Distortion from the input stage

The motivation for using a differential pair as the input stage of an amplifier is usually its low DC offset. Apart from its inherently lower offset due to the cancellation of the Vbe voltages, it has the important added advantage that its standing current does not have to flow through the feedback network. However a second powerful reason, which seems less well-known, is that linearity is far superior to single-transistor input stages. Figure 4.1 shows three versions, in increasing order of sophistication. The resistor-tail version at 1a has poor CMRR and PSRR and is generally a false economy of the shabbiest kind; it will not be further considered here. The mirrored version at 1c has the best balance, as well as twice the transconductance of 1b.

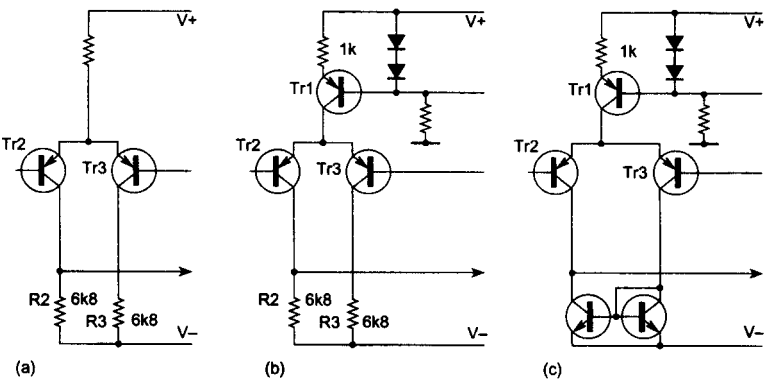


Figure 4.1
Three versions of an input pair. a Simple tail resistor. b Tail current-source. c With collector current-mirror to give inherently good Ic balance

At first sight, the input stage should generate a minimal proportion of the overall distortion because the voltage signals it handles are very small, appearing as they do upstream of the VAS that provides almost all the voltage gain. However, above the first pole frequency P1, the current required to drive Cdom dominates the proceedings, and this remorselessly doubles with each octave, thus:

$$i_{pk} = w \times C_{dom} \times V_{pk} \qquad \text{Equation 4.1}$$

where W = 2 × pi × freq

For example the current required at 100 W (8 Ω) and 20 kHz, with a 100 pF Cdom is 0.5 mA peak, which may be a large proportion of the input standing current, and so the linearity of transconductance for large current excursions will be of the first importance if we want low distortion at high frequencies.

Curve A in Figure 4.2 shows the distortion plot for a model amplifier (at +16 dBu output) designed so the distortion from all other sources is negligible compared with that from the carefully balanced input stage; with a small-signal class A stage this reduces to making sure that the VAS is

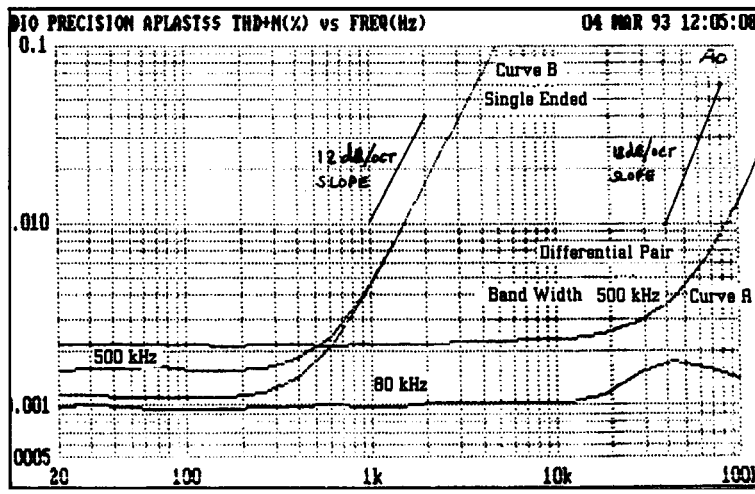


Figure 4.2
Distortion performance of model amplifier-differential pair at A compared with singleton input at B. The singleton generates copious second-harmonic distortion

properly linearised. Plots are shown for both 80 kHz and 500 kHz measurement bandwidths, in an attempt to show both HF behaviour and the vanishingly low LF distortion. It can be seen that the distortion is below the noise floor until 10 kHz, when it emerges and heaves upwards at a precipitous 18 dB/octave. This rapid increase is due to the input stage signal current doubling with every octave, to feed Cdom; this means that the associated third harmonic distortion will quadruple with every octave increase. Simultaneously the overall NFB available to linearise this distortion is falling at 6 dB/octave since we are almost certainly above the dominant-pole frequency P1, and so the combined effect is an octuple or 18 dB/octave rise. If the VAS or the output stage were generating distortion this would be rising at only 6 dB/octave, and so would look quite different on the plot.

