Mixed Signal IC Design
Notes set 3:
More High Frequency Amplifier Design

Mark Rodwell
University of California, Santa Barbara

rodwell@ece.ucsb.edu  805-893-3244, 805-893-3262 fax
Standing waves on transmission lines cause gain/phase ripples of the form \((1 - \Gamma_S \Gamma_L \exp(-j2f\tau))^{-1}\). Either we must have short transmission lines, or the lines must be well-terminated.
Why do we care about impedances matched to 50 Ohms?

If IC has small dimensions, we can use high - z internal interfaces, and just interface to e.g. 50 Ohms at chip boundaries or (in some cases) between major circuit blocks on wafer.
Why do we care about transmission lines?

There are times when we would like to ignore transmission lines and S-parameters.

If the IC has interconnects which are short, then we can treat all interconnects as **lumped**.

This means we can ignore standing waves on the lines.

But, we must be aware that \( C_{\text{line}} = \tau_{\text{line}} / Z_0 \), \( L_{\text{line}} = \tau_{\text{line}} * Z_0 \), hence a short line can be modeled as lumped but its parasitics **cannot be ignored**.
Here we can model these short lines as lumped. Ignoring BJT parasitics,

\[ C_{\text{line}} = \tau_{\text{line}} / Z_o, L_{\text{line}} = \tau_{\text{line}} * Z_o, \]

And we get an RC charging time of

\[ \tau_{\text{wiring}} = C_{\text{line}} R_L \]

Since the gain is \( g_m R_L = (kT / qI)R_L \), designing for low power (low I) means designing for high \( R_L \) and hence interconnect capacitance becomes more important
Once we have studied EF frequency response in detail, we will understand that

\[ L_{\text{line}} = \tau_{\text{line}} \times Z_o, \]

...can have substantial effect upon the EF stability and pulse response
Getting more bandwidth

At this point we have learned basics (MOTC etc)

How can we get more bandwidth…. ?
   Resistive feedback
   transconductance-transimpedance
   emitter-follower buffers, benefits and headaches
   emitter degeneration…basics
   emitter degeneration and area scaling
   ft-doubler stages..
   distributed amplifiers
   ft-doubler stages…simple, RL, power, with Darlington
   broadbanding / peaking with LC networks….
   distributed amplifiers
   examples….
Broadband gain blocks: "worlds simplest amplifier"

Examine a cascade of simple differential pairs. For the differential signal path, this becomes a CE stage cascade. If we model each BJT as a simple $g_m$ & $C_{in}$ combination, the gain per stage is $-g_m R_C$, the bandwidth is $1/2\pi R_C C_{in}$, and the gain - bandwidth product is $g_m / 2\pi C_{in} = f_\tau$. This is the origin of the "$f_\tau$ limit". We can do much better.
Broadband gain blocks: "worlds simplest amplifier"

Let us think this through. Transistors are made to provide transconductance $g_m$. An input voltage is converted into an output current. Our simplest design strategy is to convert this output current back into a voltage using a resistor...hence the gain relationship. Transconductance *always* comes at the price of input capacitance, by the ratio $C_{in} = g_m / 2 \pi f_c$.
This is a more sophisticated way of converting $g_m$ into voltage gain....nodal analysis to be done in lecture....

Per-stage gain of $A$ (negative) with input and output resistances of $R_{in/out}$ are obtained if we set

$$g_m = \frac{1 - A}{R_{in/out}} \text{ and } R_f = (1 - A) \cdot R_{in/out}.$$
Resistive feedback stages, I: bandwidth and gain-bandwidth

Since $g_m = (1 - A) / R_{in/out}$ and $C = g_m / 2\pi\tau$,

$$f_{3dB} = 1 / (2\pi (R_{in/out} / 2)C) = g_m / (\pi C (1 - A))$$

$$f_{3dB} = 2f_\tau / (1 - A)$$

Note that for $A >> 1$, this is *twice* the bandwidth of a simple CE stage!
Why have we nearly doubled the bandwidth?

The resistive feedback has provided *output *impedance of $R_{in/out}$ without losing significant output current in a physical loading resistance. We need roughly $1/2$ the $g_m$ to obtain the desired gain at a given impedance....so we get $1/2$ the $C$, and hence twice the bandwidth...
Variations on resistive feedback.

We can use sets of transistors to build the $g_m$ block. This block can then be used with either resistive loading or resistive feedback to build an amplifier.
Feedback with buffer

This is often used for ease in biasing. The required $g_m$ is also reduced to $g_m = -A/R_{in/out}$.

