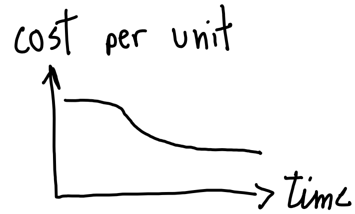
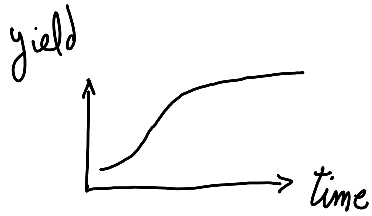
Cost



Manufacturing costs drop as expertise grows, for that process

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



Yield = 1 - waste

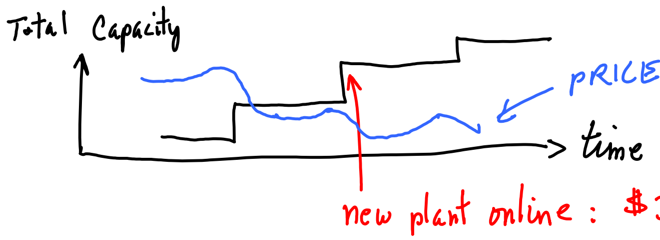
-- #(devices sellable) versus #(devices produced)

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

E.g. DRAM ⇒ price = $\propto$ cost

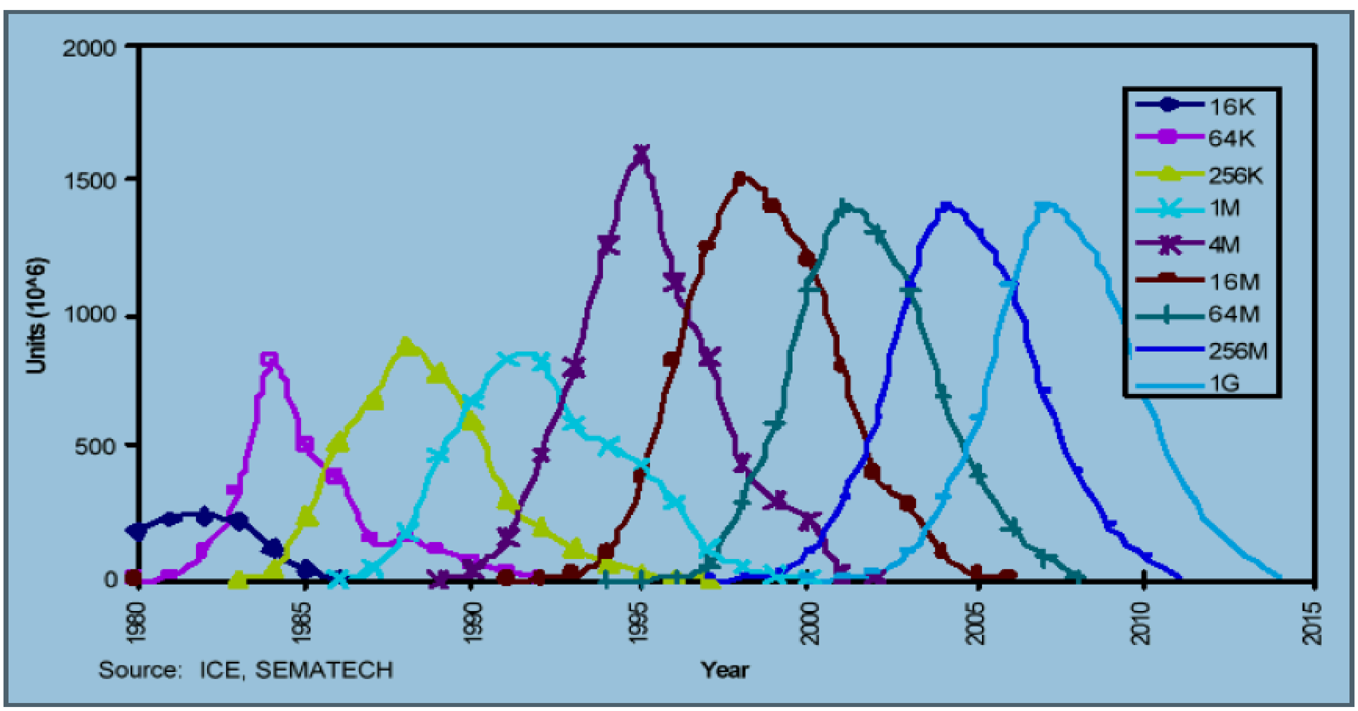
80% contract sales to large Equipment makers (hidden)

