700 Series
20V BIPOLAR ARRAY DESIGN MANUAL

Last Revision Date: 2 December 2005

The 700 Series Design Manual has been originated and is maintained by Hans Camenzind, Array Design Inc. San Francisco. Feedback is welcome. Array Design offers design assistance for users or potential users of the 700 Series (generally free of charge), as well as full custom designs in bipolar, CMOS and BICMOS technology.

See also the book "Designing Analog Chips" at http://www.designinganalogchips.com/, a free download. This book can also be purchased in printed form.

Array Design Inc.
332 Virginia Ave.
San Francisco CA 94110-5151
415-308-6090
camenzind@arraydesign.com

Blank page.

Chapter 1

Semicustom Chips

When you begin the design of an electronic product, you have the choice of putting it together from standard (or discrete) parts available off a distributor's shelf or to employ **A**pplication **S**pecific **I**ntegrated **C**ircuits (ASICs). In this first chapter we will explore and dissect the various kinds of ASICs and compare them with discrete designs.

The oldest among ASICs is the full *custom* IC. Here all the layers are specifically designed for you. The supplier will, most likely, use an already proven process, but all the masks (up to perhaps 15) must be specifically designed and made and are used for only one product: yours.

The great advantage of a custom IC is its density. It has been designed to do *one* specific job. All components on the chip are there for a purpose and the designer need not expand the capability of the chip to include as many applications as possible, as is the case for a standard IC.

There are, however, four major disadvantages to custom ICs. First, the development time is the longest among all ASICs, mainly because all layers have to be designed from ground up and wafers have to be processed all the way through. Second, the development is the riskiest: one mistake in any of the layers and your circuits won't work. Third, the development cost is high, owing to the extended design time, the fabrication of many masks and the pilot processing of wafers. And fourth, using a special mask for each processing step makes fabrication of small quantities difficult and costly.

The next category in ASICs uses *standard cells*. In this approach the *design* is economized by dividing the circuit into repeatable cells. Once these standard cells are developed, a layout can be put together very rapidly and the chances of making a mistake are greatly reduced. This approach has also been dictated by the use of the computer: a full custom circuit, where each device is hand-designed, requires an inordinate amount of memory;

in standard cell ICs a cell must be described fully only once and can be repeated with simple coordinates.

However, one of the major disadvantages of full-custom ICs remains for the standard cell design. While the cells may be standardized, their placement and routing is not. Therefore, a full set of custom masks is still required and is reflected in high cost for anything but large manufacturing quantities.

Also, standard cells are well-suited only for digital ICs, where most designs can be based on gates. For linear designs it is much more difficult to create a cell which can be regarded as standard. Just about any design requires a different speed, power consumption, precision, operating voltage, input impedance, output drive, slew rate, offset voltage etc. In addition there are many functional blocks which are unique to a particular application.

The third category within ASICs are *semicustom ICs*. Here the gates (in digital ICs) or components (in linear ICs) are pre-designed, they are in fixed locations on the chip. All the designer needs to do is to interconnect them in his or her own way.

In the simplest form only one layer, the next to last one, needs to be customized. Thus the manufacturer needs only one mask, the metal mask, to produce a unique circuit on a standardized wafer.

The chief advantages of semicustom ICs are:

1. **The development cycle is far shorter compared to custom or standard cell ICs**. The wafers can be pre-processed and inventoried at the point were customization begins; the finishing of the wafer then requires only a small fraction of the total processing time.

2. **The development cost is much lower compared to custom or standard cell ICs**. Only one mask needs to be custom-designed.

3. **Semicustom is the lowest-risk approach**. The chances of making a mistake are reduced by the number of layers. The devices upon which the design is based are well characterized and can be used to breadboard or computer-simulate the design. Changes, if necessary, can be made easily and at minimum cost.

4. **It is the only approach which makes it possible to design an integrated circuit without previous IC design experience.** Having the devices pre-designed lets you concentrate on circuit design.

5. **The semicustom approach works well for even small production quantities.** It is a relatively easy task to draw a fully diffused, standard wafer from inventory and pattern it with a custom mask.

The main *disadvantage* of semicustom ICs is usually perceived to be their larger required area, because not all of the devices on the chip are used. Thus one is tempted to conclude that semicustom ICs only have a cost advantage at small and perhaps medium quantities. This conclusion is not necessarily warranted. Consider the following points:

A. In a *series* of semicustom chips the excess number of components is never larger than the gap between the chips. For example, the bipolar linear series described in this manual consists of nine chips, covering the range from 37 to 630 transistors. Each succeeding chip increases in area by 30%, thus at least 70% of the components are used, otherwise one would used the next smaller chip. Improvement in device design and chip architecture has made it possible to use close to 100% of the components.

B. The cost of an integrated circuit is not determined by chip area alone. An equally important factor is packaging cost. A 30% higher chip area, for example, will be reduced to a much smaller percentage after assembly.

C. Because semicustom chips almost always have *some* excess components, changes can be made easily. While this feature is important during the development phase, it is often more welcome during the initial marketing phase to incorporate last minute flaws or customer wishes. It is not uncommon for a chip which had been planned to be converted to full custom to stay in semicustom form for several years.

Blank page.

Chapter 2

How These Chips Are Made

Semiconductors, as the name indicates, are a group of materials located in between conductors and insulators. There are only a few of them, mainly the elements silicon and germanium and some compounds such as gallium-arsenide.

A semiconductor is, however, not just a poor conductor or poor insulator. In contrast to resistive materials (such as carbon) it conducts electricity only if other atoms are also present. By itself it is an insulator.

Let's take the example of silicon. It is element number 14 in the periodic table. Its 14 electrons are in three distinct orbits: two in the first, eight in the second and four in the third. It is the outermost orbit which determines the *valence* of an element. Silicon, therefore, has a valence of four.

In pure form (and in a temperature range around room temperature) all four outer electrons cling tightly to the silicon atom. Thus, unlike in conductors where the outermost electrons will jump from atom to atom under the influence of an electric field, the electrons of pure silicon stay in place and very little electric current can flow.

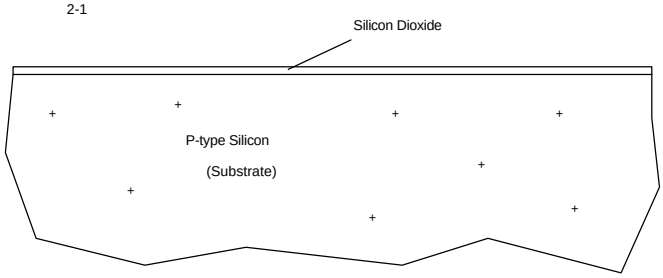
If we insert an atom with a different valence into silicon, however, things change. If this atom has an additional electron (a valence of five), four of the five electrons in its outermost shell are tied up with the four electrons of the silicon atoms (providing the cohesive force); the fifth electron is unused and is free to travel. In order for this electron to travel easily however, the silicon must be a crystal, i.e. all silicon atoms must be aligned in perfect order.

Elements with a valence of five are, among others, phosphorus, arsenic and antimony. If these elements are inserted, silicon becomes *n-type* (n denoting the negative charge of the excess electrons). The more heavily silicon is "doped" with these elements, the more conductive it becomes. Even at the maximum doping level there are comparatively few dopant atoms, fewer than one for every 10,000 silicon atoms.
Similarly we can also introduce elements with a valence of three into silicon, such as boron or gallium. When we do that there is an electron missing, forming a positive charge. This charge, or *hole*, also moves under the influence of electric charge, albeit a bit more

awkwardly. Such a semiconductor is said to be *p-type*.

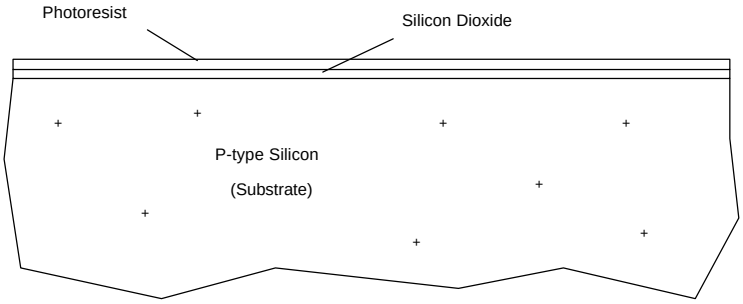
How do we insert these dopants into silicon? Diffusion is the most common method. Atoms do not stay in place, they move, or diffuse. In gasses diffusion is rapid, in liquids quite slow and in solids, at room temperature, imperceptibly minute. If we heat a solid, however, the speed of diffusion increases. If we heat silicon to red-heat (about 1200°C) a gas on the outside diffuses into the surface of the solid. Even at this temperature it may take an hour or so to reach a depth of a few micrometers (microns).

2-1



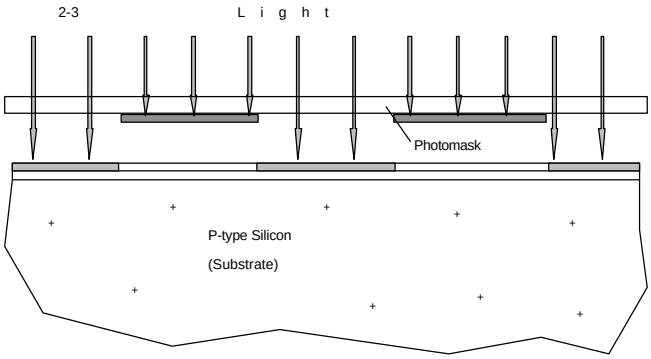
The starting point for our integrated circuit is shown in figure 2-1. It is a single-crystal wafer about 0.02 inches (0.5mm) thick, doped p-type. Silicon covers itself with an oxide layer when exposed to air. This oxide (silicon-dioxide to be precise, one form of glass) plays a crucial role in the fabrication and operation of an integrated circuit. For this reason it is grown on purpose under controlled conditions.

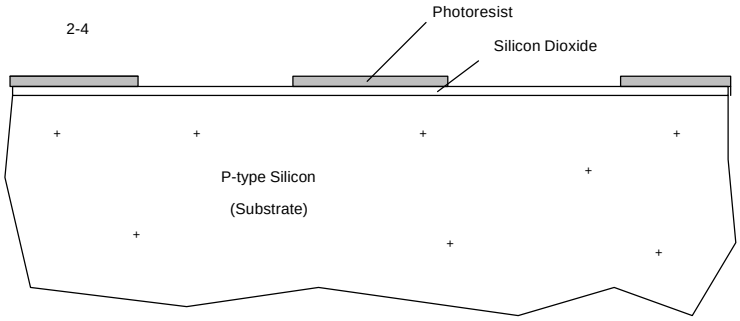
2-2



Next (figure 2-2) we spread a thin layer of photoresist on top of the oxide. This is the same kind of light-sensitive emulsion as is used on a photographic film.

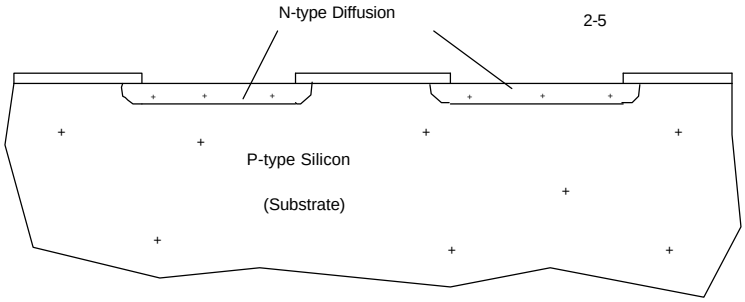
When the photoresist is dry it is exposed to light through a mask. The mask - a plate of glass - contains the first pattern for the integrated circuit. This can either be a contact print (where the mask is laid on top of the wafer), or a projection exposure (with the mask in close proximity but not touching the wafer). The dark patterns on the mask prevent the light from reaching the photoresist; in the remaining area the photoresist is exposed, developed and subsequently removed.



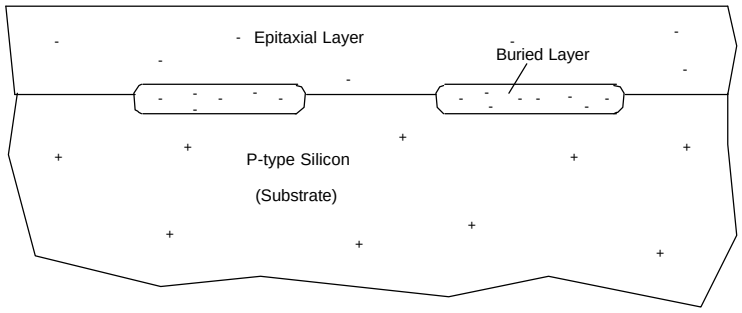


We now have the pattern of the mask directly on the wafer (figure 2-4). With an etchant we can remove the silicon-dioxide where it is not protected by the photoresist. The etchant only attacks the oxide; etching action stops at the silicon. After the etching step the photoresist is removed.

Next (figure 2-5) we expose the wafer to a gas at high temperature. The gas contains n-type dopant, which diffuses into the silicon. The concentration of dopants in these pockets is as high as we can make it, to result in the lowest possible resistance.



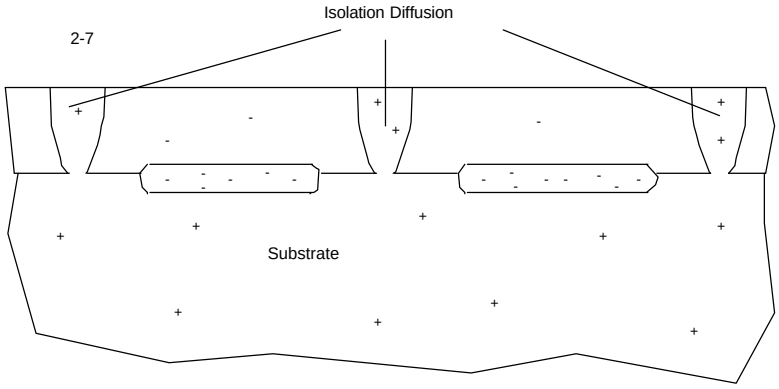
2-6



After this first diffusion we remove the remaining silicon-dioxide and grow an additional layer of silicon (figure 2-6). Called the *epitaxial layer*, it is lightly doped n-type and its silicon and dopant atoms continue in the perfect crystal structure of the substrate. Since the highly-doped n-layer is now located underneath the epitaxial layer, it is named the *buried layer*.

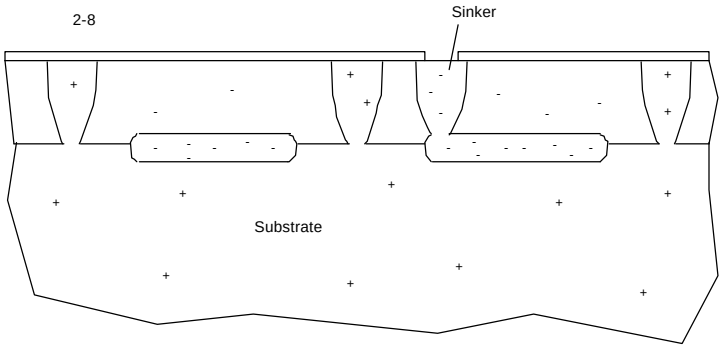
The growth of the epitaxial layer is followed by the *isolation diffusion*. Figure 2-7 shows the result (for this and all following diffusions we have omitted showing the detailed steps of applying the photoresist, mask exposure and oxide etching). The isolation diffusion is a deep one, it penetrates the epitaxial layer. The concentration of p-type dopants in these areas is higher than the previous n-type level, so that the region is converted from n to p-type.

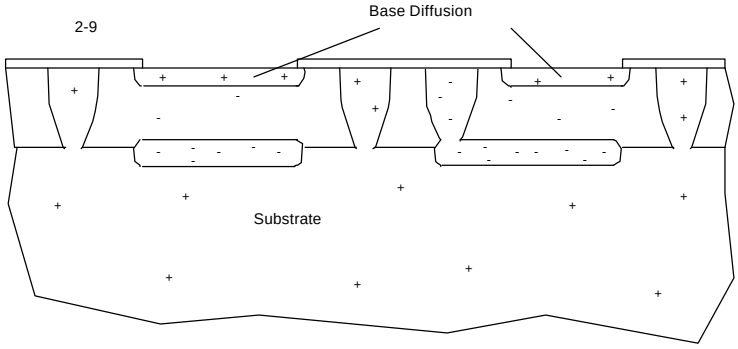
2-7



Notice that we now have two n-type pockets, entirely surrounded by p-type silicon. As long as these pockets are held at a voltage more positive than that of the substrate (which is also connected to the regions created by the isolation diffusion), its boundaries form a reverse-biased diode, with little or no current flowing. Thus, these regions are electrically separated by *junction isolation*.

After each diffusion the oxide layer is regrown, so that we can pattern it in new areas. The next diffusion (figure 2-8) gives us access to the buried layer. Called the *sinker* it provides a low-resistance path through the epitaxial layer. The sinker and buried layer provide a low-resistance path for the collector of the NPN transistor.

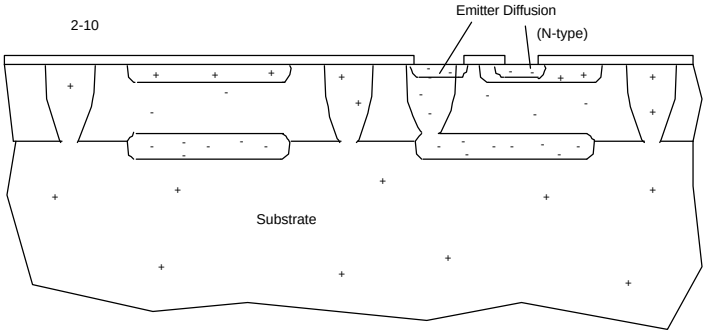


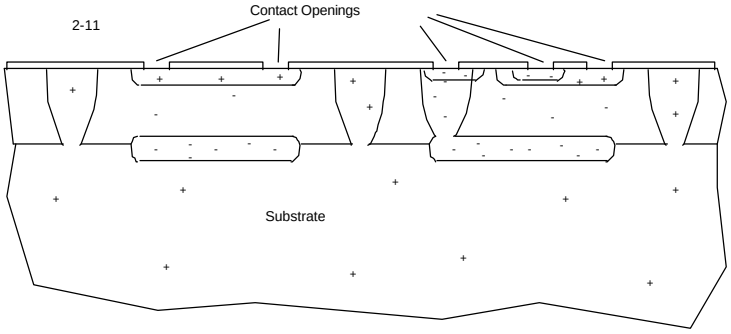


For the center layer of the NPN transistor, the base, we diffuse P-type impurities to a depth of about 1 micron.

Since a doped semiconductor has a certain resistance (which varies inversely with the dopant concentration), it can be shaped into a resistor. The base diffusion is the layer best suited for this job.

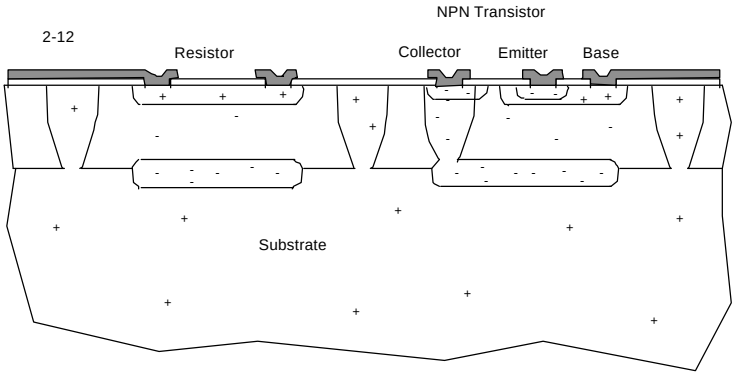
The last diffusion in the process forms the emitters of the NPN transistors (figure 2-10). It is also used to lower the collector access resistance even further. The difference in depth between the base and emitter diffusions forms the actual base of the NPN transistor. For high current gain it must have a very fine and well-controlled width (on the order of 0.1um).





After the diffusions are complete we regrow the oxide layer and pattern the contact openings (figure 2-11).

The wafer is finished at this point by depositing a thin (1um) aluminum layer over the entire surface, patterning and etching it so that it forms the interconnection pattern (figure 2-12). The wafer is then protected with a thick oxide layer, which is removed only in the bonding pad areas.



Chapter 3

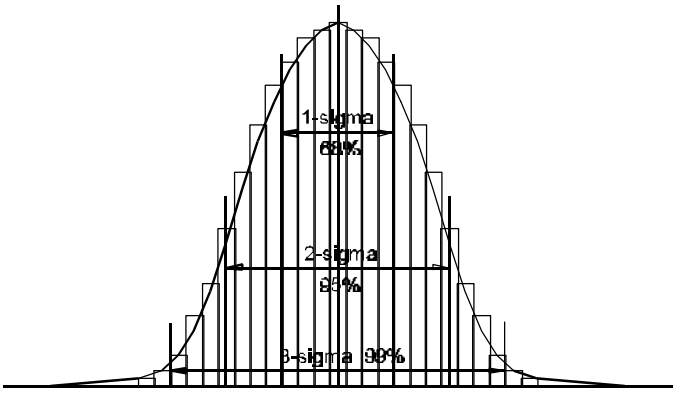
Devices

Statistical Distributions

Nothing we attempt to do is ever perfect. We aim at a goal and we are certain to miss it. If we are good at it (or lucky) the error is so minute that we can hardly measure it, but there is always an error.

Let's take the case of making an ordinary resistor. You deposit some resistive material (say carbon) on a ceramic tube and measure the resistance. You aimed at 100 Ohms and got 110. So you adjust your machine and try again. The result is 99 Ohms. You are satisfied and start production. The next day the line voltage is a little high and your resistors are running 95 Ohms. You adjust for it and the following day you measure 103 Ohms. Then the barometer changes and the resistance walks up to 104 Ohms. And so on.

If you plot your measurements day in and day out you are more than likely to find a Gaussian distribution. With diligence you will most often get a measurement of 100 Ohms (within the measurement accuracy), followed by ever smaller quantities with higher and lower resistances. The Gaussian distribution is a handy tool to describe variation.



The area under the curve contains all the measurements. Most scientific calculators can determine sigma, the deviation (in one direction); ±sigma (in both directions) contains 68.3% of all measurements. If we double this value (2-sigma) we get a deviation which contains 95.4% of all measurements; by tripling it (3-sigma) we get the deviation for 99.7% of all measurements. It is this 3-sigma limit which is most commonly used in IC design.

All parameters in an IC have a distribution, though not all of them are Gaussian. For example, hFE has a skewed distribution (more values toward the high end), but the error is small enough to include it in the Gaussian family.

A discrete resistor can be tested after fabrication. If we find a resistor with a value outside certain limits (say 3-sigma), we can throw it away or put it in a different bin. The distribution of the final resistors is then truncated: there will be no values outside 3-sigma.

In an IC we cannot test each component before use, we have to accept what is there. Thus, even if we design a circuit to 3-sigma limits, a small percentage of the parameters will be outside the limits and must be eliminated by testing the entire circuit. For this reason the 3-sigma limits given in this chapter should be accepted as the minimum standard. Be also aware that these percentages multiply. If you have 10 parameters which are critical, the total percentage of circuits inside the 3-sigma limits is not 99.7% but 97% (0.997^{10}).

Here is the entire listing of tolerance and the resulting percentages:

Tolerance	Pass	Fail
1-sigma	68.26%	31.75%
2-sigma	95.44%	4.56%
3-sigma	99.73%	0.27%
4-sigma	99.994%	0.006%
5-sigma	99.99994%	0.00006%

Parameter Correlation

The overwhelming majority of IC parameters are independent of each other. For example, there is no correlation at all between the hFE of an NPN transistor and that of a PNP transistor. The first one is determined mostly by the difference in thickness between the base and emitter diffusions, the second by the distance between two base regions.

Four parameters, however, show some relationships:

- The higher the NPN hFE, the lower the current of the base pinch resistor.
- The lower the resistor value, the lower the Zener voltage.

Isolation

If you follow two simple rules, all components in the 700 series chips will be properly isolated:

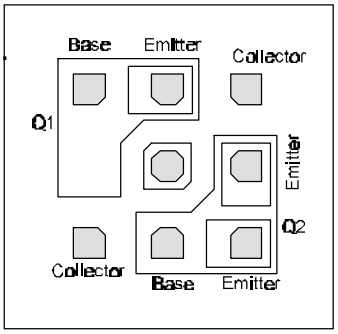
1. Connect the substrate to the most negative voltage. The connection to the substrate is the MNUS metal run around the entire chip.

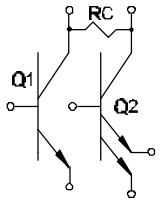
2. Connect at least one +V contact to the most positive voltage. This is the connection to the (epitaxial) N-layer surrounding all resistors.

If you have multiple supply voltages and it is possible for any one to be the most positive at different times, connect a Schottky diode from the N-layer to each supply (cathode at the N-layer). In this way it is always the highest supply which will bias the N-layer. This, however, has some limitations at high frequencies.

NPN Transistor

All the 700 Series chips contain a transistor which is both NPN and PNP; you decide which it is to be.





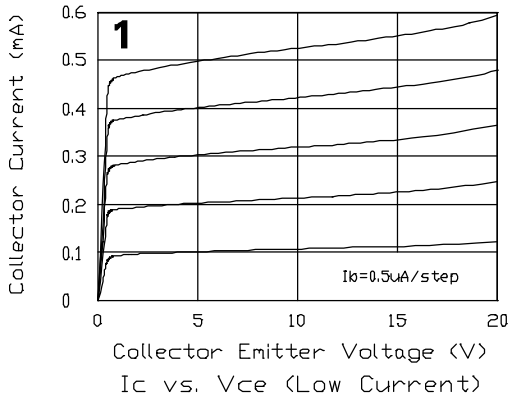
Used as an NPN, the structure contains two separate bases, a total of 3 emitters and two collector contacts. Thus, you can connect it as:

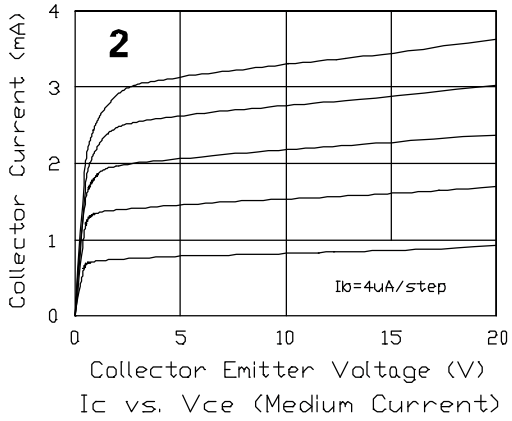
- two NPN transistors with a common collector,
- one NPN transistor with 1, 2 or 3 emitters (connect the bases together for three emitters, or two in different bases), or
- one NPN transistor and one Zener diode.

The multiple emitters can be used to create current ratios or to increase the current handling capability.

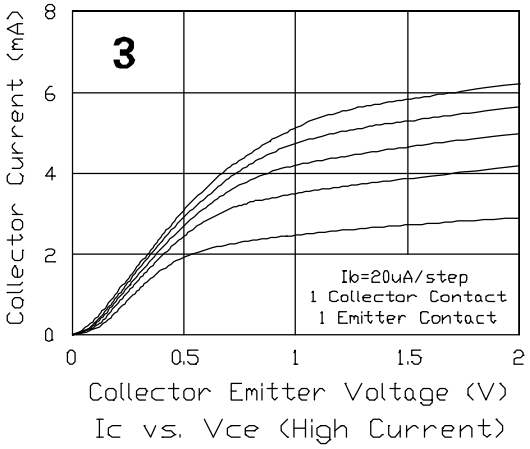
The two horizontal emitters match best since they have identical orientation.

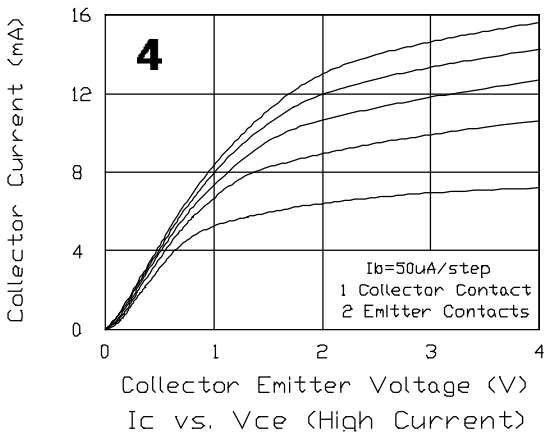
Graphs 1 and 2 Notice the Early effect. The lines slope upward with increasing voltage, especially close to 20 Volts Vce.

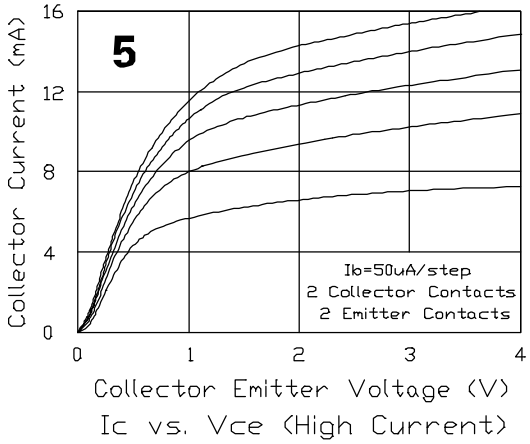


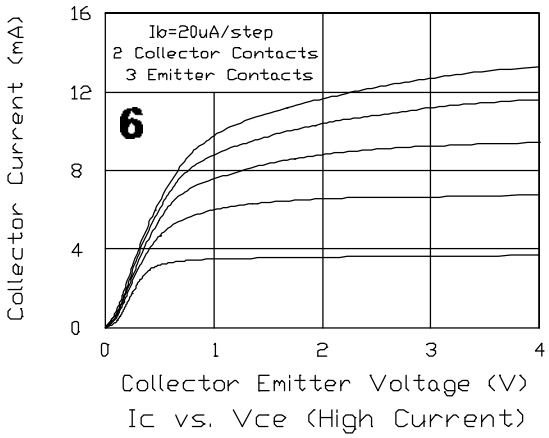


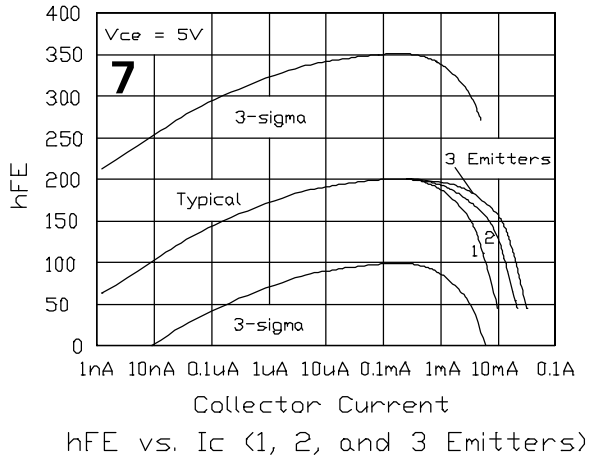
Graphs 3 to 6 Using both collector contacts results in a lower saturation voltage; additional emitters increase the upper current range (i.e the gain at high current).



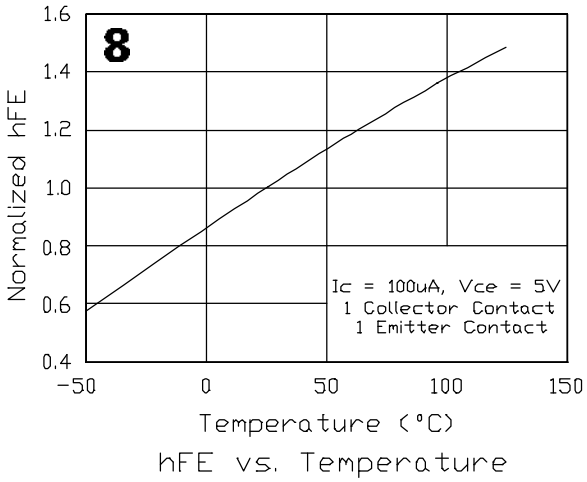




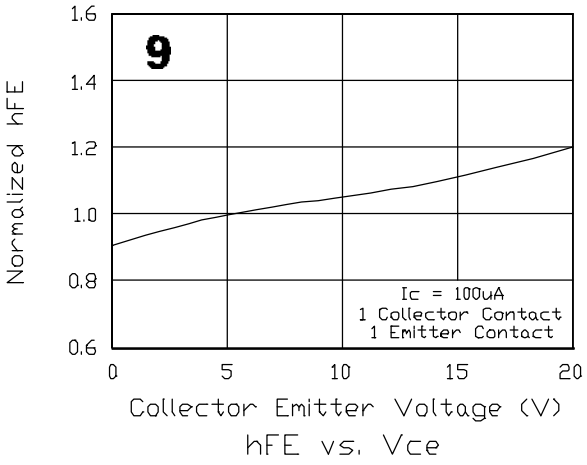




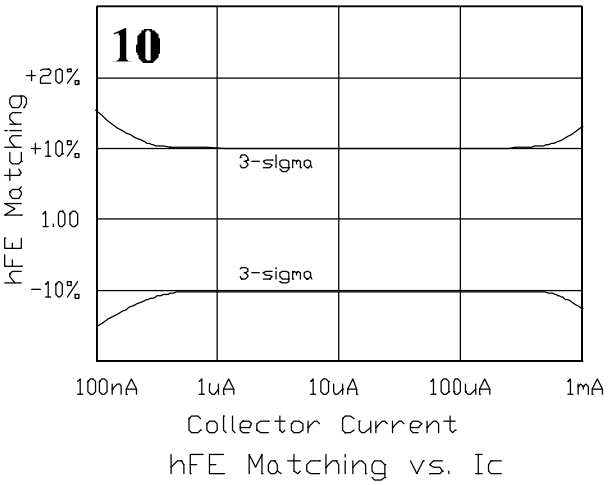
Graph 7 hFE vs collector current. At about 300uA the nominal hFE is 200, with a 3-sigma range of 100 to 350. hFE drops off both at low and high current. You can extend the upper current range by using 2 or 3 emitters.



Graph 8 Normalized hFE is multiplied with the hFE of graph 7 to get the hFE at temperatures other than 25ºC.

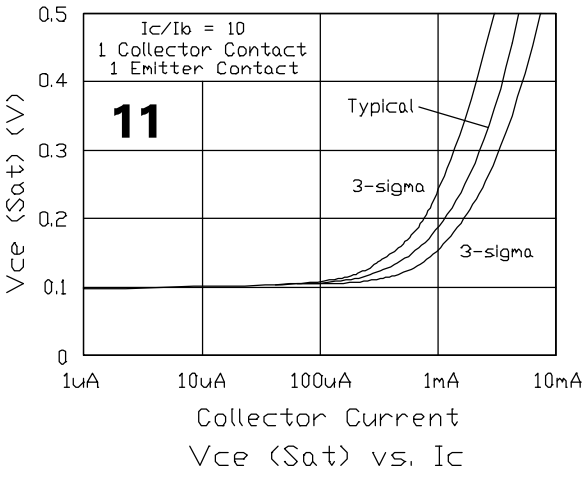


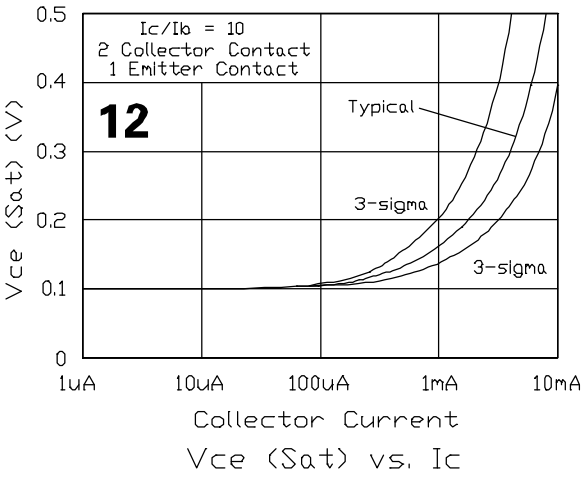
Graph 9 Normalized hFE is again multiplied with the hFE of graph 7. This graph shows the Early effect.



Graph 10 3-sigma matching of hFE between NPN transistors with single emitters is ±10%, except at very low and very high current. We find no significant difference between neighboring devices or devices clear across a die, except when there is a heat source on the chip and one device gets heated more then the other. Matching improves slightly if two or more devices are connected in parallel. See also VBE matching under NPN diode.

Graphs 11 and 12 Using 2 collector contacts reduces the saturation voltage in the upper current range.



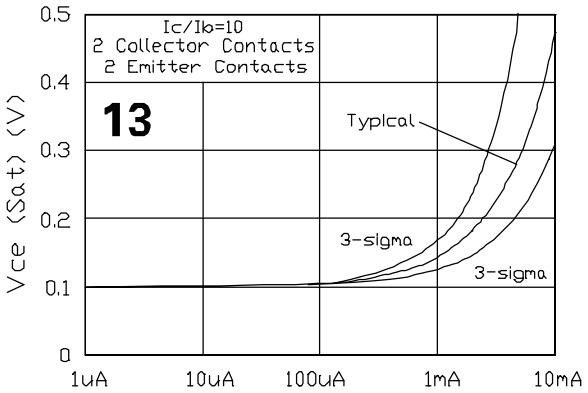


CAUTION: When you use this device as a dual NPN transistor and one of the two transistors saturates, a low-gain PNP transistor is created between the two. This stray device works as follows: the base of the saturating NPN becomes the emitter (with a potential of about 0.65 Volts); the collector of both NPN transistors is the base (in saturation it is near ground); the base of the second NPN transistor now acts as the collector. The gain of this device is about 0.15.

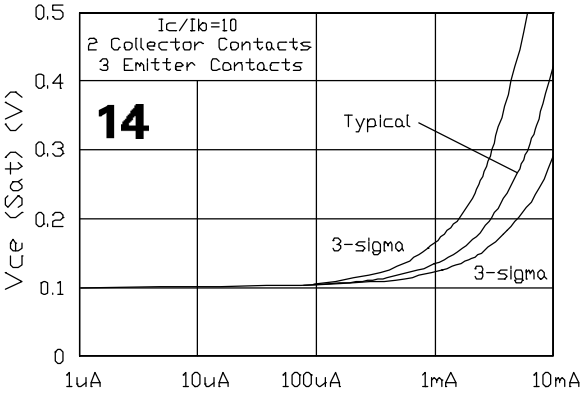
The effect of this stray PNP is only noticed *if the base of the second NPN transistor is open and the base of a third NPN is also connected to this point*. In this case the collector current of the stray PNP flows out of the base of the second NPN and turns on the third.

When using a dual NPN and one of the two transistors may saturate, you should take two precautions: 1) Ground the center contact (the PNP emitter); this reduces the gain of the stray PNP to about 0.02. 2) Provide a leakage path from the base of the non-saturating NPN to ground (e.g. a base or epi pinch resistor or a current sink).

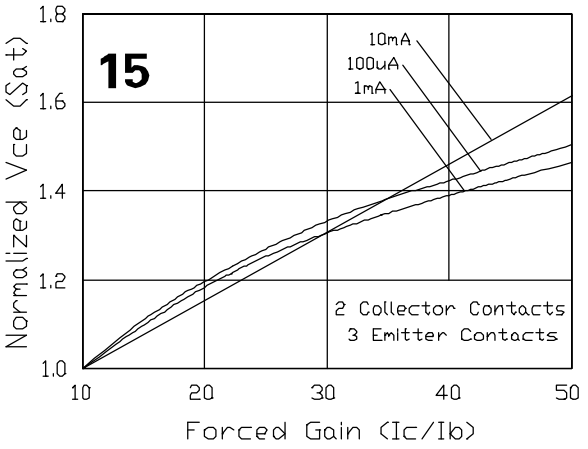
Graphs 13 and 14 Saturation voltage with 2 and 3 emitters. Unless the transistor is operated at low current levels only, connect both collector contacts.



Vce (Sat) vs. Ic



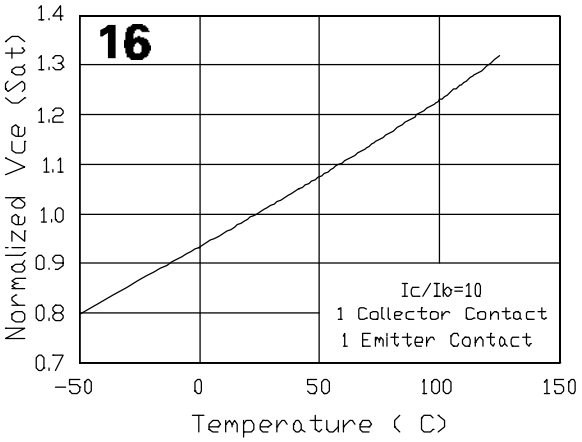
Vce (Sat) vs. Ic



Saturation Voltage vs. Base Drive

Graph 15 Graphs 11 to 14 are shown with a base current which is 10% of the collector current. If you don't want to drive the transistor that hard, use graph 15 as a multiplier for saturation voltage.

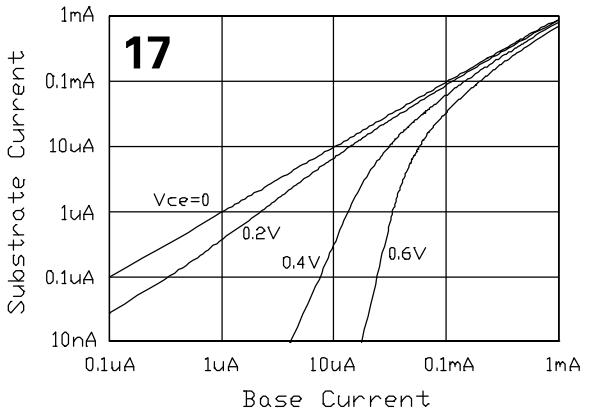
Graph 16 multiplies the saturation voltage for temperatures other than 25ºC.



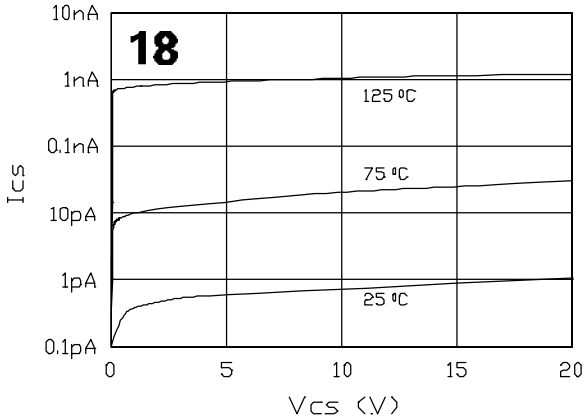
Saturation Voltage vs. Temperature

Graph 17 When the NPN transistor saturates a rather substantial substrate cur- rent flows. This current is collected at the edge of the chip by the -V run. A large total substrate current (greater than about 10mA) could, conceivably, for- ward bias some junctions in its path, especially if its source is in the center of the chip.

The substrate current is directly dependent on the amount of base current and is a strong function on how low you let the collector-emitter voltage drop.

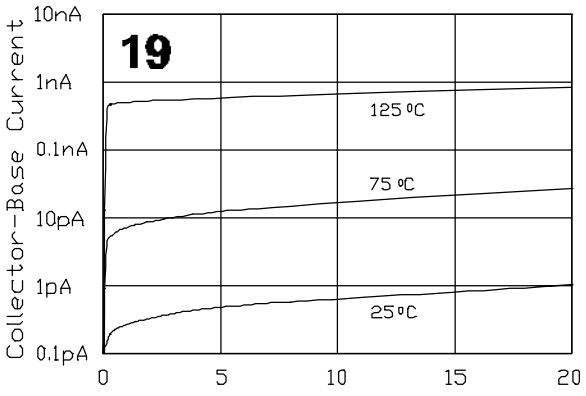


Substrate Current vs. Base Current



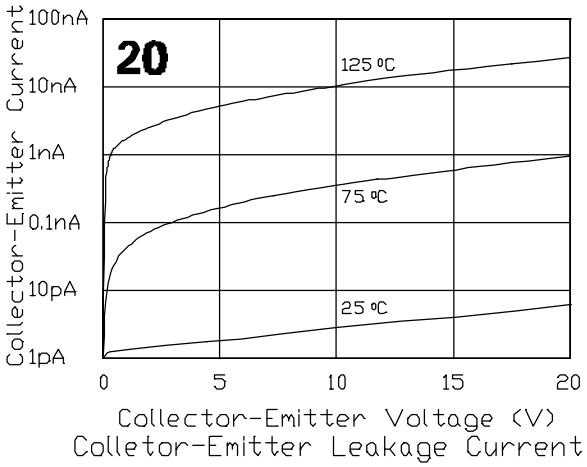
Collector Substrate Leakage Current

Graph 18 The junction isolating the NPN transistor from the substrate has some leakage current, which can be a factor at the upper temperature range in low- current application.



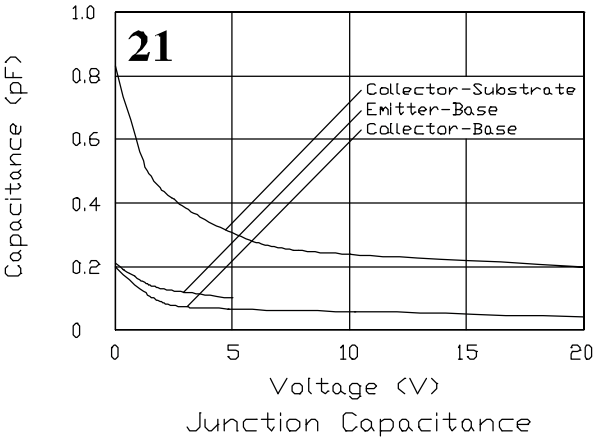
Collector-Base Leakage Current

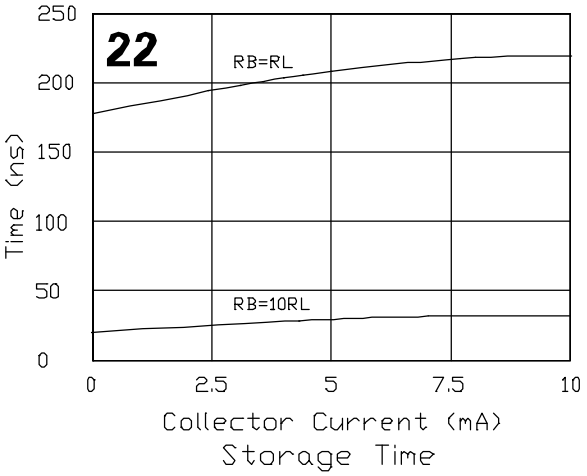
Graph 19 Leakage current of the base-collector junction. For almost all applications insignificantly small.



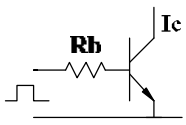
Graph 20 Leakage current from collector to emitter. This is the collector-base leakage current multiplied by the hFE. At high temperature this leakage current becomes significant in low-current applications.

Graph 21 Capacitance of the three junctions vs applied (reverse) voltage. These capacitances are considerably smaller than the capacitances in other 20 Volt semicustom chips, owing to the small (4 micron) dimension used. Band- width (ft) is approx. 800 MHz.

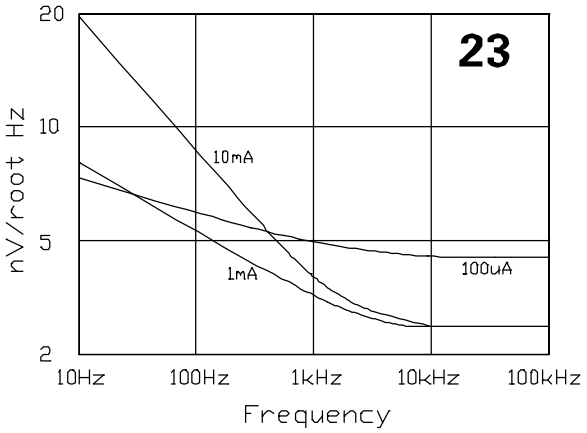


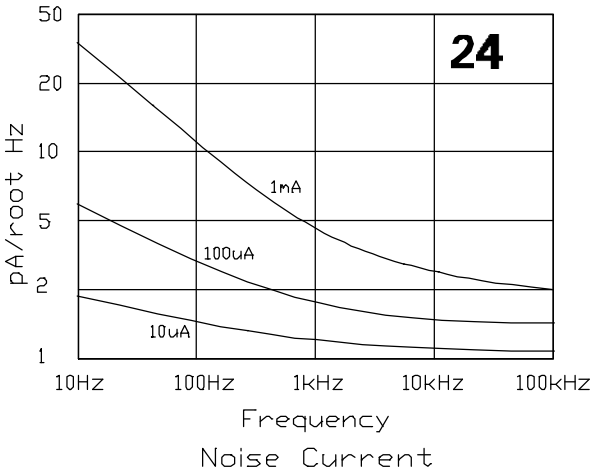


Graph 22 Storage time - the time it takes for the transistor to come out of saturation. If you have a high impedance at the base it takes far more time for the transistor to turn off than with a path (resistor or current sink) which can dis- charge the base. Also, the higher the ratio of Ic/Ib, the faster the transistor will turn off.



Graphs 23 and 24 Noise is plotted in nano-Volts or pico-Amperes per root- Hertz. Also compare these figures with the graphs for the large NPN transistor.





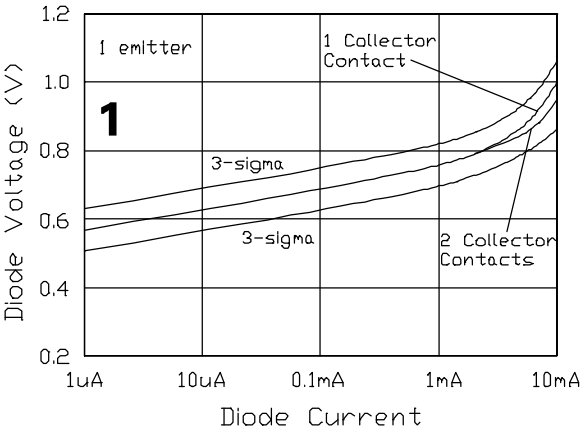
NPN Diode

There are no pure diodes in most ICs, you make them either out of an NPN or a PNP transistor. In the NPN transistor you connect base and collector terminals together (which becomes the anode); the other terminal is the emitter. This diode has a low reverse breakdown voltage (that of the Zener diode, about 5.9 Volts), but is very predictable in the forward direction.

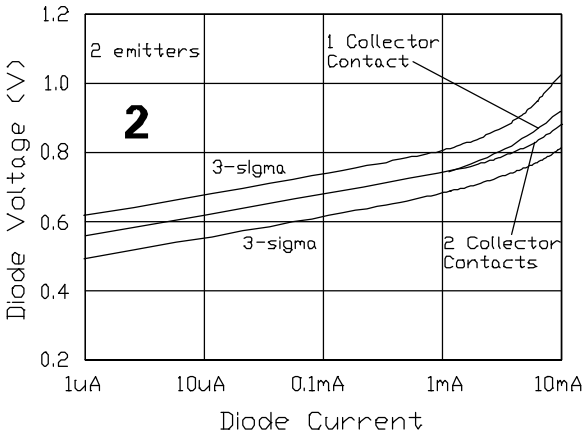
Graphs 1 to 3 The forward voltage drop is highly linear over about 8 decades of current, up to at least 1mA. It follows the equation

Delta-VBE = (kT/q)ln[(A2I1)/(A1I2)]

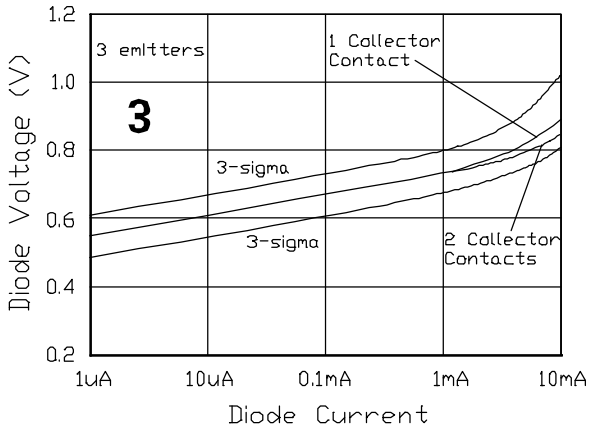
(see also chapter 5). At room temperature kt/q amounts to 26mV, so a ten-fold increase in current will increase VBE by 60mV. Going from 1 to 2 emitters re- duces VBE by 18mV. Note that, for best performance at the upper current range, you need to connect both collector contacts.



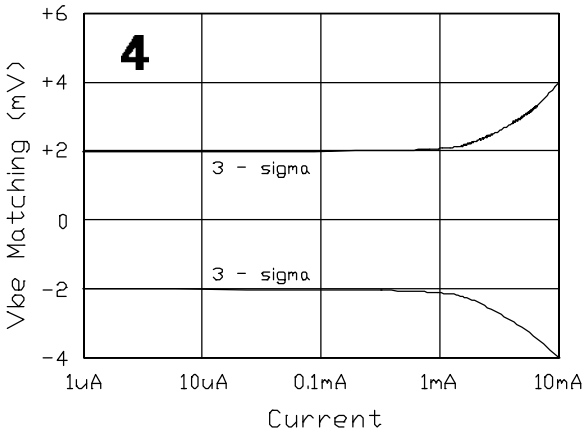
Diode Voltage vs. Current (1 emitter)



Diode Voltage vs. Current (2 emitters)



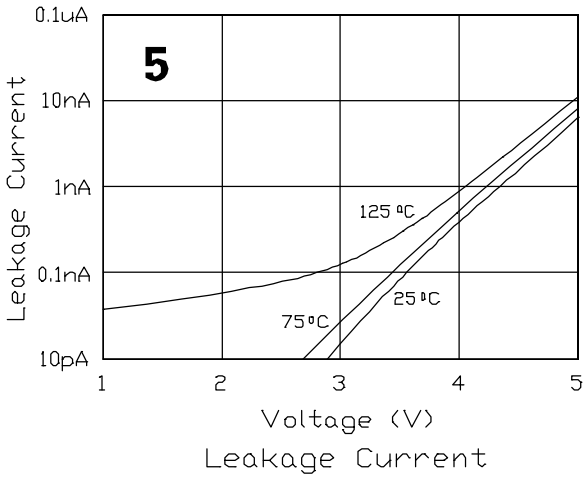
Diode Voltage vs. Current (3 emitters)



Vbe Matching vs. Current

Graph 4 Up to about 1mA 3-sigma VBE matching (between transistors with the same number of emitters) is ±2mV. For best matching use emitters 2 and 8 (emitter 5 has a different orientation). It appears to make very little difference in matching if the devices are adjacent or in other places on the same chip.

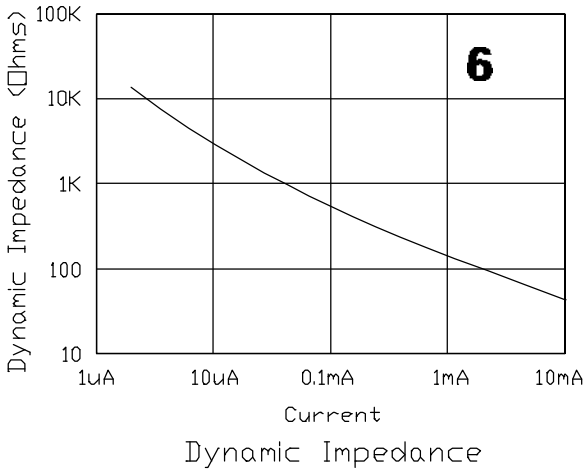
Zener Diode



Leakage Current

The connection for the Zener diode is identical to that of the NPN diode (you should not leave the collector floating). 3-sigma voltage spread is 5.6 to 6.1 Volts, with a nominal breakdown of 5.9 Volts. Temperature coefficient is approx +200ppm/°C.

Graph 5 Leakage current of the base-emitter junction.



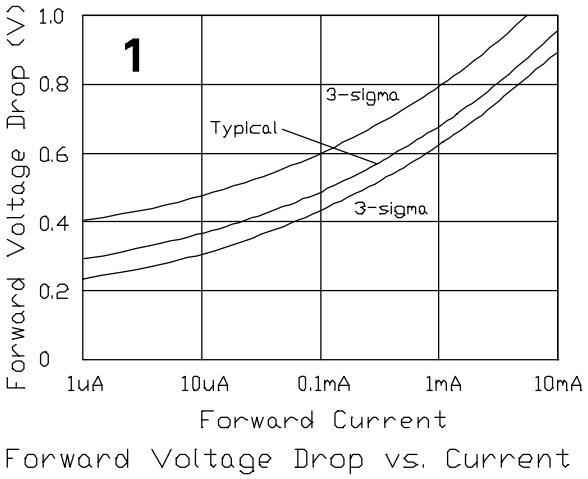
Graph 6 The slope of the Zener diode breakdown curve.

Schottky Diode

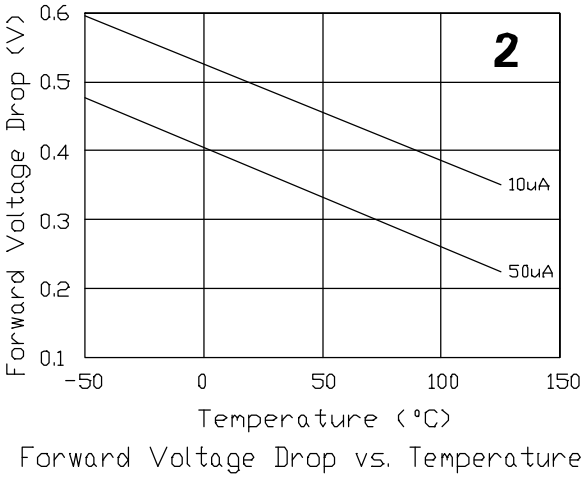
Some transistors in the 700 Series carry a Schottky diode instead of a second base. A Schottky diode is simply the junction between the aluminum and the (high resistivity) epi layer. Compared to silicon p-n junctions, the Schottky diode produces a lower forward voltage drop.

You can use this transistor as:

- a Schottky diode (ignoring the base and two emitters)
- a one or two emitter NPN transistor (ignoring the Schottky diode)
- a Schottky-clamped NPN transistor (connecting the aluminum of the Schottky diode to the base). Such a transistor does not saturate and thus turns off more rapidly.



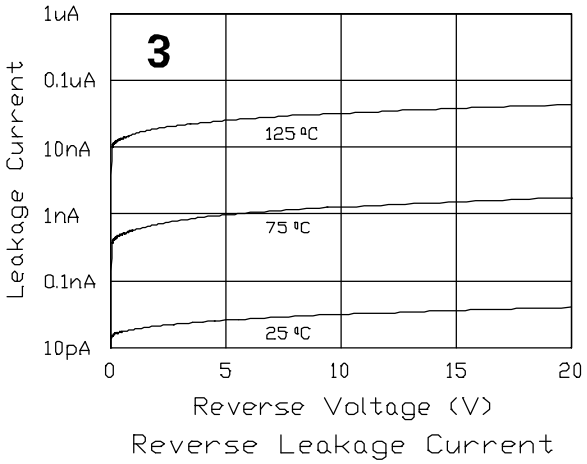
Graph 1 Forward voltage drop vs current. Note that this Schottky diode is de- signed to operate at low current only (about 50uA max).



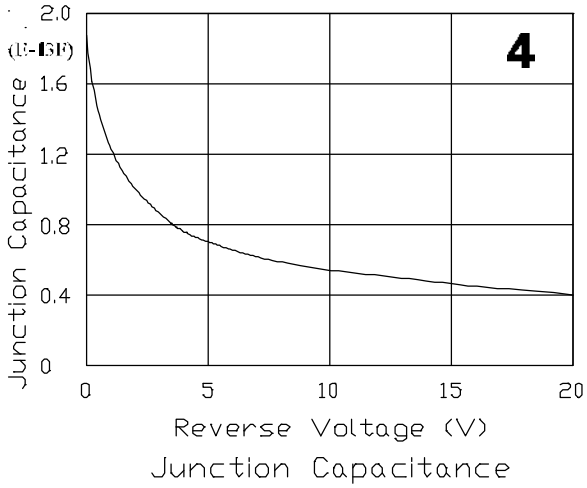
Forward Voltage Drop vs. Temperature

Graph 2 Voltage drop at 10 and 50uA vs. temperature.

Graph 3 Beware that the Schottky diode has a higher leakage current than p-n junctions.



Reverse Leakage Current

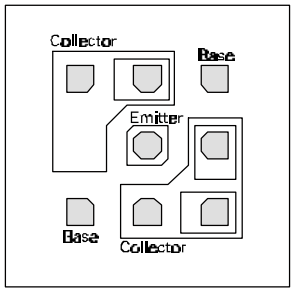


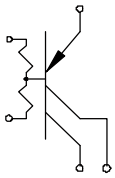
Junction Capacitance

Graph 4 Junction capacitance vs applied (reverse) voltage.

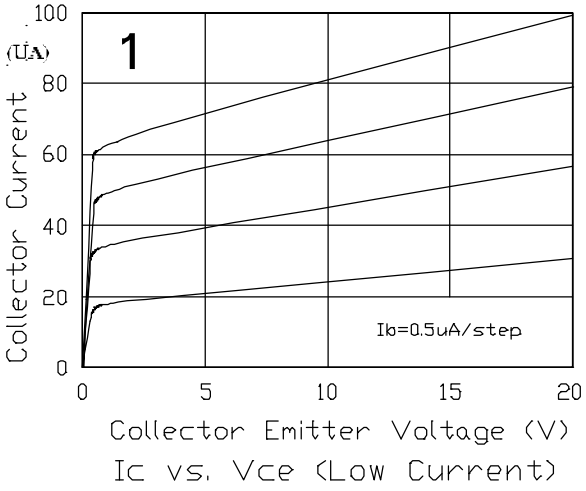
PNP Transistor

The structure that houses NPN devices can also be used as a PNP transistor. The center p-type region, unused in the NPN mode, becomes the PNP emitter. It emits current radially or laterally through the base (the epi layer) and this current is collected by two identical L-shaped regions (the NPN bases). Each collector gets half the current. The NPN emitters are ignored.



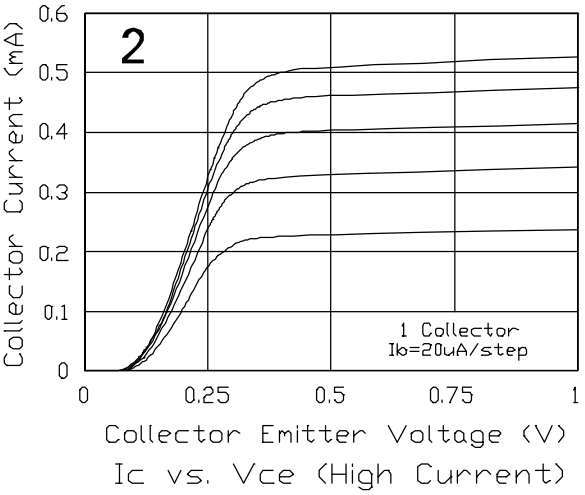


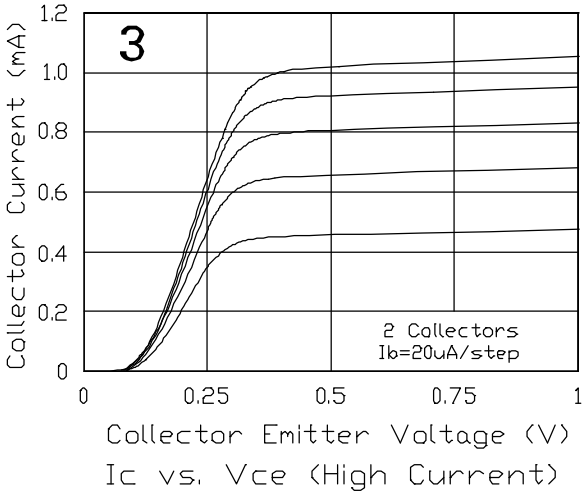
Since the base is fairly wide, the lateral PNP transistor has a lower current gain than the NPN device and its speed is inferior (ft is approx. 30MHz).

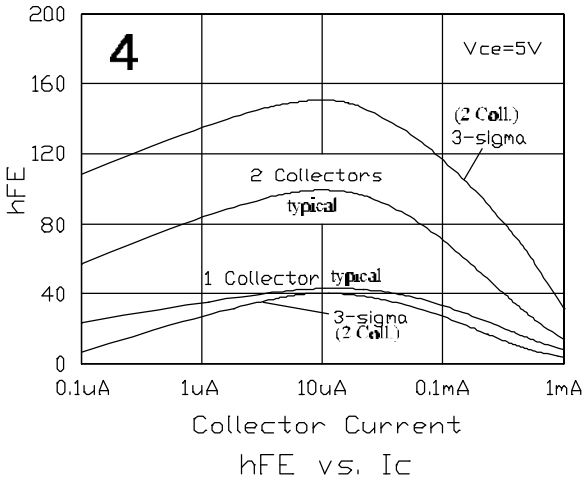


Graph 1 Notice the Early effect (the upward sloping curves).

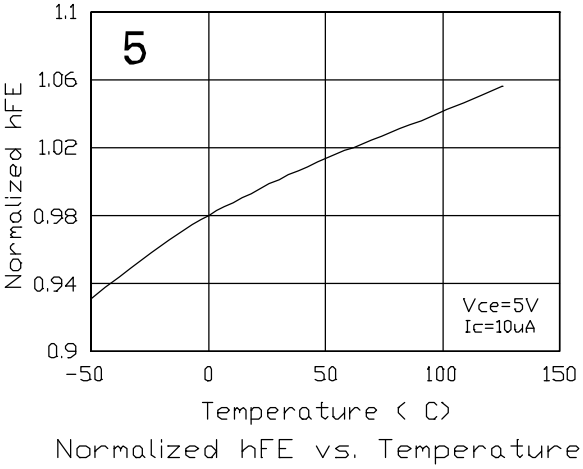
Graphs 2 and 3 The PNP transistor has a collector offset voltage of about 100mV, below which the saturation voltage cannot fall at any current level.



