

# **HMI?** Program Webinar:

## **Project Plan, Methods, Discussion and Examples of Growth Estimates**

**July 9, 2008**

*James Short and Roger Bohn (UCSD)*





# What Are The Questions We Want To Answer?

- How much new information is created each year?  
(For “all information”, for enterprises, for people?)
- What factors drive growth?
- What, where are the emerging information applications?
- What do data flows, stored data and data growth look like in the big users of storage, networks, database applications, and processor technologies?

# What Dataset Are We Building?

- A fifteen year dataset on data flows, stored data, and data growth in “all” data, enterprise data, and individual data, with case estimates of “information in use” (data plus use = information)
  - Starting with Berkeley 2000, Berkeley 2003, used as baseline data on aggregate counts
  - 2008 recollects aggregate data and directs measurement attention to “counting by use” to derive “information growth” from data growth counts
  - Fifteen years (2000-2015):
    - 2000 – 2010 data (Berkeley 2000, 2003, HMI 2008, 2010)
    - 2010 – 2015 forecasted numbers\* for key growth sectors

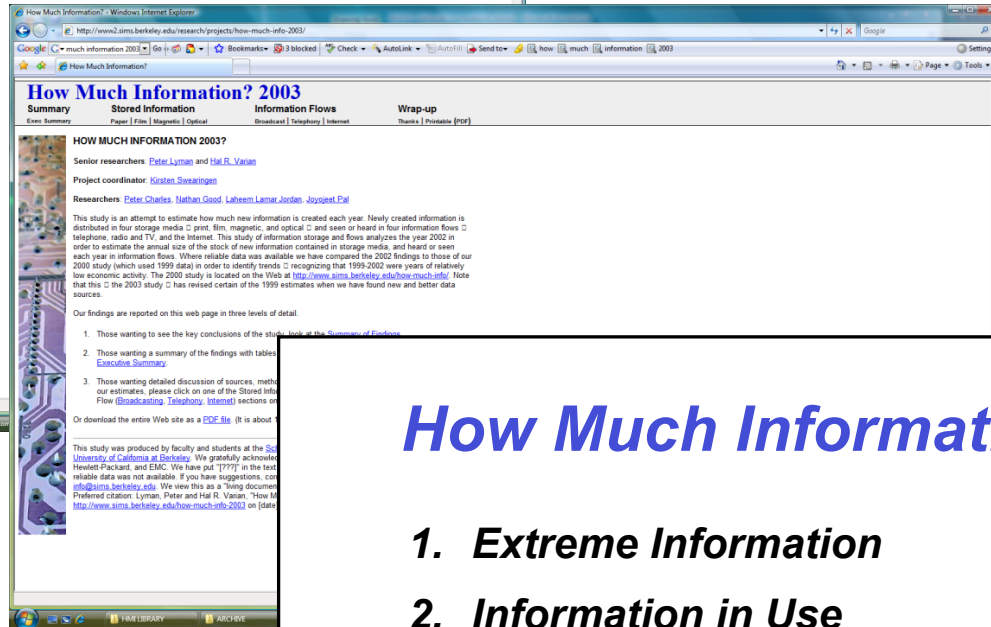
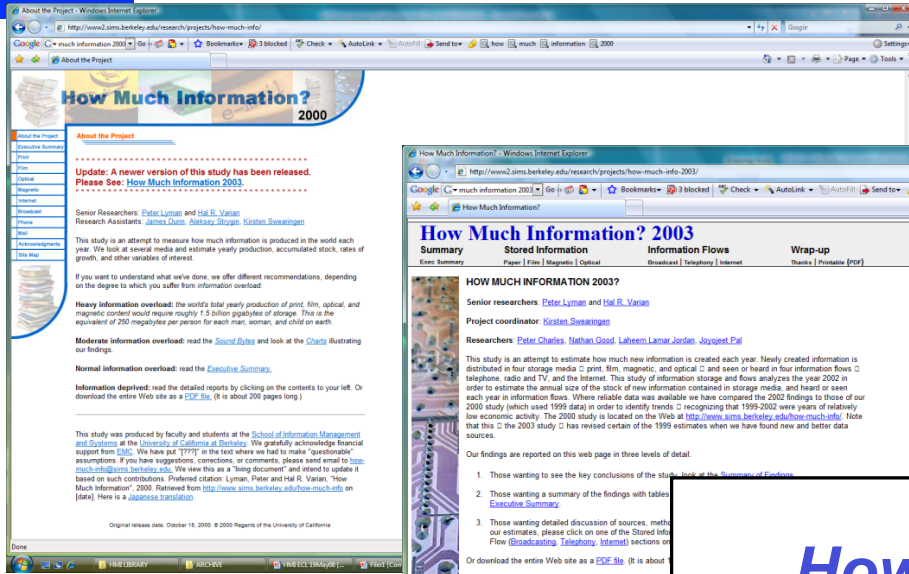
*\* Based on statistical estimates of use characteristics and growth, not extrapolations of equipment sales and installed base*