20% open market



Commodity market

--- lots of vendors
--- selling 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

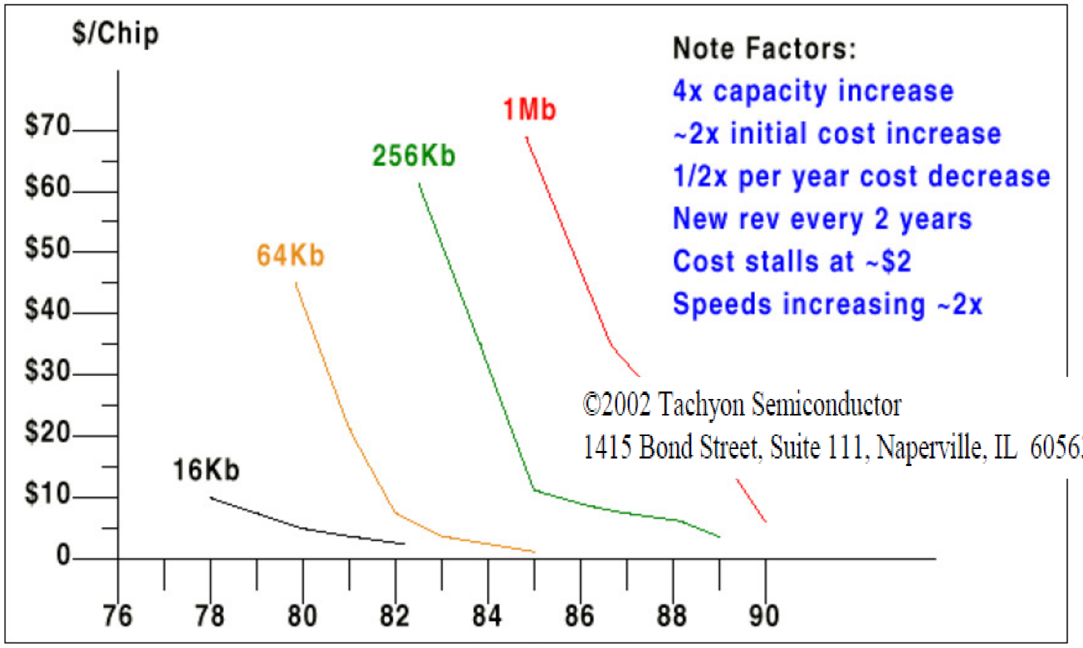
4x capacity

2x speed

Cost per bit

per bit/sec

Drop faster



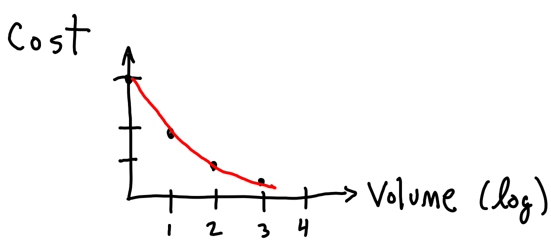
DRAM Costs [45]

Changes   Telecom (routers/switches)   20% ↑ ⇒ 50%

latency ↓ vs bandwidth ↑

(PCs)

