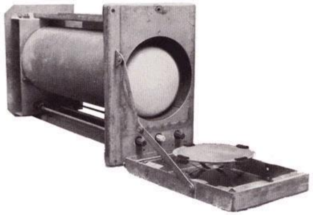
Data hopes for the Third World

Modern data storage devices transported by rickshaw in India

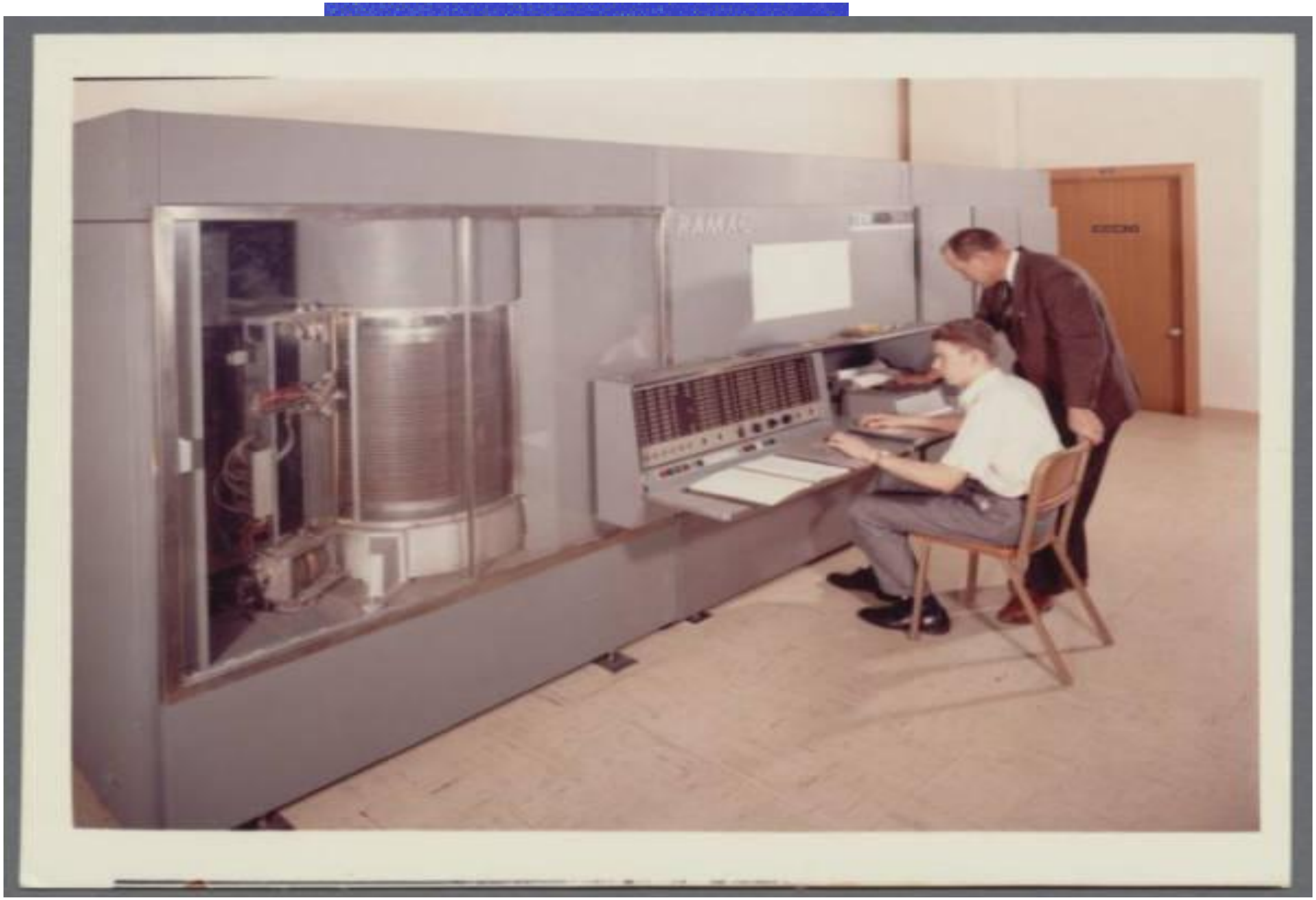
Evolution of Data Storage

- Writing Invention – 6000 years ago
- Printing invention (Gutenberg) – 600 years ago
- Paper Tape (1857)
- Punch Cards (1890)
- Vacuum Tubes (1940’s)
- Acoustic Delay Line (1944)
- William’s Tube - Cathode Ray Tube (1946)
- Magnetic Drums (1947)
- Magnetic Cores (1948)
- Magnetic Tape (1952)
- Magnetic Disk (1956)

