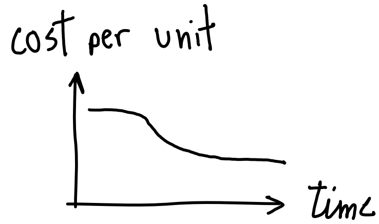
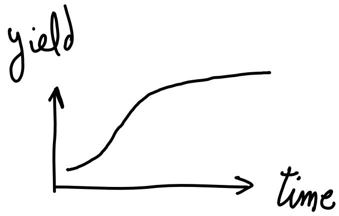
Cost



Manufacturing costs drop as expertise grows, for that process

-- better **methods**
-- better **equipment**
-- **less waste** (time, materials)



Yield% = (1 - waste%)

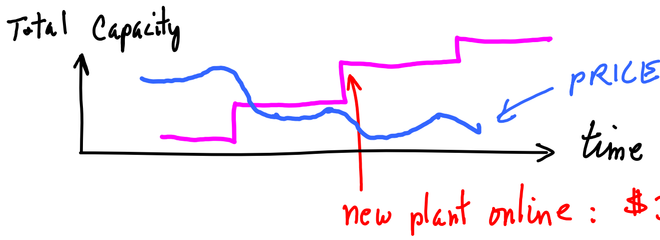
-- #(devices **sellable**) **versus** #(devices **produced**)

-- #(devices **sellable**) **versus** (**cost** to produce them)

E.g. DRAM ⇒ price = $\propto$ cost

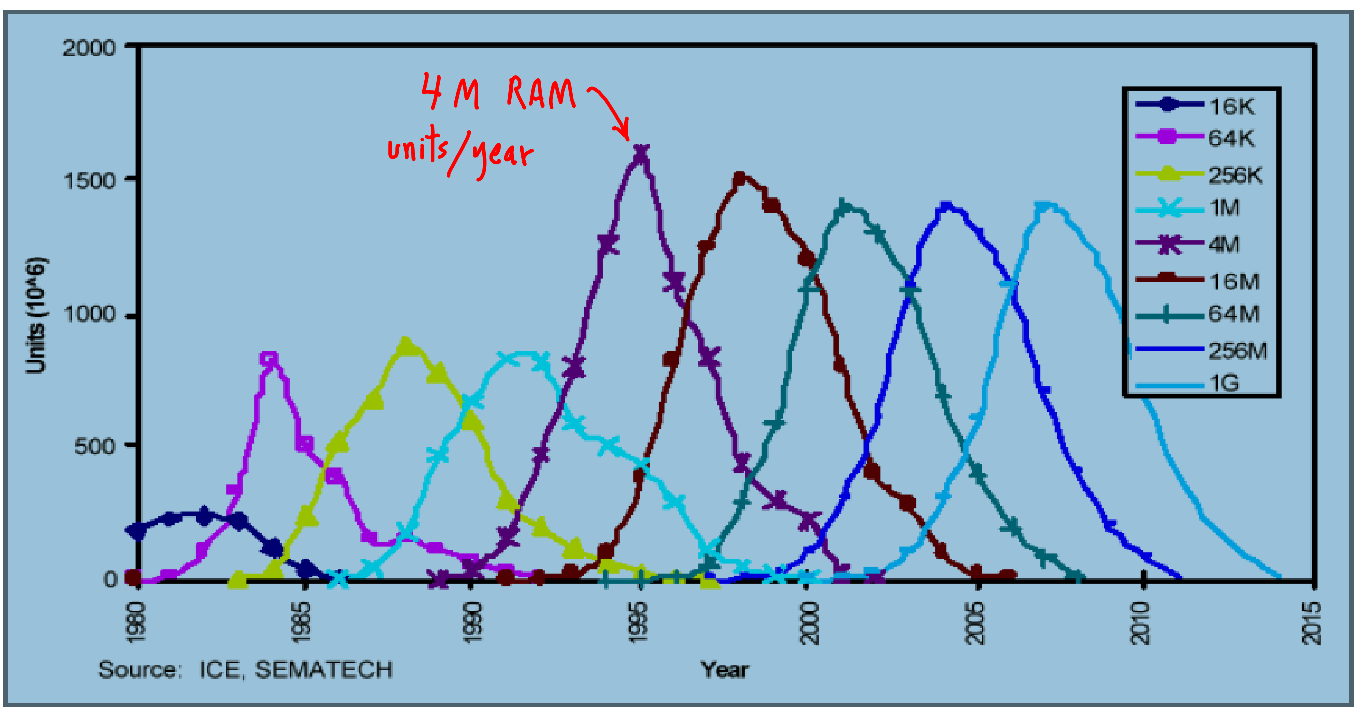
80% contract sales to large Equipment makers (hidden)

20% open market



Commodity market

---**many vendors**
---**same items**



DRAM Unit Volume by Generation [37]

Invest in largest **demand ==>** production **cost amortized ==> larger profit**
hot-new ==> high price / low volume **old-standard ==> low price**

Costs Drop

$50/64

$65/256

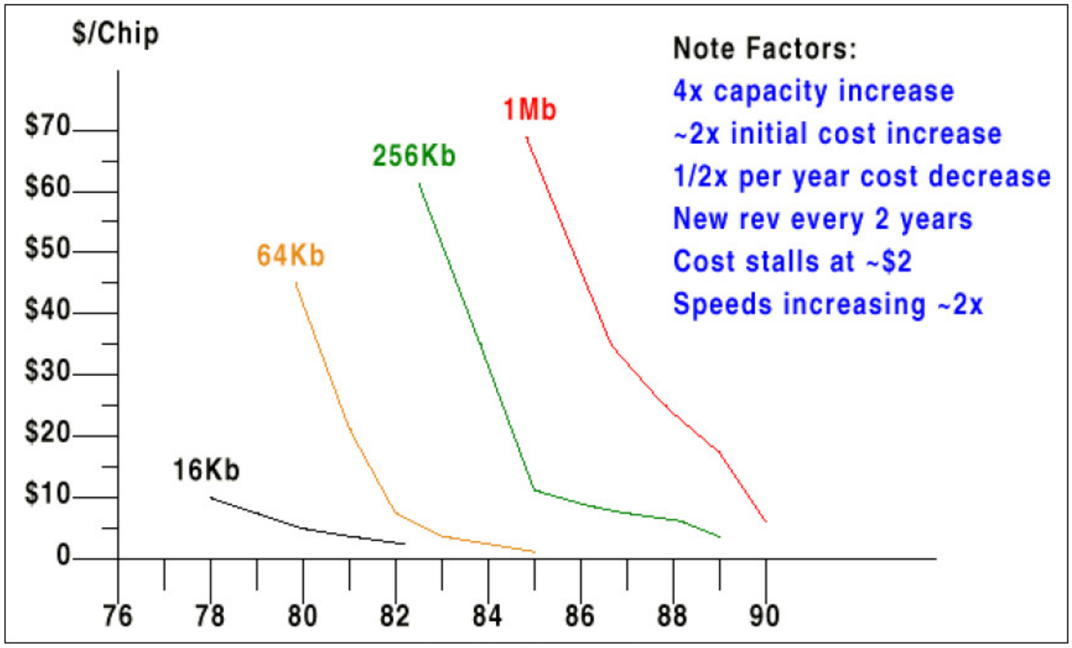
4x capacity

2x speed

Cost per bit

per bit/sec

Drop faster

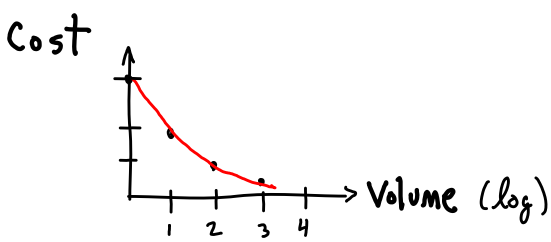


DRAM Costs [45]