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



½x Cost : 2x Volume

Volume → supplier competition → lower Cost

⇒ Low-end Market   (Price/Performance) ↓

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

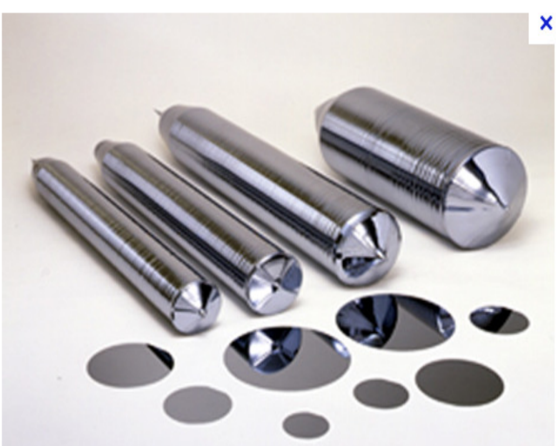
Cloud Pricing | AWS   Combined efficiencies

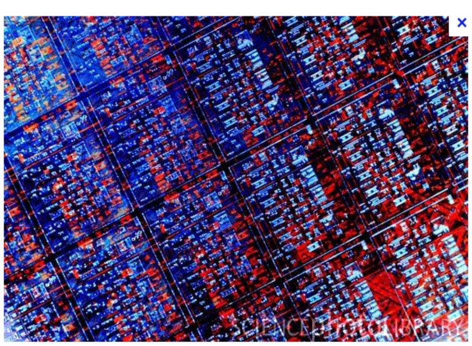
Description	Type	CU	Original $ / CU / Hour	Current $ / CU / Hour	% Reduction	Aug 2006	Oct 2007	May 2008	Oct 2009	Feb 2010	July 2010	Sep 2010	Nov 2010	Nov 2011
Small - "the original"	m1.small	1	$0.10	$0.085	15%	$0.10			$0.09					
Large	m1.large	4	$0.10	$0.085	15%		$0.40		$0.34					
Extra Large	m1.xlarge	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