# How Will This Data Be Analyzed for Industry Insight?

- Through interpretation of aggregate counts as they affect different industries, enterprises, and people
- Through selected use cases where drivers, raw growth, ratios, and management practices are studied and compared
- Through bottom up measurements (“direct measurement”) of application domains and comparison of growth measurements with aggregate numbers
- With forecasts based on simulations and statistical estimations derived from use case analysis

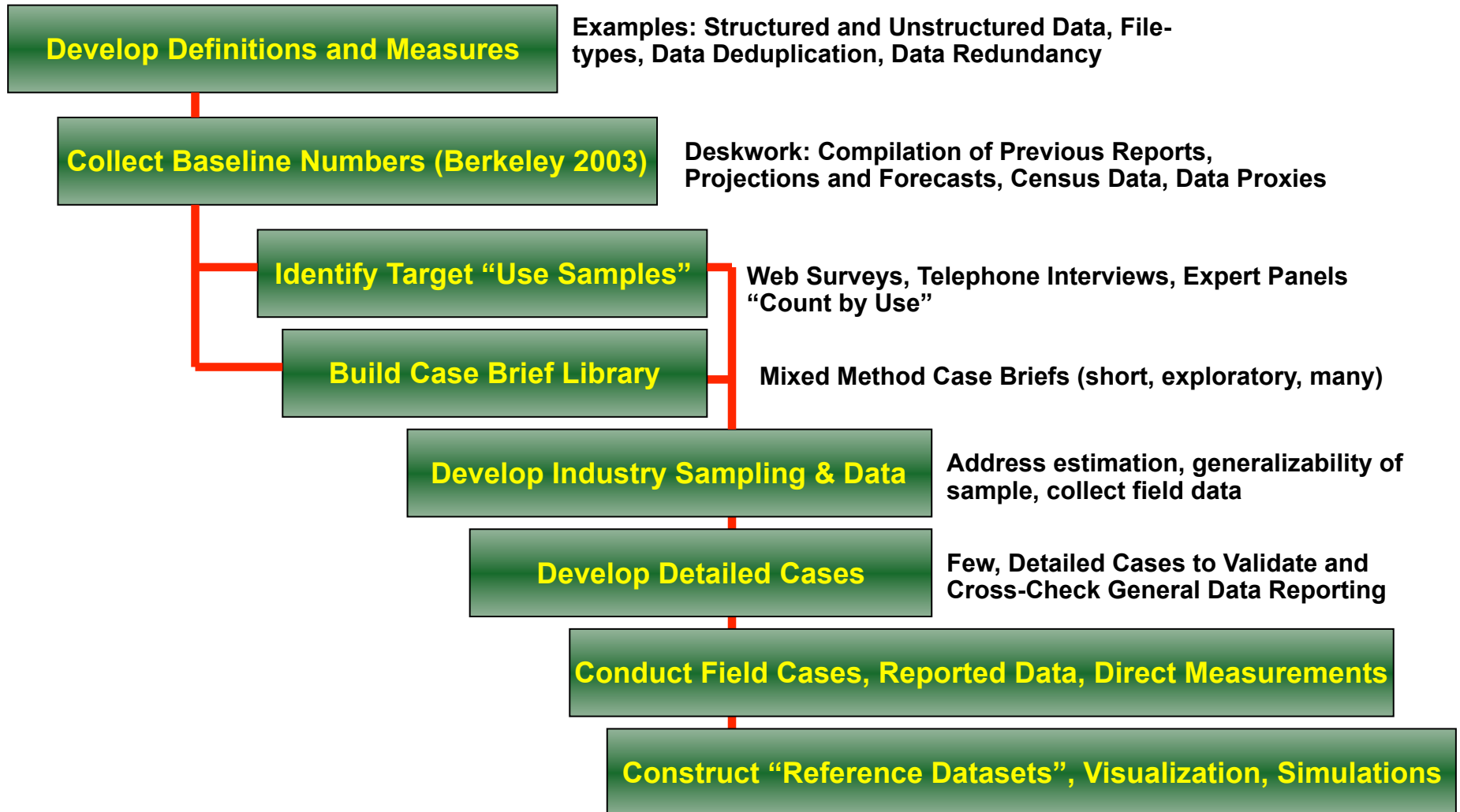
# Objective



## How Much Information? 2009

1. *Extreme Information*
2. *Information in Use*
3. *The Data on Information (Berkeley+)*
4. *Methods*

# HMI? 2009 How Are We Counting?





# Spotlight: How Much Information is there in the Universe?

- Answer: 10 raised to the power of 120 bits
- Calculated by Seth Lloyd in “Programming the Universe”
  - <http://www.technologyreview.com/Infotech/17091/?a=f>
  - <http://www.randomhouse.com/kvpa/lloyd/>
- *"Something else has happened with computers. What's happened with society is that we have created these devices, computers, which already can register and process huge amounts of information, which is a significant fraction of the amount of information that human beings themselves, as