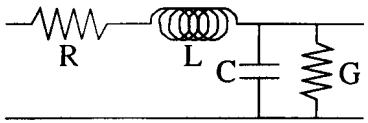
Interconnections as Transmission Lines

As integrated circuits increase in speed to the point where rise and fall times become comparable to the speed of light delay along the interconnection, we can no longer model interconnections as RC circuits. In addition, we will need to add in the effects of inductance of the wire. This effect is required for all packaging technologies and becomes important on chip for very high speed technologies such as scaled ECL, BiCMOS and GaAs. (The speed of light is about 1 ft/nS or 300μm/pS). Signal transmission generally has voltages changing rapidly and is thus limited in speed by capacitive effects for internal connections. Power supply connections are at a constant voltage but often have large current surges -- leading to large voltage shifts from L(dI/dt) effects. Power supply connections and noise issues are also strongly effected by inductance effects.

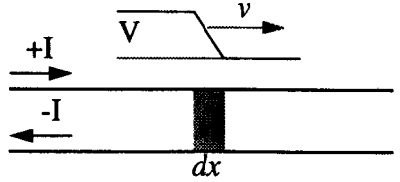
- **Fundamentals**

Inductive effects surface in a wire when the current changes fast enough with respect to the inductance such that L(dI/dt) = V is comparable to the shift in voltages driving the wire. This effect places an upper limit on the current change per time that is allowed by a wire. In effect, increasing the size of the driver beyond a certain point will not cause a larger initial current to flow. A simple model for a wire (transmission line) is shown below:



It is important to remember that both connections of the signal and the "ground" be modeled explicitly. Inductances and inductive effects are based on magnetic fields which accumulate over the entire circuit loop. Inductances on the ground return must be added to those on the signal wire itself.

- **Lossless Transmission Lines**



We are interested in particular in the lowest order mode of transmission on a transmission line as this mode is energetically favored for any finite resistance and this mode will comprise the bulk of the effects for frequency bound cases. Consider the case above, we assume a propagating step of width *dx* and velocity *v*. Initially, the line is quiescent with zero current and zero voltage between the conductors. After the step has passed a given point on the line, the current in the top conductor is +I while the current in the bottom conductor is -I, and there is a fixed voltage difference V between the top and bottom conductors. From Maxwell's equations we have:

$$\Phi = \iint_S (B \cdot dS)$$

where B is the magnetic field and dS is the normal surface component and Φ is the flux through the surface. For a closed surface, there is a constant of proportionality between the current traversing the surface and the flux. This constant is called the In-

ductance L:

$$L = \frac{\Phi}{I}$$

On a transmission line, the inductance and the magnetic flux are both defined per unit length of the line. At the front of the wave, the flux is changing, we have: $d\Phi = d(IL) = ILdx$. From Faraday's law, there is a voltage potential equal to the time change of flux. Thus the voltage across the line after the wave has passed must be equal to the average rate of change of flux:

$$V = \frac{d\Phi}{dt} = IL\frac{dx}{dt} = ILv$$

At the same time, there is a change in voltage in the line, there must be sufficient current flowing down the line to charge the capacitance per unit length. We know that $Q = CV$, so the current I flowing down the line must be:

$$I = \frac{dQ}{dt} = \frac{d(CV)}{dt} = CV\frac{dx}{dt} = CVv$$

Combining with the inductance relation above we get:

$$v = \frac{1}{\sqrt{LC}}$$

where L and C are both defined in terms of unit length. In addition, the impedance Z0 of the line (analog of the resistance) is:

$$Z_0 = \frac{V}{I} = \sqrt{\frac{L}{C}}$$

The derivation above assumed that the electric and magnetic permeabilities were unit quantities. In MKS units, we have the following relation:

$$v = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c}{\sqrt{\varepsilon_r\mu_r}}$$

where ε and μ are the dielectric constant and magnetic permeability of the (uniform) transmission medium and c is the speed of light in vacuum = 29.972 cm/nS. For non magnetic materials i.e. those with no ferromagnetic or diamagnetic effects, the value of μ is very close to its vacuum value: 1. Accordingly, we have:

$$Z_0 = \frac{\sqrt{\varepsilon_r}}{cC}, L = \frac{\varepsilon_r}{c^2C}$$

So we can describe a lossless transmission line in terms of its capacitance per unit length C, and its dielectric constant.

$$v = \frac{c_0}{\sqrt{\epsilon_r}} = \frac{30\ cm/nsec}{\sqrt{\epsilon_r}}$$

Dielectric	Relative Dielectric Constant (ϵ_r)	Propagation Speed (v) (cm/nsec)
Polyimide	2.5–3.5	16–19
Silicon dioxide	3.9	15
Epoxy glass (PC board)	5.0	13
Alumina (ceramic)	9.5	10

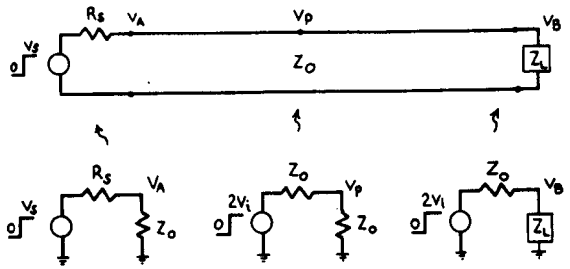
Note that the total delay time for a wave on such a conductor is:

$$t_d = \frac{l}{v} = l\sqrt{LC} = \sqrt{L_TC_T}$$

where L_T and C_T are the total inductance and capacitance for the entire transmission line.

- **Circuit Models**

We have calculated the characteristic impedance for a wave traveling down a transmission line. The value found for Z was completely real (lossless case). Thus it acts as a pure resistance and we can model the transmission line instantaneous boundaries using an analog of a resistor voltage divider. Consider the line below:



Using the resistor model, if a step input is applied by the source of magnitude Vs, we should see an instantaneous change in potential at point Va of magnitude:

$$Va = \frac{Z_0}{R_s + Z_0} Vs$$

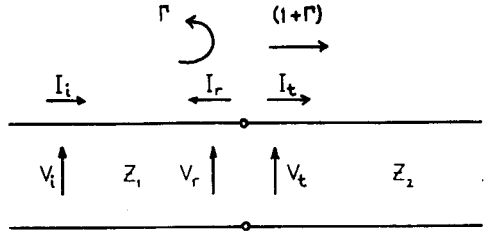
At the end of the transmission line, we have a step-like wave impinging on a load Z_L. We model this as a source of potential 2Vi across a similar voltage divider. Thus the voltage step across the load is:

$$V_L = 2Vi \frac{Z_L}{Z_0 + Z_L}$$

The factor of two in the proceeding equation reflects the two components of the traveling wave, one going in each direction along the line. In fact, if a discontinuity is found along the transmission line, some amount of the wave will be reflected, and some transmitted.

