Part 7: Bipolar

BJT Advantages

BJT small signal model

The HBT

HBT applications

BiCMOS

Bipolar logic

Bipolar Junction Transistor

First solid-state amplifier device (Bardeen, Brattain and Shockley, Bell Laboratories 1948). Was used to create the first integrated circuits (TTL, of more specifically ECL, see later)

Largely replaced by MOSFETs in logic and switching applications, since BJTs consume more power in their on-state. Yet BJTs still have a major role in small signal amplification, analogue control and power switching.

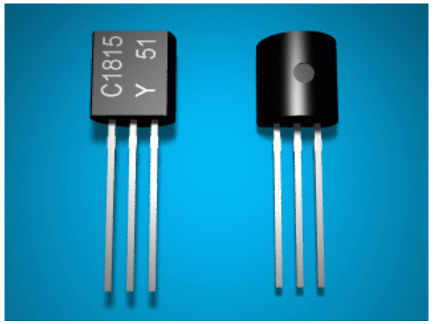
BJT Advantages

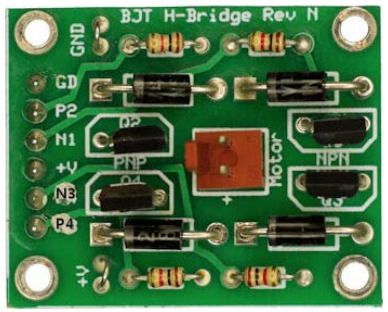
- Tend to have larger transconductance and output resistance values than MOSFETs.
- Tend to have better power handling capabilities
- Can offer excellent high frequency performance (see HBT, later)

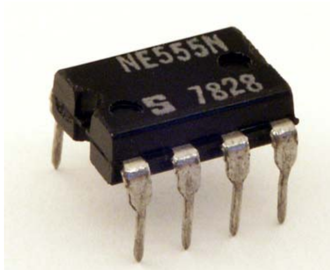
Bipolar Junction Transistor

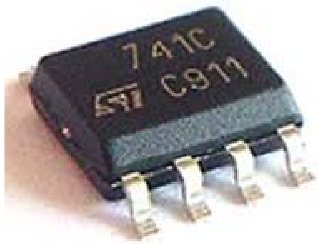
- Can exhibit lower noise
- Available in many types for custom circuit design

For these reasons the BJT is still a very common device used in discrete circuits and operational amplifiers

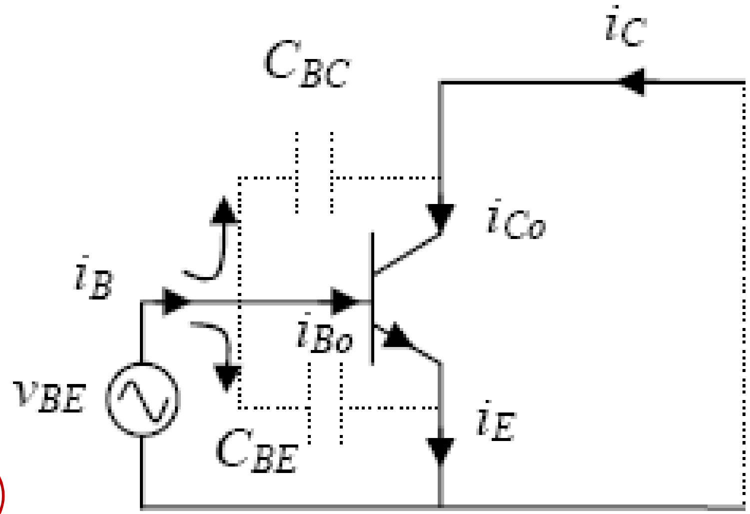








BJT Small-signal model



Need a small signal model for the BJT (in a similar way to the MOSFET)

$$i_B = i_{B0} + \frac{V_{BE}}{1/j\omega C_T} = i_{B0} + j\omega C_T V_{BE}$$

Electron injection into the base

Displacement current from the capacitor charging/discharging

Total capacitance : $C_T = \frac{Q_T}{v_{BE}}$, where Q_T = total stored charge in the device due to V_{BE}

C_T represents delays associated with all stored charge in the device (includes charge in transit across the base and the collector depletion region).

BJT Small-signal model

Small signal current gain, h_{fe}

$$\left|h_{fe}\right| = \left|\frac{i_c}{i_b}\right| = \left|\frac{i_{c0}}{i_{b0} + j\omega Q_T}\right| = \frac{1}{\left|\frac{1}{\beta} + j\omega Q_T / i_{C0}\right|}$$

At low frequency, with $I_c \sim I_{c0}$ and $I_b \sim I_{b0}$ then $h_{fe} \approx \beta$

At high frequencies $\frac{j\omega Q_T}{i_{Co}} >> \frac{1}{\beta}$ so $\left|h_{fe}\right| \approx \frac{i_{Co}}{j\omega Q_T}$

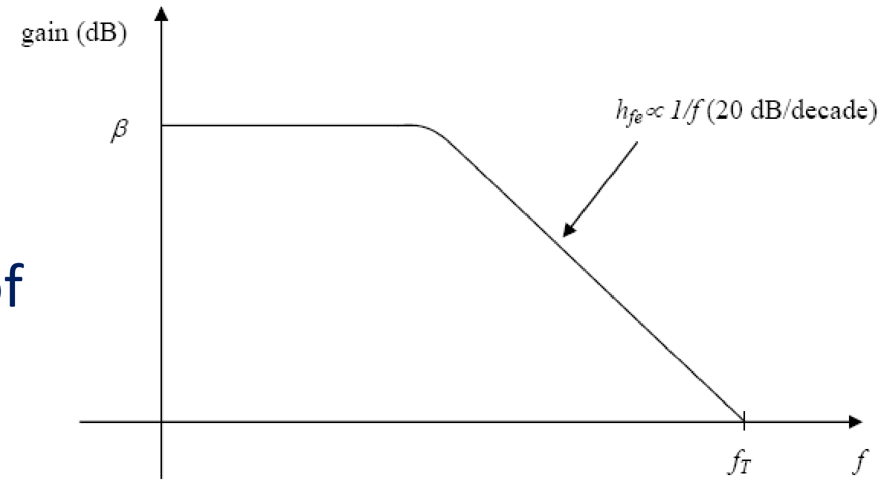
The charge in the transistor, Q_T, is swept through the device every τ_{EC} seconds. This τ is the total transit time plus the charging time for of C_{BE} and C_{BC} which are due to the excess charge in the base and charge in transit in the collector depletion region

So $i_{Co} = \frac{Q_T}{\tau_{EC}}$ and therefore $h_{fe} = \frac{1}{2\pi\tau_{EC}} = 1$ when $f = f_T$

BJT Small-signal model

Hence $f_T = \dfrac{1}{2\pi\tau_{EC}}$

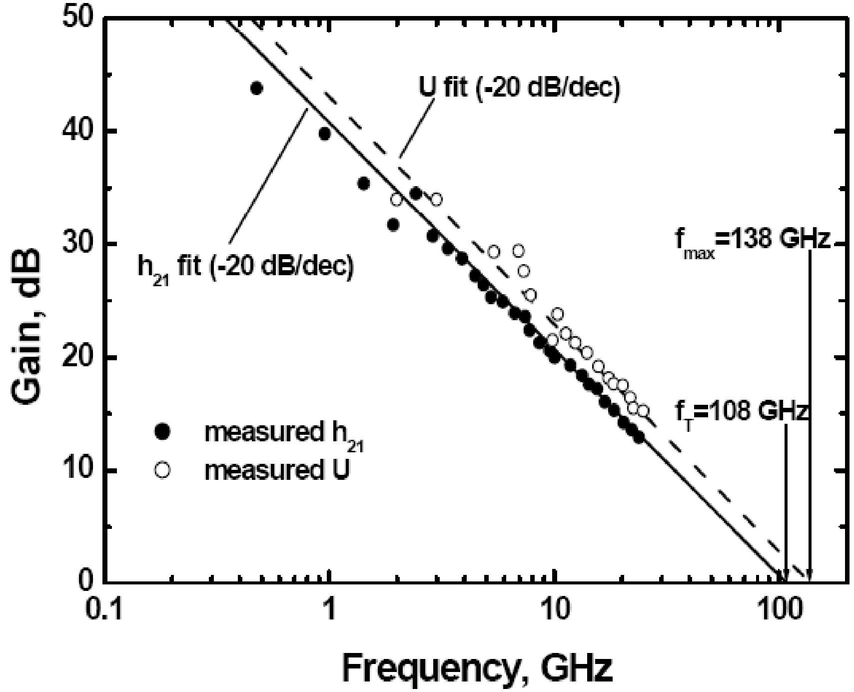
Typical variation with frequency of the current gain h_{FE}



