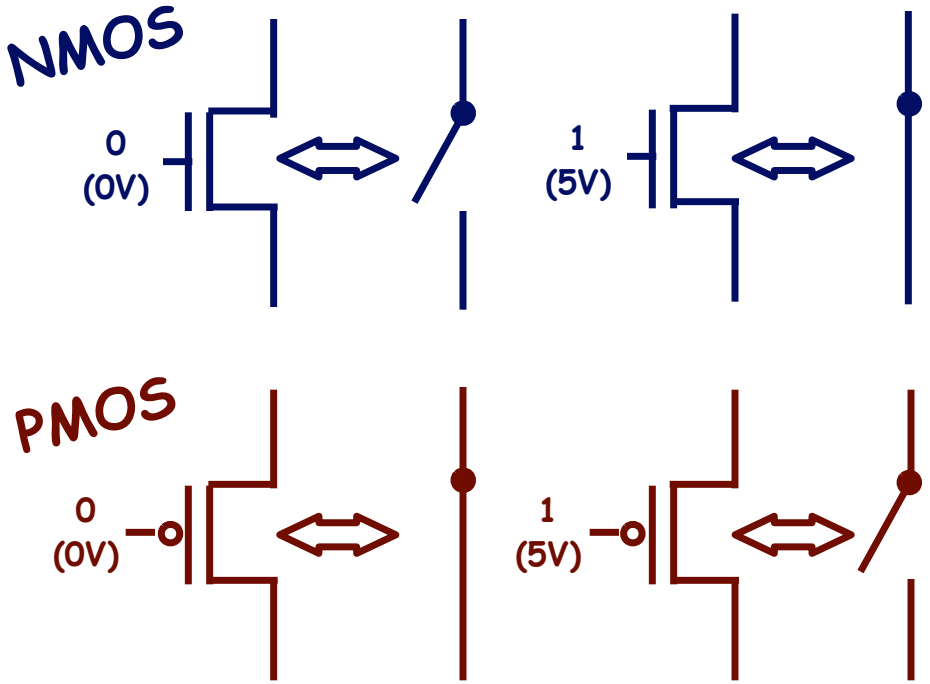
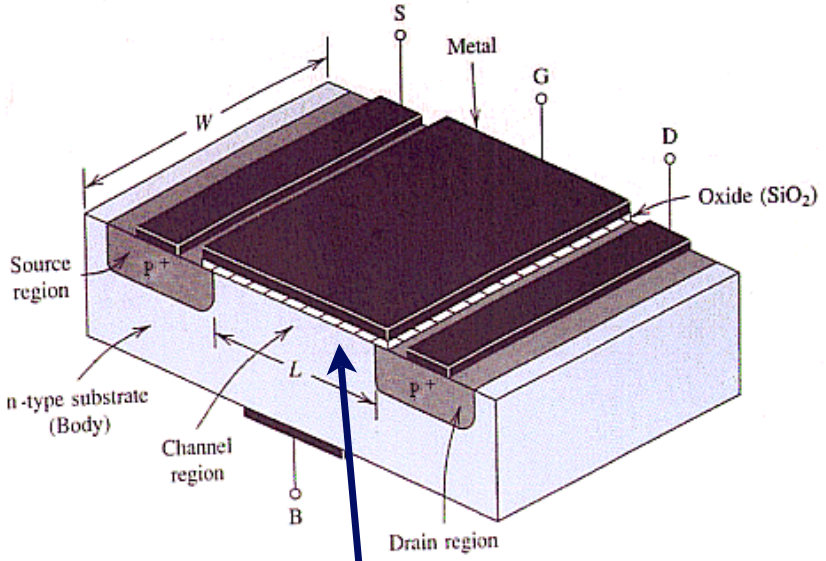
Transistors used to manipulate/store 1s & 0s

Switch-level representation



Cross-sectional view



(can act as a capacitor storing charge)

Using above diagrams as context, note that (with NMOS) if we (i) apply a suitable voltage to the gate & (ii) then apply a suitable voltage between source and drain, current will flow.

A few important relationships...