Changes Telecom (routers/switches) 20% ↑ ⟹ 50% of market

latency ↓ vs bandwidth ↑

routers	PCs
SRAM	DRAM
12 ns	40 ns
$50/~2MB	$200/GB



$\frac{1}{2}$x Cost : 2 x Volume

Volume → supplier competition → lower Cost

⟹ Low-end Market (Price/Performance) ↓

Standardization / Volume ====> market acceptance of innovations

Cloud Pricing

AWS

Combined efficiencies

bottomed out

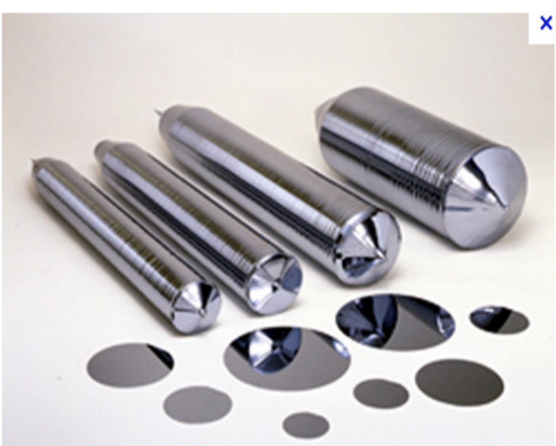
Description	Type	CU	Original $ / CU / Hour	Current $ / CU / Hour	% Reduction	Aug 2006	Oct 2007	May 2008	Oct 2009	Feb 2010	July 2010	Sep 2010	Nov 2010	Nov 2011
Small - "the original"	m1.small	1	$0.10	$0.085	15%	$0.10			$0.09					
Large	m1.large	4	$0.10	$0.085	15%		$0.40		$0.34					
Extra Large	m1.xlarge	8	$0.10	$0.085	15%		$0.80		$0.68					
High-CPU Medium	c1.medium	5	$0.04	$0.03	15%			$0.20	$0.17					
High-CPU Extra Large	c1.xlarge	20	$0.04	$0.03	15%			$0.80	$0.68					
High-Memory Double Extra Large	m2.2xlarge	13	$0.09	0.077	17%				$1.20			$1.00		
High-Memory Quad Extra Large	m2.4xlarge	26	$0.09	0.077	17%				$2.40			$2.00		
High Memory Extra Large	m2.xlarge	6.5	$0.12	0.077	33%					$0.75				
Cluster Compute	cc1.4xlarge	33.5	$0.05	$0.04	19%						$1.60			
Cluster Compute Eight Extra Large	cc2.8xlarge	88	$0.03	$0.03	0%									$2.40
Micro	t1.micro	0.9	$0.02	$0.02	0%							$0.02		
Cluster GPU Instance	cg1.4xlarge	33.5	$0.06	$0.06	0%								$2.10	