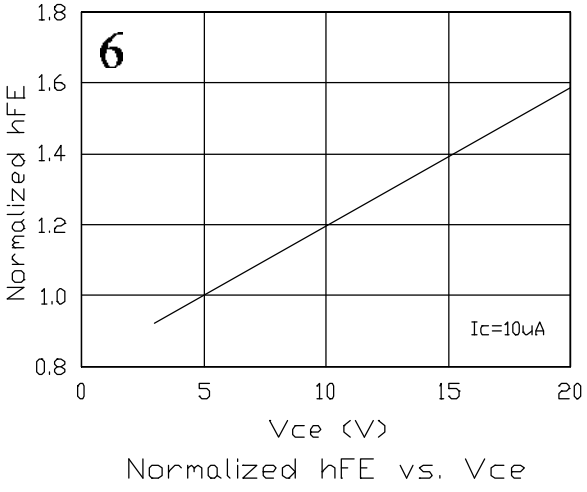




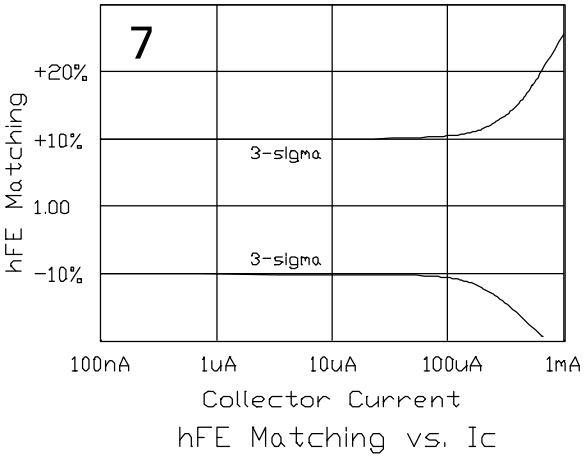
Graph 4 There is a significant hFE drop-off above about 100uA. In general, this device is rarely useful above 300uA (i.e. 150uA per collector). Both collectors are always active; you cannot simply ignore one collector. If you don't need to split the current into two equal parts, connect both collectors together.



Graph 5 Multiply this curve with the curve of graph 4 to get hFE at temperatures other than 25°C.

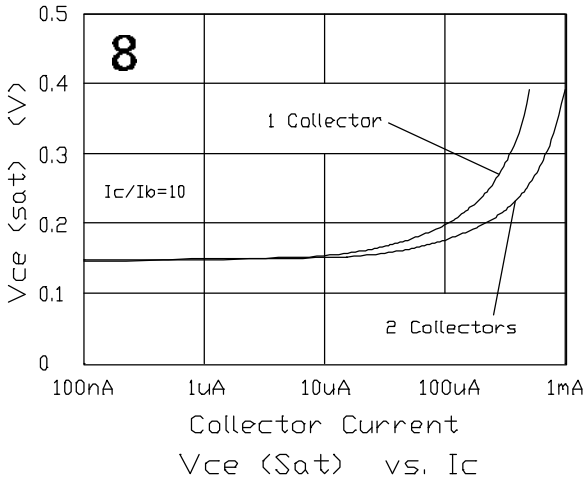


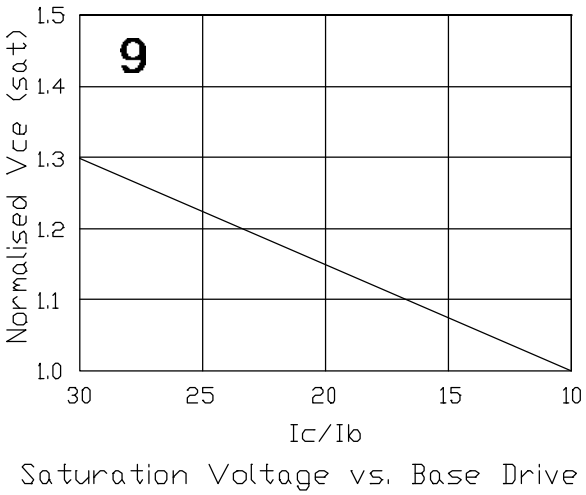
Graph 6 Multiply this curve with the curves of graphs 4 and 5 to get hFE at voltages other than 5 Volts. It depicts the Early effect.



Graph 7 The 3-sigma matching of hFE is ±10% and increases at the upper current range. We find very little difference in matching between neighboring de- vices and other devices on the same chip, except in the presence of heat sources. See also VBE matching under PNP diode.

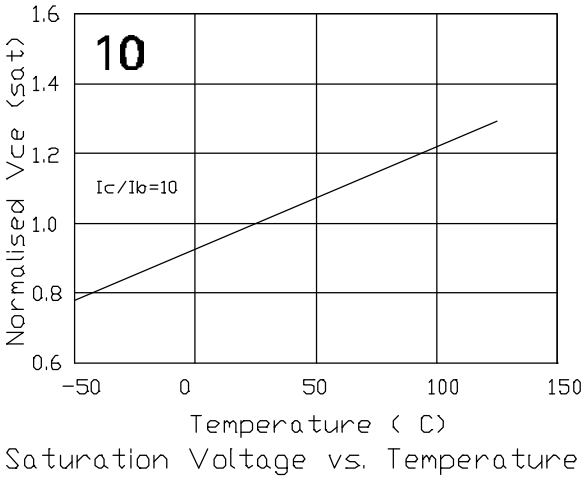
Graph 8 Saturation voltage vs collector current. You can, of course connect PNP transistors in parallel to get lower saturation voltage and higher current capability.

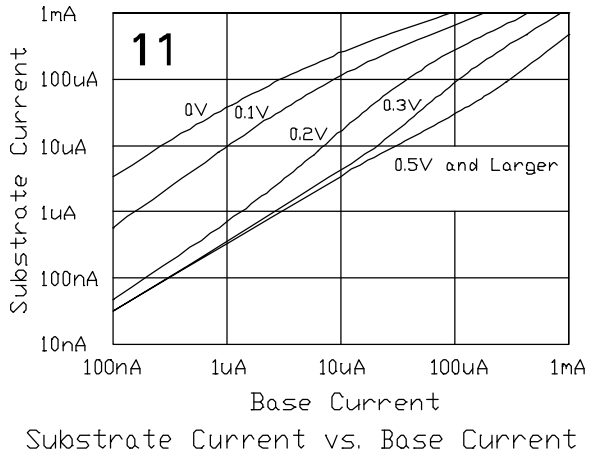




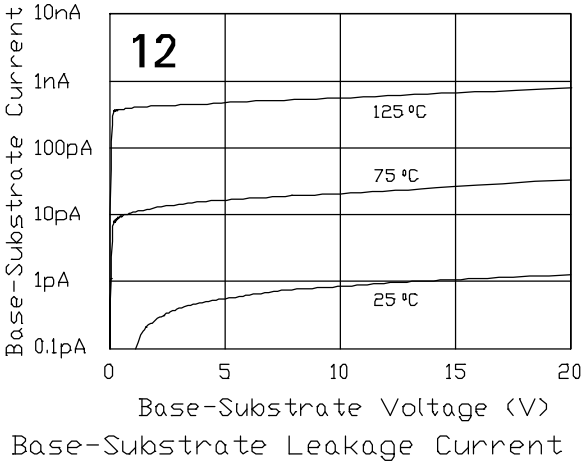
Graph 9 Multiply this curve with the curve of graph 8 to get saturation voltage if you use a base current which is less than 10% of the collector current.

Graph 10 Multiply this curve with the curve of graph 8 to get the saturation voltage at temperatures other than 25°C.

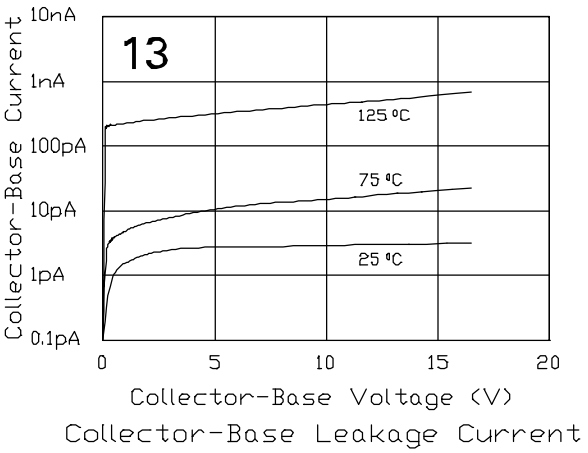




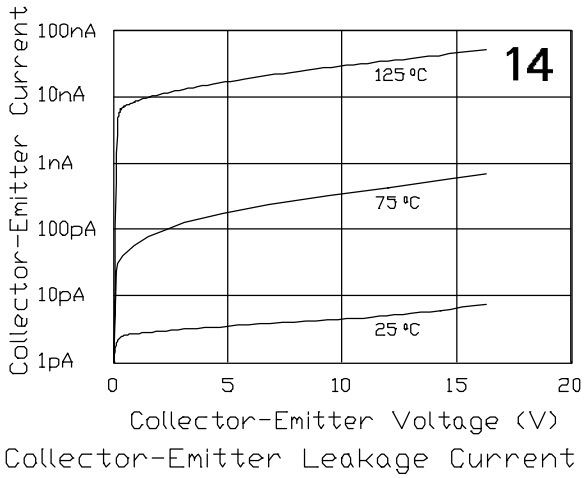
Graph 11 The PNP transistor always has some current flowing to the substrate, a current which is proportional to its base current. When any of its collectors saturate this flow of current becomes substantial. It is collected at the edge of the chip by the -V run. A large substrate current (greater than about 10mA) could, conceivably, forward bias some junctions in its path, especially if its source is in the center of the chip. So, be careful about letting too many PNP transistors saturate with a high base current. You can measure this current in the substrate pins of the kit-parts.



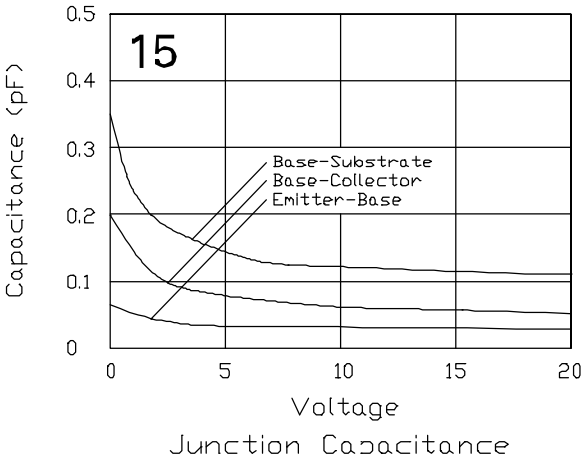
Graph 12 Leakage current from base (the epitaxial layer) to substrate. Can be significant at high temperature with very low base currents.



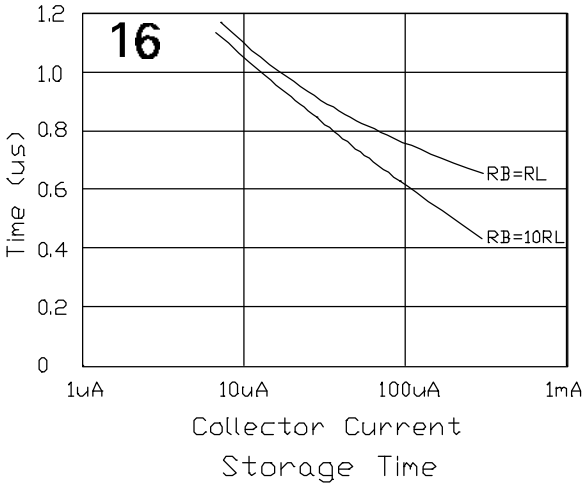
Graph 13 Collector-base leakage current. Usually too small to worry about.



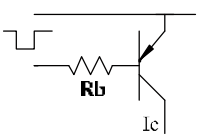
Graph 14 Leakage current from collector the emitter. This is the collector-base leakage current multiplied by the hFE. At high temperature this leakage current becomes significant in low-current applications.



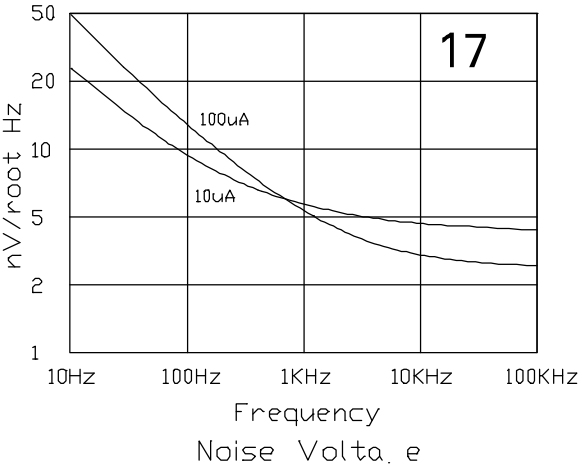
Graph 15 Junction capacitances vs applied (reverse) voltage. These capacitances are small, owing to the small (4 micron) dimensions used. The PNP transistor is still a slow device, however, due to its wide base width.

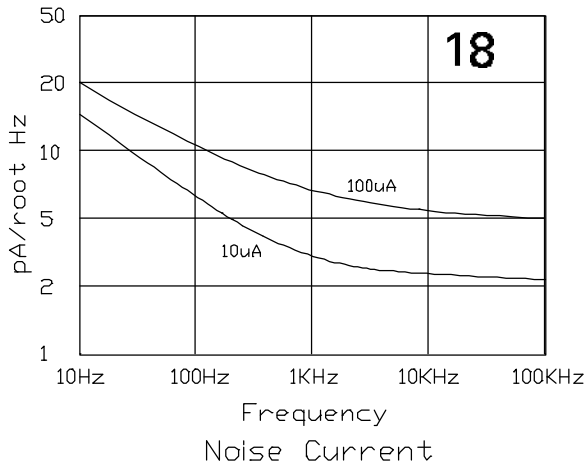


Graph 16 Storage time (the time it takes the transistor to turn off after you remove the base signal) vs collector current. You can speed up storage time by providing a path to discharge the base (a resistor or current source) and using a high Ic/Ib ratio. You should not use lateral PNP transistors for speeds of less than 100nsec, the models lack accuracy at extreme speed.



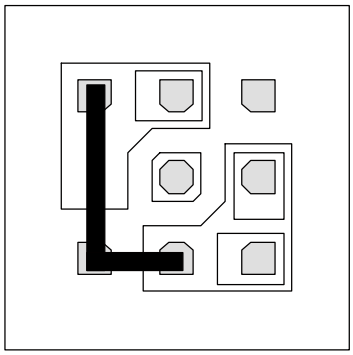
Graphs 17 and 18 The lateral PNP transistor is not particularly noted for low noise. Compare this with the noise of the NPN (and particularly the large NPN) transistor.

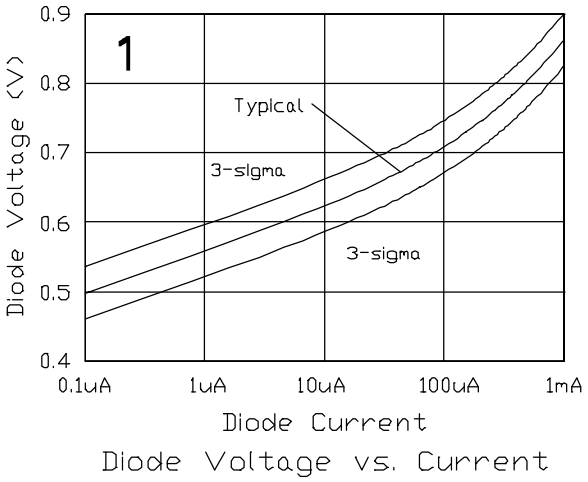




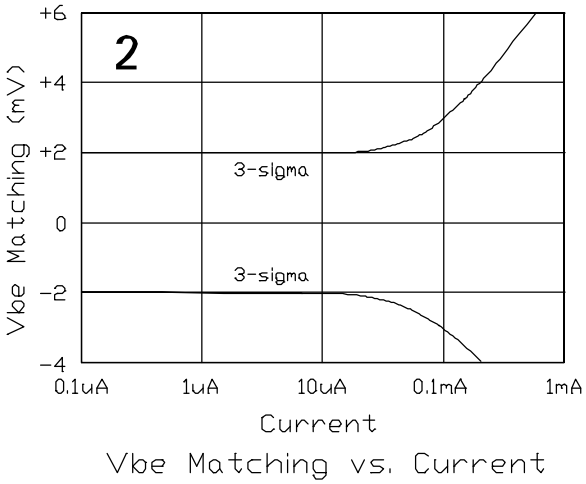
PNP Diode

As we noted before, the NPN diode has a low breakdown voltage (the Zener voltage, about 5.9 Volts). If you need a diode with a breakdown voltage of up to 20 Volts, use the PNP transistor, connecting together the collectors and the base.





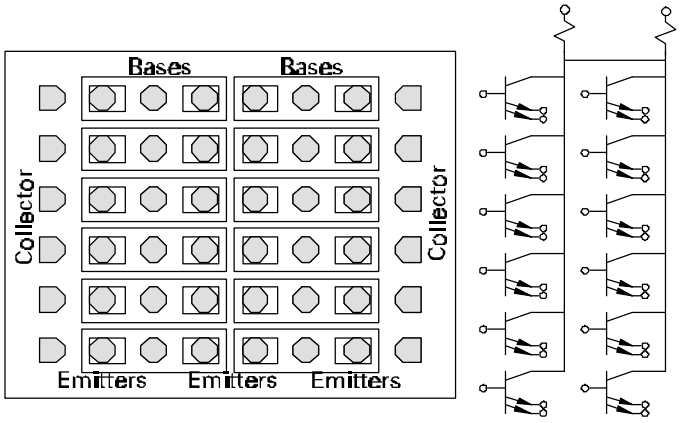
Graph 1 The current capability of the PNP transistor is limited to about 300uA. Also not that, if you go higher in current, there will be significant substrate cur- rent.



Graph 2 Matching of VBE (in PNP transistors or diodes) is ±2mV, increasing toward the upper current range as the device runs out of current gain. It appears to make very little difference in matching if the devices are adjacent or in other places on the same chip.

Large NPN Transistor

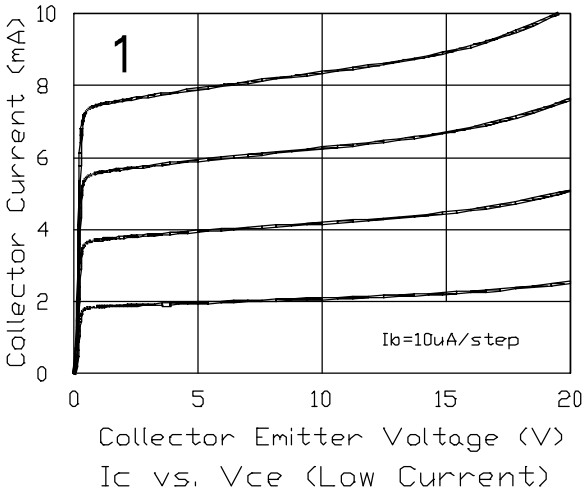
The large NPN transistor, located on the periphery of each chip, consists of 24 emitters, each identical to the emitter of the (small) NPN transistor. There are 12 separate base regions, each containing 2 emitters. All of these emitters and bases are in a single collector region with two opposite contacts.

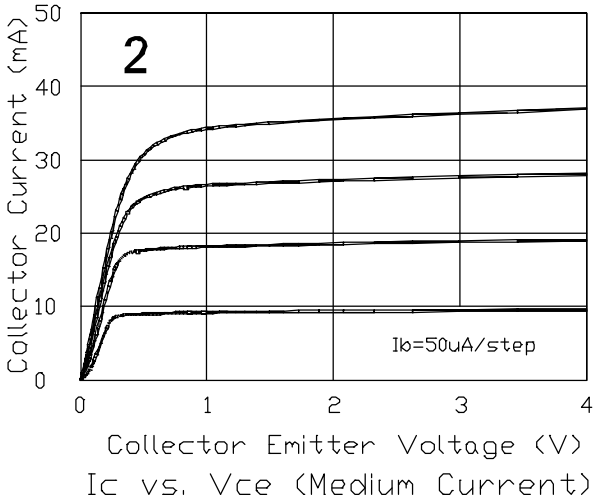


Connect as many emitters and bases as you like (however, if an emitter is used, the base containing it must be connected too). In this way you can:

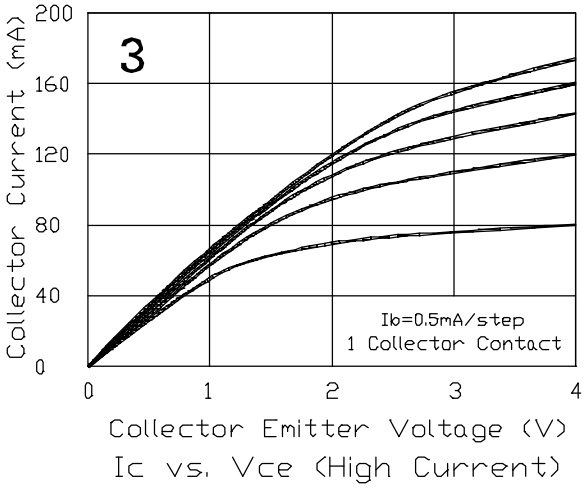
- create current ratios between 1 and 24,
- get an increased current handling capability (up to 200mA) and
- create a multiple base / multiple emitter device with a common collector.

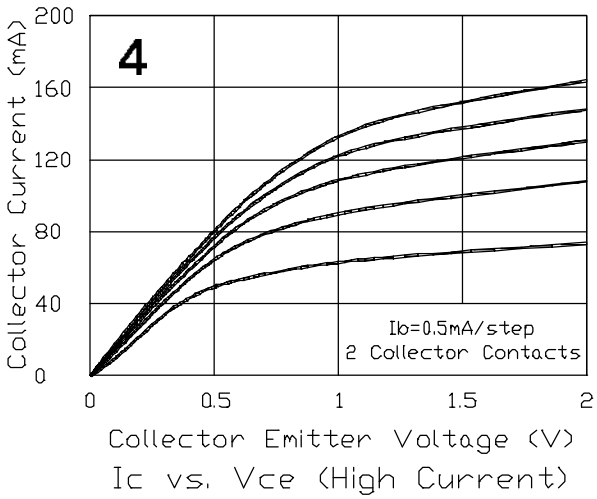
Graphs 1 and 2 The large NPN transistor has the same Early effect (the in- creased current gain toward higher collector-emitter voltage) as the (small) NPN transistor.





Graphs 3 and 4 Connecting both collector contacts results in a lower saturation voltage. You can, of course, connect a second large NPN transistor in parallel, but be aware of wide metal runs necessary for the collector and emitter leads.





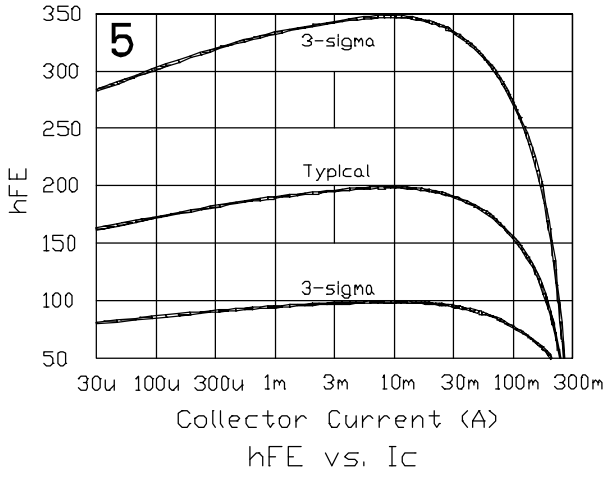
A Note on Metal Widths

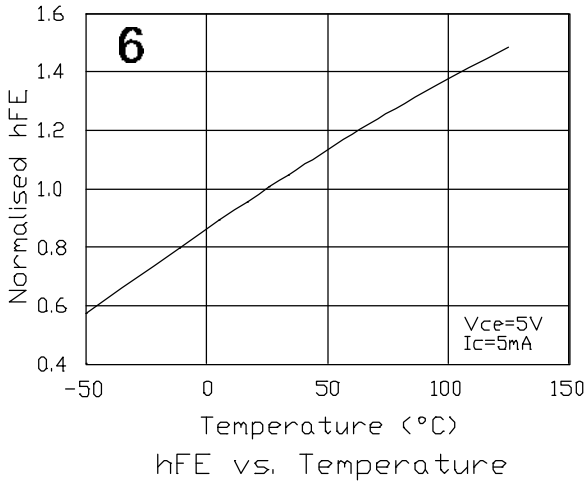
If you exceed a certain current density in the aluminum, you can cause migration or even melting. We use as a safe upper limit $4x10^5$ Amperes/cm2, which corresponds to 4 mA per micron width. The normal (and minimum) metal width is 4 microns, so you can run 16mA safely through the smallest metal run.

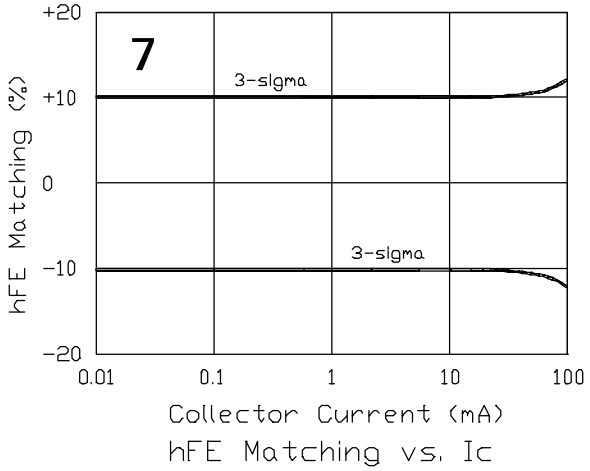
To avail yourself of the 200mA capability of the large NPN transistor you need to use a metal width of 50 microns.

You should also be aware of the resistance of the aluminum layer which, although small, can add up to significance in a long, narrow run. The nominal resistance of the aluminum is 40mOhms/square, with an upper (3-sigma) limit of 80mOhms/square. To calculate the resistance of a run you simply divide its length by its width to get the number of squares. A 4-micron wide, 80-micron long run, for example, has 20 squares, amounting to 0.8 Ohms nominal.

Graphs 5 and 6 The current gain drops off long before you reach 200mA, but is still high enough to be useful. Consider a compound transistor (see chapter 5) if you need a higher current gain. The curve of graph 6 is multiplied with graph 5 to get hFE at temperatures other than 25ºC.

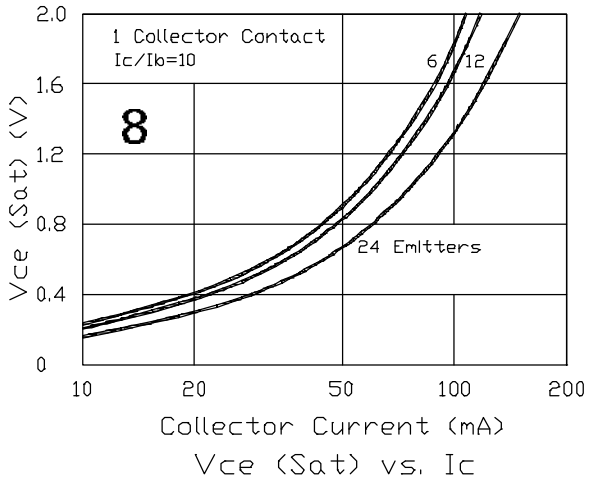




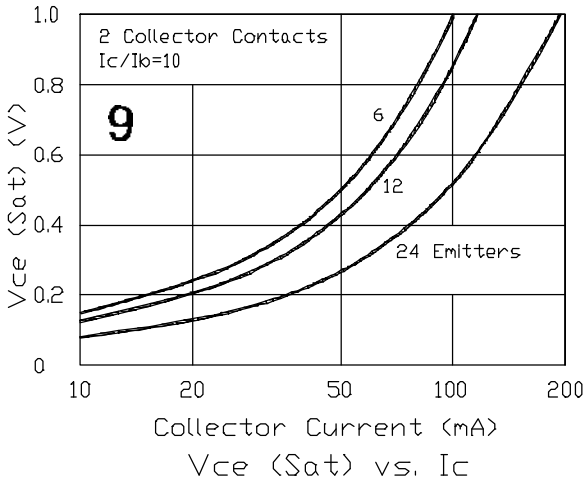


Graph 7 Matching of hFE for single emitter or current ratios to single emitters. The more emitters you match (e.g. 12:12) the better the matching.

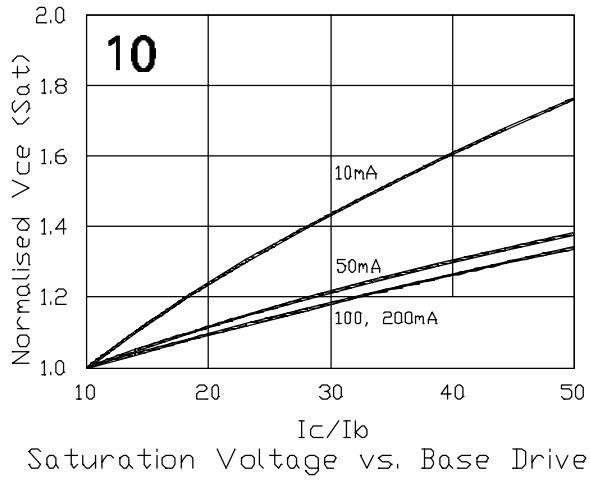
Graphs 8 to 11 Current handling depends strictly on the number of emitters you use. Note the difference in saturation voltage between one and two collector contacts. If you drive the base with less than 10% of the collector current, use graph 10. Graph 11 multiplies the other graphs for temperatures other than 25ºC.



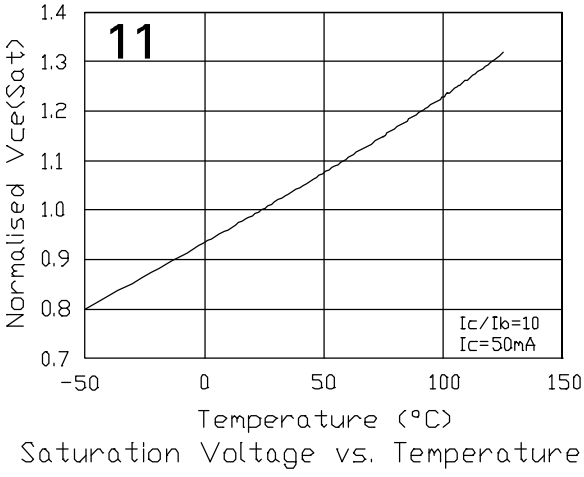
Vce (Sat) vs. Ic



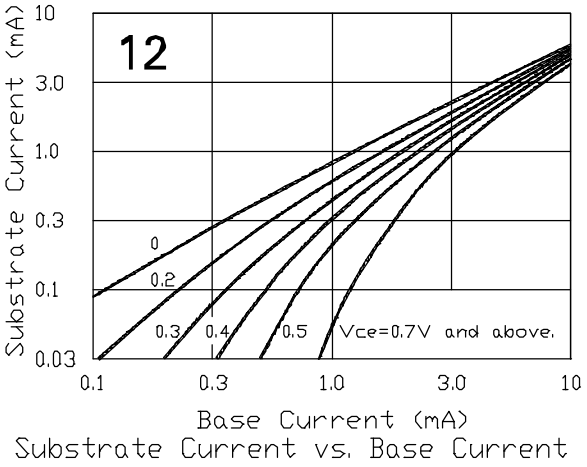
Vce (Sat) vs. Ic



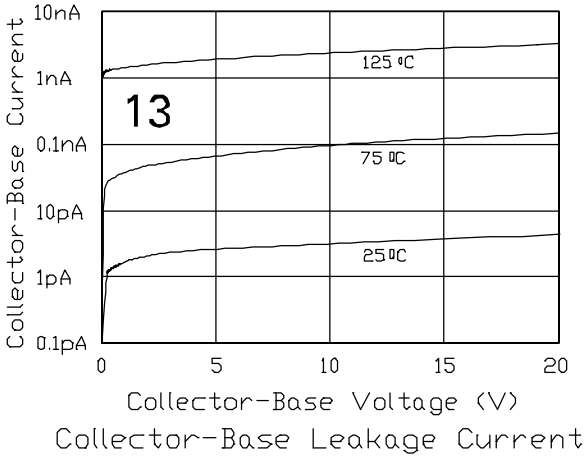
Saturation Voltage vs. Base Drive



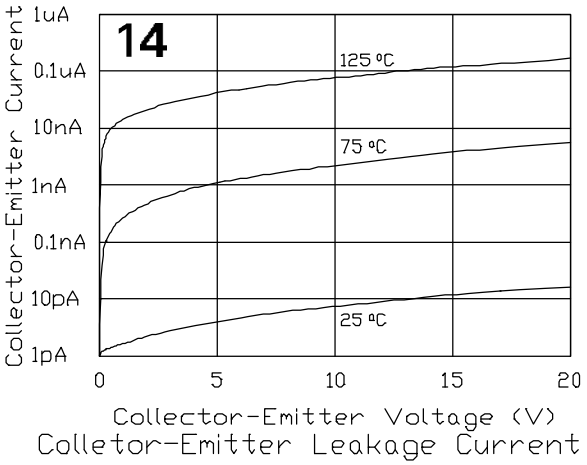
Saturation Voltage vs. Temperature



Graph 12 In saturation there is a substantial substrate current. That is why the large NPN transistors are located close to the edge of the chip where the substrate current is collected.

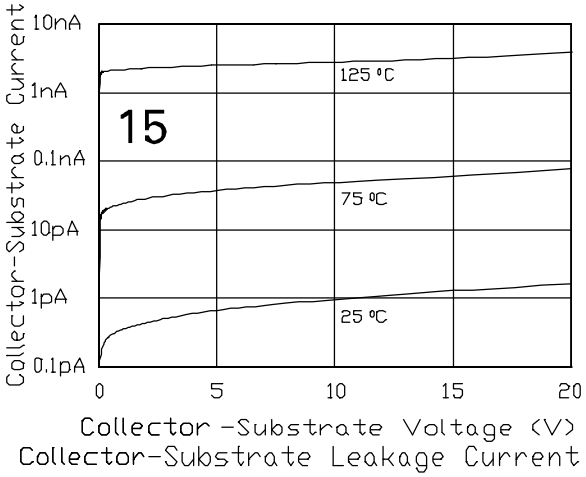


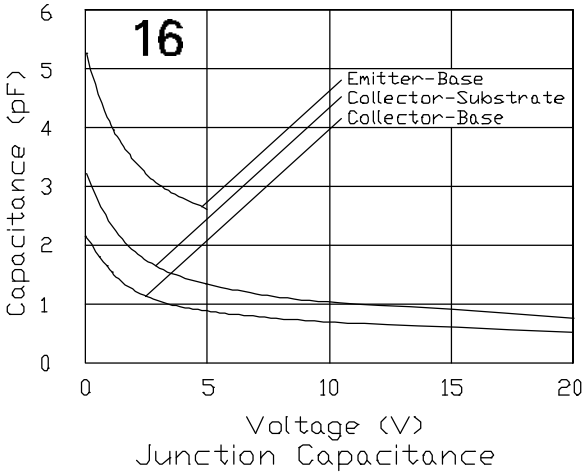
Graph 13 Due to its large size (comparatively, that is), even the collector-base leakage current can become significant at high temperature.



Graph 14 Don't ignore the collector-emitter leakage current of the large NPN transistor, especially at high temperatures. If you provide a path between base and emitter, it will drop to the level of the collector-base current.

Graph 15 There is also a leakage current between the collector (the epitaxial layer) and the substrate.

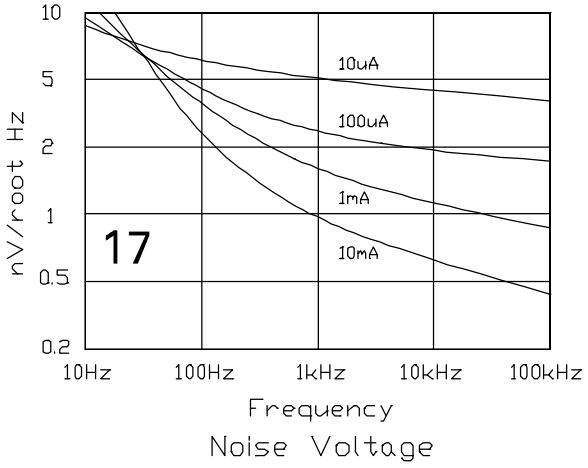


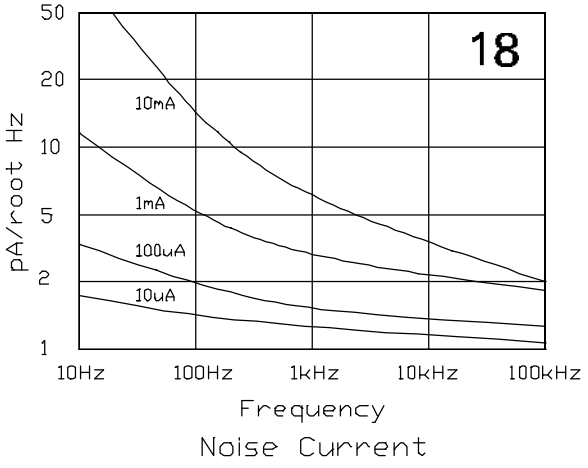


Graph 16 Junction capacitances vs applied (reverse) voltage. Larger than those of the small device. The large NPN transistor has an ft of about 400MHz.

.

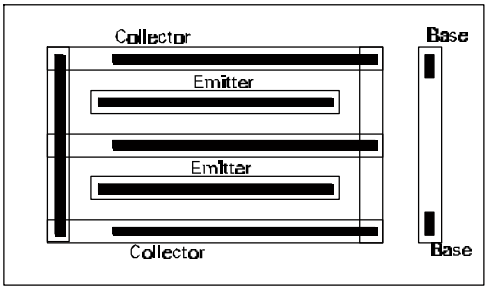
Graphs 17 and 18 The large NPN transistor, operated at the current levels of the (small) NPN transistor, makes an excellent low-noise device.





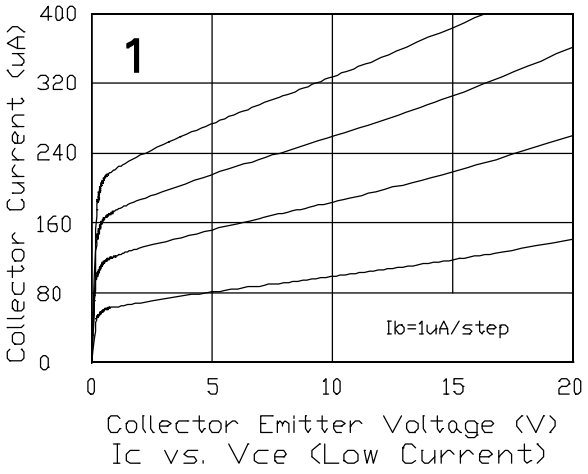
Large PNP Transistor

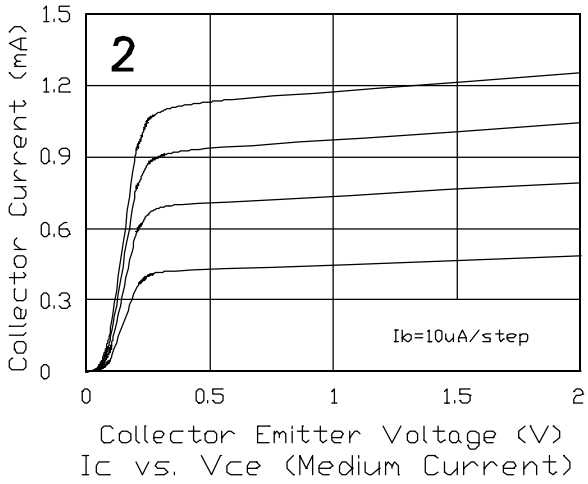
The large PNP transistor uses the same layers as the (small) PNP transistor, but there are two long emitters, fully enclosed by the collector. This device is capable of carrying a larger current (6mA), though it is still limited in frequency response and current gain.

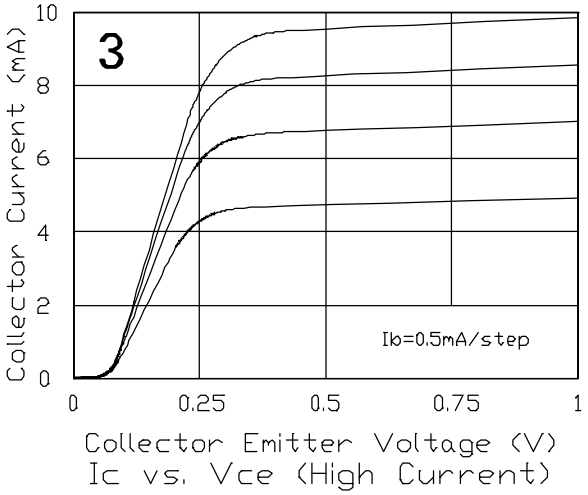


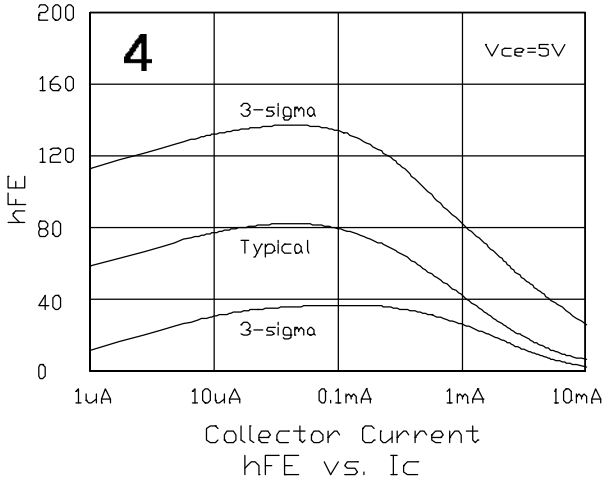
The two emitters are connected together and all three of the parallel collector-metal lines must be used. The transistor has an optional cross-under in the base connection.

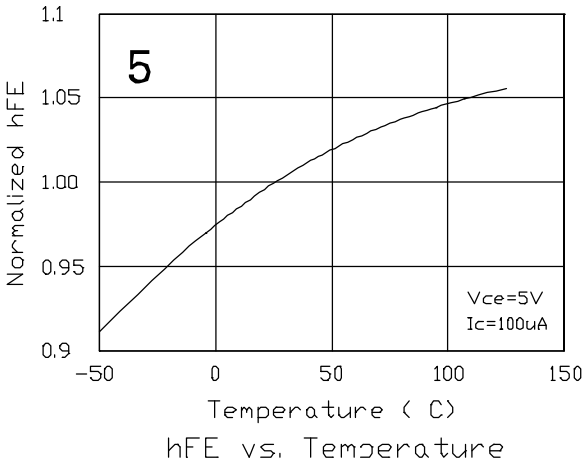
Graphs 1 to 4 The large PNP transistor has the same Early effect as the smaller device and, at about 200uA, its hFE starts to drop off. Up to about 6mA the gain is still high enough to be used in a compound transistor connection (see chapter 5) together with the large NPN transistor or as an emitter follower in a class B output stage.



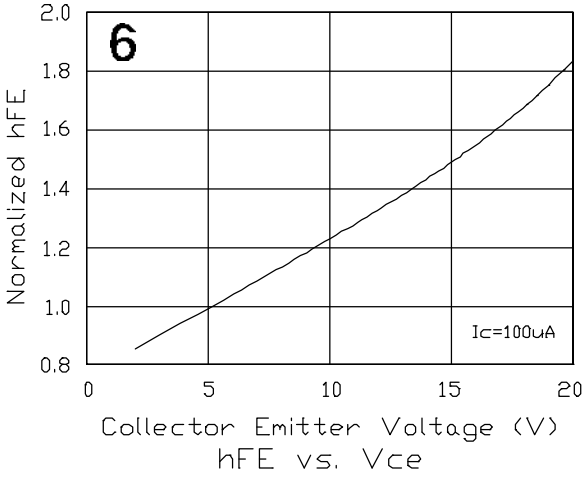




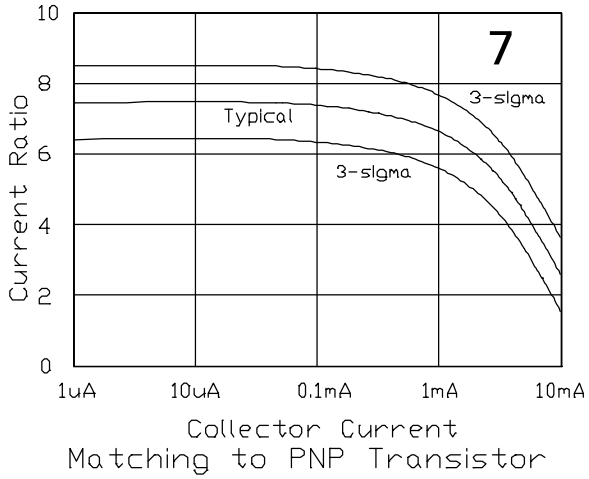




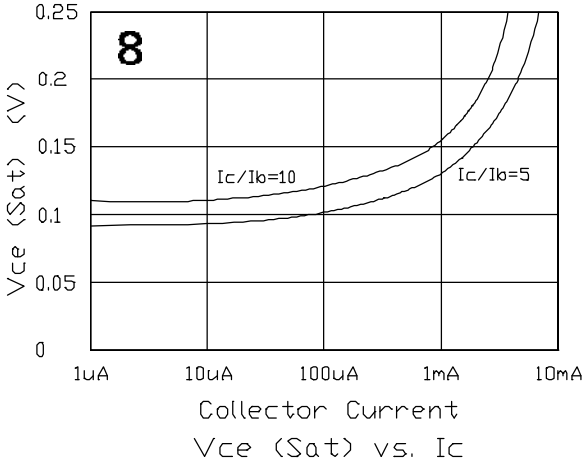
Graph 5 Multiply this curve with graph 4 to get hFE at temperatures other than 25oC.



Graph 6 Multiply this curve with graphs 4 and 5 to get hFE at collector-emitter voltages other than 5 Volts. This graph depicts the Early effect.

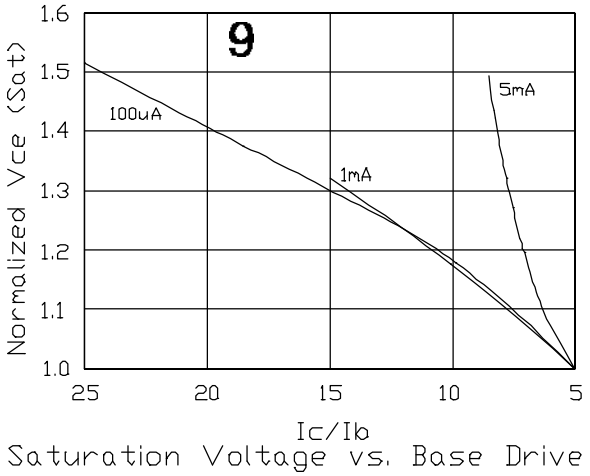


Graph 7 You can count on some fixed current ratio between the large and the small PNP transistors (about 7.5), though its accuracy is not as good as that between identical devices.

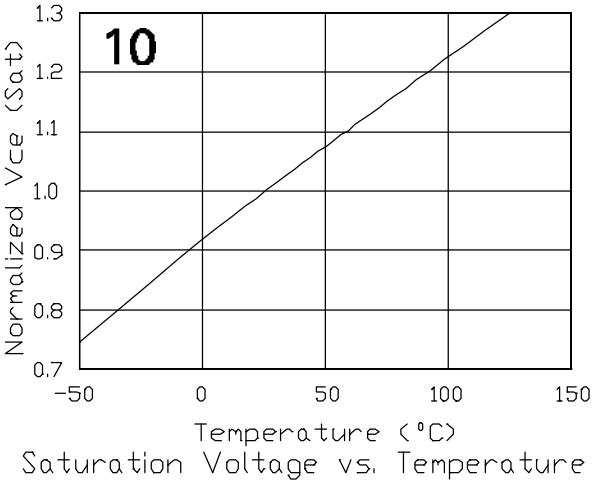


Graph 8 You will need to drive the large PNP transistor hard to get the best saturation voltage and current capability.

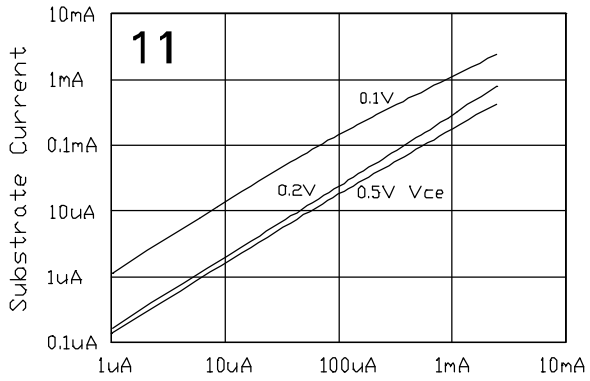
Graphs 9 and 10 These graphs work together (as multipliers) with graph 8.



Saturation Voltage vs. Base Drive



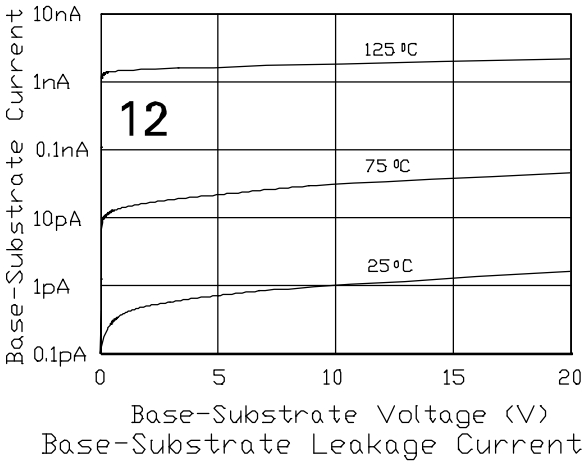
Saturation Voltage vs. Temperature



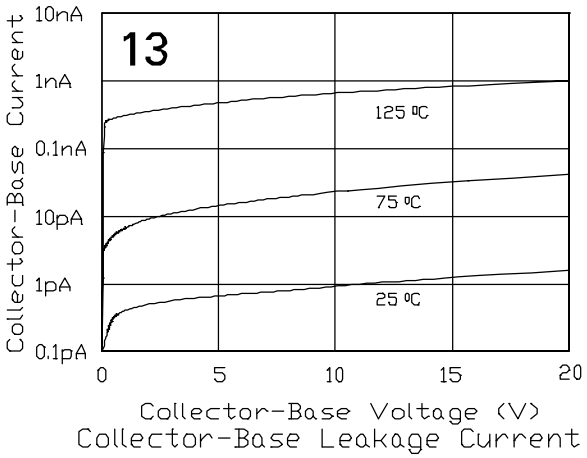
Substrate Current vs. Base Current

Graph 11 The large PNP transistor creates a substantial substrate current, especially when it saturates. That is the reason we located it near the edge of the chip, where this current can be collected by the -V trace.

Graph 12 There is leakage current between the base (the epitaxial layer) and the substrate, especially at high temperature. Usually too small to matter, unless you operate the device with very low base currents.

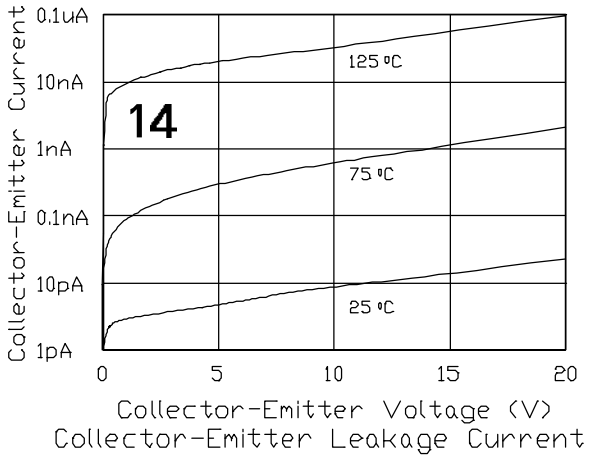


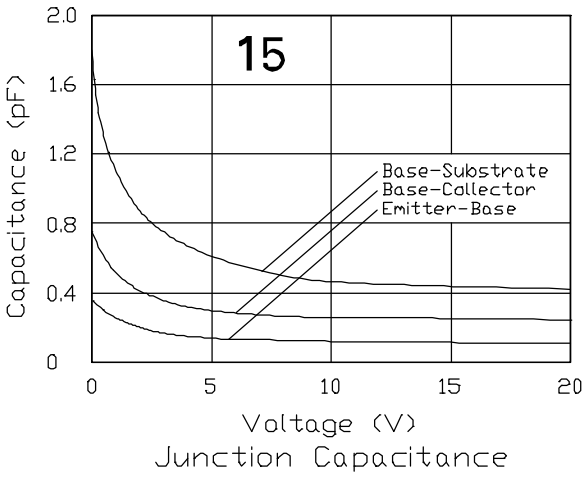
Base-Substrate Leakage Current



Graph 13 Leakage current between collector and base. Usually no bother, un- less you run the transistor below 100nA.

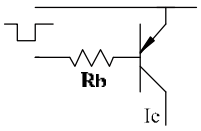
Graph 14 The leakage current between collector and emitter is that of graph 13, multiplied by hFE. You can avoid this multiplication by providing a path between emitter and base.

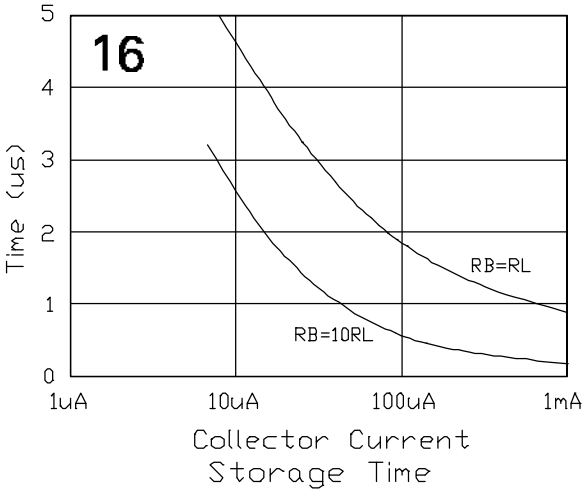




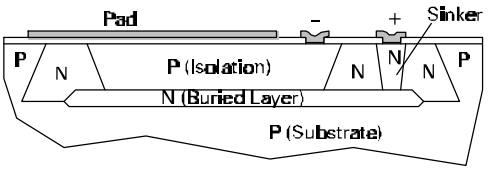
Graph 15 Capacitance of the three junctions vs applied (reverse) voltage.

Graph 16 The large PNP transistor, as is its smaller counterpart, is not a fast device. You can speed up turn-off somewhat by providing a discharge path for the base (a resistor or current source).



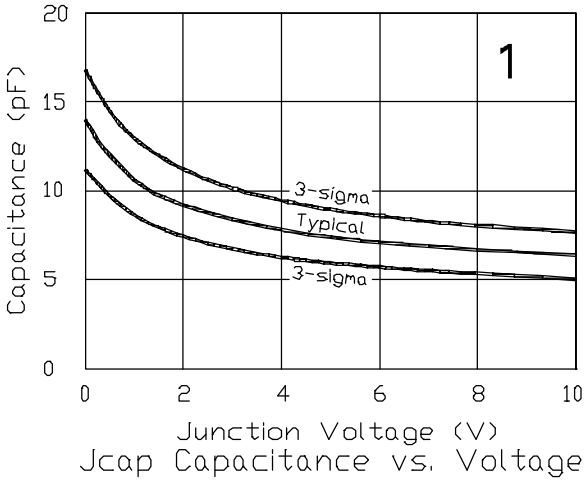


Junction Capacitor



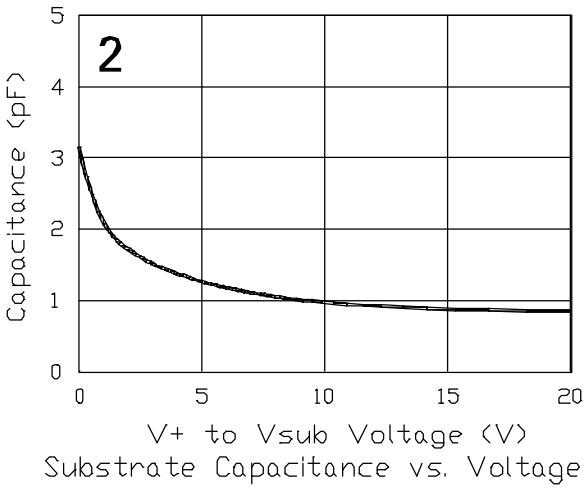
Underneath some of the pads there are structures which utilize the relatively large capacitance between the isolation diffusion and buried layer (both of these layers have a high doping concentration). Minimum breakdown voltage is 9 Volts, which limits their use to low-voltage nodes. The pad above can still be used, it is isolated from the capacitor by the oxide layer.

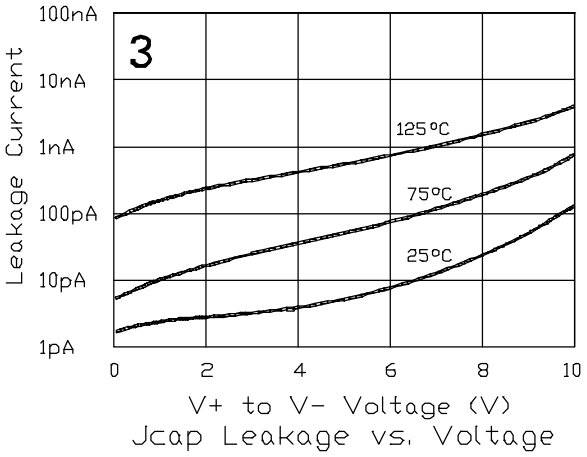
Since this is a reverse-biased junction, one terminal must be consistently more positive than the other.



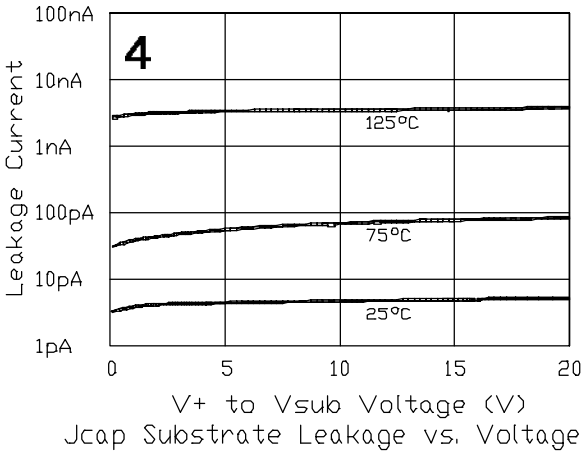
Graph 1 The capacitance of a junction is voltage dependent. You get the highest capacitance at 0 Volts.

Graph 2 There is also a stray capacitance between the positive terminal and the substrate.



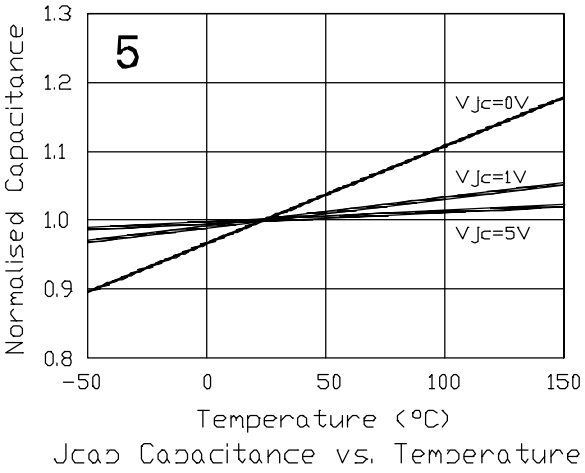


Graph 3 Leakage current between the two capacitor terminals.



Graph 4 Leakage between the positive terminal and the substrate.

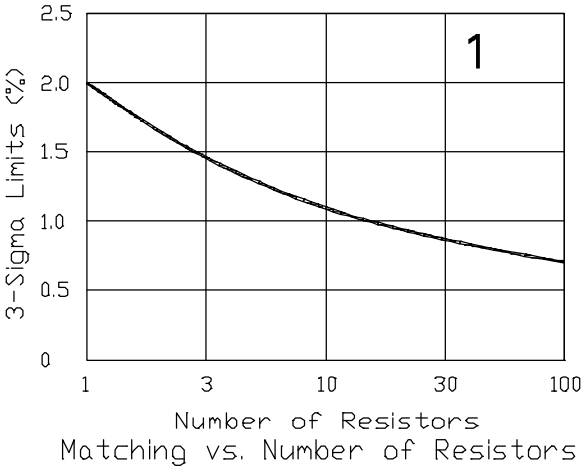
Graph 5 The temperature coefficient of the capacitor varies with the voltage across it.



Resistor

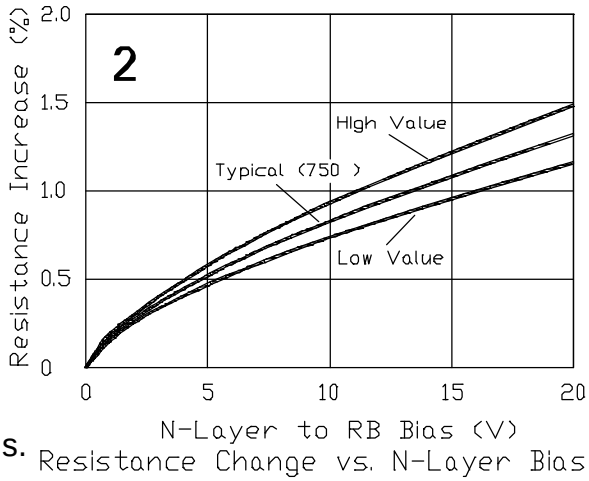
There is just one resistor value in the 700 Series: 750 Ohms. We chose this approach because the matching between resistors with different geometries is far inferior. You can make a large number of different resistor values by connecting these devices in series and parallel.

The absolute-value (3-sigma) variation for the resistor is ±20% (600 to 900 Ohms).



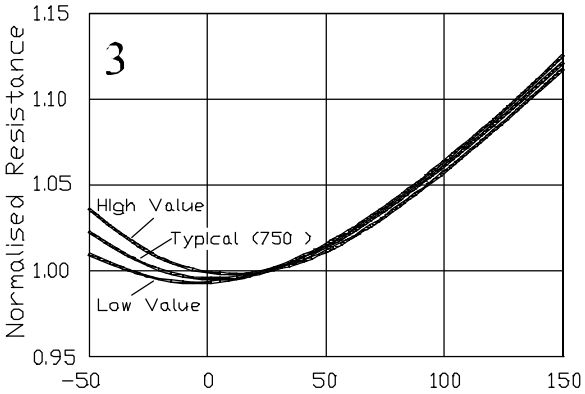
Graph 1 Two resistors match within 2%. The more resistors you use, the bet- ter the matching gets.

Graph 2 All resistors are located in a common island (the epitaxial layer), which is connected to the most positive voltage. The voltage between the resistor and the island reduces in effect the thickness of the resistor (due to the depletion layer), so there is a small change in resistance. If possible, therefore, matching resistors should be located at the same DC potential.



Resistance Change vs. N-Layer Bias

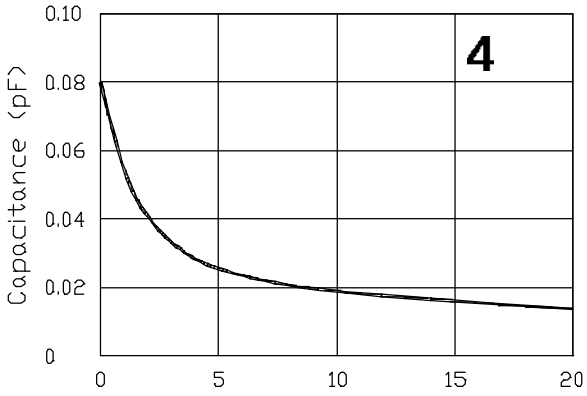
High Value: A resistor running close to the maximum of 900 Ohms. Low Value: A resistor with a value below the nominal, close to 600 Ohms.



Resistance vs. Temperature

Graph 3 The value of a diffused resistor changes with temperature and its temperature coefficient is considerably larger than that of even a carbon resistor.

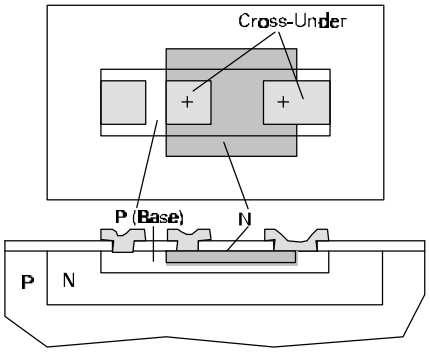
Graph 4 Each resistor has a small capacitance to the island. For a single resistor this results in a -3dB frequency of several GHz. Hardly ever the frequency bottleneck, unless you have a very long string.

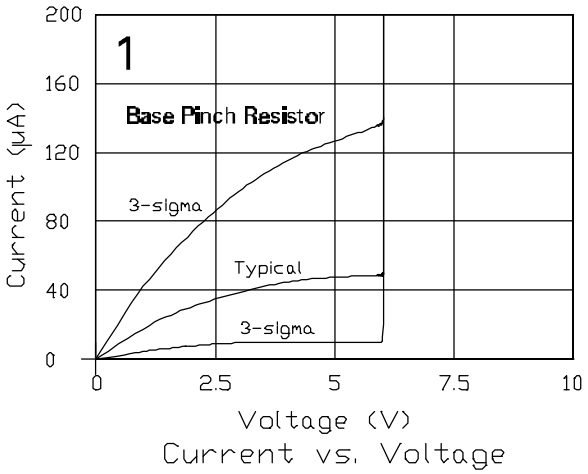


Capacitance vs. N-Layer Bias

Base Pinch Resistor

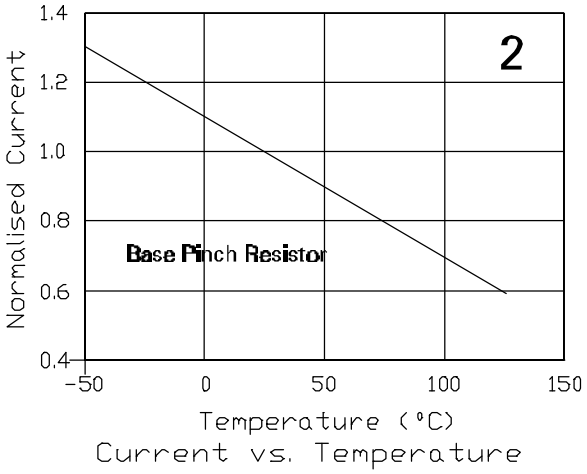
There are a few resistors on each chip in which the emitter layer covers most of the base strip. This, together with the applied voltage, "pinches" off the cross- section, so that the resulting resistance is high and non-linear (it is more of a current source). Applied voltage is limited to 5 Volts.