- One way to think about switching time:
  - Charge is carried by electrons
  - Carrier velocity is proportional to the lateral E-field between source and drain
    - i.e. $v = mE$
      - m = carrier mobility (and can be though of as a constant)
  - Electric field defined as: $E = V_{ds}/L$
  - Time for charge to cross channel = length/speed
    - (i.e. meters / (meters/s) = seconds)
    - $= L/v$
    - $= L/(mE)$
    - $= L/(m*(V_{ds}/L))$
    - $= L^2/(mV_{ds})$

Thus, to make a device faster, we want to either increase V_{ds} or decrease feature sizes (i.e. L)

A few important relationships

- What about power?
  - First, need to quickly discuss equation for capacitance:
    - $C_L = (e_{ox}WL)/d$
      - e_{ox} = dielectric, WL = parallel plate area, d = distance between gate and substrate
  - Then, dynamic power becomes:
    - $P_{dyn} = C_L V_{dd}^2 f_{0-1}$
      - Dynamic power is a function of the frequency of 0 to 1 or 1 to 0 transitions (as this involves the movement of charge)
        - » Note frequency in this context is NOT clock frequency
      - Note that as W and L scale, C_L decreases which in turn will cause a decrease in P_{dyn}.
      - Note that while an increase in V_{dd} will *decrease* switching time, it will also cause a quadratic *increase* in dynamic power.

A few important relationships

Parameter	Relation	Full	General
W, L, t_{ox}		1/S	1/S
V_{dd}, V_t		1/S	1/U
N_{SUB}	V/W_{depl}^2	S	S^2/U
Area/device	WL	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S
C_{gate}	$C_{ox}WL$	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S
I_{sat}	$C_{ox}WV$	1/S	1/U
Current Density	I_{sat}/Area	S	S^2/U
R_{on}	V/I_{sat}	1	1
Intrinsic Delay	$R_{on}C_{gate}$	1/S	1/S
P	$I_{sat}V$	$1/S^2$	$1/U^2$
Power Density	P/Area	1	S^2/U^2

In theory, as scaling continues, devices should get faster and power density should stay the same.

But will it?

Moore's Law

- "Cramming more components onto integrated circuits."