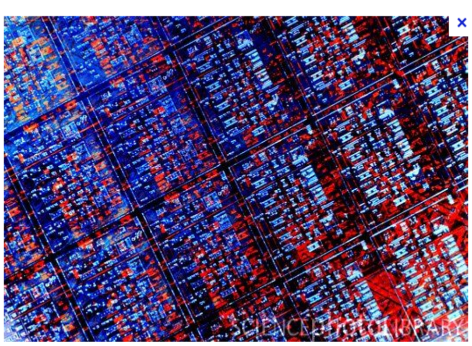
Price Reductions

CPUs, Chips, SoC

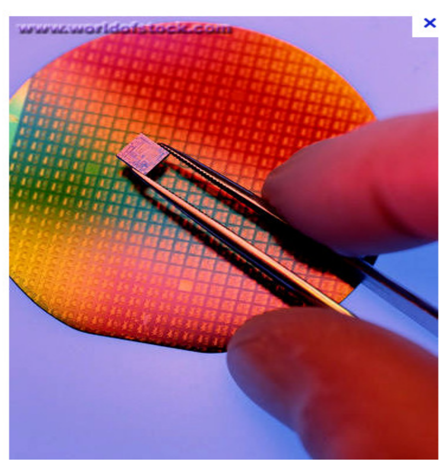
Si ingots slicing → wafers

masking, etching, doping

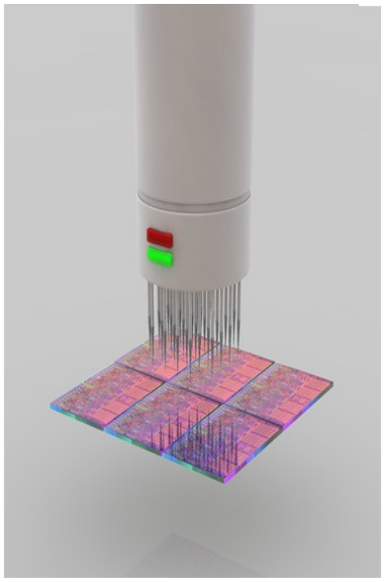


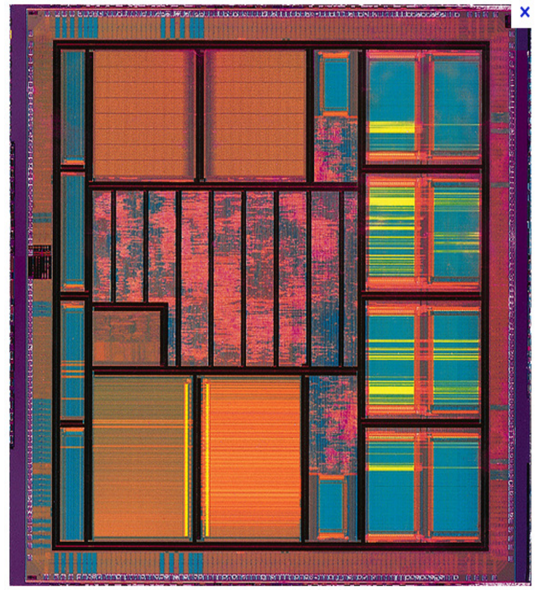


dicing →

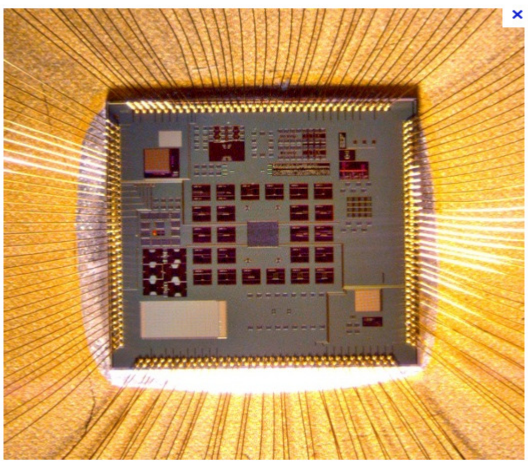


Circuit testing

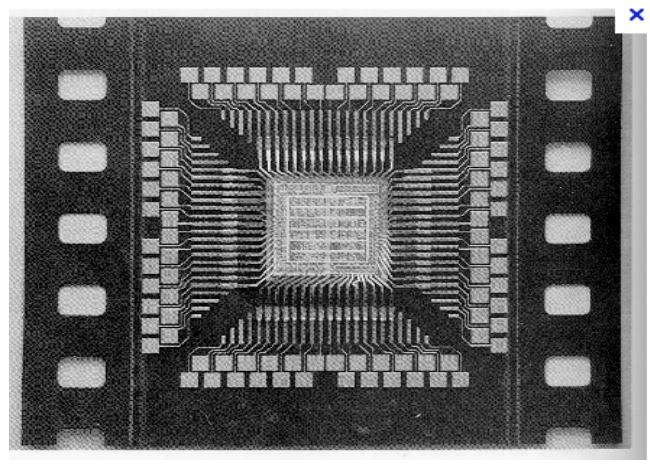




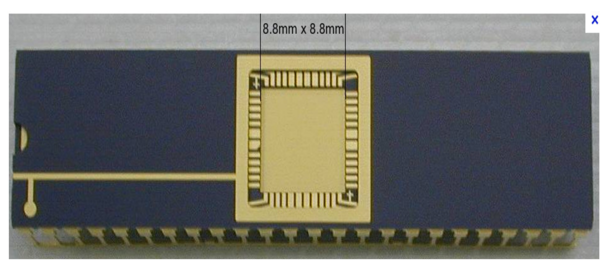
Pad Bonding

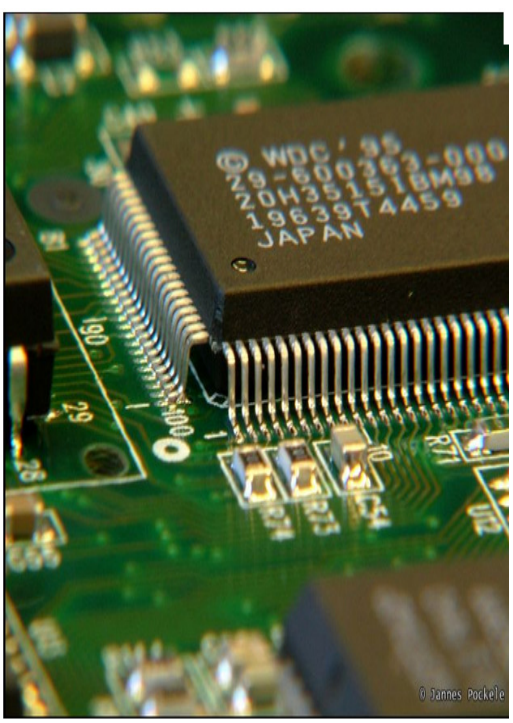


Pin Packaging



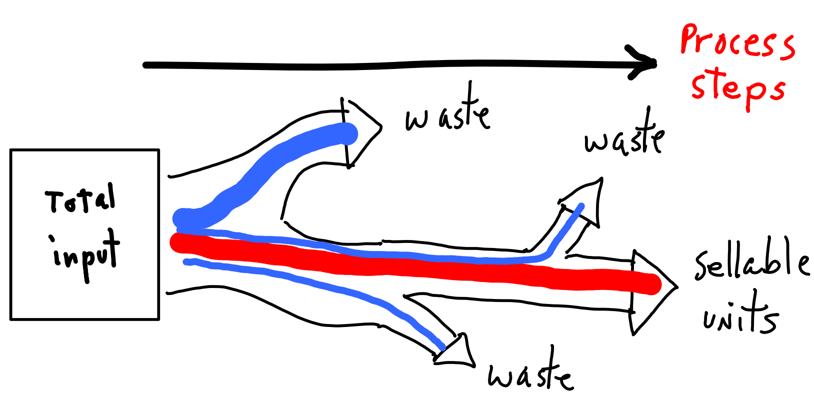
encasing



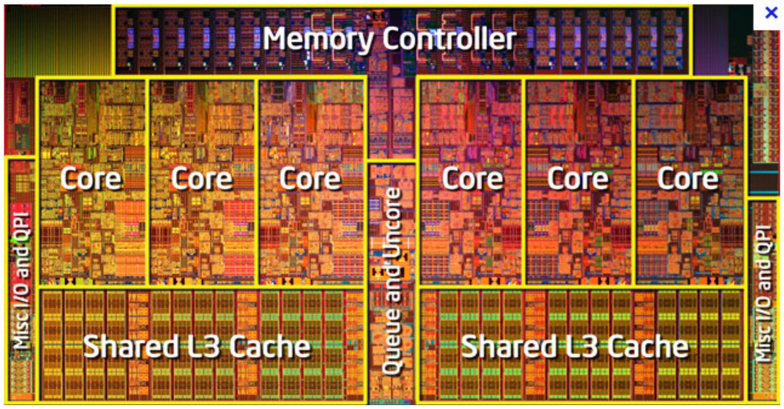


board printing

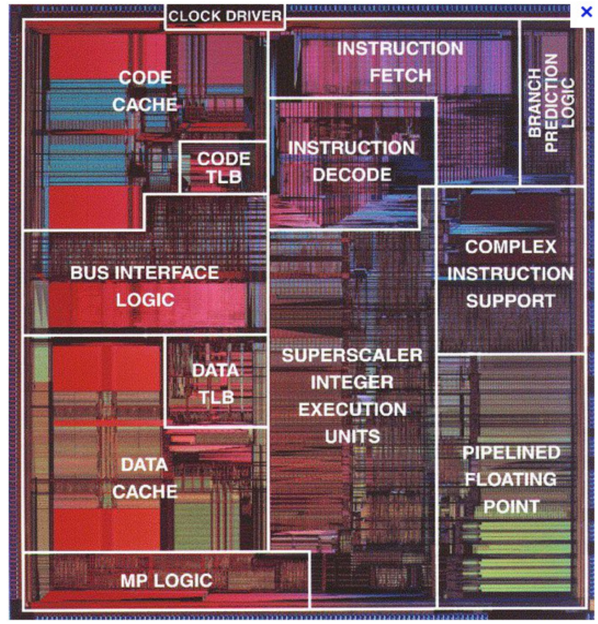
mounting



what's inside?