- **Impedance Changes**

The simplest case of a discontinuity is that of a transition into a new transmission line with differing Z. We can calculate the reflection and transmission coefficients of the wave by assuming conservation of the charge, and matching of the fields across the boundary. Consider the following system:



On both sides of the line we must have: Vx = Zx Ix as this condition guarantees conservation of current. In addition, at the junction we must have:

$$I_i = I_r + I_t, V_i + V_r = V_t$$

After the wave has passed, the reflected and transmitted waves will share the energy of the incident wave and will propagate away from the discontinuity, each in its own characteristic impedance. Thus:

$$\frac{V_i}{Z_1} = \frac{V_r}{Z_1} + \frac{V_t}{Z_2}$$

From which we can find that the potential of the waves are:

$$V_r = V_i \frac{Z_2 - Z_1}{Z_2 + Z_1} = \Gamma V i$$

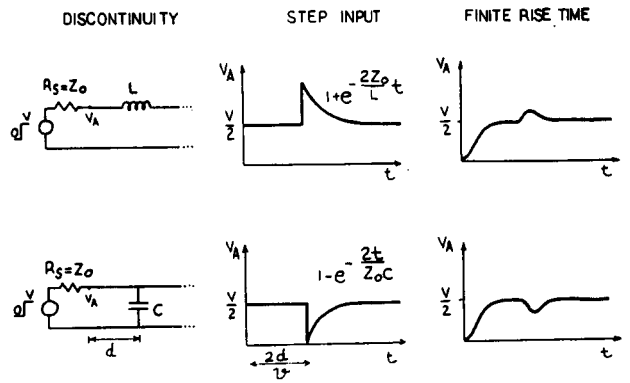
$$V_t = V_i \frac{2Z_2}{Z_2 + Z_1} = (1 + \Gamma) V_i$$

Note that these equations give the proper results for the case when $Z_2 = Z_1$, namely no reflection and 100 per cent transmission. It is interesting to examine the previous case of a terminated line with a load impedance of Z_L. Since the load is real (lossless) we have the case above, where the transmitted wave is the portion which appears across the resistor and is dissipated. The reflected wave can have 3 cases:

1. if $Z_L = Z_0$ then the line is (broadband terminated) and no reflected wave is observed.
2. if $Z_L > Z_0$ then the reflected wave has the same polarity as the incident wave.
3. if $Z_L < Z_0$ then the reflected wave has the opposite polarity as the incident wave.

- **Inductive and Capacitive Discontinuities**

It is often more accurate to model the effects of connections or of package pins etc. as abrupt discontinuities (since they are usually much smaller than the expected wave length) of capacitance or of inductance or both. In the figure below the effect of such a discontinuity is plotted. We assume here that the source impedance is equal to that of the transmission line and that the line is either infinite or is terminated on the far end.



An inductor acts initially as an open circuit (the current cannot change instantaneously fast) so the reflected wave is of the same polarity and magnitude and the voltage at Va doubles. Later, as the current through the inductor increases, the reflected wave subsides and eventually the inductor acts as a short. The time constant of the charging of the inductor is: $L/2Z_0$. (This is the cumulative current from the transmitted wave). Note that the reflected wave will travel backwards along the transmission line until it too reaches a termination.

A capacitor across a transmission line initially acts as a short circuit. Thus the reflected wave is equal in magnitude and of opposite polarity to the incident wave and the voltage at Va drops to zero. As the capacitor charges, the reflection subsides until the capacitor acts as an open circuit. The time constant of this relaxation is: $Z_0C/2$. (Again, the potential of the transmitted wave across the capacitor sets this value).

If the step input has a slow (finite) rise time, the peak values of the reflected wave are not reached since the discontinuity is partially relaxed before the incident wave reaches maximum potential. Indeed, if the rise time is slow enough, the voltage at Va assumes an average value made of several reflections of the reflected component, and appears to charge slowly with the capacitor. (Circuit theory case). However, if the transit times are sufficiently fast, the reflections are observed and cause problems in the system. For pure inductive or capacitive discontinuities, the amplitude of the reflection is:

$$V_r = \frac{LV_i}{2Z_0 t_r}, V_r = -\frac{CZ_0 V_i}{2t_r}$$

where t_r is the rise time of the input transition. Below is a small table of characteristic discontinuities.

Component	Capacitance (pF)	Inductance (nH)
68 pin plastic DIP pin†	4	35
68 pin ceramic DIP pin††	7	20
68 pin SMT chip carrier lead†	2	7
68 pin PGA pin††	2	7
256 pin PGA pin††	5	15
Wire bond	1	1
Solder bump	0.5	0.1

† No ground plane; capacitance is dominated by wire-to-wire component.
†† With ground plane; capacitance and inductance are determined by the distance between the lead frame and the ground plane, and the lead length.

- **Applications**

Transmission line effects will become significant when the rise time of the signal (10% to 90%) swing occurs in a time comparable to the time of transit across the transmission medium. That is:

$$t_r < 2.5t_f = \frac{l}{v_{tm}}$$

where t_f is the transmission group velocity, l is the length and v_{tm} is the velocity of the transmission line propagation. On the other hand, we can assume the line will act as a lumped capacitor (i.e. simpler RC analysis) whenever:

$$t_r > 5t_f$$

As an example: 0.5nS rise time with SiO_2 ($\varepsilon = 3.97$) gives l_{max} = 3cm. This means that line shorter than about 1.5cm can be treated as RC delays, lines longer than 3cm will show transmission line effects-- i.e. should be terminated etc. Below is a table of interconnection lengths, technologies and dimensions.

The actual rise time of a gate is a function of two factors:

1. The rate (and gain) of the driver turn-on. (Sometimes the driver is slowed on purpose.)
2. The effective driving current (set by the ratio of driver source resistance to transmission line impedance).

$$v = \frac{c_0}{\sqrt{\epsilon_r}} = 15 \text{ cm/nsec}$$

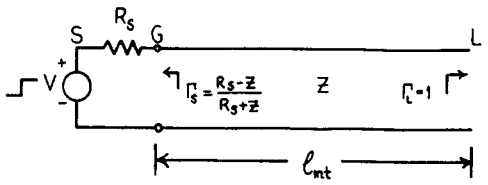
Rise Time t_r (psec)	Critical Line Length $l_{crit} = vt_r/2.5$ (cm)
50	0.3
100	0.6
250	1.5
500	3.0
750	4.5
1000	6.0

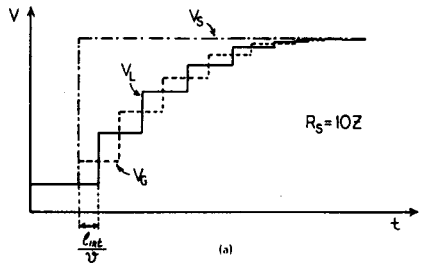
Technology	On-Chip Rise Times	Off-Chip Rise Times
CMOS	0.5–2 nsec	2–4 nsec
Bipolar	50–200 psec	200–400 psec
GaAs	20–100 psec	100–250 psec

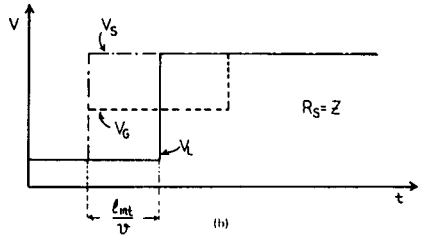
Component	Dimension
Die	1 cm
Chip carrier (single-chip)	3 cm
Module (multichip)	10 cm
PC Board	50 cm

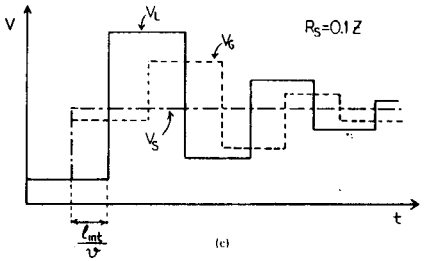
If the driver is turned on rapidly, then the effective rise time has two cases: if the source resistance is large compared to the line impedance, the initial step wave will be small and take several iterations to charge the lumped line capacitance. If, on the other hand, the source resistance is comparable to the line impedance, the step will be large, and the rise time will be determined by the parasitic and local line capacitances and any series inductances.

In the figure below, we show the effects of a step driving a transmission line in the case of several different driving impedances. Here we assume an unterminated transmission line shown below.

