2x.4 BJT Amplifier Distortion: a SPICE Exploration

Bipolar transistors are remarkably useful when used as signal amplifiers, but when used in simple circuits they do have serious imperfections, such as distortion. We know that feedback around multiple stages of a high-gain amplifier works well to reduce distortion, but it's useful to explore the BJT-amplifier distortion scene without the use of feedback. Transistor operation is well explained by semiconductor theories such as the Ebers–Moll base-voltage collector-current equations, the Early effect, and the Gummel–Poon charge equations. Formulas from these theories have been incorporated into the SPICE circuit analysis engine developed at UC Berkeley. SPICE can be used for accurate analysis of transistor-circuit performance, provided accurate component models are used. In this section we'll do some "lab experiments," but without the lab bench. Instead we'll use the free demo version of IntuSoft's SPICE program, which they call ICAP/4; download it and follow along with us! Readers unfamiliar with SPICE are advised to read the brief SPICE Primer in Appendix J.

2x.4.1 Grounded-emitter amplifier

Here we take up in more detail the subject first explored in §2.3.4 and Figure 2.46, and especially the circuit we introduced in §2.3.5B and Figure 2.52A.

We've chosen the 2N5088 transistor (the 2N5088 and 2N5089 are popular parts in good audio preamplifiers) in a simple common-emitter (CE) amplifier circuit. We biased it at I_C=1 mA and V_C=10 V (half the supply voltage), and evaluate the amplifier's gain and distortion. Figure 2x.15 shows the circuit entered into the SpiceNet schematic capture and SPICE management program. As shown, we assigned values of AC=1 and DC=20 to the V1 and V2 voltage generators.

First let's check the frequency response of the amplifier. We click the SIMULATION SETUP button (pencil over wavy line; or ACTIONS | SIMULATION SETUP | EDIT) and select AC ANALYSIS. We'll use 20 points per octave, from 1 Hz to 10 MHz, and we'll SAVE the project at this point. NOTE: If you enter "10M" for the ENDING FREQUENCY, you will get an error telling you that F_{ending} must be greater than $F_{starting}$; that's because SPICE (which doesn't distinguish upper and lower case) interprets "M" as milli: you must enter "10meg."

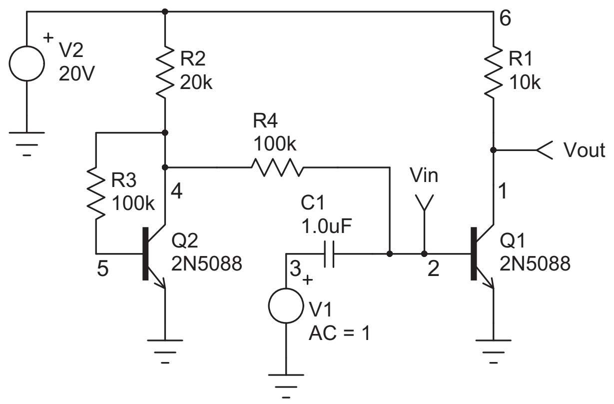

Figure 2x.15. First circuit: grounded-emitter amplifier.

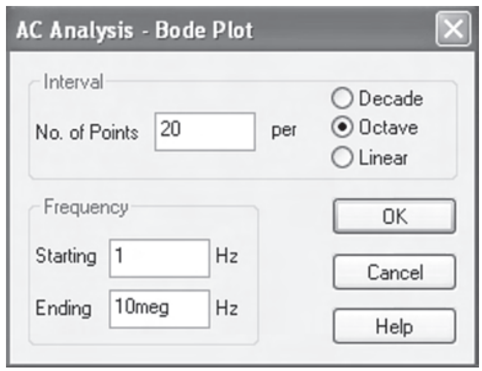

Figure 2x.16. Frequency response setup screen.

Clicking the RUN SIMULATION button (running person button) starts the IsSpice4 SPICE engine, and launches a new window (ISSPICE4) and associated subwindows (SIMULATION STATUS, SIMULATION CONTROL, OUTPUT, and ERRORS AND STATUS); it may also launch the IntuScope display window, with associated subwindows ADD WAVEFORM and SCALING. Next we'll press EDIT TEXT (SpiceNet window: pencil over paper, or ACTIONS | TEXT EDIT), which launches a new window (ISED), and select the OUT file (use the WINDOW tab), so we can check the DC OPERATING CONDITIONS that SPICE calculated for our circuit before doing the AC ANALYSIS. Part way down we see

```
    Node    Voltage
***
 V( 1 )    9.564778e+000
 V( 2 )    4.817968e-001
```

(The node numbers may differ, depending on the order in which you placed the parts.) This tells us the collector is sitting at 9.56 V, close enough to our 10 V goal, and the collector current must be close to 1 mA. The base-emitter voltage is biased at 482 mV.

Further down in the OUT file we see hundreds of frequency response datapoints, but it's more convenient to plot the data than to examine these numbers, so we click the SCOPE button (crosshairs with sinewave, or ACTIONS | SCOPE) to bring up the IntuScope display program. Once there we click ADD WAVEFORM (if the dialog hasn't started automatically), check the box labeled TEST PTS ONLY, select the VOUT signal, and click ADD. We adjust the plot by clicking OPTIONS and selecting LOG-LOG under GRAPH TYPES. Expand the outer window enough so you can resize the graph itself by dragging the corner. Rename the vertical axis (double click on the default label, then uncheck AUTO GENERATE LEGEND). You should get something like Figure 2x.17.

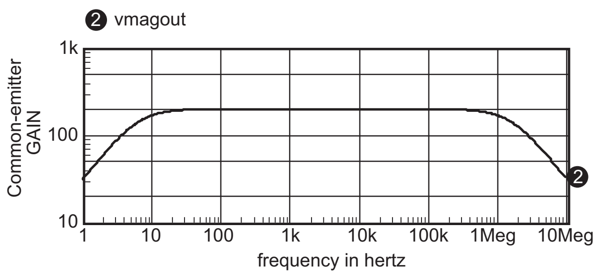

Figure 2x.17. Frequency response, 1 Hz–10 MHz, of the circuit of Figure 2x.15.

The frequency response is flat from about 20 Hz to 500 kHz. The grounded-emitter amplifier gain formula is $G{=}g_m R_L/r_e$ (where the dynamic resistance $r_e{=}1/g_m{=}V_T/I_c$, with $V_T{=}kT/q{=}25$ mV at room temperature). Since $r_e{=}25\,\Omega$ at 1 mA (a handy number we should all have memorized), we expect the gain to be about 400, rather than 200 on the plot above. That's a factor-of-two error; something must be seriously wrong!

2x.4.2 Getting the model right

SPICE can provide accurate answers, but only if it is given accurate component models. The SPICE engine has an accurate built-in processing model for BJTs, which it runs based on parameter values it gets from a component library. Let's use EDIT TEXT to examine the transistor model for the 2N5088 in our SPICE library. The part we used is identified by IntuSoft as a generic "amplifier" part, labeled QN5088. The model looks like this:

```
.MODEL QN5088 NPN BF=780 BR=4 CJC=7.83P
+ CJE=11.8P IKF=30M IKR=45M IS=21.0P ISE=41.8P
+ NE=2 NF=1 NR=1 RB=92.6 RC=9.26 RE=23.1
+ TF=3.18N TR=127F VAF=98.5 VAR=18 XTB=1.5
```

The terms RB, RC, and RE are external resistances that SPICE adds to the three transistor pins in its internal model. Do these values make sense? Looking at a 2N5088 datasheet, we see that the transistor works well to 30 mA or more. We know that the dynamic emitter resistance r_e should be less than an ohm at 30 mA, so how are we likely to fare if an additional 23 Ω is added by the model? It's going to be a disaster!

