The Design and Implementation of Gallium Arsenide Digital Integrated Circuits

A thesis submitted to
the Faculty of Science of
the University of Edinburgh,
for the degree of
Doctor of Philosophy

by

A.D. Welbourn, MA, CEng, CPhys

Department of Electrical Engineering

March 1988

Acknowledgements

In undertaking a research project of this nature, one has cause to be grateful to innumerable people, and I must begin by apologising to those who are omitted from this formal acknowledgement, and by offering them my thanks for their help and support.

There are those who are deserving of especial mention, and I am most grateful for the guidance and encouragement given by Alec Livingstone, Bob Merrett and Peter Mellor, from the Gallium Arsenide Integrated Circuit section at British Telecom's research laboratories. Some of the measurements, specifically identified in the body of the text, were performed by my colleagues Graham Blau, Ian Sturgess and Bob Merrett, and I add my thanks to them by name here.

To Professor Mavor and Professor Jack, my supervisors at Edinburgh, I also owe a great deal of thanks, not the least for their patience in awaiting this manuscript, and for their most helpful and constructive comments.

I also acknowledge the Directorate of British Telecom Research and Technology, firstly for permission to undertake this study, and then for permission to publish this thesis.

Greatest thanks of all must be reserved for my wife Elizabeth, without whose love, patience, suffering and encouragement, this thesis would never have been completed.

List of Contents

Contents: Page:

List of figures

List of tables.

List of symbols used in equations.

Symbol:	Definition:
a	channel thickness
A	voltage amplitude
A_{eff}	effective cross-over area
C	capacitance or capacitance per unit area
C_{ds}	drain-source capacitance
C_{dsx}	drain-source metal capacitance
C_{eff}	effective capacitance of cross-over
C_{gd}	gate-drain capacitance
C_{gdx}	gate-drain metal capacitance
C_{gs}	gate-source capacitance
C_{gsx}	gate-source metal capacitance
C_L	capacitance of a metal disc on first interconnect level
C_o	capacitance at zero bias (for a diode junction)
C_p	peripheral capacitance (per unit length)
C_U	capacitance of a metal disc on second interconnect level
$d_{1,2}$	system delay times
E	electric field
f	clock frequency
f	effective fan out
f_{stop}	frequency at which latched data disappears
g_m	mutual transconductance
g_{meff}	available transconductance in a real device
I	current
I_{ds}	drain-source current
I_{dss}	FET saturation current at zero gate bias
I_o	saturation current of a diode
I_{sat}	FET saturation current
k	Boltzmann constant
k_1	arbitrary constant used in exponential roll-off equation
m	power-law fitting parameter for FET saturation region
n	ideality factor of diode
n	integer multiplier
N_d	doping level in channel
q	electron charge
Q_{tot}	total impurity charge in a conducting channel
r_1, r_0	depletion layer thickness at a given reverse bias: used in periphery conditions
R_o	diode resistance
R_s	source resistance
R_x	slew rate during period x
t	propagation delay
$t_{1,2}$	pulse widths
T	absolute temperature
v	electron velocity
v_s	saturated drift velocity
V	voltage
V_1, V_2	node voltage
V_{bg}	back-gating threshold voltage
V_{bi}	built-in potential of Schottky barrier

Symbol:	Definition:
V_{cc}	Most positive power supply voltage in bipolar circuits (common collector voltage)
V_{dd}	Most positive power supply voltage in FET circuits (common drain voltage)
V_{ds}	drain-source voltage
V_{ee}	Most negative power supply voltage in bipolar circuits (common emitter voltage)
V_{gg}	Most negative power supply voltage in FET circuits (gate pull-down bias)
V_{gs}	gate-source voltage
V_{of}	arbitrary offset voltage in FET model
V_{p}	pinchoff voltage
V_{sat}	saturation voltage
V_{sat0}	peak saturation voltage
V_{sg}	back-gating bias applied to a side electrode
V_{ss}	Zero bias supply voltage in FET circuits (common source voltage)
V_{t}	threshold voltage
V_{tt}	Terminating voltage in ECL circuits
w	width of FET channel or diode anode
w_1, w_2	track widths
w_{1eff}, w_{2eff}	effective track widths of equivalent parallel plate capacitor
z	length of FET gate or diode anode
z	scaling parameter in Lehovec and Zuleeg's FET model
ϵ	relative permittivity / dielectric constant
ϵ_o	absolute permittivity of free space
μ	electron velocity at low field
ψ	arbitrary parameter used to solve non-linear equation
w	angular frequency

1. Introduction

This thesis is the culmination of three years part time research work, and, I confess, three harder years of writing. The prime objective of this research was to develop the necessary infrastructure to be able to design high speed digital integrated circuits in gallium arsenide (GaAs). At the same time, colleagues were developing the necessary expertise to undertake the fabrication of the circuits. This thesis therefore describes the development of new models, the investigation of new logic circuits, and the adoption of these in constructing a circuit of medium scale integration (MSI) complexity. Although one circuit, an 8:1 multiplexer, was chosen as the vehicle for this work, the aim of the project was to develop the necessary techniques to undertake a number of different designs. In presenting the results, quite clearly the design of this circuit figures large (Chapter 5), but it has been treated as an exemplar design, rather than as a unique circuit development. The reader is also asked to bear in mind that details of the processing technology were also changing throughout the duration of this work, demanding a very close interaction between the circuit and process design engineers, coupled with a highly flexible approach to the development of the circuit elements.

I am aware that the passage of time during the second (writing) phase has had a major influence on the content of this work. On the negative side, a proportion of the work reported has become out-dated, where others have repeated and bettered the contribution I have made. Much more significant, however, is the positive side, where I believe that time has lent maturity to my views and analyses. This is particularly so, as I have spent this period employed in research in the parallel, competing field of silicon bipolar technology. I have therefore been forced to balance the advantages of the two technologies and, from this experience, I have found that there can be no better way to a true comparison of technologies than to be employed in one, contemporaneously with extolling in print, the virtues of the other. This document then is the outcome of this balancing of ideas.

The research leading to this thesis was performed at British Telecom Research Laboratories during 1981 to 1984, and I am particularly indebted to the management of these laboratories for the opportunity to conjoin both study and employment. As a "part time" degree study in industry, some constraints not always present in University research have been in force, one of which requires elaboration.

As in all research, the entrance to many interesting avenues of exploration has been opened throughout the work. Whilst many graduate students would have had the freedom to explore these, at least to their own satisfaction, the direction of industrial research is generally single-minded, and this project proved no exception. Some of the steps along the path toward the final goal have therefore been left with remaining question marks. This is nowhere more clear than in Chapter 4, which describes the development of computer models to aid the understanding and prediction of circuit performance. The models were developed to be of use in the overall scheme, not to be a complete piece of work in themselves. As shown from the results presented in Chapter 5, the detail was sufficient to achieve the desired objectives, but an expert in modelling will recognise that there is ample scope for others to pick up the threads. In searching for completeness in the work, it is in Chapter 5 that the reader should look, for here is contained the main objective of the exercise, the design and successful implementation of a chosen MSI IC.

As this work was undertaken as a part of a team of some ten professional engineers, I should take a moment to underline those areas for which I can claim sole responsibility, and on which I would expect my own contribution to be judged. I begin in Chapter 2 by claiming for my own the interpretation of the background material, and it is here especially that the work benefits from the passage of time during the writing phase. I hope in here there is some useful linking of facts, hitherto treated only in isolation. In Chapter 3 I rely on claiming only the analysis, but once again, here is a compilation of material not hitherto collated. As the thesis progresses through Chapters 4-6, I may lay claim to an increasing

proportion of the work as original, with the circuit design and chip development being the culmination of this.

At the beginning of this research project, the idea of using capacitor coupled logic (CCL) for GaAs ICs had been suggested by the head of the GaAs IC section at BTRL. A breadboard mock-up using silicon transistors had shown the principle to be sound, and the first integrated test structures had been designed. Elsewhere in the world, a few renowned laboratories had reported successes in the field, but a great number of problems were still unsolved. CCL purported to offer solutions to some of these, and it is from here that my research began. Although I cannot claim the CCL technique as my own, almost all of the more detailed understanding of its behaviour, arose out of my work. It is gratifying to know that although BTRL was alone in using the technique for many years, many of the early, commercially available GaAs IC components employ a very similar basic cell.

There is no single logical structure for a work of this size which surpasses all others, and I have endeavoured to create a structure which makes the fewest number of assumptions in the early chapters, only to prove them in the later text. In so doing, I have been forced to break the circuit description into two distinct Chapters. In Chapter 3, the reader is led to an intuitive understanding of the behaviour of the more basic functional elements, and in Chapter 5, these elements are put together into a useful circuit of much more complex behaviour. These two are separated by Chapter 4 which introduces the detailed account of device and component modelling not required for the intuitive circuit descriptions of the earlier Chapter, but essential to the completion of the IC design. By selecting this sequence, the simple description of CCL flows more naturally from the historical overview of Chapter 2. Furthermore, as the thesis progresses from the overview of Chapter 2, to the climax of the complete IC design in Chapter 5 and onto a brief glimpse of the future in Chapter 6, there is a gradual focussing of attention toward the unique contribution of this research project, as already described.

This focussing is also a mirror of that within the overview of Chapter 2, which begins by introducing GaAs as an engineering material, and moves on to describe its range of applications, placing in context the use employed here, before discussing the techniques which enable it to be used. In the final Section, the history of its use in IC applications is treated, leading ultimately to its use in CCL.

Finally in Chapter 6, two further possible configurations are discussed. One of these is now in widespread use, but it developed out of the work contained here, and the other is a completely new possibility, not yet published in open literature.

Chapter 2. An Introduction to Gallium Arsenide

It is the intention of this Chapter to introduce the reader to the technology and circuits in gallium arsenide (GaAs). Clearly this is a large subject, and the chapter is therefore necessarily large. It is hoped that for all there will be something new, either new material, or just a new treatment of the facts.

The Chapter begins from a material scientist's point of view, introducing the fundamental properties of GaAs as a crystalline solid. Some basic understanding of the physics of such materials is assumed but, wherever possible, the underlying assumptions have been kept to a minimum and some reasoning has been presented to explain why the material has certain properties. The consequence of these properties is also explored. Following this important background lies a link between the materials physics and both the commercial and engineering aspects of GaAs. A brief history of the development and exploitation of GaAs is presented, referring to the materials properties and how both an electronics engineer and a technologist may put these properties to work. A separate section is devoted to "back-gating", not a fundamental property of GaAs, but a phenomenon often encountered in a practical circuit. As such, it is only fully understood in the context of both the materials and applications sections.

The second major division within this Chapter outlines the technology of GaAs integrated circuit (IC) manufacture, again taking a somewhat historical perspective. Where appropriate, comparisons are drawn with the more widely known and understood technology of silicon IC production. This Section is intended to be helpful to the silicon technologist interested in GaAs and to the design engineer wishing to understand the technology.

The final Section leads the reader towards the specific subject of digital integrated circuits, and presents a critical history of the way in which GaAs circuits have progressed from 1976 until the present day. Again, reference is made to the comparative development

of silicon digital integrated circuits, to set the work in its context.

2.1 Fundamentals of GaAs.

2.1.1 GaAs Material Properties

To many accustomed to the "schoolboy" description of the operation of semiconductors, the concept of a compound semiconductor seems quite alien. However, gallium arsenide is just one (probably the best known) of a wide range of semiconducting compounds [1]. The crystal structure is very similar to both silicon and germanium [2], but with one of the atoms of the unit cell replaced by gallium and the other by arsenic. Instead of being entirely covalent the crystal bonding is therefore partially ionic, but this causes only marginal changes to the material properties. Perhaps the most significant change to arise from the chemistry is that the {111} cleavage planes leave a pure surface of either gallium or arsenic. As these arrangements are not energetically the most favourable, some surface re-ordering takes place, leaving the surface region with an entirely different energy-band structure from that in the bulk. Similar re-ordering takes place to a lesser extent on the other exposed crystal faces. It is responsible for a high density of surface states [3] property which has significant bearing on the performance and design of GaAs semiconducting devices.

GaAs is much more brittle than silicon, and this too could be caused by the ionicity. Thus if a {100} [4] plane is disturbed laterally in a <010> direction the usual ionic attraction force across the plane will be replaced by a repulsive force separating the two planes and breaking the crystal. This property makes thin wafers of GaAs more difficult to handle than the equivalent silicon wafers. Whilst this does not present an insurmountable problem, it does cause some concern when using equipment designed to handle the more robust silicon.

The presence of two atomic species in the crystal widens the range of crystal defects which are found compared with that containing a single element [5]. The most obvious additional defect being due to the substitution of the antitype. Of the two alternatives, the

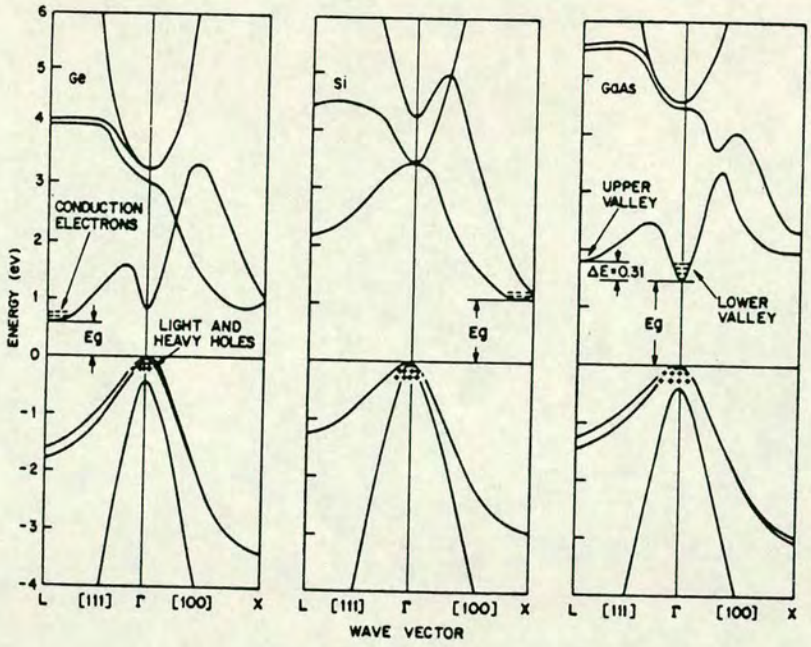


Figure 2.1 Band structures of Ge, Si and GaAs.

presence of arsenic on a gallium site is the more important. This defect gives rise to a mid-gap energy level denoted "EL2" [6], which, at least in part, contributes towards the semi-insulating property discussed later. The influence of EL2 is described in Section 2.2.

Apart from these properties arising from the chemistry, most of the material properties of particular interest in semiconductors derive from the electronic band structure of the crystal [7]. Ignoring the ionicity, GaAs crystallises in a form very similar to both germanium and silicon, and the similarity extends to the energy band diagram [8] of figure 2.1, in which the energy of the electrons is plotted against the wave vector [9]. The wave vector represents the electron momentum within the crystal. The small differences in the wave functions of the constituent atoms are reflected by similar differences in the energy-band diagrams of the crystals, and silicon, germanium and gallium arsenide all have different band structures. Nevertheless, all three materials show similarities in the lowest conduction band (the band immediately above the energy gap), with three minima corresponding to three different wave vectors in the

crystal. These lie in the X and L directions (along the <110> and <100> directions respectively), and at the centre of the Brillouin zone [10] (corresponding to a zero in the crystal momentum). The separation between the top of the valence band and the lowest of these minima is the energy gap [11].

Unlike both germanium and silicon, the lowest minimum in GaAs lies at the zone centre, at the same value of momentum as the top of the valence band. Transitions between the two bands are thus made significantly easier. Energy is conserved by absorption or emission of a photon, and no momentum change is required. This transition is termed "direct", in contrast to the "indirect" transition in both silicon and germanium, in which the interaction demands the presence of a phonon as well as the photon [12]. The direct transition allows GaAs to be used for both light emitting diodes (LEDs) [13] and lasers [14] with a useful energy conversion efficiency.

A second property follows from the direct transition. Because the electron-hole pairs recombine as a two-body, rather than a three-body interaction, the probability of recombination is much higher, and the minority carrier lifetime is much reduced. The lifetime is typically five orders of magnitude smaller in GaAs than in silicon [15]. This property alone prevents the exploitation of bipolar junction transistors in GaAs.

The other important property of GaAs is the very high electron mobility at room temperature - at 8500 cm^2/Vs in undoped material [16], it is some 5 or 6 times greater than in silicon. Quite clearly, since switching speed is related to mobility [17], this large increase in mobility should be reflected in faster circuits.

However, the electron mobility alone does not dictate the total carrier behaviour. Strictly, the parameter termed mobility, derived as the slope of the velocity versus electric field characteristic, should be referred to as the low-field mobility, because it is not independent of the applied field. In any semiconductor, the carrier velocity tends to saturate at higher fields [18]. This is most

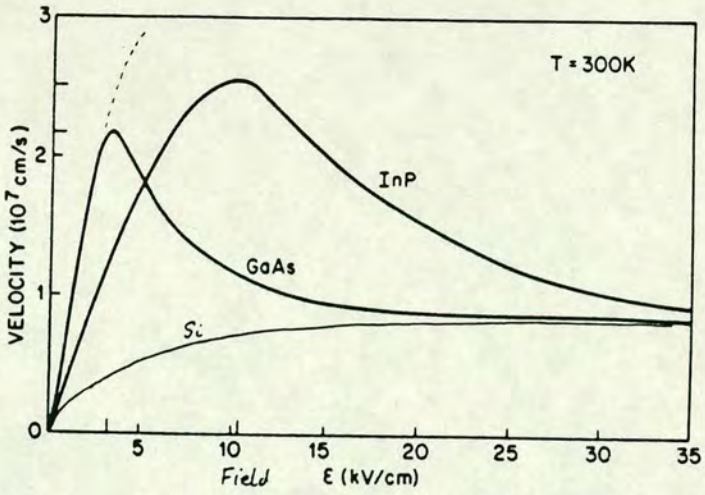


Figure 2.2 Velocity-field characteristics of electrons in various semiconductors.

readily explained as a loss of energy from the carriers to the lattice, because the effective temperature of the carriers is raised above that of the crystal. In GaAs a second, much more pronounced effect is seen [19], as shown in figure 2.2. As the electrons are given more energy from the electric field, a point is reached when the electrons cease to increase their velocity, instead slowing down significantly.

This behaviour can once again be explained directly from the band structure. The high electron mobility is a consequence of the tight radius of curvature on the bottom of the conduction band (actually, the effective mass is inversely proportional to the second derivative of the energy with respect to the wave vector [20]). A further consequence of this is that the density of states near the edge of the conduction band is small [21]. When the electron is given more energy by acceleration, it will eventually have sufficient energy to occupy one of the other minima in the conduction band. As the radii

of curvature of these other minima are large, the effective mass is high, leading to a lower velocity for a given energy. The density of states is also high in each of these minima and its effect is further increased because there are six silicon-like and eight germanium-like minima, all at a similar energy (see figure 2.1) [22]. As the

probability of occupation of a given state is proportional to the density of states, an electron will transfer preferentially to the high mass, low velocity state.

It should be noted that the transfer between valleys requires a phonon interaction [23], and scattering is therefore not instantaneous even when the energy is sufficiently high. Thus an electron may continue to accelerate for a short time, as depicted by the dotted line in figure 2.2. This behaviour is variously termed "ballistic transport" [24] or "velocity overshoot" [25].

Electron behaviour as outlined above is clearly susceptible to variations in temperature. At high temperature the electrons will obtain the required energy to change bands more readily, and the necessary phonon interaction will be more probable. These two factors combine to give a drastic reduction in mobility, and a lowering of the peak velocity as the temperature is increased [26]. At reduced temperature, the converse is true.

The effect of temperature on GaAs device performance further differs from that on silicon devices because of other differences in material properties. The much larger bandgap in GaAs [27] allows device operation up to high temperature, and the lower ionisation energies of both electrons and holes reduces the temperature at which carrier freeze-out occurs [28]. Together these two properties offer the potential for operation of GaAs devices over a very wide temperature range [29]. One practical consideration for operation at high temperature is that the vapour pressure of GaAs is relatively high, with surface decomposition and release of arsenic vapour above 600°C [30]. This limit is independent of the actual device being made, but there is commonly a much lower temperature limit because of subsequent fabrication steps; the latter is clearly dependent upon the type of device being made and the processing sequences adopted. For example, the metal used to form a Schottky junction with the GaAs may interdiffuse above a certain temperature, causing degradation of the junction [31].

In marked contrast to the superior electron mobility, the hole mobility in GaAs is similar to that obtained in silicon [32]. Again, this can be inferred directly from the band structure diagrams, there being little difference in the valence bands of the materials. The light-hole band [33] in silicon has a somewhat tighter curvature, and it thus exhibits a slightly higher hole mobility. Whilst the two or three to one ratio of electron to hole mobilities may be tolerated in complementary devices in silicon [34], the twenty to one ratio renders such considerations in GaAs impractical for most applications, although it has found use in very low power memory cells [35].

There are other properties of interest which are related directly to the choice of GaAs as the substrate material, but which do not follow directly from a simple interpretation of the band structure. These include the thermal conductivity and thermal expansion coefficient, in addition to the ability to form a semi-insulating variety.

The most notable of these is that GaAs has a semi-insulating form [36], in which the free carrier concentration is so low that the material acts as an adequate insulator. Devices formed in separate islands within a semi-insulating (SI) GaAs substrate are electrically isolated from each other inherently. The factors associated with the SI nature of GaAs are explored further in Section 2.2. Under certain circumstances this insulating property breaks down, according to the back-gating effect, which is described more fully in Section 2.1.3.

The most widely discussed of the thermal properties is the thermal conductivity. At 45 W/m.K at room temperature in GaAs [37], it is only one third of that of silicon. This is often claimed to be a serious disadvantage [38], it being argued that in a power limited chip (nearly always the case for ultra-high speed operation [39]) a lower thermal conductivity implies the imposition of a lower limit on the allowed power dissipation and, therefore, on chip complexity. Thus the potential speed-power advantage of GaAs must immediately be forfeit. In reality, only a fraction of the total chip-to-air thermal resistance is accounted for in the chip itself [40], and the basic value of the substrate conductivity is almost immaterial.

Less commonly cited as a problem is the higher thermal expansion coefficient in GaAs [41]. In practice, this may well prove more important than the poorer thermal conductivity. There are three basic problems; firstly in the registration between different stages of the processing, secondly from stresses induced during die bonding and thirdly from variations in parameters caused by stresses arising during fabrication.

Firstly, when fabricating integrated circuits, the thermal expansion coefficient of the glass chosen for the lithographic mask plates should be similar to that of the substrate material. As the minimum dimensions are shrunk and the wafer diameter is increased, differences in the expansion rates of the mask and substrate become comparatively more important [42]. Most readily available optical mask plates are designed for use with silicon and are thus non-ideal for GaAs.

Secondly, the dice are heated during the packaging operation in order to achieve a uniform die bond [43]. On cooling, the substrate might crack if the stresses are too great.

Thirdly, GaAs, being more brittle than silicon, is more susceptible to damage by stresses induced by temperature cycling during IC fabrication. Various surface dielectrics are used during the manufacturing, and these can have widely different expansion from that of the underlying semiconductor [44]. As any stress arising is close to the active device, device parameters can be widely changed as a result [45]. The larger expansion coefficient of GaAs demands that more care is taken in process design, in order to avoid these problems.

A final property of GaAs which is often quoted is the radiation hardness [46]. This is of particular interest to the military, who have invested a great deal of money in research on GaAs for this reason. There are several forms of radiation damage; a slow degradation leading to eventual failure, caused by total accumulated dose, and sudden catastrophic failure caused by different types of single event (e.g. by neutrons or by gamma rays). The advantage of

GaAs here is sometimes overstated. It is now possible to get good protection to most of these radiation events in both MOS and bipolar silicon technologies, but special techniques are required to obtain this [47]. In GaAs this proof against damage appears to be available "free". It might be argued that the hardness of GaAs might be improved with effort, in much the same way that the hardness of silicon has already been improved. Radiation hardness will not be considered further in this thesis.

2.1.2 Historical exploitation of GaAs

As discussed in the previous Section, a major advantage of GaAs is the direct bandgap. This leads immediately to the exploitation of GaAs for both light emitting diodes (LEDs) and lasers. A high quantum efficiency [48] is achieved because the recombination of the carriers and the attendant release of a photon may be directly stimulated by a photon; no phonon interaction is required. Even with this special property, GaAs lasers would not have developed so rapidly, but for two additional factors.

Firstly, the infra-red emission of the GaAs laser (at a wavelength of 0.85 um [49]) was suitable for use in early optical fibre transmission systems [50].

Secondly, a second semiconductor, aluminium arsenide (AlAs), shares an almost identical lattice parameter with GaAs [51]. A family of materials, $Al_xGa_{1-x}As$, can be formed from a mixture of the two (here 'x' denotes the mole fraction of AlAs), and because the lattice parameters of both components are equal, layers of different composition can be grown without strain induced defects forming in the interface [52]. The bandgap of the material changes according to its composition [53]. By using layers of differing composition (and doping) a whole gamut of useful devices can be built [54] which exploit the change of bandgap at the interface. The term heterostructure or (heterojunction) is used to describe these interfaces.

The technology required for the fabrication of such heterostructure lasers emerged contemporaneously with the explosion of interest in

optical fibre transmission. The continued development of the GaAs laser became dominated by market potential, rather than by the more common, but less urgent driving force of future promise. Consequently, both the availability and quality of GaAs material advanced rapidly.

