Middlesex University London

Middlesex University Research Repository:

an open access repository of
Middlesex University research

http://eprints.mdx.ac.uk

Camp, Robert Paul, 1986.
Characteristics of UHF Transistors
Using Autoregistered Structures.
Available from Middlesex University's Research Repository.

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this thesis/research project are retained by the author and/or other copyright owners. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge. Any use of the thesis/research project for private study or research must be properly acknowledged with reference to the work's full bibliographic details.

This thesis/research project may not be reproduced in any format or medium, or extensive quotations taken from it, or its content changed in any way, without first obtaining permission in writing from the copyright holder(s).

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:
eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

ABSTRACT

The basis of a novel bipolar transistor structure was proposed by Dr R. Aubusson of Middlesex Polytechnic in 1977. The novelty lies in replacing the conventional overlay transistor's P+ base grid with a refractory metal grid, in order (a) to lower the base resistance and (b) to autoregister the emitter. It was claimed that the linearity of the transistor would also be improved. A number of questions raised by this idea have been investigated, the methods and conclusions of which are presented here.

Plausible structures, using the metal base grid, are proposed and compared with conventional structures. Some advantages are seen to be possible.

The current understanding of distortion analysis applied to transistors is reviewed. The main ideas are presented in a unified manner and are extended to higher order. A number of the transistor's second order effects are analysed in a novel fashion. The metal base grid transistor is analysed and compared with conventional transistors, with favourable results.

Practical aspects of fabricating the metal base grid transistor were investigated. A procedure for deposition has been determined and is presented here along with the

film physical and electrical characteristics. Analysis of the tungsten-silicon interface shows the suitability of the metallization as a base grid. Suitable means of delineating the tungsten film have been assessed and a working procedure determined. Subsequent deposition of various insulators has been investigated and the problems associated with the readily oxidized tungsten film have been overcome. Formation of the emitter, requiring further high temperature processing, has been assessed in view of the limitations imposed by the preformed base metallization.

In summary, it has been shown that the novel structure can be constructed and that significant performance improvement is to be expected, although a full realization was not possible within the resource constraints of the project.

ACKNOWLEDGEMENT

The author would like to extend his appreciation to the following people:

Mr A. Cross of BT for the donation of an Edwards HF sputtering kit.

Mr E. Michaelowicz of BOC for much practical advice on setting up and using the sputtering equipment and also the supply of spare parts and other high voltage items.

Dr K. Das of Middlesex Polytechnic and Dr P. Hemment of Surrey University for their help and advice on the use and interpretation of the Rutherford Back Scattering analysis.

Dr Ashburn of Emulsitone for advice on the use of spin-on silica film.

Prof J. Butcher of Middlesex Polytechnic and Dr L. Kennedy of GEC as project supervisors.

Dr R.Aubusson of Middlesex Polytechnic for the original idea of the novel transistor and supervision in the initial phase of the project.

Mr J. Linnell, Mr D. Court and Mr G. Shorthouse laboratory technicians at the Middlesex Polytechnic Microelectronics Laboratory.

Special thanks to Jane for help in typing the thesis, advice on presentation and for an excellent job in proof reading the final draft.

GLOSSARY

a Attenuation constant (cm)

f_{α} The frequency at which α has fallen to 0.707 of the low frequency value

f_{β} The frequency at which β has fallen to 0.707 of the low frequency value

f_{MAX} The transistor's maximum frequency of oscillation

f_{T} The frequency at which β has fallen to 1

q Electron charge (C)

C Specific contact resistance (Ω-cm^{2})

D Diffusion sheet resistivity (Ω/square)

dBm Power relative to 1mW (dB)

K Boltzmann constant (J/K)

L Length of the contact (cm)

M Magnitude of distortion component

M Metal sheet resistivity (Ω/square)

T Absolute temperature (K)

V_{T} Thermal voltage, given by kT/q (V)

W Width of contact (cm)

α Common base current gain

β Common emitter current gain

τ_{F} Transit time of intrinsic transistor (s)

τ_{T} Total transit time (s)

ϕ Built in junction voltage (V)

ε Permittivity of free space (F/cm)

ε_{R} Relative permittivity

INDEX

ABSTRACT

ACKNOWLEDGEMENT

GLOSSARY

CHAPTER 1 - INTRODUCTION — Page

1.1 The Need For Improved Device Linearity 1
1.2 High Frequency Power Transistors 2
1.3 Applications Using UHF Transistors 8
1.4 Disadvantages of Bipolar Junction Transistors (BJTs) 11
1.5 The Power FET 14
1.6 New Life for the BJT 18
1.7 Scope of This Thesis 20

CHAPTER 2 - UHF TRANSISTOR STRUCTURES

2.1 High Frequency Power Transistors 24
2.2 Interdigitated Transistors 26
2.3 Overlay Transistors 31
2.4 Mesh Transistors 33
2.5 Novel Types of Transistor 34
2.6 Effects of Misalignment 37
2.7 Comparison of Structures 41

2.8 Fabrication Techniques 55

CHAPTER 3 - MODELLING

3.1 Transistor Models 62

3.2 Small Signal Model 62

3.3 Large Signal Model 70

3.4 Charge Control Model 77

3.5 Base Push-out 84

3.6 Collector Capacitance 90

3.7 Early Effect 92

3.8 Emitter Capacitance 93

3.9 Package and Extrinsic Components 94

3.10 Extrinsic Base and Emitter 97

CHAPTER 4 - DISTORTION

4.1 Review of the Literature 106

4.2 Low Frequency 107

4.3 Harmonic Distortion 108

4.4 Intermodulation Distortion 110

4.5 Cross-Modulation 112

4.6 High Frequency 113

4.7 Distortion in the Transistor 115

4.8 Exponential Non-Linearities 117

4.9 Early Effect Non-Linearities 125

4.10 High Injection Non-Linearities 133

4.11 Kirk Effect Non-Linearities 138

4.12 Advantages of the Novel Structure 144

CHAPTER 5 - A NOVEL TRANSISTOR

5.1 Terms of Reference 150
5.2 Autoregistered Transistors 151
5.3 The Novel Metal Base Transistor 157
5.4 Emitter Fabrication 166
5.5 Proposed Process Schedule 171

CHAPTER 6 - EXPERIMENTAL WORK

6.1 Practical Considerations 173
6.2 Mask Set 173
6.3 Deposition of Tungsten 178
6.4 Etching of Tungsten Film 186
6.5 Making Ohmic Contact 193

CHAPTER 7 - EXPERIMENTAL MEASUREMENTS

7.1 Terminations 201
7.2 Intermodulation Measurement 202
7.3 Choice of Intermodulation Product 206
7.4 Measurement in a 50Ω System 210
7.5 Setting up the Equipment 216
7.6 Distortion Measurements 219

CHAPTER 8 - COMPUTER SIMULATION

8.1	Program Limitations	228
8.2	SPICE	229
8.3	Fourier Analysis Using SPICE	234
8.4	Analysis at Low Frequency	243
8.5	Analysis at High Frequency	245

CHAPTER 9 - CONCLUSIONS

9.1	Summary	249
9.2	Conclusions	251
9.3	Recommendations	254

BIBLIOGRAPHY

APPENDICES

I Contact Resistance

II Current Distribution Along the Base Finger

III Volterra Series

IV High Frequency Distortion Limit

V Fifth Order Expansion of Volterra Kernels

VI Comparison of Approaches at H.F.

VII Figures of Merit for Linearity

VIII Fourth and Fifth Order Exponential Distortion

IX Distortion Due to Base Push-out

X Mask Making

XI Sputtering Procedure

XII Bipolar III Process Schedual

XIII S-Parameters

XIV Fourier Analysis of Initial Conditions

CHAPTER 1 - INTRODUCTION

1.1 The Need for Improved Device Linearity

UHF power transistors are used in many applications related to communications. Typically the requirement is for a simple, low cost amplifier, often using a single transistor. Distortion arising in these amplifiers is of concern as the spurious signals so produced are difficult to filter out. Some of the common forms of distortion are discussed in Chapter 4. Generally it is not the distortion of the wanted signal which is the problem but more usually the spurii which constitute unwanted signals in a nearby band. An example of this is in a cable TV repeater which is carrying many channels. Non-linearities can cause a carrier signal to be modulated by a signal in a different channel and can also modify the depth of modulation. In this case the variation in modulation depth would probably not be noticeable, whereas the crosstalk between channels could be, even though the unwanted signal may be small.

Circuit techniques are used where possible to minimize distortion products and these include the use of filters and optimal biasing of the transistor. The use of feedback is of limited use since it trades gain for linearity and the available gain is often not large enough to accommodate this. Typically a device with f_T of 2GHz

may be required to have linear gain at 400MHz. The available gain in this case is approximately 15dB.

The technique common at low frequency, using feedback around a multi-stage amplifier is not applicable since it cannot be compensated to provide stable operation and sufficient gain at high frequency. A technique which can be used to advantage is that of two transistors used in push-pull. This has the effect of cancelling the even-order distortions. It will later be seen, however, that the odd order distortions are the most objectionable. In the final analysis it is the transistor which produces the distortion and, since it is difficult to compensate this, much effort should be directed towards making the device itself more linear.

1.2 High Frequency Power Transistors

The UHF transistor is often required to exhibit useful gain from tens of megahertz to above a gigahertz. Power handling varies from the half Watt medium power device to hundreds of Watts. Generally the power gain is required above the f_β of the device, which means that the available power gain falls off at approximately 20dB per decade of frequency. The gain-bandwidth product, f_T, is therefore one of the most important parameters for these devices, describing the h_{fe} fall-off with operating frequency, as shown in Figure 1.2.1.

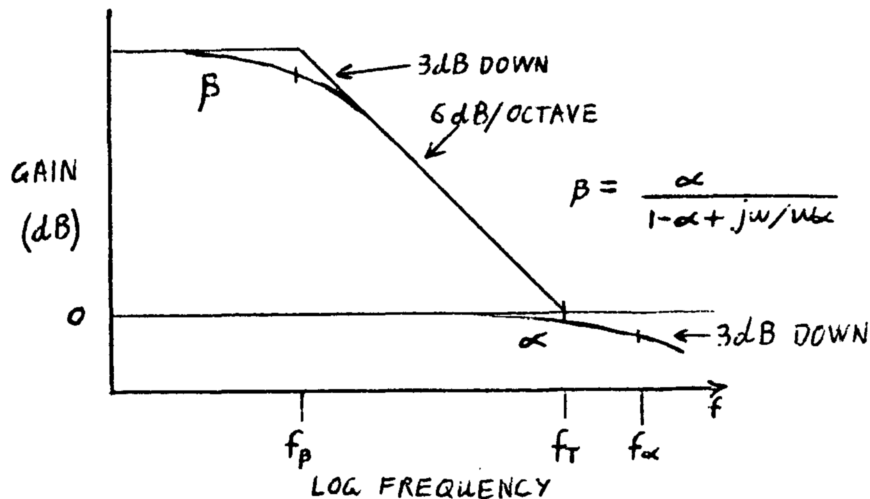


Figure 1.2.1. Idealized Gain-Bandwidth Diagram

The transistor's f_T may be described in terms of the forward transit time, τ_F, by the reciprocal relationship $f_T = 1/2\pi\tau_F$. One component of the forward transit time is due to the time taken for base minority carriers to reach the collector, after having been injected from the emitter. For silicon the electron mobility is approximately three times that of the hole mobility, which means that an n-p-n transistor offers a speed advantage over a similar p-n-p device. The choice of silicon is attributable in the main part to ruggedness and maturity of process. Germanium and gallium arsenide both exhibit higher mobilities than silicon but they suffer from several disadvantages. Neither material has a natural oxide which makes the photolithography steps more complicated than for silicon. In addition, germanium devices do not have the thermal ruggedness of silicon,

being prone to thermal runaway, while gallium arsenide substrates are difficult and expensive to prepare. At higher currents, collector current is directly proportional to the emitter periphery. Due to current crowding the central region of the emitter is biased off and contributes little. Current densities can exceed 2000A/cm^2 [100] or 60μA/μm of emitter periphery [78]. Measurements carried out by the author, on commercially available transitors confirm this. Detailed measurement of a BFW16A's geometry shows that under recommended maximum current conditions the current density in the emitter would be 2700A/cm^2 while the current density in the periphery would be 45μA/μm.

Current density in the metallization can reach 10^5 to 10^6 A/cm^2 [94] when electromigration will be a problem in aluminium, although it can be alleviated by using aluminium alloys which tend to prevent the atoms being moved from their lattice sites. The previously mentioned transistor operates with a current density of 4mA/μm in the emitter electrode metallization. For an assumed typical thickness of around 1μm this would be 4×10^5 A/cm^2.

Modern high frequency geometries maximize emitter periphery while maintaining minimal area. To achieve this three structures are commonly used: interdigitated, overlay and mesh, Figure 1.2.2. In the first two the

emitter is constructed from many narrow stripes, which may be minimal length in the overlay, while in the third the emitter is in the form of a mesh.

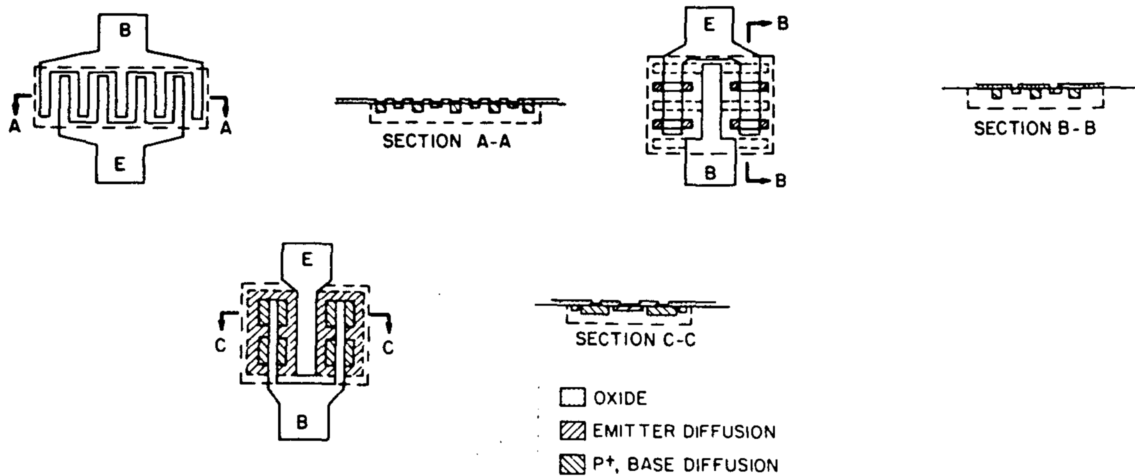


Figure 1.2.2. Common HF Transistor Structures

Interdigitated devices are used for small signal and medium power applications only, since the long thin emitter fingers are not well suited to carry large currents. The metal base fingers help to give a low base resistance, which coupled with the low operating currents tends to give a more linear operation, than overlay transistors. Overlay transistors have a much wider emitter metallization, which makes them suitable for operation with large emitter current densities. The base resistance however is larger, since part of the extrinsic base is formed from a diffused grid. This coupled with a larger base to collector junction area, due to the area required by the diffused base grid, tends to produce a less linear transistor. Overall the distortion tends to be worse in overlay transistors [162].

The ultimate performance that can be expected from a silicon bipolar device has been approached by various authors [161]. On the basis of a stripe geometry an assessment of the maximum frequency of oscillation, f_{max}, has been obtained. For silicon this is;

$$f_{max} = \frac{40}{s+2t} \text{ GHz}$$

where s is the stripe width and t is the spacing, both being in microns. Maximum power and maximum frequency can be related to the extrinsic base resistance by;

$$P_{max} f^2_{max} = \frac{6.25 \times 10^{20}}{r_b} \text{ WHz}^2$$

On the assumption that the transistor is properly matched to the load. Notice that the maximum power can be doubled by halving the base resistance, this being achievable by doubling the emitter periphery. In effect this is the same as putting two transistors in parallel. This is the fundamental method of increasing the power handling capability of any given type of transistor structure. It should be noted that in principle this does not alter the frequency characteristics of the transistor, this being a function of the type of structure used.

Apart from power handling, the maximum collector voltage

that can be handled is of importance. This is very much the case in radio transmitter applications where the load is usually inductive. The maximum voltage that can appear across the, reverse biased, collector base junction is twice the supply voltage due to the back emf of the load. Maximum voltage of operation can be related to the unity gain frequency [161] by;

$$V_{max} f_T = 2 \times 10^{11} \, V/S$$

The maximum frequency of oscillation can be related to the gain bandwidth product [78, 135] by;

$$f_{max} = \left[\frac{f_T}{8 \Pi r_b C_c} \right]^{1/2} \text{Hz}$$

where C_c is the collector transition capacitance. Some typical figures for a medium power device might be;

$s \approx t \approx 3 \mu m$

$r_b \approx 10 \Omega$

$V_{max} \approx 25V$

$C_c \approx 1pF$

giving;

$$f_{max} \approx 5GHz$$

$$P_{max} \approx 3W$$

$$f_T \approx 5GHz$$

Modern integrated circuit transistors with sub-micron line-widths, have f_T approaching 20GHz. The super, self-aligned process technology (SST) device was reported [54] as having an f_T of 17GHz when working at a current density of $30kA/cm^2$, the emitter being 0.35μm x 16μm and working at 2mA. This works out to 62μA/μm of emitter periphery. Power devices can achieve f_{max} greater than 15GHz with peak pulse power greater than 100W at 1GHz.

1.3 Applications Using UHF Transistors

The UHF transistor is used in applications ranging from low power integrated circuits [43, 54, 65] to high power transmitters of hundreds of watts output power. For high frequency power gain transistors are used in common emitter. Class AB operation is often employed for good linearity and low quiescent current. Class B push-pull is slightly more complicated but is effective in cancelling even order distortions, due to its balanced operation.

Medium power, general purpose, wide bandwidth amplifiers are available off the shelf. An example is the Mini-circuits ZHL-2-12, which will provide a minimum of 24dB ± 1dB gain over the frequency range of 10MHz to 1.2GHz. Maximum output power is at least 28.5dBm at the top end of the frequency range. The noise figure is typically 10dB and the third order distortion intercept point (the output power at which fundamental power and third order distortion power are projected to be equal) is typically 38dBm, which corresponds to IM3 of -64dBm (Chapter 4). Power supply requirements are 24V at 0.75A. A circuit of similar performance over the bandwidth 0.4MHz to 150MHz, ZHL-3A, requires 80% of the current and costs 40% of the price. The premium for performance above 1GHz is apparent.

A typical application for medium power devices is that of CATV (community antenna television). Bias conditions might typically be I_c of 80 to 120mA and supply voltage of 10 to 15V with gain at 500MHz greater than 13dB [105]. Good linearity for these devices is essential since the cross-modulation generally sets the upper limit of signal level. The input devices are not so critical due to lower signal levels.

An example of performance obtainable with a 2N5109 transistor over the frequency range 50 to 300MHz [161] is 52dBm output with cross-modulation distortion of -57dB.

Sono-buoy applications require typical output power of between 0.25W and 1.5W at 165MHz. Harmonic output is required to be 40dB down from the carrier. For a one watt output power an IM2 of -64dBm is required and for third order an IM3 of -84dBm.

For mobile radio 12V or 20V operation is normal. The transistors may be required to withstand twice the supply voltage if a transformer output is employed. Single side-band transmission is normally employed as it exhibits a small channel width. That is, it utilizes a small bandwidth thus enabling a large number of radio transmitters to be operated in the same area. Signal-to-noise ratio is good since most of the transmitted power is used to encode the information, with suppressed carrier operation. To make best use of the available bandwidth and to take advantage of the narrow channel of this type of transmission requires that no transmitted power is outside of the allotted frequency band. Cost precludes the use of multi-pole filters and so the devices themselves are required to be linear.

An example of a suitable amplifier uses a 2N5070 transistor as the output device, able to deliver 5W peak envelope power over the 2MHz to 30MHz bandwidth. With two tones applied, gain is greater than 40dB and intermodulation products are more than 40dB down on either signal. This corresponds to an IM3 of approximately

108dBm. In order to achieve this the output device employs emitter degeneration to trade some of the available bandwidth for an improved linearity.

1.4 Disadvantages of Bipolar Junction Transistors (BJTs)

The bipolar junction transistor is fundamentally a non-linear device. The basic gain mechanism involves the exponential law of the base-emitter p-n junction. Additionally the variation of current density within the devices causes the current gain to be current dependent. At higher frequencies the junction capacitances become important due to their lower reactance. These are functions of the doping profiles and are also non-linear functions of the terminal voltages.

For maximum power gain the transistor is used in the common emitter configuration. The power gain is approximately $R_L h_{fe} gm$, where R_L is the collector load resistance and gm is the small signal transconductance. Current gain, h_{fe}, is nearly constant at moderate collector currents but varies in inverse proportion at higher currents. Transconductance is approximately I_E/V_T at moderate currents and $I_E/2V_T$ under "high-level injection", when the injected minority carrier density is comparable to the majority carrier density in the base. The thermal voltage V_T is given by kT/q where the symbols have their usual meanings. Power gain is therefore seen

to be a function of the emitter current.

Input impedance at low frequency is approximately $(h_{fe}+1)V_T/I_E$. The parallel diffusion capacitance is given by $MI_E/2\Pi f_T V_T$, where M is a doping gradient factor between 0.1 and 1 [169]. Input impedance is thus both current and frequency dependent. The frequency dependence can be compensated relatively simply by suitable matching networks [162]. Current dependence, on the other hand, is non-linear and difficult to accommodate. It represents a non-linear load to the driving circuit causing distortion at the input of the transistor.
Output of the transistor is via the collector current which develops power in the (linear) collector load. The basic transfer function, $I_C = V_{BE}gm$, is non-linear due to the current dependent nature of the transconductance. At higher frequencies the reactances of the emitter and collector capacitances become important and can modify the distortion performance. Under these conditions feedback by the non-linear collector-base capacitance tends to reduce the second order distortion caused by the non-linear transconductance. The generally more objectionable third order distortion, however, tends to be increased.

Thermal problems with bipolar devices cause much concern. One of the basic problems is that the base-emitter voltage reduces by about 2mV/°C. As a result of this any local

hot spot tends to have an increased base drive and allows more current to flow, thus making it hotter still. This current hogging mechanism means that a large emitter transistor must be designed with care to ensure adequate current sharing. The most satisfactory method is to use emitter ballasting, whereby the emitter is broken into many small sites each with a small series resistance. Extra current flowing in a resistor tends to bias the associated emitter off, thus preventing hogging. This resistance is in series with the emitter and reduces the available transconductance to $gm/(1 + gmR_E)$ where R_E is the resistance in the emitter. For a collector current of 100mA the gm is approximately $4\Omega^{-1}$ at room temperature and an R_E of only 0.25Ω will halve the available transconductance. As the current requirement increases, the number of emitter sites are increased. In effect more parallel transistors are used for higher current operation and so the problem does not worsen.

Second breakdown can cause destruction of the bipolar transistor if it is operated outside the safe working conditions. Breakdown occurs due to avalanche carrier multiplication within the transistor. Collector current then increases rapidly with small increase in collector voltage once the breakdown voltage is reached. Part of the multiplication can be caused by the transistor action itself, but reducing the base drive impedance reduces this and allows a higher collector voltage to be sustained.

Once breakdown has occurred second breakdown can occur. This appears as a large increase in current associated with a rapid reduction in collector voltage. The mechanism is caused by hot spot formation [158] and is now understood sufficiently to be avoided by designing for operation within the safe operating conditions.

1.5 The Power FET

The FET has a number of fundamental advantages over the BJT: high input impedance, negative temperature coefficient of carrier mobility, square law transfer function and majority carrier operation.

Higher input impedance eases matching of the input and enhances the power gain. Typically the high frequency input impedance is several hundreds of ohms for an FET compared with a few ohms for a bipolar transistor. The input impedance is dominated by the gate-substrate capacitance. The DC bias current is essentially the gate leakage, and can usually be ignored. Absence of DC current flow in the gate of the transistor prevents de-biasing due to IR drops, consequently current crowding does not occur.

The negative temperature coefficient of carrier mobility makes the FET self-regulating. The current density decreases with increase in temperature and prevents

current hogging. The FET is thus immune to thermal runaway and second breakdown. This in turn simplifies the construction of large area devices. Ideal MOS transistor operation obeys a square law which is advantageous for communications applications as a reduction can be achieved in the third order intermodulation distortion products. Essentially the drain current varies as the square of gate voltage and so distortion products due to the gain mechanism are even order. This ideal picture however is clouded by two other effects. In the first place high frequency operation requires a short channel length since carriers must not take an appreciable time to drift from source to drain. In this case channel length modulation becomes an important factor, that is, the channel length is modified by the depletion region around the reverse biased drain. Transistor operation then departs from ideal square law. This mechanism is in fact very similar to the Early effect in bipolar transistors. The second mechanism is due to the non-linear drain-channel and drain-substrate capacitance which becomes important under high frequency operation.

Majority carrier operation means that the FET does not suffer from any significant charge storage problems. N-channel transistors are almost exclusively employed because electron mobility is greater than hole mobility in a given material.

The main delay components of FETs are the transit time of carriers in the channel and the CR time constant arising from the gate capacitance and transconductance. Transit time is given by;

$$\tau = \frac{L^2}{\mu_n V_D} \quad S$$

where;

L is the channel length

μ_n is the carrier mobility

V_D is the drain-source voltage

In silicon the scattering-limited velocity, V_{sc}, of about 10^7 cm/s is reached when the electric field strength exceeds 5×10^4 V/cm. For a sub-micron channel length this figure is exceeded with a 5V drain voltage. Under these conditions the transit time is;

$$\tau = L/V_{sc}$$

However, the transit time is often smaller than the C-R time constant [158]. The gain-bandwidth product of the device is then given by;

$$f_T = \frac{gm}{2\Pi C_{GS}} \quad \text{or} \quad \frac{v_{sc}}{2\Pi L}$$

The channel length, L, is important because both gm and f_T vary in inverse proportion. An f_T of 30GHz can be expected [158] for a silicon MESFET with a channel length of 0.5µm.

Similar devices in GaAs and InP might be expected to give 40GHz and 70GHz respectively. Shorter channels are required to be shallower for adequate control. This requires a larger doping concentration in the channel, which in turn compromises the drain breakdown voltage. It is projected that gate lengths of 0.1µm could be achievable giving a silicon MESFET an f_T of 80GHz. Gate width can be determined from a projected maximum drain current of 5A/cm of gate width. State-of-the-art GaAs MESFETs have a power-(frequency)^2 figure of around 7×10^{20} WHz^2, which is about twice that for state-of-the-art silicon bipolar devices [158].

From a cost viewpoint the bipolar transistor still holds the advantage, benefitting from a high yielding, mature process. The GaAs MESFET on the other hand suffers from materials and process problems which lead to low yields and high cost.

1.6 New Life for the BJT

Improvements to the BJT are still being made. The well known goals are the reduction of: base-width, base resistance and base area to emitter periphery ratio. The reduction in base-width has probably reached a minimum, subject to constraints imposed by breakdown. Reduction of base area is obtained by both reducing the emitter to base electrode spacing, which also reduces the base resistance, and the base electrode size.

Both of these objectives are being achieved by a combination of new techniques and processes, and also by improvements in established processes.

Improvements in photolithography allow sub-micron line-widths, which allow smaller geometries to be fabricated. The use of polysilicon allows electrode fabrication to be followed by oxidation, thus providing isolation. It also offers the advantages of: isolating thin diffusions from the possibility of aluminium spikes, autodoping and distributed emitter ballasting [31]. Ion implantation allows a high degree of control over the doping profile and total number of impurity atoms. Being a low temperature process it enables a large variety of materials to act as a mask which would not be possible with furnace deposition. Dry etching offers both isotropic and anisotropic etching to be performed.

Isotropic etching allows the autoregistration of contact holes to the underlying surface topography, as in the super self-aligned process [63] for example. Metallization improvements have meant that multiple metal systems can be used to improve reliability by reducing electromigration. Use of silicon nitride allows better passivation of shallow junctions and can also be used as a secondary etch mask.

All of these processes have helped in the improvement of bipolar transistor performance. Some improvements have been directed towards integrated circuit application. The SST, for example, [63,43] has 0.3μm emitter-base spacing and an f_T of 12GHz. It uses polysilicon emitter and base electrodes with 0.5μm and 0.3μm contact sizes. This structure could conceivably be used as an interdigitated structure with emitter pitch of about 3μm. Other structures use a side-wall contact to the base and [154, 158] an f_T of 14GHz is achievable due to the reduced base resistance of these structures.

Some improved devices are interdigitated [113] with 2μm pitch emitter and 0.5μm emitter opening. Devices with up to 40μmx10μm long fingers have been manufactured and show a theoretical f_{max} of 30GHz. Biased at 1mA per emitter with V_{CE} of 10V it has an f_T of 10GHz, giving 15dB gain at 4GHz. Small signal base resistance is 7Ω. The stepped electrode transistor (SET) [31] (see section 2.5) has an

emitter pitch of 4μm. With 1.5μm base contact and 1.7μm emitter contact, the emitter base spacing is 0.4μm. This device has an f_T of 8.4GHz.

Current state-of-the-art includes 500W pulse power at 1GHz [158] and continuous power of 60W at 2GHz, 6W at 5GHz and 1.5W at 10GHz. These have power-(frequency)^2 products of 240, 150 and 150W(GHz)^2. An improvement of up to three times might be expected with future improvements in fabrication and process techniques.

1.7 The Scope of this Thesis

This chapter has shown that there is a requirement for bipolar transistors with improved linearity. Work described in this thesis is based on the assumption that bipolar transistors will continue to be important in applications requiring high frequency gain with good linearity.

The subject of this thesis is a novel overlay bipolar transistor structure that had been proposed to the author. Part of the novelty was in the proposed use of a metal base electrode to replace the conventional diffused grid. Additional to this was the novel idea of constructing the base electrode prior to the emitter formation. The idea being to subsequently autoregister the emitter within the base grid in order to achieve good current sharing and

minimal use of silicon area. Such a transistor structure might be expected to exhibit both good frequency performance and good linearity.

The novel transistor structure is compared to conventional transistor structures in Chapter 2. It is shown that, for a given process, improved performance can be achieved by means of reducing the feature size. This point is discussed in terms of mask alignment restrictions and it is shown that the new structure has some advantage. An additional advantage of improved current distribution is also analysed and results presented.

Device modelling is discussed in Chapter 3, where the idealized models for small and large signal conditions are reviewed. This is followed by models with application to non-linear analysis and modelling of the main second order effects. A model for the base electrode series resistance is then developed and applied to the novel structure in order to determine a reasonable geometry for the novel transistor.

The fourth Chapter gives a short review of the history of distortion modelling and then describes the types of distortions produced by the various non-linear mechanisms within the transistor. Analysis published in the literature is presented and then extended to fifth order terms. Distortion due to several of the major second

order effects is then analysed and results presented.

The novel transistor structure is discussed in Chapter 5 where, the individual process steps are critically discussed. This leads to the development of a target structure, and a tentative process schedule.

Experimental work was carried out in order to develop a complete processing schedule to manufacture the novel transistor. This work is described in Chapter 6 where a number of technical problems are discussed and their solutions presented.

A measurement procedure was prepared for the evaluation of distortion performance. Sample measurements where taken on conventional production devices and the results presented in Chapter 7 for completeness.

An investigation was made into the usefulness of SPICE for distortion analysis. Fairly simple closed form solutions exist for first order distortion performance in both the high and low frequency limits. The more complicated results between these limits suggest the use of a computer. Results of this and the conclusions drawn are presented in Chapter 8.

Production of a working version of the novel transistor was not possible within the resource constraints of the

project. The unusual process sequence inhibited some potential offers of resources due to the fear of possible contamination. Conclusions drawn from the work carried out are presented in Chapter 9.

2.1 High Frequency Power Transistors

High frequency power transistors operate at large current densities, often above 2000A/cm² [100]. This fact, coupled with the negative temperature coefficient of the base-emitter voltage at constant current can lead to an uneven current distribution. A well-known method of avoiding this 'current hogging' is to break the emitter into a number of isolated segments, introducing a resistor between each and the emitter contact. Separate currents then flow in each emitter-resistor combination. Any tendency towards hogging, with its associated increase in current density, is counteracted by the increase in the voltage drop across the resistor which tends to bias off the emitter. Thermal coupling between emitters and the form of the external bias circuitry have an influence on the actual resistance required in each emitter. The problem has been investigated by Bosch [17].

Lateral current flow in the intrinsic base region causes a reduction in junction bias towards the inner regions of the emitters. The resultant preferential biasing of the emitter periphery causes current crowding. Under these conditions the centre of the emitter is a very poor injector of carriers and it is thus preferable to make the emitters with a large periphery-to-area ratio. Typical

current density in the emitter periphery is 1.5mA/mil (60μA/μm) [78].

With a requirement to make many narrow emitters three structures have become prevalent. These are the interdigitated, overlay and mesh structures.

For high frequency operation the parasitic components must be kept small. Those in the extrinsic regions can be reduced, to some extent, by means of the doping profiles and the geometry whilst in the active region the current densities tend to dictate the physical size. Later it will be demonstrated that an autoregistered structure has a potential advantage in reducing the intrinsic and the extrinsic parasitic components by ensuring a more efficient use of the emitter, thus allowing it to be smaller.

The vertical structure of the transistor is made small, sub-micron, to reduce the total delay and ensure a high f_T. Current flow is from the emitter on the top surface of the die to the collector, substrate, contacted from the back. The need for a low resistance contact to the collector, for a low saturation voltage, and a low doping density at the base junction, to give high breakdown voltage and low capacitance, is met by use of an epitaxial structure. An n-p-n transistor, which could be any of the previously mentioned structures, would normally be made

in a thin n-type epitaxial layer on an n+ substrate.

2.2 Interdigitated Transistors

This structure is so called because of the alternate emitter and base, finger shaped, contacts. A typical geometry is shown in Figure 2.2.1.

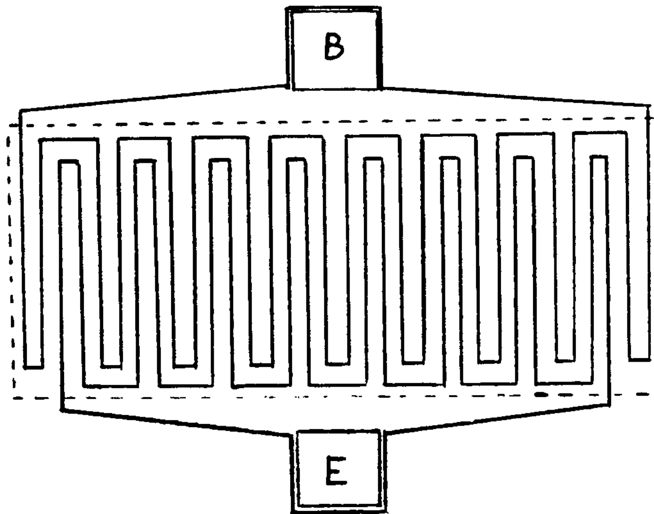


Figure 2.2.1. Interdigitated Transistor

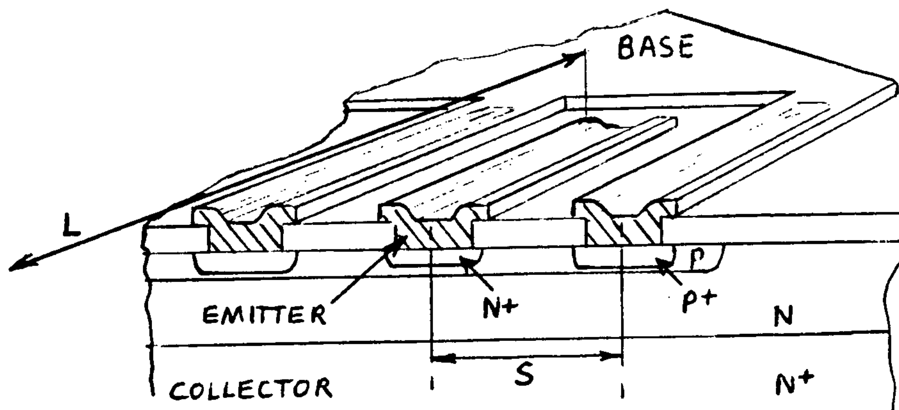


Figure 2.2.2. Cross-section of Interdigitated Transistor

The structure, shown in cross-section, Figure 2.2.2, is typically made in an epitaxial layer of 3 to 10μm

thickness. The base contact diffusion is not required if the base surface concentration of dopant is large enough, $N_A \geq 10^{18}$.

This type of structure has been well analysed and useful equations relating the device performance to the geometric sizes are well known [158, 161, 162, 166, 169]. The forward transit time, τ_F, and thus the f_T is determined by the delay experienced by the carriers as they pass from emitter to collector. Forward transit time is therefore primarily a function of the vertical structure, and to first order is independent of the stripe width, S and length, L. The maximum frequency of oscillation, f_{MAX}, can be related to f_T by;

$$f_{MAX} = \left[\frac{f_T}{8\pi r_{bb'} \cdot C_{cb}}\right]^{1/2} \quad \text{Hz}$$

and since $r_{bb'}$ and C_{cb} are dependent on S and L, to first order, according to;

$$r_{bb'} = \frac{r_o S}{L}$$

and;

$$C_{cb} = C_o SL$$

We find;

$$f_{MAX} = \frac{1}{S}\left[\frac{f_T}{8\Pi R_0 C_0}\right]^{1/2} \quad \text{Hz}$$

Which indicates that the maximum frequency of operation is predominantly a function of the stripe width, S. Here S has been used in a loose sense, in that for the case of the base spreading resistance, it represents the spacing between the emitter and base electrodes while in the expression for the collector-base junction capacitance, it represents the pitch of adjacent emitter electrodes. The model is correct to first order for the case of a scaled geometry where S represents the scaling factor. The expressions would be expected to lose accuracy for small geometries when the doping profiles begin to be modified due to the, then excessively disproportionate, sideways diffusion.

The power gain may be expressed by;

$$G = \frac{f_T}{f^2\, 8\Pi r_{b'b} C_{c'b}}$$

which is related to the stripe width by;

$$G = \frac{f_T}{s^2 f^2 8\Pi R_o C_o}$$

These equations assume negligible voltage drops along the length of both the emitter and base stripes. This is reasonable for an L:S ratio of typically 20 or less.

It is clear that for high frequency performance the device dimensions should be as small as possible. A lower limit is fixed by the resolving ability of the lithographic process used and also the ability to align successive process layers.

Current flow at moderate to high levels will be concentrated at the edges of the emitter. The amount of current, for a given base-emitter terminal voltage, is dependent on the extrinsic resistances of the base and emitter. When the emitter is not centrally located between the adjacent base fingers, the base resistance to each side of the emitter will be different leading to different current densities. In order to balance the currents on the two sides of each emitter, and over the whole device in general, an autoregistered structure may be used. This technique enables the base and emitter areas to be defined with the same mask which largely eliminates asymmetry.

The p+ regions serve two purposes. First, they improve

the ohmic contact between the base metallization and the intrinsic base, desirable because metal to lightly doped silicon tends to make a rectifying junction. Second, the sideways diffusion of the p+ reduces $r_{bb'}$ by moving the high conductivity base contact closer to the emitter. Downward diffusion of p+ and up diffusion of the n+ extrinsic collector put a lower limit on the epitaxial thickness, because if the n+ actually contacted the p+, a very low V_{cbo} would result, typically making $V_{cbo} \approx 5V$.

Current sharing may further be encouraged by interposing a polysilicon layer under the emitter metal [159]. Such a structure is shown here in Figure 2.2.3.

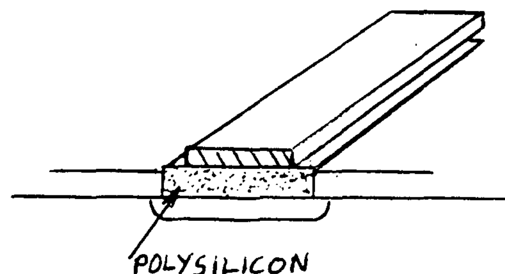


Figure 2.2.3. Structure with Polysilicon Emitter Contact

Where each side of the emitter diffusion is seen to have a series resistance i.e. it forms an emitter ballast resistor. This polysilicon layer has the added advantage of preventing aluminium, if used, spikes penetrating to the base. These can occur by the formation of

intermetallic compounds under severe hot spot conditions or during alloying of the contacts [32], when the aluminium can diffuse through the shallow emitter. Additionally the polysilicon may be used as a diffusion source for the emitter formation [31].

2.3 Overlay Transistors

This type of transistor takes its name from the emitter metallization which is over the base electrode instead of beside it, as in the interdigitated structure. A typical structure is shown in Figure 2.3.1.

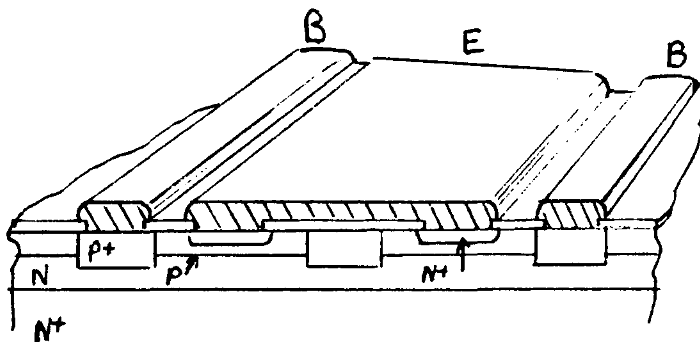


Figure 2.3.1. The Overlay Transistor

The main advantages of this type of structure lie in its handling of the emitter current. The emitter is broken up into many small sites which gives it a large periphery to area ratio allowing high current handling. Additionally the emitter metallization is much wider than the narrow fingers of the interdigitated structure. This gives

improved reliability from the point of view of electromigration, owing to the lower current densities in the emitter electrode. Processing is made easier as a result of the coarser metallization and the yield is consequently larger.

Base contact is made by an additional p+ grid which serves to distribute the base current more evenly and also enhances the ohmic contact.

Compared with the interdigitated transistor the overlay will give a larger periphery/area ratio but it has larger parasitic components associated with the extrinsic base [28]. For requirements where maximum power is needed the overlay transistor is preferred. The interdigitated transistor on the other hand is preferred when maximum frequencies are required.

The parasitic components associated with the extrinsic base tend to limit the frequency performance. These components can be large due to the need to allow for misalignment of the emitter with respect to the p+ base grid. Actual contact between these two highly doped regions is to be avoided as it compromises breakdown voltage and emitter efficiency. Under these conditions normal avalanche breakdown would be reduced to a lower zener breakdown. Emitter efficiency would be reduced in the locality due to the smaller ratio of resistivities.

Typical geometry for an overlay transistor contemporary with the start of this research work is shown in Figure 2.3.2.

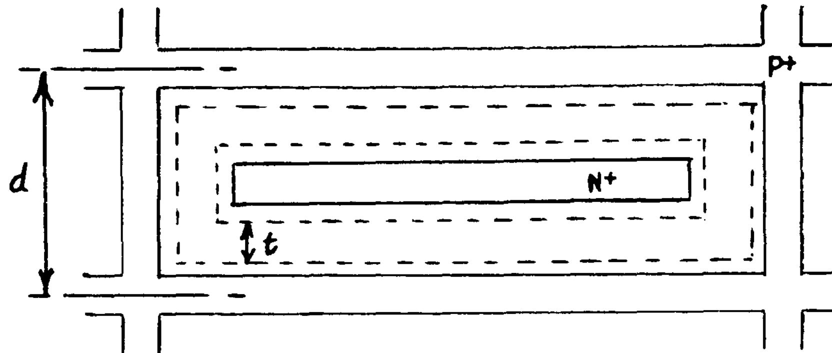


Figure 2.3.2. Emitter and Base Grid of Overlay Transistor

Here the p+ grid is drawn as 2.5μm wide and the emitter as 1μm wide. Sideways diffusion of the relatively deep p+ grid is about 2.5μm while that of the emitter is about 0.5μm. The misalignment tolerance, t, is about 2 to 2.5μm. This gives a repeat distance, d, of 12.5μm.

2.4 Mesh Transistors

This type of structure [78] forms the emitter into a mesh around the base contact areas, a typical structure is shown in Figure 2.4.1.

It is most suitable as a power device, having a lower $r_{bb'}$ but a larger R_e. Current sharing in the emitter will not be optimum.

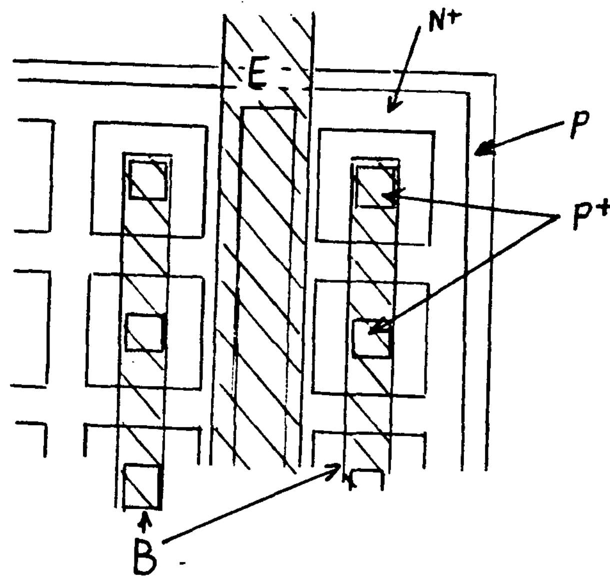


Figure 2.4.1. Mesh Transistor

2.5 Novel Types of Transistor

The previously described transistors are normally fabricated by means of a double diffused process, that is to diffuse both the emitter and base. The epitaxial layer forming the active part of the collector. Aluminium, actually an alloy of aluminium and silicon to reduce electromigration and to eliminate contact pitting, is used to form contacts to the base and emitter on the top surface while a collector contact is made to the back of the die.

The last ten years have seen dramatic advances in the techniques available to the transistor manufacturer. Use of ion implantation is now common place, allowing the doping profiles to be closely controlled while retaining

very shallow junctions. Base and emitter junction depths of 0.4μm and 0.2μm respectively are readily achievable. The key to this ability is that impurities can be placed directly into the silicon without significantly moving impurities already sited. Crystal damage is caused by the implantation process and must be annealed out [55] by heating the silicon, but the impurities need not undergo significant diffusion. Total dose of impurity can be controlled to allow transistors with very small base storage time.

Polysilicon is another useful option for transistor manufacture; this may be used simply to separate the emitter from the aluminium metallization to prevent the metal penetrating to the junction, which can occur under normal operating conditions. In addition the resistivity can be varied over a wide range which allows the polysilicon to be used as a distributed emitter ballasting resistor.

A further possibility is to use the doped polysilicon layer as the impurity source for a subsequent junction diffusion. Both ion implantation and polysilicon can provide for differential etch rates which can be used to advantage. Etch rates for Si, SiO_2 and Si_3N_4 increase with implantation dose rate [107, 31] while the implant energy determines the depth to which the etch enhancement extends. Polysilicon has an etch rate determined by the

doping density allowing control of the edge profiles by varying doping density throughout the thickness. Advances have been made in the use of plasma and ion enhanced etching, these allow isotropic or anisotropic etching to be carried out at will. Structures with large height to width ratios can be obtained by these measures which are unobtainable with wet etching.

Multiple metallization schemes are now used, in some cases to avoid electromigration which could reduce reliability in aluminium metallized transistors.

The stepped electrode transistor (SET) [31] is an example of use of some of the newer techniques, Figure 2.5.1. The n on n+ wafer has a p-type base diffused in the normal fashion, two layers of polysilicon are deposited and photoengraved to produce the inverted trapezoid over the emitters, using the differential etch rate of the two layers.

Figure 2.5.1. The Stepped Electrode Transistor

Conformal layers of SiO_2 followed by Si_3N_4 are deposited by chemical vapour deposition (CVD). The emitters are formed by using the emitter contact as an, As, impurity source. The surface of the wafer is then ion implanted to enhance the etch rate of the SiO_2 and Si_3N_4 layers everywhere except where shadowed by the emitter contact. The base windows are opened by chemical etching and B diffused into the base contact region. The emitter contact windows are opened by unmasked etching. Metallization is carried out to form both the emitter and base electrodes, which are isolated because the metal cannot cover the step formed by the emitter contact. The resulting geometry is small because it does not rely on the accuracy of mask alignment to obtain separation between base and emitter.

2.6 Effects of Misalignment

A misalignment tolerance must be allowed when emitter and base contacts are defined by separate process steps. This accommodates the uncertainty of each mask's location with respect to the part processed wafer. The author has taken detailed measurements of a commercially available transistor, BFW16A, with a travelling microscope. These measurements reveal misregistration of up to 1.7μm. The emitter pitch was found to be 9μm with an average distance between base and emitter of 1μm. Examples have been measured where the misalignment is such that the emitter

to base spacing is 1/3μm on one side and 5/3μm on the other. This is a ratio of 5:1 which, even allowing for measurement tolerances, serves to show that large ratios can be caused by fairly small misalignments if the geometries are small. Misalignment is to be avoided if possible since it limits the minimum emitter pitch that can be achieved and may lead to excessive current flow in parts of the emitter. This latter consideration is now analysed more fully.

Structures which have alternate emitter and base areas can be considered as an emitter flanked by two base contacts. The comparative spacing of the base contacts to emitter on either side will be determined by the accuracy of registration of the emitter mask to the previously defined base contacts.

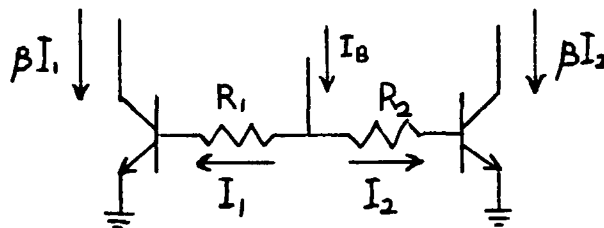


Figure 2.6.1. Equivalent Circuit for B-E-B Structures

An equivalent circuit shown in Figure 2.6.1, consists of two similar transitors with variable base spreading

resistances, the sum of which is approximately constant with varying degrees of misalignment.

Now defining;

$$R_B = (R_1 + R_2)/4$$

$$K = I_2/I_1$$

$$M = R_2/R_1$$

where;

R_B is the base spreading resistance for the transistor with no misalignment, i.e. the parallel combination of R_1 and R_2 when they are both equal.

K is the ratio of base currents, ideally 1, as a result of the misalignment.

M is the ratio of the actual base spreading resistances, due to misalignment, which may also be expressed as $(R - r)/(R + r)$, where r is the deviation from the nominal base resistance, R.

From the diagram we see that;

$$I_1R_1 + V_T\ln(\beta I_1/I_S) = I_2R + V_T\ln(\beta I_2/I_S)$$

Rearranging and substituting gives;

$$\frac{4I_BR_B}{V_T} = \frac{(1 + M)(1 + K)\ln K}{(1 - KM)}$$

which relates the current sharing ratio, K, to the misalignment, M, and the voltage drop, I_BR_B, in the base spreading resistance. We see that at small currents K approaches 1 while at high currents it is asymptotic to 1/M. The graph, Figure 2.6.2, shows K as a function of I_B with M as a parameter.

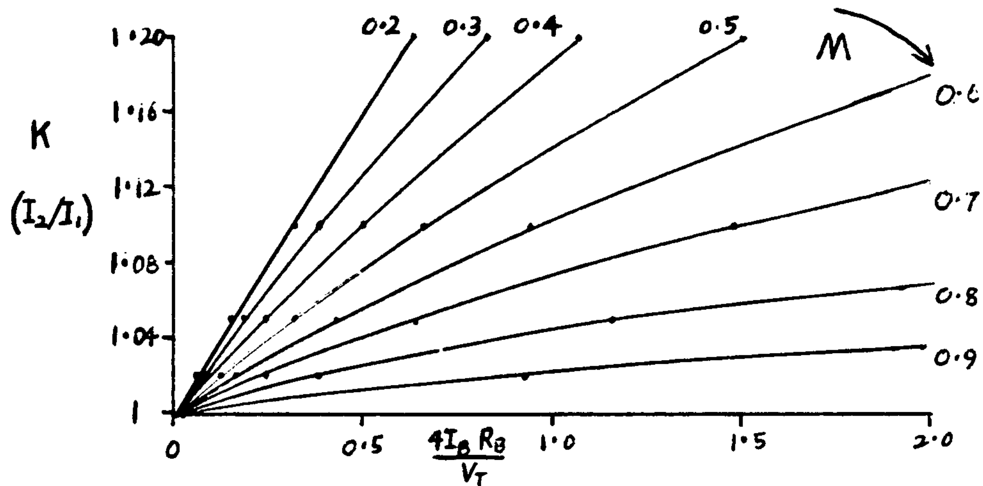


Figure 2.6.2. Current Sharing Related to Misalignment

It is clear that the worst current anomalies occur for M much less than one i.e. most mismatch. It is also clear that the mismatch in current symmetry is worse both for larger base currents and larger base spreading resistance.

2.7 Comparison of Structures

In order to make a comparison between the main structures in common use it is necessary to make some assumptions about doping profiles and manufacturing capability. Common emitter cut-off frequency, f_T, is determined by the forward transit time, τ_F, by;

$$f_T = 1/2\pi\tau_F$$

where τ_F is the total forward transit time of the transistor and f_T, the cut-off frequency, is that frequency at which h_{fe} has fallen to unity. Forward transit time, τ_F, in turn is determined by the vertical structure of the transistor. It can therefore be seen that f_T is determined by the structure (base-width, etc.) below the emitter periphery. This is true at low current densities when current flow is predominantly vertical but may not be true at high currents when base push-out can occur as a three dimensional effect. This, however, is a high current effect and the transistor would not normally be used under these conditions since the f_T would be considerably reduced. It will, therefore, be assumed that a similar process can be used with all the transistor structures considered, in order to provide the same f_T in each case. The maximum frequency of operation, f_{MAX}, may in fact exceed f_T and is defined as the frequency at which the power gain has fallen to unity, assuming the

transistor is used in the most favourable passive circuit.

The relationship can be expressed as;

$$f_{MAX} = \frac{1}{4\Pi}\left[\frac{1}{r_{bb'} C_{cb} T_F}\right]^{1/2} \quad Hz$$

This is an idealized expression and assumes the frequency response is described by a dominant pole formed by the base resistance and the collector-base capacitance. Notice that this dominant pole relationship implies

$$f_{MAX} = (PG)^{1/2} f \quad Hz$$

above the f_β of the transistor, where PG is the maximum power gain available at the frequency of operation, f. Now because it has been assumed that a similar process is used for each structure, both $r_{bb'}$ and C_{cb} are determined, to first order, by the geometry of the structures alone. Figure 2.7.1, shows a general stripe geometry where it can be seen that C_{cb} is proportional to L.P and $r_{bb'}$ is proportional to S/L, where P is the emitter pitch, S is the base emitter separation and L is the emitter stripe length. It can be seen that the product $r_{bb'}.C_{cb}$ is proportional to P.S and f_{MAX} is increased by a reduction in P or S. To compare stripe geometries it is sufficient to compare the P-S product, the smaller the better.

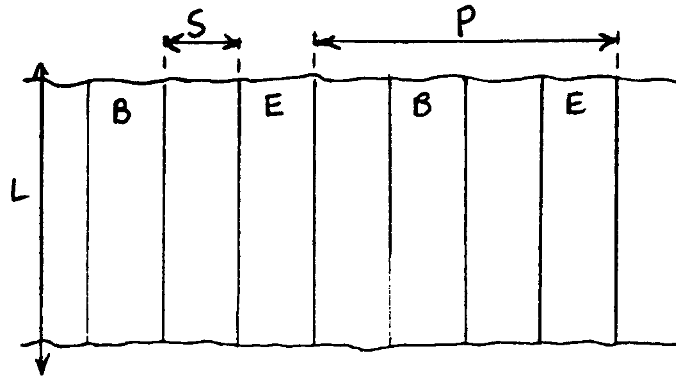


Figure 2.7.1. General Stripe Geometry

More generally, however, the geometry must be considered as a two-dimensional array, Figure 2.7.2, and C_{cb} taken as the total base area required for a given emitter periphery and $r_{bb'}$ as the total base resistance for a given emitter periphery.