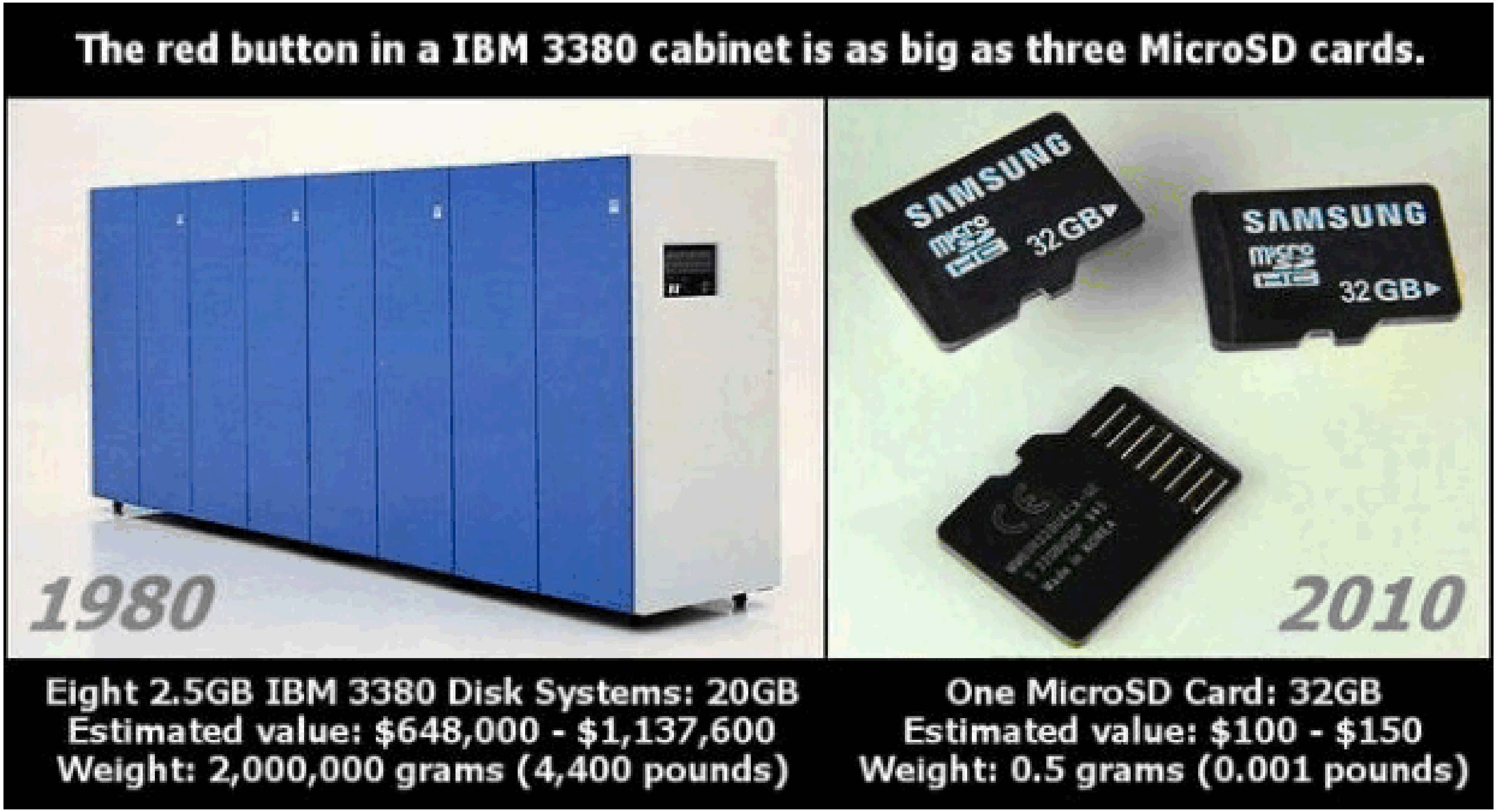




Evolution of Storage Media



IBM 3380 Disks



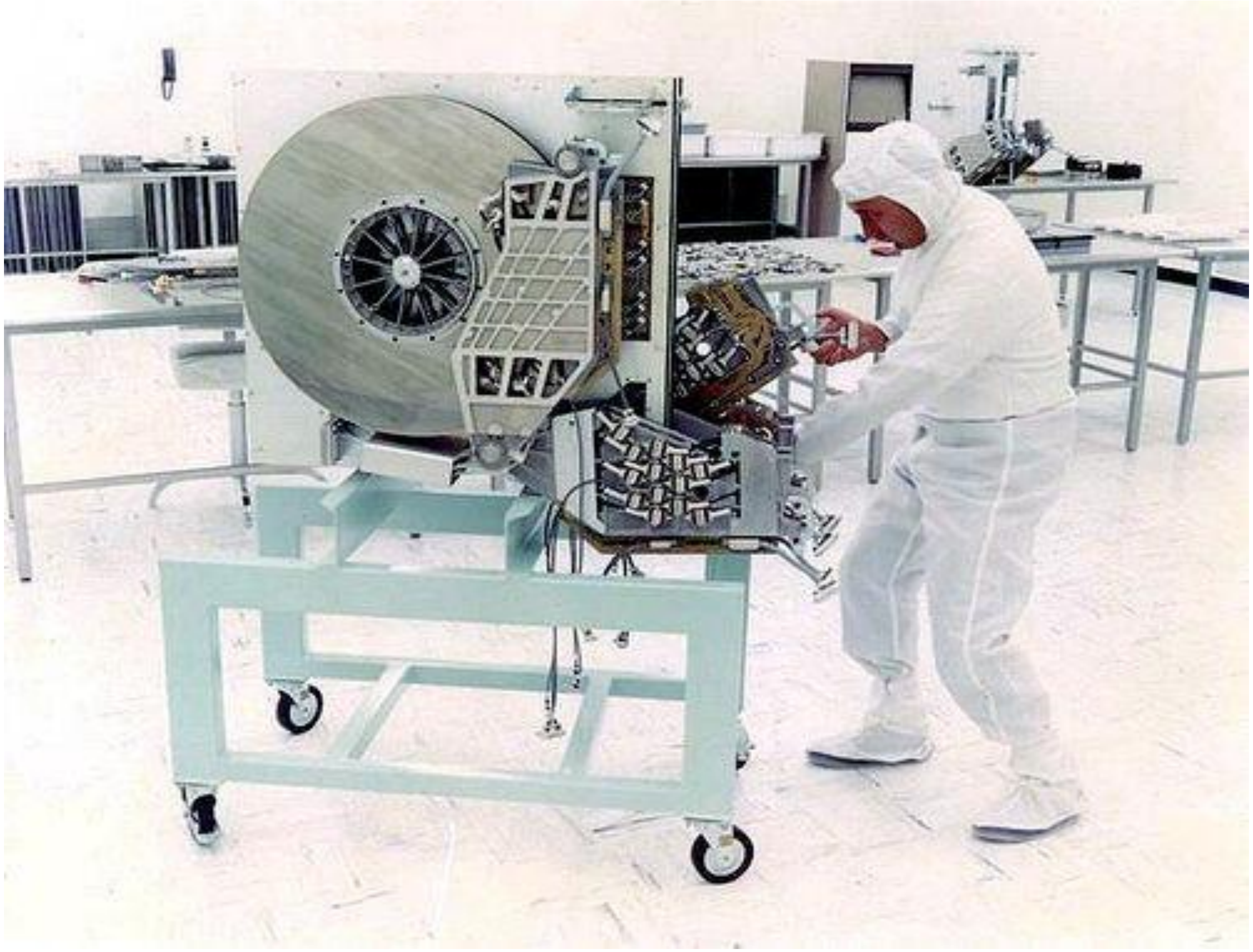
The red button in a IBM 3380 cabinet is as big as three MicroSD cards.

Eight 2.5GB IBM 3380 Disk Systems: 20GB
Estimated value: $648,000 - $1,137,600
Weight: 2,000,000 grams (4,400 pounds)

One MicroSD Card: 32GB
Estimated value: $100 - $150
Weight: 0.5 grams (0.001 pounds)