a species, can process. When I think of all the information being processed there, all the information being communicated back and forth over the Internet, or even just all the information that you and I can communicate back and forth by talking, I start to look at the total amount of information being processed by human beings — and their artifacts — we are at a very interesting point of human history, which is at the stage where our artifacts will soon be processing more information than we physically will be able to process."*

SETH LLOYD — HOW FAST, HOW SMALL, AND HOW POWERFUL?: MOORE'S LAW AND THE ULTIMATE LAPTOP  
[7.23.01]

# Spotlight: How Much Data? Human Conversation

*Question: Are Women Really More Talkative Than Men?*

**Women 16,215 ( $\pm 7301$ )**

**Men 15,669 ( $\pm 8633$ ) per day**

	<b>Text</b>	<b>Audio</b>	<b>Video</b>
Bytes/word	5	2.5KB	30KB
Bps@200 wpm	16.5	8KB	100KB
Per day @16000 words per day	90KB	40MB	480MB
Per person-year	33MB	14.6GB	175GB
Per Earth-year @6.6B people per planet	<b>0.2 E18</b>	<b>97 E18</b>	<b>1.2 E21</b>

**1E18 =  $10^{18}$  = 1 exabyte**

*Mehl, Matthias R, Vazire, Simine, Ramierz-Exparza, Nairan, Slatcher, Richard B., and Pennebaker, James W., "Are Women Really More Talkative Than Men?" 6 July 2007, Vol 317 Science.*