It soon became clear that alternative materials (based on indium phosphide, InP [55]) would allow better transmission because they could operate at 1.3 and 1.55 um wavelength, where attenuation in the fibre could be significantly reduced [56]. The techniques necessary for working in these newer compound semiconductors (containing three or four of In, Ga, As, and P) followed directly from those developed for GaAs and, in turn, GaAs has since benefited from work on these materials. Despite the fact that the technological forefront in optoelectronics has moved away from GaAs, other applications have followed, no doubt stimulated both by the availability of devices and by the example set in optical transmission. Most notable of these is in the compact disc (CD) system [57], well established in domestic HiFi systems and now being developed as a mass storage medium for computing [58]. This consumer application, in which low cost is the key to success, is entirely dependent upon the GaAs laser, and probably represents the largest commercial market for GaAs at the moment.

It is probable that GaAs would still have been developed for conventional electronics applications, even in the absence of the special optical properties, but the rapid deployment of effort on GaAs for optoelectronics certainly helped to advance this. The combination of the high electron mobility and semi-insulating substrate is the key factor here. The electron behaviour alone is sufficient for interest to be expressed in GaAs for high frequency active devices [59] and the insulating property allows the fabrication of high quality, high frequency passive components, such as inductors and transmission lines [60]. Quite clearly, in those applications where GaAs might be considered simply because of its high electron mobility, silicon devices of reduced geometry might also be considered [61]; the increased difficulty of the reduced geometry being balanced against the additional experience of silicon

technology. However, most high frequency passive components on silicon are ruled out because of the lossy substrate [62], leaving GaAs unchallenged for those applications requiring both active and passive devices. Perhaps it is not surprising then that the exploitation of GaAs has followed the path of discrete microwave devices [63] into analogue microwave integrated circuits (firstly hybrid and then monolithic) [64] and finally into digital ICs [65].

The highest frequency active devices are generally two terminal structures, as the parasitic feedthrough from input to output of a three terminal structure tends to restrict the bandwidth. Thus, diodes for use both as mixers [66] and oscillators [67] extend in frequency performance well beyond transistors. When GaAs was first explored, the anomalous electron behaviour allowed a completely new type of diode to be made; the Gunn diode [68]. The action of this device follows directly from the slowing down of the electrons at the critical electric field, and the so-called "negative differential mobility". In order to accommodate the reduction in velocity, a space charge layer develops to change the field distribution. When this space charge layer becomes mobile, a Gunn domain [69] is formed. The arrival of this domain at the electrode releases the charge packet. A new domain will form rapidly, because the discharging of the first domain once again allows the electric field to exceed the threshold field again. The overall effect is a periodic pulsing of the current - an oscillator. Such devices can oscillate at several tens of gigahertz, delivering quite high power [70].

In addition to this and other special devices, there is much interest in conventional diodes for high frequency operation. In the simplest analysis, the higher mobility of GaAs allows a higher current to flow in the material for a given electric field. This translates directly into a lower resistance for a given size and shape of junction, giving both a lower noise figure [71] and a lower capacitance. These are both advantages for high frequency mixers.

This same principle of low resistance and noise figure for a given current would be expected to follow in three terminal devices and there is quite clearly an interest in GaAs transistors. The range of

GaAs transistors is more limited than the range of silicon transistors. Again the material properties explain why.

The most useful transistor for high frequency use is the bipolar transistor [72]. Several factors account for this, but the two key points are the very large transconductance [73] which can be achieved and the vertical nature of the device. The latter allows the critical dimension of the bipolar transistor (the vertical base width) to be much smaller than the equivalent, lithographically defined, dimension of an FET (the horizontal channel length). However, as already stated, the minority carrier lifetime in GaAs is too short to allow conventional bipolar transistors to be made with a usable current gain.

Recently, bipolar transistors have been made using a heterojunction comprising both GaAs and AlGaAs to overcome these problems [74]. The technology for this so called HBT (heterostructure bipolar transistor) is similar to that developed for the heterostructure laser devices already discussed. There is great interest in HBTs, both as a means of coupling the mobility advantage of GaAs with the attractive device properties of the bipolar junction transistor, and also because of the numerous additional advantages which accrue from the use of the heterojunction [75]. It is very much a device of the future, being still in the early development stages.

Having eliminated the homojunction bipolar transistor as a means of exploiting GaAs, attention is turned to the family of unipolar or field-effect devices. The most obvious FET to consider is the metal oxide semiconductor transistor (the MOSFET or MOST) [76], which has proved itself to be most successful in silicon. This success is in part attributable to the advent of CMOS [77], in which both p- and n-channel devices are fabricated together, making a complementary circuit in which the DC power dissipation is virtually zero. This configuration is clearly not available in GaAs, as already stated, because of the grossly dis-similar electron and hole mobilities.

Clearly then one would like to make use of the single polarity MOS transistor, which in practice means the n-channel (or nMOS)

transistor (again the low hole mobility of GaAs offers a poorer performance prospect for the p-channel device than is already available with silicon). The MOSFET is selected in preference to the junction FET (or JFET), because of the higher transconductance [78] and the more controllable and simpler fabrication procedure. Sadly, the MOSFET too is unavailable in GaAs, because of the very large numbers of trapping centres at the interface between the semiconductor and the oxide [79]. As the band-bending at the semiconductor surface adjusts with changes in the applied gate voltage, the occupation of surface traps changes depending on the relative energy of the trap and the Fermi-level [80]. Relaxation of any trapped electrons or holes results in a change in the surface potential of the semiconductor, and a consequent change in the magnitude of any accumulation or inversion charge. In silicon, the traps are few in number and the surface charge is stable with time, but in GaAs, the trap density is significant, and the net surface charge is unstable [81]. The DC and AC behaviour of such an MOS gated transistor in GaAs are therefore very different. The MOS gate effectively acts as a leaky capacitor at the input to an ideal FET [82].

Many different surface treatments and alternative dielectrics have been tried to enable MOSFETs to be made in GaAs [83]. (Because the GaAs embodiment of the MOS structure does not necessarily use either a native oxide, or in some cases an oxide at all, the term MISFET is preferred, denoting the metal-insulator-semiconductor construction.) Some success has been claimed, but it has always proved short lived. Probably the most successful attempt to use GaAs MISFETs applied the model of a leaky capacitor to design circuits capable of accommodating the imperfect FET behaviour [84], but this approach too has its shortcomings. This approach is very similar to that employed in the bulk of the work in this thesis, the concept actually being derived from the philosophy underlying capacitor coupled logic [85].

Having eliminated the widely used transistor types, only the JFET remains [86] as a possible means of widely exploiting GaAs amplifiers or switches. The GaAs JFET is indeed a viable transistor and, as expected, the performance of this device does bear out the initial

expectation that the higher electron mobility of GaAs would be reflected in a better high frequency performance, whether measured as a digital switching speed [87], or as an amplifier noise figure [88].

However, this comparison is true when like devices are compared. The GaAs JFET is better than the silicon JFET, but since the latter is rarely used, this comparison is of little benefit. The comparison between the GaAs JFET and the silicon bipolar transistor (BJT) is a complex issue, still unresolved [89]. However at the onset of this research work, the actual performance of a silicon BJT was very much restricted by parasitic elements and remained inferior to the GaAs JFET. Of immediate interest however is the comparison between the GaAs JFET and the silicon MOSFET.

Most of the arguments used in this comparison are again ongoing, and more complex than can be discussed in this Section. In summary the GaAs JFET has an edge in performance, although by a smaller margin than first expected [90]. The inherent mobility advantage of GaAs over silicon is in fact strengthened because of the reduced mobility in the inversion layer of the MOSFET [91]. However, the sharper effective channel doping of the MOSFET gives it a higher transconductance [92] and, therefore, a better switching performance for a given mobility. The clearest distinction between the devices is that the channel length (the key dimension) can be much shorter for the MOSFET than for the JFET [93]. In the JFET the size of the p-gate region defines the channel length and a metal contact must be patterned inside this [94]. In the MOSFET, the channel length is generally somewhat smaller than the patterned gate length, especially using a self-aligned technique [95].

If this were the end of the story, the possible gains for GaAs would be insufficient to merit anything other than academic study. However, in 1966 Mead proposed a new transistor structure [96], behaving exactly like the JFET (and therefore available to GaAs), but utilising a Schottky gate electrode [97]. This device was termed the MESFET (MEtal Semiconductor FET). As in the MOSFET, the channel length of the MESFET is determined by the minimum patterned feature size, and there is a significant advantage available to GaAs. Almost

all the research on GaAs transistors and integrated circuits has been performed on the MESFET rather than the JFET. A small effort still remains concentrated on the JFET, as there are still some advantages from using this device [98].

The evolution of GaAs devices does not stop with the MESFET. As already mentioned, the HBT is actively being studied as a means of incorporating the advantages of the bipolar transistor. Equally, the heterostructure can be used in a field effect device to improve the performance [99]. The principle is to use the bandgap discontinuity at a heterojunction to segregate the ionised donors and the free carriers. This allows a high free carrier concentration without the carriers suffering the usual mobility degradation due to scattering from the ionised impurities [100]. Devices based on this structure are now being utilised quite successfully in both digital circuits [101] and high frequency amplifiers [102]. In the latter case, an exceptionally low noise figure has been coupled with a high gain at the same operating point, opening up a very large potential market. Various names have been suggested; HEMT (High electron mobility transistor) [103], TEGFET (2D electron-gas FET) [104], SDHT (selectively doped heterostructure device) [105] and MODFET (modulation doped FET) [106] are all in common usage.

2.1.3 Back-gating

As has already been intimated, the insulating substrate property of GaAs can fail under certain circumstances. The phenomenon by which this can take place is variously termed "back-gating" [107] or side-gating [108], because a voltage applied to a side- or back-electrode close to the device under consideration can modulate the current flowing in the device. However, the side-electrode must be either an n-channel region, or an alloyed contact. A Schottky electrode exerts no influence at all [109]. Further, the voltage on the electrode must be negative and must exceed a threshold value. Superficially, this phenomenon affects the device in a similar manner to the "body effect" in an MOS transistor [110], although both the detailed behaviour and the mechanism involved are very different.

A more closely related phenomenon is that from which the name derives, the back-gating phenomenon in discrete GaAs devices grown on epitaxial layers [111]. Here, a change in the potential on the back surface of the substrate affects the channel region, despite being several hundred microns away. This phenomenon was found to be removed when a high quality buffer layer was grown between the substrate and the epi-layer [112]. Clearly the occupation of traps [113] at the interface was being modulated by the substrate voltage.

Why should the same (or a similar) phenomenon reappear in integrated circuits, even in the absence of the high trap density interface between epi-layer and substrate, and why should surface voltages, even at tens of microns separation, contribute to the effect?

It is apparent that the implantation damage [114] gives rise to a moderate trap density at the channel to substrate interface even following the anneal. This reproduces the effect of the traps between the epi-layer and the substrate found in the discrete FETs. To answer the second question, a model for the effect must be developed.

Firstly, it is necessary to study the conduction in an insulator to explain this behaviour. The electric field between the device and the secondary electrode at the onset of back-gating is substantially below the expected breakdown field of the substrate. However, a detailed inspection of the leakage current between two isolated Ohmic regions formed in the substrate (shown in figure 2.3) reveals that a rapid increase in the current takes place beyond some threshold voltage V_{bg} [115]. The mechanism postulated for this current flow is one of electron hopping between the localised deep level traps in the substrate, in a manner analogous to the method of current flow in amorphous materials [116].

A model (shown in figure 2.4) can be developed to explain the back-gating phenomenon based on this weak substrate conduction [117]. The presence of these deep level traps and this conduction mechanism renders the substrate weakly p-type. Thus a negatively biased Ohmic

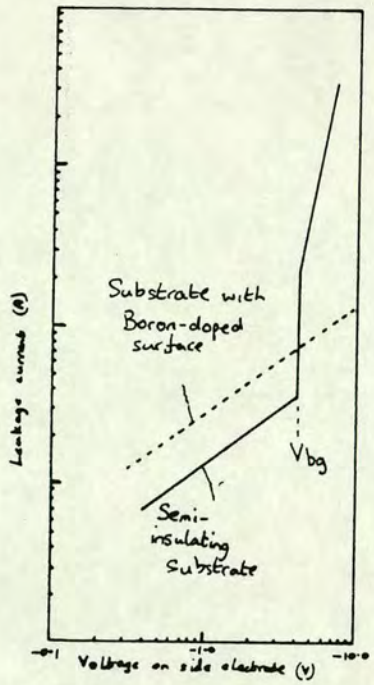


Figure 2.3 Leakage currents in GaAs.

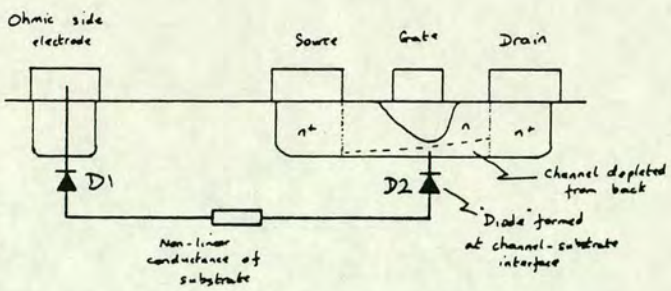


Figure 2.4 Model to explain the effect of back-gating on MESFETs.

region can act as a source of current injection into the substrate, the contact acting as a forward-biased diode, D1. When an n-type channel is formed in the substrate, there is effectively a second diode, D2, between the channel and the substrate. As the reverse-biased voltage on the adjacent contact is increased, whilst remaining below V_{bg}, the field in the substrate remains uniform, and the small leakage current can be supported through the reverse-biased diode, D2. However, when the voltage is increased beyond V_{bg}, the substrate will conduct a rapidly rising current, but diode D2 will not. In consequence, the voltage across the substrate is clamped, and all excess voltage appears across D2. Thus, once beyond the threshold, a small change in voltage on the back-gate electrode can cause a significant change in the channel conduction (and therefore threshold voltage), by depleting the channel from the back.

This model explains why an Ohmic injector is needed for the back-gating electrode, and why a positive voltage has no influence. The reduction of the back-gating threshold under illumination is also explained by this model. The wide variation in the magnitude of the back-gating effect can also be seen. Historically, the poor

uniformity and quality of substrate material led to large spreads on the value of V_{bg}, and thus differences in the reduction in saturation current and threshold voltage.

The simplest cure for back-gating is to move the back-gating threshold voltage outside the operating region [118]. Depending upon the substrate quality, this voltage may vary from a few volts to several tens of volts. Clearly no problem is envisaged for the DCFL circuits (see Section 3.3), in which the total voltage may be less than 1.5-2 V [119]. However, in BFL circuits, the total voltage in the circuit may be 7-10 V [120], and the problem may be quite serious. Because of the large charge storage capacitances associated with the back-gating model, the apparent back-gate voltage at any node will be both a time and a spacial average of the voltages on all nearby Ohmic electrodes [121]. This can act to eliminate the back-gating problem even where a single voltage exceeds the back-gating threshold, provided that the average back-gate voltage does not. However, it can also act to make the circuit performance both pattern and frequency sensitive. It has been claimed that careful layout of circuits can eliminate the problem [122], but this is only a solution for small, hand-crafted circuits. A much safer strategy is to ensure that the back-gate threshold voltage exceeds the total circuit voltage by some suitable margin.

It has been found that implantation of mid-gap species (protons [123], boron [124] or oxygen [125]) into the surface of the substrate (but masking the active areas) improves the back-gating threshold significantly, especially when that threshold is low. This is not to be expected, as the model so-far described is a bulk model, and the implant only affects at most the top 1 μm. On studying the leakage characteristics before and after this step, the conventional leakage current worsens as a result, but the threshold voltage is significantly increased [126]. The model shows that the leakage current is not a problem, but that the voltage clamping effect of the rapid turn-on is the major culprit. By moving this turn-on, even at the expense of the leakage current, the problem has been eradicated. For the surface implant to be effective, the back-gate threshold of the bulk of the wafer must be much higher than the surface region,

otherwise the breakdown path would simply be pushed deeper in the substrate. Outdiffusion of the mid-gap traps (Cr, C or EL2) near the surface is blamed for reducing the threshold of the surface region [127]. This would occur during the high temperature cycling of the wafer, (e.g. during an implant anneal). Other evidence supports the suggestion that the necessary outdiffusion can occur [128].

Perhaps a more radical solution to the back-gating problem is to prevent the diode D2 from having any effect. Various methods have been suggested to add a p-type guard ring, either around the drain [129], or beneath the whole transistor [130]. In the latter case, both the transistor turn-off and control of the threshold voltage is improved dramatically.

This is not the only explanation of back-gating [131]. It is possible that surface conduction mechanisms can modulate the interface states found in the gaps between the gate and source or drain, or that surface charges in these regions can be modulated by the gate potential. This alternative mechanism is clearly only affected by side-gate electrodes, and this has led to the insistence of some workers that the alternative name be used. In the majority of cases the problem has been caused by the true back-gating effect, but poor control of the interface between the overlying dielectric and the GaAs can allow the second effect to confuse the picture in a few cases.

The effects of back-gating were discovered during the course of the work of this thesis [132], but the necessary model leading to the understanding of the phenomenon did not come until near the end of the research. Now, the addition of an isolation implant step is almost universal and, although of academic interest, the importance of the back-gating problem has been diminished.

2.2. GaAs IC Fabrication

This Section treats various aspects of the integrated circuit fabrication process. Following the wafer preparation, individual

sections discuss doping, device isolation, junction formation and metalisation.

2.2.1 The preparation of GaAs Wafers.

Ingots of GaAs are grown on a seed crystal placed in a melt using one of two techniques: the Horizontal Bridgeman (HB) technique [133], or the Liquid Encapsulated Czochralski (LEC) process [134].

Growth by the HB method has historically yielded crystals with fewer dislocations [135], but the technique tends to leave a residual p-type impurity. Chromium, a deep level impurity, has often been added to the melt in order to counteract this inherent acceptor level [136]. The impurity level due to chromium pins the Fermi-level near the centre of the bandgap, rendering the wafer semi-insulating (SI) [137]. Ingots of HB material are also of flattened cross section, and wafers are cut along a diagonal to obtain the circular wafers with which the industry (and most processing equipment) is familiar. Each wafer thus shows a gradation in "age" across its surface. The chromium tends to deplete from the melt as the growth proceeds, giving a steady variation in the chromium concentration across the wafer, with the top edge showing the highest concentration. Subsequent processing is affected by this variation, and considerable non-uniformity of device parameters can result [138].

The LEC technique does not suffer such variations and it is now almost universally favoured. The rotating seed crystal is drawn vertically out of the melt at a controlled rate. The resulting ingot has a circular cross section, and when sawn into wafers, each wafer represents a zone of isochronous crystallisation from the melt. Originally chromium was also added to LEC material, generally in large quantities. Batch to batch variations of the activation of subsequent dopants were found as the differing quantities of residual chromium gave an uncontrolled degree of compensation [139]. The chrome was also found to redistribute very rapidly during processing, giving peaks of Cr at the surface, with sub-surface regions almost totally depleted [140]. Even the type and method of formation of the surface capping layer could have a major influence on the device

parameters [141]. The properties of devices prepared on such material were almost totally unpredictable, giving very low circuit yields.

By reducing the amount of chromium added to the melt to be just sufficient to compensate the residual impurities exactly, many of these problems were at least reduced in scale [142]. More recent studies have shown that SI material can be obtained without recourse to such deliberate overdoping, provided that the melt constituents are tightly controlled [143]. This controls the concentration of the native defect EL2 [144], which is itself a mid-gap trapping level. Carbon is another deep level impurity which is inevitably present in the melt [145]. Like EL2, this helps to give a SI substrate, but it must be controlled accurately to prevent excessive (or inadequate) compensation of the wanted shallow impurity levels.

Even with such careful control, there is still some residual variation in properties across the wafer, arising from the inevitable temperature gradient found across the cooling boule. The distribution of sheet resistivity across the diameter of a slice shows a characteristic "W" pattern [146], but the amplitude is much less than for the equivalent HB grown sample. A similar pattern is found in the distribution of etch pit densities (EPD), showing that the resistivity variations are linked to the numbers of dislocations found across the wafer [147]. These are a consequence of the large stresses experienced during cooling. Further studies have shown that it is not the dislocations which cause the resistivity variations, but the segregation of impurities or defects (Carbon or EL2) to these dislocations [148]. Thus, the clustering of these defects around centres of high dislocation density results in localized variations in the Fermi-energy. It is this clustering effect which causes the non-uniform electrical characteristics. Long (several hours), high temperature anneals following growth, but preceding further processing allow these mid-band traps to distribute randomly, giving a much more uniform Fermi-level, despite the still high dislocation density of the material [149].

An alternative technique for improving substrate quality relies on eliminating the dislocations, thereby preventing segregation of the important mid-band centres. In this technique, indium is added to the melt [150]. Unlike the earlier addition of chromium, the intention is not to change the electrical characteristics, but to change the physical properties. Indeed indium is an isotype of gallium and thus has no "doping" influence. The incorporation of indium into the lattice at about the 0.1 to 1% level strengthens the crystal allowing fewer dislocations to form during the cooling phase. There is an obvious analogy with the strengthening of steel by the incorporation of carbon, although here the carbon occupies an interstitial site. One may expect a reduction in the electron mobility due to the electron scattering from the localized energy levels around the indium site, but in practice, no such effect is observed [151]. Naively, one may recognise that the compound InGaAs is a semiconductor whose electron mobility is still higher than that of GaAs [152], and suggest that perhaps the mobility in indium-doped GaAs might be expected to better that of more conventional material. Such an argument is totally spurious, ignoring as it does the physical origin of the band structures and associated scattering mechanisms in crystals. In view of the differing sizes of the indium and gallium ions it appears somewhat fortuitous that no degradation in mobility is found.

Recently, GaAs ingots with low defect concentrations have been grown in a strong magnetic field [153]. It is claimed that the temperature of the cooling crystal may be more uniform under these conditions, but the mechanism for this is not clear.

2.2.2 Doping in GaAs

Obtaining intentionally doped layers in semiconductors is arguably the most important step of an integrated circuit fabrication process. The same three techniques adopted in silicon IC processing are to be found in use within GaAs technologies. These are: epitaxial growth with in-situ doping; diffusion and ion-implantation.

FETs fabricated on epitaxially grown GaAs layers tend to have better microwave [154] and low noise [155] properties, particularly where an

undoped buffer layer is grown between the GaAs substrate and the doped epi-layer [156]. Traditionally, the isolation between devices has been achieved in this structure by etching away the surface (doped) layer. This mesa-etching process leaves islands of active areas standing above an insulating substrate, giving a highly non-planar topology. Isolation may also be achieved by doping the non-active regions with a mid-gap impurity, to compensate the shallow donor concentration. The "undoped" buffer layer, if used, will generally be weakly p-type, giving a relatively large isolation capacitance compared with the alternative techniques, when this overdoping ploy is adopted. Epi-layers tend to be restricted to the monolithic microwave ICs, in which the analogue performance is all important and where the relatively low integration level can tolerate loss of yield due to non-planarity.

Diffusion is uncommon for digital ICs, for similar reasons to those prevailing in silicon, where it is no longer fashionable because of the poor control of the final profile, and the lack of sharp interfaces. Both of these are critical to achieve devices of the highest performance. Impurity diffusion in GaAs also suffers some loss of control because the diffusion temperature has a natural ceiling at the dissociation temperature of the GaAs, unless extreme caution is taken.

Ion-implantation is thus the dominant technique for layer doping [157]. Using this technique, the dopant can be reliably delivered into a uniform shallow layer, with a sharp fall-off in concentration into the substrate.

The channel beneath the gate electrode is the most important region of the transistor. As n-channel devices are required, this region must be doped with a donor impurity, the obvious choice for which is one of the group VI elements, as the ionicity of the crystal will preferentially force such an impurity to the arsenic site. Oxygen is unsuitable as it forms a deep rather than a shallow impurity [158], but sulphur [159], selenium [160] and tellurium [161] have all been reported. Sulphur is highly mobile, and is too easily guided into the "channels" in the crystal giving a long impurity tail [162], with

an ill-defined junction depth. Whilst this is inappropriate for the accurately defined channels, it may still be useful for contact regions or for low resistance diodes [163]. Tellurium is a somewhat heavy ion, requiring a very powerful accelerator to dope anything other than the shallowest of layers. Although not optimum, selenium is therefore the most commonly used of these elements by default. Even here, the required accelerating energy may be beyond the capability of many commercial machines (up to 400 keV [164] may be required, with only 200-250 keV a typical maximum available). Considerable damage is caused to the substrate by such a large ion, and this may be difficult to remove. The most successful ploy is to heat the substrate to some 3-400°C during the implantation [165], under which conditions the worst damage is self annealing.

Perhaps the use of selenium would have been explored further had there not been an alternative. Surprisingly, silicon can be used as a reliable n-type dopant [166]. One may expect a silicon atom to be an amphoteric impurity, depending on its incorporation on an arsenic or a gallium site. In practice, the gallium site is strongly favoured, leading to free electrons in the doped material. Unlike selenium, the damage caused by implantation of silicon is not excessive, nor the required accelerating potential too high; and unlike sulphur, silicon does not diffuse too rapidly. Consequently, silicon is used throughout the industry as the main donor impurity.

Although p-type GaAs has little use for device channels, it is assuming an increasingly important role. Once used only for the p-gate by those interested in JFETs [167], it has become apparent that excellent n-channel FET performance can be obtained by burying a p-type region beneath the device channel [168]. Two useful purposes are served: firstly to give a well defined cut-off between channel and substrate, and secondly to increase the sharpness and reproducibility of this interface by allowing the peak of the p-layer to compensate the poorly controlled implant tail.

When diffusion was the major doping method, zinc was the most commonly used acceptor species [169]. With implantation, magnesium is the most common [170], although beryllium is a much more

favourable atom [171], especially for the tightly controlled deep buried layer described above. The main inhibition here lies with the safety hazard: beryllia (the oxide) is a highly dangerous material [172].

The implantation of the ions is only the first stage in producing the required conducting layers. Here the distinction is being drawn between impurity concentration and free carrier concentration. The two may not be equal, as an impurity may be electrically inactive (the expected hole or electron remaining bound to the parent atom). Implantation produces an accurate distribution of the chemical species, most of which are still electrically neutral. Accompanying this distribution of impurities will be a further distribution of damage to the crystal; this damage may be sufficiently extensive to amorphise the surface region. A high temperature anneal must accompany the implantation, to serve two purposes. Firstly, the damage centres can move freely, allowing the surface to recrystallise if necessary. Secondly, the dopant can be incorporated into the lattice, allowing the carriers to be freed from the localized impurity atom. This second process is called activation.