i7



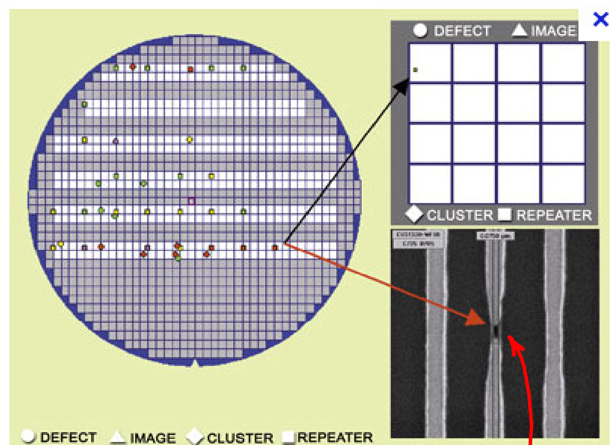
P5

$$\text{Cost} = \frac{C_{die} + C_{Test_1} + C_{package} + C_{Test_2}}{\#(\text{sellable units})}$$

$$C_{die} = C_{wafer} / (\# \text{dies})(\text{yield})$$

$C_{wafer} \approx \$5000$

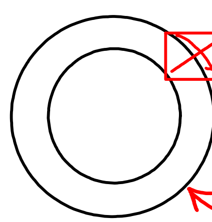
$$\Longrightarrow \quad C_{die} \propto 1/(\# \text{dies})$$



wire defect

$$\#(\text{dies}) = \left(\frac{\text{Area}_{wafer}}{\text{Area}_{die}}\right) - \left(\frac{\text{Circumference}_{wafer}}{\text{Diagonal}_{die}}\right)$$

$$= \frac{\pi r^2}{A_{die}} - \frac{2\pi r}{\sqrt{2}\sqrt{A_{die}}}$$



$$\text{yield} \cong \frac{\#(\text{good wafers}) / \#(\text{wafers})}{\left[1 + \frac{\#(\text{defects})}{cm^2}(A_{die}\, cm^2)\right]^N}$$

Curve fitting for particular process $\Rightarrow$ $N \in [11.5, 15.5]$

$$\frac{\#(\text{defects})}{cm^2} \approx \underline{0.04}$$

function of time + volume

300 mm Wafer

$A_{die} = 2.25\ cm^2 \Rightarrow 109$

$A_{die} = 1\ cm^2 \Rightarrow 424$

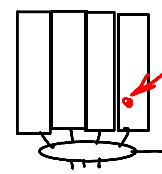
P5 Sandy Bridge	embedded CPU, 32b	printer Controller
$2\ cm^2$	$0.1\ cm^2$	$0.04\ cm^2$
\$50	\$13	\$0.1

Amortized Costs

Mask = \$1M

$\Rightarrow$ reconfigurable gate arrays

Redundancy, e.g.



defect

Select 3 banks
1 spare

$$\left.\begin{array}{l}\text{Die size} = \#(\text{Transistors}) + \#(\text{pins}) \uparrow \\ + \\ \text{Volume} \downarrow \\ + \\ \text{Customization} \uparrow\end{array}\right\} \text{cost}$$

Warehouse-Scale Costs

$$\text{Cost}_{\text{computing}} = \text{Cost}_{\text{equipment}}/\text{unit Time} + \text{Cost}_{\text{power}} + \text{Cost}_{\text{structure}}/\text{Time} + \text{Cost}_{\$}/\text{Time} + \text{Cost}_{\text{repair}}$$

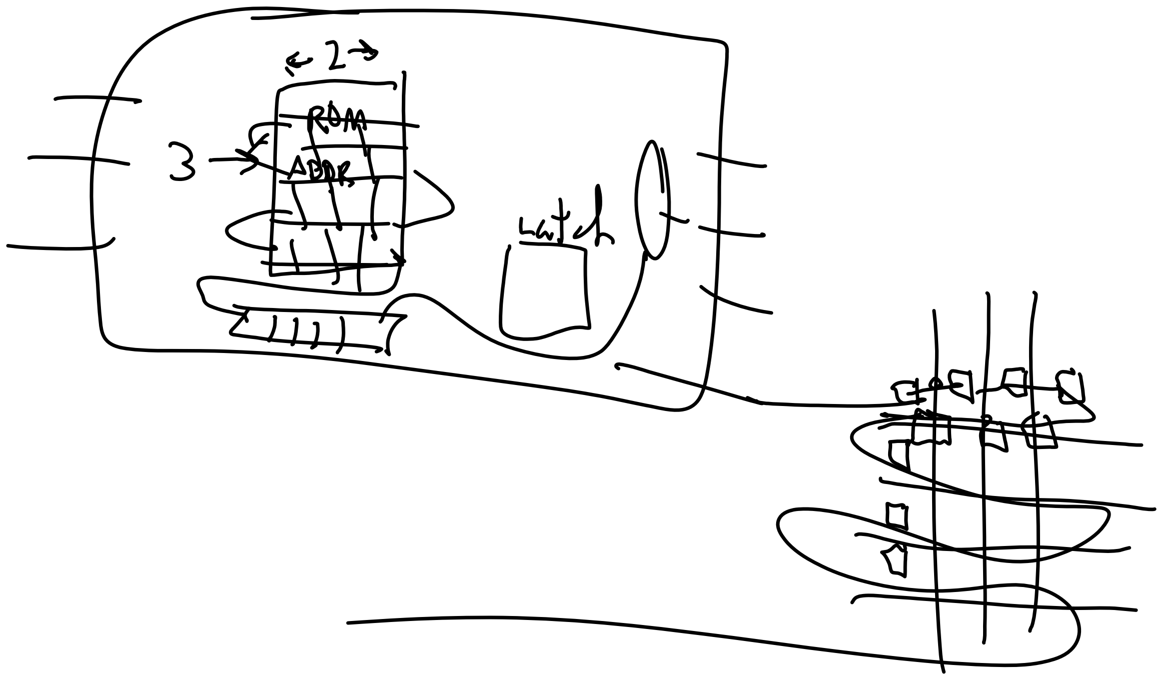
$$\underset{(60\%)}{\left(\frac{\text{Computers + networks}}{3\ \text{yr}}\right)} + \underset{(40\%)}{\left(\quad \text{other} \quad\right)}$$

For our purposes

$$\$/\text{die} = \frac{\$/\text{wafer}}{\underbrace{\#(\text{dies/wafer})}_{\left(\frac{A_{wafer}}{A_{die}}\right)}\ \underbrace{(\%\ \text{good dies})}_{\text{yield}}}$$

$$\text{yield} = \frac{1}{\left[1 + \left(\frac{\text{defects}}{cm^2}\right)\frac{A_{die}}{2}\right]^2}$$

FPGA



E.G.

$\$/\text{wafer} = \$1{,}500$ wafer size $= 200\text{ mm} \Rightarrow A = \pi r^2 \cong 3\times10^4\text{ cm}^2$

$\#(\text{defects}/\text{cm}^2) = 0.031$ die size $= (1\text{ cm})\times(1\text{ cm}) = 1\text{ cm}^2$

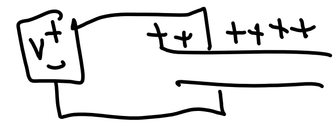
$$\text{yield} = 1/\left[1 + (0.031)A_{die}/2\right]^2 = 1/\left[1+(0.031)50\right]^2 = 1/\left[2.55\right]^2 = 1/2.4$$

$$\#\text{dies}_{good} = \left(\frac{A_{wafer}}{A_{die}}\right)\left(1/2.4\right) = \left(\frac{3\times10^4}{10^2}\right)\left(1/2.4\right) = \frac{300}{2.4} = 125$$

$$\$/\text{die}_{good} = \frac{\$/\text{wafer}}{\#(\text{dies}_{good})/\text{wafer}} = \frac{\$1{,}500}{125} = \$12/\text{die}$$