At high frequency, the gain in dB = 20 log h_{FE}

$$= 20\log\frac{1}{2\pi f\tau_{EC}} = 20\log\frac{f_t}{f}$$

Or in other words a factor of 10 change in f gives a 20 db change in h_{FE}

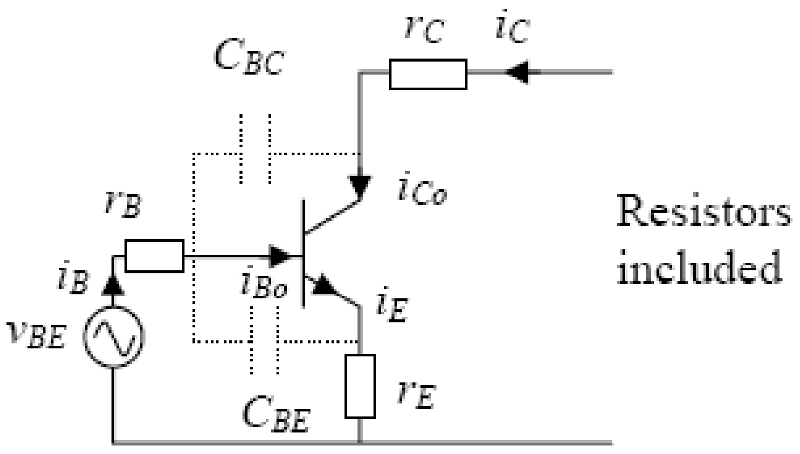


K. Onodera et al., IEEE Trans. ED 38, p. 429 (1991)

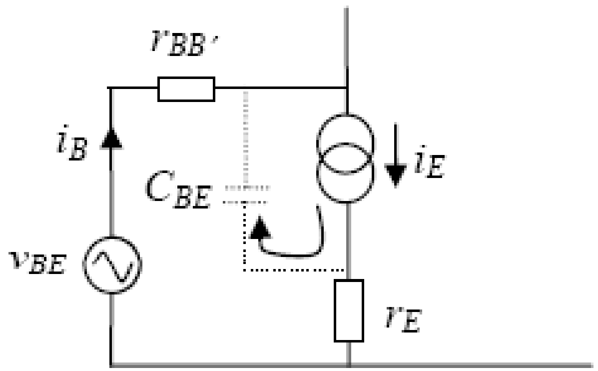
BJT Small-signal model

What are the components of the Total transit time, τ_{EC} ?

-Just a combination of the individual delays $\tau_{EC} = \tau_{BE} + \tau_{BC} + \tau_B + \tau_C$



First lets look at the base-emitter junction which gives τ_{BE}



C_{BE} needs to be charged before the electrons can be injected into the base. i.e. a charge of $C_{BE}\delta v_{BE}$ is required for a voltage change δv_{BE}.

BJT Small-signal model

C_{BE} is charged by i_E

$$\therefore \tau_{BE} = \frac{\delta Q_{BE}}{\delta i_E} = C_{BE}\frac{\delta v_{BE}}{\delta i_E}$$

The emitter current, I_E is related to the voltage V_{BE} through the diode equation

$$I_E = I_{Eo} e^{\frac{qV_{BE}}{kT}}$$

so

$$\frac{\partial V_{BE}}{\partial I_E} = \frac{1}{\frac{\partial I_E}{\partial V_{BE}}} = \frac{1}{I_{E0}\frac{q}{kT}e^{\frac{qV_{BE}}{kT}}} = \frac{1}{I_E\frac{q}{kT}} = \frac{kT}{qI_E}$$

$$\tau_{BE} = \frac{kT}{qI_E}C_{BE} \approx \frac{kT}{qI_C}C_{BE}$$

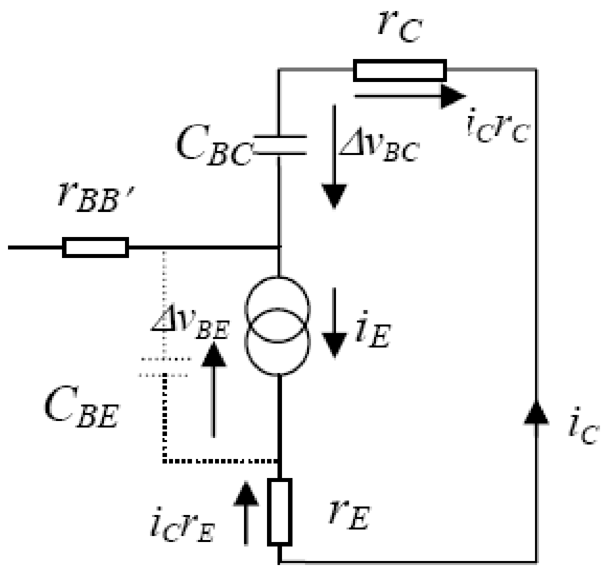
because $I_C \approx I_E$ (Base transport factor usually ~1)

The parameter $\frac{kT}{qI_C}$ can be seen as the dynamic resistance of the emitter-base diode. This resistance reduces as I_C increases.

BJT Small-signal model

Note that V_{BE} and v_{BE} are diode voltages rather than terminal voltages and hence r_E and $r_{bb'}$ are not involved.

Now lets look at the base-collector region, to derive τ_{BC}



Applying Kirchoff's law around the circuit gives:

$$\Delta v_{BC} = \Delta v_{BE} + \Delta i_C (r_E + r_C)$$

$$\tau_{BC} = \frac{\partial Q_{BC}}{\partial i_C} = C_{BC}\frac{\partial v_{BC}}{\partial i_C} \qquad = C_{BC}\left(\frac{\partial v_{BE}}{\partial i_C} + r_E + r_C\right)$$

Using previous derivation: $\tau_{BC} = C_{BC}\left(\frac{kT}{qI_C} + r_E + r_C\right)$

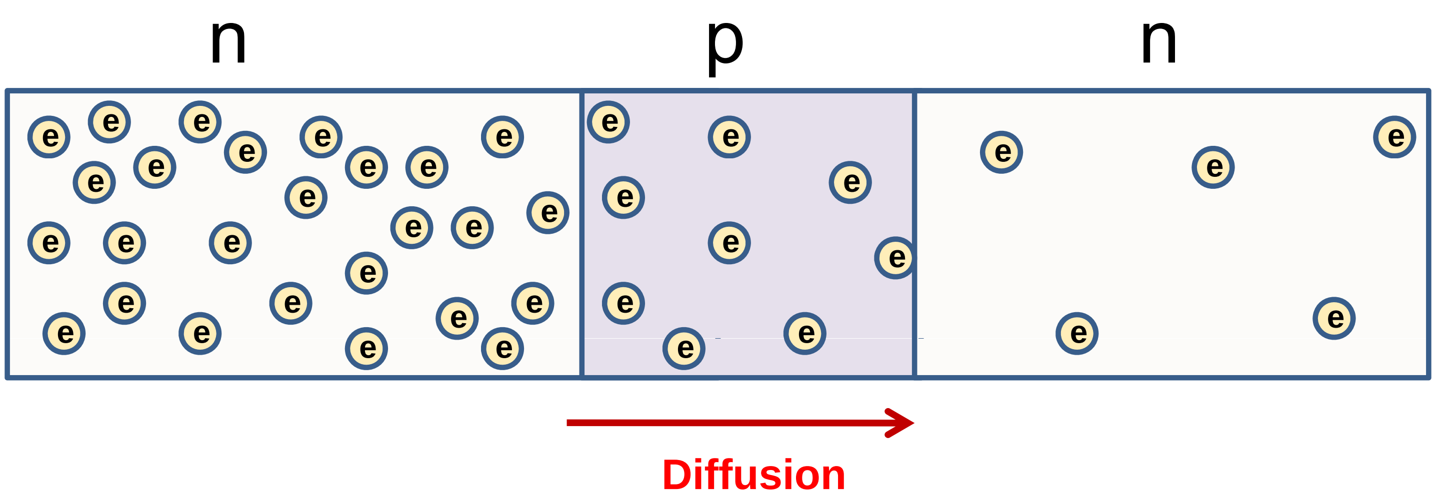
BJT Small-signal model

Base transit time: τ_B $\tau_B = \frac{W_B^2}{2D_e}$ Where W_B is the base width

and D_e is the electron diffusion coefficient $D_e = \frac{\mu_e kT}{q}$

Why dependent on diffusion?

Electrons (*for npn*) or holes (*for pnp*) in injected into the base will be minority carriers. They will diffuse across the base from a region of high concentration to a low one



BJT Small-signal model

The flux of electrons is proportional to the gradient of the concentration

$$Flux = -D_e \frac{\partial n}{\partial x} \qquad J_{e,diff} = qD_e \frac{\partial n}{\partial x}$$

So the base transit time is related to the minority carrier diffusion time across the base

Consider the minority carrier charge in the base

$$Q_B = qA\int_0^W \Delta n(x,t).dx = \frac{qAW\sigma(0,t)}{2}$$

Integrate all the sheet charge in the base

$$i_c = qAD_e \frac{\partial n(x,t)}{\partial x} = \frac{qAD_e n(0,t)}{W} = \frac{Q_B}{W^2/2D_e}$$

BJT Small-signal model

But I_c is just $i_c = \frac{\partial Q_B}{\partial t} = \frac{Q_B}{\tau_B}$ $\qquad \frac{Q_B}{\tau_B} = \frac{Q_B}{W^2/2D_e}$

So $\tau_B = \frac{W^2}{2D_e}$

Collector transit time: τ_C.
This is the time taken for electrons to (rapidly) transit the B-C depletion region.