Such a post-implant anneal is well known in silicon IC processing, and presents few problems. A temperature of 800-1000°C is typical [173], this being held for up to thirty minutes. Clearly, the higher the temperature, the shorter the anneal time which is required. A great deal of attention is now being paid to rapid transient annealing (RTA) [174], in which the temperature is ramped in a matter of seconds up to 11-1300°C where it is held for a similar short period, before being rapidly cooled. In this short time, little diffusion of the impurities can take place to distort the sharp as-implanted profile. The species is activated because rearrangement of the atoms is only necessary in the vicinity of the impurity. Similarly, if the damage caused during the implant is only minor, it should be removed by this localized rearrangement permitted in this high temperature, short time regime of diffusion.

Several techniques have been employed to produce the necessary transients, including the rapid scanning of an energy beam (electron-

beam [175] or laser [176]) across the wafer in a raster pattern. Because tiny areas are annealed at any time, the effective thermal mass is small, and very rapid temperature transitions may be recorded. Equally significant however, is the large localized stress to which the wafer may be subjected, and which may induce defect centres of its own [177]. By far the most widely used technique uses a large battery of flash lamps and mirrors, with the wafer sitting on a carbon susceptor at the centre of focus of the beams [178]. This is a whole wafer (and potentially a whole batch) process, with obvious advantages for production. The temperature gradients are however somewhat less spectacular, especially during cooling. Nevertheless, this process offers an excellent compromise.

As we turn to the logistics of annealing the implant damage in a GaAs process, a completely new problem emerges. The annealing task is identical to that just described for silicon, therefore requiring the same conditions. However, the arsenic vapour pressure is such that there is considerable dissociation of the wafer when the temperature exceeds 600°C [179]. The surface may be completely distorted, with nodules of metallic gallium being found [180].

Several solutions have been found to this problem of wafer degradation, amongst which are:

i. performing the anneal in arsine gas to create an over-pressure of arsenic and prevent further outdiffusion [181].

ii. Producing a localized over-pressure of arsenic by the proximity annealing technique. Generally an arsenic doped silicon wafer is placed in contact with the surface of the GaAs wafer [182].

iii. Preventing the arsenic escaping by capping the wafer with a suitable pin-hole free dielectric (silicon dioxide [183], silicon nitride [184] and aluminium nitride [185] have all been used successfully).

Increasingly, RTA is being used as the solution to the problem. By preventing excess diffusion, arsenic cannot be lost, other than from the surface monolayer.

All these techniques can be used quite successfully in IC processing. It is still true, however, that obtaining a working process in the first instance is very much a black-art, with two apparently identical recipes producing widely different results. The capped anneal is probably still the most common method in use.

2.2.3. Isolation between Devices.

As well as producing doped layers for active devices it is important to isolate the regions between devices. Here GaAs starts off with an advantage over silicon because of its semi-insulating (SI) form [186]. However, the earliest isolation technique to be used in ICs followed the mesa etching principle. Here, a uniformly doped surface region is formed on an SI substrate, either by epi-layer growth, or by blanket implantation into the substrate. The unwanted regions are then etched away, through the doped surface region and into the SI region below. This leads to a highly non-planar wafer, giving obvious problems of step-coverage when metalisation is later added.

This technique was soon superseded by the planar technique, in which the active areas are defined in the SI substrate by selective doping [187]. The area between devices remains insulating. Significant problems have been encountered with this type of isolation due to back-gating, as explained in Section 2.1.3.

Implantation with one of a number of deep-level impurities (oxygen [188], boron [189] or hydrogen [190] are used) improves the isolation by moving the back-gating threshold well beyond the operating region of the circuit [191]. This practice was introduced to supplement the inherent isolation between selectively doped islands following the identification of the back-gating problem. It has since been realised that the channel doping can be fully compensated by this technique, so that blanket doping can be used for the active areas, with a selective area overdoping for the isolation.

2.2.4. Junction Formation.

Following the doping of the active regions, suitable junctions must be formed to give the desired device characteristics. Two types of interface are readily identified: the rectifying junction and the Ohmic junction.

There are two different types of rectifying junction to be considered: the p-n junction and the metal-semiconductor junction. As discussed earlier, the metal-oxide semiconductor junction is of little interest in GaAs, because of the unstable charge trapped at the interface.

The p-n junction [192] is formed by diffusion or implantation of one type of impurity into a region of the opposite type. In this respect it does not represent a further technological development over those discussed earlier in this Chapter.

In GaAs, the metal-semiconductor (or Schottky) junction [193] is the more common of the two types. Here, a suitable metal must be placed in contact with the clean semiconductor surface. The surface state density [194] is such that the built-in potential in the semiconductor is virtually independent of the choice of metal [195]. This is in marked contrast to the silicon Schottky barrier where the metal work function determines the built-in voltage [196]. Consequently, many different metal systems can be used for the GaAs Schottky junction, and the choice of metal is dictated by the requirements placed on it by the subsequent processing steps, by the regard for long term reliability, and by the capabilities of the metal deposition equipment.

Early work saw much use of titanium-gold (Ti/Au), with titanium forming a highly adherent layer and gold providing a high conductivity and corrosion resistance. With more sophisticated deposition tools (e.g. electron-beam evaporator [197], magnetron sputterer [198], etc.) the more refractory metals could be deposited to advantage [199]. Following the deposition of Ti/Au, there is an upper temperature limit of 200°C to avoid interdiffusion of the two metals and penetration of Au into the GaAs, the latter leading to

degradation of the junction [200]. By interposing platinum as a barrier layer between the titanium and the gold, the allowed temperature range can be extended to well over 300°C, giving a useful extra margin [201].

Even this higher temperature stability is far below that required for some of the more sophisticated process sequences now being used. One such example arises when the gate metal is deposited prior to the source and drain extrinsic implant to minimise the parasitic resistance [202]. In this case, the post-implant anneal must be performed with the Schottky contact already formed. This places rigorous demands upon the Schottky metal, with the junction being subject to 8-900°C. Various solutions to this problem have been suggested, very many of these by one laboratory (Fujitsu), suggesting that each is in some way inadequate and that the optimum solution is still to be found. Most of the solutions revolve around tungsten [203], its silicide [204] or a titanium tungsten alloy [205] or co-silicide [206]. The battle is to find the compromise between the high temperature stability and the low electrical resistivity.

The second type of junction is the Ohmic contact. This is required simply to be able to contact to the active regions, and inject current into (or extract it from) the channel region. As the name implies, the voltage drop across such a junction should be linearly related to the current. In reality this type of junction is unlikely to obey Ohm's law exactly, but for all practical purposes, the linear approximation should be valid over the current range of interest. Generally a much more useful definition would require that the voltage drop at the interface is small, and that current is not inhibited in either direction.

In silicon an Ohmic contact can be formed by using a metal which gives a very low built-in potential, provided that there is a moderately high doping level in the semiconductor [207]. When the doping is degenerate, there are sufficient dopant species such that the impurity energy level forms a band merged with the conduction band (or valence band) [208], and the Fermi-level lies fully within this band. The built-in potential of such a metal-silicon junction

is reduced so that the Schottky junction becomes more like the Ohmic junction. The high doping level also means that the band bending is spread over a much shorter distance, the barrier becoming almost transparent to tunnelling electrons, thus forming a good Ohmic contact. Very often, the Ohmic contact is formed in this way unwittingly. For example, the aluminium (Al) contact to silicon arises (at least in part) because the Al diffusing into the silicon dopes the surface [209]. In p-type material, this is a p^+p junction, and on highly doped n-type this is a p^+n^+ barrier, which behaves like a short circuit because of its very low breakdown voltage.

In contrast to the wider choice of suitable metals for a Schottky contact to GaAs, the choice of Ohmic metal is much restricted. The high interface state density precludes the option of a low barrier height junction. Although the use of a highly doped surface region helps to lower the resistance of the contact, only the alloyed junction is in widespread use [210]. Throughout the industry only one metallurgy has any favour, and this is the gold-germanium-nickel system. It is thought that the nickel moves rapidly during the alloying (typically 450°C for 15 min), transporting the germanium with it [211]. The germanium alters the band structure to allow the contact to be made, and the gold remains largely on the surface to prevent oxidation, allowing a subsequent metal layer to make good contact.

The different constituents in the Ohmic metal often segregate into small islands, some of these being prone to corrosion during subsequent processing [212]. Where the Schottky metal is formed at a later stage of the processing, it is convenient to overlay the Ohmic contact to improve both the morphology and the resistance to corrosion, and also to reduce the contact resistance [213]. Where the processing does not follow this sequence, for example when the gate metal is used as an auto-registered mask for implantation [214], the Ohmic regions are covered at the earliest opportunity. Often, in these structures, the first interconnection level is distinct from the gate metal [215], and this would be used to protect the Ohmic contacts.

2.2.5 Interconnection.

In any IC technology, at least two levels of interconnection are required to enable the active devices to be joined into circuits. In early silicon ICs, only one of these interconnections was metal. Deep diffusions, isolated from the rest of the substrate, were used for very short sections and typically one or two metal tracks could be crossed in this way [216]. Later, polysilicon became common as one of the connections [217] and because of its much reduced capacitance by virtue of the oxide underlayer, tracks in this could be run somewhat further. However, both of these measures are inappropriate where ultimate speed is required, and two (or more) levels of metal, separated by a dielectric layer (typically 1 um thick) are now used extensively [218].

In GaAs, neither of the inferior solutions is available anyway. Clearly, there is no polysilicon layer already present in the device structure. A diffused underpass is inappropriate, because the Schottky metal (forming the other interconnect level) would form a junction to such an underpass, there being no intermediate dielectric. Equally, the absence of a dielectric, coupled with the morphology problems discussed earlier prevents the Ohmic and Schottky metals from forming the two interconnection materials. Clearly, the two-level metal solution is the only option.

The Schottky metal is required to have a low resistivity to ensure a low resistor-capacitor time constant for the long thin gate finger. It can also be placed directly onto insulating GaAs without forming a junction and it therefore has a low capacitance. Both of these features ensure that the Schottky metal may normally be used for one of the necessary routing levels.

When a refractory self-aligned gate is used, the Ohmic contact is formed after the Schottky metal is deposited. If the gate metal is now used as the first level of interconnection, then the source and drain regions must be contacted with the second level of metal. It is unwelcome practice to demand that the second level should contact directly to the Ohmic metal (after cutting a window in the dielectric), and the gate level is therefore not used as an

interconnection. Instead, a further metal (still representing the first level of interconnection) is added directly to the wafer to make contact to both the Ohmic contact and the Schottky metal, no intermediate dielectric having been deposited and cut.

Following the first level of interconnect, a dielectric is deposited, small windows (called contacts or "vias") are cut where a connection between the two levels is required, and the second level deposited and patterned. In silicon technologies, this interlayer dielectric is commonly silicon dioxide of some form (deposited rather than thermally grown), although silicon nitride might be contemplated. In the latter case, the higher dielectric constant [219] is not attractive for high speed operation, as the capacitance at a crossover is much increased. A more recently adopted material is polyimide [220], which can be spin-deposited like a photoresist, with useful planarisation properties to aid in the metal step coverage. Similar materials are employed for the dielectric in GaAs processes.

The advent of GaAs technologies highlighted a potential problem with quality of the interlayer dielectric. For the first time the high frequency loss factor of this layer might be called into question, especially with the adoption of some of the newer materials (e.g. polyimide). In fact no problem has been found except where the cure-temperature of the polyimide was insufficient to eliminate all the moisture [221]. Several other reliability hazards have been linked with polyimide, again often associated with poor curing cycles [222].

However, the effect of the interlayer dielectric can still be important in optimising circuit performance. In analogue circuits designed to operate at tens of gigahertz, a very low inductance interconnect system is required [223]. This can be achieved by forming the metal patterns in the usual way, before thickening it by electroplating to some three or four microns [224] compared with the more conventional 0.5-1 μm. In order to pattern high-Q inductors, spirals are printed in the metal [225], and it is necessary to make a low capacitance connection to the centre of the spiral. The most successful technique is to use this thick metal as a self-supporting bridge structure, using air as the low capacitance dielectric [226].

A conventional dielectric is usually used to support the bridge during fabrication, and this is later etched away [227]. This same technique has been copied into a digital technology, claiming a reduction in track capacitance of some 60% [228]. There is a further saving because such bridges may pass directly over the active transistor, where tracks are normally prohibited (with a conventional dielectric, there is always the possibility of anomalous charge build-up in the dielectric which will change the surface potential in the gaps between the electrodes, changing the transistor threshold voltage, as in a standard MOS device).

2.3 MESFET Logic

2.3.1. Introduction to GaAs MESFETs

It was shown in Section 2.1 that the MESFET is the most important active element for GaAs digital ICs, with the JFET being the alternative. Both these devices, however, behave in like manner, the channel of the transistor being constricted by the depletion region associated with the gate-source junction. In the case of the MESFET this is a Schottky junction, and for the JFET a p-n junction. Only n-channel transistors are considered, as the greatly inferior hole mobility of GaAs renders the p-channel device no advantage over the silicon counterpart.

A distinction is made between those devices in which the channel between source and drain is conducting when no gate-bias is applied, and those in which a source-drain current can flow under the same conditions. These two types are respectively described as depletion-mode [229] and enhancement-mode [230] transistors, in keeping with the terminology used for MOS transistors [231]. However, this description is misleading, as the conductance of the channel is modulated by changing the degree of encroachment of the neutral depletion region into a fixed size conducting path. This is in marked contrast to the operation of a MOST where the amount of inversion charge forming the conducting channel is itself either 'depleted' or 'enhanced'. In order to preserve this distinction, an alternative notation may be used [232], in which the terms 'normally-on' and 'normally-off' replace 'depletion' and 'enhancement'. Using this notation, the actual mode of operation of the MESFET device is

more accurately described. Throughout this thesis, the latter terminology will be used.

Theoretical considerations as to the mechanisms involved in MOST operation do dictate this division into two types. A similar break is not apparent in the case of the MESFET, but it is still convenient to maintain some distinction, because the type of circuit which may be designed with these devices does indeed show such a division. Again the normally-on and normally-off descriptors reflect this practical consideration. When studied from the circuit viewpoint, a third description may be used: namely 'quasi-normally-off' [233]. This new category reflects the fact that circuits can be designed which do not demand a fully cut-off channel in the absence of a gate bias, but which will still work with a weakly conducting channel.

The difference between these categories of transistors (whether or not the source-drain channel will conduct with no voltage applied to the gate) is controlled by the channel doping, the physical channel thickness and the Schottky barrier (or built-in) potential [234]. In GaAs, totally unlike silicon, the surface potential is strongly pinned by the surface states [235], such that the work-function of the metal used for the barrier has almost no effect on the built-in potential. This lies between 0.7 and 0.8 V, with the actual value being dependent upon the surface preparation. The distinction between normally-on and normally-off devices is thus entirely physical, being controlled only by the magnitude and depth-variation of the donor density in the channel.

The effective channel thickness, including allowance for the dopant distribution within the channel, is most conveniently described in terms of the voltage appearing across a depletion width equal to the actual channel thickness. As with other FET devices, this voltage is always positive and is defined as the "pinchoff voltage" [236]. The difference between the built-in potential and the pinchoff voltage is then defined as the threshold voltage. This is negative for normally-on FETs and positive for those which are normally-off. Again care should be exercised because the term "pinchoff voltage" is often used in the literature to describe the threshold voltage of a

normally-on FET, with "threshold voltage" being reserved for normally-off FETs. The more rigorous definition will be adhered to throughout. Because it is usually negative, some difficulty can be experienced when using a qualitative description for the threshold voltage. The most natural description, in which a highly negative threshold voltage is described as a "large" threshold voltage, is adopted here.

A further difference between the familiar MOST and the less common MESFET (or JFET) is that the positive gate-source voltage of the MOST is limited only by the onset of avalanche multiplication near the drain [237], whereas in the MESFET or JFET it is limited by the onset of forward-biased conduction between the gate and source electrodes. The former limit can be modified by altering the fine detail of the fabrication process, but the latter is fixed by the physics. The major consequence of this is that the useful logic levels are restricted to lie between the built-in voltage (+0.8 V), and the threshold voltage. This causes no problems for normally-on FETs, but imposes severe constraints on the processing of normally-off FETs where the resultant voltage swing is limited to a few hundred millivolts [238]. Consequently the threshold voltage has to be controlled to an accuracy of some tens of millivolts in order to obtain a useful yield of LSI circuits [239].

All the early experience of GaAs FETs was gained with discrete microwave transistors, in which the low noise and high gain properties were paramount. Such requirements demanded normally-on devices, usually several hundred microns wide, and often with a large threshold voltage [240]. Accurate control of neither the channel thickness nor its doping was required for these transistors for two reasons. Firstly, small variations only produce a small effect in such large transistors, and secondly, with discrete devices, the circuit components could be individually tuned, to compensate for the remaining parameter variations.

For digital circuits, significant advantages accrue from the use of a normally-off IC process. These include simpler circuits, simpler layout, and lower power dissipation [241]. Initially, however, the

controlled and reproducible reduction in electrical channel thickness required to transfer from a normally-on process suitable for discrete devices, into one suitable for normally-off transistors proved too difficult. Consequently, early research concentrated on the much more readily fabricated, but much more cumbersome circuits of the normally-on transistor. Only more recently have the rapid improvement in GaAs wafer quality and the adoption of new processing techniques allowed the greater potential (within a digital domain) of the normally-off circuits to be exploited.

2.3.2 Critical History of Normally-on MESFETs in GaAs ICs

The principle limitation of the normally-on FET for digital ICs is the incompatibility of the input and output signals from the conventional single transistor switch circuit. These signals must be made directly compatible in order to cascade logic stages, and extra circuitry has to be introduced to transpose the voltage level at either the input or the output. Alternative methods of achieving this voltage-level-shifting have given rise to several different logic families, each with its own advantages and disadvantages. Three basic techniques can be identified, but the picture becomes clouded by several derivative circuit types.

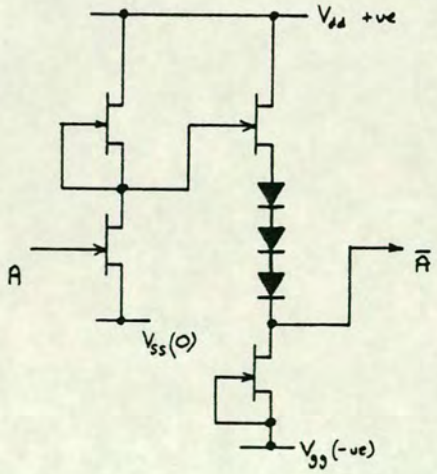


Figure 2.5 Buffered FET Logic (BFL) Inverter.

The first of these techniques, Buffered FET Logic (BFL) [242], uses a two-stage logic cell, as shown in figure 2.5. The conventional switch stage drives into a source follower transistor used as an

output buffer. Forward-biased diodes in the buffer stage produce a constant voltage difference between the output of the first and second stages, thereby giving the necessary level-shifting. Both stages use FETs operated in the familiar "depletion load" configuration instead of resistors. There are sound reasons for preferring a FET load device over a resistor, depending upon the circuit voltage, the voltage swing, and the FET saturation voltage. More pressing, however, at least in the early development of the GaAs IC technology, was the comparative ease of fabricating the FET (an almost identical device to the switch transistor), compared to the resistor. Resistors could be fabricated in a completely separate process step (e.g. a thin film layer [243]), or from a doped GaAs layer. In the latter case the addition of the gate to form the FET improves the tolerancing between the switch FET and its load device. Schottky junctions are used for the diodes, giving a superficial compatibility with the FET gate process - superficial because the doping requirements for the diode are somewhat different from those of the transistor [244].

The first GaAs digital IC employed a BFL circuit design [245]. This operated as a frequency divider at a (then) spectacular 4 GHz clock frequency, albeit at a power dissipation of 40 mW per gate. Unfortunately, this spectacular entrance of GaAs into the digital arena was to be widely misinterpreted, rendering some disservice on two counts towards the future acceptance of GaAs as a viable IC medium.

Firstly, although BFL cannot be described as a low-power technology, the very high power dissipation exhibited by that circuit is not typical of the more modern results [246]. Indeed, a complex chip employing a low-power derivative of BFL has been fabricated with a power dissipation of less than 1 mW per gate [247]. Nevertheless, BFL has been permanently branded as a high power logic family, and has often been dismissed from consideration, when in fact it may have offered a viable solution to the problem at hand.

More importantly for the long term, this first real evidence of the capabilities of GaAs yielded a circuit performance far in excess of

any contemporary result in silicon [248]. This added weight to the extravagant claims which had long been made for the ultimate performance of GaAs [249]. Since that early result, however, the maximum clock frequency of a divider circuit has only quadrupled for GaAs [250], as compared to the near tenfold increase for the fastest silicon circuits [251]. Since the major advantage of using GaAs is its potential for high speed circuits, it can quite reasonably be argued that the performance of GaAs has not warranted the claims made on its behalf. Many of the erstwhile advocates of GaAs have turned prophets of doom as a result of this apparent poor showing.

It is instructive to look at the differences between this first GaAs circuit and the contemporary silicon circuits which led to such wild inaccuracies in predictions based on a direct comparison of performance.

The most obvious of these differences is that only a single result was available for GaAs, and it was unclear how representative this result was, whereas the predictions for future silicon IC performance were based on a wealth of experience. This single GaAs result was in fact based on a circuit used mainly as a demonstrator. The excessive power dissipation was due in part to the very large transistors used. Another significant factor was the large threshold voltage employed for these transistors, and consequently the large power supply voltages required (two supplies are necessary for BFL).

To make a reasonable comparison between the performance of GaAs and silicon based on this one GaAs result, that result should have been scaled to take account of the narrower transistors which would have to be employed in a more practical IC. Unlike the familiar experience of scaling silicon transistors where the capacitances vary in almost exact proportion with the area, the semi-insulating nature of the GaAs substrate leads to a much more complex relationship between node capacitance and device area. Further, when comparing a MESFET with a bipolar transistor, the MESFET "active" capacitance is much smaller, making the "parasitic" capacitance more important. Thus in reducing the transistor dimensions from those used in the demonstrator circuit, the current (and therefore the power) is

reduced with area, but the total node capacitances are reduced by a much smaller fraction, giving a much reduced speed. This GaAs result was therefore totally unrepresentative of the technology at the time.

If a modern CMOS IC process is compared with that used to make the first MOS ICs, it is readily apparent that the modern process has been highly refined in order to achieve the maximum performance. Similarly it was assumed that a refinement of the GaAs processing would pay dividend in improved performance, and that this first result was very much below what could be achieved. Whilst a modern GaAs process has many refinements over that early process which do indeed offer improved speeds, the reduction in performance when moving from demonstrator to real circuit must be offset against this. It should also be recognised that the minimum feature size of that first GaAs IC was 1 um, which was very advanced for its time. More recent results have used sub-micron gate-lengths but, even today, the more useful GaAs IC products are still produced with a 0.7 to 1 um minimum geometry. This feature size marks the approximate boundary between what can be achieved using the relatively simple optical lithography techniques, and that which needs the much more difficult and costly X-ray and electron-beam techniques [252].

Also implicit in this question was the assumption that there was plenty of room for refinement of the new GaAs process, but not much scope for major improvement of the silicon process. This may have been true if applied to the fabrication of MOS circuits, but was far from true for bipolar technology, with which GaAs is invariably compared. For a long period, the technology used for bipolar circuits was relatively stagnant, with the major effort being employed in the rapidly expanding MOS arena. The last five years have seen a radical change, with major improvements in bipolar technology [253]; the adoption of oxide rather than junction isolation [254], and the incorporation of polysilicon [255] being the two most important steps. Both of these have had a major impact on the speed performance of the bipolar technology, most notably by enabling fully self-aligned transistors to be made [256]. Before the introduction of these techniques, the majority of the device area was "inactive". This inactive area was necessary to be able to make

contact to the devices and give isolation between them, but otherwise only contributed unwanted capacitance resulting in a degradation of performance.

Thus, continued (and perhaps unexpected) refinement of the process for the fastest silicon ICs, coupled with the relentless reduction of minimum geometries, has maintained, or even accelerated the increase in speed performance. For GaAs, however, continued refinement of the process has concentrated on improving circuit yields and on increasing the overall complexity. This has been accompanied by some performance improvements, but without an associated reduction in minimum dimensions and, after accounting for the non-representative nature of the first result, this improvement has been much slower than predicted.

The apparent failure of GaAs to live up to expectations is thus attributable to the engendering of false expectations, rather than to a real failure of the material. If the improvements in some of the alternative figures-of-merit (e.g. the power-delay product) are monitored, the apparent anomaly disappears [257].

This criticism must be reserved for the interpretation which has been placed on the resultant performance of this first GaAs IC, and not on the result itself. This was still a highly creditable performance from a relatively simple technology (albeit at 1 um), and the result represented an excellent demonstrator of the capabilities of GaAs. The actual performance was not bettered for some considerable time in any technology.

Following on rapidly from this initial breakthrough, a second technique was employed successfully to produce the Schottky Diode FET Logic (SDFL) circuits [258]. This technology offered a somewhat slower circuit performance, but with the advantage of a much reduced power dissipation. A simple inverter is depicted in figure 2.6.