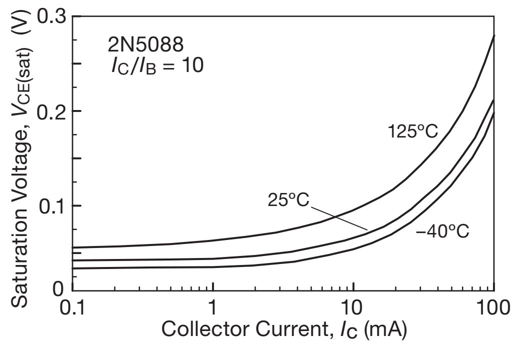

Figure 2x.18. Checking the model: datasheet's $V_{CE(sat)}$.

What value should RE have? We can get an idea from a datasheet plot of the 2N5088's saturation voltage, $V_{CE(sat)}$ (Figure 2x.18). We see that V_{CE} is about 210 mV at 100 mA, and perhaps 75 mV of that is due to simple transistor action. So the sum of RC and RE should be no more than $R{=}135\,\text{mV}/100\text{mA}$, or 1.35 Ω. Compare that to 32.4 Ω in the model, whew! We could modify the values for the QN5088 part, but rather than trust the rest of the model, let's see if we can use another part in the library. Under the Fairchild heading in IntuSoft's library, we find both a 2N5088 and a 2N5089. Here's what you see if you drop those parts into the schematic, then look at the *.out file:

```
.MODEL 2N5088F NPN BF=1.122K BR=1.271
+ CJC=4.017p CJE=4.973p EG=1.11 FC=.5
+ IKF=14.92m IKR=0 IS=5.911f ISC=0
+ ISE=5.911f ITF=.35 MJC=.3174
+ MJE=.4146 NC=2 NE=1.394 RB=10 RC=1.61
+ TF=821.7p TR=4.673n VAF=62.37 VJC=.75
+ VJE=.75 VTF=4 XTB=1.5 XTF=7 XTI=3
.MODEL 2N5089 NPN BF=1.434K BR=1.262
+ CJC=4.017p CJE=4.973p EG=1.11 FC=.5
```

```
+ IKF=15.4m IKR=0 IS=5.911f ISC=0 ISE=5.911f
+ ITF=.35 MJC=.3174 MJE=.4146 NC=2 NE=1.421
+ RB=10 RC=1.61 TF=822.3p TR=4.671n VAF=62.37
+ VJC=.75 VJE=.75 VTF=4 XTB=1.5 XTF=7 XTI=3
```

The first thing we see is that there's no RE at all in these models (SPICE assumes RE=0), and RC is 1.61 Ω, so that looks pretty good. The only difference we see between the 2N5088 and 2N5089 models are BF=1.122K and BF=1.434K, the transistor's beta values. Both of these parts are on the same datasheet, which shows h_{FE} current gain (or beta) values of 350 and 450 respectively. Only minimum values are shown at 1 mA (although the typical values shown for 0.1 mA are 3× higher than the minimum values). Many manufacturers provide SPICE models for their parts, and they often choose the worst-case values. Here we are given optimistic values, but since beta isn't critical in our circuit, we'll accept that and substitute Fairchild's 2N5088 model in our circuit. (If the value of beta mattered much in our circuit, we would edit the SPICE model to match reality.)

Using the new transistor model (highlight old transistor, then DELETE key or EDIT | CLEAR; then place new part), we try the frequency response plot again, getting the plot of Figure 2x.19. Aha! Now we get a small-signal gain of 340, much closer to expectations. We also have better high-frequency response, which makes sense given Fairchild's lower capacitance values in their model.

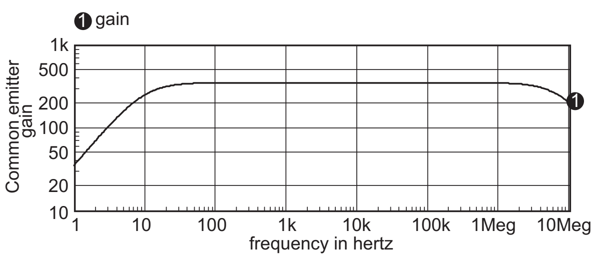

Figure 2x.19. Frequency response with better transistor model.

Rechecking the DC operating conditions, we see

```
V( 1 )    1.045805e+001
V( 2 )    6.690726e-001
```

We have 10.45 V on the collector, OK, that's fine. But it's interesting that now we have V_{BE}=669 mV, compared with 482 mV earlier. This is a more sensible value, and is due to a more realistic value I_S=5.911 fA, compared with 21 pA in the first (generic) model. Fairchild's datasheet plot (Figure 2x.20) shows about 630 mV at 1 mA, but we're happy enough with this model.

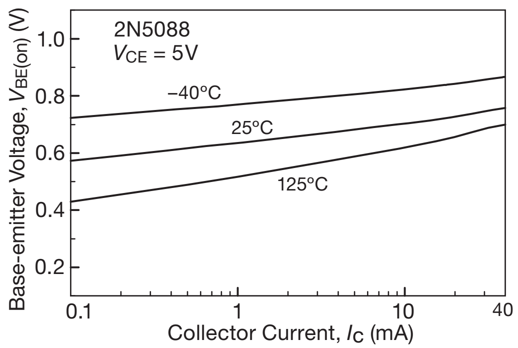

Figure 2x.20. Checking the model: datasheet's $V_{BE(on)}$.

2x.4.3 Exploring the linearity

Now we're ready to continue and evaluate the nonlinearity of the common-emitter amplifier. Going back to our schematic in SpiceNet, we double-click the V1 voltage source. In the dialog box, under TRAN GENERATORS, click on the button showing "none," and select PWL (Figure 2x.21). We'll create a triangle test waveform by entering a few time–voltage datapoints; a triangle wave input makes it easy to spot deviations from linearity.

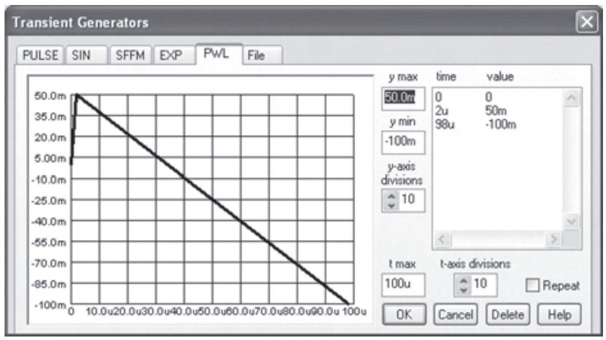

Figure 2x.21. Voltage sweep setup screen.

The gain of a common-emitter transistor stage is $G=-g_m R_L$, where $g_m=V_T/I_C$. Our concern with this equation is that a changing output requires a changing collector current, and the equation (see §2.3.4A and Figures 2.45 and 2.46) tells us to expect a substantial gain variation from this changing collector current, thereby causing high distortion. We're looking to see if the amplifier produces a rounded response to the linear triangle wave input, an indication of changing gain over the waveform, with corresponding distortion.