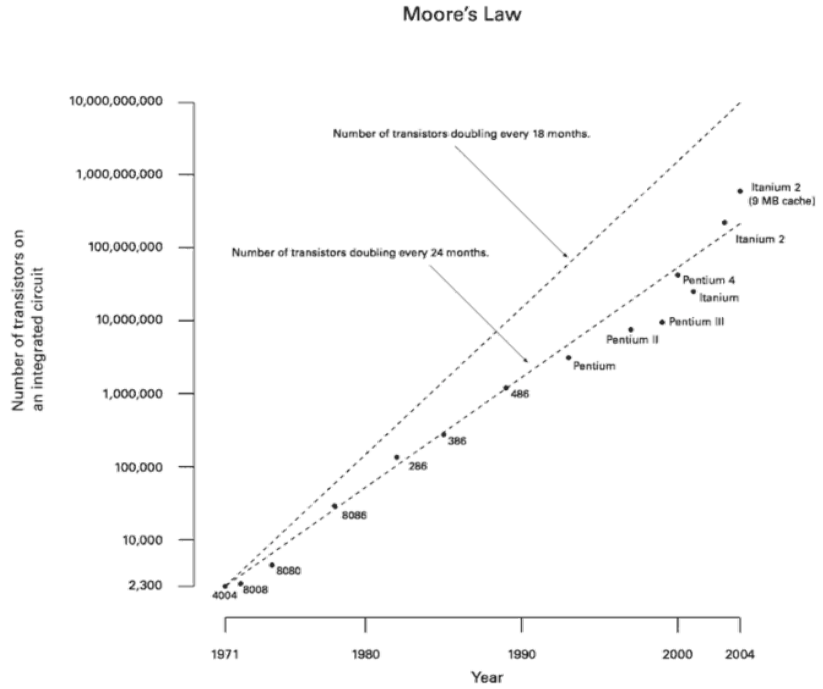
  - G.E. Moore, Electronics 1965

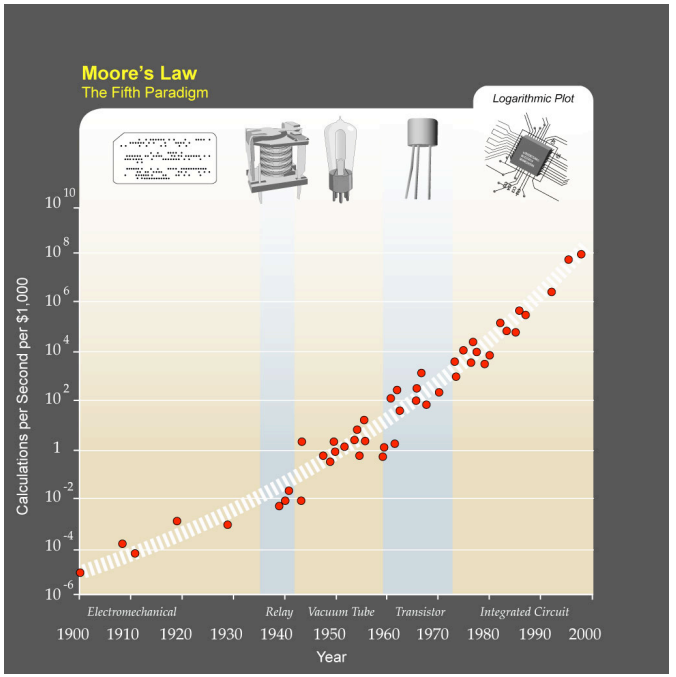
  - Observation: DRAM transistor density doubles annually
    - Became known as "Moore's Law"
    - Actually, a bit off:
      - Density doubles every 18 months
      - (in 1965 they only had 4 data points!)
  - Corollaries:
    - Cost per transistor halves annually (18 months)
    - Power per transistor decreases with scaling
    - Speed increases with scaling