Here W_c = collector depletion thickness
And V_{sat} is the electron saturation velocity.

The factor of 2 represents the average delay time

$$\boxed{\tau_C = \frac{W_C}{2v_{sat}}}$$

'2' not expected!

Without the factor of 2 you would have the delay time for the slowest carriers.

BJT Small-signal model

Summing all these contributions

$$\tau_{EC} = \frac{kT}{qI_C}C_{BE} + C_{BC}\left(\frac{kT}{qI_C} + r_E + r_C\right) + \frac{W_B^2}{2D_e} + \frac{W_C}{2v_{sat}}$$

Simplifying this gives $$= \underline{\frac{kT}{qI_C}(C_{BE} + C_{BC})} + (r_E + r_C)C_{BC} + \frac{W_B^2}{2D_e} + \frac{W_C}{2v_{sat}}$$

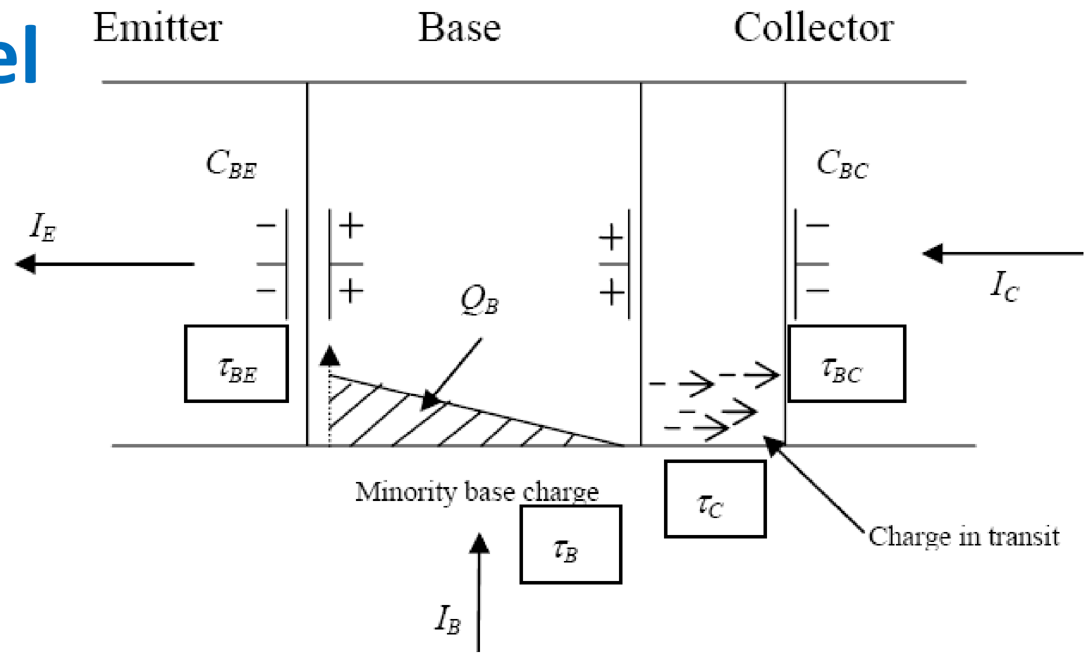
Usually this term dominates. If it does, then τ_{EC} α $1/I_C$, or in other words f_T α I_C

This leads to some interesting conclusions regarding high frequency operation

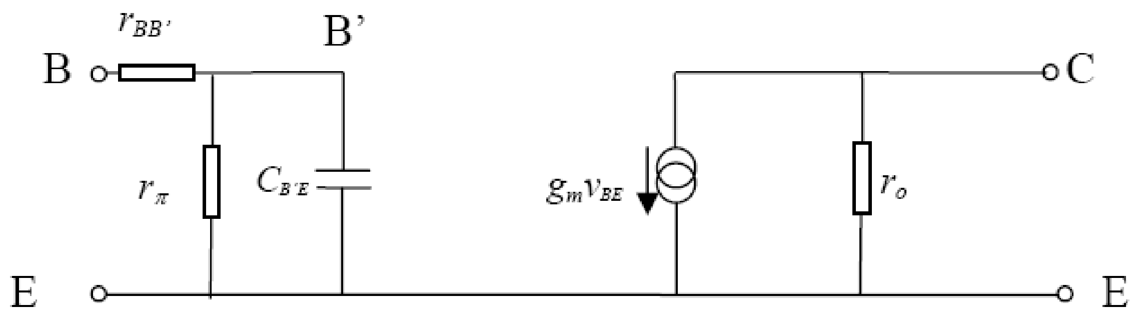
- High speed is achieved through a high collector current (BJT works best if driven hard)
- High speed needs **Small capacitance-** small device area. **Reduced parasitic resistances**- high doping & **Thin base & collector regions**

BJT Small-signal model

Overall picture of charge in a BJT



Simplified small signal equivalent circuit diagram (similar to MOSFET)



$$g_m = \frac{\partial I_C}{\partial v_{BE}} = \frac{q}{kT} I_{EO} \exp\left(\frac{qV_{BE}}{kT}\right) = \frac{qI_C}{kT} \qquad I_C \approx I_E$$

BJT Small-signal model

R_π is the input dynamic resistance of the EB junction $$= \frac{\partial I_E}{\partial V_{BE}} = \frac{kT}{qI_E}$$

This reduces with increase drive current

C_{BE} represents the capacitance of the depletion region and the diffusion charge due to minority electrons

R_o is the output resistance (as discussed before for the MOSFET where it is due to the short channel effect

$$r_o = \left(\frac{\partial I_C}{\partial V_{CE}} \right)^{-1}$$

In BJTs. R_o will vary due to the 'Early effect' . The base width can increase as a function of the applied base-to-collector voltage.

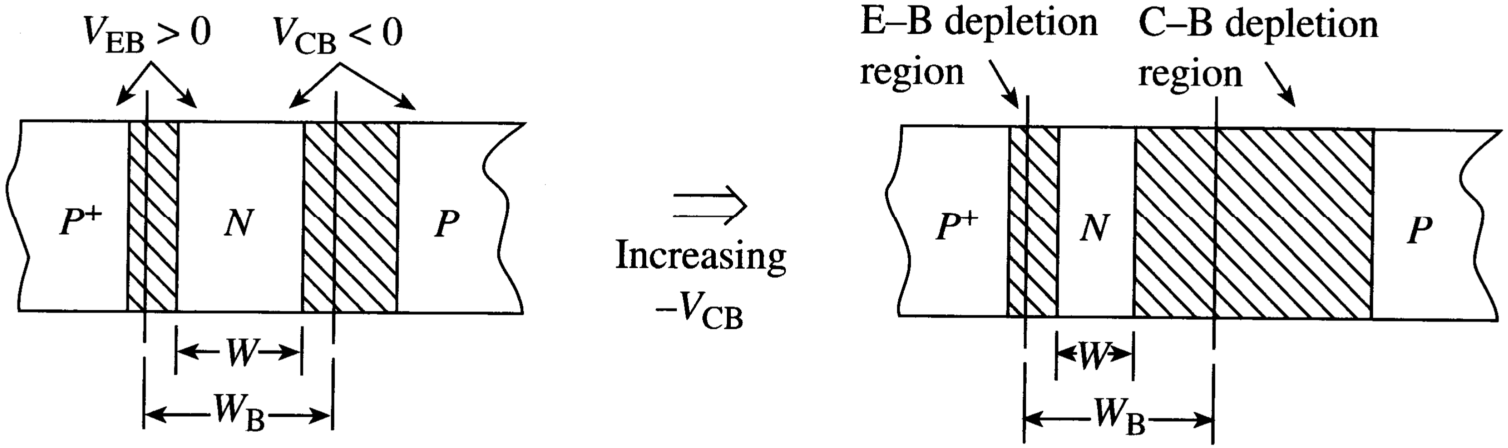
Very similar to short channel effect in MOSFETs, but mechanism is different

BJT Small-signal model

Named after J.M. Early (does not mean the effect is early!)