# Spotlight: At present, most data stays local

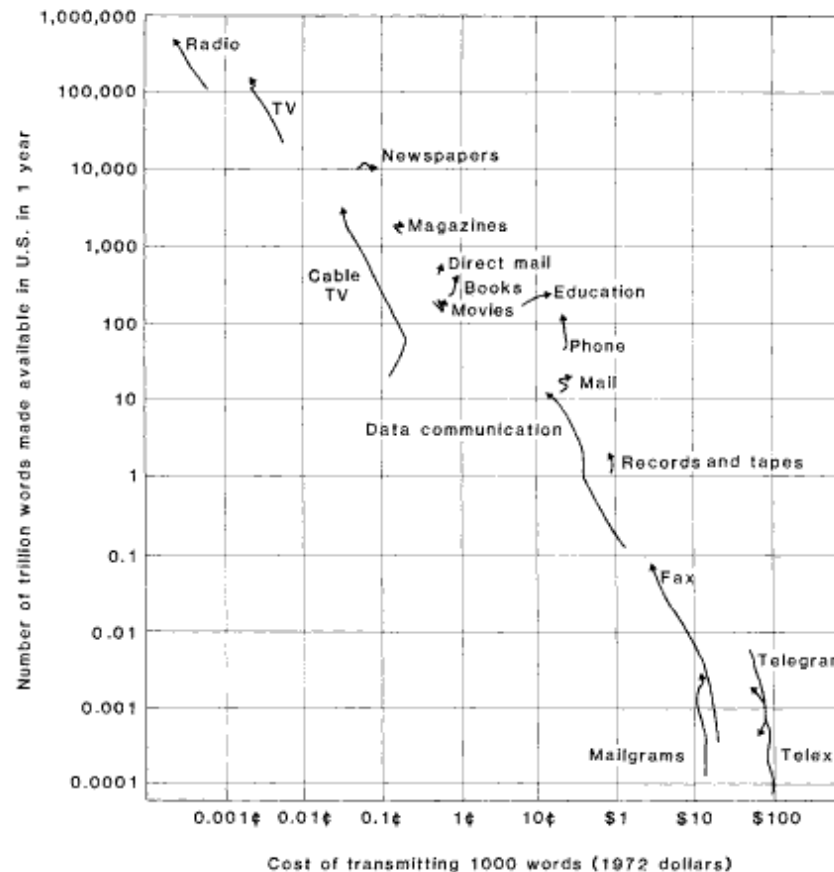
Data transformed	Data stored?	Example
Events		speech
Measured (analog)		Video surveillance
	Stored locally	VHS recorder
A to D		Camera
	Stored locally	Aircraft black box
Sent externally		Email photos
Broadcast/downloaded	Replicated	You-tube
	Backed up	CERN data