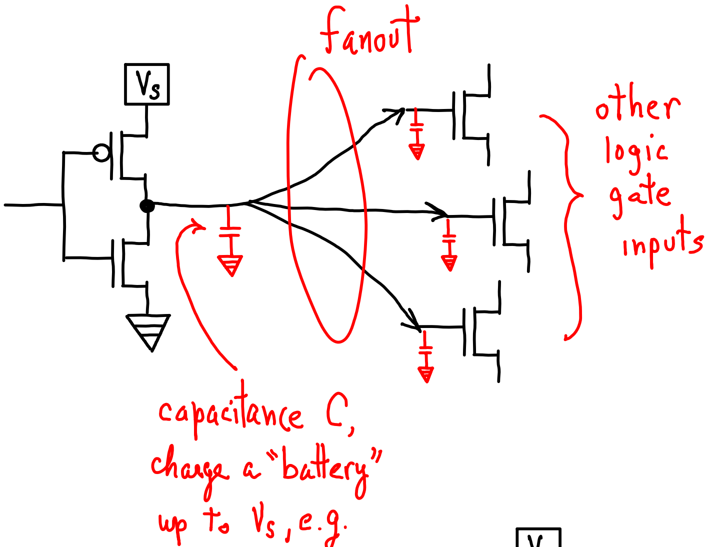
CMOS power and energy consumption

1. **Dynamic:** energy converted to heat due to switching a logic gate's output (0-1 or 1-0).
2. **Static:** energy converted to heat due to (steady) leakage currents.



Dynamic

$\boxed{V_s}$

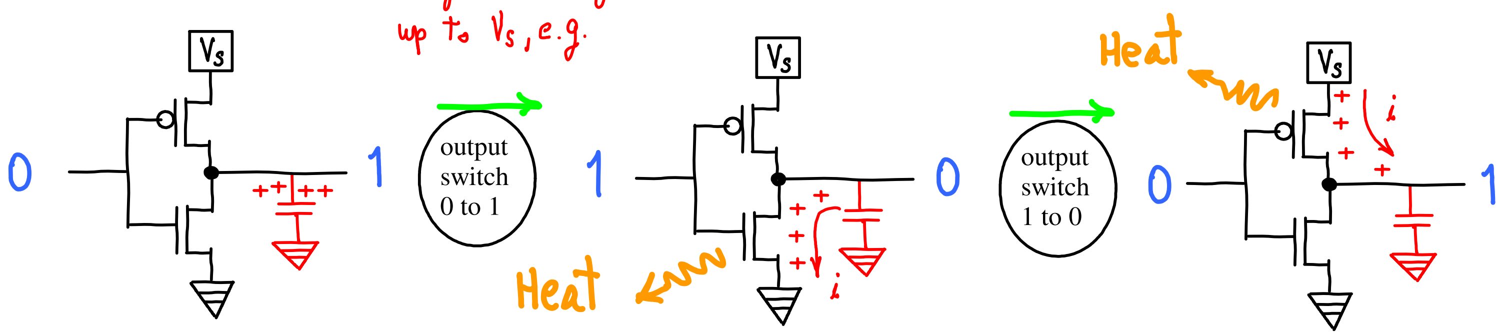


$CR_{max} \propto V$

Speed of charging C

$\longrightarrow E = V/d$

$e^- \longrightarrow$ acceleration $\sim$ gravity



$$\frac{\text{Joules}}{\text{sec}} = \text{power} = V\left(\frac{+}{\text{sec}}\right) = Vi = (iR)i$$

$R_{\text{Transistor}}$

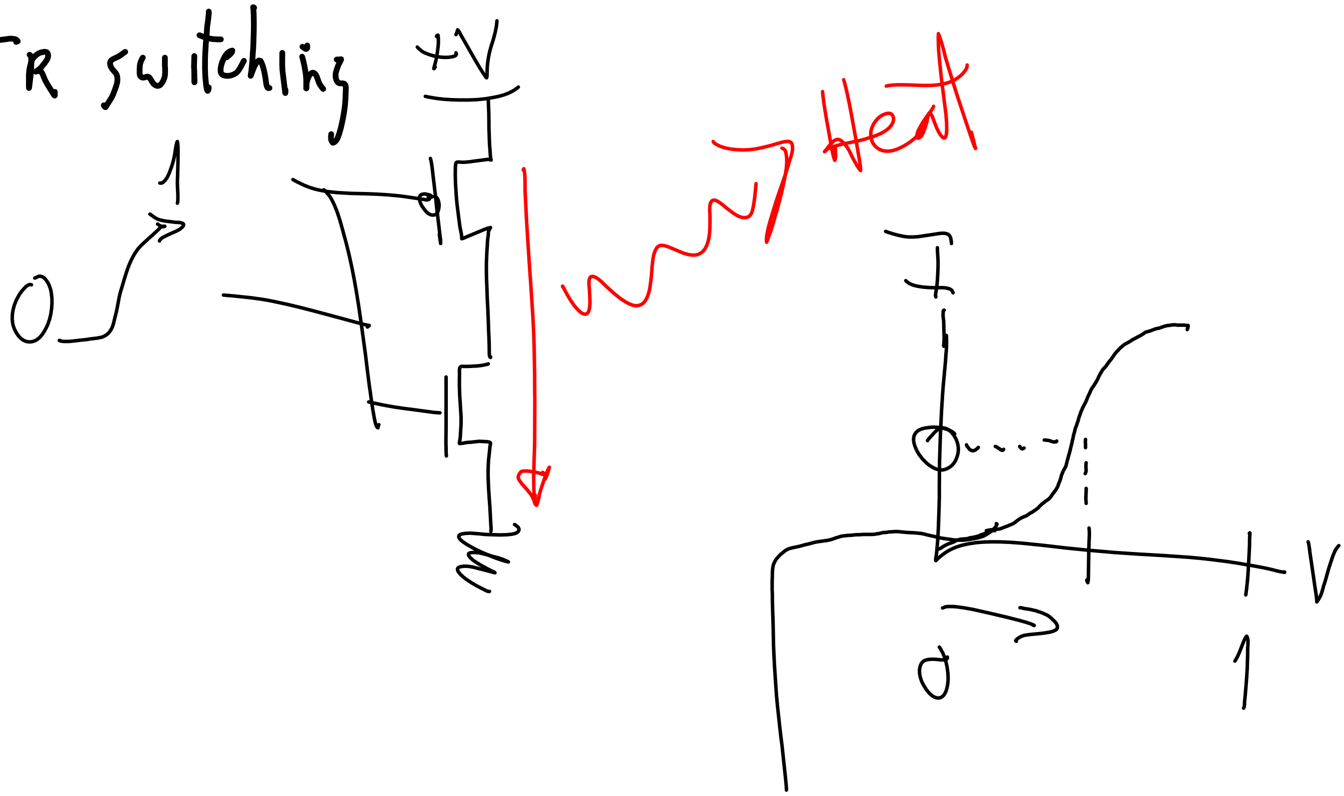
$$E = \frac{\text{Joules}}{\text{Sec}}(\Delta t \text{ sec}) = \frac{1}{2} C V^2$$

$$\Longrightarrow \frac{\left(E/\text{Transistor}\right)}{\text{0-1-switch}} = \frac{1}{2} C_{\text{Transistor}} V^2$$