When bias voltages change, depletion widths change. The effective base width will be a function of the bias voltages

Most of the effect comes from the C-B junction since the bias on the collector is usually larger than that on the E-B junction



BJT Small-signal model

W_B is increased by the E-B and C-B depletion regions

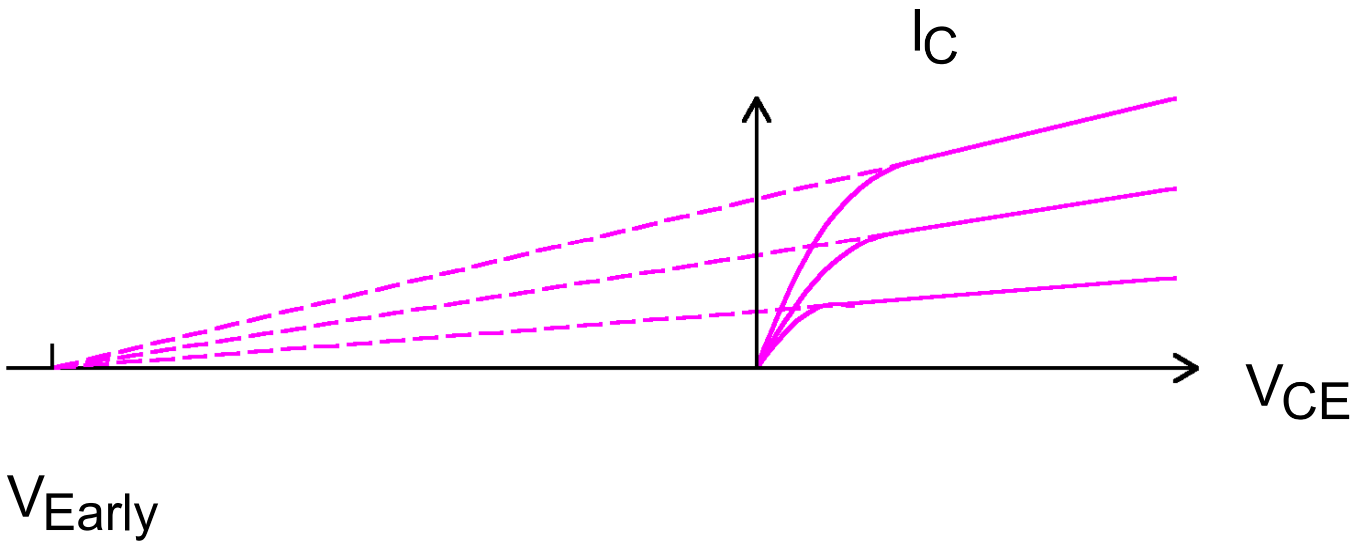
This is dependent on the relative doping concentrations (here N can be n or p)

$$W_{CB} \approx W_B \frac{N_C}{N_C + N_B}$$

So if $N_C << N_B$ then this effect can be minimised

Since $I_C \propto \frac{1}{W_B}$ then any increase in W_B tends to reduce I_C

Curves tend to converge at a single point known as the Early voltage



BJT Small-signal model

Ideal situation- very high Early voltage indicating the absence of **base width modulation**

More typical values 700-10,000

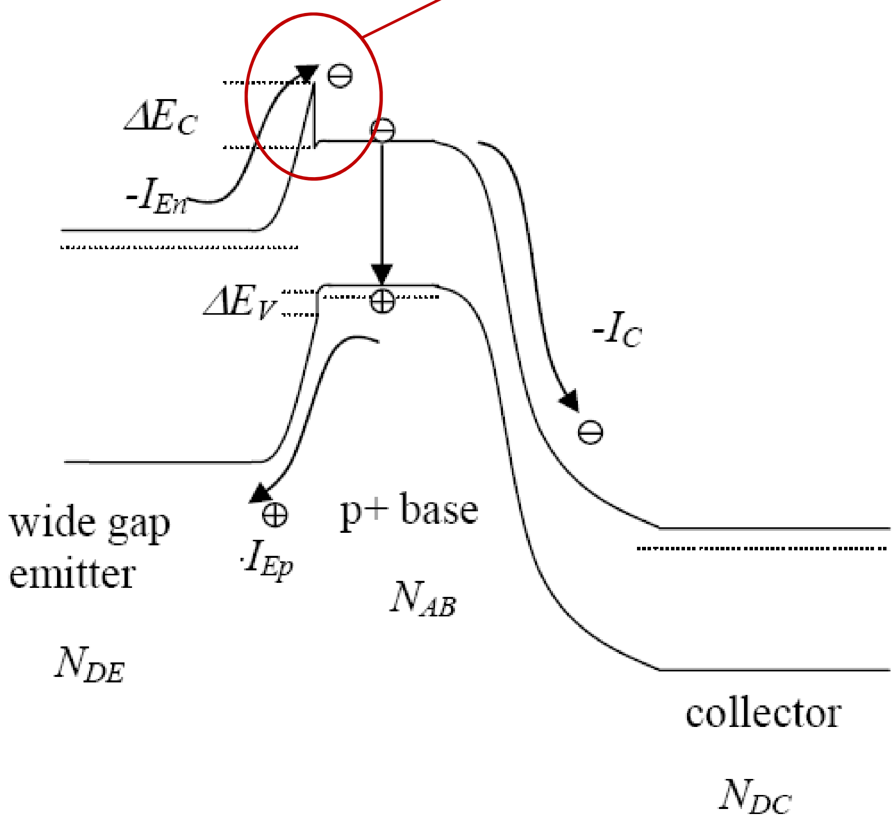
Type	N_D cm^{-3}	N_E cm^{-3}	N_c cm^{-3}	W μm	L_n μm	GU_I cm^{-4} · s	Calculated V_A volt	Measured V_A volt
T_1	3E20	1.0E16	3.4E14	17.3	124	2.7E13	8.3E2	7.0E2
					160		1.0E3	9.4E2
T_2	5E20	2.5E17	1.0E14	6.6	157	3.1E13	1.0E4	8.3E3
T_3	6E20	3E16	8E13	8.0	170	3.5E13	2.2E3	2.1E3
T_4	3E20	6E16	6E13	23.6	300	3.7E13	1.0E4	9.7E3
T_5	2E20	2E17	6.6E13	13	250	4.5E13	1.65E4	1.3E4
					256		1.66E4	1.4E4
T_6	8E19	1.7E17	5.6E13	13.5	154	6.3E13	1.52E4	1.3E4
T_7	8E19	8.5E16	1.0E14	15	173	7.9E13	6.4E3	5.2E3
					176		6.5E3	6.8E3

$r_{BB'}$ – base access resistance. This is usually quite high, due to the lower doping levels needed in the base to prevent minority carrier flow. (See HBT later)

HBT

The heterojunction bipolar transistor (HBT) is a variation on the simple BJT structure with much improved performance

Note this 'spike' in the conduction band can be graded out



A wide gap emitter is used to ensure the barrier for holes is much larger than the barrier for electrons at the emitter-base junction. This can give this device clear advantages over the BJT.

HBT

HBTs can be made using many different heterostructure systems, e.g:– Si/SiGe, AlGaAs/GaAs, GaInP/GaAs, InP/InGaAs

Characteristics of a HBT:

Wide gap emitter - e.g: Si, AlGaAs, GaInP, InP . This gives us greater design flexibility by removing one of the major limitations of the BJT.

Narrow gap base - e.g: SiGe, GaAs, InGaAs. This can have a high electron mobility.

High base doping- Overcomes limit on base doping in the BJT. Gives low base resistance

HBTs are high speed bipolar devices. For a Si BJT, f_T ~ few GHz (max 100GHz)

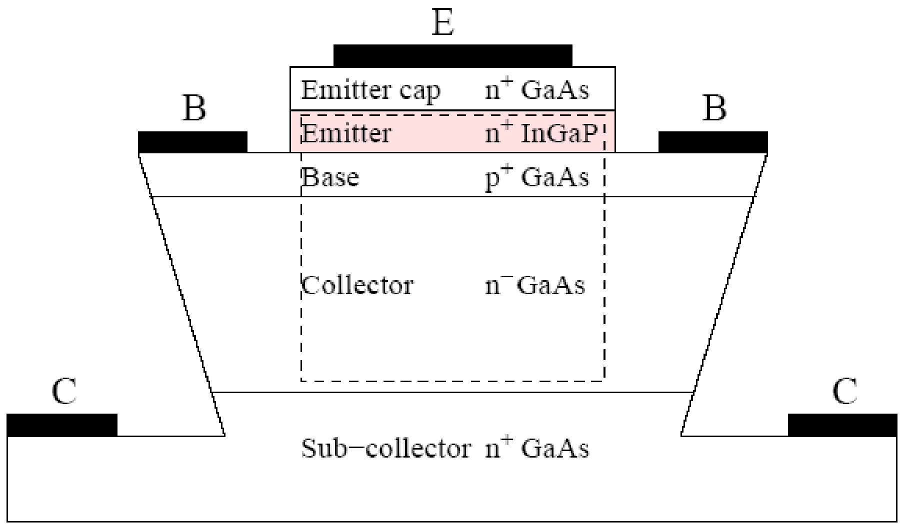
SiGe HBTs $f_T > 100GHz$ InP/InGaAs $f_T > 500GHz$

HBT

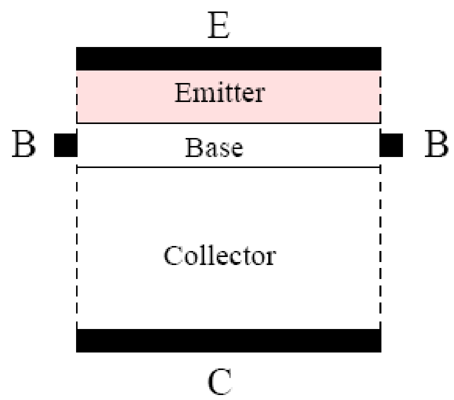
Basic structure of the HBT.