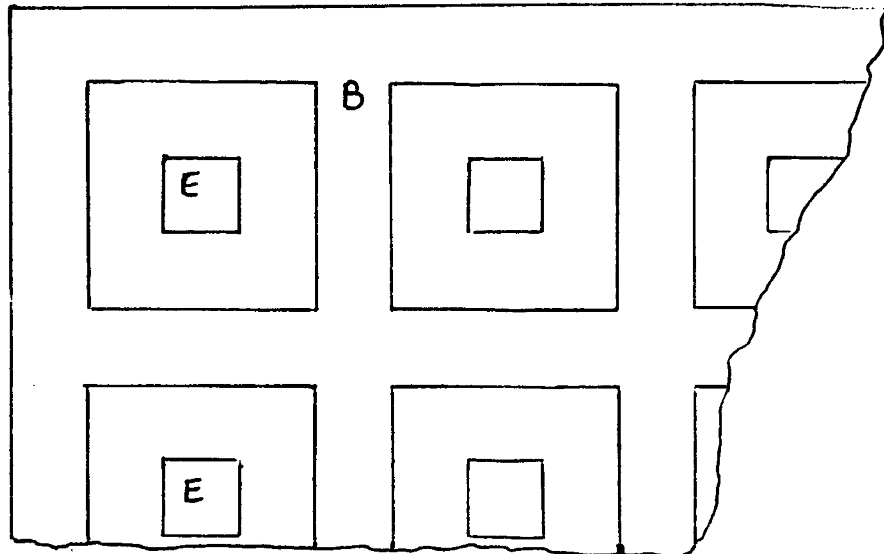


Figure 2.7.2. General Array Geometry

Once again the $r_{bb'}.C_{cb}$ product is independent of emitter length and the analysis can be carried out on one cell.

Notice that the stripe geometry is directly comparable since it is made of long cells and, also, the end effects may be ignored if the stripe is long enough.

The minimum size geometry that can be achieved is determined by three factors: the sideways diffusion of the emitter, and base contact if used, the photolithography limits and the alignment achievable between mask layers. For each structure minimum sizes are determined by various combinations of these factors. These factors will now be considered in detail.

The interdigitated structure is shown in Figure 2.7.3.

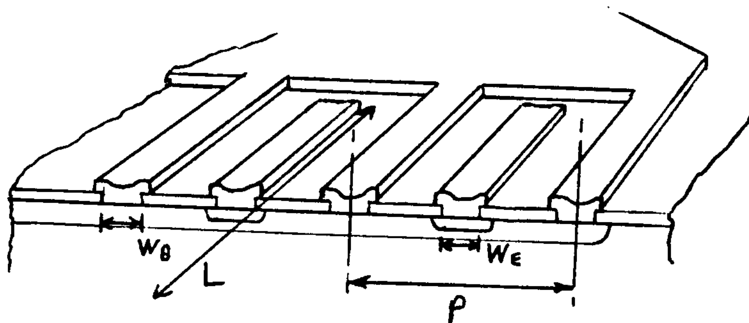


Figure 2.7.3. The Interdigitated Structure

where;

L = Length of emitter, and base, fingers

P = Pitch of emitters, and bases

W_e = Width of emitter contact

W_b = Width of base contact

N = Number of emitter fingers

Consider the fabrication, step by step, for the recut emitter structure. The first critical alignment is the emitter, and base, contact with respect to the emitter diffusion. Since the emitter contact must not extend beyond the emitter diffusion, it must be one misalignment tolerance, M, smaller all round, Figure 2.7.4.

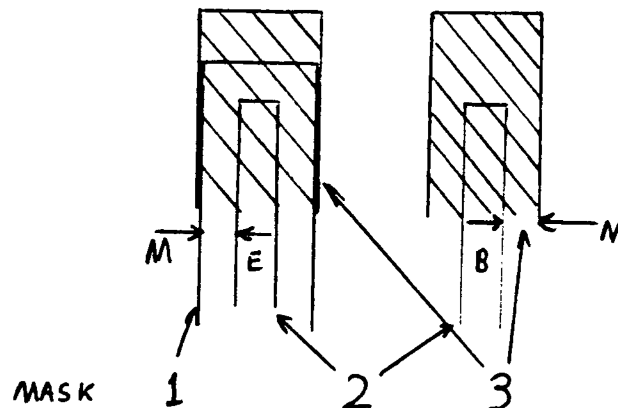


Figure 2.7.4. Recut Emitter Alignment Sequence

The second alignment is that of metal with respect to the contacts, again one misalignment tolerance must be allowed all round. In addition to this a minimum overlap of metal over contact may be required, since coincidence of the metal edge with contact window edge can cause problems during the photoengraving, due to the surface topography.

The pitch, P, between adjacent emitters, in it's simplest form, is thus given by;

$$P = W_e + W_b + 2C + 4a + 4M$$

where;

C is the minimum metal-metal separation

a is the minimum overlap of metal over contact

M is the maximum misalignment allowed.

The emitter diffusion width, W_d, is given by;

$$W_d = W_e + 2M + 2S_e$$

where S_e is the sideways diffusion of the emitter, approximately 0.7 times the junction depth.

Fabrication of the washed emitter structure is a little different, Figure 2.7.5. Here the alignment of interest is the base contact with respect to the emitter contact and one misalignment tolerance must be allowed for. Metal may now be aligned to either the base or emitter contact to give a final emitter pitch of;

$$P = W_e + W_b + 2C + 4a + 6M$$

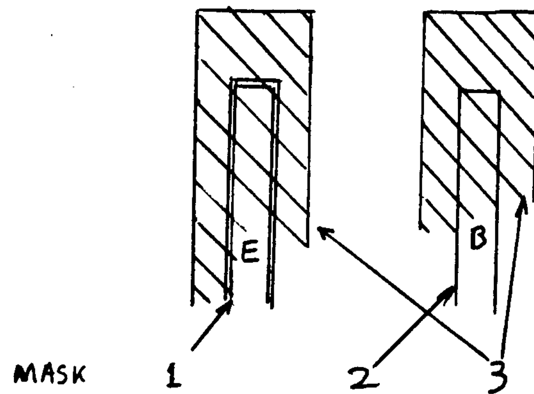


Figure 2.7.5. Washed Emitter Alignment Sequence

and the emitter diffusion width is given by;

$$W_d = W_e + 2S_e$$

The position of the emitter with respect to the two adjacent base contacts determines the current flowing in each side of the emitter. The largest and smallest distances for the recut emitter are;

$$C + 2a + 2M - S_e$$

and;

$$C + 2a - S_e$$

respectively, and for the washed emitter the corresponding distances are;

$$C + 2a + 4M - S_e$$

and;

$$C + 2a + 2M - S_e$$

Other important dimensions for an N emitter transistor (recut or washed emitter) are;

$$\text{Total area of base diffusion} = (L + 2L_e)(NP + 2L_b)$$

where;

L_e is the overlap of base over the end of the emitter stripe and L_b is the overlap of the base over the last base contact.

$$\text{Total emitter periphery} = 2N(L + 2W_d - W_e)$$

$$\text{Total emitter area} = N.W_d(L + W_d - W_e)$$

$$\frac{\text{Emitter Periphery}}{\text{Base area}} = 2/P \quad \left| \begin{array}{l} L \gg 2L_e \\ L \gg 2W_d - W_e \\ NP \gg 2L_b \end{array} \right.$$

A general form of the overlay structure is shown in Figure 2.7.6. The emitter areas may form fingers as shown or may be minimal length forming squares.

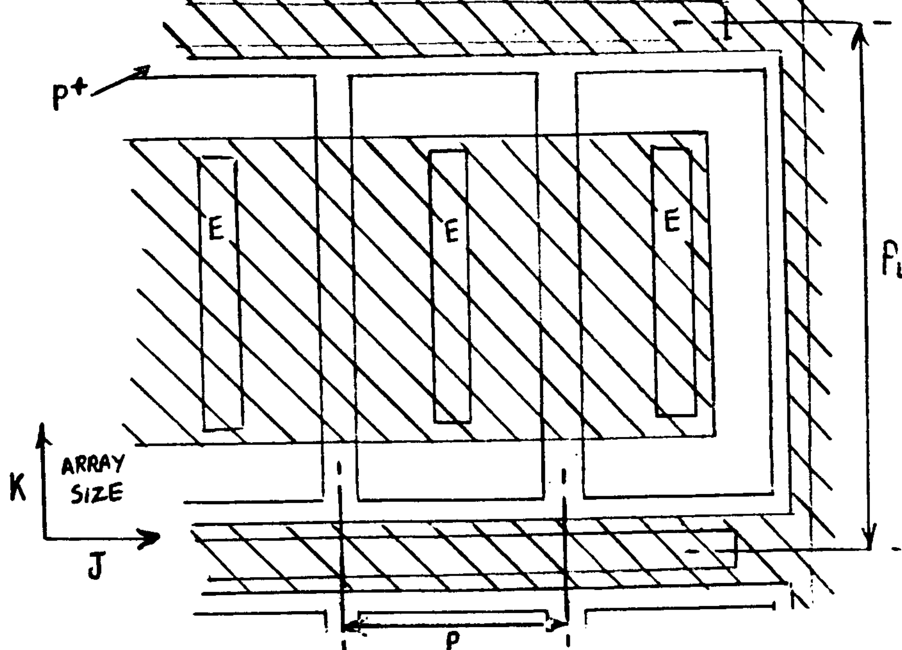


Figure 2.7.6. Overlay Transistor

A typical cross-section is shown in Figure 2.7.7, where it is seen that the p+ base grid and the p-type base are the two deepest diffusions.

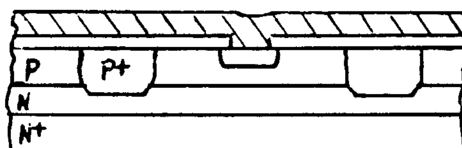


Figure 2.7.7. Cross-section of Overlay Transistor

The order in which these diffusions are made is dependent on the relative junction depths, although both will be made prior to the emitter. The first critical alignment step is the emitter, which must be aligned with respect to

the base grid, Figure 2.7.8.

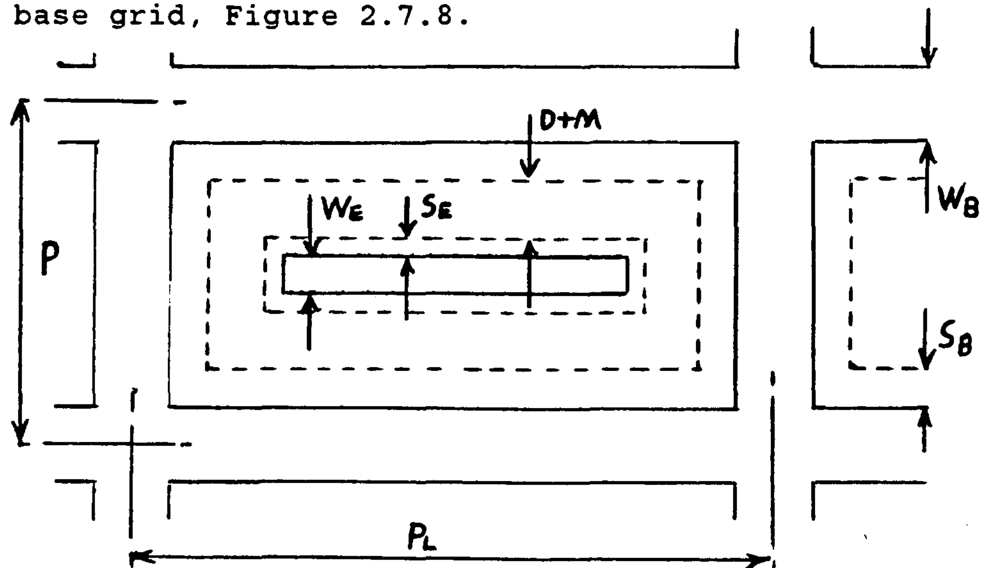


Figure 2.7.8. Emitter Alignment to Base Grid

For a washed emitter the pitch, P, is given by;

$$P = W_e + W_b + 2S_b + 2S_e + 2d + 2M$$

where d is the minimum separation between the emitter and base grid.

The emitter width is given by;

$$W_d = W_e + 2S_e$$

For a recut emitter an additional misalignment tolerance must be allowed for, giving:

$$P = W_e + W_b + 2S_b + 2S_e + 2d + 4M$$

and;

$$W_d = W_e + 2M + 2S_e$$

Notice that P is independent of the metal layer since the base contact is made via the p+ base grid.

For the pitch along the length of the emitter, P_L, there are two cases, Figure 2.7.9, the minimum length emitter and the long emitter.

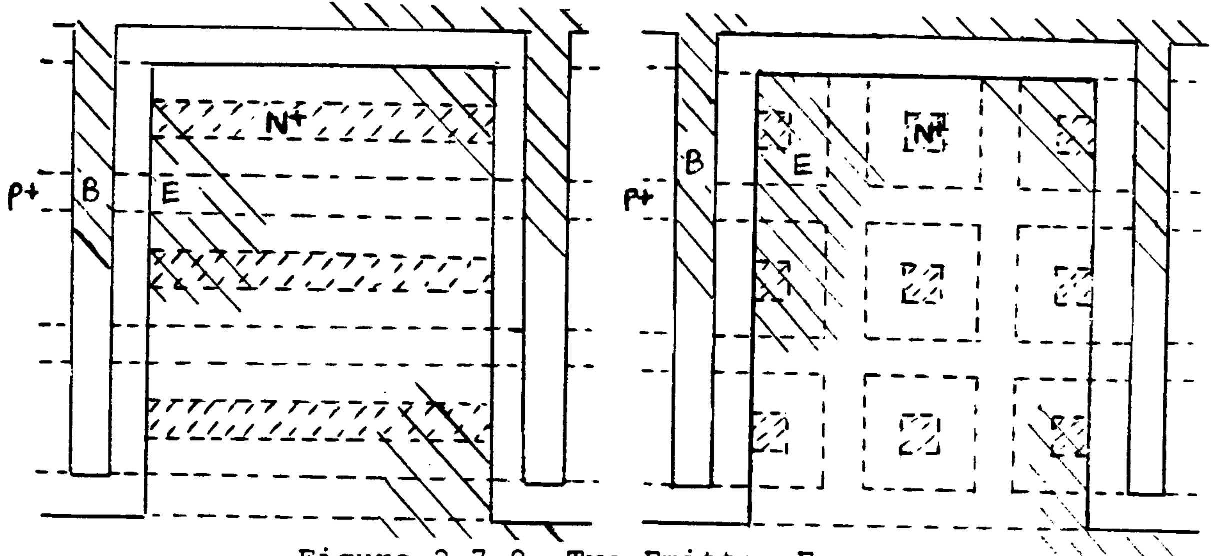


Figure 2.7.9. Two Emitter Forms

The short emitter has a pitch given simply by $P_L = P$ and the long emitter pitch is given by $P_L = P - W_e + L$, both relationships hold for washed or recut emitters. On average these pitches may be slightly larger due to any restriction imposed by metal to metal spacing between the base and emitter metal fingers.

$$\text{Total area of base diffusion} = (KP_L + 2L_b)(JP + 2L_b)$$

$$\text{Total emitter periphery} = 2N(L + 2W_d - W_e)$$

or $4NW_d$

for the minimum length emitter where $L = W_e$

Total emitter area $= NW_d(L + W_d - W_e)$

or NW_d for $L = W_e$

$$\frac{\text{Emitter Periphery}}{\text{Base area}} = 2L/PP_L \quad \left| \begin{array}{l} L \gg 2W_d - W_e \\ KP_L \gg 2L_b \\ JP \gg 2L_b \end{array} \right.$$

for the long emitter and for the short emitter is;

$$\frac{\text{Emitter Periphery}}{\text{Base area}} = 4W_d/P \quad \left| \begin{array}{l} KP_L \gg 2L_b \\ JP \gg 2L_b \end{array} \right.$$

These ratios may be compared to that for the interdigitated transistor, 2/P, and we see that for the same pitch, P, the long emitter overlay is only L/P_L as good while the short emitter overlay is $2W_d/P$ as good. Both factors are less than unity, in the first case L/P_L approaches unity as L increases, although this is limited by an increasing $r_{bb'}$, and in the second case $2W_d/P$ approaches unity as S_b, the base sideways diffusion, and M, the misalignment tolerance, reduce. In addition to a reduced emitter periphery to base area ratio the overlay transistor has a larger, than the interdigitated transistor, base diffusion to emitter metal capacitance.

Both factors imply, and practice bears out, that the conventional overlay structure cannot attain as high an f_{MAX} as an interdigitated transistor made with a similar process.

The mesh transistor is similar to the minimum length emitter overlay transistor, with the emitter and base diffusions exchanged, Figure 2.7.10. It can be seen that there are many similarities with the previous structures. The pitch, P, is governed by metal to metal separation in the same fashion as with the interdigitated transistor and

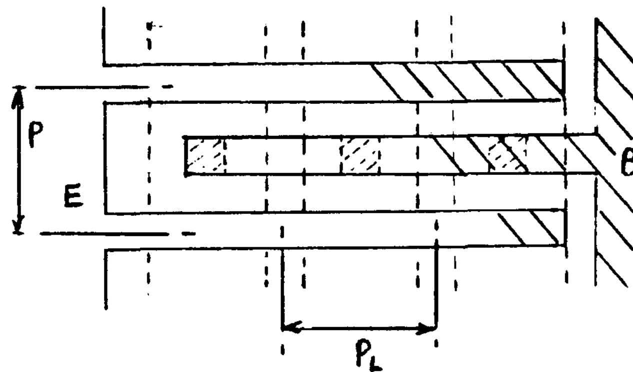


Figure 2.7.10. The Mesh Transistor

the pitch is the same.

$$P = W_e + W_b + 2c + 4a + 4M$$

For a recut emitter the pitch, P_L, is not dependent on the metal layer and is determined by the diffusions in the same fashion as the overlay structure giving;

$$P_L = W_e + W_b + 2S_b + 2_e + 2d + 4M$$

as for the washed emitter, since the emitter does not have a contact window below the base metal.

Emitter width under the contact $W_d = W_e + 2M + 2S_e$

while that between the contacts $W_{dL} = W_e + 2S_e$

The emitter periphery for each base contact is, therefore;

$$2(P + P_L - W_d - W_{dL})$$

and;

Total area of the base $= ((1 + J)P + W_b)((1 + K)P_L + W_b)$

where $P + W_b$ and $P_L + W_b$ are the overlap of the emitter mesh in the J & K directions respectively.

$$\text{Total emitter periphery} = 2KJ(P + P_L - W_d - W_{dL}) + 2(KP_L + W_{dL} + JP + W_d)$$

where the second term is the periphery of the outside of the emitter mesh.

$$\text{Total emitter area} = (JP + W_d)(KP_L + W_{dL}) - JK(P - W_d)(P_L - W_{dL})$$

$$\frac{\text{Emitter periphery}}{\text{Base area}} = \frac{2(P + P_L - W_d - W_{dL})}{PP_L} \Bigg|_{\substack{K \gg 1 \\ J \gg 1}}$$

2.8 Fabrication Techniques

Conventional overlay transistor structures require six masks for fabrication, these are;

I Base well

II Base grid

III Emitter

IV Emitter and base contacts

V Metal

VI Passivation

The transistors are fabricated on a silicon wafer by means of photoengraving. Typically the starting material for an n-p-n transistor is n+ with a thin, 3 to 15μm, epitaxial layer of n material. Processing proceeds by first

oxidizing the wafer by furnacing in an oxygenated ambient. This has two major purposes: to form a barrier to impurities and to remove the top surface of the silicon along with any impurities and crystal damage. After furnacing a photoresist is deposited, thickness must be large to reduce defects and small to allow adequate resolution to be obtained. Thickness of less than 0.5µm is usually required which can be achieved with good uniformity by means of spinning the wafer after the wafer surface has been flooded with photoresist. A low temperature bake is then used to dry the photoresist.

The first mask is used to expose the photoresist, either by contact printing or proximity printing, and the photoresist developed to obtain an etch resistant mask. An oxide etch is then used to pattern the oxide as defined by the photoresist. Once the photoresist has been removed the base well can be diffused in, this occurring only in the areas not covered by oxide. The surface is oxidized during the drive in, leaving the wafer prepared for the subsequent photoengraving of the base grid. Alignment of the base grid to the base well is obtained by use of alignment marks, which are included on all layers, and a mask aligning machine. Modern machines can achieve registration accuracy of about 0.5µm. Once the base grid has been diffused in, the wafer is once more oxidized in readiness for the emitter photoengraving and subsequent diffusion. The possible misregistration of the emitter

within the base grid must be allowed for in the design, making the pitch between emitters larger.

Contact must now be made to the emitter and base grid. This is achieved by photoengraving contact holes in he oxide above the base and emitter and then depositing a metallization layer, which is then photoengraved to form base and emitter contacts. Depending on device application and intended package, a passivation layer may be used. This prevents impurity from penetrating the silicon during operation at elevated temperature, although hermetically sealed packages can also prevent this. In the event that passivation is used, the areas over the bond-pads must be opened, to allow bonding, by means of a photoengraving using the sixth mask.

A number of variants of this basic fabrication technique are possible. The base well and base grid may often be processed in the reverse order, depending on the required depth of the grid, usually the deepest diffusion is done first. Two examples of this are the early 2N3375 and 2N3866 transistors [135] the former of these uses small square emitters, Figure 2.8.1, and the base is diffused before the base grid. The second transistor uses longer emitters, Figure 2.8.2, the base grid is diffused before the base well, allowing a slightly deeper diffusion of the grid. Modern transistor structures often make use of ion implantation to produce repeatable shallow junctions.

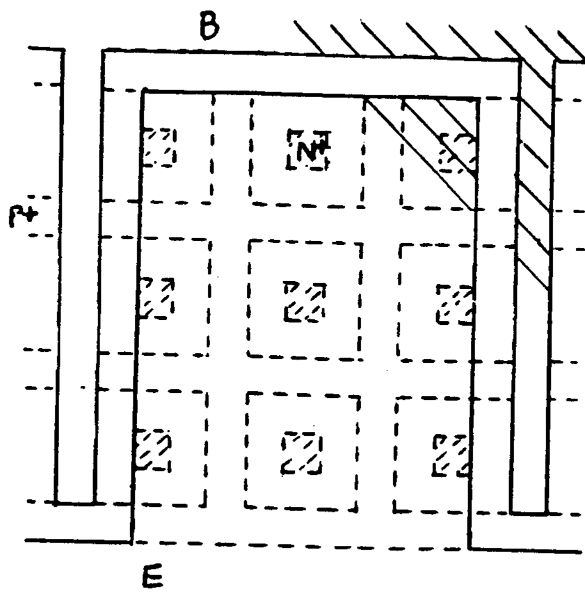


Figure 2.8.1. 2N3375 Geometry

These are difficult to achieve by diffusion due to the short furnacing times required. In addition masking can be achieved by the photoresist alone without the need for an underlying oxide. Insulators other than oxide are available and may be deposited in a variety of ways: spun on, sputtered and chemical vapour deposition (CVD) for example.

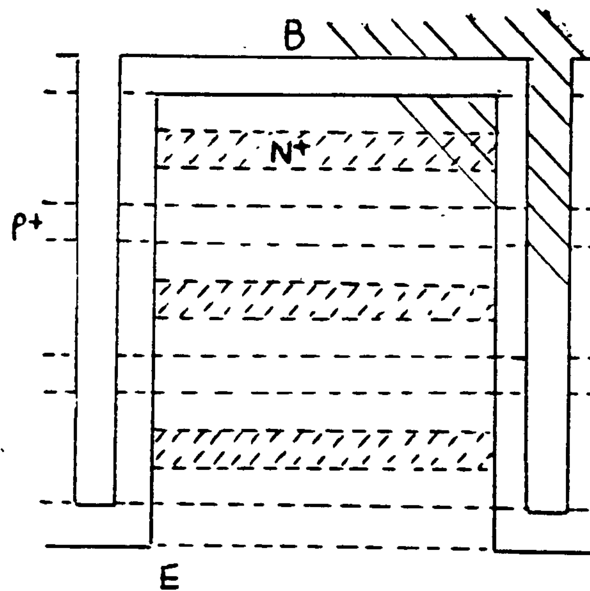


Figure 2.8.2. 2N3866 Geometry

An important variant of the basic structure is that of the washed emitter, as opposed to the recut emitter. Inspection of the basic geometry will show that the alignment of the emitter to the base grid and contact to emitter are both critical and affect the minimum emitter pitch attainable. Metal to contact alignment is semi-critical while base well to base grid and passivation to metal are non-critical. It is, therefore, evident that the relationship between the base grid, emitter and contact layers are important, Figure 2.8.3, shows such a situation with the effect of misalignments shown.

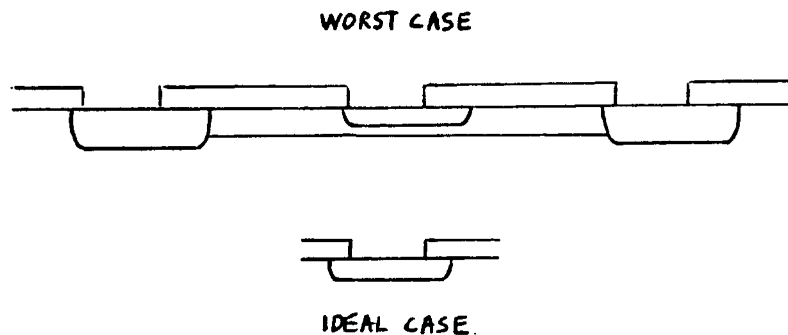


Figure 2.8.3. Effect of Misalignment

In this case both contact and emitter masks are aligned with respect to the grid and in consequence the emitter must be large enough to accommodate two misalignment tolerances in each direction. Now this situation is obviously improved for the emitter if the contact is aligned with respect to the emitter, in this case only one misalignment tolerance in each direction need be

accommodated by the emitter, although two tolerances must be accommodated by the base grid.

A further improvement is possible by making use of the surface topography, Figure 2.8.4, prevailing after the emitter fabrication.

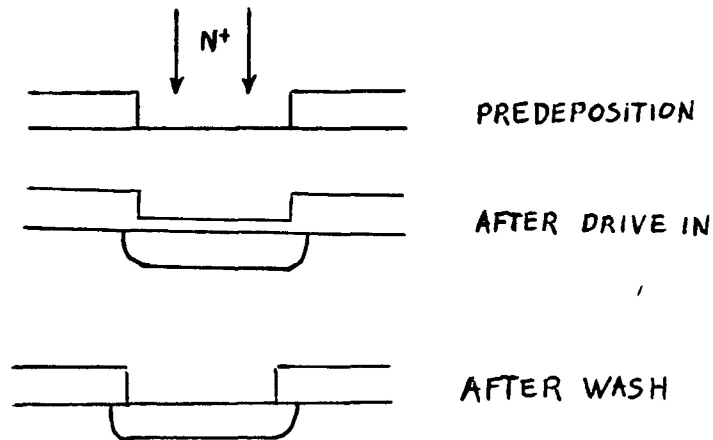


Figure 2.8.4. The Washed Emitter

The oxide covering the emitter is much thinner than elsewhere, as a result the window may be opened without the use of a mask by a short duration etch, to wash the emitter. The recut and washed emitter geometries are compared in Figure 2.8.5, where it appears that the washed emitter geometry is smaller, however the base and emitter contact windows are not in a fixed relationship to each other and this impinges on the metallization mask alignment, which then becomes critical. The dimension X_{min}, Figure 2.8.5 must be at least twice the misalignment tolerance, 2M, plus twice any overlap of metal over contact that may be required.

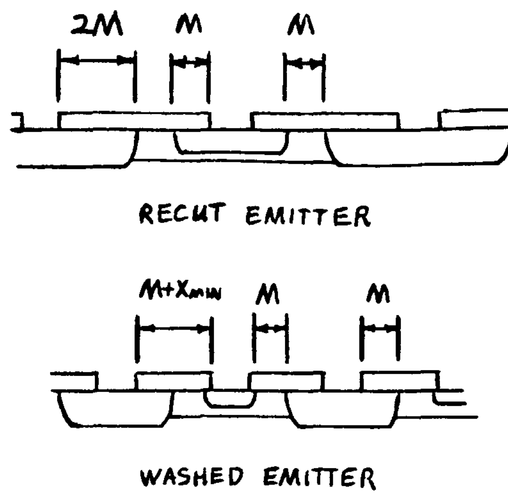


Figure 2.8.5. Recut and Washed Emitter Comparison

The minimum sizes obtainable with recut and washed emitters depends on the details of the particular process and the required width of the base grid, i.e. a wide base grid allows advantage to be taken of the washed emitter, and smaller geometry to be achieved.

3.1 Transistor Models

This chapter will cover modelling of bipolar transistors with a view to analysing distortion performance. In order to form a basis for a non-linear model, the well known small signal linearized models are reviewed and then the large signal non-linear Ebers-Moll model. The integral charge model is then discussed, with particular attention as to how the most important of the second order effects are implemented.

Package and extrinsic components are identified and the modelling of the extrinsic base resistance investigated.

The main contributions to non-linear distortion within the transistor are identified and discussed.

3.2 Small Signal Model

Ideally, for low distortion, large signal transistor operation should be linear. Real transistors, however, are fundamentally non-linear and only show near linear operation under small signal conditions, unless embedded in a linearizing network. In addition to this, real transistors also show deviations from the idealized model of transistor action, i.e. ideal diode law and constant

current gain. In terms of linear transistor design, the objective is to obtain a device which has minimal deviations from a linear action. Ultra-linear transistors therefore may be considered as devices with small deviations from the ideal model. With this in mind, a good starting point to model non-linear action is seen to be a linear model. From this model each non-linearity can then be analysed in isolation. The rationale being that, since each non-linearity is small any interaction between effects will be negligible. Transistors under consideration in this thesis are normally used in the common emitter configuration, in order to achieve the largest power gain, which makes a hybrid model most appropriate due to the low input and high output impedances. A common linear model, which presents a suitable starting point, is the small signal linearized h-parameter model, Figure 3.2.1.

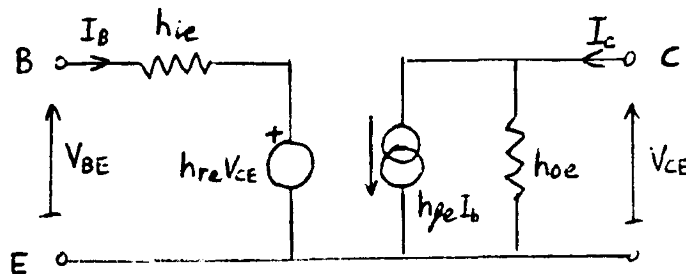


Figure 3.2.1. h-Parameter Model

This model, which uses two dependent generators, can be

rearranged as the T-model which uses only one dependent generator, Figure 3.2.2.

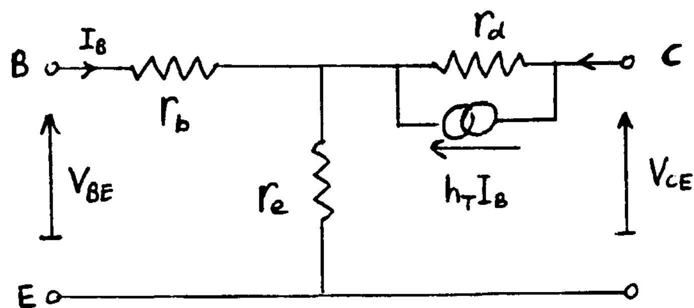


Figure 3.2.2. T-Model

The T-model is to some extent related to the physical operation of the device but is not easily modified to model the excess phase effect [169]. This is an effect seen in graded base transistors where the built in field increases the beta cut-off frequency, f_β, by more than the accompanying change in the phase response would suggest. The beta roll-off then no longer conforms to a dominant pole response but shows an excess phase at higher frequencies i.e. for $f > 2f_\beta$. This effect is not easy to implement in the T-model. The current controlled current source h_T is driven by I_b and the required phase shift, to model excess phase, cannot be modelled in a simple fashion. A model which is more suitable is the hybrid-Π Figure 3.2.3. The dependent gm generator of the hybrid-Π model is controlled by $V_{b'e}$, the phase of which can be modified with respect to I_b by means of a reactive

network [169].

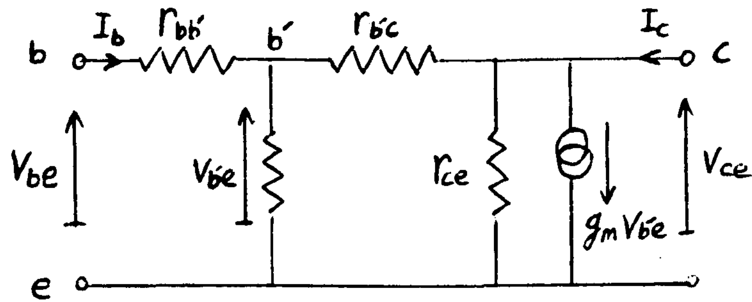


Figure 3.2.3. Hybrid-Π Model

The hybrid-Π model uses one more component than is mathematically needed for a h-parameter model, but does more accurately resemble the physical device. The resistor $r_{bb'}$ represents the base spreading resistance or the extrinsic base resistance. The b node of the hybrid-Π model represents the extrinsic base i.e. the base electrode, which can be contacted, and b' the intrinsic base which cannot be accessed. Nodes e and c represent the intrinsic emitter and collector respectively.

The current gain of the transistor can be represented as either a common base or common emitter h-parameter, where h_{fb} the common base forward gain is defined by;

$$h_{fb} = \left.\frac{\delta I_c}{\delta I_e}\right|_{V_{b'c} = 0} \approx \frac{i_c}{i_e} = -\alpha$$

and is commonly represented by α, which is defined here as a positive quantity. The common emitter forward current gain, h_{fe}, is defined by;

$$h_{fe} = \left.\frac{\delta I_c}{\delta I_b}\right|_{V_{b'c} = 0} \approx \frac{i_c}{i_b} = \beta$$

and is commonly represented by β. These two parameters are related by;

$$\alpha = \frac{\beta}{1 + \beta}$$

The T-parameters are obtained from the h-parameters by the four equations

$$h_T \approx h_{fe}$$

$$r_d \approx \frac{1}{h_{oe}}$$

$$r_e = \frac{h_{re}}{h_{oe}}$$

$$r_b = \frac{|h|_e - h_{re}}{h_{oe}}$$

where;

$$|h|_e = h_{ie}h_{oe} - h_{fe}h_{re}$$

The hybrid-Π parameters are obtained from the h-parameters plus $r_{bb'}$ by the four equations;

$$g_m \approx \frac{h_{fe}}{h_{ie} - r_{bb'}}$$

$$r_{b'e} \approx h_{ie} - r_{bb'}$$

$$r_{b'c} = \frac{h_{ie} - r_{bb'}}{h_{re}}$$

$$\frac{1}{r_{ce}} = h_{oe} - \frac{h_{re}(1 + h_{fe})}{h_{ie} - r_{bb'}}$$

The transconductance, g_m, and $r_{b'e}$ are obtainable from h_{fe} and the incremental emitter resistance r_e by noting;

$$g_m = \left.\frac{\delta I_c}{\delta V_{b'e}}\right|_{V_{b'c} = 0} \approx \alpha\cdot\frac{\delta I_e}{\delta V_{b'e}}$$

which gives;

$$g_m \approx \frac{\alpha}{r_e}$$

also;

$$i_c = g_m v_{b'e}$$

$$\approx g_m r_{b'e} i_b$$

since;

$$r_{b'c} \gg r_{b'e}$$

which gives;

$$r_{b'e} \approx \frac{\beta}{g_m} \approx (1 + \beta) r_e$$

Thus;

$$g_m \approx \frac{h_{fe}}{(1 + h_{fe}) r_e}$$

and;

$$r_{b'e} \approx (1 + h_{fe}) r_e$$

where;

$$r_e \approx \frac{V_T}{|I_e|}$$

and V_T the thermal voltage is given by KT/q.

At high emitter current densities, when the transistor is operating in the high injection region, that is the excess carrier concentration in the base is of similar magnitude to the doping concentration, the value of r_e may approach twice the value expected [159].

Typical values for a mW device working at room temperature with a 1mA emitter current are;

$g_m \approx 40m\Omega^{-1}$

$r_{b'e} \approx 2.5k\Omega$ for $\beta = 100$

$r_{bb'} \sim 250\Omega$

$r_{b'c} \sim 2.5M\Omega$

$r_{ce} \sim 50k\Omega$

For a medium power device working at 100mA the values will typically be;

$g_m \approx 4\Omega^{-1}$

$r_{b'e} \approx 25\Omega$ for $\beta = 100$

$r_{bb'} \sim 2.5\Omega$

Transconductance and base-emitter input impedance reduce as a direct result of the increase in bias current, while the extrinsic base resistance decreases due to the larger emitter periphery that is required to handle the increased current i.e. equivalent to using several transistors in parallel.

At higher frequencies the diffusion mechanism must be examined in more detail. Equations suggest a lossy transmission line model [173] which, although accurate, is unwieldy. The hybrid-Π model continues to give good accuracy with simplicity of modelling at higher frequencies. The model can be extended by the addition of intrinsic junction and diffusion capacitances in parallel with $r_{b'e}$ and $r_{b'c}$. The reactive network formed by these capacitances, which are assumed independent of frequency, model the frequency dependence of the device. In fact the component values can vary at higher frequency due to modifications of the carrier distributions, for example the emitter's injection into the base. Under these conditions it is still usually adequate to assume frequency independence of components measured at or near the operating frequency [169].

3.3 Large Signal Model

A suitable large signal model is the Ebers-Moll model, Figure 3.3.1. Basically the intrinsic transistor is

modelled as two junction diodes, with their interaction, via the base, modelled by two current generators. The junction capacitances can, of course, be included directly.

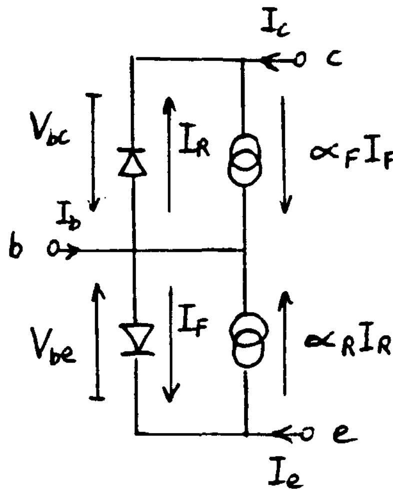


Figure 3.3.1. Ebers-Moll Model (Injection Version)

Here the reference currents are;

$$I_F = I_{es} \exp(V_{be}/V_T - 1)$$

and;

$$I_R = I_{cs} \exp(V_{be}/V_T - 1)$$

Where I_{es} and I_{cs} are the emitter and collector saturation currents respectively.

The terminal currents are;

$$I_C = \alpha_F I_F - I_R$$

$$I_e = \alpha_F I_R - I_F$$

$$I_b = (1 - \alpha_F) I_F + (1 - \alpha_R) I_R$$

and the betas are defined by;

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F}$$

and;

$$\beta_R = \frac{\alpha_R}{1 - \alpha_R}$$

The two reference currents, I_{es} and I_{cs}, and the two large signal common base current gains, α_F and α_R, are related by the reciprocity relation [158, 159].

$$\alpha_F I_{es} = \alpha_R I_{cs} = I_s$$

Where I_s is the transistor saturation current.

Now all the currents can be expressed in terms of I_s by putting;

$$I_{cc} = \alpha_F I_F$$

and;

$$I_{ec} = \alpha_R I_R$$

so that;

$$I_{cc} = I_s \exp(V_{be}/V_T - 1)$$

$$I_{ec} = I_s \exp(V_{bc}/V_T - 1)$$

where I_{cc} is the collected part of the emitter current. Similarly I_{ec} is that part of the collector current collected by the emitter. This can be considered as the transport model, Figure 3.3.2, as the reference currents are those transported across the base.

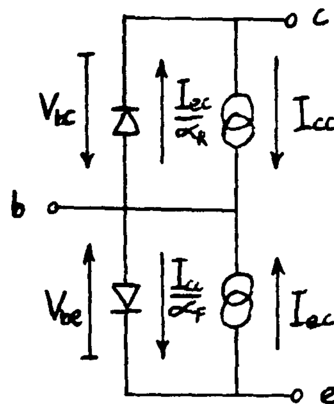


Figure 3.3.2. Ebers-Moll Model (Transport Version)

The Ebers and Moll model, which has two current sources, can be reconfigured to obtain a large signal non-linear hybrid-Π model, Figure 3.3.3, which uses a single current source. Here $I_{cc} - I_{ec}$ represents the common component of emitter and collector current. While I_{ec}/β_R represents

the common component of base and collector current. Similarly I_{cc}/β_F represents that part of the base current that also flows in the emitter.

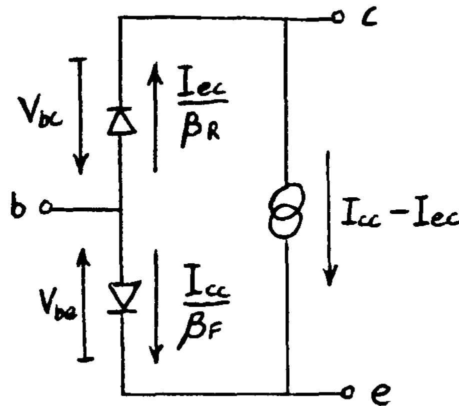


Figure 3.3.3. Large Signal Non-linear Hybrid-Π Model

The terminal currents are given by;

$$I_b = I_{ec}/\beta_R + I_{cc}/\beta_F$$

$$I_e = I_{ec} - I_{cc} - I_{cc}/\beta_R$$

$$I_c = I_{cc} - I_{ec} - I_{ec}/\beta_R$$

It should be noted that this form of the model is directly reducible to the small signal linear hybrid-Π model, see Figure 3.2.3.

At this stage it is apparent that the model can be extended to include the extrinsic resistances and the two sets of junction and diffusion capacitances along with any

parasitic components, Figure 3.3.4.

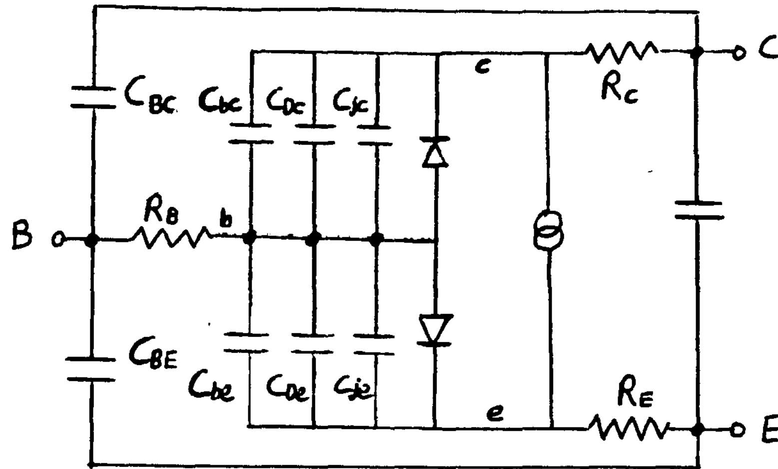


Figure 3.3.4. Non-Linear Model with Extrinsic Components

Junction capacitances C_{jc} and C_{je} can be approximated by;

$$C_{je} = \frac{C_{jeo}}{(1 - V_{be}/\phi_e)^{1/me}}$$

and;

$$C_{jc} = \frac{C_{jeo}}{(1 - V_{bc}/\phi_c)^{1/mc}}$$

where C_{jeo} and C_{jco} are the emitter and collector capacitances with $V_{be} = 0$ and $V_{bc} = 0$ respectively.

These expressions have singularities at $V_{be} = \phi_e$ and $V_{bc} = \phi_c$ respectively, although this is purely a function of the model since the capacitance does not increase without limit when the junction voltage approaches the

built in potential. This detail of the model is generally not important since the diffusion capacitance is dominant when the junction is forward biased. For linear operation the transistor is biased in the forward active region and must not enter the saturation region. Under these conditions $V_{bc} < \phi_c$, since typically $\phi_c > V_{be} - V_{cesat}$, and the emitter capacitance is dominated by the diffusion capacitance.

Diffusion capacitances are due to the mobile charges associated with the various currents. In the case of the reference current, I_{cc}, the charge is in four main regions: emitter and base neutral regions and the emitter and collector space charge regions. Charge can be expressed as a product of delay time and current giving;

$$Q_{De} = (\tau_e + \tau_{ebscl} + \tau_b + \tau_{cbscl}) I_{cc}$$

$$= \tau_F I_{cc}$$

where τ_F is the forward transit time of the intrinsic transistor, and;

$$\tau_b = \frac{W_b^2}{2D_n} \qquad \text{homogeneous base}$$

$$\tau_{cbscl} = \frac{W_{cbscl}}{2V_{scl}} \qquad \text{triangular distribution}$$

in most cases, the collector-base scatter limited transit time, τ_{cbscl}, is negligible.

The reverse current, I_{ec}, can be expressed in a similar fashion but in this case the components are in the collector and base neutral regions and the two space charge regions, giving;

$$Q_{De} = (\tau_c + \tau_{cbscl} + \tau_{bR} + \tau_{ebscl}) I_{ec}$$

$$= \tau_R I_{ec}$$

The capacitors can then be defined as;

$$C_{De} = \frac{dQ_{De}}{dV_{be}} = \frac{\tau_F dI_{cc}}{dV_{be}}$$

$$C_{Dc} = \frac{dQ_{Dc}}{dV_{bc}} = \frac{\tau_R dI_{ec}}{dV_{bc}}$$

When the transistor is in the forward active region the reverse current is several orders of magnitude below that of the forward current and is therefore negligible.

3.4 Charge Control Model

The original charge control concept was proposed by Beaufoy and Sparkes in 1957. The basic idea is that

control charge divided by transit time equals controlled current, $I_c = Q_b/\tau$. This idea was extended by Gummel [121] in 1969 to include the effects of the junction bias voltages. The basic equation relating the variables for an n-p-n transistor is;

$$I_c = C \frac{\exp(V_{be}/V_T) - \exp(V_{bc}/V_T)}{Q_G}$$

where;

Q_G is normalized mobile charge (holes) associated with the base terminal, the Gummel number.

$$C = \frac{(qn_i A)^2 D_0}{a_{av}}$$

D_0 is the low field diffusivity

a_{av} is the average ratio of electron current to collector current

A is the cross-sectional area of the transistor

Assumptions made for this derivation are;

i) There is negligible avalanche multiplication

ii) The Einstein relation holds, $D = V_T \mu$

iii) An idealized field dependence of the mobility which is independent of the doping concentration

iv) One dimensional current flow and a number of simplifications in the mathematics.

With C and Q_b held constant, this basic equation relates the idealized collector current to the junction voltages. The bias dependence of Q_b and C model the high level effects. Variation of C with bias is limited to changes in a_{av}, which for large h_{fe} is almost unity.

Further refinement of the charge control model has produced a model of sufficient accuracy to allow calculation of the intermodulation distortion [38]. Base charge is modelled by;

$$Q_b = Q_{bo} + Q_e + Q_c + B\tau_f I_f + \tau_r I_r$$

where;

Q_{bo} is the equilibrium charge stored in the base

Q_e is the capacitively stored charge in the emitter junction [159]

Q_c is the capacitively stored charge in the collector junction

B describes the base push-out

$B\tau_f I_f$ represents the forward diffusion capacitance stored charge

$\tau_r I_r$ represents the reverse diffusion capacitance stored charge

I_f is the component of the collector current due to V_{be}

I_r is the component of the collector current due to V_{bc}

This type of model is available in SPICE, the circuit simulation program. A number of high level effects are modelled by Q_b, Early effect by Q_c, the equivalent effect in the emitter by Q_e. Base push-out and high level injection is modelled by the forward diffusion capacitance. The main advantage of this model is the unified manner in which base-width modulation, transit time and current gain are handled. Additional second order effects are modelled by the addition of non-ideal diodes and a current dependent base resistance.

The actual model used by SPICE and the way that it is implemented varies according to the version number. In some versions, which use the Ebers-Moll model, the base push-out mechanism is assumed to be predominantly that proposed by van der Ziel and Agouridis. Base widening, B, is then described [159] by

$$B = 1 + \frac{L_E^2}{4W_B^2}\left[\frac{I_C^2}{3I_{CO}^2} - \frac{I_C}{I_{CO}} + 1 - \frac{I_{CO}}{3I_C}\right]$$

for $I_C \geq I_{CO}$ and;

$$B = 1$$

for $I_C \leq I_{CO}$

The system of equations used in version 2G.0 is as follows [175].

DC Model

$$I_C = \frac{I_S}{Q_B}(\exp(V_{be}/(N_F V_T)) - \exp(V_{bc}/(N_R V_T)))$$

$$- \frac{I_S}{\beta_R}(\exp(V_{bc}/(N_R V_T))-1) - I_{SC}(\exp(V_{bc}/(N_C V_T))-1)$$

$$I_B = \frac{I_S}{\beta_F}(\exp(V_{be}/(N_F V_T))-1) + \frac{I_S}{\beta_R}(\exp(V_{bc}/(N_R V_T))-1)$$

$$+ I_{SE}(\exp(V_{be}/(N_E V_T))-1) + I_{SC}(\exp(V_{bc}/(N_C V_T))-1)$$

Where the last two terms in the equation for I_B model recombination in the emitter and collector transition regions. In the equation for I_C the base charge is normalized by dividing by Q_{BO} and putting $C/Q_{BO} = I_S$.

$$Q_B = \frac{Q_1}{2}(1 + (1 + 4Q_2)^{1/2})$$

where;

$$Q_1 = \frac{1}{1 - V_{bc}/V_{AF} - V_{be}/V_{AR}}$$

$$Q_2 = \frac{I_S}{I_{KF}}(\exp(V_{be}/(N_F V_T))-1) + \frac{I_S}{I_{KR}}(\exp(V_{bc}/(N_R V_1))-1)$$

where;

V_{AF} and V_{AR} are the forward and reverse Early voltages

I_{KF} and I_{KR} determine the current gain fall-off at high currents

I_{SE}, N_E, I_{SC} and N_C determine the current gain fall-off at low currents.

The base resistance, which is current dependent, is given by;

$$R_{bb'} = R_{BM} + \frac{R_B - R_{BM}}{Q_B}$$

AC Model

$$C_{be} = \frac{d}{dV_{be}}\left[\frac{I_S}{\tau_{FF} Q_B}(\exp(V_{be}/(N_F V_T))-1)\right] + \frac{C_{je}}{(1 - V_{be}/V_{je})^{mje}}$$

$$C_{bc} = \frac{\tau_R I_S}{N_R V_T}\exp(V_{bc}/(N_R V_T)) + \frac{C_{jc}}{(1-V_{bc}/V_{jc})^{mjc}}$$

where;

$$\tau_{FF} = \tau_F\left[1 + X_{TF}\left[\frac{I_F}{I_F + I_{TF}}\right]^2 \exp(V_{bc}/1.44V_{TF})\right]$$

and;

$$I_F = I_S(\exp(V_{be}/(N_F V_T)) - 1)$$

In this system of equations the base push-out is modelled as a function of the forward current and the collector-base voltage by X_{TF}, I_{TF} and V_{TF}.

The collector-base capacitance is shared between the internal and external base terminals by

$$C_{B1} = C_{BC}(1 - X_{CJC})$$

$$C_{B2} = C_{BC}X_{CJC}$$

Where X_{CJC} is a factor between zero and one.

3.5 Base Push-Out

Kirk [26] proposed a one dimensional model to describe the gain fall-off at higher current levels. The large number of carriers in the collector overwhelm the collector doping density and form a current induced base. This effective increase in the base-width at higher current produces the fall-off of current gain and gain-bandwidth product.

Van der Ziel and Agouridis [47] proposed a two-dimensional model where a certain critical current density cannot be exceeded in the collector. Any additional current must be collected over a larger collector area, necessitating the extra carriers to be injected from the emitter edge and

travel a diagonal path across the base. This diagonal path is an effective increase in the base-width.

Whittier and Tremere [26] in a tour de force have developed these ideas in a unified manner. Figure 3.5.1 shows a general epitaxial structure, while Figures 3.5.2 and 3.5.3, show the one dimensional and two dimensional models respectively.

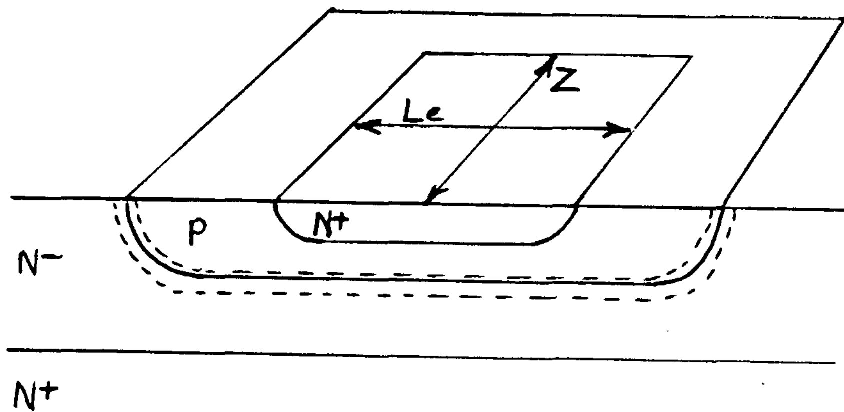


Figure 3.5.1. General Epitaxial Structure

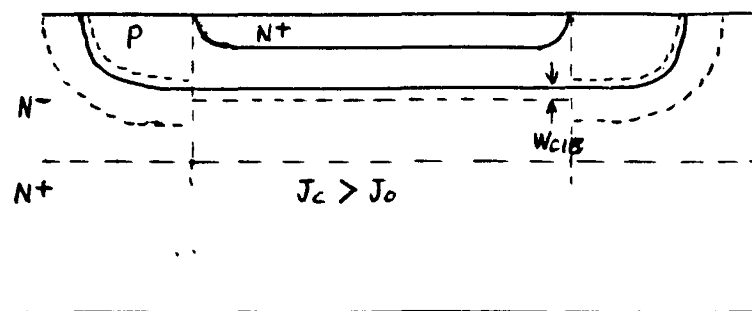


Figure 3.5.2. One Dimensional Model

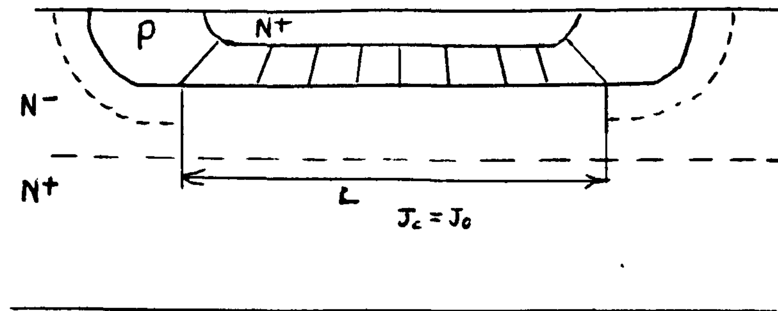


Figure 3.5.3. Two Dimensional Model

The critical current density for the space charge limited conditions is;

$$J_o = qV_{lim}\left[\frac{2\varepsilon_r\varepsilon_o|V_{cb}|}{qW_{EPI}^2} + N_{Dc}\right]$$

Where V_{lim} is the scatter limiting velocity and N_{Dc} is the donor concentration in the epitaxial layer. This equation is obtained from Poisson's equation and the assumption that V_{lim} is the average carrier velocity in the collector transition region.