This non-linearity, which depends on the rate-of-change of the output voltage, is the nearest thing that exists to the late unlamented TID (Transient Intermodulation Distortion), an acronym that now seems to be falling out of fashion. SID (Slew-Induced-Distortion) is a better description of the effect, but implies that slew-limiting is responsible, which is not the case.

If the input pair is *not* accurately balanced, then the situation is more complex. Second as well as third harmonic distortion is now generated, and by the same reasoning this has a slope nearer to 12 dB/octave; this vital point is examined more closely below.

BJTs vs FETs for the input stage

At every stage in the design of an amplifier, it is perhaps wise to consider whether BJTs or FETs are the best devices for the job. I may as well say

at once that the predictable Vbe/Ic relationship and much higher transconductance of the bipolar transistor make it, in my opinion, the best choice for all three stages of a generic power amplifier. To quickly summarise the position:

Advantages of the FET input stage

There is no base current with FETs, so this is eliminated as a source of DC offset errors. However, it is wise to bear in mind that FET gate leakage currents increase very rapidly with temperature, and under some circumstances may need to be allowed for.

Disadvantages of FET input stage

1 The undegenerated transconductance is low compared with BJTs. There is much less scope for linearising the input stage by adding degeneration in the form of source resistors, and so an FET input stage will be very non-linear compared with a BJT version degenerated to give the same low transconductance,
2 the Vgs offset spreads will be high. Having examined many different amplifier designs, it seems that in practice it is essential to use dual FETs, which are relatively very expensive and not always easy to obtain. Even then, the Vgs mismatch will probably be greater than Vbe mismatch in a pair of cheap discrete BJTs; for example the 2N5912 N-channel dual FET has a specified maximum Vgs mismatch of 15 mV. In contrast the Vbe mismatches of BJTs, especially those taken from the same batch (which is the norm in production) will be much lower, at about 2–3 mV, and usually negligible compared with DC offset caused by unbalanced base currents,
3 the noise performance will be inferior if the amplifier is being driven from a low-impedance source, say 5 kΩ or less. This is almost always the case.

Singleton input stage versus differential pair

Using a single input transistor (Figure 4.3a) may seem attractive, where the amplifier is capacitor-coupled or has a separate DC servo; it at least promises strict economy. However, the snag is that this singleton configuration has no way to cancel the second-harmonics generated in copious quantities by its strongly-curved exponential Vin/Iout characteristic [1]. The result is shown in Figure 4.2 curve-B, where the distortion is much higher, though rising at the slower rate of 12 dB/octave.

The input stage distortion in isolation

Examining the slope of the distortion plot for the whole amplifier is instructive, but for serious research we need to measure input-stage

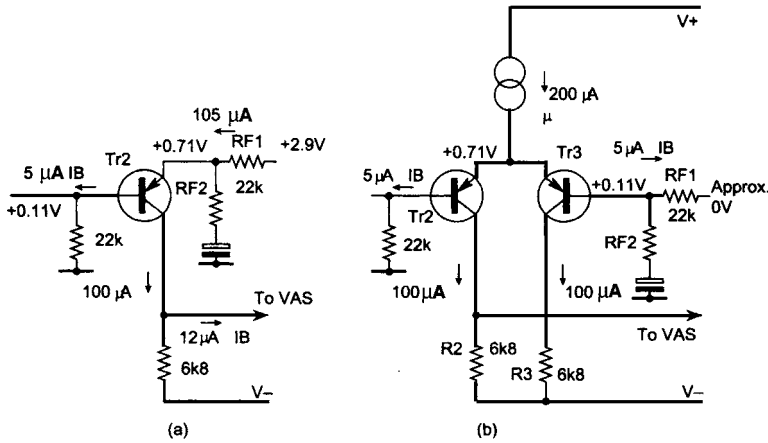


Figure 4.3
Singleton and differential pair input stages, showing typical DC conditions. The large DC offset of the singleton is mainly due to all the stage current flowing through the feedback resistor RF1

non-linearity in isolation. This can be done with the test circuit of Figure 4.4. The op-amp uses shunt feedback to generate an appropriate AC virtual-earth at the input-pair output. Note that this current-to-voltage conversion op-amp requires a third –30 V rail to allow the i/p pair collectors to work at a realistic DC voltage – i.e. about one diode's-worth above the –15 V rail. Rf can be scaled as convenient, to stop op-amp clipping, without the input stage knowing anything has changed. The DC balance of the pair can be manipulated by VR1, and it is instructive to see the THD residual diminish as balance is approached, until at its minimum amplitude it is almost pure third harmonic.