Source: http://www.crunchgear.com/2010/03/20/before-and-after/

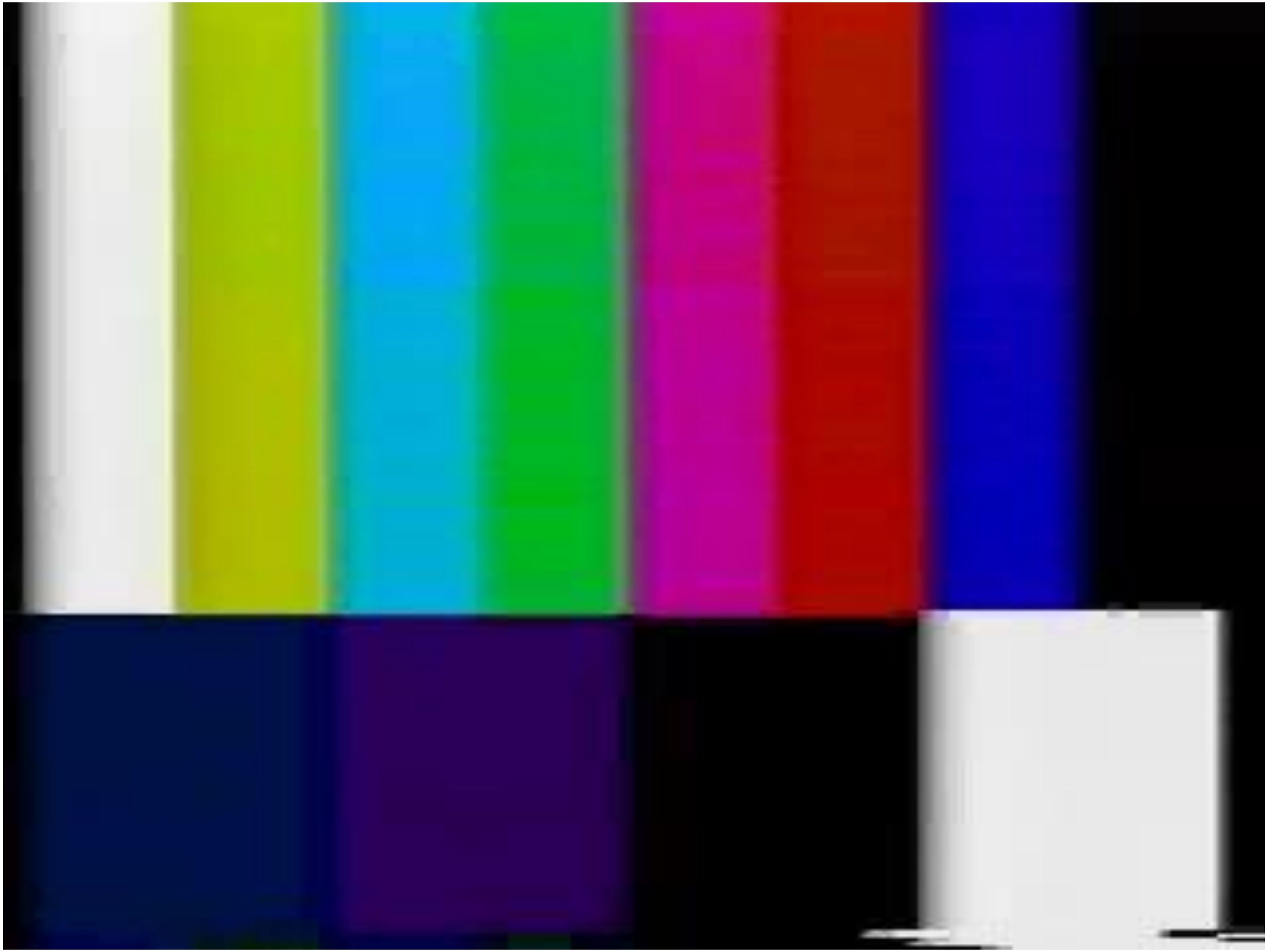
Hard Drives in 1975

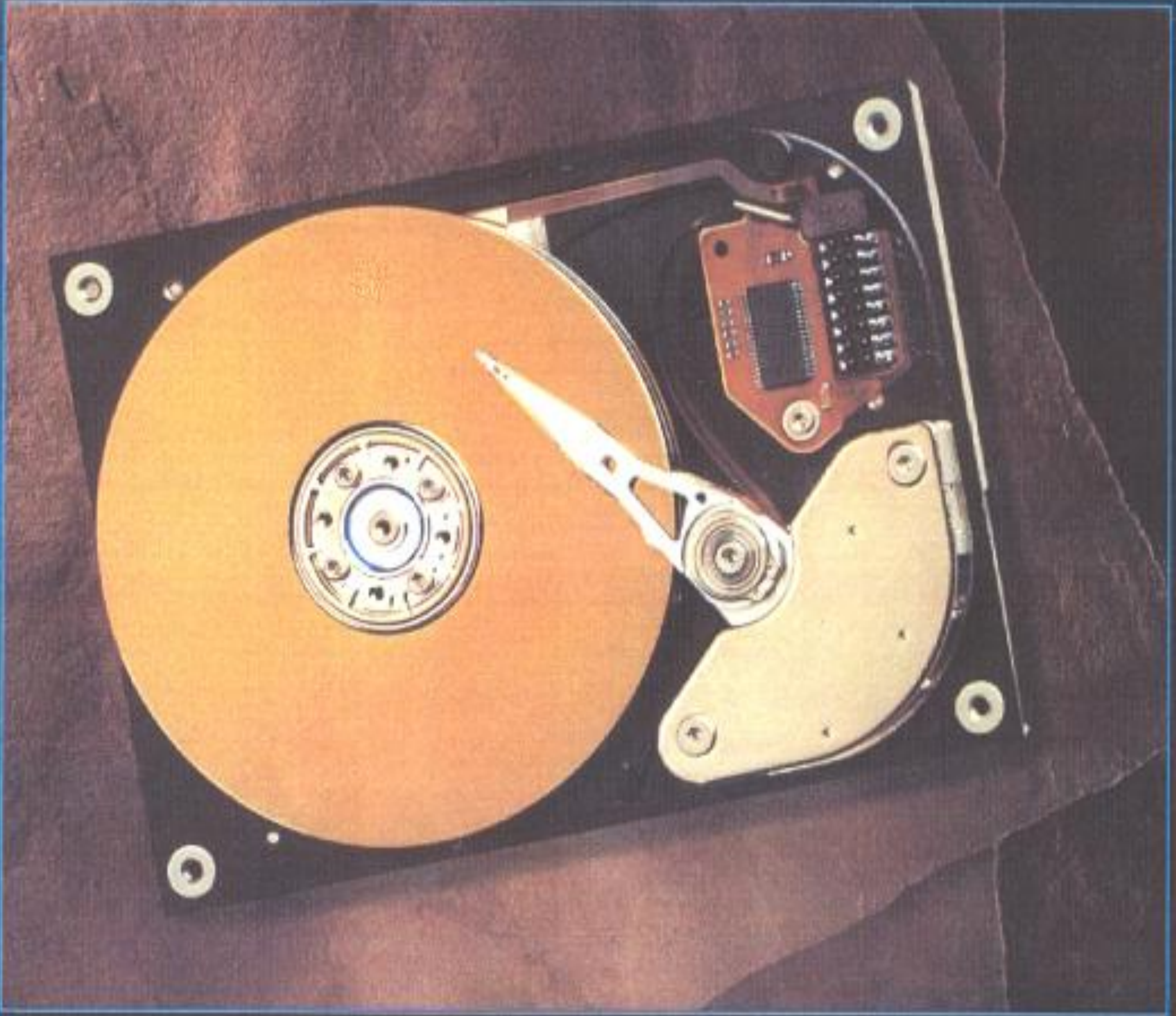


Size Shift: 1975-1992

Copyright © 2010 - Moshe Yanai

RAMAC Introduction Movie 1956





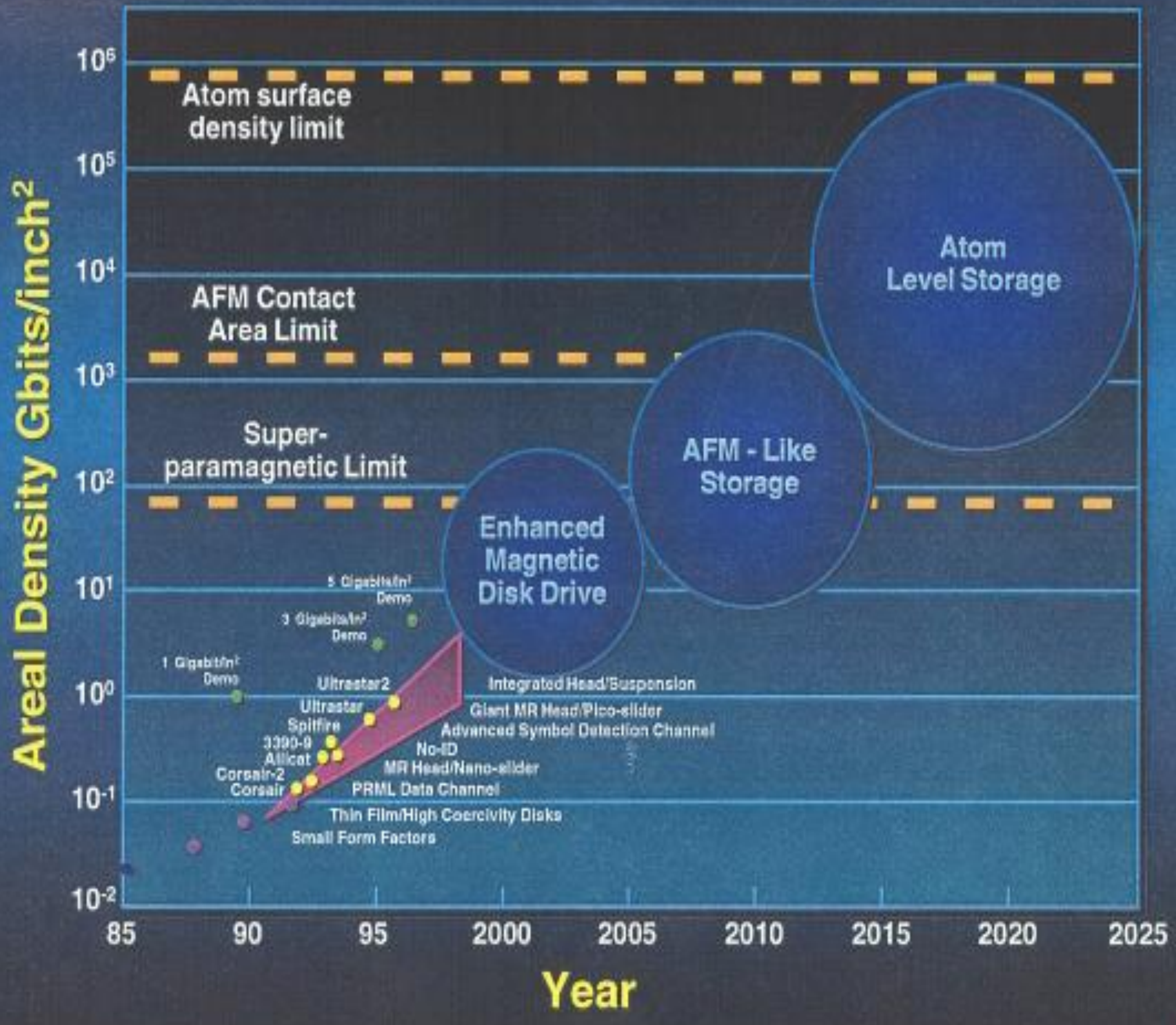
IBM

Seagate

Modern Data Storage

- Magnetic Tape. In 1952, IBM pioneered the magnetic tape, realizing that both punch cards and ticker tape were far too slow.
  - In this year alone, they created the first ever "clean room" to manufacture the tape, and NRZI (Non-Return-to-Zero Inverted) encoding for data storage.
- Magnetic Disk. In 1956, a small team of IBM engineers in San Jose introduced the first computer disk storage system.
  - The 305 RAMAC could store five megabytes of data on 50 disks, each 24 inches in diameter. RAMAC's revolutionary recording head could go directly to any location on a disk surface without reading all the information in between.

Advanced Storage Roadmap



IBM
Advanced
Technology

A. Mobile - 12/10/96

There's Plenty of Room at the Bottom!

Assuming 1 Year Production of 100 Million Drives at 1 Gigabyte Each, Number of Bytes to Be Stored

$$=10^8 \text{ x } 10^9 \text{ bytes} = 10^{17} \text{ bytes}$$

- The world production of storage at the present time fits in a volume of 1mm^3!