For the one dimensional model the expression describing the charge induced base (W_{CIB}) is;

$$W_{CIB} = (1 - I_o/I_c)W_{EPI}$$

For the low field asymptote where conduction in the collector is ohmic. Under high field conditions the carriers are moving at a saturated drift velocity at the onset of space charge limited current. The ensuing constant space charge density in the epitaxial region gives a triangular field distribution with a constant integral at a given V_{cb}. This leads to the high field expression;

$$W_{CIB} = \left[1 - \left[\frac{I_o - I_{cx}}{I_c - I_{cx}}\right]^{1/2}\right] W_{EPI}$$

where;

$$I_{cx} = qN_{DC}V_{lim}A_e$$

The expressions for the forward transit time and current gain for $I_c \geq I_o$ are then;

$$\tau_F = \frac{V_T C_{Te}}{I_c} + \frac{W_{BO}^2}{nD_{nb}} + \frac{W_{CIB}^2}{4D_{nc}} + \frac{W_{EPI} - W_{CIB}}{2V_D}$$

$$h_{fe} \approx \left[\frac{Q_B D_{PE}}{Q_E D_{NB}} + \frac{W_{BO}^2}{nL_{nB}^2} + \frac{W_{CIB}^2}{4L_{nC}^2}\right]^{-1}$$

Where n is a factor accounting for the doping distribution

in the base and V_D is the average drift velocity in the collector epitaxial region.

The two dimensional model produces two similar equations when $I_C \geq I_0$;

$$\tau_F = \frac{V_T C_{TE}}{I_C} + \frac{W_{B0}^2}{nD_{nB}} + \frac{W_{LB}^2}{4nD_{nB}} + \frac{W_{EPI}}{2V_D}$$

$$h_{fe} \approx \left[\frac{Q_B D_{pE}}{Q_E D_{nB}} + \frac{W_{B0}^2}{nL_{nB}^2} + \frac{W_{LB}^2}{12nL_{nB}^2} \right]^{-1}$$

Where the lateral base spread at the collector junction, W_{LB}, is given by;

$$W_{LB} = (I_C/I_0 - 1)L_E$$

Both models describe an effective variation on base-width which occurs when $I_C > I_0$ whilst for $I_C \leq I_0$ the base-width is constant. The expressions are both continuous at $I_C = I_0$ but are not analytic at this point i.e. there is a discontinuity in the slope of the W_B - I_C curve. It is clear from any f_T - I_C curve that there is not an abrupt change in the derivative of base-width with current although the approximation may be very close. From the point of view of the f_T curve this is not a problem but it is a problem for distortion analysis since the abrupt

change in the derivatives prevent analytic solutions. This discontinuity is also a problem in simulation programs such as SPICE where it can cause computational problems in converging to an iterated solution. Some versions of SPICE assume that the van der Ziel and Agouridis effect is dominant [159] and models the part of the forward transit time due to the diffusion capacitance stored charge, τ_f, by;

$$\tau_f = B\tau_{fl}$$

Where τ_{fl} is the low current value of τ_f and;

$$B = 1 + \frac{L_E^2}{4W_B^2}\left[\frac{I_C^2}{3I_{CO}^2} - \frac{I_C}{I_{CO}} + 1 - \frac{I_{CO}}{3I_C}\right]$$

for $I_C \geq I_{CO}$ or;

$$B = 1$$

for $I_C \leq I_{CO}$.

These expressions both have; $B = 1$, $dB/dI_C = 0$ and $d^2B/dI_C^2 = 1$ at $I_C = I_{CO}$, the critical current. This aids the convergence within SPICE since the program employs both the value and the first derivative at each point in the analysis. When the analysis is carried out in the

vicinity of a discontinuity in the derivative the program has to switch between two different regimes, and it is this abrupt change which can cause difficulties.

3.6 Collector Capacitance

For a transistor operating in class A, the collector junction is normally reverse biased. Collector-base capacitance is then dominated by the junction capacitance, while the diffusion capacitance is small by comparison. The one-sided junction approximation is good for the typical n+np+n+ epitaxial transistor structure. The equation describing this type of junction is well known [159];

$$C_j(V_{bc}) = \frac{C_j(0)}{(1 - V_{bc}/\phi)^m}$$

where;

ϕ is the junction barrier potential

m is the capacitance gradient factor

Typically ϕ is 0.7V and m is 1/2 for a step junction or 1/3 for a linearly graded junction.

Transistors built in n or n+ substrates use the n material

as a collector. The lower doping density of the n material is required to achieve a suitable collector-base breakdown voltage, while the highly doped n+ substrate reduces the collector series resistance. Under normal bias conditions the depletion region, due to the collector junction, extends predominantly into the n region since the p+ base region has a much greater doping density. The junction capacitance is in inverse proportion to the width of this depletion region if a parallel plate model is assumed. Any increase in the collector voltage causes an increase in the depletion width which may reach the n+ substrate if the epitaxial layer is sufficiently thin, the collector is then said to be fully depleted. Further increase in the collector voltage will cause the depletion region to extend into the n+ substrate. However, the penetration will be very small due to the high doping level. Under these conditions capacitance will vary more slowly with collector voltage, the corresponding capacitance gradient factor may then fall to 1/10 or less.

Transistors can be manufactured which will fully deplete with a few volts applied to the collector [29, 105]. This gives a reduction in m which allows an advantage to be gained in the distortion performance of the transistor. Variation of m with collector voltage is a function of the abruptness of the n - n+ junction.

3.7 Early Effect

Increasing the collector voltage of a normally biased transistor causes the collector depletion region to extend into both the collector and base regions. The actual width of the depletion region, which determines the collector-base capacitance, is dominated by the extension into the collector. The small extension into the more highly doped base has little effect on the capacitance but can be important compared to the base-width. A reduction in base-width generally results in an improvement in the base transport factor and a decrease in the base transit time. This is the Early effect and to first order it can be modelled by [159];

$$I_s(V_{bc}) = I_s(0)/(1 + V_{bc}/V_A)$$

$$h_{fe}(V_{bc}) = h_{fe}(0)/(1 + V_{bc}/V_A)$$

$$\tau_B(V_{bc}) = \tau_B(0)(1 + V_{bc}/V_A)^2$$

where the Early voltage V_A is defined by;

$$V_A^{-1} = \frac{1}{W(0)} \frac{dW}{dV_{bc}} \bigg|_{V_{bc} = 0}$$

For constant V_{be} the transistor exhibits an output resistance of $(V_A + V_{be})/I_c$. This is also true for a

constant base current since $I_B = I_s(\exp(V_{be}/V_T) - 1)/h_{fe}$ and is independent of V_{bc} to first order.

3.8 Emitter Capacitance

With the transistor used in a common emitter configuration the emitter base junction is normally forward biased. Variation of the base-emitter voltage causes a change in the minority carrier charge in the base region. This gives rise to a diffusion capacitance which is given by dQ/dV. Calculations for the diffusion transistor [166] yield an approximate expression of;

$$C_D = \frac{I_E W^2}{2 V_T D_n}$$

while for a drift transistor the diffusion capacitance is approximately;

$$C_D = \frac{I_E W}{V_T \mu_n \varepsilon_s}$$

where ε_s is the built in field due to the doping profile in the base.

In both cases the capacitance is proportional to the emitter current.

The junction capacitance follows the same form as that of the collector transition capacitance. However, the diffusion capacitance will normally dominate at higher emitter currents. Emitter delay is given by the product of the small signal emitter resistance, r_e, and the small signal emitter diffusion capacitance, c_e. The incremental emitter resistance, r_e, is approximately V_T/I_E or twice this value under high injection conditions and the emitter delay is;

$$r_e C_e = \frac{W^2}{2D_n}$$

for the diffusion transistor and;

$$r_e C_e = \frac{W}{\mu_n \varepsilon_s}$$

for the drift transistor.

3.9 Package and Extrinsic Components

The intrinsic transistor requires connections to its emitter, base and collector regions. At least part of these connections are by necessity within the silicon die and form parasitic components. A typical structure is shown in Figure 3.9.1, with some of the parasitic components indicated.

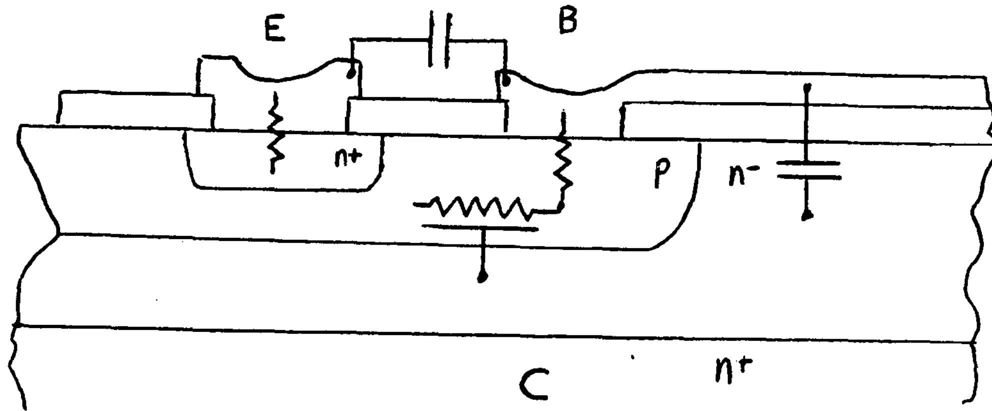


Figure 3.9.1. Typical Transistor Structure

Packaging of the device will normally add a small amount of series resistance and series inductance into each of the three connections and capacitances between each pair of terminals. The reactive components are more important at higher frequencies and can result in unacceptable losses at a few gigahertz, stripline packages are required at these frequencies and above.

Of the three terminals the collector is most tolerant of series resistance and the emitter is least tolerant. While the transistor is out of saturation the collector has a high output impedance and the collector resistance is of little consequence. Any IR drop, i.e. collector current multiplied by the collector load resistance, in the collector will, however, cause premature entry into the saturation region and compromise the saturation voltage. Resistance in the emitter or base leads has the

effect of reducing the transistor's available gain. The emitter is most critical due to the larger currents flowing, i.e. βI_B.

All of the components due to packaging tend to be linear and do not contribute directly to non-linearities produced within the transistor.

Some of the parasitic components due to the transistor structure can be non-linear. One important component is the base spreading resistance with its distributed non-linear junction capacitance. Base resistance decreases with current [24, 72] due to current crowding in the emitter and conductivity modulation in the base. Interdigitated transistors show the least variation in base resistance. Distributed capacitance associated with the base resistance is usually modelled as a lumped C-R network. SPICE, for example, allows the capacitance to be shared between the internal and external base nodes.

The base-emitter and base-collector junctions both appear at the top surface of the die where surface effects modify the characteristics. These effects are seen at low currents and can be modelled by non-ideal diodes between base-emitter and base-collector of the intrinsic transistor. MOS capacitors between metallizations and substrate can be non-linear due to any trapped charge. It has been reported [29] that the device performance can be

compromised by the oxide capacitance between the base metallization and the collector substrate. In this particular case a general improvement of about 10dB in distortion performance was reported when the oxide capacitance was replaced with junction capacitance. This being achieved by extending the base diffusion under the base contact.

3.10 The Extrinsic Base and Emitter

Both conventional interdigitated and the novel metal base overlay transistor have base and emitter contacts as shown in Figure 3.10.1.

Figure 3.10.1. Base and Emitter Contacts

Base current flows along the base finger through the metal-silicon contact and then through the extrinsic base to the adjacent emitters. Emitter current is seen to follow a similar pattern, along the emitter finger through the metal-silicon contact and then predominantly to the

emitter periphery. Calculations for the base and emitter series resistance can, therefore, be carried out in a similar manner. Current flow is assumed to be along the finger when in the metal and normal to the finger when in the interface or the silicon, which is quite reasonable for a metal sheet resistivity below one tenth of the diffusion's sheet resistivity. The discussion that follows is given for the base contact, but are also applicable to the emitter contact.

A cross-section of the base contact is shown in Figure 3.10.2.

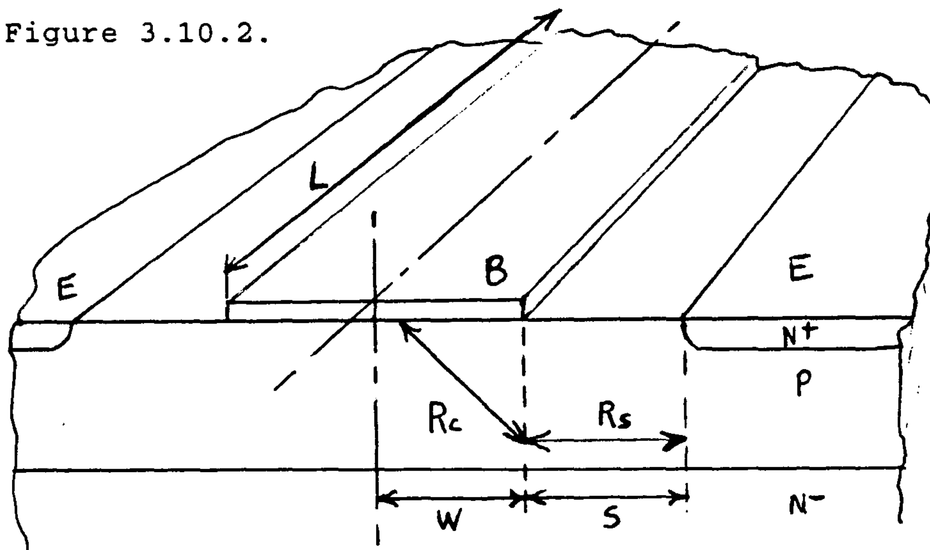


Figure 3.10.2. The Base Contact

Here the base resistance is simply;

$$R_B = R_C + R_S$$

where R_C is the contact resistance and R_S is the base spreading resistance. Only half of the base contact is

associated with the calculation of R_C due to symmetry, W is half of the base contact width.

The spreading resistance part is simply;

$$R_S = R_D S/L$$

where R_D is the sheet resistivity of the base diffusion, in Ω/square. The contact part can be evaluated [81], assuming a transmission line model and zero metal resistivity to give;

$$R_C = \frac{a}{L} R_D \coth(W/a)$$

where;

$a = (R_{CS}/R_D)^{1/2}$ is the attenuation constant (cm)

R_{CS} is the specific contact resistance of the metal-silicon interface (Ωcm^2).

The effect of non-zero metal resistance has been evaluated [114] for W/a >> 4, i.e. large contact size or small specific contact resistance, and gives;

$$R_C = \frac{a}{L} \cdot \frac{R_M 2m + R_D 2D + R_M R_D W/a}{Rm + RD} \Bigg|_{W/a \gg 4}$$

where;

$$a = (R_{CS}/(R_M + R_D))^{1/2} \qquad \text{(cm)}$$

and;

R_m is the metal sheet resistivity (Ω/square).

The devices discussed here have narrow contacts which can give W/a of less than one. Appendix I, shows how the calculus can be carried out without any assumption of W/a to give;

$$R_C = \frac{a}{L} \, \frac{(R_M^2 + R_D^2)\coth(W/a) + R_M R_D (2/\sinh(W/a) + W/a)}{R_M + R_D}$$

where;

$$a = (R_{CS}/(R_M + R_D))^{1/2} \qquad \text{(cm)}$$

Notice that putting $R_M = 0$ reduces this to the first equation, and making $W/a >> 4$ reduces it to the second equation. Various other conditions lead to additional simplifications of this equation, and are given in Appendix I.

The total resistance due to one base stripe and one emitter stripe can be calculated in a similar fashion,

consider an interdigitated transistor.

Figure 3.10.3 shows a single base-emitter pair, each stripe having only half of its width associated with the other. Notice also that current multiplication occurs at the emitter junction and all emitter resistances are multiplied by (1+β) to account for this. The emitter base junction is forward biased and the resistance seen looking into the base terminal is evaluated assuming the simplified current flow models previously discussed. All resistances are assumed ohmic, although the small signal emitter resistance is included.

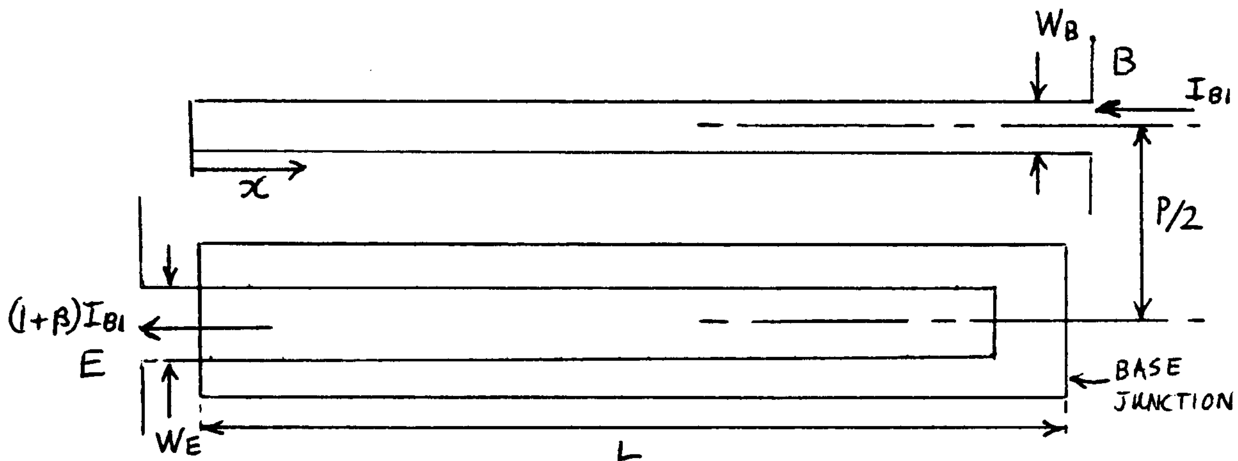


Figure 3.10.3 Total Resistance of One Stripe

The total base input resistance for one emitter-base pair is then given by substituting into the previous equation to give;

$$R_{B1} = \frac{a}{L} \cdot \frac{(D^2 + M^2)\cosh(n) + nDM\sinh(n) + 2DM}{(D + M)\sinh(n)}$$

where;

$M = 2R_{MB}L/W_B$ (Ω)

R_{MB} is the sheet resistivity of the base metallization (Ω/square)

$D = 2R_{ME}L(1 + \beta)/W_E$ (Ω)

R_{ME} is the sheet resistivity of the emitter metallization (Ω/square)

$n = L/a$

$a = (C/(M + D))^{1/2}$ the attenuation constant (cm)

$C = (R_B + (1 + \beta)R_E + V_T/I_{B1})L^2$ (Ω/cm^2)

$R_B = R_{CB} + R_{SB}$

$R_E = R_{CE} + R_{SE}$

R_{CB}, R_{CE} are the base, emitter contact resistances (calculated by application of the above equation)

R_{SB}, R_{SE} are the base, emitter spreading resistances

V_T/I_{B1} is the small signal emitter resistance seen from

the base

The base current in the base metal stripe is given, Appendix I, by;

$$I_B(x) = I_B \frac{D\sinh(x/a) + M\sinh(L/a) - M\sinh(L/a - x/a)}{(D + M)\sinh(L/a)}$$

and the emitter current, similarly by;

$$I_E(x) = (1+\beta) I_B \frac{D\sinh(L/a) - D\sinh(x/a) + M\sinh(L/a - x/a)}{(D + M)\sinh(L/a)}$$

To extend this analysis to an overlay transistor (metal base or conventional) consider:

a) The emitter metal forms a stripe of width L and R_{ME} will fall, maybe to zero.

b) The base contact may have a low resistance drive at each end, the analysis is then applicable to each half length finger.

Finally consider a conventional overlay transistor, Figure 3.10.4. The emitter is analysed in the same fashion as before, except the base contact is slightly different. There is no contact resistance to calculate since the base grid is co-planar with the extrinsic base.

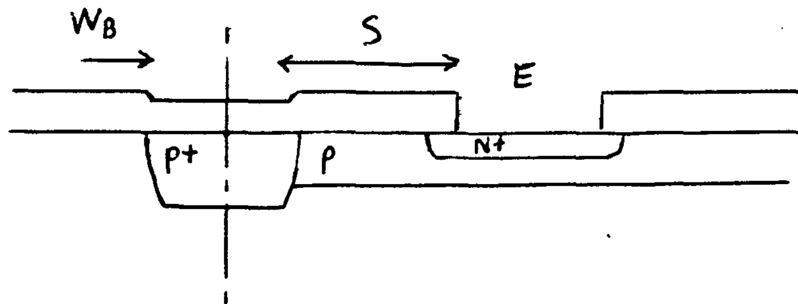


Figure 3.10.4 Section of Conventional Overlay Transistor

The previous equation is applied by setting;

$$R_{CB} = 0$$

and;

R_{MB} = sheet resistivity of the base grid.

In all cases the total resistance seen from the base terminal, R_{BT}, is given by;

$$R_{BT} = R_{B1}/N_R$$

where N_R is the number of repeats of the element analysed.

The base resistance, R_{BB}, can be estimated by setting the emitter resistance components, R_E and R_{ME} to zero and

using;

$$R_{BB} = R_{BT} - V_T/I_B$$

Notice that the term V_T/I_{B1} is not removed from the equation for R_{B1} since it has a very strong influence on the current distribution along the base finger. This point is discussed further in Appendix II.

CHAPTER 4 - DISTORTION

4.1 Review of the Literature

Up to the mid-1950s few papers were written concerning distortion in bipolar transistors. Most effort had been directed towards semiconductor theory by authors such as W. Shockley. L. J. Giacoletto published his classical paper in 1954 and laid the foundations of the hybrid-Π model, while W. M. Webster published on the variation of current gain with emitter current.

With the scene set the study of distortion started in earnest, N. I. Meyer being a notable author. Many of these authors were referenced by H. Lotsch [44] when he considered many types of distortion due to the exponential emitter-base junction. This he analysed by means of a Taylor expansion. Later, in 1968 he considered distortion in a diode and produced an extensive bibliography.

Until 1965, when J. Reynolds [46] produced results in terms of frequency dependent Y-parameters, published work remained confined to low frequency. It was at this time that S. Narayanan was presenting his PhD thesis. Two years later, in what appears to be an extension of this work, he published [40] an analysis using Volterra series.

The published work since 1967 has generally made reference

to one or more of the trio Lotsch, Reynolds and Narayanan. Some papers refine the results for particular cases or show simplified analysis [30, 38, 117] while others apply the results to produce devices with lower distortion [29, 148, 105].

Relationships between distortion products at low frequency have been dealt with by K. Simons [82] while S. Perlow [3] has considered the frequency dependence.

This chapter reviews distortion analysis, in particular the analysis of exponential distortion due to the emitter base junction and the high frequency distortion due to the non-linear collector capacitance and emitter resistance. The analysis is here extended up to fifth order to determine the relative effect of higher order distortions. Further analysis is carried out here to determine the distortion caused by the major second order effects. These being; variation in output conductance due to Early effect, variation in transconductance due to high injection effect and variation in base-width due to Kirk effect. Results of this analysis are then used to show the potential advantage of the novel structure.

4.2 Low Frequency

At frequencies less than $1/2\Pi\tau_F\beta$ the device can be considered as memoryless. Under these conditions the

analysis can be performed on a DC transfer curve.

If the output variable, y, is a function of the input variable, x, the transfer function $y = f(x)$ can be expanded into a power series by a Taylor expansion. This expansion can be taken around any point. In this case the quiescent operating point, x_0, is most appropriate.

We obtain;

$$y = a_0 + xa_1 + x^2 a_2 + x^3 a_3 + +$$

where;

$$a_n = \frac{f_n(x_0)}{n!}$$

Here a_0 is the DC operating point and a_1 is the small signal gain. Higher order coefficients express the non-linearities of the device. The input signal level, x, is referred to x_0, i.e. a deviation from x_0. The derivatives can be used to give various figures of merit as described in Appendix VII.

4.3 Harmonic Distortion

If the input is of frequency f_1 components of $2f_1$, $3f_1$ etc. can appear in the output. To relate this type of

distortion to the power series the substitution $x = b\cos wt$ is made, then using the identity $\cos^2 wt = (1 + \cos 2wt)/2$ the various frequency terms can be calculated, i.e. the square term produces a 2f component and a DC component. The distortions produced by the higher order coefficients are summarized in Figure 4.3.1. It will be noted that even order terms produce a DC term; this is known as rectification distortion. Odd order terms produce a component at the frequency, f; this is known as gain compression distortion and can either increase or decrease the linear gain. The amplitude of any nth order component is proportional to the nth power of the input level, i.e. if $x = b\cos wt$ the third order distortion components are proportional to b^3.

n	DC	1f	2f	3f	4f	5f	6f	7f	8f
0	1								
1		1							
2	1/2		1/2						
3		3/4		1/4					
4	3/8		1/2		1/8				
5		5/8		5/16		1/16			
6	5/16		15/32		3/16		1/32		
7		35/64		21/64		7/64		1/64	
8	32/128		7/16		7/32		1/16		1/128

Figure 4.3.1. Harmonic Components Produced by coswt

4.4 Intermodulation Distortion

When the input consists of a sum of two or more frequencies, the distortions of the device can give sum and difference frequencies at the output. To investigate this the substitution $x = b_1 \cos w_1 t + b_2 \cos w_2 t$ is used. Second order distortion gives frequencies $f_1 \pm f_2$, $2f_1$ and $2f_2$ while third order distortion gives frequencies of $2f_1 \pm f_2$, $2f_2 \pm f_1$, $3f_1$, etc. As discussed earlier 2f and 3f components are harmonic distortion, the sum and difference frequencies are intermodulation distortions.

A summary of all the distortion products up to fifth order is given in Figure 4.4.1. Frequency components at the input are identified as a, b ... e and the output components are identified as sum and difference frequencies. Where a term of the form (a - a) appears this represents a DC component and does not affect the frequency of the particular expression. Amplitude at the output of any component is obtained by multiplying the coefficient by the magnitude (M) and the amplitude of all of the input signals that contribute. As an example we see that the component a ± (b - b) has frequency, a, and an amplitude proportional to the input, a, and the square of the input, b. It should also be noted that this is one of three components of frequency a due to third order distortion.

Coeff-icient	Number of Fre-quencies at Input (N)	Type of Product	Mag-nitude (M)	Permu-tatio-ns for N Sig-nals	Total Number of Comp-onents
a_1	1	a	1	1	1
a_2	2	(a - a)	1/2	2	2
		2a	1/2	2	2
		a ± b	1	1	2
a_3	3	a ± (a - a)	3/4	3	3
		a ± (b - b)	3/2	6	6
		3a	1/4	3	3
		2a ± b	3/4	6	12
		a ± b ± c	3/2	1	4
a_4	4	(a - a)±(a - a)	3/8	4	4
		(a - a)±(b - b)	3/2	6	6
		2a ± (a - a)	1/2	4	4
		2a ± (b - b)	3/2	12	12
		a ± b ± (a - a)	3/2	12	24
		a ± b ± (c - c)	3	12	24
		4a	1/8	4	4
		2a ± 2b	3/4	6	12
		3a ± b	1/2	12	24
		2a ± b ± c	3/2	12	48
		a ± b ± c ± d	3	1	8
a_5	5	a±(a - a)±(a - a)	5/8	5	5
		a±(a - a)±(b - b)	15/4	20	20
		a±(b - b)±(b - b)	15/8	20	20
		a±(b - b)±(c - c)	15/2	30	30
		3a ± (a - a)	5/16	5	5
		3a ± (b - b)	5/4	20	20
		2a ± b ± (a - a)	5/4	20	40
		2a ± b ± (b - b)	15/8	20	40
		2a ± b ± (c - c)	15/4	60	120
		a±b±c±(a - a)	15/4	30	120
		a±b±c±(d - d)	15/2	20	80
		5a	1/16	5	5
		3a ± 2b	5/8	20	40
		3a ± b ± c	5/4	30	120
		2a ± 2b ± c	15/8	30	120
		2a ± b ± c ± d	15/4	20	160
		4a ± b	5/16	20	40
		a ± b ± c ± d ± e	15/2	1	16

Figure 4.4.1. Frequency Components

Generally the most objectionable products are those which take the form of either 2a - b or a + b - c, these being of similar frequency to a, b and c. These types of products are caused by all odd order coefficients, although third order effects would normally be the largest in magnitude.

4.5 Cross-Modulation

Where the input consists of an amplitude modulated signal and an unmodulated signal, the distortions can cause the transfer of the modulation to the unmodulated frequency.

In particular the third order distortion component of form a ± (b - b) is responsible for this type of modulation. It can be seen that the amplitude of frequency a will be controlled by the amplitude of frequency b. With an input of the form;

$$B(1 + M\cos mt)\cos bt + A\cos at$$

The aforementioned distortion produces the relevant terms of;

$$3a_3 AB^2 M(\cos mt)\cos at$$

$$(3/2)a_3 AB^2 (1 + M^2/2)\cos at$$

while a ± (a - a) produces;

$$(3/4)a_3 A^3 \cos at$$

and the linear term produces;

$$a_1 A \cos at$$

Combining terms gives a resulting cross-modulation of;

$$A_C (1 + M_C \cos mt) \cos at$$

where;

$$A_C = A(a_1 + (3/4)a_3 (B^2 (2 + M^2) + A^2))$$

and;

$$M_C = \frac{12a_3 B^2 M}{4a_1 + 3a_3 (B^2 (2 + M^2) + A^2)}$$

4.6 High Frequency

At high frequency the delay of the transistor, τ_F, becomes a significant part of the signal period. Under these conditions the device cannot be considered as memoryless. In effect the stored charge in the transistor at a particular time will influence the output at a later time.

A Volterra series can be used to analyse this type of system, see Appendix III.

The Taylor series can be considered as a special case of the Volterra series and the coefficients as a special case of the Volterra kernels. Volterra kernels are like weighting functions which describe the effect of past events on the present, they are in fact multi-dimensional impulse responses. The kernels are functions of frequency, as well as bias conditions. Several authors [30, 29, 42] have analysed the transistor with this type of approach, the results in general being quite complicated. In the high frequency limit, however, these results reduce to a simple form.

A simple approach to this type of analysis is to start from a charge control model having made the high frequency approximations. Appendix IV shows how the generator current (I_S) can be related to the collector current (I_C) by;

$$I_S = \tau(I_C)\frac{dI_C}{dt}$$

From this relationship the high frequency kernels can be obtained, as shown in Appendix V.

An outline of this procedure is as follows. The total

delay, τ_F, is expressed as a power series in I_C by means of a Taylor expansion about the quiescent point. Thus expressing the input signal as a function of the output signal (I_C). Next the output is expressed as a Volterra series in terms of the input current (I_S) where the kernels as algebraic quantities. The Volterra series is then substituted for the output signal in the power series, to produce an equation relating the Volterra kernels to the input current. The kernels are then determined by first expanding the series and then collecting terms of like order to produce a set of simultaneous equations. These are then solved to give the values of the kernels. Results, given to fifth order in Appendix V, show that the kernels can be expressed in terms of the derivatives of $\tau(I_C)$, which in turn can be expressed in terms of the transistors structure.

Poon [30] has shown that in the high frequency limit the distortions can be related to the derivatives of $f_T(I_C)$. Appendix VI shows that these two approaches give identical results.

4.7 Distortion in the Transistor

Distortion is caused by any non-linear effect within the transistors' operation. These may be put into three broad groups: exponential $I_C(V_{BE})$, gain variations $\beta(I_C, V_{CE})$, capacitance $C_{CB}(V_{CB})$.

Exponential non-linearity is fundamental to bipolar operation. The ideal equation, $I_C = I_S(\exp(V_{be}/V_T) - 1)$ has been analysed by various authors [44, 46, 48] under various conditions by a Taylor expansion around the bias point.

At low frequencies emitter degeneration can be used. A resistor placed in the emitter circuit improves the linearity by negative feedback. In effect the variation in r_e occurs in a resistance of $r_e + R_e$, thus has less effect than when it occurs in r_e alone. At higher frequencies, however, when the h_{fe} is lower, negative feedback is not always usable.

Gain variations are caused by the distribution and movement of charge within the transistor. These effects include: Early effect, base push-out, current crowding, conductivity modulation and avalanche multiplication. The Early effect and avalanche multiplication being a function of the collector voltage, while the others are due to the current densities. As may be expected, these effects can be reduced by proper control of the doping profiles. However, in high frequency devices this can be compromised by the need for small geometries to achieve the required f_T.

Capacitance and charge storage effects are the non-linear memory of the device and give rise to a frequency

dependence. In effect these may be ignored at low frequencies, but at high frequency they begin to have a larger effect. The analysis becomes more involved, due to the memory, a Volterra approach can then be used [40, 153].

4.8 Exponential Non-linearity

Considering this in isolation and under small signal conditions when β can be assumed constant, Figure 4.8.1.

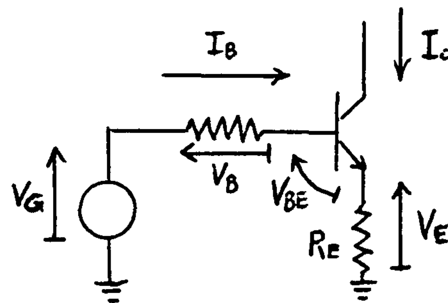


Figure 4.8.1. Ideal Transistor at Low Frequency

Now;

$$V_G = V_B + V_{BE} + V_E = I_B R_B + V_T \ln(I_C/I_S + 1) + I_E R_E$$

and now using $I_E = (1 + \beta)I_C/\beta$ and $I_B = I_C/\beta$

we get;

$$V_G = (R_B + (1+\beta)R_E)I_C/\beta + V_T \ln(I_C(1+\beta)/(I_S\beta) + 1)$$

This can be changed into a form whereby I_C is a function of V_G by using a Taylor expansion. To obtain the first derivative;

$$\frac{dV_G}{dI_C} = \frac{R_B + (1+\beta)R_E}{\beta} + \frac{V_T(1+\beta)}{I_C(1+\beta) + I_S\beta}$$

now since I_S is many orders of magnitude below the normal operating value of I_C the second term reduces simply to V_T/I_C to give;

$$\frac{dI_C}{dV_G} = \frac{1}{R_B/\beta + (1+\beta)R_E/\beta + V_T/I_C} = \frac{1}{R_X(I_C)}$$

Now differentiating by V_G and substituting for dI_C/dV_G the second derivative is obtained;

$$\frac{d^2 I_C}{dV_G^2} = \frac{V_T}{R_X^3 I_C^2}$$

similarly;

$$\frac{d^3 I_C}{dV_G^3} = \frac{3V_T^2}{R_X^5 I_C^4} - \frac{2V_T}{R_X^4 I_C^3}$$

and so on for higher derivatives. If V_Q is taken as the

quiescent input and V the signal about this point, then $V_G = V_Q + V$. The quiescent collector current is then $I_Q(V_Q)$. The Taylor series can now be expanded about the quiescent point.

$$I_C(V_Q + V) = I(V_Q) + I'(V_Q)V + I''(V_Q)V^2/2! + +$$

Now denoting $I_C(V_Q + V)$ by I_0 and $I(V_Q)$ by I_Q we have;

$$I_0 = a_0 + a_1 V + a_2 V^2 + a_3 V^3 + +$$

where;

$$a_0 = I_Q$$

$$a_1 = \frac{1}{R_X(I_Q)} = \frac{1}{R_X}$$

$$a_2 = \frac{V_T}{2R_X^3 I_Q^2}$$

$$a_3 = \frac{V_T^2}{2R_X^5 I_Q^4} - \frac{V_T}{3R_X^4 I_Q^3}$$

etc.

With similar expressions for higher order coefficients.

From the definitions of M2, M3 we obtain the figures of merit, which indicate the amount of distortion;

$$M2 = \frac{V_T}{2R_X I_Q}$$

$$M3 = \frac{V_T^2}{2R_X^2 I_Q^2} - \frac{V_T}{3R_X I_Q}$$

Appendix VIII shows how the analysis is extended to fifth order to give;

$$M4 = \frac{5V_T^3}{8R_X^3 I_Q^3} - \frac{5V_T^2}{6R_X^2 I_Q^2} + \frac{V_T}{4R_X I_Q}$$

$$M5 = \frac{7V_T^4}{8R_X^4 I_Q^4} - \frac{7V_T^3}{4R_X^3 I_Q^3} + \frac{13V_T^2}{12R_X^2 I_Q^2} - \frac{V_T}{5R_X I_Q}$$

As stated earlier, the transistor is a non-linear device and any R_B or R_E would tend to linearize it. Thus we may set $R_B = R_E = 0$ to find the maximum distortion and on substitution find;

$$M2 = \frac{1}{2!} \qquad D2 = \frac{I_L}{2!I_Q}$$

$$M3 = \frac{1}{3!} \qquad D3 = \frac{I_L^2}{3!I_Q^2}$$

$$M4 = \frac{1}{4!} \qquad D4 = \frac{I_L^3}{4!I_Q^3}$$

$$M5 = \frac{1}{5!} \qquad D5 = \frac{I_L^4}{5!I_Q^4}$$

It is easily seen that distortion caused by lower order coefficients is greater than that caused by higher order coefficients. It is also apparent that all distortions are reduced at higher quiescent currents and, the higher order terms reduce faster than the lower order terms. It should be noted that the figures of merit reduce due to a smaller generator voltage being required for the same output at higher quiescent current. The details of this can be seen more clearly by noting the expressions for a_0, a_1, etc. with $R_B = R_E = 0$.

$$a_1 = \frac{I_Q}{V_T}$$

$$a_2 = \frac{I_Q}{2!V_T^2}$$

$$a_3 = \frac{I_Q}{3!V_T^3}$$

etc.

We see that all coefficients are proportional to the bias current, as would be expected from the nature of the exponential function. The third order distortion, for example, increases as the third power of the input level, i.e. as $(V/V_T)^3$. It is now obvious that doubling the bias current will double both the linear signal and the third order, and in fact all other, distortion components at the output. Now to return the linear output component to its original level, we simply halve the input signal. In doing this the third order component will, at the output reduce eight fold i.e. $1/2^3$. The resulting third order component is therefore one quarter of it's original level, that is (original level) x 2 x $1/2^3$, which is exactly what D3 tells us i.e. third order components are proportional to $1/I_Q^2$. In effect the distortion is given with the interfering frequency components at a current level, I_L, at the output. This is the same information as IM3 which is a decibel relationship using a reference power, 1mW, at the output. While the figures of merit, M2 etc, are defined in terms of the maximum linear current that can be obtained at the output, again this is the same information as IM3. Here for example M2 tells us that if we double the linear output current and also double the quiescent current then the distortion effects will also double, i.e. there is no penalty in increasing the output level.

Distortion caused with non-zero base and/or emitter resistors is also a function of the quiescent current. At

higher bias levels the term V_T/I_Q, that is r_e, becomes smaller and effects of the ohmic R_E and R_B become more important. As I_Q becomes larger the last term in each figure of merit first begins to dominate and then in the limit tends to zero. In fact the distortion performance will show nulls for the correct biasing conditions and, in fact cancellation can be achieved. Towards the limit the figures of merit are;

$$M2 = \frac{V_T}{2R_X I_Q} \qquad D2 = \frac{V_T I_L}{2R_X I_Q^2}$$

$$M3 = -\frac{V_T}{3R_X I_Q} \qquad D3 = -\frac{V_T I_L^2}{3R_X I_Q^3}$$

$$M4 = \frac{V_T}{4R_X I_Q} \qquad D4 = \frac{V_T I_L^3}{4R_X I_Q^4}$$

$$M5 = -\frac{V_T}{5R_X I_Q} \qquad D5 = -\frac{V_T I_L^4}{5R_X I_Q^5}$$

In practice the distortion decreases with an increase in bias current, which can be viewed as a consequence of the reduction in r_e. However, there is a limit to I_Q since at higher current levels other effects serve to increase the distortion. Several authors [30, 41] have noted this point and come to the conclusion that increasing the collector current and emitter area to maintain a constant

current density results in an improvement of third order distortion of $-40\log R$, where R is the ratio of areas (and currents). This being a 12dB improvement for a doubling of current, power and capacitance considerations given an upper limit for this technique. The figure of merit M3 ($R_B = R_E = 0$) shown here exactly agrees with this. In general the nth order distortions improve by a factor of $20(n-1)\log R$ dB where R is the ratio of the two currents. In the event that R_B and R_E are not zero the effect is more pronounced, asymptoting to $20n\log R$ dB as the ohmic resistance increases.

Take the M3 as an example;

$$M3 = \frac{V_T^2}{2R_X^2 I_Q^2} - \frac{V_T}{3R_X I_Q}$$

using $R_X = R_Y + V_T/I_Q$, after some manipulation we find;

$$M3 = \frac{V_T/(I_Q R_Y) - 2}{6(V_T/(I_Q R_Y) + 1)^2} \times \frac{V_T}{I_Q R_Y}$$

$$\approx -\frac{V_T}{3I_Q R_Y}\Bigg|_{I_Q R_Y \gg V_T}$$

In fact this limit can be rapidly approached for small values of R_Y, i.e. $R_Y = 2\Omega$, $I_Q = 100mA$, $V_T = 25mV$ gives $V_T/I_Q R_Y$ a value of 0.125. A graph produced in

[41] Figure 4.8.2 to demonstrate the 12dB per doubling of current rule in fact shows a 15dB per doubling of current indicating the presence of some ohmic resistance.

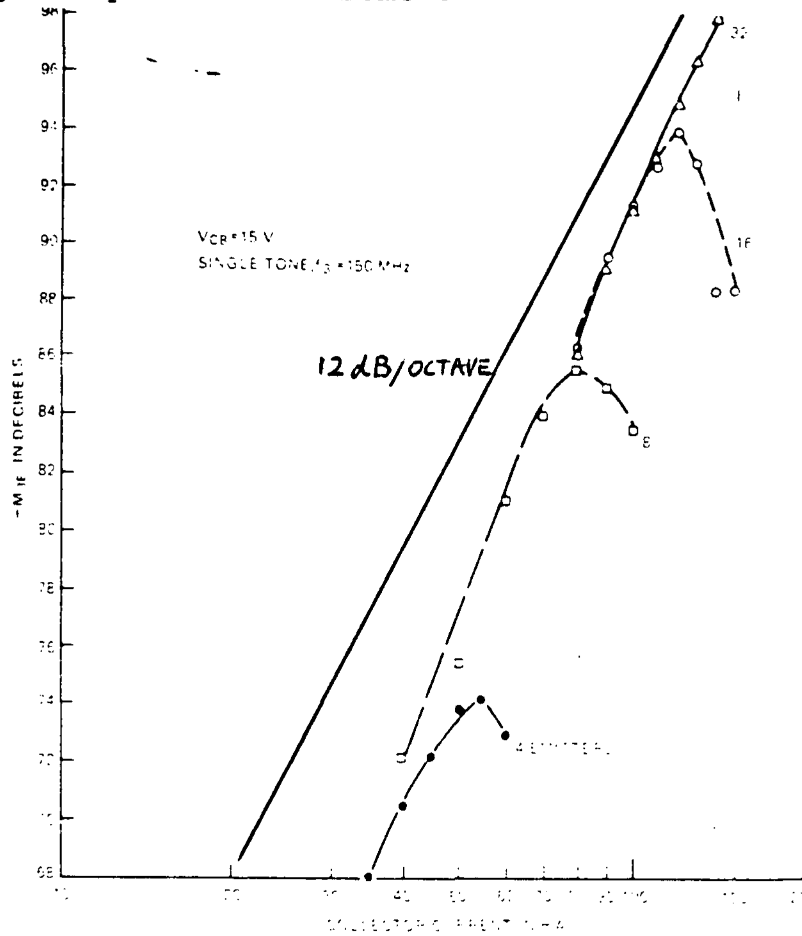


Figure 4.8.2. Variation From the Ideal 12dB/octave

4.9 Early Effect Non-linearities

To first order the Early effect can be modelled by [159];

$$I_c'' = \frac{I_c(V_{be})}{1 - V_{cb}/V_A}$$

where $I_c(V_{be})$ is the collector current at $V_{cb} = 0$ and V_A is the Early voltage.

To investigate the distortion caused by Early effect we may use the circuit shown in Figure 4.9.1. With the capacitor made large enough, only the DC current will flow in R. Now using $V_{cb} = V_S - RI_C$ in the previous equation, we find the DC conditions are given by a quadratic equation.

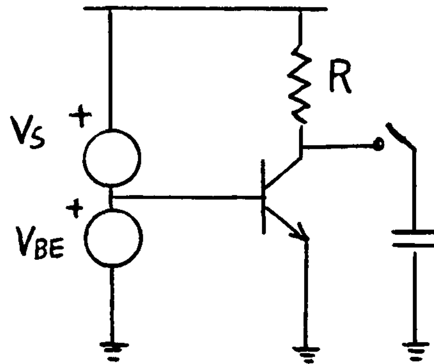


Figure 4.9.1. Circuit to determine Early effect distortion

Solving this gives the bias current as;

$$I_{CO} = \frac{V_S - V_A + ((V_A - V_S)^2 + 4RI_C(V_{be})V_A)^{1/2}}{2R}$$

which is larger than would be expected by consideration of V_{be} alone.

Analysis of the distortion will be carried out by comparing the current that would flow in the collector with the capacitor connected and without the capacitor. This novel technique will determine the distortion due to the Early effect in isolation. With the capacitor the collector current can be represented by;

$$I_{cc} = b_0 + b_1 x + b_2 x^2 + +$$

where b_0 is the DC component given by the previous equation and x is the arbitrary input variable.

With the capacitor out of circuit all current will flow in the resistor which will cause an AC variation in the collector voltage; this in turn modifies the AC signal. The resultant signal is represented by;

$$I_c = a_0 + a_1 x + a_2 x^2 + +$$

where $a_0 = b_0$ since the DC conditions have not changed.

We now have;

$$I_{cc} = \frac{I_c(V_{be})}{1 - (V_S - b_0 R)/V_A}$$

and;

$$I_c = \frac{I_c(V_{be})}{1 - (V_S - I_c R)/V_A}$$

which combine to give;

$$KI_{cc} = (K + gi)(a_0 + i)$$

where;

$$K = 1 - (V_S - a_0 R)/V_A$$

$$g = R/V_A$$

and;

$$i = I_C - a_0$$

is the small signal current. It should be noted that the form of this equation is a direct result of the first order model used. The more general case can be considered as having V_A as a function of V_{cb}. The resulting equation is the same as before except the left-hand side has $V_A(V_{cb})$ substituted for V_A while the right-hand side uses $V_A(V_{cb})$ at the bias voltage.

It is possible to proceed by taking a Taylor expansion of K and g and following the same line as will be used here. This analysis, however, will be restricted to the first order model of the Early effect for several reasons. The first order model is generally adequate for circuit analysis and would be expected to give a reasonable distortion model. This is justifiable since V_A is a weak function of collector voltage and any expansion of V_A would have small coefficients compared to the expansion point. Additionally, in order to relate any expansion to

the transistor structure specific knowledge of the doping profile would be needed (possibly from C-V data) making the analysis specific to transistor type.

The expansion of I_{cc} may be chosen at will, for instance the expansion due to the exponential base-emitter junction could be used. In this event the result obtained for I_c would indicate the distortion due to the combination of the two non-linearities.

To examine the Early effect in isolation we need to see the effect on a linear signal, hence we use;

$$I_{cc} = b_0 + b_1 x$$

This is substituted into the last equation along with the expansion for i.

Now terms of equal order, power of x, are equated to give, to fifth order, the set of equations;

$$b_0 = a_0$$

$$b_1 = a_1 h/K$$

$$0 = a_2 h + g a_1^2$$

$$0 = a_3 h + g 2 a_1 a_2$$

$$0 = a_4 h + g(2a_1 a_3 + a_2^2)$$

$$0 = a_5 h + g2(a_1 a_4 + a_2 a_3)$$

where;

$$h = K + ga_0$$

Note that $h = 1$ when $2a_0 R = V_S$, a typical bias point.

The equations are now solved by successive substitution to give;

$$a_0 = b_0$$

$$a_1 = b_1 K/h$$

$$a_2 = -gb_1^2 K^2/h^3$$

$$a_3 = 2g^2 b_1^3 K^3/h^5$$

$$a_4 = -5g^3 b_1^4 K^4/h^7$$

$$a_5 = 14g^4 b_1^5 K^5/h^9$$

which are the coefficients describing I_C. The figures of merit are;

$$M2 = -d$$

$$M3 = 2d^2$$

$$M4 = -5d^3$$

$$M5 = 14d^4$$

where;

$$d = ga_0/h = \frac{1}{(V_A - V_S)/Ra_0 + 2}$$

By inspection is can be seen that distortion is reduced by a larger V_A, smaller supply voltage or a smaller voltage across the load resistor. Typical values, say, $V_A = 50V$, $V_S = 20V$, $R = 50\Omega$, $a_0 = 0.2A$ give: $M2 = -0.2$, $M3 = 0.08$, $M3 = -0.04$ and $M5 = 0.02$.

This analysis has been carried out on a common base configuration with voltage drive, that is;

$$I_C(V_{be}) = I_S(\exp(V_{be}/V_T) - 1)$$

To extend the analysis to other configurations is simple. For the common emitter with voltage drive it is sufficient to note that V_S must be replaced by the supply voltage with respect to the base voltage.

For current drive at the base it is sufficient to note that I_C and h_{fe} are modified by V_{cb} in the same way. The resultant expression $I_C = h_{fe} I_B$ indicates that the analysis is identical, to first order.

For current drive in common base, cascode, we see the major advantage this configuration has by noting;

$$I_C = \frac{h_{fev} I_E}{1 + h_{fev}}$$

where;

$$h_{fev} = \frac{h_{fe}}{1 - V_{cb}/V_A}$$

which when substituted gives;

$$I_C = \frac{I_E h_{fe}/(1 + h_{fe})}{1 - V_{cb}/((1 + h_{fe})V_A)}$$

i.e. V_A has been replaced by $(1 + h_{fe})V_A$ and the numerator is $I_C(I_E)$ which is independent of Early effect because h_{fe} is the value at $V_{cb} = 0$.

It should be noted that this result is also applicable for the case of voltage drive to a common emitter configuration with some ohmic resistance, R_E, in the

emitter circuit. If $R_E \gg r_e$ then the applied voltage appears across R_E, which is equivalent to a current drive into the emitter.

4.10 High Injection Non-linearities

The transition from low injection to high injection is well modelled by the SPICE model, Chapter 3.

$$I_c = \frac{I_S}{Q_b}(\exp(V_{be}/(N_F V_T)) - \exp(V_{bc}/(N_R V_T)))$$

$$- \frac{I_S}{\beta_R}(\exp(V_{bc}/(N_R V_T))-1) - I_{SC}(\exp(V_{bc}/(N_C V_T))-1)$$

Where;

$$Q_b = \frac{Q_1}{2}(1 + (1 + 4Q_2)^{1/2})$$

and;

$$Q_1 = \frac{1}{1 - V_{bc}/V_{AF} - V_{be}/V_{AR}}$$

$$Q_2 = \frac{I_S}{I_{KF}}(\exp(V_{be}/(N_F V_T))-1) + \frac{I_S}{I_{KR}}(\exp(V_{bc}/(N_R/V_T))-1)$$

I_c can be expanded as a power series in V_{be}, with $V_{cb} = 0$ the equations reduce to

$$I_c = \frac{I_s}{Q_b} \exp(V_{be}/(N_F V_T))$$

$$Q_1 = 1$$

and;

$$Q_2 = \frac{I_s}{I_{KF}} \exp(V_{be}/(N_F V_T))$$

Where the current due to the collector base voltage is negligible compared to the current due to the emitter base voltage and the reverse Early voltage is large compared to the emitter forward bias junction voltage.

The collector current is described by;

$$I_c = \frac{2I_s \exp(V_{be}/(N_F V_T))}{1+(1+I_s/I_{KF} \exp(V_{be}/(N_F V_T)))^{1/2}}$$

It can be seen that at low currents, $I_c << I_{KF}$, I_c is given by;

$$I_s \exp(V_{be}/(N_F V_T))$$

while at high currents, $I_C >> I_{KF}$, it is given by;

$$2(I_S I_{KF})^{1/2} \exp(V_{be}/(2N_F V_T))$$

In both cases the ideal diode law is obeyed, although with different saturation current and emission coefficient. The figures of merit for both of these asymptotes are given by;

$$M(n) = 1/n!$$

as detailed earlier. During the change from one regime to the other the ideal diode law is not followed, this can be viewed as a change in emission coefficient. For an increase in current, at around the knee current I_{KF} the gm of the transistor does not increase in proportion to I_C/V_T but at a reducing rate as the gm tends to $I_C/2V_T$. This extra modification of the gm then modifies, and in fact reduces the distortion generated by the transistor due to its linearizing effect.

A Taylor expansion can now be used to analyse the transfer characteristic of a transistor biased close to the knee current. The first five derivatives of I_C with respect to V_{be} are given in Appendix VIII by;

$$\frac{dI_C}{dV_{be}} = gh_1$$

$$\frac{d^2 I_C}{dV_{be}^2} = g^2 (h_1 - h_2/2)$$

$$\frac{d^3 I_C}{dV_{be}^3} = g^3 (h_1 - 3h_2/2 + 3h_3/4)$$

$$\frac{d^4 I_C}{dV_{be}^4} = g^4 (h_1 - 7h_2/2 + 9h_3/2 - 15h_4/8)$$

$$\frac{d^5 I_C}{dV_{be}^5} = g^5 (h_1 - 15h_2/2 + 75h_3/4 - 75h_4/4 + 105h_5/16)$$

where;

$$g = \frac{1}{N_F V_T}$$

$$h_n = \frac{I_S \exp(nV_{be}/(N_F V_T)) \ (I_S/I_{KF})^{n-1}}{(1 + I_S/I_{KF} \exp(V_{be}/(N_F V_T)))^{n-1/2}}$$

The figures of merit are given by;

$$M2 = \frac{(h_1 - h_2/2) I_C}{2! h_1^2}$$

$$M3 = \frac{(h_1 - 3h_2/2 + 3h_3/4) I_C^2}{3! h_1^3}$$

$$M4 = \frac{(h_1 - 7h_2/2 + 9h_3/2 - 15h_4/8)I_C{}^3}{4!h_1{}^4}$$

and;

$$M5 = \frac{(h_1 - 15h_2/2 + 75h_3/4 - 75h_4/4 + 105h_5/16)I_C{}^4}{5!h_1{}^5}$$

These can now be evaluated at the knee current I_{KF}, i.e. $I_S \exp(V_{be}/N_F V_T) = I_{KF}$, which gives;

$$I_C = \frac{2I_{KF}}{1+2^{1/2}} = 0.828I_{KF}$$

$$h_n = \frac{I_{KF}}{2^{n-1/2}}$$

and;

$$M2 = 0.4392 \qquad D2 = 0.4392 I_L/I_Q$$

$$M3 = 0.09997 \qquad D3 = 0.09997 I_L{}^2/I_Q{}^2$$

$$M4 = 9.401 \times 10^{-3} \qquad D4 = 9.401 \times 10^{-3} I_L{}^3/I_Q{}^3$$

$$M5 = 5.548 \times 10^{-5} \qquad D5 = 5.548 \times 10^{-5} I_L{}^4/I_Q{}^4$$

These show an improvement over the figures of merit for

the exponential base - emitter junction alone, for which $M2 = 0.5$, $M3 = 0.167$, $M4 = 0.0417$ and $M5 = 8,33 \times 10^{-3}$. This being due to the smaller variation in gm that occurs at around the knee current.

4.11 Kirk effect non-linearities

It is unclear in many transistors as to which mechanism is responsible for the reduction in h_{fe} at high currents. Models due to Kirk and Van der Ziel both describe observable results and it is expected that most transistors conform to one or other theory in some respects. Apart from this it is clear that base push-out is to be avoided in normal use. The large change in base-width that occurs has a profound effect on transistor action, reducing both the f_T and the base input impedance.