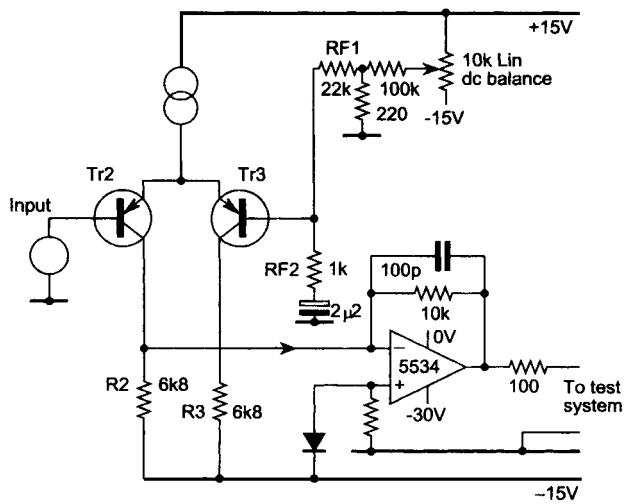


Figure 4.4
Test circuit for examining input stage distortion in isolation. The shunt-feedback op-amp is biased to provide the right DC conditions for TR2

The differential pair has the great advantage that its transfer characteristic is mathematically highly predictable[2]. The output current is related to the differential input voltage Vin by:

$$I_{out} = I_e.\tanh(-V_{in}/2V_t) \qquad \text{Equation 4.2}$$

(where Vt is the usual *thermal voltage* of about 26 mV at 25°C, and Ie the tail current).

Two vital facts derived from this equation are that the transconductance (gm) is maximal at Vin = 0, when the two collector currents are equal, and that the value of this maximum is proportional to the tail current Ie. Device beta does not figure in the equation, and the performance of the input pair is not significantly affected by transistor type.

Figure 4.5a shows the linearising effect of local feedback or degeneration on the voltage-in/current-out law; Figure 4.5b plots transconductance

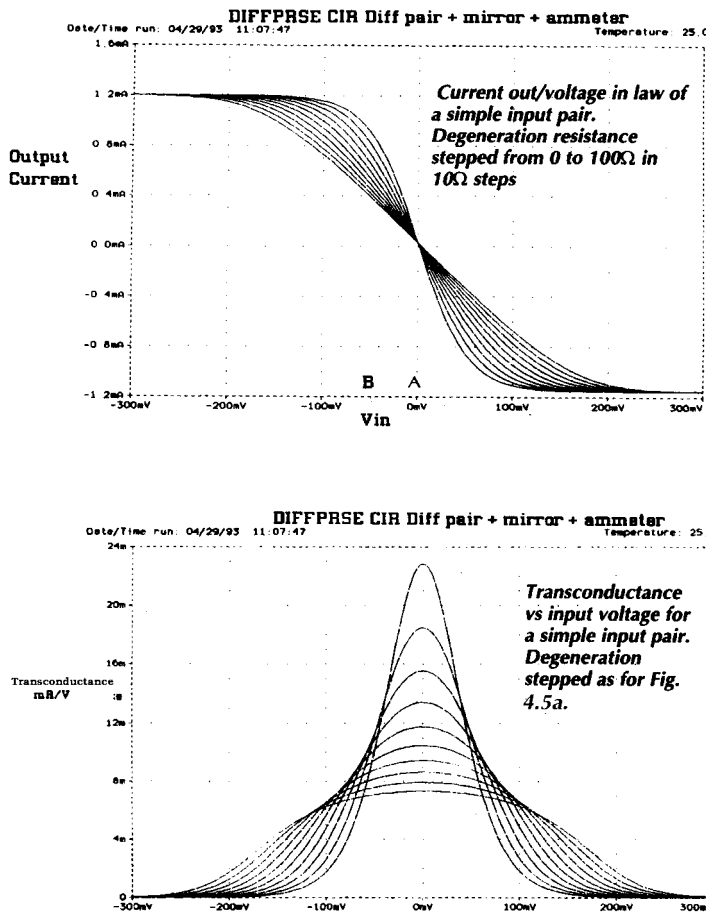


Figure 4.5
Effect of degeneration on input pair V/I law, showing how transconductance is sacrificed in favour of linearity. (SPICE simulation)