We have to decide what our test-voltage range should be. With a gain of 350, a +50 mV input should easily drive the collector voltage down to zero volts before we start the ramp; and we end the ramp at −100 mV to drive the output close to the positive supply rail. We select 2 μs for the initial step to 50 mV, and 98 μs to complete the ramp. This is fast enough to avoid droop from the ac coupling capacitor, and slow enough to avoid errors from the high-frequency roll-off. We enter the values with commas between the number pairs, and a CR after each pair. We select scaling factors (Y MAX, T MAX, etc.) for the plot, as shown, so the waveform will display nicely.

Next we select TRANSIENT ANALYSIS under SIMULATION SETUP, and use a 100 μs TOTAL ANALYSIS TIME and a 0.1 μs DATA-STEP TIME (1000 points) to capture our waveform.

We run the SPICE engine again, and then go to IntuScope and select FILE | NEW GRAPH. Using the ADD WAVEFORM dialog's TRAN1 mode, we select vout and click ADD. If the amplifier were linear, we'd see a linear output ramp – an inverted, amplified version of the input. Instead we see a curving, squashed output waveform (Figure 2x.22).

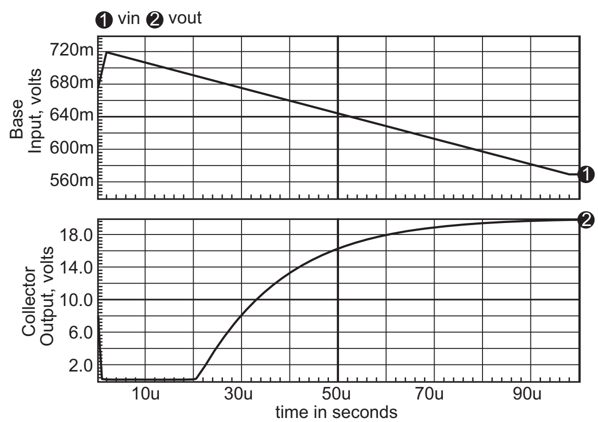

Figure 2x.22. Output response to an input ramp for the circuit of Figure 2x.15.

First, we see that a +50 mV signal (669 mV(quiescent) + 50 mV = 719 mV) was more than adequate to drive the output to zero volts; in fact the output doesn't start to rise until the base voltage falls to about 690 mV, or +21 mV input signal. This is further confirmation that the gain increases with I_c.

Looking at the shape of the output curve, we see that the first part looks reasonably linear, but the gain falls dramatically at high collector voltages (i.e., at low collector currents).

A. Input–output transfer function

Ideally we'd like to have a plot of output voltage directly versus input voltage, rather than each plotted separately versus time. We can do this on a new graph, by changing the X-AXIS choice in the ADD WAVEFORM dialog, from default−time, to the input signal V1 (we added a test point V1 on the input). Figure 2x.23 shows the result.

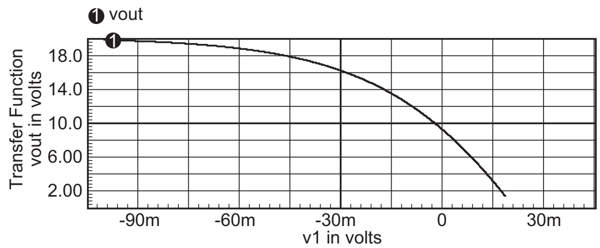

Figure 2x.23. Transfer function of the circuit of Figure 2x.15. The output goes (nonlinearly) from 20 V down to 1 V for inputs going from −100 mV to +25 mV.

At this point we also made a few more changes: we set the INITIAL CONDITION (in the SpiceNet schematic) of coupling cap C1 to 669 mV (double-click the cap and enter IC=−669m), and we checked USE INITIAL CONDITIONS (UIC) in the TRANSIENT ANALYSIS dialog. We set V1's initial value to +21 mV (in the transient generator PWL setup). We set the MAXIMUM TIME STEP to 0.02 μs, to reduce noise in the output data. Finally, we set the TIME TO START RECORDING DATA to 0.3 μs to skip past the startup transient. We refined the plot with the SCALING dialog to adjust the graph's x-axis ("x scale" = 15m, "x-offset" = −30m).

B. Gain versus input

IntuScope lets us do some valuable post-processing analysis. By differentiating the V_{out}-vs-V_{in} plot with CALCULATOR | CALCULUS, we get a plot of the amplifier's gain versus output voltage, shown in Figure 2x.24. (We adjusted the graph's x- and y-axis for appearance. If you want to follow along, highlight the gain trace, then go to the SCALING window: we used 15m for the x-scale with −30m offset, and 75 for the y-scale with −300 offset.)

Line #1 is the output voltage (from 20 V down to 0.38 V) versus input voltage, and line #2 is the amplifier's gain versus input voltage. The gain varies from nearly 0 to more than −500; the gain at the quiescent point (v1 = 0 V) is −341.6 (drag the cursor until x=0 in the display at the bottom of IntuScope; or just enter 0 there), as we saw in the small-signal frequency response plot. To take these measurements, we used IntuScope's CURSOR tool, which

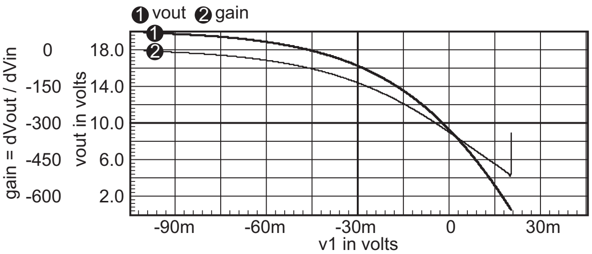

Figure 2x.24. Gain (trace 2) calculated from the transfer function (trace 1) of Figure 2x.23; the gain goes from zero to −500, ouch!

shows us X and Y values for the selected trace, as shown in Figure 2x.25. A gain varying from 0 to 500 isn't very good, but we know how to do much better.

2x.4.4 Degenerated common-emitter amplifier

Clearly a simple grounded-emitter amplifier has a great deal of distortion, but we can trade off higher gain for lower distortion by adding an emitter resistor (recall §2.3.4B and Figure 2.52B), Figure 2x.26. Let's aim for a gain of ten.

Some like to think of emitter degeneration as negative feedback. But no explicit external feedback path is involved, so purists who worry about transient response issues, etc., with ordinary feedback, needn't be concerned. Furthermore, time delays, phase shifts, and feedback stabilization issues at high frequencies are avoided with emitter degeneration.

For the simulation we set v1's PWL signal for a +1000 mV to −1000 mV range, and plotted transfer function and gain versus v1 as before. In terms of percentage gain variation, the result (Figure 2x.27) is much better (note expanded scale), but it's still a mediocre amplifier, with the gain (plot 2) changing from about G=9.8 at an output voltage of 2 V, down to 9.68 at 10 V (about 1.2% lower), and then continuing down to 8.81 at 18 V (another 9% lower).

Furthermore, the drop in gain is not symmetrical about the quiescent point, so the amplifier produces second-harmonic distortion, insulting the ears of the audiophile. If we could balance the positive and negative gain losses, the even harmonic distortion products would be eliminated. That can be done by going to a balanced (differential) circuit configuration.