The reduction in power was not entirely due to the changed circuit configuration, but was (perhaps more importantly) attributable to a completely redesigned fabrication process [259]. The earlier BFL

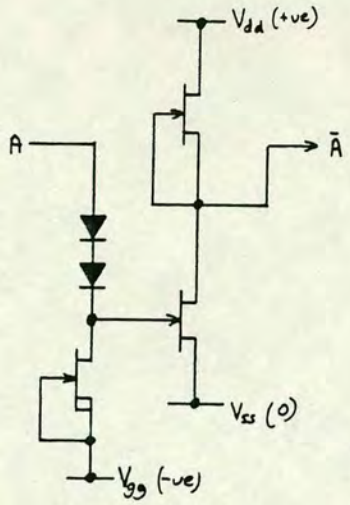


Figure 2.6 Schottky Diode FET Logic (SDFL) Inverter.

result had been achieved using a process essentially identical to that used for discrete transistors [260]. In introducing SDFL, the Rockwell team had "tuned" a process much more closely to the needs of an IC technology. The threshold voltage had been cut significantly to suit the SDFL circuit, such that only a small voltage-level shift was required. Selective ion-implantation into the semi-insulating substrate replaced mesa etching as an isolation technique, resulting in a planar process more amenable to integration. The use of multiple ion-implantation schedules allowed different regions of the devices to be individually optimised. Thus diodes were fabricated with low series resistance whilst the more appropriate threshold voltage of -1 to -1.5 V was used for the FETs. Much was made of the need for tight process control [261] in order to operate with small voltage swings and a minimum degree of level-shifting. The Rockwell laboratories were particularly successful in this aspect.

With hindsight, one can again identify popular misconceptions about the capabilities of SDFL. Because of the claim that fine tolerances were an essential feature of the SDFL process, other research teams were slow to adopt SDFL as a potential circuit technology, preferring instead to follow the HP lead with BFL, which could be made more easily. It is true that the absence of the low impedance buffer stage in SDFL does impose some restriction, especially on the fan-out capability of the circuit. However, at that time the principle

difference between the two technologies was that Rockwell had selected very low power as their main target, paring down all the margins and device parameters to achieve this, whilst the HP team had opted for a more conventional process (similar to the discrete devices in relatively wide circulation), which gave much more tolerance to variations in device parameters, but at the expense of a significantly higher power dissipation. This was a managerial decision rather than an inherent consequence of adopting the respective technologies. However, the respective performance has long been wrongly attributed directly to the design of the respective logic gates.

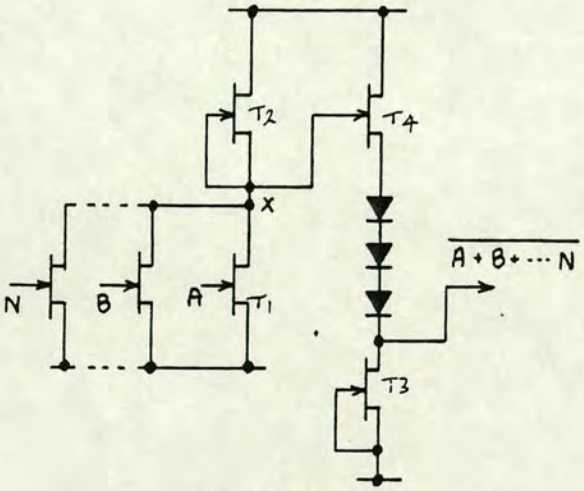


Figure 2.7 BFL NOR gate.

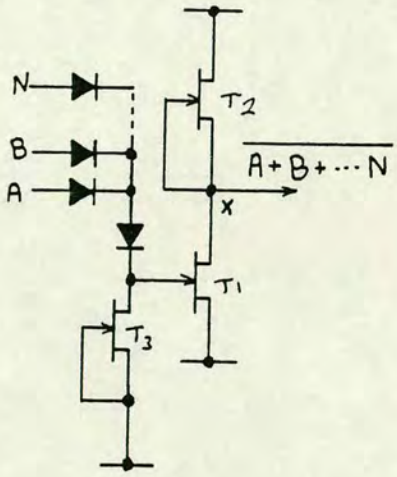


Figure 2.8 SDFL NOR gate.

As well as employing different level-shifting techniques, BFL and SDFL tend to differ in the implementation of more complex logic functions. The method of increasing the fan-in of BFL is indicated in figure 2.7, whilst that for SDFL appears in figure 2.8. In both cases, the expansion of the inverter to a NAND function is achieved by the replacement of the transistor T1 by a dual-gate transistor. Expansion to a NOR function requires an additional transistor in parallel with T1 for BFL, but only the addition of a diode for SDFL. As this diode is small, with a small capacitance, a large fan-in can be readily tolerated for an SDFL NOR gate [262], whereas the equivalent BFL NOR gate demands the addition of several FETs and

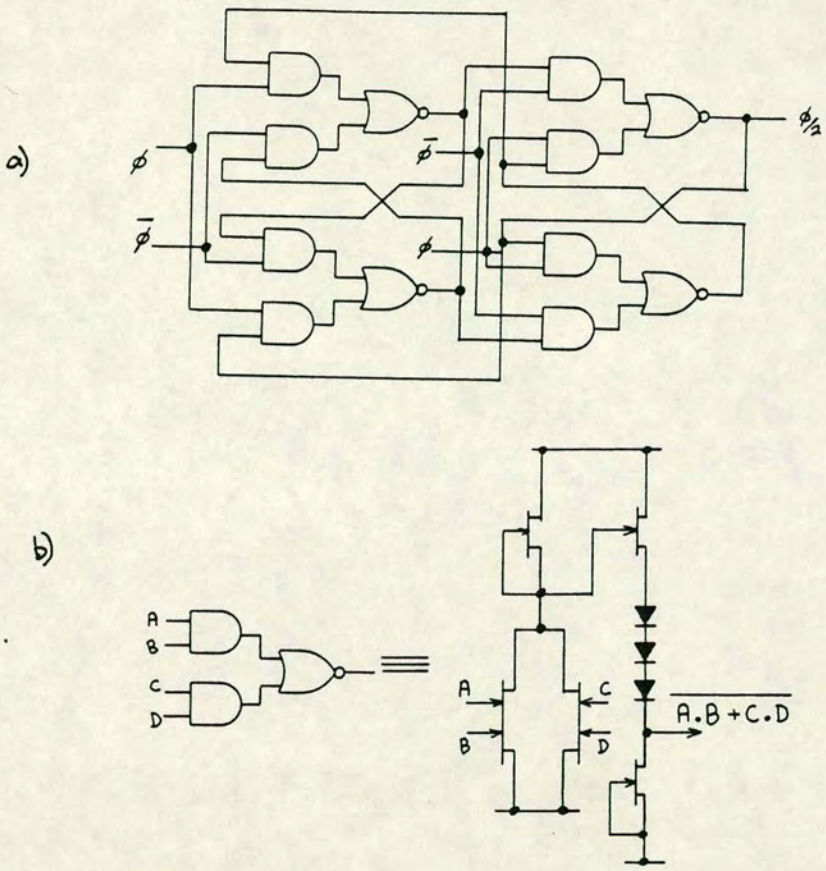


Figure 2.9 Efficient AND-NOR divider in BFL:
a) logic diagram,
b) gate circuit diagram.

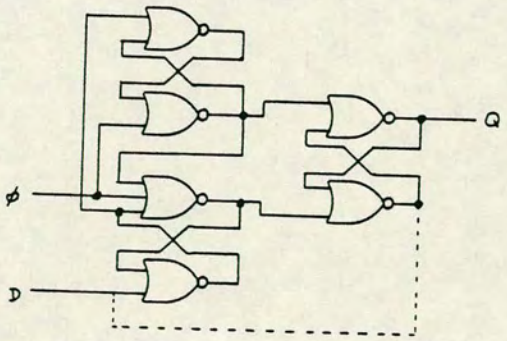


Figure 2.10 Edge triggered D-latch (dotted line shows connection as a divider).

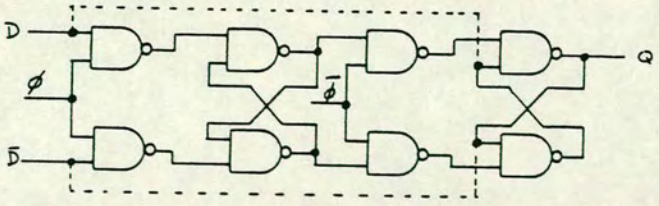


Figure 2.11 Traditional master-slave latch (dotted line shows connection as a divider).

complex wiring. In MESFET logic, unlike the situation in CMOS logic, the NAND configuration introduces a penalty of logic swing, with the maintenance of a consistent output low at node X being particularly difficult. Consequently, NOR logic is preferred in SDFL [263], especially in the Rockwell implementation in which the minimal operating margins are used. On the other hand, the large voltage swing and noise margins adopted in the HP implementation of BFL allow the more layout-efficient NAND gate to be employed.

One consequence of this is that the highly efficient divider circuit shown in figure 2.9 is normally used in BFL, but is rarely adopted in SDFL. This configuration has a maximum clock frequency of 1/2t (where t is the propagation delay of the AND-NOR gate of figure 2.9b). The more familiar edge-triggered D-type of figure 2.10 is commonly used in SDFL circuits, but this has a smaller upper limit on clock frequency of 1/4.85t. A direct comparison between maximum published clock frequency of SDFL and BFL circuits again tends to present a distorted picture because of this difference in usage. A relaxing of the operating margins for SDFL would allow the faster circuit to be used, whilst a tightening of the margins for BFL may force the adoption of the slower circuit. It is interesting to note that the fast divider used in BFL circuits has the same logic implementation as used in most emitter coupled logic (ECL) bipolar circuits [264], but prior to its use in HP's first GaAs divider, this circuit had never been employed with FETs. All earlier FET circuits had used either the slower 1/4t circuit of figure 2.11, or the edge-triggered circuit (figure 2.10).

Although it remains true that SDFL is somewhat slower and less power hungry than BFL, fundamentally the two circuit types share a greater commonality than is generally acknowledged. This became more apparent with the introduction of a modified version of BFL [265] in which the source follower was removed from the voltage-level shifting stage, as depicted in figure 2.12. If the constituent elements of this new circuit are regrouped, as shown in figure 2.13, then the modified BFL inverter is transformed into an SDFL inverter, after making the necessary allowances for different supply voltages and

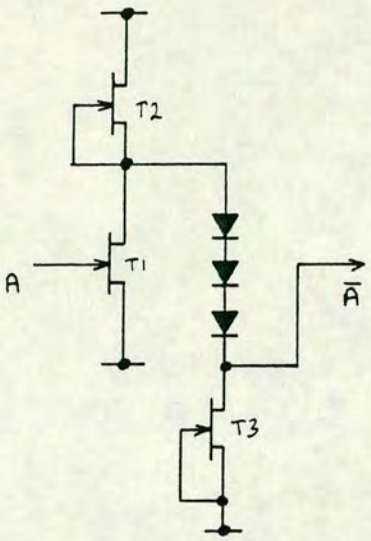


Figure 2.12 Modified BFL Inverter.

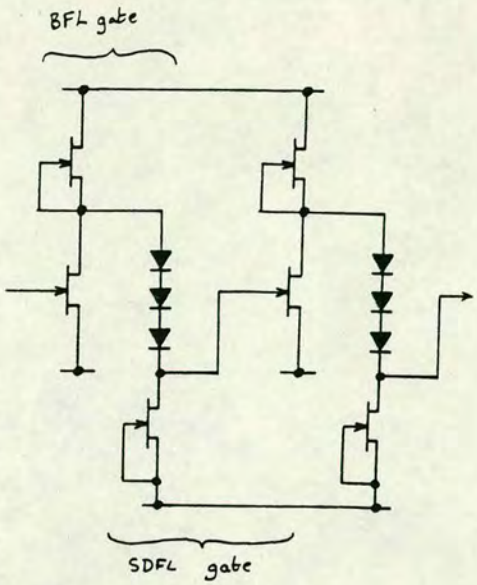


Figure 2.13 Common origins of BFL and SDFL gates.

voltage-level shift. Gates more complex than inverters do, however, retain the distinct identity of the BFL logic type.

The switching behaviour of these modified BFL gates is different from that of the original BFL circuit. Switching from high to low becomes more efficient, but this is at the expense of the opposite transition. To retain symmetrical waveforms in a circuit, the transistor sizes have to be modified to offset this change. This results in a smaller transistor T3, which reduces the quiescent current flowing in the level-shift branch, and gives a reduced power consumption. (Note that there is still a constant current in the level shifting branch of the circuit, but that this flows through T2, rather than through T4 as in the standard BFL gate.) The new gate also offers a faster switching time when it is lightly loaded, because the deleterious saturation of T4 is avoided [266].

In a second circuit derived from the basic BFL concept, the influence of the SDFL configuration can again be seen. This circuit, shown in figure 2.14, builds in greater layout flexibility by allowing the inputs to be distributed between more than one switch branch. These

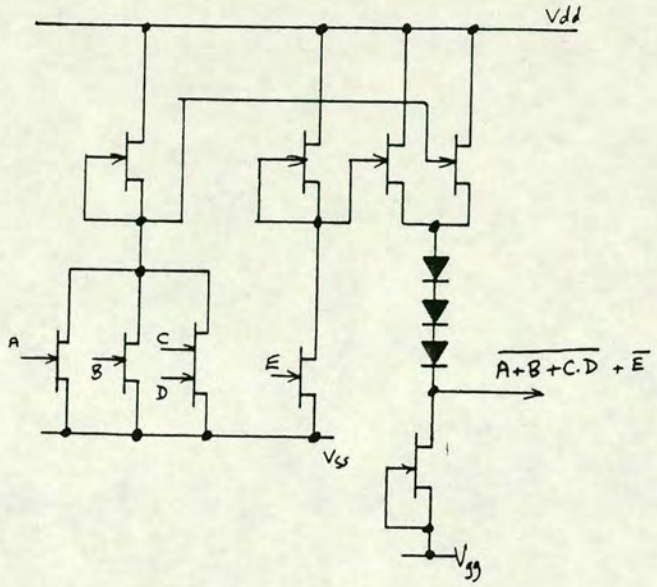


Figure 2.14 BFL circuit with combinatorial logic in buffer as well as switch.

are then merged in the source follower branch (the SDFL input plane). This flexibility was put to good use in a programmable logic array [267].

The third type of level-shifting used in GaAs normally-on MESFET circuits uses capacitive coupling between stages [268], leading to the name CCL (Capacitor Coupled Logic). The circuit is shown in figure 2.15. This type of level-shifting has been employed in the work described in this thesis, with the capacitors being implemented as reverse-biased Schottky diodes. In its most basic form, there is only a single capacitor between logic gates and the switching action relies upon all such capacitors being pre-charged. The charge on the nodes will decay with time, setting a minimum frequency at which every node must be exercised. The advantages which accrue to this technique include the use of only one power supply, an extremely large noise margin with consequent high tolerance to parameter variations [269], and reduced component count leading to simplified layout. The principal disadvantage is the minimum operating frequency, which is typically 1-100 kHz. There are also some

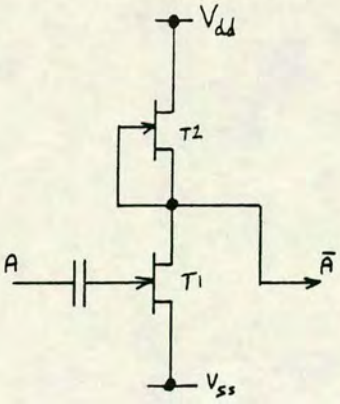


Figure 2.15 Capacitor coupled logic (CCL) inverter.

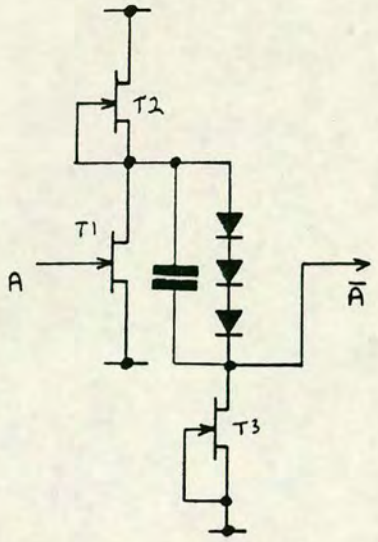


Figure 2.16 Capacitor diode FET logic (CDFL) inverter.

restrictions of freedom in the conversion of circuit function into logic gates [270].

It should be noted that this circuit technique is not the same as that used in dynamic MOS logic [271]. In dynamic logic, the charge stored on a capacitor is different in the two logic states. When the power is removed from the stage the leakage currents are small and the capacitor retains its charge. On readdressing the stage, the original logic state can be reconstituted from the stored charge. Often, the parasitic nodal capacitances are sufficient and the capacitor need not be a physical component. In CCL however, the charge on the capacitor differs only slightly in the two states and the circuit power must be retained whilst it is operating. The function of these capacitors can be more closely likened to that of 'speed-up' capacitors which are added to discrete bistable circuits [272].

By adopting this notion of a speed-up capacitor, the basic CCL circuit can be merged with BFL to form the capacitor diode FET logic (CDFL) [273], as depicted in figure 2.16. This configuration has also been called Feed-Forward Static Logic (or FFSL) [274] to indicate that there is now no lower limit on the frequency of

operation of this type of logic gate. The capacitor can be added to either of the two BFL configurations, with or without the source follower, although the operation is somewhat different in the two cases, as discussed in Chapter 6. In either case, the switching is controlled by the capacitor, with only a small DC current required to flow in T3 to maintain the charge on the capacitor. The power dissipation of this circuit is only slightly larger than in CCL, still much lower than in BFL for comparable performance [275].

2.3.3 Development of Normally-off Logic

The adoption of a process capable of manufacturing normally off FETs leads to a number of advantages. In particular, if the supply rail is chosen with care, the input and output levels are mutually compatible and the switching stages may be directly interconnected. This is the so-called direct-coupled FET logic (DCFL) configuration [276]. Assuming similar transistor sizes to those which would be used in a normally-on circuit, there is a clear advantage in packing density arising both from a reduced component count and from the removal of one power bus. The remaining power supply can be typically 1-2 V, giving a considerably lower power consumption than that of the normally-on circuit.

The large-signal transconductance plays a major role in determining the circuit speed [277]. For a given transistor the transconductance increases with "useful" signal swing [278], and one may therefore expect the larger voltage swing of the normally-on transistor to offer the higher transconductance. This is not always the case. The transconductance is a measure of the "efficiency" of modulating the channel current by the gate voltage. In the higher threshold transistor the screening effect of the depletion region (particularly near cut-off) reduces this efficiency quite significantly. For a given threshold voltage, the transconductance is highest for a delta function of charge at the surface, and lowest for a thick but low-doped channel [279]. The peak doping level which may be used in the channel is limited by the onset, near the gate-drain junction, of avalanche multiplication leading to breakdown [280]. As the voltage in a normally-off transistor circuit is lower, the circuit can tolerate devices with a lower breakdown voltage, and a higher peak

doping level may be used. This in turn allows the dopant distribution to be closer to the ideal of the delta function, yielding the higher transconductance.

It is worth digressing to note that both an MOS inversion channel [281], and a 2D electron-gas layer in a HEMT [282] correspond to this ideal of a delta function and offer this optimum switching performance.

A second phenomenon also contributes to a higher transconductance in the normally-off transistor. In the high electric field region beneath the gate electrode, carriers can be forced into the substrate [283], allowing a small drain current to remain when the channel should be cut-off. This current changes relatively slowly with the gate potential and thus leads to a poor transconductance in the near-threshold region. The sharper the transition from channel to substrate, the more the carriers are confined, and the smaller this effect. The shallower channel of the normally-off FET, therefore, allows a sharper sub-threshold characteristic. Again the HEMT also scores heavily, as the potential barrier between the channel and the substrate confines the carriers [284].

Excluding any technological problems, the significant disadvantage of the DCFL logic arises from the rapid increase in the gate-drain capacitance as the gate-source junction becomes forward biased [285]. This capacitance forms the unwanted feedback capacitance, the effect of which is magnified by the voltage gain of the stage (the Miller Effect [286]). This therefore represents a significant contribution towards the gate propagation delay. During most of the logic transition of a normally-on logic gate, this capacitance is small (typically one tenth of the input capacitance [287]) because the junction is highly reverse biased. This capacitance becomes appreciable only during the final portion of a rising edge (or the initial portion of a falling edge), and this coincides with a small instantaneous voltage gain. In a DCFL logic gate however, the gate-drain capacitance is relatively large for the whole transition and, more importantly, it is large during the period of largest voltage gain [288].

A further cause for concern with the conventional DCFL design can arise when the input of the transistor is overdriven. When the circuit is designed primarily for speed, a relatively large supply voltage is employed, such that the load transistor is nearly always in the saturation regime. When a stage is switched off, the load current is being forced into the gate of the succeeding transistor to give the largest gate voltage possible. This also gives the highest transconductance, and generally the shortest switch time. However, the current density in the gate metal can be extremely high, giving some worry for device lifetime. This configuration is commonly used to achieve "world best" ring oscillator delays, even though it would never be employed in a useful circuit. Even when not intentionally overdriving the stage, a not inconsiderable gate current must flow, and thus it is very difficult to maintain adequate control of the circuit operating points over a large circuit.

It is difficult to quantify individually the impact of each of these advantages and disadvantages of the normally-off transistor. Nevertheless, there is an overall gain in performance over the normally-on transistor. If high speed alone is being sought, then a normally-off circuit cannot match the best speeds of BFL. However, clock rates of 3 to 4 GHz are attainable at less than half the power dissipation of BFL [289]. For slower circuits (around 1-2 GHz), power dissipation may be reduced to about one tenth that of BFL, whilst at the same time offering sufficient packing density to allow the fabrication of 16K RAMs [290].

With such impressive gains already demonstrated, we may ask why there has been so much concentration on the development of a normally-on technology. The answer to this is quite clear. The fabrication of normally-on transistors is much easier than for normally-off devices.

It has already been stated that most normally-on circuits have a reasonable tolerance to variations in the device parameters. They also work well over a range of supply voltages, and offer reasonable noise immunity. In the early days of GaAs ICs, variations of threshold voltages across a wafer could easily show a standard deviation of some 200 mV or more [291]. Wafer to wafer variation

could be much larger still. With a typical logic swing of only 500-700 mV, no DCFL circuit could have been made successfully.

Such variations were common in both epitaxially grown layers and ion-implanted layers, despite the more accurate dosimetry associated with implantation. Evidently the large variations in the threshold voltage were not being caused by poor control of the dopant distribution, but by incomplete activation of these dopants. Gradually, the material quality has been improved, as described in Section 2.2, such that uniform activation can now be achieved.

In addition to the problems of consistent quality of starting material, the specific fabrication steps used for wafer preparation have also been shown to contribute to these variations, causing some workers to identify phenomena which others do not see. In particular, stress in the overlying dielectric has been shown to influence both the rate and position dependence of the dopant redistribution and activation [292]. This can in turn lead to an orientation dependence of device parameters because of the anisotropy of the stress tensor for GaAs [293]. A gate-length dependence of threshold voltage is also observed for channels of less than 2 um in length, but this again is a consequence of the type of processing employed, rather than an intrinsic property of GaAs [294]. Even though some foundries may not experience all of these effects, an understanding of their origins is important to achieve long term reliability and products of high quality.

After expending considerable effort, firstly in identifying these effects and, secondly, in eliminating them where possible, variations in threshold have been reduced to a few tens of millivolts, with a best reported result for "dislocation-free material" of some 8 mV [295]. The pattern of this work has followed much the same path as that already trodden in improving the quality of silicon wafers. There is still a large gulf separating silicon and GaAs in terms of wafer quality, but this "new" dislocation free GaAs material looks promising for the future. Early experimental evidence saw the best GaAs results from laboratories with in-house crystal growth, but this

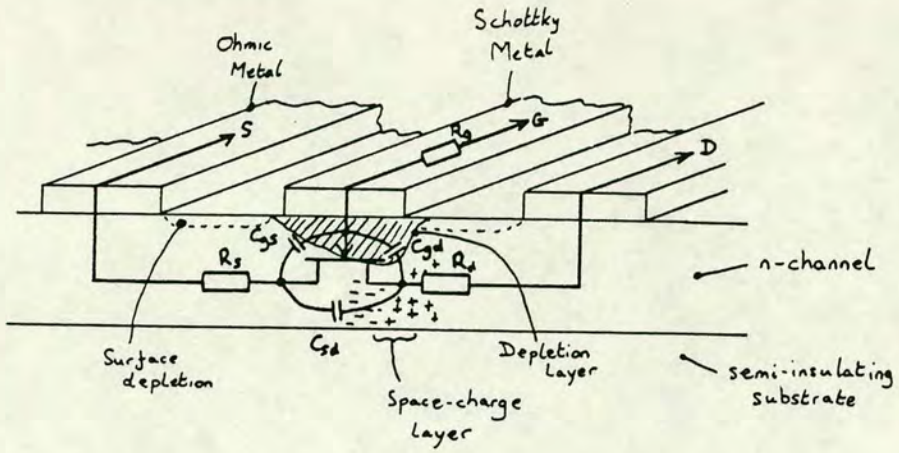


Figure 2.17 MESFET cross-section and electrical model.

contrast has diminished with continued improvements in the quality of commercially supplied material.

Another totally unrelated technology-dependent problem found in GaAs affects the viability of a normally-off process much more seriously than it influences the normally-on process. This is the problem of surface depletion of the channel in the space between the gate and source/drain electrodes. The strong pinning of the surface results in a depletion region extending down into the channel, in much the same way as that beneath the gate electrode. This surface potential can vary between 0.3 and 0.6 V, with the actual value depending upon the dielectric material being used and on the surface treatment during the processing [296]. Its effect is to increase the resistance of the "extrinsic" region. Figure 2.17 shows a simple electrical model for a practical FET, and illustrates the way in which the "extrinsic" regions modify the behaviour of a real device as compared with the ideal behaviour of the "intrinsic" FET. These extrinsic regions are passive (i.e. non voltage-dependent), but influence the behaviour of the FET over the whole of the I-V characteristic because of the reduced terminal voltages at the intrinsic FET. The source resistance is particularly important as it acts as a feedback element, reducing the transistor gain. This change is reflected in the transconductance, g_m, according to:

$$g_{meff} = \frac{g_m}{1 + g_m.R_s}$$

The source resistance, R_s, of a normally-on transistor may reduce the available transconductance by some 30% or so, having a significant impact on speed [297]. However, for a normally-off transistor the surface depletion may correspond quite closely to the electrical channel thickness, giving an almost open-circuit channel and a very low effective transconductance. Although the problem is well known in all FET devices, it is particularly acute for GaAs because of the very high surface-depletion potential.