    - Reliability increases with scaling
      - Of course, it depends on how small you try to make things
        - » (I.e. no exponential lasts forever)

Remember these!

Moore's Law

- "Performance doubles every 18 months"
  - This is a common interpretation of Moore's Law - but not the original intent
    - Actually, performance doubles ~ every 2 years
- Here are 2 nice pictures...





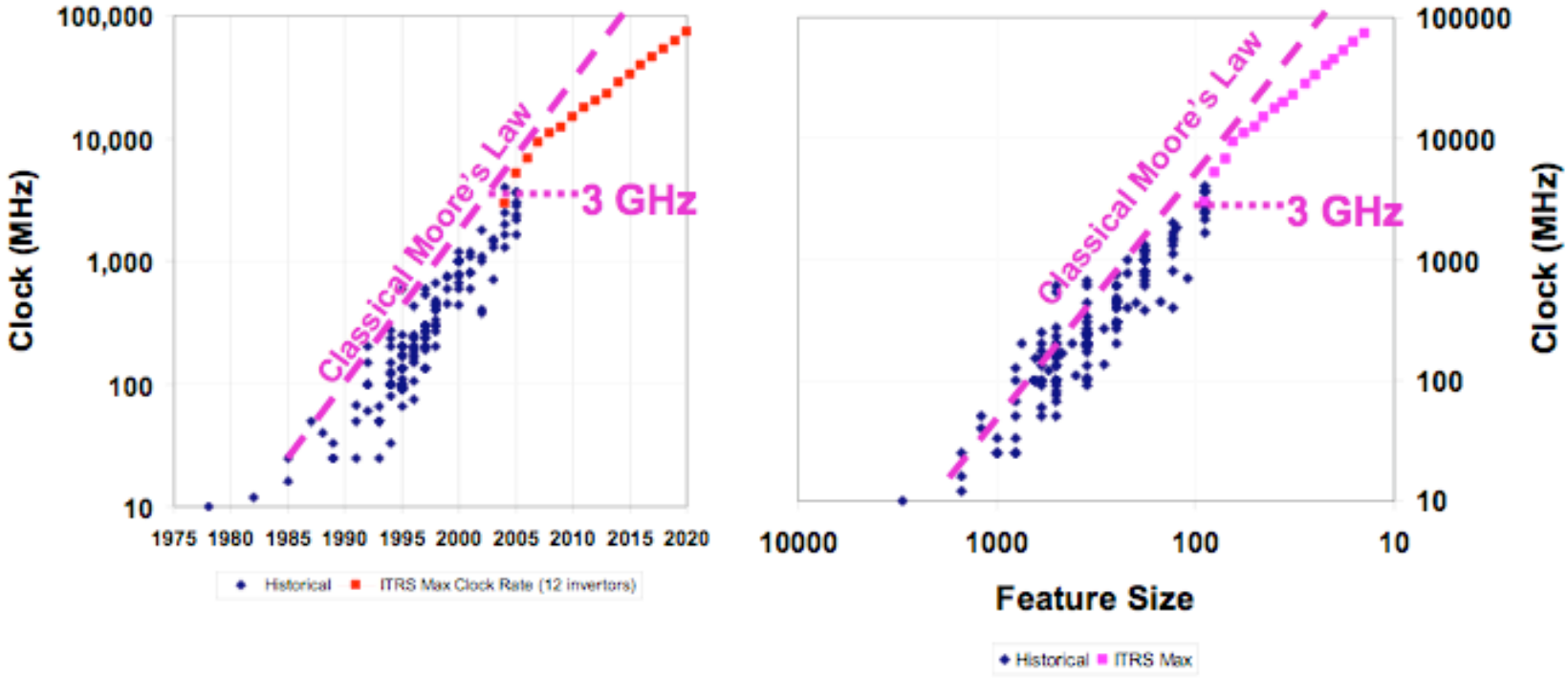
Moore's Law (continued)