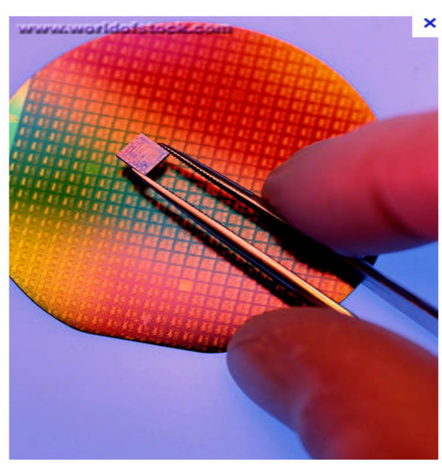
CPUs, Chips, SOC

Si ingots   slicing → wafers   masking, etching, doping

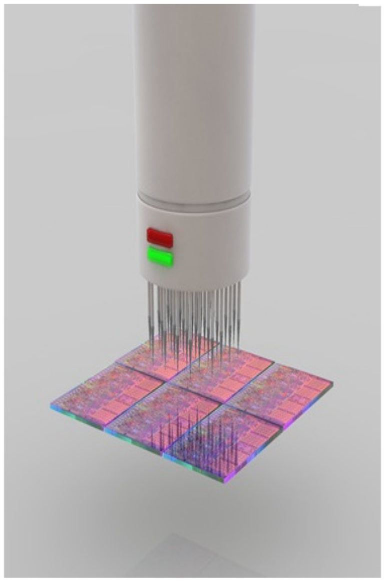


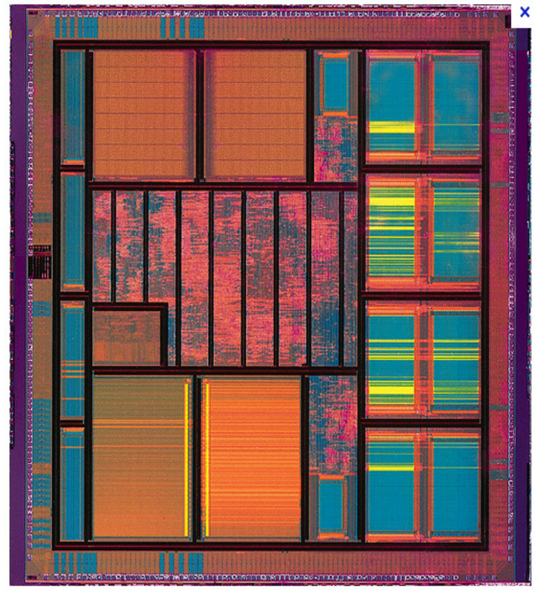


dicing →



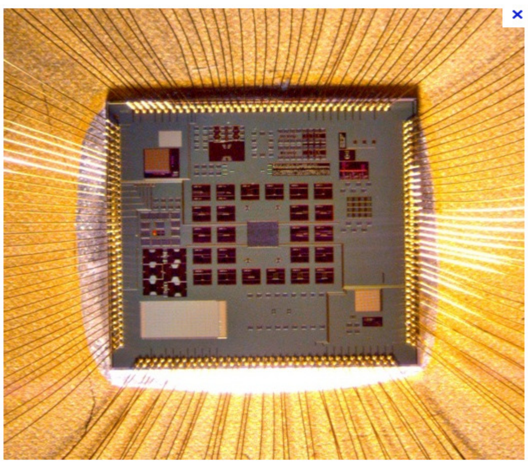
Circuit testing

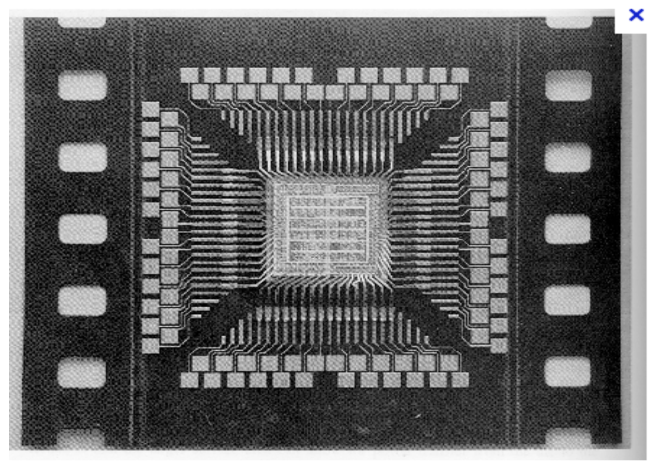