- 1000 years of production of storage products at the current rate fit in a volume of 1cm^3!

IBM

Seagate

Data bits have no physical limitations…

How much data does mankind store?

- IDC said about 161 exabytes in 2006
- In 2010, we'll reach 988 exabytes
- That's 600% growth in 4 years

161,000 PB

We must provide a simple solution for the storage needs of the modern enterprise

988,000 PB

5 Key Attributes for Enterprise Storage Solutions

- **Reliability –** Business data is more critical than ever, with no tolerance for downtime for most applications. Requirements are now greater than 5 nines – overcoming the human factor!
- **Performance –** Consistent performance under all conditions eliminating hot spots and staying consistent during rebuilds after hardware failures
- **Functionality –** Tier 1 functions (e.g. replication, thin provisioning) that scale with no performance penalty and are inherently built-in to the architecture
- **Manageability –** Total system virtualization with emphasis on ease of use
- **Cost –** Reasonable cost so business can concentrate on its core business rather than IT

All these key attributes -- with unlimited scalability

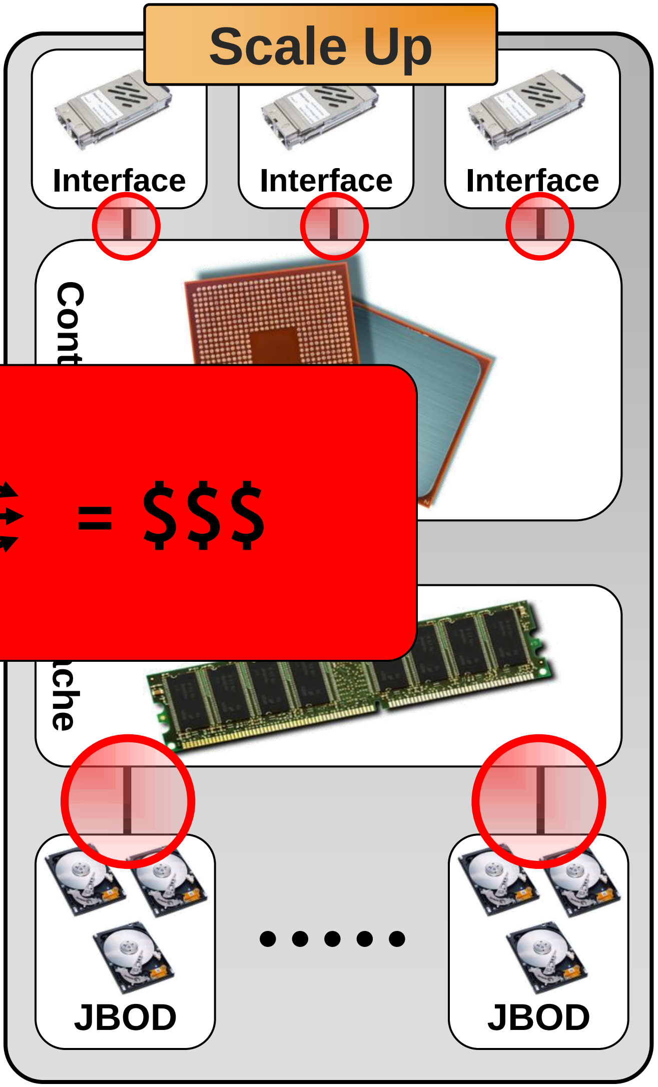
Current Enterprise Storage Solutions

Building blocks:

- Disks
- Cache
- Controllers
- Interf
- Interc

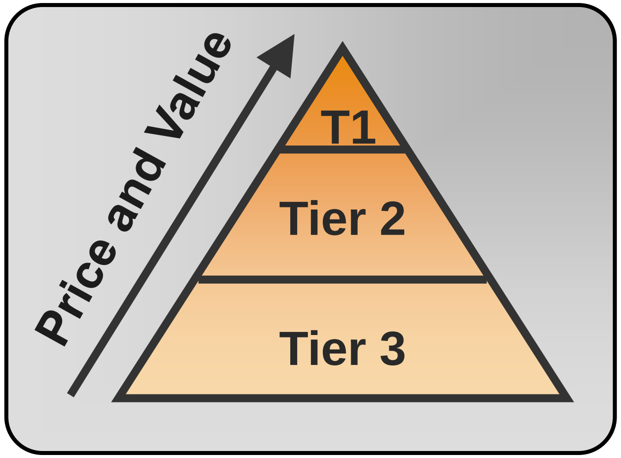
PERFORMANCE
RELIABILITY
SCALABILITY
→ = $$$

With this current architecture, scalability is achieved by using more powerful (and more expensive) components



Available Solutions Add Cost and Complexity -- Creating the Need for ILM

- ILM attempts to cope with storage pains via multi-tiered storage
  - Tiered storage management and data classification are costly and complex
  - Excessive data movements create reliability and performance issues
  - Utilization rates remain low (50% or less), with limited ability to execute thin provisioning



Imagine prioritizing electricity at home…

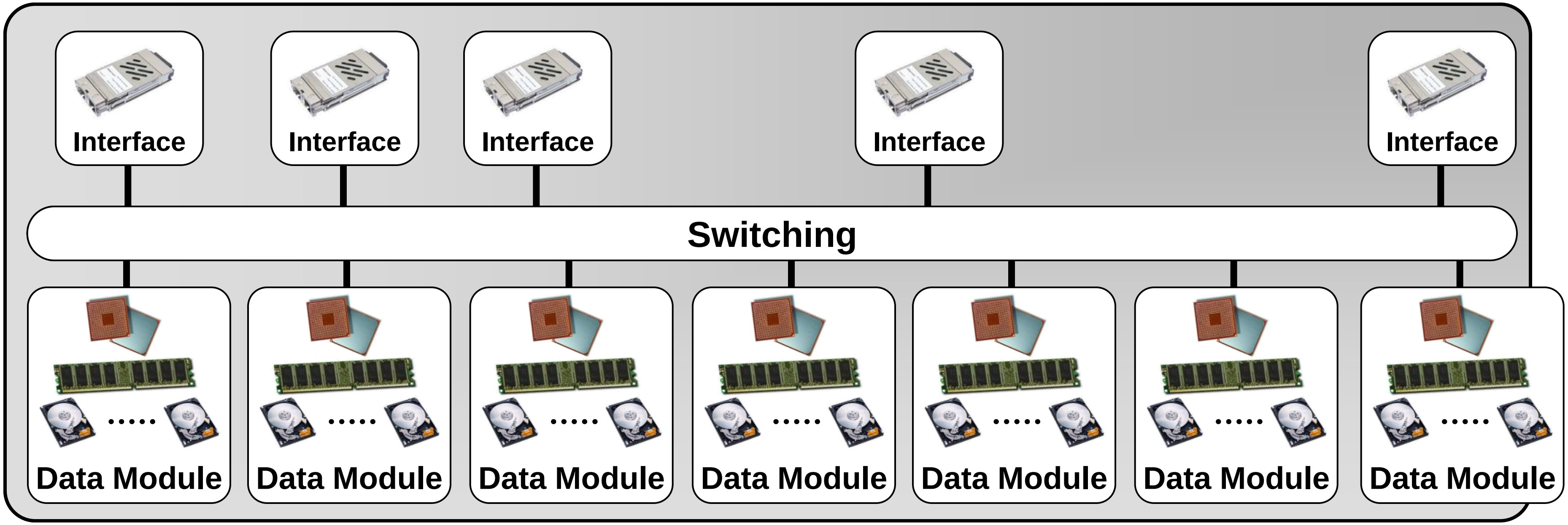
Laundry Power? Lamp Power? TV Power?

XIV – Example of 21st Century Architecture

Design principles:

- Massive parallelism
- Granular distribution
- Off-the-shelf components
- Coupled disk, RAM and CPU
- User simplicity

Scale Out



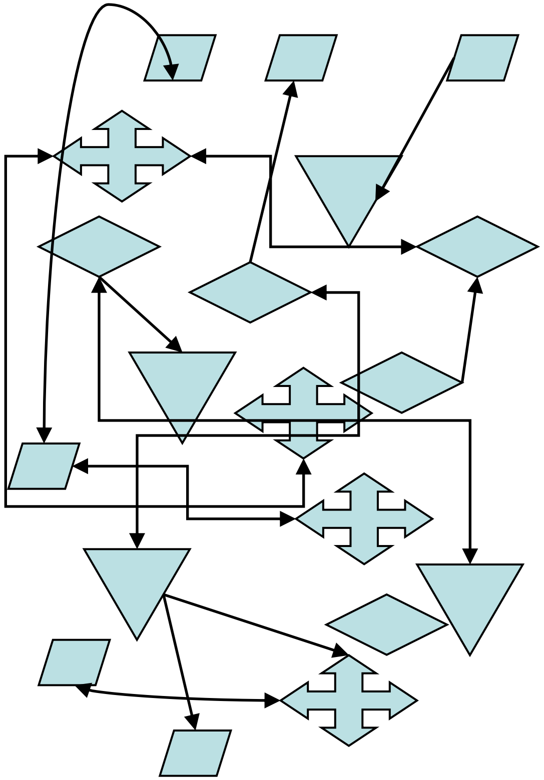
Only a novel architecture can close the gap between:

the exponential growth of information and
the limitation of the container

Data

0101001011101010111
0110101010111010101
0110101100110101011
0101110101010101011
0101110110101010101
0101010101010111101
0101010111010101010
1110101101011101010
1010111010101011110
1010101011011101010
1010111111101010111
0101101010011101010
0010101010100010101

