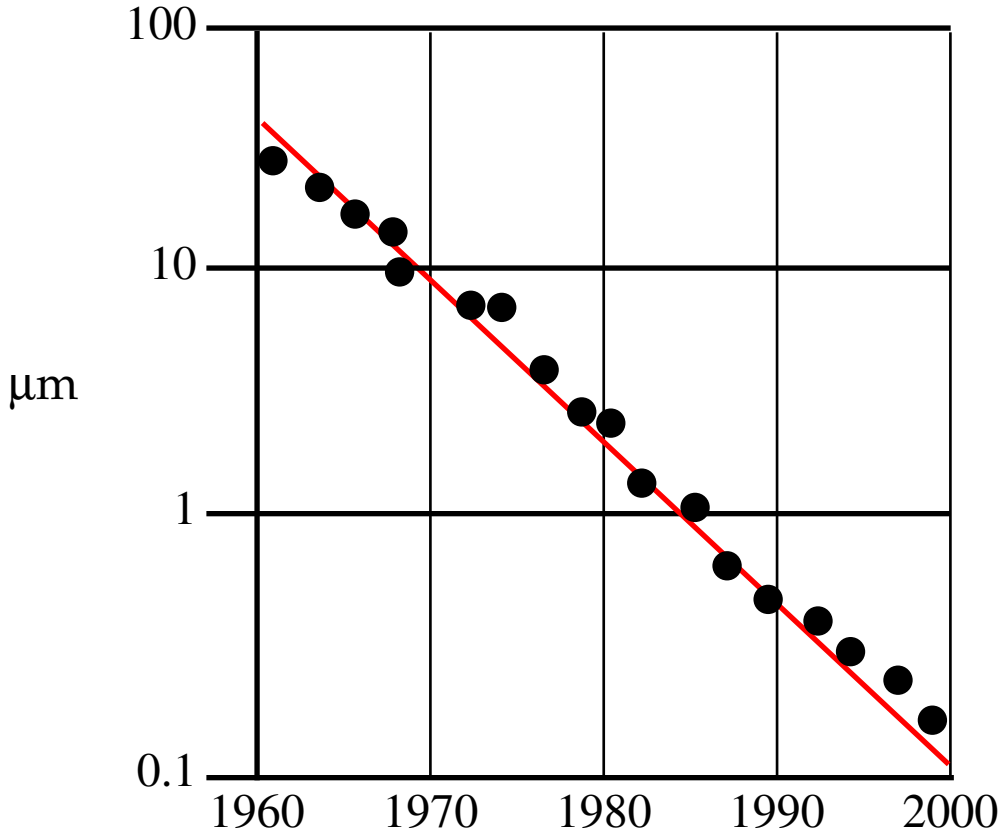
Scaling

Technology scaling rate is approximately 13%/year, halving every 5 years.



The size of the circuits also continues to increase.

Besides increasing the number of devices, scaling has had a profound impact on both speed and power.

Full Scaling (*Constant Electrical Field Scaling*)

In the ideal model, all the dimensions of the MOS devices, e.g., the voltage supply level and depletion widths are scaled by the same factor *S*.

Keeping the electric field patterns constant avoids breakdown and other secondary effects.

This leads to greater device density, higher speed and reduced power consumption.

R_{on} remains constant -- performance is improved because of the *reduced capacitance*.

Circuit speed increases linearly while the power scales down quadratically!

Both clearly indicate the benefits of scaling.

Scaling

Parameter	Relation	Full Scaling	General Scaling	Fixed-V Scaling
W, L, t_{ox}		***1/S***	***1/S***	***1/S***
V_{DD}, V_T		***1/S***	***1/U***	***1***
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{sat}	$C_{ox}WV$	1/S	1/U	1
Current density	I_{sat}/Area	S	S^2/U	S^2
R_{on}	V/I_{sat}	1	1	1
Intrinsic Delay	$R_{on}C_{gate}$	**1/S**	**1/S**	**1/S**
P	$I_{sat}V$	**$1/S^2$**	**$1/U^2$**	**1**
Power density	P/Area	1	S^2/U^2	S^2

See text for assumption used to derive this table.

Dimensions are scaled by *S* while voltages are scaled by *U*.

Fixed Voltage Scaling

Full scaling is not a feasible option.

For example, to keep new chips compatible with existing chips, voltages cannot be scaled arbitrarily.

Providing multiple voltage supplies is expensive.

5V was used up through the early 90s.

Voltages of 3.3 and 2.5 used since the introduction of 0.5 µm.

The change from a **fixed-voltage scaling** model to the **general scaling** model used today can be justified by reviewing the rightmost column.

In velocity saturated device, keeping the voltage constant while scaling the device dimensions:

- Does **not** provide a performance advantage over full scaling model (1/S vs. 1/S)
- But has a major power penalty associated with it ($1/S^2$ vs 1).

Fixed Voltage Scaling

Note the gain of increased current level is offset by the **higher** voltage swing, which only hurts power dissipation.

Also note that this is very different from the situation when transistors were operating in the *long-channel* mode.

Here, current was a **quadratic** function of the voltage.

In this scenario, keeping voltage constant gave a performance advantage (net reduction in "on" resistance"

Other reasons for scaling the supply voltage include hot-carrier effect and oxide breakdown.

These latter reasons played a significant role in the trend we see today.

Bear in mind that this is a *first order* analysis -- in reality, there is a (small) performance benefit with *fixed* voltage due to, e.g., channel length modulation.

General Scaling

Supply voltage is now being scaled, but at a slower rate than feature size.

For example, from 0.5 µm to 0.1 µm, supply voltage reduced from 5 V to 1.5V.

Then why not stick with *full scaling* model if there is no benefit to keeping the supply voltage higher.

- Some device voltages, e.g., *silicon bandgap* and *built-in junction potential*, are material parameters and cannot be scaled.
- V_T scaling is limited since making it too low makes it difficult to turn off the devices completely.

  This is aggravated by large process variations.

A more general scaling model is needed, where **dimensions** and **voltages** are scaled independently using *S* and *U* respectively.

Under *fixed voltage* scaling, **U = 1** as shown in the last column of the table.

General Scaling

Under *general scaling* model, performance scenario is identical (1/S) to other models but power dissipation lies between the two models, S > U > 1.

Recent CMOS technologies and projections of the future.

Year intro ->	2001	2003	2005	2007
Drawn channel L (nm)	90	65	45	35
Physical channel L (nm)	65	45	32	25
Gate oxide (nm)	2.3	2.0	1.9	1.4
V_{DD} (V)	1.2	1.0	0.9	0.7
NMOS I_{Dsat} (μA/μm)	900	900	900	900
NMOS I_{leak} (μA/μm)	0.01	0.07	0.3	1