The base-width is given by the various equations in Chapter 3 along with expressions for τ_F and h_{fe}. For the one dimensional, Kirk, model the forward transit time is given by;

$$\tau_F = \frac{V_T C_{TE}}{I_C} + \frac{W_{BO}^2}{N_B D_{NB}} + \frac{W_{CIB}^2}{4D_{NC}} + \frac{W_{EPI} - W_{CIB}}{2V_D}$$

where W_{CIB} represents the charge induced base. This being zero for $I < I_0$ and increasing in magnitude with $I > I_0$. Derivatives of this with respect to I_C determine the

distortion in the high frequency limit. We have;

$$\frac{d\tau_F}{dI_C} = -\frac{V_T C_{TE}}{I_C^2} + \frac{1}{2}\left[\frac{W_{CIB}}{D_{NC}} - \frac{1}{V_D}\right]\frac{dW_{CIB}}{dI_C}$$

and;

$$\frac{d^2\tau_F}{dI_C^2} = \frac{2V_T C_{TE}}{I_C^3} + \frac{1}{2}\left[\frac{W_{CIB}}{D_{NC}} - \frac{1}{V_D}\right]\frac{d^2 W_{CIB}}{dI_C^2} + \frac{1}{2D_{NC}}\left[\frac{dW_{CIB}}{dI_C}\right]^2$$

where these equations are valid for the case of zero collector load resistance. For non-zero load resistance, an additional term $R_C C_C$ is required in the expression for τ_F. The original expression given by Kirk [27] included a term which was the product of the collector transition capacitance and the resistance of the epitaxial part of the extrinsic collector. Later this term was taken to be approximately zero and dropped from the expression. For the grounded collector case this is correct since the internal collector series resistance must be small in order to avoid premature saturation. Ideally the product of maximum collector current and internal collector resistance should be in the order of tens of millivolts or less. A typical load resistance will be therefore be two or three orders of magnitude larger than the internal collector resistance. Now the product of this load

resistance and the collector transition capacitance must be of similar order to, or less than, other delays in the transistor, because if this is not the case then it is not a high frequency transistor owing to an excessively large collector capacitance. It is now easily seen that the original approximation, that the collector delay due to the internal resistance is zero, is correct. Collector delay due to the load resistance is however not generally negligible, since it may be arbitrarily large.

The full expression is therefore;

$$\tau_F = \frac{V_T C_{TR}}{I_C} + \frac{W_{BO}^2}{N_B D_{nB}} + \frac{W_{CIB}^2}{4D_{nc}} + \frac{W_{epi} - W_{CIB}}{2V_D} + \frac{C_{co} R_C}{(1-V_{BC}/\phi)^{1/n}}$$

and the first two derivatives are;

$$\frac{d\tau_F}{dI_C} = - \frac{V_T C_{Te}}{I_C^2} + \frac{1}{2}\left[\frac{W_{CIB}}{D_{nc}} - \frac{1}{V_D}\right]\frac{dW_{CIB}}{dI_C} + \frac{C_C R_C^2 \, 1/n}{V_{CE}}$$

$$\frac{d^2 \tau_F}{dI_C^2} = \frac{2V_T C_{Te}}{I_C^3} + \frac{1}{2}\left[\frac{W_{CIB}}{D_{nc}} - \frac{1}{V_D}\right]\frac{d^2 W_{CIB}}{dI_C^2} + \frac{1}{2D_{nc}}\left[\frac{dW_{CIB}}{dI_C}\right]^2$$

$$+ \frac{C_C R_C^3 \, 1/n(1/n+1)}{V_{CE}^2}$$

where C_c is the collector capacitance at the collector bias voltage and V_{CE} is taken as approximately $\phi - V_{BC}$. The relevant derivatives of W_{CIB} are given in Chapter 3 and are;

$$\frac{dW_{CIB}}{dI_C} = \frac{I_0 W_{epi}}{I_C^2}$$

$$\frac{d^2 W_{CIB}}{dI_C^2} = - \frac{2 I_0 W_{epi}}{I_C^3}$$

for the low field case and;

$$\frac{dW_{CIB}}{dI_C} = \frac{(I_0 - I_{CX})^{1/2} W_{epi}}{2(I_C - I_{CX})^{3/2}}$$

$$\frac{d^2 W_{CIB}}{dI_C^2} = \frac{3(I_0 - I_{CX})^{1/2} W_{epi}}{4(I_C - I_{CX})^{5/2}}$$

for the high field case.

A similar set of equations exist for the two dimensional model, viz;

$$\tau_F = \frac{V_T C_{Te}}{I_C} + \frac{W_{BO}^2}{N_B D_{NB}} + \frac{W_{LB}^2}{N_B 4 D_{NB}} + \frac{W_{epi}}{2V_D} + C_C R_C$$

$$\frac{d\tau_F}{dI_C} = -\frac{V_T C_{Te}}{I_C^2} + \frac{W_{LB}}{N_B 2D_{NB}} \cdot \frac{dW_{LB}}{dI_C} + \frac{C_C R_C^2 1/n}{V_{CE}}$$

$$\frac{d^2\tau_F}{dI_C^2} = \frac{2V_T C_{Te}}{I_C^3} + \frac{W_{LB}}{N_B 2D_{NB}} \cdot \frac{d^2 W_{LB}}{dI_C^2} + \frac{1}{N_B 2D_{NB}}\left[\frac{dW_{LB}}{dI_C}\right]^2$$

$$+ \frac{C_C R_C^3 1/n(1/n+1)}{V_{CE}^2}$$

where the derivatives of W_{LB} the lateral base spreading are given by;

$$\frac{dW_{LB}}{dI_C} = \frac{L_E}{I_0}$$

$$\frac{d^2 W_{LB}}{dI_C^2} = 0$$

Appendix IX shows a worked example with typical figures for a 6GHz transistor where the values of M2 and M3 are calculated at just below and just above the critical current for each model, Figure 4.11.1. Of the one dimensional models, high field is most appropriate since the device will have small dimensions, typically 5V or more across an epitaxial thickness of 2μm, as in this example, is 25KV/cm.

Dist	$I_c \leq I_o$	$I_c \geq I_o$	Model
M2	0.0104	0.168	1D Low Field
M2	0.0104	3.55	1D High Field
M2	0.0104	0.0104	2D
M3	0.0455	0.114	1D Low Field
M3	0.0455	12.99	1D High Field
M3	0.0455	9.54	2D

Figure 4.11.1. Figures of Merit for Base push-out

The corresponding drift velocity is about 8.5×10^{6} cm/S [174].

The figures presented here need some explanation. It has been pointed out in Chapter 3 that these models have a discontinuity in their derivatives at $I_c = I_o$ and that in practical devices this tends to be smoothed out. As presented here the figures indicate the amount of distortion that would be expected if the transistor were biased just above or below the critical current I_o assuming a sharp transition. For a smooth transition the distortion at I_o would be somewhere between the two figures, in fact the smoother the transition the lower the current at which the figures of merit would increase. A second point to note is that the transistor would not normally be biased near the critical current but at about half this value, for class A operation. As the signal amplitude increases from a small value the figures of merit will stay constant until the peak of the current

approaches the critical current when the figure of merit will start to increase. From the point of view of the distortion level the nth order distortions would increase as the nth power of the fundamental until the peak current approached the critical current when the distortion would increase at a greater rate. This idea is comparable with the rule of thumb that distortion is acceptable until the value of S_{21} falls by 0.1dB.

The main point to note is that the theory developed here is in agreement with the generally observed effect that, there is some collector current above which the distortion increases at an unacceptable rate.

4.12 Advantages of the Novel Structure

The novel structure has several advantages over conventional overlay structures;

1) Even sharing of current along the emitter fingers due to the low resistance of the base metallization, which tends to give evenly distributed base current.

2) Even sharing of current to the two sides of the emitter finger due to the autoregistering of the emitter diffusion.

3) Smaller collector area due to the reduced geometry

advantage of the autoregistering technique.

4) Smaller emitter area due to the more efficient use allowed by the even current distribution.

All these advantages can affect the distortion performance, the extent of which is determined by the dominant distortion mechanism.

Types of distortion are summarized in Figure 4.12.1 with their dependence on various parameters and operating conditions indicated.

Distortion mechanism	Parameters
Exponential	I_C (J_C)
Early effect	V_A
Transition time	C_C C_E I_C
Kirk effect	J_C
High injection	J_C

Figure 4.12.1. variations of distortion

Low and high frequency are here taken as being below and above that frequency at which the common emitter gain has fallen 3dB below the low frequency value respectively.

The novel transistor structure allows an evenly distributed collector current and smaller, emitter and

collector; transition capacitances. Early voltage V_A is a function of the doping profile and as such is not improved by the novel structure. It is concluded that the exponential, transition time, Kirk effect and high injection distortion mechanisms are affected by the novel structure. While the Early effect distortion is not.

Exponential distortion is due to the variation of gm with collector current in the ideal exponential region of operation. As such it depends only on I_C and is not a function of emitter area. Distribution of current within the transistor is important, this is determined by the distributed base and emitter resistances. Variations of distortion with current distribution is a function of the bias current and the associated base and emitter resistances. The case of the distribution to the two sides of the emitter due to misalignment between base and emitter is discussed in Chapter 2. The exponential distortion may increase or decrease with misalignment according to the resistances in the base and emitter terminals and the bias current. Variation of current along the emitter finger is more complicated but is reduced by the increased attenuation constant, Chapter 3, that the novel transistor exhibits.

High injection distortion is due to a variation in gm due to current density. This manifests itself as a reduction in gm by a factor of two over some current range. It

occurs when the density of mobile carriers in the base becomes comparable to the fixed acceptor density. As such it will start to occur at lower collector currents when the current distribution is uneven within the transistor.

Base push-out occurs when the mobile carriers in the collector overwhelm the fixed charges changing part of the n-type collector to p-type base. Since the base is very thin the current in the collector is distributed in a similar fashion to that in the emitter end base, i.e. current flow is predominantly normal to the top surface of the die. Base push-out thus occurs at lower collector currents when the distribution is uneven.

It is seen that high injection and base push-out are affected in similar fashion by base contact misalignment, both occurring at lower currents with increased misalignment.

Early effect distortion is essentially unaffected both by variations in current distribution and collector area. The output impedance is due to a modulation of current gain with collector voltage and since the collector series resistance is small the local collector voltage is essentially independent of the local current but is set by the total collector current, and the load resistor. Distortion due to early effect can be reduced, in the transistor, only by modifying the doping profile and not

by any changes in geometry.

Collector capacitance is obviously reduced by a reduced collector area and this is advantageous for reduction in third order distortion. Application to reduction in second order distortion is unclear since M2 shows that cancellation occurs between effects due to collector and emitter capacitances. It is interesting to note that when the collector delay $C_c R_c$ dominates the forward delay, the distortions are almost independent of the size of the collector capacitance.

For the type of transistors under discussion i.e. medium power, it is normal to operate in the high injection region, but below base push-out. The latter is to be avoided since the input impedance falls, due to a lower h_{fe}, and the distortion shows a large increases.

The advantages of the novel structure, excluding any modifications in the doping profiles, with regard to improved distortion performance stems in the first place from the improved current sharing. This is both due to the autoregistered emitter and the low resistance of the base metallization, which together give an improved uniformity in junction bias throughout the transistor. This can partly take the place of emitter ballasting to allow a higher transconductance to be achieved. With all parts of the emitter having similar current densities base

push-out is held off to higher currents. The h_{fe} vs I_C curve would thus be expected to remain flat to higher currents. Due to this a given maximum, linear, collector current can be achieved with a smaller transistor area. A further reduction is also achieved due to the autoregistered fabrication technique, making for a sizable reduction in collector-base capacitance.

CHAPTER 5 - A NOVEL TRANSISTOR

5.1 Terms of Reference

The basis of the experimental device was described to the author by Dr. R. Aubasson of Middlesex Polytechnic in 1976, although it had been conceived some years earlier. The idea, as expounded, was as follows:

A base well is fabricated in the substrate and the window, through which the base was diffused, is then reopened. A refractory metal layer is next deposited and narrow slits are then etched in this to form an array above the base well. Multiple emitters are then diffused in using the metal base electrode as a mask, the resulting structure is shown in Figure 5.1.1, where the emitters are seen shorted to the base electrode.

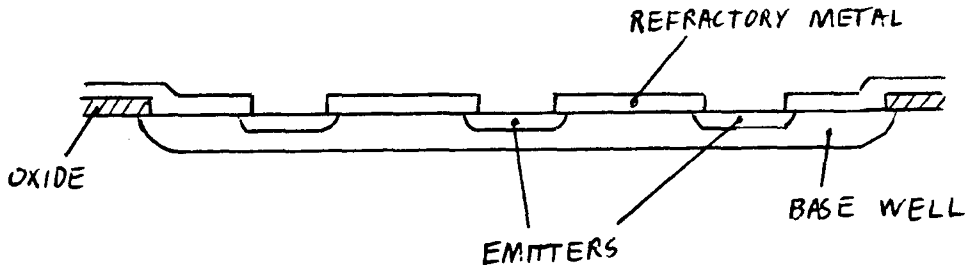


Figure 5.1.1. Transistor After Emitter Formation

The emitter junction must now be recovered by cutting back

the base metal. A suitable dielectric is then deposited and the emitter contacts opened by use of the underlying topography. The emitter electrode metallization is then carried out and the result should be a very linear high frequency transistor.

This chapter considers the above statement from both a manufacture and performance point of view. A feasible structure is proposed and compared to other contemporary structures.

5.2 Autoregistered Transistors

Previous chapters have shown that high frequency gain increases as the ratio of emitter periphery to base area increases. Power transistors with their large emitter peripheries are normally constructed with alternate base and emitter contacts with the collector contact on the back of the die. The repetitive pattern formed by the base and emitter contacts can then be considered as an array of single emitter transistors. Each taking the form of a long narrow emitter accompanied by two base contacts in close proximity. This is true for both the overlay and interdigitated structures. To first order the ratio of emitter periphery to base area is 2/P, where P is the pitch of the emitters. It is, therefore, desirable to make the emitter pitch as small as possible. A cross-section of an interdigitated transistor is shown in

Figure 5.2.1.

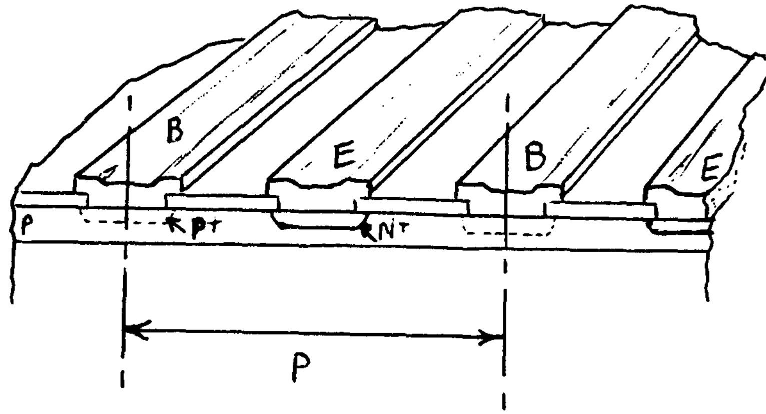


Figure 5.2.1. Interdigitated Transistor

The P+ diffusions are optional depending on the details of the processing. Their purpose is to reduce the resistivity of the extrinsic base region and to give an increased concentration at the surface to improve the ohmic contact at the metal-silicon interface. It is important that the P+ diffusion should not come into contact with the emitter diffusion as it could lead to a reduced breakdown voltage and reduced emitter efficiency. In principle the P+ diffusion can be as shallow as required and a relatively high sheet resistivity tolerated since current flow along the base finger is carried in the metal. The emitter pitch is thus dominated by problems associated with the photoengraving and mask alignment. One of the main features of the structure is that the base and emitter metallizations are coplanar and they must be isolated from each other while covering their respective contact windows. The minimum pitch is determined from

consideration of:

i) Minimum size contact windows for emitter and base

ii) Minimum width of metal to ensure coverage of the windows while tolerating misalignment and etch tolerances (E, B)

iii) Minimum gap between the base and emitter metal that can be reliably etched (G)

iv) When a washed emitter process is used there is also an emitter misalignment tolerance (M)

The pitch is then E + B + 2G for a recut emitter structure or E + B + 2G + 2M for a washed emitter structure, although the emitter will be considerably smaller. The analysis is shown in more detail in Chapter 2. Typical values would give a pitch of 12μm or 14μm for a washed emitter and 10μm or 12μm for a recut emitter. A refined process tuned for a particular structure would be expected to give a smaller pitch.

The overlay transistor uses an emitter metallization which overlays a diffused base grid. A cross-section is shown in Figure 5.2.2. The wide emitter metallization which provides for large emitter currents makes this type of structure ideal for power applications. The base metal

lies adjacent to the emitter metal and perpendicular to the emitter stripe. The emitter pitch is, therefore, not dependent on the limits of the metallization.

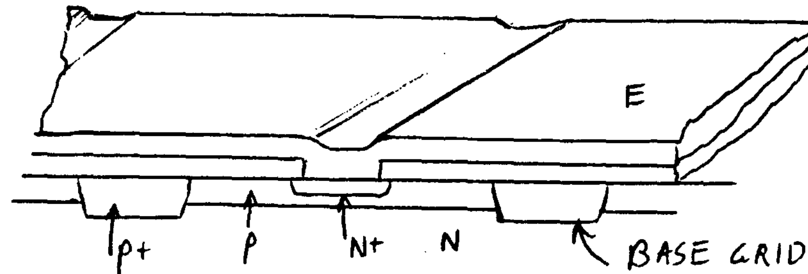


Figure 5.2.2. Overlay Transistor

Base current flowing parallel to the emitter stripe must be carried by the P+ base diffusion. The ramification of this is that the P+ grid must be a deep diffusion of high concentration in order to reduce the base series resistance. The width of the diffusions are necessarily wide when driven in because sideways diffusion is similar to the downwards diffusion and also the diffusion window in the oxide must be large enough to allow sufficient carriers to be deposited. Here then the emitter pitch is dominated by the base grid diffusion and is determined by:

i) Minimum window width for deposition of required total of carriers (W)

ii) Junction depth of the base grid after drive in (J)

iii) Minimum separation between emitter and base grid diffusions (S)

iv) Minimum size contact for the emitter (E)

v) When a washed emitter process is used there is also an emitter misalignment tolerance (M)

The pitch is then W + 2J + 2S + E for a recut emitter or W + 2J + 2S + E + 2M for a washed emitter. Typical values are 9.5μm and 11.5μm respectively and again a tuned process would be expected to give a reduction in these figures.

The mesh transistor can be considered as a combination of both the interdigitated and overlay structures, as shown in Chapter 2. Pitch in one direction is governed by the metal line width and separation while the other direction is governed by the oxide windows and separation. Some structures may show an advantage approaching 2.2 times the emitter length of an equivalent interdigitated device.

The ideal cross-section of the metal base transistor is shown in Figure 5.2.3. The structure is that of an overlay transistor with the base grid diffusion replaced by a metallization. As with the conventional overlay, emitter pitch is determined by the base grid and emitter formation. This structure has the advantages of a base

grid which does not require a deep diffusion and is therefore narrower.

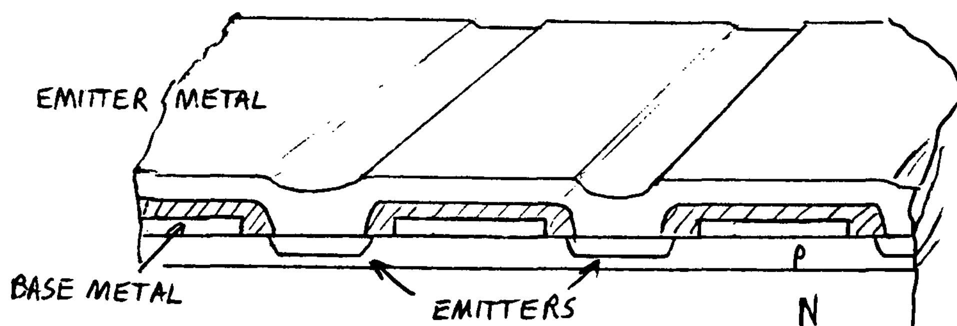


Figure 5.2.3. Metal Base Transistor

Washed emitter structures effectively auto-align the emitter contact and avoid the use of a misalignment tolerance, allowing the emitter size to be smaller. The novel structure of the metal base transistor takes this idea further by using the base metal to define the base ohmic contact and auto-align the emitter. Emitter pitch is, therefore, independent of the alignment accuracy.

The emitter pitch is determined by:

i) Minimum width of the base grid metal (B)

ii) Minimum emitter width (E)

iii) Minimum separation of emitter and base considering the sideways diffusion of the emitter and

encroachment of the base metal contact (J)

The pitch is then B + E + 2J, which could typically be 4μm, as shown in Chapter 2. This is an advantage of 2.5 over the conventional, washed emitter, interdigitated structure and 2.4 over the overlay structure.

It is clear that auto-alignment offers a reduction in device size. The metal base structure attempts to utilize this fact to the full. With the advent of sub-micron line widths it is conceivable that an emitter pitch of 2μm is achievable.

5.3 The Novel Metal Base Transistor

The properties required of the base metallization are:

i) Practical to deposit

ii) Able to be delineated to fine limits

iii) Have low resistivity

iv) Make good ohmic contact with silicon

v) Have good adhesion to silicon and silicon dioxide

vi) Be able to withstand the emitter fabrication

This last requirement implied a refractory metal was required. Properties of the metals considered are shown in Figure 5.3.1.

Element	Resistivity (Ωcm x 10^{-6})	Melting Point (°C)
Mo	5.7	2610
Ni	7.8	1453
Pd	10.8	1552
Pt	9.8	1769
Rh	4.7	1966
Ta	13.5	2996
W	5.5	3410

Figure 5.3.1. Properties of Refractory Metals

The lowest resistivity material is rhodium; however, it is expensive and was discounted on this basis. The next lowest resistivity is tungsten, which also has the highest melting temperature. It appears to be slightly more difficult to deposit than molybdenum [7], when deposited by sputtering, due to the high compressive stresses that develop. Molybdenum, however, oxidizes quite readily at room temperature, about three times faster than tungsten. It was felt, on balance, that tungsten offered the best advantages and was, therefore, selected. A number of multiple metal systems are reported using platinum silicide [147] and titanium silicide [4] as an interface between the silicon and tungsten, but there appears to be no advantage for contacts to heavily doped silicon.

Tungsten can be deposited in several ways: electron beam evaporation, chemical deposition and sputtering being three common methods. Delineation of high resolution patterns is achievable [2, 6]. Resistivity is about twice that of aluminium (2.74 x $10^{-6}\,\Omega cm$). Contacts to heavily doped silicon are ohmic [1, 2] and adhesion is good to silicon and silicon dioxide. The lowest melting eutectic in the tungsten-silicon binary system is 1400°C activation energy being one of the highest known for metals [2]. Additionally it does not suffer from electromigration and improved reliability has been reported [149] for power transistors using tungsten metallization for base and emitter contacts.

An as deposited resistivity of 25 x $10^{-6}\,\Omega cm$ has been reported [152] by RF diode sputtering, with reductions to 7.5 x $10^{-6}\,\Omega cm$ after short heat treatments above 950°C, specific contact resistance can be 6 - 9 x $10^{-5}\,\Omega cm^2$ for boron surface concentrations of 1.5 x $10^{-19} cm^{-3}$ [1] and 8 - 15 x 10^{-6} Ωcm^2 for 2 x $10^{19} cm^{-3}$ arsenic.

It appeared that tungsten films could be deposited relatively easily and also exhibit the desired properties. The likely specific contact resistance should be less than 1.5 x $10^{-4}\,\Omega cm^2$ which is quoted [1] for a boron surface concentration of 5.5 x $10^{18} cm^{-3}$ after a one hour anneal at 700°C. With a surface concentration of 1.5 x $10^{19} cm^{-3}$, 6 - 9 x $10^{-5}\,\Omega cm^2$ as deposited has been reported which

falls to 2 - 3 x $10^{-5}\Omega cm^2$ with a one hour anneal at 700°C.

A comparison can now be carried out between the conventional P+ grid and the proposed metal grid. With concentrations of 3 x $10^{20} cm^{-3}$, the solid solubility of boron in silicon at 1100°C, resistivity at room temperature is about 5 x $10^{-4}\Omega cm$ [158].

In a modern high frequency transistor the base junction depth would be about 0.5μm and the base grid would probably not extend much beyond this. An estimate made on the basis of a Gaussian and complementary error function profiles in a background doping of $10^{14} cm^{-3}$ (P-type in n-type) give average conductivities [8] of 6 x $10^2 (\Omega cm)^{-1}$ and 4 x $10^2 (\Omega cm)^{-1}$ respectively for surface concentrations of 3 x $10^{20} cm^{-3}$. For a 1μm junction depth the corresponding sheet resistivities are 16.7Ω/square and 25Ω/square. Surface concentration for a complementary error function is unlikely to be as high as 3 x $10^{20} cm^{-3}$ since the impurities diffuse away from the surface during the drive in. The estimate of the sheet resistivity for this diffusion is, therefore, an underestimate, 20Ω/square is a likely value. Resistivities for a P-type diffusion through the base well, Figure 5.3.2, is assumed to be similar. This can be justified since the base well has a small junction depth and a lower surface concentration than the base grid.

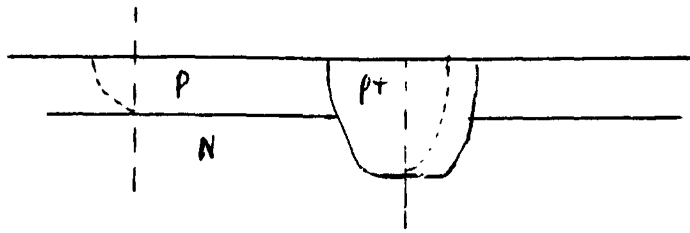


Figure 5.3.2. Profile of Base Grid

The thickness of a tungsten film required to give a sheet resistivity of 20Ω/square is 0.0125μm, assuming an as deposited resistivity of 25 x 10^{-6} Ωcm; more realistically it might be expected to be a third of this. On this basis a film thickness of about 0.05 - 0.1μm would be most acceptable, giving a sheet resistivity of typically 0.8 - 1.6Ω/square.

Contact resistance can be calculated, assuming a base sheet resistivity of 200 - 600Ω/square, from the equation given in Chapter 3. The attenuation constant is given by;

$$a = \left[\frac{R_{CS}}{R_M + R_D} \right]^{1/2}$$

where;

R_{CS} is the specific contact resistance, less than

$1.5 \times 10^{-4} \Omega cm^2$ (3×10^{-5} typical)

R_M is the metal sheet resistivity,

0.8 - 1.6Ω/square

R_D is the diffusion sheet resistivity,

200 - 600Ω/square

Calculations are given for the worst (typical) case specific contact resistance.

a = 8.64μm to 4.99μm (3.86 to 2.23μm typical)

If the base metal width, W_B, is taken as 2μm or less, then the w/a ratio is less than 0.2 (0.45 typical) and the contact resistance approximates, Appendix I, to;

$$R_C = \frac{R_{CS}}{WL}$$

where;

W is the metal width associated with the contact ($W_B/2$)

and;

L is the length of the contact

It is concluded that for small geometries the resistance of a tungsten contact to the base diffusion is determined by the specific contact resistance and the contact area only. On the assumption of: 2μm base metal, base sheet resistivity of 600Ω/square and emitter-base spacing of 1μm, the base resistance is 15.6kΩ/L (3.6kΩ/L typical), where L is the length of the emitter periphery in microns. Maximum current density in the emitter periphery is 60μA/μm which, assuming a current gain, β, greater than 20, corresponds to a base current density of less than 3μA/μm. Voltage drop in the base contact is then 46.8mV (11mV typical), which is acceptable.

Consideration must be given to the length of the base contact, Figure 5.3.3, the attenuation constant along the contact is given, Chapter 3, by;

$$a = \left[\frac{(R_b + R_j) W_B}{R_{MB} \times 2} \right]^{1/2}$$

where $R_j = V_T.L/I_{B1} \approx V_T/3 \times 10^{-6}$ giving a = 122μm to 173μm (86 to 122μm typical). Resistance for the base contact is given, Chapter 3, by;

$$R_{B1} = \frac{a2R_{MB}\coth(L/a)}{W_B}$$

and the ohmic part, R_{BB1}, is given by;

$$R_{BB1} = R_{B1} - V_T/I_{B1} = 333\Omega \text{ (}96\Omega\text{ typical)}$$

$$I_B R_{BB1} = 50mV \text{ (14mV typical)}$$

for L = 50µm. Under these conditions the specific contact resistance of the tungsten silicon interface dominates the determination of the base resistance. Maximum voltage drop in the base metal finger can be approximated, Appendix II, by;

$$V_{MB} = V_T \ln(\cosh(L/a))$$

$$= 1 - 2mV \text{ for } L = 50\mu m \text{ (2 - 4mV typical)}$$

for the operating conditions previously assumed.

As a comparison, consider the conventional overlay transistor. Assuming similar geometry and extrinsic base, R_B is simply the base spreading part, 600Ω, and R_{MB} is replaced by the diffused base grid sheet resistivity, 20Ω/square. The attenuation constant is 21µm, much smaller, and notice that it is determined by R_J, 8.3kΩ. Assuming L = 20µm (the emitter is twice this length) and

following the same analysis gives;

$$R_{BB1} = 108\Omega$$

$$I_B R_{BB1} = 6.5mV$$

$$V_{MB} = 10mV \text{ for } L = 20\mu m$$

It is evident that the metal grid has advantages over the diffused grid for making long base contacts with even distribution of current.

The ohmic resistance of the base is seen to be higher, typically a factor of two on the assumptions made here, due to the specific contact resistance of the tungsten-silicon interface. Higher contact resistance actually helps improve the current distribution, in fact it is like emitter ballasting except that it is in the base. Earlier it was shown that with β=20 the assumed conditions gave a worst case voltage drop, in the base resistance, of 46.8mV, or 11mV typical, in the base lead. It is easily seen that these figures are equivalent to an emitter series resistance of 1.8 x r_e and 0.4 x r_e respectively. The actual gm achievable would then be 0.36 x gm_0 and 0.71 x gm_0 respectively, where gm_0 is the gm available with no emitter series resistance. The typical figure is quite reasonable but is based on a conservative value of beta, namely 20, if a beta of 50 is assumed then these

figures improve to 0.58 x r_e and 0.86 x r_e respectively. In addition the equivalent emitter series resistance, due to the specific contact resistance of the base metallization, can be made smaller by making the base contact wider. Emitter pitch would then increase toward that achieved with the conventional structures. The width of the base electrode therefore has a considerable influence the characteristics of the transistor, and typically allows for a considerable advantage.

5.4 Emitter Fabrication

The emitters are required to be autoregistered by, but electrically isolated from, the base metallization (the main point of the novel structure). This requires two main processing steps, namely the emitter formation and the electrical isolation of the base metallization. As has been mentioned, the original concept of the transistor involved using the base metal directly as a diffusion mask, Figure 5.4.1. Further processing would then be required to widen the opening over the emitters. There are several good reasons why this is not a good way to proceed. The base metallization is in an unprotected state and, to prevent oxidation, must remain below a few hundred degrees centigrade if any oxygen is present. With no oxygen present the maximum temperature must remain below about 600°C [1] in order to prevent silicide formation at the metal-silicon interface. In this

structure the formation of silicide needs to be avoided as it would extend over the emitter-base junction, forming a short circuit.

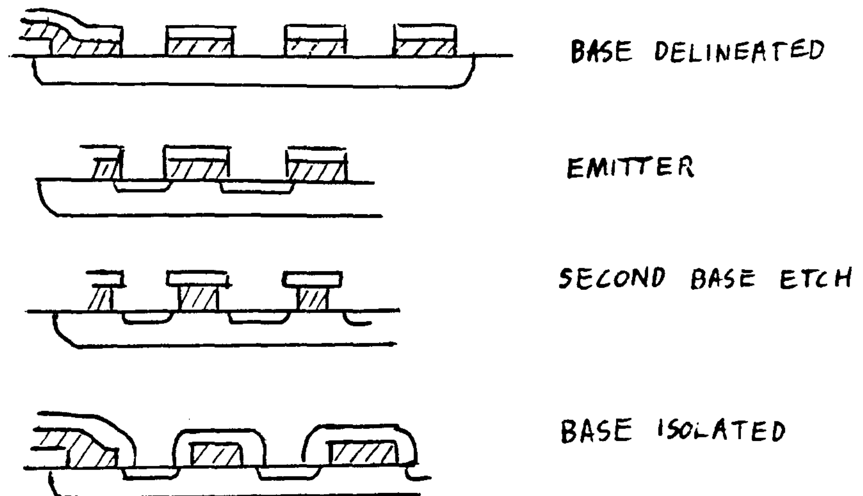


Figure 5.4.1. Original Scheme for Emitter Formation

Both problems can be avoided by ion implanting the emitter and annealing after the further etching of the metal. However, there is no need to carry out any additional etching. A major requirement of the structure is that the emitter must be narrow to give good high frequency performance, which in turn requires the emitter contact hole to be narrow and in registration. The implication is that the emitter contact can be used to register the emitter, since it must itself be registered by the base metallization. This is the case since it would not be possible to align the emitter contact with the, minimal width, emitter by conventional means. The point is, that the emitter contact must be formed anyway and if it is done before the emitter it obviates the need for a second etching of the base metallization, Figure 5.4.2.

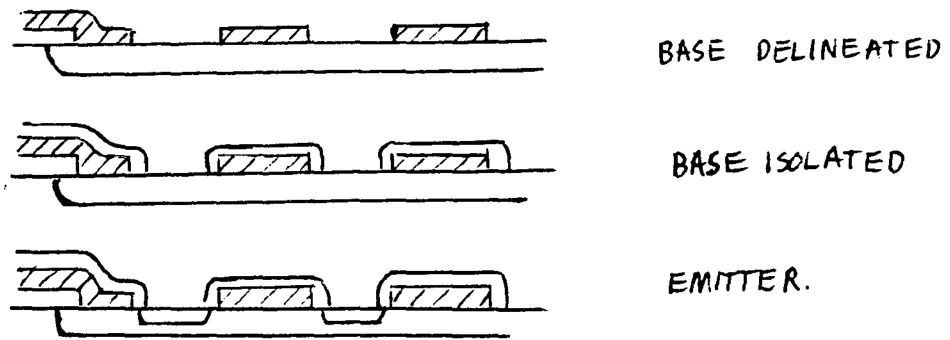


Figure 5.4.2. Revised Scheme for Emitter Formation

In addition the tungsten is covered during the emitter formation which lessens the process restrictions. In either eventuality the structure, prior to the emitter metallization, requires that both the top surface and the side-walls of the base metallization are coated with some form of insulator. This structure must, by necessity, be achievable without recourse to a mask for emitter window definition in order to obtain a minimal emitter pitch. The base metallization and the overlapping surface topography must be used to define the emitter windows directly. One possible procedure is to deposit a conformal layer of some insulator over the delineated base metal and then carry out an anisotropic etch, Figure 5.4.3, by means of plasma or ion beam enhanced etching. It can be seen that the covering of the side-walls presents a greater thickness to the etching process and in consequence may be retained if the etch duration is

correctly timed. The insulator above the base metal will be etched at a similar rate as the insulator over the emitter sites and must, therefore, be thicker.

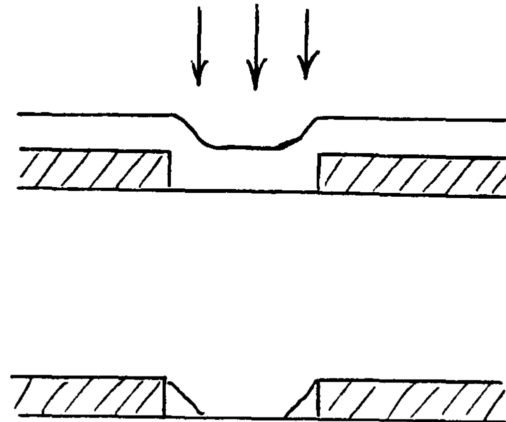


Figure 5.4.3. Anisotropic Etching

Such a structure could be achieved by covering the base metal with an insulator prior to delineation, which is then carried out by first delineating the insulator and then using this as a secondary mask to etch the metal. A conformal coating may then be applied and anisotropically etched, Figure 5.4.4.

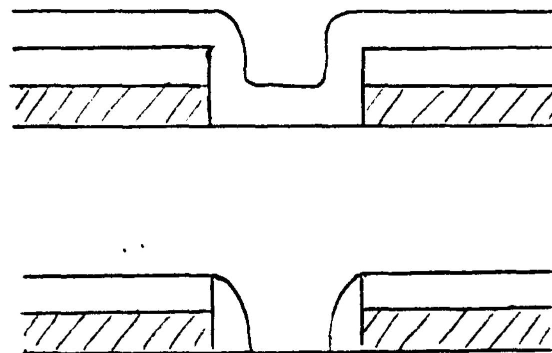


Figure 5.4.4. Modified Structure Prior to Etching

A further possibility is to undercut the metal by etching as shown in Figure 5.4.5, although the step must be suitable for continuous metal coverage by the second layer metal.

Figure 5.4.5. Undercut Metal

Both the structures in Figures 5.4.4 and 5.4.5 would be created before the emitter was fabricated. There is, however, a further alternative which may be used on the second structure. The top coat must be silicon dioxide and the lower one nitride. The emitters are then formed by ion implantation before the emitter windows are opened. Silicon dioxide over the emitters and over the nitride will suffer damage from the high energy particles, whilst that covering the metal side-walls will not since it is sheltered by the overhanging nitride. The faster etch rate of implant damaged silicon dioxide [31] can then be employed to open the windows by a conventional wet etch. Any wet etch must be carried out prior to the emitter being annealed since the differential etch rate will be

lost in the process.

5.5 Proposed Process Schedule

Five masks are required for manufacturing the novel transistor, although the final passivation mask may be omitted if a suitable hermetically sealed package is used.

I) Base well definition

II) Base contact definition (Metal 1)

III) Via (Metal 1 to Metal 2)

IV) Emitter contact definition (Metal 2)

V) Passivation (windows over bond pads)

The process sequence, Figure 5.5.1, commences with an initial oxidation. Base wells are defined by Mask I and the base formed by diffusion or ion implantation in the conventional manner. Base metallization is followed by a nitride or oxide coat, both are then delineated using Mask II. A conformal coating of oxide is then applied and anisotropically etched to open the emitter sites, emitters are then fabricated. The area over the metal 1 bond pad site is opened using Mask III. Aluminium or other metal 2 is deposited and delineated by Mask IV in the conventional

manner. Finally passivation is deposited and windows for the bond pads opened by use of Mask V.

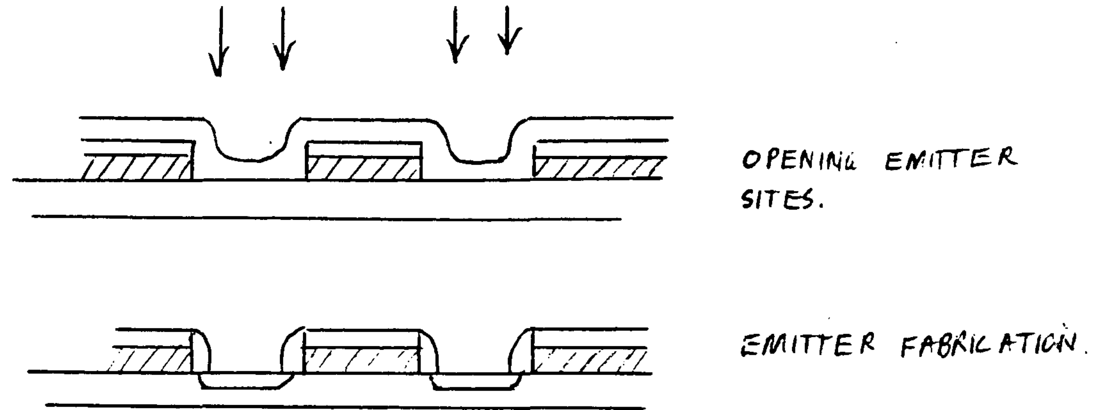


Figure 5.5.1. Process Sequence

CHAPTER 6 - EXPERIMENTAL WORK

6.1 Practical Considerations

This chapter describes investigations into the processing steps required to make the device proposed in Chapter 5. A set of masks made for this purpose are described and results obtained from the test structures presented.

The setting up of a tungsten metallization system is described and results are presented for deposition rate and film characteristics. High temperature processing of the tungsten film was investigated along with the deposition and use of various insulating layers.

Both silicide and ohmic contact formation at the silicon-tungsten interface were evaluated and results are given here.

Finally the problems involved in fabrication of the emitter after the base metallization was in place have been investigated.

6.2 Mask Set

A mask set was designed and made to enable investigation into various process steps. The set comprised five masks:

1) Base definition

2) Metal 1

3) Emitter definition

4) Base contact

5) Metal 2

The masks were hand cut in Rubylith, a cut and strip film, at 200x magnification. First photoreduction was 20:1, followed by a second 10:1 photoreduction at the step and repeat stage. Details of this are given in Appendix X. Finished masks are on 2.5" glass plates with 31 rows of 25 exposures, giving a total of 775 die per wafer. The die size was 0.035" x 0.044".

Structures comprised:

1) Alignment marks

2) Resolution test pattern

3) General purpose test grids

4) Resistivity measurement structures

5) Interdigitated transistor structure

6) Novel overlay transistor structure

The alignment marks allowed either light or dark field masks to be used as required. These are shown in Figure 6.2.1.

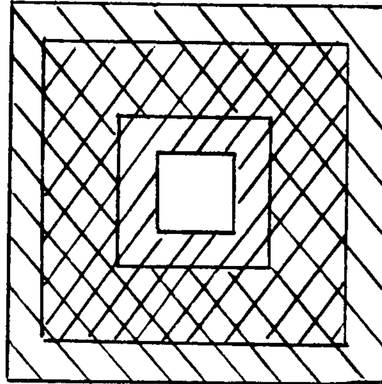


Figure 6.2.1. Alignment Marks

The resolution test pattern Figure 6.2.2, was used to check the photographic resolution and the photoengraving resolution.

200x

Figure 6.2.2. Resolution Test Pattern

Two general purpose test grids, Figure 6.2.3, enabled evaluation of the first metal photoengraving, step coverage and process results of the overlaying deposited insulator.

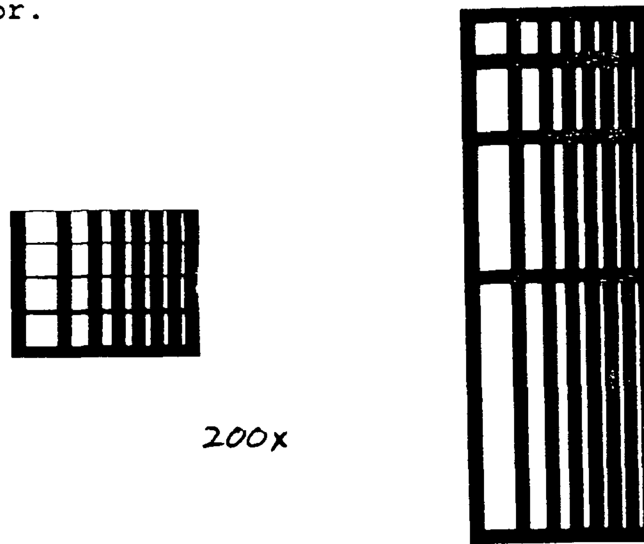


Figure 6.2.3. General Purpose Test Grids

Resistivity structures, Figure 6.2.4, enabled the sheet resistivity of n+, p+ and metal 1 to be evaluated along with contact resistance to metal 2 and also p+ to metal 1.

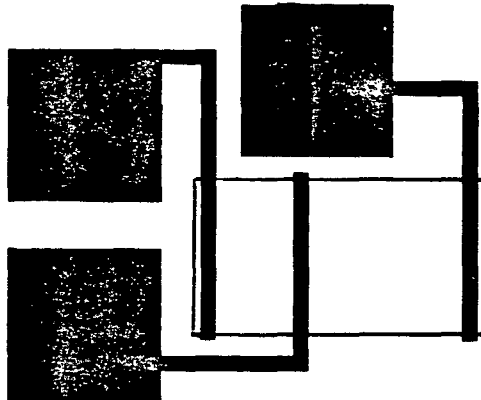


Figure 6.2.4. Sheet Resistivity and Contact

Resistance Structures

Care was taken to make contacts of equal size and also to match the series resistance of the probe pads.

A simple interdigitated transistor was included, Figure 6.2.5. This used quite a large feature size, 10μm contact, consistent with the available equipment.

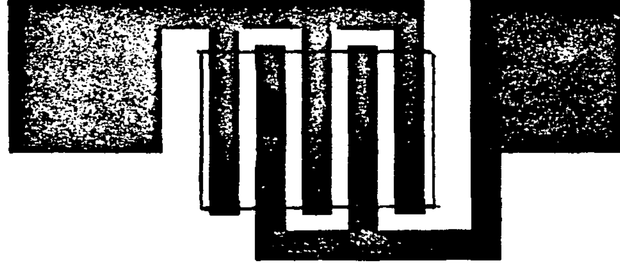


Figure 6.2.5. Interdigitated Transistor

A single cell overlay transistor, Figure 6.2.6, completed the structures on the process evaluation die.

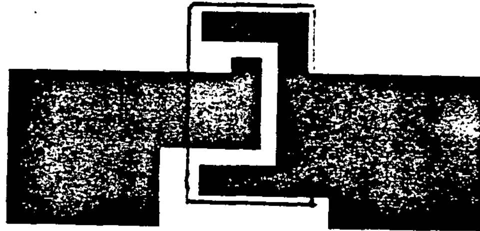


Figure 6.2.6. Overlay Transistor

Process steps that involved these masks are illustrated in Figure 6.2.7. Mask 1 was used to define the base wells, which was then followed by the base diffusion. The base wells were then deglassed and the first (refractory) metal deposited, mask 2 having been used to delineate this. An insulator was next deposited and mask 3 was used to open the emitter windows. Subsequent emitter diffusion was followed by either washing out the emitter windows or recutting them with the use of mask 3. Base contacts were then cut, using mask 4, and this was followed by the second metallization, which was delineated by mask 5.

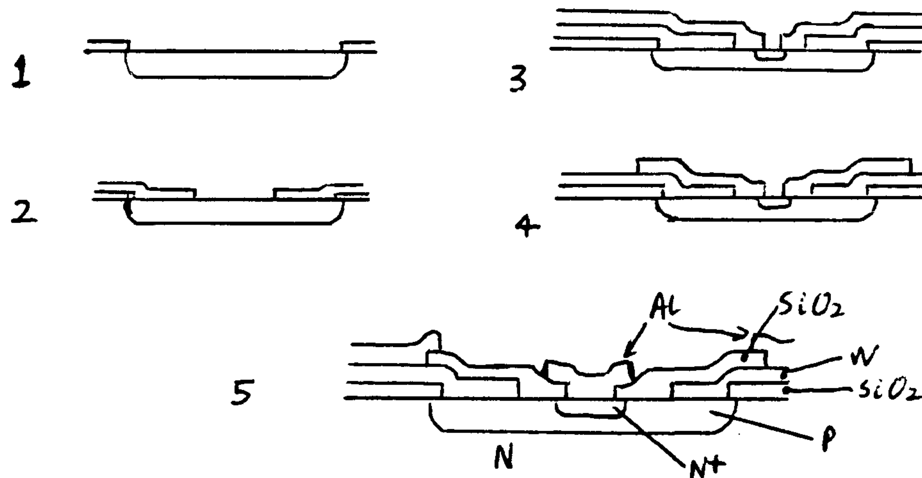


Figure 6.2.7. Process Steps

6.3 Deposition of Tungsten

There were several deposition techniques suitable for the deposition of tungsten metal, these included vapour phase deposition, electron beam evaporation and sputtering.

Chemical vapour deposition appeared to have an advantage,

as reported in [2]. Using this technique the native oxide covering the oxide windows can be etched off prior to the deposition phase. A similar process may be carried out with sputtering. In this case the top surface of the doped silicon may be removed to reach the more highly doped region below the surface. This happens due to the segregation of the dopant (boron) into the oxide. In each case the oxide may be removed and the metal deposited without allowing an oxidizing ambient to reach the silicon surface. This is important since the metal-silicon contact is affected by even small amounts of oxide. Electron beam evaporation would not allow this procedure as standard. The most suitable equipment available for the research work presented here was a high frequency sputtering system, Edwards E19E with an EHW5/1RF generator. Initially this was a dual electrode machine, suitable for sputtering dielectric material, which used a pair of concentric disc and annulus targets.

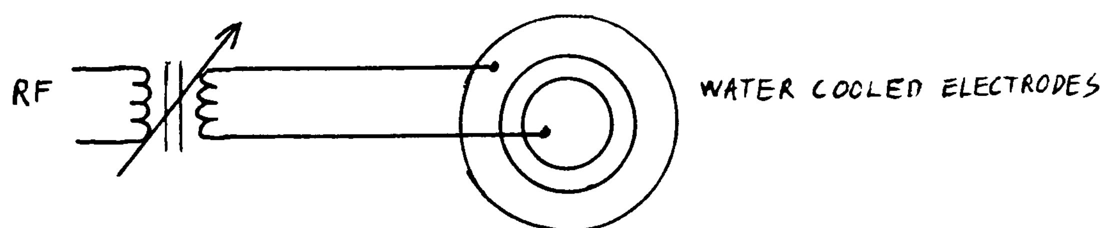


Figure 6.3.1. The Dual Target Scheme

Due to the cost of targets it was decided to use a 2" diameter, one eighth inch thick target, 99.95% purity. This entailed modifying the electrical circuit of the sputtering machine.

The high frequency generator used a transformer output, power control being achieved by means of a mechanical linkage which allowed the, transformer, coupling to be varied. Initially this was configured as an isolated balanced output, Figure 6.3.1, to drive the dual, dielectric, targets in a balanced fashion. This was modified to a grounded single ended drive, Figure 6.3.2, in order to drive the, single, tungsten target.

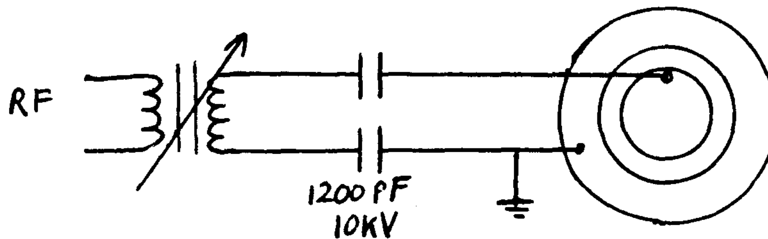


Figure 6.3.2. Single Target Scheme

The tungsten target was supplied on a copper backing plate and had to be attached to the disc electrode so as to make good thermal and electrical contact. The simplest method was by means of a silver loaded epoxy resin. Both target and electrode were prepared with wire wool, to clean off

any oxide and form a key. Both parts were then degreased with acetone, which was wiped dry with clean paper wipes and not allowed to dry solely by evaporation. Silver loaded epoxy resin was prepared and applied to both surfaces. The target was then placed in the vacuum chamber, which was then pumped down to de-aerate the resin. After this the resin was smoothed down with a spatula and the target rubbed into place. The assembly was then held in position while the resin cured.

During sputtering a plasma forms around the target and the disc electrode and, if allowed, it would sputter the electrode. In order to prevent this the target had to be screened in the places other than where a plasma was required.

The curious fact, at first sight, about plasma is that it will form only if the electrodes are greater than a certain distance apart. This distance being the Crookes dark space which forms between the cathode and the surrounding discharge glow at low pressures. To prevent sputtering occurring in specific areas it is necessary to reduce the electrode spacing in these areas to less than the size of the dark space.

For the system used, i.e. the pressure of the gas, a distance of a quarter of an inch was small enough to prevent plasma formation. This dimension was therefore

used for the separation of the screen from the electrode assembly. The screen was fabricated in stainless steel in such a way that it fitted inside the annular electrode, which was grounded, and enclosed all of the circular electrode except the front face of the tungsten target. A cross-sectional view of the assembly with the fabricated screen is shown in Figure 6.3.3.

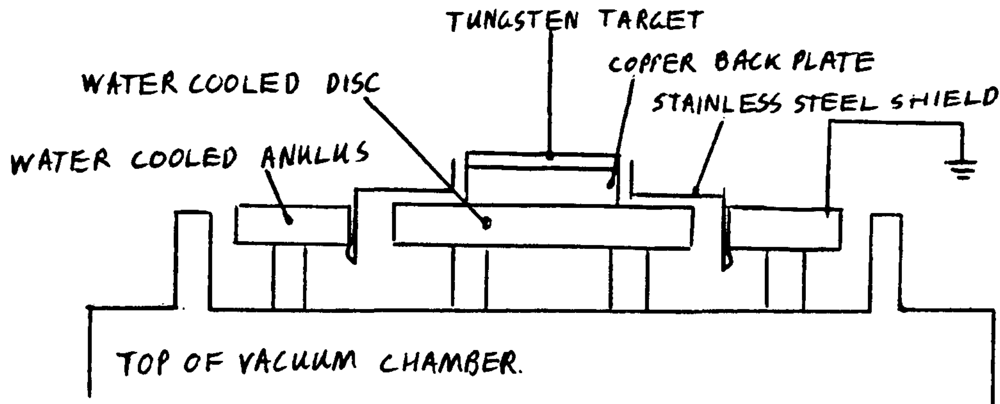


Figure 6.3.3. Cross-section of Target Assembly

Initially adhesion of the tungsten film was found to be a problem, but then it was discovered that heating the substrate was advantageous. A heater was made to run from a low voltage power supply available on the equipment, Figure 6.3.4. Heating elements were formed in nichrome wire of suitable resistance, then supported and isolated in glass tubing. Six of these were held in a frame with suitable electrical contacts. The heated wafer support was three inches square and placed about two inches below the target.

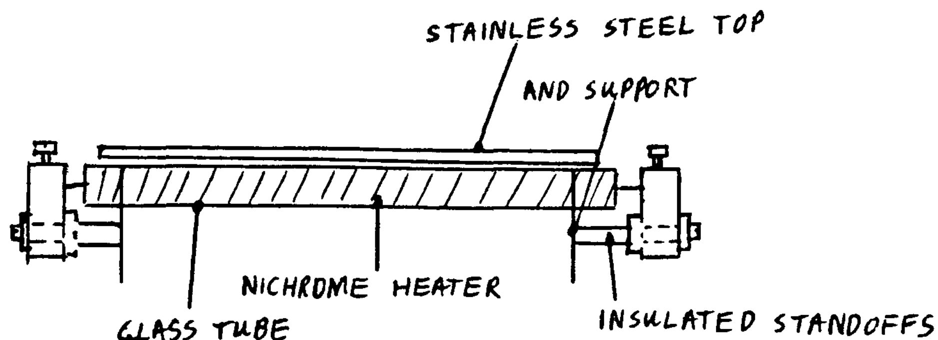


Figure 6.3.4. Heater Assembly

After some experimentation a reasonably reliable procedure was established to enable an 0.2μm layer to be deposited in about twenty minutes, Appendix XI. This procedure was found to be repeatable but there were reliability problems.

Often the film was seen to ripple or erupt into small flakes when the chamber was opened. It was discovered that; as deposited the tungsten films were under compressive stress and they were thus in an unstable equilibrium. Any irregularities caused stress concentrations. Under these conditions the film either fractured, in which case a small flake broke off, or it formed a ripple which propagated across the wafer, Figure 6.3.5.

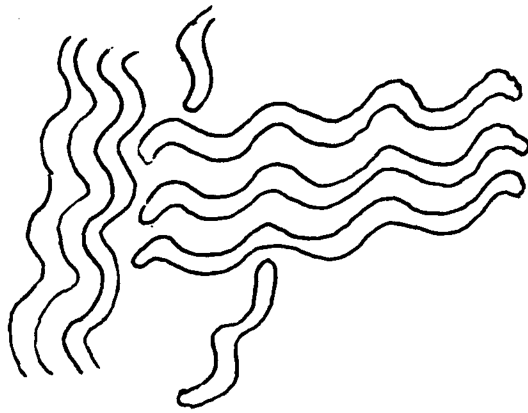


Figure 6.3.5. Ripples in the Tungsten Film

Deposition rate has been plotted here as a function of the anode current of the HF generator, Figure 6.3.6. Variation of film thickness across the slice could be as much as 100%. This was due to the small size of the tungsten target and the small separation from the wafers to be coated.