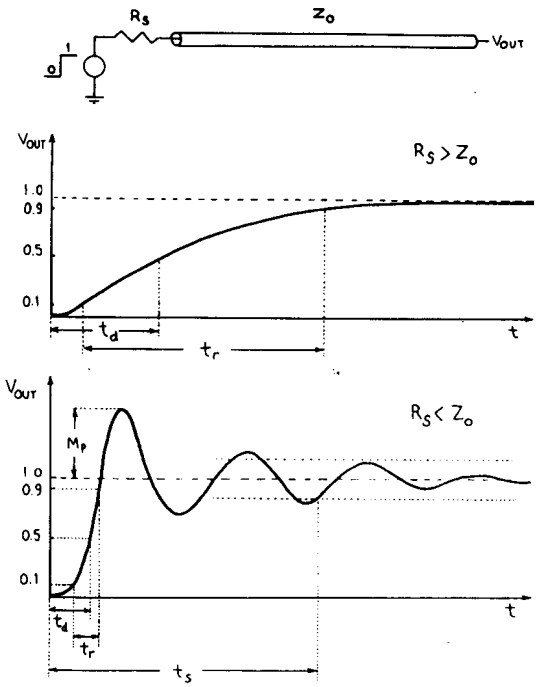








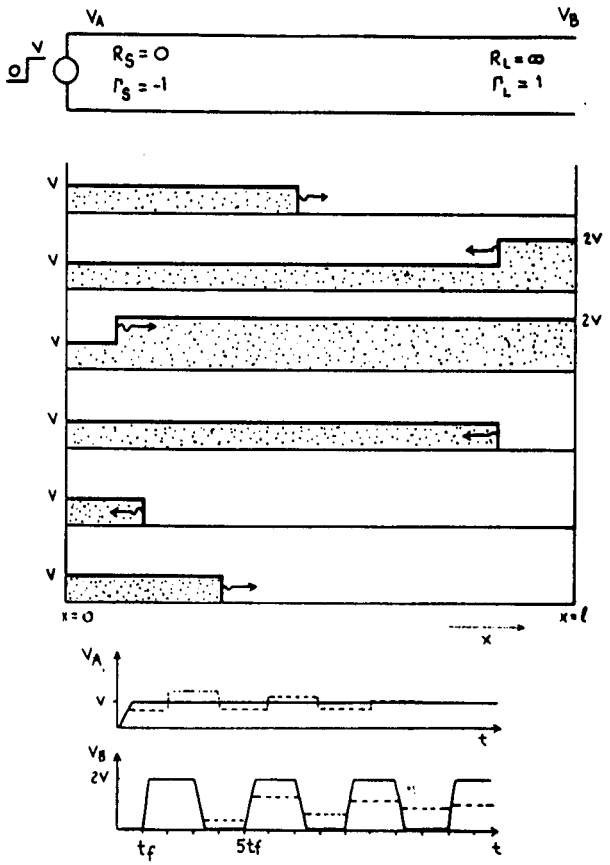
In any real circuit, the effects are somewhat mitigated by having a finite rise time driver-- However, there is still plenty of problems in the case of lines where t_f is larger than several times t_r



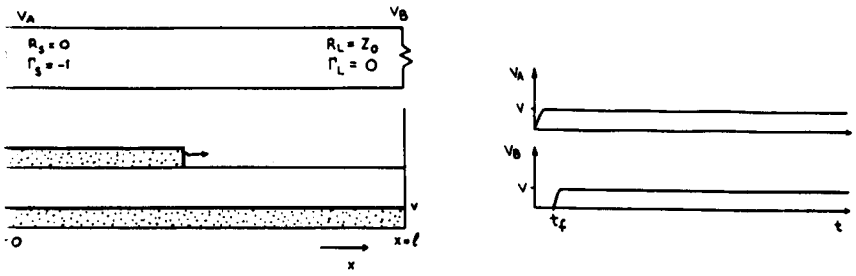
- **Circuit Termination**

We will now examine characteristic cases of circuits with interconnections modeled in the cases where transmission line phenomena dominate. We first consider resistive cases:

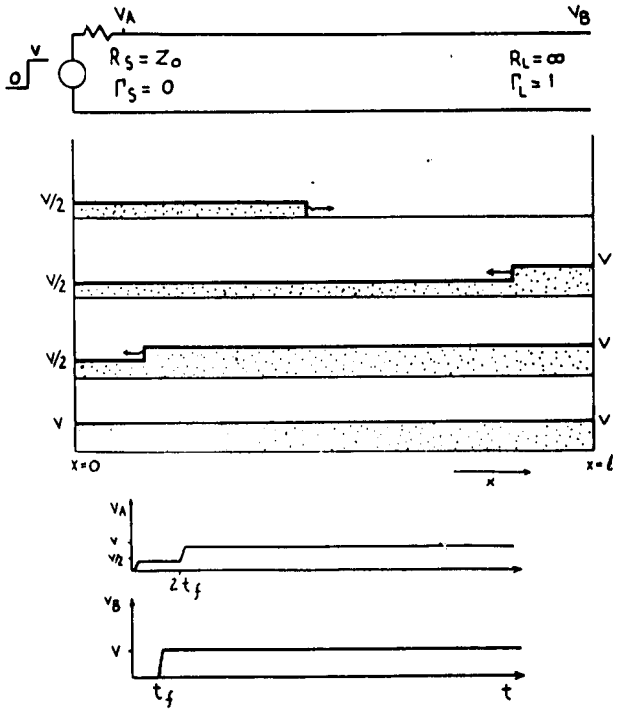
Unterminated: $R_S = 0, R_L = \infty$



Parallel Termination: $R_S = 0, R_L = Z_0$



Source Termination: $R_S = Z_0, R_L = \infty$



- **Worst Cases for Source Termination**

If a transmission line is terminated at the source end (which as we shall see is often preferable in CMOS), we would like to know the limits of voltage excursion for both under and over terminated cases. In general, both the source resistance and the line impedance vary due to process and fabrication variations. If Rs is significantly smaller than Z_0, we expect to see ringing of the signal. The amplitude of the ringing for the first reflection (at the receiver end) is set by the ratio of Rs to Z_0:

$$1 + \Gamma = \frac{2Z_0}{R_S + Z_0} = \frac{2}{\left(\frac{R_S}{Z_0}\right) + 1}$$

So the receiver sees a sequence (For $Rs=Z_0/3$):

	t_f	$2t_f$	$3t_f$	$4t_f$	$5t_f$
V_R:	1.5V	0.75V	1.12V	0.94V	1.03V
V_S:	0.75V	1.12V	0.94V	1.03V	0.98V

(you should work this out to be sure you know what is going on) In the case where Rs is larger that Z_0, we expect that the timing will take several trips to achieve a voltage large enough so that the receiver will trigger. If we assume a worst case over and undershoot of 0.75V, we find that the ratio between Rs and Z_0 should be:

$$\frac{1}{3} < \frac{R_S}{Z_0} < \frac{5}{3}$$

- **Capacitive Loading**

If a source terminated transmission line is driving a capacitive load C_L, we expect that the capacitor charging will be of order $C_L Z_0$. i.e. The line behaves as if it is a resistor with resistance Z_0. This will be the case when the capacitor charging time is much smaller than the time of flight along the transmission line. This condition is equivalent to requiring that: $C_L << C$. i.e. the load capacitance is much smaller than the total line capacitance of the entire line. When this is not the case, the capacitor acts as a direct short on the line until it is charged, with a time constant of:

$$t_{\frac{1}{2}} = t_f + \frac{Z_0 C_L}{2}$$

i.e. we see the capacitor charging after an initial delay.