2x.4.5 Differential amplifier

Realizing that our single-ended common-emitter amplifier circuit already requires two transistors (one to bias the other), we are motivated to ask why we shouldn't instead use these same two transistors in a differential amplifier. In essence, one will still be biasing the other, but in a more useful manner. If the long-tail pair (see §2.3.8) is biased with a 2 mA current, each side carries 1 mA, so the output load resistor will be biased at half the positive supply voltage (+10 V) as before. This allows for a symmetrical output swing up to ±8 V or so. Let's test[9] the circuit (Figure 2x.28) by driving it with a ±100 mV dc-coupled ramp of 100 μs duration, via the v1 PWL dialog.

The output transfer function exhibits a nice symmetry about the quiescent point, reflected in the plot of calculated gain; the symmetry implies absence of distortion at even harmonics. But there remains a large variation in gain: more than a factor of two over the 2 V to 18 V output range. That is, the amplifier continues to have substantial distortion, in fact more than that of the amplifier it replaces. We should not be surprised, though, because this circuit lacks the linearizing emitter degeneration of the previous circuit (Figure 2x.26); we'll fix that, presently.

A. Estimating the distortion

We can evaluate the amplifier's performance analytically. The derivation of the gain of a differential amplifier involves the ratio of two Ebers–Moll exponentials, which leads to the appearance of the hyperbolic tangent function.[10] For differential current $I_{\rm out}$ and long-tail current $I_{\rm E}$ we have

$$I_{\rm out}/I_{\rm E} = 2\tanh\frac{V_{\rm in}}{2V_{\rm T}},$$

where $V_{\rm T}$=kT/q=25 mV at room temperature. A series expansion gives us

$$\begin{aligned} I_{\rm out}/I_{\rm E} &= \frac{V_{\rm in}}{V_{\rm T}} - \frac{2}{3}\left(\frac{V_{\rm in}}{2V_{\rm T}}\right)^3 + \cdots \\ &= \frac{V_{\rm in}}{V_{\rm T}}\left[1 - \frac{2}{3}\frac{V_{\rm T}}{V_{\rm in}}\left(\frac{V_{\rm in}}{2V_{\rm T}}\right)^3 + \cdots\right] \end{aligned} \tag{2x.1}$$

The first term is $I_{\rm out}/I_{\rm E}$=$V_{\rm in}/V_{\rm T}$, equivalent to the familiar G=$R_{\rm L}/r_{\rm e}$. The second term shows us how this drops off for inputs greater than 25 mV or so (i.e., $V_{\rm T}$), and should be enough to evaluate the nonlinear gain dropoff.

A single-ended differential amplifier like ours has half the gain of a full differential output, or G=$R_{\rm L}/2r_{\rm e}$. At the

[9] Here we've lazily adopted the parlance of our time, referring to a pure numerical simulation as a "*test*"! It gets worse – you'll hear people say something like "I built this circuit, and measured . . . ," when in fact they built nothing, and measured nothing. The starry-eyed delusions of the SPICE-obsessed designer.

[10] Written tanh x," and pronounced "tansh." The tanh function goes from 0 to 1 as its argument goes from 0 to infinity, and tanh(x)≈x for $x \ll 1$).

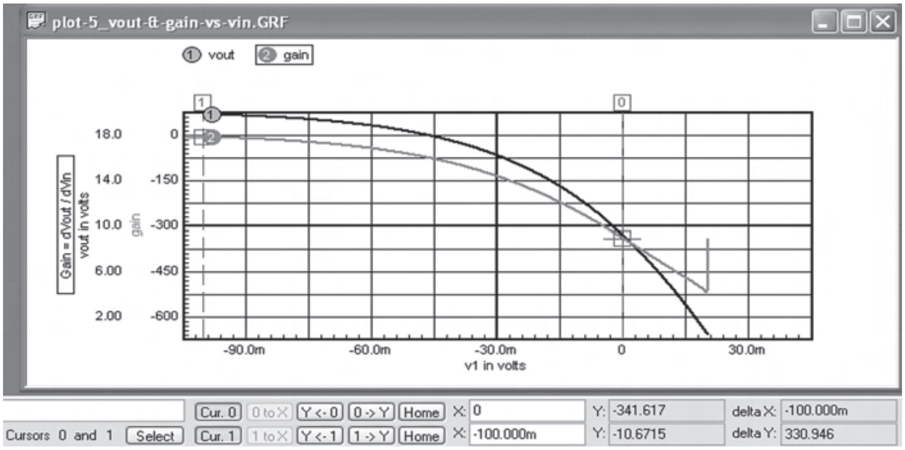

Figure 2x.25. Reading gain values from the plot with a cursor.

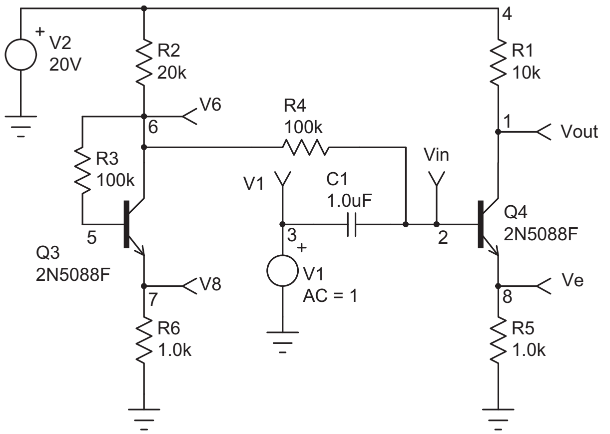

Figure 2x.26. Second circuit: degenerated common-emitter amplifier.

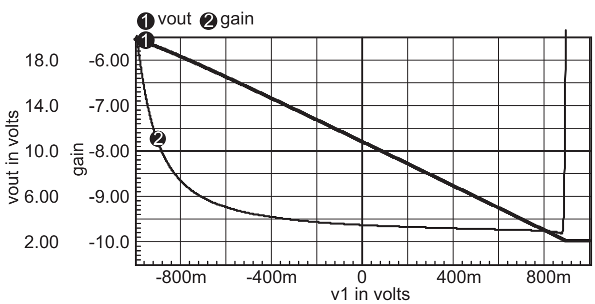

Figure 2x.27. Transfer function and gain of the circuit of Figure 2x.26. For this circuit the gain changes from −5.5 to −9.8, about 80%.

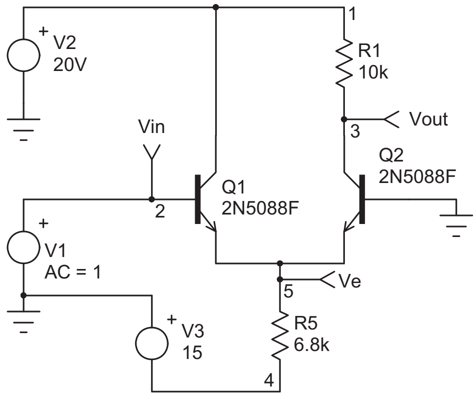

Figure 2x.28. Third circuit: differential amplifier.