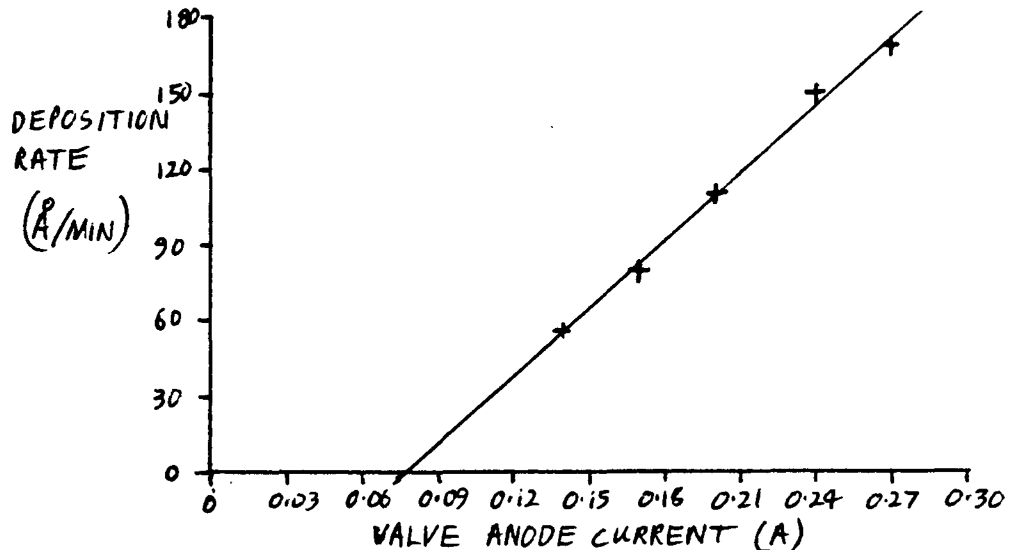


Figure 6.3.6. Deposition Rate

Sheet resistivity was found not to be inversely

proportional to the film thickness, Figure 6.3.7, which tends to indicated that the first strata of the films to be deposited had higher resistivity than later strata.

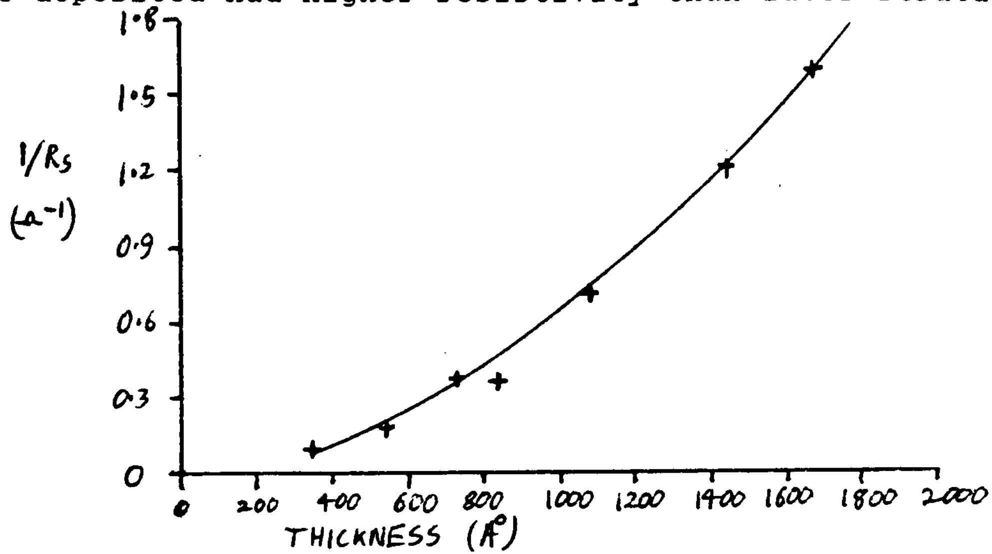


Figure 6.3.7. Sheet Resistivity

This could be due to heating of the slice which occurs due to kinetic energy transfer during the initial phase of the sputter deposition. Average resistivity decreased with an increase in film thickness, as shown in Figure 6.3.8.

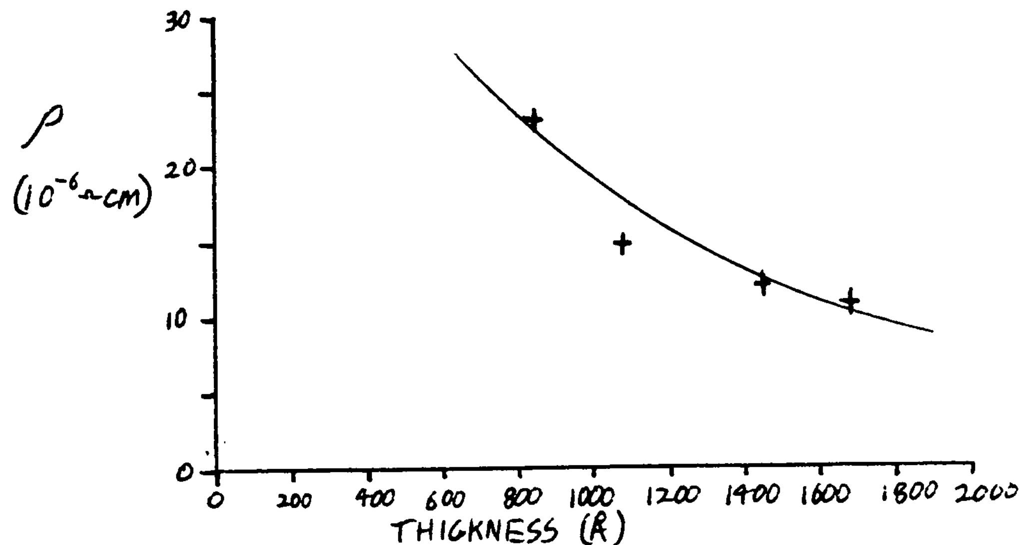


Figure 6.3.8. Average Resistivity

6.4 Etching of Tungsten Film

Initially, delineation of the tungsten film was attempted with a solution of potassium ferricyanide and potassium hydroxide [6]. Used with Kodak negative photoresist it was found that, although the etching was reasonable, it loosened the photoresist. Other etchants were tried, such as hydrofluoric acid and nitric acid mixture, and also undiluted hydrogen peroxide. All caused the resist to lift. An adhesion promoter hexamethyldisilazane (HMDS) was investigated but this gave no improvement. Shipley positive resist was later found to adhere well and to allow reliable delineation. This required a new mask to be made since the previously used resists had been negative.

Of all the etchants tried hydrogen peroxide was found to be best, and easiest to prepare. Etch rate was about 160Å per minute at room temperature and higher at elevated temperature. Gaps of about a micron could be etched in 0.2μm films. Results were generally limited by the resolution of the contact masks and were adequate for the feature size of the test masks.

The major problem found was the residual compressive stress in the sputtered tungsten film. The rippling has already been described. Etching of the film was sometimes able to act as a catalyst for this.

Attempts were made to anneal the film to prevent this problem. It was considered that the combination of silicide formation, which would be expected to improve the adhesion of the film, and stress relief in the film would improve the film stability. Annealing at 600°C in a reducing atmosphere, necessary to prevent the tungsten oxidizing, was found to have no effect. A temperature of 800°C was sufficient, samples that had been raised to this temperature for about five minutes were never seen to ripple at a later stage.

High temperature processing subsequent to tungsten deposition was required for the emitter fabrication and silicide formation for ohmic contact in addition to any annealing. Emitter formation was initially intended to be by diffusion. This, however, required an oxidizing ambient but the tungsten oxidizes readily at about 300°C and so some barrier would have been required to keep all oxygen away from the tungsten, this barrier would have been later required to separate the emitter and base metallizations, Figure 5.2.3.

In order to investigate this possibility three dielectrics were tried: silicon dioxide (silox), silicon nitride and spin-on silica film. Eventually it was found that all three could be deposited onto a tungsten film.

The silox deposition was carried out by a reaction between

silane and oxygen, this deposited silox on to the wafer, which was heated to 400°C. When carrying out this deposition it was found that the normal procedure could not be used as it caused the tungsten to oxidize. A modified procedure was devised, as follows.

The susceptor block must be cold when the wafer is placed on it. This is because the chamber is open to atmosphere and oxygen is present. The chamber is sealed and a nitrogen ambient is then introduced. The susceptor block is then raised to the required deposition temperature. With the deposition temperature prevailing, silane is introduced, but oxygen cannot be allowed in at this stage. The nitrogen gas flow is then switched off and the flow of silane is allowed to flush the chamber free of the residual gas. Only now can the oxygen be let in since it will now react with the silane before it can reach the tungsten. After deposition the susceptor block should be allowed to cool to about 100°C before the chamber can be opened.

As deposited the silox had low density; it etched with poor definition and was also porous. To be of use it needed to be densified at 800°C to prevent oxygen, if present in the gas ambient, reaching the tungsten surface. If oxygen reached the tungsten surface the resulting oxide growth tended to crack the silox coat, allowing a faster reaction rate. Silox could be densified with some success

in a nitrogen ambient, which had small traces of oxygen, for ten minutes. Longer times in the furnace resulted in the loss of all metal due to oxidation. It appeared that the silox was never dense enough to protect the metal surface completely.

Later work was carried out in a silox reactor of larger capacity. The substrate heater had a large thermal mass which was inconvenient for cold loading. In an effort to circumvent this problem, hot loading was attempted while flooding the area with nitrogen. This method was never satisfactory. The only way was to load the chamber while it was cold and then heat it with a nitrogen ambient present.

Nitride deposition was carried out at atmospheric pressure in a furnace. Wafers were loaded into a boat in the usual manner and then put into the mouth of the furnace. This particular furnace would be open at the mouth during loading. To avoid back diffusion of oxygen into the furnace the nitrogen flow rate was turned as high as possible while wafers were loaded into the mouth of the furnace and then pushed slowly in. This coupled with a narrow, three inch, tube and low temperature, of 800°C, prevented back diffusion from being a major problem.

The process was not free of problems however. Microscopic examination of the nitride film revealed a number of small

circles scattered about. These sometimes appeared as conical eruptions in the nitride deposition and it was under these areas that the tungsten had oxidized. With longer deposition times the circles were found to be larger. The nitride formed was of good quality and generally sound.

The mechanism responsible for the circular features was eventually determined. In the first place the nitrogen used (BOC white spot) contained minute traces of oxygen and/or water. This was later proved to be the case in a better furnace, which had an "elephant" for loading the boats. A tungsten coated slice was loaded into this furnace with adequate purging and raised to 800°C for 20 minutes. Subsequent microscopic examination showed a number of small areas where oxidation had occurred. The sites were small and widely spaced. This formation of small, well spaced, areas of tungsten oxide is the first part of the mechanism. Wafers entering the nitride furnace would therefore have had a small number of minute oxidation sites formed as the first layer of nitride was deposited. Once a nitride layer had formed it had sufficient integrity to prevent further oxygen reaching the tungsten surface. The small oxidized sites, however, formed irregularities which the nitride could not seal. Further oxygen was then able to penetrate and cause further oxidation. The oxide, being of larger volume, lifted and cracked the nitride film in an expanding

circle. The process was relatively slow, about 2×10^{-3} inch/minute (about 8×10^{-4} mm/S), due to the combination of low temperature and low concentration of oxygen. It was expected that a purer nitrogen supply would remedy this problem.

Spin-on silica film is a suspension of silicate in an alcohol carrier. It was applied at low temperature to 800A film of tungsten and densified at about 250°C. Some problems were found with the film. Reactions started from small sites and spread out into larger circles that increased with time, similar to the silox results. Microscopic examination of the silica film after deposition showed that foreign bodies were the problem. Where silica film covered a dust particle it thinned and on densifying was seen to crack in the vicinity. Further layers of silica film did not rectify the situation.

These breaches in the silica film formed the sites where tungsten oxidation could occur. The silica film appeared to be dense enough to prevent oxygen from reaching the metal surface. When applied to patterned metallizations, step coverage was seen to be good.

It was found that all three materials could be deposited on to tungsten films with reasonable success. All three could be raised to 800°C in contact with a tungsten film, in a non-oxidizing ambient. Nitride and silica film could

be used at elevated temperatures in an oxidizing atmosphere. Both had sufficient integrity to prevent oxygen reaching the tungsten surface in the absence of any defects. Any breach in the film, however small, would quickly propagate across an underlying tungsten film. This mechanism could be restricted by delineating the tungsten into the required pattern. Loss was then restricted to metal areas that initially contained the film defect. Thus for the best yield the insulating film should be deposited after the tungsten has been delineated, which should itself be done after the annealing.

Annealing of the uncovered tungsten film can be carried out in a forming gas (90% nitrogen, 10% hydrogen) ambient. The hydrogen reduces any tungsten oxide back to tungsten, giving off water. The majority of the early work carried out on the high temperature processing investigation was problematic due to back diffusion of oxygen into the furnace. These furnaces were of the open end type and a great deal of effort was directed to avoiding back diffusion, although at that time it had not been proven that this was the problem. Using the highest gas flow rates during loading gave some success, very slow and careful loading, to avoid turbulence, gave no advantage. Partial capping also gave no effect. Later work used a furnace with an "elephant" i.e. a semi-sealed loading tube. Results with this were excellent since back diffusion into

the furnace could be eliminated. For best results the slices should be loaded, well forward, in the elephant which is then placed over the furnace mouth. The furnace is then allowed to purge for long enough to sweep all of the, original, air out of the elephant. Only then can the boat be pushed to the furnace mouth and then into the furnace centre. An interesting example occurred when a defective elephant was used. This was chipped such that it left a small hole, about 1cm square, at the mouth of the furnace. This small hole was sufficient to thwart every attempt to avoid oxidizing the tungsten film.

It cannot be over emphasized as to the importance of keeping oxygen away from the hot tungsten surface.

6.5 Making Ohmic Contact

To investigate silicide formation at the silicon-tungsten interface wafers were prepared and then analysed by means of Rutherford back scattering (RBS) at Surrey University. Where appropriate the processing follows the Bipolar III schedule, Appendix XII. The starting material was 1 to 10Ωcm, n-type, <111> silicon of 1.5" diameter. A boron diffusion was made over the whole surface. Junction depth was 3μm and sheet resistivity was 70Ω/square. Native oxide was removed by an HF dip followed by a rinse in de-ionized water and then blow dried with hot nitrogen. Sputter deposition of 0.16μm of tungsten followed

immediately. One wafer was retained and the rest had 0.3μm of silox deposited. This followed directly after removing the wafers from the sputtering chamber. Annealing at 800°C for 20 minutes in a nitrogen ambient followed. In addition one and a half minutes were taken to push the boat to the centre of the furnace and one and a half minutes to remove it afterwards. A sample was retained and the remaining wafers were stripped of silox. A further sample was retained and the last wafer was stripped of metallic tungsten in hydrogen peroxide. All samples were then diced into 5mm squares suitable for the RBS jig.

The samples that were analysed were tungsten in silicon prior to annealing and the annealed samples with silox and tungsten removed. The first gave results as shown in Figure 6.5.1.

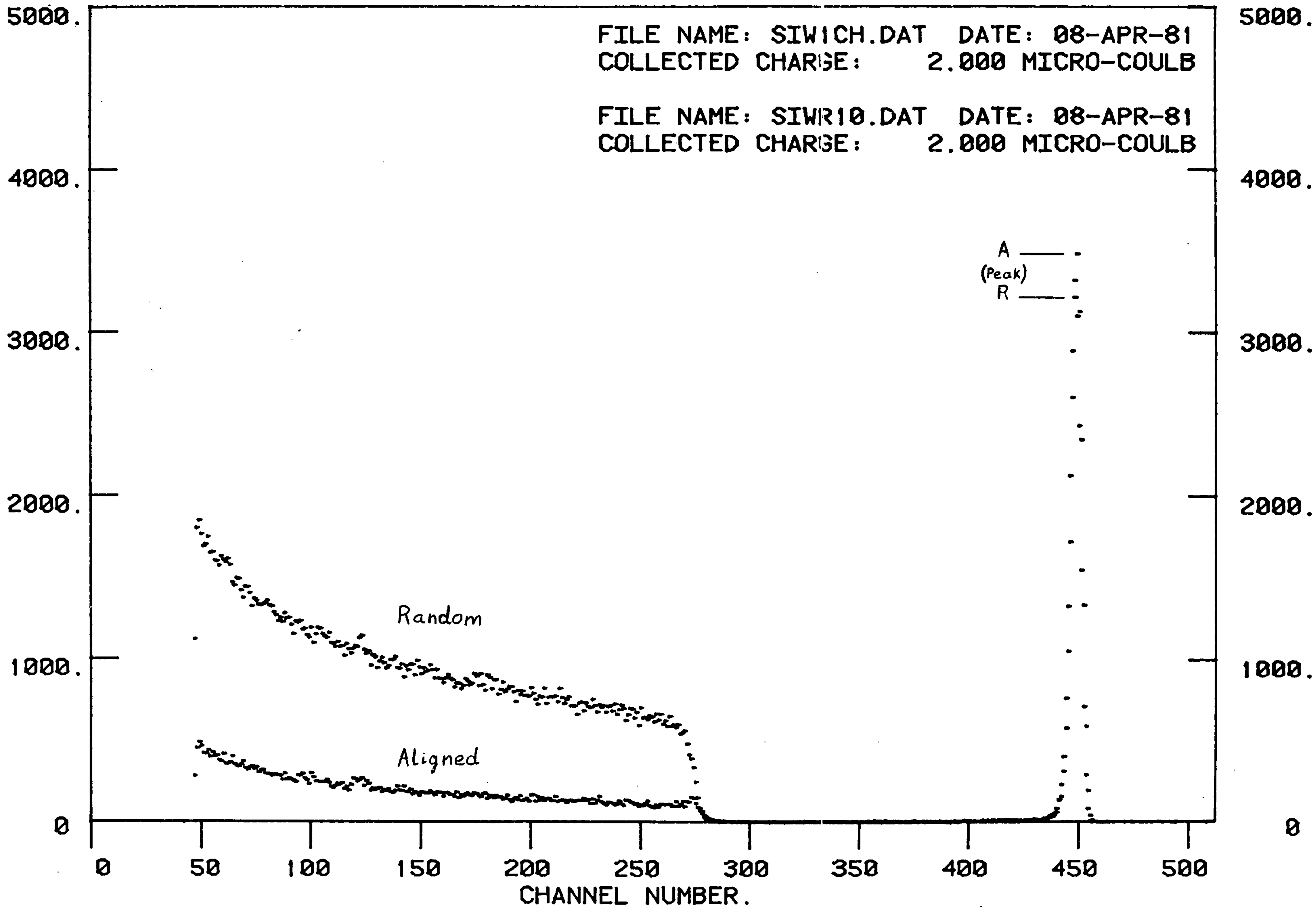

FILE NAME: SIWICH.DAT DATE: 08-APR-81
COLLECTED CHARGE: 2.000 MICRO-COULB
FILE NAME: SIWR10.DAT DATE: 08-APR-81
COLLECTED CHARGE: 2.000 MICRO-COULB
A
(Peak)
R
Random
Aligned
5000.
4000.
3000.
2000.
1000.
0
0
50
100
150
200
250
300
350
400
450
500
CHANNEL NUMBER.

Figure 6.5.1. RBS Plot of Tungsten Film on Silicon Before Annealing

These plots show the number of back scattered particles against their energy. Two plots are shown, the larger is for random scattering of the particles and the smaller for the crystal lattice aligned with the particle beam direction. The ratio of these giving a measure of the crystallinity of the sample. When aligned, a crystal has less chance of scattering a particle because the atoms are effectively hidden behind each other, Figure 6.5.2.

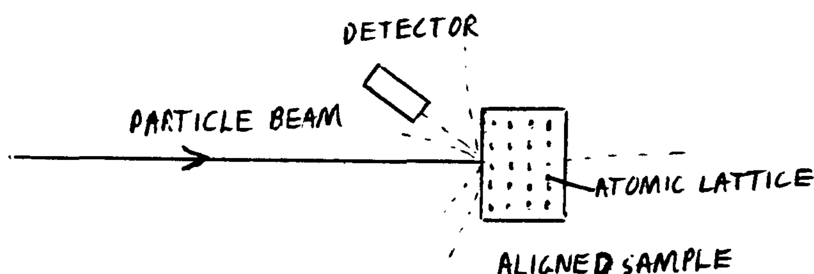


Figure 6.5.2. Rutherford Back scattering

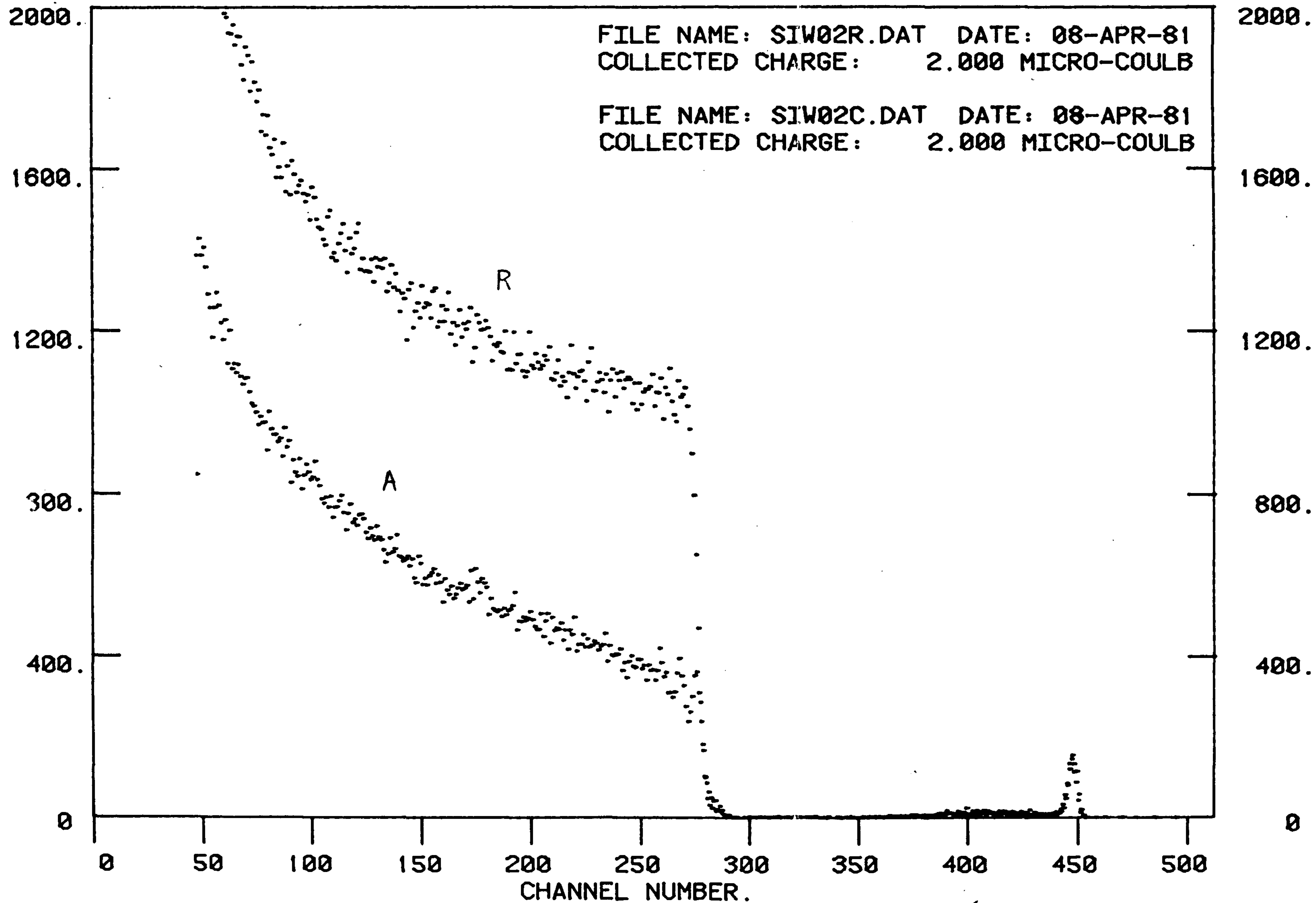
From Figure 6.5.1 the sample can be seen to consist of two parts, the plateau to the left is caused by the silicon and the sharp peak to the right is due to the tungsten film. The energies of the scattered particles are characteristic of these elements. A gap appears between the silicon plateau and the tungsten peak due to the different atomic weights of the materials. In addition the silicon is crystalline whereas the tungsten is not. Notice that the plateau to the left has risen for lower energies. This is because the probability of a collision

increases and the energy decreases as the path length of the particles inside the sample increases. The silicon is thick compared to the tungsten.

The results for the annealed sample are shown in Figure 6.5.3.

The vertical scale has altered but the silicon plateau is much the same as previously. The tungsten peak is smaller, which was not surprising since all the metallic tungsten had been etched off this sample. Notice that just to the left of the tungsten peak is a small hump, shown expanded in Figure 6.5.4, which was due to the silicide formed at the interface during annealing.

The peak of this hump occurs around channel 400 and corresponds to a depth of 0.1 to 0.2μm beneath the silicon surface. The tail of this diffusion may extend twice this distance. The silicide does not appear crystalline so the tungsten atoms are not generally occupying silicon lattice sites. Note that this did not mean that the silicide was not crystalline but that any silicide lattice was not aligned with the silicon lattice. The amount of tungsten represented by the two peaks amounts to some 8% of the first sample's metallization, there being 4% being at the surface and 4% in the silicide layer.



FILE NAME: SIW02R.DAT DATE: 08-APR-81
COLLECTED CHARGE: 2.000 MICRO-COULB
FILE NAME: SIW02C.DAT DATE: 08-APR-81
COLLECTED CHARGE: 2.000 MICRO-COULB
R
A
2000.
1600.
1200.
300.
400.
0
2000.
1600.
1200.
800.
400.
0
0
50
100
150
200
250
300
350
400
450
500
CHANNEL NUMBER.

Figure 6.5.3. RBS Plot of Tungsten on

Silicon After Annealing

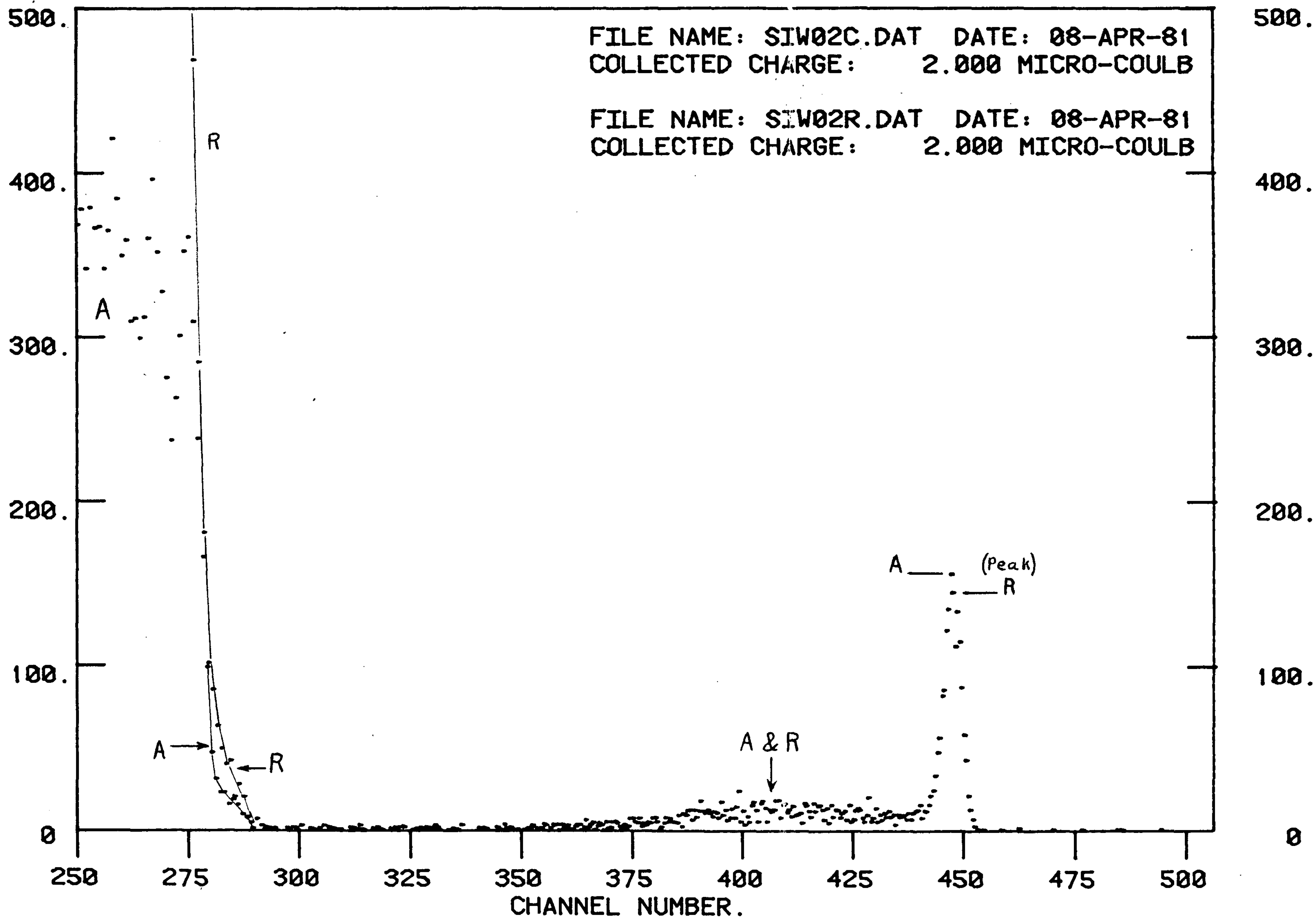

FILE NAME: SIW02C.DAT DATE: 08-APR-81
COLLECTED CHARGE: 2.000 MICRO-COULB
FILE NAME: SIW02R.DAT DATE: 08-APR-81
COLLECTED CHARGE: 2.000 MICRO-COULB
500.
400.
300.
200.
100.
0
250
275
300
325
350
375
400
425
450
475
500
CHANNEL NUMBER.
A
R
A & R
(Peak)

Figure 6.5.4. RBS Plot of Tungsten Silicide

Some RBS plots show peaks at around channel numbers 120, 180 and 280, which indicate traces of oxygen and carbon at the surface. It is not clear if the plots shown here have peaks at these places.

CHAPTER 7 - EXPERIMENTAL MEASUREMENTS

7.1 Terminations

Device characterization should be carried out in a circuit configuration which reflects the needs of the application. For the case of UHF medium power transistors the need is best met by s-parameter measurements taken on a common emitter configured transistor. The 50Ω terminations used for this type of measurement simplify the impedance matching at both the input and output. Open circuit termination, required for h-parameter measurement are difficult to achieve at higher frequency. The y-parameters require only short circuit terminations, which are easier to implement, but these too are difficult to achieve accurately at higher frequencies. This can lead to inaccurate measurements and instability, oscillation, in the device under test.

Where the device is to be characterized for distortion performance terminations other than 50Ω may be required. Additionally the signals which characterize the distortion of the transistor can be small compared with signals caused by distortions in the measuring equipment. These problems are discussed and the setting up and verification of the measuring system is described. Finally measurements are presented and conclusions drawn.

7.2 Intermodulation Measurement

The most appropriate measure of distortion is that which directly relates to the particular application. In most cases this will be the third order modulation of type: 2a - b or a + b - c, where a, b and c represent frequencies at the input and the expressions represent frequency components at the output.

The basic experimental equipment to carry out this kind of measurement is shown in Figure 7.2.1.

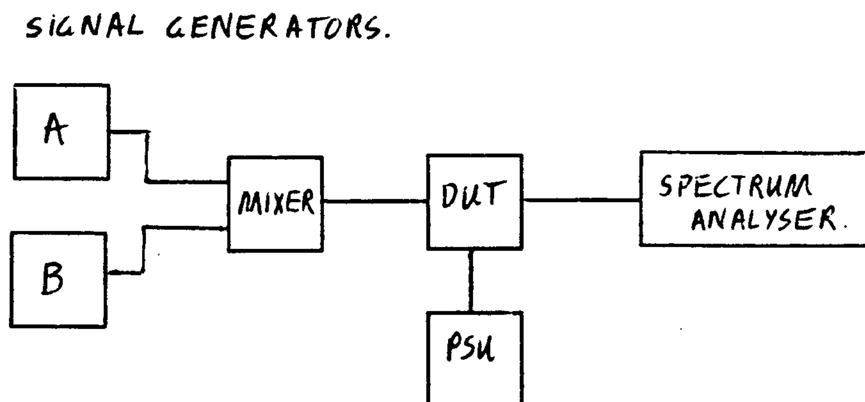


Figure 7.2.1. Distortion Measurement Configuration

This configuration which is applicable for both high and low frequency measurement. For high frequency measurements a 50Ω coaxial system is used with suitable jigs for the device under test.

A system of this type was set up for the measurement of distortion.

The equipment used was as follows:

Signal generators:

Marconi HP3654 Function generator

Spectrum analyser:

Marconi 110MHz TF 2730

Marconi 50Hz - 200MHz TK2374 zero loss probe

Mixer:

Mini circuits ZSC 4-3

Power supplies:

Thurlby PL320

Farnell L30B

Meters:

Fluke 8050A

Network analyser:

HP3577A 5Hz - 200MHz network analyser

HP35677A 100kHz - 200MHz S-parameter test set

The equipment was set up for two tone distortion measurement, with inputs at 20MHz and 22MHz. Initial measurements were then made on the system alone to verify the system performance. In order to do this the test jig was replaced by a connector, i.e. a very linear device, and a set of measurements taken, Figure 7.2.2.

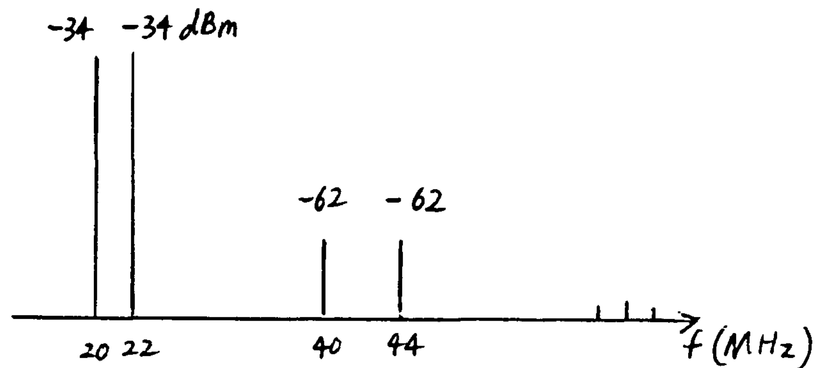


Figure 7.2.2. Response of Measuring System Alone

The system showed an obvious non-linearity, measurements were then made on each piece of equipment to ascertain the cause.

The frequency spectrum of each of the signal generators was measured in turn and found to contain harmonics;

frequency (MHz)	power (dB)	normalized
f = 20, 22	-28	0
2f	-56	-28
3f	-68	-40
4f	-75	-47
5f	-75	-47

It was ascertained that the available oscillators had double frequency components of about 30 to 40dB down from the fundamental.

The mixer was found to have the following port to port attenuations;

I/P - O/P	- 6dB
I/P - I/P	-32dB
O/P - I/P	- 6dB

and was also found to be quite linear.

With both signals applied to the spectrum analyser, using the mixer, the frequency components were seen to consist only of the two fundamentals and their harmonics. No intermodulation components could be seen, indicating that the system was quite linear from the the mixer onward. As an aside, it was initially intended to use an HP853A spectrum analyser. However, an a + b component was identified with components 2a and 2b at 6dB down. These

components showed the classical second order distortion effects when a or b were varied, identifying the machine as in need of repair.

7.3 Choice of Intermodulation Product

Considering first the second order distortion, the distortion products are at 2a, 2b and a + b. When a single input, a, is supplied one would anticipate that the a and 2a components at the output would be sufficient to evaluate the second order kernel. This indeed is the case. However, the application of a single frequency is not straightforward. The signal generators which are generally available give a substantial first harmonic, 2a, component. This complicates the measurement since the 2a component measured at the output will consist of a linearly amplified 2a component plus the required distortion product. Extraction of the required component is difficult since it is generally small compared to the measured 2a component i.e. the signal generator distortion will be much larger than the device distortion.

This problem can be reduced by the use of a high purity signal generator and/or using a filter with a sharp roll off, 80dB/decade for example, at the signal generator output. The filter method is of limited application because the second harmonic is only 0.3 of a decade, of frequency, away from the fundamental. Even an 80dB/decade

filter will only reduce any unwanted second harmonic by about 24dB. Apart from this both methods are complicated by the availability of such equipment and the limited bandwidths which are typical of this type of equipment. Fortunately a very simple method exists to circumvent all of these problems.

This method requires the application of two components at the input and the measurement of the a + b or a - b component. This simple solution relies on the fact that all distortions due to the signal generators are harmonic components. The components a + b and a - b involve only the fundamentals and therefore are not affected by the harmonic content of either input. This is only true, of course, so long as one frequency is not a multiple of the other.

It should be noted, however, that the a and b components measured at the output are affected by the gain compression distortion of the device (a + a - a), see Chapter 4, and the intermodulation of the incident components (2a - a). The first of these is a third order intermodulation component which should be negligible compared to the linear component, a, at the output. The second component should also be negligible since it is a second order intermodulation component involving the, small, second harmonic component at the input. With the generators used, for example, it should be 30dB down on

the device distortion.

When two input signals are of similar frequency the a + b component at the output is close to the second harmonic while the a - b component is close to zero. Consequently it is often better to use one frequency, a, in the frequency band of interest and a much lower frequency, b. The two components a + b and a - b then appear in or around the band of interest, actually appearing as sidebands around frequency a.

In the case of third order distortion the component 2a - b is the most appropriate two tone measurement, but this cannot be measured under the same conditions as the second order components. The problem is that 2a, which for this component is formed inside the device, is in fact incident at the input. As a consequence the same frequency is also generated by the second order distortion and can be of similar or greater amplitude.

Once more an elegant solution exists, this is simply to use three signals at the input. So long as there is no frequency component common to two or more of the input signals the ouput will be free of any components due to harmonic components at the input. When the three input signals, a, b and c, are of similar frequency then the component a + b - c will also be around this frequency. This particular component is of much interest as it is

generally the most objectionable in most applications of 'linear' transistors. This is because it is an in-band signal and nothing can be done about it once it has been generated, i.e. it cannot usually be differentiated from the wanted signal.

Gain compression for any order distortion measurement can be argued in the same manner as for second order and consequently is negligible for measurements of small distortions.

Fourth and fifth order distortion can be measured in a similar fashion to second and third order if four or five input frequencies, respectively, are used.

It should be noted that the fourth order component $a + b + (a - a)$ is at the same frequency as the second order distortion, $a + b$. Generally all components produced by second order distortion are also produced by fourth order distortion. As a consequence any second order distortion component that is measured will include an additional part due to fourth order. In fact the result is more general than this. Any even (odd) order distortion components are also produced by all higher even (odd) order distortions. Isolation of any particular order distortion from higher order distortion must therefore be carried out mathematically. However, the distortions are generally such that higher order

influences on lower orders can be ignored.

When making measurements it is important to be aware of which component is being considered as an adjustment is required to obtain an absolute value. The magnitude, Figure 4.4.1, Chapter 4, of the nth order distortions involving n different frequencies is m = n!/2n-1 giving an adjustment to IM(n) of 20log(m).

7.4 Measurement in a 50Ω System

Distortion measurements are carried out in dBm, which means that the measurement system must measure absolute power. The transistor is to be measured in the common emitter configuration, Figure 7.4.1, as shown, under varying bias conditions.

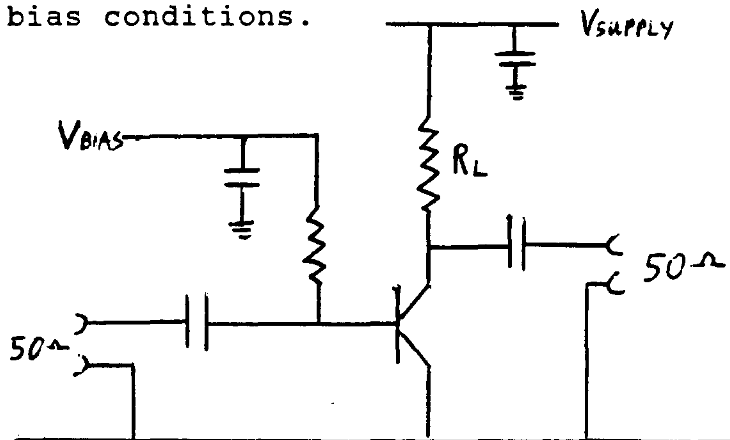


Figure 7.4.1. Common Emitter Measurement Circuit

Collector current and collector load resistor will be varied and the input impedance will also vary with

collector current. An ideal situation is to use impedance transformers in the input and output. This situation can be simplified a great deal with due consideration of the measurement requirements.

Measurements for distortion compare the wanted part of the signal with the unwanted parts. Noting this it is then only necessary to keep the output impedance resistive to avoid the need for conjugate matching and allow simple correction factors to be calculated for the power measurements. A similar argument can be used for the input at lower frequencies when the input impedance is resistive. In addition no phase information is needed and so the measurements are independent of line length. Losses in lines will need to be calibrated out if they are excessive but this is straightforward.

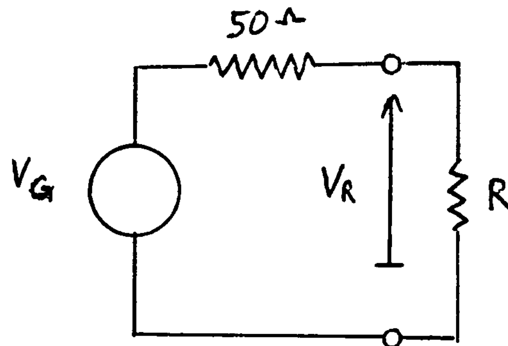


Figure 7.4.2. Equivalent circuit of generator side

Consider the generators, these are calibrated for the power they can deliver into a properly matched load. In

the event that a mixer or an attenuator is used in the signal path the reduced output is again specified for a matched load. The equivalent circuit is simply a voltage generator with a 50Ω output impedance,Figure 7.4.2.

With a matched load, R = 50Ω, the calibrated output power of the generator, P_G, is;

$$P_G = \frac{V_G^2}{800}$$

For an arbitrary load we have;

$$V_R = \frac{V_G R}{R + 50}$$

and combining we get the output voltage in terms of the load resistor and the calibrated power, viz;

$$V_R = \frac{R(800P_G)^{1/2}}{R + 50}$$

The power delivered into R is;

$$P_R = \frac{R400P_G}{(R + 50)^2}$$

Care must be taken if a generator is used with an

improperly matched load. The generator may not be well behaved, for unmatched conditions, due to the VSWR set up in the output line. To avoid these problems the generator should be padded out with a passive attenuator. This allows the attenuator output to be badly matched while still allowing the generator to see a well matched load.

Similar analysis can be done on the measurement circuit. The circuit output impedance is R_L, in parallel with any slope resistance of the transistor, the AC power in the load, P_L, is simply;

$$P_L = i_c{}^2 R_L / 2$$

and the output voltage is simply;

$$V_L = i_c R_L$$

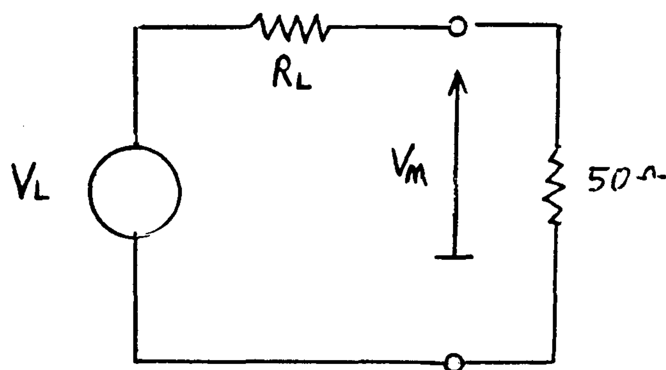


Figure 7.4.3. Equivalent circuit of Measurement Side

The high frequency equivalent circuit when the capacitor

is effectively a short circuit is shown in Figure 7.4.3.

The measured power, P_M, (as seen by the analyser) is;

$$P_M = \frac{V_M^2}{100}$$

where;

$$V_M = \frac{50V_L}{R_L + 50}$$

combining we obtain the power delivered to R_L without the analyser attached in terms of the monitored power;

$$P_L = \frac{(R_L + 50)^2 P_M}{50R_L}$$

Monitoring in this fashion modifies the output voltage swing, collector voltage, due to loading. Distortions due to the collector-base capacitance will thus be modified. To avoid this we can simply consider the parallel combination of R_L and the input impedance of the analyser as the load resistor. This allows the DC collector voltage to be set by the collector current and R_L but does restrict the AC load to about 50Ω. A better way to proceed is to monitor the voltage at the collector output with a high impedance probe and use appropriate calculations to

determine the power. A purpose built probe is available for the Marconi spectrum analyser. Conversion to power is simple, the equivalent circuit is shown in Figure 7.4.4.

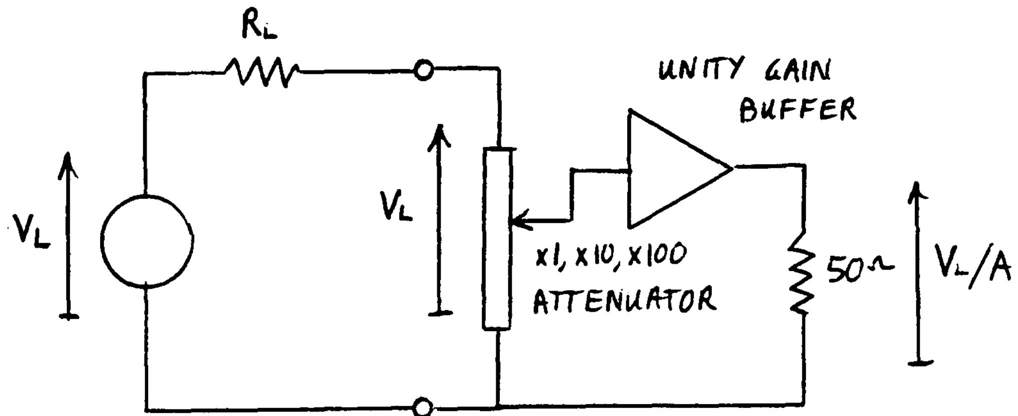


Figure 7.4.4. Equivalent Circuit Using Voltage Monitoring

The power measured by the analyser is;

$$P_M = \frac{V_L^2}{2A^2 50}$$

and the power in the load is;

$$P_L = \frac{V_L^2}{2R_L}$$

giving;

$$P_L = \frac{A^2 50 P_M}{R_L}$$

The measuring techniques described here allow the transistor distortions to be measured over a wide range of load conditions.

7.5 Setting Up the Equipment

Initial checks were carried out to see if the equipment was performing as required. Each signal source was swept from 1 to 100MHz (18 - 100MHz for HP3654) and observed on the spectrum analyser.

The Marconi TF 2730 performed well, showing less than 1dB variation over the whole range. The generator from the network analyser showed a 2dB fall over the range.

The zero loss probe was used to measure the unterminated outputs of each generator, the outputs showed similar variations to when correctly terminated. When terminated with a 50Ω through terminator (GE 84013A) measurement with the zero loss probe showed the correct -6dB drop. This measurement being the same as that with the generator connected directly to the analyser.

With three generators driving via the mixer the spectrum showed the fundamentals 6dB down and the expected first harmonic about 30dB down on the fundamental. The equipment was then configured as indicated in section 7.2. A BFW16A transistor was put in circuit as shown in

section 7.4.

Two tone measurements were taken under conditions;

I_c = 10mA

R_L = 1KΩ

V_{supply} = 20V

Frequencies at input;

a = 20MHz

b = 24MHz

Power at generator outputs; -30dBm

Measurements were made with a 50Ω input, frequency components, in dBm, at output were;

a = - 9

b = - 9

a + b = -37

2a = -42

2b = -44

2a - b = -57

Amplitudes of the inputs were varied and each component in the output was observed to act as the theory predicts. The input frequencies are close to the f_β of this device, 28MHz, and the second order components are well above. This being the case we would expect to see the second order components falling off at 20dB/decade. Theory gives 2b as 1.6dB down from the 2a component and 6.8dB down from the a + b component, while 2a should be 5.2dB down from the a + b component.

To check the gain in the transistor requires some calculation. The device is running at 10mA, which gives an r_e of 2.5Ω, from the data sheet beta should be 28, so r_{in} is about 70Ω typical. The voltage delivered to the base is 8.27mV, from section 7.4, allowing for 6dB loss in the mixer. Voltage gain of the transistor is 50/2.5, giving an output voltage of 165mV into 50Ω. Output power is 0.27mW or -5.7dBm. The measured value of -9dBm is about right since the device will have both a finite R_e and $r_{bb'}$ also the device is operating close to the f_β, in fact s_{21} is 2dB down from its low frequency value at 20MHz.

Further measurements were made at higher frequency which also followed the theory very closely. Similar measurements were made with three tones at 20MHz, 22MHz and 22.7MHz. The three fundamentals and the a + b - c components in the output were measured as; -9, -9, -9 and

-57dBm respectively.

7.6 Distortion Measurements

Distortion measurements were made, on BFW16A transistors, at low frequency, that is below the f_β of the device. The frequency response of the 50Ω terminated transistor is shown in Figure 7.6.1, where it can be seen that the bandwidth was 29MHz.

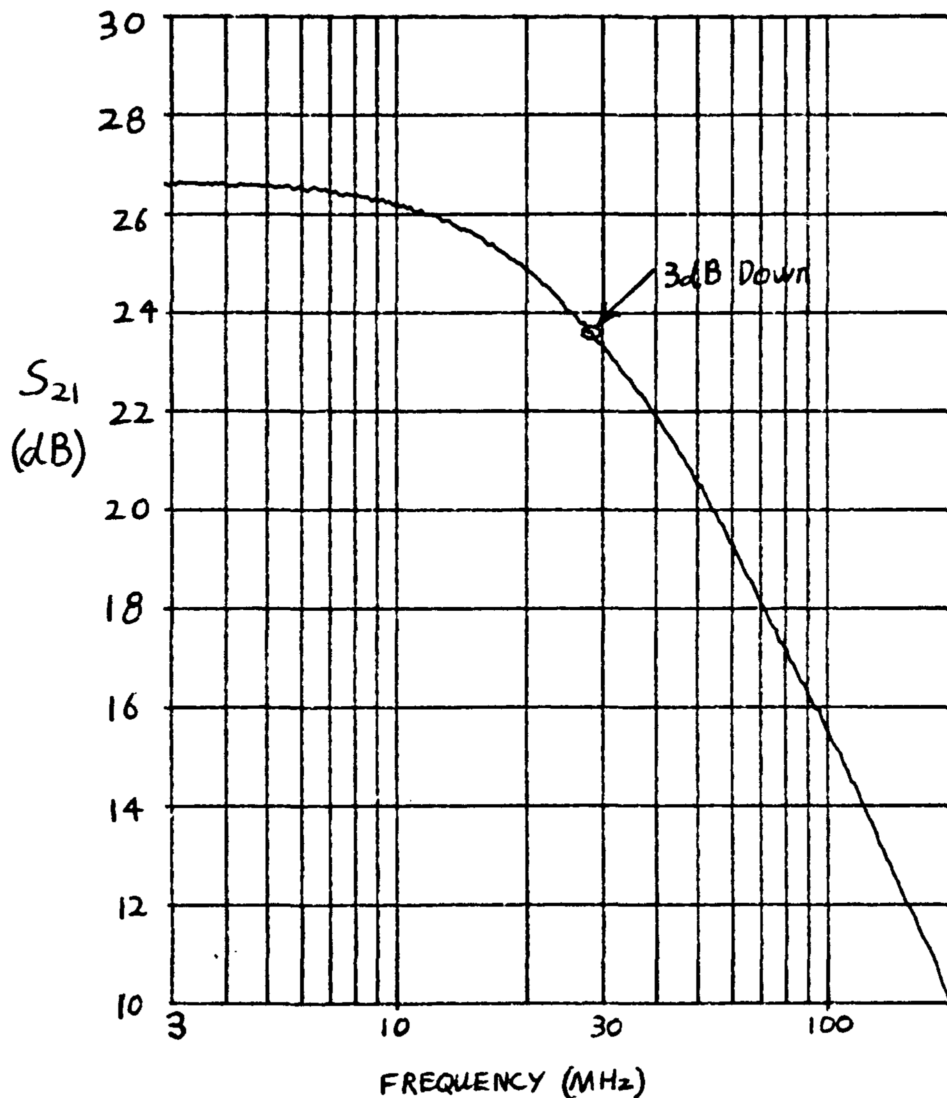


Figure 7.6.1. Frequency Response of BFW16A

Measurement of S_{21} was made over over a range of collector currents, Figure 7.6.2, where it can be seen that the forward gain peaks at around 100mA. These measurements where taken under constant collector-emitter voltage conditions, which was simply achieved by modifying the supply voltage, i.e. $V_{ce} = V_{supply} - I_c R_L$.

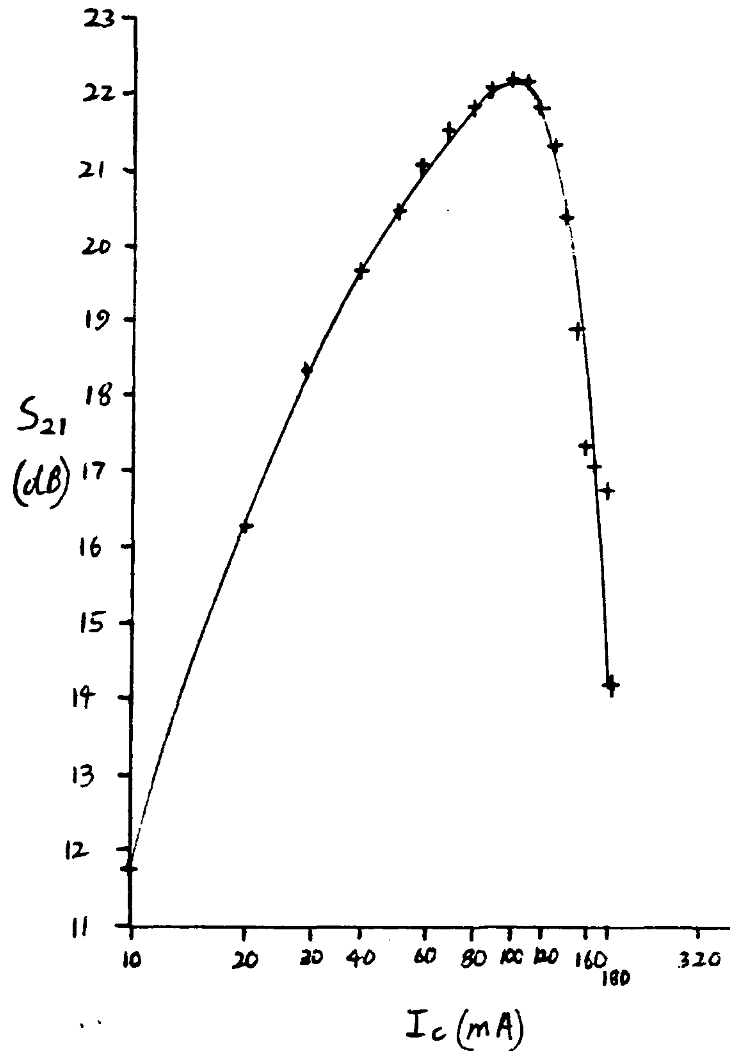


Figure 7.6.2. S_{21} Dependency on I_c

Three frequencies were applied at the input, via a mixer, of 9.6MHz, 9MHz and 4MHz. This gave an a + b - c component in the output of 3.4MHz, which was easy to identify. Each of the three frequencies was varied over a small range to verify that the a + b - c component was frequency independent, which it was. It will be noticed that the a + b + c component is 22.6MHz which is below the 3dB frequency of the circuit. Measurements were then made on a number of BFW16A transistors, Figure 7.6.3, with an effective collector load resistance of 8.33Ω. Plots of IM3 and power gain are shown. Theoretical IM3 is shown in dotted line for comparison. A theoretical gain slope is also shown but this is not an absolute value since there is an offset due to an uncalibrated input circuit, which is of no consequence. Both IM3 and the power gain follow the theoretical curves well up to 80mA when a sharp increase in IM3 is seen to occur. This corresponds to a fall in the power gain, as discussed in Chapter 4. IM3 is seen to be some 13dB above theoretical at 2mA and falls to about 8dB above it at 10mA. After this it is seen consistently rise above theoretical until 80mA is reached, when the large increase, in IM3, occurs.