The solution to this problem is to compensate for the surface depletion by increasing the thickness of the extrinsic channel compared to the intrinsic channel. This can be achieved either by recessing the gate and making a physically thicker region, or by increasing the concentration of free carriers beyond the gate region to increase the electrical thickness. Both methods have been adopted with success, each with its own advantages and disadvantages.

Gate recessing may be achieved by two techniques. The most common form is to etch a groove locally into the GaAs before depositing the gate material. The actual procedure adopted may vary, but self-aligning the gate with the recess is easily achieved by using a lift-off technique [298] to pattern the gate metal, with a single photoresist mask being used for the two steps. Here the main problem is to control the etching uniformity and retain the desired control over the threshold voltage. The second such technique uses a platinum gate electrode [299]. This can be sintered into GaAs at a modest temperature (c 250°C), with the position of the metal semiconductor junction burying deeper into the material as the temperature or time are increased. Fortuitously the junction so formed retains its rectifying properties with a high degree of ideality. Gate recessing was the earliest technique to be used because of its relative simplicity, and because it was already well established as a technique for making discrete microwave transistors. However, control of the threshold uniformity is inadequate to the needs of VLSI. Doubts over the long term stability of gates formed

using the platinum sintering technique must also be raised because of the low activation energy of the sintering process.

The major difficulty of increasing the doping level in the extrinsic region is that the additional implant must be self-aligned with the gate electrode if its effect is to be optimised. The simplest method of self alignment is to use the gate as the mask during the extrinsic implant step [300]. This is fraught with some dangers however, and required the introduction of stable refractory gate metals, as described in Section 2.2.

Even having found a suitable material, the implant itself can cause problems. The transistor is most susceptible to avalanche breakdown in the high field region at the drain end of the gate, and the introduction of a high doping level at this point can give an abnormally low breakdown voltage [301]. There are two alternative measures which can be taken to prevent this occurring.

Firstly, instead of using a simple gate, a T-shape can be formed [302]. The highest field region occurs at the edge of the contact with the GaAs, i.e. where the gate is narrow. The highly doped extrinsic region is aligned with the broader top of the gate, and the two regions are therefore separated by the degree of overhang or undercut. This process suffers additional complexity in order to define the T-gate accurately.

The second of the two methods uses a conventional gate, but relies on the tailoring of the extrinsic implant to avoid breakdown. The extrinsic dopant is implanted with a lower peak concentration, but with the peak buried beneath the surface, much below the peak of the channel implant [303]. Because of the lateral scatter of the implant, this also has the effect of shortening the gate-length [304]. A shorter gate may be beneficial in improving performance, but has the penalty of increasing the variability of the threshold voltage, because much of the reported gate-length dependence of threshold voltage is caused specifically by this type of buried extrinsic implant [305].

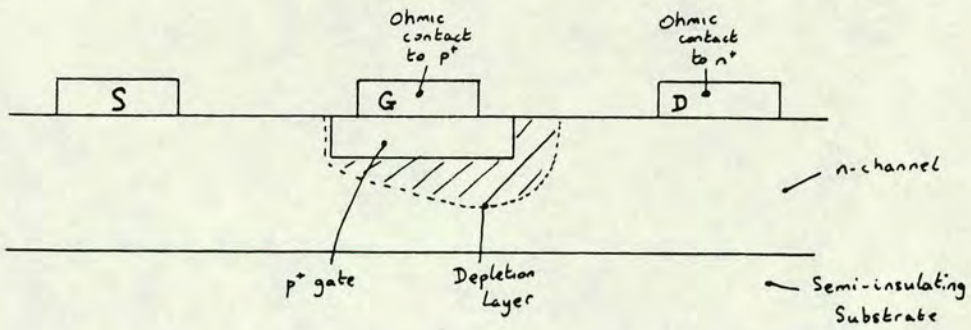


Figure 2.18 Cross-section through a JFET.

Another quite separate technique for self-aligning the extrinsic region to the gate relies on a substitutional technique [306]. This SAINT (Self-Aligned Implantation of N^+ Transistor) technique uses a multi-level dielectric to produce an effective T-mask for the implantation. A conventional gate electrode is later substituted for the dielectric following the anneal. Three steps are required for the gate deposition, and the process consequently suffers in a poor yield [307]. It was however the first technique to be successfully employed for a 1K RAM [308] in GaAs.

It is worthwhile comparing the JFET with the MESFET here. A cross-section through a typical JFET is shown in figure 2.18, and it is immediately apparent that the problem of high extrinsic resistance is diminished in just the same way as with the Pt-buried gate MESFET already described. The junction is buried beneath the surface by the thickness of the p-region, giving a much thicker extrinsic conduction channel. Even under the gate, the channel can be somewhat deeper because the built-in potential of the p-n junction is 1.1 V [309], compared with 0.8 V for the Schottky junction. The advantage of the larger built-in voltage is more apparent because of the additional 0.3 V allowed in the logic swing. This in turn gives an increase in noise immunity, and reduces the susceptibility to threshold voltage variations. There is some claim that such variations are minimised by the JFET structure because both donor and acceptor ions will respond similarly to variations in activation [310]. This argument should not be pressed too far however.

These advantages were all key factors leading to the early success of JFET normally-off circuits [311]. Nevertheless, with the improved control of MESFET uniformity, the more complex JFET process is less favoured.

The main problem is again a technological one. If the input signal to the JFET is end-fed, the relatively high resistivity of the p-region (compared to a metal) leads to a large time constant for the gate, which is modelled as a distributed RC network. In most cases, this time constant would be much larger than the required circuit switching times. It is therefore necessary to overlay the p-gate with a low resistance Ohmic contact, as shown in figure 2.18. The metal (or the contact window) must be printed within the p-region. The gate-length (defined by the size of the p-region) is therefore not the smallest patterned feature. Thus half-micron lithography is required for a 1 um JFET gate-length, where the equivalent MESFET would require only a one micron process. This is not a unique problem - it is found at the emitter of a bipolar transistor - and does not render the JFET totally impractical. In the case of the bipolar transistor, the extra process complexity is worthwhile because the bipolar device has unique properties [312]. The JFET, competing with the MESFET on almost equal terms, offers less justification for the adoption of the more complex process.

A subsidiary problem is encountered with JFETs: one of safety. The most useful acceptor impurity is beryllium. The extreme toxicity of its oxide has led to a great deal of resistance to its widespread adoption. Routine maintenance of the ion-implanter also becomes very difficult if it has been contaminated with beryllium. Such problems, whilst not technical cannot be dismissed.

A number of circuit developments has taken place alongside these advances in fabrication technology. An early development from DCFL used the so-called quasi-normally-off FETs in Low Pinchoff FET Logic or LPFL [313]. It was recognised that the (then) imprecise control of threshold voltage presented major problems, given the narrow range allowed within the conventional DCFL circuit. A family of circuits were devised in which the allowed range of threshold voltage was

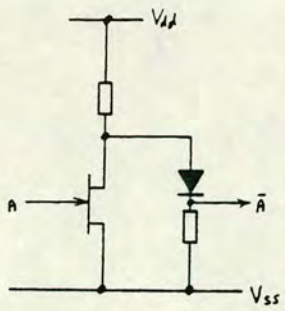


Figure 2.19 Low-pinchoff FET logic (LPFL) inverter.

increased thus allowing marginally normally-on transistors to be used (up to a threshold voltage of -0.2 V). One such circuit is reproduced in figure 2.19. It is clear that each logic gate requires a higher component count. However, a more complex logic function can be performed by each gate [314], in much the same way as in the normally-on logic families.

In fact, the origins of the different LPFL circuits can be seen to lie in either BFL or SDFL, with the number of level-shifting diodes reduced to one and the negative supply rail merged with the V_{ss} rail. Compared with DCFL, this type of logic suffers from relatively high power consumption, and it has now fallen from favour, in the light of the improved threshold voltage uniformity.

Another new circuit technique has been developed from the current-switched logic familiar in bipolar technologies. By analogy with Emitter Coupled Logic (ECL) [315], this circuit is described as Source Coupled FET Logic (SCFL) [316]. Like its ECL counterpart, SCFL can either be driven fully differentially or "single-ended", in the latter case by tying the second input to a reference voltage. It can operate with an even wider range of threshold voltages than LPFL by selecting the degree of level shifting introduced in the source follower. However, when used with a normally-on FET, the performance rapidly degrades with increasing threshold voltage, particularly when driven from a "single ended" signal source [317]. If normally-off transistors are used, SCFL has a number of advantages. In particular the transistor can be made to operate almost entirely in the saturation regime thereby minimising the gate-drain

capacitance [318]. At the same time, the voltage swing is easily defined by the available current and the load resistor values, rather than relying on the forward conduction of the gate to clamp the output of the previous stage. The problem of "overdriving" the gate, common with DCFL and discussed earlier, is thus eliminated.

In DCFL, the very small operating window renders NAND gates inoperable without a very severe penalty in transistor width (and thus in switching speed). As with SDFL, this renders the fastest of the toggle circuits impractical with DCFL. However, the "natural" divider circuit for SCFL is once again that using the very efficient AND-NOR combined gate. All these advantages have combined to give a rather impressive performance for an SCFL divider [319], operating at an input clock frequency of 11 GHz.

Chapter 3. Simple GaAs IC design using capacitor coupled logic

This Chapter is designed to take the reader onward from the background material of the previous chapter. The focus is placed on gaining understanding of the operation of capacitor coupled logic (CCL) circuits, in preparation for the detailed design work discussed in Chapter 5. In this Chapter, the emphasis is very much on the intuitive understanding of operation, supported where necessary by results of simulation and experiment.

The first Section sets the scene by describing the state of the work when the author took over the project, introducing the simple simulation models and the existing fabrication process, finishing with a description of the first test circuit designed prior to this work.

The second Section is devoted to building an understanding of the fundamental logic blocks, whilst the third section extends this to more complex and function specific elements.

3.1. Starting position

3.1.1. Device models

The early stages of GaAs IC development were frustrating in many ways. It was very tempting to join the bandwagon of computer aided design (CAD) [320] and assume that GaAs technology, like silicon, should be well modelled. Transistor level simulation was widespread, and simulators were becoming more common. One such simulator, ASTAP [321], was available for helping in this design work. At the beginning of this project work, ASTAP was probably the only simulator allowing user-defined device models to be incorporated, in this case by means of conventional FORTRAN subroutines. Differences in FET behaviour between GaAs and silicon were known, and the ability to define models suited to GaAs was thought paramount. However, out of expediency at the start of the programme, a MESFET model was developed based on simple JFET equations. The two gate junctions were represented by non-linear capacitances paralleled by diodes. Figure 3.1 shows the model components used to describe the

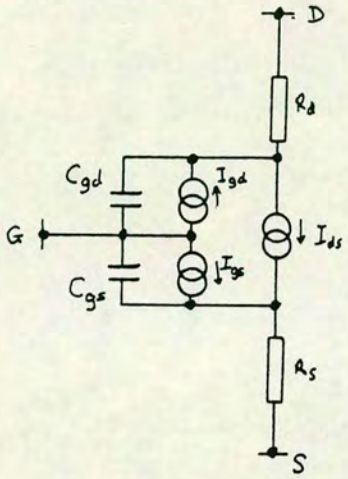


Figure 3.1 FET electrical model, (initial computer model excluded R_s and R_d).

transistor. The parameters used in this model were obtained from a small sample of discrete devices.

This model, together with the best fit parameters, was inherited at the outset of this research programme. It soon became apparent that the GaAs process produced widely different devices both across and between wafers. This model could not be used with any certainty to describe the transistor behaviour. Not only were the parameters derived from the early measurements unrepresentative of the whole process, but the model itself did not adequately represent the device behaviour. Any attempts to improve on the model were thwarted by the very wide and apparently uncontrollable spread on device parameters. Even links between related parameters (e.g. threshold voltage and saturation current) could not be established reliably.

Thus it was that representative models were not available for adequate computer simulations, yet the design tools were present. In an ideal world, the processing would have been improved and optimised until repeatable. Models would then have been developed and circuits built with certainty. In practice, and in a commercial environment,

external pressures dictated that work directed towards the production of working circuits had to be the first priority. This in turn indicated that the process development had to be performed without serious reliance on the design tools which had enabled the continued and rapid growth of silicon technology. (It must be said that no computer tools at all were available during the equivalent development phase of silicon devices). Only later in the development cycle would it be possible to employ CAD profitably.

However, despite the inadequacy of the models, considerable reliance was placed on them to indicate the feasibility of some circuit configuration. This was particularly important to capacitor coupled logic, where apparently functional logic circuits were rendered inoperable because of an unexpected stable state in which no capacitors were charged (this is discussed in more detail in section 3.2.3). Only following the charging of the capacitors could the appropriate logic function be guaranteed. Whilst paper exercises could be (and were) used to ensure some charging path, the confirmation of this behaviour using transistor-level simulations was reassuring. Even with the poor transistor models, some trend analysis was also possible and useful.

An approximate idea of device performance could be obtained by using these imperfect models and assuming a fixed error between simulation results and experiment. The earliest results indicated that the maximum operating speed of the circuits was about one quarter of that predicted with these most basic models.

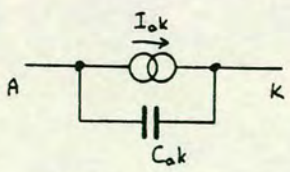


Figure 3.2 Diode model.

In the fabrication process used for this work, only two components, the MESFET and the Schottky diode were available. Figure 3.1 depicts the MESFET model elements, and 3.2 shows the diode model. Only three subroutines are used, as the diode model re-uses the subroutines employed for both the gate-source and gate-drain elements of the MESFET.

The diode current is adequately represented over the sensible operating range by the conventional exponential voltage dependence [322]:

$$I = I_o \cdot \exp\left(\frac{q.V}{kT}\right)$$

in which I_o is related to the dimensions and material properties of the device, q and k represent the normal physical constants and T is the junction temperature in Kelvin.

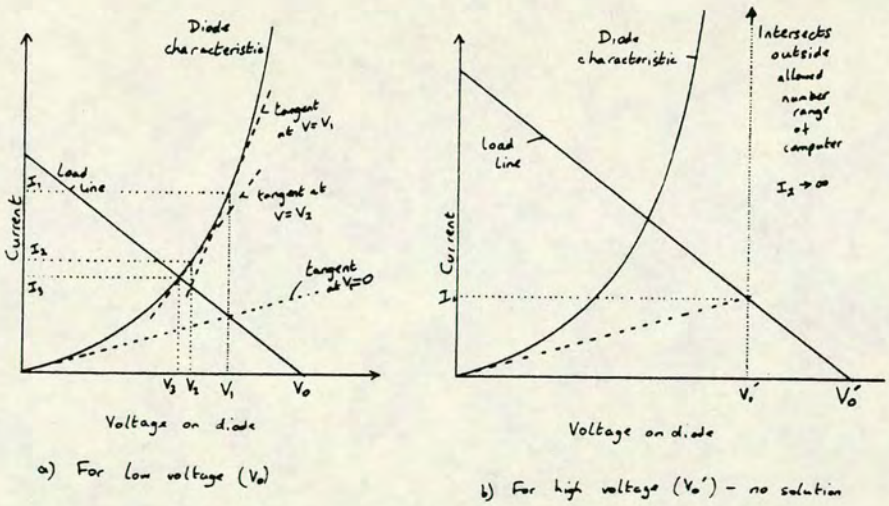


Figure 3.3 Schematic solution of the current flowing in a series combination of diode and resistor.

Problems are always encountered with the diode equation when used in numerical analysis programs, because the current can over-range very easily [323]. This is best illustrated by a graphical representation of the numerical method used to solve for the current flowing in a series combination of diode and resistor under a given applied bias voltage. Figure 3.3a shows the approach to solution when the voltage

is small and 3.3b shows how there is no solution when a large voltage is applied. To avoid this problem, a maximum slope is allowed on the forward current versus voltage curve. This model is actually quite realistic, as it is equivalent to assigning a series resistance to the diode. Although the facility is provided in the model for varying the value of this slope, no attempt was made to scale this with the diode shape, and it is thus only a numerical convenience rather than a real device parameter.

Under reverse bias, the diode is modeled as a resistor in parallel with a small fixed leakage current. Experimentally, the reverse leakage current may take an extremely wide range of values depending upon the exact processing. The voltage dependence of the leakage current is also unpredictable, and it is therefore impossible to model this component accurately. The influence of leakage, although not negligible in practice, may be ignored in simulations, as it is swamped by the displacement current in the parallel capacitor. However, to improve the DC stability of the simulations, a finite leakage should be used, and the linear expression is the most straightforward to implement, but there is no physical basis for assuming this form for the reverse leakage. As with the forward resistance, there is no attempt to scale the value according to the shape and size of device used.

The forward diode current is made temperature dependent, both through the conventional Boltzmann factor, and by making the pre-exponential term proportional to the square of the temperature [324]. As none of the other model elements incorporates temperature dependence, the inclusion of the temperature as a separate parameter here is somewhat misleading.

Parameters of this model are: area, length and width, temperature, applied voltage, forward current multiplier, ideality factor, forward series resistance and reverse leakage resistance. To enable the model to be applied consistently alongside the capacitor and FET current models, a flag is supplied to identify either a transistor or a diode. In the former case, the supplied value of area is ignored,

instead the area being computed from the length and width data. In the latter, the converse is true.

In a similar manner, the diode capacitance model relies heavily upon the conventional formula [325]:

$$C = \frac{C_o}{(V_{bi} - V - kT/q)}$$

in which V_{bi} is the built-in potential of the barrier and C_o may be derived from the appropriate physical constants and diode construction.

However, this equation must be modified to account for the finite channel thickness. When the reverse bias exceeds the threshold voltage of the material, the capacitance drops nearly to zero, with only the lateral spreading of the depletion region continuing to contribute. This is modelled by setting the diode area equal to the product of the channel thickness and the length of the periphery when the threshold voltage is exceeded.

The conventional formula is further modified for the case of ion implanted regions. A purely empirical formula is used to reduce the capacitance still further with increasing reverse bias by linearly reducing the doping level to zero at the threshold voltage. This compensates for the gradual fall-off in doping in an ion-implanted channel [326] (when compared with an epitaxial channel). This empirical capacitance formula has the added advantage for the ion implanted devices of not containing a discontinuity at the threshold voltage. If the ion-implant flag is not set, numerical stability problems are likely when the transistor is biased near the threshold. During the course of this work, the capacitor model has only been used to simulate ion implanted diodes, so this problem did not arise.

Under forward bias, the conventional formula for the depletion layer capacitance approaches a singularity at the built-in potential. To overcome this, the capacitance is made linearly dependent upon the voltage for all positive biases, with a continuous gradient at the

zero bias crossover. Although error in the calculated capacitance increases as the voltage approaches the built-in potential, the onset of current flow in the parallel ideal diode renders negligible any errors so-caused. The built-in potential should be temperature dependent, but this allowance was not made.

The parameters of the diode capacitor model are: area and periphery, width and length, channel thickness, threshold voltage, built-in potential and surface doping level. Two flags are also used, the ion implant flag as already described, and a second flag to allow either a transistor or a diode structure to be selected. The latter controls the area and periphery calculations in much the same way as its equivalent in the diode current model. However, when the transistor is being used, the area is set to half the actual area, and the periphery is set equal to the width. This assumes that the length is very much less than the width (always valid), and that half of the area is used for the gate-source diode and half for the gate-drain. This model is clearly inadequate and will be developed further in Chapter 4. The latter assumption is actually contradicted in use, when the gate-drain capacitor is set equal to one tenth of the computed value. This derives from studies of discrete transistors used for small signal analogue circuits which always have a large drain bias [327]. These do indeed show a gate-drain capacitance about one tenth of the gate-source value.

The final model element required is that for the drain-source current. This element is calculated according to three distinct conditions, defining the linear, saturation and cut-off regions. The value of current is dependent upon both the gate-source and drain-source voltages. In the cut-off region, defined by the gate-source voltage being less than the threshold voltage, the current is set to zero. The boundary between the linear and saturation regions is determined by the threshold voltage and the gate and drain biases. A saturation voltage is defined equal to the gate voltage minus the threshold voltage. When the drain voltage exceeds this saturation voltage, the transistor is operating in the saturation region, and the current is set proportional to the square of the saturation voltage. The constant of proportionality is a user defined parameter

representing the transistor gain. The current in the saturation regime can be further modified by a fixed output conductance. The current in the linear region is modelled by a parabola passing through the origin and with its peak at the saturation current and voltage.

There are a number of discrepancies between this model and a real device, but as already explained, the "typical device" did not exist at this time, and this model represented an ideal device. The saturation current is that predicted from simple theory for a JFET, in which a constant mobility model is used. The current in the linear region is a simplification of that given by the same model. The simplification avoids excessive use of non-integer power laws, and therefore allows more rapid computation.

The parameters of the drain current model are: gate and drain voltage, width and length, threshold voltage, gain factor and channel modulation factor (related to output conductance). Again the parameters were not temperature dependent.

3.1.2. The Fabrication Process.

The fabrication process in use at the beginning of this study was very unsophisticated. Standard circular wafers were used, although in many parts of the industry, scribed squares of GaAs were still in use. By silicon standards, the 3 cm wafers were very small, but two inch wafers were beginning to be available at this time and these promised a larger area of usable material. Ion implantation was used for doping, but this was not performed selectively. A CVD silicon nitride cap was deposited on both sides of the wafer prior to the anneal, which was performed on a graphite heated susceptor. Following the implant and anneal, the device areas were delineated by mesa etching, with the top half micron or so of GaAs being etched away between the doped regions. This simple doping procedure did not allow diodes and transistors to be fabricated on separate materials, nor did it allow the contact regions to be more heavily doped than the channel regions.

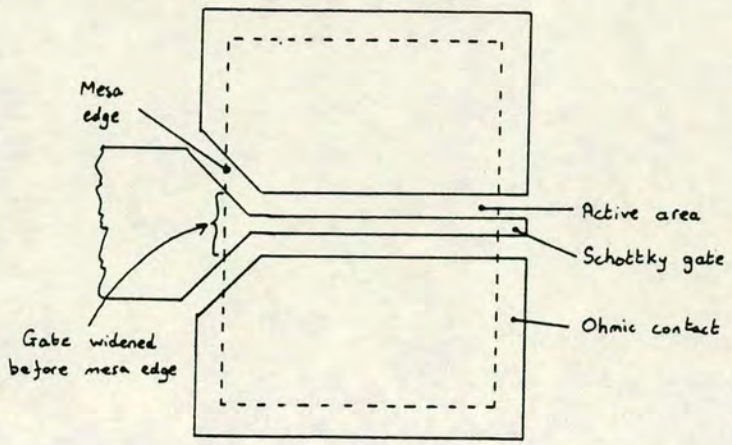


Figure 3.4 Design of tapered gate transistor.

A further consequence of the mesa isolation process is the problem of ensuring that the metal gives adequate coverage on the step between the isolated and active regions. This problem is most acute for the narrow gate tracks, only 1-1.5 μm wide. To overcome this problem, the gate tracks were tapered as shown in figure 3.4, such that the typical width of the track where it crossed the step was some 3 μm or so. A similar taper is necessary on the source and drain edges, to maintain the process tolerance.

The Ohmic contact used in this process was the conventional alloy of gold, germanium and nickel, but this was not fully overlayed by the Schottky metal, as the potential corrosion hazard had not been recognised. The Schottky metal was formed from a two step evaporation of titanium and gold. Polyimide was used for an interlayer dielectric, before a further layer of identical metal was used to form the second interconnection.

The minimum printed dimension in this process was 1 μm, used solely for the patterning of the gate metal. Although this is smaller than used in most silicon technologies, certainly at the time, the remaining dimensions of the process are quite coarse, with most other layers printed at around 5 or 6 μm. The conventional measure of the process, the width of and separation between metal tracks being quite coarse at 6 and 8 μm respectively. The 1 μm patterns were printed

using a conventional contact lithography and the lift-off technique was employed for defining the metal tracks on both layers.

In order to pattern the small gate features accurately, without problems of wing formation, the thickness of the first level of metal was restricted to about 0.25 μm, with the second layer of metal being double this value. The thickness of the polyimide interlayer dielectric was typically 1 μm.

During the course of this work, some changes to this basic process were found necessary. Details of these changes are reserved for Section 3.2.4, when the reasons for the alterations can be justified against the desired improvements of circuit operation.

3.1.3. A simple test chip.

A small test chip containing a frequency divider as the most complex element was designed to prove the principle of capacitor coupled logic. In addition to the divider circuit, ring oscillators were included, together with individual logic gates - 2 and 3 input NOR and NAND gates and bistable latches. At the beginning of this study the test chip had been designed but no wafers had been processed.

The test chip was split into quadrants, each exactly 1 mm square, with 16 bonding pads for each quadrant. Because of the simplicity of the test structures, even using the four different units, the chip size was bond-pad limited. Thus many of the structures had long tracks connecting them to the pads. These caused some problems because no allowance had been made for the voltage drop in these tracks. In particular, the power tracks should have been substantially increased in size.

One of the quadrants contained test structures to examine the feasibility of using capacitor coupled logic for four-phase dynamic logic. These features were never studied, because of the practical difficulties of implementing multi-phase dynamic circuits at high speed.

The remaining three quadrants were very similar. In one, a divider and bistable latch were both driven from the same external clock signal. The other two also contained the divider and latch, but the clock signals were derived from ring oscillators of different lengths, the oscillation signal also being available as an output. Two useful purposes were served by this combination. Firstly, correct operation of the divider shows that the internal signal swing of the oscillator is close to saturation and large enough to drive the logic which follows. Secondly, the divider is driven at relatively high speed, eliminating the difficulty of introducing a clean clock signal onto the chip. In the case of the latch, only one external signal (either the data or the clock) was required, instead of two.