This examples uses InGaP as the wide gap emitter ($E_g \approx$ 1.9eV) with a GaAs ($E_g \approx$ 1.42eV) base and collector.

An n+ emitter cap and sub-collector give reduced contact resistances

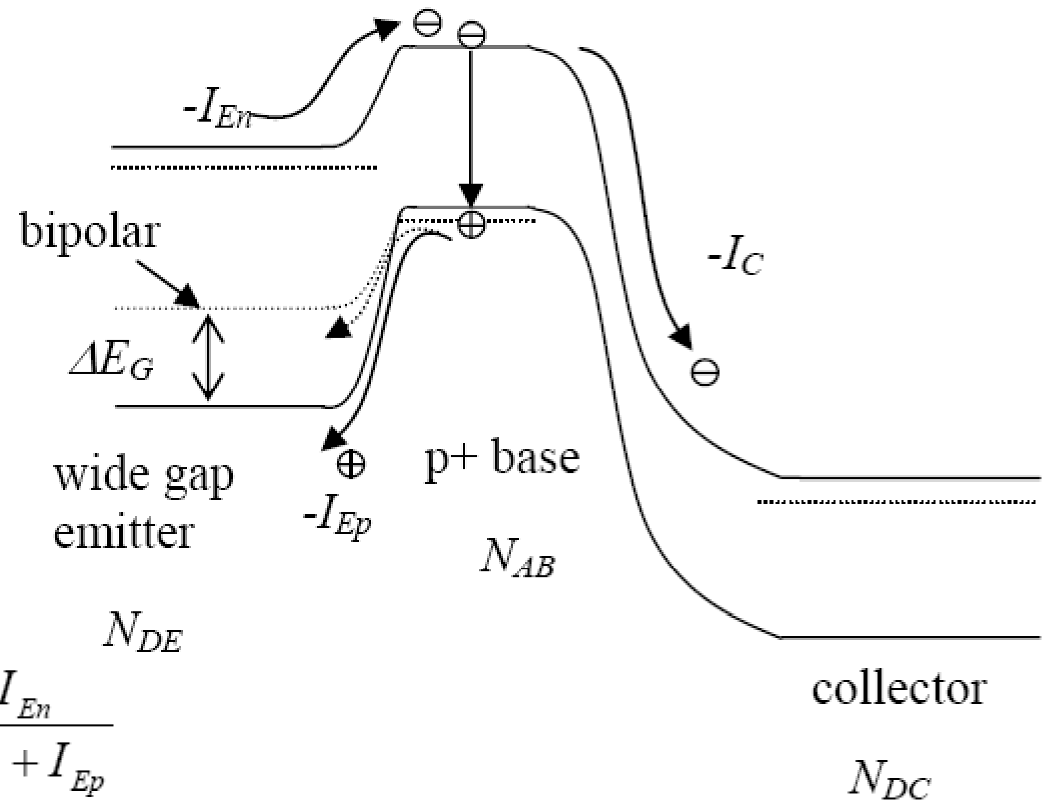


Simplified 1-D structure



HBT

The emitter base heterojunction gives a number of benefits



BJT injection efficiency $\gamma = \dfrac{I_{En}}{I_{En} + I_{Ep}}$

In the HBT, the holes experience a much bigger barrier than holes. I_{Ep} can be very low and $\gamma \rightarrow 1$.

Note in the simple BJT, I_{Ep} can be made low only by reducing the base resistance. This reduction (in base access resistance) has the effect to reduce the speed.

HBT

At the emitter-base heterojunction

where ΔE_G – bandgap difference

$$\frac{I_{Ep}}{I_{En}} \sim \exp\left(-\frac{\Delta E_G}{kT}\right)$$

Previous example: InGaP-GaAs HBT

E_g(InGaP)= 1.9eV E_g(GaAs) =1.42eV so ΔE_g=0.48eV

$$\frac{I_{Ep}}{I_{En}} \approx \exp\left(-\frac{0.48}{0.025}\right) = 4.5x10^{-9}$$

This is a huge factor

kT at room temperature ~ 0.025 eV

I_{Ep} and I_{En} are also proportional to their respective doping levels. This means that the base hole concentration could be increased by a large factor and still I_{Ep} would be significantly smaller than I_{En}

This higher base doping can significantly reduce the base access resistance $r_{bb'}$

HBT

Advantage of HBTs over the Si BJT

A higher injection efficiency due to the heterojunction.

- Higher gain, due to a higher γ
- Or can allow higher base doping, giving reduced $r_{BB'}$ C_{BC} or can use a thinner base and keep $r_{BB'}$ constant, resulting in a reduced base transit time τ_B

Reduced emitter doping N_{DE}

- Reduced C_{BE} and hence reduced τ_{BE}

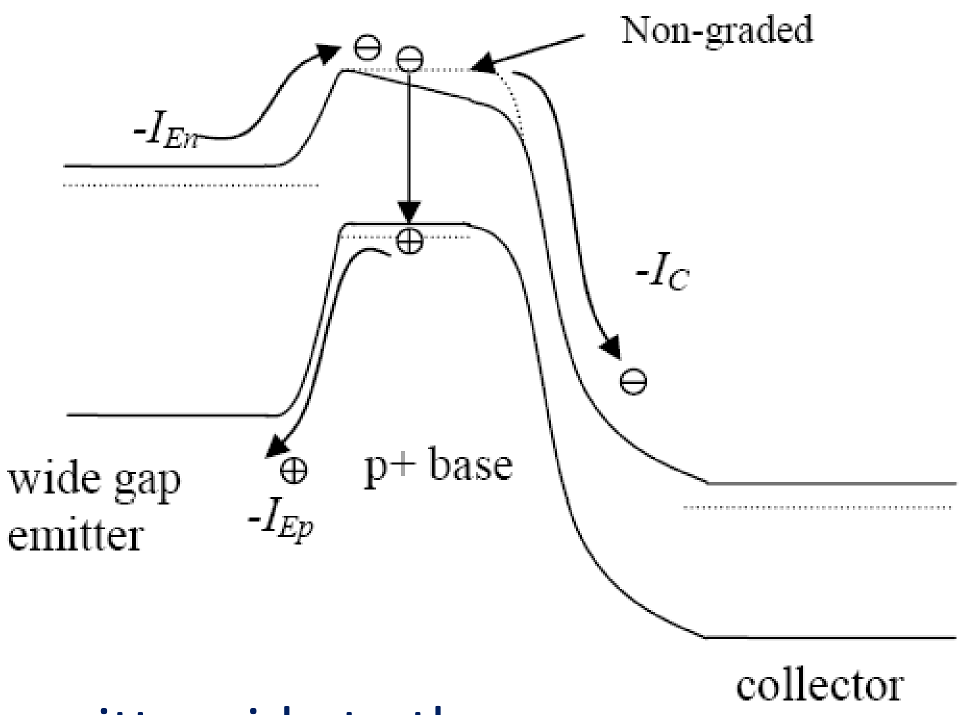
Use of high mobility base material

- Reduced τ_B and $r_{BB'}$ (eg: SiGe instead of Si, or III-V materials such as InGaAs)

HBT

Further options

Graded base



A graded base is achieved by reducing the band gap of the base from the emitter side to the collector side e.g: by increasing the Ge content in a SiGe base.

For example, assume a 25% variation in the Ge content along a 25nm base (could be from $Si_{.75}Ga_{.25}$ to $Si_{.5}Ge_{.5}$).
This compositional variation is equivalent to a ΔE_g~ 0.2eV

HBT

This band gap change will induce an equivalent electric field across the base

$$= \frac{\Delta E_G}{qW_B} = \frac{0.2}{2.5\times10^{-8}} = 80kV/cm$$

At this field strength, electrons travel at their saturation velocity rather than diffusing across. The base transport is very much improved.

Base grading is also possible in the III-V material systems such as AlGaAs/GaAs and InP/InGaAs

Maximum Frequency of Operation, f_{max}

This is a so-called figure of merit – frequency at which power gain reaches unity. The full derivation of the power gain is complicated. Here we will consider only a simplified version.

Power gain, G_P = output/input $\propto \frac{\mathrm{Re}(Z_{out})i_C^2}{\mathrm{Re}(Z_{in})i_B^2}$

HBT

Now $\frac{i_C}{i_B} = h_{FE}$ so it turns out that

Origin of this factor is beyond this course

$$G_P = \frac{1}{4}\frac{\mathrm{Re}(Z_{out})}{\mathrm{Re}(Z_{in})}\left|h_{FE}\right|^2$$

$\mathrm{Re}(Z_{in}) \approx r_{bb'}$ (the base access resistance)

$$\mathrm{Re}(Z_{out}) = \frac{1}{2\pi f_T C_{BC}}$$

Also beyond this course. See text books!