Graph 1 The variation of a base pinch resistor is large (10 to 150uA). Use it only if you need some current, but the exact amount of current is not very important.

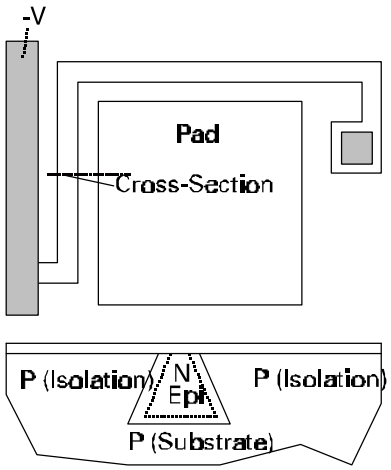
Graph 2 Temperature coefficient of the current. Multiply this curve with graph 1.

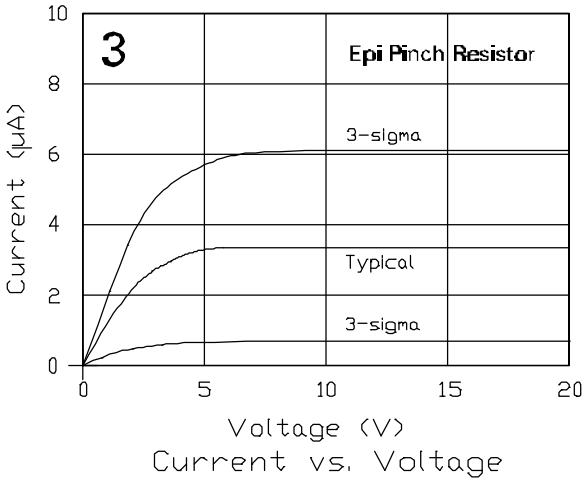


Epi Pinch Resistor

This device, located near a pad, uses the epitaxial layer. But we have made it long and very narrow, so that the resistance is high and the applied voltage (through the depletion layer shown dotted) pinches off the cross-section further. This results in a steady current rather than a resistance.

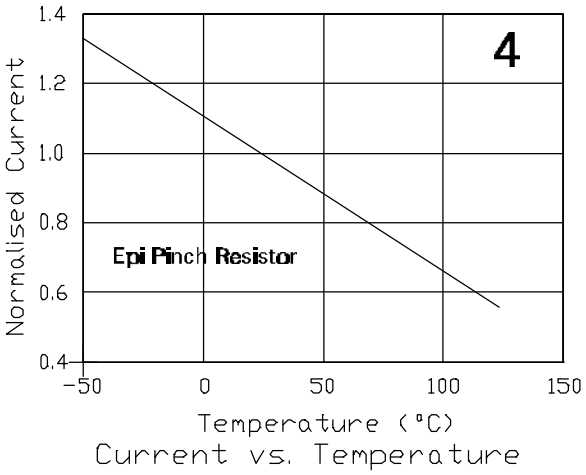
One terminal of the epi pinch resistor is permanently connected to the substrate; the free terminal, therefore, sinks a current to the most negative voltage.





Graph 3 At about 5 Volts the pinch-off is complete. Between 5 and 20 Volts the epi pinch resistor is a current source. Notice the large variation (1 to 8uA). Use this device only if you need some small current, but the level of the current is not important. Ideal for circuit startup (see chapter 5).

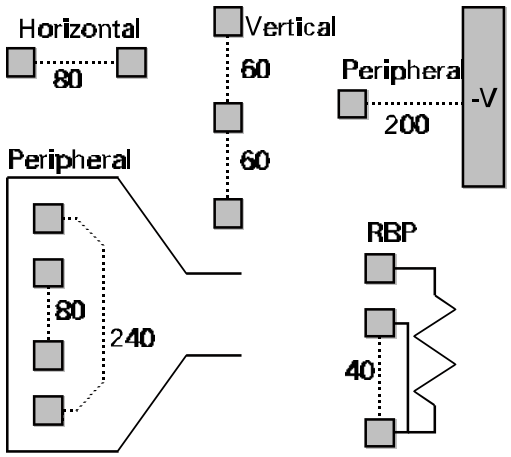
Graph 4 Temperature coefficient of the current. Multiply this curve with graph 3.



Cross-Unders

We have placed cross-unders (using the emitter diffusion) in large numbers around the chips. The typical values (in Ohms) are given here; their variation is ±25%.

Occasionally you can use these cross-unders as low-value resistors, though they do not match or track the base-diffused resistors.



Electrostatic Discharge Protection

Next to most pads there is a small contact box, which leads to a large buried- layer/substrate diode underneath the pad. Although bipolar devices are largely immune to electrostatic discharge, you may increase the amount of protection by covering this box with metal connected to the pad. This adds about 3pF of capacitance to the pad.

This protection diode absorbs negative-going voltage spikes. To reduce positive transients we recommend that you use a wide metal path leading from the pad and let it cross over several resistors and/or cross-unders.

The 700 Series Chips

There are nine chips in this series forming a smooth progression in size; except for the smallest one (which was designed to fit into the SOT-23 package) each chip is approximately 30% larger in area than the next smaller one.

Chip Series	**710**	**711**	**712**	**713**	**723**	**724**	**734**	**736**	**747**
Die Size mils	33x33	44x40	47x53	53x66	70x66	74x79	98x79	98x106	118x119
Die Size square mils	1089	1760	2491	3498	4620	5846	7742	10388	14042
Pads	4	8	17	22	25	30	30	41	48
NPN/PNP Transistors	14	22	27	39	60	80	120	180	280
Schottky NPN Trans.	4	6	10	11	12	16	24	36	56
Large NPN Transistors	1	1	1	2	3	4	9	5	9
Large PNP Transistors	0	1	1	2	3	3	6	4	5
Total Transistors	19(33)	30(52)	39(66)	54(93)	78(138)	103(183)	159(279)	225(406)	350(630)
750 Ohm Resistors	122	190	210	411	623	895	1268	1798	2487
Total Base Resistance	91k	142k	157k	308k	467k	671k	951k	1.34M	1.86M
Base Pinch Resistors	2	2	9	9	7	11	8	14	16
Epi Pinch Resistors	1	1	2	2	2	2	2	2	2
Junction Capacitors	1	2	2	4	7	7	9	12	10
Cross-Unders	40	68	70	160	200	300	450	650	950

The main features of the 700 Series are:

- All devices use an advanced, small-geometry process, which results both in smaller chip size and increased complexity. This is the smallest possible geometry for an operating voltage up to 20 Volts.

- Each of the small transistors can either be NPN or PNP *with no degradation in performance.* In the NPN mode the transistor has three separate emitters and two separate bases, allowing the creation of current ratios and multiple use of a single device.

- Years of experience in the semicustom field have taught us how to design a chip which can be interconnected easily. You will find the 700 Series to be by far the easiest to route.

- All critical components, such as the resistors and small transistors have not only identical size but also identical orientation, giving the best possible matching.

Each chip contains the same basic components and is based on an identical architecture. In the center portion of the chips are islands of 12 transistors each, 10 of which are convertible from NPN into PNP and 2 are Schottky NPN transistors. These islands are surrounded by a field of resistors. In this series the number of resistors and the total resistance is extra large, which makes it easier to design an IC, especially for first-time designers.

These islands of transistors are arranged in columns and rows. The *number* of the chip tells you how many columns and rows there are. The 712 chip, for example, contains a single column and two rows.

Between the bonding pads along the periphery are all other devices: large (200mA) NPN transistors, large (6mA) PNP transistors, pinch (high-value) resistors and junction capacitors. Sprinkled throughout each chip are low-value cross-under resistors.

Outside the bonding pads there is space for three interconnection lines and a wide metal stripe for the most negative potential. The positive potential can be bussed through the chip in special lanes. These two potentials automatically take care of the proper junction isolation of all the components.

Selecting Chip Size

To determine the size of the chip you need consider three factors:

1. How many bonding pads do you need?

2. What is the requirement for special devices? Looming largest here is the high-current NPN transistor. Occasionally you may have to consider the higher-current PNP transistor, Schottky-clamped NPNs or capacitors.

3. What is the small-transistor count?

The last factor is the hardest to estimate before doing some actual design. However, you should be aware that *chip size is not the only ingredient in the cost of an IC.* There are three other factors which are equally important:

A. The yield of a wafer (how many chips actually work) depends both on the chip size and how well the circuit is designed.

B. To the chip cost the packaging cost has to be added. This means that a 30% increase in chip area does *not* result in a 30% increase in the cost of the assembled IC; the increase in cost is always smaller.

C. In addition to the chip and packaging cost there is a cost associated with testing. Both a wafer test and a package test is required.

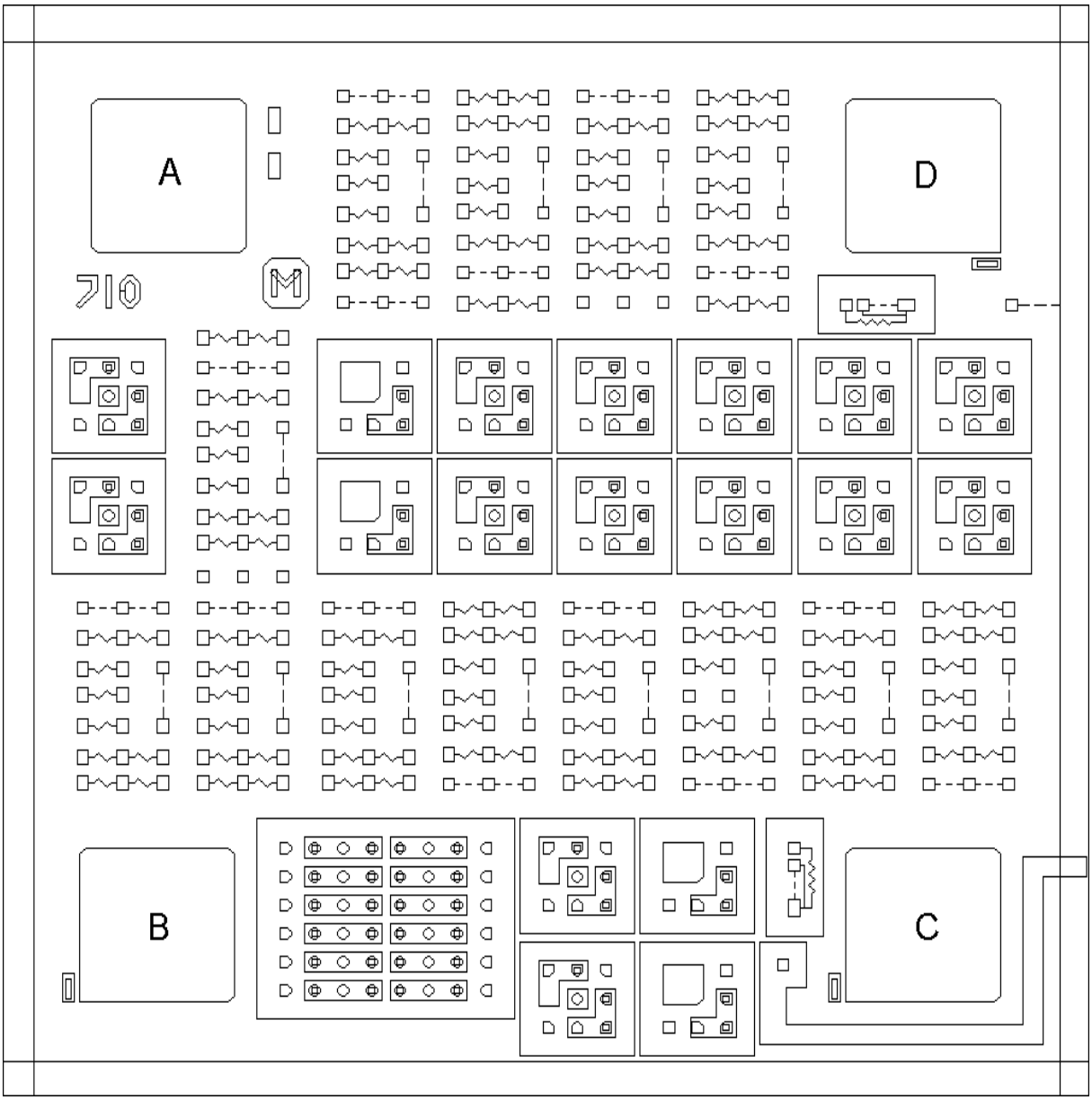
How much of a chip can you actually use? Certainly more than 70%. If you use less than 70% consider going to the next smaller chip. Because of the advanced architecture a utilization factor of 85% is not too difficult, even with a single metal layer. If you are willing to take on a bit of challenge it is quite possible to use all - 100% - of the *transistors* and you are still likely to be left with some of the resistors.

Die Sizes / Potential (Unyielded) Die per 150mm (6") Wafer

Die	Size (mils)	Die per Wafer
710	33x33	23373
711	44x40	13695
712	47x53	9027
713	53x66	6363
723	70x66	4745
724	74x79	3703
734	98x79	2763
736	98x105	2061
747	118x119	1480

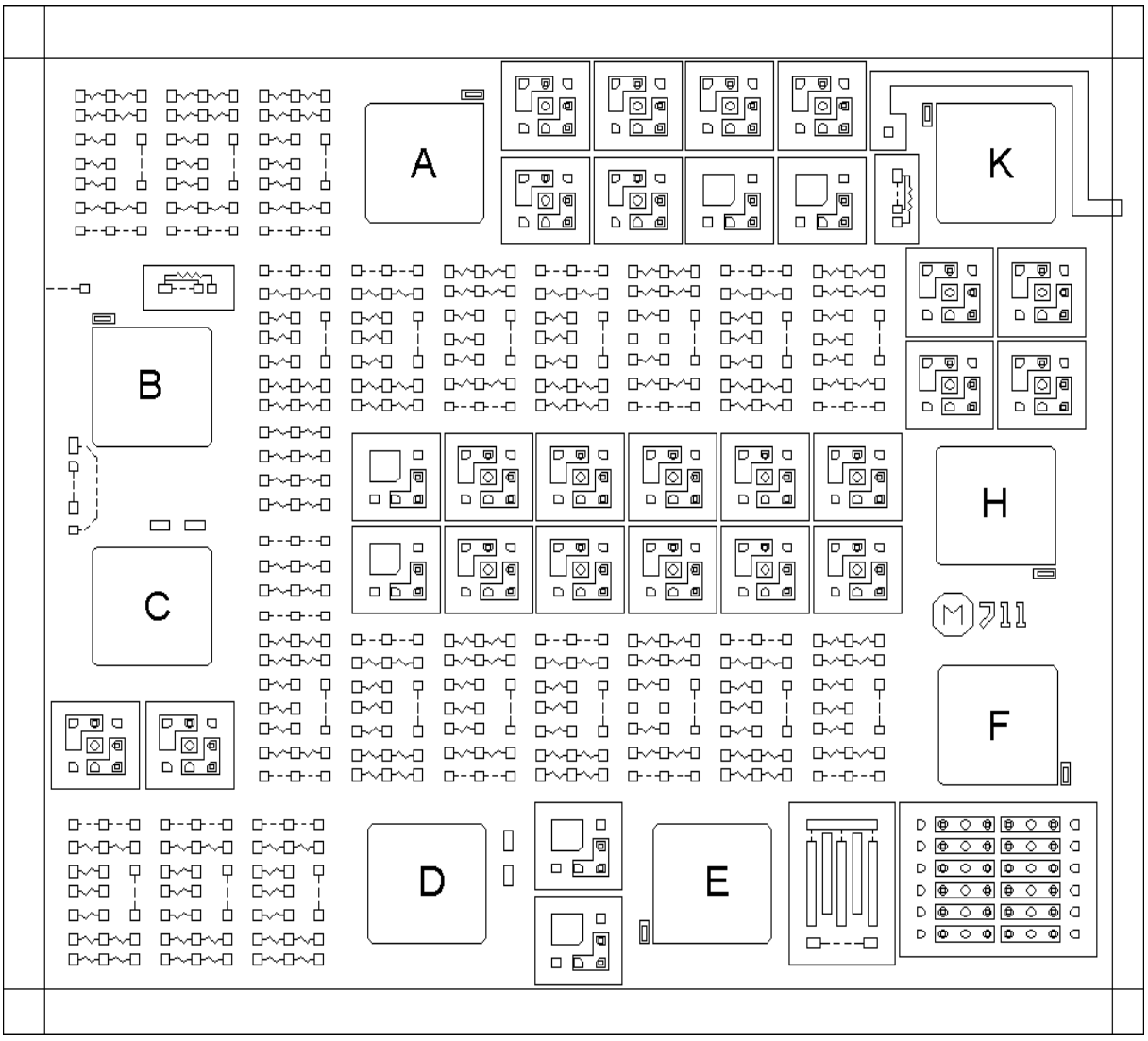
710 33x33 mils (0.84x0.84mm)

4 pads - 14 small NPN/PNP transistors - 4 small Schottky NPN transistors - 1 large NPN transistor - 91kOhms base resistance - 2 base pinch resistors - 1 epi pinch resistors - 1 junction-capacitor - 40 cross-unders - Fits smallest surface mounted packages (SOT)



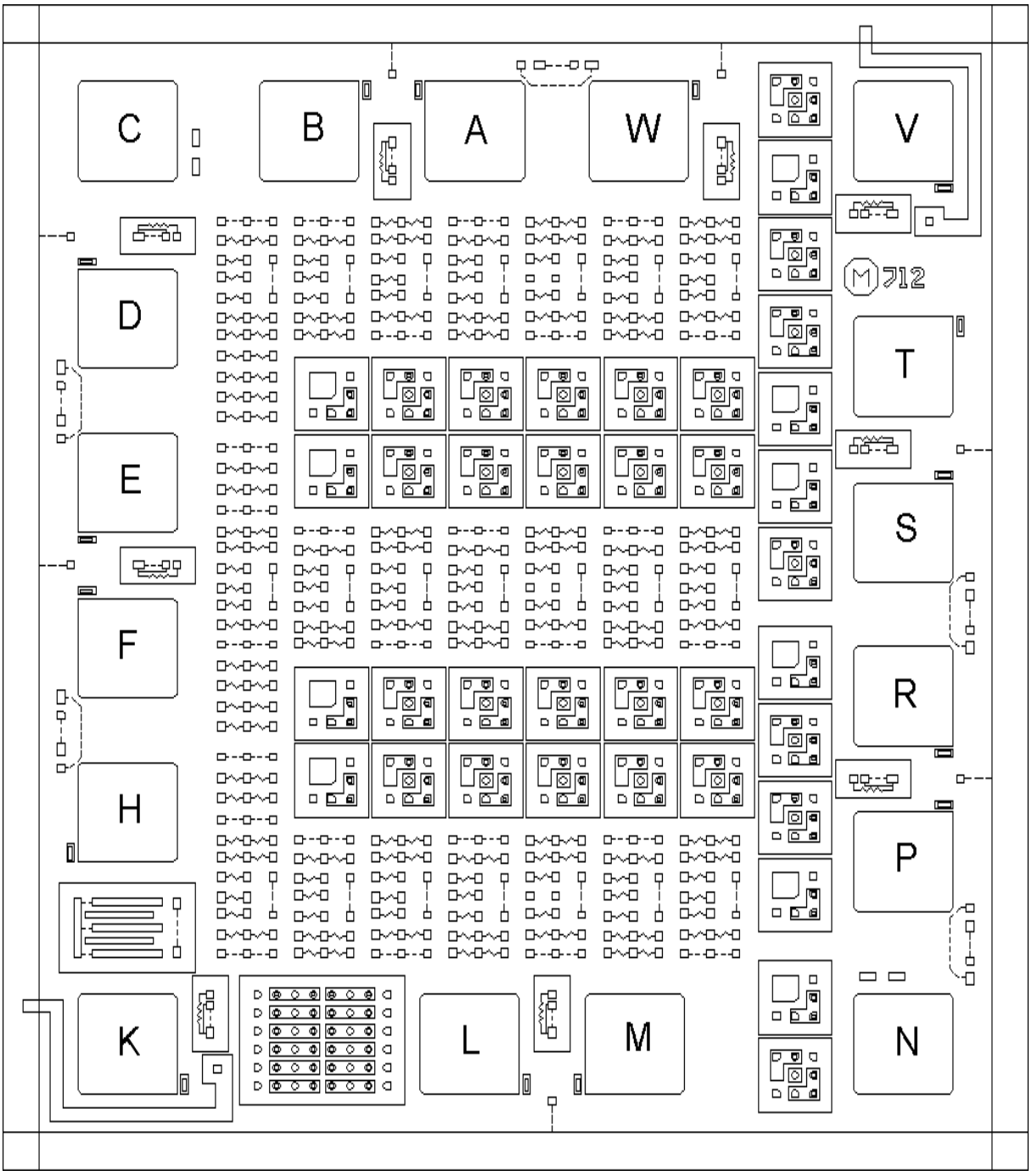
711 44x40 mils (1.1x1mm)

8 pads - 22 small NPN/PNP transistors - 6 small Schottky NPN transistors - 1 large NPN transistor - 1 large PNP transistor - 142kOhms base resistance - 2 base pinch resistors 1 epi pinch resistor - 2 junction-capacitors - 68 cross-unders



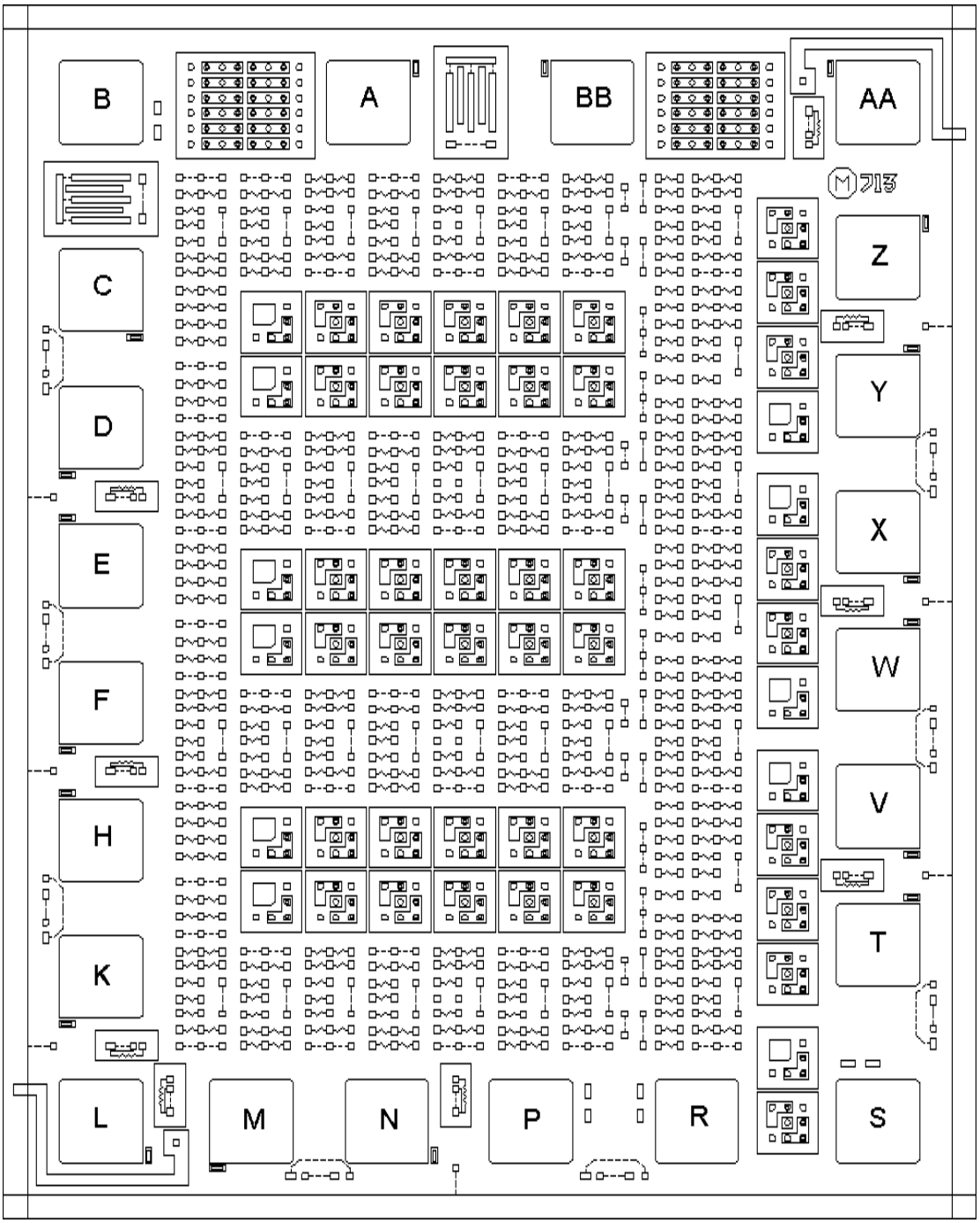
712 47x53 mils (1.2x1.35mm)

17 pads - 27 small NPN/PNP transistors - 10 small Schottky NPN transistors - 1 large NPN transistor - 1 large PNP transistor - 157kOhms base resistance - 9 base pinch resistors
2 epi pinch resistors - 2 junction capacitors - 70 cross-unders



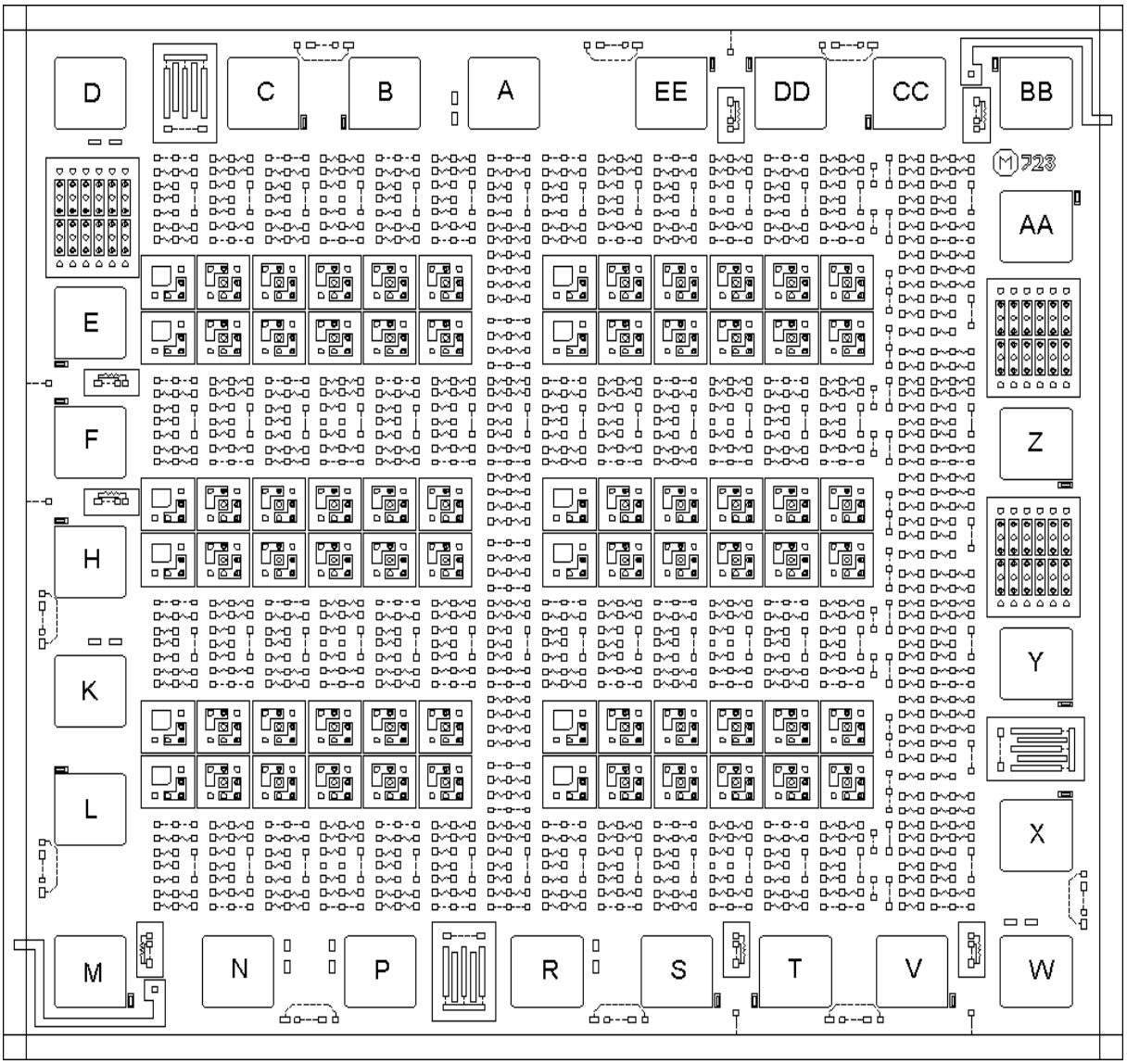
713 53x66 mils (1.35x1.7mm)

22 pads - 39 small NPN/PNP transistors - 11 small Schottky NPN transistors - 2 large NPN transistors - 2 large PNP transistors - 308kOhms base resistance - 9 base pinch resistors
2 epi pinch resistors - 4 junction capacitors - 160 cross-unders



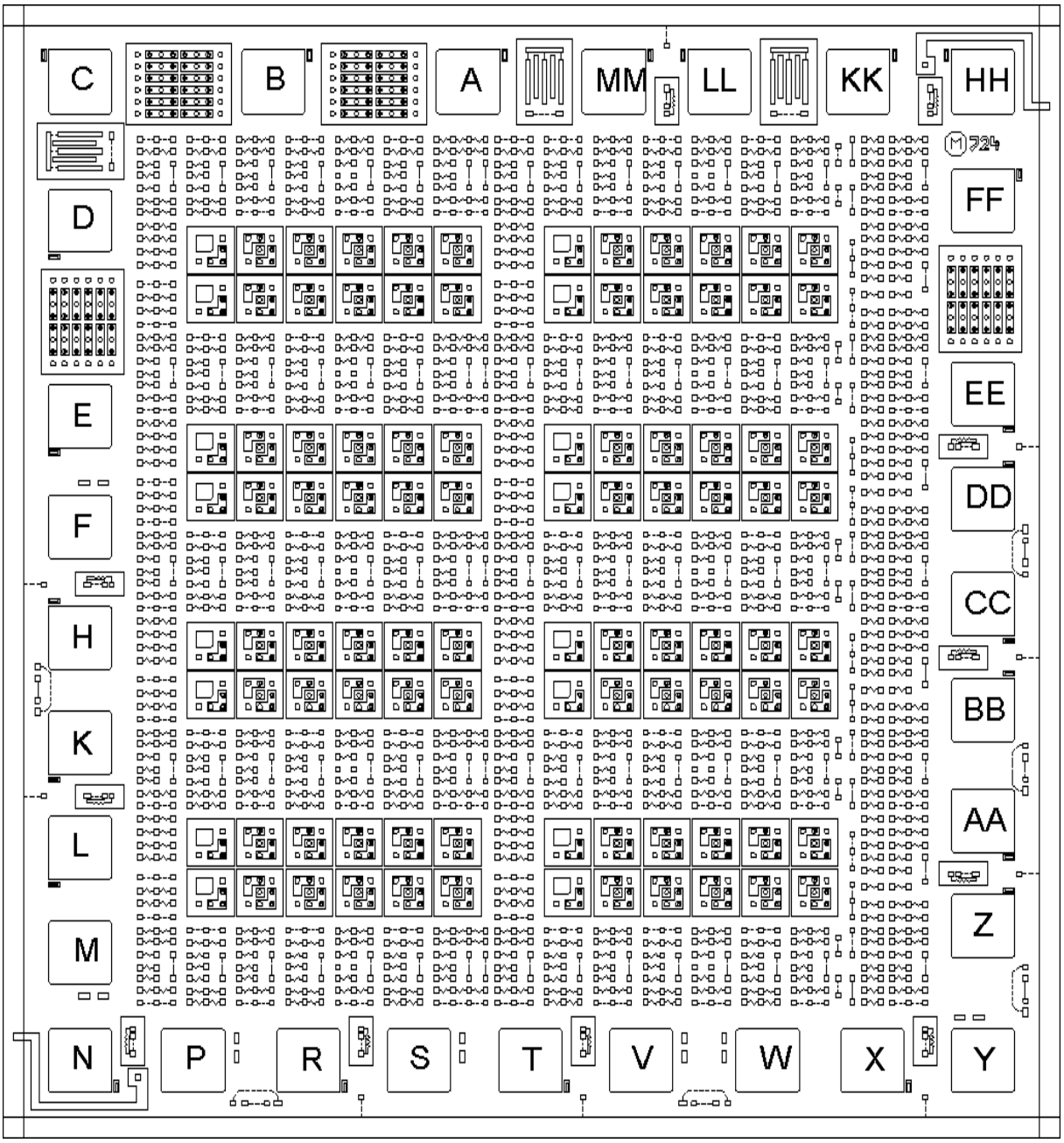
723 70x66 mils (1.8x1.7mm)

25 pads - 60 small NPN/PNP transistors - 12 small Schottky NPN transistors - 3 large NPN transistors - 3 large PNP transistors - 467kOhms base resistance - 7 base pinch resistors
2 epi pinch resistors - 7 junction capacitors - 200 cross-unders



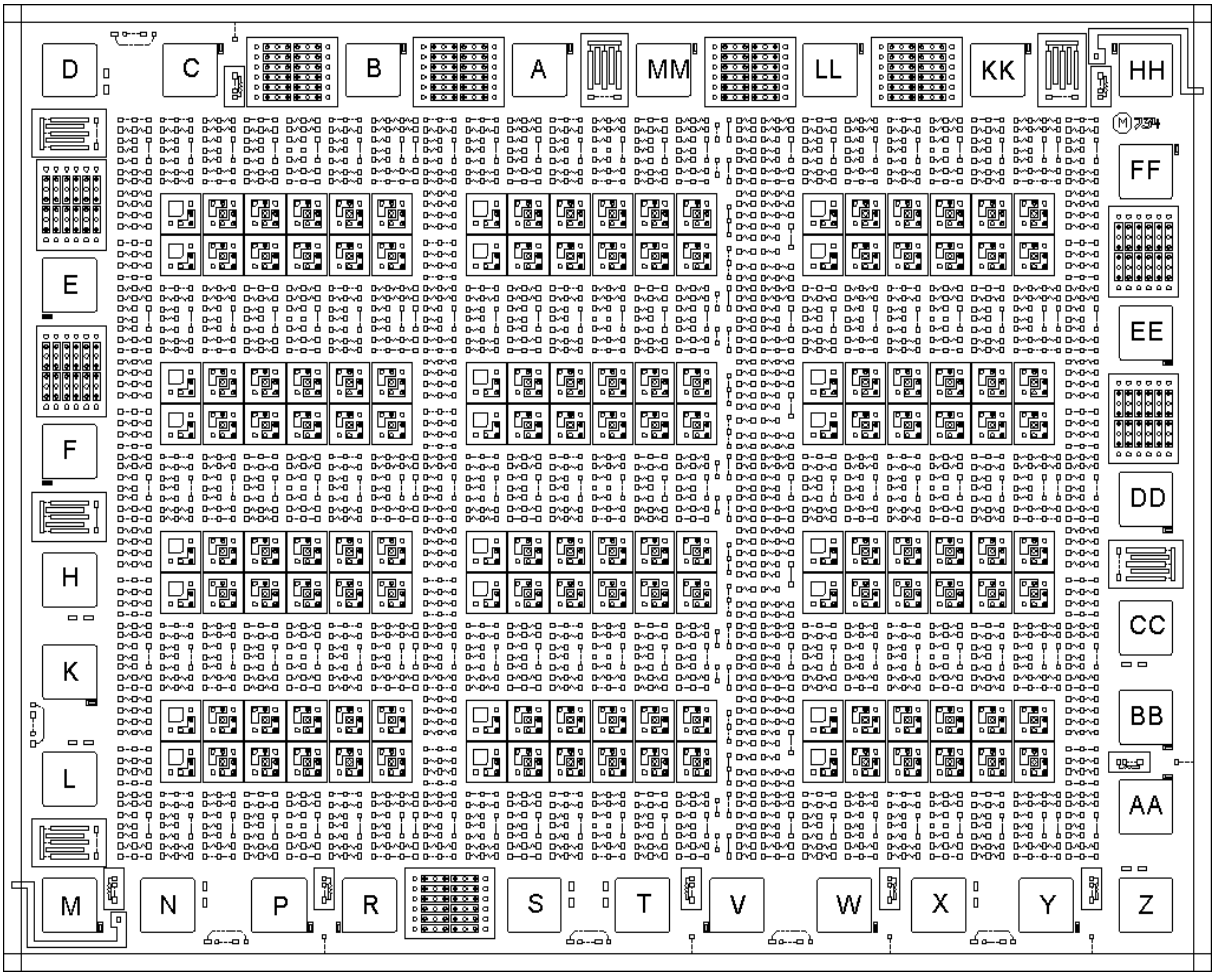
724 74x79 mils (1.9x2mm)

30 pads - 80 small NPN/PNP transistors - 16 small Schottky NPN transistors - 4 large NPN transistors - 3 large PNP transistors - 671kOhms base resistance - 11 base pinch resistors
2 epi pinch resistors - 7 junction capacitors - 300 cross-unders



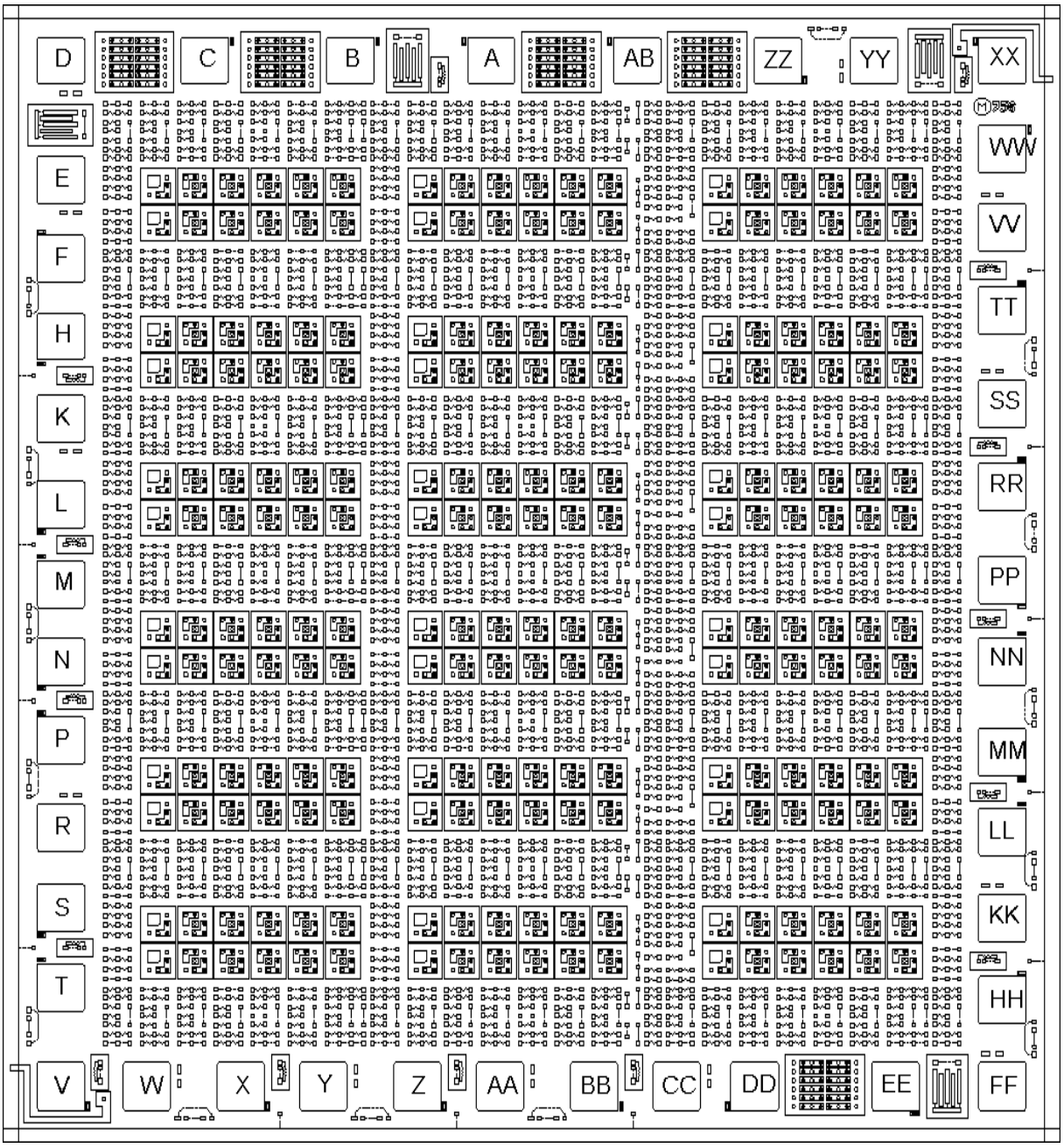
734 98x79 mils (2.5x2mm)

30 pads - 120 small NPN/PNP transistors - 24 small Schottky NPN transistors - 9 large NPN transistors - 6 large PNP transistors - 951kOhms base resistance - 8 base pinch resistors
2 epi pinch resistors - 9 junction capacitors - 450 cross-unders



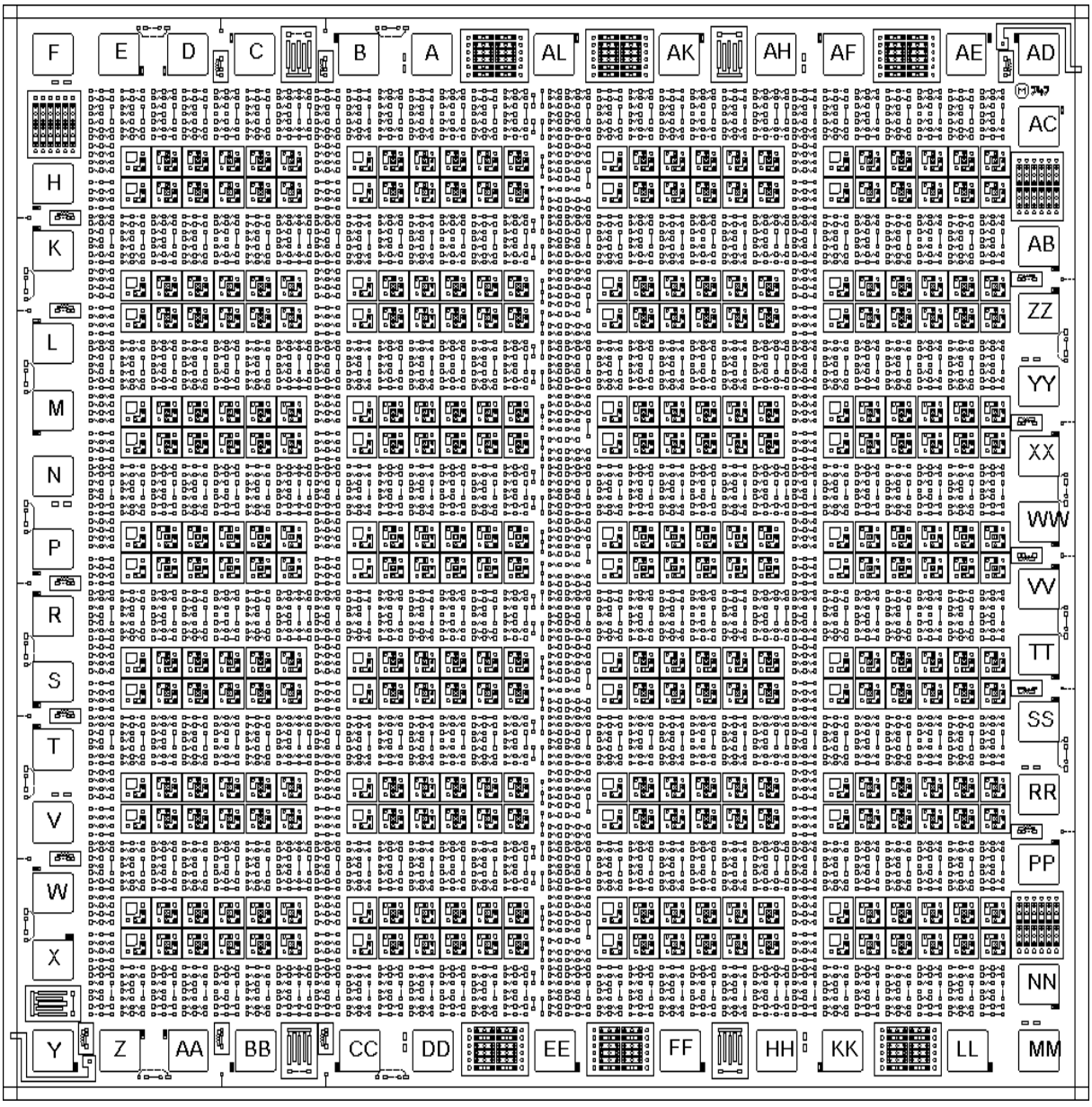
736 98x105 mils (2.5x2.7mm)

41 pads - 180 small NPN/PNP transistors - 36 small Schottky NPN transistors - 5 large NPN transistors - 4 large PNP transistors - 1.34MOhms base resistance -14 base pinch resistors
2 epi pinch resistors - 12 junction capacitors - 650 cross-unders



747 118x119 mils (3x3mm)

48 pads - 280 small NPN/PNP transistors - 56 small Schottky NPN transistors - 9 large NPN transistors - 5 large PNP transistors - 1.86MOhms base resistance - 16 base pinch resistors 2 epi pinch resistors - 10 junction capacitors - 950 cross-unders



Linear IC Design

Designing linear (bipolar) integrated circuits is first of all different. Different from digital design and different from designing with discrete components and standard ICs.

There are some principles - we counted six of them. We will go through them first, slowly and carefully.

Using these design principles a lot of clever linear IC designers have come up with bits and pieces of circuits which form useful functions. If you take a close look at the schematics of some of the standard linear ICs you will recognize many of them. Even the most expert of IC designers use them. It is because of these circuit *elements* that linear IC design is manageable. You select the elements best suited for your design and connect them together so that they perform the function you desire.

That is why we have put together a large library of these linear IC design elements in the second section of this chapter, with detailed explanations and comments.

In the third section of this chapter we have collected some design examples, which we call *functions*. We intend to add to this section gradually ourselves, but most of the additions will come from the users of the 700 Series. So, if you have created a useful function or circuit, send it to us. We'll add it to the collection, with credit, of course.

All of the circuits in this chapter are contained in a zipped file. You can pull them directly into your design.

The following two books we consider the best for linear IC design:

Gray and Meyer: "Analysis and Design of Analog Integrated Circuits", John Wiley & Sons, 1984, ISBN 0-471-87493-0
Grebene: "Bipolar and MOS Analog Integrated Circuit Design", John Wiley & Sons, 1984, ISBN 0-471-08529-4

Part of 700 Series Manual

Design Principles

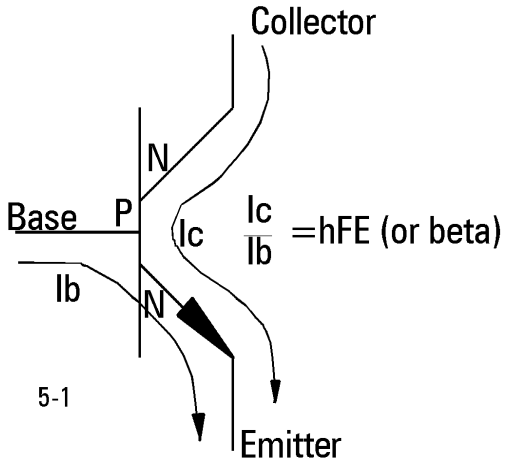
Examining the Bipolar Transistor

How does a transistor really work? Well, there are three layers: in the case of the NPN transistor the outer ones are p-type; sandwiched in between is a thin n-type region.

P-type means that, in the silicon crystal lattice, there are also a few atoms which have one less electron in the outermost shell than silicon (which has 4). Thus Boron, Aluminum or Gallium, which have three electrons in the last orbit, fill the bill. This makes for a *deficiency* of electrons. Since electrons have a negative charge, the resulting charge is positive, hence p-type. Likewise, the presence of atoms with 5 electrons in the outermost shell (Phosphorus, Arsenic or Antimony) causes an oversupply of electrons, i.e. the silicon becomes n-type.

These "dopant" atoms can be inserted into the silicon by diffusion (at very high temperature, which causes the dopants to penetrate even solid materials, or by ion implantation, i.e. "shooting" the atoms into the crystal lattice).

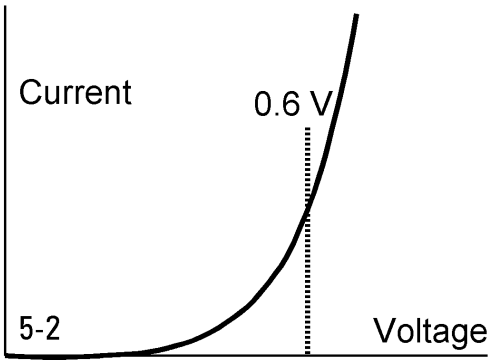
Current can flow from p to n, but only if a certain voltage (about 0.6 Volts for silicon) is exceeded (this is the conventional current flow; the electrons actually flow the other way). There is no (or very little) current-flow from n to p, except for one very important case:



If we apply a positive voltage to the base (p) and a negative one to the emitter (n) current will flow from base to emitter once about 0.6 Volts is exceeded. The electrons now flow through the thin base region. With a positive voltage applied to the collector, some of these (negatively charged) electrons are attracted by the positive collector voltage and flow toward it, rather than out the base. This flow takes place despite the fact that the collector-base junction is reverse biased; it does so solely because the base region is thin, which forces some of the electrons too close to the collector region. The thinner we make the base, the more of the electrons end up at the collector terminal rather than the base. With the base thickness used in an average transistor (about 0.2 microns) more than 100 times as many electrons flow to the collector as do to the base. The ratio of the two we call the current gain:

hFE (or beta) = Ic/Ib

This is a (fairly) linear relationship. If we increase the base current by a factor of 10 we can expect the collector current to increase by a factor of 10 also. On the other hand, look at the base-emitter voltage. This is a forward-biased diode. Hardly any current flows below about 0.3 Volts. At about 0.6 Volts there is a substantial amount of current and at higher voltages the current increases drastically. In fact, this is an exponential relationship: at any point on the curve, an increase of about 60mV (at room temperature) causes the current to increase tenfold. If we were to plot this curve with the current on a logarithmic scale, it would be a straight line.



The formula for a diode is:

$V = kT/q \ln (I1/Is)$

where
V is the diode voltage in Volts
k is the Boltzman constant (1.38E-23 Joules/Kelvin)
q is the electron charge (1.6E-19 Coulombs)
T is the (absolute) temperature in Kelvin
ln is the natural logarithm
Is is the diffusion current, which depends on the doping concentrations of the two layers, and the diode area
and I1 is the diode current

Note: 1.38E-23 is the notation for 1.38×10^{-23}. Since it is required for Spice, we will be using this notation throughout.

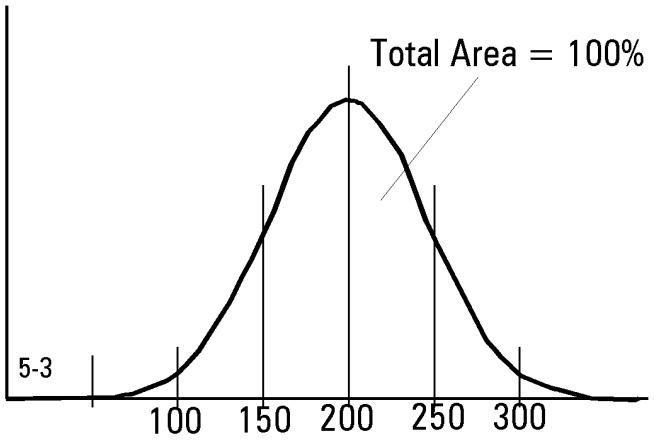
At first glance one might think that the diode voltage has a positive temperature coefficient because of the presence of T. However, "Is" has a much stronger *negative* temperature coefficient, so that the overall change is about -2mV/^{o}C

We are forced to conclude from all of this that the bipolar transistor is a good current amplifier but, as a voltage amplifier, it is very non-linear.

Let's look at another aspect of the current gain, though. As we noted above, to obtain a current gain of 100 or more, the base region must be very thin. How is this base region created? Well, first the (p-type) base region is diffused in. Then the (n-type) emitter follows and is made a little shallower. The difference between the two creates the actual base region (see chapter 2). Now, all this diffusing happens at a very high temperature, above 900^{o}C. It needs to be this high to get the dopants to move into the silicon within a reasonable time.

At this high temperature the red-hot silicon is no longer a semiconductor. There are no diodes and there is no transistor action. So we need to let the wafer cool down before be can tell how accurate the emitter diffusion was, how much base-width we have created and what hFE has resulted.

The upshot of all of this is that we cannot expect to get the same hFE every time. Sometimes we diffuse a little too deep, sometimes a little too shallow. Summing it up in one sentence: we get a *distribution*. Shown here is the Gaussian distribution, where the total number of measurements is represented by the area under the curve. Most wafers will show an hFE around 200, but some will be 100, some 300. Since individual transistors on a wafer are not accessible (and we couldn't do much to change the hFE anyway), we are forced to design a circuit so that is works well with an hFE ranging from 100 to 300. In fact, as the curve shows, even this does not include *all* transistors; a small (hopefully very small) percentage will have an hFE beyond the limits.

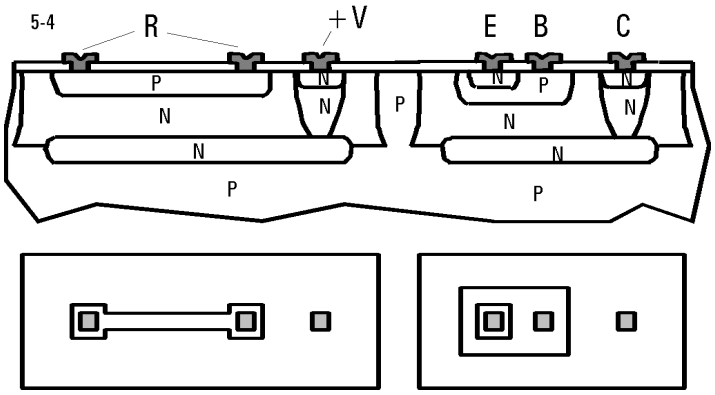


Not all distributions are Gaussian. The current gain distribution, for example, is slightly asymmetrical. But it is usually accurate enough to assume that we are dealing with a Gaussian distribution only.

For the Gaussian distribution it is handy to use sigma deviations. A deviation (from the average, or peak of the curve) of one sigma (in both directions) includes 68% of all measured values. Twice that deviation (2 sigma) includes 95% and 3-sigma 99.7%. Thus, if we give a 3-sigma range of hFE (for the NPN transistor) of 100 to 300, 0.3% of the transistors may have a current gain outside this range. If you cannot live with a larger range you will need to eliminate the 0.3% through testing. More on statistical distributions in chapters 3 and 10.

This is a fact of life in ICs and applies to all devices and almost all parameters. Consider, for example the resistor. There are three semiconductor layers in the 700 series which could be used as resistors: the epitaxial layer, the base diffusion and the emitter diffusion. Their resistance depends on how heavily the region is doped - the more dopant atoms (either p or n) the lower the resistance. The base diffusion gives the most practical range of resistance and is, therefore, almost universally used.

We can delineate (with a mask) a strip of base diffusion, place contacts at either end and we have a resistor. If we make the strip long and narrow, the resistance will be high; short and wide gives a low resistance value. This part is simple; what's not so simple is that the base diffusion wasn't really made to be a resistor. First of all, it is a semiconductor material and as such doesn't have the best properties for a resistor; its temperature coefficient, for example, is considerably higher than that of a metal-film resistor. Second, the NPN transistor has first call on the properties of the base layer. Third, as with the hFE, we cannot measure the resistance while it is being made. At the high diffusion temperature the base layer no longer even resembles a semiconductor.

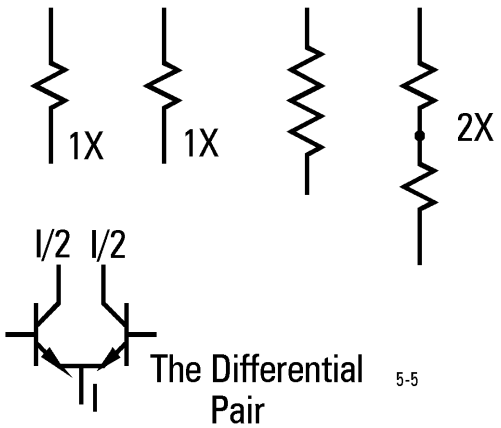


So, again, there are variations. About ±25% (3-sigma), to be specific, much more than even the least expensive carbon resistor. And, of course, unlike the carbon resistor, we cannot test each resistor and put it in separate bins according to value. Nor can we trim it like a metal-film resistor.

But there is a hidden advantage, which has become one of the most powerful design tool in linear IC design. While the resistors (and hFEs and most other parameters) may vary a great deal from wafer to wafer, *they are likely to be the same on a wafer*. We may miss the target but, if we do, it will affect all devices on a wafer by (very nearly) the same amount. This leads us to the first design principle:

Principle #1: Use Matching

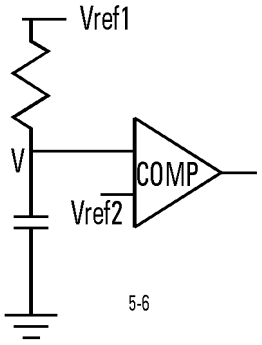
While the absolute variation of almost all parameters in an IC is large, their relative values tend to be close together because they have undergone the same treatment and are in close proximity.



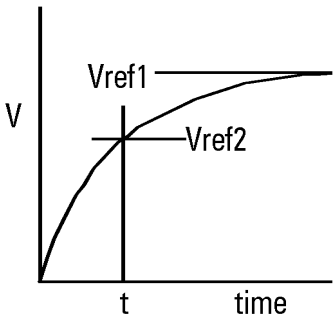
For example, (base diffused) resistors have a (3-sigma) variation of ±25%, but their *ratios* vary by only 2% (3-sigma), even less if you use many of them together. Thus, you can count on two equal values to be very nearly the same, not matter what they may be. Furthermore, you can create quite accurate *resistor ratios other than 1*. In a full-custom IC this is normally done by giving two resistors different lengths; in the 700 Series you do it by connecting identical resistors in series or parallel (which results in the best possible matching accuracy).

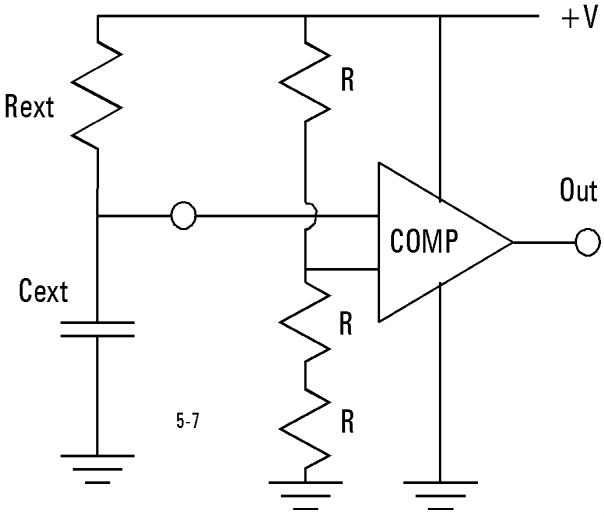
But matching goes far beyond the resistor. All transistor parameters (most importantly their hFEs and VBEs) match well. This property, as you will see soon, leads to the ubiquitous *differential pair*.

To illustrate this principle, we are going to analyze an example: the 555 timer. Before its appearance, a timing function was generally designed as shown on the right. A capacitor is charged through a resistor from a (regulated) reference voltage Vref1. The voltage across the capacitor is sensed by a comparator. When this voltage reaches the level of a second (regulated) reference voltage Vref2 (somewhat lower that Vref1), the comparator switches. It's not a very elaborate circuit, except that we need two accurate reference voltages; both of them influence the timing.



5-6

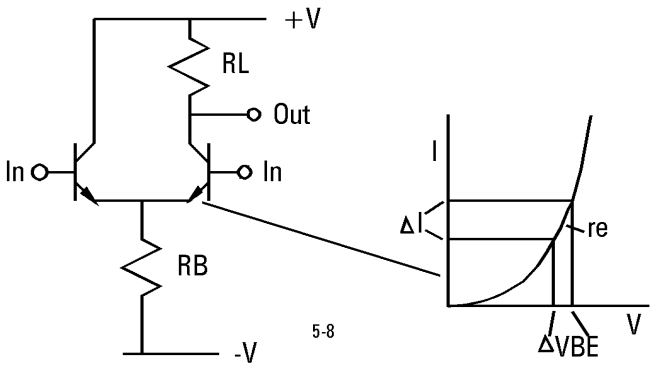




5-7

Enter the principle of matching. We throw away both reference voltages and use a string of matched resistors to create Vref2. At first glance you might think that we have a much less accurate circuit now. But observe that, with three equal resistors, the trigger point is 2/3 of the supply voltage; the only tolerance here is determined by the matching accuracy of the three resistors. Now increase the supply voltage: the trigger point increases proportionally, *but so does the charging current*. The two effect cancel precisely. If the input current of the comparator is small and the matching accurate, the time is only a function of the external capacitor and resistor. We have eliminated the need for absolute accuracy within the IC.

Let's extend this principle to the transistor. An example that is easy to grasp is the comparator. When one input moves past the other, its output switches. For a well-designed comparator, this happens within a few millivolts. With single transistors this would be very difficult to achieve; with pairs it is a simple task.



Our comparator example here is still very primitive, using only the first of the linear design principles. A pair of NPN transistors share the same resistor, RB. With the two inputs at the same voltage, the current flowing through RB splits evenly between the two transistors. Thus one-half the total current flows through RL, creating a voltage drop at the output. Notice that, so far, there was no need to talk about transistor parameters. No matter what they are, the current will split into two equal parts, as long as the transistors match.

Now increase the right-hand input. Through the base-emitter diode you force its emitter to a higher voltage. Therefore, it takes a larger share of RB's current and the output voltage drops. If you decrease the right-hand input, the opposite happens.

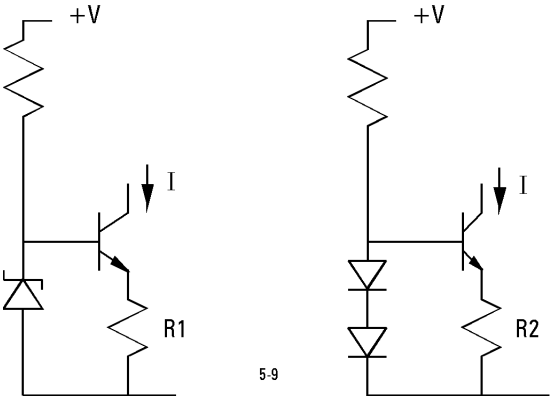
Because the base-emitter diode has an exponential voltage-current characteristic, this switching of currents from one side to the other happens gradually, over a range of a few hundred millivolts. So, what we have here is a rather poor comparator, one with a soft threshold. In other words: the comparator has very little gain.

Principle #2: Replace Resistors with Current Sources

There is also something else wrong with this design: As we increase the higher of the two inputs, the voltage drop across RB increases by nearly the same amount (the input voltage minus the diode drop). Thus the total current is set by the higher of the two input voltages and varies a great deal.

Let's tackle this second problem first. Why is there a resistor used to deliver the current to the emitters of the differential pair? Part of it is tradition. A resistor is the least expensive discrete component. For an IC, it might well be the most expensive component, especially if is needs to be a large value and thus take up a lot of IC area. Also, it is the *wrong* component.

What we really want is a *current source*, a device which delivers a current independent of voltage (Actually, the proper term here is current *sink*, but it has become common practice to call all these elements current sources, regardless of their actual position.) There are no discrete components which are, by themselves, current sources, but in an IC we can make one, usually taking up less area than an equivalent resistor.



5-9

There are two designs shown here (more - many more - later in this chapter). The one on the left uses a Zener diode to set up a voltage at the base of a transistor. The current through the transistor is simply given by the Zener voltage (say 6 Volts), minus the base-emitter voltage, divided by R1. Now, as long as the collector voltage is a bit higher than that of the emitter, the transistor current is fixed. Disadvantage: you need some extra voltage room.

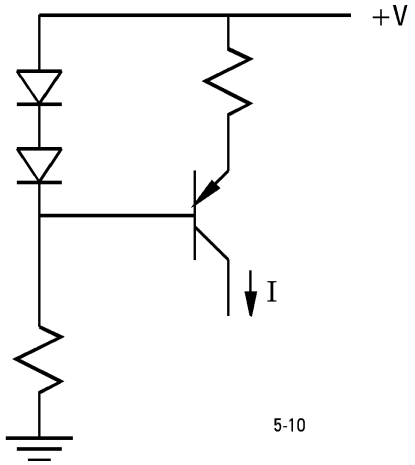
The current source on the right has a slightly poorer performance, but it doesn't waste much supply voltage. Here the base voltage is set by two diodes (about 1.3 Volts). From base to emitter we drop one diode voltage, so the voltage across R2 is one diode voltage. Again, the current through the transistor is this voltage divided by R2 and the collector voltage can drop to about one diode drop above the minus supply. Since diodes (in the forward direction) are exponential devices (i.e. not nearly as sharp as a Zener diode), the current is a bit dependent on the supply voltage.