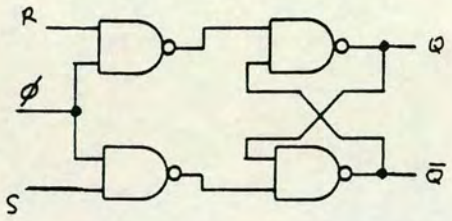


Figure 3.5 Basic bistable circuit.

The 4-NAND gate circuit of figure 3.5 was chosen for the bistable circuit. In its basic form, this circuit requires both a true and complementary signal on both the clock and data signals. In both cases, the complementary input was derived from the true signal using a single inverter stage. Two of these bistables were assembled to form the divider (or T-latch), by feeding the first stage into the second and the inverted output of the second back to the first, in the conventional manner.

In order to ensure correct operation of the ring oscillators, a NAND gate was incorporated into the ring (see Section 3.2). Because of a complete uncertainty of the viability of the fabrication process, and

especially of the associated yield, the number of stages in the ring was kept as low as practical. With a ring shorter than seven stages, saturated logic swings could not be guaranteed, so a seven stage ring was chosen. Because the NAND gate represents a larger than unity value of both fan-out and fan-in, and because one other stage is loaded by the output buffer, the gate delay can not be extracted from such a short ring with any accuracy. Thus a second ring, with fifteen stages, was patterned. The difference in oscillation period of the two rings may be used to derive the propagation delay per gate eliminating all the loading effects.

The incorporation of the NAND gate into the ring oscillator loop also eases the testing. A sampling oscilloscope is required to measure these structures because of the high oscillation frequency anticipated. Obtaining a clean display using a sampling 'scope can be problematic, because of the difficulty of triggering. However, by applying a relatively slow square wave to the second input of the NAND gate, the oscillator is gated on and off, with the oscillation being synchronised to the gating signal. By also triggering the oscilloscope from the gating signal, the measurement integrity and simplicity are ensured.

It was apparent that GaAs circuits would have to show input and output (I/O) compatibility with existing high speed logic elements in order to achieve acceptance. In practice this implied compatibility with emitter coupled logic (ECL), in which only a small logic swing of about 0.8 V is used, centred about -1.3 V [328]. The GaAs internal logic swing is much larger than this, and some amplification at the input is clearly required, together with some reduction in swing at the output.

An amplifier had been designed using the models described in section 3.1.1, in which a small hysteresis had been included to improve the input noise margin. This circuit was also included in the test chip for evaluation.

Fewer problems were anticipated at the output. A source follower capable of carrying a large current was used, but with no load

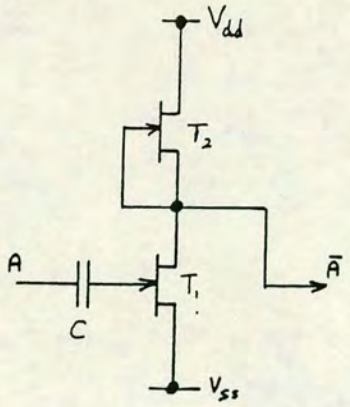


Figure 3.6 CCL Inverter.

resistor on the chip. By connecting an external load to a suitable voltage, some degree of compatibility with ECL logic levels was anticipated. This required the main circuit to operate with the drain supply grounded and a negative supply voltage for the source. In order to drive the required output current, a large (150 um wide) transistor was used. To prevent an excessive loading effect on the main circuit, this transistor was itself driven by a separate buffer stage consisting of an inverter. This two stage output buffer was incorporated into all the main test elements, the D and T latches and the ring oscillators.

However, use of the two-stage buffer on the outputs of the individual gates was inappropriate, as the buffer, designed as a logic buffer rather than as a linear amplifier, would have completely masked their behaviour. Instead, these gates were designed using the normal size of transistor (about 10 um wide), but with a small (only 50 um^2) output pad positioned as close to the transistors as possible. This pad was deliberately chosen to be small, to minimise the capacitive loading effect on the logic gate. It was intended that a manually manipulated probe needle attached to a high impedance oscilloscope probe would be used to measure the performance of these gates. Similar pads were used for the inputs to these gates, and again the signals were to be injected via a manually manipulated probe. This design was adopted to avoid degradation of the input signals.

This test chip was suitable for very simple characterisation, and to prove the principles of CCL, but it was soon apparent that many improvements were required, and that a testing methodology needed to be thought through as an integral part of the chip design. More details of the results will be presented in Section 3.2.4.

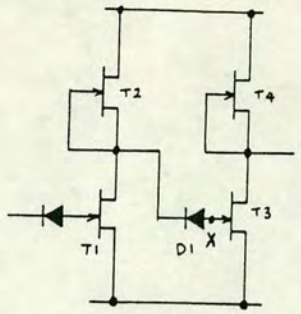
3.2. Capacitor Coupled Logic

Capacitor coupled logic (CCL) is a digital IC technique developed by the GaAs IC research team at British Telecom just before the commencement of this research effort [329]. The basic principal of CCL is quite straight forward, and simple demonstrations of its feasibility require little or no understanding of the differences between CCL and more conventional logic techniques. The test chip already discussed was designed on this basis. However, to put CCL to work at a more complex level, a much more detailed understanding of circuit behaviour is needed. This understanding, which forms the basis of this section, has been developed throughout this programme of work. Although the development in modelling and chip design are best presented in a chronological fashion, it seems appropriate to group together the detailed knowledge on gate behaviour at the outset, even though this understanding has been built up throughout the study.

3.2.1. The Inverter

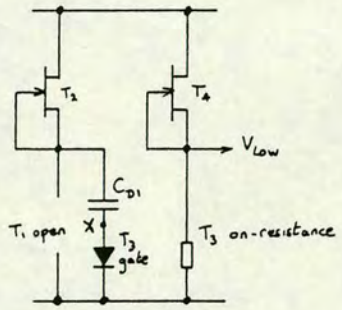
As in all logic systems, the fundamental building block is the inverter. In CCL, the inverter comprises three components, the switch transistor (T1), the load device (T2) and the coupling capacitor (C), as shown in figure 3.6. Although not essentially so, it is a beauty of the technology that the load device may be a transistor identical in property to the switch transistor. The two are thus conveniently made with the same process steps. The capacitor presents a somewhat greater problem for co-integration.

For this particular application, the DC bias on the capacitor is unidirectional, and neither the exact value of capacitance nor its linearity are critical factors. It does however need to have a reasonably high Q [330]. Of the possible capacitor technologies, the planar dielectric capacitor is unattractive because of potential



a)

b)



c)

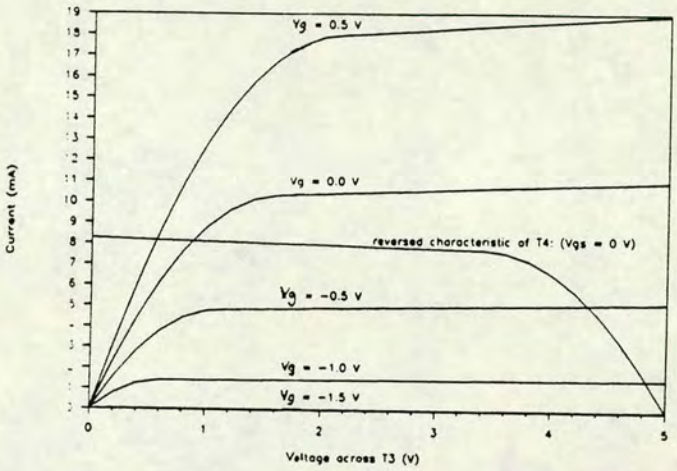


Figure 3.7 Determining operating points of an inverter:
a) Circuit diagram
b) equivalent diagram with T1 off
c) adding a load line to FET output characteristics.

yield problems given the requirement for large numbers of capacitors. The interdigitated dielectric capacitor is generally physically large requiring very thick metal [331]. This leaves the junction capacitor as the most suitable choice. In fact this is an excellent choice, because in GaAs, the most suitable junction is the Schottky diode, and such a device mates well with the surrounding circuit elements. In the CCL circuit, the anode of the diode connects to the transistor gate (both Schottky metal), and the cathode to the drain of the previous transistor (both Ohmic metal). At first sight, the capacitor can thus be integrated with the transistors with ease, although their respective sizes need to be established to confirm this.

To study the operation of the inverter it is more convenient to use two cascaded stages as shown in figure 3.7a. Assuming that the logic is saturating, it is possible to switch-off transistor T1 and figure 3.7b shows a part of the circuit under these conditions. In particular, the drain of T1 appears to be an open circuit, and the

gate of T3 is replaced by a diode. Diode D1 is shown as a capacitor. This capacitor will charge-up, initially with a constant current set by the saturation current of T2. During this charging cycle, the input to T3 is clamped at one diode voltage drop (V_{bi}) above the V_{ss} rail. The circuit will eventually equilibrate, with the output of the first inverter at V_{dd}. The voltage on the internal node 'X' will be somewhat less than V_{bi}, as the diode current will be near zero. However, a worst case calculation assumes that the capacitor is charged to a voltage given by (V_{dd} - V_{bi}).

The output voltage of the second inverter can best be obtained graphically from figure 3.7c. Here the normal transfer characteristic of T3 is drawn and a load line [332] is added by drawing the transfer characteristic of T4 with its origin at V = V_{dd}, and with increasing voltage drawn to the left. The intersection of the two curves define both the output voltage and the current drawn through the inverter. As in NMOS [333], the quiescent operating point is dependent upon the relative size of the transistors T3 and T4, and on the output conductance.

When capacitor D1 begins to charge, there is a large charging current which applies a high gate voltage to T3. As the charging current drops, the forward bias on T3 gate diminishes. Thus, throughout the charging cycle the output voltage of the second inverter will rise from V_1 towards V_2 shown in figure 3.8. The actual operating point is therefore dependent upon the recent history of the logic gate, as well as on the process and design parameters of the transistors.

In order to make the circuit operational over a wide range of these variables, it is desirable to increase the width ratio of T3 and T4 to a higher value than would be used in NMOS, thereby minimising the spread in output voltage during the charging cycle. A somewhat arbitrary (but realistic) specification of 1 V was chosen for the worst case output low.

If the voltage levels of the two inverter stages are compared, a specification for the capacitor can now be defined. When the input to T1 has changed to the logic one state, the cathode of D1 will now

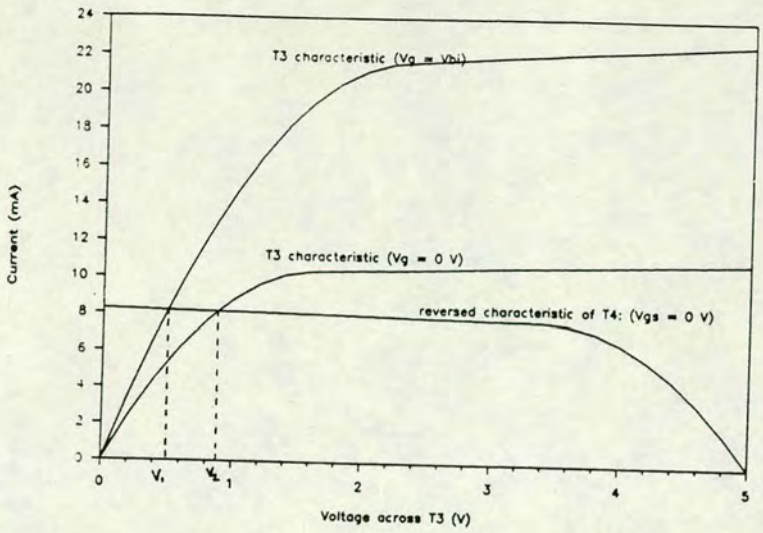


Figure 3.8 Shift in logic-0 as capacitor charges up; V_{ol} moves from V1 to V2.

be at 1 V, and the voltage change at this node between the two states will be $(V_{dd}-1)$. In order to turn T3 completely off, the voltage at X must change from $+V_{bi}$ to V_t, the threshold voltage of the transistor. A target threshold voltage of -1.5 V was initially chosen, although this implies a likely range of -1 to -2.5 V given the variability of implant activation in GaAs especially in its early days. With a 5 V supply, and allowing for the worst case threshold and output levels, the capacitor therefore must be 80% efficient, i.e. a swing of 4 V at the capacitor cathode must translate into a minimum corresponding swing of 3.2 V at the anode.

Figure 3.9 indicates why the coupling capacitor is not 100% efficient in transferring a voltage swing at the input into an equivalent swing at node X. When the input is held close to V_{dd}, the gate of T3 is conducting and almost no charge is stored on the T3 gate-source capacitor (C_{gs3}). Instead the charge is all mobile, i.e. a current. When the input is brought low, the charge which was stored on D1 is now shared between D1 and C_{gs3}'. This charge sharing is only achieved by reducing the voltage across D1. Using the equivalent circuit model of figure 3.9, it is apparent that the voltage changes can be calculated from a knowledge of the areas, and hence the capacitances, of D1 and T3. However, this calculation is complicated by the non-linear nature of the C-V characteristic of both capacitors. It is also necessary to reconcile the apparently "infinite" gate

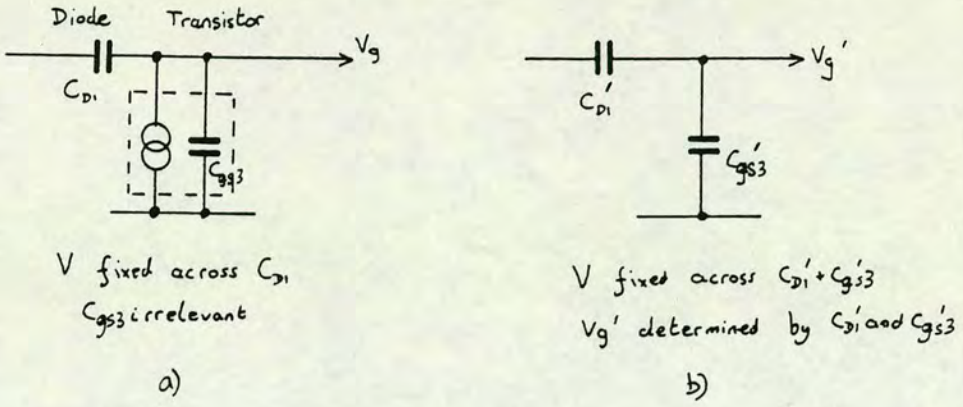


Figure 3.9 Model to indicate loss of coupling efficiency: a) input-high state, b) input-low state.

capacitance of T3 when under full forward bias, even though it effectively carries no voltage.

A much more useful understanding is obtained by thinking about the physical models of the diode and transistor rather than the electrical ones. Ignoring the initial charging cycle and the displacement current which flows during a transition, no gate current actually flows, and T3 gate switches from threshold to the onset of gate conduction. Thus, in the input high state, T3 has virtually no depletion layer whilst D1 has a large depletion region, bearing nearly the full supply voltage. Although the very shallow depletion layer of T3 makes a calculation of the capacitance somewhat fraught, the calculation of its stored charge is quite straightforward. After switching, sufficient charge has been transferred into T3 to deplete the layer fully. This charge must have been removed from the depletion layer in D1, as no current has been allowed to flow through the reverse biased junctions.

The required area ratio of D1:T3 can be calculated from the applied voltage and the worst case boundary condition, namely that T3 changes from having no depletion layer to being fully cut-off (i.e having a depletion layer thickness equal to the channel thickness). A value for the area ratio of around 3:1 is adequate for the data specified, but after allowing for additional stray and metalisation capacitance

a value of 5:1 offers a more suitable margin. In the case of uniformly doped channels, the charge in the depletion layer is readily calculated from [334]:

$$Q_{tot} = q.N_d.a.z.w = 2.\epsilon\epsilon_o.V_p.z.w/a$$

where N_d is the doping level in the channel, $\epsilon\epsilon_o$ is the permittivity of the semiconductor, V_p is the pinchoff voltage defined in Chapter 2, and a, z and w represent the channel thickness, length and width respectively.

Using this picture, even the non-uniform doping of the devices can be readily accounted for by integrating the charge in the depletion layer between the two limits.

If the electrical model is used, and a full circuit is drawn, there is some confusion as to the manner of treating the gate-drain capacitance. However, in this physical model, there is no quandary. The equivalent problem is to know how to treat the shape of the depletion layer of the transistor. Predicting the shape during the transition may cause problems (see figure 3.10a), but fortunately only the two end-points are of interest. In one state, the channel is fully open, with only a small drain voltage. This produces minor distortion of the depletion region as shown in figure 3.10b. In the other state, the channel is fully cut-off and no debiasing current flows under the gate, and the channel is thus depleted as in figure 3.10c. Both of these states are adequately modeled assuming a depletion layer parallel to the channel.

In a treatise on conventional logic elements, a transfer function for the switch could now be defined. The transfer function is produced by ignoring the "digital" nature of the circuit, instead treating it as an analogue circuit [335]. The voltage level at the output is plotted as a function of the DC voltage at the input, as shown in figure 3.11a. By reversing the role of the input and output axes in this graph, and again plotting the transfer function, as shown in figure 3.11b three common points (X, Y and Z) can be identified. If a host of identical circuits are now cascaded, such that the output

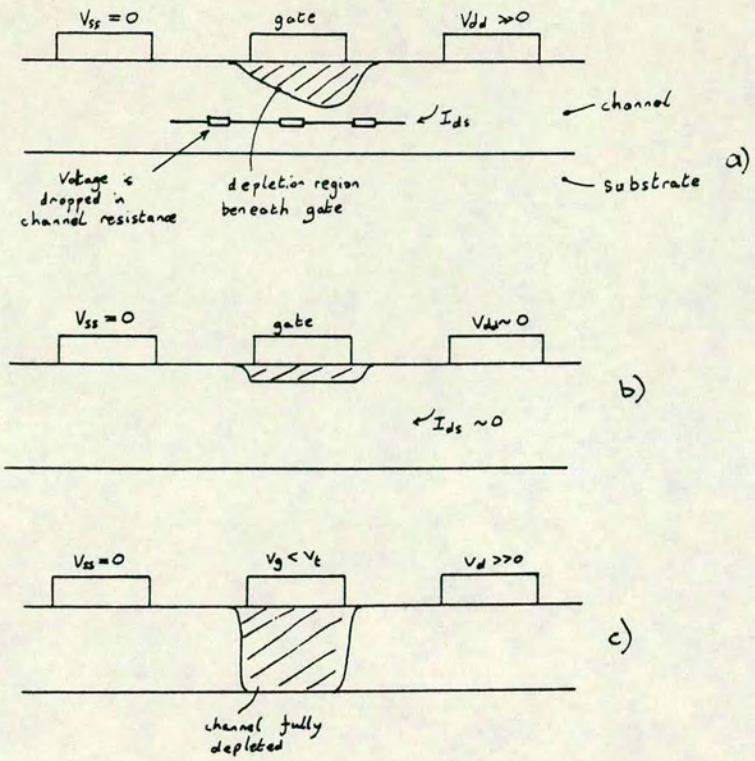


Figure 3.10 Depletion in a MESFET:

a) arbitrary condition - field increases depletion thickness towards the drain

b) small drain voltage - depletion thickness nearly constant

c) No current flowing i.e. ($V_g < V_t$) - channel is fully depleted.

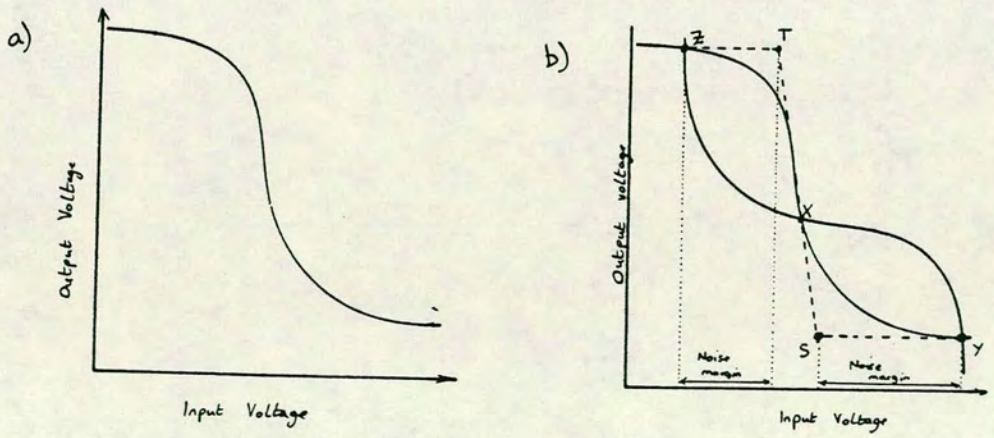


Figure 3.11 Transfer characteristics of an inverter:
a) single gate, b) multiple cascaded gates (showing operating points and noise margins).

of the first forms the input to the second etc. these three points represent the possible voltages at the output of the final stage. If the input to the first stage is at voltage X, then the output is also X. This is the unstable position, because any slight deviation from X is amplified in each stage, until the output has shifted to either of the points Y and Z after a few stages. Because of this, Y and Z represent the logic levels of the circuit. If the input voltage exceeds S or is less than T, only a single stage is required to force the output to one of the logic levels. The voltages (Y-S) and (T-Z) are defined as the noise margins at the input [336].

In CCL, such a definition of transfer function and noise margins is not sensible. Any DC voltage lying between the two logic levels applied to the input produces a fixed output voltage, somewhere between 0 and 1 V, which corresponds to the logic zero output. The nearest equivalent to a DC transfer function is obtained by plotting the output voltage versus the input voltage for a fixed value of charge on the coupling capacitor (this is equivalent to fixing the voltage across the capacitor). As the chosen amount of charge is varied, a family of characteristics is produced, as shown in figure 3.12. For any input condition, the output voltage can be predicted knowing both the input voltage and the voltage across the capacitor. However, during a real (dynamic) switching transient, the voltage across the capacitor will change as its charge is transferred into the transistor depletion layer. Figure 3.13 shows the locus of the dynamic transient superimposed on the characteristics of figure 3.12. The type of curve shown in figure 3.12 is therefore not helpful, particularly as it cannot be measured for a real logic element.

A much more helpful definition is required. An alternative (dynamic) transfer function can be defined in terms of the output level immediately after a defined change. Practically, the input can be held at a fixed positive voltage, and then a falling edge can be applied to the input, the output level immediately after this change being monitored. This then gives rise to a new family of transfer characteristics, one curve for each of the possible pre-charge conditions.

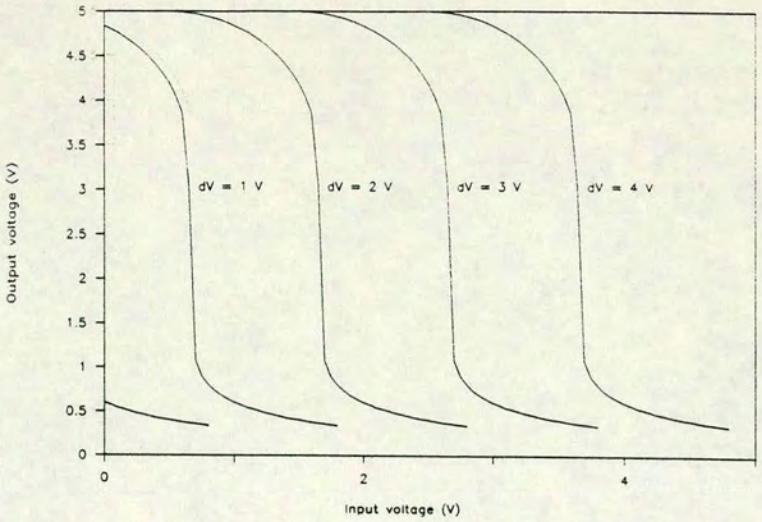


Figure 3.12 Transfer characteristics of a CCL inverter with a fixed voltage level shift.

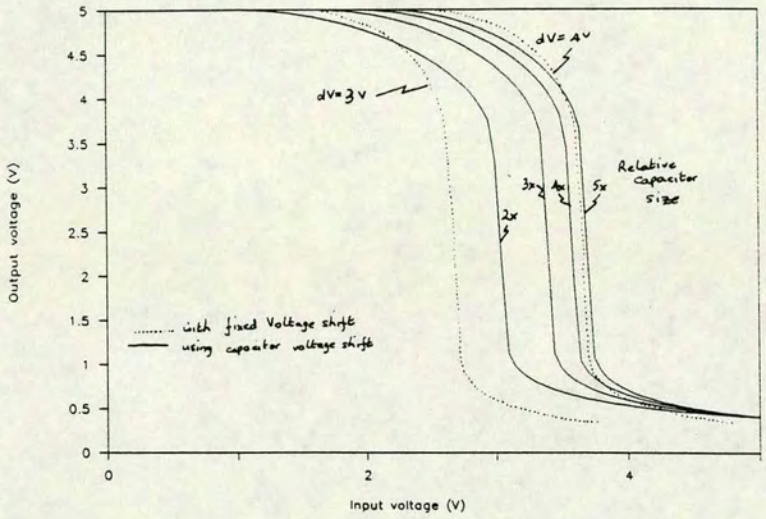


Figure 3.13 Transfer characteristics of a CCL inverter for a logic-1 to logic-0 input transition for different sizes of capacitor.