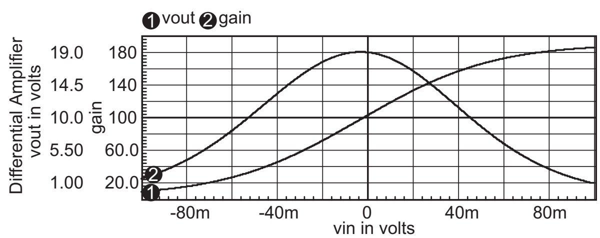

Figure 2x.29. Transfer function and gain of the circuit of Figure 2x.28. The gain error involves the hyperbolic tangent, see text.

default SPICE temperature of 27°C (where V_T=25.8 mV) this predicts our amplifier should have a gain of 10k/51.7, or G=193, in reasonable agreement with the SPICE result of G=180 for input signals less than 10 mV. Let's try a larger input, say 50 mV. Equation 2x.1 predicts the gain should decrease by a factor of $2/3\times25.8/50\times(50/51.7)^3$, or 31%. Going back to the graph, and allowing for the $-4\,\text{mV}$ offset voltage, we read off $G{\approx}96$, which is a decrease of -47% from the small-signal gain. Though the formula gives us a good idea of what to expect, it's likely that SPICE, with its Early-effect corrections, etc., gives us a more accurate answer.

At this point it's worth doing a sanity check to validate the stories that SPICE is telling us. We breadboarded the differential amplifier of Figure 2x.28 and measured[11] its transfer function (using "XY" mode on our Tektronix lunchbox-style 'scope to plot V_{out} versus V_{in}), producing the screenshot of Figure 2x.30. The real-life circuit is a pretty good replica of the SPICE plot above (hmmm..., or should it be the other way around? Hard to think of the real thing as a "replica"!)

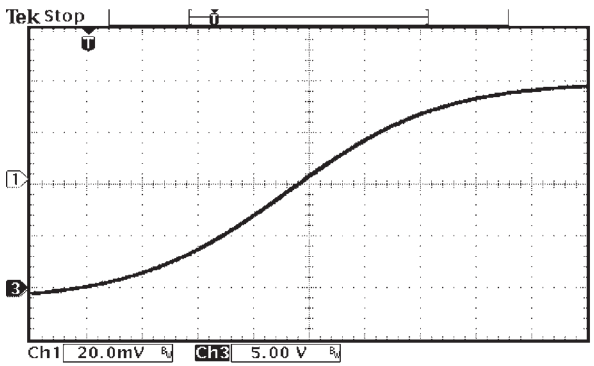

Figure 2x.30. Measured transfer function of the circuit of Figure 2x.28.

With renewed confidence in SPICE, let's try some variations. An obvious improvement is to replace the emitter tail resistor with a 2 mA current source, and to improve the symmetry by using equal 10k resistors on the collector of both transistors. With these modifications SPICE drops the gain to 161, and the offset voltage largely disappears (Figure 2x.31).

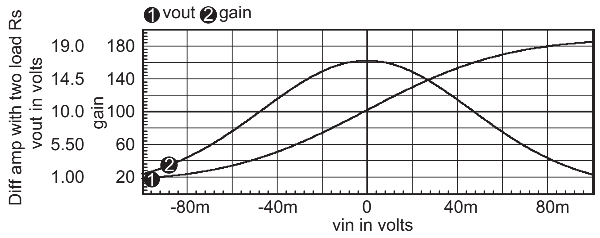

Figure 2x.31. Transfer function and gain of the circuit of Figure 2x.28, again over an input range of $-100\,\text{mV}$ to $+100\,\text{mV}$, modified with a 2 mA current sink replacing R_5.

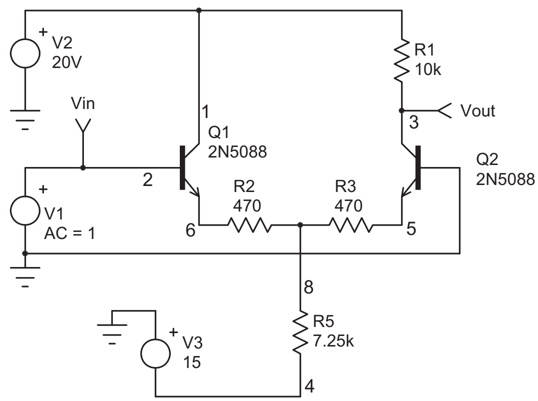

Figure 2x.32. Fourth circuit: differential amplifier with emitter degeneration.

2x.4.6 Differential amplifier with emitter degeneration

As with the simple single-ended common-emitter amplifier, we can improve differential amplifier performance with emitter degeneration, as in Figure 2x.32. In the simplified gain formula $G{=}R_1/(R_2+R_3+r_{\text{e}1}+r_{\text{e}2})$ it's primarily the current-dependent r_{e} terms that degrade the linearity.

Figure 2x.33 shows the greatly improved linearity, with $G{=}9.654$, close to our goal of $10.0\times$. The output waveform looks nice and straight, and the gain plot has the nice symmetry that indicates an absence of second-harmonic distortion. Nonetheless, the distortion at large amplitudes is still high (by audiophile standards), with a gain reduction of 1.2% (9.537/9.654) at output levels of 4 V and 16 V (and soaring to -12.5% at output levels of 2 V and 18 V, near clipping). To obtain these numerical values, slide one of the graph's cursors until it's aligned with our desired output voltage, and then read the gain value as displayed in the box.

2x.4.7 Sziklai-connected differential amplifier

The circuit in Figure 2x.34 improves the linearity, and it is popular among microphone preamp designers. The basic idea is to maintain a constant current (and hence constant g_m) for the matched *npn* input transistors by making them into Sziklai pairs.[12] In this way the complementary

[11] For real: on the bench this time, with wires and stuff!

[12] See §§2.4.2A and 2x.6 for a discussion of the Sziklai connection.

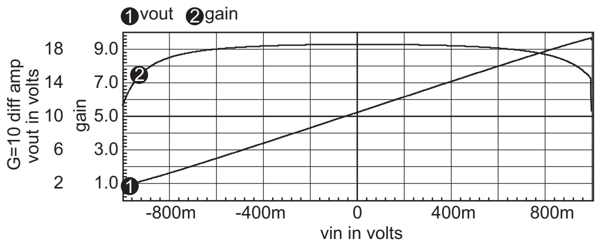

Figure 2x.33. Transfer function and gain of the circuit of Figure 2x.32. The input voltage range is −1 V to +1 V for this circuit of reduced gain.

pnp transistor of each Sziklai pair will do the work of servicing the changing output current. The new gain formula is $G{=}R_1/(R_2{+}R_3)$, without any bothersome r_e terms. Figure 2x.35 shows the resulting transfer function and gain. Now we're beginning to see some seriously good low-distortion performance! The gain is 9.988 near zero volts.

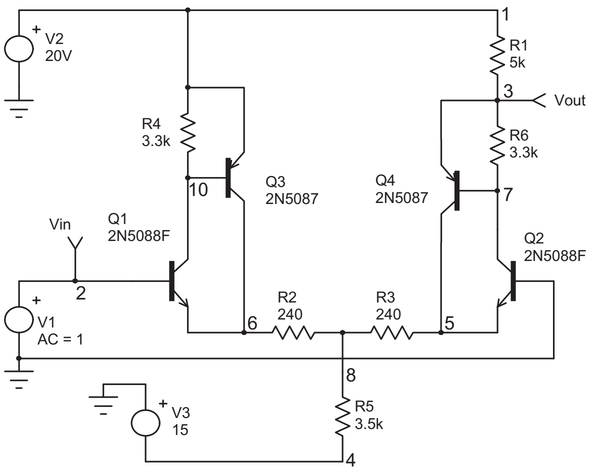

Figure 2x.34. Fifth circuit: Sziklai-connected differential amplifier with emitter degeneration.