The previous two examples are current sinks. You can make an equivalent current source with a PNP transistor.

If, for example, we wanted to create a current of 50 microamperes, the required resistor value would be approx:

R = Diode Voltage(VBE) / I = 650mV / 50uA =13kOhms

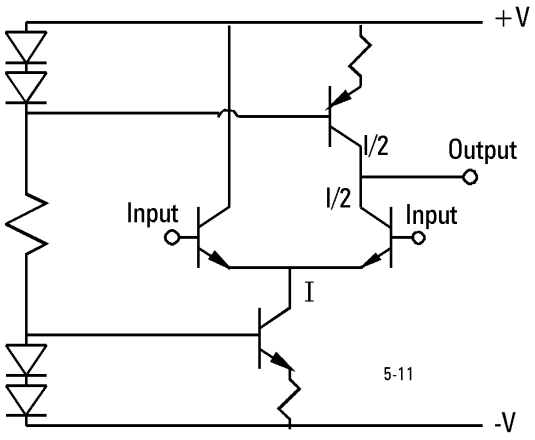
The equivalent discrete resistor with a 10 Volt supply would be 200kOhm, which would require an area many times that of this current source.



5-10

Lets go back to our comparator. In the next figure we replace both resistors with current sources.

The one at the bottom provides the operating current for the differential pair, which now changes very little as the inputs are moved up and down. The one at the top replaces the load resistor; since its impedance is very much higher than that of a resistor, we now get a large amount of gain.



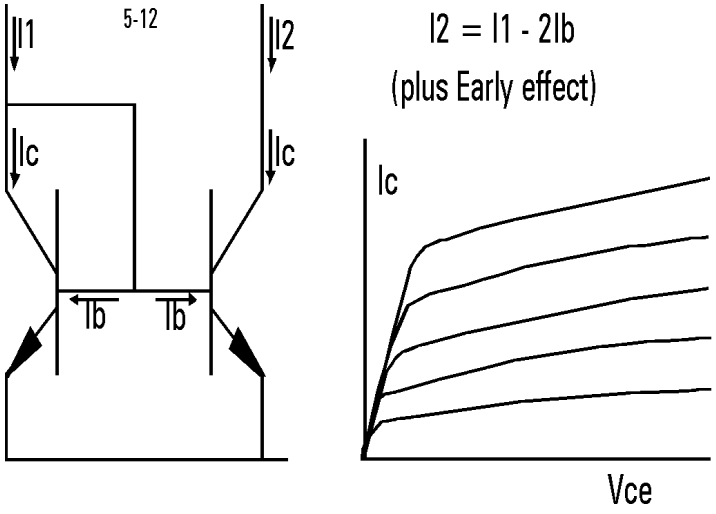
The impedance of an ideal current source is infinite. In most practical current sources the current always increases somewhat with increasing voltage. But the resulting impedance (i.e. slope) ranges from hundreds of kilo-Ohms to a hundred Meg-Ohms, an increase of 100 to 10000 compared to a simple resistor.

The voltage gain of a transistor stage is roughly given by the load impedance divided by the impedance in the emitter. The latter is the slope of the base-emitter diode, roughly 26 Ohms/mA at room temperature (more on this under differential stages). If a transistor is operated at say 100uA, the emitter resistance is 260 Ohms. With a load impedance of 1MOhm, the resultant voltage gain is about 3800. Using a resistor as load (say 40kOhms, allowing a 2 Volt DC drop), the voltage gain is only 154. Besides that, three transistors and one low-value resistor take up much less area than one 40kOhm resistor.

But there is still one problem with our comparator. To work as designed, the load current must be exactly one-half of the operating current. We can come close, but there is a better way:

Study the picture on the right carefully, you are going to see it (and numerous derivatives of it) many times in linear IC design.

First note that the transistor on the left has its base connected to the collector, i.e. it is connected as a diode. Its collector voltage is controlled by the base, which has the already mentioned exponential (and very accurate) voltage-current characteristic. There is still transistor action though - the collector voltage is well above saturation - and Ic is hFE times larger than Ib.

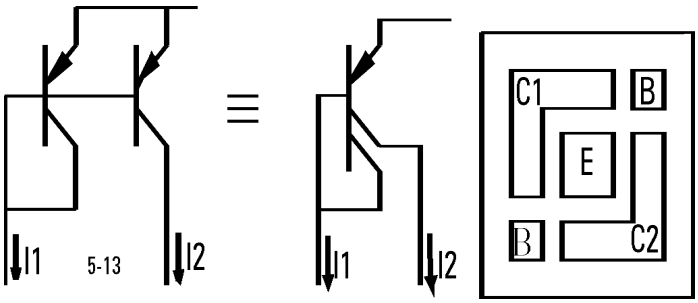


Now the transistor on the right. Its base and emitter are connected to the same terminals as that of the first transistor. Therefore, it must draw the same base current. Since the two devices match, they also must have the same hFE. Therefore, the two collector currents are the same and I1=I2. We have *mirrored* I1.

Well, not quite. There are two errors which make this current mirror a bit less than ideal. The first one: both base currents are supplied by I1. Thus the collector current of the left-hand transistor is I1-2Ib, and it is this current that is mirrored. With an hFE of 100, I2 is 2% smaller than I1.

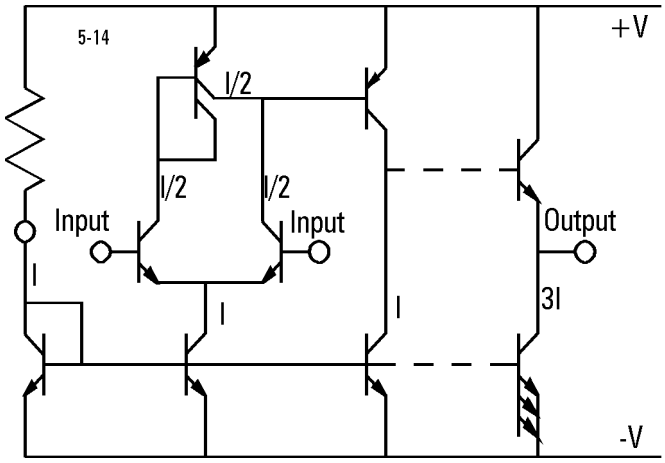
The second error is more serious. As the curves in chapter 3 show, hFE is voltage dependent. This "Early effect" arises from the fact that each junction has a region where the electrons and holes are separated by the voltage (the "space charge region"). As the voltage is increased this region widens and is pushed further into the thin base region. This decreases the base-width, thus increasing the gain.

The upshot of this long-winded explanation is that, the higher the voltage of I2, the more it will increase over I1. This error can amount to as much as +20%. Much more on this under current mirrors later in this chapter.



Exactly the same configuration is possible with a PNP transistor. Here we have a further advantage. As shown in chapter 3, in a (lateral) PNP transistor the collector can easily be split into segments. In the 700 Series the PNP transistors come with two identical collector halves. Thus a single device can be connected as a complete 1:1 current mirror.

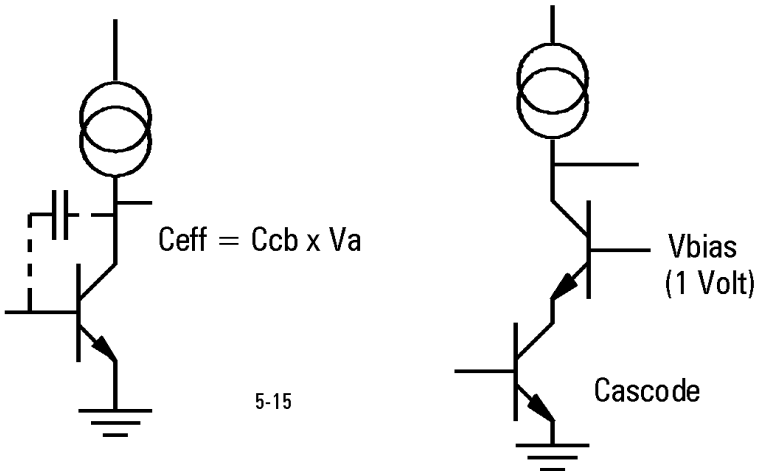
Now back to our comparator, and this time the design is complete. We have employed a PNP current mirror as the load of the differential stage. Now even if the operating current I varies, one-half of it will be mirrored.



Notice the four transistors at the bottom. The left-most transistor is diode-connected, steering a total of *three* other transistors: the first delivers the operating current to the differential pair, the second forms a load for the second stage (more high gain) and the third supplies the bias current for the output stage. Since the output stage normally requires more current (to drive loads), we have simply given that transistor 3 emitters, so that it draws three times the base current and delivers 3I. In the 700 Series the NPN transistors can have one, two or three emitters, just for that purpose.

Principle #3: Current Amplification

As we have noted above, the voltage gain in the bipolar transistor is anything but linear. Current gain (hFE), on the other hand, is a naturally linear parameter. For this reason alone it is easier to achieve high performance stressing current rather than voltage amplification.



But there is a second reason. Each junction has a capacitance (created by the "space-charge region"). Of particular bother is the collector-base capacitance. Not that it is especially large (it isn't), but it is badly situated. Using the transistor as a voltage amplifier, base and collector terminals move in opposite directions (i.e. they are 180^{o} out of phase). Since the transistor is capable of a large voltage gain (especially with a current source load), the voltage swing at the collector can be several hundred to several thousand times as large as that of the base. Looking into the base, the far end of the collector-base capacitance has a large voltage swing, and thus its capacitance appears to be far larger. This "Miller effect", simply put, *multiplies* the C-B capacitance by the voltage gain. So, if the junction capacitance is say 0.5pF and the voltage gain 1000, a 500pF capacitance is presented to the input. With an input impedance of 10kOhms, the cutoff frequency is a mere 32kHz, a far cry from the real capability of the transistor.

The quick fix to this problem is the so-called Cascode stage. An additional transistor, with its base connected to a fixed bias voltage just high enough to keep the first transistor out of saturation, shields the input. The Miller capacitance now works against the bias voltage which, presumably, has a low impedance.

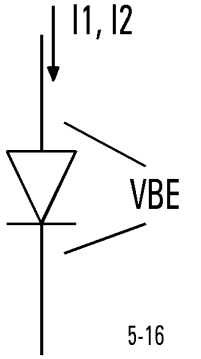
The cascode stage is only a halfhearted use of current amplification. A better approach (at least for high-frequency performance) would be to avoid converting to a voltage altogether.

Principle #4: Delta VBE

As we have seen above, the forward conduction of a diode is an exponential curve. Its absolute magnitude of the voltage contains the expression:

ln (I1/Is)

where ln is the natural logarithm, I1 is the current at which you are measuring the voltage and It is the saturation current. This latter parameter is a problem; apart from the material (silicon) it depends on the doping levels of both sides of the junction and the area of the diode. Thus, an NPN transistor connected as a diode, for example, has a different *absolute* forward characteristic compared to a PNP transistor.



$$\Delta VBE = \frac{kT}{q} \ln \frac{I1}{I2}$$

at room temp.: $\Delta VBE = 26mV \ln \frac{I1}{I2}$

I1/I2 = 2 ΔVBE = 18mV

I1/I2 = 10 ΔVBE = 60mV

I1/I2 = 100 ΔVBE = 120mV

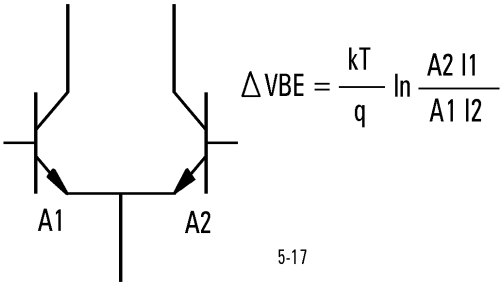
However, if we forget about the absolute level and concentrate on *relative* levels, we find that the manufacturing parameters drop out completely. The beauty of an exponential curve is its constant slope for *ratios*. If we start at a very low current and increase the current by say a factor of 10, we find that the voltage changes by a certain amount. If we now move to a much higher current and change it again by a factor of 10, we get exactly the same voltage change. The expression for this is:

Delta-VBE = (kT/q) ln (I1/I2)

The term VBE is commonly used in bipolar IC design for the diode voltage. It signifies the voltage between base and emitter, the most practical diode junction in an IC. All of the diodes we are discussing here are in fact transistors. By far the best connection consists of tying base and collector together. Since the current through the base in this connection is only 1/hFE of the total current (the major portion flows through the collector), the resistance of the base plays only a minor role (the collector resistance is not factor at all until the current is so high that the transistor saturates). The exponential behavior of such a diode-connected transistor is very accurate over about 8 decades of current.

At room temperature kT/q is approx. 26mV (a handy value to remember).

There is a second part to this formula: The change in voltage (delta VBE) can be created by changing either the current (as we have done so far) or the *area*. If we double the area of the diode, its voltage will drop 18mV at room temperature (at any current). If we increase it by a factor of 10, the drop is 60mV. What is important here is the *current density*, i.e., for example, milliamperes per square centimeter of diode area, but only in a relative way. It is never necessary for us to know the actual area of the diode.



The complete formula is shown in this figure, applied to a differential pair. If we double A2, for example (connect two emitters), we drop its base-emitter voltage by 18mV. In this way we can create a precise offset voltage. However, notice that this delta VBE is *proportional to absolute temperature*. If 18mV is our choice at room temperature, we get 24mV at 125°C and 13.2mV at -55°C. How do you counteract that? Make the current proportional to absolute temperature too.

So, we can create a delta-VBE with a current ratio or an area ratio, or *both at the same time*.

Delta-VBE is one of those parameters that comes along once in a generation. It is not affected by the process and it is highly predictable over a very wide range of current. True, it is dependent on temperature, but that again can be turned into a significant advantage, as we shall see next.

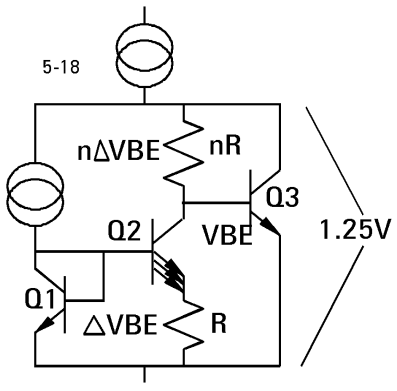
Principle #5: Delta-VBE vs VBE

Delta-VBE has a positive temperature coefficient. It is not just *some* positive temperature coefficient; it is strictly related to absolute temperature. At zero Kelvin delta-VBE is zero and increases from there in a straight line.

We have also noticed, casually, that VBE has a negative temperature coefficient. As it happens, this temperature coefficient is not just some arbitrary number either. As we lower the temperature, VBE increases linearly until it reaches a very specific voltage at absolute zero. For silicon this voltage is 1.25 Volts, the bandgap voltage, which tells us how much voltage does it take for electrons to cross the energy gap. This is a fundamental atomic constant and it not dependent on the process.

So, if we take a VBE and add it to a voltage related to delta-VBE we get less of a temperature coefficient. In fact, if we add enough delta-VBE voltage to the VBE, the two temperature coefficients cancel. The total voltage at which this happens is - surprise - 1.25 Volts, the bandgap voltage.

Delta-VBE tends to be small; as we have seen a 10:1 current density ratio produces about 60mV. We need to add about 600mV of delta-VBE voltage to a VBE to come up with 1.25 Volts total.



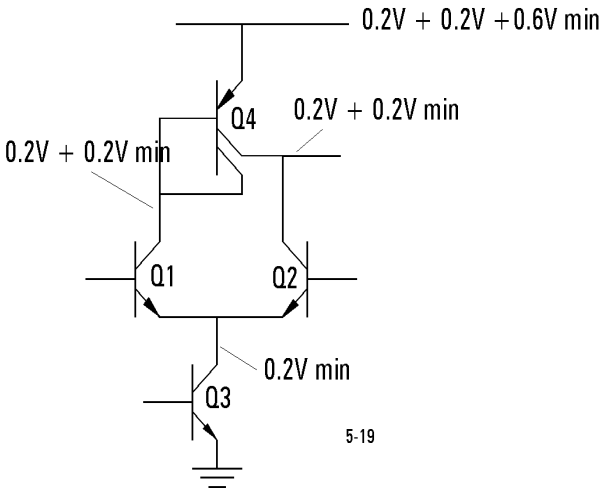
One way this can be accomplished is shown here. A delta-VBE is created between Q1 and Q2 by giving Q2 more emitter area and running it at a smaller current. This delta-VBE appears across R. If we ignore the base current, the emitter current of Q2 is the same as the collector current.

If we now give nR about 10 times the value of R, then 10 times delta-VBE appears across nR. Q3 then regulates the voltage so that it amounts to the voltage across nR plus its own VBE.

This circuit is a primitive bandgap voltage reference. It has a few errors, among them the base currents which we chose to ignore. But it illustrates the principle. You will find more (and better) bandgap circuit later in this chapter.

Principle #6: Low-Voltage Design

This linear IC design principle doesn't quite rank with the others; it only applies if you need to operate your circuit with very low supply voltages, say a battery voltage which can go as low as 1 Volt.



Each junction (in silicon) requires about 0.6 Volts to be forward biased. For a sophisticated design you quite often need to stack several devices between supply rails. So, at first glance, it seems that anything with more than one device would require more than one volt. But that is not the case.

The secret here is that the collector voltage can drop *below* the base voltage by a significant amount. At low currents, the collector does not saturate until it gets to within about 200mV of the emitter. This is illustrated with the basic differential stage here. The collector of Q3 (and emitters of Q1 and Q2) can operate down to about 0.2 Volts, the saturation voltage of Q3. The collectors of Q1 and Q2 have their own saturation voltage, so we are up to 0.4 Volts so far. Then comes then only diode in the bias chain (the half of Q4 which is diode-connected); it requires about 0.6 Volts (at low current). So the total supply voltage required to keep the devices out of saturation is 1 Volt. The output can then swing between about 0.4 and 0.8 Volts

Design Elements

General Remarks

The circuits described in this section are, generally, not complete by themselves. They are hardly ever more than a few transistors large and perform specific functions, useful within an IC design. They have been created over the years by a small cadre of clever designers and we have given the source when known.

Whenever possible or practical we have chosen specific values and given you performance data. But don't feel hemmed in by these values, regard them as a starting point only. It is exceedingly rare that you can use a circuit as designed, linear IC design just doesn't work this way. In almost all cases you will need to change the resistor values or change an emitter ratio here and there. We hope we have given you a good enough explanation with each element so that you can optimize it for your own use. We suggest you use Spice to help you do this; its a powerful tool.

All transistors in these circuits start with the letter Q. In a Spice simulation (see chapter 7) the model of an integrated transistor must be made into a sub-circuit to be accurate enough, and a subcircuit carries the letter X. In both Simetrix and P-Spice transistor numbers are automatically converted to begin with Q. Internal base resistors (RB) are in multiples of the unit resistor (750 Ohms); all others are in Ohms.

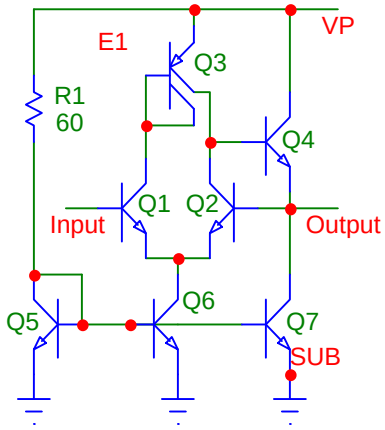
Similarly we call the positive supply VP. For most cases we have chosen just one supply, connected to VP and ground, to make things simple. This is an arbitrary choice on our part; these elements work equally well in an environment with positive and negative supplies or just a negative one.

The organization of this collection is alphabetical. In this section all schematics are numbered starting with the letter E. The numbering sequence carries no weight. In future revisions we will insert numbered elements at random, but we will not change any numbers assigned. In this way the "Widlar Current Source" will always be E5, no matter what revision you have.

If you know of any other elements, let us know. We want to make this the most comprehensive handbook for linear bipolar IC design ever written.

Buffer

A buffer (or voltage follower) has a voltage gain of exactly 1, a high input impedance and a low output impedance. E1 is probably the most simple buffer you can design, we haven't seen one with fewer transistors yet.



Q5 is a diode-connected transistor which receives the basic bias current from R1. This current is mirrored by Q6 and Q7. Q6 provides the operating current for the differential pair Q1 and Q2 and Q7 delivers the pull-down current for the output stage.

Q3 is an "active load", a current mirror which provides a high output impedance of the only gain stage. Q4 closes the loop and is capable of sourcing a large amount (say 5mA) of pull-up current.

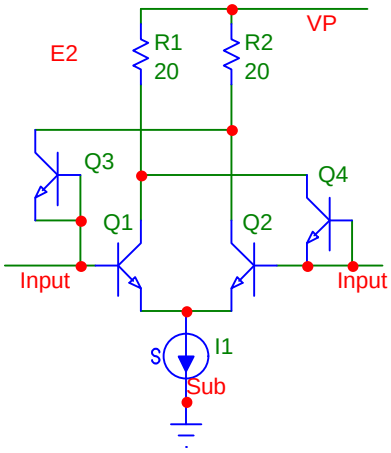
The higher the bias current, the faster the speed. Be careful not to exceed about 200 uA for the differential stage, Q3 (a lateral PNP) cannot handle more than that.

Features: Simple circuit. Current consumption is determined by a single resistor. Stable, requires no compensation capacitor. Low-voltage operation (down to at least 1.5 Volts)

Drawbacks: There is an error created by Q3, which is the most simple (and fairly poor) current mirror (see current mirrors). This error can translate into an additional 5 mV offset. The input current is not very low, it is the base current of Q1, or I/2hFE.

Improvements: You can increase the number of emitters in Q7, thus increasing the pull- down current at the output (but don't go too far, remember that Q3 needs to supply the base current for Q4, and pulling too much current at this point creates an additional offset voltage for the differential pair). A current source instead of R1 makes the bias current more supply-independent. A more sophisticated current mirror instead of Q3 reduces the offset voltage.

Capacitance Neutralization

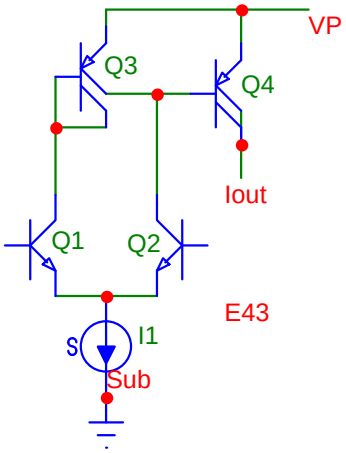


As explained in the first part of this chapter, the "Miller Capacitance" reduces the frequency response of a transistor by a large factor. The Miller Capacitance is the collector-base capacitance, multiplied by the voltage gain.

In a balanced differential pair the collector-base capacitance can be neutralized, as shown in E2. Here the collector-base junctions of identical transistors Q3 and Q4 are cross-connected. Now a voltage swing of the opposite phase counteracts the effect.

With 10kOhm loads and input impedances and an operating current of 1 mA, the extra two transistors increase the -3dB frequency from 2.2Mhz to 9.6Mhz. Care must be taken to assure that the differential pair is balanced at the inputs and outputs; this scheme will not work for a single-ended output.

Comparators



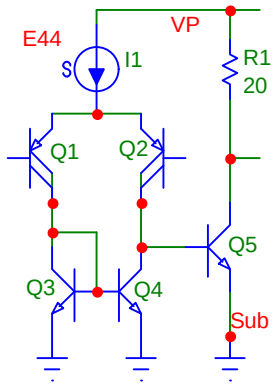
E43 (we told you the circuit numbers did not have to be in order) is the most simple comparator you can design. Q1 and Q2 form a differential pair; if the two inputs are at the same level, current I1 splits into two equal parts. The left-hand half, carried by the collector of Q1, drives the current mirror Q3 (see current mirror E15, here used as a so-called "active load"); the other half, carried by the collector of Q2, bucks the current returned by the second collector of Q3. Ideally these two bucking currents are equal (they are not quite).

Now, if input 2 is a little higher than input 1, Q2 gets a little more than 50% of I1, Q1 a little less. So there is a difference between the two bucking currents, which comes out of the base of Q3, turning it on.

This comparator has a small input offset voltage: as shown in E15, the base currents for Q3 cause an error (about 5%) in the current mirror. This translates into an input offset voltage of 1.3mV (V=(kT/q)ln(1.05) or (26mV)ln(1.05), at room temperature).
Be careful not to make I1 too high. The first reason for this is that the hFE of Q3 drops toward the upper current range, which can increase the offset by a large factor. The second reason: both Q3 and Q4 can saturate, which not only causes substrate currents (which can be equal to I1), but makes Q4 run at a high current (up to about 2mA, unless you limit Iout with a current source load).

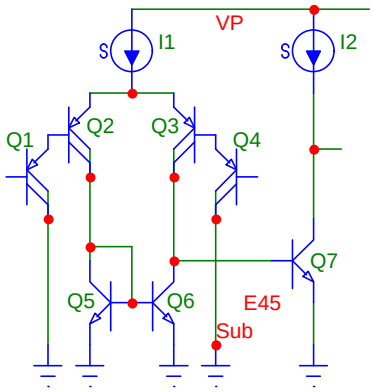
The performance of all comparators shown here depends greatly on the operating current (the more, the faster), so we won't give you representative performance figures.

You can simulate or breadboard these circuit elements in just a few minutes and get the figures for your specific conditions.



E44 is the same as E43, except that NPN and PNP transistors have been interchanged. Notice that this takes two separate transistors (Q3 and Q4) for the current mirror (active load). The collectors of each input transistor must be connected together; if you leave one floating it will saturate, which reduces the effective hFE and causes substrate current.

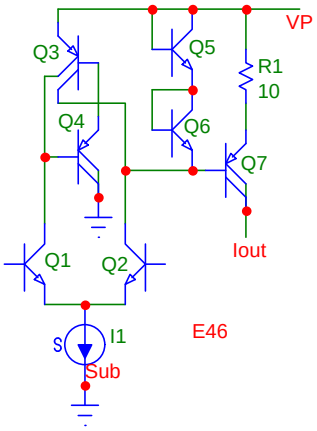
Here the NPN output transistor Q5 can saturate (if its emitter is at the most negative voltage) without causing much of a substrate current. If you use a current source instead of R1 the gain of the comparator increases drastically (i.e. the switching point is much sharper).



In the two previous examples the input currents are simply the base currents of the differential pair, or I1/(2hFE). For 200uA of operating current this amounts to about 1uA worst case. If this is too high for your application, consider using an extra pair of transistors, as shown in E45. Now the input current is divided by another hFE and thus drops to about 10nA (for an operating current of 200uA). The penalty is speed; compound transistors are considerably slower than single ones.

There is an advantage in E45 which could be significant: with the collectors of Q1 and Q4 connected to ground, Q2 and Q3 have more headroom. In other words, the inputs can move clear down to ground without any of the transistors being cut off or saturating. A great feature for single-supply operation.

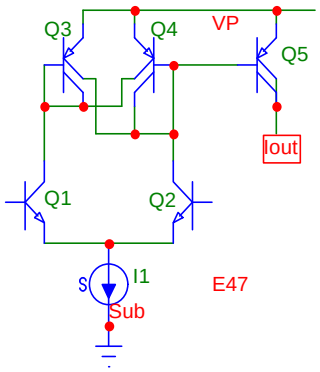
E46 is a more accurate version of E43 and also avoids over-current in the output transistor. First, we have added Q4, which avoids the base current error. Second, we limited the current of Q5 by adding R1 and the two diode-connected transistors Q6 and Q7. Now the current of Q5 cannot be larger than VBE/R1.



Even with E46 you should keep I1 low (say below 200uA), because Q3 can still saturate and thus cause a substrate current. The saturation of Q5 is much less of a problem now because its base current is limited by the presence of the two diodes.

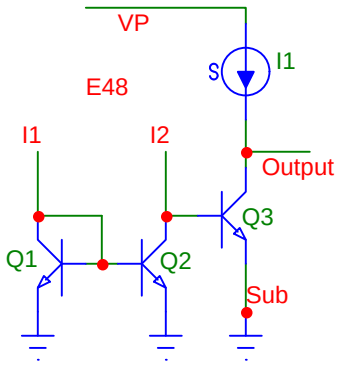
In E47 two current mirror transistors (Q3 and Q4) are used and their collectors are cross connected. At the switching point this creates positive feedback, which causes a snap- action.
In other words, within less than a millivolt of balance this comparator has infinite gain and a very small offset voltage (the base currents of Q3 and Q4 cancel out).

We like this comparator best, not only because of its clean switching point but also because there is very little substrate current, even with I1 as high as 500uA. Also, the current of Q5 doesn't go much above I1.

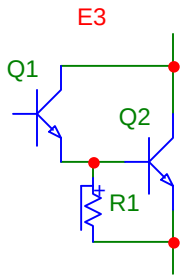


You can also compare currents, rather than voltages. In E48 Q1 and Q2 form a simple Widlar current mirror (see E5). If I2 is larger than I1, the current difference will flow into the base of Q3, turning it on. If I1 is larger than I2, Q3 is cut off.

The Widlar current mirror has some drastic errors, but here they are largely avoided. Notice that the collectors of Q1 and Q2 operate at the same collector voltage (VBE), so there is no Early effect error. If I3 has twice the magnitude of the input currents, that base current error is canceled too.



Compound Transistors



If you find that the current gain of a transistor is insufficient to do a specific job, you might try multiplying the gain with a second device.

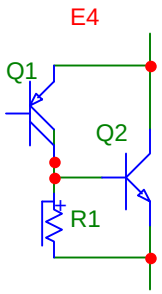
The first such scheme is the well-known Darlington connection, shown in E3. The emitter current of Q1 feeds into the base of Q2, thus the combined current gain is hFE1xhFE2. With a minimum HFE of 100, that results in a minimum combined current gain of 10'000!

There are some disadvantages though: 1. The combined collector cannot drop below one VBE (plus Vsat). 2. The leakage current of Q1 is multiplied by hFE2, a factor you need to carefully check when your circuit is to operate at 100°C or higher, and 3. The turn-off (storage) time is very slow because there is no way to remove current out of the second base. To speed up the turn-off you might want to connect a resistor as shown (an ideal application for a base pinch resistor). Note, however, that this also reduces the combined current gain.

The collector of Q1 can be connected directly to the supply, which makes the saturation voltage as low as that of a single transistor. But make sure the current of Q1 does not go sky-high (if Q1 is switched, put a resistor in series with the collector).

This circuit works equally well for PNP transistors.

E4 shows the combination of a PNP with an NPN transistor. At the base of Q1 the compound transistor looks like a PNP, but it has the current capability of an NPN.



E4 has limitations identical to E3: 1. The output cannot drop below one VBE (plus Vsat of Q1), 2. The leakage current of Q1 is multiplied by hFE2, a factor to be considered at high temperature, and 3. Turn-off is slow; a resistor as shown does wonders to speed up the circuit, but reduces the combined current gain.

You can, of course also make Q1 NPN and Q2 PNP. The compound transistor then will be NPN with a current gain of hFE(NPN) x hFE(PNP), but its current capability is limited by the PNP transistor.

Current Mirrors

Current mirrors are powerful design tools for linear ICs. For this reason we list, discuss and compare here all known bipolar current mirrors.

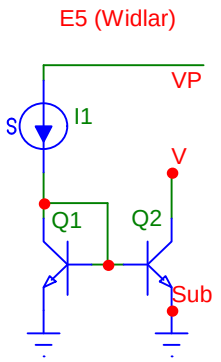
For each current mirror we have plotted its accuracy, i.e. the ratio of the secondary to the primary current, derived by sweeping the supply voltage in a Monte Carlo Analysis. This will show you at a glance what level of error you can expect.

All of these parameters have been measured using a uniform current of 100 uA for NPN and 50uA for PNP, an arbitrary choice. Use these figures only to compare the various current sources.

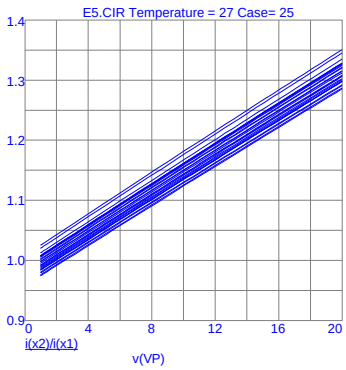
In all of these current mirrors we have assumed that you want to produce a current ratio of exactly 1, i.e. the output current is the same as the input current. This need not be so. By using emitter ratios (see the first part of this chapter) and, for some of the circuits, resistor ratios, you can produce any ratio of currents you desire (well, almost any). These ratios do not have to be integers; ratios of 2 to 3, or 5 to 4 are easily possible, though they tend to consume a few more transistors.

You can design current mirrors using either NPN or PNP transistors. Purely NPN current mirrors sink current at the output (the current goes to ground or the negative terminal); those with PNP transistors source current (the current comes from the positive terminal). We start with NPN current mirrors:

In this first and most simple current mirror (E5), Q1 is connected as a diode, the voltage at the collector is the same as that of the base. Mind you, it is still an active transistor, collector and base currents are strictly related by hFE (unless you are so high in current that the transistor saturates).



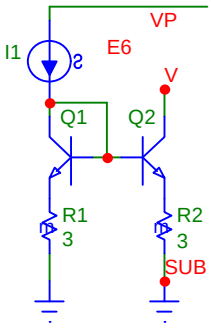
The base and emitter of Q2 are connected to the same points as the base and emitter of Q1. So, the two devices have the same base voltage and, therefore, the same base current. Assuming the two devices match perfectly, the two identical base currents produce identical collector currents in the two transistors. If we ignore the base currents, I2 is the same as I1. What we are doing here is duplicate a current and, at the same time, convert it from a sourced current into a current sink.



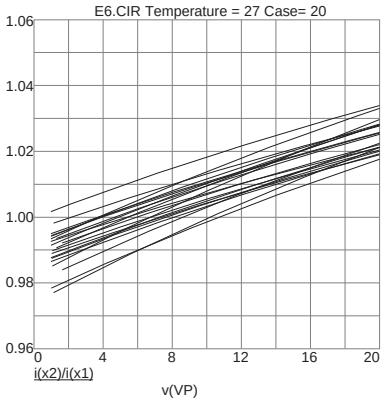
There are two errors in this circuit. First, we cannot really ignore the base currents. With a minimum hFE of 100, each base current amount to 1%. I1 not only has to deliver the collector current for Q1, but two base currents. I2, therefore, is smaller than I1 by about 2%.

The second error is more serious. The collector voltages of Q1 and Q2 are not necessarily the same. The one at Q1 is always VBE, about 0.6 or 0.7 Volts. The one at Q2 depends on you application, it may be as low as 0.2 Volts or as high as 20 Volts. The hFE of Q2 increases with increasing voltage (the "Early Effect"), causing the collector current of Q2 to be higher than that of Q1, by as much as 20%.

References: R.J. Widlar: "Some circuit design techniques for linear integrated circuits," IEEE Transactions on Circuit Theory, Dec 1965, pp. 586-590. Widlar: "Biasing scheme especially suited for integrated circuits," US Patent 3,364,434, 1968.

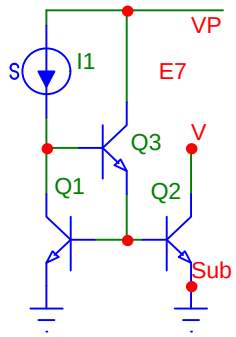


The accuracy of the Widlar current source can be improved by adding identical emitter resistors, as shown in E6. The voltage drops in the resistors provides local feedback, forcing the currents to be more equal. The larger the resistors, the better the matching (we dropped 200mV to get the performance shown).

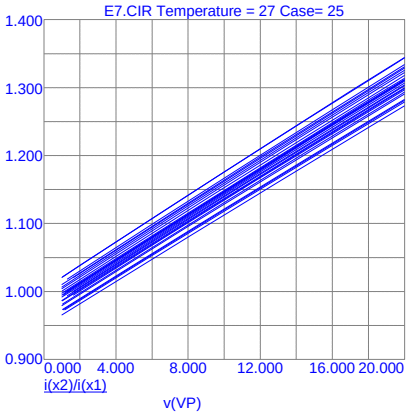


But there is a limitation here. The scheme only works if I1 stays within a narrow range. If I1 varies, say, over two decades, the voltage drop across the resistors at the low end of the current is too small to make a difference. Also note that the resistor drop (and Vsat) determines how low the collector of Q2 can move.

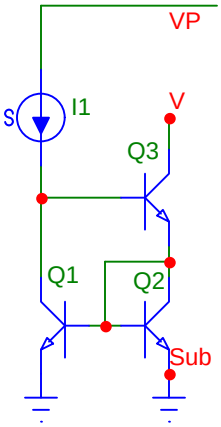
A general remark about current ratios, before we proceed. To get twice the output current in E5, for example, you give Q2 two emitters (i.e. the base-emitter junction has twice the area). To make it one-half, you give Q1 two emitters. I2 will be 2/3 of I1 if Q1 has three emitters and Q2 two. In E6 you must give the resistors the inverse ratio. For example, if I2 is to be one-half of I1, Q1 gets two emitters and R2 is made 2xR1.



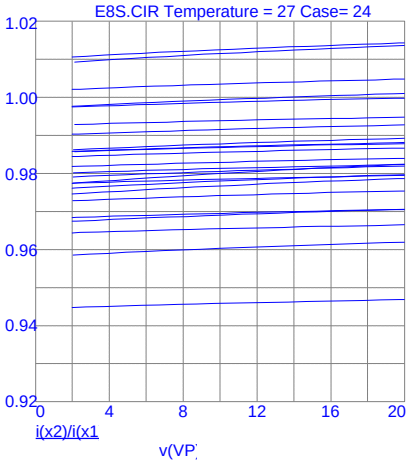
We can remove the base-current error of the Widlar current mirror with an extra transistor, shown as Q3 in E7. The base currents are now delivered directly from the positive supply rather than through I1. The Early Effect error, however, remains. Not much of an improvement, unless you know that the output voltage will be at about a VBE.



E8 (Wilson)



A big step up in accuracy is the Wilson current mirror, shown in E8. It costs you an extra transistor and requires a higher minimum operating voltage, but the improvements are dramatic (the voltage at the input must be at least two VBE (about 1.3 Volts) and the output cannot drop below one VBE plus Vsat).



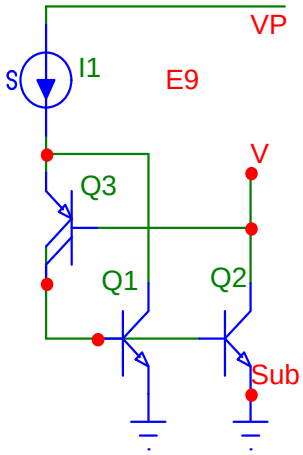
If you care to do the analysis, you will find that all the base currents exactly cancel. More importantly, there is a feedback loop which tends to counteract some of the voltage-related hFE variation in Q3. The nice thing about the Wilson current mirror is that this feedback loop is stable without requiring a compensation capacitor.

References: George R. Wilson: "A monolithic junction FET-NPN operational amplifier," IEEE Journal of Solid State Circuits, December 1968, pp. 341-348. G.R. Wilson: "Regulator controlled by voltage across semiconductor junction device," U.S. Patent 3,886,435, 1971.

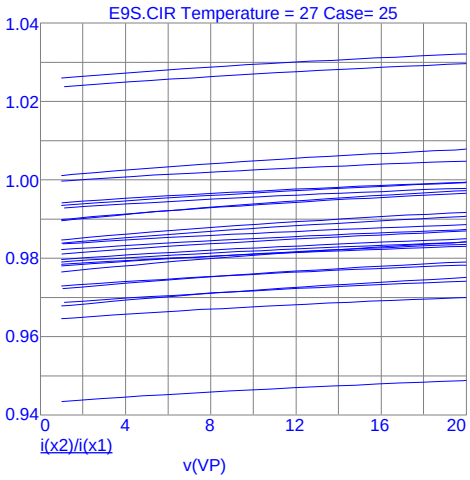
A General Note on Multiple Current Mirrors:

You can easily produce more than one mirrored current. For example, if you connect a fourth transistor with its base and emitter parallel to those of Q2 in E8, it will receive the same base current as Q2 and its collector current will be the same (more or less because of the Early effect) as I2. Using two transistors in series, with their bases connected to the bases of Q2 and Q3, results in a better second current output (more on this under current sources).

But: every time you add another output you add some base-current error.

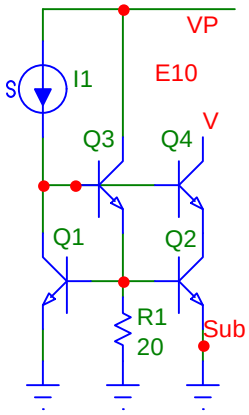


E9 adds a PNP transistor to the basic Widlar current mirror. The base of Q3 senses the output voltage and causes its collector to deliver just enough base current to Q1 and Q2. This base current, however, comes from I1, so the first (and usually smaller) error remains. The Early Effect error is greatly reduced, since Q3 holds the collector voltage of Q1 and Q2 with one VBE. The output voltage can drop to as low as two saturation voltages (about 400 mV) and the input must be at least one VBE higher than that.



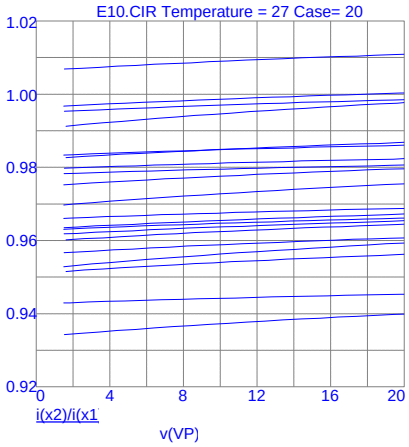
In E9 the input does not stay at a constant low voltage, it is always one VBE higher than the output. So, for some applications this scheme (and some of the following current mirrors) won't work. In other applications, as for example in an "active load" for a differential pair, it is actually desirable to have input and output voltages tracking (ideally: the same), so that the transistors to which the current mirror is connected, also have the same collector voltages.

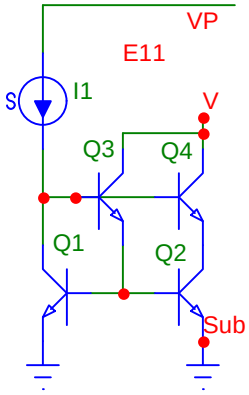
The remaining NPN current mirrors require four transistors:



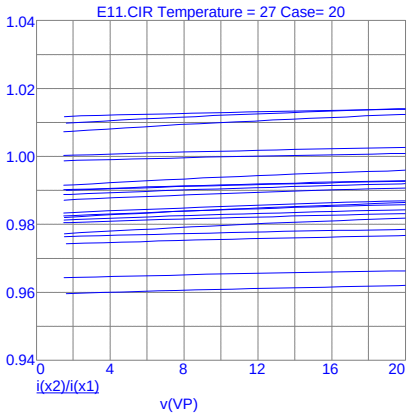
E10 is a derivative of E7; it adds a "Cascode" transistor, Q4, to shield Q2 from the voltage variation at the output. Resistor R1 speeds up turn-off time and bypasses possible leakage current of Q3. The output cannot be lower than a VBE plus the saturation voltage of Q4 and the input is fixed at two VBE.

There is still a base current error due to Q4, which is delivered through I1; if you draw a substantial amount of current through Q3 by making R1 low, you will also have to add the base current of Q3 to the error.





E11 is a modification of E10. The base current for Q1 and Q2 is taken by Q3 from I2. Since the base current for Q3 comes from I1, there is some compensation. But note that you can no longer afford to connect a resistor between bases and emitters of Q1 and Q2, the resulting current would produce too much of an error. Input and output voltage limitations are unchanged from E10.



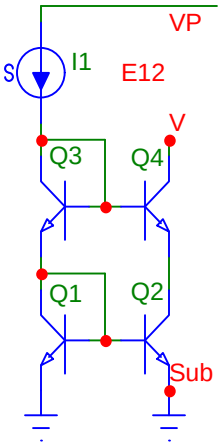
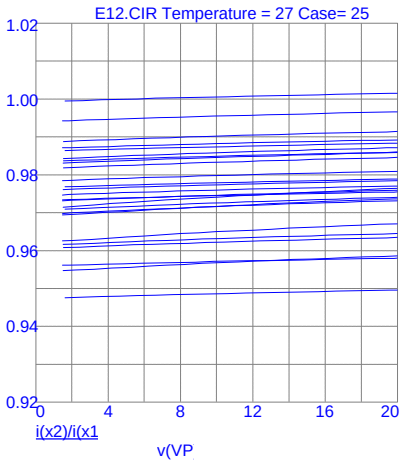
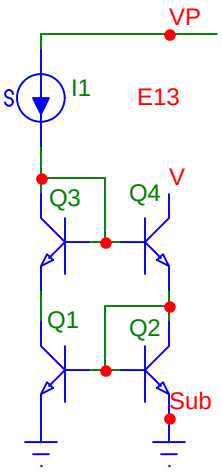


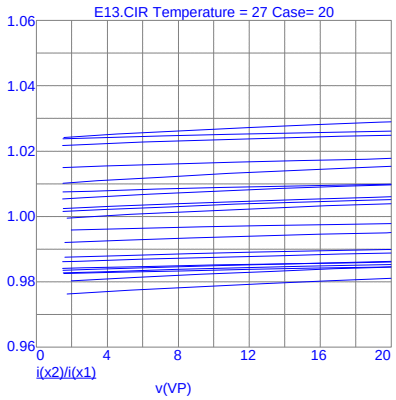
Figure E12 shows an alternate connection for a "Cascode" stage. The collectors of Q1 and Q2 are both at VBE and Q4 shields Q2 from the output voltage variation. Its output cannot be lower than a VBE plus the saturation voltage of Q4 and the input is fixed at two VBE.

There is no compensation for base currents in E12. Because the following current mirror is so much better in performance with only a small change, E12 does not have very much to recommend itself.

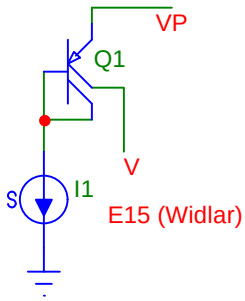




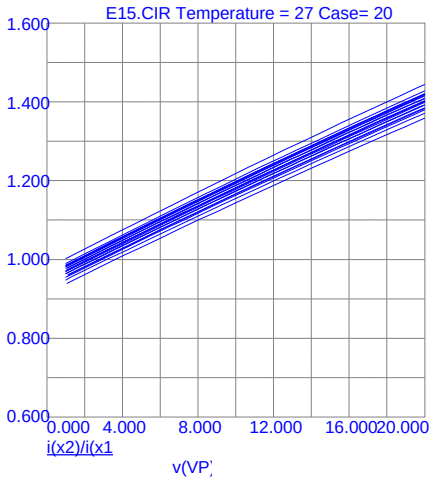
If we connect Q2 as a diode (rather than Q1), we get both base-current cancellation and feedback action. If you look at E13 closely you will find that it is very similar to the Wilson current mirror, but that Q1 and Q2 now have identical collector voltages. The voltage requirements remain the same: the input is at a fixed two VBE, the output can drop down to VBE plus the saturation voltage of Q4.



Following now are the PNP equivalents of the current mirrors just shown. With one exception (E23), their upper current handling capability is limited by the PNP transistor. You get nearly full performance up to about 100 uA per collector. Above this current the hFE drops rather rapidly. While these current mirrors still work at 500 uA or even higher, their errors become substantially larger. You can, of course, operate at a higher current by connecting additional PNP transistors in parallel.

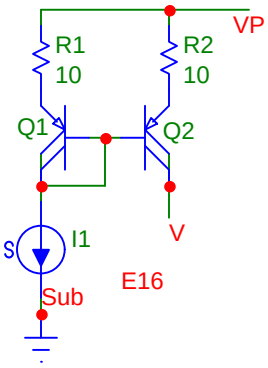


E15 is the Widlar current mirror implemented in PNP (see E5). The big advantage here is that a single (lateral) PNP transistor can do the entire job, since this collector can be naturally split into two identical regions, each carrying half the current.

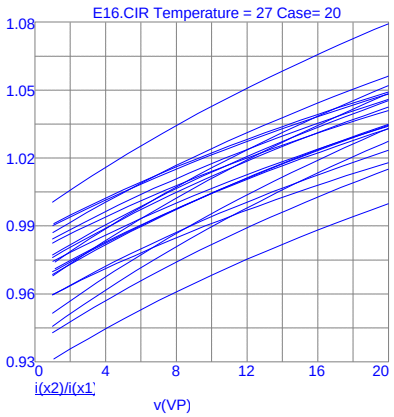


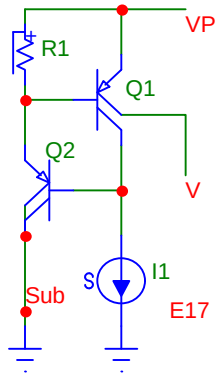
The performance in PNP, however, is quite a bit inferior. First, the PNP transistor has a lower current gain, resulting in a larger base-current error. Second, the (lateral) PNP transistor has a more severe Early Effect, the hFE increases even more at the collector-emitter voltage is increased. Third, the lateral PNP transistor is much slower than the (vertical) NPN transistor.

So, when you use the PNP Widlar current mirror (as you will, quite often), be aware of the errors and the speed limitation. A 36% current error translates in an 8mV offset ((26mV)ln(1.36)).

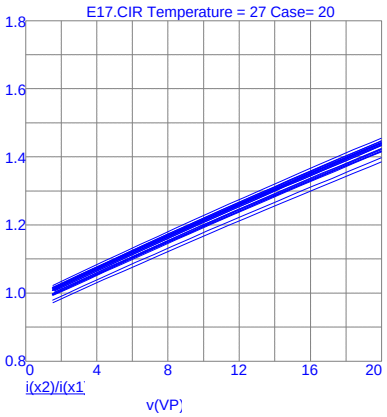


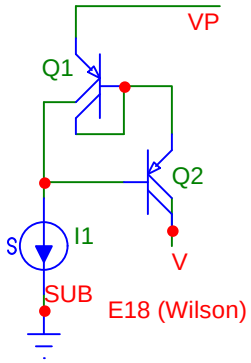
As in E6, you can improve the accuracy of the Widlar current mirror with identical emitter resistors, which even out the differences in hFE, but you lose the advantage of the single transistor. For our analysis we have dropped 200 mV across the resistors, but the larger, the better. Be aware, however, the such resistors work well only over a narrow current range; if I1 drops to a low level, the voltage drop across the resistors becomes too small to matter and the matching gets worse again.



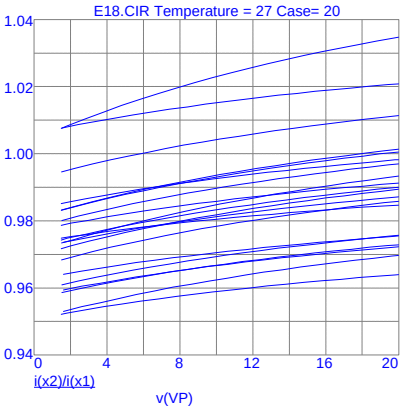


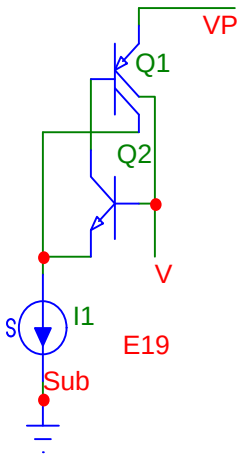
E17 is the PNP equivalent of E7, with R1 as an optional bypass for improved speed and high-temperature performance (to avoid having the leakage current of Q2 multiplied by Q1, which can be a bothersome amount at 125°C). For our analysis we used a base pinch resistor. Notice the slow switching speeds; there is also a substantial amount of ringing during turn-on.





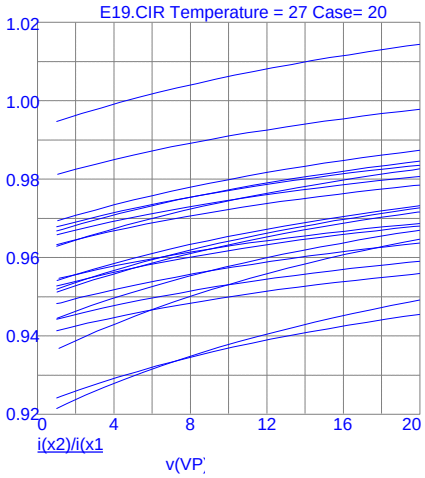
The PNP Wilson current source (E18) requires just two transistors, but compare the performance with that of E8; its output impedance is quite a bit lower. Something to keep in mind: because of the lower current gain and stronger Early Effect, PNP current mirrors tend to be poorer and usually require more transistors. You may want to consider turning the circuit upside down, so that the current mirroring can be done by NPN transistors.



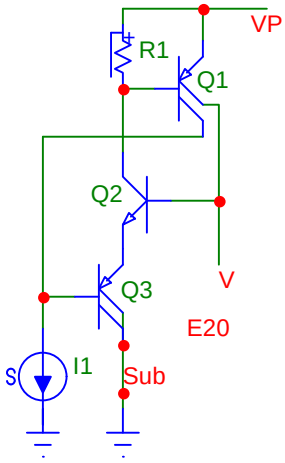


In E19 the two collector voltages are forced to approx. the same level by Q2 (the input is one VBE lower than the output). This removes the major error, caused by the Early effect, but the base current for Q1 is still supplied by I1. This circuit is one in which the input is not at a constant voltage. Read the remark under E9.

Notice the painfully slow turn-off time. You could improve it with a resistor between base and emitter of Q1, but you can't afford it because it would cause a major error (the current would have to come from I1). E19 also shows considerable ringing during turn-off.

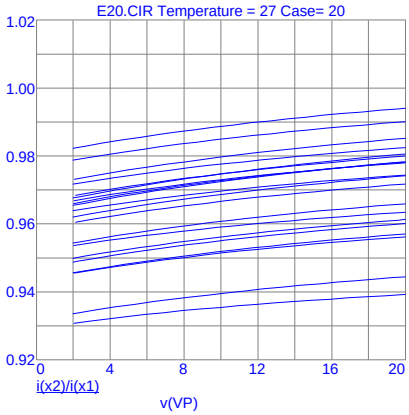


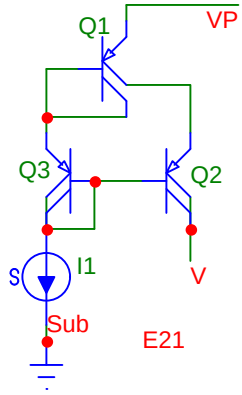
Reference: Ahmed: "Current Amplifier," US Patent 3,843,933, 1974.



With an additional transistor (compared to E19) we can shunt the base current (and the current consumed by the optional resistor R1) to ground. The two collectors are now two VBEs apart, but they still track. Note, however, that the output cannot be below 2VBE, otherwise Q2 and Q3 get cut off.

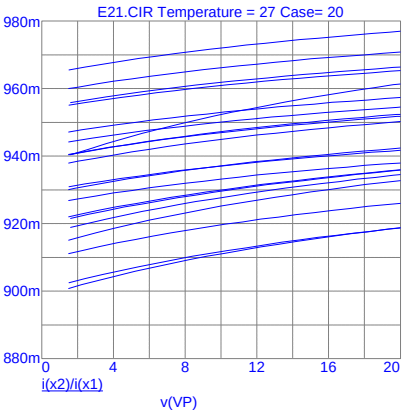
Reference: Ahmed: "Current Amplifier," US Patent 3,843,933, 1974.

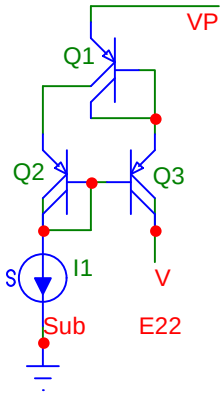




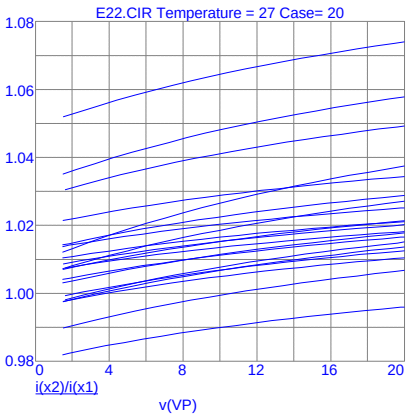
E21 is the PNP equivalent of E12 and, as before, there is no base-current compensation. Compared to the circuit immediately following it is inferior, except for a slightly higher output impedance. Notice especially the drastic difference in turn-off time.

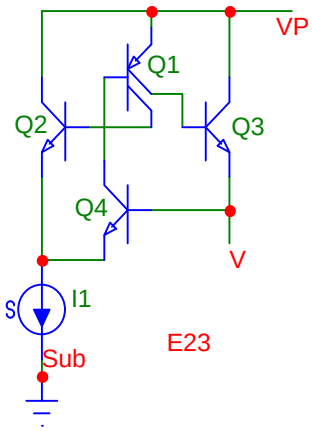
Reference: Witlinger: "Current amplifier," US Patent 3,835,410, 1974.





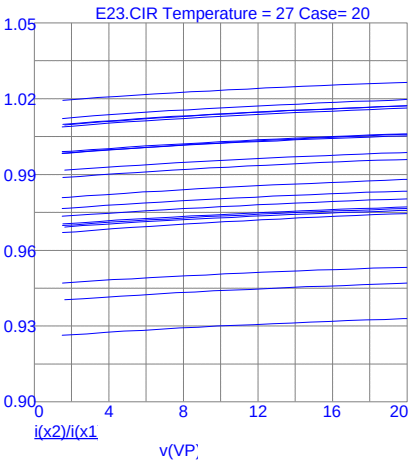
E22 corresponds to the NPN circuit E13. This is the Wilson current mirror (E18) with transistor Q3 added. The sole purpose of Q3 (which is connected as a diode) is to increase the collector voltage of Q1 so that the two collector voltages are identical. In an PNP transistor, which has a strong Early effect, this small difference results in a significant improvement.



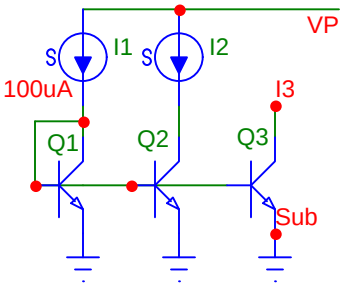


If you need a current higher than PNP transistors can deliver, E23 may do the job. Here Q2 and Q3 form com pound transistors with Q1. The lion's share of the current flows through the NPN transistor; Q1 only has to deliver their base currents.

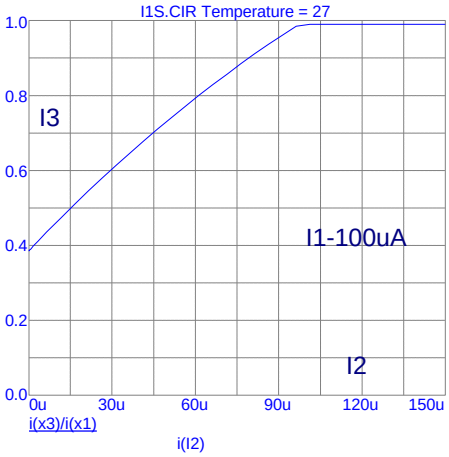
As in E19, Q4 holds the input and output voltages at approx. the same levels (the input is always one VBE lower than the output) and delivers the base current for Q1 from I1. The base current error is now very small because Q1 runs at a low current. This low operating current, however, has a telling effect on switching times.



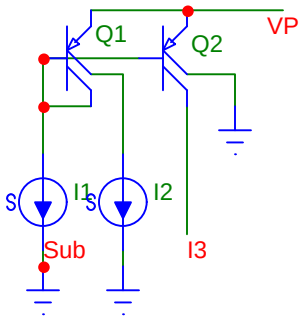
What Happens When a Current Mirror Saturates?



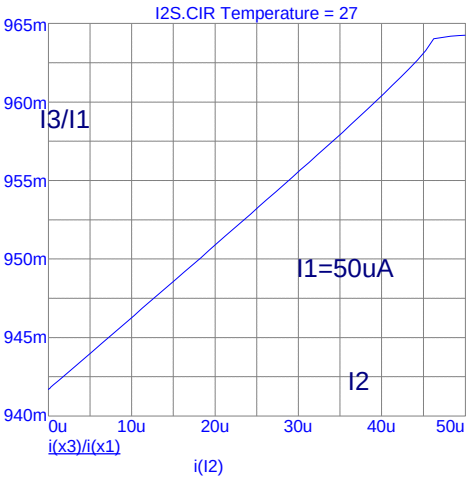
If you have several dependent current mirrors and one of the transistors saturates, it disturbs the others; the collector, dropping in voltage to near the emitter potential, forms an additional diode parallel to the diode-connected transistor.



Here is our measurement for the NPN current mirror. The ratio I3/I1 is measured (with Q3 kept out of saturation), as I2 is reduced and Q2 saturates. As you can see, the effect is significant: at low current I3 drops to as little as 40%.

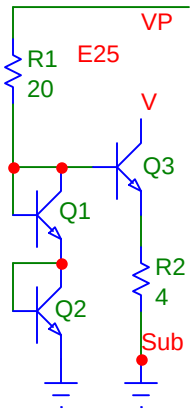


The PNP transistor is considerably less sensitive to this effect. As I2 is reduced and one collector of Q1 saturates, the currents in the collectors of Q2 drop very little.

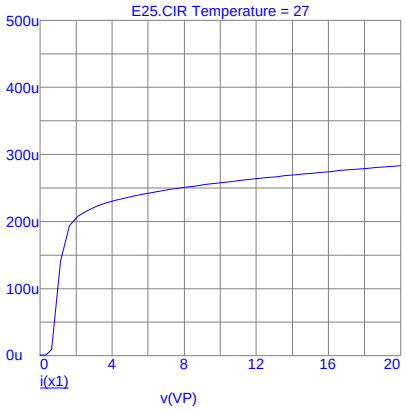


Current Sources

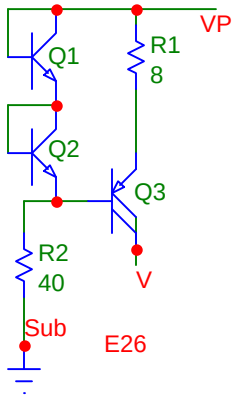
Just to make sure we don't have a misunderstanding, let's re-iterate: strictly speaking half of the current sources discussed here are actually current sinks. But we simply go along with convention and call all these circuits current sources. (If you want to be clever about it: there are no real current sources in ICs. The power supply is the actual source of the current; the so-called "current source" merely defines the current drawn).



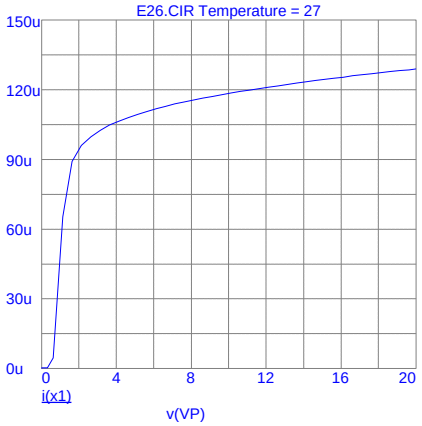
Our first example, E25, is a simple one. A diode (in our case a diode-connected transistor) has a very non-linear (exponential) curve when forward biased. Here we connect two of these diodes (Q1 and Q2) in series and forward-bias them with R1. The base of Q3 picks up the two-diode voltage, which leaves a net voltage of one diode (VBE) at its emitter. The current through Q3 is simply VBE/ R2. As you can see from the curve, the current is reasonably well regulated. Practical current range: 20uA to 4mA.

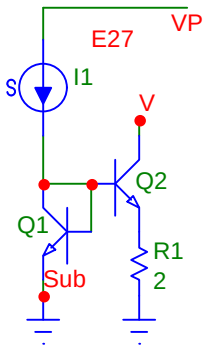


The negative temperature coefficient of VBE and the positive one of the resistor result in a current change of about -0.33%/°C. The variation is mostly that of the resistor, increased slightly by the variation of the 3 VBEs. You can expect 3-sigma to be ±30%.

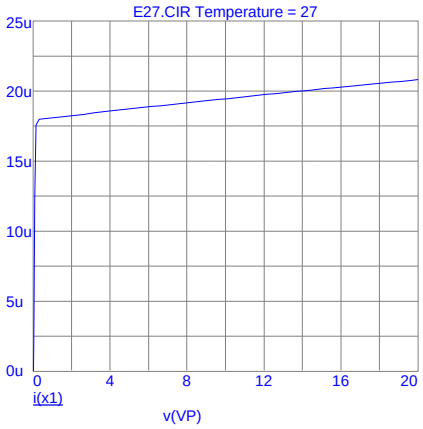


You probably guessed what's coming: we can make a PNP equivalent of E25. In E26 we chose diode-connected NPN transistors; we could have chosen PNP transistors, it makes very little difference in the value of I1 but you have to be careful not to exceed their current limit and saturate them (watch out for the substrate current if you do!).





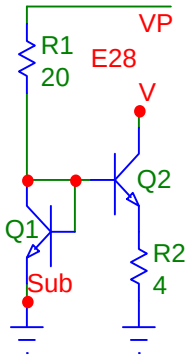
We don't really need to bias the base of the current source transistor from 2VBE, a single VBE is enough. In the "Widlar current source" of E27 the current of Q2 is reduced (compared to I1) by R1. You can create a low current with this circuit without requiring large-value resistor.