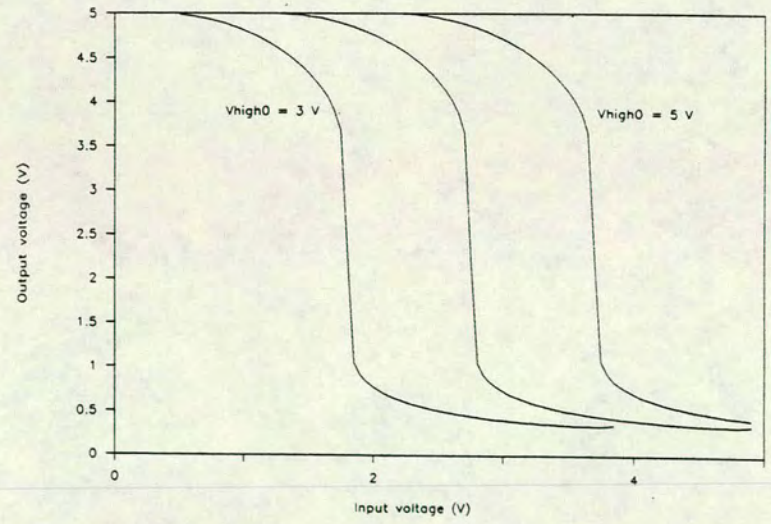


Figure 3.14 Transfer characteristics of a CCL inverter for a logic-0 to logic-1 input transition:- input voltage prior to logic-1 to logic-0 transition as parameter.

One such set of curves is shown for an inverter in figure 3.14, with the input voltage to the diode prior to the application of the switching edge as the parameter. These curves were produced using the models presented in Section 3.1.1. During a real logic one to zero switching transient, the output voltage of the logic gate should then follow one of these characteristic curves, depending upon the starting voltage.

Strictly, such a transfer characteristic is only applicable for the falling edge transition. In this case, the positive voltage prior to the transient forces the capacitor into a known state of charge. However, on the rising edge transition, there is no easily defined state of pre-charge, as the internal node voltage can neither be measured nor forced into an exactly known state. The leakage currents in the circuit are acting to reduce the charge on the capacitor, so this voltage is a function of the elapsed time since the last downward transition. Following an upward transient, the state of the circuit is therefore a function of both this rising edge and the previous falling edge, as well as their separation in time. In order to plot a dynamic transfer characteristic for the upwards transient it is therefore necessary to specify both the voltage at the input prior to the transient and the total charge stored on the capacitor prior to the transient (the latter is equivalent to specifying the voltage at the internal node). Even if it were possible to represent such a family of transfer characteristics in any sensible pictorial fashion, given two independent parameters, it would be of little practical help, because only one of these parameters is accessible experimentally. We must therefore conclude that there is no general definition of transfer characteristic for the rising edge input.

It is possible to define a special case which may be of importance, and this allows at least some comparison between the transfer functions on the rising and falling edges. The special case is obtained by assuming that the input to the logic element prior to the application of the rising edge is at the low logic level. This is a perfectly reasonable assumption, as any DC positive input applied to a logic gate driving the one under test will result in a logic low at

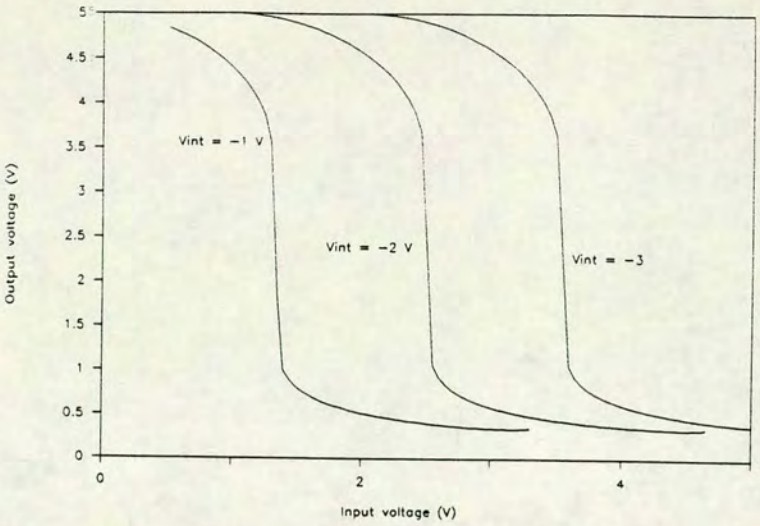


Figure 3.15 Transfer characteristics of a CCL inverter for a logic-0 to logic-1 input transition:- intermediate node voltage V_x as parameter.

the input in question. Having made this assumption, a family of curves can be produced (figure 3.15) with a single distinguishing parameter, namely the voltage at the intermediate node prior to switching. Although this family of transfer characteristics cannot be verified experimentally, the form of the simulated curves can be compared with those for the falling edge (figure 3.14). Over a wide range, the two sets are sufficiently similar to apply the curves for the falling edge to all transitions. The theoretical curves obtained for the rising edge can be used to identify the limiting bounds, outside which the logic performance is impaired, and eventually inhibited.

Although obtained dynamically, it should be stressed that the transfer characteristics of figure 3.14 should be treated in like manner to the conventional DC transfer characteristics of an inverter. No high frequency or phase effects, such as transit time, have been taken into account [337].

3.2.2. Other simple logic gates.

This section will deal with the extension of these principles to both NAND and NOR gates, as well as to the buffered inverter.

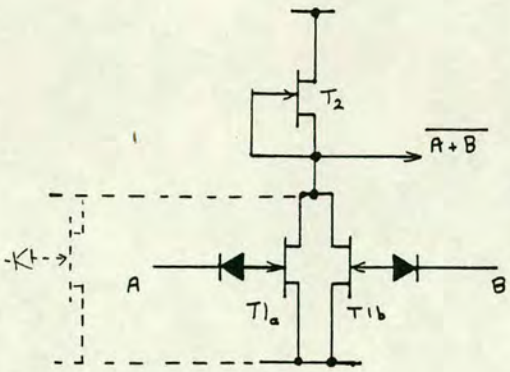


Figure 3.16 Two input CCL NOR gate.

Converting the inverter into a NOR gate is probably the easiest step to take: an extra transistor in parallel with T1 performs the necessary function. Using a positive logic notation [338], the switching of either or both of these transistors on (input = 1) produces a low (0) at the output. To complete the logic element, this second input must also be coupled with a capacitor. Figure 3.16 shows the completed circuit of the 2-input NOR gate. Extension to more than two inputs is clearly possible as shown by the dotted elements, and follows exactly the same route.

The rules governing the capacitor and transistor sizes are similar to those already given for the inverter. Turning firstly to the NOR gate circuit in the logic-1 state. In contrast to the inverter, in which transistor T1 may be allowed to be partially conducting, it is essential to demand that each of the inputs is fully cut off in the input low state, as the effect of any residual drain current in T1 is multiplied by the number of inputs (the fan-in). In practice, and for convenience, the same demand will be placed on the transistors in the inverter, resulting in a more rigid specification than that already expressed.

To meet the output low condition, the requirements for the NOR gate are identical to those for the inverter because the NOR gate must be capable of switching from the high to low state in response to just a single active input. In this special case where the input pattern to the NOR gate has a single logic one with the remaining inputs all

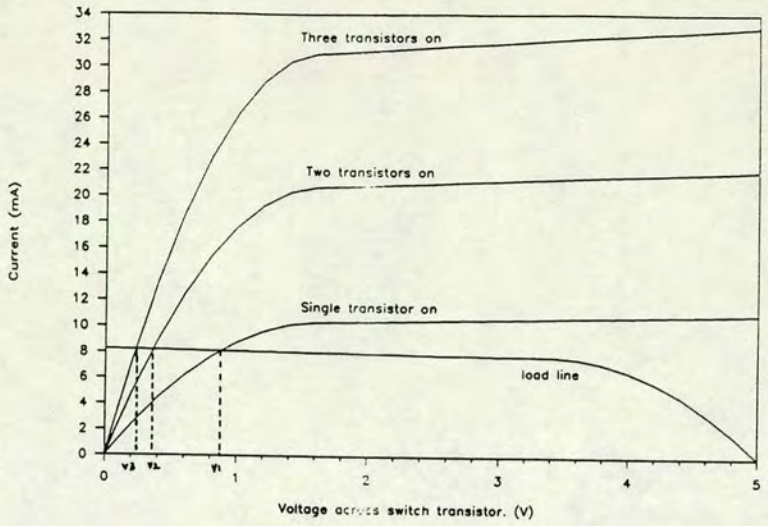


Figure 3.17 Dependence of V_{ol} on input states for NOR gates.

zero, the gate can be modeled by an inverter onto the output of which has been added an additional capacitor representing the open circuited transistors of the unused inputs.

When all the inputs are at logic one, the output of the gate will be closer to 0 V than that for an inverter, and there will necessarily therefore be some pattern ripple on the output low. Because any succeeding logic stage is fully switched off in this state, none of this ripple is transmitted for more than one stage. Figure 3.17 shows the three different voltages which represent a low output from a three input NOR gate. To produce this figure, the transfer characteristic of the transistor T1 is scaled by 1,2 or 3, (depending how many such transistors are switched on) whilst the load line remains constant. It can be seen that the resulting ripple voltage is small.

The main limitation to the fan-in of a NOR gate is a practical one, determined by the layout. Whilst the transistors T1a,b, can be widely separated, therefore allowing a very large number, they should all be placed as close to the load device as possible. The track connecting the drain of T1 to the source of T2 has to carry a DC current, whereas the track linking this same node to the input of the next logic gate only has to carry the switching current, generally very much smaller. If the transistors T1 are widely dispersed from

T2, the different inputs can be severely affected by voltage drops causing different effective logic thresholds. Particularly in high speed logic, the impedance of the return current path should also be considered.

Perhaps a more minor limitation should be considered; each additional input represents a slowing down of the logic gate, because of the load source-drain capacitance of each transistor. To date, an arbitrary limit of four has been set, although with experimentation, this could probably be extended to eight.

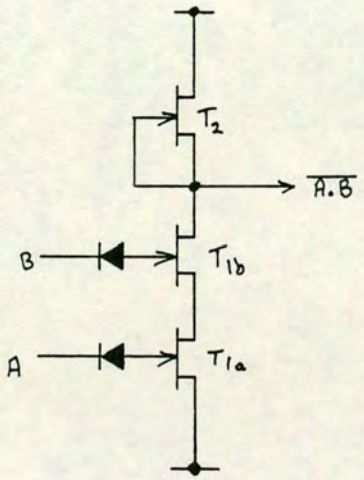


Figure 3.18 CCL NAND gate.

The other natural development from the inverter is the NAND gate. Again the development is superficially simple, but there are possible pitfalls. Figure 3.18 shows the least complicated extension, in which the two inputs are placed in series rather than in parallel. Quite clearly, in order for current to flow in this circuit, both of the transistors must be switched on, i.e. both inputs must be high. A capacitor is again required on each input to perform the level-shifting.

The DC logic levels can be derived in the same way as before. Achieving the output high again demands that the input transistor is (nearly) fully turned off, but this time only one of the two must be off. For the lower input to be fully turned off, its input must fall to the threshold voltage. The condition on the upper transistor is then irrelevant. If, however, the lower transistor is on, the source

of the upper transistor, T1b, will lie at (say) 0.5 V whilst it is switched on. In order for it to turn off, therefore, the gate apparently only needs to be reduced to (V_t + 0.5 V) and the upper input appears to switch at a larger voltage than that required to switch the lower. This is a misleading picture. The source voltage also falls during the transition, because the current through T1a is reducing. In fact, whilst the input to T1a remains constant, T1a can be replaced by a resistor in the source of T1b. This acts as a feedback element, reducing the effective transconductance of the transistor. On the conventional logic transfer curve this is manifest as a reduced gain, giving smaller noise margins (see figure 3.11). The DC requirement at the upper input is thus the same as at the lower, namely that the input must fall to the threshold voltage.

If we also demand from the NAND gate the same logic zero output voltage as from the inverter, it immediately becomes apparent that the transistor sizes for T1 must be changed if the size of T2 remains fixed. The capacitor coupling allows the gate source voltage of each switch transistor to find its "natural" level. Thus both T1a and T1b will have equal input voltages despite the higher source voltage of T1b. The impedances of the two transistors are therefore identical. A combined transistor T1′ can be used to represent the joint effect of the two series transistors. The output curve of T1′ is identical to that for each of the constituents, except that the axes have been scaled appropriately as shown in figure 3.19. Thus, for a similar load current, the switch transistor of the NAND gate must be wider than that of the inverter or NOR gate.

The definition of a suitable logic transfer characteristic for the inverter presented some rather difficult conceptual problems when dealing with capacitor coupled logic. In the case of the NAND gate, these problems are heavily compounded, and are not even considered. It is worth pursuing some qualitative description of the behaviour however.

As a first consideration, assume that the lower input is switched on, and remains so. As already intimated, T1a then acts as a resistor

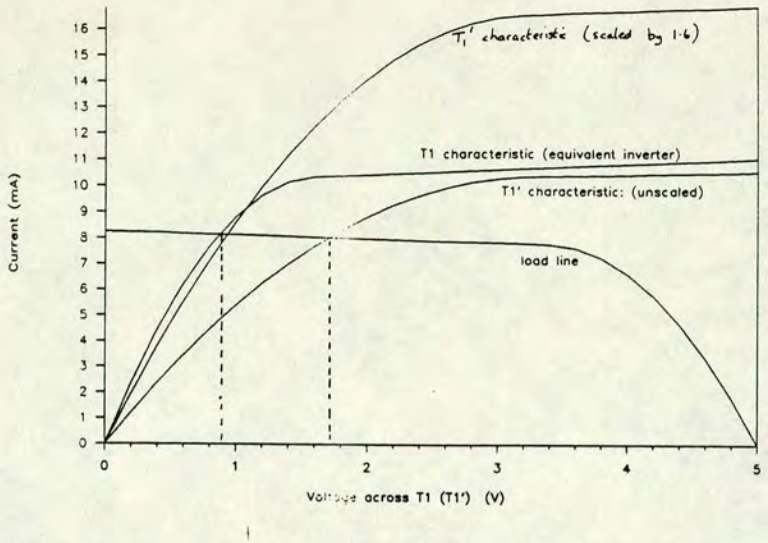


Figure 3.19 Shift in NAND gate V_{ol} for scaled and unscaled transistors.

which is nearly Ohmic because it remains well within the linear operating region being very much wider than T2. The circuit then responds to changes in the upper input, exactly as if the circuit were an inverter, but with a reduced effective transconductance for T1b.

Turning to the opposite case, where the upper input remains on, and the lower input is switched. The circuit behaviour is that of a somewhat modified cascode amplifier. As T1a turns off, the source of T1b rises, but the gate does not, as it is held by the charge on the capacitor. T1b therefore also switches off, improving the switching characteristic. When T1a is now switched on, T1b is again switched, because its source voltage falls. Note that any charge lost from D1b must flow through T1a as well as through the gate of T1b, and the charging time constant for this will initially be high, whilst T1a is only partially on. The turn-on may thus be worsened by this action, depending upon the charge lost from D1b. More seriously, the switching behaviour is again related to the circuit history. An additional diode, as shown in figure 3.20, will cure this problem, but the additional input capacitance offsets any advantage, and it has proved unnecessary.

It thus appears that the switching behaviour of the two inputs to the NAND gate is dissimilar, with the rising and falling edges being affected in a different way. Allowance must be made for this

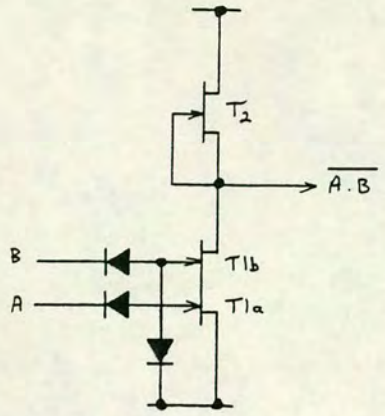


Figure 3.20 CCL NAND gate with guaranteed charging paths.

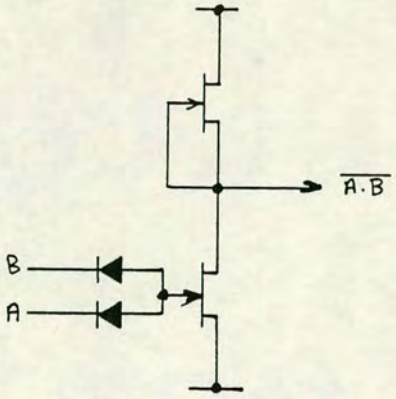


Figure 3.21 Multiple capacitor NAND gate.

difference when simulating system performance, and when deciding operating margins.

Although some useful circuits would demand three input gates, two factors mitigate against such logic elements impractical. Firstly, if a three input NAND gate were made by placing three switch transistors (T1a,b,c) in series, the spread in performance between the different inputs would be greater than that experienced for a two input gate, rendering design even more difficult. Secondly, in order to meet the criterion established for the output low voltage of an inverter, these transistors would be even larger than those used for the two input NAND, thus occupying a larger chip area and presenting an effective loading to the previous stage which is much greater than a fan-out of unity. Consequently, the three input NAND has not been tried in a real application.

An alternative method of fabricating a NAND gate is simply to duplicate the input capacitors without adding an extra transistor. This technique has the potential to allow a fan-in greater than two. Figure 3.21 depicts the circuit in question. A qualitative analysis is sufficient to dissuade the would-be user from adopting this combination.

Assume a starting configuration with both diodes fully charged, and both inputs high. If one of the two inputs (say A) now falls to the low state, the transistor gate must follow, and T1 is switched off. However, as well as charge transferring from D1a into T1, some extra charge must be transferred into D1b to allow it to carry the extra voltage. This charge flow into D1b will have two consequences. Firstly, the input voltage at B will dip as the load T3 carries a current. Thus, a state change at one input to the NAND produces a noise pulse at the other input, which can also be transmitted to other logic gates connected to this node. Secondly, the transfer efficiency of D1a is drastically reduced, and it therefore needs to be larger in comparison with T1 than is required in the standard inverter.

If D1b is also now switched, its starting voltage is larger than normal, and T1 is switched off even harder. The normal voltage change at input D1a may now be insufficient to turn T1 back on again.

These two dangers of diode pumping, and unwanted cross coupling between inputs combine to prevent the use of this configuration.

One further logic element remains, namely the buffer. Although the buffer may be added to any of the foregoing, it is most conveniently added to the inverter. In a few instances it is necessary for a logic gate to drive a particularly large load, either if a clock line is being driven, or if driving off-chip. The inactive pull-up of the conventional inverter gives a marked difference between the rise and fall times, both of which are excessive when the load capacitance is large. (Although a transistor is used in preference to a resistor, it is still considered inactive as its gate is not driven during the transition). A quasi-complementary driver (figure 3.22) can be used in much the same way as the CMOS gate is used. When the input shifts from zero to one, the load transistor T4 switches off, allowing all the current of T3 to be pulled out of the load capacitor. Similarly, on switching the other way, T4 is being switched hard on as T3 is switching off. The normal inverter comprising T1 and T2 is used purely to switch T4. As this inverter does not experience the large

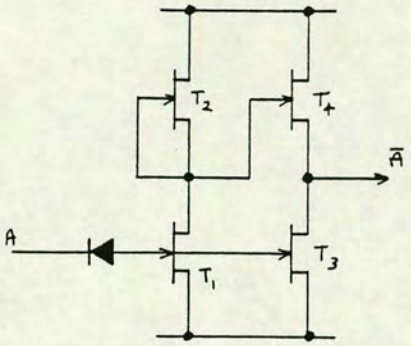


Figure 3.22 Quasi-complementary buffer.

load capacitance, these transistors can be smaller than T3 and T4. Because the full saturated current of the transistors is available for switching (there is no "DC bias current"), the transistor T3 can be half the size of that required for a normal inverter. The load presented to the preceding stage is thus reduced.

The penalties for using such a buffer are: slightly less tidy layout; and much larger switching transients on the power supply. The latter arises because momentarily during the transition, both transistors can be conducting quite heavily. However, the quiescent current of the stage is small. Unlike the case with CMOS, this is not zero, because T4 is always partially conducting, but it is less than that taken by an inverter capable of driving the same load.

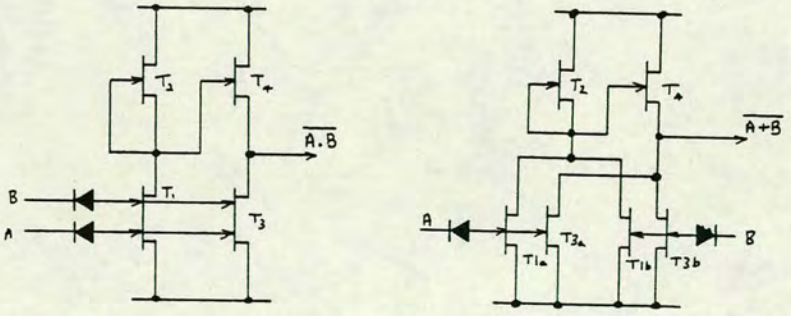


Figure 3.23 Quasi-complementary gates:
a) NAND buffer
b) NOR buffer.

The buffer could be added to both the NAND and NOR gates, as shown in figure 3.23, but the wiring complexity of such gates is problematic. Often the better practice is to use a small logic gate, which has a

lower input capacitance, and to drive a larger buffered inverter with this gate. The resulting switching delay may be only marginally greater than that from the much more complex multi-input buffer.

3.2.3. Precharging simple gates.

In the foregoing analysis it has been assumed that there is direct access to each input under consideration, and in this way, it has been possible to ensure that the coupling capacitors are fully charged. The dynamic transfer characteristic (figure 3.14) showed that the capacitor charge state was as important as the applied voltage in determining the output state of a circuit. Even though the circuit was nominally designed to operate as an inverter, when operated outside the normal bounds (i.e. those defined by the designer), it may cease to perform the "logically" defined function. Such a failure would be expected of a logic element from any logic family. However, with CCL, as well as the voltage and current conditions, the boundary conditions must also include some statement of the capacitor pre-charge condition. At power-on, the system may not (and probably will not) have every logic element within these so-called normal operating bounds. In order to bring the system inside its specification (as derived by a "logical" analysis of the interconnection of the gates) it is necessary to pre-charge each capacitor, such that each logic gate does truly behave as a logical element.

In very large scale integration (VLSI), a similar problem has been given much exposure, in the quest for so-called "design for testability" [339]. In this situation, the problem is not one of ensuring the correct state of charge on a series of capacitors, but is simply a problem of shear scale - how to know the voltage at each and every node within a system of many thousands of nodes. Some form of link between each internal node and the outside world is required to be able to address this question. Although the complexity in CCL is much reduced, the need is exactly the same, to design for accessibility of all internal nodes.

Figures 3.14 and 3.15 show this situation clearly. On the falling edge (figure 3.14), when the voltage immediately prior to the

transition has fallen below some fixed value, the logical operation ceases. On the rising edge (figure 3.15), the same is true when the intermediate node voltage prior to the transition becomes too high. Because the leakage currents cause this intermediate voltage to rise towards zero under static input conditions, there is a maximum time for which the input to the logic gate should be held low. This time depends critically upon both operating temperature and exact wafer processing, so there is no theoretical prediction for this parameter. Experimentally, circuits have been shown to exhibit a lower cut-off frequency around 1-100 kHz. This is sufficiently low to give a very wide range of operation - the upper frequency limit being a few gigahertz. However, it should be stressed that this lower cut-off frequency is applicable at every node on a chip, not just for selected nodes. In particular, this does not simply mean the more readily measured frequency such as that of the strobe or clock.

This situation is made worse at power-on as best illustrated by the case of a long "inverter" chain (figure 3.24). When power is first applied to the circuit, none of the capacitors is charged, so each

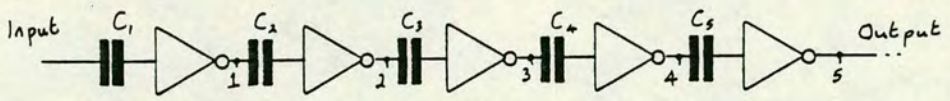


Figure 3.24 Charging a long inverter chain.

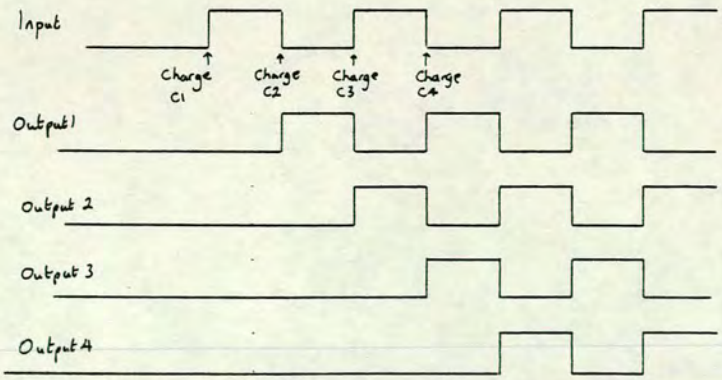


Figure 3.25 Charging cycle for figure 3.24.

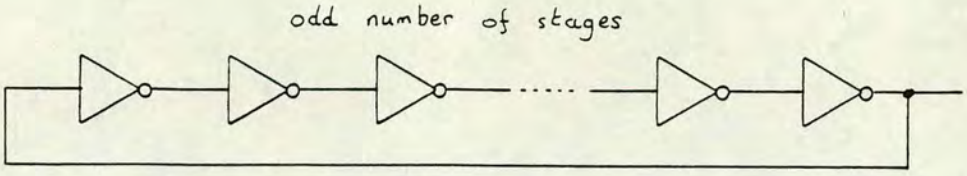


Figure 3.26 Conventional ring oscillator.

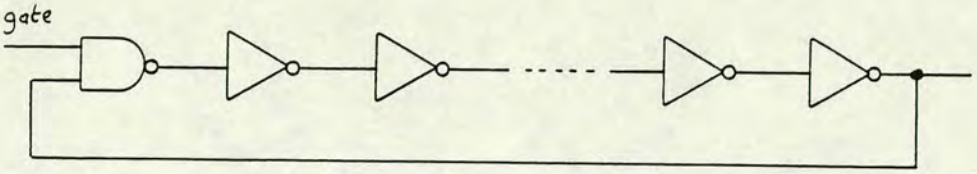


Figure 3.27 Gated ring oscillator.

stage in the chain has its output "stuck-at" logic zero. When a logic 1 is applied to the input, the first capacitor is charged, and the first stage now loses its "stuck-at-0" status. The next zero at the input switches the first stage to a 1, which in turn pre-charges the capacitor of the succeeding stage, releasing the second stage from its frozen condition. The signals continue to ripple down the chain, figure 3.25 showing a schematic charging cycle for this inverter chain. Although an unrealistic circuit, the example adequately illustrates the problem.