- **Termination in Resistive Transmission Lines**

In a lossy transmission line the amplitude decreases exponentially with the length of the wire. The attenuation due to resistance goes as: $R/2Z_0$, the amplitude at the receiving end is:

$$V_r = 2e^{-\frac{R}{2Z_0}}$$

Again, we wish the value of R to be set so that the voltage Vr is large enough to trigger on the first impulse, yet small enough to quickly reduce the amplitude of the reflections. If $R*l = 2Z_0/3$, we have:

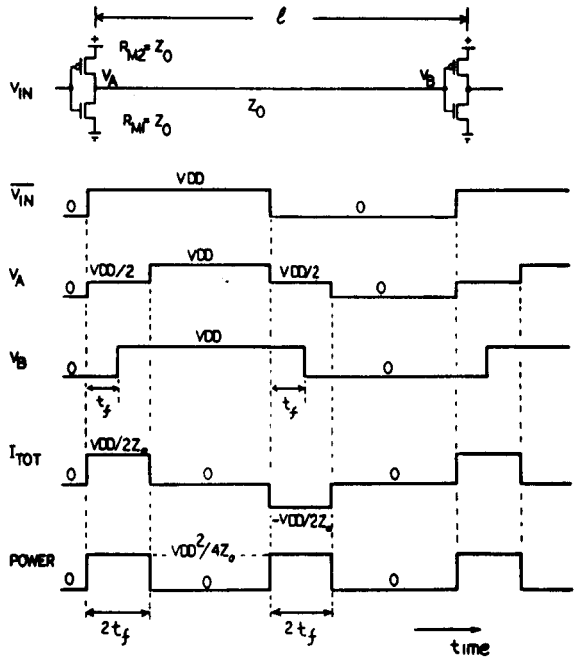
$$V_{r1} = 2Ve^{-(1/3)} = 1.43V, V_{r2} = V(1.43 - e^{-1}) = 1.06V$$

On the other hand, if the resistance is too large, the transmitted pulse will be too small -- to ensure that it is at least 0.75 of the total amplitude we must have: $R*l < 2Z_0$. We can re-write these conditions in terms of length of a wire with fixed resistance per length and impedance:

$$\frac{2Z_0}{3R} < l < \frac{2Z_0}{R}$$

- **Termination of CMOS driven lines**

Consider the line below: here the channel resistance of the transistors is matched to the line impedance:



The circuit achieves the ideal for CMOS switching -- no power is lost except during switching. We can calculate the size of the drivers to achieve this matching of resistance to impedance as follows: Under ideal conditions, 1/2 of the voltage swing appears on the transmission line. This amplitude requires a current I = $Vdd/2Z_0$ as the line is charged. Since Vds = Vdd/2 the transistor is in the linear region so:

$$i = \frac{V_{dd}}{2Z_0} = \mu C_{ox}\frac{W}{L_{eff}}\left((V_{dd}-V_T)\frac{V_{dd}}{2}-\frac{V_{dd}^2}{8}\right)$$

Thus:

$$\frac{W}{L_{eff}} = \frac{1}{\mu C_{ox}Z_0\left(\frac{3}{4}V_{dd}-V_T\right)}$$

Although this seems to be an excellent technique, we must have very accurate control of both the sizes and the thresholds of the transistors for it to work. It is often easier (although much more costly) to source terminate using a resistor in series with the driver. This increases the size required for the driver, but resistors can often be fabricated with much tighter tolerances than transistors.

Another interesting phenomena of series source termination is that the drivers closest to the source receive the data last. This is due to the shift of Vdd/2 in the initial pulse. The driver is not turned on fully until the return of the reflection from the end of the wire. Points closest to the source will have to wait longest for the reflection.

- **Busses**

In general, ultra high speed systems produce an output for each separate connection that is required. This is very costly and often vastly increases the size of the system. A more common strategy is to make terminated busses with fanout. Although these structures make sense from the cost standpoint, they are often the cause of difficulties in high speed circuits when not carefully designed.

Generally, a bus can be modeled as a tree of connections away from the source. Source series terminations cannot be used unless the source is either at the middle or one of the ends of the bus. Commonly, large (low Rs) drivers are used with termination at the farthest extent of the bus. This works well if the bus is similar to a single line with some short and long attached stubs. (The termination should be at the end of the longest stub...) In this case at each long stub junction, we have a reflection and two transmitted waves. If the stub is long compared to the wavelength, the reflection will be -1/3 the transmission will be +2/3 in each continuing line. (This is due to the effective impedance of $Z_0/2$ at the junction).

If the stub is short, we can consider it as an incremental capacitive load-- this reduces the effective impedance of the line by adding to the capacitance. A stub is short if both its internal length is shorter than the effective rise time, and the spacing between adjacent stubs is closer than the rise time of the signal. In this case: The impedance decreases due to the added capacitance, the velocity of propagation decreases as well.

$$Z_{eff} = \frac{Z_0}{\sqrt{1 + \frac{C_{stub}}{C}}}$$

while the additional delay time due to the stubs is:

$$\Delta T_D = T_0\left(\frac{Z_0}{Z_{eff}} - 1\right)$$

If, however, the stubs are spaced widely apart so that the time of flight is larger than the rise time, we must calculate the effects of the stubs individually-- using the capacitive discontinuity formula described earlier.

- **Loss**

Physical interconnects in non-superconducting technologies have resistive loss in the conductors. If the resistance is large enough, the resistance dominated the inductance -- and transmission line effects disappear. Losses in transmission lines are caused by factors other than the wire resistance -- in particular, dielectric loss.

The resistance of the transmission line itself is subject to skin effect -- which increases the effective resistance of the wire. In such a system, we expect the amplitude of the wave to decrease exponentially with the distance propagated. That is:

$$V(l) = e^{-\alpha l} V(0)$$

where α is the attenuation and is:

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} = \frac{R}{2Z_0} + \frac{GZ_0}{2}$$

where R, C, L, G, and Z_0 are the resistance, capacitance, inductance, conductance (loss) per unit length, and the impedance.

- **Conductor Loss**

We have the following relations for a wave on a lossy transmission line:

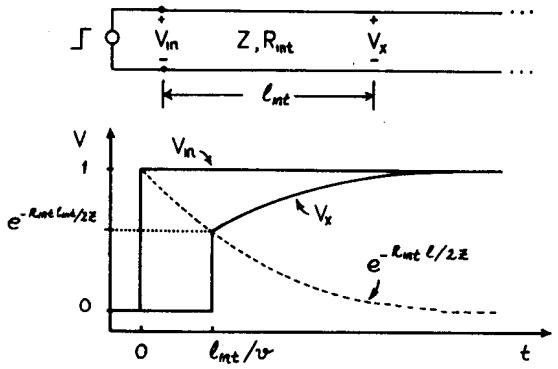
$$\frac{\partial V}{\partial x} = RI + L\frac{\partial I}{\partial t}$$

$$\frac{\partial I}{\partial x} = C\frac{\partial V}{\partial x}$$

We can find the response to a unit step at time t=0 at any position on the line as:

$$V(x,t) = \Theta(vt-x)\left\{ e^{-x\frac{R}{2Z_0}} + \frac{Rx}{2Z_0}\int_{\frac{x}{v}}^{t} e^{-t\frac{R}{2L}} \frac{I_1 R}{2L} (\text{stuff...}) \right\}$$

(Note that the expression in the text is wrong--) However, the basic structure of the pulse above is relatively simple:



If Rl >> $2Z_0$ the impulse of the line can be neglected -- an the line appears as an RC line with the proviso that no response occurs until the wave can propagate to the receiver. If we wish to receive the first edge of the transmission, we must have the attenuated step close to full swing at the receiver. i.e. we want:

$$R < \frac{Z_0}{2}$$

Since this comes from an exponential relation, it is sharp: for

$$R > 5Z_0$$

we can ignore the transmission line effects.

(Note: if we divide the line into an infinite number of circuit segments and apply our circuit model from before:

$$V_{i+1} = \left(\frac{2Z_0}{2Z_0 + \frac{R}{n}}\right) Vi$$

we get:

$$V_n = \left(\frac{2Z_0}{2Z_0 + \frac{R}{n}}\right)^n V_0$$

In the limit of large n, this is just:

$$V(l) = V(0)\, e^{\frac{R}{2Z_0}}$$

We can use the strong convergence of this limit to model the transmission line effects as a network in Spice. (A useful technique when accurate calculations are necessary).)

- **Skin Depth**

High frequency currents propagating on a resistive conductor tend to flow along the outside of the conductor. (In a superconductor, the flow is entirely along the surface). As the frequencies increase for a finite resistance conductor, the current flows more and more on the outside of the wire. -- This implies that the resistance loss of the wire is related to its surface area, not its volume.