There is no direct calculation of I2, you need to use an iterative approach, simulate the circuit or breadboard it. Here is how you estimate the current: R1 reduces the current in Q2, which creates a delta-VBE (=(26mV)ln(I1/I2), at room temperature). This delta-VBE appears across R1, creating a current I2 = (delta VBE)/R1. You now need to go back and plug in the new ratio of I1/I2 and so on. Spice can probably do it a lot faster.

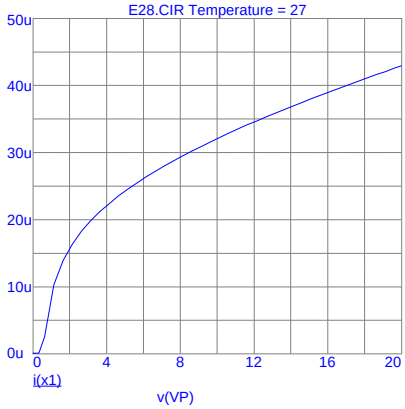
The temperature coefficient of E27 is determined by the delta-VBE (positive) and R1 (negative) and is about +0.2%/^{o}C at low temperature and +0.1%/^{o}C at high temperature. This circuit works down to less than 1 Volt. The 3-sigma variation is that of R1, or about ±25%.

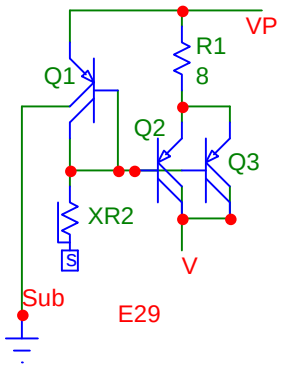
References: Widlar: "Some Circuit Design Techniques for Linear Integrated Circuits," IEEE Transactions on Circuit Theory, Dec. 1965, pp. 586-590. Widlar: "Low-value current source for integrated circuits," US Patent 3,320,439, 1967.



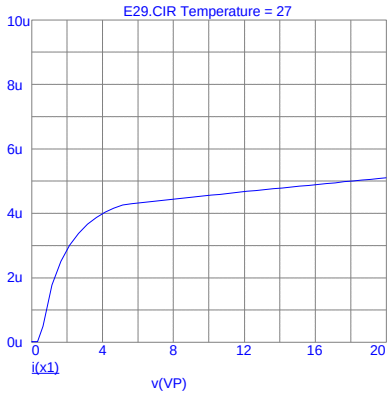
If you don't have a current source for the primary current, a resistor will work fairly well, as shown in E28. Of course the supply rejection is not nearly as good, but there is some regulation.

To change the output current you can vary R1 and R2 or increase the number of emitters for Q2 (more current) or for Q1 (less current).



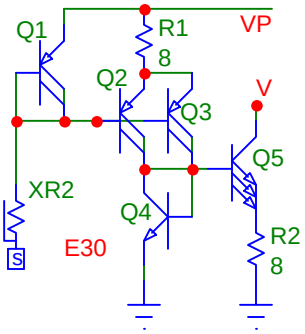


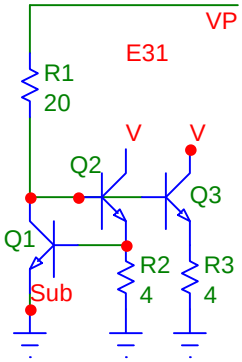
Yes, a PNP equivalent again (E29), but with some wrinkles. This current source has a very small primary current, that of the epi pinch resistor. We double it by grounding one of the collectors of Q1. Then we double Q2, which results in a delta-VBE of 18mV. The resulting current begins to depend a little more on the delta-VBE and R2 and less on the epi pinch resistor, leading to a reduction in variation (by 40%)



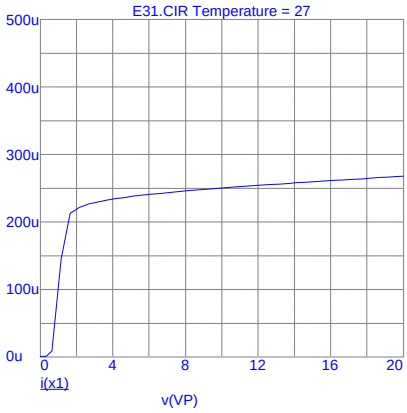
and a smaller temperature coefficient (-0.6%/^{o}C at low temperature, -0.4%/^{o}C at the high end).

If you need a current sink we can carry this approach even further. In E30 the current is mirrored once more by Q4 and Q5. We have given Q5 three emitters, which results in a delta-VBE, i.e. an additional positive temperature coefficient and a further reduction in the variation caused by the epi pinch resistor (by about 65%). The total temperature coefficient is now -0.1%/°C at low temperature and -0.2%/°C above room temperature. Practical current range: 1 to 10uA.

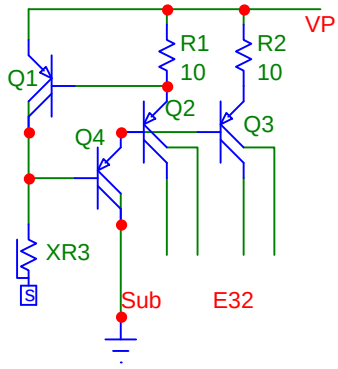




Now we are going to get slightly more sophisticated. In E31 a primary current is created by R1. The voltage at the collector of Q1 and base of Q2 is lifted until the emitter of Q2 drives just enough current into the base of Q1 to counteract the primary current at the collector. At this point the cur rent through Q2 is VBE/R2 (plus a small amount of base current for Q1). Additional transistors can be connected as shown with Q3 and R3 can be different from R2.

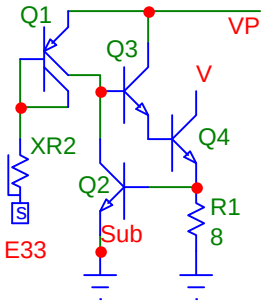


The temperature coefficient of I1 (and I2) is that of a VBE and a resistor (the first negative, the second positive), or about -0.35%/°C. Practical current range: 20uA to 4mA.

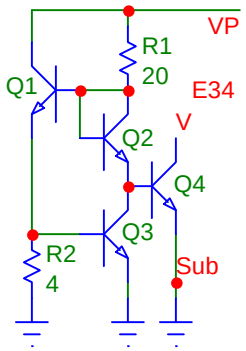


E32 shows a PNP version of E31. Here we use an epi-pinch resistor, which wastes little current. To assure that Q2 and Q3 have enough base current, we make them into compound transistors with Q4. Note that with a PNP transistor the output current can be split into two equal parts (I1+I2=VBE/R1). Again, R2 can be different from R1. Practical current range: 20uA to 200uA (limited by the current range of the lateral PNP transistor). Temperature coefficient and 3-sigma variation are similar to those of E31.

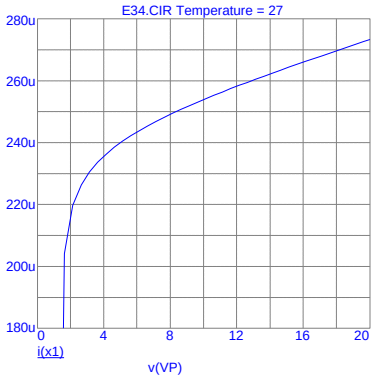
E33 is identical to E31, except that R1 has been replaced by an epi pinch resistor and Q1. Less than 3uA (twice the current of REP) is wasted to create a current of 92uA. If you tried to achieve this with a base resistor, it would take about 3MOhms with VP=10V! Again, the compound Q4 is necessary to assure that Q3 receives enough base current. Practical current range: 20uA to 4mA. Temperature coefficient and 3-sigma variation are identical to E31.



E34 is a current source especially designed to work down to very low output voltages. It's a bit difficult to understand, so bear with us.

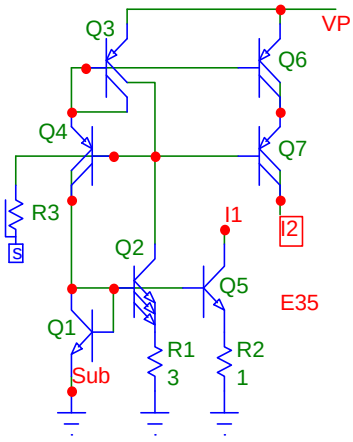


Q1, Q2 and Q3 form a feedback loop. The current through R1 (the primary current) lifts the base of Q1. The emitter current of Q1 flows into the base of Q3, turning it on. The collector current of Q3 flows through the diode-connected Q2 and then counteracts the current coming from R1. So far we have established an operating point for the three transistors.

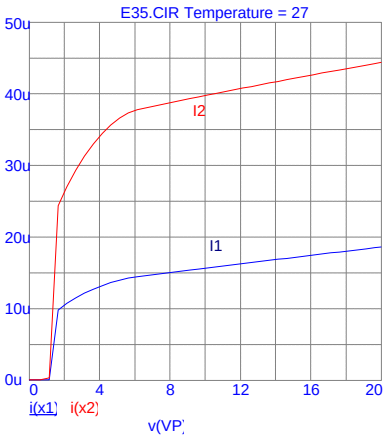


Now, the current through R2 must be VBE/R2, since R2 is connected between the base and emitter (the VBE) of Q3. This is also the current for Q1. Q2 and Q3, on the other hand, run at whatever current is supplied by R1, a current that changes with supply voltage. And here comes the important part: to get the VBE at the base of Q4, we add the VBEs of Q3 and Q1 and subtract the VBE of Q2. Thus the two poorly controlled VBEs cancel and we end up with a voltage at the base of Q4 which is identical to the VBE of Q1. The current through Q4, therefore, is VBE/R2.

Since the emitter of Q4 is at ground, its collector can operate down to the saturation voltage. The temperature coefficient of E34 is about -0.35%/^{o}C and the 3-sigma variation ±28%. Practical current range: 20uA to 4mA.



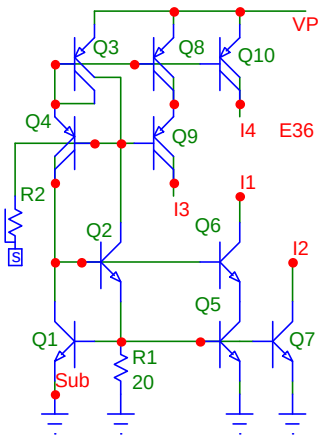
In all of the current sources discussed so far there is a primary current from which the secondary current is derived. In E35 no such primary current is needed. The collector current of Q2 is returned to the diode-connected Q1 by the Wilson current mirror Q3/Q4. In a way this current source creates its own primary current. (See E27 for the Q1/Q2 Widlar current source)



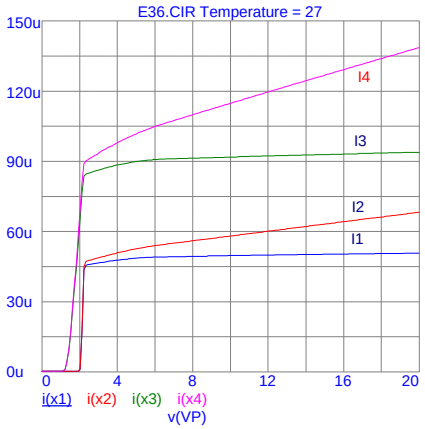
But you have to be very careful about these kind of circuits: somehow they must be started up (see Startup circuits). This can happen by itself because of noise or leakage, but you don't want to leave it to chance. So in this case we have added R3, an epi pinch resistor. The epi pinch resistor draws current out of the base of Q4, which in turn draws base current from Q3, which starts the loop; the epi pinch resistor has a sufficiently high current level to get the loop going but is not so high as to disturb the operation of the loop in a significant way at full current.

The current source itself is a closed loop, so it cannot deliver current directly. But you can connect other strings (more than shown) as we have done with Q5/R2 and Q6/Q7. R2 can be quite different from R1 and Q5 can have more than one emitter (for higher current). Q7 is not absolutely necessary, it just provides a Cascode effect and makes I2 less dependent on voltage.

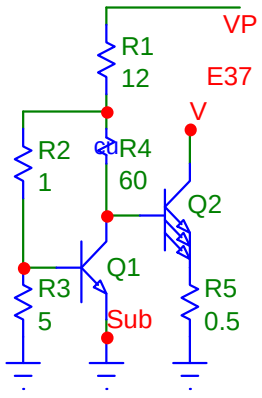
Note that I2 has roughly twice the magnitude of I1. This is because Q3 has to deliver two collector currents. You can, of course, use the collectors of Q7 (and Q6) separately. Temperature coefficient is about +0.25%/°C and 3-sigma variation ±30%. Practical current range: 1 to 200uA.



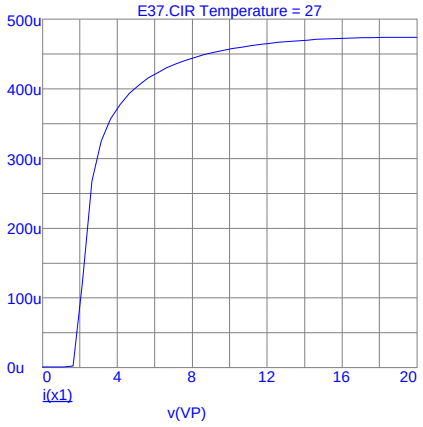
E36 looks complicated, but isn't. Q1 and Q2 form the same current source as in E31 and the collector current of Q1 is mirrored back to Q1 the same way as in the previous circuit (and also uses the same startup scheme). It is simply a combination of two circuit elements different from the combination in E35.



Notice that, as is almost always the case, the outputs with cascode transistors (I1 and I3) have some ten times the impedance of the ones without the extra transistor. Temperature coefficient is about -0.5%/°C and 3-sigma variation ±28%. Practical current range: 20 to 300uA.



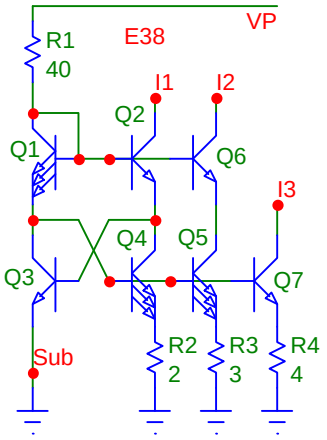
E37 contains two features that you can apply to many of the other current sources as well. The two transistors basically form a Widlar current source (see E28), but Q1 is not simply connected as a diode. Instead, R2 and R3 form a voltage divider so that, at the junction point of R1, R2 and R4 there is a voltage about 17% higher than a VBE. This node has a voltage characteristic identical to that of a diode, only the voltage is multiplied by 1.17.

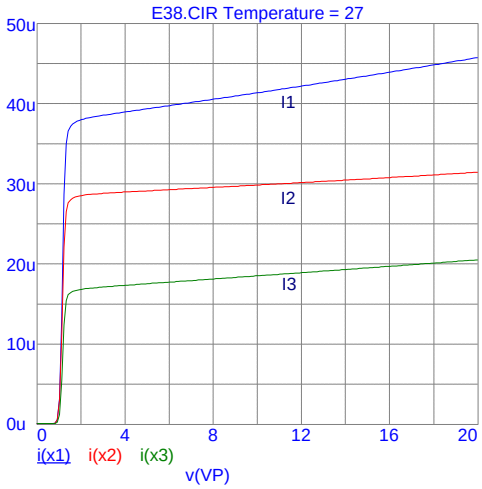


This voltage is dropped slightly by R4 (a cross-under, in Ohms) and connected to the base of Q2. As the supply voltage is increased and R1 delivers more current, the collector current of Q1 increases as well, creating a larger voltage drop across R4. Thus the voltage increase in the diode curve is counteracted by the voltage across R4 and the base of Q2 stays at a fairly constant level.

However, this scheme (sometime called gm compensation) only works over a fairly narrow current range (but it works well enough to keep the output current within a remarkably narrow range from about 8 to 20 Volts). If you want to run Q1 at a different current, you will need to re-dimension R4.

Useful current range is about 5uA to 4mA, temperature coefficient (for the values shown) is -0.25%/°C and 3-sigma variation ±27%.

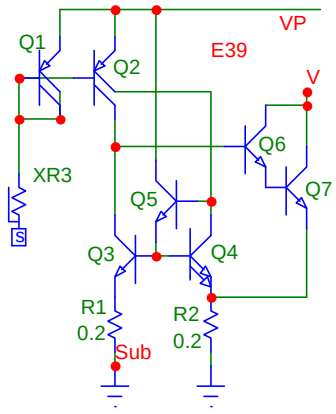


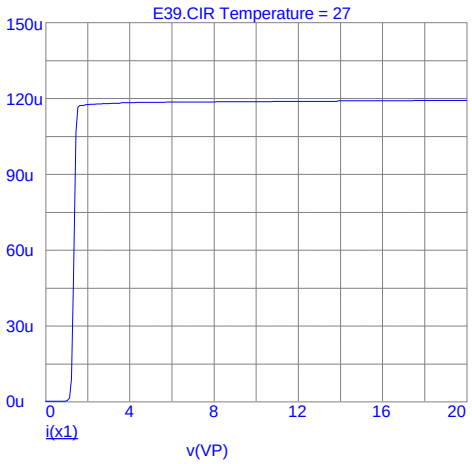


E38 is an example of how a small change in the hands of a clever designer can make a large difference. At first this circuit looks vaguely similar to one of the more sophisticated current mirrors (E13, perhaps), except that Q4 has a resistor in its emitter and there are some additional output strings. But then you notice that the bases of Q3 and Q4 are cross-coupled. As in E13 there is a feedback loop, only here it is in the form of a figure 8. It is because of this sophisticated feedback loop that the output current is substantially independent of supply voltage.

Transistors Q1 and Q4 must have two or more emitters to run at a current density lower than Q2 and Q3. This creates a delta-VBE, which appears across R2. I1 thus amounts to delta-VBE/R2, or (kT/q)ln(a)/R2. The number of emitters is represented by "a" and kT/q is 26mV at room temperature.

Additional currents can be tapped with more strings; two transistors (Q5/Q6) result in a slightly higher output impedance than just one (Q7). The temperature coefficient is +0.28%/°C at low temperature and +0.15%/°C at high temperature. 3-sigma variation is ±26% and practical current range 10 to 200uA.





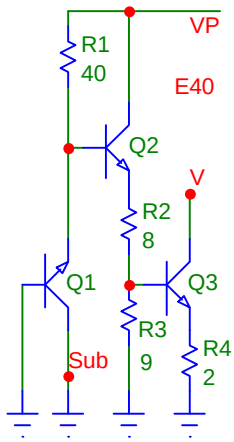
An even higher performance can be obtained with E39. Here two identical currents are generated by the collectors of Q2, mirrored from the epi pinch resistor (REP). These current varies greatly, but it doesn't matter.

Q3, Q4 and Q5 form another current mirror, out of balance by the use of two (or more) emitters in Q4. The Darlington pair Q6/Q7 provides feedback.

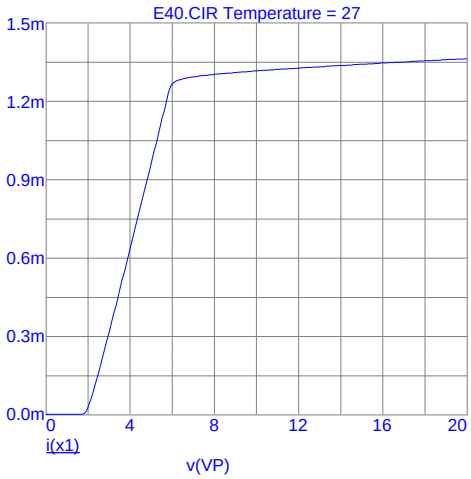
Now, the collector currents of Q3 and Q4 must be equal to match the collector currents of Q2. But they can be equal only if the output Darlington pair provides enough current to create an additional voltage drop across R2 to make up the difference in VBEs between Q3 and Q4. Therefore,I1=(kT/q)(1/R)
ln(a), where "a" represents the number of emitters in Q4 and (kT/q) amounts to 26mV at room temperature. Note that R1 and R2 must be equal (in the example 0.2 (RB) consists of five 750 Ohm resistors connected in parallel).

As shown, the output impedance of I1 is nearly 50 MOhms. The temperature coefficient is +0.25%/^{o}C at low temperature, +0.17%/oC at high temperature and the 3-sigma variation is ±26%. You can remove Q5 (and connect Q4 as a diode) and use a single transistor instead of the Darlington pair, but you get reduced performance.

Reference: George Erdi: "Starting to Like Electronics in Your Twenties", p 172, in Williams: "Analog Circuit Design", Butterworth-Heinemann, Stoneham, MA, 1991. Erdi: US Patent 4,837,496, 1989.

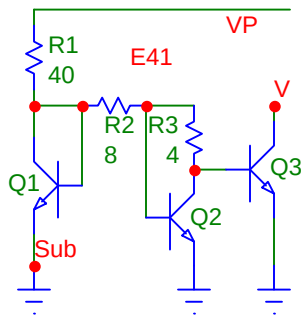


E40 provides a fairly high current with a relatively small temperature coefficient. The approach is somewhat crude, but effective: Q1 is connected as a Zener diode, providing about 5.9 Volts with a temperature coefficient near zero. At the emitter of Q2 we have approx. 5.1 Volts with a slightly positive temperature coefficient. After the voltage divider R2/R3 we are left with about 2.7 Volts, still with the same temperature coefficient. The VBE of Q3 drops another 0.7 Volts and adds its own temperature coefficient, so that the voltage across R4 amounts to 2 Volts and has a fairly strong positive temperature coefficient. The current I1 (ignoring the base current of Q3) is then simply 2 Volts divided by R4.

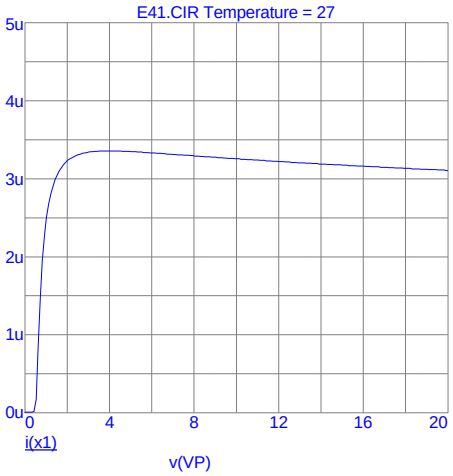


From -50^{o}C to room temperature the current rises at a steady +0.1%/^{o}C. At high temperature it gently curves downward at a rate of -0.03%/^{o}C. The 3-sigma variation is ±30%. Practical current range: 50uA to 5mA.

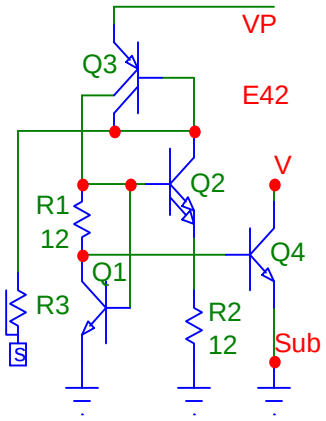
The current source elements following are especially designed to provide very low currents.



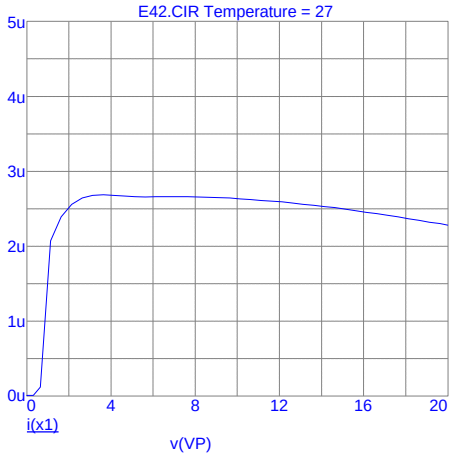
In E41 the diode-connected Q1 runs at a fairly high current set by R1. R2 delivers a much smaller current to Q2, which is connected as a diode, except that it has R3 in its collector circuit. As the supply voltage is increased the VBE of Q1 increases, delivering more current to Q2. But, as the collector current of Q2 increases, so does the voltage drop across R3. If R3 is dimensioned just right, the voltage at the base remains more or less constant, resulting in a steady and very low output current without requiring large value resistors.



The temperature coefficient of this current source is about +0.3%/^{o}C and the 3- sigma variation ±25%. You can create current in the range from about 1 to 10 uA by varying resistor values and emitter ratios.

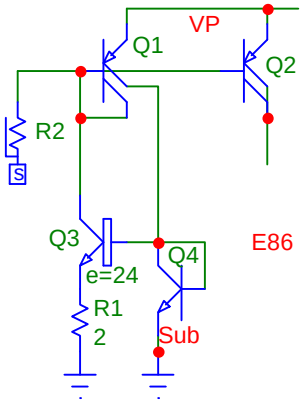


The previous current source has one disadvantage that could be serious in some applications: compared to what it produces it consumes a lot of current. E42 avoids this by the use of a feedback loop (similar to E35, but without making the PNP part a Wilson current mirror). Startup of this loop is accomplished through R3, identical to the startup of E35 and E36.

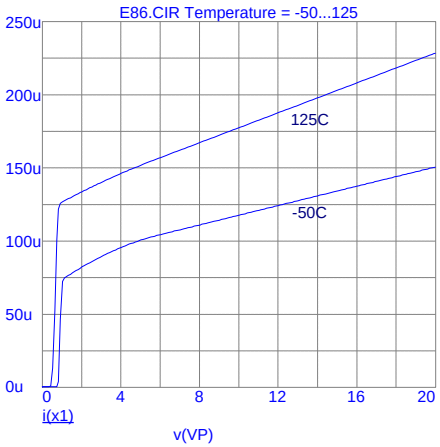


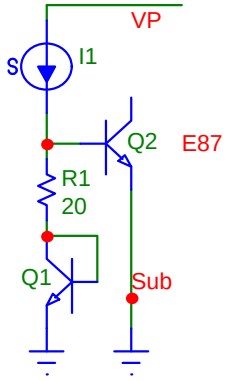
The base of the output transistor Q4 could be connected directly to the base of Q1, but we have added gm compensation with R2, similar to E37.

Temperature coefficient is about +0.3%/^{o}C, 3-sigma variation ±25% and practical current range 1 to 50uA.



Here are two additions to our current source collection. E86 uses a large NPN transistor and an epi- pinch resistor. The 24:1 emitter ratio of Q3 and Q4 creates a delta VBE of 83mV (at room temperature), which is dropped across R1. The current, therefore, has a positive temperature coefficient. REP (the epi- pinch resistor) starts the Q1/Q3/Q4 loop.





In E87 R1 and Q1 form a current sink, bypassing the leakage current of Q2. If you can spare the transistor, this is a good way to shunt the base-emitter diode because Q1 has the same temperature coefficient as the base-emitter diode of Q2 (a pinch resistor, unfortunately, has the wrong temperature coefficient).

A Note on IC Device Symbols

It has become common practice in IC design to draw the base lead right through a transistor. When you see this in a circuit diagram it means that the base of this transistor is connected to the line, even though there is no dot.

Also, two collector lines for a PNP transistor means there are two separate collectors. In the 700 Series the two collectors are equal, so each collector carries half the current.

Occasionally you will see the collector lines exit in both directions. For clarity we prefer to draw a the collectors on the same side as the emitter.

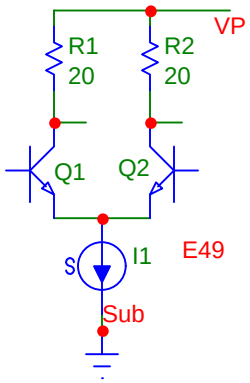
... and a Reminder

The values of all base resistors in this chapter are in multiples of the basic resistor (RB, 750 Ohms). R1 10, for example, denotes a string of 10 resistors, amounting to 7.5kOhms. R2 0.2 signifies 5 parallel-connected resistors.

Differential Stages

"Differential stage" is a slightly overblown name for a circuit which amplifies the difference of two signals. For example, in E49 the inputs could be (and often are) connected to the outputs of another differential stage. Each output has its own, independent signal with an average DC level somewhere between ground and VP. Assuming that these average DC levels are identical for the two inputs (or outputs), the difference between the two is pure AC.

In E49 the operating current, I1, is shared between the emitters of Q1 and Q2. If Vin is zero I1 splits into two equal halves. The voltage drops across the load resistors (properly regarded as load impedances) are now identical and Vout is zero also. Thus, there is no DC shift between input and output, if both are regarded as difference signals.

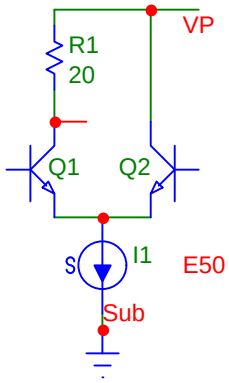


Now, increase the signal to the base of Q1 and decrease the signal to the base of Q2 by an identical amount. In this way Q1 gets more of I1, Q2 less. The left-hand output, therefore decreases, while the right-hand one increases by an identical amount. If you increase Vin more and more, Q1 will eventually get the entire I1, while Q2 will be shut off.

The changeover in current from one transistor to the other has as its basis the equation we have found in many other places: delta-VBE=kT/q [ln (I1/I2)]. This translates into a voltage gain AV. There is handy parameter: re=I(q/kT), which amounts to I/26mV at room temperature. Picture "re" as the built-in emitter resistance (i.e. the slope of the base-emitter diode), which gets smaller and smaller as the current through the transistor is increased. The figure to remember is 26 Ohms per mA. At 1mA "re" is 26 Ohms, at 10mA 2.6 Ohms, at 100uA 260 Ohms etc. The voltage gain is then simply the ratio of the two impedances, RL/re.

The larger the operating current the more gain we get. With resistors as loads, however, a larger operating current also means a larger DC voltage drop at the output and thus the achievable voltage gain remains constant, no matter how high you make the operating current. Only if the load is some other impedance (an inductance or a current source, for example), the achievable voltage gain becomes truly a function of current.

The formula for gain holds for only a very small input voltage range. Beyond about 20mV the curve deviates markedly from a straight line and the voltage gain becomes quite non-linear. The curves shown also tell you one other characteristic of the differential stage: 100% or zero current in one side is approached gradually (asymptotically), it takes a rather large input voltage to switch currents completely. Again using the formula delta-VBE = kT/q [ln (I1/I2)], we get (at room temperature) 60mV for a current ratio of 10:1, 120mV for 100:1 and 180mV for 1000:1, i.e. for every 60mV the ratio increases by a factor of 10.

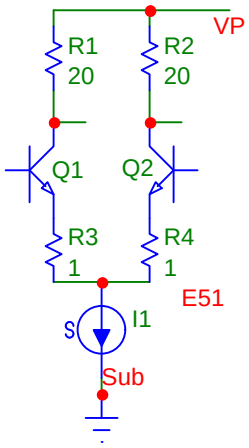


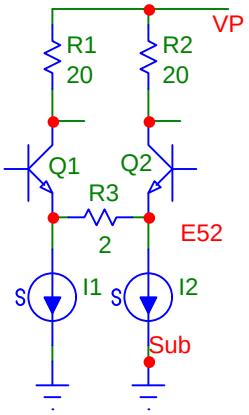
It is not necessary to take the signal differentially at the output, one output alone contains the entire signal, as shown in E50. However, there are two important differences: the voltage gain is cut in half and there is also a DC potential which has nothing to do with the input signal but is simply determined by the operating current I1, the DC resistance of RL and VP.

Note that, in any differential amplifier, only the difference in voltage between the two input terminals is important. Thus you will get the same

output if you move both terminals (i.e. one up, the other one down) or hold one steady and move only the other one.

You can improve the linearity of the differential stage by adding emitter resistors, as shown in E51. Now the current-independent RE adds to "re", which linearizes the voltage gain but reduces it too. As in E50, you can make the output "single- ended", which cuts the gain by a factor of 2 and leaves a DC bias.





E52 is an alternate connection for the differential stage with additional emitter resistance. Two current sources are used, each running at half the current compared to E51 and RE is connected between the two emitters. This circuit is, in fact, identical in performance to E51, with the resistor in E52 representing both resistors in series in E51.

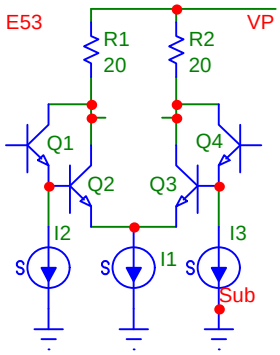
A Note on Offsets

IC components are closely matched, but the matching is never ideal. In a transistor there are two main parameters which affect matching (or mismatching): VBE and hFE. You can expect two VBEs to be within±2mV on a chip (it makes very little difference if they are neighboring devices or a few devices removed, except when you have a heat source on the chip). The current gain, hFE, matches within ±10%. As a comparison, individual diffused resistors match within ±2%; this figure goes down to about 1% if you use many of them in series or parallel.

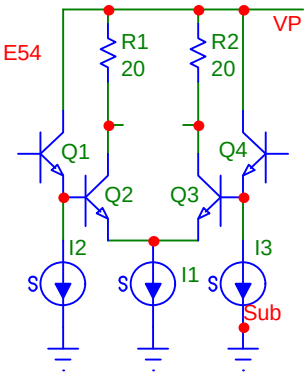
Do you get smaller offsets using emitter resistors, as in E51 and E52? Figure it this way: without them the offset voltage is a plain 2mV (3-sigma). Assuming that the VBE is 600mV, that amounts to 0.33%. Resistor matching is between 1 and 2%, which adds to the offset. Therefore, for best matching, use no resistors.

But don't forget the hFE mismatch, which can sometimes cause a larger error. Because of it the input (base) currents can be different by ±10%. If there is a large input resistance, the difference in voltage drops can easily amount to more than ±2mV.

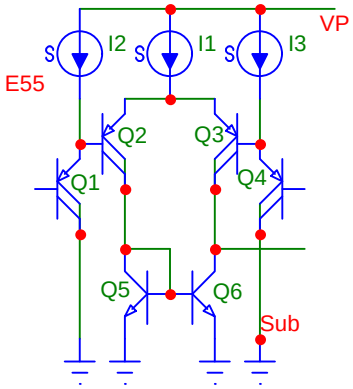
If you desire a very low input current (i.e. high input impedance) and you cannot run the differential stage at a low enough operating current, consider using a Darlington connection as shown in E53. Without the two additional (dotted) current sources the input impedance is now multiplied by another hFE. With a minimum hFE of 100 and an operating current of 1mA, the input current is down to 50nA.



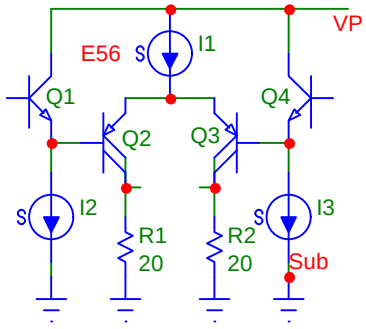
As explained in E3, the Darlington connection is a bit slow. Not only do Q1 and Q4 run at a very low current (as low as 1.7uA with an operating current of 1mA), but there is no effective path to discharge the bases of Q2 and Q3. So, for speed, you may have to use the two additional current sources I2 and I3 (perhaps 10 to 50uA), which reduces the advantage of the Darlington connection somewhat.



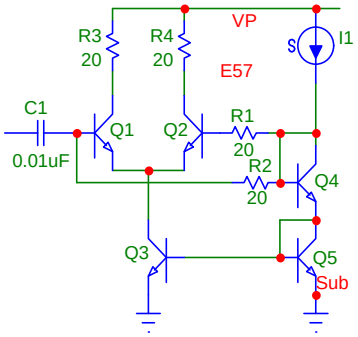
You don't need to connect the additional two transistors in true Darlington fashion. Moving their collectors to the positive supply removes a small fraction of the current (1% with an hFE of 100) but it results in a significant advantage for some applications. In the differential stage of E54 the common-mode range extends clear up to VP, though at this voltage the outputs can swing only 2 VBE (minus the saturation voltages of Q2 and Q3) below VP.



E55 is the PNP version of E54 with an "active load" (current mirror) added. With the inputs all the way down at ground the collector of Q2 has a "headroom" of 2 VBE - Vsat, more than enough for the one VBE required by Q5. The output can swing 2 VBE minus the saturation voltages of Q3 and Q6, again more than enough to connect the base of another NPN transistor.



You can also mix NPN and PNP transistors, as shown in E56. Here you drop down one (NPN) VBE to the bases of Q2 and Q3 and move up one (PNP) VBE to their emitters, which can be helpful in tight supply-voltage situations. Contrary to the previous examples, I2 and I3 are mandatory in this circuit; they are the only sources of current for the bases of Q2 and Q3 and provide an operating current for Q1 and Q2.



Our last differential stage shows a self-biased arrangement. In E57 Q5 and Q3 form a current mirror to provide the operating current, derived from I1. The additional diode-connected transistor, Q4, forms a 2 VBE bias voltage of fairly low impedance. The bases of the differential stage are connected to this point through two equal large value resistors (e.g. 10 to 50kOhms). Now an AC signal can be connected to one input through an external capacitor.

To maximize the input impedance you can expand Q1 and Q2 to Darlington connections (see compound transistors). In this case the input impedance will almost entirely be determined by R2. For small AC signals (less than about 0.5 Volts) consider using base pinch resistors for R1 and R2 (with the positive end at Q4).

A Note on Common-Mode Range and Common Mode Rejection

How high and how low can you move the input signals to a differential stage? Look at E49, the most simple example. Assume that Vin is zero, but that both inputs together are at some DC level. If you move this level down to one VBE (about 0.65V) above ground, the input transistors become cut off and there is no more voltage left for I1. So the lower end of the common- mode range is one VBE plus whatever voltage the current source I1 requires (see also Current Sources).

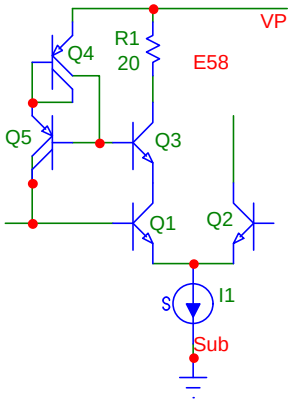
Moving the inputs (together) toward VP, the first obstacle is created by the voltage drops across RL. Assume a VP of 15 Volts, an operating current of 1mA and RL of 10kOhms. Each RL drops 5 Volts. As the inputs are moved above about 10.5 Volts, Q1 and Q2 saturate and the gain of the stage drops to near zero.

An "active load" (current mirror) greatly extends the common mode range; it requires only one or two VBE, depending on the type of current mirror used. With the Darlington-type circuits shown in E54 and E55 this extends the common mode range all the way to one supply rail, a great advantage in single-supply operation. (And remember that, even though we have drawn the diagrams with a positive supply, VP, and ground, all of them can be used with a negative supply and ground or both a positive and negative supply).

The common-mode rejection of these differential stages is almost entirely determined by the performance of the operating current source. No current source is ideal, their currents change a little bit with voltage. If you need a large common-mode rejection (and supply voltage rejection) choose a current source with a large output impedance (see Current Sources and Current Mirrors).

Input Current Compensation

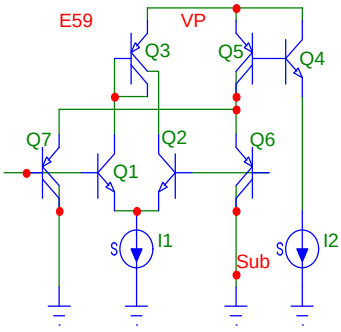
A Darlington stage is an easy way to achieve high input impedance (and low input current), but it suffers from greatly reduced frequency response. There are ways of getting nearly the same result without the loss of speed.



In E58 Q1 and Q2 are part of an ordinary differential stage. Q3 is connected in series with Q1 and thus runs at (almost) the same current. Thus Q1 and Q3 have (almost) the same base currents. The base current of Q3 is mirrored by Q4 and Q5 (this is the Wilson current mirror of E18) and fed back into the base of Q1.

In this circuit the major portion of the base current required by Q1 is provided by the current mirror, the input only has to supply the amount lost due to errors. Unfortunately the errors are fairly substantial: Q1 runs at a current higher by one base current and the collector-emitter voltages of Q1 and Q3 are quite different, Nevertheless you can expect a drop in input current by a factor of about 10.

There is also a frequency effect. The PNP transistors, slow devices right out of the starting gate, run at very low current levels (the base current of the NPN transistors) and are thus slowed down even more. Thus, the input current increases markedly beyond about 100kHz.

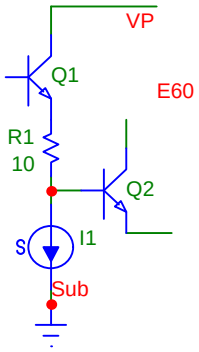


A more complex circuit with different performance is E59. Q1 and Q2 is the differential pair, Q3 is the "active load" (not essential, shown here simply as an example) and I is the operating current. A second, identical operating current I runs through Q4. The base current of Q4 is supplied by the base of Q5 which, therefore, runs at a current I, divided by hFE(NPN) and multiplied by hFE(PNP). This current splits between Q6 and Q7, so that their base currents amount to I/2hFE(NPN), exactly the same as the base currents of Q1 and Q2. Since the base current of an NPN transistor flows into the base and that of a PNP transistor out of it, the two currents cancel.

The base current cancellation of this circuit is quite accurate, but only with a very small (<10mV) input signal. As one input is moved high and the other one low, their collector currents change and so do their base currents, a fact which is lost on Q4. But for circuits which naturally have small input signals, such as operational amplifiers, the input current cancellation is much higher (a factor of about 100) than in E58. Be careful not to make I too high. The current of Q5 is at almost the same level and lateral PNP transistor cannot carry more than about 200uA without a drastic reduction in hFE.

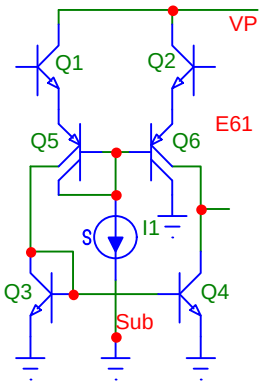
Reference: Kikuchi: "High input impedance circuit", US Patent 4,602,172, 1986

Level Shifting



When amplifying or operating on a signal you invariable move toward one supply rail in DC level and you need a scheme to shift this level in a predictable manner. E60 is such a scheme, employing an emitter follower (Q1), a resistor and a current source. The base of Q1 picks up the high DC level. At the emitter the level is one VBE lower and at the base of the following transistor the level is lower still by R1 times I1. A change in DC level at the base of Q1 (i.e. the signal) is translated in an identical change at the input of the following change because of the high impedance of the current source.

In almost all current sources (see Current Sources in this chapter) the current is determined by a resistor, which will match R1, making the voltage drop constant. The only unwanted parameter is the VBE of Q1, a small enough factor which can usually be tolerated. Assuming you make I1 reasonably high (e.g. >100uA) the circuit has hardly any loss up to at least 50MHz.



A second scheme to shift DC levels is shown in E61. Here the input stage, Q1 and Q2, works merely as a differential emitter follower. The PNP transistors Q5/Q6 pick up the signal at their emitters and deliver a level-shifted current to the active load (Q3 and Q4, the Widlar current mirror of E5). The PNP transistors have no gain, in fact they produce a small loss.

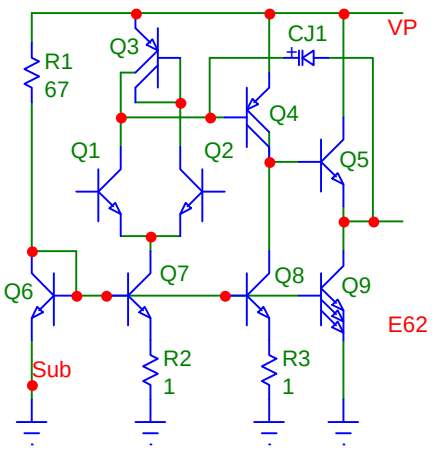
The operating current is set by I1, which flows into a diode connected PNP section and produces (more or less) identical currents in to other collectors of the PNP transistors (another Widlar current mirror). Observe that each collector of the PNP transistors carries (approximately) a current of I1. Because of the PNP transistors, you should not make I1 higher than about 100uA.

Operational Amplifiers

Before you dive headlong into the design of an operational amplifier, ask yourself the question: do I really need an op-amp to do the job. Op-amps are great discrete devices: with a few (well, lets be honest, a few hundred) basic op-amps you can cover an enormous range of applications by modifying the circuit around them. But in an IC the op-amp usually does one specific job and, while a universal building block might be convenient for the designer, it is not necessarily efficient nor does is always give the best performance. If you need a gain of 4 it might be easier, better and far more efficient to design a one-stage circuit with a gain of 4 rather than build one with a gain of 500000 and then throttle it down to 4 with feedback, using dozens of extra devices, (including compensation capacitors) and extra current.

This point is especially grave when considering frequency compensation. Operational amplifiers have a built-in tendency to oscillate. You couple an output with a very high gain back to an input. This feedback is supposed to be negative in phase, so that the output signal counteracts the input signal, which does not cause oscillation. But each stage produces delay and at some frequency the total delay is long enough for the feedback signal to still move positive when it is already intended to go negative. If there is still gain at this frequency the circuit will oscillate. The method employed to avoid this - frequency compensation - is to reduce (and control) the frequency response of the open loop gain with one or several capacitors so that the gain is less than one before the delay has caused a 180o phase shift. Such a frequency compensation is, to say the least, a tricky affair.

Operational amplifiers are a step or two beyond being mere circuit elements (they all employ several of the elements described in this chapter). So we have included only one design here, one of the most simple operational amplifiers which has the welcome feature that it is remarkably easy to compensate.



In E62 Q1 and Q2 form a differential stage with Q3 as an active load (a simple Widlar current mirror described in E5). Q4 is the second stage and Q5, an emitter follower, the third.

The operating current for the entire circuit is derived from R1 which, at 15 Volts, produces about 300uA. Q7, Q8 and Q9 are all current mirrors, slaved to Q6. The current of Q7 (the operating current for the first stage) is reduced to about 60uA by R2. The same amount of current in Q8 forms a high-impedance load for the second stage. For the last stage we usually want a higher current to pull down a load; for this design we chose to simply triple the number of emitters in Q9, making it carry about 900uA.

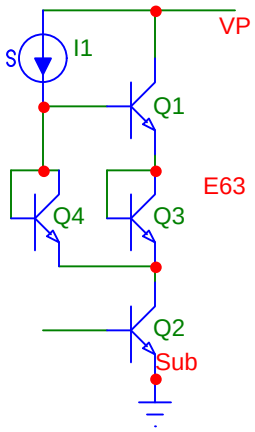
For frequency compensation you have two choices: If your supply voltage or output swing is sufficiently low, use a junction capacitor; for a full 20 Volt range you can use the collector-base junction of a large PNP transistor (base at the output).

Open loop gain of this circuit is 75dB (5600), gain-bandwidth product 9MHz and input current 60uA/2hFE, or 300nA worst case. The offset is remarkably small, owing to two features: 1) Q1 and Q2, as well as the two collectors of Q3 are operating at identical collector-emitter voltages (no Early effect), and 2) Q3 take a base current from one side of the differential pair, but Q4 takes an almost identical current from the other side (if the current of Q7 and Q8 are the same). Apart from the normal mismatch of VBEs (±2mV) there is a maximum offset of 200uV.

As shown, the power-supply rejection ratio is 78dB. You can improve this greatly by using a current source in place of R1.

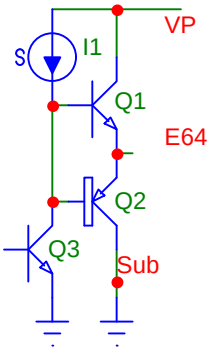
Output Stages

E63 is the so-called "totem pole" output stage, useful mostly for switching bi-directional loads. With the input signal (to the base of Q2) high, Q2 is "on" and the load current flows through the diode- connected Q3 and Q2. The collector of Q2 is now near ground and the base of Q1 is one VBE higher, i.e. sufficiently low to hold Q1 off.



With Q2 off, I1 flows into the base of Q1 and the output moves high (it can move to within one VBE of VP, plus whatever voltage drop is produced in the current source I1). The two diode-connected transistors Q3 and Q4 prevent current from flowing simultaneously in Q1 and Q2.

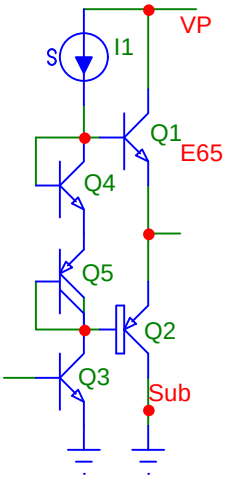
Note that three devices carry the output current: Q1, Q2 and Q3. If, for example, you need to switch 200mA, all three devices must be large NPN transistors.

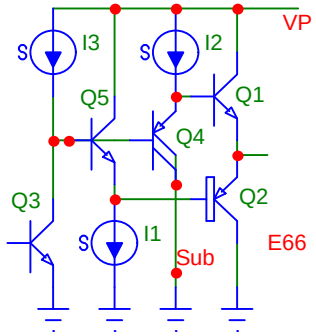


An NPN/PNP combination for an output stage is employed in E64, again useful mostly for switching. There is a "dead-band" of 2 VBE between the bases of Q1 and Q2. When Q1 is on (i.e. Q3 is off), the base-emitter junction is necessarily reverse-biased by 2 VBE. In the same way, when Q3 is "on", pulling current our of the base of Q2, the base-emitter junction of Q1 is reverse biased. In this way Q1 and Q2 can never be on at the same time, preventing a short-circuit current.

Q2 is the kind of application for which the large PNP transistor was intended, carrying a load current of up to 6mA. The circuit has two subtle advantages: 1) Q3 provides plenty of base current for the (low gain) PNP transistor and 2) Q2 cannot saturate (and thereby produce bothersome substrate currents) because its emitter is always at least one VBE above ground.

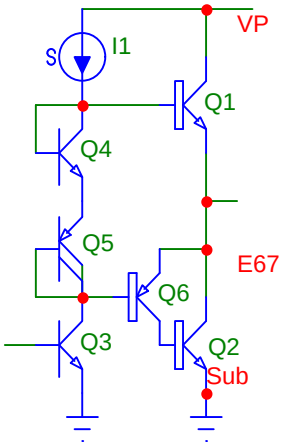
E65 is similar to E64, but the "dead-band" has been reduced to zero by the insertion of two diode connected transistors. Now there is a current flowing simultaneously through Q1 and Q2, but this current is controlled by the ratios of the emitter areas of Q4 + Q5 to those of Q1 + Q2. For example, if Q4 and Q5 are small diode-connected transistors (one NPN, the other PNP) and Q1 and Q2 are large NPN and PNP transistors, the current in the output stage will be roughly six times I1. Having such a "quiescent" or standby current makes this circuit applicable to a linear (i.e. a class B amplifier) output stage without causing undue distortion.



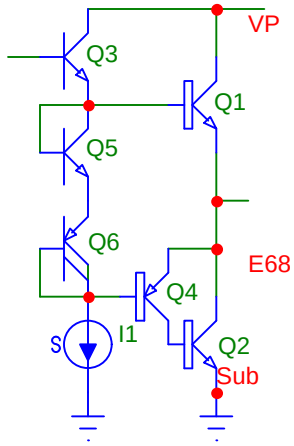


E66 also closes the "dead-band" with two VBEs, but the two transistors doing this, Q4 and Q5, are biased individually and work as emitter followers. The operation of the circuit is identical to the previous one. Its disadvantage is the requirements of two additional current sources, its advantage the gentle loading of I3 and Q3. Thus I3 can be a low current, i.e. it can be part of an amplifier stage.

An output stage similar to the two previous ones (class B or AB, no dead- band) but for a much higher current (at least 200mA) is shown in E67. Q6 and Q2 form a compound PNP transistor (the large PNP transistor was dimensioned so it could provide enough base current to a large NPN device for 200mA operation). The dead-band between the output devices is still 2 VBE, which is closed with Q4 and Q5. As in E65, the biasing of the large transistors (Q1, Q2, Q6) by the small ones (Q4 and Q5) produces a quiescent current in the output of about 6xI1.



As discussed under Compound Transistors in this section, the combination of a lateral PNP transistor (especially the large one) with an NPN makes for a rather slow device and, at high temperatures, larger than usual leakage current. This problem can be reduced with a bypass resistor or current source from the base of Q2 to ground.



Compared to E67 the circuit of E68 interchanges the places of Q3 and I1. This is usually the more convenient arrangement, considering that the combination of Q4 and Q2 has a higher gain than Q1 alone. This means that I1 can be smaller, while Q3 is able to deliver plenty of current to the base of Q1. Notice, however, that the output cannot be pulled any higher than VP minus 2VBE.

The same caution for the compound Q4/Q2 applies as in E67.

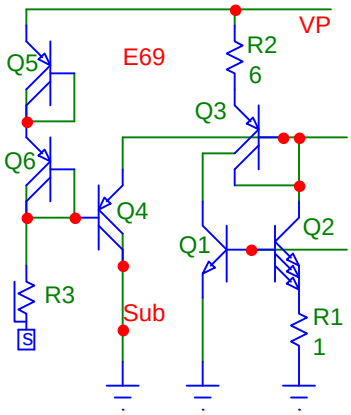
Startup

Many of the circuits discussed here feed on themselves, that is to say they produce a current which is returned by a current mirror in a feedback loop. Once such a circuit is operating it reaches a stable operating point without any problem. But there has to be something in the circuit which brings it up to this operating point. If there is no current produced in the first place, there is no current to be fed back. Unless this problem is recognized, you can end up with a circuit which sometimes starts up (e.g. when you have a high-enough leakage current) and sometimes does not.

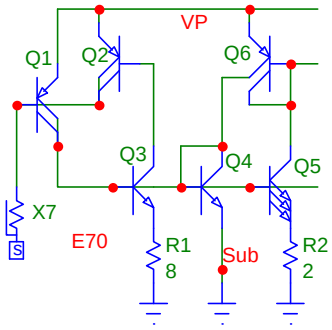
There are several methods which can start such circuits reliably. The first and most simple is to feed in a small amount of current, say produced by a pinch resistor. This will decisively start a circuit but, since the current remains, will also result in some degradation of performance. We have used this scheme in E35, E36, E42 and E79.

The second method is to use an capacitance (usually external). If you can count on a rapid rise of the power supply, a capacitive surge current can easily bring the loop over the point where it starts feeding on itself. There is no introduced error after startup, but the price is almost always and external capacitor and a pin.

The third method is shown in E69. The circuit for which we are assuring startup is the current source Q1/Q2/Q3, where Q2 feeds Q3 and Q3, in turn, feed Q1. The diode-connected transistors Q5 and Q6 are biased from an epi pinch resistor, so that they set up a voltage 2 VBE below VP. Q4 moves this voltage one VBE higher and, at startup, pulls a small amount of current out of the base of Q3. Once the loop reaches full operating current there is a voltage drop of about 340mV across R2, which cuts off Q4. The secret here is to find or create a node which moves by at least 300mV after startup and thus disconnects the loop from the startup devices.



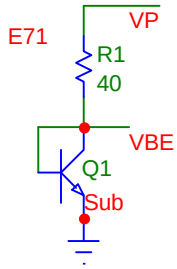
A similar scheme is used in E80.

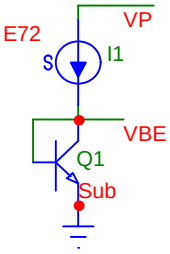


In E70, showing the fourth method to achieve reliable startup, the epi pinch resistor turns on Q1, which feeds a substantial amount of current (about 15uA) into the bases of Q3, Q4 and Q5. As the loop Q4/Q5/Q6 reaches a current level more than enough to sustain operation, Q3 turns on Q2, which shuts of Q1. The (wasted) current of the epi pinch resistor remains, but Q1's current ceases, thereby leaving the loop undisturbed at full current.

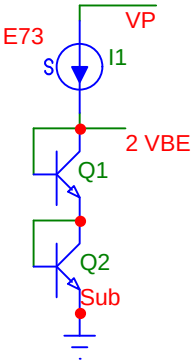
Voltage References

Let's start with the most simple reference: a VBE. It may not be very accurate (say ±10%) and it certainly has a pronounced temperature coefficient (about -0.3%/oC), but it has a reasonably low impedance (25 Ohms at 1mA, 250 Ohms at 0.1mA etc.). VBE biasing is underrated: it is in fact ideally suited to bias other transistors; it matches their absolute (base-emitter) voltage and temperature coefficient.

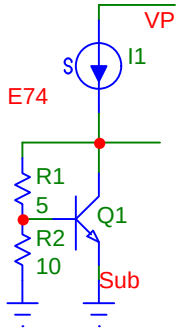




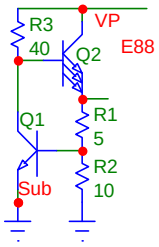
If you would like VBE biasing to be independent of the power supply, use a current source (E72). There are plenty of examples in this chapter. You can also choose a current source with a positive temperature coefficient to counteract the negative tempco of the diode.



You don't need to be satisfied with a single VBE, stack them up two (as in E73), three or four high. By the way, some designers prefer to make the two transistors a Darlington connection. We find no significant difference in performance and simple stacking is easier to interconnect during layout.

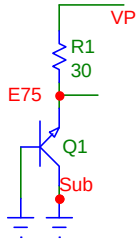


There is also no need to stick to integer VBEs. A voltage divider (E74) can produce fractions (but they need to be greater than one), such as 1.5 VBE, 1.1 VBE, 2.6 VBE etc. Note, however, the I1 will need to be large enough to feed the resistors.

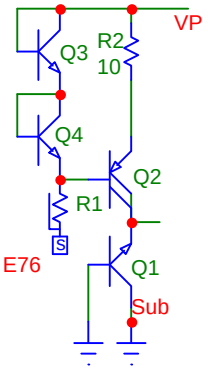


E88 shows an improved version of E74. The current for R1 and R2 is delivered by a separate transistor, Q2. This not only improves accuracy but gives the output node a low impedance. As shown the circuit delivers 1.02V (at room temperature), which drops to 0.97V with a load current of 10mA.

Contributed by Beat Seeholzer, ITR/Microswiss, Rapperswil, Switzerland.



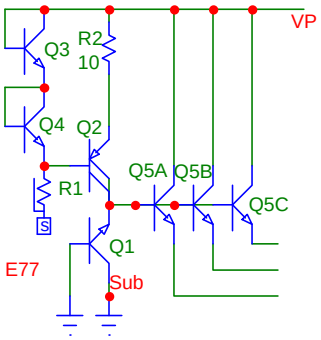
The next step up in performance and voltage is the Zener diode (E75). Any NPN transistor is also a Zener diode (the emitter/base junction in the reverse direction). Connect the (unused) collector to the base to avoid any stray effects. The 3-sigma range of Zener voltage is 5.6 to 6.2 Volts, with an average of 5.9 Volts and its temperature coefficient is near zero. The dynamic resistance of a single emitter is about 200 Ohms.



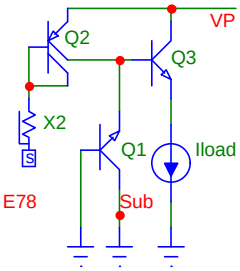
E76 shows a simple way to make the Zener voltage almost entirely independent of the supply voltage. An epi pinch resistor feeds an operating current into two diode-connected transistors, Q3 and Q4 (aha: diode-biasing). This voltage is dropped (toward VP) by the base-emitter junction of Q2 and one VBE remains across R2. Q2, therefore runs at a constant (but temperature dependent) current of VBE/R2 and feeds the Zener diode Q1 and whatever is connected to it.

With an epi pinch resistor only a small amount of current can be produced but supply rejection is excellent. For larger current use a base resistor in place of R1.

E77 is the same as the previous circuit, with the addition of Q5, an emitter follower. A much larger amount of current can now be obtained, at a voltage which one VBE lower than the Zener voltage (and now has a slight positive temperature coefficient (about +2mV/oC).



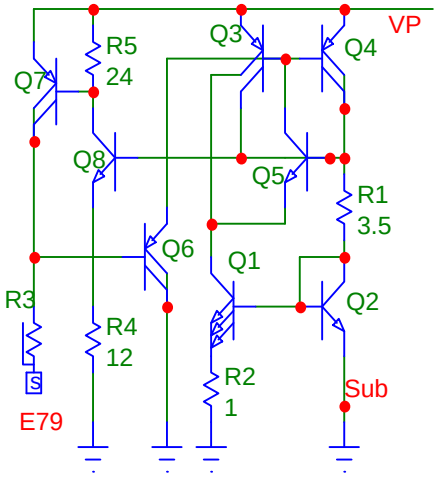
Notice the three separate emitter of Q5. If you have a noisy load it is of advantage to connect different portions of your circuit to separate emitters. In this way you get isolation, i.e. the noisy load cannot feed back into the other emitters.



A different scheme for biasing the Zener diode is shown in E78. The epi pinch resistor is really a good enough current source by itself (at the voltage required to power a Zener diode), so Q2 simply returns this current. Q1 is the Zener diode and Q3 the emitter follower.

Note that Iload is quite limited. The current of the epi-pinch resistor can be as low as 1uA and the minimum hFE of Q3 is 100, so keep the load current below 100uA.

Now we are getting a lot more sophisticated. E79 is a bandgap reference (more information on the basics of bandgap references in the first part of this chapter). Q2 is diode-connected NPN transistor with one emitter. Q1, whose base is connected to Q2, not only has three emitters but also an emitter resistor.

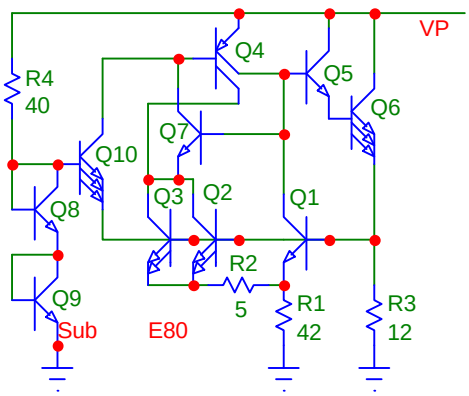


Q3 through Q5 form a current mirror similar to E19. Notice that three collectors (of Q3 and Q4) are connected together on the right-hand side. Thus the current generated by Q1 is tripled and returned (through R1) to Q2. This is one of those loops which needs to be started up (see Startup), which is accomplished by the epi pinch resistor R3 and the emitter follower Q6 pulling current out of the bases of Q3 and Q4. When the output voltage reaches about 0.9 Volts Q8 and Q7 turn on and shunt the epi pinch resistor to VP.

Now, Q2 is running at three times the current (about 300uA) compared to Q1. Also, Q1 has three times the number of emitters, so that the current density of Q2 is nine times that of Q1. This results in a delta-VBE of 57mV (at room temperature) dropped across R2. The current of Q1 (57mV/750 Ohms, or 76uA), which has a positive temperature coefficient, is tripled by the current mirror and dropped across R1.

With the correct value of R1 (i.e. the correct ratio of R2:R1) the positive temperature coefficient of the voltage across R1 cancels the negative temperature coefficient of Q1 and, at the output we have a voltage with is very nearly temperature independent.

For this bandgap reference our measurements show a 3-sigma range from 1.18 to 1.32 Volts, a temperature coefficient of 30ppm/oC and a 3mV change with a supply voltage from 2.5 to 20 Volts. This is one of the most simple bandgap references you can build. Its disadvantage is that you cannot load down the output, it is strictly a voltage reference, not a voltage regulator.



Our second bandgap reference has the advantage that some load current can be drawn. In E80 Q2 and Q3, connected in parallel, have four times the number of emitters than Q1 (here we used only the two emitters per transistor which match best) and R2 is their emitter resistor. Q4 and Q7 form a 1:1 current mirror (E19), which forces Q1 and Q2/Q3 to run at identical currents. A delta-VBE of 36mV (at room temperature) is dropped across R2 and appears in multiplied form across R1 (it is, again, the ratio of R1:R2 that counts). The voltage dropped across R1 has a positive temperature coefficient, to which is added the negative temperature coefficient of the VBE of Q1. The loop is closed by Q5 and Q6. Load current is now supplied by Q6 without disturbing the loop.

This loop, too, needs to be started up. Here we have chosen a different scheme. The diode-connected transistors Q8 and Q9 provide a potential of 2 VBE, which is reduced to one VBE at the emitters of Q10. This potential is sufficient to cause a small amount of current in the loop. Once the loop is operating at full current, the potential at the emitter of Q10 rises to 1.2 Volts and Q10 becomes completely cut off, no longer influencing the loop. The load current is limited by the amount of current the start-up circuit can provide.

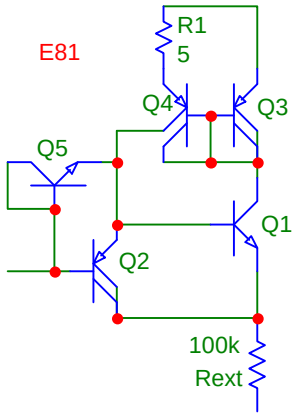
The loop needs frequency compensation, which can be accomplished with a junction capacitor between collector (+) and base of Q7. For supply voltages below 10 Volts this can a junction capacitor; for higher voltages use the collector- base junction of a large NPN transistor (collector is +).

For this circuit we measured a 3-sigma voltage range of 1.2 to 1.3 Volts, a temperature coefficient of 30ppm/^{o}C and a variation of 7mV from 4 to 20 Volts.

References: Brokaw: "A Simple Three-Terminal IC Bandgap Reference", IEEE Journal of Solid State Circuits, December 1974, pp. 388-393. Brokaw: "Solid-state regulated voltage supply", US Patent 3,887,863, 1975.

Voltage-to-Current Converters

Using an external precision resistor, you can design remarkably accurate voltage-to- current converters. We will start with the most simple (and least accurate) example.

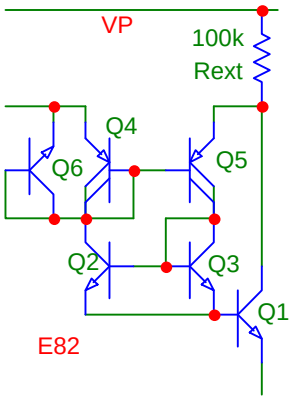


In E81 the input voltage appears between ground and a terminal. It is then increased by the VBE of Q2 and decreased by the VBE of Q1 (and here is the basic weakness of this circuit already: NPN and PNP VBEs don't match well). Thus, the voltage across Rext is roughly equal to Vin and the output current is simply Vin/Rext.