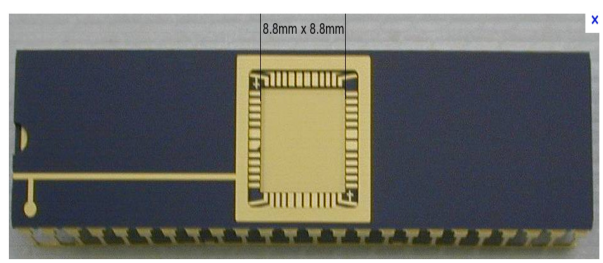
Pad Bonding

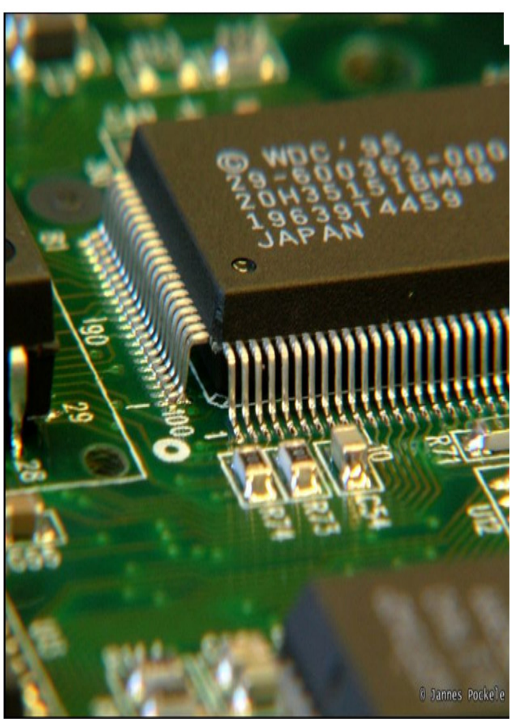
Pin Packaging





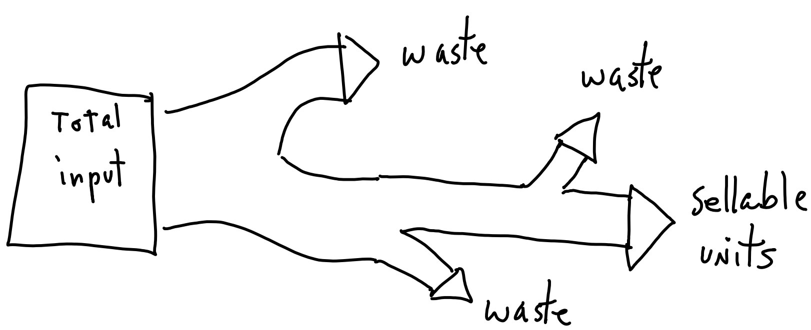
encasing



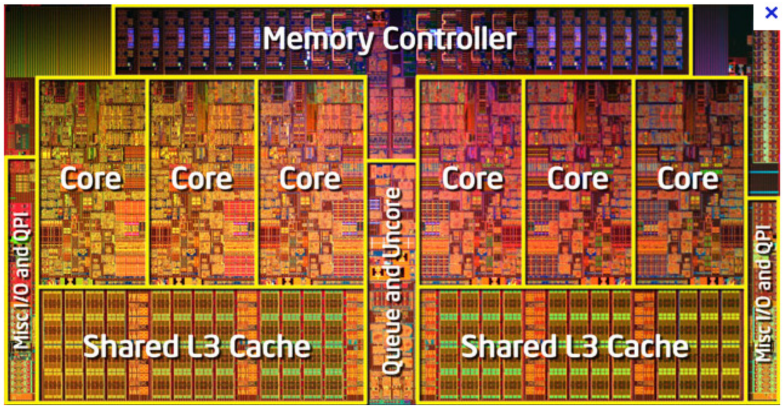


board printing

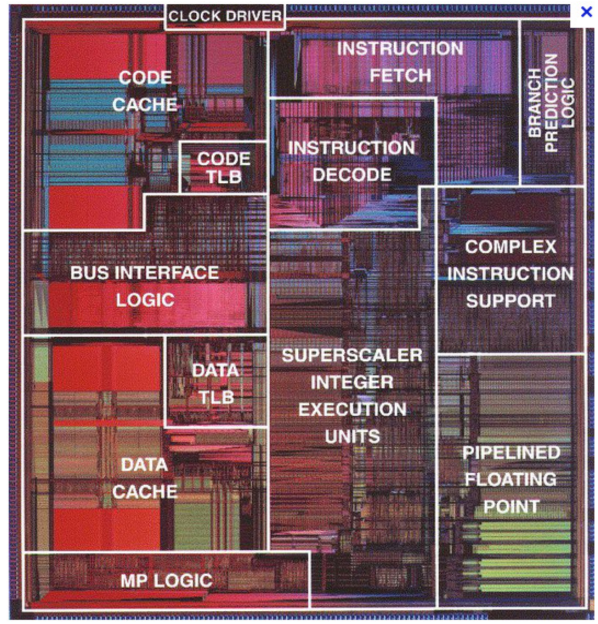
mounting



what's inside?