- Moore's Curve is a self-fulfilling prophecy
  - 2X every 2 years means ~3% per month
    - I.e. ((1 X 1.03) * 1.03)*1.03… 24 times = ~2
  - Can use 3% per month to judge performance features
  - If feature adds 9 months to schedule…it should add at least 30% to performance
    - ($1.03^9 = 1.30 \Rightarrow 30\%$)

A funny thing happened on the way to 45 nm

- Speed increases with scaling...

Remember these!

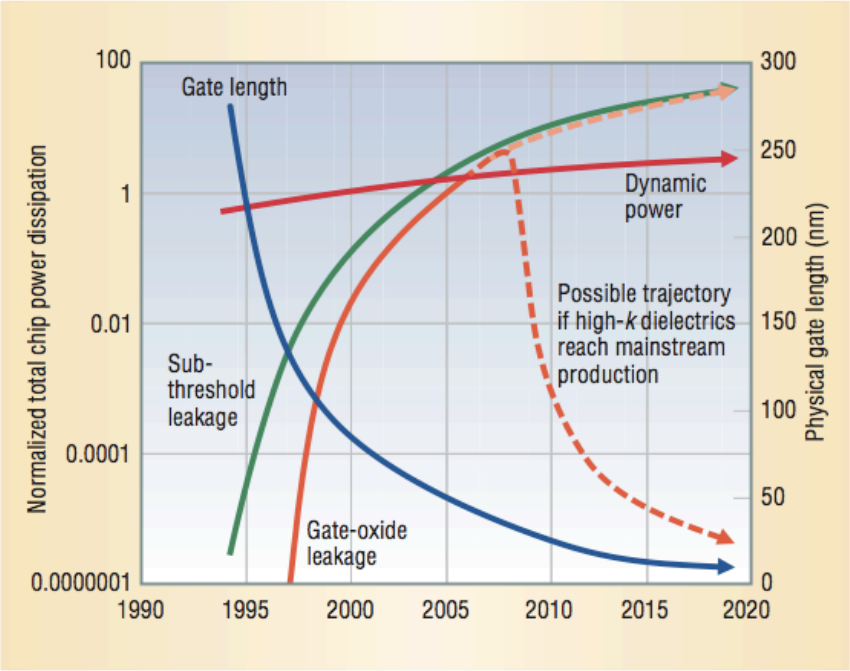


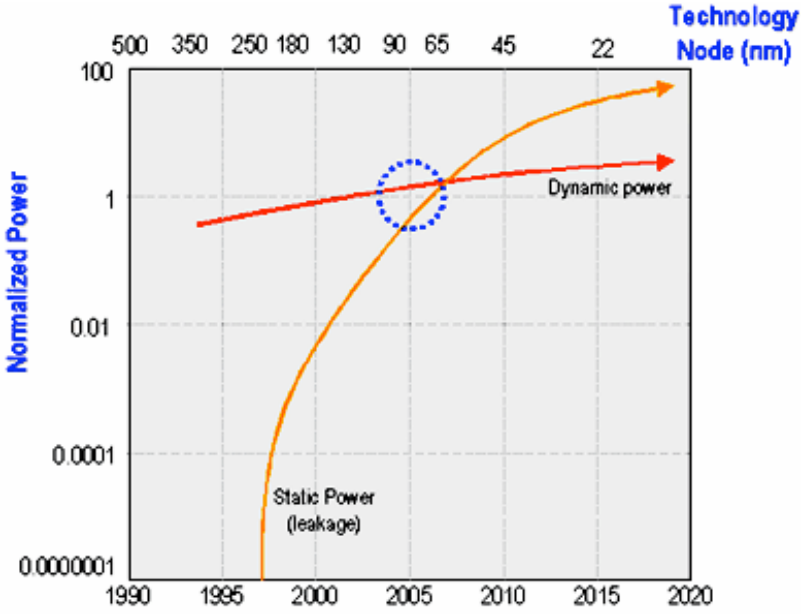
2005 projection was for 5.2 GHz - and we didn't make it in production. Further, we're still stuck at 3+ GHz in production.

A funny thing happened on the way to 45 nm

- Power decreases with scaling...

Remember these!





A funny thing happened on the way to 45 nm

- Power decreases with scaling...

Remember these!

This is not the trend that you want to see:

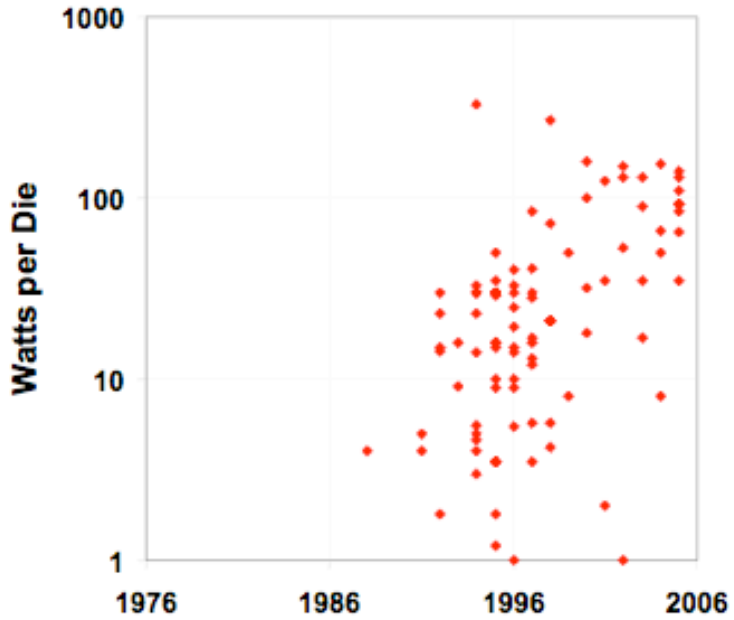
Decreasing Supply Voltage, Increasing Power