With the current mirror formed by Q3 and Q4 (a 3:1 fraction, further reduced by the emitter resistor R1) a small portion of the current is returned to feed the base of Q1. The circuit generally starts by itself because of leakage currents, but if it doesn't, the diode-connected transistor Q5 will kick it into action (but, in this case, nothing will happen until the input voltage exceeds about 1.2 Volts the first time).

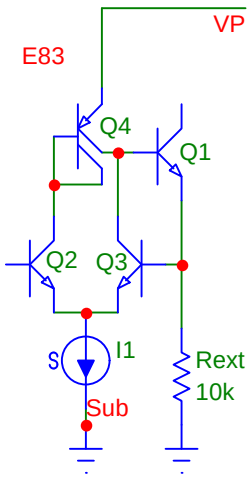
Because of the PNP transistor in the path the maximum current cannot exceed about 0.3mA. Accuracy is limited by the offset between PNP and NPN VBEs, which can be as large as 100mV (with 10 Volts full scale this amounts to a 1% error).

In E82 the input voltage is applied between VP and a terminal. Here the required matching is between two PNP transistors (Q4 and Q5), which is considerably better than E81. With a dual current mirror loop these two transistors are made to run at identical currents (to get matching). The loop runs at the base current of Q1. Again we have a potential startup problem, which is prevented by the diode connected Q6.



In this circuit half the base current for Q1 comes from Vin, which not only loads down Vin slightly but also causes a small error (you can reduce this error greatly by using a Darlington pair for Q1). On the other hand there are no PNP transistors in the path and the converter can run up to the current limit of Q1, at least 10mA.

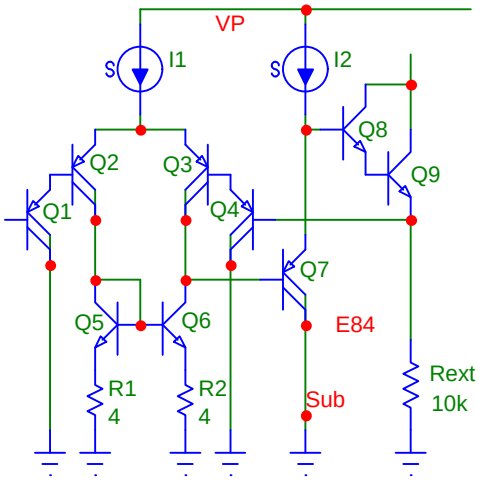
Because the current mirror transistors Q2/Q3 and Q4/Q5 have different collector- emitter voltages (they are both just simple Widlar current mirrors as shown in E5), the offset between the voltage across Rext and Vin can be as much as 15mV. This, plus the current escaping through Vin amounts to an error of 0.65% with 10 Volts full scale, 2% with 1 Volt full scale.



In E83 we adapt the voltage follower of E1 by taking the current out of the collector of Q1. With the differential pair Q2/Q3, the current mirror Q4 and the emitter follower Q1 closing the loop, the voltage across Rext is forced to match Vin and the output current is then simply Vin/Rext. Keeping I1 reasonably low (say about 5% of the maximum output current), the accuracy is a respectable 0.2% on 10 Volts full scale. The converter works up to the current limit of Q1.

You can improve the accuracy by using a Darlington pair for Q1, a more sophisticated current mirror in the place of Q4 and then set I1 very low (say 1% of the maximum output current). Notice, however, that Vin cannot go to ground unless you have a negative supply for I1.

E84 is the best- performing voltage-to-current converter and has the advantage that Vin can range clear down to ground. The modified Darlington differential stage Q1 through Q4 (see also E55) has an active load (current mirror E6 formed by Q5 and Q6). It is important not to drop more than about 300mV across R1 and R2 to prevent Q6 from saturating. With Vin at ground the collector voltages of Q5 and Q6 are at one VBE, which is ideal for matching and provides sufficient headroom for Q2 and Q3 (their emitters are at 2 VBE at this point). The output of the active load is buffered and moved up one VBE by Q7. The Darlington emitter follower then closes the loop.

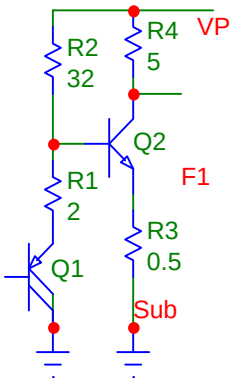


Functions

Amplifiers

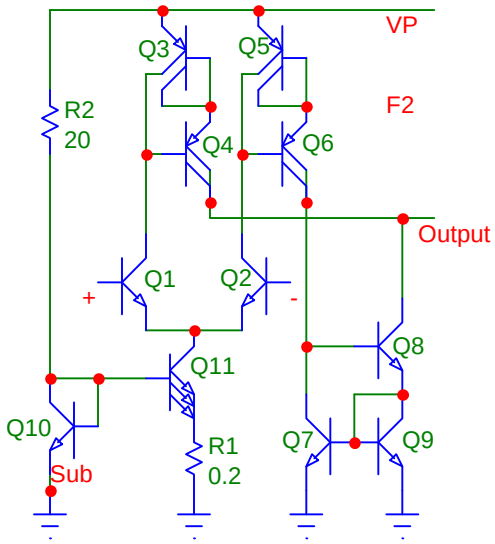
F1 Fixed Gain AC Amplifier

The input of this amplifier must be connected to a device with a DC path, such as a dynamic microphone, with the second terminal connected to ground. No input coupling capacitor is required. The base of Q1 is at DC ground and its emitter one VBE above it. A voltage drop across RB2 is created by the resistor ratio RB1/RB2. Subtracting the VBE of Q2, this voltage drop is reproduced across RB4, thus setting the operating current of Q2. RB3 is dimensioned to create a DC voltage drop of approx. half the supply voltage.



The gain of the amplifier is given by the ratio of RB3 to RB4 (plus re), minus the ratio of the first divider RB2/(RB1+RB2), or 18.6dB with the values shown. Gain variation from -55C to 125C is 0.3dB and the three-sigma variation ±0.4dB. The -3dB frequency is 12MHz; adding a cascode stage to Q2 brings this up to 18MHz.

F2 Transconductance Amplifier



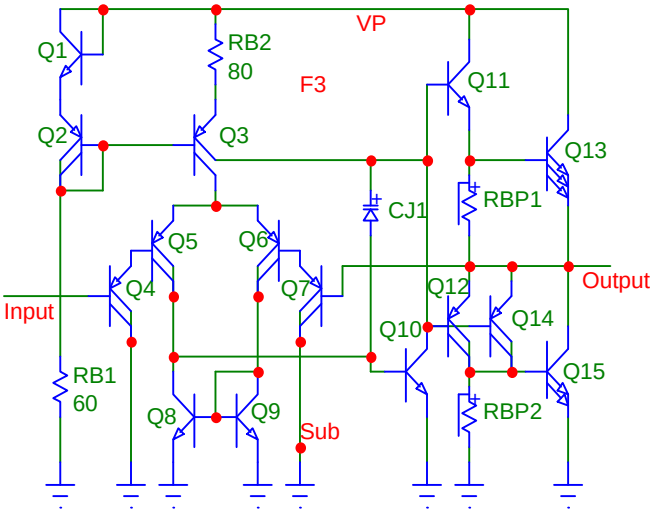
The differential stage Q1/Q2 is used to convert the input voltage to current (see E49); it is linear only over a small voltage range, less than ±50mV.

The left-hand current is mirrored once, by Q3/Q4, while the right-hand current is mirrored twice (the Wilson current mirrors Q5/Q6 and Q7/Q8/Q9) and subtracted from the left-hand current at the output. Thus, with zero input voltage, the two currents cancel.

RB1 sets the full-scale current, which must not exceed the capability of the PNP transistors (about 100uA). The current source Q10/Q11/RB2 is particularly well suited for this application; its positive temperature coefficient (see E27 and E28) compensates for the temperature dependence of gm, the transconductance of the differential stage.

Frequency range is approx. 20MHz (there is some peaking at 9MHz). Change from -55C to 125C is -2dB and the three-sigma variation ±0.5dB.

F3 Buffer



Q4 through Q7 form a Darlington differential pair with the conventional active load (current mirror Q8/Q9). Q10 is the second stage and Q11 through Q15 are connected as a class B output stage.

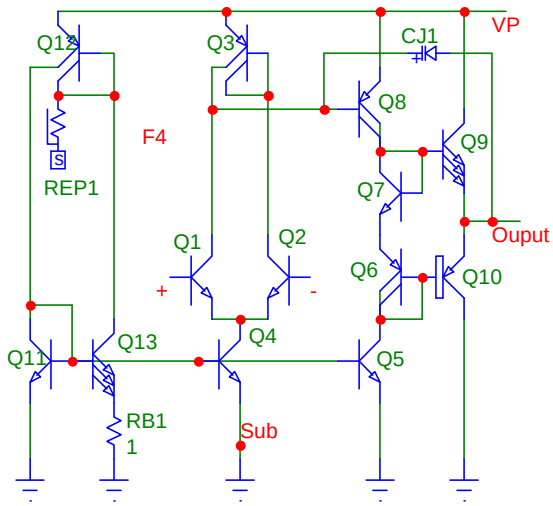
Thanks to the Darlington configuration, the input current is very low: with a bias current of 5uA (as shown) the worst-case base current of Q4 is 5uA/40/40, or about 3nA.

The particular bias scheme chosen is VBE-dependent, which results in a strong negative temperature coefficient. Combined with the positive temperature coefficient of hFE, this results in a fairly steady maximum current limit for the upper half of the output stage (30mA at -55C, 45mA at 125C). The parallel connection of Q12/Q14 makes the current limit of the lower half almost identical.

With a junction capacitor the output swing is limited the about 10 Volts. If you need a larger voltage, use a large NPN transistor (base +, collector -) instead or an external 30pF capacitor.

F4 Operational Amplifier

The biasing of this amplifier is a simplified E35 current source and we set the current (of Q4 and Q5) at about 90uA. This results in a worst-case input current of about 0.5uA. Common-mode range is within 1 Volt of supply rails.



The second stage consists of Q8, which feeds a class B output stage: Q6 and Q7 bias Q9 and Q10 (a current ratio of about 1:3). The output stage is capable of sourcing 20mA and sinking 6mA.

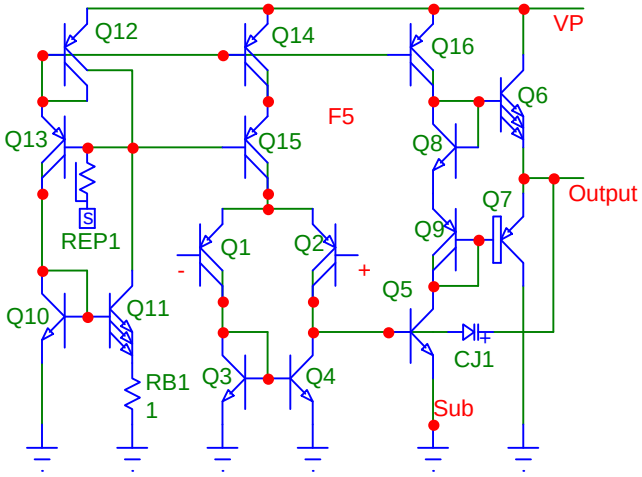
Open loop gain is 88dB and ft 10MHz. With an internal compensation capacitor (a JCAP up to 10V supply, a large NPN for higher supply voltages) the amplifier is just barely stable at unity gain. Note, however, that such an amplifier need not be unity-gain stable if you intend to use it only at a certain fixed and higher gain.

F5 Operational Amplifier

A PNP equivalent of F4. Here we use a slightly better biasing current source and cascode the biasing for the input stage (Q15) to get maximum common-mode rejection. With 90uA of bias current the worst- case input current is just over 1uA.

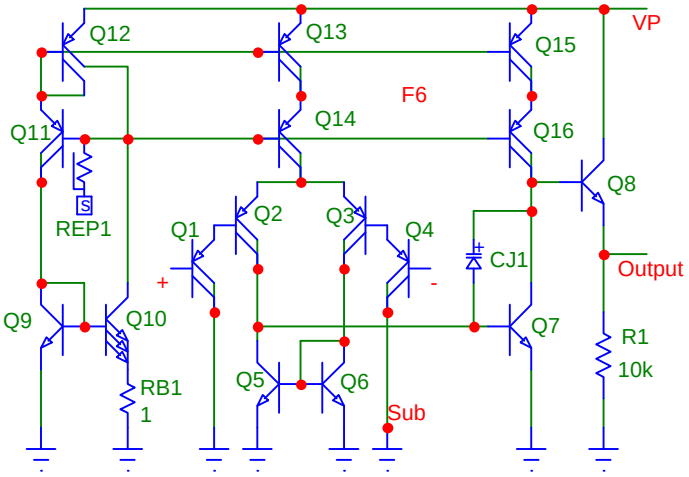
Common-mode range (with a 15V supply) is 1V to 13.5V. The supply voltage can vary from 2.5 to 20 Volts.

Open loop gain is 95dB, falling off to 0dB at 20MHz. The same consideration as in F4 applies: with an internal compensation capacitor (a JCAP up to 9V supply, a large NPN for higher supply voltages) the amplifier is just barely stable at unity gain. Note, however, that such an amplifier need not be unity-gain stable if you intend to use it only at a certain fixed and higher gain.



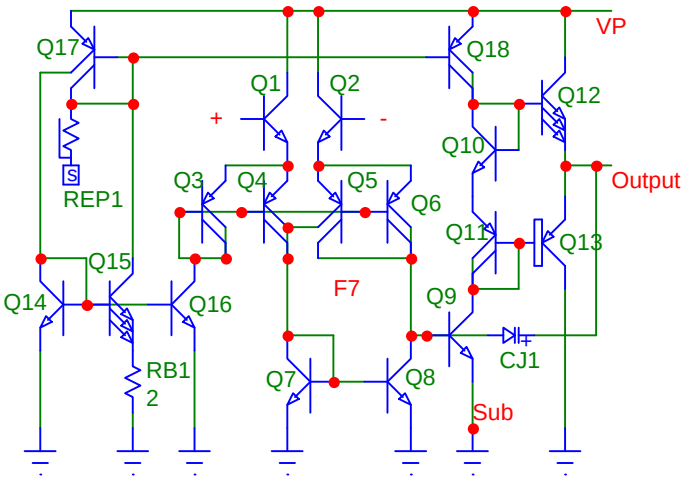
F6 Operational Amplifier

A circuit almost identical to F5, except we upgraded the input stage to a Darlington configuration and made the output stage a simple emitter follower. Now the common-mode range extends all the way to ground and the input current is 30nA worst case.



The power supply rejection ratio is -130dB due to the high-performance current source and the cascode stages Q14 and Q16. Open loop gain is 100dB, dropping to 0dB at 8MHz. See compensation notes for F4. For operation above 9V and with minimum closed-loop gain of 100 we recommend that you compensate this circuit by simply using a large NPN for Q7, eliminating CJ1.

F7 Operational Amplifier

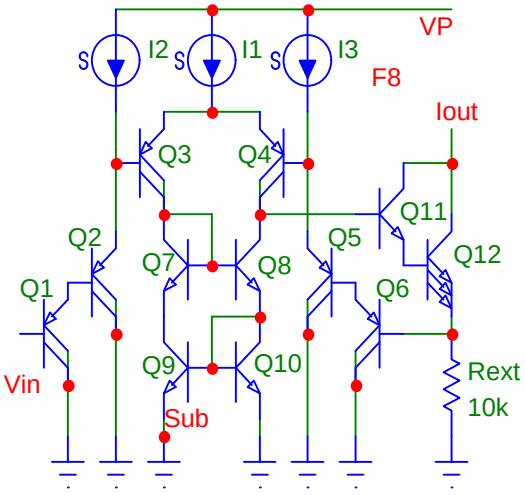


Here we took parts of the F4 and F5 circuits and gave the combination a different input stage. Q1 and Q2 are merely emitter followers and the four PNP transistors Q3 through Q6 act as a grounded-base differential stage. The biasing of Q4/Q5/Q6 (and, therefore, Q1/Q2) comes from the diode-connected Q3. Q16 pulls approx. 20uA out of Q3. This current is duplicated in Q4, Q5 and Q6, for a total of 80uA, or 40uA in each differential half. Three PNP collectors each feed Q7 and Q8 with 30uA. Thus, the PNP transistors actually produce a loss (the gain is 0.75), but the signal is shifted to ground and the notoriously slow PNP transistors are operated in their fastest connection.

Read the compensation notes for F4. No compensation capacitor is required for a closed-loop gain of 40dB or higher, in which case the open loop gain of 92dB drops to 0dB at 15Mhz.

Converters

F8 Voltage to Current Converter



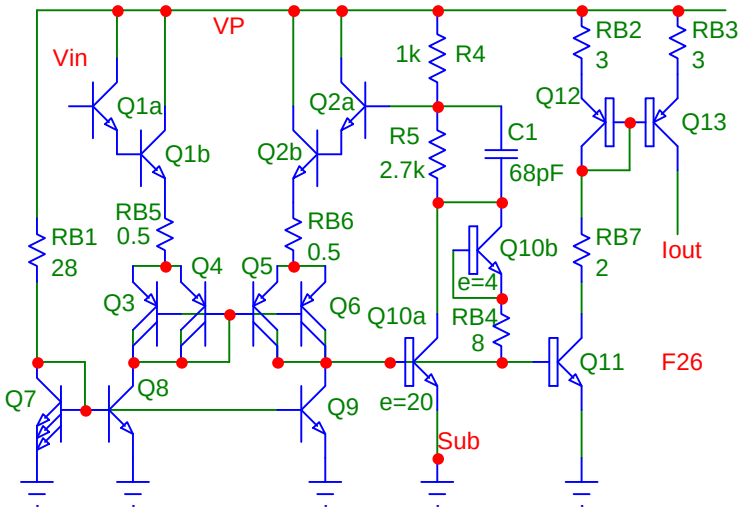
Q1/Q2 form a Darlington buffer, designed to minimize the input current and give the following stage some operating headroom. An identical buffer (Q5/Q6) measures the voltage across the external resistor used to set the current. These two voltages, buffered and elevated in DC level, are compared by the differential stage (Q3/Q4, with active load Q7 to Q10) and the difference is applied to the Darlington output pair Q11/Q12.

We haven't drawn in the details for the biasing currents, because they are not critical. I1 must have a high enough level so that the base current drawn by Q11 is comparatively small. I2 and I3 must be identical and large enough to safely operate the input transistors.

With a 10V/10mA range we measured an accuracy of ±0.07%.

F26 Voltage to Current Converter

In this converter an input voltage referenced to the positive supply causes a current to be sourced toward ground. The Darlington transistors Q1 and Q2 form an input buffer, steering the emitters of Q3/Q4 and Q5/Q6. The operating current for the input stage is set by RB1 and Q7; one-third of this current is used by Q8 to operate the diode-connected Q3/Q4. This current is mirrored by Q5/Q6 and the opposed by the collector current of Q9. Ignoring base currents, the collector currents of Q5/Q6 and Q9 cancel when the input pair is balanced.



Any current difference is picked up and amplified by Q10a. The collector current of Q10a creates a voltage drop across R1. With the feedback loop closed, this voltage equals the input signal. R2 reduces and regulates the loop gain and C1 is the main compensating capacitor. Additional compensation is achieved with Q10b (2 bases and 4 emitters in the same large NPN transistor as Q10a), in series with RB4.

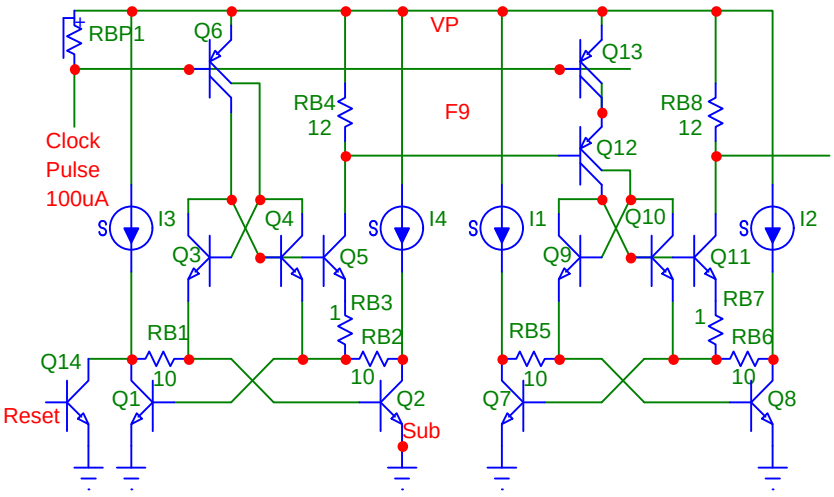
The base-emitter junction of Q11 is in parallel to that of Q10a, which duplicates the feedback current. This current is then mirrored by Q12 and Q13. Here emitter resistors improve accuracy; Their values have been chosen for a maximum output current of 1mA (2.25 Volts nominal voltage drop). Moving the DC level out the output by 1 Volt changes the current by 0.33%.

Notice the 20:21 emitter ratio of Q10a and Q11. This 5% difference is used to compensate for base-current errors.

With a 5 Volt supply, the maximum input voltage is a little over 1 Volt. Response time is on the order of 150nsec.

(Contributed by Garry Brown, Rover Group, Coventry, UK)

F9 Frequency Divider



There are two almost identical stages shown here, representing the first two stages of a chain. Each stage has two flip-flops: the first, Q1/Q2 (Q7/Q8), is permanently powered through current sources. The second, Q3,Q4 (Q9/Q10), is only powered for a brief time.

Assume the reset signal (high at the base of Q14) has briefly lowered the collector potential of Q1 so that Q1 is on and Q2 is off. Now a clock pulse arrives at the base of Q6, which supplies power to the flip-flop Q3/Q4. Since the base of Q2 is lower by a VBE than that of Q1, Q3 turns on first, which sets the state of the second flip-flop. Rising in potential, Q3 now forces Q2 to turn on, which turns off Q1. Thus, with each clock pulse, the Q1/Q2 flip-flop changes states.

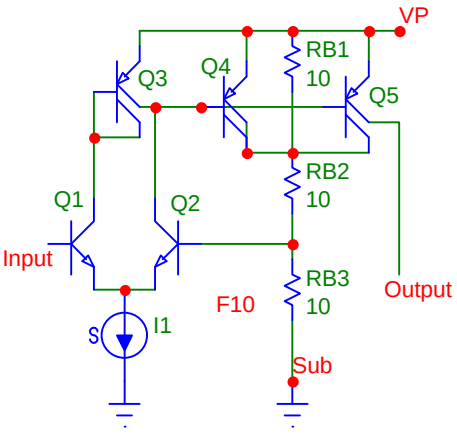
During the clock pulse the state of the Q1/Q2 flip-flop is also sensed by Q5 and, at every second clock pulse, Q5 turns on Q12. This signal is ANDed with the clock pulse (Q13) and fed to the second stage. (It is important to use RB4. There is a small signal produced by Q5 during all clock pulses, which is weeded out by the resistor).

You can set or reset each stage as shown by Q14. Note that Q1 and Q14 have a common collector; they can be combined into one transistor on the IC.

The 10uA operating currents are arbitrary; this circuit works well over a wide current range and with supply voltages from 2.5 to 20 Volts. Should you change the operating current, make the clock pulse at least five times the operating current and change the resistor values proportionally.

Hysteresis

F10 Schmitt Trigger



Assume that the resistor string consists of three identical resistors and that the input voltage is high. The voltage at the base of Q2 is then one-third of VP. Now lower the input voltage gradually. At one- third VP Q1 starts turning off and Q2 turns on. When the current in Q2 is high enough, Q4 and Q5 turn on, moving the upper point of the divider to near VP. The base of Q2 is, therefore, no longer at one-third VP but at one-half VP and Q1 is suddenly and completely turned off. To turn

Q1 back on, you now have to increase the input voltage to one-half VP, at which point the procedure reverses and the voltage at the base of Q2 snaps back to one-third VP.

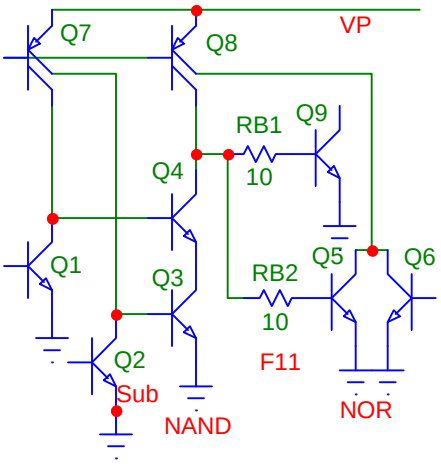
The switching points are:

Lower:	VP x RB3 / (RB1 + RB2 + RB3)
Upper:	VP x RB3 / (RB2 + RB3)

However, there is some error produced by the saturation voltage of the PNP transistors. A computer analysis will bring this out nicely and lets you set the resistor values accurately.

Logic Functions

F11 Gates



Connecting NPN (or PNP) transistors in series or parallel results in simple logic functions. These bipolar schemes are not nearly as area efficient or speedy as equivalent CMOS designs, but they serve well in the incorporation of a minor amount of logic in a predominantly analog IC.

The series connection of Q3 and Q4 forms a NAND gate: only if the bases of both transistors are high (Q1 and Q2 off) is the collector of Q8 low.

The parallel configuration of Q5/Q6 is a natural NOR gate: a high potential at either base brings the collectors low. Note that Q5 and Q6 have a common collector; on the IC they can be combined into one transistor.

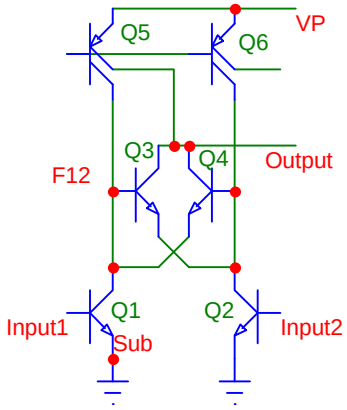
If the output of a gate feeds more than one base, you must use balancing resistors (RB1, RB2) to distribute the current evenly. Dimension them so that their voltage drop is at least 50mV.

You can power such logic circuits either from current sources as shown (Q7 and Q8) or through resistors. Current sources tend to take up less space at low operating currents.

F12 Exclusive OR Gate

Picture input 1 high, so that the collector of Q1 is near ground. One collector current of Q6 (a current source) flows into the base of Q4, turning it on. The output, therefore, is low.
Now, if input 2 goes high too, the collector of Q2 moves low as well, which cuts off both Q3 and Q4 and moves the output high.

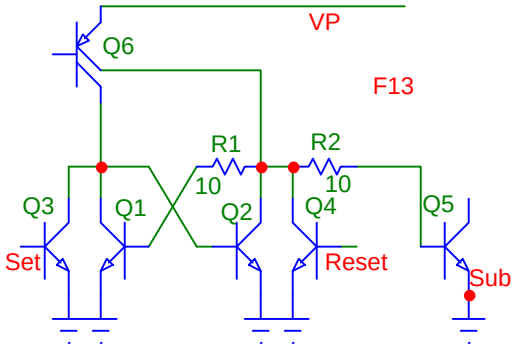
You can use either current sources to power this circuit (Q5 and Q6 - just about any of the PNP current sources listed earlier in this chapter will work) or resistors. Current sources take up less space at low operating currents.



This is an ideal circuit to optionally invert a signal. Remove Q2 and put a pin in its place. With the pin shorted to ground, the propagation from input 1 to the output is normal. With the pin unconnected, the propagation is inverted.

Note that Q3 and Q4 have a common collector. On the IC they can be combined into one device.

F13 RS Flip-Flop



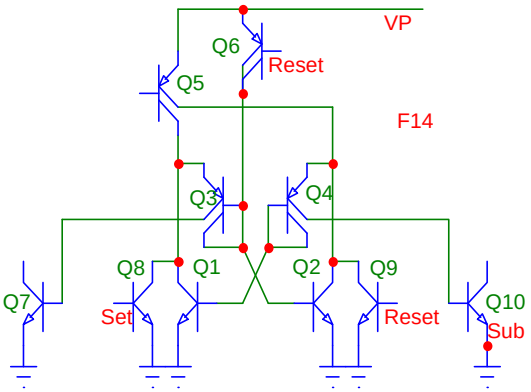
Q1 and Q2 form the basic flip-flop. Only one device can be on at any given time, the base of the other one is held near ground. Set or reset is accomplished by temporarily moving the high collector low.

In this scheme it is essential to use the balancing resistors RB1 and RB2. Anytime two bases are connected to the same node and either transistor can saturate, you must drop at least 50mV across a base resistor to balance the currents delivered to the bases.

Q6 is intended the represent a current source (any PNP current source discussed earlier in this chapter). For reasonably high operating currents resistors work equally well.

F14 RS Flip-Flop

A slight improvement over F13, especially for low operating currents. Instead of using balancing resistors, the base currents are split by the PNP transistors Q3 and Q4, delivering 50% to each base.

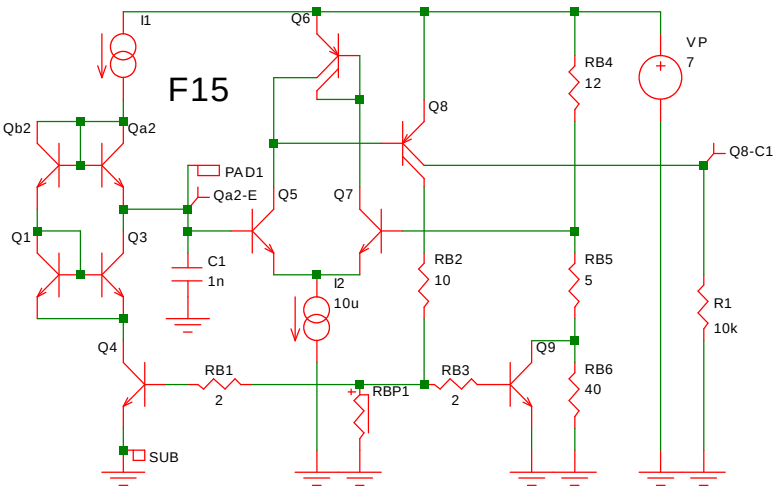


Since the bases of the basic flip-flop (Q1/Q2) are no longer shunted directly by the opposite collector, there is now always a high impedance at this point. This means you can set or reset each side by either shunting a collector to ground (Q8, Q9) or feeding a small amount of current directly into the base (Q6).

Q5 is intended to represent a current source (any PNP current source discussed earlier in this chapter). For reasonably high operating currents resistors work equally well.

Oscillators

F15 Square-Wave Oscillator



The circuitry to the right of the capacitor forms a Schmitt trigger. Start with the voltage divider RB4/RB5/RB6 and assume the resistors to be equal for now. Thus the voltage at the base of Q7 is 2/3 of VP. Now picture the voltage across the capacitor increasing gradually until it reaches 2/3 of VP. At this point Q5, Q8 and Q9 all turn on and Q9 shunts RB6 to ground.

The reason the voltage across the capacitor moved high was the current I1, which charged the capacitor through the diode Q2a. Simultaneously with Q9 (note the balancing resistors RB1 and RB3) Q4 turns on. This shunts I1 to Q2b and Q1 (Q2a becomes reverse-biased) and produces an identical current in Q3, which discharges Cext. The voltage across the capacitor now gradually (and linearly) drops until it reaches the lower threshold produced by the shunting of RB6. At this point the cycle starts again.

The two thresholds are:

Upper: VP (RB5 + RB6) / (RB4 + RB5 + RB6)
Lower: VP (RB5) / (RB4 + RB5)

The indicated resistor values (in RB, the basic 750 Ohm resistor) are only suggestions.

You have two choices concerning I1. Make it proportional to VP and the frequency will be independent of the supply voltage. Or use a voltage-to-current converter and regulate VP, thereby creating a voltage- controlled oscillator.

Keep I2 as low as possible (we suggest 10uA).

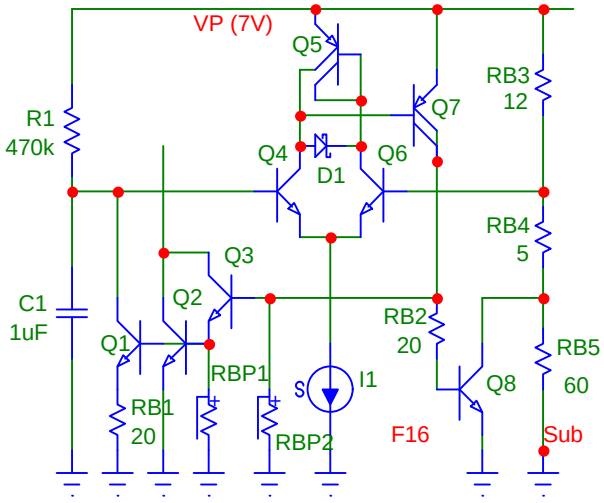
F16 Pulse Generator

The right-hand part of the circuit is similar to F15. We have added a Schottky diode (D1) to make sure Q7 cannot be cut off. A resistor in series with the collector of Q7 is no longer necessary: Q7 is going to be turned on for only short periods.

What is different is the charging and discharging of the capacitor. First, we use a resistor directly for the charging; the waveform (exponential rather than linear) is not important.

Second, we discharge the capacitor through Q1, an independent (and less accurate) current sink. Since the duty cycle is short (say less than 10%), the accuracy of the discharge current is less important.

During the discharge of the capacitor we also turn on the Darlington transistor Q2/Q3, which powers an LED. Don't be fooled by the fact that we haven't used a large NPN transistor here or at least one with more emitters. In the Darlington configuration this little device can supply a large amount of current, at least 50mA. In addition, Q2 and Q3 can be combined into one device on the IC since they have a common collector.

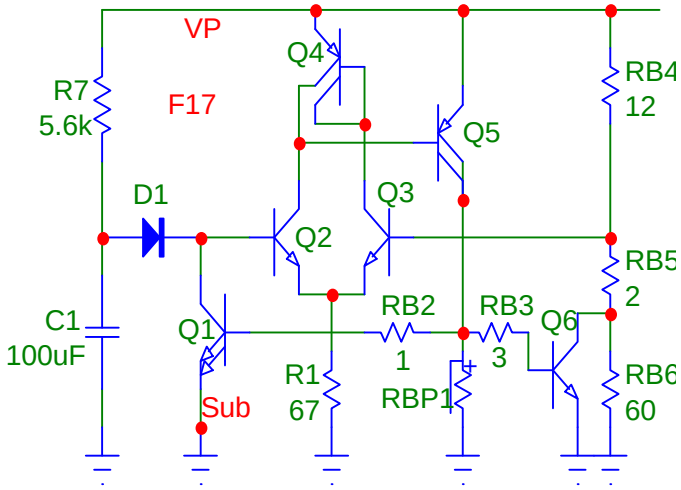


The supply voltage can range from 2.5 to 20 Volts if you adjust the LED series resistor. Make I1 about 50uA; you can also use a resistor here, perhaps with a base-pinch resistor in series.

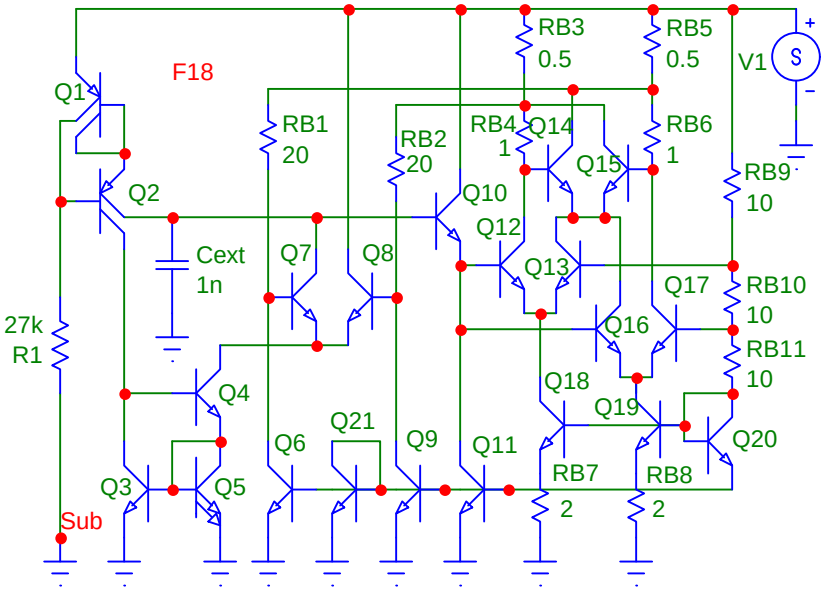
F17 Single Pin LED Flasher

This is a refinement (or specialization) of F16. Instead of having a separate pin for the LED, we power it with the discharge current of the capacitor. This not only saves a pin, it also reduces current consumption. The duty cycle naturally is very short, but the flash is very visible.

You need to use a large capacitor and there is a small error produced by the fact that the LED is between the capacitor and the threshold voltage (a 5% change in frequency from 5 to 20 Volts VP).



F18 Triangle Wave Oscillator



Here is a considerably more complex oscillator, which operates at up to 20MHz. The high speed is achieved by using only NPN transistors in the switching path and letting none of them saturate.

Lets start with the current source. The external resistor produces a current for the Wilson current mirror Q1/Q2. The mirrored current is split into two equal parts in Q2. On half directly charges Cext, the other half is mirrored and doubled in the Wilson current mirror Q3/Q4/Q5 (note the two emitters in Q5). This current is then switched by Q7/Q8. If Q8 is on, the capacitor is charged; with Q7 on there is a net (and equal) discharge current.

The voltage across the capacitor is sensed and buffered by Q10 and fed to the inputs of two comparators: Q12/Q13 sense at the upper threshold, Q16/Q17 at the lower one. The threshold voltages for the comparators come from the divider RB9/RB10/RB11.

The output currents of the comparators steer the flip-flop Q14/Q15 and the voltage drops across resistors RB3 to RB6 are small enough so that the flip-flop does not saturate.

Now comes the tricky part. We must drop the DC level of the flip-flop signal down to where Q7 and Q8 can handle it, below the capacitor voltage. This is accomplished with the resistors RB1 and RB2 and the current sources Q6 and Q9 (see E60).

There are three features in this circuit which combine to make it work accurately over the military temperature range and with supply voltages from 5 to 20 Volts:

1) The voltage applied to both the current source resistor (Rext) and the threshold resistor string (RB9/RB10/RB11) is VP - 2VBE, which cancels out the effect of both the VBEs and changes in supply voltage.

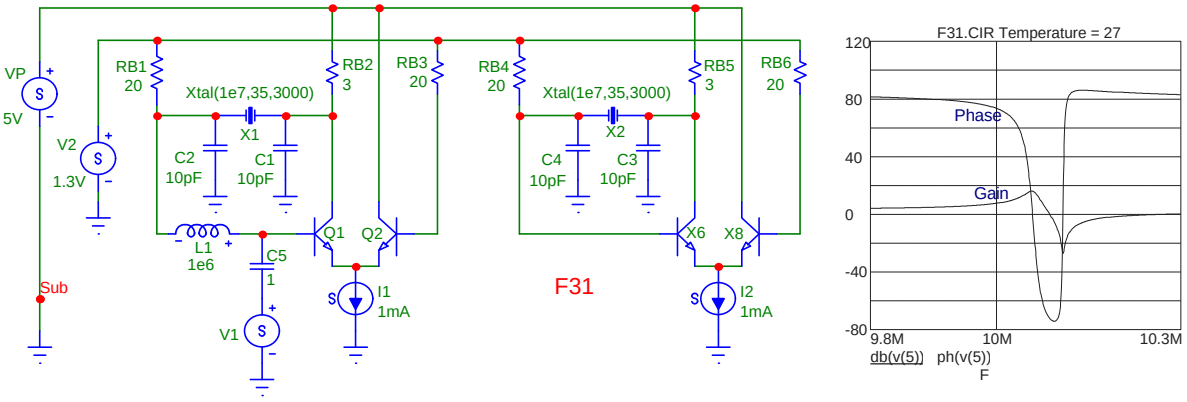
2) The flip-flop is powered from VBE-dependent currents (Q18/Q19, RB7/RB8) which allows its collector to move a fraction of a VBE at any temperature.

3) The level-shifting scheme operates with current sources which are dependent on the supply voltage (minus 2VBE) and resistors only.

The triangle wave output moves between the voltage levels indicated in the schematic with A and B (1/3 of VP - 2VBE). Thus, if you use a comparator with one input at the triangle wave and the other at a tap of RB10, you get a rectangle output. With a center tap on RB10 this will be an

exact square wave. Moving the tap higher or lower produces a widely varying duty cycle without any effect on the frequency.

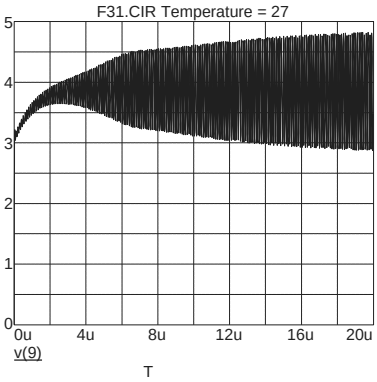
F31 Crystal Sine-Wave Oscillator



A crystal changes phase abruptly by 180 degrees at resonance. It can be made to oscillate by placing it in the feedback path of an amplifier. Either positive or negative feedback works, but approx. 90 degrees of additional phase-shift must be provided. The circuit will then oscillate at the frequency where the loop-phase is zero degrees and the loop-gain above one.

We have shown the circuit twice; on the right L1 , C5 and Vac have been inserted so that the gain and phase of the loop can be measured (at C4). This is quite important to assure reliable operation as part of the phase-shift (at high frequency) is caused by the transistors.

The amplitude of the sine-wave is determined by I1 and RB2. V2, the biasing for the bases, is kept low enough so that Q1 does not saturate.

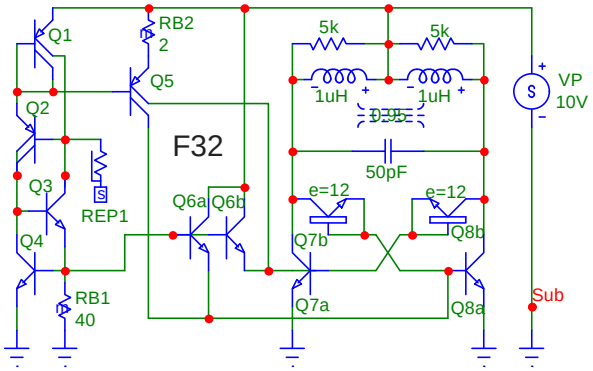


Simulating crystal oscillators can be frustrating and time-consuming. Crystals can have a very high Q (e.g. 30000), which means that they take a long time to start up - up to one second. To catch the actual start (at a minute level), the maximum time step must be kept very small. This translates into long simulation times and you can easily get the impression that your circuit is not working because nothing appears to be happening.

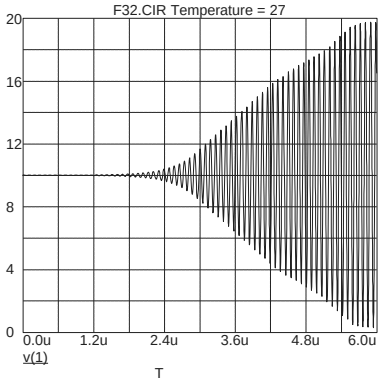
F32 L-C Oscillator

Here two opposite-phase 12MHz sine-waves are produced, each with a peak-to peak voltage of twice the supply voltage. The inductor is center-tapped, represented by two mutually coupled (0.95) coils. The two 5kOhm resistors adjust the Q (about 50).

Q7 and Q8 are large NPN transistors, but only one section each (a) is used as an active device, while the base-collector capacitances of 12 sections (b) provide positive feedback. In this way only these two transistors are exposed to high voltage and the N-layer of the chip can be biased at VP. 12 sections represents an optimum for distortion and current consumption at 12MHz.

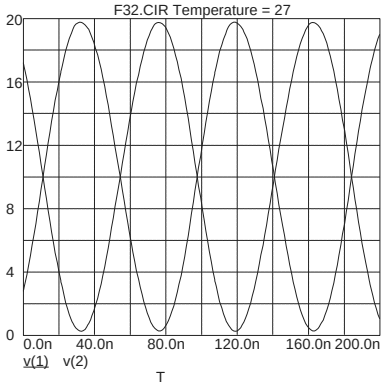


Q5 injects a small bias current into the bases of the oscillating transistors, to place them in their linear operating region for the initial buildup of the oscillation. These currents are derived from current source Q1-Q4, which is the circuit E36. Q6a and Q6b catch the negative swing of the coupling capacitors Q7b and Q8B.



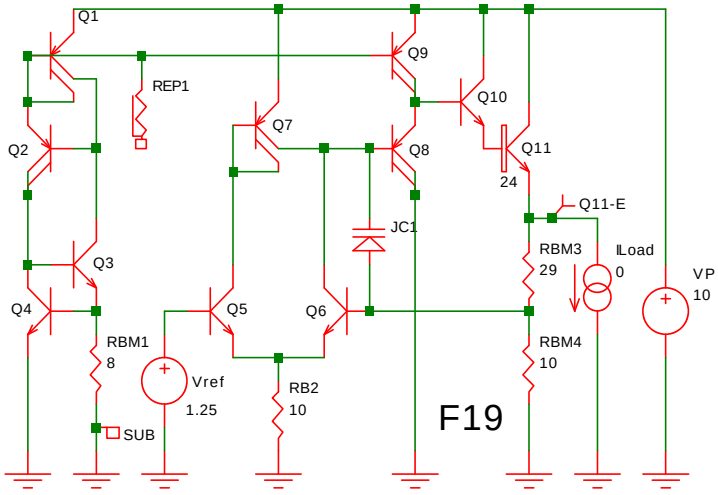
Though the Q of an LC circuit is much lower than that of a crystal, the problem described in F31 above exists here too. The oscillation builds up slowly and the maximum time step in the simulation must be set to a small value (about 1nsec) so that the very small initial amplitude is not ignored.

The circuit consumes about 4mA average. Sine-wave distortion is less than 1% (the higher the Q the lower the distortion) and the amplitude is tightly controlled by the supply voltage.



Regulators

F19 Voltage Regulator



This regulator presumes the presence of a 1.25V reference (see below and in the previous section).The differential amplifier Q5/Q6/Q7 senses the difference between this reference voltage and the tap on resistors RB3/RB4 (here set for a 5V output). Q8 buffers this difference voltage and removes whatever current is not needed at the base of the Darlington output transistor Q10/Q11 to produce the desired output voltage.

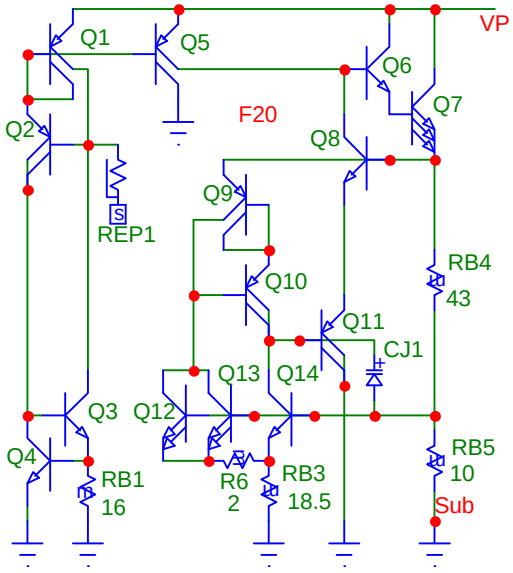
The bias current comes from the (E36) current source formed by Q1 to Q4. We have chosen this type of current source deliberately; its negative temperature coefficient compensates nicely for the positive temperature coefficient of hFE, so that there is a fairly constant current limit (150mA at -55C, 220mA at 125C). The current limit is set by the natural drop-off in hFE with current and tends to be very predictable

Idle current is about 200uA at room temperature, supply rejection -95dB and output impedance 0.02 Ohms. Because of the junction capacitor, output voltage is limited to 10 Volts.

F20 Voltage Regulator

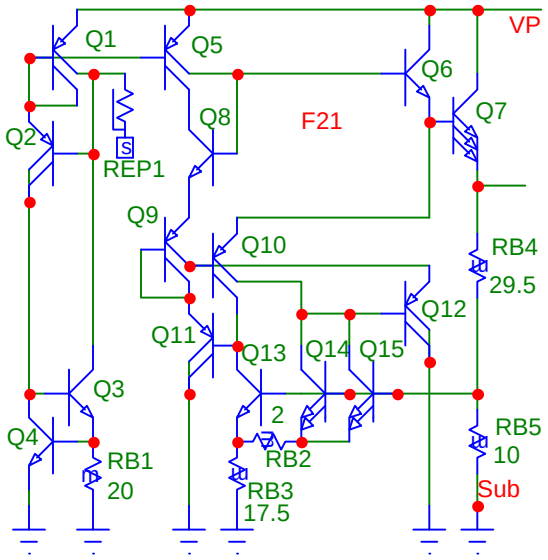
This voltage regulator has a built-in reference voltage similar to E80. At the bases of Q12/Q13/Q14 the required voltage for the feedback loop to stabilize is the bandgap voltage, or about 1.25 Volts. The reference is powered from the regulated voltage and the feedback transistors, Q11 and Q8, remove just enough current from the base of the output Darlington transistor (Q6/Q7) to bring the output to the desired level (set by the resistor divider RB4/RB5, 7 Volts for this example).

The bias current source is identical to that of F19, except that we have reduced the current, first by increasing RB1 and second by connecting one of the collectors of Q5 to ground. The output current is now limited to 70mA at -55C and 100mA at 125C. Output impedance is less than 1 Ohm.



Because of the junction capacitor, output voltage is limited to 10 Volts.

F21 Voltage Regulator



The only difference you will find between this circuit and F20 is the reference. We found a scheme which is stable without requiring a compensation capacitor. A load capacitor further increases stability.

Powered from the emitter of Q6 (which also provides a useful bypass current for Q7 at high temperature), Q10 delivers two equal currents to the basic bandgap cell Q13, Q14 and Q15. Any imbalance in these currents is sensed by Q11 and Q12 which, through Q9, divert just enough bias current from the base of Q6 to produce the desired output voltage. Q8 merely drops one VBE to prevent Q10 from saturating.

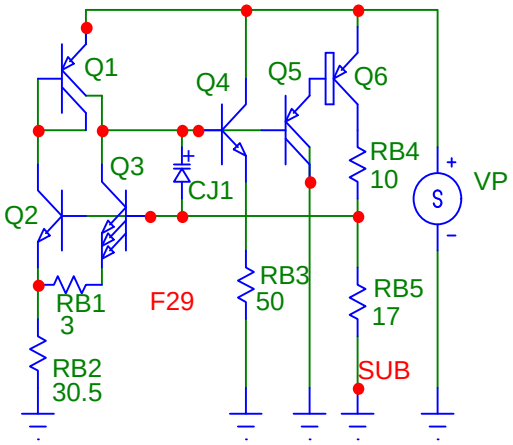
The divider string RB4/RB5 has been set here for 5 Volts, but you can set the ratio for any voltage from 3 to 18 Volts. Output impedance is 0.3 Ohms and the current limits at 90mA at -55C and 120mA at 125C.

F29 Voltage Regulator

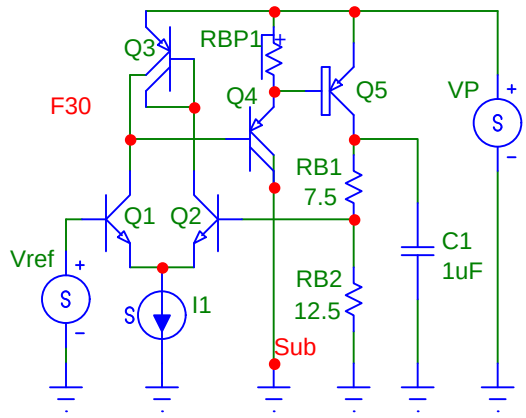
The advantage of this regulator is its low drop-out voltage, using a large PNP transistor in the output stage; this, however also limit the maximum current to about 5mA.

Q1, Q2 and Q3 form a basic Brokaw bandgap reference; the voltage at the base of Q3 is approx. 1.25 Volts. The error voltage is buffered by Q5 and fed to the output transistor, Q6. The junction capacitor CJ1 provides the compensation but also limits the maximum supply voltage to about 10 Volts.

The minimum supply voltage is just 0.1 Volts above the regulated output, with 2.3 Volts an absolute minimum at low temperature. The output impedance is 10 Ohms and the power supply rejection -37dB up to 1kHz. If you need better power supply rejection at high frequency, connect a 1uF capacitor to the output.



F30 Voltage Regulator



Here we are using a separate 1.25 Volt bandgap reference and more gain for error correction. The output stage is the same as in F29, but the differential pair with active load (Q1, Q2 and Q3) add considerable gain.

With a 2-Volt output as shown, the minimum supply voltage is 2.1 Volts, at any temperature. Again, the large PNP transistor at the output (Q5) limits the current to about 5mA, but the output impedance is less than 1 Ohm at 2mA.

C1, the capacitor across the output, is essential for compensation. It also holds the power supply rejection to -77dB at any frequency.

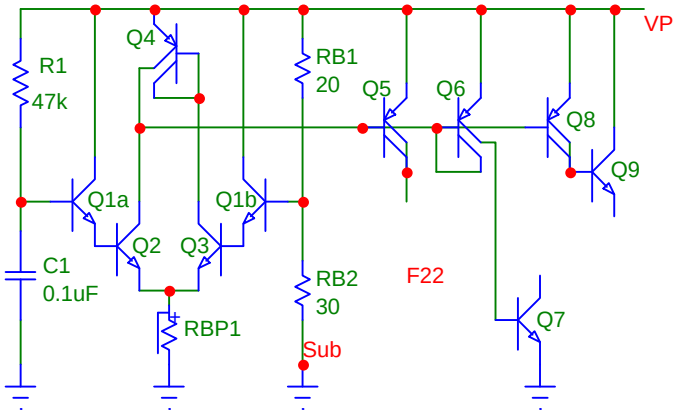
Timers

F22 Startup Timer

A simple yet accurate scheme to generate a time delay upon application of a supply voltage. R1 and C1 are external components and produce the time constant. Q1 to Q4 form a comparator which works with an input of up to VP. The threshold is set by the resistor divider (RB1, RB2). The resistor ratio needs to be set so that the voltage is above 2VBE (the cutoff for the differential stage) and below VP. We recommend that you go no higher than 0.8VP, otherwise the time becomes noise sensitive.

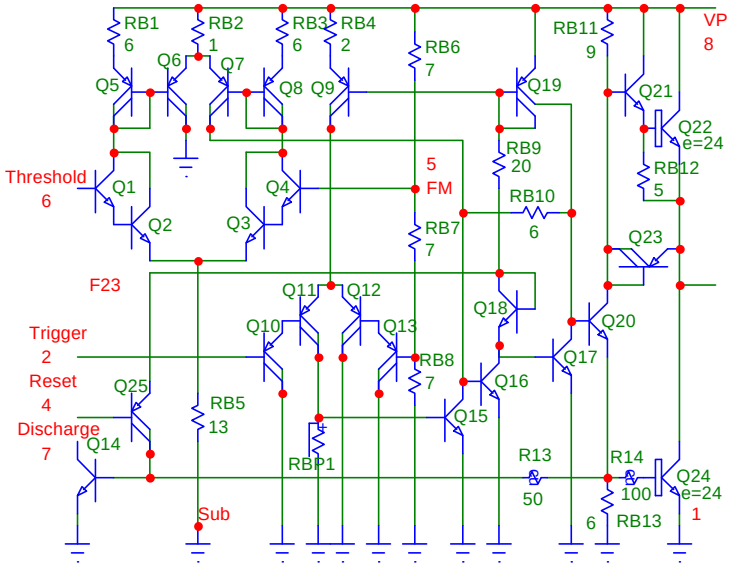
The operating current for the comparator can be supplied by a current source or simply a pinch resistor. Use a base-pinch resistor if its voltage does not exceed 5 Volts; for higher voltage an epi-pinch resistor will work.

We showed you several choices of outputs. Q5 can switch up to 300uA to VP (provided you have at least 10uA of operating current in the comparator). Q6 delivers the operating current of the comparator to the base of Q7, which switches a load to ground. And Q8/Q9 are capable of sourcing a large amount of current, though the output voltage is at least on VBE below VP.



F23 555 Timer

Many of you have inquired about the possibility of incorporating a 555 timer in an IC, so here we will explain the entire schematic, implemented with 700 Series components.



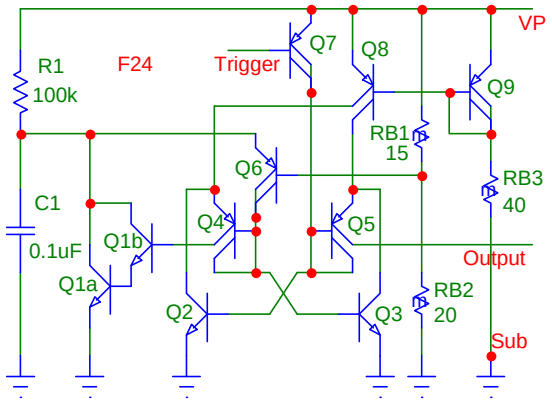
There are two comparators, one for the upper threshold, one for the lower. Q1-Q4 form the upper comparator, with a rather complicated scheme (Q5-Q8) as the active load. With today's processes the use of resistors and separate devices is no longer necessary; use the NPN equivalent of E45 and save 3 devices. The lower comparator (Q10-Q13), on the other hand, could stand some improvement by replacing RBP1 with an active load.

The two comparators set and reset a flip-flop, formed by Q16 and Q17. The output of the flip-flop is Q20, which powers both the output stage and the discharge transistor Q14.

The output stage is capable of conducting a great deal of current in either state, but there is a brief period when both halves are on. If you don't need to source current, for example, simply leave out Q21, Q22, Q23, RB11 and RB12 and connect the collector of Q20 to VP.

F24 Timer

This is a somewhat more modern scheme for a simple timer, using the flip-flop of F14. As R1 charges C1, a threshold determined by VP x RB2 / (RB1+RB2) and the VBE of Q6 is reached. Q6 now pumps current into the base of Q3, Q3 turns on and Q2 turns off, which causes current to flow into the emitter of Q4. This current splits into two equal parts: one half flows into the base of Q3, keeping it on. The other half turns on the Darlington transistor Q1 (two transistors with a common collector).



Note that Q1, because it is a Darlington transistor, has a voltage drop of one VBE. This cancels out the VBE created by Q6 and the swing across the capacitor is that given by the voltage divider RB1/RB2.

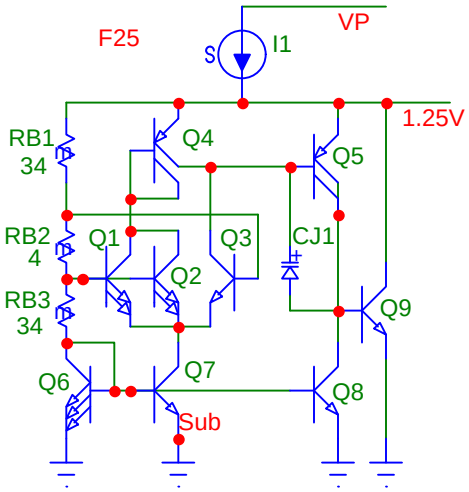
The timer is triggered by re-setting the flip-flop. This can be done by Q7 (as indicated) or by an NPN transistor shunting Q2.

The bias current for the flip-flop is produced by the current source Q8/Q9/RB3. Any other current source discussed earlier will work here.

Voltage References

F25 Voltage Reference

F25 is intended to be a general-purpose reference, producing the basic bandgap voltage of 1.25 Volts. Q1 and Q2 together have four emitters, Q3 only one. Thus the delta VBE amounts to 36mV (at room temperature), which exists across RB2. Q4 holds the collector current of Q1/Q2 and Q3 equal; any imbalance is amplified by Q5 and Q9.



The feedback loop now adjusts itself so that the voltage across RB2 is 36mV. The total voltage across all three resistors amounts to 658mV and has a positive temperature coefficient. Together with the VBE of Q6 (which has an equal but negative temperature coefficient), the total voltage adds up to 1.25 Volts.

The current sources Q7 and Q8, powered from Q6, provide the operating currents. The ratio of RB1 to RB3 is chosen to provide the maximum operating headroom.

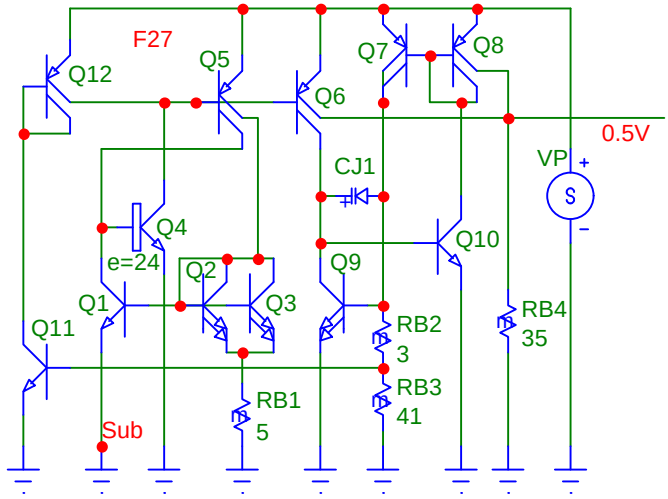
F27 Low-Voltage Reference

In the general approach to designing bandgap references, two elements with opposite temperature coefficients are placed in series. Since the fundamental bandgap voltage is 1.25 Volts, these circuit will not work with a single battery.

However, it is not necessary to place these elements in series; the two opposite effects can be combined through current rather than voltage.

In F27 Q1 through Q5 produce the positive temperature coefficient. The ratio of 4:1 emitters (Q2/Q3 and Q1) causes a delta-VBE of 36mV (at room temperature), which is dropped across RB1. Thus the current in Q5 (and also Q6) has the positive temperature coefficient of the delta-VBE, moderated by the temperature coefficient of RB1.

We have chosen a somewhat daring approach to provide start-up for this element: Q4 is a large NPN transistor, thus its leakage current is larger than that of Q5 and Q12. As long as there are a few picoamperes flowing out of the base of Q5, the current will be amplified.



One collector of Q6 is used to provide both the start-up and operating current for the element with the negative temperature coefficient, Q7 through Q10. The current of Q7 increases until the voltage drop across RB2/RB3 reaches the VBE of Q9. At this point the feedback loop (through Q10) will keep the current steady; CJ1 provides loop compensation. The current delivered by Q7 (and thus Q8) has the negative temperature coefficient of a VBE, but is also moderated by the temperature coefficient of RB2/RB3.

The base voltage for Q11 is about 40mV lower than that of Q9, produced by the RB2/RB3 tap. Q11, therefore runs at about one-fifth the current of Q9 and, through Q12, provides leakage bypass for Q5 after start-up.

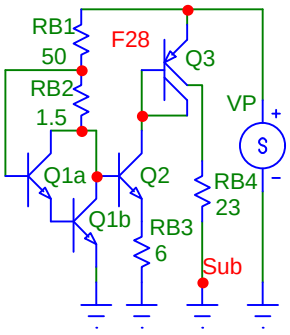
Now, one collector current each of Q6 (positive temperature coefficient) and Q8 (negative temperature coefficient) are summed and run through RB4. Two effects combine: 1) the temperature coefficients of RB1, RB2 and RB3 are canceled in RB4, and 2) if the ratio of RB1 to (RB2 + RB3) is chosen correctly, the temperature coefficient of the voltage across R4 is zero.

An additional advantage of this approach: you can vary the voltage drop across R4 independently, simply by selecting the value of the resistor. As shown, this voltage is 500mV and deviates less than 1mV over the military temperature range. With a minimum supply voltage of 0.9 Volts, you can go as high as 0.8 Volts at the output. The circuit works with a supply voltage of up to 20 Volts.

Note, however, that this is only a voltage reference; it has the output impedance of RB4.

F28 Voltage (and Current) Reference

For many applications it is not necessary to use a bandgap reference. If you merely need a bias point or operating current with good power supply rejection, consider F28.



Q1 (a single transistor in a Darlington configuration) provides a voltage of 2VBE. Thus the voltage across RB3 is a VBE and the current through Q2 (and Q3) is VBE/RB3, with a strong negative temperature coefficient.

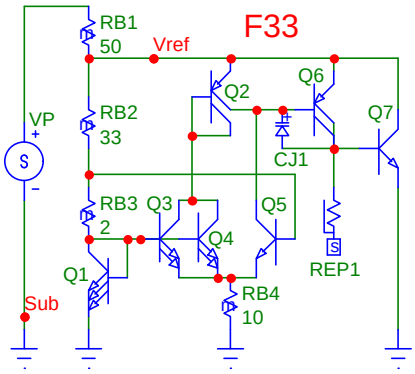
The performance advantage of this circuit is in RB2. As the supply voltage increases, the current through RB1 and RB2 increases as well.

At the junction of the two resistor, the base of Q1a holds the voltage more or less constant. This means that the collector voltage of Q1a/Q1b decreases with increasing supply voltage, lowering the current through Q2. With the right value for RB2 the collector voltage can be made to remain quite steady over a fairly wide supply voltage range.

With the values shown the circuit is optimized for 5-Volt operation and delivers 45dB of power supply rejection. You will find that, for different supply voltages, you will need to change that ration of RB1 to RB2.

F33 Anatomy of a Bandgap Reference

Here is a relatively simple shunt voltage reference. We will use it to go through the design and analysis steps in some detail.



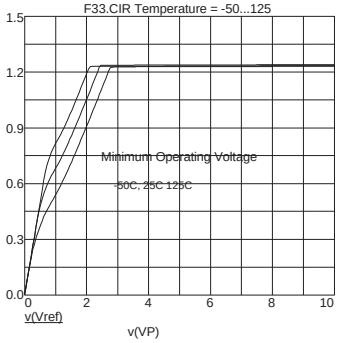
The VBE portion of the reference voltage is provided by Q1, the delta-VBE part by Q3-Q5 (multiplied by RB2 and RB3). The collector currents of Q3/4 and Q5 are kept equal by the active load Q2 and any difference is amplified by Q6/Q7. CJ1 is the loop compensation capacitor and REP1 a leakage current bypass. RB1 simply provides the operating current.