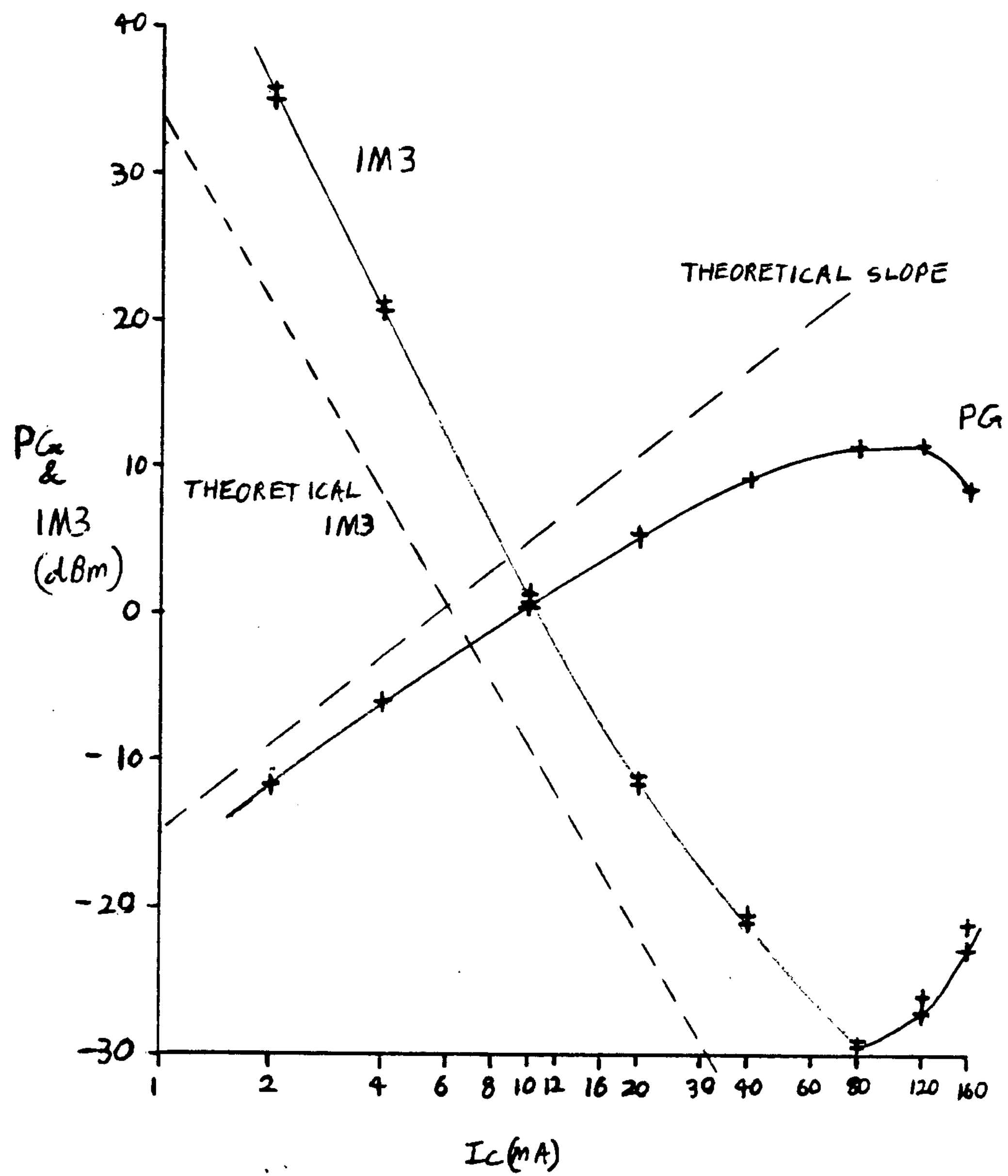


Figure 7.6.3. Measured IM3

The power gain is also seen to deviate from the ideal towards higher currents. This is due to the onset of base push-out. Further measurements were made with input signals of 2MHz, 2.2MHz and 9.6MHz which gave IM3 of -37dBm with a collector current of 40mA. This is very close to the theoretical, -35dBm, showing that the model is very good when the low frequency approximation holds. Both 9MHz and 9.6MHz, as used in the previous measurements, are at a frequency where the gain has already started to fall off. Sum and difference frequencies are below the 3dB frequency which gives a frequency independent output, but the operation does not strictly conform to the low frequency approximation. These lower frequency measurements were repeated for each combination of input frequencies raised by 10dB and IM3 was seen to be identical for each of the six input combinations. The actual value of the distortion component, of course, actually varies i.e. goes down by 10dB when any one input signal is reduced by 10dB, by 20dB when any pair of input signals are reduced by 10dB and by 30dB when all the input signals are 10dB down.

Further measurements were made above the 3dB frequency. With two input frequencies of 20MHz and 24MHz. Measurements were made on the a + b, 2a and 2b, i.e. 40MHz, 44MHz and 48MHz components in the output, Figure 7.6.4.

Inputs (dBm)		Outputs (dBm)				
a	b	a	b	a+b	2a	2b
-30	-30	-9	-9	-37	-42	-44
-20	-30	1	-9	-27	-21	-44
-20	-20	1	1	-17	-22	-24
-30	-20	-9	1	-27	-42	-23

Figure 7.6.4. Second Order Distortions at High Frequency

It can be clearly seen that the output components follow the theory very well. For both inputs at -30dBm the a + b component is 5dB above the 2a component and 7dB above the 2b component. Theoretical figures are;

$$20\log(2\times40MHz/44MHz) = 5.19dB$$

and;

$$20\log(2\times48MHz/44MHz) = 6.77dB$$

In each case the factor 2 appears due to the magnitude of 1/2 for each of the 2f components. The difference between the 2a and the 2b component is 2dB and theoretically it should be;

$$20\log(48MHz/40MHz) = 1.58dB$$

These figures are consistent with the accuracy of measurement, ± 0.5dB. It will be noticed that a reduction of 10dB in either input signal reduces the a + b component by 10dB and the double frequency component of the altered input reduces by 20dB while the second double frequency component does not change. Similar measurements were made for third order distortions and showed similar results.

Variation of high frequency distortion with collector current was investigated.

I_c (mA)	Output (dBm) 19.3MHz
1	-62
2	-56
4	-56
6	-56
8	-57
10	-57
12	-57
14	-58
16	-59
18	-58
20	-63

Figure 7.6.5. HF Third Order Distortion with Current

With input frequencies of 20MHz, 22MHz and 22.7MHz the a + b - c, 19.3MHz, component at the output was measured, Figure 7.6.5. All frequencies involved are close to the 3dB frequency which means the reactive components tend to be of importance in the distortion performance. Third order distortion for this transistor, under these conditions, is seen not to be a strong function of collector current. A similar set of measurements were made at half this frequency, i.e. 10MHz, 11MHz and 11.3MHz, Figure 7.6.6.

I_c (mA)	Output (dBm) 9.7MHz
1	-59
2	-54
4	-52
10	-50
16	-49

Figure 7.6.6. LF Third Order Distortion with Current

For these conditions some of the components will be above the 3dB frequency and thus the results will not conform to a low frequency approximation model. The a - b + c, 10.3MHz, component for example was markedly frequency dependent. It is noticeable that the the current

dependence is greater at low frequency, third order components appearing to fall monotonically with increase in I_c. The relationship however is more complicated than this, as discussed in Chapter 4.

This chapter has shown that there is a general agreement between the theoretical distortions produced by transistors and actual distortions.

CHAPTER 8 - COMPUTER SIMULATION

8.1 Program Limitations

Computer simulation of circuits is now widely used. SPICE is the classic example of a simulation program that has facilities for modelling most types of transistor. Being free and usable although with a number of 'bugs' and shortcomings it has become the de facto standard.

The DC model is generally good with facilities for modelling many of the second order effects. The AC model accounts for junction capacitances and storage delays and again is quite good.

The capabilities of SPICE for analysing distortion performance are not clear. This chapter concerns itself with whether SPICE is suitable for distortion analysis and, if so, what conditions must be met for meaningful results to be obtained.

A small signal distortion analysis command is available, .DISTO, which causes an analysis to be performed at the DC operating point. This then gives a breakdown of the contributions of each of the nonlinear components in the transistor model. In the version of SPICE used, this option had been corrupted by modifications made to the MOS model by other users. Incidentally, this appears to be a

common problem, one needs to be alert as to which version is in use and also whether attempts have been made to fix the bugs. This chapter therefore is confined to investigating how large signal distortion analysis can be evaluated.

For large signal analysis an output signal is obtained in the form of samples at regular time intervals. This may then be analysed by means of a Fourier analysis, which is available within the program.

8.2 SPICE

The program has facilities for modelling most electronic components and an extensive range of voltage and current sources. These include independent, various functions of time, and dependent sources, both linear and non-linear, which can be either current controlled or voltage controlled. The transistor model employed is the Gummel-Poon model described in Chapter 3 which is able to model many second order effects and is suitable for distortion analysis.

A variety of analyses are available allowing operating point information to be obtained, DC sweeps, frequency sweeps and transient analyses. The analysis of most interest here is the transient analysis as it allows the large signal distortion to be investigated. This is

certainly true for gross distortions, however the situation is not so clear for the analysis of ultralinear transistors where the distortion products may not be significantly larger than the "numerical noise" floor of the program.

The basic transient analysis procedure within SPICE starts with the initial conditions set up by the operating point analysis. From this starting point a new state is determined for the end of an incremental time step, this procedure is repeated many times until the analysis is finished. The time steps have an upper bound which is set by a maximum change in voltage or current that the program will allow, in this way fast transients do not lose accuracy due to too few time steps. In addition to these limits there is additional control which the user may exercise through use of the .TRAN command. This allows the start and stop time of the simulation to be set along with the time step size which sets the time increment for the output data. An additional facility is to set a maximum time step independently of the other controls. For good accuracy it is important to keep the maximum time step small, as a rule of thumb, less than one tenth of the smallest C-R time constant of interest for general analysis and much smaller for good accuracy in distortion analysis.

The smaller the time step the closer the analysis follows

the continuous time result. As an example consider a simple C-R network with the capacitor charging. The voltage on the capacitor is shown in Figure 8.2.1 for both the continuous time case and the discrete time case.

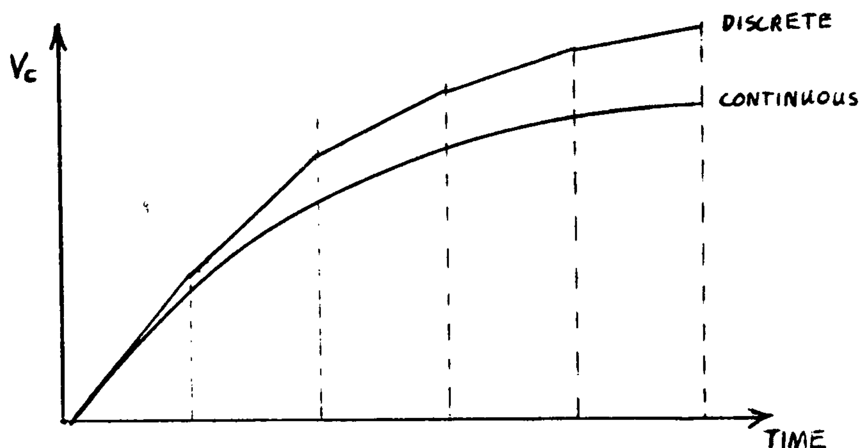


Figure 8.2.1. Capacitor charging curve

It can be seen that for the discrete case the charging time is smaller. The reason for this is that the derivative is continually reducing with time but for the discrete time case the derivative remains fixed for the duration of the time step, therefore on average it is too large, i.e. charges too quick. This effect can be clearly seen on SPICE simulations when analysing propagation delays, if the time step is increased the delay apparently decreases by several percent. It is apparent that the discrete time model deviates from the continuous time model and when a Fourier analysis is performed this deviation appears as harmonic distortion. In addition to this effect there is some uncertainty in the actual quantity because of the finite accuracy of the

calculations within SPICE. This can be considered as numerical noise.

The frequency analysis available is restricted to a, discrete, Fourier transform up to the ninth harmonic, i.e. up to the 9f component. In principle it is possible to do an intermodulation distortion analysis with this type of analysis, that is an actual analysis rather than calculating from a harmonic analysis. The fundamental condition for this type of analysis, Fourier, to be valid is that the waveform under scrutiny must be repetitive with period T, where T is the time duration of the sample. To carry out an intermodulation analysis thus requires all of the frequency components to have an integer number of cycles within the sample period T. Put another way each, frequency must comply with;

$$T = I_n / f_n$$

Where f_n is the frequency component and I_n is an integer. Now since the input frequencies will appear at the output they also must conform to this equality.

Now consider f as some arbitrary reference frequency, then we may choose, for example, two input frequencies of 3f and 4f. The output will consist of sum and difference frequencies, namely, for second order components; 1f, 6f, 7f, 8f and a DC component. Third order components will

consist of; 2f, 3f, 4f, 5f, 10f and 11f, Chapter 4.

Notice that inputs of 1f and 2f are of no use since the second order components will be DC, 1f, 2f, 3f and 4f, while the third order components will be DC, 1f, 2f, 3f, 4f, 5f and 6f. Which means that there is no easily obtainable second order component because there will also be a third order component at the same frequency.

The first case cited does not suffer from this problem. For this example all of the information is held in just two of the frequency components, one second order and one third order. The components due to second order effects cannot be produced by third order effects so, in principle, any of these components holds the information from which the second order coefficient or kernel can be evaluated. The DC component in fact will consist of two components one due to the 3f input and one due to the 4f input plus, of course, the DC bias. In the case of third order components, we would not use 3f or 4f since we would need to subtract the fundamental component of that frequency and also the gain compression components, which are vectors. In fact the 1f component is a good choice, to see why, consider the sample period that is required, this is simply one period of the frequency f.

If a Fourier analysis were carried out, the largest components would be 3f and 4f, the input frequencies,

while the distortion information would be given by 1f and 2f. Alternatively the same information can be determined from the harmonic components of a single input frequency. This is marginally simpler and also allows fourth, fifth and even sixth order distortions to be evaluated.

8.3 Fourier analysis using SPICE

The accuracy of SPICE for Fourier analysis can be gauged by analysing a SPICE signal source, Figure 8.3.1, the 1Ω resistor is required to avoid nodes with single connections.

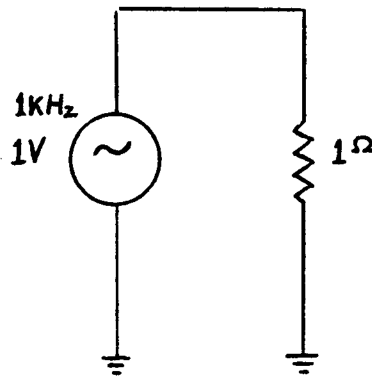


Figure 8.3.1 Analysis of SPICE signal source

The SPICE manual recommends at least 100 points for good accuracy, results for this and for 1000 points are shown in Figure 8.3.2, analysis being carried out in the first millisecond of the simulation.

Harmonic number	frequency (KHz)	Fourier components	
		100 point	1000 point
1	1	1.000	1.000
2	2	1.124×10^{-4}	1.078×10^{-8}
3	3	3.102×10^{-5}	2.932×10^{-9}
4	4	1.480×10^{-5}	1.514×10^{-9}
5	5	9.016×10^{-6}	1.139×10^{-9}
6	6	5.931×10^{-6}	1.332×10^{-9}
7	7	4.144×10^{-6}	2.673×10^{-9}
8	8	3.290×10^{-6}	9.096×10^{-9}
9	9	2.588×10^{-6}	9.998×10^{-7}
THD		0.011825%	0.0001%

Figure 8.3.2 Results of SPICE analysis

Note that in SPICE terminology the fundamental frequency has harmonic number 1, this terminology will be used throughout. The main harmonics of interest are second and third with fourth and fifth of next importance. A maximum dynamic range of 1 in 10^8, 160dB, may be expected by inspection of the second harmonic. An analysis of 5000 points gave a second harmonic of 9×10^{-9} reducing to a ninth harmonic of 8×10^{-11}, the total harmonic distortion being 10^{-6}%. Current flowing from the signal source is simply related by I = V/R and gives practically identical Fourier components.

Replacing the resistor by a capacitor, 1F or 1μF and using a 1000 point analysis gives a total harmonic distortion of 0.009648% for the voltage source and 5.715988% for the current. The problem here is due to the initial conditions at the start of the analysis. The voltage source has zero internal impedance which means that the the terminal voltage in this case is purely a function of the generator voltage and is not affected by the load. The current on the other hand is a function of both the terminal voltage and the load. In this case the voltage is zero at time zero and so also is the current, but it should be AwC, where A is the voltage amplitude and w is the signal angular frequency. When the analysis is carried out in the second cycle, i.e. between 1ms and 2ms, the resulting total harmonic distortion is 0.000100% for the voltage and 0.000135% for the current as shown in Figure 8.3.3. This improvement occurs because, by the start of the second cycle the current is in a steady state relationship with the voltage. The time actually required for this state to be achieved is related to the time constant of the the circuit which is very much less than one period of the voltage generator frequency.

Harmonic number	frequency (KHz)	Fourier components(1000 points)	
		V	I
1	1	1.000	6.283×10^{-3}
2	2	1.085×10^{-8}	5.625×10^{-11}
3	3	2.975×10^{-9}	1.103×10^{-11}
4	4	1.462×10^{-9}	1.158×10^{-11}
5	5	1.090×10^{-9}	5.758×10^{-9}
6	6	1.390×10^{-9}	1.438×10^{-11}
7	7	2.721×10^{-9}	9.728×10^{-12}
8	8	9.058×10^{-9}	4.567×10^{-11}
9	9	9.998×10^{-7}	6.259×10^{-9}
THD		0.000100%	0.000135%

Figure 8.3.3 Distortion due to capacitor

The voltage analysis for the capacitor is virtually the same as for the resistor while the current analysis is worse by a third. This may be due to the relation $I = CdV/dt$ where dV/dt is arrived at numerically within SPICE i.e. it may be obtained from a small difference of two large values. The worst case components are the fifth and ninth, each of which are about 10^6 below the fundamental, i.e. 120dB of dynamic range. Some improvement might be expected from a reduction in the tolerances in the .OPTIONS command but at the cost of longer simulation times.

The resistor and capacitor are reasonably linear, a simple C-R network Figure 8.3.4 when analysed over 1000 points produced 0.000100% THD for voltage across the capacitor and current flowing (R = 1Ω, C = 1μF, f = 1KHz). The THD reduced to 10^{-6}% for an analysis of 5000 points and the computing time increased five fold.

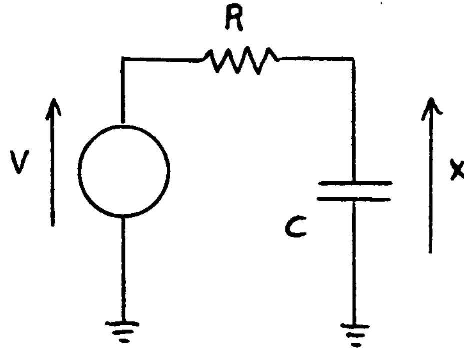


Figure 8.3.4. Simple C-R network

The length of simulation time required depends on the relationship between the signal frequency and the C-R time constant, because generally the initial conditions do not correspond to the steady state conditions. This point is now analysed more formally.

The response of the C-R network is described by;

$$x + CR\frac{dx}{dt} = V$$

where;

x is the voltage across the capacitor

and;

$$V = A\sin wt$$

is the sinusoidal input signal which will be used within SPICE.

The general solution for x is;

$$x = ae^{-t/CR} + \frac{A\sin wt}{1+(wCR)^2} - \frac{AwCR\cos wt}{1+(wCR)^2}$$

and for SPICE the analysis will start with zero charge in the capacitor giving the boundary condition x = 0 at t = 0. The specific solution is;

$$x = \frac{A(wCRe^{-t/CR} + \sin wt - wCR\cos wt)}{1 + (wCR)^2}$$

It can be seen, Appendix XIV, that a Fourier analysis of this function over a period of $2\Pi/w$ seconds will lead to a frequency component, $w/2\Pi$, due to the input signal and additional frequency components $nw/2\Pi$ due to the exponential term. Where n is an integer 0,1,2 etc. carrying out a Fourier analysis over the period T to $T + 2\Pi/w$ produces Fourier components;

$$\frac{(wCR)^2 A e^{-T/CR}(1-e^{-2\pi/wCR})(g\sin nwT + h\cos nwT)}{\pi(1 + (wCR)^2)(1 + (nwCR)^2)}$$

where;

$$g = -nwCR$$

and;

$$h = 1$$

for the cosine coefficients, while;

$$g = 1$$

and;

$$h = nwCR$$

for the sine coefficients.

It should be noted that these terms are not true harmonic frequency components in the output, but are a direct result of the, exponentially decaying, initial conditions. The discrete Fourier analysis requires a periodic function, and this condition does not hold while the exponential term is significant.

The exponential components are close to zero for the conditions 1 >> wCR, T >> CR and wCR >> 2Π and approach a maximum around w = CR.

The variation of total harmonic distortion with various C-R products is shown in Figure 8.3.5, while Figure 8.3.6 shows the variation in total harmonic distortion with sampling time.

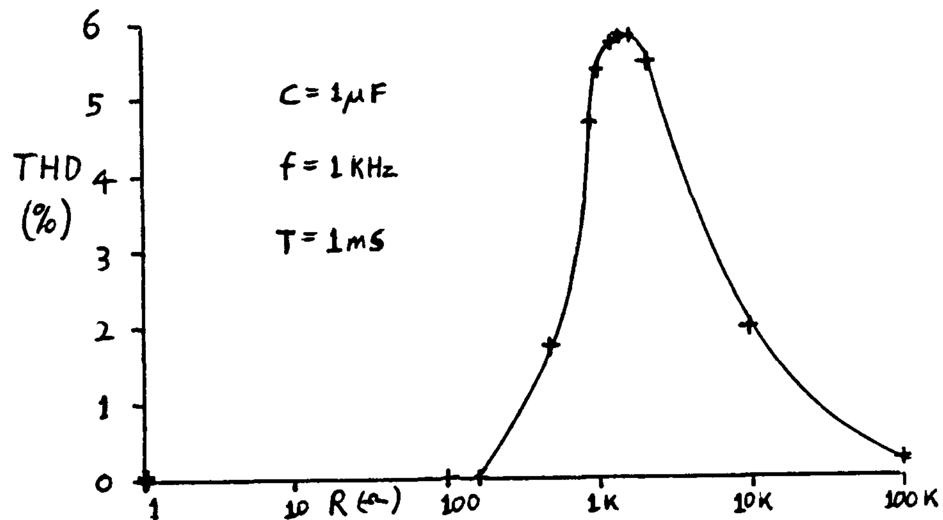


Figure 8.3.5 Variation of THD with C-R

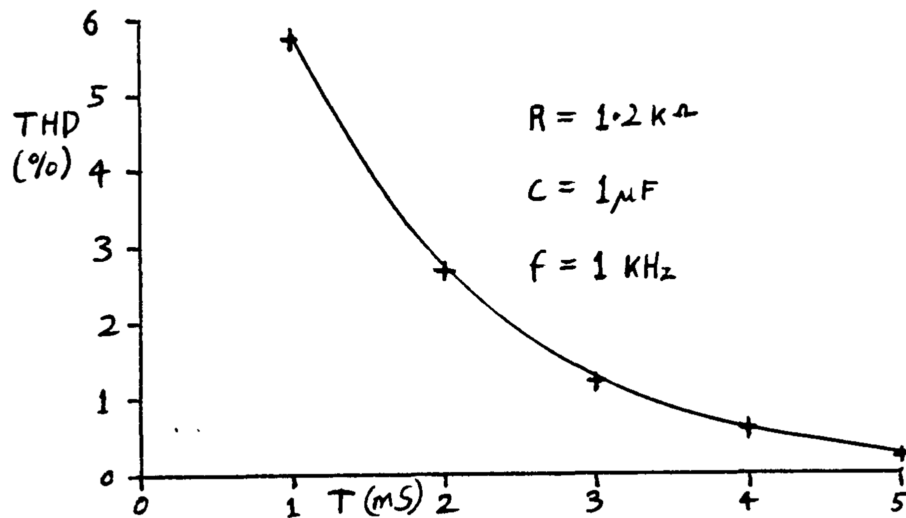


Figure 8.3.6 Variation of THD with T

For this case T = 1ms and Tw = 2Π, therefore the sine term

is zero and the cosine term is 1.

The sine coefficients are given by;

$$\frac{(wCR)^2 A e^{-T/CR}(1 - e^{-2\pi/wCR}) nwCR}{\Pi(1 + (wCR)^2)(1 + (nwCR)^2)}$$

and the cosine coefficients by;

$$\frac{(wCR)^2 A e^{-T/CR}(1 - e^{-2\pi/wCR})}{\Pi(1 + (wCR)^2)(1 + (nwCR)^2)}$$

The modulus of the nth component is;

$$\frac{(wCR)^2 A e^{-T/CR}(1-e^{-2\pi/wCR})}{\Pi(1 + (wCR)^2)(1 + (nwCR)^2)^{1/2}}$$

which is in good agreement with SPICE, Figures 8.3.5 and 8.3.6.

In broad terms the distortion contributed by the exponential term is negligible under conditions where the stimulus periodicity is much larger or much smaller than any circuit time constants. This is because if the time constants are small the exponential term decays fast and is soon negligible. If the time constants are long on the other hand they will not vary much during the analysis period and therefore not contribute much distortion. For

time constants of similar magnitude to the stimulus periodicity the simulation must be allowed to continue until the exponential term has decayed sufficiently i.e. T >> CR. Under these conditions the term $e^{-T/CR}$ becomes small and reduces the effective distortion.

8.4 Analysis at low frequency

In order to evaluate simulation accuracy at low frequency an idealized transistor model was used, as shown in Figure 8.4.1. Analysis using one hundred

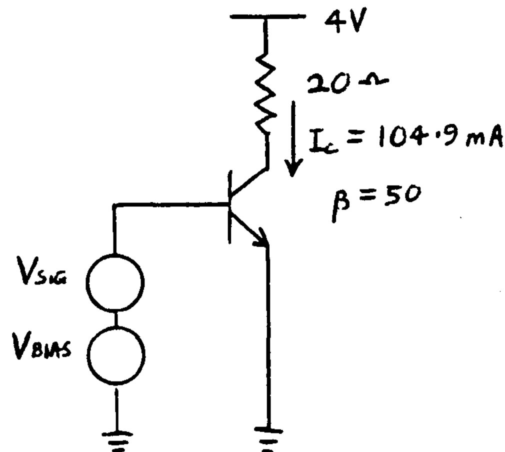


Figure 8.4.1 Ideal transistor circuit

points for the Fourier analysis gave rather poor results. With Vsig of 1mV amplitude the fundamental is correct within 0.5% but second harmonic is in error by 2.9% and third harmonic by 32%. For an input of 10mV amplitude the results are slightly better with the fundamental within 0.5% and second and third harmonics 2.5% and 6.4% respectively, fourth harmonic is in error by 23%

When one thousand points are used for the analysis the gives a much improved accuracy. An input of 1mV peak gives inaccuracy of the third order component of 0.2% and fourth harmonic of 5.8%, fundamental and second harmonic are better than 0.1%. A larger input of 10mV gives a fifth harmonic accuracy of 5%.

The corresponding purity of the signal source is 0.01% and 0.0001%for the hundred point and thousand point analysis respectively. For the hundred point analysis the distortion is predominantly second and third order while for the thousand point analysis it is predominantly second and ninth order.

Inaccuracies in the Fourier analysis are due to "numerical noise" which occurs in the first place due to the quantization of the discrete samples which form the input signal. In principle this noise can be removed by over sampling the signal, i.e. taking more samples, noise is then averaged out within the Fourier analysis, and the final accuracy is determined by the numerical accuracy of the transform. Achievable dynamic range is thus determined by the number of points used for the transform, one thousand points for example allows a range of five to six orders of magnitude with five percent accuracy for the smallest components.

For simulation purposes the amplitude ratio of the

frequency components must be kept within the dynamic range limits of the program. Higher order distortions increase in magnitude faster than low order distortion, i.e. f_n increases approximately in proportion to the nth power of the input amplitude. With these considerations in mind it can be seen that best accuracy is obtained by simulating with the largest possible input amplitude, which will be set by the onset of saturation.

Further simulations have been carried out using base and emitter resistors in the circuit shown in Figure 8.4.1 in each case the results compare with theory. It is seen that an accuracy of 100 - 120dB is obtainable with a one thousand point analysis.

8.5 Analysis at high frequency

Analysis at high frequency is considerably more complicated than at low frequency because there are more components with significant impedances, some of which are nonlinear. Thus there is more calculation required and potentially more numerical noise. In addition to this the harmonic components are attenuated by the circuit itself this increases the required dynamic range of the analysis.

In order to investigate the high frequency analysis a simplified ideal model was used as shown in Figure 8.5.1.

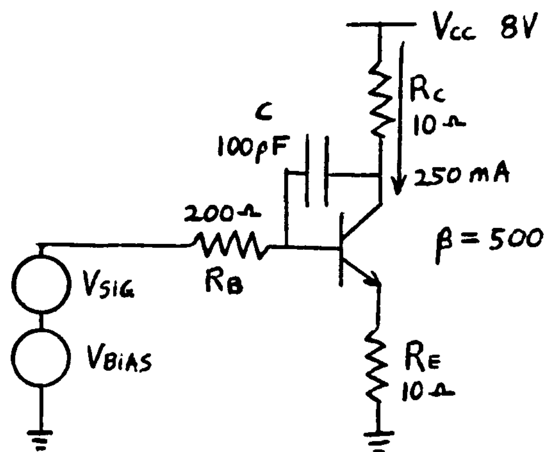


Figure 8.5.1 Simplified Ideal High Frequency Circuit

Here the low frequency distortion is kept small by using emitter degeneration and running the transistor at high current to maintain a small r_e. The dominant pole of the circuit is set by R_B and C, the frequency is approximately;

pole frequency $= 1/2\pi R_B C(1 + R_C/R_E)$

and for the values used is approximately 4MHz. This circuit was analysed at 1KHz and at 100MHz. Calculation at low frequency gives an M2 of 4.94×10^{-3} the 1KHz component at the output has a magnitude of 9.03×10^{-4}V peak. Expected 2KHz component will then be, Appendix VII, 5.1×10^{-9}V peak. The dynamic range required of SPICE is therefore at least $9.03\times10^{-4}/5.1\times10^{-9} = 1.77\times10^{5}$ which, as has just been shown, is about the limit available. Higher order distortions will need even more dynamic range than this. This simulation gave normalized harmonic components

of zero.

Calculation at high frequency gives an M2 of $6.3x10^{-10}$ this being due to the collector capacitance alone. The fundamental, 100MHz, component at the output has a magnitude of $6.34x10^{-5}$V peak. Expected 200MHz component will then be, Appendix VII, $6.29x10^{-10}$V peak. The dynamic range required of SPICE is therefore at least $1x10^{5}$ which, is fairly close to the limit available. A normalized second harmonic component of around $1x10^{-5}$ is therefore expected. With a simulation carried out over 20nS, i.e. 10nS to let the initial conditions subside, the analysis was carried out over the second 10nS period and gave a second harmonic normalized component of 0.035. This is much higher than expected. Further simulations were then carried out with progressively longer periods allowed to elapse before the Fourier analysis was carried out, the resulting normalized second harmonic is shown in Figure 8.5.2.

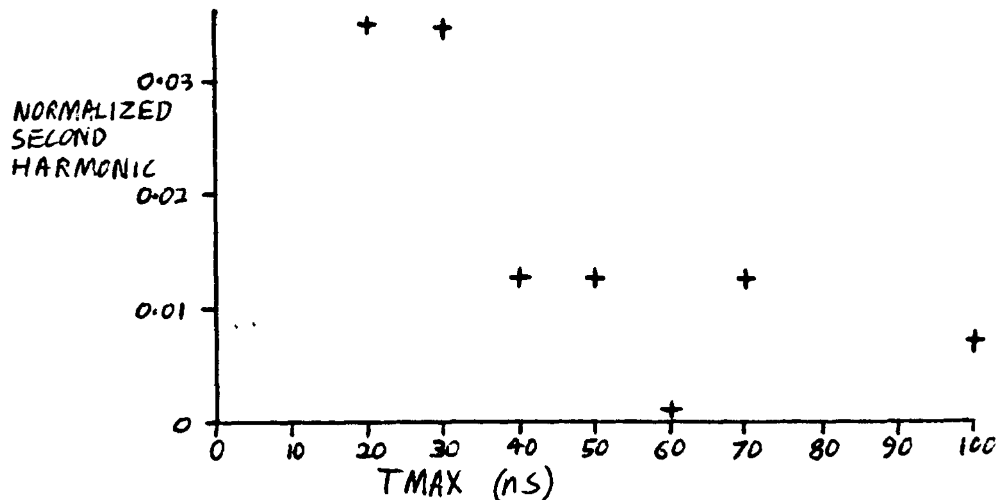


Figure 8.5.2. Second Harmonic Component versus TMAX

The graph appears to show a trend toward zero with increasing simulation time. This trend, however, is not consistent, at one place it increases with an increase in simulation time. Even the 100nS simulation shows a figure which is several orders of magnitude larger than expected. It is expected that the problem is caused by the non-linearity of the collector-base capacitor.

It is concluded that SPICE is of little use in the analysis of small non-linearities in transistors at high frequencies. The program does, however, appear to be of use for low frequency analysis. Dynamic range appears to be limited to around six orders of magnitude. Analysis of circuits with fixed capacitances can be achieved, but it is important in such cases to ensure that the circuit has achieved a steady state condition before the Fourier analysis is carried out.

CHAPTER 9 - CONCLUSIONS

9.1 Summary

The basis of a novel bipolar transistor structure was proposed by Dr R. Aubusson of Middlesex Polytechnic in 1977. The novelty lies in replacing the conventional overlay transistor's P+ base grid with a refractory metal grid, in order (a) to lower the base resistance and (b) to autoregister the emitter. It was claimed that the linearity of the transistor would also be improved. A number of questions raised by this idea have been investigated.

The need for such a transistor was investigated in terms of the requirements for improved performance and whether other devices showed clear advantages.

Modelling of the high frequency bipolar transistor was reviewed. Detailed modelling of the base extrinsic resistance was carried out. Plausible structures, using the metal base grid, were proposed and compared with conventional structures.

The current understanding of distortion analysis applied to transistors was reviewed. The main ideas were presented in a unified manner and then extended to higher order. A number of the transistor's second order effects

have been analysed in a novel fashion. The metal base grid transistor in particular was analysed and compared with conventional transistors.

The setting up and use of distortion measurement equipment has been described. Results of measurements made to check suitability of equipment have been presented.

The suitability of SPICE, the well known circuit simulation program, for large signal distortion analysis has been investigated.

Practical aspects of fabricating the metal base grid transistor were investigated. A procedure for tungsten deposition has been determined and was presented here along with the films physical and electrical characteristics. Analysis of the tungsten-silicon interface has been carried out by means of Rutherford back scattering. Various means of delineating the tungsten film were assessed and a working procedure determined. The deposition of various overlaying insulators and the problems caused by the, readily oxidized, tungsten film have been investigated. Formation of the emitter, requiring further high temperature processing, has been assessed in view of the limitations imposed by the, already formed, base metallization.

9.2 Conclusions

The bipolar transistor was found to be most suitable for continued research. FETs are seen to offer some advantages but these are not overwhelming.

The metal base grid transistor compared favourably with conventional transistors and allows for a smaller emitter pitch. The major difference with respect to the conventional overlay transistor structure lies in the use of a metal base electrode. This has the effect of maintaining a constant base-emitter voltage over all parts of the junction, due to the small sheet resistivity of the metal. The overall base series resistance however does not reduce by as much as might be expected, although it is typically smaller than for a conventional overlay transistor with a larger emitter pitch. The base series resistance tends to be limited due to the fairly large specific contact resistance of the metal to silicon interface. This point is not of much consequence since multi-emitter power transistors of this type conventionally employ emitter ballasting in order to maintain thermal stability. Any specific contact resistance in the base in fact has the same effect, except it carries the base current as opposed to the emitter current. Specific contact resistance can, in principle, be traded with emitter ballasting, i.e. emitter series resistance, in the ratio of $1:(1 + \beta)$.

Any advantage accruing from the autoregistering of the emitter to the base electrode, other than due to a direct reduction in pitch, is seen to be marginal. Analysis of the current sharing irregularities in the emitter, due to misalignment of the base electrode, shows that typically there would be less than 10% difference between the currents flowing in each side of an emitter finger. This can be translated into a corresponding saving in emitter area.

Theoretical distortion performance of transistors has been investigated and the theory shown to be correct by means of distortion measurements made on commercial devices. The novel transistor would be expected to give an improved distortion performance. This advantage is due to the size reduction that can be achieved, which gives a smaller, non-linear, collector-base capacitance.

Computer simulation of distortion using SPICE has been investigated and shown to be of limited accuracy for general usage. Good accuracy is seen to be obtainable for DC analysis only.

Practical aspects of fabricating the metal base grid transistor were investigated. High frequency sputtering was used as a means of deposition. This was found to be reasonably usable but suffered from unreliability in making ohmic contact. Problems were caused by the

compressive stress in the film formed during deposition, which caused the film to wrinkle. This can be minimized by clean operating conditions, i.e. reducing the irregularities which form catalysts for the formation of wrinkles. Annealing the tungsten film at around 800°C was found to be the only reliable way of preventing wrinkles forming. Analysis of the tungsten-silicon interface has shown the suitability of the metallization as a base grid. Delineating the tungsten film could be carried out with good resolution, around 1μm, with Shipley positive resist and using H_2O_2 as an etchant. Subsequent deposition of nitride, silox and spin on silica film was found to be possible. In the case of silox a modified procedure was used whereby the oxygen was allowed into the reaction chamber last in order to avoid oxidizing the tungsten film. All of these films were found to be unreliable for high temperature processing, where oxygen had to be kept from the tungsten. It was discovered that any local reaction lifted the overlaying film, thus allowing further reaction to take place. The only reliable way to carry out high temperature processing after the tungsten is in place is to do it in an oxygen free ambient. It should be noted that "White Spot" nitrogen is not oxygen free and causes oxidation of the tungsten film if used.

Formation of the emitter, requiring further high temperature processing, has been assessed in view of the limitations imposed by the preformed base metallization.

It appears that emitter formation is possible after the base metallization has been carried out by means of ion implantation and subsequent activation in an oxygen free ambient.

In summary, it has been shown that the novel structure can be constructed and that a significant performance improvement is to be expected, although a full realization was not possible within the resource constraints of the project.

9.3 Recommendations

A number of the processes used for this research work were dictated by the resource constraints and are not necessarily the most appropriate.

Tungsten deposition was carried out by means of RF sputtering. This caused two problems firstly ohmic contact was not guarenteed and secondly the film had a compressive stress as deposited which made the film unstable, this has to be seen to be believed. Both problems can in principle be overcome while sputtering. The ohmic contact problem is due to the native oxide that forms, almost immediately, on silicon when oxygen is present. This film can be removed inside the sputtering chamber by first reverse sputtering the sample, i.e. use the sample as the target. Then without breaking vacuum

proceeding with the deposition from the tungsten target. This was not implemented on the equipment the author used since it constituted a major modification to the equipment requiring a method to switch the high tension generator output plus various high voltage gas tight feed-throughs. The second problem, compressive stress, can be overcome by heating the sample either during or after deposition to 800°C in order to anneal the film. This is quite difficult to achieve inside the vacuum chamber since the heater will "out gas" at a significant rate under these conditions. Better, is to heat the sample outside the chamber in a non-oxidizing ambient, as done here. A possible alternative is chemical deposition which can be carried out in a controlled ambient and allows, as with sputtering, the contact holes to be cleaned of oxide. It is not clear however whether the compressive stress problem is overcome.

The three types of insulator investigated were found not to be reliable protection for the tungsten, against oxygen, at high temperatures. With spin on silica film and nitride the problem is not to do with the intrinsic film properties but is attributable to dust particles. In principle only one defect is required in the cover of a tungsten film in order for it to be totally oxidized, the reaction rate of course depends on the temperature. For this reason, any high temperature processing that is to be carried out in an oxidizing ambient after the tungsten

deposition should be done after the metal has been delineated, since this localizes the destructive mechanism.

BIBLIOGRAPHY

1 Kumar V., Fabrication & Thermal Stability of W-Si Ohmic Contacts, J. Electrochem Soc: Solid-State Science & Tech., Vol 123, No. 2, 1976, pp 262-269

2 Shaw J. M., Amick J. A., Vapour Deposited Tungsten for Silicon Devices, Solid-State Technology, Dec. 71, pp 53-57

3 Perlow S. M., Third-order Distortion in Amplifiers and Mixers, RCA Review, Vol 37, June 76, pp 234-266

4 Danley L. W., A Metallization System for Microwave and UHF Power Transistors, Solid-State Technology, June 75, pp 35-39

5 Microwave & RF Handbook, AEI Semiconductors 1979

6 Kern W., Shaw J. M., Electrochemical Delineation of Tungsten Films for Microelectronic Devices, J. Electrochem Soc: Electrochemical Technology, Vol 118, No. 10, Oct 71, pp 1699-1704

7 Rodriguez A., Misra M., Hesselbom H. J., Tove P.A., Fabrication of Short Channel MOSFETS with Refractory Metal Gates using RF Sputter Etching, Solid-State Electronics, Vol 19, 1976, pp 17-21

8 Data book compiled by Middlesex Polytechnic, approximate date 1978

9 Parrott J. E., Leonidou L. P., A General Theory of p-n Junction Capacitance, Phys. Stat. Sol (a), Vol 25, 1974, pp 231-240

10 Ilieva M., Kamenova M., Popova A., MOS Structures with "Chlorine" Grown Oxide and Tungsten-Gold Metallization, Thin Solid Films, V30, 1975, pp 281-285

11 Daw A. N., Choudhury N. K. D., Sinha T., On the Variation of Cut-Off Frequency at High Injection Level with Emitter End Concentration of a Diffused Base Transistor, Solid-State Electronics, V17, 1974, pp 1108-1110

12 Kennedy D. P., Spreading Resistance in Cylindrical Semiconductor Devices, Journal of App. Physics, Vol 31, No. 8, Aug 80, pp 1490-1497

13 Maheshwari L. K., Ramanan K. V., The Effect of Doping Dependent Mobility on Base Transit Time in Transistors, Solid-State Electronics, Vol 18, 1975, pp 1142-1144

14 Conti M., Corda G., Diffusivity at High Injection in Epitaxial Power Transistors, Solid-State Electronics, Vol 20, 1977, pp 563-566

15 Zarabi M. J., Satyam M., Some Investigations on Secondary Breakdown in p-n Junctions Considering the Effect of Thermally Generated Carriers, Solid-State Electronics, Vol 20, 1977, pp 407-412

16 Li S. S., Thurber W. R., The Dopant Density and Temperature Dependence of Electron Mobility and Resistivity in n-type Silicon, Solid-State Electronics, Vol 20, 1977, pp 609-616

17 Bosch G., Anomolous Current Distributions in Power Transistors, Solid-State Electronics, Vol 20, 1977, pp 635-640

18 Rey G., Bailby J. P., Marty A., Base Widening into the Emitter Region of an n+npn Bipolar Transistor, Solid-State Electronics, Vol 20, 1977, pp 545-550

19 de Graaff H. C., Slotboom J. W., Schmitz A., The Emitter Efficiency of Bipolar Transitors, Solid-State Electronics, Vol 20, 1977 pp 515-521

20 Guckel H., Thomas D. C., Iyengar S. V., Demirkol A., Transition Region Behaviour in Abrupt Forward-Biased pn-Junctions, Solid-State Electronics, Vol 20, 1977, pp 647-652

21 Wilson B. L. H., The Emitter Efficiency of Silicon Transistors, Solid-State Electronics, Vol 20, 1977, pp 71-74

22 Baccarani G., Jacoboni C., Mazzone A. M., Current Transport in Narrow Base Transistors, Solid-State Electronics, Vol 20, 1977, pp 5-10

23 Barret C., Vapaille A., Interfacial States Spectrum of a Metal-Silicon Junction, Solid-State Electronics, Vol 20, 1977, pp 73-75

24 Unwin R. T., Knott K. F., Comparison of Methods used for Determining Base Spreading Resistance, IEE Proc., Vol 127, Pt 1, No. 2 April 80, pp 53-61

25 Rey G., Bailbe J. P., Some Aspects of Current Gain Variations in Bipolar Transistors, Solid-State Electronics, Vol 17, 1974, pp 1045-1057

26 Whittier R. J., Tremere D. A., Current Gain and Cut-off Frequency Fall-off at High Currents, IRE Trans Electron Devices, Vol ED-16, No. 1, Jan 69, pp 39-57

27 Kirk C. T., A Theory of Transistor Cut-off Frequency (f_T) Fall-off at High Current Densities, IRE Trans Electron Devices Vol ED-9, March 62, pp 164-174

28 Pringle R. D., Microwave Transistor and Monolithic Integrated Circuit Technology, Microelectronics, Vol 6, No. 4, 1975, pp 33-41

29 Abraham H. E., Meyer R. G., Transistor Design for Low Distortion at High Frequencies, IEEE Trans Electron Devices, Vol ED-23, No. 12, Dec 76, pp 1290-97

30 Poon H. C., Implication of Transistor Frequency Dependence on Intermodulation Distortion, IEEE Trans Electron Devices, Vol ED-21, No. 1, Jan 74, pp 110-112

31 Sakai T., Sunohara Y., Sakakibaba Y., Murota J., Stepped Electrode Transistor: SET, Jap Journal of Applied Physics, Proc 8th Conf 1976 Solid State Devices, Tokyo, Vol 16, 1977, Supp 16-1, pp 43-46

32 Tsuchimoto T., Kumagaya F., Itoh K. H., Double Implanted UHF Power Transistor with Epitaxial Overcoat Contact to Shallow Emitter, Solid-State Electronics, Vol 19, 1976, pp 1042-1043

33 Review, Fine Geometry Improves Transistors, Electronics, 2nd Sept, 1976, pp 32-34

34 TRW Semiconductors, UHF Transistor Puts Out 80 W, Electronics, 3rd Oct., 1974, p 151

35 Kwok S. P., UHF Broadband Amplifier Design, Motorola Application Note, AN-406

36 Norstrom H., Experimental and Design Information for Calculating Impedance Matching Networks for use in RF Sputtering an Plasma Chemistry

37 Ning T. H., Isaac R. D., Solomon P. M., Yu H. N., Tang D. D. L., Feth G. C., Wiedmann S. K., Self Aligned Bipolar Transistors for High Performance and Low Power Delay VLSI, IEEE Trans on Electron Devices, Vol ED-28, No. 9, Sept. 81, pp 1010-13

38 Poon H. C., Modelling of Bipolar Transistors using Integral Charge Control Model with Applications to Third Order Distortion Studies, IEEE Trans on Electron Devices, Vol ED19, No. 6, June 72, pp 719-731

39 Mueller O., Ultralinear UHF Power Transistors for CATV Applications, IEEE Proc., Vol 58, No. 7, July 70, pp 1112-1121

40 Narayanan S., Transistor Distortion Analysis using Volterra Series Representation, Bell System Tech J, May-June 1967, pp 991-1024

41 D'altroy F. A., Jacobs R. M., Nacci J. M., Panner E. J., Ultralinear Transistors, Bell System Tech J, Vol 53, No. 10, Dec 74, pp 2195-2202

42 Narayanan S., Poon H. C., An Analysis of Distortion in Bipolar Transistors using Integral Charge Control Model and Volterra Series, IEEE Trans Circuit Theory, Vol CT 20, No. 4, July 73, pp 341-351

43 Ohara M., Akazawa Y., Ishihara N., Konaka., Bipolar Monolithic Amplifiers for a Gigabit Optical Repeater, IEEE Journal Solid State Circ, Vol SC 19, No. 4, Aug 84, pp 491-497

44 Lotsch H., Third Order Distortion and Cross Modulation in a Grounded Emitter Transistor Amplifier, IRE Trans Audio, March 61, pp 49-58

45 Agourdis D. C., Van der Ziel A., Noise Figure of UHF Transistors as a Function of Frequency and Bias Conditions, IEEE Trans on Electron Devices, Vol ED 14, No. 12, Dec 67, pp 808-816

46 Reynolds J., Nonlinear Distortions and Their Cancellation in Transistors, IEEE Trans Electron Devices, Vol ED 12, No. 11, Nov 65, pp 595-599

47 Van der Ziel A., Agourdis D., The Cut-off Frequency Fall-off in UHF Transistors at High Currents, Proc IEEE (letters), Vol 54, March 66, pp 411-412

48 Lotsch H. K. V., Theory of Nonlinear Distortion Produced in a Semiconductor Diode, IEEE Trans Electron Devices, Vol ED 15, No. 5, May 68, pp 294-307

49 Shimizu M., Kitabayashi H., BEST (Base Emitter Self-aligned Technology) A New Fabrication Method for Bipolar LSI, IEDM Tech Digest, 1979, pp 332-335

50 Amantea R., Interpreting the Beta versus Collector Current and Temperature Characteristics of a Transistor, RCA Review, Vol 43, June 82, pp 375-390

51 Wilson P. R., The Emitter-Base Breakdown Voltage of Planar Transistors, Solid-State Electronics, Vol 17, 1974, pp 465-467

52 El-Shandwily M. E., Transit Time in the Base Region of Drift Transistors Considering Recombination and Variable Mobility, Solid-State Electronics, Vol 17, 1974, pp 507-509

53 Van Vliet K. M., Min H.S., Current-Voltage Relations and Equivalent Circuits of Transistors at High Injection Levels, Solid-State Electronics, Vol 17, 1974, pp 267-284

54 Suzuki M., Hagimoto K., Ichino H., Konaka S., A 9-GHz Frequency Divider Using Si Bipolar Super Self-Aligned Process Technology, IEEE Electron Device Letters, Vol EDL-6. No. 4, April 85, pp 181-183

55 Lee D. H., Mayer J. W., Ion-Implanted Semiconductor Devices, IEEE Proc, Vol 62, No. 9, Nov 74, pp 1241-1255

56 News Item, UHF Transistor Offers Flat Gain, Electronics, Vol 50, Feb 17, 77, p 142

57 News Item, Tektronix Develops pnp Microwave Transistor, Electronics, Vol 50, Jan 77, p 42

58 Kronquist R. L., Fourrier J. Y., Pestie J. P., Bilman M. E., Determination of a Microwave Transistor Model Based on an Experimental Study of its Internal Structure, Solid-State Electronics, Vol 18, Nov 75, pp 946-963

59 News Item, Microwave Bipolar Devices Combine Ion Implantation with Isoplanar, Electronics, Vol 46, April 26, 1973, pp 30-31

60 Beale J. R. A., Slatter J. A. G., The f_T of Bipolar Transistors with Thin Lightly Doped Bases, Solid-State Electronics, Vol 19, 1976, pp 549-556

61 Rey G., Dupuy F., Bailbe J. P., A Unified Approach to the Base Widening Mechanisms in Bipolar Transistors, Solid-State Electronics, Vol 18, 1975, pp 863-866

62 Review, Microwave Devices Withstand VSWR, Electronics, Vol 48, March 20, 1975, pp 36 & 38

63 Sakai T., Yamamoto Y., Kobayashi Y., Kawarada K., Inabe Y., Hayashi T., Miyanaga H., A 3ns 1-kbit RAM using Super Self Aligned Process Technology, IEEE J Sol-State Devices, Vol SC 16, No. 5, Oct 81, pp 424-429

64 Mock M. S., Transport Equations in Heavily Doped Silicon, and the Current Gain of a Bipolar Transistor, Solid-State Electronics, Vol 16, 1973, pp 1251-1259

65 Nakata T., Miyazaki S., Shirotori K., 0.5-2.6GHz Si-Monolithic Wideband Amplifier IC, 1985 IEEE MTT-S Digest, pp 55-59

66 Basavaraj T. N., Bhattacharyya A. B., The Effect of Retarding Field on the Base Transport Characteristics of Planar Transistors, Solid-State Electronics, Vol 16, 1973, pp 921-929

67 Daw A. N., Choudhury N. K. D., Gupta N. S., Sinha T., On the Variation of Cut-off Frequency with Emitter End Concentration of a Diffused Base Transistor, Solid-State Electronics, Vol 16, 1973, pp 669-673

68 de Graaff H. C., Collector Models for Bipolar Transistors, Solid-State Electronics, Vol 16, 1973, pp 587-600

69 Barnoski M. K., Loper D. D., Microwave Characteristics of Ion Implanted Bipolar Transistors, Solid-State Electronics, Vol 16, 1973, pp 441-451

70 Bouma B. C., Roelofs A. C., An Experimental Determination of the Forward-Biased Emitter-Base Capacitance, Solid-State Electronics, Vol 21, 1978, pp 833-836

71 Saraswat K. C., Meindl J. D., Breakdown Walk-out in Planar p-n Junctions, Solid-State Electronics, Vol 21, 1978, pp 813-819

72 Berbalk G., Unger B., Experimental Investigations of the Current Dependence of the Base Resistance, Solid-State Electronics, Vol 21, 1978, pp 794-796

73 Choo S. C., Ling M. S., Hong H. L., Li L., Tan L.S., Spreading Resistance Calculations by the Use of Gauss-Laguerre Quadrature, Solid-State Electronics, Vol 21, 1978, pp 769-774

74 Oosaka F., Temperature Dependence of f_T of Silicon Double Diffused Bipolar Transistors, Solid-State Electronics, Vol 21, 1978, pp 665-666

75 Shacter S. B., Van der Ziel A., Chenette E. R., Sutherland A. D., The Effect of Hot Spots on the Noise Characteristics of Large-Area Bipolar Transistors, Solid-State Electronics, Vol 21, 1978, pp 599-602

76 Van der Ziel A., High Injection Noise in Transistors, Solid-State Electronics, Vol 20, 1977, pp 715-720

77 Lan S. S., Mills R. H., Muth P. G., Temperature Rise During Film Deposition by RF and DC Sputtering, J. Vac Si Technol, Vol 9, No. 4, July-Aug 72, pp 1196-1202

78 Cooke H. F., Microwave Transistors: Theory and Design, IEEE Proc, Vol 59, No. 8, Aug 71, pp 1163-1181

79 Ashburn P., Bull C., Nicholas K. H., Booker G. R., Effects of Dislocations in Silicon Transistors with Implanted Bases, Solid-State Electronics, Vol 20, 1977, pp 731-740

80 White M. H., Thurston M. O., Characterization of Microwave Transistors, Solid-State Electronics, Vol 13, 1970, pp 523-542

81 Schuldt S. B., An Exact Derivation of Contact Resistance to Planar Devices, Solid-State Electronics, Vol 21, 1978, pp 715-719

82 Simons K. A., The Decibel Relationships Between Amplifier Distortion Products, IEEE Proc., Vol 58, No. 7, July 70, pp 1071-1085

83 Srivastava A., Bhattacharyya A. B., Determination of Base Recombination Lifetime and Surface Recombination Velocity at N-N+ Interface of Epitaxial Transistors, Solid-State Electronics, Vol 21, 1978, pp 1089-1090

84 Li S. S., The Dopant Density and Temperature Dependence of Hole Mobility and Resistivity in Boron Doped Silicon, Solid State Electronics, Vol 21, 1978, pp 1109-1117

85 Lo T. C., Base Transport Factor Calculations for Transistors with Complementary Error Function and Gaussian Base Doping Profiles, IEEE Trans on Electron Dev, Vol ED 18, No. 4, April 71, pp 243-248

86 Product Feature, 40 W - 2 GHz Power Transistor, Microwave J, Vol 20, May 77, p 59

87 Brown D. M., Cady W. R., Sprague J. W., Salvagni P. J., The P-Channel Refractory Metal Self-Registered MOSFET, IEEE Trans Electron Dev, Vol ED 18, No. 10, Oct 71, pp 931-940

88 Jacoboni C., Canali C., Ottaviani G., Quaranta A. A., A Review of Some Charge Transport Properties of Silicon, Solid-State Electronics, Vol 20, 1977, pp 77-89

89 Hawkins R. J., Limitations of Nielson's and Related Noise Equations Applied to a Microwave Bipolar Transistor and a New Expression for the Frequency and Current Dependent Noise Figure, Solid-State Electronics, Vol 20, 1977, pp 191-196

90 Verwey J. F., Heringa A., de Werdt R., Hofstad W. V. O., Drift of the Breakdown Voltage in p-n Junctions in Silicon (Walk-out), Solid-State Electronics, Vol 20, 1977, pp 689-695

91 Grung B. L., An Analytical Model for the Epitaxial Bipolar Transistor, Solid-State Electronics, Vol 20, 1977, pp 977-984

92 Basararaj T. N., Jindal S., Bhattacharyya A. B., Collector Junction Modelling of Planar Transistors, Solid-State Electronics, Vol 20, 1977, pp 977-984

93 Jaeger R. C., Brodersen A. J., Self Consistent Bipolar Transistor Models for Computer Simulation, Solid-State Electronics, Vol 21, 1978, pp 1269-1272

94 Sideris G., Power from Transistors, Electronics, Vol 46, May 10, 1973, pp 68-69

95 Development Note, Sony Starts Converting Bipolar Devices to New Structure, Electronics, Vol 47, June 27, 1974, p 3E-4E

96 Development Note, Transistor Chips Attain UHF Levels, Electronics, Vol 46, Dec 6, 1973, pp 62-63

97 Product Note, Power Transistor Elevates Emitters, Microwaves, Vol 16, June 77, p 17

98 Max L. B., Balanced Transistors: A New Option, Microwaves, Vol 16, June 77, pp 42-44

99 Product Feature, VMOS: The Next Microwave Power FET, Microwaves, Vol 16, June 77, p 74

100 Peliotis S., MOSFET Development up Power and Frequency to Challenge Bipolars at VHF to Microwave, Microwaves, Vol 16, Oct 77, p 10+

101 Veloric H. S., Presser A., Wozniak F. J., Ultra-thin RF Silicon Transistors with a Copper-Plated Heat Sink, RCA Review, Vol 36, Dec 75, pp 731-743

102 Mitchell B., FETs Challenge Bipolars for Microwave Supremacy, Electronic Engineering, Vol 44, Dec 72, pp 54-55

103 Monroe J. W., de Koning J. G., Kelly W. M., Tokuda H., Spot Compression Points with Equal-Gain Circles, Microwaves, Vol 16, Oct 77, pp 60-77

104 Schetzen M., Nonlinear System Modelling Based on the Wiener Theory, IEEE Proc., Vol 69, No. 12, Dec 81, pp 1557-73

105 Nakata T. A., Kushiyama H., Suzuki H., Low Distortion Microwave Linear Power Transistor, NEC Research & Development, No. 63, Oct 81, pp 54-58

106 Matsumoto S., Okamoto Y., Hashimoto A., Nojima T., 6 GHz SSB-AM Radio Relay System, Jap Telecoms Review, Vol 23, No. 4, Oct 81, pp 361-368

107 Gibbons J. F., Hechtl E. O., Tsurushima T., Ion Bombardment Enhanced Etching of Silicon, App Physics Letters, Vol 15, No. 4, Aug 69, pp 117-119

108 Arora N. D., Hauser J. R., Roulston D. J., Electron and Hole Mobilities in Silicon as a Function of Concentration and Temperature, IEEE Trans on Electron Devices, Vol ED 29, No. 2, Feb 82, pp 222-5

109 Krishna S., The Influence of Post Emitter Processing on the Current Gain of Bipolar Transistors, IEEE Trans on Electron Devices, Vol ED 29, No. 3, March 82, pp 430-435

110 Yoshii A., Kitazawa H. Tomigawa M., Horiguchi S., Sudo T., A Three Dimensional Analysis of Semiconductor Devices, IEEE Trans on Electron Devices, Vol ED 29, No. 2, Feb 82, pp 184-189

111 Dumke W. P., On the Additivity of Ohmic and Space Charge Limited Currents, Solid-State Electronics, Vol 25, No. 2, Feb 82, pp 101-103

112 Guckel H., Thomas D., Demirkol A., Iyengar S., The Forward Biased, Abrupt p-n Junction, Solid-State Electronics, Vol 24, No. 2, Feb 82, pp 105-113

113 Leung C.C., Snapp C.P., Grande V., A 0.5μm Silicon Bipolar Transistor for Low Phase Noise Oscillator Applications up to 20GHz.