Architecture



Disks

Storage Architecture Revolutions

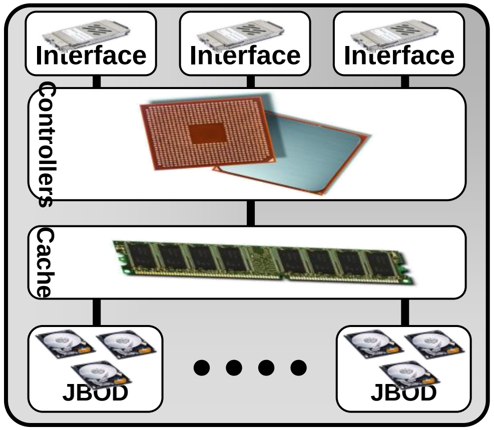
1970

- Mainframe
- Monolithic
- Gates design
- Very expensive

- Complex
- Downtimes
- On site technician
- Manual

1990

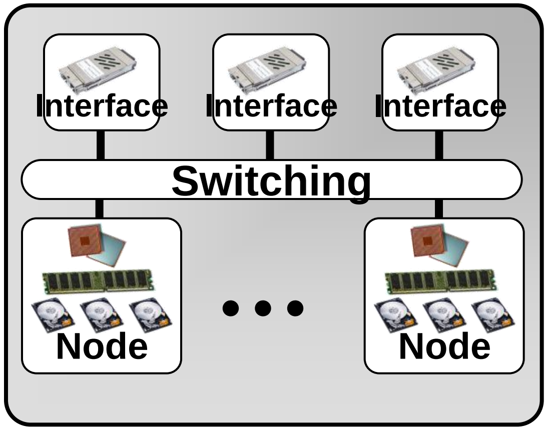
- Cluster architecture
- Tightly coupled
- Custom HW design
- Expensive components



- Long, complex development cycles
- System exposed on failures (the human factor)
- Complex reactive service
- Requires tuning for optimal performance

2010

- Scalable grid architecture
- Node independent
- Commodity H/W building blocks
- Off-the-shelf low cost components



- Fast, efficient development cycles
- WEB resiliency
- Self healing
- Data layout eliminate hotspots

THANK YOU

myanai@us.ibm.com

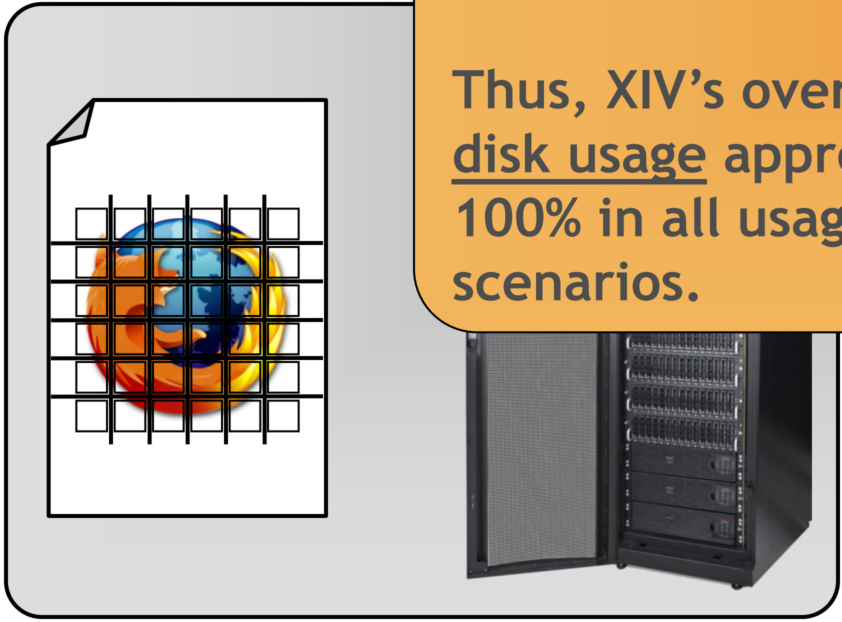
- Backup Slides

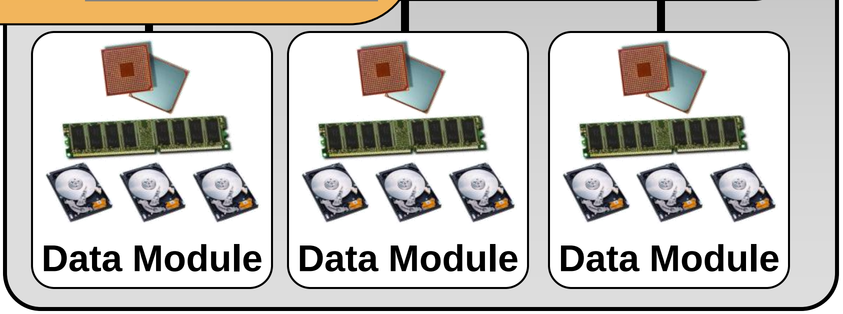
XIV - System Distribution Algorithm

- Each volume is spread across all drives
- Data is “cut” int
- XIV’s distribut ross **all** disks in the system ps

XIV disks behave like connected vessels, as the distribution algorithm aims for constant disk equilibrium.

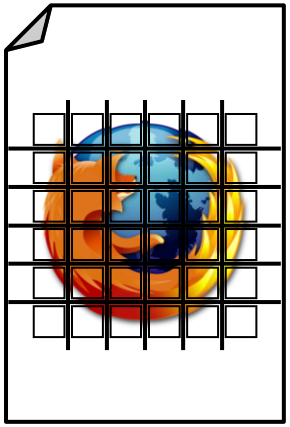
Thus, XIV’s overall disk usage approaches 100% in all usage scenarios.