The decrease into the conductor is exponential with coefficient:

$$\delta = \sqrt{\frac{\rho}{\pi\mu f}}$$

where f is the frequency, μ is the permeability, and ρ is the resistivity of the material. To get proper units: μ is about $4\pi x10^{-9}$ H/cm, so f should be inverse seconds, and ρ should have units of Ω-cm ($2x10^{-6}$ for Al). The effect of skin depth is to decrease the area available for current conduction -- this increases the apparent loss. At low frequencies, the attenuation per unit length for a stripline is:

$$\alpha = \frac{R}{2Z_0} = \frac{\rho}{2whZ_0}$$

which is inversely proportional to both the width and the conductor thickness. At high frequencies, however, only the surfaces of the conductor and the ground plane conduct-- leading to:

$$\alpha = \frac{R_{skin}}{Z_0} = \frac{\rho}{W\delta Z_0} = \frac{\sqrt{\pi\mu_0 f\rho}}{wZ_0}$$

Which depends inversely only on the width. Note, that as frequency increases, the attenuation also increases -- and the amplitude is related to the negative exponent of the attenuation factor.

- **Dielectric Loss**

A wave propagating in a partially conducting medium causes motion of carriers which scatter and slowly attenuate the energy of the wave. This loss is called dielectric loss and is related to the conductivity and dielectric constant. We define:

$$G = \omega C \tan(\delta_D)$$

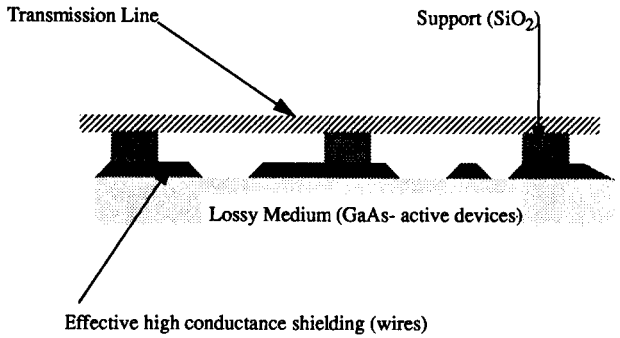
and G has the value:

$$G = \frac{\sigma_D C}{\varepsilon}$$

where σ_D is the conductance per unit length, C is the capacitance per unit length, and ε is the dielectric constant. The attenuation due to dielectric loss α_D is:

$$\alpha_D = \frac{GZ_0}{2} = \pi f C \tan(\delta_D)\sqrt{\frac{L}{C}} = \frac{\pi\sqrt{\varepsilon} f \tan(\delta_D)}{c}$$

Dielectric loss is negligeable in many materials, however, it is not negligeable in doped (non-compensated) GaAs or any Si substrate. High frequency line in both materials are often run on air bridges above a ground plane to minimize the wave penetration and hence the loss.



Noise and Crosstalk

Three basic sources of noise in IC's are:

- **Reflections**

Noise produced by discontinuous transmission lines.

- **Cross Talk**

Cross talk is produced by rapidly changing fluxes in close proximity of the target line. The coupling is represented by mutual inductances and/or mutual capacitances. The magnitude of the induced emf is proportional to the time rate of change and to the flux coupling.

- **Power coupled Switching Noise**

Switching large currents on power supply lines cause other effects than resistive drops. In particular, the large current swing causes back emf= L dI/dt. This has the effect of further limiting the current that can flow to supply the components.

Noise in a synchronous system has two possible effects: it can delay the proper recognition of the signal by causing transitory noise which eventually decay away. It can also cause latching errors of data in cases of pre-charged or triggered logic. In this case the error is permanent in that simply slowing the clock cannot remove the error. These errors are especially critical in asynchronously switched signals, as the effective state cannot be recovered after the error. A particularly pernicious form of this fault is cross talk on clock lines--

Please note that these errors can occur at either the receiver or the driver gate. If the driver gate is subject to a reflection from an earlier signal, the output pulse can be weakened or missing. At the receiver end, a glitch in the power supply can masquerade as a glitch in an input signal since the switching point is defined relative to the supply voltage.

Delay faults are prevalent for parts coming from the slower end of the speed distribution, while logic errors often occur in parts at the faster end of the speed distribution. This is because the slower parts do not generally switch fast enough for the glitches to be a problem, they merely increase the time before the pulse gets accepted. Very fast circuits are often very sensitive to transient glitches.

- **Line Impedance**

There are several trade-offs with line impedances that directly effect the performance of the system. High impedance lines (large Z_0) have the following characteristics:

1. Small Capacitances
2. Low power, smaller drivers
3. Generate smaller switching transient
4. Have higher susceptibility to cross talk

5. Cannot rapidly drive large capacitive loads.

Low impedance lines (small Z_0):

1. Have large capacitances (lower velocity).
2. Can supply large currents.
3. Are less suspectable to cross talk.
4. Generate larger switching transients
5. Have smaller reflections from capacitive discontinuities.

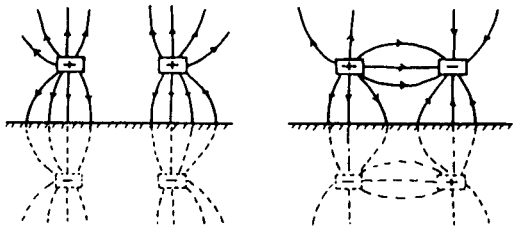
Optimal line impedances are different for CMOS and Bipolar technologies due the differences in noise immunity and drive capability. For bipolar systems impedances of 60-80Ω are ideal, while for CMOS ideal values are 80-100Ω. This large impedance trades on CMOS's higher noise immunity in return for lower power dissipation. Driving large capacitive loads in CMOS is difficult in any case as the drivers must be very large to make the source resistance low compared to the line impedance.

Although we can make large CMOS buffers to drive the low impedance lines-- these buffers require several stages of amplification -- nullifying the effect of the more rapid charging. It is necessary to consider both the source and line impedance in calculations of the voltage steps in CMOS drivers.

- **Cross Talk: Coupled transmission lines**

(Pure capacitive charge sharing based cross talk was discussed at the beginning of the course-- this model will also take into account the effects of inductance as well.)

A two wire transmission line has two basic (superposible) modes of propagation: equal potential on both lines with opposite charged ground plane images, and opposite charged potentials on both lines. These two modes have differing impedances and propagation velocities.



These two modes are referred to as "even" and "odd". Even modes have the two lines propagating waves of the same potential. We can find the impedances of the two systems by summing the impedances of the pairs of lines and image charges which are in phase. In particular, the even mode acts as a larger wire (twice as large) above the ground plane, but the field lines couple only

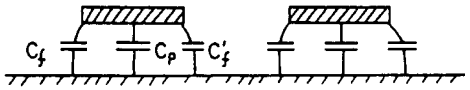
to the ground. In the odd case, the field lines couple to both the ground and to the other line-- thus lowering the effective impedance. In general:

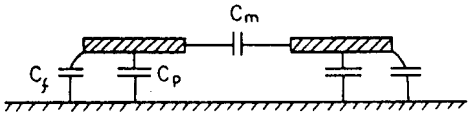
$$Z_{odd} < Z_0 < Z_{even}$$

for lines which are loosely coupled:

$$Z_0 \cong \sqrt{Z_{odd} Z_{even}}$$

In the case of coupled microstriplines we have:



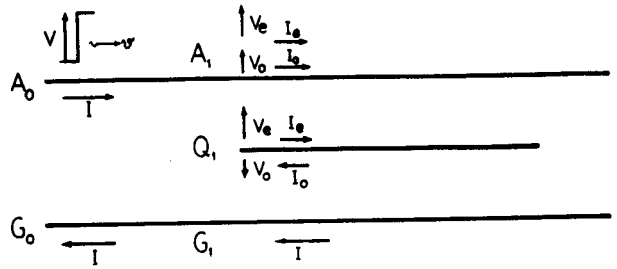


$$C_i = C_p + 2C_f$$
$$C_e = C_p + C_f + C'_f$$
$$C_o = C_p + C_f + C_m$$

where C_p is the plate capacitance, C_f is the fringing capacitance, Cf' is the modified fringing, and Cm is the capacitive coupling.