114 Marlow G. S., Das M. B., The Effects of Contact Size and Non-zero Metal Resistance on the Determination of Specific Contact Resistance, Solid-State Electronics, Vol 25, No. 2, Feb 82, pp 91-94

115 Rhoderick E. H., Metal Semiconductor Contacts, IEE Proc., Vol 129, Pt 1, No. 1, Feb 82, pp 1-14

116 Meyer R. G., Eschenback R., Chin R., A Wide Band Ultralinear Amplifier from 3 to 300 MHz, IEEE J. Solid-State Circuits, Vol SE 9, No. 4, Aug 74, pp 167-175

117 Duff D. G., Poon H. C., An Analysis of Low Frequency Second Order Distortion in Bipolar Transistors Applied to an Amplifier, IEEE J. Solid-State Circuits, Vol SC 8, No. 6, Dec 73, pp 447-453

118 Narayanan S., Intermodulation Distortion of Cascaded Transistors, IEEE J. Solid-State Circuits, Vol SC 4, No. 3, June 69, pp 97-106

119 Meyer R. G., Shensa M. J., Eschenbach R., Cross Modulation and Intermodulation in Amplifiers at High Frequencies, IEEE J. Solid-State Circuits, Vol SC 7, No. 1, Jan 72, pp 16-23

120 Tsukamoto K., Akasaka Y., Kijima K., Thermal Diffusion of Ion Implanted Arsenic in Silicon, Jap Journal of App Physics, Vol 19, No. 1, Jan 80, pp 87-95

121 Gummel H. K., A Charge Control Relation for Bipolar Transistors, Bell System Tech J, Vol 49, No. 1, Jan 70, pp 115-120

122 Gummel H. K., Poon H. C., An Integral Charge Control Model of Bipolar Transistors, Bell System Tech J, Vol 49, No. 5, May-June 70, pp 827-852

123 Nielsen E. G., Behaviour of Noise Figure in Junction Transistors, Proc IRE, Vol 45, No. 7, July 57, pp 957-963

124 Zandveld P., Crystal Damage and the Properties of Implanted p-n Junctions in Silicon, Solid-State Electronics, Vol 21, 1974, pp 721-727

125 Editorial, Degradation in Microwave Bipolar Transistors, Solid-State Technology, Vol 18, Feb 75, p 29

126 Pellegrini B., Current Voltage Characteristics of Silicon Metallic Silicide Interfaces, Solid-State Electronics, Vol 18, 1975, pp 417-426

127 Roulston D. J., I_C vs V_{BE} Law in Double Diffused Bipolar Transistors, Solid-State Electronics, Vol 18, 1975, pp 427-429

128 Roulston D. J., Low Current Base Collector Boundary Conditions in GHz Frequency Transistors, Solid-State Electronics, Vol 18, 1975, pp 845-847

129 Sandberg I. W., Volterra Expansions for Time Varying Nonlinear Systems, Bell System Tech J, Vol 61, No. 2, Feb 82, pp 201-225

130 Sandberg I. W., Expansions for Nonlinear Systems, Bell System Tech J, Vol 61, No. 2, Feb 82, pp 159-199

131 Livermore R., High Current Testing of Power Semiconductors, Electronic Engineering, April 82, pp 51-59

132 Borrego J. M., Temple V. A. K., Adler M. S., Extension of Gummel's Charge Control Relation, Solid-State Electronics, Vol 20, 1977, pp 441-442

133 Besser L., Swenson S., Update Amplifier Design with Network Synthesis, Microwaves, Oct 77, pp 50-56

134 Coughlin J. B., Gelsing R. J. H., Jochems P. J. W., Van der Laak H. J. M., A Monolithic Silicon Wide Band Amplifier from DC to 1 GHz, IEEE J of Solid-State Circ, Vol SC 8, Nol. 6, Dec 73, pp 414-419

135 Carley D. R., McGeough P. L., O'Brien J. F., The Overlay Transistor, Part I: New Geometry Boosts Power, Electronics, Aug 23, 1965, pp 71-77

136 Donahue D. J., Jacoby B. A., Part II: Putting the Overlay to Work at High Frequencies, Electronics Aug 23, 1965, pp 78-81

137 Eimbinder J., Part III: Combining the Field for Ways to Match Overlay's Performance, Electronics Aug 23, 1965, pp 82-84

138 Kakihara S., Wang P. H., Simple CAD Techniques to Develop High Frequency Transistors, IEEE J Solid-State Circuits, Vol SC 6, No. 4, Aug 71, pp 236-243

139 Ashburn P., Soerowirdjo B., Arsenic Profiles in Bipolar Transistors with Polysilicon Emitters, Solid-State Electronics, Vol 24, 1981, pp 475-476

140 Rustagi S. C., Chattopadhyaya S. K., Emitter Space Charge Layer Transit Time in Bipolar Junction Transistors, Solid-State Electronics, Vol 24, 1981, pp 367-370

141 Elmasry M. I., Roulston D. J., Base Component of Gain and Delay Time in Base Implanted Bipolar Transistors, Solid-State Electronics, Vol 24, 1981, pp 371-375

142 Yu A. Y. C., Electron Tunneling and Contact Resistance of Metal Silicon Contact Barriers, Solid-State Electronics, Vol 13, 1970, pp 239-247

143 Rideout V. L., A Review of the Theory and Technology For Ohmic Contacts to Group III-V Compound Semiconductors, Solid-State Electronics, Vol 18, 1975, pp 541-550

144 Chang C. Y., Fang Y. K., Sze S. M., Specific Contact Resistance of Metal Semiconductor Barriers, Solid-State Electronics, Vol 14, 1971, pp 541-550

145 Gummel H. K., On the Definition of the Cut-off Frequency f_T, Proc IEEE (Letters), Vol 57, Dec 69, p 2159

146 Basset R., McCombs M. D., Production Bipolars Edge Out FETs at 4 GHz, Microwaves, Vol 20, No. 2, Feb 81, pp 43-49

147 Sinha A. K., Electrical Characteristics and Thermal Stability of Platinum Silicide to Silicon Ohmic Contacts Metallized with Tungsten, J. Electrochem Soc: Solid State Science & Technology, Vol 120, Dec 73, pp 1767-1771

148 Kato K., Hivai M., A Design Method of Low Distortion High Frequency Transistors, Electron and Commun. Jap., Vol 58, No. 6, June 75, pp 118-123

149 Hakim E. B., Reich B., RF Power Transistor Reliability Advancement using Refractory Metal Contacts, IEEE Proc., Vol 59, No. 10, Oct 71, pp 1542-3

150 Tang D. D., MacWilliams K. P., Solomon P. M., Effects of Collector Epitaxial Layer on the Switching Speed of High Performance Bipolar Transistors, IEEE Electron Devices Letters, Vol EDL 4, No. 1, Jan 83, pp 17-19

151 Hart B. L., Modelling the Early Effect in Bipolar Transistors, IEEE Journal Solid State Circ, Vol SC 18, No. 1, Feb 83, pp 139-140

152 Tang C. C., Chu J. K., Hess D. W., Plasma Enhanced Deposition of Tungsten, Molybdenum and Tungsten Silicide Films, Solid State Technology, March 83, pp 125-128

153 Narayanan S., Application of Volterra Series to Intermodulation Distortion Analysis of Transistor Feedback Amplifiers, IEEE Trans Circuit Theory, Vol CT 17, No. 4, Nov 70, pp 518-527

154 Naakazato K., Nakamura T., Nakagawa J., Okabe T., Nagata M., A 6GHz ECL Frequency Divider using Sidewall Base Contact Structure, 1985 IEEE International Solid-State Circuits Conference, pp 214-215

155 Javitz, A. E., Materials Science & Technology for Design Engineers

156 Tietz, T. E., Wilson, J. W., Behaviour & Properties of Refractory Metals

157 Smithels, C. J., Tungsten, Chapman & Hall 1952

158 Sze, S. M., Physics of Semiconductor Devices, 2nd Ed., Wiley 1981

159 Getreu, I., Modelling the Bipolar Transistor, Tektronix Inc. 1976

160 Paul, R., Measurement of Transistor Parameters, ILIFFE 1969

161 Shurmer, H. V., Microwave Semiconductor Devices, Pitman Publishing 1971

162 R.C.A., R. F. Power Transistor Manual, R.C.A. 1971

163 Adam, S. F., Microwave Theory and Applications, Prentice-Hall 1969

164 Kuo, F. F., Network Analysis and Synthesis, Wiley 1966

165 Maissel, L. I. ed, Glang, R. ed, Handbook of Thin Film Technology, McGraw-Hill 1970

166 Burger, R. M., Donovan, R. P., Fundamentals of Silicon Device Technology, Prentice-Hall 1968

167 Mittleman, J., Circuit Theory Analysis, ILIFFE 1964

168 S-Parameter Design, Hewlett Packard, App note 154 April 1972

169 Hunter, L. P., Handbook of Semiconductor Electronics, McGraw-Hill 1970

170 Liao, S. Y., Microwave Devices and Circuits, Prentice-Hall 1980

171 Van der Wiele, F., Engle, W. L., Jaspers, P. G., Process and Device Modelling for IC Design, Noordhoff, Nato Advanced Studies Series

172 Smullin, L. D., Haus, H. A., Noise in Electron Devices, Wiley

173 Millman, J., Halkias, C. C., Integrated Electronics, McGraw-Hill 1972

174 Grove, A. S., Physics & Technology of Semiconductor Devices, Wiley 1967

175 SPICE Version 2G.0 User's Guide, 22 Sept 1980

APPENDIX I

Contact Resistance

This appendix addresses the modelling of a metal to silicon contact and is applicable to both the emitter and base contact. The problem is formulated in [114] assuming a transmission line model, Figure A.I.1.

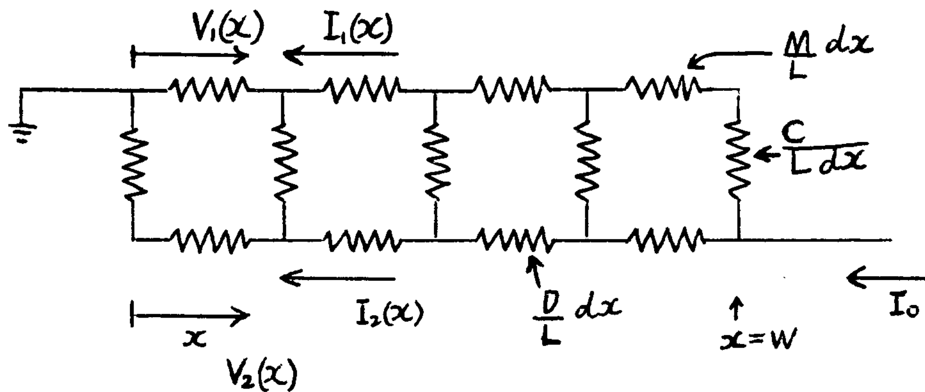


Figure A.I.1 Transmission Line Model with

Non-Zero Metal Resistance

Both the metal and the diffusion are assumed resistive as well as the interfacial layer. Resistance are assumed to be independent of the currents flowing in the contact. The currents and voltages throughout the contact are then related by a pair of simultaneous partial differential equations;

$$D \frac{\delta^2 I_2(x)}{\delta x^2} - DI_2(x) + MI_1(x) = 0$$

$$\frac{\delta I_1(x)}{\delta x} + \frac{\delta I_2(x)}{\delta x} = 0$$

Where;

L is the length of the contact (cm)

M is the metal sheet resistivity (Ω/square)

D is the diffusion sheet resistivity (Ω/square)

C is the specific contact resistance (Ωcm^2)

The general solution to these equations is;

$$I_1(x) = C_1 D/M - C_2 \exp(x/a) - C_3 \exp(-x/a)$$

$$I_2(x) = C_1 + C_2 \exp(x/a) + C_3 \exp(-x/a)$$

where, a, the attenuation constant is given by;

$$a = \left[\frac{C}{M + D}\right]^{1/2} \quad \text{(cm)}$$

Now using boundary conditions;

$$I_1(0) = I_0$$

$$I_1(W) = 0$$

$$I_2(0) = 0$$

$$I_2(W) = I_0$$

Then putting;

$$n = W/a.$$

The solution in [114] assumes n >> 4, and is therefore based on a simplification, but the exact solution is;

$$C_1 = I_0 M/(D + M)$$

$$C_2 = \frac{C_1 (D/M + e^{-n})}{(e^n - e^{-n})}$$

$$C_3 = - \frac{C_1 (D/M + e^n)}{(e^n - e^{-n})}$$

Now substituting these coefficients back into the general solution gives expressions for the currents flowing at any point in the metal or diffusion. Now using the hyperbolic identities, the two reference currents can be expressed as;

$$I_1(x) = I_0 \frac{D\sinh(W/a) - D\sinh(x/a) + M\sinh(W/a - x/a)}{(D + M)\sinh(W/a)}$$

$$I_2(x) = I_0 \frac{D\sinh(x/a) + M\sinh(W/a) - M\sinh(W/a - x/a)}{(D + M)\sinh(W/a)}$$

The voltage drop across the contact is then given by;

$$V_2(W) = \frac{C}{L} \cdot \left. \frac{dI_2(x)}{dx} \right|_W + \frac{M}{L} \cdot \int_0^W I_1(x)dx$$

which, when evaluated, gives;

$$R_C = \frac{a}{L} \cdot \frac{(D^2 + M^2)\cosh n + nDM\sinh n + 2DM}{(D + M)\sinh n}$$

where;

$$R_C = V_2(W)/I_0$$

This is the general solution, notice that it is symmetrical in D and M, the diffusion and metal sheet resistivities. Several special cases can now be considered.

1) Zero metal sheet resistivity, $M = 0$. The equation reduces to;

$$R_c = \frac{a}{L} D\coth n$$

2) Small metal sheet resistivity, D >> M. The equation is then approximately;

$$R_c = \frac{a}{L}\left[D\coth n + \frac{2M}{\sinh n} + Mn\right]$$

where;

$$a \approx (C/D)^{1/2}$$

3) Large contact width (W) or small specific contact resistance (C), W/a = n >> 4. The equation then approximates to;

$$R_c \approx \frac{a(D^2 + M^2 + nMD)}{L(D + M)}$$

the result given in [114].

4) Zero specific contact resistance, C = 0 and a = 0. The equation reduces to that of parallel resistors;

$$R_c = \frac{WDM}{L(D + M)}$$

5) Large specific contact resistance (C) or small contact width (W), W/a = n ≈ 0. The equation reduces to the resistance due to the specific contact resistance alone;

$$R_c = \frac{C}{WL}$$

This has 1% error for W/a = 0.24 and 10% error for W/a = 0.77. In Appendix II it will be shown that a reasonable value for W is the attenuation length, a, and under these conditions the above approximation can give a significant error.

APPENDIX II

Current Distribution Along the Base Finger

Figure A.II.1 shows an idealized base finger where the ohmic resistance in the emitter is regarded as negligible.

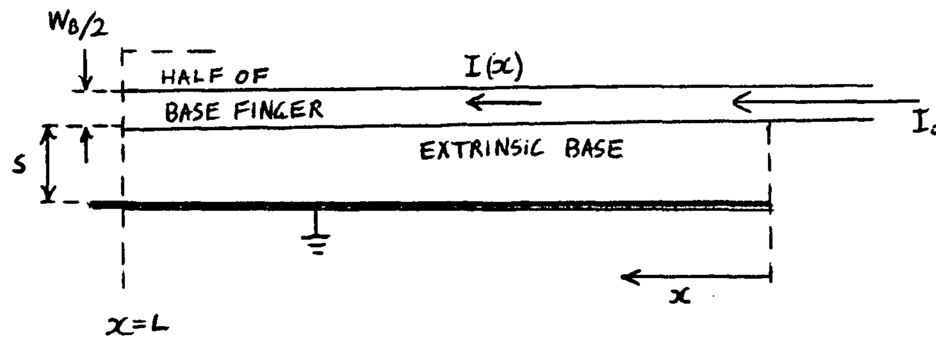


Figure A.II.1 Idealized Base Finger

Now with reference to Appendix I and Chapter 3, the current I(x) is given by;

$$I(x) = I_0 \frac{\sinh(L/a - x/a)}{\sinh(L/a)}$$

and the resistance (R_0) by;

$$R_0 = \frac{aM\coth(L/a) - V_T/I_0}{L}$$

where;

$M = R_{MB} L2/W_B$

R_{MB} = sheet resistivity of base grid

$a = (C/M)^{1/2}$, the attenuation constant

$C = (R_B + V_T/I_o)L^2$

R_B is the sum of the base contact and spreading resistances.

The attenuation constant, a, is a characteristic length describing the loss of current along the base finger. For a base finger of infinite length the current is described by;

$$I(x) = I_o \exp(-x/a)$$

At a distance, x = a, along the finger only 37% of the current flows in the finger, the other 63% having crossed to the emitter-base junction. It is clear, therefore, that the current is not linearly distributed along the length of the junction and little purpose is served by the base finger after a few attenuation constant lengths. A finite finger length has a modified current distribution due to the boundary condition I(L) = 0 and is described by the previous equation.

Current distribution can be characterized by comparing the current in a short length of the junction (dx) at x = 0

and $x = L$.

Differentiating gives;

$$\frac{dI(x)}{dx} = I'(x) = -\frac{I_0 \cosh(L/a - x/a)}{a\sinh(L/a)}$$

and defining the ratio of currents at the ends of the base finger, X_B, as;

$$X_B = I'(0)/I'(L)$$

we get;

$$X_B = \cosh(L/a)$$

Ideally X_B should be unity, which implies an even distribution of current along the length of the junction. In practice, however, some variation is inevitable and an upper limit of 1.5 is acceptable. This corresponds approximately to $L = a$, and requires the sustaining voltage across the junction to vary by about 10mV along the length of the base finger. On this basis the maximum finger length is defined by the attenuation constant;

$$L = a = (C/M)^{1/2}$$

where;

C/M is the ratio of specific resistance between finger and junction (Ωcm) to finger resistivity (Ω/cm).

$$M = R_{MB} L2/W_B$$

and;

$$C = (R_B + V_T/I_o)L^2$$

Largest attenuation constants can be expected with structures using metal base fingers, due to the lower sheet resistivity, while lowest values are given by the conventional overlay. The value of C needs to be large, compared to M, which is counter-intuitive at first sight. For a conventional overlay the two components of C are given by;

$$R_B = R_{BS} S/L$$

$$V_T/I_o > V_T(h_{fe} + 1)/(L \times 60 \times 10^{-6})$$

Where R_{BS} is the sheet resistivity of the extrinsic base and a maximum emitter current of 60µA/µm has been assumed. Now taking;

$R_{BS} > 600\Omega/\text{square}$

$s > 2\mu m$

$V_T = 0.026V$

$h_{fe} + 1 = 50$

we find;

$R_B < 1200/L$

$V_T/I_o > 21666/L > 18xR_B$

It is apparent that the small signal emitter resistance, seen from the base, is the dominant component and R_B can in fact be small in comparison.

An estimate of the maximum base grid length for a conventional overlay transistor can now be obtained from the foregoing equations, viz;

$$L = (10833xW_B/R_{MB})^{1/2}$$

which for;

$R_{MB} = 30\Omega/\text{square}$

and;

W_B = 2μm

gives;

L = 27μm (total base grid length is twice this)

For a transistor using a metal finger for the base contact the lengths will be correspondingly larger, viz, tungsten base metal; R_{MB} < 2Ω/square, L > 104 and interdigitated (aluminium); R_{MB} < 0.05Ω/square, L > 658. The latter figure gives some indication of why the base resistance of an interdigitated transistor is relatively bias independent.

The variation of base resistance with attenuation constant can be determined by differentiating R_0, noting that dM/da = -2M/a for C held constant, to give;

$$\frac{dR_0}{da} = \frac{M\,\mathrm{sech}^2(L/a)}{a} - \frac{M\coth(L/a)}{L}$$

which for L = a gives -0.589M/a showing that the base resistance reduces with an increase in attenuation constant.

Variation with current can be obtained by differentiating with respect to I_0, using;

$$\frac{da}{dI_0} = - \frac{L^2 V_T}{2aMI_0{}^2}$$

to give;

$$\frac{dR_0}{dI_0} = \left[1 - \frac{L^2 \operatorname{sech}^2 (L/a)}{2a^2} - \frac{L\coth(L/a)}{2a} \right] \frac{V_T}{I_0{}^2}$$

which for L = a gives $-0.0185V_T/I_0{}^2$. For larger values of L the resistance will reduce at a faster rate with increasing I_0 which means that the current will be crowded to a greater extent. Under these conditions base push-out will occur at lower currents than would be expected from consideration of the emitter size alone. It is clear that a contact which is longer than one attenuation length adds additional parasitic components in greater proportion than the effective increase in junction size.

APPENDIX III

Volterra Series

A rigorous mathematical discussion of Volterra series is beyond the scope of this thesis. This appendix however reviews the subject and reports the results which are of most importance to the nonlinear analysis of bipolar transistors. For a comprehensive introduction to the mathematics the reader is recommended to read [104] which is an invited paper forming a tutorial on nonlinear modelling methods. Alternatively [30, 40, 118, 119] show the application of the technique to transistor analysis. The important result that eventually crystalized out was that in the high frequency limit the nonlinear performance is related to the derivatives of the f_T versus I_C curve. This point is discussed in Appendix IV and has much simplified the distortion analysis of transistors.

The major stumbling block to understanding nonlinear effects at high frequency is the fact that the instantaneous output is a function of the history of the input. If a nonlinear system has no memory element then the instantaneous output will only be a function of the instantaneous input and the static transfer characteristic will completely describe the system. In circuit terms this is equivalent to a circuit which is completely described by a DC transfer curve. Such a circuit could be

analysed by evaluating the output that corresponds to an input signal which consists of one or more sinusoidal components. A Fourier transform of this would then yield the frequency components present in the output, and so determine the harmonic distortion due to the circuit. If the transfer function can be represented as a polynomial the analysis may be carried out analytically. This is done by substituting a sinusoidal function of an independent variable, i.e. time, into the transfer function and then using the trigonometric identities to resolve the frequency components. This method is discussed in some detail in Chapter 4 where it is used to evaluate various harmonic and intermodulation distortions produced by transistors. This type of approach however is of no use if the system has some form of memory.

The type of memory under discussion is generally some form of storage such as an inductor or a capacitor with a finite storage time. Latch type elements are not included in this definition, such devices have infinite storage time and constitute a different class of problem.

If a square wave were to be applied to the input of such a system the output would be distorted because it would depend on the recent history of the input and not the instantaneous input. In order to characterize the system a parallel may be drawn from the no memory case. Consider first a linear system, here the system would be

characterized by the output for a unit input, i.e. if the input doubled then the output would also double. The output is then simply a linear multiple of the output due to the unit input. In the same way a time varying input can be considered as a linear combination of unit impulses. In order for this to be the case the impulse must be of zero duration, however if it had a finite amplitude then the area it enclosed would be zero. For this reason it is defined in terms of the area enclosed, and thus has infinite amplitude. In order to evaluate the output for a given input at any instant the individual outputs for each impulse in the input wave must be summed. This process is called convolution and is quite simply the sum effect at the output due to each earlier input impulse. The integral actually performs a summation of the instantaneous products of the input and the impulse response i.e. the sum effect at the output due to the input at every instant. It is in fact a multiplication, of the input and the impulse response, in the transform domain and is also known as a product in the mean.

For nonlinear systems a single parameter is no longer enough to fully describe the system. For the case where no memory is involved the system can often be described by a polynomial. It should be noted however that this is not always the case, it is a necessary condition that the function is both continuous and differentiable at every point. Transfer characteristics of transistors do however

fall into this class and may therefore be represented by polynomial expressions. The first order coefficient then represents the linear gain while the higher order coefficients represent higher order nonlinear gain components. These ideas are discussed in more detail in Chapter 4 where the Taylor expansion is used to obtain polynomial expansions for various transfer characteristics.

When a system has memory the linear impulse response is insufficient to describe it and a higher order expansion must then be used. The linear impulse response, which is called the first order kernel, must be augmented with higher order kernels i.e. higher order impulse responses. The expression thus formed is equivalent to the polynomial expression with each coefficient replaced by a kernel and each product replaced by a product in the mean. In other words it is a polynomial in the transform domain. The series is called a Volterra series and can be considered as the general case of the Taylor series. In exactly the same way as for the Taylor expansion there are conditions which must be met before the Volterra expansion is valid. The system must be time invariant for example, which means that the kernels do not vary with time. The series must be convergent to be valid, which in simplified terms requires the higher order terms to diminish in magnitude. Both of these conditions must in fact hold for the Taylor expansion to be valid. The Volterra series may be written

as;

$$y(t) = h_0 + \int_{-\infty}^{\infty} h_1(\tau_1)x(t-\tau_1)d\tau_1$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_1\tau_2$$

$$+ \quad + \quad +$$

Where the nth degree kernel is represented by h_n and the nth order product in the mean is the nth order integral. The system output with no input is h_0, and the linear response of the system is described by the first degree kernel h_1 i.e. the impulse response. The linear output of the system at any time is the convolution of the first degree kernel, h_1, and the input signal, x(t). In the same fashion the second degree kernel, h_2, describes the systems 'square term' the two dimensional convolution of h_2 with the two dimensional input describes the second order distortions. Generally an nth order convolution is carried out on the h_n kernel with an n dimensional input. The variables of integration, τ, can be considered as filters or sieves, any positive value of τ selects a value of the input variable which occurred in the past and a value of the kernel, the product of which gives the part of the present output which is due to the input at time τ in the past.

Electrical circuits can be described [104] with a reduced Volterra series. Firstly, the output is not dependent in any way on future inputs to the system. Such a system is known as a causal system and the integrals need only be evaluated from a lower bound of zero, i.e. positive values of τ, since the integral to this point is zero. Secondly, the memory is not infinite so that the upper bound can be replaced by some finite time which will correspond to the system settling time. Thirdly, the kernels are symmetrical, i.e. $h_2(\tau_1, \tau_2) = h_2(\tau_2, \tau_1)$.

The series may be expressed in the transform domain by;

$$y(t) = A_1(\tau_1)*x(t) + A_2(\tau_1,\tau_2)*x^2(t) + +$$

where A_n represents the nth degree kernel and * represents the product in the mean. Here the parallels with the Taylor series can be clearly seen, and in fact the response of the system may be analysed in the same way. If the input signal, x(t), is replaced by a sum of sinusoidal signals then a Fourier transform will enable the resulting harmonic and intermodulation components in the output to be evaluated. Such a transform may be effected prior to any substitution by use of a Fourier transform or a Laplace transform, the latter being the general case of the former. The system response may then be expressed in the frequency domain by;

$$y(w) = A_1(jw_1)*x(w) + A_2(jw_1, jw_2)*x^2(w) + +$$

where the nth degree frequency domain kernel, A_n, is an n dimensional vector which results from the n dimensional transformation of the nth degree time domain kernel, h_n. Its effect is to alter all the sinusoidal components of $x^n(w)$ by its magnitude and phase. Generally evaluating this expression is quite complex but in Appendix IV it is seen that for a transistor operating above it's f_β the result can be relatively simple.

APPENDIX IV

High Frequency Limit

Using the high frequency hybrid-Π model [119], Figure A.IV.1.

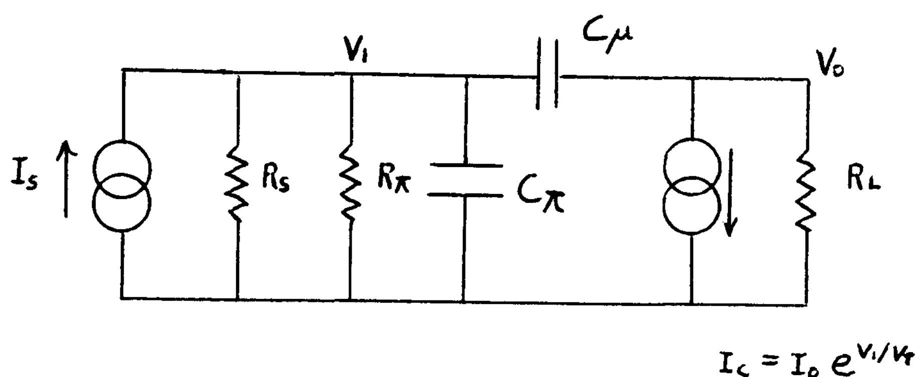


Figure A.IV.1. High Frequency Hybrid-Π Model

We may write;

$$i_1 = \frac{v_1}{R_S} + \frac{v_1}{r_{in}(I_C)} + \frac{dQ_B}{dt} + C_C\frac{d(v_1 - v_C)}{dt}$$

Where Q_B is the mobile base charge plus charge in the emitter transition region. At sufficiently high frequency, $f >> 1/2\Pi R_{EQ}C_C$, where $R_{EQ} = R_S // R_{in}$ is the equivalent input impedance due to the the parallel combination of R_S and R_{in}, the reactive terms dominate, giving;

$$i_1 \approx \frac{dQ_B}{dt} + C_C \frac{d(v_1 - v_C)}{dt}$$

Now using this and the charge control relationship [122];

$$\tau_F \frac{di_C}{dt} = \frac{dQ_B}{dt} + C_C \frac{dv_1}{dt}$$

Where τ_F is the transit time of the intrinsic device gives;

$$i_1 = \tau_F \frac{di_C}{dt} - C_C \frac{dv_C}{dt}$$

then substituting $v_C = -i_C R_C$ we have [116, 119, 29];

$$i_1 = (\tau_F(I_C) + C_C(I_C) R_C) \frac{di_C}{dt}$$

Which relates the input signal current, i_1, to the collector current, i_C, by $\tau_T(I_C)$ the total forward delay of the transistor.

An appropriate application of this is to use;

$$\tau_T = C_E r_E + \tau_B + C_C R_C$$

where;

$$C_C = C_{C0}/(1 + (V_{CB} + v_C)/\phi)^{1/n}$$

$$r_E = \alpha V_T/I_C \approx V_T/I_C$$

$$v_C = -i_C R_C$$

and;

τ_B the base transit time is taken as constant (until the onset of base push-out).

Then τ_T can be expanded as a Taylor series in i_C about the quiescent point I_Q to give;

$$i_1 = (\tau_0 + 2\tau_1 i_C + 3\tau_2 i_C^2 + +)\frac{di_C}{dt}$$

where;

$$\tau_0 = \tau_T$$

$$\tau_1 = \frac{1}{2!}\cdot\frac{d\tau_0}{di_C}$$

$$\tau_2 = \frac{1}{3!}\cdot\frac{d^2\tau_0}{di_C^2}$$

$$\tau_3 = \frac{1}{4!} \cdot \frac{d^3 \tau_0}{di_c^3}$$

and generally;

$$\tau_n = \frac{1}{(1+n)!} \cdot \frac{d^n \tau_0}{di_c^n}$$

The derivatives are found from;

$$\tau_T = C_E r_E + C_C R_C + \tau_B$$

$$= \frac{C_E V_T}{I_C} + \frac{C_{C0} R_C}{(1+V_{CB}/\phi)^{1/n}} + \tau_B$$

and now using $dV_{CB} \approx -R_C dI_C$ we get;

$$\frac{d\tau_F}{dI_C} = -\frac{C_E V_T}{I_C^2} + \frac{C_C R_C^2 1/n}{(\phi+V_{CB})}$$

Or, since $\phi + V_{CB} \approx V_{CE}$ we may write;

$$\frac{d\tau_F}{dI_C} \approx -\frac{C_E V_T}{I_C^2} + \frac{C_C R_C^2 1/n}{V_{CE}}$$

$$\frac{d^2 \tau_F}{dI_C^2} \approx \frac{2C_E V_T}{I_C^3} + \frac{C_C R_C^3 1/n(1/n+1)}{V_{CE}^2}$$

$$\frac{d^3 \tau_F}{dI_C^3} \approx -\frac{6C_E V_T}{I_C^4} + \frac{C_C R_C^4 1/n(1/n+1)(1/n+2)}{V_{CE}^3}$$

$$\frac{d^4 \tau_F}{dI_C^4} \approx \frac{24C_E V_T}{I_C^5} + \frac{C_C R_C^5 1/n(1/n+1)(1/n+2)(1/n+3)}{V_{CE}^4}$$

Where V_{CE} is the quiescent voltage of the collector with respect to the emitter and C_C is the collector junction capacitance at this bias.

Collector current can be expressed as a function of the input current by a Volterra series;

$$I_C = A_0 + A_1 * i_1 + A_2 * i_1^2 + A_3 * i_1^3 + +$$

where the high frequency Volterra kernels are related to the derivatives of the forward delay as shown in Appendix V.

Substituting gives

$$A_1 = \frac{1}{\tau_0 jw_1} = \frac{1}{\tau_F jw_1} = \frac{1}{jw_1 (C_E r_E + C_C R_C + \tau_B)}$$

$$A_2 = \frac{\tau_1}{\tau_0^3 w_1 w_2} = \frac{-C_E V_T / I_C^2 + C_C R_C^2 / (nV_{CE})}{2w_1 w_2 (C_E r_E + C_C R_C + \tau_B)^3}$$

$$A_3 = \frac{\tau_2 \tau_0 - 2\tau_1^2}{\tau_0^5 jw_1 w_2 w_3}$$

$$= \frac{\dfrac{C_E V_T}{3I_C^3} + \dfrac{C_C R_C^3 (1/n+1)/n}{6V_{CE}^2}}{(C_E r_E + \tau_B + C_C R_C)^4 jw_1 w_2 w_3} - \frac{\dfrac{1}{2}\left[\dfrac{C_E V_T}{I_C^2} - \dfrac{C_C R_C^2/n}{V_{CE}}\right]^2}{(C_E r_E + \tau_B + C_C R_C)^5 jw_1 w_2 w_3}$$

The figures of merit are given by;

$$M2 = A_2 A_0 / A_1^2 = A_0 \tau_1 / \tau_0$$

$$M3 = A_3 A_0^2 / A_1^3 = A_0^2 (\tau_2 \tau_0 - 2\tau_1^2) / \tau_0^2$$

$$M4 = A_4 A_0^3 / A_1^4 = A_0^3 (5\tau_0 \tau_1 \tau_2 - 5\tau_1^3 - \tau_0^2 \tau_3) / \tau_0^3$$

$$M5 = A_5 A_0^4 / A_1^5$$

$$= A_0^4 (14\tau_1^4 + 6\tau_0^2 \tau_1 \tau_3 + 3\tau_0^2 \tau_2^2 - 21\tau_0 \tau_1^2 \tau_2 - \tau_0^3 \tau_4) / \tau_0^4$$

In terms of circuit parameters the figures of merit are;

$$M2 = \frac{I_C (C_E V_T / I_C^2 - C_C R_C^2 / (nV_{CE}))}{2(C_E V_T / I_C + C_C R_C + \tau_B)}$$

$$M3 = \frac{I_C^2}{3} \left[\frac{C_E V_T / I_C^3 + C_C R_C^3 (1/n+1)/(2nV_{CE}^2)}{C_E V_T / I_C + C_C R_C + \tau_B}\right]$$

$$- \frac{I_C^2}{2} \left[\frac{C_E V_T / I_C^2 - C_C R_C^2 / (n V_{CE})}{C_E V_T / I_C + C_C R_C + \tau_B} \right]^2$$

where;

$$C_C = \frac{C_{C0}}{(1 + V_{CB}/\phi)^{1/n}}$$

$$V_{CC} = V_{CC} - I_C R_C$$

and;

$$V_{CB} \approx V_{CC} - I_C R_C - \phi$$

APPENDIX V

Fifth Order Expansion for Volterra Kernels

The basic procedure carried out in this appendix is to start from a non-linear relationship between the generator current or voltage and the output current or voltage, i.e.

$$i_1 = \tau_T(I_C)\frac{di_C}{dt}$$

The non-linear function, τ_T, is then expanded as a power series of the current. Then a Volterra series form is assumed for the current of the form;

$$i_C = A_1(jw_1)*i_1 + A_2(jw_1, jw_2)*i_1^2 + +$$

This equation is then substituted into the first equation and expanded. Since the powers of i_C form an orthogonal set, like terms can be collected together to give equations for each power. The first order kernel is evaluated and substituted into the expression for the second order kernel. The higher order kernels can be expressed in terms of the lower order terms and can then be evaluated by means of substitution.

The input current may be related to the collector current by the non-linear equation [116, 119, 29];

$$i_1 = (\tau_F(I_C) + C_C(I_C)R_C)\frac{di_C}{dt}$$

or more generally;

$$i_1 = \tau_T(I_C)\frac{di_C}{dt}$$

expanding $\tau_T(I_C)$ about a quiescent current I_Q gives;

$$i_1 = (\tau_T(I_Q) + \tau_T^{(1)}(I_Q)i_C + \tau_T^{(2)}(I_Q)i_C^2/2! + +)\frac{di_C}{dt}$$

$$= (\tau_0 + 2\tau_1 i_C + 3\tau_2 i_C^2 + +)\frac{di_C}{dt}$$

where;

$$i_C = I_C - I_Q$$

$\tau_T^{(n)}(I_Q)$ is the nth derivative of τ_T at the bias current I_Q.

and now defining;

$$\tau_n = \frac{1}{(1+n)!}\frac{d^n\tau_T}{di_C^n}$$

and then using;

$$(n+1) i_c^n \frac{di_c}{dt} = \frac{di_c^{n+1}}{dt}$$

we get;

$$i_i = \tau_0 \frac{di_c}{dt} + \tau_1 \frac{di_c^2}{dt} + \tau_2 \frac{di_c^3}{dt} + +$$

Now assume i_c can be expressed in terms of Volterra kernels by;

$$i_c = A_1(jw_1)*i_i + A_2(jw_1, jw_2)*i_i^2 + A_3(jw_1, jw_2, jw_3)*i_i^3 +$$

where A_n is a vector which alters all the frequency components of I_i^n by its magnitude and phase. Now substituting into the former expression, expanding and collecting terms of equal order, we get for first order

$$i_i = \tau_0 \frac{d(A_1 i_i)}{dt} = \tau_0 A_1 \frac{d(\exp(jw_1 t))}{dt} = \tau_0 A_1 jw_1 i_i$$

giving;

$$A_1 = \frac{1}{\tau_0 jw_1}$$

For second order;

$$0 = \tau_0 \frac{d(A_2 i_1^2)}{dt} + \tau_1 \frac{d(A_1 i_1)^2}{dt}$$

$$= (\tau_0 A_2 + \tau_1 A_1^2) \frac{di_1^2}{dt}$$

giving;

$$A_2 = - \frac{\tau_1}{\tau_0} A_1^2 = \frac{\tau_1}{\tau_0^3 w_1 w_2}$$

For completeness, the expansions of i_1 are given here up to fifth order. We have;

$$i_1 = A_1 i_1 + A_2 i_1^2 + A_3 i_1^3 + A_4 i_1^4 + A_5 i_1^5 + +$$

$$i_1^2 = A_1^2 i_1^2 + 2A_1 A_2 i_1^3 + (2A_1 A_3 + A_2^2) i_1^4$$

$$+ (2A_1 A_4 + 2A_2 A_3) i_1^5 + +$$

$$i_1^3 = A_1^3 i_1^3 + 3A_1^2 A_2 i_1^4 + (3A_1^2 A_3 + 3A_1 A_2^2) i_1^5 + +$$

$$i_1^4 = A_1^4 i_1^4 + 4A_1^3 A_2 i_1^5 + +$$

$$i_1^5 = A_1^5 i_1^5 + +$$

and now gathering third order terms gives;

$$0 = (\tau_0 A_3 + 2\tau_1 A_1 A_2 + \tau_2 A_1{}^3) \frac{di_1{}^3}{dt}$$

giving;

$$A_3 = -\frac{\tau_2 A_1{}^3}{\tau_0} - \frac{2\tau_1 A_1 A_2}{\tau_0}$$

and substituting gives;

$$A_3 = \frac{\tau_2 \tau_0 - 2\tau_1{}^2}{\tau_0{}^5 jW_1 W_2 W_3}$$

The fourth and fifth order terms may be evaluated to give;

$$A_4 = \frac{5\tau_0 \tau_1 \tau_2 - 5\tau_1{}^3 - \tau_0{}^2 \tau_3}{\tau_0{}^7 W_1 W_2 W_3 W_4}$$

and;

$$A_5 = \frac{14\tau_1{}^4 + 6\tau_0{}^2 \tau_1 \tau_3 + 3\tau_0{}^2 \tau_2{}^2 - 21\tau_0 \tau_1{}^2 \tau_2 - \tau_0{}^3 \tau_4}{\tau_0{}^9 jW_1 W_2 W_3 W_4 W_5}$$

APPENDIX VI

Comparison of Approaches at High Frequency

According to Abraham and Meyer [29] the Volterra kernels, to third order, are given by;

$$A_1 = \frac{1}{jw_1 \tau_0}$$

$$A_2 = \frac{\tau_1}{2w_1 w_2 \tau_0^2}$$

$$A_3 = \frac{\tau_2/3 - \tau_1^2/(2\tau_0)}{jw_1 w_2 w_3 \tau_0^4}$$

where;

$$\tau_n = \left. \frac{\tau_F^{(n)}(I_C)}{(n+1)!(-R_C)^{n+1}} \right| n = 0,\ 1,\ 2$$

and the load resistor R_C appears due to a conversion from collector current to collector voltage. This type of analysis starts from the high frequency model after the high frequency simplifications have been made.

Narayanan and Poon [30, 42] started from a general case considering variation in f_T, h_{fe} and exponential non-

linearity. The general solutions were then simplified for the high frequency limit to give;

$$i_1 = w_T$$

$$i_2 = \frac{1}{2w_T w_T^{(1)}}$$

$$i_3 = \frac{1}{6w_T (w_T w_T^{(2)} + (w_T^{(1)})^2)}$$

where;

$$w_T = 1/\tau_T$$

and;

$w_T^{(n)}$ is the nth derivative of w_T with respect to I_C.

It is now a fairly simple exercise to make the substitution $\tau_T = 1/w_T$ in the previous expressions to obtain the kernels in terms of w_T and its derivatives. This is done without making the conversion to voltage, since the output is considered as a current by Poon. The result of this substitution is;

$$i_1 = jw_1 A_1$$

$$i_2 = w_1 w_2 A_2$$

$$i_3 = j w_1 w_2 w_3 A_3$$

The ratios used in [30] are identical to those used in [29] viz;

$$\frac{i_1^{\,2}}{i_2} = \frac{A_1^{\,2}}{A_2}$$

and;

$$\frac{i_1^{\,3}}{i_3} = \frac{A_1^{\,3}}{A_3}$$

The two approaches are seen to give identical results.

APPENDIX VII

Figures of Merit for Linearity

When a transfer function can be represented by a power series;

$$I_C = a_0 + a_1 v + a_2 v^2 + +$$

the a_1 term can be seen to represent the linear gain, while higher order terms can be considered as non-linear distortions.

To determine the distortion performance to the transfer function it is necessary to compare the second and higher order coefficients to the first order coefficient, a_1. Due to the non-linear nature of the function the ratio of any non-linear output to linear output is not constant. For such a ratio to be meaningful it must be defined with reference to some standard value. It is common therefore to use as a figure of merit the ratio of non-linear output power to linear output power when the linear power is at some reference level. The reference level is conventionally 1mW. If the output, y, represents collector current then the small signal collector current, i_c, is;

$$i_c = I_C - a_0$$

and, for a sinusoidal signal, the linear output power is;

$$P_o = \frac{i_c^2 R_c}{2}$$

where R_c is the collector load resistance. The corresponding input voltage, v, is given by;

$$v = \frac{2P_o}{a_1 R_c^2}$$

Now to investigate the second order component assume the general case where the input signal is the sum of two sinusoids, f_1, f_2, of amplitude, A, B, the a_2 coefficient produces components in the output current of

$a_2 AB$ at frequency $f_1 \pm f_2$

$a_2 A^2/2$ at frequency $2f_1$

$a_2 B^2/2$ at frequency $2f_2$

The intermodulation component, $f_1 \pm f_2$, is proportional to each of the two linear terms;

$a_1 A$ at frequency f_1

$a_1 B$ at frequency f_2

If the input amplitudes A and B are both at the level to give the reference power at the output then the second order intermodulation component has amplitude;

$$P_{o2} = a_2 \left[\frac{2P_o}{a_1 R_c^2} \right]^2$$

The ratio can be defined as;

$$\frac{P_{o2}}{P_o^2}$$

which as a decibel relationship is;

$$IM2 = 20\log\left[\frac{a_2}{a_1^2} \left[\frac{0.002}{R_c} \right]^{1/2} \right] \quad dBm$$

The equivalent third order relationship is;

$$IM3 = 20\log\left[\frac{a_3\, 0.002}{a_1^3 \; R_c} \right] \quad dBm$$

This being for the case when the power series describes the output current. The figures of merit are defined in dBm since the ratios represent actual powers, the

reference power is 1mW and R_c is the load resistance. In the event that the outputs are expressed as voltages R_c is replaced by $1/R_c$ in the expressions.

The expressions can also be written as;

$$\begin{aligned} IM2 = &\ (\text{power at } f1 \pm f2)\ \text{dBm} \\ &- (\text{power at } f1)\ \text{dBm} \\ &- (\text{power at } f2)\ \text{dBm} \end{aligned}$$

with a similar expression for IM3 which can easily be extended to IM4 and higher order.

The various frequency components have different magnitudes, M, as detailed in Figure 4.4.1, Chapter 4, consequently the measured power will in general be IM(N) + 20log(M).

The simplest way of expressing the ratio of coefficients is seen inside the brackets of the above expressions. The factor;

$$\left[\frac{0.002}{R_c}\right]^{1/2}$$

has units of current and represents the 1mW reference power level. For convenience a second figure of merit is defined, for use in this thesis, which replaces the above expression with a_0 to give;

$$M2 = a_2 a_0 / a_1^2$$

$$M3 = a_3 a_0^2 / a_1^3$$

$$M4 = a_4 a_0^3 / a_1^4$$

$$M5 = a_5 a_0^4 / a_1^5$$

These are normalized, to the bias current, figures of merit and for a given order of distortion allow a relative comparison between causes.

IM2 is related to M2 by;

$$IM2 = 20\log\left[\frac{M2}{a_0}\left[\frac{0.002}{R_c}\right]^{1/2}\right] \quad \text{dBm}$$

and similarly;

$$IM3 = 20\log\left[\frac{M3}{a_0^2}\frac{0.002}{R_c}\right] \quad \text{dBm}$$

It is useful to be able to relate the distortion to the linear output signal. In the case of second order distortion we have $a_2 x^2 / a_1 x$. The substitution $I_L = a_1 x$ can be made to give, $a_2 I_L / a_1^2$, where I_L is the magnitude at the output of each of the linear components involved. This may also be expressed in terms of M2 to give;

$$D2 = \frac{M2 I_L}{a_0} = \frac{\text{Second order distortion}}{\text{M . Linear output } (I_L)}$$

Where M is the magnitude of the distortion considered, as indicated in Figure 4.4.1, Chapter 4.

Similar terms can be written for the higher order distortions;

$$D3 = \frac{M3 I_L^2}{a_0^2}$$

$$D4 = \frac{M4 I_L^3}{a_0^3}$$

$$D5 = \frac{M4 I_L^4}{a_0^4}$$

These relationships clearly demonstrate the disproportionate increase in distortion level as the amplitude of the fundamental increases. It should be

noted that the linear output term does not include any gain compression effects, the total expression is $a_1 x + 3a_3 x^3/4 + 5a_5 x^5/8$ where x is a sinusoidal input. This is a good approximation however a 1% gain compression being equivalent to third harmonic at -50dB or fifth harmonic at -60dB.

Intercept power is often quoted as a figure of merit. This refers to the output power at which the projected distortion product power is equal in magnitude to the projected linear output power. For example if the input power is increased by 2dB the linear signal at the output will increase by 2dB and the second order distortion products will increase by 4dB, in a similar fashion, nth order products increase by 2n dB.

The intercept power is simply related to the intermodulation figure of merit. Since IM(n) = nth order product power - (sum of n linear powers).

Now the intercept power P_{int} is that power at which the nth order product power, at the output, is equal to the linear power at the output. This gives;

$$IM(n) = (1 - n)P_{int} + M \quad \text{dBm}$$

where M is a factor which accounts for the actual product quoted. For example, the case of one tone at the input

and the intercept of the linear power and the third harmonic power quoted M will be 20log4. Where this term corrects for the 1/4 factor in the third harmonic, Figure 4.3.1.

All of the definitions made in this appendix are valid in the high frequency limit. Equations relating the collector current (I_c) to the generator current (I_1) are the same form for low and high frequency. Thus for the high frequency case each coefficient is simply replaced by the equivalent kernel, i.e. a_1 becomes $A_1(jw_1)$. In the case of a_0 this is the quiescent output level and is identical for both cases, i.e. $A_0 = a_0$.

For example M2 becomes;

$$M2 = A_2 A_0 / A_1^2$$

The justification for this is more subtle since we should write;

$$M2 = A_2(jw_1, jw_2) A_0 / A_1(jw_1) . A_1(jw_2)$$

Now by noting that each kernel has all of its characterizing frequencies in the denominator we see that cancellation occurs. This leads to the observation that the distortion is independent of frequency when the high frequency approximation is valid.

The relationship between frequency components obtained by Fourier analysis can easily be related to the figures of merit.