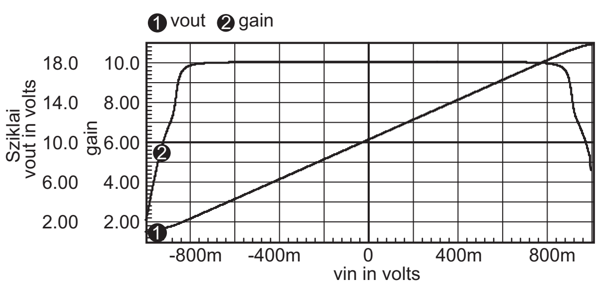

Figure 2x.35. Transfer function and gain of the circuit of Figure 2x.34.

The same data is shown in Figure 2x.36, expanding the gain axis to show only the top 5%. Compared with the unadorned diff amp with degeneration (Figure 2x.32), we see a smaller −0.4% gain variation over a 4 V to 16 V output or nearly 10× better, and less than −2% at 2 V to 18 V outputs. We could play with the values of R_4 and R_6 to improve on that 2% value (at the low currents necessary for large positive output swing, a low value of R_6 combined with a relatively large current in Q_2 leaves insufficient current to operate the Sziklai-transistor Q_4).

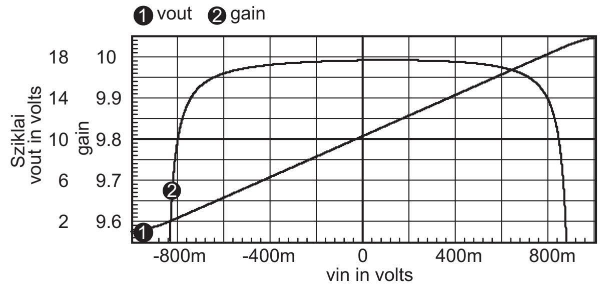

Figure 2x.36. Transfer function and gain, expanded scale.

2x.4.8 Sziklai-connected differential amplifier with current source

We can further improve the circuit by replacing R_5 with a current source, as in Figure 2x.37. Now we're enjoying an even lower −0.25% gain-reduction distortion over the 4 V to 16 V output range (Figure 2x.38). While it's true that many folks don't consider 0.25% to be "low distortion," it's also true some prefer soft limiting to the hard limiting that one experiences with conventional feedback circuits. That's one argument made for vacuum-tube amplifiers. This circuit has only 0.1% peak gain distortion over a 6 V to 14 V output swing (half of full output range). Note that what we've called "distortion" is the *peak-to-peak* gain deviation; the more usual measure is *rms* distortion, which is smaller typically by a factor of five, as seen in the measured harmonic distortion plots of Figure 2x.49.

If we increase the gain of this circuit, say to about $G{\sim}50$ by substituting 50 Ω emitter resistors, we might expect the distortion to increase. As Figure 2x.39 shows, we see gain reductions of 1% over a 4 V to 16 V swing, and 0.4% over 6 V to 14 V.

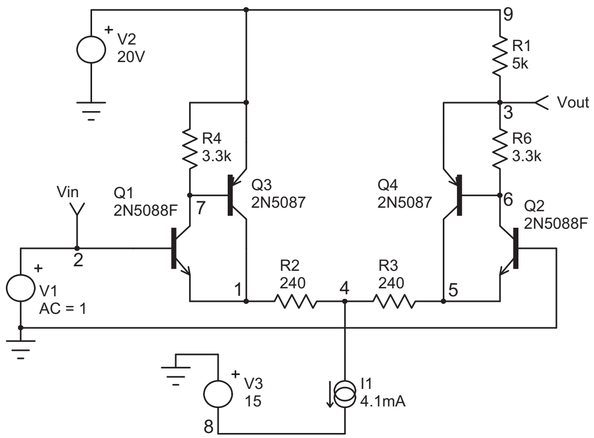

Figure 2x.37. Sixth circuit: Sziklai differential amplifier with emitter current source.

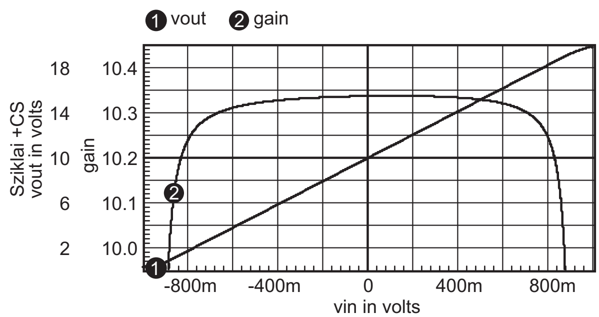

Figure 2x.38. Transfer function and gain of the circuit of Figure 2x.37.

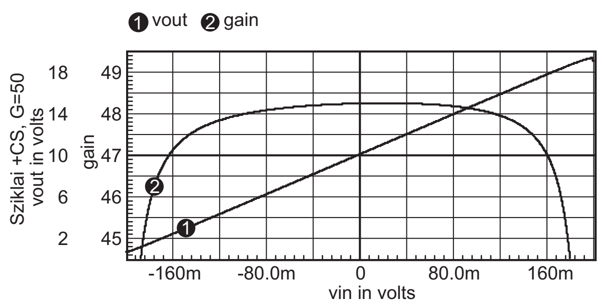

Figure 2x.39. Transfer function and gain, with reduced emitter resistors.

2x.4.9 Sziklai-connected differential amplifier with cascode

We've dealt with several sources of distortion while improving our circuit, but we haven't yet dealt with the Early effect, which describes the variation of V_{BE} with changing V_{CE} (see §§2.3.2 and 2x.5). We can try to reduce this effect with a cascode circuit, to eliminate the changes in collector voltage across the differential pair during large output swings. Figures 2x.40 and 2x.41 show the trial circuit and simulation results.

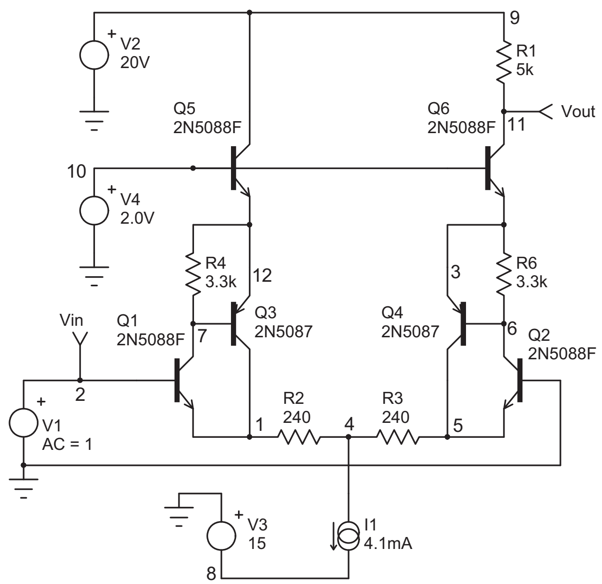

Figure 2x.40. Seventh circuit: Sziklai differential amplifier with cascode.