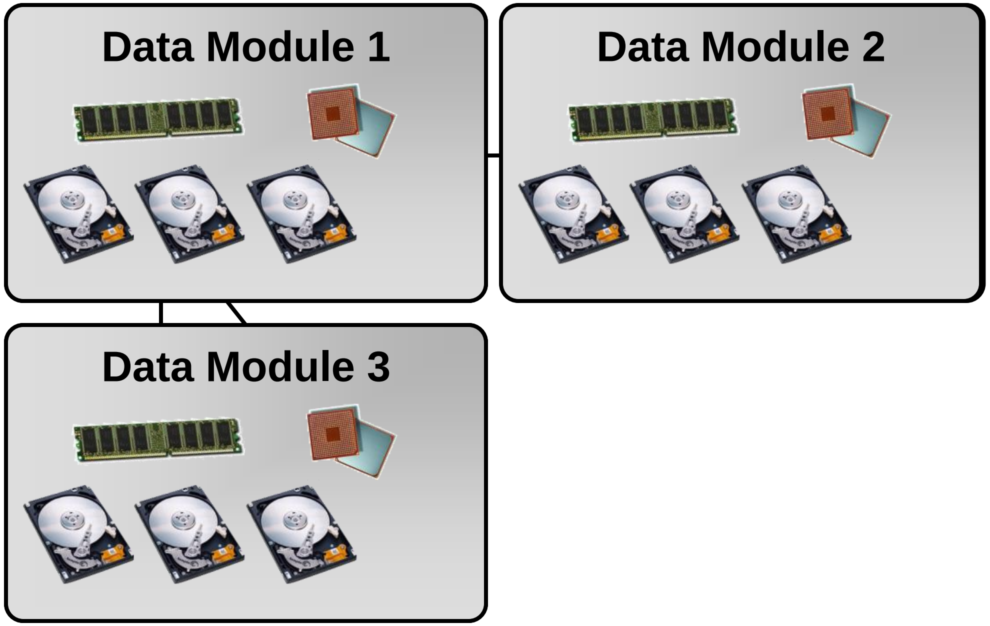




XIV - Distribution Algorithm on System Changes

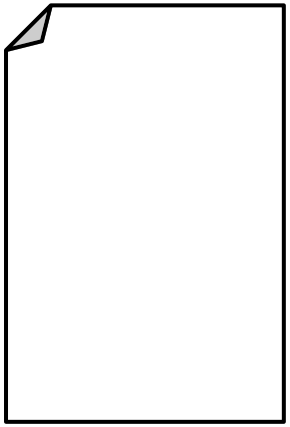
- Data distribution only changes when the system changes
  - Equilibrium is kept when new hardware is added
  - Equilibrium is kept when old hardware is removed
  - Equilibrium is kept after a hardware failure

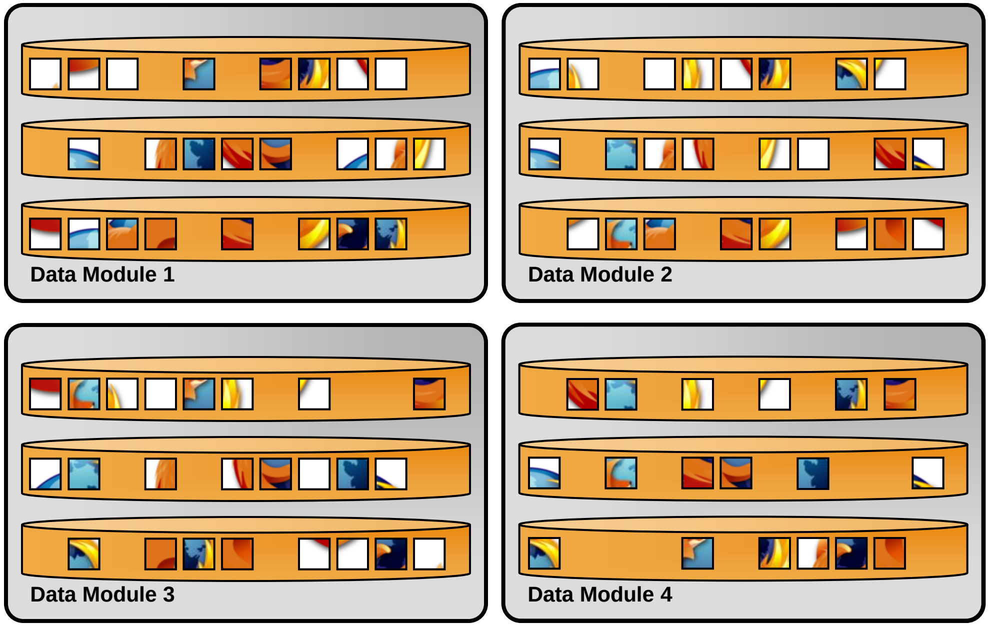




XIV - Distribution Algorithm on System Changes

- Data distribution only changes when the system changes
  - Equilibrium is kept when new hardware is added
  - Equilibrium is kept when old hardware is removed
  - Equilibrium is kept after a hardware failure





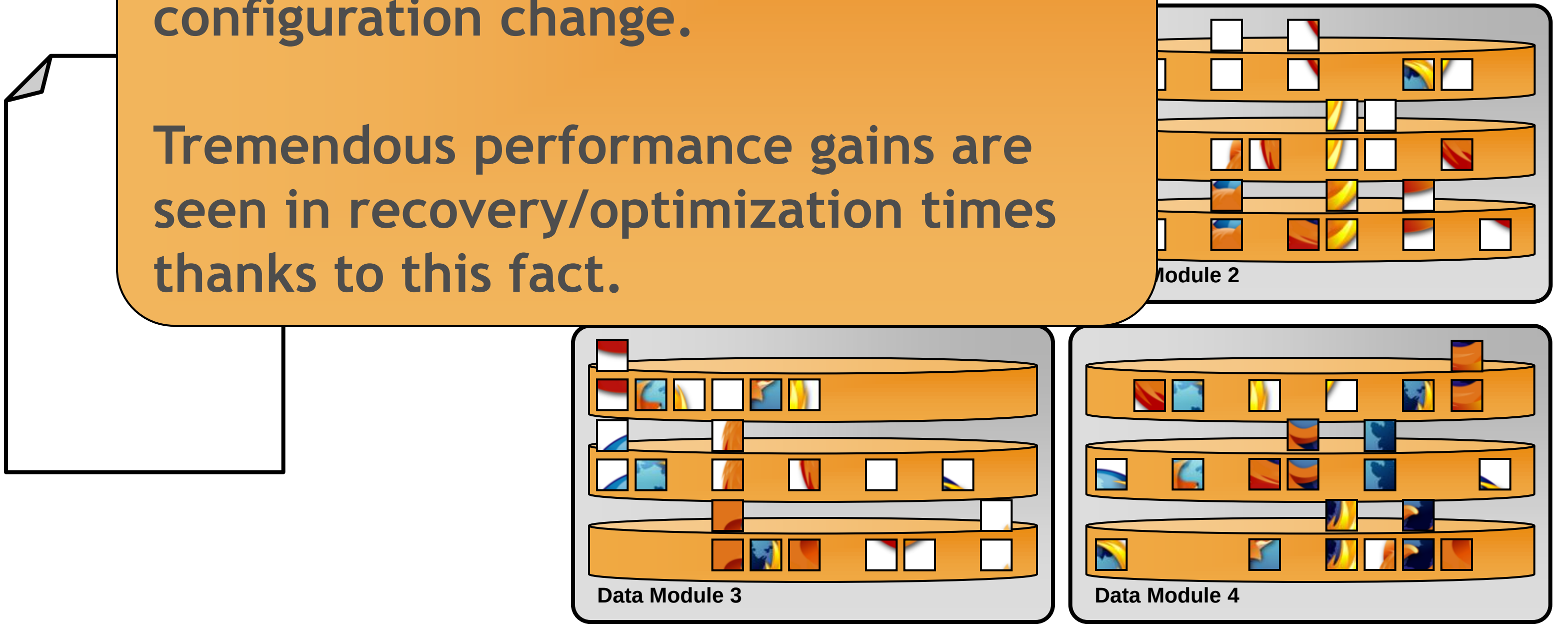
[hardware upgrade]

XIV Distribution Algorithm on System Changes

- Data distribution only changes when the system changes
  - Equilibrium is kept when new hardware is added
  - Equilibrium
  - Equilibrium

The fact that distribution is <u>full</u> and <u>automatic</u> makes sure all spindles join the effort of data re-distribution after configuration change.

