Scaling:

- As the transistors become smaller, they switch faster, dissipate less power and are cheaper to manufacture. Despite the ever increase is challenges process advances have actually accelerated in the past decade.
- Such scaling is unprecedented in the history of technology. However scaling also excessive noise and reliability issues and introduces new problems.
- Designers need to be able to predict the effect of this feature size scaling on chip performance to plan future products, ensures existing products will scale gracefully to future processes for cost reduction and anticipate looming design challenges.

Transistor Scaling:

The characteristics of an MOS device can be maintained and the basic operational characteris. Can be preserved if the critical parameters of a device are scaled by a dimensionless factor . These parameters include.

º All dimensions (x,y, z directions)

º Device voltages

º Doping concentration densities.

Another approach is **lateral Scaling** , in which only the gate length is scaled. This is commonly called a gate shrink because it can done easily to an existing mask database for a design.

For **constant field scaling**, all devices dimension including channel length L, width W and oxide thickness t_{or} are reduced by a factor of 1/s. The supply voltage VDD and the threshold voltages are also reduced by1/s.

- The substrate doping WA is increased by.
- Because both distance and voltage are scaled equally, the electric field remain constant.
- A gate shrink scales only the channel leng leaving other dimensions, voltages and doping levels unchanged.
- This offers a quadratic improvement in gate delay according to the first order model.

- The gate delay improvement is closer to linear because velocity saturation keeps the current and effective resistance appronimately constant.
- The constant voltage scaling increased the electric fields in the devices. By the 1 μ m generation velocity saturation was servere enough that decreasing feature size no longer improved device current.

Inter connect Scaling:

- Two common approaches to interconnect scaling are to either scale all dimensions or keep the wire height constant.
- Wire length decreases for some types of wires, but may increase for others? Local are scaled wires are those that decrease in length during scaling.
- Example: A wire across 64 bits ALU is local because it becomes shorter as the ALU is migrated to finer process. A wire across a particular micro processer is scaled because when the microprocessor is shrunk to the new process the wire will also shrink.
- Un repeated interconnect delay is remaining about constant for local interconnect and increasing for global interconnect . This presents a problem because transistor are getting faster, So the ratio if interconnect to gate delay interconnect with scaling .
- In moders process with aspect ratios 1-5-22 fringing capacitance accounts for the majority of the total capacitance.
- Scaling spacing but not height interconnect the fringing capacitance enough that the extra thickness scarcely improves delay.
- Observe that when wire thickness is called the capacitance per unit length remains constant. Hence, a reasonable initial estimate of the capacitance of a

  minimum-pitch were is about 0.2fF/ μm, independent of the process.

- Wire capacitance is roughly 1/10-1/6 of gate capacitance per unit length.

Impacts on Design:

- One of the limitations of first order scaling is that it gives the wrong impression of being able to scale proportionally to zero dimensions and zero voltage.

Improved performance and cost:

- The most positive impact of scaling is that performance and cost are steadily improving. System architects need to understand the scaling of CMOS technologies and predict the capabilities of the process several years into the future, when a chip will be completed.

Interconnect :

Scaling transistors are steadily improving in delay but scaled wires are holding constant or getting worse.

- The wire problem motivated a number of papers predicting the demise of conventional wires.
- The plot is misleading in two ways.
- First the gate delay is shown for a single unloaded transistor rather than a realistically loaded gate. Second, the wire delay shown for fixed but as μ technology scales, most local wires connecting gates within a unit also become shorter.

Power:

In classical constant field scaling, power density remains constant and overall chip power increases only slowly with die size.

- Power density has sky rocketed because clock frequencies have increased much faster to classical scaling would predict and V_{DD} is some what higher than constant field scaling would demand.
- Dynamic power consumption will not continue to increase at such rates

because it will become uneconomical to cool the chips.

- The static power consumption caused by sub threshold leakage was historically negligible but becomes important for threshold voltage below about 0.3 to 04v.