against input voltage and shows clearly how the peak transconductance value is reduced, but the curve made flatter and linear over a wider operating range. Simply adding emitter degeneration markedly improves the linearity of the input stage, but the noise performance is slightly worsened, and of course the overall amplifier feedback factor has been reduced, for as previously shown, the vitally-important HF closed-loop gain is determined solely by the input transconductance and the value of the dominant-pole capacitor.

Input stage balance

Exact DC balance of the input differential pair is essential in power amplifiers. It still seems almost unknown that minor deviations from equal Ic in the pair seriously upset the second-harmonic cancellation, by moving the operating point from A in Figure 4.5a to B. The average slope of the characteristic is greatest at A, so imbalance also reduces the open-loop gain if serious enough. The effect of small amounts of imbalance is shown in Figure 4.6 and Table 4.1; for an input of –45 dBu a collector-current imbalance of only 2% gives a startling worsening of linearity, with THD

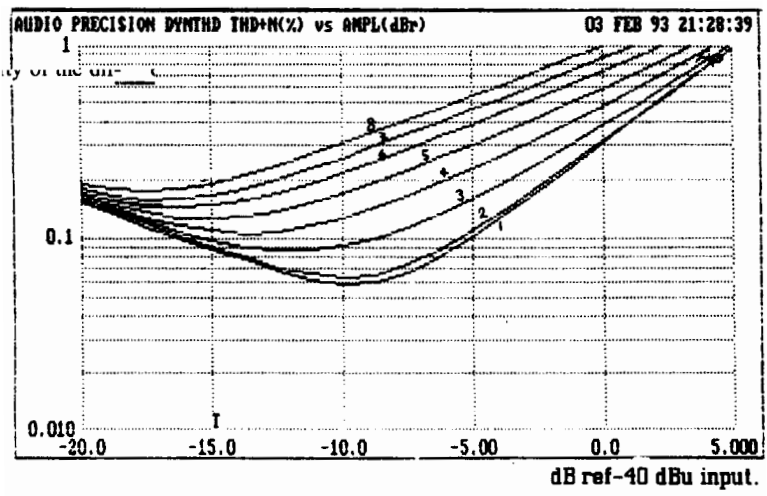


Figure 4.6
Effect of collector-current imbalance on an isolated input pair; the second harmonic rises well above the level of the third if the pair moves away from balance by as little as 2%

Table 4.1
(Key to Figure 4.6)

Curve No.	Ic Imbalance	Curve No.	Ic Imbalance
1	0%	5	5.4%
2	0.5%	6	6.9%
3	2.2%	7	8.5%
4	3.6%	8	10%

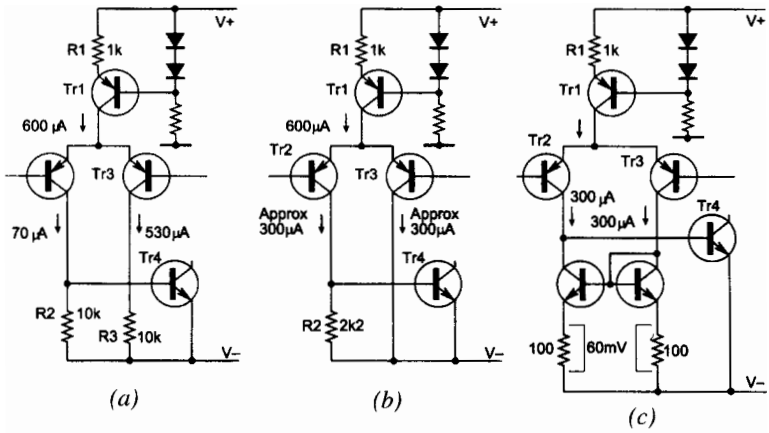


Figure 4.7
Improvements to the input pair. a Poorly designed version. b Better; partial balance by correct choice of R2. c Best; near-perfect Ic balance enforced by mirror

increasing from 0.10% to 0.16%; for 10% imbalance this deteriorates badly to 0.55%. Unsurprisingly, imbalance in the other direction (Ic1>Ic2) gives similar results.