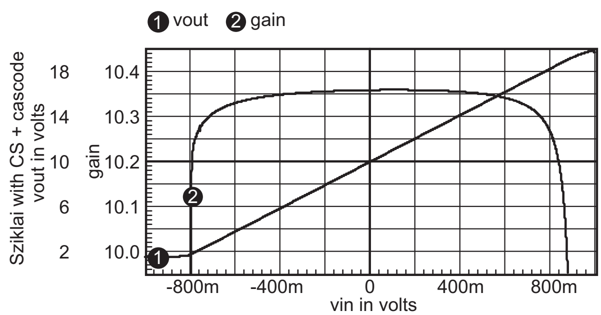

Figure 2x.41. Transfer function and gain of the circuit of Figure 2x.40.

Well, that doesn't seem to have made much improvement – still 0.1% for half scale. It may be we've suffered an offsetting degradation from operation at such a low V_{CE}=1.4 V. Another way to combat the Early effect, at least for small signals, is to balance the load resistances, so we tried a 5k resistor in the Q_1Q_3 collector, in place of the cascode. Aha, a bit better, we get 0.07% gain falloff at half of full swing (not shown).

2x.4.10 Caprio's quad differential amplifier, with cascode

Caprio's quad[13] is a unique configuration (Figure 2x.42) in which normal changes in V_{BE} are canceled. Here's how to understand this diabolically clever circuit: the voltage drop from Q_1's base to the right-hand side of R_3 is the sum of two base–emitter drops, one corresponding to the left-hand collector current, and the other corresponding to the right-hand collector current. But exactly the same statement goes for the voltage drop from Q_2's base to the left-hand side of R_3. So, even when the collector currents become unbalanced (from an input signal excursion), the input signal differential is faithfully conveyed across the gain-setting resistor R_3. In other words, the exact input signal voltage appears across R_3, without distortion caused by changing transistor base–emitter voltages. Cute!

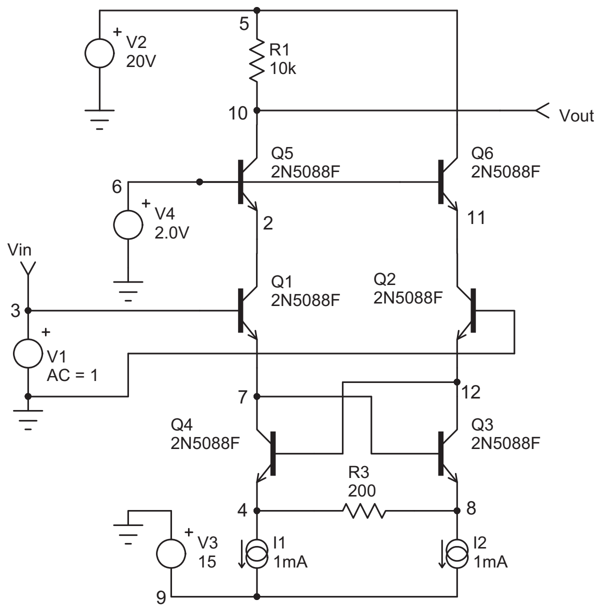

Figure 2x.42. Eighth circuit: "Caprio's quad."

Caprio's quad is limited to operation with small input signals, say under 400 mV, to avoid saturating transistor Q_3 or Q_4. In our circuit the gain has been set to 50, so that ±160 mV can drive the output over a swing of ±8 V. Figure 2x.43 shows the results.

The gain reduction is 0.36% at 4 V and 16 V outputs, and 0.12% at 6 V and 14 V. This is three times better than the 0.4% we observed for the Sziklai-connected amplifier

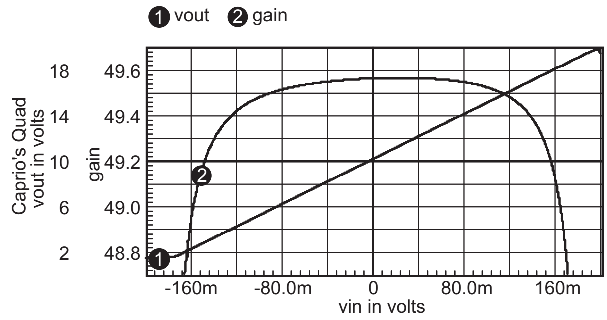

Figure 2x.43. Transfer function and gain of the circuit of Figure 2x.42, with inputs from −200 mV to +200 mV.

at the same gain. Barrie Gilbert points out[14] that Caprio's quad has another issue to worry about, namely a negative input resistance, leading to instabilities with slightly reactive input sources.

With lower gain and high input signals, the circuit exhibits a phase inversion when the input is overdriven. Figure 2x.44 shows what happens when you drive a Caprio quad designed for a gain of 25 with a 1 Vpp input triangle. It has a nice linear range, as we saw, but gets into big trouble if overdriven.

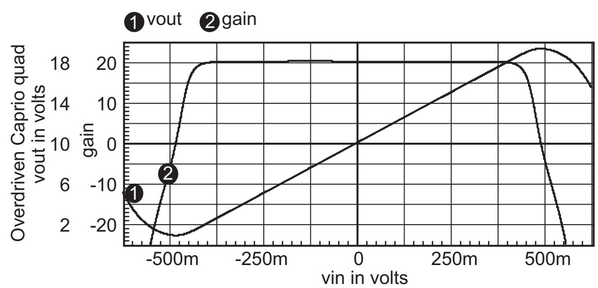

Figure 2x.44. Overdriving Caprio's quad.

2x.4.11 Caprio's quad with folded cascode – I

The Caprio quad with cascode has admirably low distortion (this is open-loop, mind you!), owing to the quad's cancellation of V_{BE}s in the input stage, and the cascode's suppression of Early effect. But we can do even better: by "folding" the cascode we can take advantage of the full rail-to-rail supply voltage range. Our first try is a gain-of-100 circuit (Figure 2x.45), in which the *pnp* output stage

[13] R. Caprio, "Precision differential voltage–current convertor," *Electron. Lett.*, **9**, 147–148 (1973).

[14] Toumazou, ed., *Analogue IC Design: The Current-Mode Approach*, Peregrinus Ltd. (1990), page 72.

is biased to mid-supply when Q_1's quiescent collector current of 1 mA leaves 0.5 mA of emitter current for the output cascode Q_3.

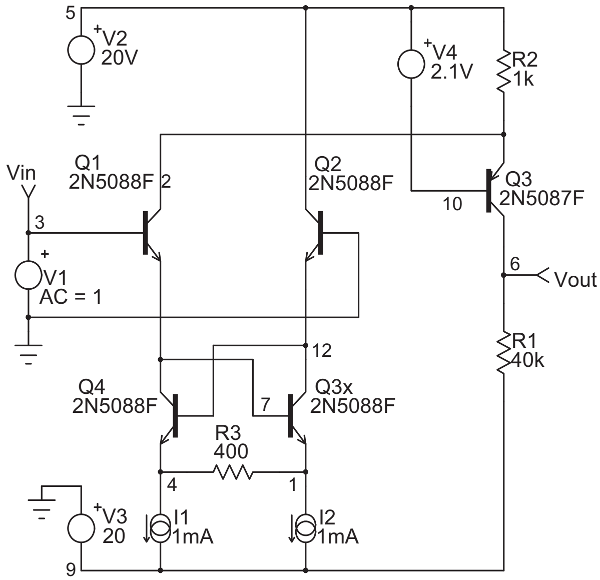

Figure 2x.45. Ninth circuit: Caprio's quad with folded cascode.