- **Coupling of Pairs of Transmission Lines**

We would like to calculate the coupling between two transmission lines as a function of the two impedances for even and odd transmission. To this end we postulate the following system:



Line A is the active line, along which a step wave is propagating. Line Q is the quiet line -- which is quiescent initially. G is the ground plane. We assume that Q is open (unterminated) at the point Q_1 adjacent to A_1. If a voltage step propagates down A, no reaction will be seen on Q until the step reaches point A_1 in the limit of fast rise times. Furthermore, if the coupling is relatively weak, the impedance of A is not altered much, so we shall assume that the wave travels unimpeded past the intersection. (We could be more accurate and model the reflection at point A1 but this effect diminishes rapidly as the coupling decreases). After A1, we have even and odd modes excited in the pair of lines. The superposition of these components at A1 must give the appropri-

ate continuous boundary condition to the incoming wave. Thus we have:

$$V_{A1} = V_e + V_o$$
$$I_{A1} = I_e + I_o$$
$$V_{Q1} = V_e - V_o$$
$$I_{Q1} = I_e - I_o$$

Since the wave is essentially unchanged on A, we have:

$$V_{A1} = V_e + V_o = V$$

Finally at the point Q_1, we must have $I_e = I_o$ for the total current to be zero (current conservation). We define the (voltage) coupling to be:

$$K_V = \frac{V_{Q1}}{V} = \frac{V_e - V_o}{V_e + V_o} = \frac{I_e Z_e - I_o Z_o}{I_e Z_e + I_o Z_o} = \frac{Z_e - Z_o}{Z_e + Z_o}$$

This fixes the coupling as a function of the coupled impedances, which can be estimated for striplines as shown before. If Q is terminated at Q_1 or if the coupling causes a reflection, the value of K_V is only decreased-- we have calculated a realistic worst case.

- **Matrix Formulation of Multi-Conductor Lines**

We can represent the coupling capacitances between 2 or more conductors in a compact form as follows:

$$\vec{Q} = C\vec{V}$$

$$\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} C_{11} & -C_{12} \\ -C_{21} & C_{22} \end{bmatrix} \bullet \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

Where C_{11} and C_{22} are C in the symmetric case, C_{12} and C_{21} are C_m. The minus signs in the matrix indicate the change in potential from the induced charge. Similarly, we can define an inductance matrix:

$$\vec{V} = L\dot{\vec{I}}$$

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \bullet \begin{bmatrix} I_1 \\ I_2 \end{bmatrix}$$

again $L_{11} = L_{22} = L$ and $L_{12} = L_{21} = L_m$ in the symmetric case, where L is the inductance per unit length and L_m is the mutual inductance.

For a homogenous medium (one in which the dielectric constant and the magnetic permeability are constant for all parts of the wave field) we have:

$$\frac{\hat{I}}{v^2} = LC$$

or that the L and C matrices are inverses of each other and v is the velocity of propagation. (Note this implies that the odd and even components of the wave travel at the same velocity in a homogenous medium). In an inhomogeneous medium, the odd and even components will travel at different velocities given by:

$$v_e = \frac{1}{\sqrt{(L + L_m)(C - C_m)}}$$

$$v_o = \frac{1}{\sqrt{(L - L_m)(C + C_m)}}$$

since the even and odd capacitances and inductances can be expressed as:

$$L_e = L + L_m$$

$$C_e = C - C_m$$

$$L_o = L - L_m$$

$$C_o = C + C_m$$

Finally, the even and odd impedances:

$$\left(Z_e = \sqrt{\frac{L+L_m}{C-C_m}}\right), Z_o = \sqrt{\frac{L-L_m}{C+C_m}}$$

just as one would suspect. These relations hold true for both homogenous or inhomogeneous systems. Note that Z_eZ_o is Z_0^2 in the case of weak coupling. In a homogenous system, we have:

$$K_V = \frac{C_m}{C} = \frac{L_m}{L}$$

- **Inhomogeneous Media**

In a medium where not all of the wave propagates in the same dielectric constant (or possibly the same magnetic permeability), the two modes will propagate at differing velocities. This acts to degrade the signal transmission and to increase the noise coupling. For multiple conductors:

$$K_{C_j} = \frac{\sum_{i \neq j} C_{ij}}{C_{jj}}$$

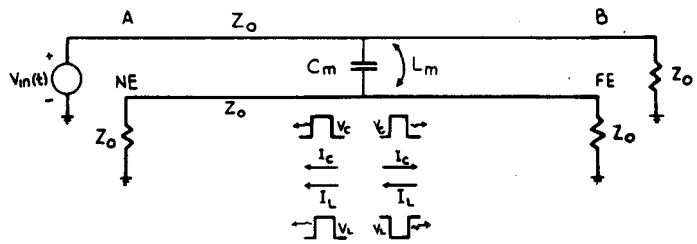
$$K_{L_j} = \frac{\sum_{i \neq j} L_{ij}}{L_{jj}}$$

and the total coupling is just:

$$K_V = \frac{K_C + K_L}{2}$$

(In the homogenous case, $K_C = K_L = K_V$). In a properly terminated line in a homogenous system, the capacitive coupling and inductive coupling are of equal magnitude but have opposite sign. Thus the total power coupled into an adjacent wire will cancel in the regions far from the disturbance-- so no cross talk will be perceived. On the other hand in an inhomogeneous system (for example a micro stripline with differing dielectric constant on the top and the bottom of the stripline) these two wave will propagate with different speeds or alternatively, the capacitive and inductive couplings will not cancel and noise will be perceived at the far end.

Consider two weakly coupled transmission lines shown below:



As usual we assume that the transmitted wave is unaffected by the coupling as it is weak. We consider the point of coupling to

be the present position of the voltage step. We shall call point A the near end and B the far end of the transmission lines. The induced current in the quiet line is of opposite magnitude for capacitive coupling and propagates in both directions from the point of coupling. The inductive coupling is based not on the voltage but on the current change, it also propagates in both directions, but carries the same sign as the excitation in the forward direction (toward the far end), and the opposite sign in the reverse direction (toward the near end).

At the near end, the coupling currents of both pulses add constructively (in direct analog to the ground plane current) and continue for the round trip time of the pulse along the transmission line. The near end pulse is just an attenuated version of the excitation pulse, with width equal to the round trip time:

$$V_A(t) = K_A(V_i(t) - V_i(t-2\tau_f))$$

$$K_A = \frac{v}{4}\left(Z_0 C_m + \frac{L_m}{Z_0}\right)$$

Note that this effect would be more complex if the line was not terminated at the near end due to the reflection of the wave in the quiet line. If the rise time is sufficiently slow, the amplitude of the coupling will be less since the returning wave will subtract potential before the incident potential has reached its peak value.

At the far end the capacitive and inductive currents are of opposite values. In the homogenous case, the currents will be equal and no current will be detected at the far end. (Note that the termination of the near end is critical to achieve this -- since the reflected near end wave will eventually arrive at the far end and be perceived as noise. In the general case, however, the two waves travel at different velocities and so will not cancel at the far end. In fact, to first order, the far end noise amplitude is proportional to the slope of the incident wave and the length of the line:

$$V_B(t) = K_B l \left.\frac{dV_i}{dt}\right|_{t-t_f}$$

$$K_B = \frac{1}{2}\left(Z_0 C_m + \frac{L_m}{Z_0}\right)$$

This is just the difference in potential between the two points on the two waves which are slowly diverging. To illustrate these formula, in a homogenous medium we have:

$$\frac{C_m}{C} = \frac{L_m}{L}$$

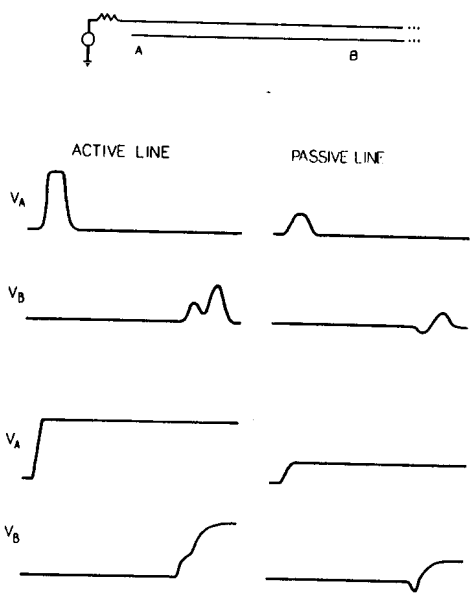
so that $I_C = I_L$ and then K_A is:

$$K_A = \frac{1}{4\sqrt{LC}}\left(\sqrt{\frac{L}{C}}C_m + L_m\sqrt{\frac{C}{L}}\right) = \frac{1}{4}\left(\frac{C_m}{C} + \frac{L_m}{L}\right) = \frac{C_m}{2C}$$

KB = 0. The factor of two in the value of the coupling reflects the termination at the source end. (The other coupling value assumed an unterminated end.)