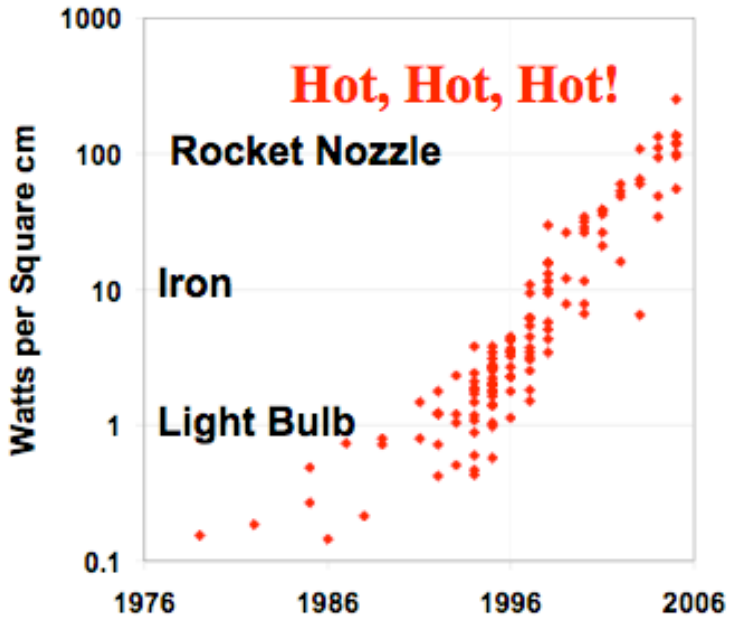
YEAR	2004	2007	2010	2013	2016
TECHNOLOGY	90 nm	65 nm	45 nm	32 nm	22 nm
CHIP SIZE	550 mm^2	550 mm^2	550 mm^2	550 mm^2	550 mm^2
NUMBER OF TRANSISTORS (LOGIC)	553 M	1 Billion	2 Billion	4.5 Billion	8.5 Billion
DRAM CAPACITY	1.0 Gbits	2.0 Gbits	4.3 Gbits	8.5 Gbits	35 Gbits
MAXIMUM CLOCK FREQUENCY	4.1 GHz	9.3 GHz	15 GHz	23 GHz	40 GHz
MINIMUM SUPPLY VOLTAGE	**0.9 V**	**0.8 V**	**0.7 V**	**0.6 V**	**0.5 V**
MAXIMUM POWER DISSIPATION	**150 W**	**190 W**	**200 W**	**200 W**	**200 W**
MAXIMUM NUMBER OF I/O PINS	3000	4000	4000	5300	7000

A funny thing happened on the way to 45 nm

- Speed increases with scaling...
- Power decreases with scaling... Remember these!

Why the clock flattening? POWER!!!!