Running the SPICE engine, we get a plot of the nearly rail-to-rail output swing, but with disappointing linearity (Figure 2x.46). The gain drops off markedly at the negative portion of the output swing; it's down about 20% from the peak gain when V_{out} is -18 V (the diagonal trace is V_{out}, and the curvy trace is the gain. (Trace 7, at top, is a preview of what comes next.)

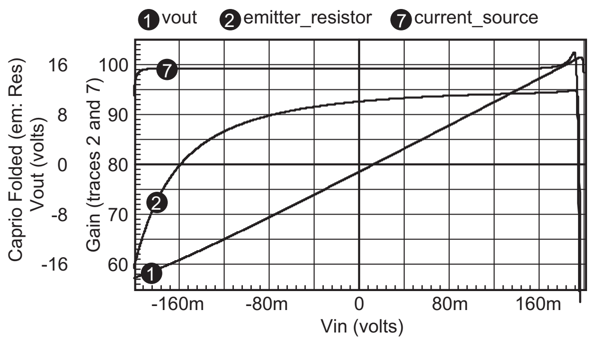

Figure 2x.46. Transfer function and gain of the circuit of Figure 2x.45.

The reason is this: although Q_1 sinks a current that is accurately linear with input signal, the residual current provided by R_2 is *not* constant (because the V_{BE} of Q_3 varies with collector current, approximately -60 mV/decade). So, for example, when the output is close to the negative rail, Q_3's V_{BE} is reduced, which increases R_2's residual current and therefore the signal V_{out}. The folded cascode has greatly degraded the distortion of Caprio's quad! Something needs to be done.

2x.4.12 Caprio's quad with folded cascode – II

Not to fret – there's an easy fix. Just replace R_2 with a current source, chosen to bias V_{out} at mid-supply after allowing for Q_1's quiescent current. Figure 2x.47 shows the circuit (where we've been lazy and used a current source symbol; in reality you'd use a BJT current source, or current mirror with small emitter resistors; we've also replaced the pair of redundant current sinks with a functionally identical single 2 mA sink).

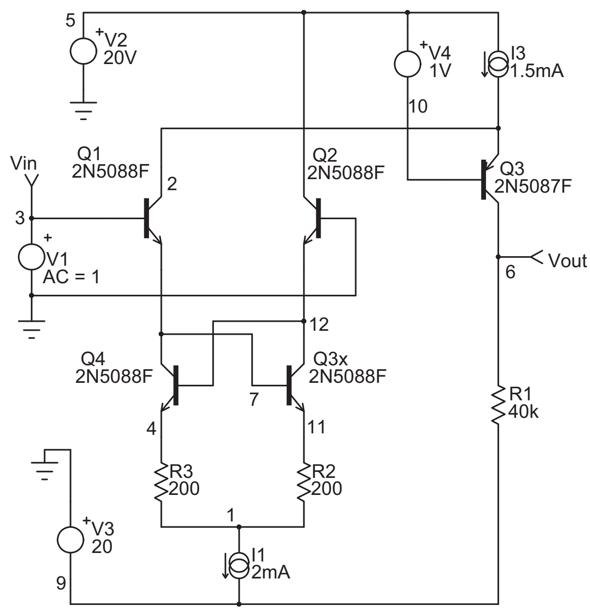

Figure 2x.47. Tenth circuit: Caprio's quad with folded cascode and emitter current source.

This circuit change produces excellent output linearity (trace 7 plots the gain versus input voltage), while preserving the nearly rail-to-rail output swing. Look at the gain plot in Figure 2x.48, where the vertical scale has been expanded to show the residual variation of gain with output swing – the peak-to-peak gain variation is less than 0.4% over the output range of -19 V to $+17$ V.

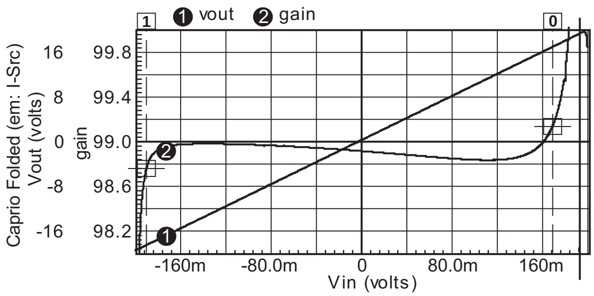

Figure 2x.48. Transfer function and gain of the circuit of Figure 2x.47.

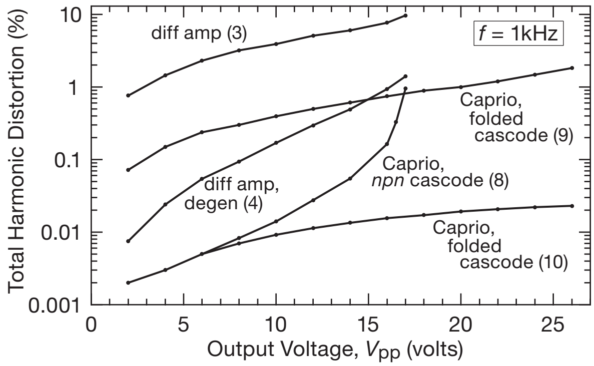

Figure 2x.49. Measured harmonic distortion at 1 kHz of the open-loop amplifier circuits of Figures 2x.28, 2x.32, 2x.42, 2x.45, and 2x.47. The numbers in parentheses refer to the *n*-th circuit iteration).

2x.4.13 Measured distortion

As we remarked in §2x.4.8, the peak "gain-reduction distortion" figures represent the worst deviation from perfect linearity over the swing, and so they're quite a bit larger than the usual measure of audio amplifier distortion – rms total harmonic distortion (THD).

Knowing the transfer function, it's possible to calculate the distortion numerically. But it's more fun to breadboard several of these circuits and *measure* their THD on the bench. It's also a chance to escape the lure of computer modeling, and get a grip on reality.

We did that, with the results shown in Figure 2x.49, where the logarithmic vertical axis tends to downplay the substantial range of measured distortions. Look closely, and you'll see that emitter degeneration produces reasonable distortion values in the differential amplifier, but the Caprio quad with cascode does an order of magnitude better; and in the current-source-fed folded cascode it reigns triumphant, with an admirably low distortion of just 0.01% (especially for an amplifier without linearizing negative feedback) even at 25 Vpp output swing. Reality nicely follows theory!

2x.4.14 Wrapup: amplifier modeling with SPICE

Our tour of distortion in BJT amplifiers has been fun, and easy. It let us explore the properties of various configurations, making changes with both little effort and great reward.

But, a *warning*: SPICE takes its models literally, and therefore goes seriously off the rails when the models are inaccurate. We saw this here, with the poor generic QN5088 model we tried first. Similar problems are found with MOSFET models, many of which fail completely in modeling the important "subthreshold region" (§3.1.4A).

It's essential to validate your SPICE models with real-world behavior before placing trust in simulations.

Readers interested in delving deeply into audio amplifier design will find inspiration in the excellent *Audio Power Amplifier Design, 6th edition* by Douglas Self (Focal Press, 2013).