**1000X more data *could* be captured**

**Potential value versus current (kinetic) value**

# Spotlight: The Costs of Broadcasting Words Has Plummeted, What About Consumption?

Words Communicated versus Costs of Communication



*Ithiel Pool, Tracking The Flow of Information, Science 1983*

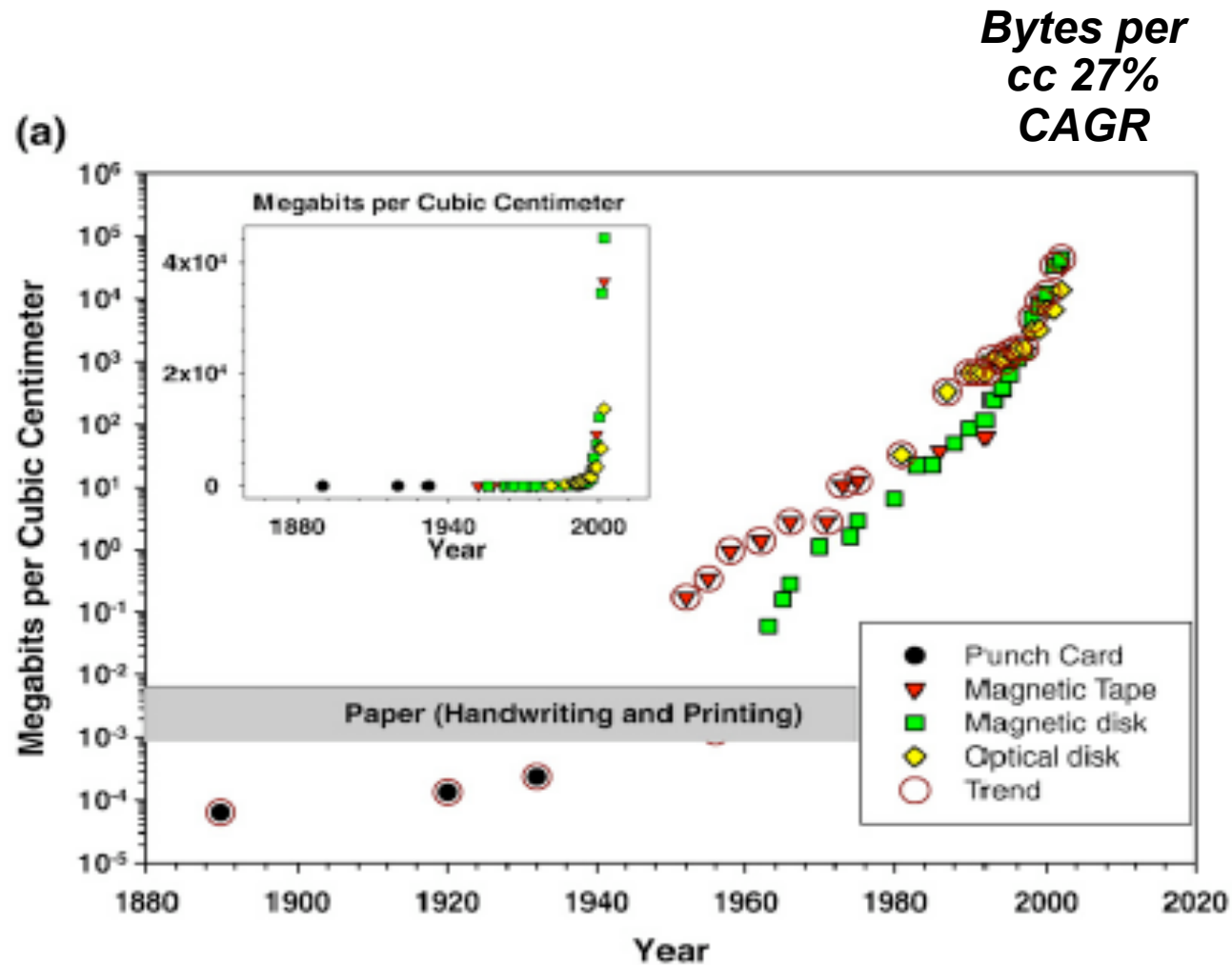
Fig. 1. Trends in volume and costs of communication for 17 media, 1960 to 1977 (plotted on log by log scales).



## **Spotlight: Chris Magee and Heebyung Koh's work on mapping IT technology changes with emphasis on storage**

- “A functional approach for studying technological progress: Application to information technology”  
Technological Forecasting & Social Change 73  
(2006)