Imbalance defined as deviation of Ic (per device) from that value which gives equal currents in the pair.

This explains the complex distortion changes that accompany the apparently simple experiment of altering the value of R2[3]. We might design an input stage like Figure 4.7a, where R1 has been selected as 1k by uninspired guesswork and R2 made highish at 10k in a plausible but misguided attempt to maximise o/l gain by minimising loading on Q1 collector. R3 is also 10k to give the stage a notional *balance*, though unhappily this is a visual rather than electrical balance. The asymmetry is shown in the resulting collector currents; the design generates a lot of avoidable second harmonic distortion, displayed in the 10k curve of Figure 4.8.

Recognising the crucial importance of DC balance, the circuit can be rethought as Figure 4.7b. If the collector currents are to be roughly equal, then R2 must be about 2 × R1, as both have about 0.6 V across them. The dramatic effect of this simple change is shown in the 2k2 curve of Figure 4.8; the improvement is accentuated as the o/l gain has also increased by some 7 dB, though this has only a minor effect on the closed-loop linearity compared with the improved balance of the input pair. R3 has been excised as it contributes very little to input stage balance.

The joy of current-mirrors

Although the input pair can be approximately balanced by the correct values for R1 and R2, we remain at the mercy of several circuit tolerances. Figure 4.6 shows that balance is critical, needing an accuracy of 1% or

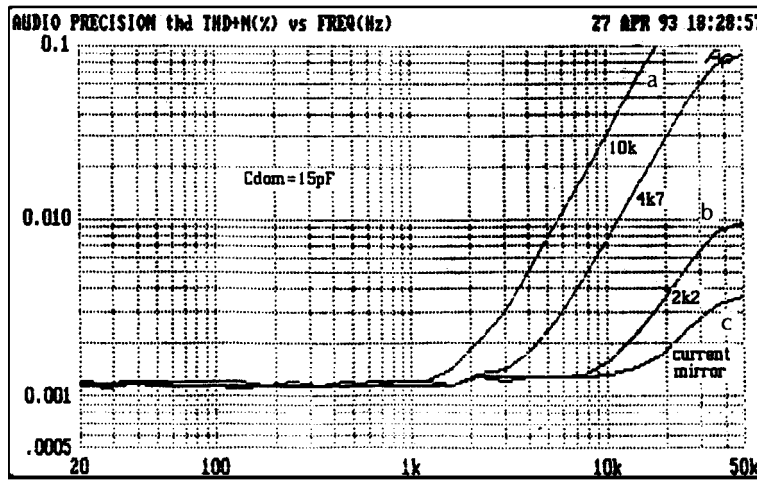


Figure 4.8
Distortion of model amplifier:
a Unbalanced with R2 = 10k. b Partially balanced with R = 2k2. c Accurately balanced by current-mirror

better for optimal linearity and hence low distortion at HF, where the input pair works hardest. The standard current-mirror configuration in Figure 4.7c forces the two collector currents very close to equality, giving correct cancellation of the second harmonic; the great improvement that results is seen in the current-mirror curve of Figure 4.8. There is also less DC offset due to unequal base-currents flowing through input and feedback resistances; I often find that a power-amplifier improvement gives at least two separate benefits. This simple mirror has well-known residual base-current errors but they are not large enough to affect the distortion performance.

The hyperbolic-tangent law also holds for the mirrored pair[4], though the output current swing is twice as great for the same input voltage as the resistor-loaded version. This doubled output is given at the same distortion as for the unmirrored version, as linearity depends on the input voltage, which has not changed. Alternatively, we can halve the input and get the same output, which with a properly balanced pair generating third harmonic only will give one-quarter the distortion. A pleasing result.

The input mirror is made from discrete transistors, regretfully foregoing the Vbe-matching available to IC designers, and it needs its own emitter-degeneration for good current-matching. A voltage-drop across the current-mirror emitter-resistors in the range 30–60 mV will be enough to make the effect of Vbe tolerances on distortion negligible; if degeneration is omitted then there is significant variation in HF distortion performance with different specimens of the same transistor type.

Putting a current-mirror in a well-balanced input stage increases the total o/l gain by at least 6 dB, and by up to 15 dB if the stage was previously poorly balanced; this needs to be taken into account in setting the

compensation. Another happy consequence is that the slew-rate is roughly doubled, as the input stage can now source and sink current into Cdom without wasting it in a collector load. If Cdom is 100 pF, the slew-rate of Figure 4.7b is about 2.8 V/µsec up and down, while 4.7c gives 5.6 V/µsec. The unbalanced pair at 4.7a displays further vices by giving 0.7 V/µsec positive-going and 5 V/µsec negative-going.