This improves bandwidth for low-gain stages, though the improvement is partially offset by the additional capacitances of the EF stage.
Resistive feedback as 50 Ohm gain blocks

Note the gain relationships...
High Speed Amplifiers

18 dB, DC-50 GHz

8.2 dB, DC-80 GHz

Gains, dB

Frequency, GHz
First stage is resistively terminated, second stage has resistive feedback...this is because first stage is allowed to limit, or can have AGC applied, either of which would violate the feedback gain/impedance relationships...
Cherry-Hooper Gm-Zt amplifier

Very popular feedback varaint. Here input/output impedances are not consistent. Input impedance of 2nd stage is

\[ R_{in} = \frac{(R_c + R_f)}{(1 + g_{m2} R_c)} \rightarrow \frac{1}{g_{m2}}. \]

Transimpedance of second stage is (derive these)

\[ Z_{T2} = R_f \left( g_m - G_f \right) \left( g_m + G_C \right) \rightarrow R_f. \]

Overall gain is

\[ A_v = g_{m1} Z_{T2} \]
Cherry-Hooper Gm-Zt amplifier

Let us work through impedances and gains both carefully and by back-of-envelope…
…why is first order time constant so small?
Cherry-Hooper Gm-Zt amplifier--limiting behavior

If $I_1 < I_2$ (derive this) amplifier will always limit on the gm stage, not the Zt stage....hence large - signal operation maintains low impedances (hence short time constants) provided by feedback.
Cherry-Hoopers: DC bias variants
Cherry-Hooper bandwidth---zero'th order

By inspection,

\[ a_1 = \left( C_{be2} + C_L \right) / g_{m2} \]

\[ a_2 = R_L C_{be2} C_L / g_{m2} \] (work more detailed example in lecture)

Note immediately that first - order time constant is small,
second - order time constant is NOT.

Much more detailed analysis in Hitachi paper....
Gm-Zt amplifier: understanding time constant behavior

This has input and output impedance of

\[ R_{in} = R_{out} = \frac{1}{g_{m1}} \]

So each capacitor has charging time constant of \( \frac{C}{g_{m1}} \)

Shorting \( C_2 \) makes the \( C_1 \) charging time constant \( = C_1 R_f \)

so terms in a2 are of the form \( (\frac{C_2}{g_m})C_1 R_f \).
Cherry-Hooper bandwidth---with Ccb

\[ a_1 = \frac{C_{be2}}{g_{m2}} + \frac{C_{be3}}{g_{m2}} + C_{cb} R_f \]

\[ a_2 = \left( \frac{C_{be2}}{g_{m2}} \right) (C_{be3} R_f) + \left( \frac{C_{be2}}{g_{m2}} \right) (C_{cb} R_f) + \left( \frac{C_{be3}}{g_{m2}} \right) (C_{cb} R_f) \]

Note that Ccb has reduced bandwidth and has improved damping.
Darlingtons 1

Why not just use emitter followers to buffer the stage input capacitances ????
Darlington 2

Taking $R_{ex} = 0$ and $\beta = \infty$...

\[ a_1 = C_{be2} (R_{bb2} + r_{e1}) \]
\[ + C_{cb2} \left( (R_{bb2} + r_{e1}) (1 + g_{m2} R_L) + R_L \right) \]
\[ + C_{cb1} (R_{bb1} + R_1) + C_{be1} (0!) \]
\[ a_2 = C_{be2} C_{be1} (R_{bb2} + r_{e1}) (R_1 + R_{bb1}) \left( \frac{r_{e1}}{(r_{e1} + R_{bb2})} \right) + \ldots \]
Note the greatly reduced charging time for $C_{be2}$:

$$a_1 = C_{be2} (R_{bb2} + r_{e1})$$

instead of

$$a_1 = C_{be2} (R_{bb2} + R_1)$$

This is the motivation for the Darlington commonly given in introductory texts. But a more detailed consideration reveals that (1) the charging time for $C_{be2}$ is at a minimum $C_{be2} R_{bb2}$ (more on this later!) and (2) that $C_{be2} (R_{bb2} + R_1)$ **must reappear** in the second-order time constant, with consequent reduction in circuit damping.
Darlingtons

Clearly, exact problem is hard....lets look at effect of $C_{be}$ alone (all other BJT parasitics are zero). Then

$$a_1 = C_{be2} r_{e1}$$

$$a_2 = C_{be2} C_{be1} r_{e1} R_1$$

If we have a cascade of stages, then note that $R_L = R_1$ and that gain $= A = R_1 / r_{e2}$. The damping factor is proportional to the ratio $a_1 / \sqrt{a_2}$. Note we can make the first - order time constant very small, but not the second - order time constant ...(work through...)...attempts to broadband response result in loss of damping
Darlington: improving the damping

Addition of $R_{ee}$ has 2 effects. The Q1 EF gain is reduced

$$a_1 = C_{be1} \left( (R_1 + R_{bb1})(1 - A_{v,ef}) + \ldots \right) + C_{be2} \left( R_{bb2} + R_{ee} || R_{out,ef} \right)$$

$$a_2 = C_{be1} \left( (R_1 + R_{bb1})(1 - A_{v,ef}) + \ldots \right) C_{be2} \left( R_{ee} || (R_1 + R_{bb1}) \right)$$

So, the terms in $a_1$ associated with $C_{be1}$ are increased. More importantly, terms in $a_2$ associated with $C_{be2}$, (e.g. $R_{22}^1$) are reduced due to the parallel loading of $R_{ee}$.
Getting more bandwidth

At this point we have learned basics (MOTC etc)

How can we get more bandwidth…. ?
  Resistive feedback
  transconductance-transimpedance
  emitter-follower buffers, benefits and headaches
  emitter degeneration…basics
  emitter degeneration and area scaling
  ft-doubler stages..
  distributed amplifiers
  ft-doubler stages…simple, RL, power, with Darlington
  broadbanding / peaking with LC networks….
  distributed amplifiers
  examples…. 