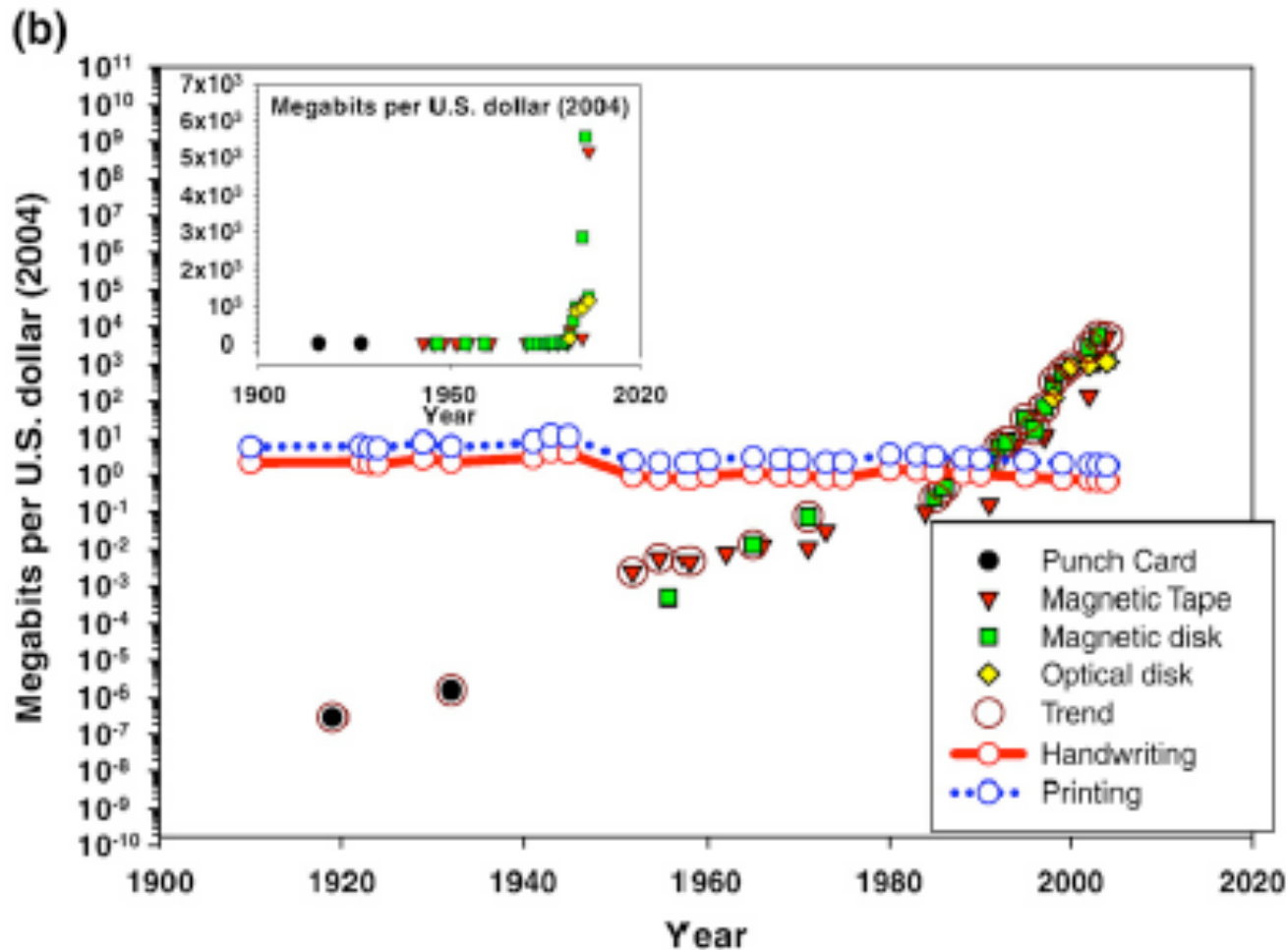
# Storage 100 year trends



Ref: Heebyung Koh and Christopher L. Magee (2006)

# Storage 100 year trends

Bytes per dollar  
27% CAGR



Ref: Heebyung Koh and Christopher L. Magee (2006)

# Exponential growth rates

	<b>Whole period</b>	<b>1940 to present</b>
<b>Bytes per volume</b>	$21 \pm 2$	$26 \pm 2$
<b>Bytes per dollar</b>	$26 \pm 3$	$27 \pm 5$
<b>Bandwidth = bits/ sec</b>	$19 \pm 3$	$35 \pm 5$
<b>(Bandwidth x mile) per \$</b>	$19 \pm 3$	$33 \pm 4$

*Ref: Heebyung Koh and Christopher L. Magee (2006)*

# Spotlight: Auto sensors - New Apps, Moore's Law



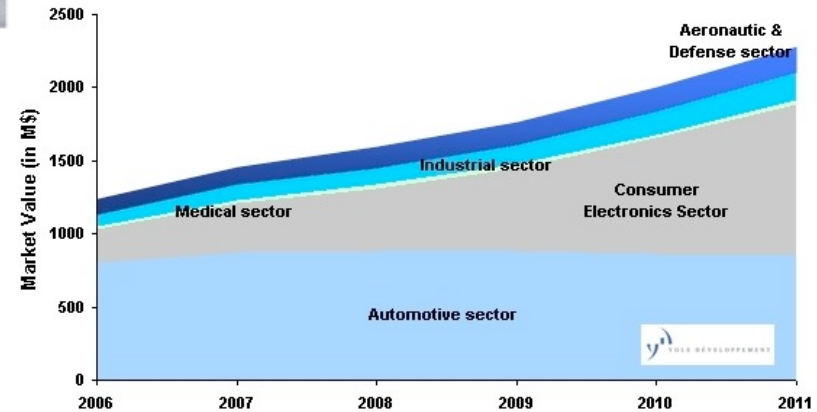
## Micro-Electro-Mechanical Systems (MEMS)