Also $h_{fe} = f_T / f$

So we end up with a power gain

$$G_P = \frac{f_T}{8\pi r_{BB'} C_{BC} f^2}$$

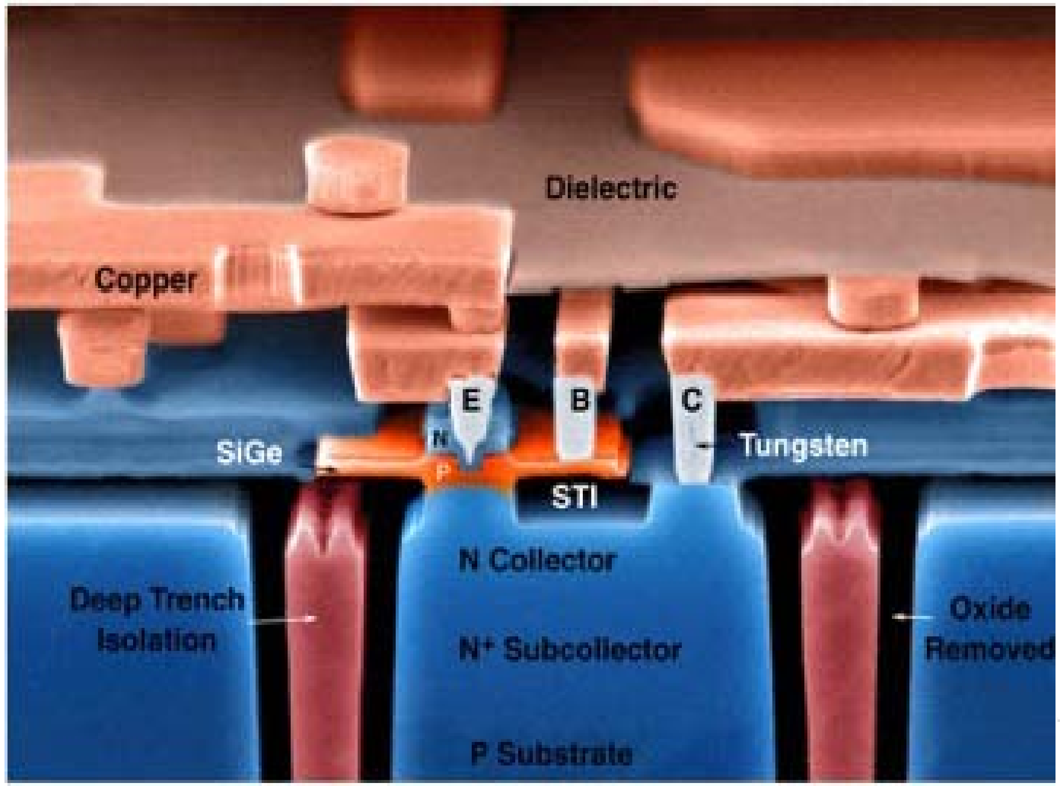
Define f_{max} as the frequency when the power gain goes to 1

$$\rightarrow f_{max} = \left(\frac{f_T}{8\pi r_{BB'} C_{BC}}\right)^{\frac{1}{2}}$$

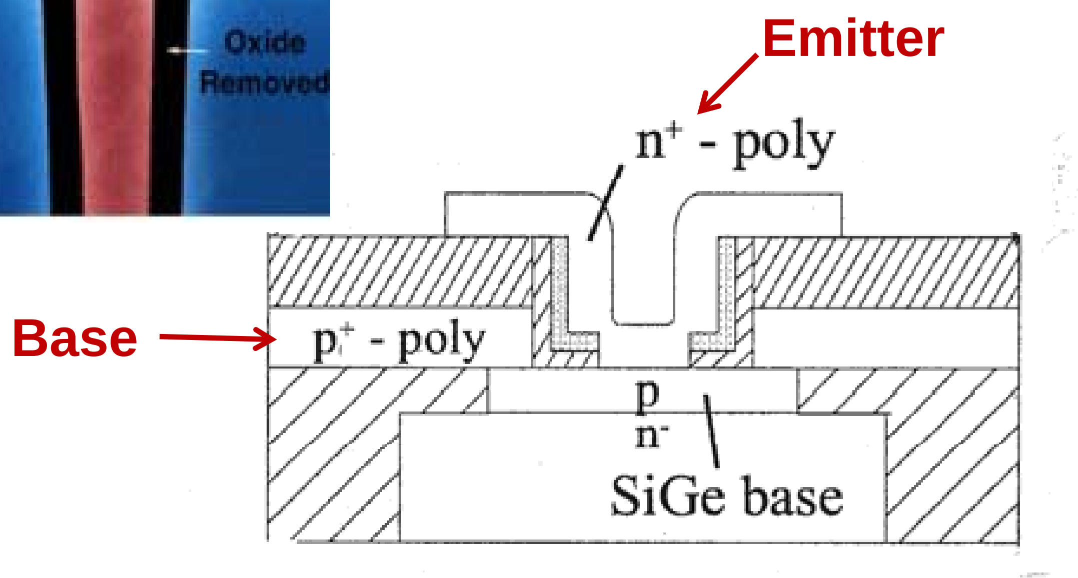
Reducing the parameters $r_{BB'}$ and C_{BC} is therefore important to ensure high frequency operation.

Reducing area reduces C_{BC}. Reducing the BE spacing reduces $r_{bb'}$

HBT



Low base-emitter spacing assured in the design of the device



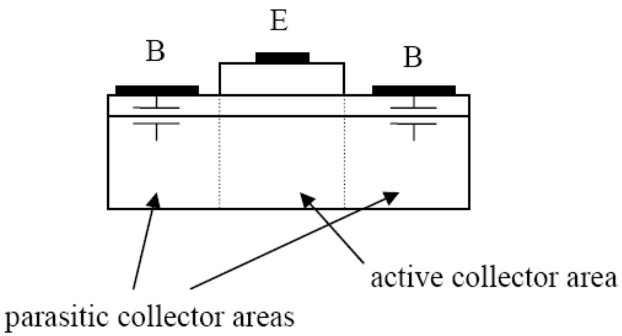
HBT

Scaling of HBTs

Dimensions: HBT is a vertical geometry device. A reduced base and collector thickness can be achieved using epitaxial growth methods. The subsequent increase in the base access resistance, $r_{BB'}$, can be countered by reducing the base-emitter contact spacing.

Ohmic contacts: Reducing the size of the base and emitter contacts increases their contact resistance. Present technologies place a limit on scaling to ~ 1 μm spacing

Parasitic collector: As the device shrinks, the effect of parasitic collector components becomes more important



Breakdown: Thin collector regions, used for fast transit, cannot support large

HBT

voltages before impact ionisation occurs and the device breaks down. The breakdown voltage of any devices must always remain above the collector supply voltage.

Current drive – increased speed can be achieved by increasing the current density, as we discussed before for the BJT .

Remember $\tau_{BC} = C_{BC}\left(\frac{kT}{qI_C} + r_E + r_C\right)$

However the extent at which we can increase the current is limited by the *Kirk effect.*

The Kirk effect occurs at high current densities and causes a dramatic increase in the transit time of a bipolar transistor. The effect is due to charge density build-up associated with the current passing through the base-collector region

HBT

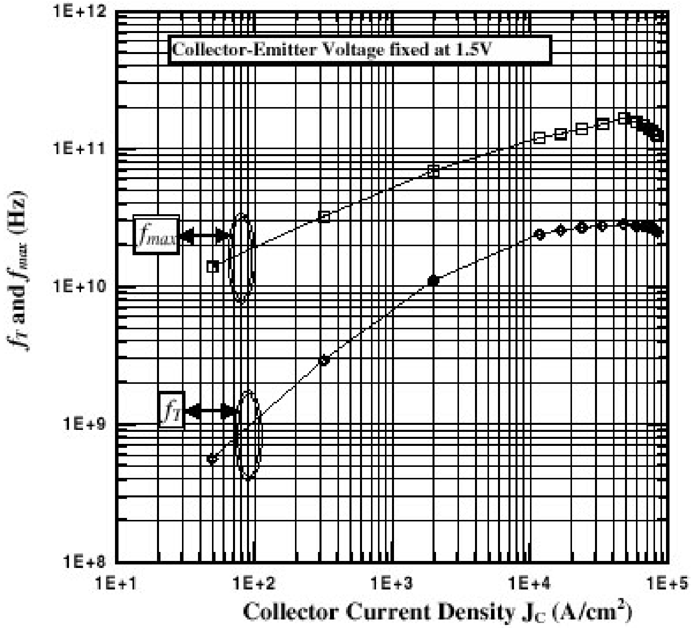
As the very high density of injected carriers starts to exceeds the charge density in the depletion region the depletion regions cease to exist. Instead a charge dipole forms and this begins to restrict the current low.

The effect occurs if the charge density associated with the current (j_c) is larger than that from the ionized impurities in the base-collector depletion region. $j_C \geq qN_{DC}v_{sat}$

At that point the effective width of the base layer then equals the width of the base and collector layer, which increases the transit time substantially. The increased transit time reduces the collector current and will increase the transit time.

HBT

Increasing the collector doping can easily eliminate the Kirk effect. However, this also increases the base-collector capacitance and decreases the collector-base breakdown voltage. As a result is a part, the Kirk effect affects both RF and power devices. A proper trade-off between these factors is part of any device design and optimization.