Improving input-stage linearity

Even if the input pair has a current-mirror, we may still feel that the HF distortion needs further reduction; after all, once it emerges from the noise floor it octuples with each doubling of frequency, and so it is well worth postponing the evil day until as far as possible up the frequency range. The input pair shown has a conventional value of tail-current. We have seen that the stage transconductance increases with Ic, and so it is possible to increase the gm by increasing the tail-current, and then return it to its previous value (otherwise Cdom would have to be increased proportionately to maintain stability margins) by applying local NFB in the form of emitter-degeneration resistors. This ruse powerfully improves input linearity, despite its rather unsettling flavour of something-for-nothing. The transistor non-linearity can here be regarded as an internal non-linear emitter resistance re, and what we have done is to reduce the value of this (by increasing Ic) and replaced the missing part of it with a linear external resistor Re.

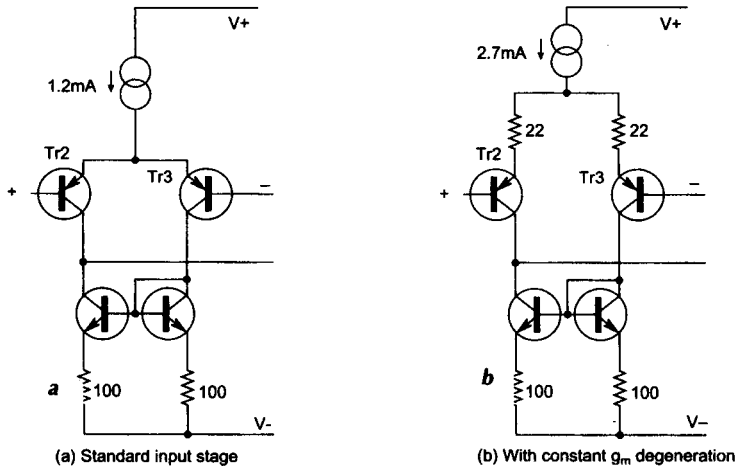
For a single device, the value of re can be approximated by:

$$re = 25/Ic \text{ ohms (for Ic in mA)} \qquad \text{Equation 4.3}$$

Our original stage at Figure 4.9a has a per-device Ic of 600 µA, giving a differential (i.e. mirrored) gm of 23 mA/V and re = 41.6 Ω. The improved version at Figure 4.9b has Ic = 1.35 mA and so re = 18.6 Ω; therefore emitter degeneration resistors of 22 Ω are required to reduce the gm back to its original value, as 18.6 + 22 = 41.6 Ω. The distortion measured by the circuit of Figure 4.4 for a –40 dBu input voltage is reduced from 0.32% to 0.032%, which is an extremely valuable linearisation, and will translate into a distortion reduction at HF of about 5 times for a complete amplifier; for reasons that will emerge later the full advantage is rarely gained. The distortion remains a visually pure third-harmonic, so long as the input pair remains balanced. Clearly this sort of thing can only be pushed so far, as the reciprocal-law reduction of re is limited by practical values of tail current. A name for this technique seems to be lacking; *constant-gm degeneration* is descriptive but rather a mouthful.

The standing current is roughly doubled so we have also gained a higher slew-rate; it has theoretically increased from 10 V/µsec to 20 V/µsec, and once again we get two benefits for the price of one inexpensive modification.

Figure 4.9
Input pairs before and after constant-gm degeneration, showing how to double stage current while keeping transconductance constant; distortion is reduced by about ten times



Radical methods of improving input linearity

If we are seeking still better linearity, various techniques exist. Whenever it is needful to increase the linearity of a circuit, it is often a good approach to increase the *local* feedback factor, because if this operates in a tight local NFB loop there is often little effect on the overall global-loop stability. A reliable method is to replace the input transistors with complementary-feedback (CFP or Sziklai) pairs, as shown in the stage of Figure 4.10a. If an isolated input stage is measured using the test circuit of Figure 4.4, the constant-gm degenerated version shown in Figure 4.9b yields 0.35% third-harmonic distortion for a –30 dBu input voltage, while the CFP version

Figure 4.10
Some enhanced differential pairs: a The Complementary Feedback Pair. b The Cross-quad. c The Cascomp