Tremendous performance gains are seen in recovery/optimization times thanks to this fact.



XIV - SNAPs with No Limitations

- SNAPs creation/deletion is instantaneous
- High P
- Unlimit

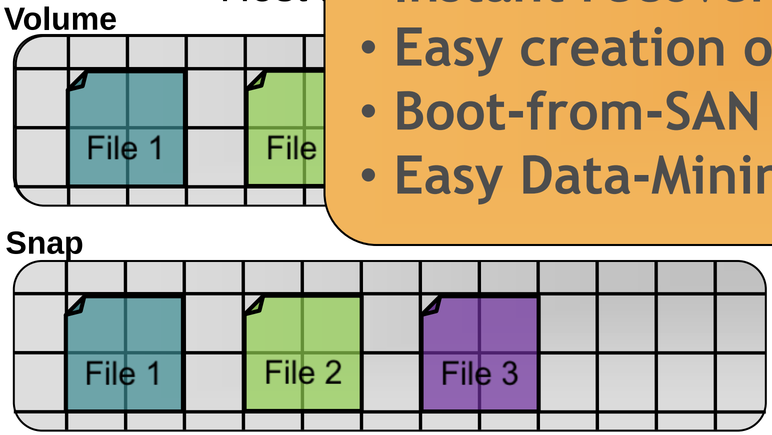
Distributed SNAP on each Server.
mory operations

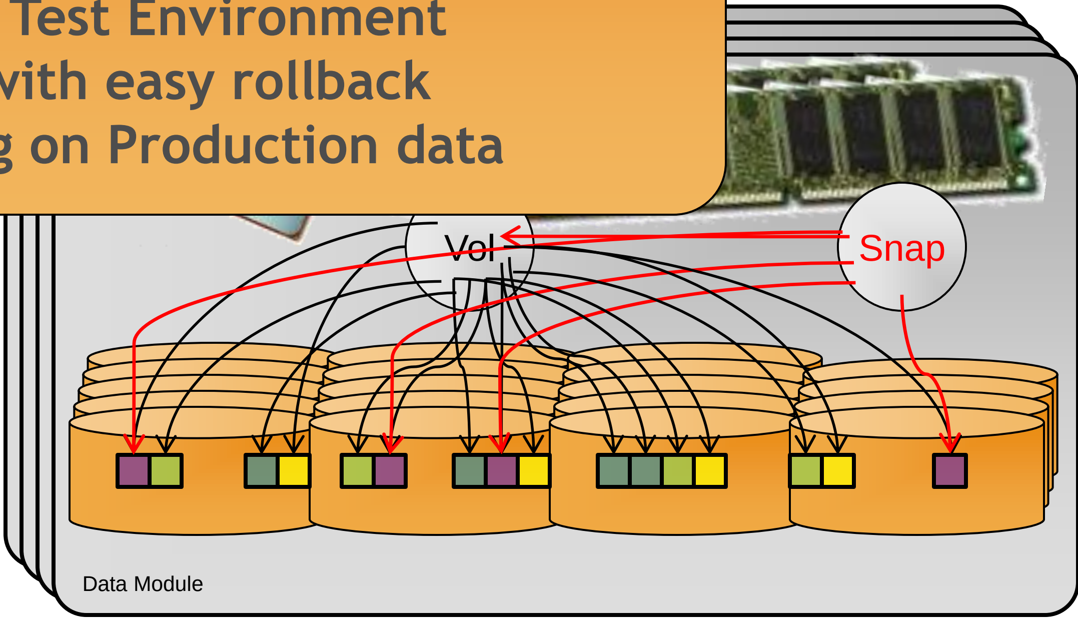
as fast as
on volumes

High Performance, Unlimited SNAPs provide:

- Easier Physical Backup to Tape
- Instant recovery from Logical Backup
- Easy creation of Test Environment
- Boot-from-SAN with easy rollback
- Easy Data-Mining on Production data

Host





Restore Volume from SNAP copy

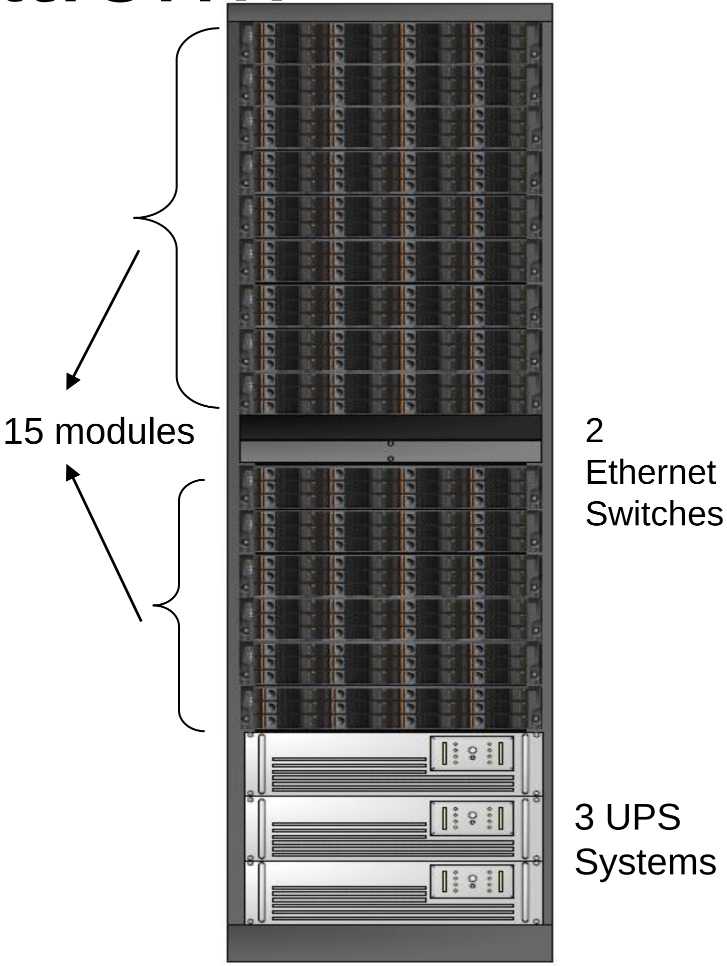
On a SNAP, each Server simply points to original volume. Memory only Operation

[illegible]

IBM XIV Storage System Hardware Platform

Machine Type: 2810-A14

- 180 disks per rack
  - 6 to 15 modules per rack
    - 12 disks per 2U module
  - 1TB or 2TB 7200RPM SATA disk drives
- 161 TB usable capacity for a single rack with 2TB drives
- 240 GB of system cache per rack (8GB per module)
- Up to 24 4GB FC host ports
- 6 1Gb iSCSI host ports
- 3 UPS systems



IBM XIV Storage