• Transmission on Inhomogeneous Lines

On inhomogeneous coupled lines, the transmission energy is split between the two modes of the wire, the odd and even modes. In an inhomogeneous line due to uneven dielectric constants of surrounding materials, the inductance of both waves is the same. However, the capacitance of the odd wave is larger than that of the even wave so the odd wave has a slower velocity. Remember that $v=(LC)^{-1/2}$. Thus the two waves diverge in both coupled wires as shown below. The Active line has both positive components, with the primary amount of energy in the even mode. Far down the line, the active line will have split into two separate modes, with the faster even mode first. In the passive coupled line, the same splitting occurs, with the even mode first as before. In this case, however, the slower odd mode is of opposite polarity and creates an odd pulse.



• Power Supply Level Noise

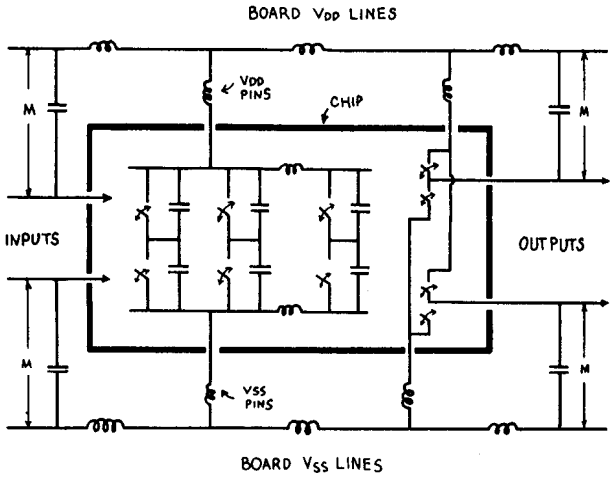
Much of the noise associated with CMOS circuitry is power supply level noise, caused by large current switching and associated back EMF. This phenomena increases with improved switching speeds and with longer interpackage wires. This noise in the power supply affects the circuits in 3 ways:

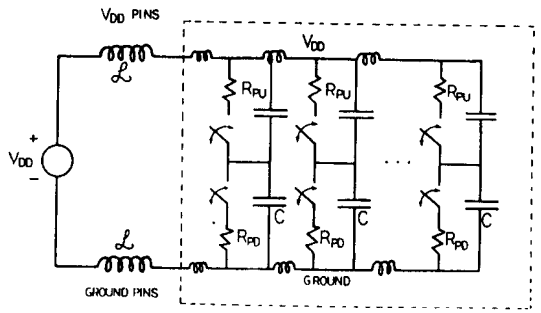
1. The decrease in available current can increase the communication delays by reducing the switching current available to drive the external connections.
2. As mentioned before, the change in available current can cause level shifting in the receiver circuits, leading to delayed of faulty reception.
3. Other gate on the sending chip can be affected by the change in voltage, leading to increased on-chip delays or internal faults as well.

The power supply noise can have both fast and possibly quiescent changes, the fast changes do to current switching, and the quiescent changes due to resistance based voltage loss on interior portions of the chip. (Properly designed chips do not usually see large resistive losses, but the effects can be often seen in gate arrays or EPLD's where the availability of arbitrary width conductors is limited. We shall be primarily concerned with the inductive component:

$$\Delta V = -L\frac{dI}{dt}$$

Below are models for off-chip and on-chip power supplies. The key difference between the two situations is the lack of local return paths for current of the outputs. For on-chip drivers, the current return is in local on-chip wiring which can be made to have very low inductance.





A key phenomena in the reduction of the transient power supply noise is the action of the internal capacitances of the chip which were charged by the supply prior to the switching event. It is this capacitance which provides the last line of defense from dropping power supply levels. Often, these circuits are also externally decoupled by the addition of a parallel capacitor. Note, that the parallel coupling of the internal capacitance and the external inductances of the pins can lead to resonances if the channel resistances are low enough-- we shall come back to this.

- **Driver turn-off**

A similar phenomena is perceived when current changes from some large value to a much smaller one. In this case, the power supply changes by increasing the potential between Vdd and Gnd. Although a common occurrence in current mode logic such as ECL and TTL, this is not a problem in CMOS as the loads are primarily capacitive-- leading to an exponential decay of current as the load is charged. Thus there is no rapid turn-off associated with internal loads. (This can still be a problem with external pins.)

- **Bounds on Inductance**

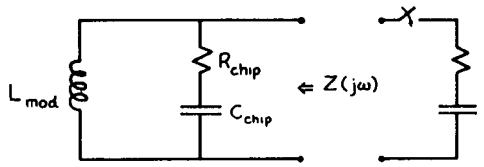
To ensure that the inductance effects are only a perturbation on the capacitive effects of the circuit, we have the following condition:

$$\frac{L}{R} \ll RC$$

This relation insures that the back emf never becomes large enough to disturb qualitatively the RC dominant circuit effects. Note, that as switching speeds improve, R is necessarily decreased, increasing the difficulty of maintaining relatively small values of L/R.

- **Power Line Resonance**

Due to scaling of components, the total chip capacitance C which is charged by Vdd and Vss tends to increase. At the same time the channel resistance of the devices is constant -- but the number of such devices is rapidly increasing. Thus the total chip resistance is decreasing. Consider the diagram below:



The inductance L is the pin inductance of the power supply-- and the total circuit is a classical lossy tank. We can calculate the effects of resonance by using a sinusoidal impedance calculation. We have:

$$Z(iw) = \frac{1}{\frac{1}{iwL} + \frac{1}{R + \frac{1}{iwC}}}$$

$$Z(iw) = \frac{-w^2RLC + iwL}{(1 - w^2LC) + iwRC}$$

The resonance condition occurs at the maxima of the impedance-- or when the magnitude of the denominator is minimal. This is close to the frequency: $w_0 = (LC)^{-1/2}$. At this frequency, the magnitude of the impedance is:

$$|Z(iw_0)| = \sqrt{\frac{L^2 + LCR^2}{R^2 + C^2}}$$

We wish to have the resonant frequency as large as possible and the resonant impedance as small as possible. Failing this, the operation of the chip for certain data inputs can provide resonant excitation of this circuit, increasing the power supply noise on each consecutive cycle until a fault occurs. Typical values for C,R, and L are: 10,000pF, 0.1Ω, and 0.5nH leading to $w_0 = 2\pi *$ (71MHz). For this case, the impedance is about 1/2 Ω. To increase the frequency, we must reduce the inductance of the pins. We can reduce the magnitude of the impedance by increasing the internal capacitance of the chip -- or by external decoupling. Alternatively, we can simply use more pins for power and ground -- sharing the current among the pins decreases the effective impedance. It is for this reason that the tiny chip (and our larger chip) have several copies of ground and Vdd pins, and they are located uniformly around the periphery of the chip.

- **Decoupling Capacitors**

Although decoupling capacitors provide help in that they shield the active chip from the inductances of the power supply lines on the circuit board before they get to the chip, these capacitors,

cannot shield inductive effects of the pins themselves as the back emf is developed across the pin structure. The value of the capacitance should be large enough to supply energy during the switching transient, but not so large that it lowers the package resonant frequency. Finally, the decoupling capacitor should be as close to the power supply terminals as possible and have as low an internal inductance as possible. (Internal loss in the decoupling capacitor is valuable as it reduces the magnitude of the resonance impedance).