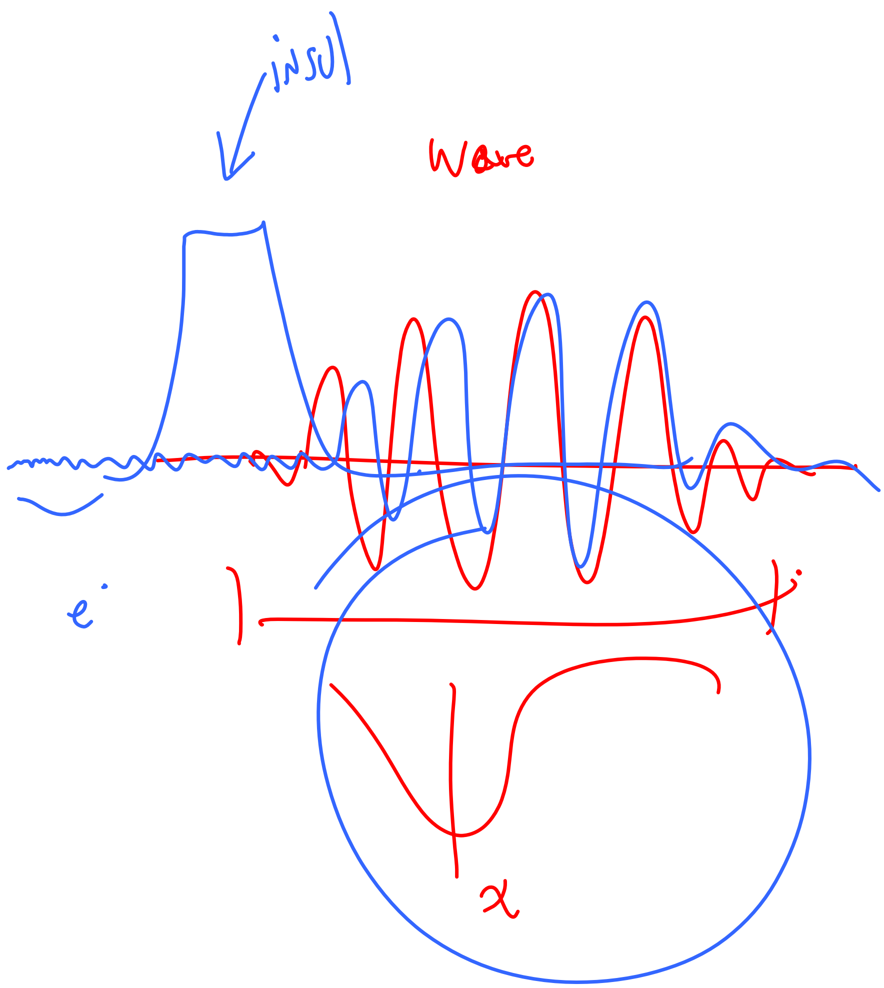
$$\Longrightarrow \frac{E_{\text{Total}}}{\text{Switch}} = \sum_{i}^{\#\text{ Transistor}} \frac{1}{2} C_i V^2 = \frac{1}{2} V^2 \sum_i C_i = \frac{1}{2} V^2 C_{\text{Total-chip}}$$

$$\Longrightarrow \text{Power} = \left(\frac{E_{\text{Total}}}{\text{Switch}}\right)\left(\frac{\text{switch}}{\text{sec}}\right) = \frac{E_{\text{Total}}}{\text{Switch}} CR$$

TR switching



QM waves



E.G. $C_{Total}^{new} = 0.85\ C_{Total}^{old}$

$V^{new} = 0.85\ V^{old}$

$$CR \propto V$$

$$\Rightarrow \frac{CR_{new}}{CR_{old}} = \frac{kV_{new}}{kV_{old}} = \frac{0.85\ V_{old}}{V_{old}}$$

$$\Rightarrow CR_{new} = 0.85\ CR_{old}$$

$$\Rightarrow \frac{Power_{new}}{Power_{old}} = \frac{(1/2)\ C_{Total}^{new}\ V_{new}^2\ CR_{new}}{(1/2)\ C_{Total}^{old}\ V_{old}^2\ CR_{old}}$$

$$= \frac{(0.85\ C_{Total}^{old})(0.85\ V_{old})^2(0.85\ CR_{old})}{C_{Total}^{old}\ V_{old}^2\ CR_{old}}$$

$$= (0.85)^4 = 52\%$$

$$\frac{E_{new}}{E_{old}} = \frac{k_{switches}\ E^{new}/switch}{k_{switches}\ E^{old}/switch} = \frac{k\ \frac{1}{2}C^{new}\ V_{new}^2}{k\ \frac{1}{2}C^{old}\ V_{old}^2} = \frac{(0.85\ C^{old})(0.85\ V_{old})^2}{C^{old}\ V_{old}^2}$$

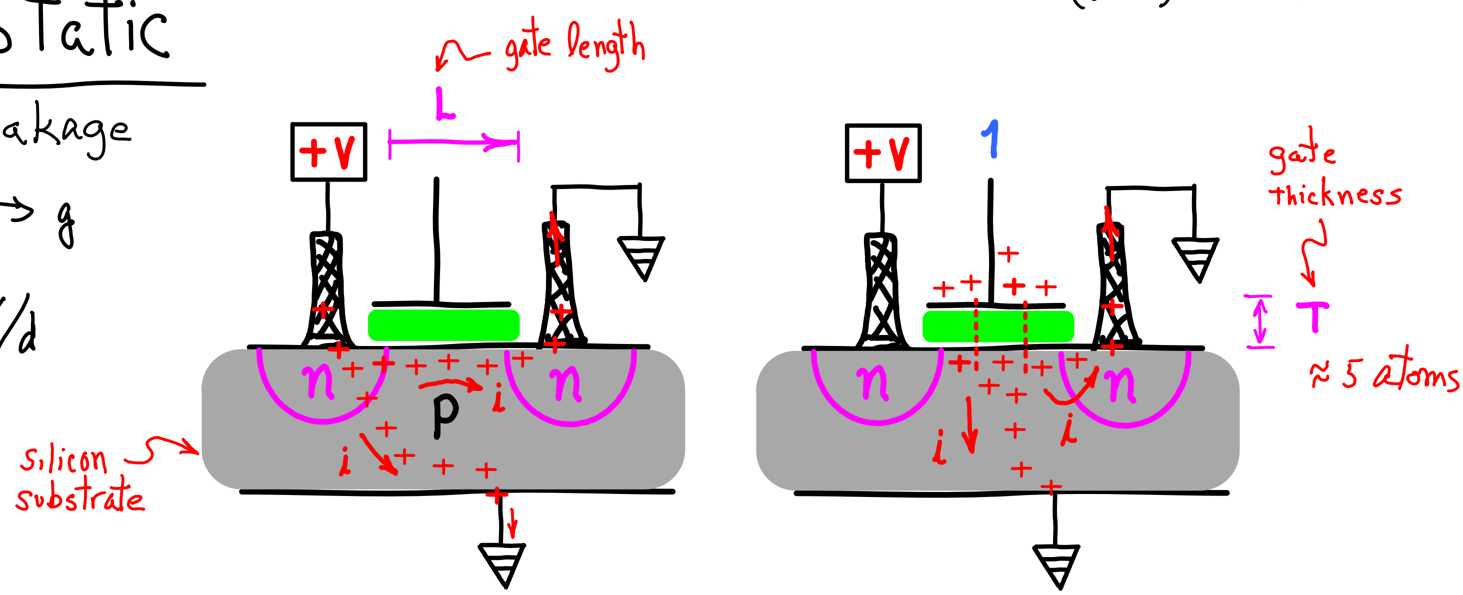
$$= (0.85)^3 = 61\%$$

Static

leakage

$\vec{E}$ g

$\approx V/d$



$i/\text{Area} \uparrow$ as $L, T \downarrow \Rightarrow Power_{leak} = i_{leak} V$

now greater than dynamic power

$$\frac{\text{STATIC POWER}}{\text{Total Power}}$$

year	1985	1990	1995	2000	2005	2010
	1%	5%	7%	20%	30%	60%

$r_1 \quad r_2 \quad r_3 \quad r_4 \quad r_5$

rate of change

Geometric Mean

what's the average rate?