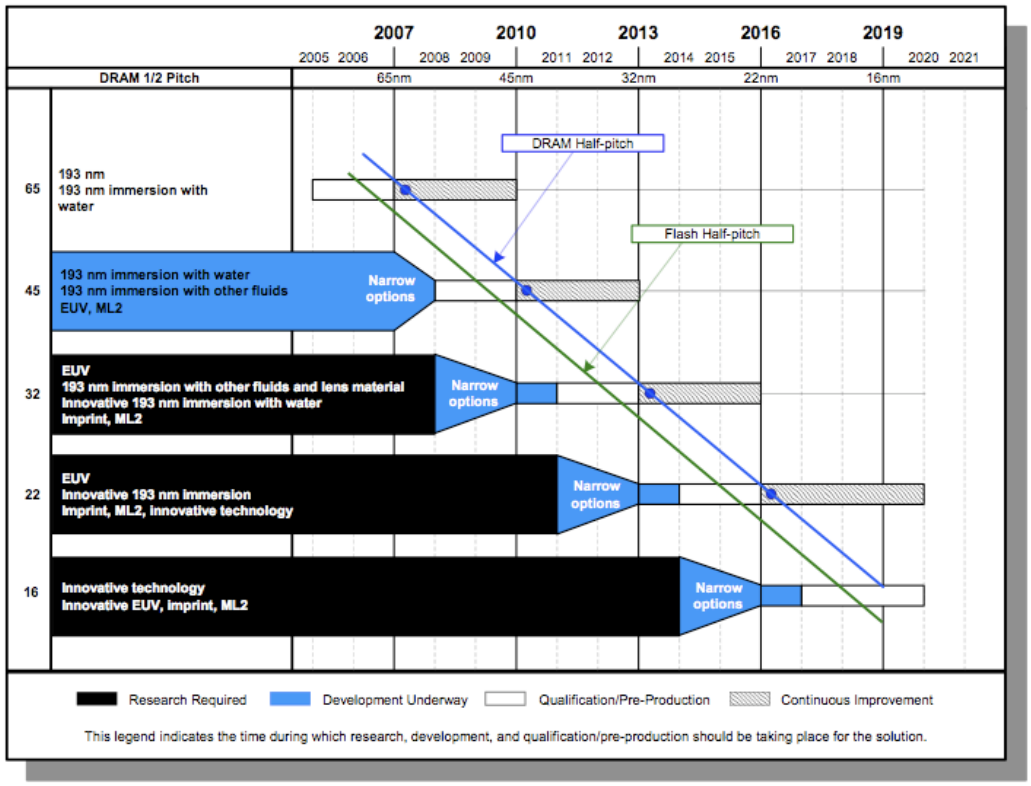


A funny thing happened on the way to 45 nm

- Reliability increases with scaling... Remember these!

Really? Lots of fabrication issues exist for which there are "no known solutions"

24 Lithography



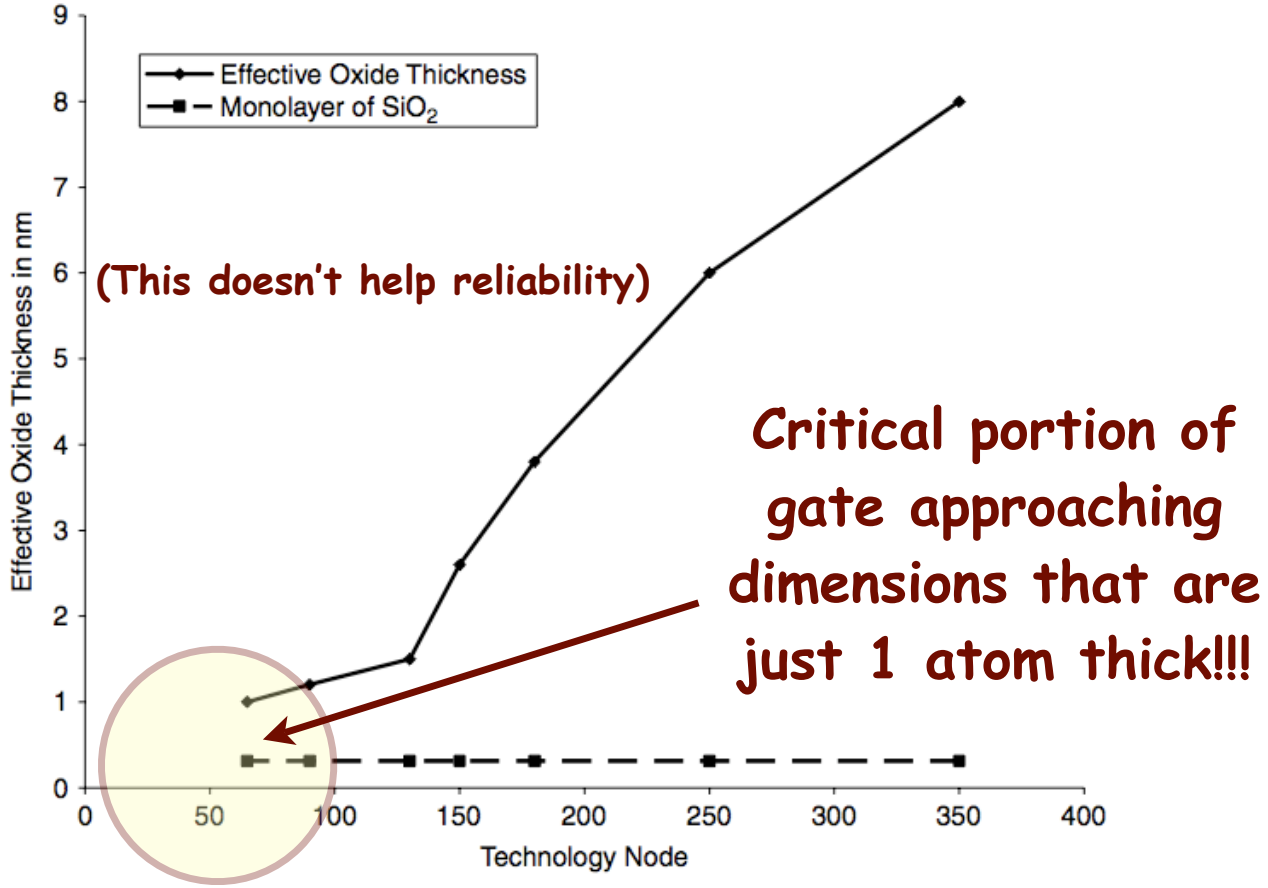
Notes: RET and lithography friendly design rules will be used with all optical lithography solutions, including with immersion; therefore, they are not explicitly noted.