i7



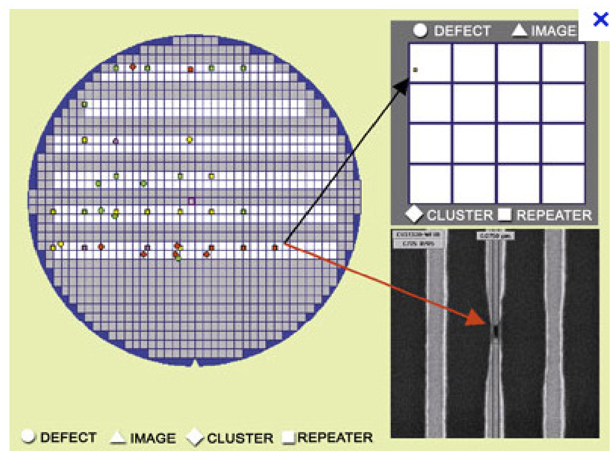
P5

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

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

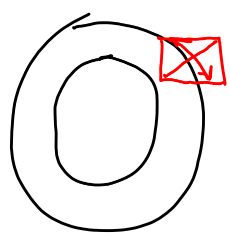
$C_{wafer} \approx \$5000$

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



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

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



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

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

$$\frac{\#(\text{defects})}{cm^2} \approx 0.04$$ ← function of time + volume

300 mm Wafer

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

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

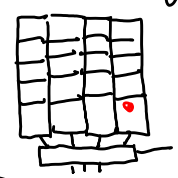
P5 Sandy Bridge	embedded CPU, 32b	printer Controller
2 cm^2	0.1 cm^2	0.04 cm^2
\$50	\$13	\$0.1

Amortized Costs

Mask = \$1M

⇒ reconfigurable gate arrays

Redundancy, e.g.



Select 3 banks
1 spare

Die size = #(Transistors) + #(pins) ↑
\+
Volume ↓
\+
Customization ↑

} cost

Warehouse-Scale Costs

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

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

For our purposes

$$\$/\text{die} = \frac{\$/\text{wafer}}{\#(\text{dies}/\text{wafer})\ (\%\ \text{good dies})}$$

$\#(\text{dies}/\text{wafer})$ → $\left(\frac{A_{wafer}}{A_{die}}\right)$

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

E.G.

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

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

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

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

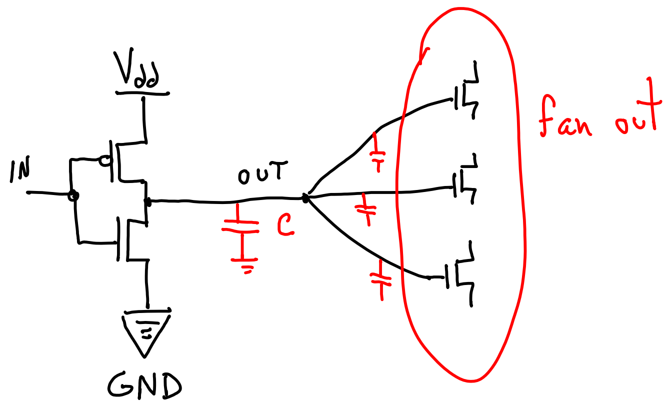
$$\$/\text{die}_{good} = \frac{\$/\text{wafer}}{\#(\text{dies}_{good})/\text{wafer}} = \frac{\$1{,}500}{125} = \$12/\text{die}$$

CMOS power and energy consumption

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

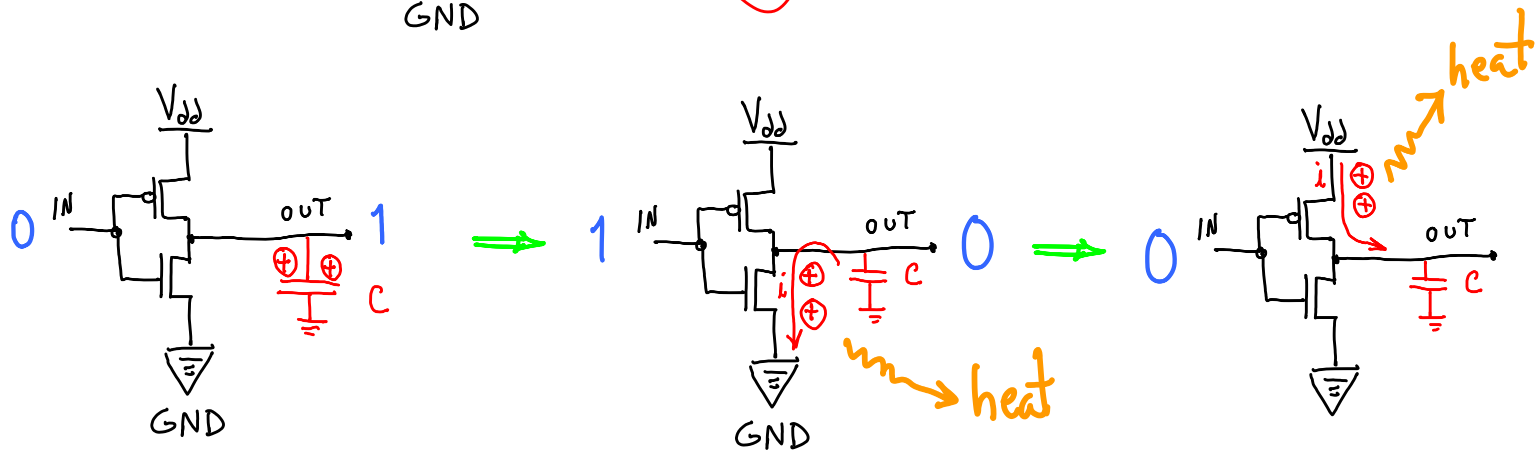
2. **Static:** energy converted to heat due to (steady) leakage currents.