Three emitters were chosen for Q1 The VBE part of the bandgap voltage is subject to absolute variation; the larger you make this transistor, the smaller the variation. The choice of three emitters is simply a compromise; only a single device is required.

The ratio of emitters in Q3/4 and Q5 is another decision that needs to be made. 4:1 results in a delta-VBE (at room temperature) of 36mV. With 10:1 this increases to 60mV and at 24:1 you get 83mV. You would like to make this voltage as large as possible, but the required area increases exponentially. For a simple device a ratio of 4:1 is adequate; for maximum accuracy you will probably end up with a ratio of 8:1.

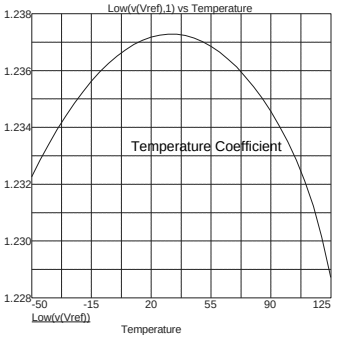
The delta-VBE appears across RB3 and is then multiplied by the ratio of RB2:RB3. To get an accurate ratio, both resistors must consist of many resistors elements. Thus, for RB2, you will need to use at least two strings of four in parallel.

Let's start analyzing this design. First we want to know if the circuit is working properly. A DC sweep will tell us that and also indicate the minimum operating voltage. Here we did this at three different temperatures.

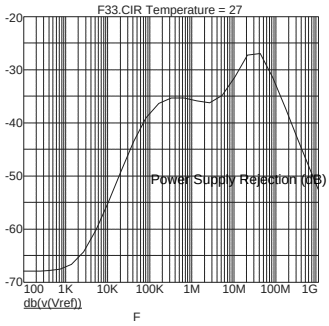


Next we perform another DC sweep, only this time we vary the temperature and keep the supply voltage constant (at say 5 Volts). You have made an initial choice for the values of RB2 and RB3, most likely dimensioning RB3 so that it sets a current which you can afford (24uA in this case) and then setting the value of RB2 so that the multiplied delta-VBE amounts to about

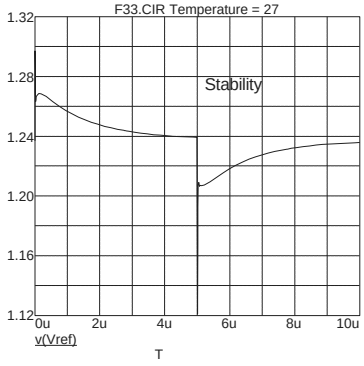
half the total bandgap voltage. In your first temperature sweep you will almost certainly see a significant temperature coefficient, but this is very easily remedied: simply adjust the value of RB3 until the temperature coefficient is zero.



You will always see a bow in the curve, because VBE and delta-VBE are based on two different mechanisms. This bow amounts to about 0.5%.



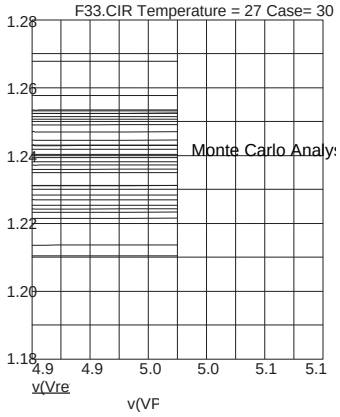
Next we might want to know how well the voltage reference regulates its output. One could sweep the supply voltage from 3 to 10 (or 20) Volts and display the output deviation, but more information can be gained with an AC sweep. Here the DC value of the supply voltage is held at 5 Volts (or whatever you choose) and an AC voltage is superimposed. This then tells you that our reference has a power supply rejection of about 28dB at the worst point without an external filter capacitor.



Now comes the most crucial question: is the circuit stable, or will it oscillate? This is best determined by abruptly stepping the supply voltage (or possible a load current) and then observing the output. If the output settles calmly, the circuit will not oscillate. You can even allow two or three cycles of damped oscillation, but any more than that is too close for comfort.

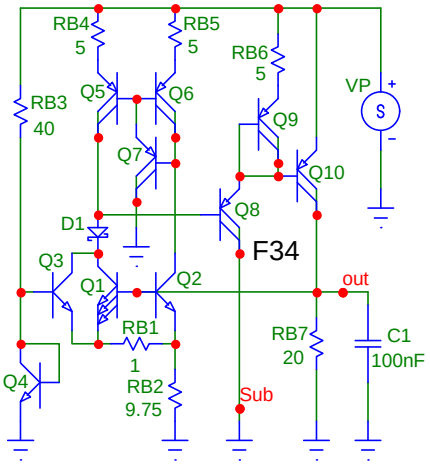
What can you do if you get oscillation? First try connecting the compensation capacitor to different nodes. Forget the theory at this point, you can quickly determine where the capacitor is effective and where it is not. You will probably find that it needs to be connected across a transistor as a Miller capacitance, so that its value is multiplied. But you may also find that a connection across several devices works best. If you cannot find a satisfactory connection, you will need to reduce the loop gain.

Two things to keep in mind here: first, a JCAP has a breakdown voltage of only 9 Volts (no problem in this circuit); second, if you intend to use an external filter capacitor, add it to the circuit first since it will change the phase.



Lastly we want to do a Monte Carlo analysis to determine how much voltage variation we will have in production. Here we sweep the supply voltage over a small range and start with perhaps 30 runs. As you can see from the graph, this is just barely adequate for this circuit. The number of runs should be large enough so that there are no widely separated results. In production, this circuit will have a 3-sigma voltage variation of ±2.5%.

F34 Bandgap Reference with High Power Supply Rejection

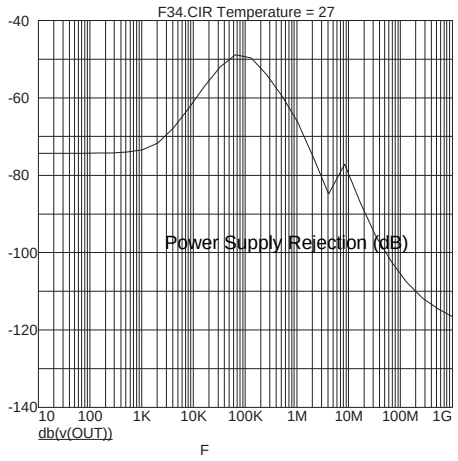


RB3, Q3 and Q4 start up the basic bandgap cell. The Schottky diode D1 is necessary to keep Q2 from saturating during this phase.

The delta-VBE is created by Q1 and Q2. If you have the devices available, you can increase accuracy somewhat by using more emitters (and transistors) for Q1 and then adjusting the ratio of RB1:RB2.

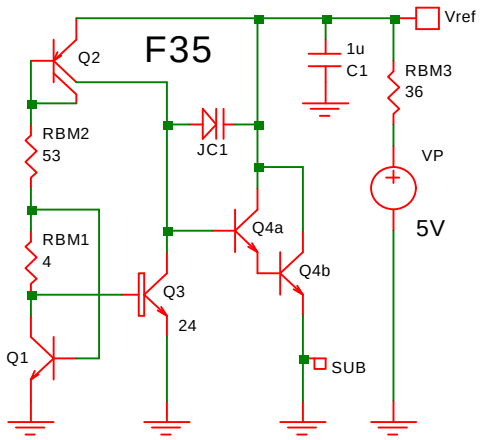
The active load (Q5-Q7) is designed not only for good matching but also to minimize the Early effect, which has a direct bearing on power supply rejection.

The gain of the Darlington pair Q8/Q10 is reduced by Q9/RB6 and C1 acts as both the loop compensation and output filter capacitor. In this way only small stray capacitances lead to the supply rail and their effects are significantly attenuated by the much larger external capacitor. Power supply rejection is greater than 48dB at 100kHz and reaches 70dB at both low and high frequency.



Contributed by Herbert Schoenke, Hagenuk Telecom GmbH, Kiel, Germany.

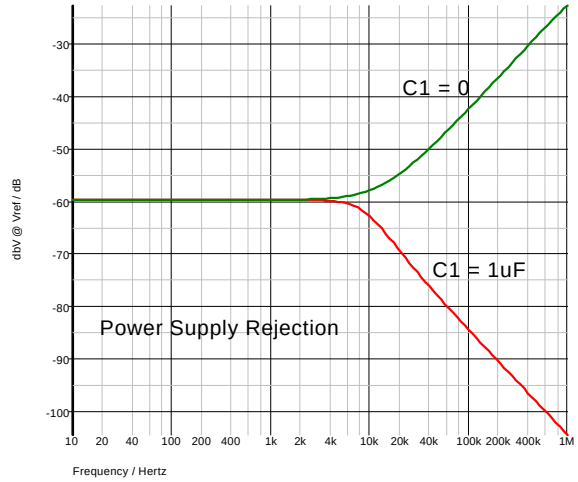
F35 A Simple and Novel Bandgap Reference

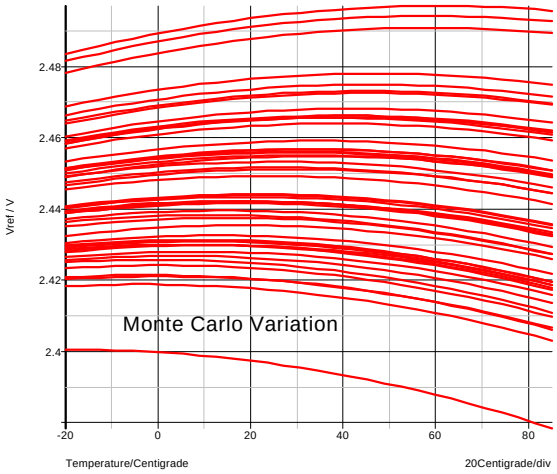


F35 requires just four transistors, yet delivers what is very likely the highest accuracy possible without trimming.

The positive temperature coefficient (or delta-VBE) is created by the difference in the number of emitters of Q3 (24) and Q1. Thus about 83mV (at room temperature) are dropped across RBM1 and increased to the required level by the series connection of RBM2. The negative temperature coefficient is provided by the VBEs of Q1 and Q2. The total voltage is 2.45V. The circuit differs from other designs in that the delta-VBE is in the collector of Q1.

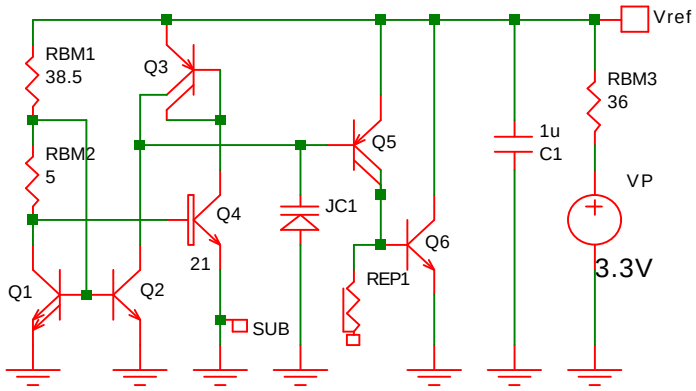
The feedback loop through the Darlington pair Q4a and Q4b (one transistor in the layout) keeps the collector currents of Q2 and Q3 equal. The loop gain is limited, resulting in an output impedance of 25 Ohms – sufficient only for light loading.



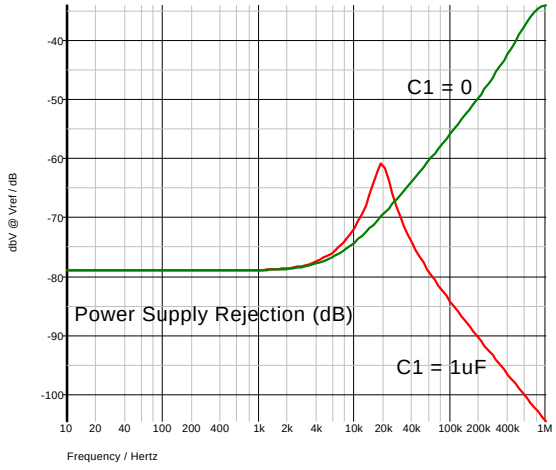


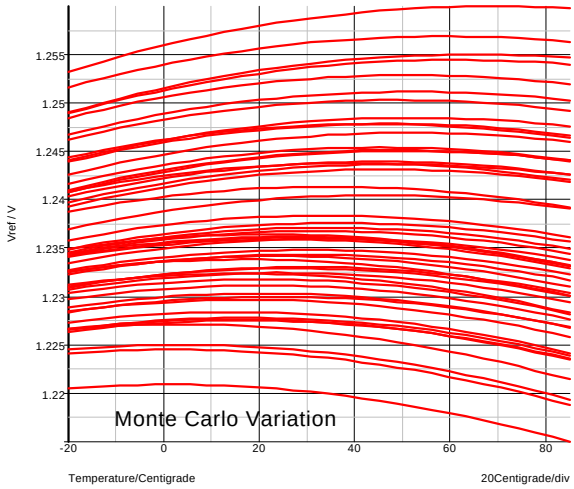
C1 is not necessary, but note that the power supply rejection increases toward higher frequencies without it. The overall variation in production is 1.8% (3-sigma). RBM3 must deliver at least 80uA at the lowest supply voltage.

F36 1.24-Volt version of F35



By adding two more transistors F35 can be converted into a 1.24-Volt bandgap reference. The loop gain is now higher, resulting in an output impedance of 2.6 Ohms. RBM3 must supply at least 60uA. The total variation from -20C to 85C is still +-1.8% in production.





Breadboarding

Integrated components differ from discrete devices in two respects: First, each component is surrounded by a junction, which can lead to some stray effects. Second, all regions and their contacts must be on one surface, which makes some of the parameters different. The integrated NPN transistor used most often, for example, has a current capability considerably below that of any discrete transistor. The integrated, lateral PNP transistor is not manufactured at all in discrete form.

For this reason it is important to verify a design. You have two methods available to do this: breadboarding and computer simulation. In some cases you may want to use both.

To breadboard your design we have provided a set of *kit parts*. These are the actual devices on the chips, wired up individually. All kit parts are in 18-pin dual in-line packages. Pin 9 is always the substrate and must be connected to the most *negative* potential in your circuit.

Kit part 7KP1 contains six identical small NPN transistors with a single emitter connected. Note that the emitter at pin 9 is permanently tied to the most negative potential.

7KP2 also contains six small NPN transistors. However, here the upper two transistors have three emitters, the middle ones two and the lower pair one emitter. This kit part lets you design with current ratios or increased currents.

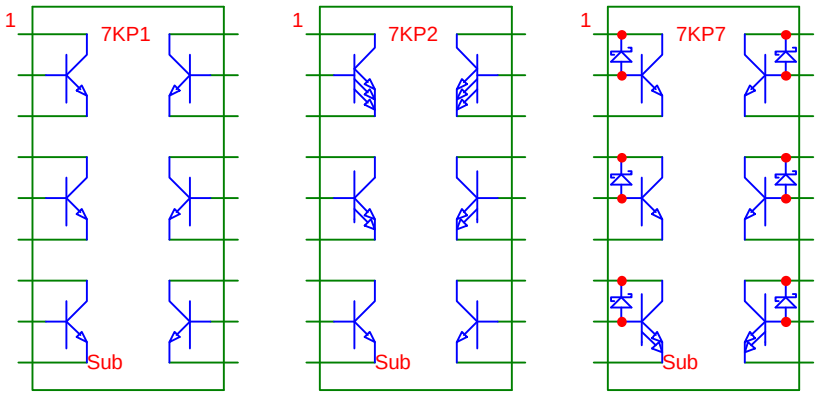
7KP3 has four small PNP transistors (with each collector brought out separately) and one epi-pinch resistor.

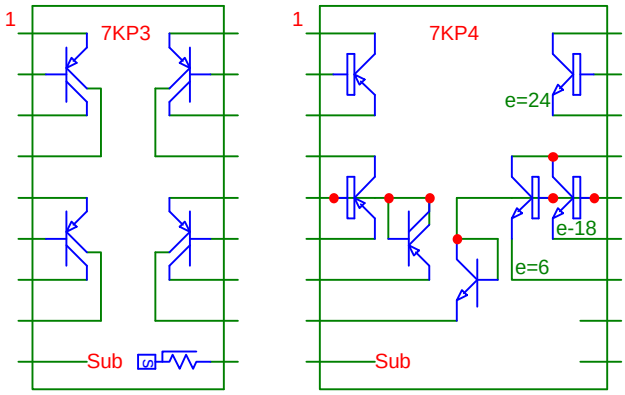
In 7KP4 we provided two each of the large NPN and PNP transistors and one each of the small transistors. One of the large NPN transistors has its emitters available in two groups.

The last kit part, 7KP7, contains six Schottky-clamped NPN transistors (with one of the emitters again permanently connected to the substrate). In this kit part the Schottky diodes are accessible between each base (anode) and collector (cathode).

We have found it too cumbersome and unnecessary to breadboard with integrated resistors. Apart from its tolerance and temperature coefficient this component more closely resembles its discrete counterpart and both of these effects can be simulated by connecting different values for different tests.

Be careful about capacitances when breadboarding. Bringing out a lead to an IC package pin (and breadboard traces) adds capacitance (2 to 6pF). This capacitance will not be present in the actual IC, which could affect both frequency response and the stability of feedback loops. If this level of capacitance is critical in your design, we strongly recommend that you also simulate at least this portion on the computer (see next chapter).





For the same reason you are usually better off breadboarding with discrete capacitors even if you intend for the capacitor to be integrated.

One additional recommendation: Separate the NPN kit-parts which you have used as Zener diodes. Although operating the base-emitter diode in reverse is certainly non-destructive, above about 10mA it changes the hFE of the transistor slightly (the amount of change is a function of current and time). Thus, if one device in the package was used as a Zener diode, it will no longer match the others well.

Naturally you will need to lay out the breadboard so that transistors which are required to match are in the same package. Matching between packages is merely accidental and can change with temperature.

Chapter 7s

Computer Simulation
Simetrix **Version 4**

This program is available from Catena Software Ltd., Terence House, 24 London Road, Thatcham, Berkshire RG18 4LQ, UK
+44 1635 866395, Fax +44 1635 868322
support@catena.uk.com
Http://www.catena.uk.com

Recommended Analysis Program

Contributor: David Bradbury, Zetex plc.

Installation

After you have installed the Simetrix program, unzip the file Sim700.zip (downloaded from the web-site) into a temporary folder. Create a new folder \Circuits within \Simetrix and then transfer the unzipped files as follows:

.lb	to \Simetrix\Models
.txt	to \Simetrix\Script
.sch	to \Simetrix\Circuits
.bat	to \Simetrix

Start Simetrix and go to **File, Scripts, Edit Startup**. Type in:

705Sim.txt
Icedmenu.txt

These are two script files and they will be run each time you start Simetrix. The first one sets up the 700 Series symbols, the second provides a link to the layout program ICED; if you don't intend to use ICED, leave it out.

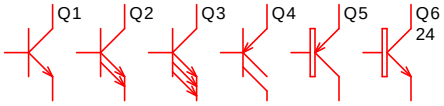
Notes: The latest 700 Series model file for Simetrix is 70025.lb (21 November 2003). If new model files are issued, make sure you remove the old one; Simetrix grabs the first model file it finds.

The introductory version offered on the web-site listed above can only handle a few transistors. Since the PNP transistor model consists of five devices and the NPN one of

two, only the smallest of the circuit examples can be analyzed (4 NPN or 2 PNP transistors). To evaluate Simetrix completely with the 700 Series examples, ask the vendor for a full version with timed expiration.

New Designs

To start a new design, click on **Parts, 700 Series** and choose say NPN1. Repeat this for the remaining five transistors, NPN2, NPN3, PNP, LPNP and LNPN, so that they are side-by- side. For an actual design you will also need to place the symbol SUB.



You will notice that these devices (and all others) snap to the grid, a feature that will enable you to hit the pins precisely when you interconnect them later on. Also, all transistors are the same size so that you can replace one with another.

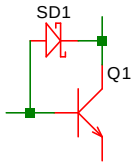
NPN1, NPN2 and NPN3 are the small devices, with 1, 2 and 3 emitters respectively. PNP is the same device (in the layout), connected as a lateral PNP transistor (make sure you connect both collectors, otherwise you will get an error message when you start the simulation).

LPNP is the large PNP transistor and LNPN is the large NPN one. The latter symbol has a number with a default setting of 24. If you use fewer than 24 emitters, select the device (left click), then right-click and change the “value”.

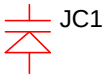
Each transistor has a hidden pin for the substrate connection. To complete the connection of this hidden net you need to connect the wire extension of the symbol SUB to the most negative voltage.

For all NPN transistors it is assumed that both collector contacts are connected. If you are unable to do that in the layout and the saturation resistance is critical for a device, we recommend that you place a cross-under (CU) in series with the collector and set the value to 50 (Ohms) for NPN1, NPN2 and NPN33 and to 6 (Ohms) for LNPN. If only a single contact is used for the bases of the PNP or LPNP no additional resistance is needed since the currents here are always small.

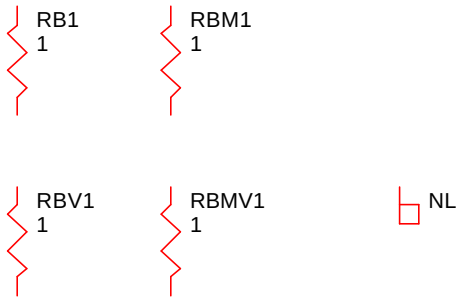
The next symbol is SD, the Schottky diode. It is part of an NPN1 transistor (or NPN2, if you use both emitters) and its cathode is always connected the collector of the transistor. If you are not worried about the capacitances of the transistor (or the P-N junction in parallel), use the Schottky diode symbol alone. If you want to include the transistor effects, connect the emitter of the transistor to the base to avoid the "floating pin" error message.



The junction capacitor (JCAP) has a unique symbol so it won't be confused with an external device. Its third pin, the substrate, is hidden and is automatically connected to the (hidden) SUB net.



You have four different symbols (and models) for the base resistor. The first one, RB, is the basic 750 Ohm resistor; its model includes the temperature coefficient, as well as the tolerance for a single resistor (used for the Monte Carlo analysis). Its default value is 1, i.e. one 750 Ohm device. Left-click on the device, then right-click to change its value to say 3, if three of them are connected in series. For a parallel connection of three base resistors the value would be 0.33 (or 0.3333 if you wish).



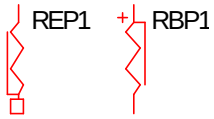
The next base resistor model is RBM (for RB, multiple). As explained in chapter 3, the more resistors you use (in series or parallel) the better the matching. The only difference in the model is the matching tolerance.

Use the RBM symbol only if you combine at least five resistors; the matching of resistors improves gradually with the quantity. Note that the matching of 1:10 resistors is not much better than 1:1, since the variation of the single resistor predominates. If you need good matching with the value of a single resistor, connect several resistors in series and parallel to get a value of one.

In the second pair of resistor symbols (RBV and RBMV) we added the voltage sensitivity (a 2% change as the voltage difference between the resistor and the surrounding N-layer changes by 20 Volts) and the stray-capacitance to the N-layer; in these models the resistance is internally divided into three segments and the result is very close to that of distributed capacitance. As before, M indicates use for multiple resistors.

If you use the second pair of resistor symbols, you will also need to get the symbol NL and connect it to bias point of the N-layer (usually the most positive voltage).

To avoid unnecessary clutter in the schematic, we have not named or numbered the symbol CU (a cross-under), it is simply a short resistor. The value here is in Ohms (rather than multiples). Resistance of cross-unders is listed at the end of chapter 3.



Among the remaining symbols are the pinch resistors. The negative terminal of REP, the epi pinch resistor, is permanently connected to the substrate; make sure you have the symbol SUB connected to the most negative supply voltage. RBP, the base pinch resistor has both terminals free (and no substrate connection), but is also polarized.

DP is the symbol for the ESD protection diode underneath most pads (and connected on the side of the pad). There is also a symbol for the pad itself.

Some hints and suggestions on drawing a Simetrix schematic:

1. After you select the part, you can rotate or flip it by clicking on the appropriate symbols in the toolbar, then place it in the desired location.

2. If you want to rotate or flip a part after is has been placed, select it and click on the appropriate symbol in the tool-bar. To avoid rubber-banding detach it first (**Edit, Detach**).

3. You can move any component after it has been placed by dragging it with the left mouse button pressed. To move entire groups of parts select them (draw a box), then place the cursor on any of the selected parts and drag with the left mouse button pressed. You can also select a component or a group of components and then **Copy** and **Paste** them. The components numbers (references) will be automatically updated to avoid duplication.

4. You can choose to automatically repeat the placement of parts or do them one at a time (Command Shell, **File, Options, General...**).

5. The device numbers (e.g. Q1) can be changed by selecting the part, **Edit, Change Reference.**

6. The most useful key in both the schematic and the plot is **Home**. It scales the content so it fills the box. With the Zoom Rectangle symbol in the tool-bar you can outline a magnified portion.

7. If you want to select a wire rather than a component, press the **Shift** key while drawing the box around the area. You can also make the wiring non-rectangular by dragging the end of a wire with the left button of the mouse depressed.

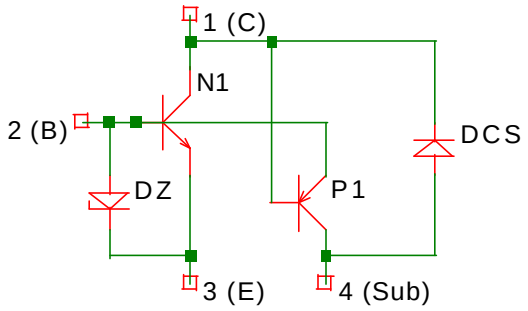
8. Simetrix has an elegant and powerful hierarchy feature, which lets you avoid ungainly schematics for complex circuits. Hierarchical schematics also have two other advantages: a) you can measure the power dissipation in an entire block or circuit (see below), b) placing the entire IC (without external components) prepares it for LVS (layout versus schematic checking) and lets you set up a variety of tests with different external stimuli or components in an efficient way.

9. **Before you start a simulation, make sure you have not forgotten the SUB symbol**. We have place a safeguard into most models: If the simulation won't start and you see an error message in the Command Shell window referring to "singular matrix" and SUB or SUB2, it is because you have forgotten to place a SUB symbol. There must only be one SUB symbol in each schematic.

The Models in Detail

You can look at the 70024.lb file with a word processor (it is an ASCII file). It starts with a description of NPN1, NPN2 and NPN3:

```
.SUBCKT NPN1 1 2 3
Q1 1 2 3 #SUB N1 lot=700SER
Q2 #SUB 1 2 #SUB P1 lot=700SER
D1 2 3 DZ lot=700SER
D2 #SUB 1 DCS
.ENDS
.SUBCKT NPN2 1 2 3
Q1a 1 2 3 #SUB N1 lot=700SER
Q1b 1 2 3 #SUB N1 lot=700SER
Q2 #SUB 1 2 SUB P1 2 lot=700SER
D1 2 3 DZ 2 lot=700SER
D2 #SUB 1 DCS
.ENDS
.SUBCKT NPN3 1 2 3
Q1a 1 2 3 #SUB N1 lot=700SER
Q1b 1 2 3 #SUB N1 lot=700SER
Q1c 1 2 3 #SUB N1 lot=700SER
Q2 #SUB 1 2 P1 2 lot=700SER
D1 2 3 DZ 3 lot=700SER
D2 4 1 DCS
.ENDS
```



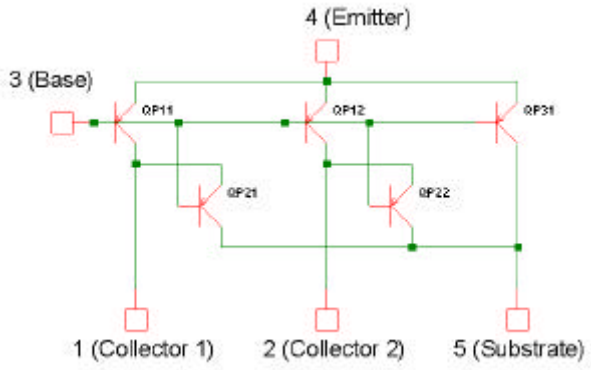
These are subcircuits, internally connecting two transistors and two diodes. The only difference between the three models is the number of times each internal device is used. Further down in the 70024.lb file are the models for the internal devices. N1 is the main NPN transistor and DZ simulates the Zener breakdown of the base emitter junction. P1 creates the substrate current as N1 saturates; the collector-substrate capacitance is contained in DCS. #SUB is a global node, recognized in and out of the sub-circuit. You connect this node to the most negative voltage with the symbol SUB.

In the model for the large NPN transistor the number of emitters used are passed through by m:

```
.SUBCKT LNPN 1 2 3
Q1 5 2 3 #SUB N1 {m} lot=700SER
Q2 #SUB 5 2 #SUB P1 {m*0.5}
D1 2 3 DZ {m} lot=700SER
D2 #SUB 5 DCS 3.5
RC 5 1 1.5
.ENDS
```

The lateral PNP transistor has two separate collectors, represented by two PNP transistors QP1 (each with half the gain). The substrate current is created by two types of PNP transistors: QP2 (turning on when the associated main transistor saturates) and QP3, which produces a small amount of substrate current out of saturation. QP3 also contains the base-substrate capacitance.

```
.SUBCKT PNP 1 2 3 4
QP11 1 3 4 #SUB QP1 lot=700SER
QP12 2 3 4 #SUB QP1 lot=700SER
QP21 #SUB 3 1 QP2
QP22 #SUB 3 2 QP2
QP31 #SUB 3 4 QP3
.ENDS
```



The model for the large PNP transistor is similar to the one above, except that only one collector is present. Both emitter stripes are assumed to be connected.

The junction capacitor (JCAP) has three nodes; in sequence: positive terminal, negative terminal and substrate. Two diodes simulate the capacitances and the one between the two terminals has its breakdown voltage set at the minimum value of 9 Volts.

The two pinch resistors (REP and RBP) are simulated as junction field-effect transistors (they are closer to current sources than resistors) and, as all other devices, contain the temperature coefficients and tolerances. The first terminal is always the positive one. In the case of the epi pinch resistor the second terminal must be at the substrate connection.

General Remarks on Schematic Entry and Simulation

In the files you downloaded and put into the folder Circuits we provided a large number of examples of Simetrix simulation. Not only can you use these Elements and Functions in your own design by copying and pasting them, but they will show you how Simetrix works. Before we begin, a few hints and suggestions:

1. Simetrix allows you to run a number of different analyses in sequence. We suggest you only do one at a time, it's far less confusing.

2. You have the choice of either placing permanent or temporary markers (probes). In E1, for example, we placed a permanent voltage probe on the output (**Parts, Probe, Voltage Probe**). In this way the result is displayed while the analysis progresses. After the analysis you can click on **Probe** or **Probe AC/Noise** and place a probe on any node you desire.

Placing a permanent trace for current it a bit more tricky. It measures current only into pins (unlike the temporary **Current** marker, which works for wires as well). When in doubt put a dummy zero-voltage source into the circuit. See for example E10.

3. You can find out what the dc bias voltages are at all nodes by right-clicking on the schematic, then clicking on **Bias Annotation**, **Auto Place Markers** and then **Update Values** after the analysis. After you have examined the voltages click on **Hide Values** and **Delete Markers**.

4. It is always a good idea to check the substrate current in an IC. If you place a wire between the SUB symbol and the most negative rail, you can do this with **Probe**, **Current in Wire** or **Probe**, **Current in Device Pin** (note that this measured current has a negative value).

If the magnitude of the substrate current is worrisome, you can find out which device creates it by placing the symbol SUB1 into the schematic and connecting it (through a wire again) to the most negative supply. Now select the device you suspect, right-click on it and go to **Edit Value/Model.**
Add sub1 at the end of the model (e.g. NPN1sub1). This changes the model for this particular transistor to one in which the hidden substrate connections don't go to SUB, but SUB1. After the run you can then measure the substrate current for this transistor alone.

Don't forget to change the model back to the original one after the measurement.

5. You can copy and paste any of the elements or functions in the directory \circuits. Afterwards simply erase any of the components not needed, such as extra power supplies. Also, there must only be one SUB (amd NL) symbol in each schematic.

6. You can change the location of the reference number (e.g. Q1) or the value (e.g. 24 for a 24-emitter LNPN) for any component. Select to component, right-click on it and go to **Properties**.

700 Series Examples of Simetrix Analysis and Probe

DC Analysis

E1 (Buffer) A permanent probe has been placed at the output so you can watch the result while the simulation is in progress. Erase the plot and go to **Probe, Voltage - Differential** and then click on the input and the output in sequence, which plots the offset. If you don't erase an existing plot, the new curve will simply be superimposed.

E9 (Current Mirror) A dummy voltage source of zero voltage is used to place a permanent current probe. Such a probe must be placed on the terminal of a single device.

E22 (Current Mirror) Here temperature is swept. You can measure the substrate current (or any other current) by going to **Probe, Current in Wire** and then clicking with the probe symbol on the wire of interest.

E25 (Current Source) The supply voltage is varied from 0 to 20 Volts to determine the minimum operating voltage. Note that the transistor models do not contain a collector-emitter breakdown voltage; the actual breakdown voltage is somewhat higher than 20 Volts, but we do not recommend using a total supply voltage greater than 20 Volts.

E31 (Current Source) Here the supply voltage is swept from 0 to 20V while the temperature is stepped. To do this you go to **Simulator, Choose Analysis** and enable **Multistep**. Then under **Define** you select **Temperature** and **List, Define List** and set the temperature steps.

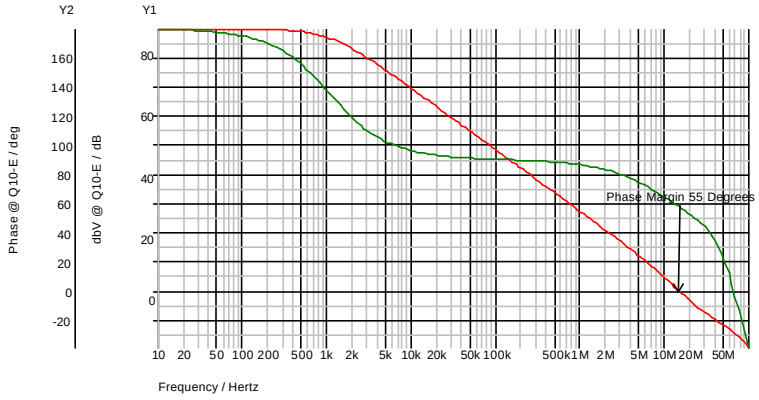
Simetrix now also lets you mark the curves. Since there are three curves in this example, you might want to have them identified when you print the plot. With the plot displayed, go to **Annotate, Add Curve Marker**. Text with an arrow affixes itself to the first curve. If you grab the point of the arrow and depress the left button on the mouse, you can move it to a different curve. You can also move the text around the same way. Double-clicking on the text lets you change the text.

AC Analysis

E57 (Fixed Gain Amplifier) There are no AC permanent markers, so we put a simple voltage marker at one of the outputs, just to see that something is happening during the analysis. Erase the plot and go to **Probe AC/Noise, dB - Differential Voltage** and click the probe symbol sequentially on the two outputs. You will now see the differential voltage gain in dB. Without erasing the plot repeat the procedure with **Probe AC/Noise, Phase - Differential Voltage**. This will add the phase shift in the circuit to the plot.

F34 (Bandgap Reference) This circuit has been set up to simulate power supply rejection. The AC source (1 Volt) is in VP (together with the DC supply). Use **Probe AC/Noise, dB Voltage** to get the result in dB.

F4 (Operational Amplifier) A very large inductor is placed in the feedback loop and the AC signal is fed in through a very large capacitor. With this addition you can measure the loop gain and phase (use **Probe AC/Noise, dB Voltage** and place the probe at the output (i.e. before the inductor) and then repeat for **Phase, Voltage**.



You should not entirely rely on this analysis to judge the stability of a feedback loop; it measures the gain and phase at one operating point only. We recommend that you add a transient analysis (e.g. put an abrupt step into the supply voltage or the load) and watch for ringing (see F4 below).

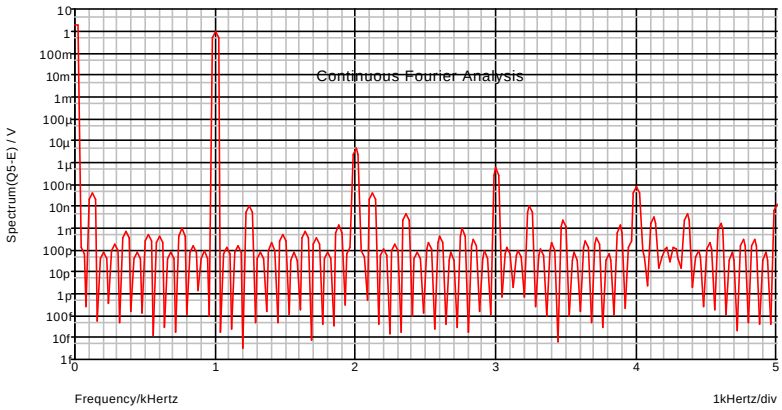
F1 (Fixed Gain Amplifier) Set up to measure noise. After the analysis go to **Probe AC/Noise, Plot Output Noise** or **Plot Input Noise** (which is the same thing minus the gain of the amplifier). The result is in nanovolts per roothertz, i.e. to get the actual rms noise voltage you have to multiply it by the square-root of the bandwidth.

For the small signal noise analysis you need to know the node number of the output. Put the cursor over the node and type **Ctrl-S**. The node number will now appear in the Command Shell window (RB4_N).
If you need to identify a current rather than a voltage, the command is **CTRL-P**.

For some circuits (e.g. mixers) small signal noise analysis is meaningless. Use the real-time noise feature now present in the Simetrix transient analysis instead.

Transient Analysis

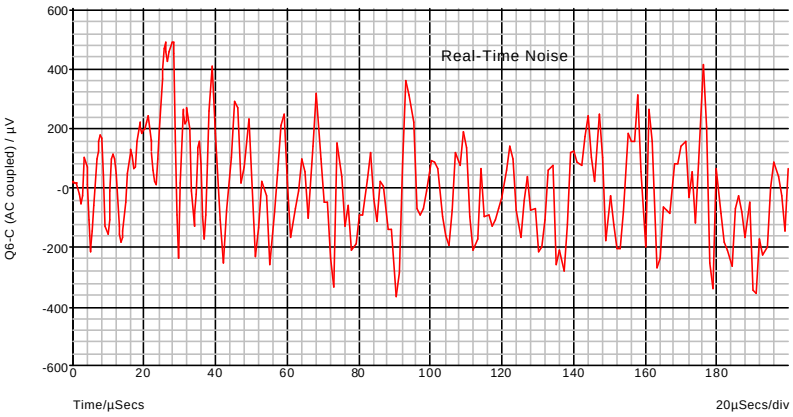
E62 (Operational Amplifier) The input in this analysis is a sine-wave, run over many cycles to obtain a good resolution. Normally the distortion at the output is measured with a Fast Fourier Transform (FFT), which occasionally shows errors due to aliasing. Simetrix has a full (or continuous) Fourier transform, which you evoke by going to **Probe, Fourier, Probe Voltage Custom**. You then can select several parameters, including resolution.



E31 (Crystal Sine-Wave Oscillator) For this analysis we have put the crystal in a sub-circuit so you can change its parameters. Be aware that high-Q crystals take a long time to reach the maximum amplitude of oscillation (up to 1 second). Fortunately the speed of the simulation has roughly been doubled in Simetrix version 4. However, a fast computer helps too.

F3 (Operational Amplifier) This amplifier receives a large-amplitude pulse at the input. The output has a relatively slow rise-time, but an abrupt fall-time. Watch for the ringing after the fall time. Up to four cycles of damped oscillation indicates that the feedback loop is stable. This, however, should also be checked with the load used.

F29 (Voltage Regulator) Here there is no voltage, current or load which changes abruptly. What we are looking for is noise at the output. Simetrix version 4 has the unique ability to activate all noise sources within the circuit (**Simulator, Choose Analysis, Transient, Enable Real Time Noise**). The permanent probe shows you the noise superimposed on the DC level. To see the noise alone erase the plot and go to **Probe, Voltage - AC Coupled**.



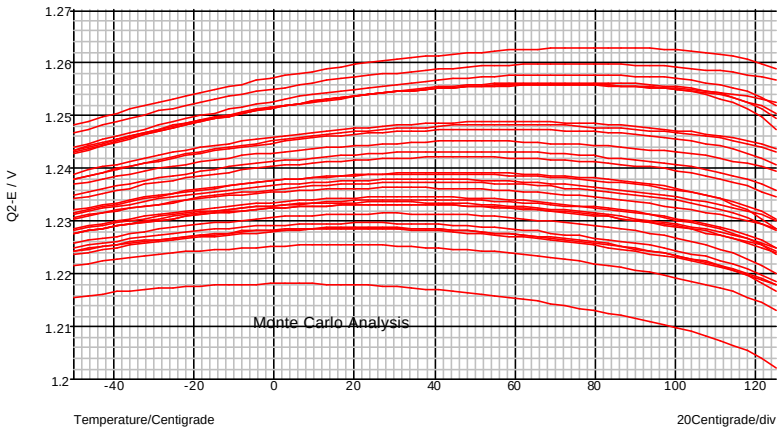
Select the **Interval** carefully, this sets the noise bandwidth. Their relationship is BW=2.5/t, where it is the interval.

Monte Carlo Analysis

The importance of the Monte Carlo analysis in analog IC design cannot be emphasized enough. Until you have subjected each parameter to be tested in production to a simulation which contains the variations occurring in production, you won't really know how well your design will fare. Setting test limits according to the results of a prototype wafer (or worse: a few assembled prototypes) is not good practice.

Each model for the 700 Series components contains the variations which will be experienced in production. This usually consists of two parts: the (absolute) variations from wafer to wafer and from run to run, and matching, i.e variations from device to device within the same chip. These variations have been checked against millions of production parts.

F33 (Bandgap Reference)
Here we make 30 Monte Carlo runs, with the temperature swept (a unique feature of Simetrix). In this way we see the total variation of the output voltage over the entire temperature range.

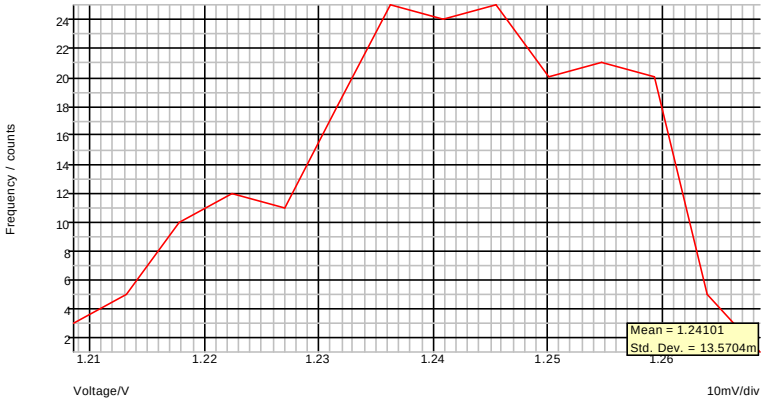


Usually 30 runs are sufficient. If you are unsure about how many runs are sufficient, simply increase the number of runs until there is no longer a change in the distribution.

F25 (Bandgap Reference) With version 4 Simetrix has added a feature which is especially valuable for simulating parameters to be tested in production: the single point histogram. You choose as the swept parameter Monte Carlo (200 points in this case). You get a plot which simply tabulates the results for the runs. Erase the plot on go to **Monte Carlo, Plot histogram**. Simply click on the node of interest and it will fill in the field **Expression** automatically.

Note that this is not a Gaussian distribution.

We have deliberately chosen a rectangular distribution for the absolute variations. These variation occur from wafer to wafer. In most 700 Series applications there are relatively few wafers involved in any production release.



You never know which wafers you will get, they may even come from different wafer lots; so there is an equal chance of getting a wafer (or a few wafers) with high or low parameters. The matching parameters, however, follow a Gaussian distribution.

You can now use this plot to set the test limits.

Other Hints and Suggestions

1. You can print a schematic or plot directly or copy them to the clipboard and then pasting the image in a word-processor. In the schematic select all the components (draw a box around them), click on **Edit** and **Copy**. In the plot right-click and select **Copy to the Clipboard**. In the word-processor paste and then enlarge the picture and place it where you want it.

With version 4 Simetrix also has the ability to save a Windows Metafile (.wmf) or Enhanced Windows Metafile (.emf). Rather than copy and paste (which allows only one image at a time), you can give the file a name and store it for later use.

2. With the feature **Probe, Power in Device** you can plot the power dissipation in any device for any analysis other than AC and Noise. This also applies to hierarchical sub-circuits (which are treated as devices). Thus, if you make a sub-circuit out of the entire IC (without the external components) and click on it with this command, you get the power dissipation for the entire chip.

LVS with ICED

When you add Icedmenu.txt to the Simetrix start-up file, a line appears at the bottom of the pull-down menu File. This provides a link to the layout program and you can perform the entire LVS sequence without actually leaving Simetrix. This applies to both the DOS and the Windows version of ICED.
In the four .bat files we have specified the path to the ICED programs as c:\iced\.... If you are using a different folder, change this with a word-processor. Also, if you are using the Windows version of ICED, change the last line to:

c:\icwin\iced %1 start=c:\icwin\ndjb32_2.cmd menu=1

assuming that you Windows ICED program in the directory \icwin.

Also, for the DOS version, in the file ic.bat the resolution of the display is set at vesa=100 (VGA, 640x480). For higher resolutions set this to:

vesa=102	(SVGA, 800x600)
vesa=104	(XGA, 1024x768)
vesa=106	(1280x1024)

You might also want to experiment with the numbers used after use= or hog=. These allocate memory (in kB) and we have set them to a minimum so they work with as little as 16MB of ram. Though the ICED programs do not require much memory, you might get a slight speed advantage setting them higher if your computer has 32MB or more.

The first item in the sub-menu of **File, ICED (for LVS)** is **Prepare Netlist**. With a schematic opened up, click on it and a netlist will be compiled under the name you specify. You can choose any name (the extension will be .net) and put it in any folder you like.

The second step is Run ICED. You will have to guide the program to the ICED .cel file which corresponds to the schematic displayed. In the ICED screen click on **File, DRC386 (DRC NT** for the Windows version) and **Full** and then **File, Exit**.

The next step is **Run DRC**, which merely checks if you have marked the convertible transistors correctly (here DRC does not check the spacings and minimum wire widths). If you see no error messages at the bottom of the DOS screen, got to the next step.

Click on **Run NLE** (netlist extractor). This part of the program will use the ICED layout file you have specified under Run ICED. You now have two netlists, one extracted from the schematic, one from the layout.

The comparison of the two netlists is done by clicking on **Run LVS**. Any differences between the two netlists are listed in two files, unmatch.lvs and param.lvs. You can look at them by clicking on the two bottom entries of the pull-down sub-menu.

ICED has a very competent feature to track down errors in the layout. Note (or perhaps print out) the layout nodes in question and then type **@nodes** in the ICED screen. The command **n0** (zero) will prompt you to type in the node you want to see and then high-lights it. Type **nd** to cancel the high-light. More of this in chapter 8.

Computer Simulation

P-Spice/Orcad/Cadence Release 9

We no longer recommend using P-Spice for the simulation of 700 Series ICs. SIMetrix (see chapter 7s) has moved far ahead of P-Spice and is available at a fraction of the price.

Users who want to continue using P-Spice should download the file PSP700.zip, which contains the symbols, models and circuits in Schematics format. You will need to download the web-update from www.pspice.com to install the Schematics program.

For information on the models and circuits read chapters 5 and 7s. Be aware, however, that P-Spice lacks many of the features present in SIMetrix.

Blank page.

Chapter 8

Layout with IC Editor

(www.ic-editors.com)

You can download the layout patterns from the 700 Series web-site (Chapter 8, Layout.zip). There are three directories within this file:

ICEDLayout	ICED files for layout
ICEDLVS	ICED files for LVS
Streamfiles	GDSII streamfiles for layout programs other than ICED

It is important to keep the latter two separate, they contain .cel files with the same names but different content (see LVS below).

In the following we have assumed that you have copied the content of ICEDLayout into a directory called c:\Layout and that your Windows ICED program is under c:\ICWIN.

At this writing the Windows ICED program still needs to be started from a DOS window. Go to c:\Layout and type:

ICWIN 711XX

You will see an example of a layout appear on the screen, based on the chip 711. The menu is in the right side of the screen. Now click your left mouse button on **File** and then **Exit**.

To create your own layout you have to use a file name which does not yet exist. Let's say we call it Trial; type

ICWIN TRIAL

After a few seconds you will see the menu on the right-hand side of the screen, but nothing else. What you have to do next is to call in one of the basic chip patterns. Click your left mouse button on:

ADD

You'll see a new menu with many different options of what to add and how. Choose the option at the very bottom, clicking the left mouse button on:

at 0,0

The menu changes again, listing all the cells which can be chosen. At the top are the nine basic chip patterns, 710 to 747 and the three examples. The cells following 747 are the devices which are needed by the chips. They are of little use to you directly, though you certainly can look at them.

Let's assume that you have decided to start with a small chip. Click your left mouse button on:

711

The screen remains almost empty at first because you are looking at a very small portion in the extreme corner of the chip. Now click on:

all (which is a sub-heading of **VIEW**)

and the entire 711 chip will come into view. You can see greater detail by clicking on **box** and choosing an area you want to see (with the mouse, clicking on the corners). If the area is now too detailed, you can click on **out** or go to **all** and then choose another view box. **Last** will show you an area that was visible just before the present one.

By clicking on **VIEW** you open up many more choices. However, we will leave the detailed explanation to the ICED manual.

You can cancel any command by pressing both the left and right buttons together.

Now choose a reasonably small window (a few transistors wide) by clicking on **box** and outlining the area. You are now ready to draw interconnections. First click on:

UseLay (The layer in use)

and you will see the choices of layers you have. **WHIT**, **RED**, **GRN**, **BLUE** and **YELL** let you do the wiring in five different colors, so that you can distinguish say supply lines from signal lines. All five layers are eventually going to be combined (together with the fixed metal pattern FMET) into a single "metal" mask.

Following the five interconnection layers is **CTEX** (for customer text). You can use this layer to identify your devices or make any kind of remark you wish; this layer will be ignored when making the mask.

Following CTEX are five layers, which are intended to be fixed (i.e. you should not alter them, though you can blank them if they bother you):

RT	Routing	BG	Background
GN	Green Legends	FMET	Fixed Metal Areas
RD	Red Legends		

Let's make some interconnections. Choose **WHIT**. Another menu will appear, which gives you the choice of a variety of geometries (and text) for this layer. Choose **WIRE**, which is preset for a 4 micron wide wire. Now place the beginning of the wire with the mouse (say on a resistor contact) and click the left button, then move the mouse and simply click at every bend. You will notice two features:

1. The movement of the cursor jumps in small steps. The entire layout is on a 2 micron grid; the wire automatically snaps to this grid.

2. You can only place the wire at right angles or in 45^{o} increments.

We have placed cyan colored routing lines all over the chip to help you place the interconnection wires. You are not compelled to follow them (and they don't show all possible interconnection routes), but you must follow two simple rules:

A. The spacing between unconnected metal areas, including the fixed metal areas (in white) already on the chip, must be at least the same (4 microns) as the width of the wire.

B. There must be at least a 16 micron spacing between the pads and other metal, i.e. you cannot go any closer to the pads than the routing lines show.

You will also notice that the screen pans automatically. If you move the cursor beyond the screen's edge, the image follows it.

Now let's go back to finishing your interconnection. When you have arrived at the point where the wire is to stop, click first the left and then the right mouse button.

You can now draw a second wire by either clicking on **Again** (which simply repeats the last command and works with any command) or on **wire**. There is no need to go through the **UseLay** routine again; the program will stay in **WHIT** unless you change the default layer.

Suppose you don't like the placement of the wire or you made a mistake. You can either delete or change it. First, deleting a wire (or anything else):

Click (your left mouse button) on **DELETE**. You'll see a small circle with an alignment cross. Move it over an edge (boundary) of the wire to be deleted and click. The wire will briefly be marked and then disappear. This is the fast, but dangerous way to delete. However, if you find you have deleted the wrong object, simply click on **UNDO**, which will reverse the last step.

The second, slower, but safer way is to select the object first. Click on **in**, which is a subheading under **SELECT**, and draw a box intersecting with any portion of the wire (or object). After you click, the entire wire will be marked. If you now click on **DELETE**, only the marked objects will be deleted. There are other choices under **SELECT**, including

the selection of an object which appears in a particular **layer**, but we will leave these details to the ICED manual.

You will notice that you cannot select (and, therefore, delete) any of the components or lines in the background. The entire chip is on the screen as one cell. However, should you select the edge of the chip, the whole chip will be marked and can be deleted.

Be careful to **UNSEL** (un-select) the object when you no longer intend to operate on it. If you leave the object selected and you go on selecting another object (perhaps in a different part of the chip) and operate on it, *all* selected objects will be affected.

Once you have selected an object you can also **MOVE** or **COPY** it if you wish. As the menus that appear when you click on these functions show, you can not only do this in the x or y direction but you can also rotate the object and produce mirror images.

Marking Components

To find your way around the chip you will need to mark or number at least the transistors. Click on **UseLay** and choose **CTEX** and then **TEXT hori s=%**. You now have the choice of how large you want to make the writing, say **10** (10 microns high). Then you type in the text, e.g. Q25, hit the **ENTER** button and place the box in a free area within the transistor. After clicking, the text will appear.
Once you select the **CTEX** layer, you will stay in it until you go through the **UseLay** routine again. So, don't forget to switch back to one of the metal layers before you place some more wires. (You will immediately notice that something is wrong, we have deliberately made the width of the CTEX layer very small).

Make sure that no text is entered in any of the metal layers. Such a text would be included in the mask streamfile.

There is also a cell called **PNP** under **ADD**, which you can use to distinguish PNP from NPN transistors. This cell can be found in the cell library using the **ADD cell** command. The first time you will need to the top level cell library using the **Next PATH**. Subsequently **ADD cell** will show PNP in the listing. After clicking on it, a box appears which fits just inside the transistor. After placing it click again and the transistor will be marked as a PNP. Similar cells named NPN and UNUSED are provided and required for LVS.

The Second Menu

You click the right-hand mouse button and a new menu appears. Starting at the top there is **SWAP**. Suppose you decide that the interconnection you just made should really have been in a different layer. Select it, click on **SWAP layers** and then click on the

two layers, the one the element is in now and the one you want it to be in. The color of the element will change and the element is in the new layer. Be sure to **UNSEL all** after you finish swapping.

The next group of commands in menu 2 is **TEMPLA**. Click on **TEMPLA screen** and you will see the ASCII file NEW32.CMD, which contains all the start-up choices. You will see, for example, that LAYER1 has the name WHIT, a width of 4 microns (for the wire), the screen color white, which is filled with pattern number 4, uses pen number 20 and will appear as layer 7 in the streamfile. You can change these choices either with a word processor (by loading the file NEW32.CMD) or through the LAYER command.

Each layout (i.e. chip) file has its own NEW32.CMD (the one present when the layout was created). If you alter the NEW32.CMD file, you will need to bring the changes into your layout by displaying it on the screen and then typing:

@NEW32

This will execute (@) the NEW32.CMD now in your directory. This file does not need the be called NEW32; for example you could copy it as LASER.CMD and change the pen setting to give different patterns for a laser printer. With the layout on the screen, you would then simply type @laser and the new command file would stay with the circuit until you change it again.

SHOW lets you see all the information of a selected item, either in an ASCII **file** or directly on the **screen**.

The next line in menu 2 is the **RULER**. With it you can measure precise distances, such as the spacing between two metal lines or the length of a metal run. If you want to measure a large distance precisely, there is a valuable feature in ICED which is best described with an example:

Let's say you want to measure the distance between a point in the upper left-hand corner and the lower left-hand corner of the chip. Click the right mouse button twice to get to the original menu. Now click on **VIEW all** to see the entire chip. Click the right mouse button to change the menu and the left button on **RULER**. You will find the cross-hairs pretty difficult to place precisely at this magnification. Now press the **ESC** button. A new menu appears and with it you can change the magnification without moving the cross-hairs. Try, for example, **BOX** which lets you enlarge the view. Once you have placed the starting point of the measurement by clicking, you can choose **ALL** through the **ESC** button sequence. This maneuver also works when drawing, moving or copying.

The **EDIT** group lets you get out of the present layout temporarily and work on a different cell.

Next down the line on the second menu is the **MERGE** command. If you want to join two wires, for example, you click on the command **MERGE wire** and then draw a box

intersecting the ends of the two wires you want to join. The two wires need not overlap or even touch. With **MERGE poly** you can join two polygons (which includes boxes); for polygons the sides must adjoin.

CUT does the opposite of merge. Here you need to select whether you want to cut along a vertical line (i.e. a wire running horizontally or in the x direction) or horizontally. You place the line which appears at the point which is to be cut and click.

The Third Menu

Click the right-hand mouse button again and you are in the third menu. We will only discuss a few of the remaining commands and leave the rest to the ICED manual:

PROTEC This feature protects parts or all of the layout from inadvertent changes. **UNPROT** releases the protection

BLANK lets you remove from view a layer or a selected item. **UNBLANK** brings them back.

LAYER allows you to change individual items in the NEW32.CMD. Suppose you don't like the fill pattern of the RED layer. Click on **LAYER**, **RED** and **pat'rn**, and choose the fill pattern you want. This change will then apply only to the present layout.

FILL lets you toggle between fill and outline.

When you click on **FILE**, **QUIT** will get you back to DOS without saving your layout. Through **FILE, EXIT** saves your work.

ICED has a number of safety features. As you draw a layout, all your moves are immediately saved in a file with the extension .jou (journal). When you EXIT, the journal file is given the extension .log, a backup file of the previous file with the extension .cl1 is created and the altered file is saved with the extension .cel. Should your work be interrupted by a power failure and thus a .jou file exists, ICED will let you recreate your work.

Printing and Plotting

After you click on **FILE**, you will also find the choice **PLOT**. Clicking on it produces a choice of printers and plotters. After selecting the printer, you choose either **none** or **sample**. The first one prints just outlines, the second fills the metal lines.

The fill pattern for each layer is chosen according to the pen number in the NEW32.CMD file or, for the displayed layout only, through the layer command. The fill patterns for some 36 pen numbers can be viewed in the file \ICED\SAMPLES\SAMPLE.DAT. A “-1” in the pen columns indicates that the layer will not be printed and 0 means outline only.

After choosing the paper size the program branches into four options. When clicking on **all** you get to choose which layers are to be printed, over how many pages you want to spread the plot (you can make large plots on an office printer by pasting the pages together), the orientation, filter size (i.e. what is the smallest feature you want to have printed), the cell depth limit (usually none) and the outline depth limit (usually 0, i.e. no cell outlines shown).

To print you exit the layout and type **MKPLOT** followed by the file name of the layout..

Suggestions for Your Layout

Near the upper right-hand corner on each chip (left side on 710) you will find the chip number. This is a 4 micron wide pattern in layer FMET. There is space for two more numbers. Contact us for the next number in sequence and write it in layer **WHIT** with the normal (4 micron) **wire** width. Note that this is not a text, but an actual metal pattern (text does not appear on the mask). We strongly suggest that you identify you chip with this number only (e.g. 73425) *and not with your product number*.

To avoid the most common errors, you might consider the following 6-point checklist:

1. In all single-emitter NPNs: is the base next to the emitter used?
2. In all dual and triple emitter NPNs: are both bases connected?
3. In a PNP, are all NPN emitters unconnected? Are all collectors (not bases) correctly connected?
4. Follow, in sequence, all path in RED, GRN, BLUE and YELL to make sure no connection of other colors cross.
5. Trace the continuity of resistor strings.
6. Check current densities for high-current paths. You should not exceed 4mA per micron width.

DRC (Design Rule Checking)

ICED has an optional DRC software package which will check for proper spacing between unconnected metal areas. We have prepared a checking-rule file to do this: METAL3.RUL. This file is in the directory \Layout.

The first time you perform a DRC, you need to type

D3RULES METAL3

This creates two additional files which are used in the actual checking.

Next, with the circuit displayed in ICED, click on **FILE**, **DRC386** and **full**. This produces a

new circuit file with a .pok extension. Exit ICED (but stay in the \Layout directory) and type:

DRC3 METAL3 XXXX ERR SLOW

The first word specifies the DRC program while the second one calls out the rules file. XXXX is the name of your layout file (e.g. TRIAL, no extension required) and ERR is the file where errors will be placed (you can give it any name). In the last word you have the choice between QUICKPASS and SLOW. Since DRC is quite fast for the relatively small 700 Series chip, we recommend the more thorough SLOW.

If the DRC program indicates that there are errors, type:

IC ERRORS (or any cell name that does not exist yet)
@ERR which places the errors on the screen (you may have to click on **View**, **all** to make them visible)

To see where exactly the rules are violated, bring in your layout:

ADD
at 0,0
XXXX

Make sure you correct the mistakes in your actual layout and not in the file ERRORS.

ICED LVS

Preparing the Schematic for LVS

To determine if your layout corresponds to the schematic you need first of all a netlist extracted from the schematic. It is recommended that you create a copy of the simulation schematic with a different name and use this copy for LVS. Then:

a. remove all external components connected to the pads. Although the LVS program will ignore such elements as power supplies or input signals, it will include any capacitor, resistor or inductor. To avoid possible confusion, it is best to simply strip off all external items.

b. remove cross-unders. In LVS they are considered shorts.

After the schematic is clean, create the netlist.

Three examples have been provided: 710xx, 711xx and 712xx.

Preparing the Layout

All NPNP transistors must be marked with one of three cells: **NPN**, **PNP** or **UNUSED**. This applies only to the transistors which can be either NPN or PNP. NPN transistors with Schottky diodes need not be marked.

The content of the file ICEDLVS needs to be copied into a separate directory (here assumed to be \700LVS). Although the cells have the same names as those in ICEDLayout, they have in fact been modified.

Now copy your layout to the directory \700LVS and bring it up on the screen in this directory. You will notice that several items have been added to the chip underneath the layout:

- All cross-unders are connected through layer 71 (a faintly visible, brown layer). This layer makes contact with layer 17, the fixed metal. Thus, the fixed metal layer should never be used for interconnections.

- All substrate terminals are connected together, also in layer 71

- The N-layer contacts are connected together.

The latter two form hidden nets.

Now click on **FILE**, **DRC386** and **full**. This creates a file with the extension .pok.

THE FOUR LVS STEPS

Note: In Simetrix the steps described below are available from a pull-down menu (see chapter 7S).

Converting the Schematic Netlist: LSTEP1

In \700LVS type: **LSTEP1** (using either upper or lower case)

The LSTEP1 program will ask you where the Schematics netlist is located and what its name is (this needs to be a P-Spice compatible netlist). There is a default listing, which is simply the last file used. Make sure you type out the entire path and the full name, with extension.

After hitting return, the program will ask you which LVS model file is to be used. If lvsmod.22 is the current one (the default), simply press return.

The program now puts the model file into the netlist, combines all components in one subcircuit and saves the new netlist (under the same name) in the \700LVS directory.

Running DRC (Design Rule Checking): LSTEP2

(Note: The first time you use the 700drc file, type: D3RULES 700drc3. This checks the rules file and creates a binary file.)

Type: **LSTEP2**

The program will ask you first which DRC program is to be used (DRC3), and then which DRC rules apply (700DRC3).

Next you are asked to identify the layout to be checked. If you have used the same name for the netlist as for the layout, the default will be correct. Press return.

The DRC program now runs a DRC check, using the rules 700DRC3 and creates a file with the same name but a .p9k extension. Note that this step only checks if all the cells have been marked properly, it is not a spacings check. You will need to run a separate DRC check using the METAL3 rules.

Running NLE (Net List Extractor): LSTEP3

(Note: The first time you use the 700nle file, type: rulesnle 700nle).

Type: **LSTEP3**

The program asks you which NLE rules are to be used. The default 700nle is an ASCII file created for the 700 Series, which defines devices, layers and connections between layers. Accept it by pressing return.

The second question identifies the layout to be extracted. The default is the one used in LSTEP2. Hit return.

The NLE program takes the file with the .pok extension and converts it into a file with the same name but a .ext extension (the extracted netlist). Both are binary files.

When running the NLE on the smallest chip, the 710 you will get a warning message, saying that there is no LPNP (large PNP transistor). All other chips have at least one of each component; the 710 does not have a LPNP.

Running LVS (Layout vs Schematic): LSTEP4

Type: **LSTEP4**

In this last step the two netlists are compared.

Reading LVS Errors

LVS errors are reported in the subdirectory \RESULTS. Of particular interest are UNMATCH.LVS, an ASCII file which describes differences between devices and nets, and PARAM.LVS, which lists discrepancies in parameters.

Perhaps the easiest way to trace errors is to note the nodes numbers in the section "The following LAYOUT NETS were not matched" and then bring the layout up on the screen and type:

@nodes

When you then type the command n0, the program will ask you for a node number (e.g. a node which does not match in UNMATCH.LVS), then highlights it and flashes it for a few seconds. Now you can see in great detail where this node does not match the corresponding on of the schematic.

The typed command **bli** blinks the highlighted node again and **nd** removes the highlight. With **nn** you can click on a connection and the program will tell you the node number.

After all the nets match, check the file PARAM.LVS to make sure the values (e.g. the number of resistors or emitters) match as well.

Note that the LVS program does not look for a particular device to match a specific device in the schematic. For example, it will only tell you that a correct combination of devices are connected the a pad, not a specific pad. If you want to make sure that specific nodes correspond (say a few of the pads), identify the metal line running to the pad in the layout by writing on top of it (say in TEXT, 3 microns tall) in the same layer as the connection.

Next open the ASCII file EQUIV.LVS. You will see that it contains the single comment "*none", i.e. it is empty. Replace it with the equivalence statements, e.g.

```
$N_0001 = PADE
$N_0007 = PADP  etc.
```

on the left side of the equation is the node identification as it appears in the netlist (the example is P-Spice) and on the right side the text of your marking (whatever you choose it to be). You can do this for as many nodes as you like.