From Figure 4.3.1 it can be seen that the amplitudes of the various frequency components are given by;

$$f_0 = a_0 + a_2A^2/2 + 3a_4A^4/8$$

$$f_1 = a_1A + 3a_3A^3/4 + 5a_5A^5/8$$

$$f_2 = a_2A^2/2 + a_4A^4/2$$

$$f_3 = a_3A^3/4 + 5a_5A^5/16$$

$$f_4 = a_4A^4/8$$

$$f_5 = a_5A^5/16$$

Where terms have been restricted to fifth order and below, A is the amplitude of the input signal, f_n is the amplitude of the n times frequency product and f_0 is defined as the DC component. The coefficients can now be given in terms of the frequency components;

$$a_0 = f_0 - f_2 + f_4$$

$$a_1 = (f_1 - 3f_3 + 5f_5)/A$$

$$a_2 = (2f_2 - 8f_4)/A^2$$

$$a_3 = (4f_3 - 20f_5)/A^3$$

$$a_4 = 8f_4/A^4$$

$$a_5 = 16f_5/A^5$$

Figures of merit can now be expressed in terms of the frequency components to give;

$$M2 = \frac{(2f_2 - 8f_4)(f_0 - f_2 + f_4)}{(f_1 - 3f_3 + 5f_5)^2}$$

$$M3 = \frac{(4f_3 - 20f_5)(f_0 - f_2 + f_4)^2}{(f_1 - 3f_3 + 5f_5)^3}$$

$$M4 = \frac{8f_4 \ (f_0 - f_2 + f_4)^3}{(f_1 - 3f_3 + 5f_5)^4}$$

$$M5 = \frac{16f_5 \ (f_0 - f_2 + f_4)^4}{(f_1 - 3f_3 + 5f_5)^5}$$

For a system with small non-linearities these reduce to;

$$M2 \approx 2f_2 f_0/f_1{}^2$$

$$M3 \approx 4f_3 f_0{}^2/f_1{}^3$$

$M4 \approx 8f_4 f_0^3/f_1^4$

$M5 \approx 16f_5 f_0^4/f_1^5$

These equations also hold for the high frequency case where f_1 is above the 3dB frequency. Under these conditions each kernel is of the form $A_n/(jw)^n$ and A_0 is defined as the DC component. The frequency dependence cancels out in the final expressions due to the dimensionless nature of the figures of merit.

APPENDIX VIII

Fourth and Fifth Order Exponential Distortion

Starting from;

$$\frac{d^3 I_C}{dV_G{}^3} = \frac{3V_T{}^2}{R_X{}^5 I_C{}^4} - \frac{2V_T}{R_X{}^4 I_C{}^3} = a_3 .3!$$

and;

$$\frac{dI_C}{dV_G} = \frac{1}{R_X}$$

where;

$$R_X = \frac{R_B}{h_{fe}} + \frac{R_E(1 + h_{fe})}{h_{fe}} + \frac{V_T}{I_C}$$

It is a simple matter to obtain the fourth and fifth derivatives at $I_C = I_Q$, these equal $a_4 .4!$ and $a_5 .5!$ respectively. Now from the definition of M4 and M5 we have;

$$M4 = \frac{a_4 a_0{}^3}{a_1{}^4}$$

$$M5 = \frac{a_5 a_0^4}{a_1^5}$$

We obtain;

$$M4 = 5U^3/8 - 5U^2/6 + U/4$$

$$M5 = 7U^4/8 + 7U^3/4 + 13U^2/12 - U/5$$

where;

$$U = V_T/(R_X I_Q)$$

When high level injection effects are included the collector current is given by;

$$I_C = \frac{2I_S \exp(V_{be}/(N_F V_T))}{1+(1+I_S/I_{KF} \exp(V_{be}/(N_F V_T)))^{1/2}}$$

The differentiation of this is much simplified if both numerator and denominator are multiplied by;

$$1-(1+I_S/I_{KF} \exp(V_{be}/N_F V_T)))^{1/2}$$

to give;

$$I_C = 2I_{KF} \; ((1+I_S/I_{KF} \exp(V_{be}/(N_F V_T)))^{1/2}-1)$$

This form of the equation is much easier to differentiate than the original quotient. Derivatives up to fifth order are given in Chapter 4.

APPENDIX IX

Distortion Due to Base Push-out

Both the Kirk and the van der Ziel models show a discontinuity in their derivative at a critical current I_0. A numerical example, using typical HF transistor parameters, is used to demonstrate the change in figures of merit that can be expected around this critical current. For the Kirk model the total transition time, τ_T, is given by;

$$\tau_T = \frac{V_T C_{TE}}{I_C} + \frac{W_{B0}^2}{n_B D_{NB}} + \frac{W_{CIB}^2}{4D_{NC}} + \frac{W_{EPI} - W_{CIB}}{2V_D} + \frac{C_{C0} R_C}{(1-V_{BC}/\phi)^{1/n}}$$

where W_{CIB} represents the charge induced base. This being zero for $I_C \leq I_0$ and increasing in magnitude with $I_C \geq I_0$. Calculations are carried out either side of I_0. Typical parameters for a high power transistor could be;

$C_{TE} = 10pF$

$C_{CD} = 0.4pF$ giving $C_C(5V) = 0.2pF$ with $n = 3$

$D_N = 2.5cm^2/S$ [174]

$R_C = 50\Omega$

$I_C = 100mA$

The respective time constants are 2.6, 10, 0, 12.5, 10 giving a total delay of 35.1pS corresponding to an f_T, for the intrinsic device, of 6.3GHz (i.e. 1/2Π25.1pS). With the resistive load the f_T is reduced to 4.5GHz.

The first derivative of τ_T is given by;

$$\frac{d\tau_T}{dI_C} = -\frac{V_T C_{TE}}{I_C^2} + \frac{1}{2}\left[\frac{W_{CIB}}{D_{NC}} - \frac{1}{V_D}\right]\frac{dW_{CIB}}{dI_C} + \frac{C_C R_C^2 1/n}{V_{CE}}$$

where;

$$C_C = \frac{C_{CO}}{(1 - V_{BC}/\phi)^{1/n}}$$

and;

$$\frac{dW_{CIB}}{dI_C} = \frac{I_O W_{epi}}{I_C^2}$$

for the low field case, or;

$$\frac{dW_{CIB}}{dI_C} = \frac{(I_O - I_{CX})^{1/2}\ W_{epi}}{2(I_C - I_{CX})^{3/2}}$$

for the high field case. Now substituting values and using;

$$W_{CIB} = 0$$

$$I_C = I_O = 2I_{CX}$$

gives respective time constant derivatives of -26, -125, 33.3, giving a total delay of -117.7pS/A, for the low field case and -26, -2500, 33.3, giving a total delay of -2492.7pS/A, for the high field case. In both cases the part due to the first derivative of W_{CIB} is the largest.

The second derivative of τ_T is given by;

$$\frac{d^2\tau_T}{dI_C^2} = \frac{2V_T C_{TE}}{I_C^3} + \frac{1}{2}\left[\frac{W_{CIB}}{D_{NC}} - \frac{1}{V_D}\right]\frac{d^2W_{CIB}}{dI_C^2}$$

$$+ \frac{1}{2D_{NC}}\left[\frac{dW_{CIB}}{dI_C}\right]^2 + \frac{C_C R_C^3 1/n(1/n + 1)}{V_{CE}^2}$$

where;

$$\frac{d^2W_{CIB}}{dI_C^2} = \frac{2I_O W_{epi}}{I_C^2}$$

for the low field case and;

$$\frac{d^2 W_{CIB}}{dI_C^2} = \frac{3(I_0 - I_{CX})^{1/2} W_{epi}}{4(I_C - I_{CX})^{5/2}}$$

for the high field case.

Substituting gives respective time constant derivatives of 520, 2500, 125, 444, giving a total delay of 3589pS/A^2, for the low field case and 520 + 3750 + 800 x 10^3 + 444, giving a total delay of 804.7 x 10^{3}pS/A^2, the largest component being due to the second derivative of W_{CIB} while for the high field case the largest part is due to the square of the first derivative.

For the two dimensional model τ_T is given by;

$$\tau_T = \frac{V_T C_{TE}}{I_C} + \frac{W_{BO}^2}{N_B D_{NB}} + \frac{W_{LB}^2}{4N_B D_{NB}} + \frac{W_{epi}}{2V_D} + C_C R_C$$

which for $W_{LB} = 0$ is the same as for the one dimensional model with $W_{CIB} = 0$.

For the case of $I_C \leq I_0$, $W_{CIB} = 0$, and which means that the distortion is caused only by variation in the emitter and collector time constants. The first two derivatives of τ_T are given by;

$$\frac{d\tau_T}{dI_C} = -\frac{V_T C_{TE}}{I_C^2} + \frac{C_C R_C^2 1/n}{V_{CE}}$$

and;

$$\frac{d^2 \tau_T}{dI_C^2} = \frac{2V_T C_{TE}}{I_C^3} + \frac{C_C R_C^3 1/n(1/n + 1)}{V_{CE}^2}$$

substitution of assumed values gives 7.33pS/A and 964pS/A^2 respectively. Notice that the two parts tend to cancel in the first derivative while they add in the second derivative. For the case of $I_C \geq I_0$, W_{CIB} is a function of I_C and must therefore be included in the calculation. Consider first the one dimensional model.

The first derivative of τ_T is given by;

$$\frac{d\tau_T}{dI_C} = -\frac{V_T C_{TE}}{I_C^2} + \frac{W_{LB}}{N_B D_{NB}} \cdot \frac{dW_{LB}}{dI_C} + \frac{C_C R_C^2 1/n}{V_{CE}}$$

where the derivative of the lateral base spreading is given by;

$$\frac{dW_{LB}}{dI_C} = \frac{L_E}{I_0}$$

Substitution for the condition $W_{LB} = 0$ gives the respective time constant derivatives of -26, 0, 33.3,

giving a total delay of 7.3PS/A, where the part due to the first derivative of W_{LB} is zero (at $W_{LB} = 0$) and some cancellation occurs between the emitter and collector components.

The second derivative of τ_T is given by;

$$\frac{d^2\tau_T}{dI_C^2} = \frac{2V_T C_{TE}}{I_C^3} + \frac{W_{LB}}{2N_B D_{NB}} \cdot \frac{d^2 W_{LB}}{dI_C^2} + \frac{1}{2N_B D_{NB}} \left[\frac{dW_{LB}}{dI_C}\right]^2$$

$$+ \frac{C_C R_C^3 1/n(1/n + 1)}{V_{CE}^2}$$

where;

$$\frac{d^2 W_{LB}}{dI_C^2} = 0$$

The respective derivatives of the time constants are 520, 0, 200 x 10^3, 444, giving a total delay of 201 x 10^3 PS/A^2, where the largest component is due to the square of the second derivative of W_{LB}.

The figures of merit can be found from;

$$M2 = \frac{A_2 A_0}{A_1^2}$$

and;

$$M3 = \frac{A_3 A_0^2}{A_1^3}$$

where;

$$A_0 = I_C$$

$$A_1 = \frac{1}{jw_1 \tau_0}$$

$$A_2 = \frac{\tau_1}{w_1 w_2 \tau_0^3}$$

$$A_3 = \frac{\tau_2 \tau_0 - 2\tau_1^2}{jw_1 w_2 w_3 \tau_0^5}$$

and;

$$\tau_0 = \tau_T$$

$$\tau_1 = \frac{1}{2!} \cdot \frac{d\tau_T}{dI_C}$$

$$\tau_2 = \frac{1}{3!} \cdot \frac{d^2 \tau_T}{dI_C^2}$$

and substituting gives (where M2 and M3 are the modules of the expressions);

$$M2 = \frac{\tau_1 I_C}{\tau_0} = \frac{\tau' I_C}{2\tau_T}$$

$$M3 = \left[\frac{\tau_2}{\tau_0} - 2\left[\frac{\tau_1}{\tau_0}\right]^2\right] I_C^2 = \left[\frac{\tau''}{6\tau_T} - \frac{1}{2}\left[\frac{\tau'}{\tau_T}\right]^2\right] I_C^2$$

where;

$$\tau' = \frac{d\tau_T}{dI_C}$$

and;

$$\tau'' = \frac{d^2\tau_T}{dI_C^2}$$

For the case of $I_C \le I_0$ the figures of merit for this example are;

$$M2 = 0.0104$$

and;

$$M3 = 0.0455$$

Once base push-out has started, $I_C \geq I_0$ the figures of merit become;

$$M2 = 0.168, 3.55, 0.0104$$

and;

$$M3 = 0.114, 12.99, 9.54$$

For the one dimensional low and high field cases and the two dimensional case respectively.

It can be clearly seen that the figures of merit increase in value when collector current exceeds I_0. The only case where this has not happened is for the two dimensional case of M2, where the part due to the variation in base-width is zero for zero change in base-width, i.e. zero for $W_{LB} = 0$. The first derivative of τ_T (7.3pS/A) for the figures used here, increases at the rate of (80×10^3 pS/A)/µm as W_{LB} increases i.e. increases M2 rapidly as W_{LB} increases.

Real transistors do not show such a sharp change in transit time at the critical current, as implied here, but the example serves to demonstrate that there is some critical current above which the transistor's performance is unacceptable.

APPENDIX X

Mask Making

The masks for photoreduction were cut on Rubylith 200x full size. Dimensions on the final masks were to be as small as 1μm, which meant that details as small as 0.2mm had to be cut, the cutting of the masks therefore required considerable care. Some of the required dimensions were smaller than the coordinatagraph had been designed to handle (0.01 inch resolution) and consequently some estimation was required. The fine adjustment screw could be used as a micrometer adjuster by counting the knurlings on the nut, which simplified estimation, enabling 0.001 inch resolution to be achieved. Guide ways and bed had to be as clean as possible if these dimensions were to be achieved, and the knife had to be sharp, properly set for depth and always used in the same orientation. Overall the results obtained were good, but where thin lines were left these tended to have been displaced sideways by the second cut. Small discrepancies were obvious on the small dimensions since the percentage error was quite high, but the photoreduction was unable to resolve the smallest of these.

Photographic Data

First Photoreduction

Camera settings:	31.130"	**and**	0.258"
Reduction:	20x		
Exposure time:	15 seconds		

Step and Repeat

Machine:	David Mann
Exposure spacing:	0.044"
Exposures per row:	25
Row spacing:	0.035"
Number of rows:	31
Start of row at:	2.5"
Filter setting:	Hi red
Filter setting:	0.095
Reduction:	10x
Exposure setting:	4 medium

Photographic Chemicals

Developer:	HRP	4:1 w/w
Stop bath:	acetic acid	56:1 w/w
Fixer:	Kodafix	3:1 w/w

Process Times

Developer:	5	minutes
Stop bath:	30	seconds
Fixer:	2	minutes

The resulting contact masks showed reasonable resolution overall, although some of the 1μm lines of grid 2 did not print well, Figure A.X.1.

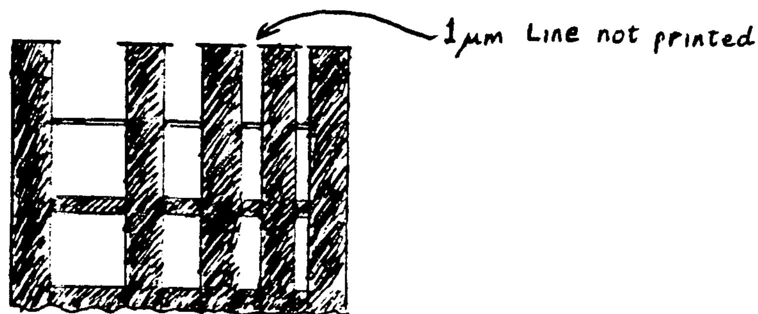
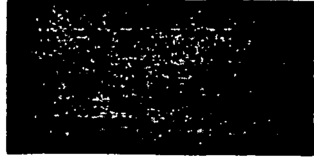


Figure A.X.1. Detail of Grid 2

The resolution check pattern indicated that resolutions of detail had started to fall off at around 2μm, although some 1μm structures had printed. Final die size was 0.044" x 0.035". Mask patterns are shown approximately 200 times magnification in Figures A.X.2 to A.X.6.



R.P.C.11

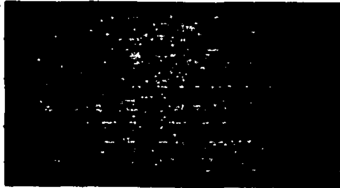


Figure A.X.2. Base Diffusion Mask 200x

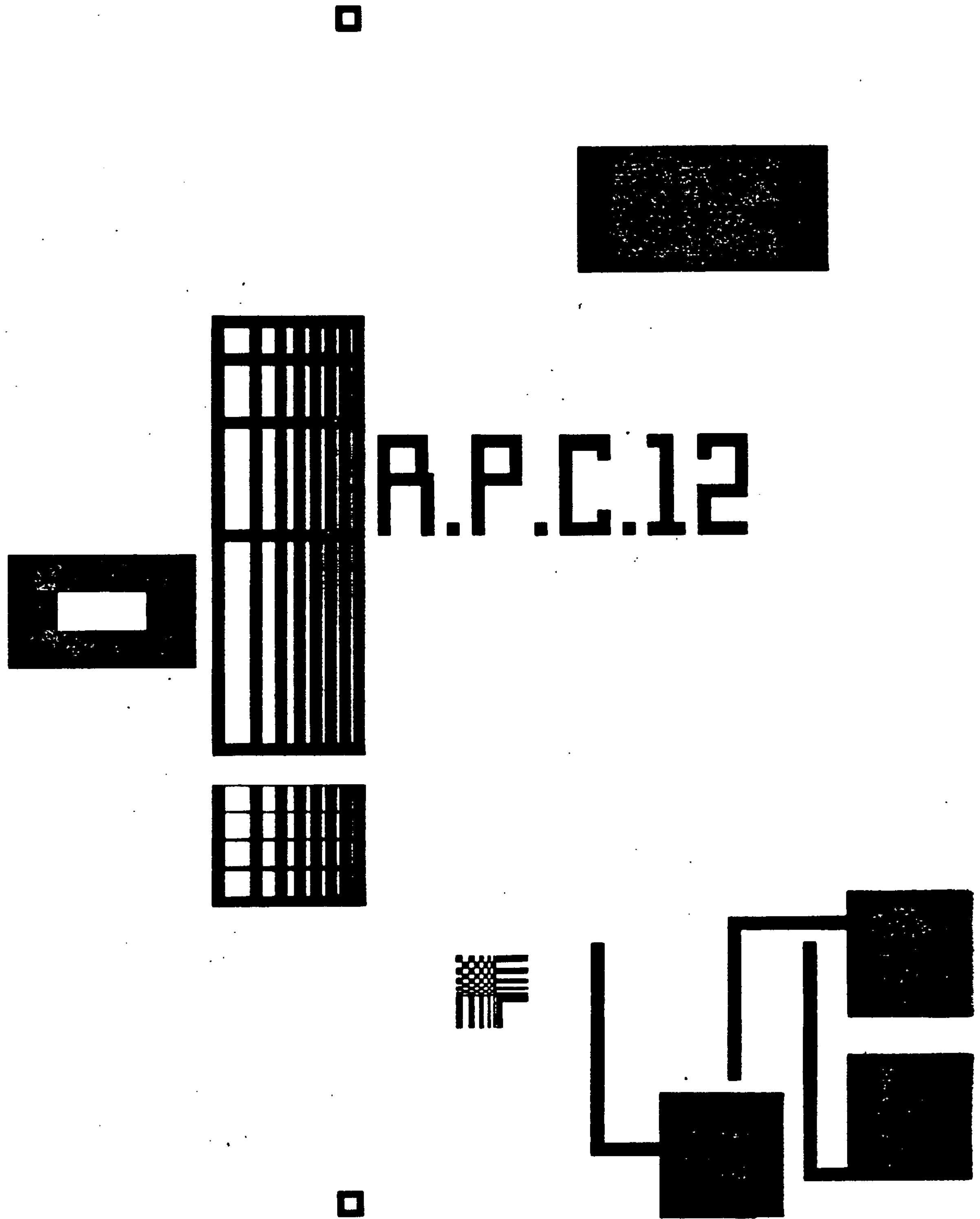


Figure A.X.3. Refractory Metal Mask 200x

Figure A.X.4. Emitter Diffusion Mask 200x

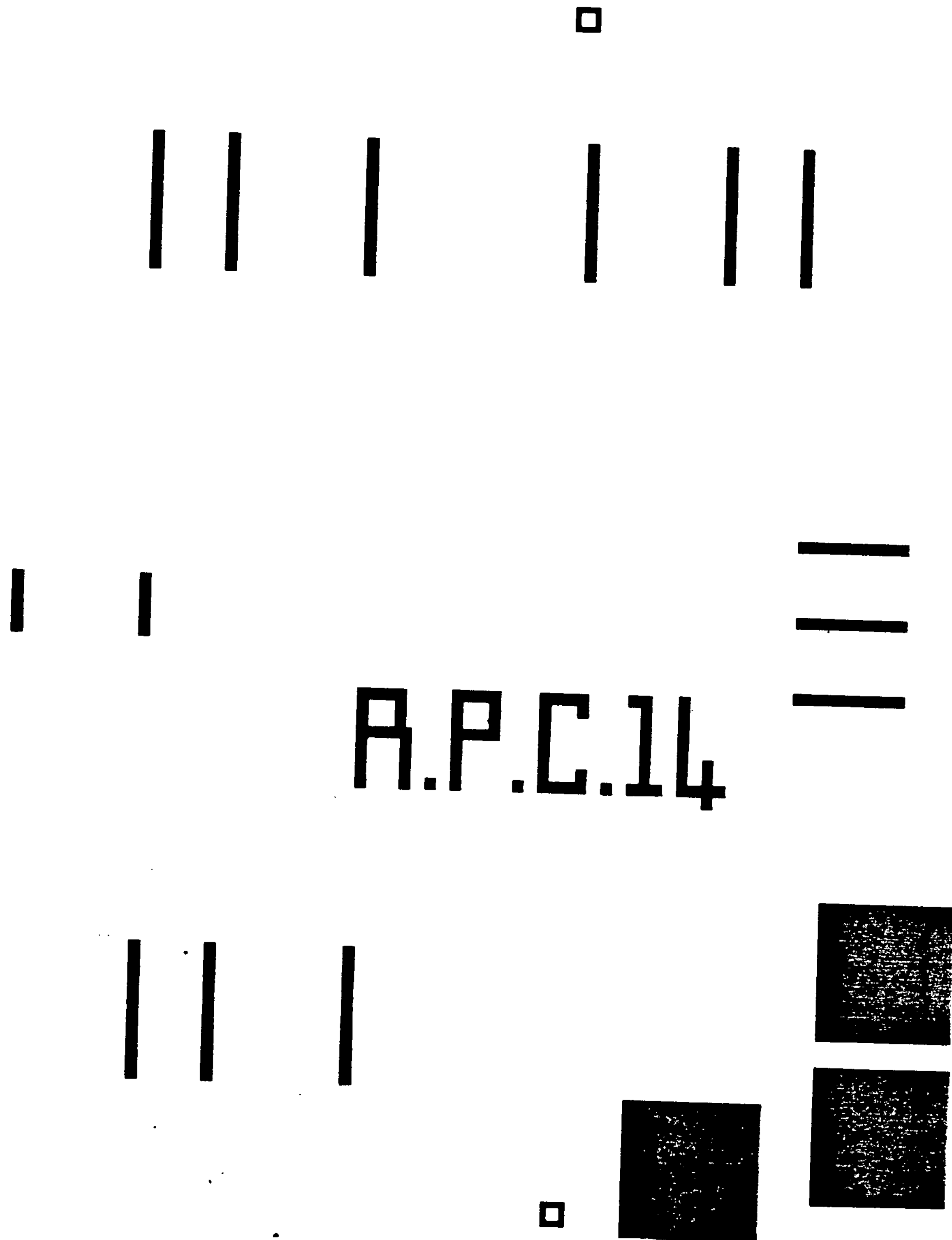


Figure A.X.5. Contact Hole Mask 200x

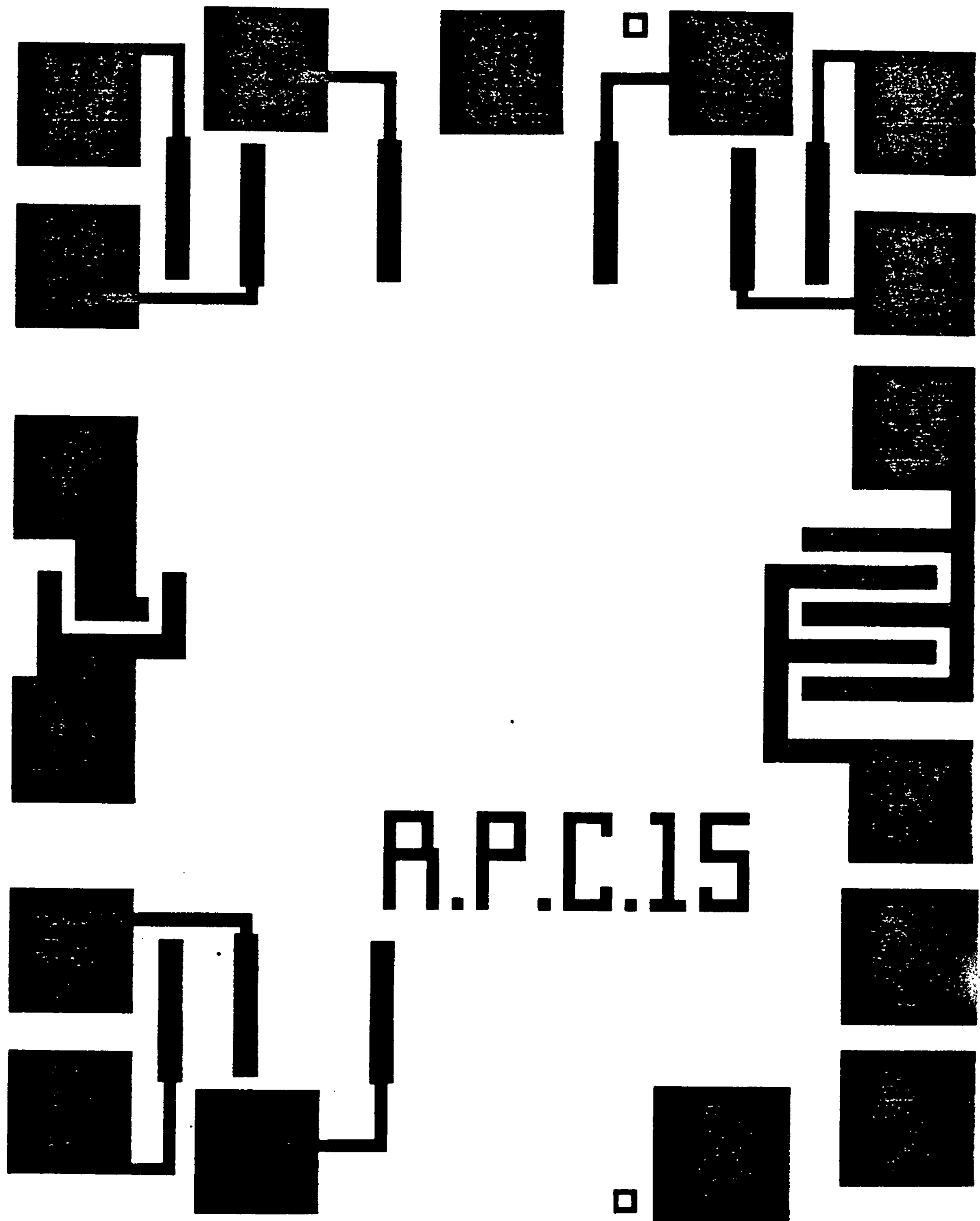


Figure A.X.6. Top Metal Mask 200x

APPENDIX XI

SPUTTERING PROCEDURE

1. Switching on the Equipment

1.1 Switch the three wall mounted electrical isolating switches to the ON position.

1.2 Turn on water at the wall mounted tap.

1.3 Switch on the vacuum station control panel via the isolating switch at the foot of the cabinet.

1.4 Check that the "WATER ON" lamp is alight, if not find cause and rectify before proceeding.

1.5 Ensure the three vacuum valves are in the closed position. Baffle valve - large handle at base of vacuum chamber, down for closed. Roughing and backing valves - on panel below vacuum chamber.

1.6 Switch on rotary pump.

1.7 Switch on diffusion pump. It will take about 20 minutes to warm up, while it does continue with:

1.8 Slowly open backing valve.

1.9 Fill the cold trap with liquid nitrogen. Top up as needed during use.

1.10 Switch on control panel of HF generator. Takes a few moments to warm up.

1.11 Switch on Pirani gauges. Backing pressure should be less than 0.05Torr.

1.12 Switch on nitrogen supply near machine. If necessary switch on the main manifold also.

1.13 Switch on argon supply at gas cylinder, adjust line pressure to 10psi.

1.14 Open inlet valve of vacuum chamber and back-fill with nitrogen. This is supplied under pressure and will eventually raise the chamber pressure slightly above atmosphere. A slight "hiss" indicates that the chamber is full and the vacuum seal is breached.

1.15 Turn off the nitrogen 1.16 Raise the top of the Vacuum chamber. Press "RAISE HOIST" on the control panel, WATCH THE CHAMBER. Sometimes the glass chamber remains stuck to the top plate by the Viton seal. If this happens press "STOP HOIST" followed by "LOWER HOIST" or simply "LOWER HOIST" and allow the chamber to return to the base plate, lest it

fall. Momentarily press "RAISE HOIST" followed by "STOP HOIST" to remove the weight of the top plate from the chamber. Break the top seal by slightly deforming the Viton seal with the thumb, pushing up the lip of the seal. ON NO ACCOUNT force anything between the seal and top plate or seal and chamber, nor must the chamber be struck in the hope of dislodging it.

The equipment is now ready for use.

2. Shutting Down the Equipment

2.1 Close chamber and shut inlet valves.

2.2 Close backing valve.

2.3 Open roughing valve and reduce chamber pressure to less than 0.1Torr.

2.4 Close roughing valve.

2.5 Open backing valve and pump down backing line to less than 0.1Torr.

2.6 Open diffusion pump baffle and pump chamber down to less than 0.1mTorr.

2.7 Close baffle.

2.8 Switch off diffusion pump. Takes about 20 minutes to cool.

2.9 Switch off gauges.

2.10 Switch off high frequency generator control panel.

2.11 Turn off nitrogen and argon supplies.

2.12 Wait for "WATER ON" light to extinguish.

2.13 Close backing valve.

2.14 Switch off rotary pump.

2.15 Switch off vacuum station control panel.

2.16 Turn isolating switches to "OFF".

2.17 Turn water supply off.

<u>The equipment is now switched off.</u>

3. Sputtering Procedure

3.1 Switch on the equipment as detailed in Section 1.

3.2 Load the chamber with the samples, placing them on the heated table.

3.3 Move the sputtering shield directly over the samples.

3.4 Blow the surface of the sample clean of any dirt using a nitrogen jet.

3.5 Lower the chamber top plate.

3.6 Switch on Pirani gauges.

3.7 Close backing valve.

3.8 Open roughing valve. Reduce chamber pressure below 0.1Torr.

3.9 Close roughing valve.

3.10 Open backing valve. Reduce backing line pressure to below 0.1Torr.

3.11 Slowly open diffusion pump baffle.

3.12 Turn low tension power supply control to minimum.

3.13 Switch low tension supply on.

3.14 Adjust Imax to 8 and I to 10. Monitor heater temperature with thermocouple thermometer. Adjust controls for 550°C.

3.15 Pump down chamber to about 1μTorr.

3.16 Move baffle chamber to the almost shut position.

3.17 Turn the argon needle valve to around 60.

3.18 Adjust needle valve and baffle to obtain a chamber pressure of about 3mTorr, backing pressure should be about 0.05Torr.

3.19 Turn RF power control to zero.

3.20 Switch on coil current supply and adjust to 6A.

3.21 On HF generator adjust, HF power to zero, other switch to "LOCAL" and "HF ON".

3.22 Turn thermocouple thermometer off.

3.23 Switch high frequency power on and adjust power dial to 2.

3.24 Strike the plasma. Do this by momentarily increasing the needle valve setting to increase the chamber

pressure slightly. Once the plasma has formed adjust needle valve and baffle as required to maintain the chamber pressure at about 3mTorr. Conditions around the plasma are different to those at the Pirani gauge, adjust for a mauve colour.

3.25 Increase power setting until ammeter reads 0.2A. Check pressure and correct if necessary.

3.26 Sputter onto the shutter for 1 minute.

3.27 Set timer to required time.

3.28 Move shutter out of the way.

3.29 Start timer. Check the machine occasionally as the pressure in the chamber can drift slightly.

3.30 When timer switches off the plasma the deposition is complete.

3.31 Switch off the RF generator.

3.32 Turn focus coil current to zero and then switch off.

3.33 Turn heater current to zero and then switch off.

3.34 Turn needle valve to zero.

3.35 Open baffle fully and allow chamber to pump down.

3.36 Switch on the thermometer and wait for temperature to drop to below 100°C.

3.37 Open chamber as detailed in 1.14 to 1.16.

3.38 Unload samples.

3.39 Continue as: 3.2 to sputter more samples or 2.1 to shut down the equipment.

Bipolar III Process Schedule

MIDDLESEX POLYTECHNIC

MICROELECTRONICS CENTRE

BIPOLAR III PROCESS SCHEDULE

NOVEMBER 1976

NS.

INDEX

Stage No.	Description.	Page
1.	Initial Slice Clean	1
2.	Slice Oxidation (1081°C)	1
3.	Micro-Neg. Isolation Photoresist.	2
4.	Isolation Boron (p+) Deposition (1154°C)	3
5.	Slice Clean and Isolation Drive-in (1184°C)	3
6.	Micro-Neg. p-type Base Photoresist	4
7.	Base (p-type) Boron Deposition (972°C)	4
8.	Slice Clean and Base (p-type) Drive-In (1184°C)	5
9.	Micro-Neg. n-type Emitter Photoresist	5
10.	Emitter (n-type) Phosphorus Deposition (1034°C)	5
11.	Slice Clean and Emitter (n-type) Drive-In(1105°C)	6
12.	Micro-Neg. Pre-ohmic Contact Holes.	6
13.	Metallisation	7
14.	Micro-Neg. Interconnection Pattern.	7
15.	Aluminium Etch	7
16.	Slice Alloy (400°C)	8
17.	Probe	8
18.	Scribe and Break	8
19.	Die Bond	9
20.	Wire Bond	9
21.	Electrical Test	10

1. Initial Slice Clean   3-6 Ω-cm

1.1 Wash the slice ultrasonically in trichlorethylene for 3 minutes.

1.2 Spray the slice in methanol.

1.3 Dry the slice with hot nitrogen.

1.4 Inspect the slice. If the defect count is high repeat stage 1.2 to 1.4.

1.5 Clean the slice in a solution consisting of 40% H_2SO_4 and 60% H_2O_2. Quench in a beaker of deionized water.

1.6 Wash the slice in running deionized water for 5 minutes.

2. Slice Oxidation (1081°C)

2.1 Etch the slice in a solution consisting of 6 volume parts deionized water and 1 volume part HF for 1 minute. Quench in beaker of deionized water.

2.2 Rinse the slice in running deionized water for 5 minutes.

2.3 Dry the slice with hot nitrogen.

2.4 Switch on N_2 and dry O_2 and load slice into furnace boat, place the boat in the mouth of the furnace tube.

2.5 Wait for 5 minutes.

2.6 15 minute dry O_2 oxidation.

Settings: N_2 Flowmeter 25cm.
O_2 Flowmeter 25cm.

2.7 Add wet gas flow for 2 hours wet oxidation.   ~ 0·5μm/Hr

Settings: N_2 Flowmeter 25cm.
O_2 Flowmeter 25cm.
O_2/N_2 mixture Flowmeter
O_2 10cm.* N_2 zero

Water bath temperature 80°C ± 3°C.   3" Tube
* 4" Tube use, 20cm and 3 Hr process.

2.8 5 minutes dry oxidation.

2.9 Pull the boat to the end of the furnace and cool for 2 minutes.

3. Micro-Neg. Isolation Photoresist.

3.1 Bake the slice for 15 minutes at 130°C ± 5°C.

3.2 Place the slice on the spinner, apply a few drops of Micro-Neg resist and spin for 30 seconds at 6000 r.p.m.

3.3 Bake the slice for 15 minutes at 95°C ± 5°C.

3.4 Align mask 1 and the slice as directed, and expose for 6.5 seconds ± 1 second, or 4 seconds if using a chrome mask.

3.5 Develop the photoresist by spraying the slice with Micro-Neg developer for 30 seconds.

3.6 Spray rinse the slice with Micro-Neg rinse for 20 seconds.

3.7 Dry the slice with nitrogen jet.

3.8 Inspect the slice for cleanliness and pattern definition.

3.9 Bake the slice for 15 minutes at 130°C ± 5°C.

3.10 Etch the oxide in a solution consisting 4 volume parts 40% NH_4F and 1 volume part HF for 5 minutes. Quench in beaker of deionized water.

3.11 Rinse the slice with running deionized water for 1 minute.

3.12 Dry the slice with hot nitrogen.

3.13 Inspect the slice (if the oxide is not fully etched repeat from step 3.11, etching for 2 minutes on each repeat).

3.14 Remove the photoresist in a solution consisting of 50% H_2SO_4 and 50% H_2O_2 for five minutes.

3.15 Repeat 3.14 with a fresh solution of H_2SO_4/H_2O_2.

3.16 Rinse the slice in running deionized water for 5 minutes.

3.17 Store the slice in a clean glass container in the drying cabinet. (30°C).

4. Isolation Boron (p+) Deposition (1154°C).

4.1 Clean the slice in a solution consisting of 40% H_2SO_4 and 60% H_2O_2. Quench in beaker of deionized water.

4.2 Wash the slice in running deionized water.

4.3 Etch the slice in a solution consisting 6 volume parts deionized water and 1 volume part HF for 15 seconds. Quench in beaker of deionized water.

4.4 Wash the slice in running deionized water for 5 minutes.

4.5 Blow the slice dry with hot nitrogen.

4.6 Nitrogen purge the boron deposition furnace tube for 5 minutes with the slice, on a quartz boat, positioned at the end of the furnace tube.

Settings: N_2 Flowmeter 25cm.

4.7 Preheat the boat and slice for 5 minutes in the boron deposition furnace tube (1154°C).

Settings: N_2 Flowmeter 25cm.

4.8 10 minutes boron deposition.

Settings: Flow meters N_2 to boron source N_2 O_2

(reading from left 4 25 8cm.

4.9 Nitrogen purge the boat and slice (at 1154°C) for 10 minutes.

Settings: N_2 Flowmeter 25cm.

5. Isolation Drive-In (1184°C)

5.1 Etch the slice in a solution consisting of 6 volume parts deionized water and 1 volume part HF for 30 seconds. Quench in deionized water.

5.2 Wash the slice in running deionized water until high resistivity effluent is obtained.

5.3 Boil the slices in concentrated HNO_3 for 15 minutes. Quench in deionized water.

5.4 Wash the slice in running deionized water until high resistivity effluent is obtained.

5.5 Repeat 5.1 and 5.2.

5.6 Blow the slice dry with hot nitrogen.

5.7 Switch on O_2 and N_2 5 minutes before loading.

5.8 15 minutes dry O_2 oxidation.

Settings: N_2 Flowmeter 25cm.

O_2 Flowmeter 25cm.

5.9 Add wet gas flow for 20 minutes wet oxidation.

Settings: N_2 Flowmeter 25cm.

O_2 Flowmeter 25cm.

Wet O_2 Flowmeter 10cm.

Water bath temperature $90^\circ \pm 5^\circ C$.

5.10 9 hour N_2 drive-in.

5.11 Pull the boat to the end of the furnace and cool for 2 minutes.

6. Micro-Neg p -type Base Photoresist.

Repeat Stage 3 using mask 2.

Thermostat settings
0·380
4·610
0·160

7. Base (p-type) Boron Deposition ($972^\circ C$)

7.1 Blow the slice dry with hot nitrogen.

7.2 Clean the slice in a solution consisting of 40% H_2SO_4 and 60% H_2O_2. Quench in a beaker of deionized water.

7.3 Wash the slice in running deionized water for 5 minutes.

7.4 Etch the slice in a solution consisting of 6 volume parts dionized water and 1 volume part HF for 15 seconds. Quench in beaker of deionized water.

7.5 Wash the slice in running deionized water for 5 minutes.

7.6 Blow the slice dry with hot nitrogen.

7.7 Nitrogen purge the boron deposition furnace tube for 5 minutes.

Settings: N_2 Flowmeter 25cm.

7.8 Preheat the boat and slice for 5 minutes in the boron deposition furnace tubes.

Settings: N_2 Flowmeter 25cm.

7.9 20 minutes boron deposition.

Flow meters	O_2	N_2	N_2 (source).	
(reading from left)	6cm.	25cm.	2.5cm.	
	10cm	25cm	5cm	3" tube.

7.10 N_2 purge for 5 minutes with the boat in the centre of the furnace.

7.11 Pull the boat to the end of the furnace and cool for 2 minutes.

7.12 Store the slice in clean glass container in drying cabinet (30°C).

8. <u>Slice Clean and Base (p-type) Drive-In (1184°C)</u>

Thermostat settings
0·89
7·55
0·29

8.1 Repeat Steps 5.1 to 5.5.

8.2 Switch the gas on as 8.3 for 5 minutes.

8.3 5 minute dry O_2 oxidation.

<u>Settings</u>: N_2 Flowmeter 25cm.

O_2 Flowmeter 25cm.

8.4 Add wet gas flow for 20 minutes, steam oxidation.

<u>Settings</u>: N_2 Flowmeter 25cm.

O_2 Flowmeter 25cm.

Wet O_2 Flowmeter 10cm. 20cm For 4" Tube

Water bath temperature 85° ± 3°C.

8.5 5 minutes dry oxidation. Settings as 8.3.

8.6 Pull the boat to the end of the furnace and cool for 2 minutes.

8.7 Remove and store the slice in a clean glass container in drying cabinet (30°C).

2 min Oxide etch to remove oxide from P-type areas.

9. <u>Micro-Neg n-type Emitter Photoresist</u>.

Repeat Stage 3 using mask 3.

10. <u>Emitter (n-type) Phosphorus Deposition (1034°C)</u>.

10.1 Repeat Steps 4.1 to 4.5.

10.2 Nitrogen purge the phosphorus deposition furnace tube for 5 minutes.

Settings: N_2 Flowmeter 25cm.

10.3 Predope the phosphorus deposition tube for 5 minutes with gas settings as 10.5.

10.4 Preheat the boat and slice for 5 minutes in the phosphorus deposition tube (1034°C).

<u>Settings</u>: N_2 Flowmeter 25cm.

10.5 10 minutes phosphorus deposition.

Settings:	Flow meters:	Source N_2	N_2	O_2
	(cm)	5	25	10

10.6 Nitrogen purge the boat and slice (at 1034°C) for 5 minutes.

Settings: N_2 Flowmeter 25cm.

10.7 Remove and store the slice in a clean glass container in drying cabinet (30°C).

11. Slice Clean and Emitter (n-type) Drive-in (1105°C)

11.1 Etch the slice in 6:1 D.I. water - HF solution for 1 minute. (To remove top surface of phospho-silicate glass).

11.2 Wash the slice in running deionized water for 5 minutes.

11.3 Blow the slice dry with hot nitrogen.

11.4 Nitrogen purge the n-type drive-in furnace tube for 5 minutes.

Settings: N_2 Flowmeter 25cm.

11.5 10 minutes steam oxidation.

Settings: N_2 Flowmeter 25cm.

O_2 Flowmeter 25cm.

Wet O_2 Flowmeter 25cm.

Water bath temperature 95°C ± 5°C.

11.6 5 minutes dry oxidation.

11.7 Pull the boat to the end of the furnace and cool for 2 minutes.

11.8 Remove and store the slice in a clean glass container in the drying cabinet (30°C).

12. Micro-Neg Pre-Ohmic Contact Holes.

Repeat Stage 3 (using mask 4) with the spin speed at 4,000 r.p.m. Etch time (stage 3.11) 3 minutes.

13. <u>Metallisation</u>

13.1 Clean the slice in 6:1 D.I. H_2O - HF mixture for 10 seconds.

13.2 Wash the slice in running deionized water for 5 minutes.

13.3 Blow the slice dry with hot nitrogen.

13.4 Place slice in evaporator and load 3 x 180 milligram charges of aluminium* (i.e. one charge per coil filament plus 33cms of wire cut into loop). 6 staples on each filament. Put in thickness monitor blank.

13.5 Pump down to 5×10^{-6}mm of Hg.

13.6 Switch on radiant heater variac control to ~~200v~~ 130V. Slice temperature should rise to 250°C in time.

13.7 Flash filaments onto shutter while substrate heating.

13.8 Adjust the slice temperature to 200°C. ~80 V setting

13.9 Run No.1 filament to its evaporating current (30 amps approx). After initial aluminium has evaporated move shutter from over filament. Evaporate all aluminium from filament.

13.10 Repeat 13.9 using No.2 filament.

13.11 Allow slice to cool to room temperature (do <u>NOT</u> let in air while the system is cooling).

13.12 Admit air through filter.

* clean aluminium charge with trichlorethylene and acetone.

14. <u>Micro-Neg Interconnection Pattern</u>

Repeat Stage 3 omitting Stage 3.1 (using mask 5) with 6,000r.p.m. organic compound and 6,000 r.p.m. resist, prebake 10 minutes at 90-95°C. Expose 1.5 seconds, develop for 12 seconds and cure for 10 minutes at 130°C.

15. <u>Aluminium Etch</u>

15.1 Etch until all excess aluminium is removed using a stirred quantity of Isoform Aluminium Etchout heated to 45° ± 5°C. Quench in deionized water.

15.2 Wash the slice in running deionized water for 5 minutes.

15.3 Strip the photoresist with J-100 Resist Strip heated to 90°. Quench in water.

10 mins. strip

15.4 Rinse the slice in running deionized water for five minutes. Blow dry.

15.5 Store the slice in a clean glass container in the drying cabinet (30°C).

16. Contact Alloying (400°C)

16.1 Blow the slice dry with hot nitrogen.

16.2 Purge the furnace tube with nitrogen for 5 minutes with the slice, on a quartz boat, positioned at the end of the furnace tube.

16.3 Alloy in aluminium; raise the boat temperature to 400°C ± 10°C by pushing to middle of the tube and withdraw immediately.

16.4 Purge the furnace tube (with nitrogen) with the boat and slice at the end for 5 minutes.

17 Probe.

17.1 Mount wafer on a vacuum chuck (diagram suppliex).

17.2 Ensure stage is down.

17.3 Swing stage over wafer and lock.

17.4 Align wafer with probes (do not adjust probes without asking for advise).

17.5 Raise stage.

17.6 Select terminals required (connections given).

17.7 Mark 'failed' chip and note best ones for bonding.

18. Scribe and Break.

18.1 With practice wafer, scribe a line and adjust graticule hairline to conincide (N.B. It is necessary to track wafer across complete number of circuits to check). 0·035" & 0·044" step size.

18.2 Mount wafer on chuck and align.

18.3 Scribe in both directions.

18.4 Blow debris clear.

18.5 Roll wafer between filter papers (face down) using $\frac{3}{4}$" copper pipe roller.

18.6 Separate chips as required.

19. Die Bond (Eutectic)

19.1 Ensure stage is up to temperature (high + low lights flashing) and that nitrogen flow meter reads about half scale.

19.2 Place header in stage and allow about 10 seconds for heating.

19.3 Place chip on header, watch formation of alloy, and with glass rods 'scrub' chip in melt orienting it with header as directed.

19.4 Remove bonded chip and header and allow to cool for a few seconds.

20. Wire Bond.

A. Ultrasonic Method.

20.1 Place circuit on magnetic chuck.

20.2 Depress foot pedal to bring tool to first search position.

20.3 With chessman, position desired pad under tool and rotate circuit to align and appropriate header post directly behind tool.

20.4 Release foot pedal to make bond on pad.

(N.B. The bonder is set for height, power and duration of bond. If the bond does not adhere, or any other malfunction occurs, ask for assistance. DO NOT attempt to adjust these conditions).

20.5 Press foot pedal for second search position and move circuit straight forward to locate header post under tool.

20.6 Release pedal to make second bond. The wire should now break off, leaving a 'stub' under the tool.

20.7 Repeat for all bonds.

20.8 Inspect the bonded chip before removal.

B. Thermocompression Method.

The thermocompression bonder is a purely manual machine. Instructions will be given when necessary.

21. Electrical Test.

Plug the finished device into socket and test for correct function.

If desired a finished circuit may be protected with a plastic lid or encapsulated with epoxy resin or silcoset.

APPENDIX XIII

S-Parameters

The motivation for using h-parameters is that it is more appropriate for circuit analysis. These parameters however are not easy to measure at higher frequencies due to the difficulty in achieving a high frequency open circuit termination. This problem is in part overcome by the use of Y-parameters which require only short circuit terminations to make measurements. A short circuit termination is easier to achieve than an open circuit termination but for high frequency measurements requires critical tuning at each frequency. Oscillation is a common problem (sic). This is a large motivation for using s-parameters which allow the device to be terminated by matched loads at both input and output.

If each part has incident port $|a|^2$ and emitter power (reflection and transmission) of $|b|^2$ then for a two port device;

$$a_1 = \frac{V_1 + I_1 R_o}{2(R_o)^{1/2}}$$

$$a_2 = \frac{V_2 + I_2 R_o}{2(R_o)^{1/2}}$$

$$b_1 = \frac{V_1 - I_1 R_0}{2(R_0)^{1/2}}$$

$$b_2 = \frac{V_2 - I_2 R_0}{2(R_0)^{1/2}}$$

Where R_0 is the generator and load impedance.

If the incident powers $|a_x|^2$ are taken as independent the scattering matrix follows;

$$b_1 = s_{11} a_1 + s_{12} a_2$$

$$b_2 = s_{21} a_1 + s_{22} a_2$$

Measurements are made by making a_1 or a_2 equal zero (by matching remote ends of the lines to prevent reflections). Basic measurement involves ratio and phase shift measurements between two equal frequency sinusoids.

Relation between h and s-parameters are given by;

$$s_{11} = \frac{(h_{11} - 1)(h_{22} + 1) - h_{12}h_{21}}{(h_{11} + 1)(h_{22} + 1) - h_{12}h_{21}}$$

$$s_{12} = \frac{2h_{12}}{(h_{11} + 1)(h_{22} + 1) - h_{12}h_{21}}$$

$$s_{21} = \frac{-2h_{21}}{(h_{11} + 1)(h_{22} + 1) - h_{12}h_{21}}$$

$$s_{22} = \frac{(1 + h_{11})(1 - h_{22}) + h_{12}h_{21}}{(h_{11} + 1)(h_{22} + 1) - h_{12}\ h_{21}}$$

and;

$$h_{11} = \frac{(1 + s_{11})(1 + s_{22}) - s_{12}s_{21}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$$

$$h_{12} = \frac{2s_{12}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$$

$$h_{21} = \frac{-2s_{21}}{(1 - s_{11})(1 + s_{22}) + s_{12}s_{21}}$$

$$h_{22} = \frac{(1 - s_{22})(1 - s_{11}) - s_{12}s_{21}}{(1 - s_{11})(1 = s_{22}) + s_{12}s_{21}}$$

Where the h-parameters are normalized to Z_0, if h'_x are actual parameters then;

$$h'_{11} = h_{11}Z_0 \qquad h'_{21} = h_{21}$$

$$h'_{21} = h_{12} \qquad h'_{22} = h_{22}/Z_0$$

APPENDIX XIV

Fourier Analysis of Initial conditions

The response of an R-C network is described by;

$$x + RC\frac{dx}{dt} = V$$

where;

x is the voltage across the capacitor

and;

V is the input signal.

For the case of a sinusoidal input signal;

$$V = A\sin wt$$

The general solution for x is;

$$x = \frac{A(wCRe^{-t/CR} + \sin wt - wCR\cos wt)}{1 + (wCR)^2}$$

Where the sine and cosine terms describe the steady state response i.e. the phase and amplitude of the fundamental

output. The exponential term describes the transition from the initial condition to the steady state condition. This is the only term that can contribute harmonic frequencies.

Fourier analysis within the simulation program is carried out numerically over one cycle of the fundamental. To discover the apparent harmonic content due to this method of analysis it is necessary to carry out a Fourier analysis on the general solution. This is now carried out over one period of the fundamental, $2\Pi/w$, starting at time T. The even terms are given by;

$$a_0/2 \quad + \quad a_1 \cos wt \quad + \quad a_2 \cos 2wt \quad + \quad +$$

where;

$$a_n \quad = \quad \frac{Aw^2 CR}{\Pi(1 + (wCR)^2)} \int_T^{T+2\pi/w} e^{-t/CR} \cos nwt\,dt$$

and the odd terms by;

$$b_1 \sin wt \quad + \quad b_2 \sin 2wt \quad + \quad +$$

where;

$$b_n \quad = \quad \frac{Aw^2 CR}{\Pi(1 + (wCR)^2)} \int_T^{T+2\pi/w} e^{-t/CR} \sin nwt\,dt$$

The general solution for either integral is;

$$\frac{Aw^2CRe^{-T/CR}(e^{-2\pi wCR} - 1)(g\sin nwT + h\cos nwT)}{\pi(1 + (wCR)^2)}$$

Where for the even terms $a_n \mid n = 0, 1, 2,,$

$$g = \frac{nw(CR)^2}{1 + (nwCR)^2}$$

and;

$$h = -\frac{CR}{1 + (nwCR)^2}$$

and for the odd terms $b_n \mid n = 1, 2, 3,,$

$$g = -\frac{CR}{1 + (nwCR)^2}$$

and;

$$h = -\frac{nw(CR)^2}{1 + (nwCR)^2}$$

It can be seen that the components are small for the conditions;

$$wCR >> 2\Pi$$

$$wCR << 1$$

and;

$$CR << T$$

To understand what these conditions mean consider them in view of the implications of the Fourier analysis. The fundamental condition that must be met before a Fourier analysis of this type is valid is that the function must be periodic, in this case a period of time T is implied. Figure A.XIV.1 shows the implied function where it can be seen that the exponential component causes a step discontinuity. It is this step, which is an artifact of the analysis, that produces the harmonic components.

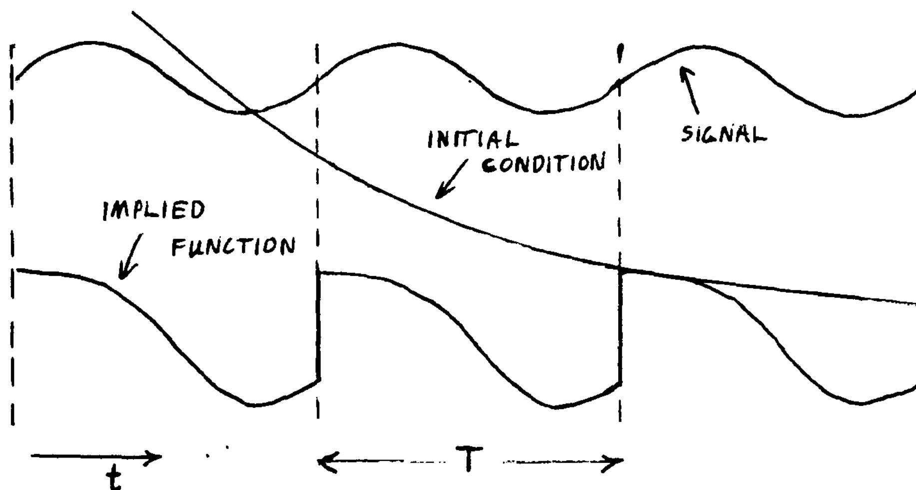


Figure A.XIV.1 Implied Function

The condition $wCR >> 2\Pi$ is the same as $CR >> 2\Pi/w$ i.e.

the C-R time constant is large compared to the period of the stimulus. Mathematically the exponential does not vary greatly over the sample period, and in circuit terms DC components take many cycles of the input frequency to decay. The other two conditions wCR << 1 and CR << T describe a C-R time constant which is short compared to the period of the stimulus and the start of the sample period respectively. Mathematically the exponential has decayed away before the analysis proceeds. Electrically this corresponds to a negligibly short C-R time constant compared to the input frequency or the duration over which the circuit has obtained equilibrium. All of these conditions imply that the change in the exponential part of the function must be small over the period of the analysis. The harmonic components are seen to be entirely due to the exponential decay of the initial conditions, and in fact are the result of an incorrect application of Fourier analysis. As such, minimum harmonic frequency components may be expected to occur for minimum size of step discontinuity.

From the point of view of circuit analysis it can be seen that there are few problems if either the resistive component is dominant or the reactive component is dominant. These cases correspond to a DC equilibrium being achieved very rapidly or very slowly respectively. In either case a Fourier analysis may be done in the first few cycles of the analysis. When both components are of

similar magnitude then the period chosen for the Fourier analysis to start must be after the DC equilibrium has been reached. This can lead to very long simulation times.