(b)

f_T and f_{max} should increase with I_c but at limited at high current due to the Kirk effect

HBT

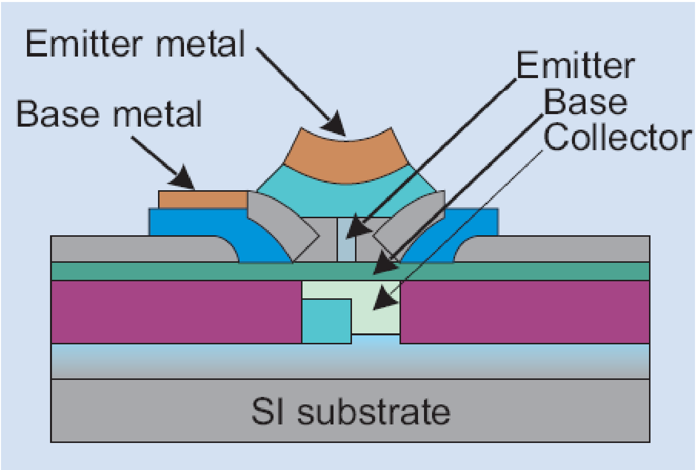
In summary HBTs offer higher f_t and f_{max} values than Si BJTs due to improved emitter injection efficiency, base resistance and base-emitter capacitance.

HBTs can generally can reach higher frequencies than MOSFETs.

The record high current-gain (f_T) and power-gain (f_{max}) cut-offs for a transistor of any type have been obtained using an InP/InGaAs graded –base HBT and a special device design

The best values to date are f_T =710GHz and f_{max}=340GHz

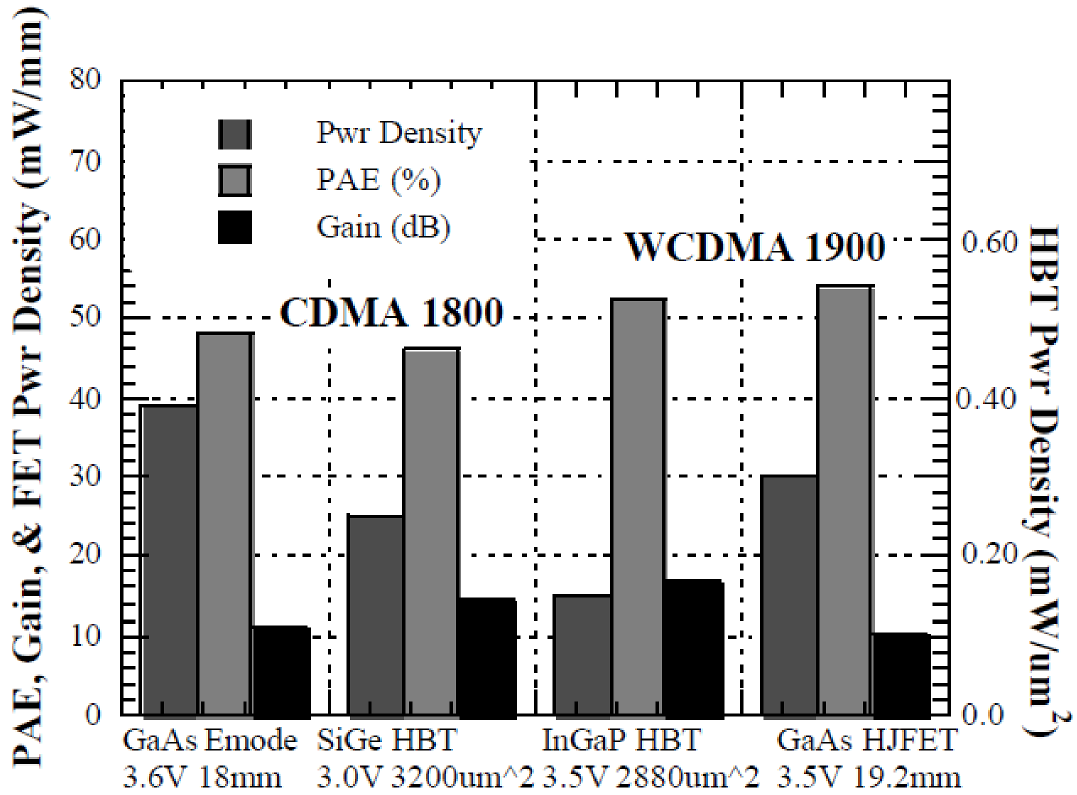
Research is now producing these metamorphically on Si substrates



.

HBT applications

HBTs are used for digital and analog microwave applications to frequencies up to 100GHz. They compete with MOSFETs and HEMT (HJFET) technology (discussed earlier)



Power density, Power added efficiency (PAE) and gain comparison for 3G mobile communications

Note: $PAE = (P_{out}^{RF} - P_{in}^{RF}) / P_{DC}^{Total}$

HBT applications

Applications of HBTs tend to lie in the K- RF bands (12-40GHz)

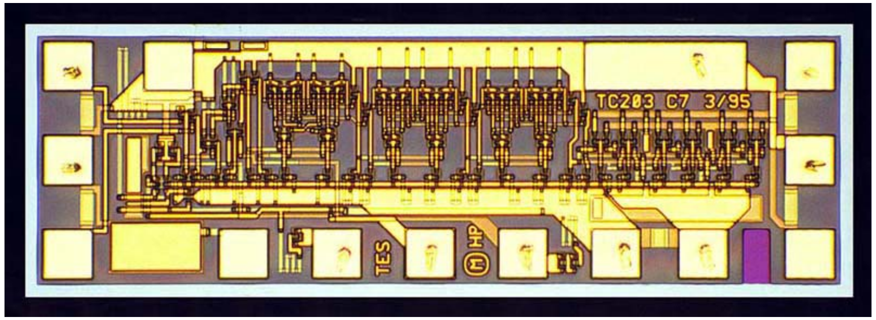
Fixed satellite uplinks, direct broadcast satellites, RADAR, Radio astronomy

Used in circuits such as Low noise amplifier, Radar Transceiver

Demodulator, Up/down converters

High speed A to D conversion.

There is strong competition between device types in terms of performance and cost.



For low-noise amplifier applications (eg: LNA in mobile phone), the PHEMT is generally recognized as the best choice.

HBTs are generally seen as more attractive for power amplification and linearity (e.g: mobile phone base-station)

BiCMOS

As the name suggests, this refers to the integration of bipolar junction transistors and CMOS technology into a single integrated circuit device.

In its simplest sense it can be thought of as BJTs whose base is controlled by a CMOS switch

The process was invented by Toshiba in 1990.

The **BiCMOS** process has commercial applications in mixed signal and discrete component logic design. More recently it has become the technology of choice for power electronics products such as voltage regulators

Why combine the two?

MOS devices are ideally suited for use in logic applications because of their low current consumption and power

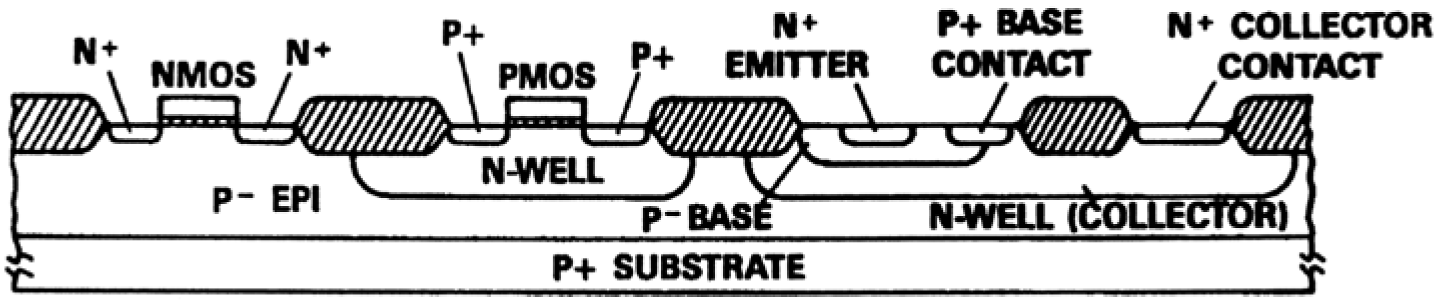
BiCMOS

The MOSFET process is also more integratable, with higher packing density and higher yields.

Yet bipolar devices are important for creating accurate voltage references and when very low noise is required. They are also good for driving large currents.

The BiCMOS process combines both transistor types. It is employed when it is desirable to put logic and analog functions together on the same chip for mixed-signal applications.