Figure 67 Lithography Exposure Tool Potential Solutions

A funny thing happened on the way to 45 nm

- Reliability increases with scaling... Remember these!

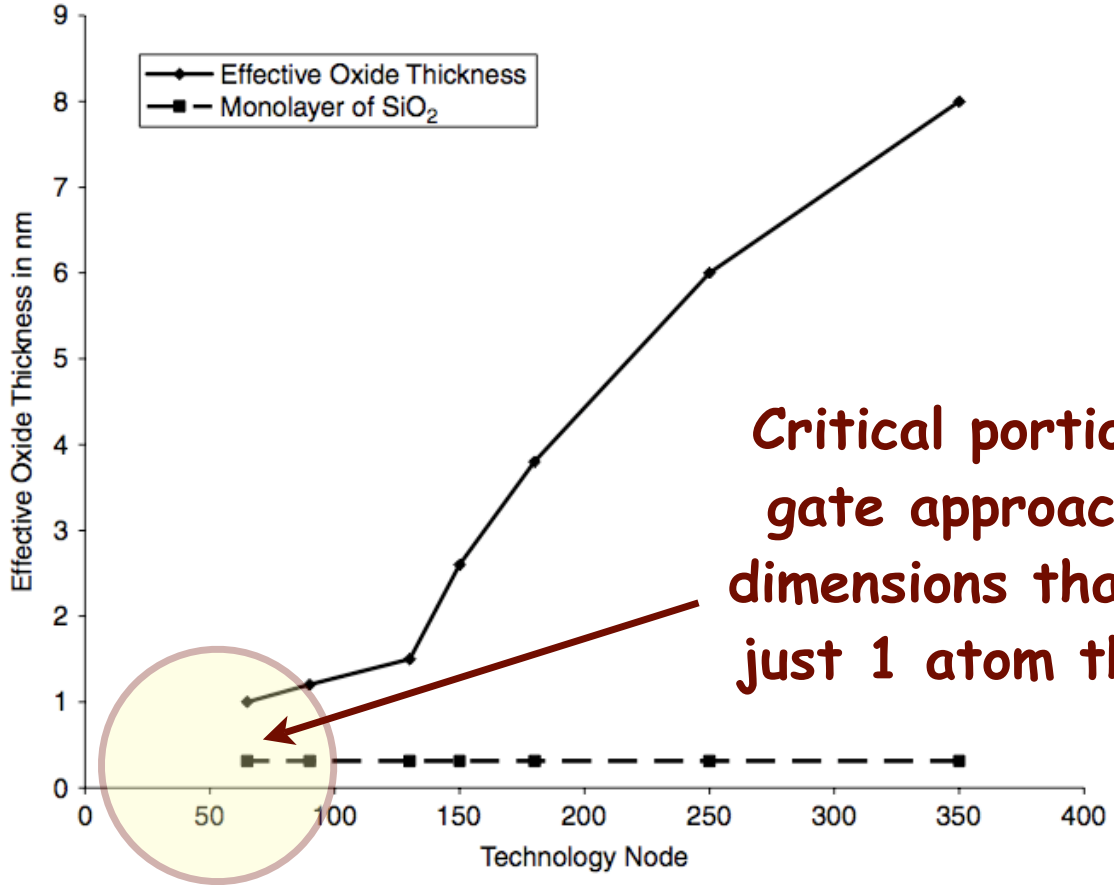
One quick example:



A funny thing happened on the way to 45 nm

- Power decreases with scaling... Remember these!

At such small dimensions, quantum mechanical effects take over and devices no longer turn off well.



Critical portion of gate approaching dimensions that are just 1 atom thick!!!