- Adaptive Optics for Ophthalmic Applications*
- Optical Cross Connects*
- Air Bag Accelerometers*
- Pressure Sensors*
- Mirror Arrays for Televisions and Displays*
- High Performance Steerable Micromirrors*
- RF MEMS Devices*
- Disposable Medical Devices*
- High Force, High Displacement Electrostatic Actuators*
- MEMS Devices for Secure Communications*

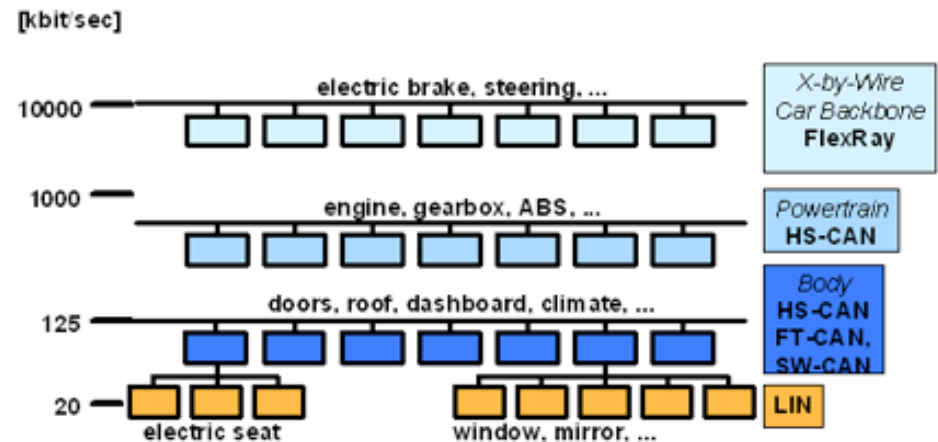
**Freescale Ships Over 100 Million Automotive Microcontrollers Annually**



**Accelerometer and Gyroscope MEMS markets**



# Spotlight: Vehicle Informatics - Auto data rates



- Bus rates > 1 Mbps; now a 10Mbps bus
- Average > 10 MCUs per new car
- 500 hours/yr x 2 Mbps = .5 TB/yr = .5E12
- US has  $10^8$  autos
- 50 Exabytes/year of data
- Save it? How long?

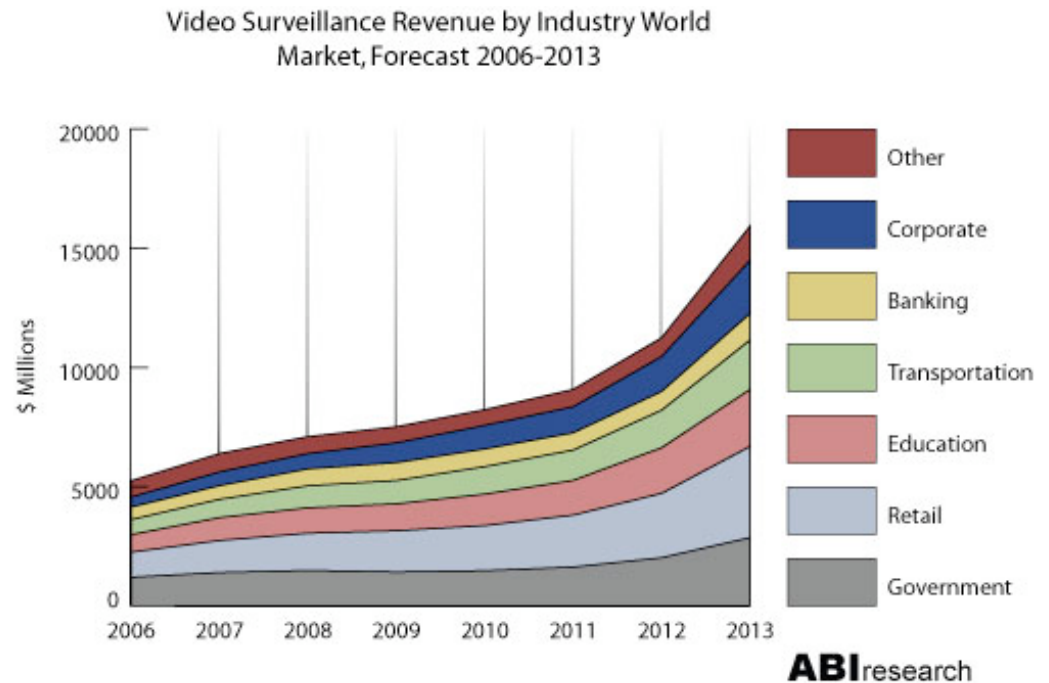
# Spotlight: Video Surveillance – New Technologies, New Players, New Apps