Making a Streamfile

To make a mask, your metal pattern will have to be in the Calma GDSII streamfile format. With your chip displayed on the screen, type:

@NEW32

Which will refresh the NEW32.CMD for the present layout and assign WHIT, RED, GRN, BLUE, YELL and FMET to streamfile layer 7 (if NEW32.CMD is set up that way). The assignment only stays as long as the chip is displayed; next time you call up the pattern all streamfile assignments are empty.

You can now check the layer assignments by going to the second menu and clicking on **TEMPLA screen**. You will see that, for these six layers, STREAM = 7 (layer 7 in the mask sequence). All other layers should say NO_STREAM.

Then click on **FILE**, **STREAM** and **none**. You will now have a file with the extension .SF in your directory, which you can copy to a floppy disk or send to us over a modem or the Internet.

It is very important that you take the streamfile from the directory \Layout, not \700LVS. The extra patterns required for LVS produce small squares in the layer FMET, which would show up on the mask.

To transfer a design to a different program, you can choose **ARCHIV** instead of **STREAM**. This also creates a GDSII streamfile, but instead of following the layer assignments in NEW32.CMD, it take the layers as they are.

If you want to convert a streamfile back into ICED format type **UNSTR32** and accept all defaults. This DOS program creates an intermediate file with an .ali extension, which can be checked if there are problems. You then type the command a second time and again accept all defaults. We strongly suggest that you unstream in a separate directory; unstreaming recreates all cells and replaces the ones you already have. You then take the top cell (the metal pattern) and copy it to your regular directory and discard the rest. The top cell will then call up your regular basic cells.

Streamfiles for Other Layout Software

If you are using layout software other than ICED, you start with the GDSII streamfiles in the directory “Streamfiles”. The layer assignments are:

1 WHIT	5 YELL	9 RD
2 RED	6 CTEX	10 BG
3 GRN	7 RT	17 FMET
4 BLUE	8 GN	

Chapter 9

Packaging

The table below lists our most commonly used packages for integrated circuits. This is, by no means an exhaustive list, there are many more package types and pin counts. Ceramic packages generally are only used for prototyping, so there is a more limited choice of these package types.

Standard Packages for 700 Series

Pins	3	8	14	16	18	20	24	28	40	44	48
Plastic											
Dual In-Line, 300 mil		8	14	16	18	20	24				
Dual In-Line, 600 mil							24	28	40		48
Small Outline, narrow	3	8	14	16							
Small Outline, wide				16		20	24	28			
Ceramic											
Dual In-Line, 300 mil		8		16							
Dual In-Line, 600 mil							24	28			48

On the following pages you will find outline drawings, blank assembly diagrams and chip outlines. The package drawings are for initial selection only, we give no tolerances and our dimensions and thermal ratings are on the conservative side.

For each package type and pin-count there is an assembly diagram, drawn at 20X. The chip outlines are drawn to the same scale. You may copy these

pages (they are excluded from the copyright and, for this reason, we have not numbered them). We suggest you compose an assembly diagram by also copying the chip outlines, cutting out the chip you need and paste it onto the center. You then draw lines between the chip pads and the package bonding areas.

There are a few rules you must follow:

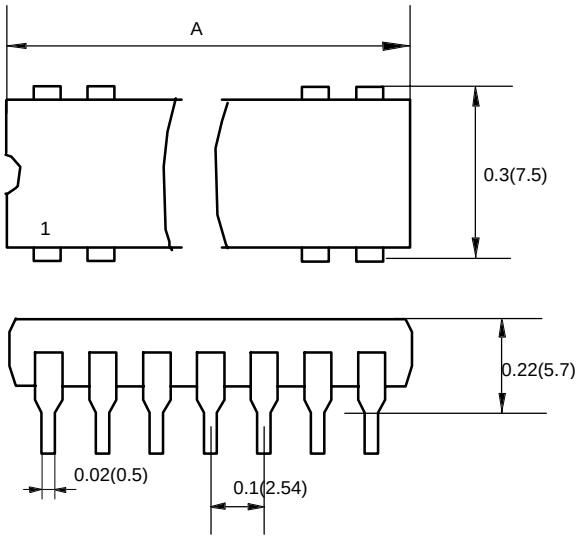
1. The pad area of the package (the rectangular area in the center) must be larger than the chip by:

without bond to pad - 10 mils
with bond to pad - 30 mils

2. You must be able to draw straight lines between near-edge chip pads and package bonding areas and these lines may never cross.

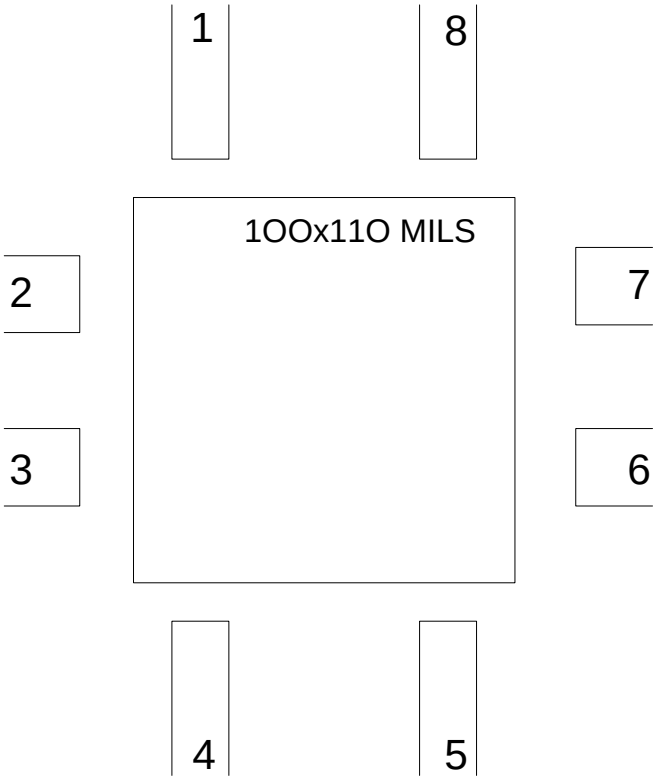
3. Orient the chip at right angles (but any orientation).

300 MIL Dual In-Line



Pins	A inch. (mm)	Thermal Resistance oC/W
8	0.4 (10)	70-120
14	0.78 (20)	50-90
16	0.78 (20)	40-80
18	0.9 (23)	35-75
20	1.03 (26)	30-70
24	1.28 (32.5)	30-70

ASSEMBLY DIAGRAM



8-PIN DUAL IN-LINE

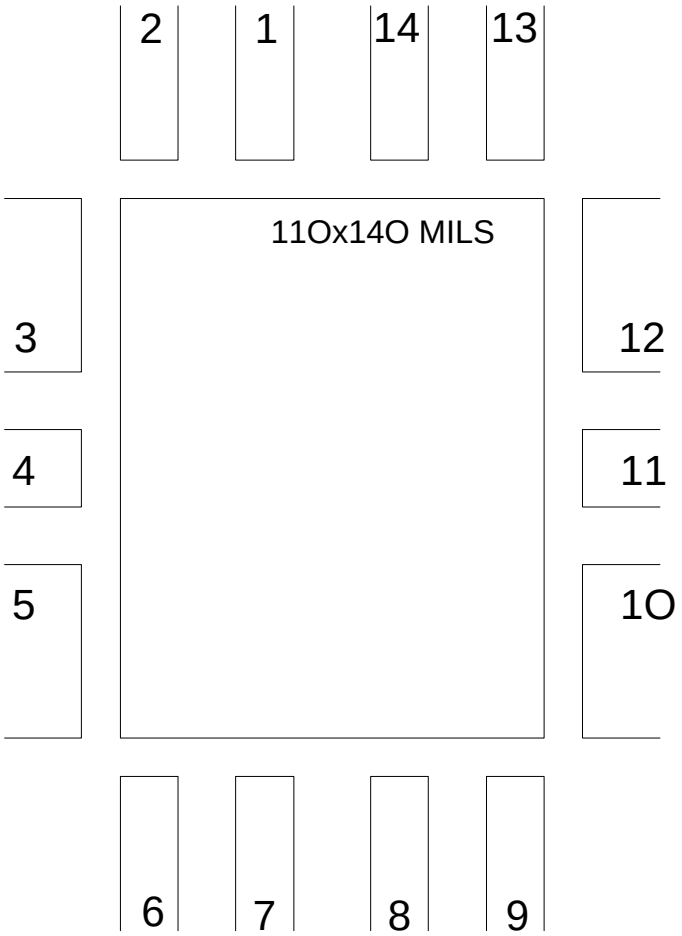
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



14-PIN DUAL IN-LINE

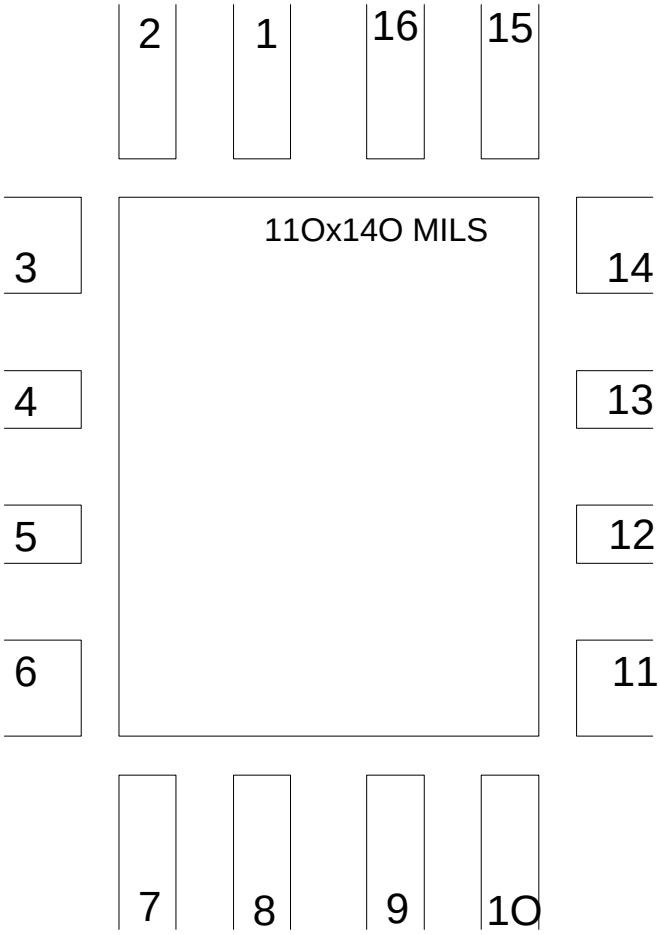
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



16-PIN DUAL IN-LINE

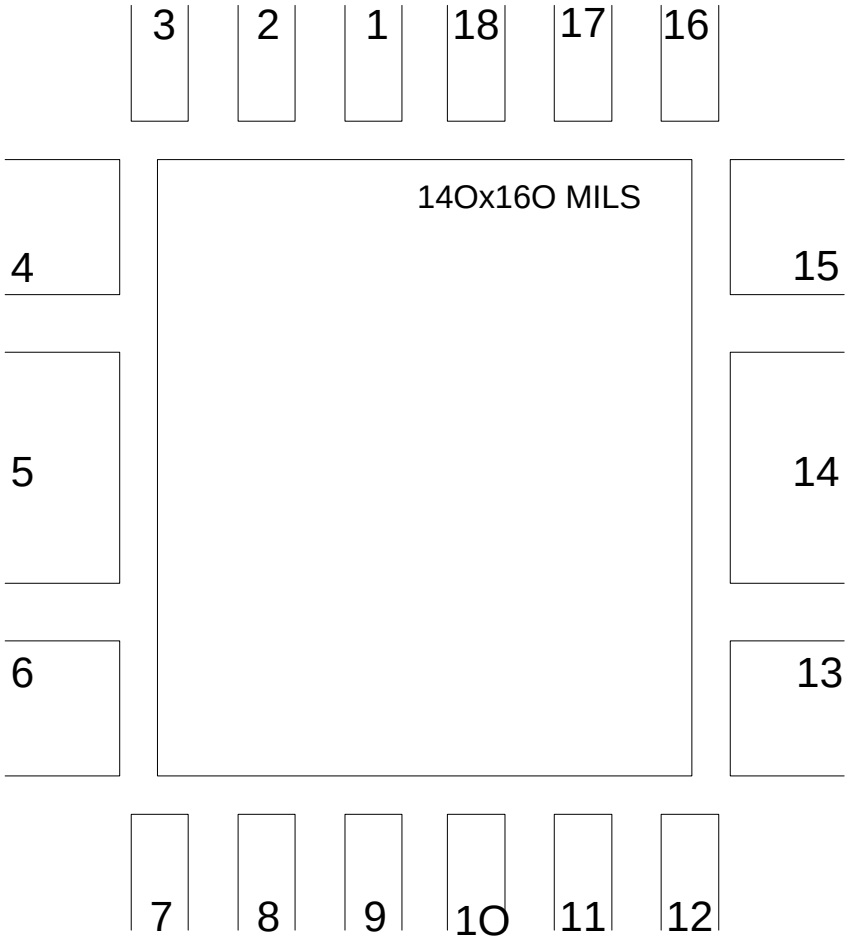
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



18-PIN DUAL IN-LINE

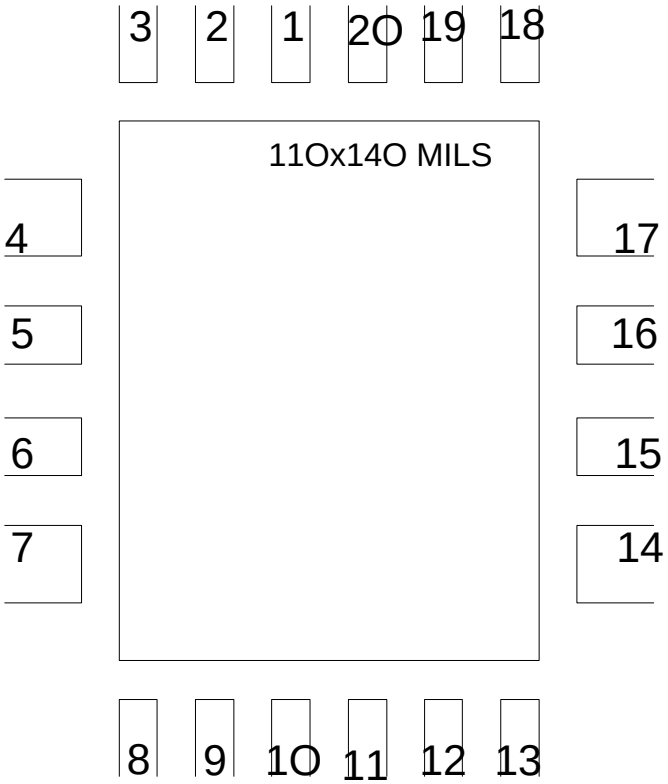
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



2O-PIN DUAL IN-LINE

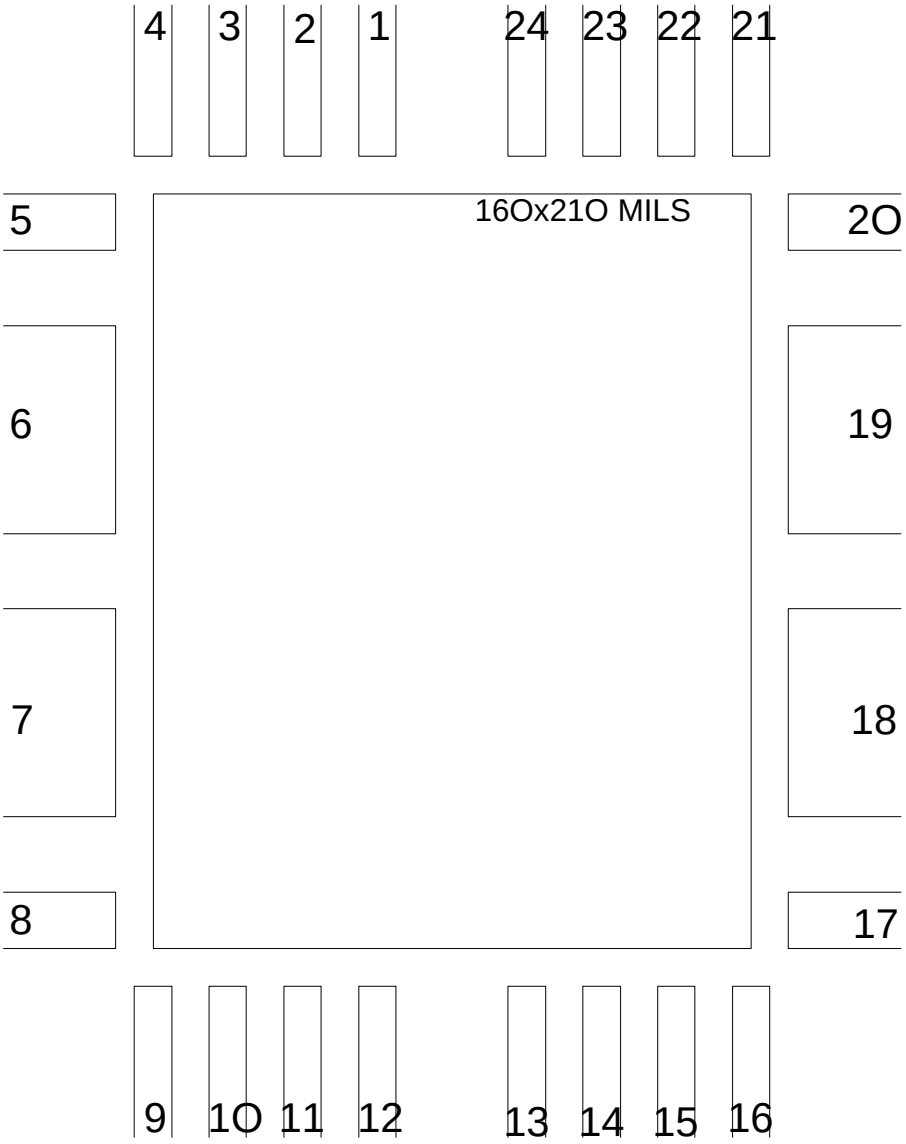
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



24-PIN NARROW DUAL IN-LINE

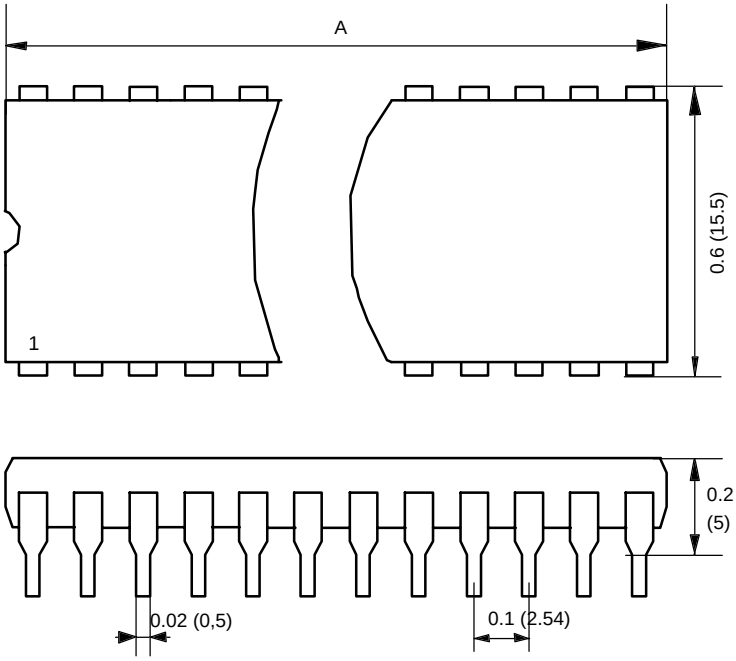
DIE IDENTIFICATION:

PACKAGE MARKING:

DATE: ORIGINATOR:

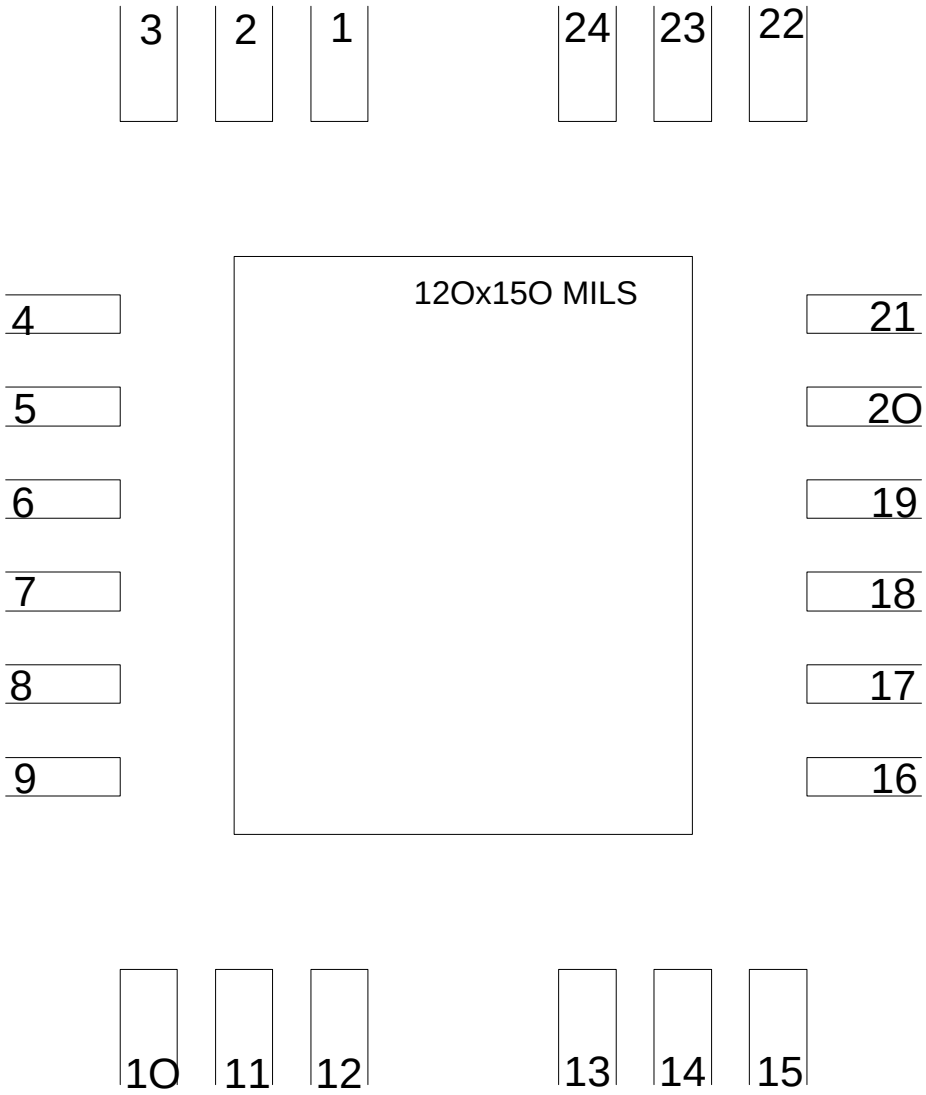
COMPANY:

600 MIL Dual In-Line



Pins	A inch, (mm)	Thermal Resistance oC\W
24	1.25 (32)	40-60
28	1.45 (37)	40-60
40	2.06 (53)	35-55
48	2.43 (62)	30-50

ASSEMBLY DIAGRAM



24-PIN WIDE DUAL IN-LINE

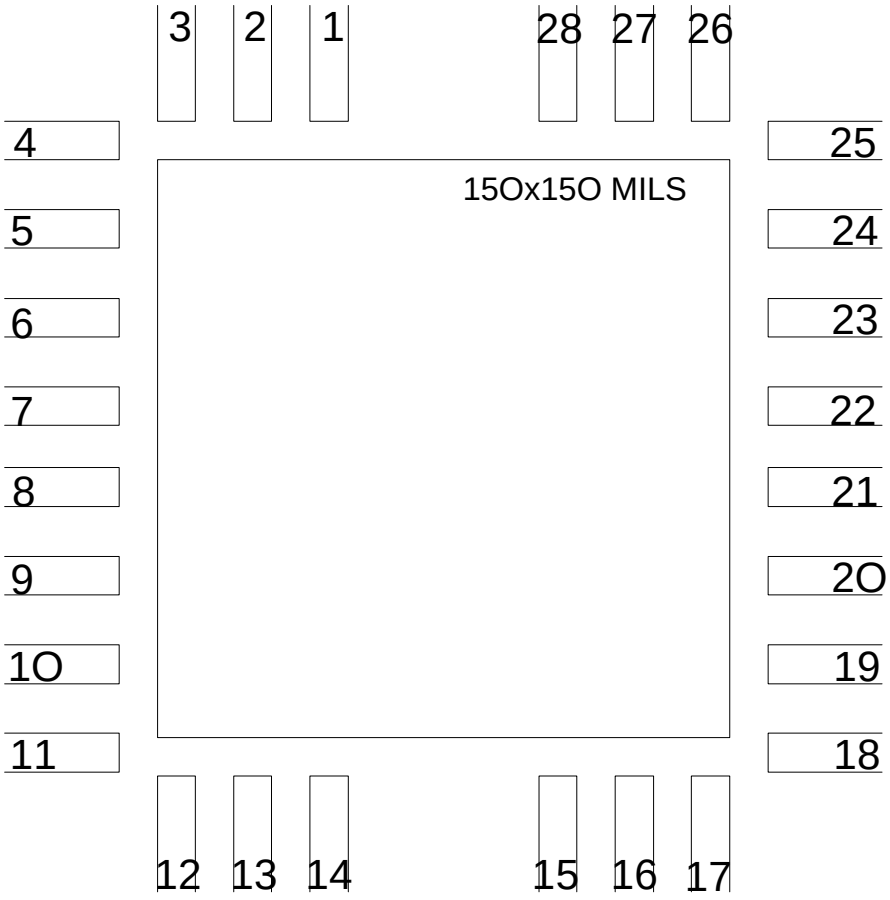
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



28-PIN DUAL IN-LINE

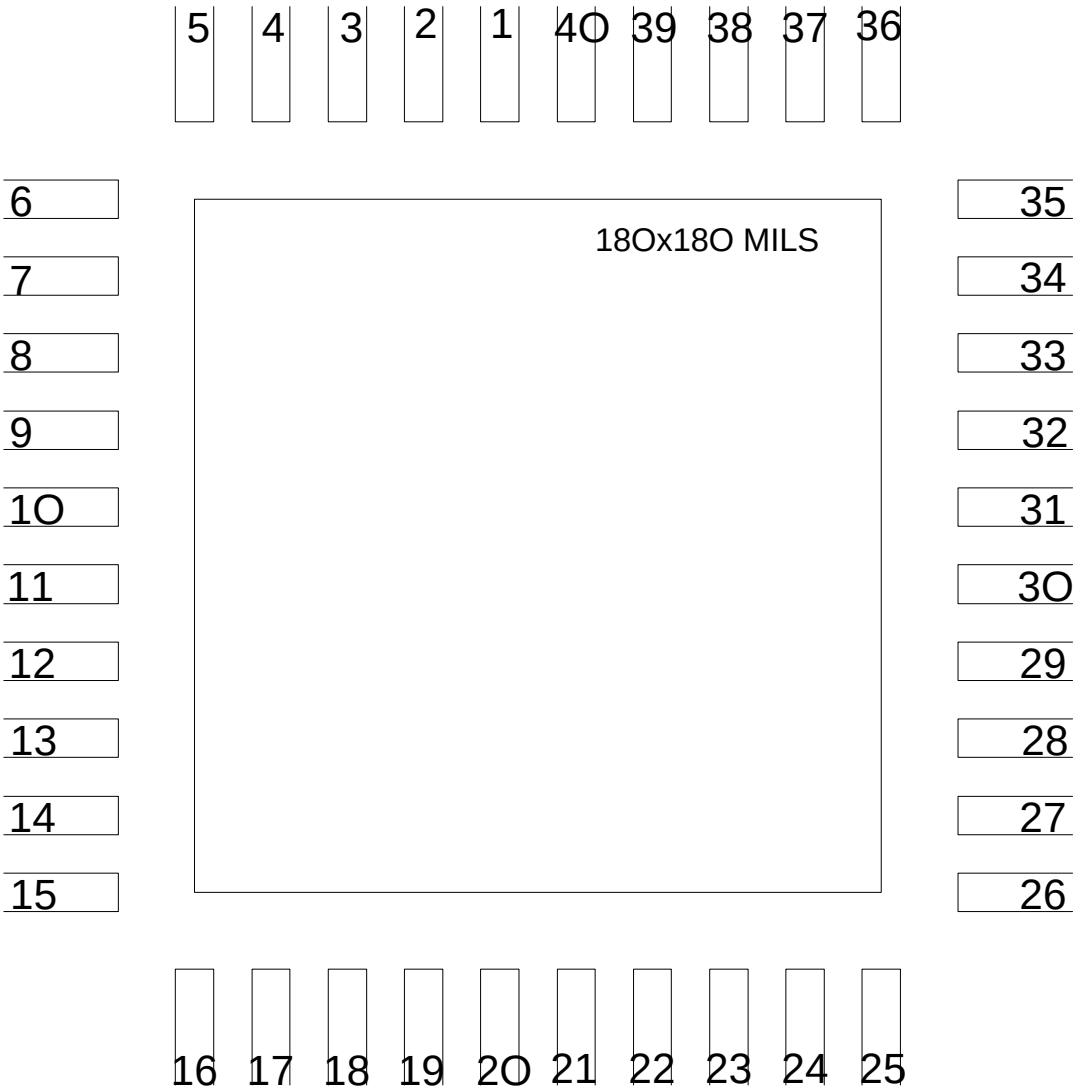
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



4O-PIN DUAL IN-LINE

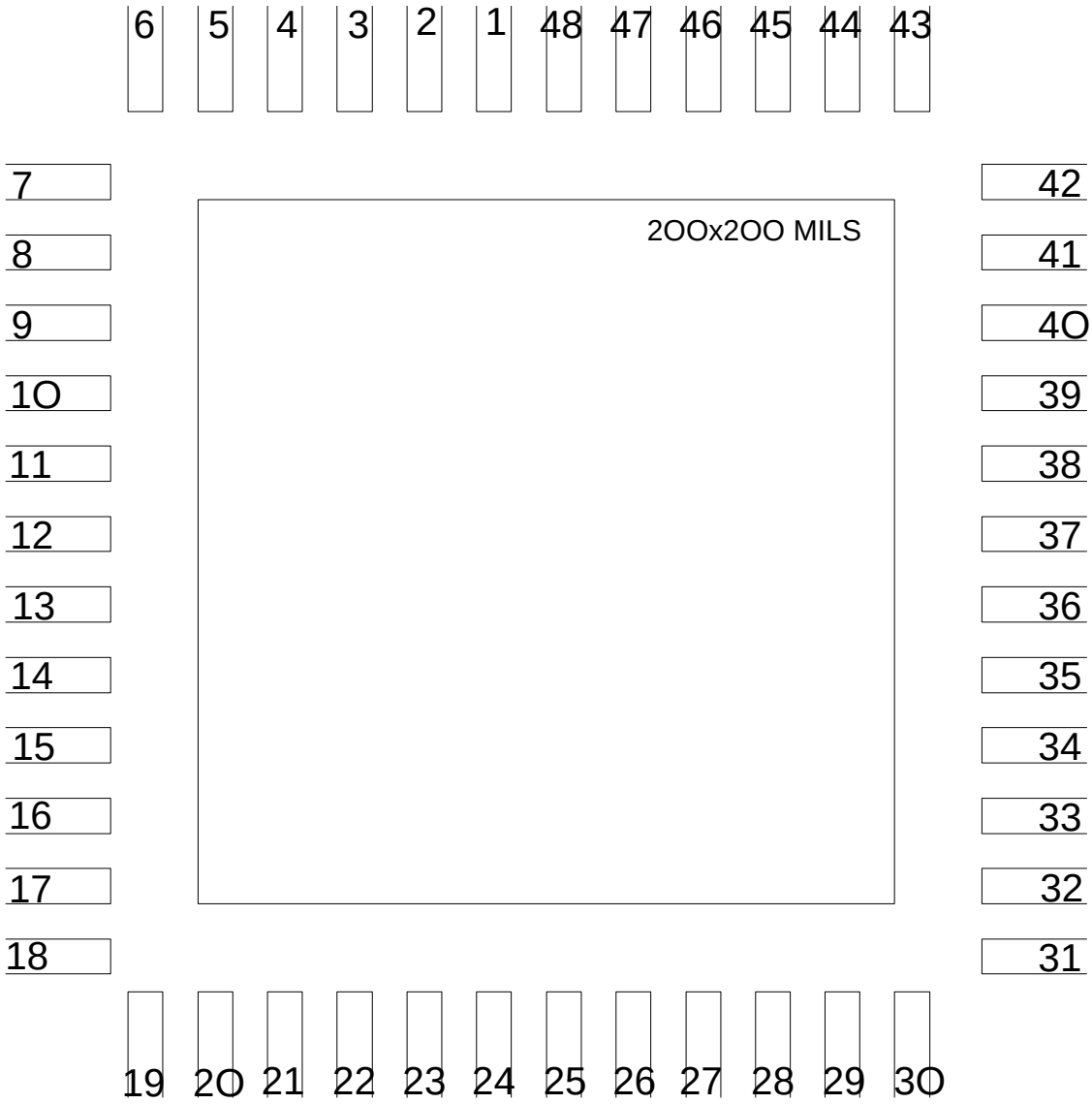
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



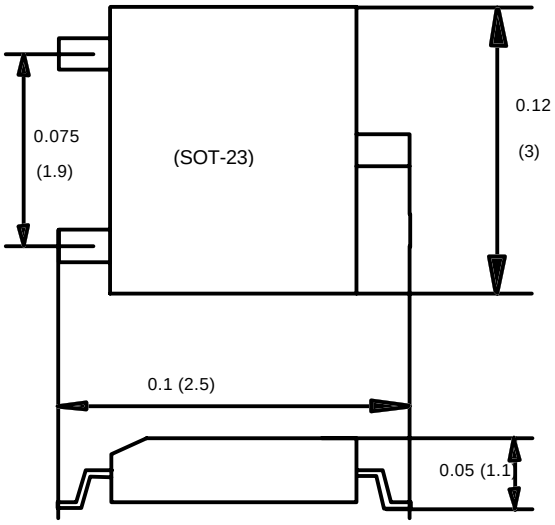
48-PIN DUAL IN-LINE

DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:



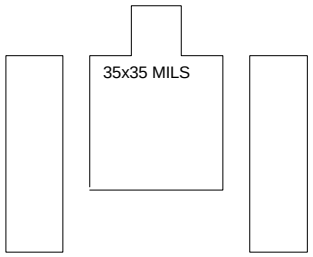
300-500 oC/W

3-Pin

Small Outline

40-70 oC/W

ASSEMBLY DIAGRAM



SOT-23

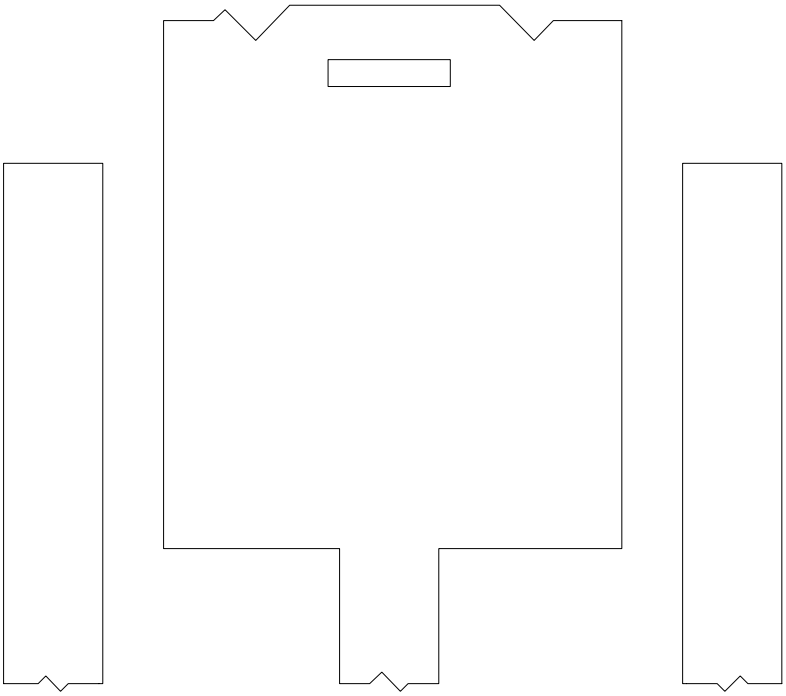
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



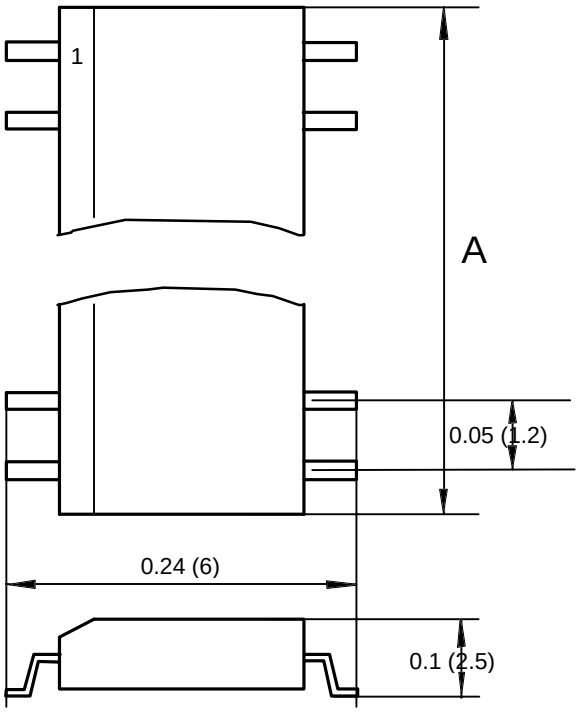
SOT-223

DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

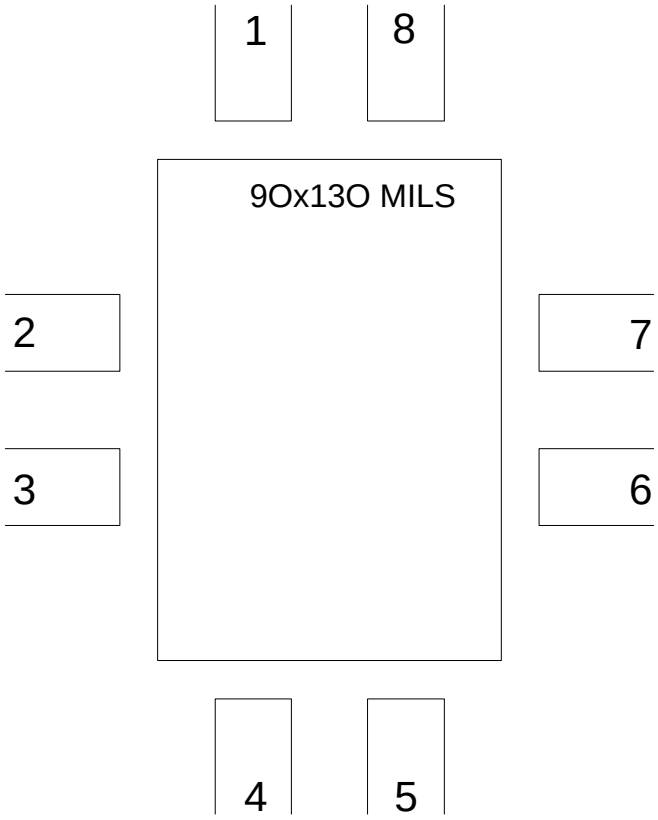
COMPANY:



Small Outline Narrow

Pins	A inch. (mm)	Termal Resistance oC/W
8	0.19 (5)	150-180
14	0.34 (8.6)	110-130
16	0.39 (10)	100-120

ASSEMBLY DIAGRAM



8-LEAD SMALL OUTLINE (SO), NARROW

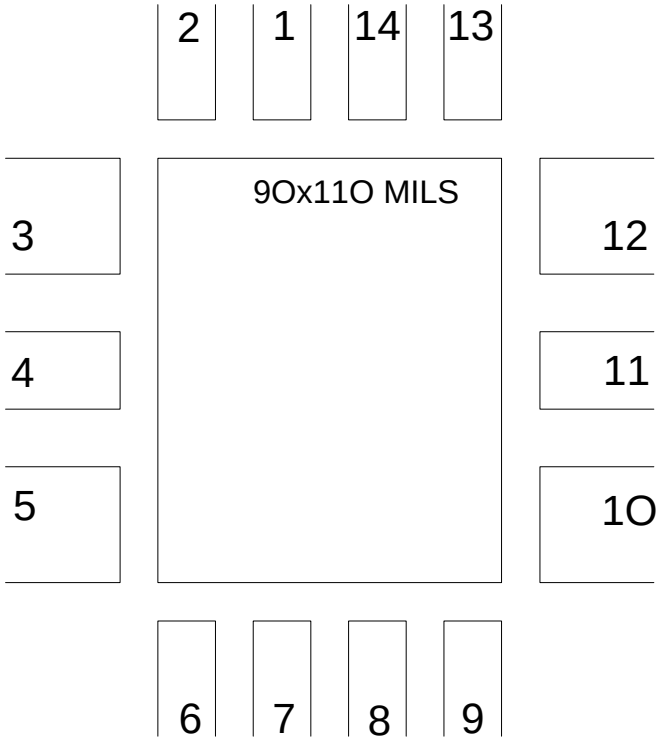
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



14-LEAD SMALL OUTLINE (SO), NARROW

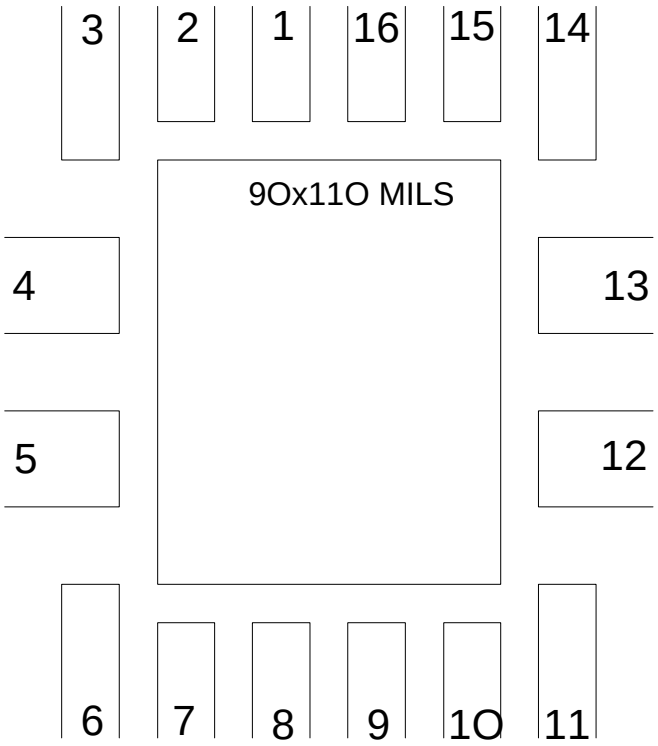
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



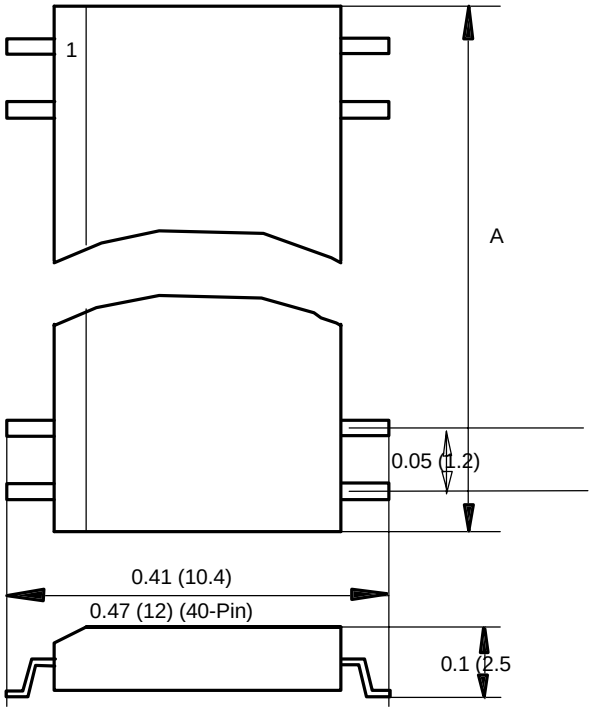
16-LEAD SMALL OUTLINE (SO), NARROW

DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

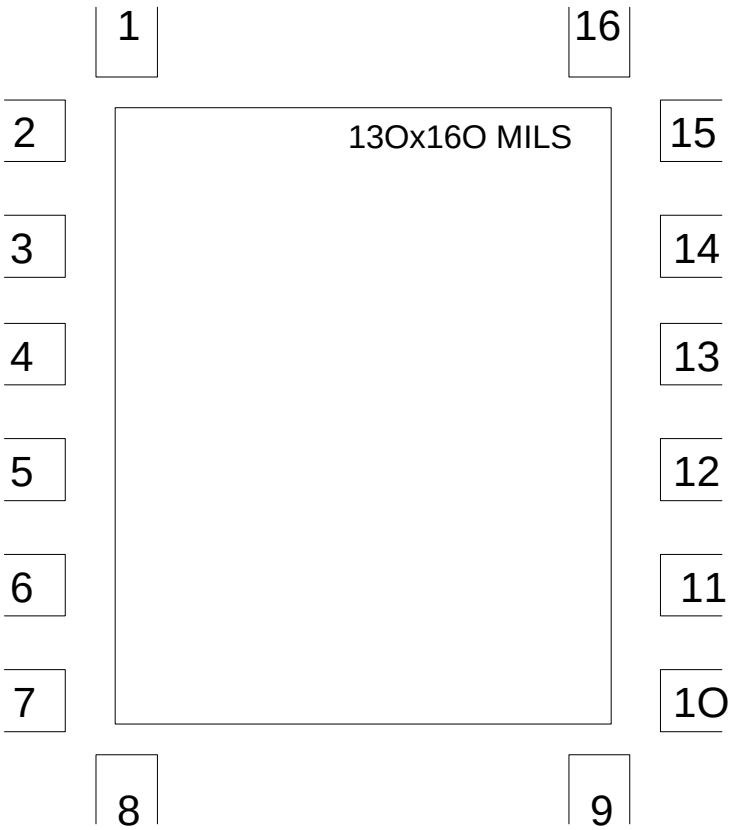


Small Outline

Wide

Pins	A inch. (mm)	Thermal Resistance oC/W
16	0.40 (10.3)	80-100
20	0.51 (12.9)	70-90
24	0.61 (15.5)	65-85
28	0.71 (18.1)	65-85
40	1.04 (26.4)	60-80

ASSEMBLY DIAGRAM



16-LEAD SMALL OUTLINE (SO), WIDE

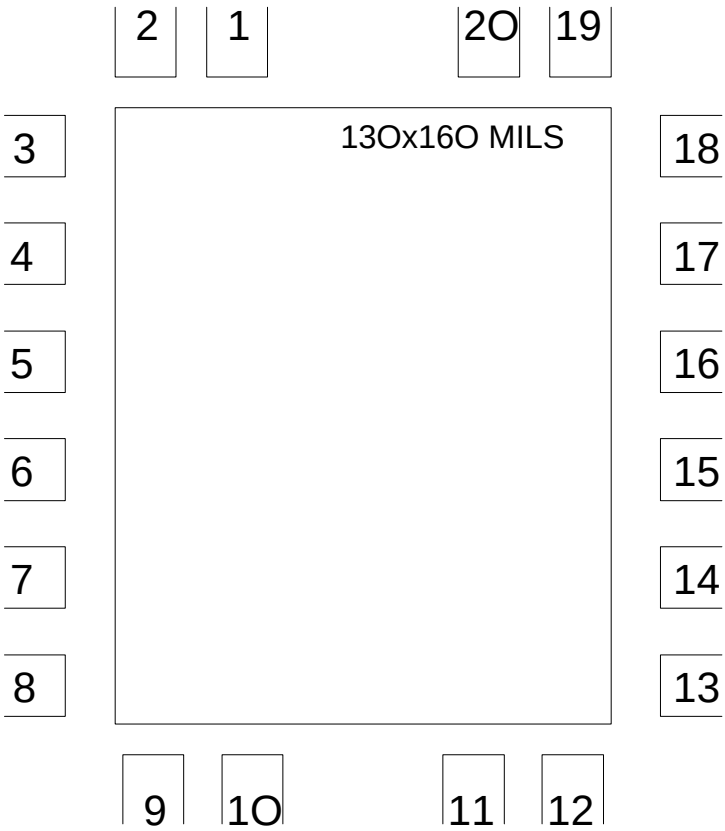
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



2O-LEAD SMALL OUTLINE (SO), WIDE

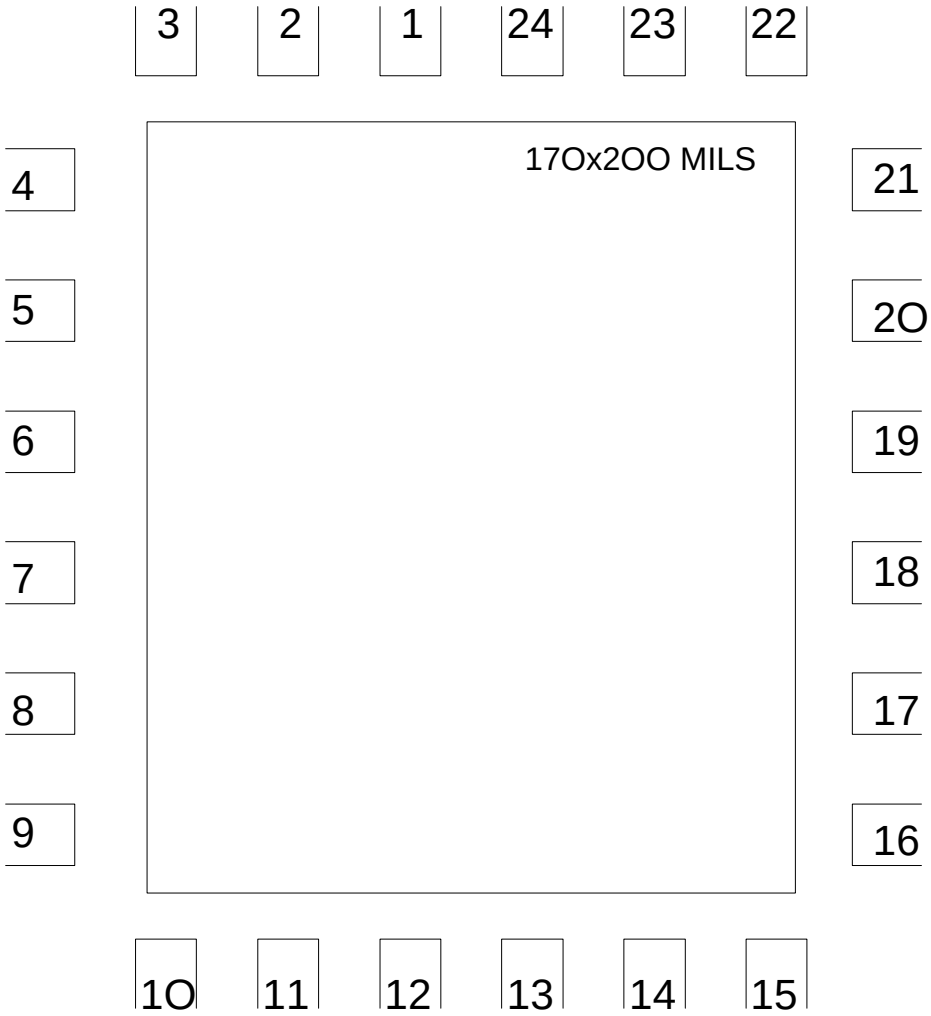
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



24-LEAD SMALL OUTLINE (SO), WIDE

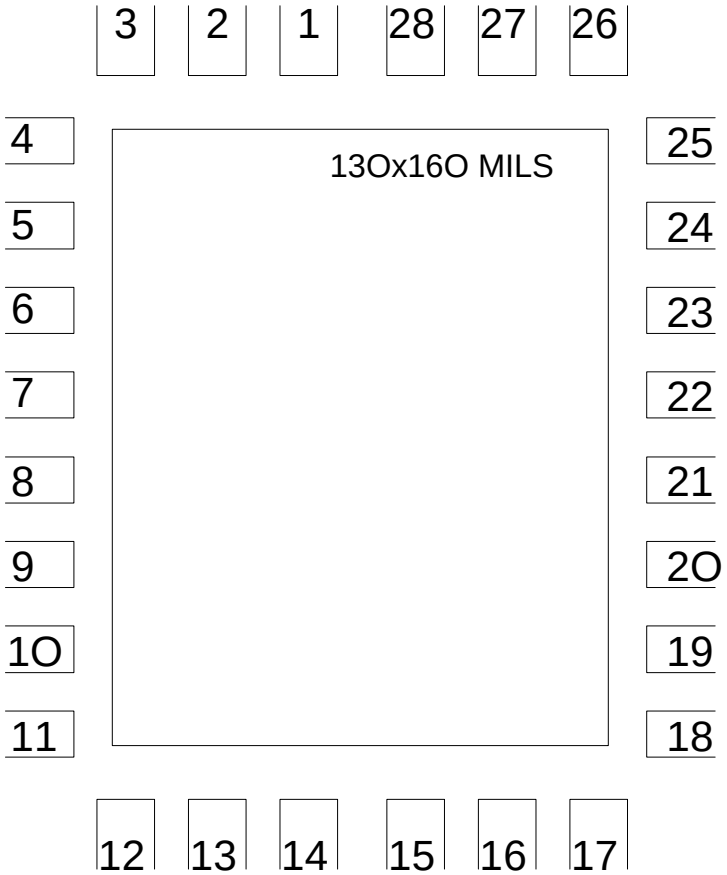
DIE IDENTIFICATION: DIE SIZE:

PACKAGE MARKING:

DATE: ORIGINATOR:

COMPANY:

ASSEMBLY DIAGRAM



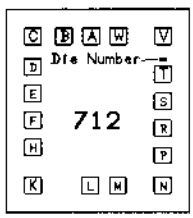
28-LEAD SMALL OUTLINE (SO), WIDE

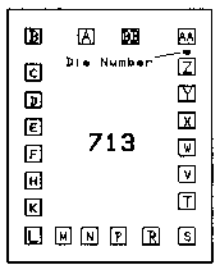
DIE IDENTIFICATION: DIE SIZE:

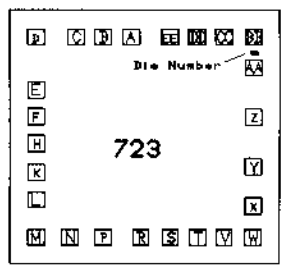
PACKAGE MARKING:

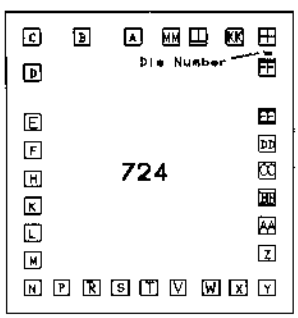
DATE: ORIGINATOR:

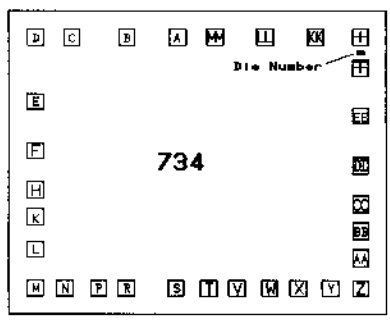
COMPANY:

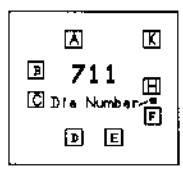

Die Number
712


Die Number
713

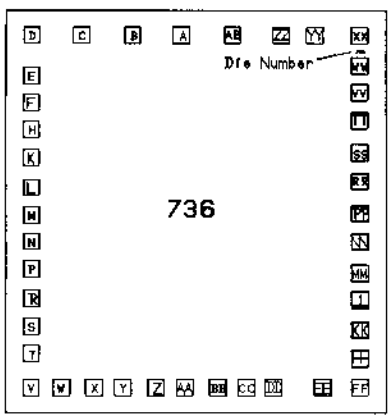

Die Number
723

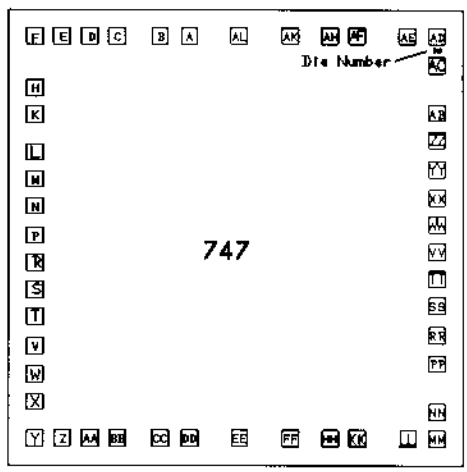

Die Number
724


Die Number
734


711
Die Number


Die Number
710


Die Number
736


Die Number
747

Prototypes and Production

The end-product of your design is a streamfile, which will be used to drive a pattern generator. This machine produces a reticle, an enlarged image of the metal pattern. This reticle will then be reduced to the final size and stepped. After your mask is complete it is used to pattern a single wafer.

At five places the image of the metal pattern will be replaced by that of a standard test pattern. This test pattern (also called PCM) is an essential part of the 700 Series concept. It uses devices identical to those in your circuit and provides access to them. The test pattern is the *only* technical interface between you and the wafer manufacturer during the prototype phase. The same test pattern is used for all nine chips and, for a wafer to be acceptable, *three out of the five patterns must meet all specifications* (shown on next page). With the purchase of a prototype wafer or packaged devices you are entitled to a printout of the test pattern measurements.

It is only in this way that we can avoid confusion when there is a problem with a circuit. If the devices on the wafer are within specifications, yet there is a problem with the prototype run, it is almost certain that the investigation should focus on the design.

To test, evaluate or trouble-shoot your circuit on the wafers, there two methods available. The first one employs a movable probe, which can be put down anyplace on the circuit (though, if there is a glass layer it becomes very difficult to penetrate it with the probe). This method is usually limited to about 4 probes and is, therefore, useful only for examination of a circuit detail.

The second method employs a probecard. This is a printed circuit board with fixed probes, aligned to make contact to all the pads. Some probers allow a small number of movable probes to be used in addition to the probecard.

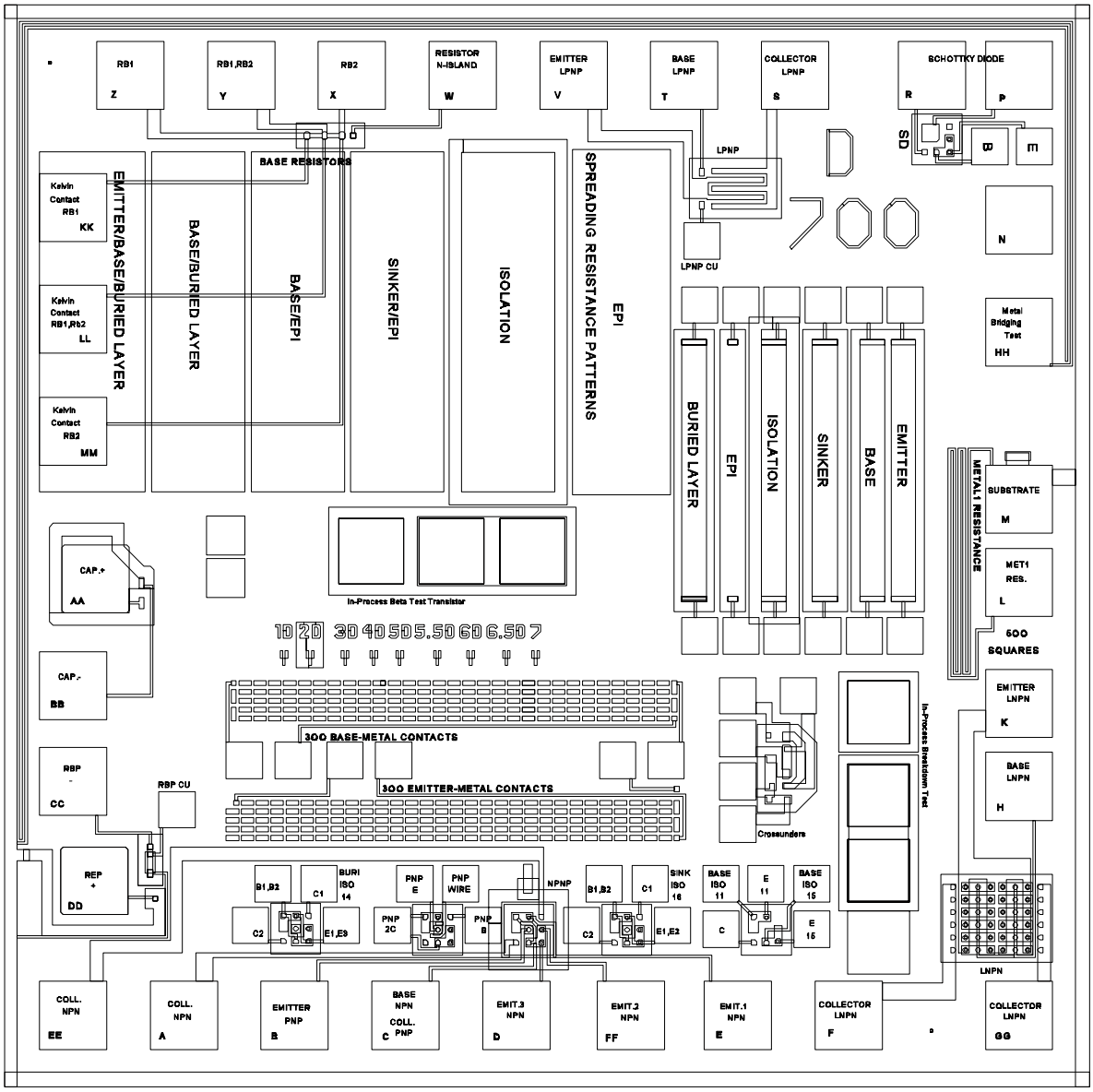
Test Pattern Evaluation
700 Series Linear Bipolar Arrays

Acceptance: All parameters in at least 3 test sites must pass.

	Device	Parameter, Conditions	Limits, Units	1top	2left	3center	4right	5bottom
1	NPN	BVCEO @ 10uA	20 min, Volts					
2		hFE, Ib= 1uA, Vce= 5V	100-350					
3		VBE @ 10uA	700 max, mV					
4		ICEO @ 20V	10 max, nA					
5		BVEBO @ 10uA	5.6-6.1, Volts					
6		RC	200 max, Ohms					
7	PNP	BVCEO @ 10uA	20 min, Volts					
8		hFE, Ib= 1uA, Vce= 5V	50-200					
9		ICEO @ 20V	10 max, nA					
10	LNPN	BVCEO @ 10uA	20 min, Volts					
11		hFE, Ib= 2mA, Vce= 5V	40-150					
12		ICEO @ 20V	50 max, nA					
13	LPNP	BVCEO @ 10uA	20, Volts					
14		hFE, Ib= 100uA, VCE= 5V	5 min					
15		ICEO @ 20V	20 max, nA					
16	RB1	Base Resistor	600-900, Ohms					
17		Delta RB1/RB2	%					
18	SD	Vf @ 10uA	300-400, mV					
19		Ir @ 20V	50 max, nA					
20	RBP	I @ 5V	10-150, uA					
21	REP	I @ 20V	1-8, uA					
22	MET	Resistance, 500 squares	40 max, Ohms					
23	JCAP	BV @ 10uA	9 min, Volts					
24		Ir @ 5V	50 max, nA					

Legend: NPN (NPN transistor with one emitter, NPN1), PNP (lateral PNP transistor), LNPN (large NPN transistor), LPNP (large PNP transistor), RB (base resistor, 750Å), SD (Schottky diode), RBP (base pinch resistor), REP (epi pinch resistor), MET (Metal interconnection layer), JCAP (junction capacitor), BVCEO (breakdown voltage collector-emitter with base open), hFE (common emitter current gain), VBE (base-emitter diode voltage with collector connected to base), ICEO (leakage current collector-emitter with base open), BVEBO (breakdown voltage emitter-base, Zener voltage), RC (resistance between the two collector contacts of the NPN transistor), Delta R1/R2 (difference between two identical resistors, close together), Vf (forward voltage drop), Ir (reverse leakage current), BV (breakdown voltage).

700 Series Test Pattern



Prototype and Production Testing

You should sharply distinguish between prototype and production testing. For prototypes it is neither necessary nor advisable to have an elaborate test. The test should eliminate circuits which are malfunctioning, but it should not discard circuits which are out of specifications because of the normal parameter variation. You will need to know what this variation is so that you can judge the success of your design. For this reason a simple DC test is sufficient for the first wafer. Often prototypes are not tested at all.

What follows next is one of the most important aspects of IC design: *You must never set the acceptance limits based on the evaluation of the prototypes*. This is one of the worst mistakes you can make. The reason is simple: Your measurements are based on a single wafer, which is too small a sample to make any kind of statistical judgement. Trust your Monte Carlo analysis, it tells you much more.

Testing production units is a different story. Here you need to eliminate the circuits whose device parameters are outside the 3-sigma limit. As we pointed out in chapters 3 and 5, the parameter distribution does not stop at ± 3 sigma, there is a small percentage outside this limit. These variations, multiplied by the number of parameters, will show up as circuits which do not meet specifications; the only way to catch them is to test every device.

We will most likely agree to generate a number of pre-production ICs. These devices will cover the normal process parameter variations and will allow the test program to be refined until an acceptable yield results. Pre-production ICs also let you check parameters which are too difficult for an automatic tester, such as noise or thermal effects.

Be aware that the setting up of a test program is quoted separately.