Advanced packages can sometimes carry their own decoupling capacitors on board-- further reducing the problem for high speed chips or modules. (SIMM modules carry their own decoupling capacitors). Finally, we can increase the chip's own internal capacitance to reduce the magnitude of the current surge. This must be done carefully, as this also reduces the frequency of the chip resonance.

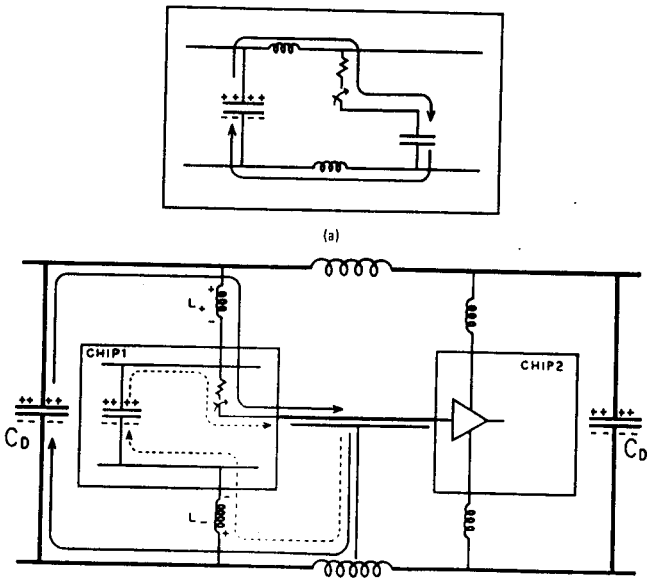
We can model the voltage drop during a transient for a decoupling capacitor by assuming that the power distribution network behind the decoupling capacitor is infinite inductance. Then the total chip switching capacitance charging current must be supplied by the decoupling capacitor:

$$\Delta V = -\frac{C}{C_D + C} V_{dd}$$

So a 10nF chip (100k gates) would need about 10x (for 9% drop) or about 100nF = 0.1μF.

- **On-Chip Decoupling Capacitors**

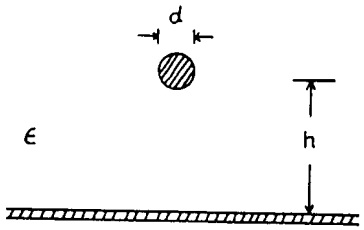
We can ameliorate the problem of pin impedance on the power supply lines by judicious use of on-chip decoupling capacitors. In this case the power and ground potential is supported by two capacitances: the source to substrate diffusion capacitors and the Vdd to Gnd metal plane overlap capacitance. Internal switching of loads then can draw charge directly from the local interconnection capacitance without leaving the chip. This source of current is not available for driving external loads on the chip periphery as these loads complete their circuit via the external ground plane and Vdd connections. These connections are shown below:



In fact, significant internal capacitance can be damaging to the external noise figure. If the chip switches large external currents and the source of charge is internal, the same inductive impedance of the pins will now produce a potential between the chip power and ground lines and the external power and ground lines. This has the effect of modulating the incoming external signals by the difference in potential. For this reason, in high performance designs, the external drivers have their own power and ground sources and their own pins, electrically separate from those of the chip interior.

- **Minimizing Pin Inductance**

Below are the electrical characteristics of a wire above a ground plane. It can be seen that shortening the wire or making it closer to the ground plane lowers the inductance.



$$C = \frac{2\pi\epsilon}{\ln\left(\frac{4h}{d}\right)}$$

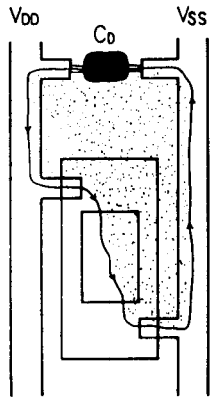
$$L = \frac{\mu}{2\pi}\ln\left(\frac{4h}{d}\right)$$

$$Z_0 = \frac{1}{2\pi}\sqrt{\frac{\mu}{\epsilon}}\ln\left(\frac{4h}{d}\right)$$

Dip pins have the highest inductance due to long lead lengths and no ground plane. Pin grid arrays have much smaller inductances (3-10nH) due to integrated power and ground planes. Multi-chip packaging using flip chip solder bumps have the lowest inductance of 0.25-1 nH. A second technique is to increase the number of power and ground connections to share the current flow. This is most effective if the pins are distributed geometrically around the die to reduce the size of the external ground loop. Since parallel inductances add like parallel resistances, we have:

$$L_t = \left(\frac{1}{L_1} + \frac{1}{L_2} + \frac{1}{L_3} + \ldots\right)^{-1}$$

So adding additional ground and power pins has the effect of reducing the inductance substantially. On the larger scale, the ground loop is the current loop between the current source and the chip power pins as shown below. To reduce the inductance, we can merely reduce the area of the current loop and thus the confined flux.



- **Example**

Consider a CMOS chip with 32 low-impedance output buffers are driven simultaneously. If the line impedance is 50Ω and the expected rise time is 2nS and we expect a 5V drive we have:

$$\frac{dV}{dt} = \frac{0.8V}{t_r} = 2\frac{V}{nS}$$

The required current (into the transmission lines) is:

$$\frac{dI}{dt} = \frac{1}{Z_0}\frac{dV}{dt} = 40\frac{mA}{nS}$$

For 32 drivers the current transient is:

$$\frac{dI_t}{dt} = 1.24\frac{A}{nS}$$

Thus to achieve a voltage drop of at most 0.5V the effective inductance of the package pins and decoupling connections must be:

$$L_{max} \leq \frac{\Delta V}{(dI)/(dt)} = 0.4nH$$

This would require 10 paris of power and ground connections with effective inductance of 4nH per pair. One can easily gauge the magnitude of this noise problem.

- **Effects of scaling on power supply noise**

The effective gate delay of CMOS components decreases by S under ideal scaling. We can not expect the chip speeds to fall at this same rate due to the increases in die size and interconnection overhead, but we can assume that the internal rise times could rise this fast in the worst case: $dt \propto 1/S$

Chip current increases rapidly with scaling since the number of components increases as: $S^2S_c^2$. The current per gate decreases since the voltage swing decreases, while the channel resistance remains relatively constant. Thus the total current drive scales as:

$$I_t = SS_c^2$$

Current demand is in the form of interconnect and gate capacitances. The amount of interconnection capacitance scales as:

$$C_I = S^2S_c^2$$

while the total gate capacitance is the number of gates times the size of gates times the capacitance per unit area:

$$C_G = S^2S_c^2 \times \frac{1}{S^2} \times S = SS_c^2$$

So the average current demand is the voltage swing times the total capacitance divided by the rise time:

$$I_D = \frac{(C_I + C_G)\Delta V}{dt} \propto S^2S_c^2$$

These factors are shown in the following table for both ideal and an improved form of scaling where the chip inductances are

scaled by use of a multiple number of power connections spread across the entire chip. To achieve this level of inductance reduction, we must nearly dedicate two interconnection levels for power and ground respectively.

Parameter	Ideal Scaling	Improved Scaling
Rise time (t_r)	$1/S$	$1/S$
Total current requirement of the chip (I_{TOT})	$S^2S_C^2$	$S^2S_C^2$
Rate of change of current ($dI/dt \propto I_{TOT}/t_r$)	$S^3S_C^2$	$S^3S_C^2$
Number of power connections (N)	1	SS_C^2
Effective inductance per connection (L)	S_C	$1/S^2$
Voltage fluctuation $[\frac{L}{N}(dI/dt)]$	$S^3S_C^3$	1
Signal levels and noise margins (V_{DD})	$1/S$	$1/S$
Signal-to-noise ratio $[V_{DD}/L(dI/dt)]$	$1/S^4S_C^3$	$1/S$

S: Scaling factor for device dimensions
S_C: Scaling factor for chip size