$\Longrightarrow$ Find $\bar{r}$ s.t. $(r_1 \cdot r_2 \cdot r_3 \cdot r_4 \cdot r_5) = (\bar{r}\ \bar{r}\ \bar{r}\ \bar{r}\ \bar{r})$

What rate would give the same overall change over the entire time period?

$r_1 = 5/1 \quad r_2 = 7/5 \quad r_3 = 20/7 \quad r_4 = 30/20 \quad r_5 = 60/30$

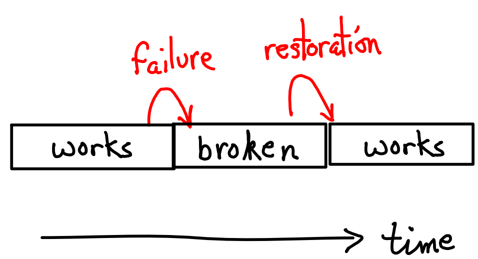
$\bar{r} = \left(\prod^{n} r_i\right)^{1/n} = (60/1)^{5} \cong 2.3$ per 5 years

Prediction for 2015? $\bar{r}(60\%) = 2.3(60\%) = 1.38\%$

Static power exceeds dynamic by 38%!

Failures, systemic, hierarchic

Service Accomplishment = it works
Service Interruption = broken



Reliability = time to failure

Mean Time To Failure (MTTF) = Avg(Reliability) = Avg(time working)
Mean Time To Repair (MTTR) = Avg(time to fix)

Failure Rate $= 1/\text{MTTF}$ (expected #failures in a year, e.g.)

Mean Time Between Failures $(\text{MTBF}) = \text{MTTF} + \text{MTTR}$

Availability $= \%$ time working $= \dfrac{\text{MTTF}}{\text{MTBF}}$

Multi-level Recovery

- error correction
- redo
- redundancy

Disk sub-system	MTTF (M hr)
10 disks	1
1 ATA controller	1/2
1 power supply	1/5
1 fan	1/5
1 ATA cable	1

Assume: exponentially dist. failures

- independent of t, time of experiment
- No dependency between components

$E(\#\text{ Failures in } \Delta T) = \mu \, \Delta T$ ← avg failure rate

1 failure → $1 = \mu \, \overline{\Delta T_1}$ ← avg time to 1st fail

$\text{MTTF} = \overline{\Delta T_1} = 1/\mu$

$$\mu = \sum_i \mu_i = \sum_i \left(1/\text{MTTF}_i\right)$$

Assume devices fail independently:
P(F1 or F2) = P(F1) + P(F2)

$$= (10 + 2 + 5 + 5 + 1)/\text{Mhr} = 23/\text{Mhr}$$

$$\text{MTTF} = 1/\mu = \frac{10^6 \text{ hr}}{23}$$

$$\cong \left(\frac{10^2}{25}\right) 10^4 = 40{,}000 \text{ hr}$$

$$\cong 4 \text{ yr.}$$

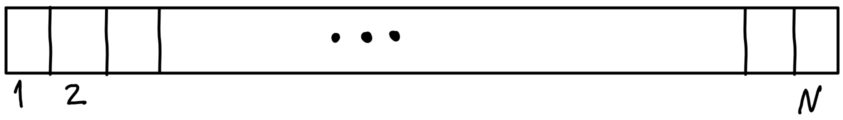
hrs/yr $= (365\text{ d})(24 \text{ hr/d})$

$\cong (400)(100/4)$

$= 10{,}000$

E.g., 2 power supplies?

Let's add a redundant power supply: if one fails we fix it while the other keeps working. How long before the system fails: it happens that before we finish the repair, the backup also breaks?



A Model

Suppose N time periods, each of unit length. N is large so that no devices every survive to age N+1.

Each time period, flip a coin:

Prob(works) = p
Prob(fails) = (1 - p)

fails at $i+1$

$$\text{Prob}(\text{fails at } i) = p^i(1-p)$$

$$E(i) = \sum_{k=1}^{N} k\,\text{Prob}(\text{fails at } k)$$

$$= \sum_{k=1}^{N} k\,p^k(1-p) = (1-p)\sum_{k=1}^{N} k\,p^k \qquad p < 1$$

$$= (1-p)\sum_{k=1}^{N} \frac{d}{dp}p^{k+1} = (1-p)\,p\,\frac{d}{dp}\sum_{k=1}^{N} p^k$$

$$\cong (1-p)\,p\,\frac{d}{dp}\,1/(1-p) = (1-p)\,p/(1-p)^2$$

$$= p/(1-p) = \text{MTTF}$$

Example

Suppose $\text{Prob}(\text{fail}) = (1-p) = 1/2^m$

$$\text{MTTF} = p/(1-p) = (1 - 1/2^m)/(1/2^m) = 2^m - 1$$

2 Power Supplies

$$\text{Prob}(\text{both work}) = p^2 \implies \text{Prob}(\text{not both working}) = (1-p^2)$$

$$\text{MTTF}_2 = p^2/(1-p^2) = p^2/(1-p)(1+p) = \frac{p}{(1-p)}\frac{p}{(1+p)}$$

$$= \text{MTTF}\,\frac{p}{(1+p)} \Big\} \text{ for } p \text{ large } (p \cong 1) \;\cong\; \tfrac{1}{2}\,\text{MTTF}$$

What's the probability the 2^{nd} fails in MTTR?

Time independent = begin as if only one power supply.

Looked at another way

Total time $= T$

N intervals, $dt = T/N$

$\text{Prob}(\text{fail in } dt) = 1/N$

$= 1 - p$

$\text{Prob}(\text{fails in interval of length } L = n\,dt)$

$$\sum_i^n \text{Prob}(\text{fail in } dt_i) = \sum^n (1/N) = n/N$$

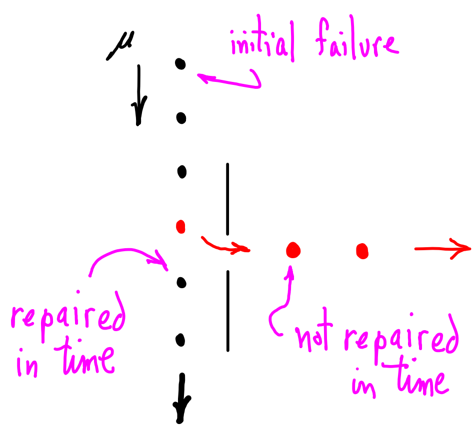
$$= \frac{L/dt}{T/dt} = \frac{L}{T} \Longrightarrow \text{uniform dist.}$$

$\text{Prob}(\text{fail in } T) = 1$

$\text{MTTF} = T/2$

$$\Longrightarrow \text{Prob}(\text{fail in MTTR}) = \frac{\text{MTTR}}{T} = \frac{\text{MTTR}}{2\,\text{MTTF}}$$