BiCMOS process



BiCMOS

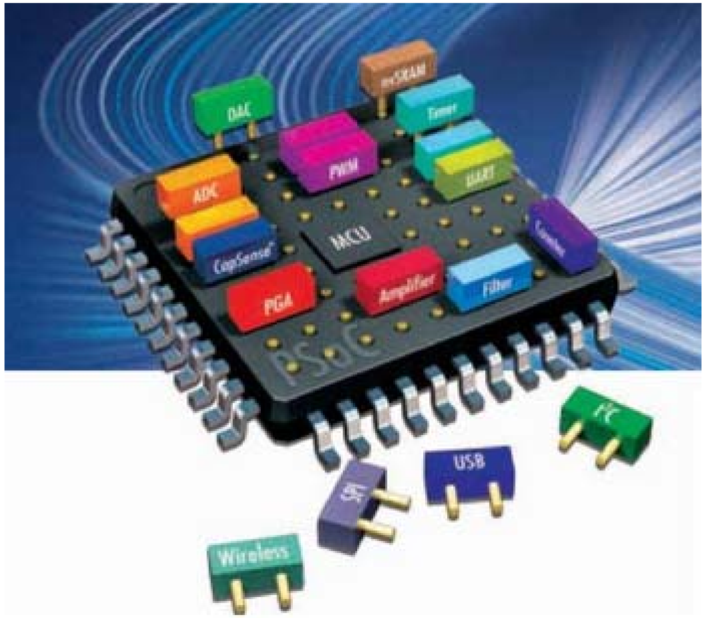
The BiCMOS process adds an NPN bipolar transistor to the standard CMOS structure. It uses the PMOS n-well as the collector of the Bipolar device and introduced a further lithographic step to create the n-base region.

A good example of a common BiCMOS IC is that used for media players or DAB radio. Here digital ciucuits to decode a data stream are used to create high quality audio frequency audio

BiCMOS is also a common feature of many system-on- a- chip approaches.

BiCMOS

A system on a chip combines different types of silicon circuitry together in one unit.

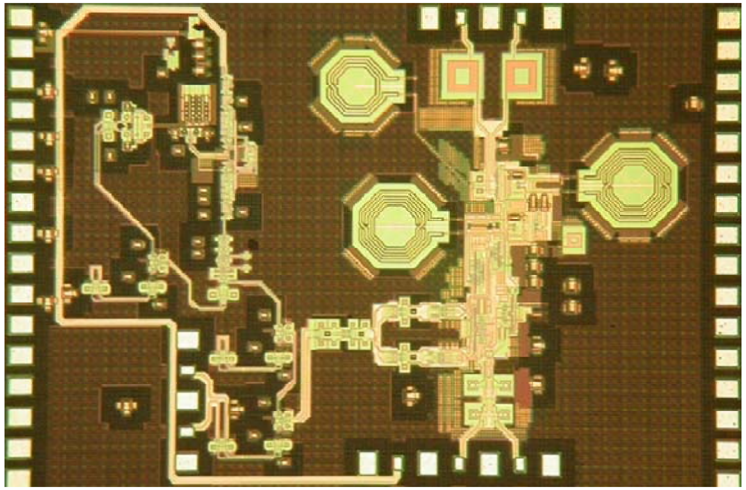


Eg: Apple A4 for the i-phone/i-pad. Combining processing, video, audio, wireless, power control etc

Some of these circuits will be CMOS, some BiCMOS

BiCMOS

SiGe BiCMOS, using HBTs, is now a major player in RF circuits



BiCMOS takes the best of both CMOS and Bipolar.

However the two types are somewhat compromised by being combined in a single process and the MOS and Bipolar circuits will not perform as well as if they were made stand-alone.

Fine tuning of both the BJT and MOS components of the process is not possible without adding many extra fabrication steps, and consequently increasing the process cost.

It does not also does not offer the low power consumption of CMOS alone or the low noise capability of Bipolar alone.

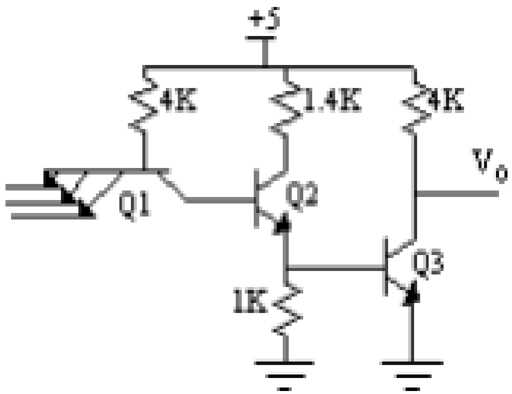
Bipolar logic

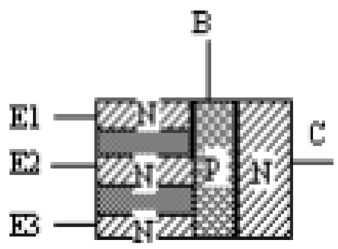
CMOS logic dominates in the modern era

However there are two bipolar logic families which are historically important and which can still find use in certain applications: **Transistor-Transistor Logic (TTL), Emitter-Coupled Logic (ECL).**

TTL is the earliest of transistor logic families, invented in 1961 and used extensively up to the 1980s. It used a special type of multiple emitter BJT to create basic logic gates for small scale ICs.

More complex systems where then assembled from several ICS





Bipolar logic

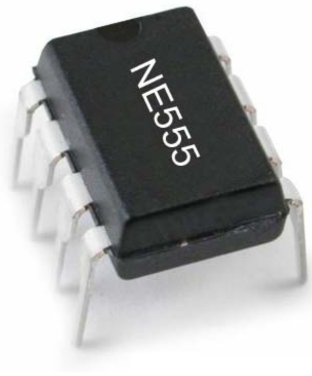
Emitter-coupled logic (ECL) was developed in the 1970s as an improvement on TTL particularly in respect to speed.

ECL in a differential-amplifier configuration in which one side of the amplifier has multiple-input bipolar transistors with their emitters tied together.

ECL has a much improved switching speed, but consumes a relatively substantial amount of power in both logic states (one or zero). A further benefit of ECL is the narrow logic level swing (approximately 800 mV) which helps to reduce noise generation.

Electronic system using TTL logic Ics (1960s)

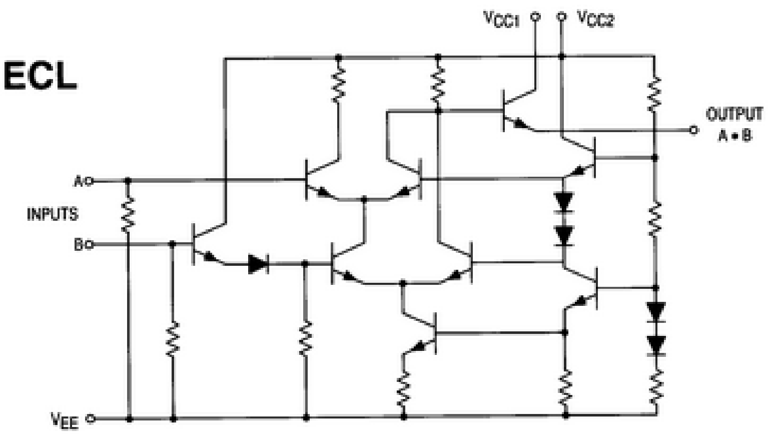
Bipolar logic



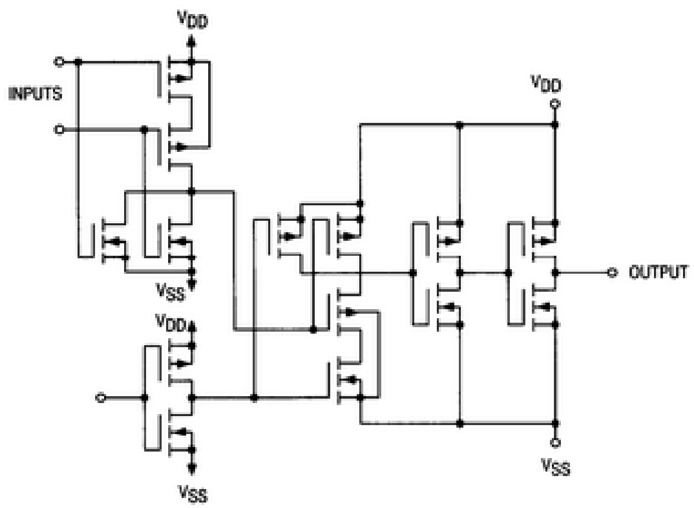
ECL was used to develop some of the mass-market logic chips we used today, eg: the NE555 timer in 1971. It has been largely superseded by CMOS (we still use the 555 today, but it is now CMOS)

An ECL logic circuit has a lower transistor count than an equivalent CMOS one. However the biasing arrangement is much simpler for the CMOS.

ECL



CMOS



Bipolar logic

One of the problems of using BJTs as switches is that they have a high static power dissipation. Additional transistors in logic circuits help to reduce this, but it is still very significant.

Yet ECL can be very fast and a circuit fashioned from high performance HBTs can be significantly faster than CMOS circuitry

However there is a caveat. The speed advantage of ECL comes only if operated in the active region (just below saturation) within which a very small change in V_{BE} can switch a large current. This limits it power handling capability.

CMOS has largely replaced bipolar logic in all but some very specialised applications. The lower power dissipation is a factor, but critical has been the manufacturability of CMOS chips and the ability to scale down in size.