Dynamic



$CR_{max} \propto V$

Speed of charging C



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

$R_{\text{Transistor}}$

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

$\Rightarrow \frac{(E/\text{Transistor})}{\text{Switch}} = \frac{1}{2} C_{\text{Transistor}} V^2$

$\Rightarrow \frac{E_{\text{Total}}}{\text{Switch}} = \sum_{i}^{\#\text{ Transistor}} \frac{1}{2} C_i V^2 = \frac{1}{2} V^2 \sum C_i = \frac{1}{2} V^2 C_{\text{Total}}$

$\Rightarrow \text{Power} = \left(E/\text{switch}\right)\left(\frac{\text{switch}}{\text{sec}}\right) = E/\text{switch} \cdot CR$

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

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

$CR \propto V$

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

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

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

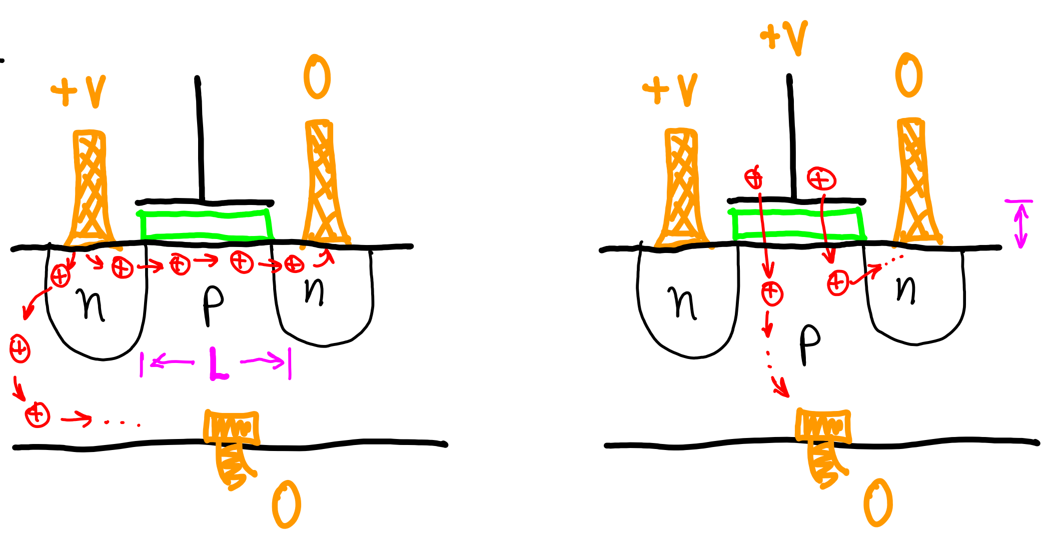
$$= \frac{(0.85\ C_{Total}^{old})(0.85\ V_{old})^2(0.85\ CR_{old})}{C_{Total}^{old}\ V_{old}\ CR_{old}}$$

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

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

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

Static



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

E.G.

GM

year	1985	1990	1995	2000	2005	2010
$\frac{\text{STATIC POWER}}{\text{Total Power}}$	1%	5%	7%	20%	30%	60%

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

? Ratio from year-to-year?

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

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

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

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