$1/\tfrac{1}{2}\text{MTTF} = \mu$, average rate of failures for 2 power supplies



$$\left\{\begin{array}{l}\text{average \% failed to} \\ \text{repair in time} \\ \text{before } 2^{nd} \text{ fail}\end{array}\right\} = \text{Prob}(\text{fail to repair}) = \frac{\text{MTTR}}{2\,\text{MTTF}}$$

$$\text{rate of system failures} = \mu(\%\text{ not repaired})$$

$$= 1/(\tfrac{1}{2}\text{MTTF})\left(\frac{\text{MTTR}}{2\,\text{MTTF}}\right) = \frac{\text{MTTR}}{\text{MTTF}^2}$$

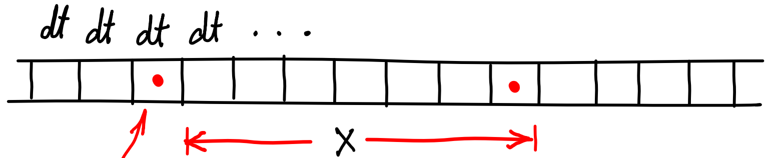
$$\Longrightarrow \frac{\text{MTTR}}{\text{MTTF}^2} = \frac{24}{(\tfrac{1}{5}M)^2} \cong \frac{25^2}{10^{12}} = \frac{(100/4)^2}{10^{12}} = \frac{10^4}{16\;10^{12}} = \frac{1}{16\;10^8}$$

$$= \frac{1 \quad \text{system failure}}{16\;10^4 \quad \text{machine yrs}}$$

$\Longrightarrow$ on average, w/ 10^4 systems, 1 fail in 16 yrs.

Poisson

Bernoulli w/ $p = \underline{\lambda\, dt}$



Probability that a failure happens in $[t, t+dt]$ $\Big\}$ flip coin w/ Prob p

inter-arrival time

$$\text{Prob}(m, N, p) = \binom{N}{m} p^m q^{N-m}$$

approx. $\Longrightarrow \dfrac{e^{-pN} (pN)^m}{m!}$

let $pN \xrightarrow[p \to 0,\, N \to \infty]{} \lambda$

Prob

$$\text{Prob}(m, \lambda) = \frac{e^{-\lambda t} (\lambda t)^m}{m!}, \quad \lambda = \text{rate} \qquad \text{Prob}(k \text{ events in time } t)$$

$$\text{Prob}(X_n \le x) = 1 - e^{-\lambda x} \qquad \text{Prob}(\text{time of next arrival} \le x)$$

$E(X) \equiv 1/\lambda$ def'n of λ

$$\text{Prob}(\text{arrival time } t \le s \mid \text{one event did occur})$$
$$= s/t \quad \text{uniform dist.}$$

Price Performance

Basic measure : operations / $

least power

server-side java ops/sec

$

benchmark

ops/$

	System 1		System 2		System 3	
Component		Cost (% Cost)		Cost (% Cost)		Cost (% Cost)
Base server	PowerEdge R710	$653 (7%)	PowerEdge R815	$1437 (15%)	PowerEdge R815	$1437 (11%)
Power supply	570 W		1100 W		1100 W	
Processor	Xeon X5670	$3738 (40%)	Opteron 6174	$2679 (29%)	Opteron 6174	$5358 (42%)
Clock rate	2.93 GHz		2.20 GHz		2.20 GHz	
Total cores	12		24		48	
Sockets	2		2		4	
Cores/socket	6		12		12	
DRAM	12 GB	$484 (5%)	16 GB	$693 (7%)	32 GB	$1386 (11%)
Ethernet Inter.	Dual 1-Gbit	$199 (2%)	Dual 1-Gbit	$199 (2%)	Dual 1-Gbit	$199 (2%)
Disk	50 GB SSD	$1279 (14%)	50 GB SSD	$1279 (14%)	50 GB SSD	$1279 (10%)
Windows OS		$2999 (32%)		$2999 (33%)		$2999 (24%)
Total		$9352 (100%)		$9286 (100%)		$12,658 (100%)
Max ssj_ops	910,978		926,676		1,840,450	
Max ssj_ops/$	97		100		145	

Figure 1.18 Three Dell PowerEdge servers being measured and their prices as of August 2010. We calculated the cost of the processors by subtracting the cost of a second processor. Similarly, we calculated the overall cost of memory by seeing what the cost of extra memory was. Hence, the base cost of the server is adjusted by removing the estimated cost of the default processor and memory. Chapter 5 describes how these multi-socket systems are connected together.

ops/watt vs load

avg utilization

≈ 10-50%

what's the message?

cpu as % performance? % cost?

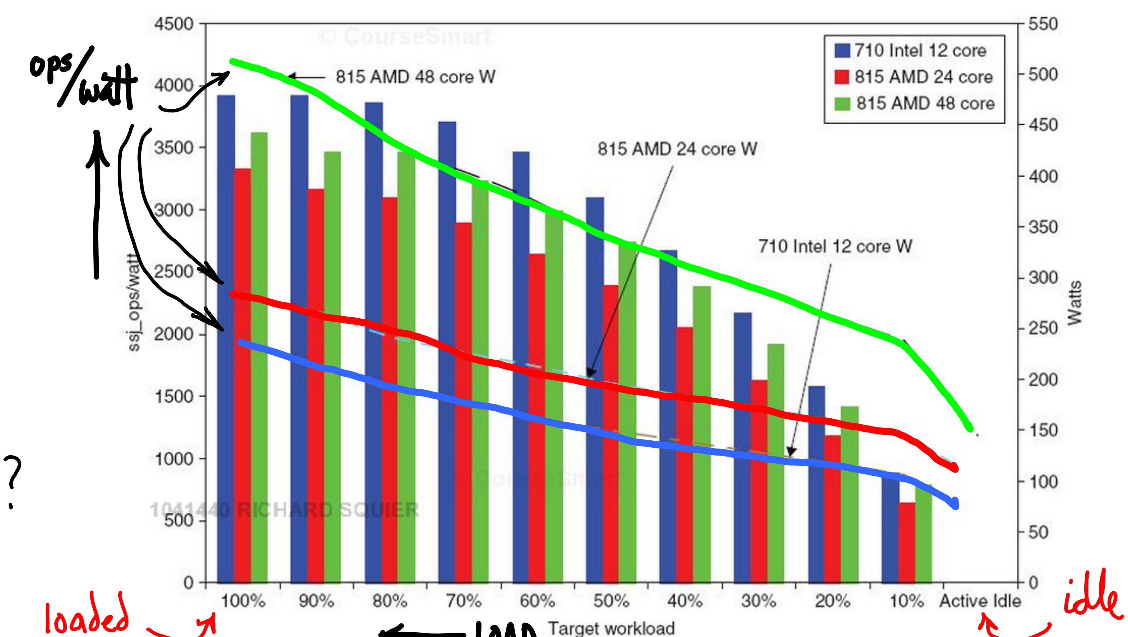


Figure 1.19 Power-performance of the three servers in Figure 1.18. Ssj_ops/watt values are on the left axis, with the three columns associated with it, and watts are on the right axis, with the three lines associated with it. The horizontal axis shows the target workload, as it varies from 100% to Active Idle. The Intel-based R715 has the best ssj_ops/watt at each workload level, and it also consumes the lowest power at each level.