- Market Drivers, Commercial & Public Sector
  - Homeland Security
  - Violence in Schools – Law Enforcement
  - Theft Detection and Avoidance
  - Revenue Enhancement
  - Improved Operations
- Obstacles to Growth
  - IT Department Reluctance
  - Costs
  - Installed Base



# Spotlight: Video Surveillance – New Technologies, New Players, New Apps

- Key Verticals
  - Banks
  - Transportation
  - Education
  - Retail
  - Casinos
  - Government
- Aggressive Overall Market Growth Forecasted
- IP/Networked video surveillance segment grew nearly 50% in 2007, to approach \$500 million worldwide



# Spotlight: Video Surveillance Use Case – License Plate Recognition (LPR)



- Syracuse, New York installs in school buses (April 2008)



- License Plate Reader (LPR)
- Mobile Plate Hunter 900
  - Several cameras on squad car connected to vehicle computer
  - Sensors work all the time
  - 90% hit rate at 150 mph closing speed
  - Traditional “running a plate” dozen a shift, LPR can run 10,000
- Market leader: Eltag NA
  - <http://www.elsagna.com/about.htm>

# Spotlight – Video Analytics

Cory Doctorow: Big Brother is not watching | Technology | guardian.co.uk - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address: http://www.guardian.co.uk/technology/2008/jun/17/surveillance.database

guardian.co.uk

Search: guardian.co.uk Search

News | Sport | Comment | Culture | Business | Money | Life & style | Travel | Environment | Blogs | Video | Jobs | A-Z


News > Technology

Series: Digital rights, digital wrongs Previous | Next | Index

## Surveillance: You can know too much

Britain has never collected more data on her citizens. But what use is all that information if we can't process it fast enough?

Cory Doctorow  
guardian.co.uk, Tuesday June 17, 2008  
Article history



Photograph: Kisty Wigglesworth/AFP/Getty Images

Technology  
Computing - Politics and technology

UK news

Series  
Digital rights, digital wrongs

Comment

larger | smaller

Guardian America.  
Find out more >>  
theguardian.co.uk

### Gadgets and gizmos

BlueTinum Internet, DAB & FM Radio  
Listen to Internet radio broadcasts from around the world. Only £99.95

Digital Luggage Scales  
Check in your luggage safe in the knowledge that you are within the required limits. Only £22.99

### Browse the Guardian Bookshop

Teach Yourself Basic Computer Skills  
£9.99 with free UK delivery

Database Management Systems  
£45.99 with free UK delivery

The Singularity is a conceit of modern science fiction: a place inside vast computers where whole universes are simulated whose reality is every bit as sharp and instantaneous as the physical world we inhabit. Books like Charlie Stross's Singularity Sky and the Matrix movie trilogy have done a great job of representing such alternative, computer-calculated realities.

Science fiction writer Rudy Rucker, however, has taken issue with the whole idea, arguing that the universe is already "calculating" reality far better than any software could manage.

After all, he says, the rock that you just dropped on your foot did a brilliant job of solving the physics of its velocity and direction, taking into account every conceivable variable from wind shear to the minute irregularities in its surface. Your nerve endings calculated, in flawless real-time, the screech of pain that was sent up your nerves into your brain, which in turn calculated precisely how much it all hurt and which cuss words flew involuntarily from your lips.

Most problems don't require such accurate solutions, so we can usefully



# Q&A