Perhaps of more interest is the ring oscillator. In its conventional form, the chain of figure 3.24 is folded around such that the output of the final stage forms the input to the first (figure 3.26). In conventional logic, provided that there is an odd number of stages in the loop, such a circuit will oscillate at a single frequency which is characteristic of the propagation delay of the constituent gates. In CCL, a ring oscillator formed in this way will remain stable, with all the capacitors uncharged. It can only be forced to oscillate by pre-charging the capacitors. This can be accidentally induced in a real circuit either by uneven application of the power supply, or occasionally by the effects of illumination. More rigorously it can be induced by using the circuit of figure 3.27 for the ring oscillator. The pre-charging is forced by the gating signal, and the resultant waveform is that of a burst oscillation, with the

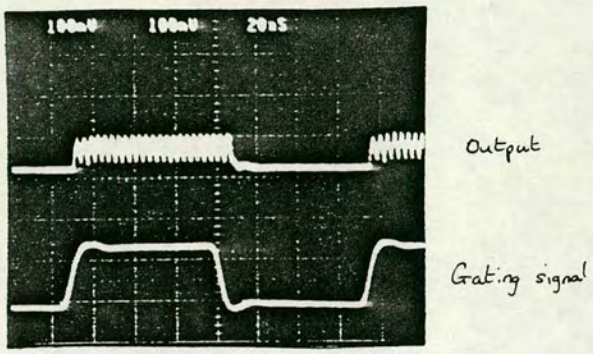


Figure 3.28 Output waveforms of a gated ring oscillator.

oscillation synchronised to the rising edge of the gating waveform. Figure 3.28 shows the output waveform from such a gated ring oscillator.

In Boolean terms, the ring oscillator circuit of figure 3.29 is identical to that of figure 3.27, but with negative rather than positive levels (viz. the high state is logic zero, and the low state logic one). Physically however, there is a significant difference, arising from the "natural" tendency of CCL to be in the switched-on state (positive logic 1, or negative logic 0). In order to break this natural tendency, just one input of a NAND gate must be forced off, whereas all inputs of the NOR gate must be similarly forced. Thus, the circuit of figure 3.27 may be pre-conditioned from a single external connection. That of figure 3.29 cannot be pre-conditioned at all, because the default state of its internal feedback disables any external input.

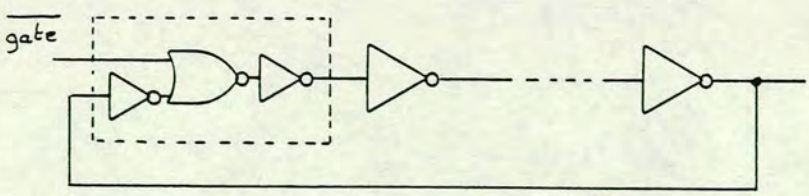


Figure 3.29 NOR gated ring oscillator, grouped to show Boolean equivalence to figure 3.27.

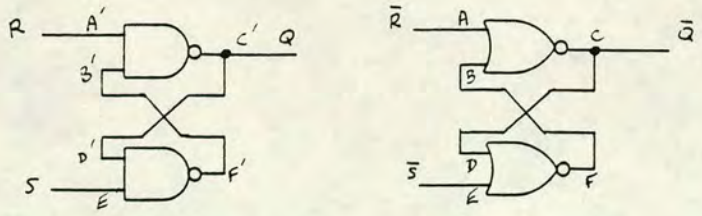


Figure 3.30 RS latches in CCL {(a) works, but (b) does not}:
a) using NAND gates
b) using NOR gates.

This principle may be extended to the latch. Figure 3.30 shows simple circuits using both negative and positive logic. The NAND gate implementation is again successful, but the circuit with NOR gates fails.

In more detail, taking the NOR gate circuit first. With no charge on the capacitors, the output C will be at 0.5 V, but input D will also be at 0.5 V. Thus the output at F will be at 0.5 V irrespective of the voltage on input E. Since F is at 0.5 V, the input B will remain at 0.5 V, thus preserving the initially assumed condition at C. No choice of pattern applied to the inputs can make the latch work.

Applying the same arguments to the NAND gate circuit. If the input A′ is forced to 5 V, the input capacitor will charge, and output C′ will be forced to 5 V when A′ is next brought low. This in turn will charge the capacitor on input D′, such that output F′can be forced high, charging the final capacitor, B′. Thus just by pulsing one of the clock inputs, the whole latch can be charged, and brought into operation.

Quite clearly, the message is that NAND gate circuits appear quite safe to use in CCL, but NOR gates should be used only with extreme caution. NOR gates in circuits containing a feedback loop simply will not operate. The exception to this rule is the combined AND-NOR gate shown in figure 3.31, in which the pre-charging may be controlled using the AND part of the function.

It is necessary always to have thought for the charging sequence of a circuit when using CCL. Some circuits are completely self charging, provided that they are left for sufficient clock cycles. Other

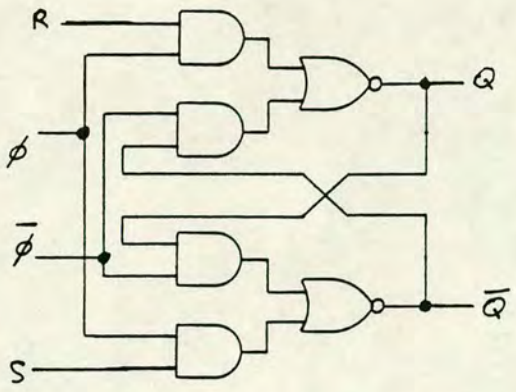


Figure 3.31 Alternative latch using AND-NOR gate, which can be made to work in CCL.

circuits may operate only following the application of a separate control sequence. As shown above, there is a third, large category of circuits to be avoided, and for which the apparent Boolean function is never actually obtained.

3.2.4 Practical application of theory.

Having established the basic principles of device operation, it is appropriate to explore the consequences of some of these details in relation to the particular fabrication process discussed in Section 3.1.2.

The physical model suggested in Section 3.2.1 highlights a major problem somewhat more starkly than does the electrical model. If the diode D1 (figure 3.7) is made from material with the same doping as the transistors, as intimated earlier, then the depletion layer will cease to grow when its applied voltage exceeds the threshold voltage. Any further voltage increase will only change the distribution of the electric field in the substrate. Thus, any removal of charge from D1 must result in its channel being less than fully depleted. As the cathode of D1 is always more positive than the source of T3, this in turn prevents T3 from being fully cut-off. Transistor T3 is switched closer to the off-state as the area of D1 increases. Thus, if D1 is made very large, it should still be possible to operate CCL with both the diode and the transistor in the same material, but this is a somewhat unwelcome constraint, as any useful circuit will require the duplication of D1 manyfold, with the obvious penalty in area.

An alternative approach was used for the early test chip to allow the single threshold voltage technology to be used. Instead of increasing the size of the diode, the anode was patterned as a series of fingers to give a large periphery to the depletion region. Now, when the voltage increases beyond threshold, the lateral spread of the edge of the depletion region will still give a significant change in the quantity of charge stored.

However, this expedient proved less successful than hoped. The actual size used for the diode was somewhat marginal. In addition, the long narrow gaps between the depleted areas resulted in a very high resistance in series with the capacitor, giving a low cut-off frequency (see Section 4.1.3). Furthermore, the first results obtained showed up a much larger pattern dependence of the transfer efficiency than was expected simply for a small coupling capacitor. After much investigation, the back-gating effect was identified [340], and the unique role that this plays in CCL was also highlighted. Figure 3.32 shows the dramatic effect of back-gating on the diode capacitance.

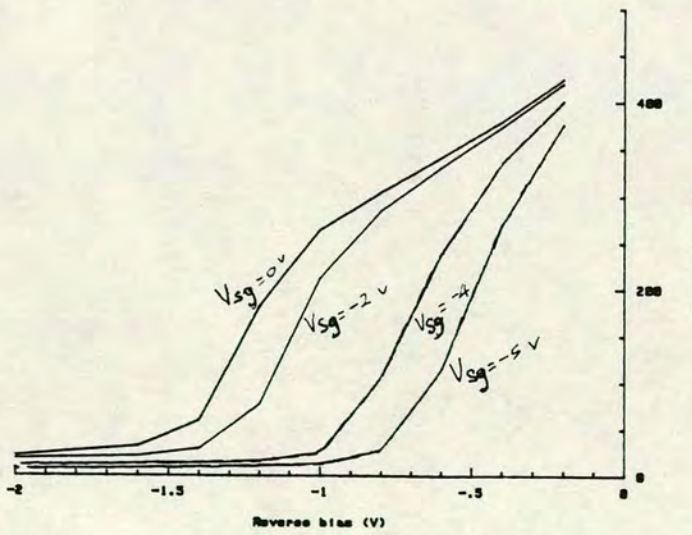


Figure 3.32 Capacitance variation of a diode with back-gating voltage (V_{sg}) applied to an adjacent Ohmic electrode.

Some discussion of the influence of back-gating on the transfer characteristic of the inverter is called for. The back-gating phenomenon causes the effective position of the substrate to channel interface of any diode or transistor to vary according to the mean

voltage on adjacent surface electrodes. The greater the average reverse bias, the more the channel is depleted from the back. As the channel thickness is a contributory factor in determining the threshold voltage, the whole transistor behaviour is changed by the influence of the back-gating voltage. This introduces some degree of pattern dependence into the logic gate behaviour, as the time-average clearly changes for different data patterns.

The different components are influenced in different ways. Transistor T3 is not grossly affected by back-gating, as the mean voltage at the centre of the channel is low, giving a low back-gating voltage. Because the voltage on the source of T3 contributes to the back-gating voltage on T4, the latter voltage is high, partially closing the channel, and reducing the saturation current. Thus the two transistors are influenced differently, and the effect on the circuit is to change the ratio of the current drive capabilities of the switch and load transistors. As the circuit was designed to cope with a large spread in threshold voltage (and therefore in this ratio), the circuit operation is relatively insensitive to this parameter. The operating points are affected marginally, as is the switching speed. The influence on D1 is much more marked, especially when the diode material has a small threshold voltage. Although the mean back-gating voltage is somewhat smaller than that for T4, it is still quite high and the back depletion can be large. Consequently, D1 reaches threshold before T3. The charge which can be stored in the fully depleted layer is thus reduced, as also is the charge stored during the lateral spread of the depletion region. The net effect is to change the transfer efficiency, in some instances by a large factor.

Having highlighted back-gating and poor transfer efficiency as two severe problems with the fabrication process already outlined, some changes were implemented to give better performance.

As discussed in Chapter 2, the threshold for the onset of the back-gating phenomenon can be moved away from the operating region by ion implantation of boron or other similar mid-gap dopant into the "isolated" regions. Boron was chosen, because it is light enough to

isolate without causing significant damage, whereas the alternative, heavier elements cause sufficient damage to require some heat treatment. The patterns used for protection of the active area during the mesa etch are also suitable for masking the isolation implant, so no extra mask design was required to implement this change.

Although poor transfer efficiency was partly due to the loss of charge because of back-gating, the low Q of the digitated capacitors, especially when operated beyond threshold, gave rise to an intolerable degradation in the circuit performance. To improve the dynamic behaviour of the capacitors, it was necessary to reduce the series resistance quite significantly. This required a diode threshold voltage significantly higher than the circuit operating voltage, giving both a low spreading resistance and near equal efficiency across the whole of the diode. The large periphery diode was no longer practical, and it was necessary to implement a process in which the diodes and transistors were fabricated separately, thus moving away from the simplest possible process.

Initially, the least complicated approach was adopted, whereby only a single type of layer was formed by direct doping, this being later modified to form the second layer. This starting dopant is chosen to optimise the low resistance capacitors, giving a very deep channel and a large threshold voltage. Shallow transistor channels are then formed by etching away the surface region before deposition of the gate metal. By measuring the current flowing between the source and drain (at saturation) before depositing the gate metal, a measure of the integrated dopant in the channel can be obtained. As this current is related to the threshold voltage of the finished transistor, the required parameters are obtained by etching the channel until the chosen current is measured. The threshold voltage of transistors formed in this way lies reasonably close to the designed value. Clearly, however, the non-uniformity of the etching process adds to the existing non-uniformity of doping, increasing the variance of the parameters across the wafer.

In order to be able to recess the transistor channels, without also recessing the anodes of the diodes, an additional masking level is required. Following the formation of Ohmic contacts to source and drain, the new pattern is printed to protect the GaAs surface except in the region between the source and drain. This area is etched away to the required depth, prior to the deposition and patterning of the Schottky metal using a lift-off technique. This new sequence has an additional critical alignment, as the 1 μm gate must lie fully within the recess (patterned at 3 μm), which in turn lies in the 5 μm gap between source and drain. These dimensional and tolerance requirements were quite demanding when first used, especially for a manual alignment technique. Even more problematic was the lift-off of the gate metal from the highly non-planar region which the active area had become. Whilst this was not an impossible task, large wings of metal were invariably present where some metal had deposited on the side of the photoresist patterns. These caused a significant yield hazard. Multi-level photoresists were developed to improve the definition and surface finish of the metal tracks, thereby also improving yield.

Following these changes to the processing, it was possible to obtain useful circuit results from the test chip. However, these pointed to further inadequacies in the process, which resulted in propagation delays far inferior to those expected. An upper frequency of 1 GHz was recorded for this technology.

This poor performance was in part attributable to the series resistance between source and gate of the transistor. The surface states cause the GaAs surface to be depleted in the gap between the source and the gate, leading to loss of transconductance, and therefore switching speed. Although this had been expected to cause trouble for devices with low threshold voltage, no significant problem had been anticipated for the deep channel associated with -1.5 to -2 V thresholds. The recessed gate technology, adopted to help overcome the twofold problem of diode Q and back-gating, is clearly an advantage in minimising this resistance, because at least a part of this "extrinsic" region is of very low resistivity. However, using the rather crude recessing technique already described

still leaves a relatively large gap of high resistivity material adjacent to the gate. In particular, the resistance is excessive in the dual gate transistor, where the recessed region must be sufficiently wide to accept both gates. The upper gate consequently experiences a very large effective component of source resistance.

The final refinement in this sequence of improving the transistor performance involved the use of a self-aligned gate recess technique. Now, the gate metal pattern is first defined in photoresist (PR). The recess etch is performed, as before using the source-drain saturated current as a measure of the end-point, and metal is deposited. The PR is then dissolved to lift-off the unwanted metal, leaving the gate metal only in the recess. The lift-off process is actually assisted by the presence of the recess, and the major additional process hazard concerns the morphology of the all-important metal to semiconductor contact at the bottom of the recess. However, no significant problem has been observed with this aspect of the device. There is no recess in any of the "extrinsic" region, and the two gates of the dual transistor lie in separate wells. Figure 3.33 shows a schematic cross section through this device. One quite major attraction of this technique is that the carrier concentration may be very high in the surface region which is etched away, thus giving a very low resistance indeed for the extrinsic region. Provided that this doping peak lies close to the surface and that both the recess and the metal are tapered as shown in 3.33, there is no danger of low breakdown voltage.

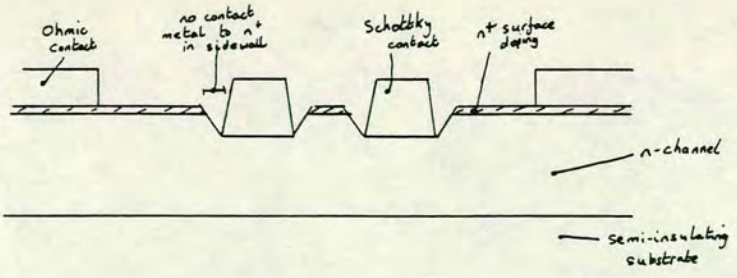


Figure 3.33 Cross-section of a "self-aligned", recess-etched dual-gate MESFET.

Adopting this technique for improving the FET performance demands a re-think of the remaining process, as the anode metal was traditional patterned at the same time as the gate metal. Clearly, this is no longer feasible, as the diode must retain the high threshold. However, in this more sophisticated process, there is no longer the constraint of using identical doping profiles in the diode and transistor, and it would be possible to dope the diodes sufficiently heavily to retain the surface recess. The alternative solution is to use separate metalisation steps for the diode and transistor. As well as forming the gate, the "first" level of metal is also used to cover the Ohmic alloy, and it is preferable to protect this alloy from the recess etch. Thus if two separate metal depositions are performed, the first is used only for the gate areas and receives the predeposition etch, but the second forms the anodes, covers the Ohmic alloy, overlays the first deposition outside the transistor area to forge the contact and also forms the first interconnection level. This use of separate metal depositions is preferred, but it should be emphasised that these are not separate layers in the conventional sense, as the second deposition immediately follows the first, without an intermediate dielectric. Note that this is uniquely possible using the lift-off patterning technique.

Having now obtained an "ideal" FET fabrication route, it is instructive to explore the remaining weaknesses of the earlier process. As well as the transistor parasitic resistance being poor, the series resistance of the coupling diodes was still not optimised. Although the threshold voltage had been improved, the doping profile could not be truly optimised in this type of process, and the digitated anodes were still being used to avoid complete mask redesign. Because of their long, thin construction the resistance was too high, even on the new material, where the switching current could flow beneath the depletion layer, rather than just in the gaps between fingers. The diodes were completely redesigned using the models presented in Chapter 4.1 to give low resistance and smaller capacitance by virtue of a much reduced area.

In reducing the area occupied by the diode capacitors, a second benefit accrued from the reduction of the stray capacitance.

Although designers originally claimed that parasitic capacitance would not be a problem with GaAs because the semi-insulating (SI) substrate would offer very low values, this was rather a hopeful and naive assumption. As a direct consequence of the SI substrate, the capacitance to substrate may be reduced, but the inter-electrode capacitance is increased. Thus the latter is significantly higher in GaAs than in silicon. This calls for great care in choice of layout, particularly where adjacent areas may be subject to the Miller effect. Of particular concern was the positioning of the diode with respect to the transistor and, in the NAND gate, the inter-relationship of the two diodes. The actual size and shape of the coupling diodes therefore proved to be important, because their parasitic loading effect on the circuit was quite significant.

With modification to both process and diode construction, it is instructive to review one aspect of the circuit operation, namely the effect of avalanche breakdown [341]. In the transistor, this breakdown first takes place at the drain end of the gate. In a circuit, this is most likely to occur in a switched-off inverter, when the gate is at its most negative, and the drain its most positive. The breakdown voltage must therefore exceed the absolute sum of the supply voltage and the threshold voltage. Strictly, the value should be larger than this, as the negative input voltage can overshoot the threshold quite considerably. The diode too has a similar breakdown condition, but its specification is more relaxed, as the diode voltage does not exceed the supply voltage.

Clearly the requirement that the transistor does not enter breakdown is rigid, as the transistor parameters (particularly the Schottky barrier parameters) would be degraded by the high fields. The specification for the diode is less obvious. If the diode begins to enter a region of soft breakdown (i.e. a condition of increasing, but still low, leakage current) only at its largest operating voltage, some benefit may obtain. This condition will only arise when the inverter is switched on, and the increased leakage current will flow through the transistor gate, forcing a well determined input state. This is in contrast to the normal situation when the leakage current is very small, and the transistor operating point is both ill-defined

and pattern sensitive. The circuit configuration provides a self-limiting breakdown condition, which will not damage the diode, and performance is actually enhanced. When the circuit switches logic states, the diode moves away from breakdown, and suffers only the normal leakage current.

This argument suggests that the ideal circuit uses a diode with zener breakdown voltage about 0.5-1 V less than the supply voltage. This is an unrealistic target, as the breakdown condition in GaAs is even less well controlled than the threshold parameters and is quite sensitive to variations in operating temperature. However, the practical consequence is that the design tolerance on the breakdown voltage may be smaller than otherwise anticipated. At the worst case value of breakdown voltage, the circuit can then be allowed to operate in this mode.

Practically, the specification of the breakdown voltage sets a limit on the peak doping level in the device, although the exact value allowed is dependent upon the depth at which the peak occurs. If the same doping is to be used for the diodes and transistors, but with the latter recessed, the surface and bulk doping levels are independently chosen by the respective breakdown criteria. Thus a profile as shown in figure 3.34 may be appropriate.

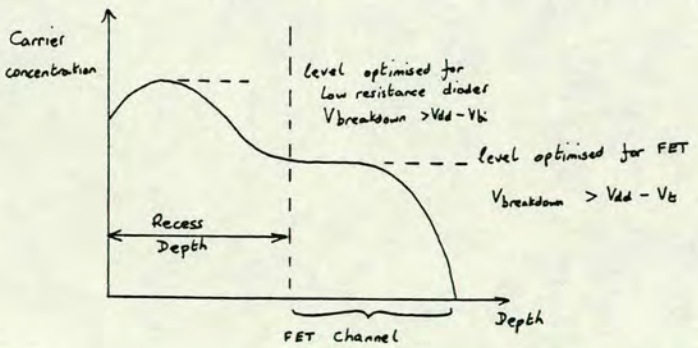


Figure 3.34 Doping profile for recessed-gate process.

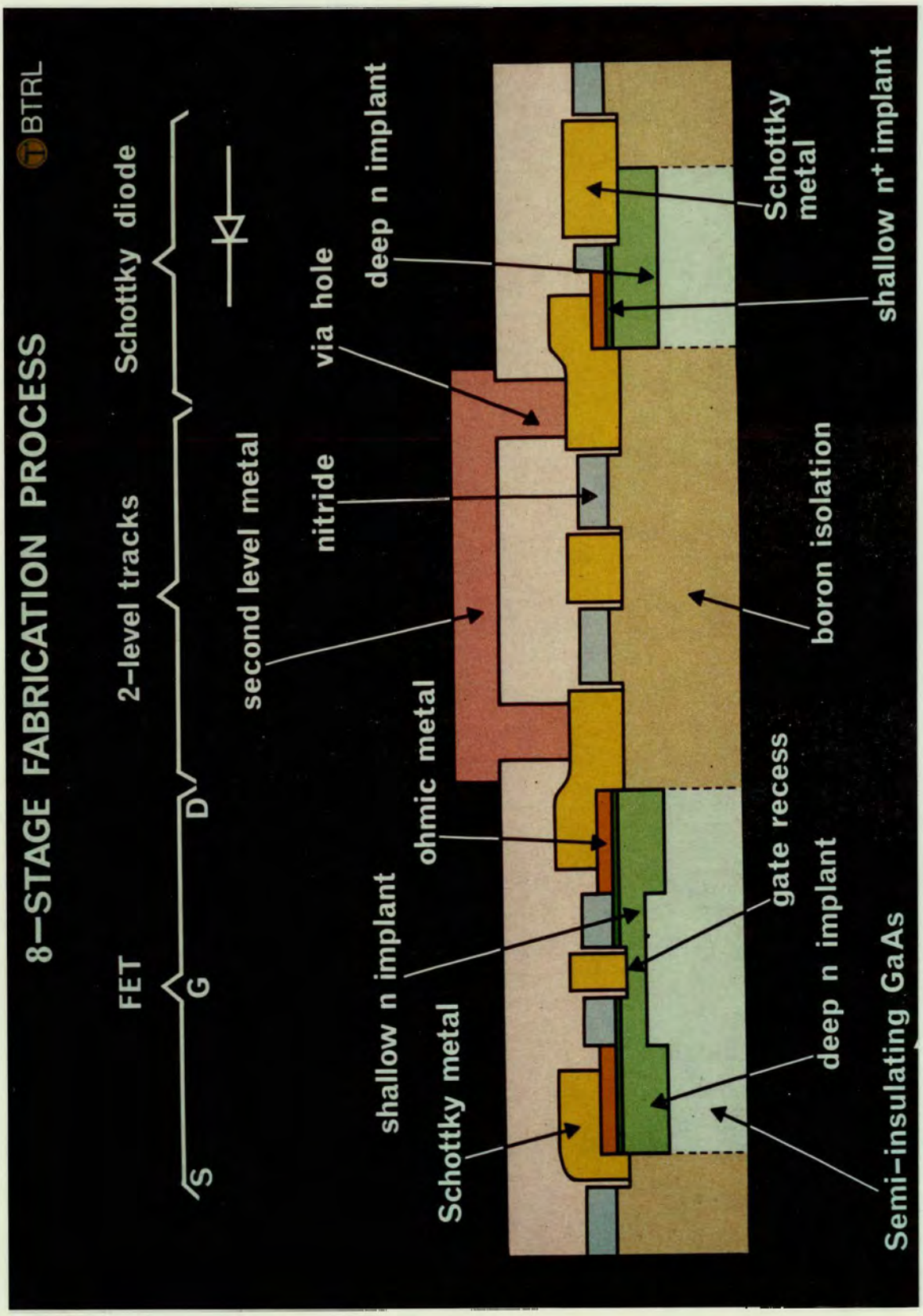
It was suggested earlier that the two active regions could benefit from individual tailoring, by ion implanting into selected areas using photoresist masks. The advantages accrue principally from the increased freedom. The shallower the transistor recess, the more

accurate the control of threshold voltage, but the worse the series resistance, unless the surface doping can be increased above the level set by the diode breakdown. Equally, the shallower the recess, the sharper can be the transition between the active area and the substrate, with attendant improvement in the transconductance near threshold. In contrast, the deeper is the diode, the higher is its quality factor. Again, the deeper below the surface is the peak, the larger the threshold for a given carrier concentration, and the higher that peak level may be. The major disadvantage of this approach is the obvious increase in process complexity.

In making this move to direct selective implantation of the active regions, it is a natural development to move away from the mesa etching technology. Two benefits accrue from this. The first and most obvious is the improvement in the topology, which eases the metalisation process, particularly for the second level. The second benefit is also related to step coverage. In the absence of a step between the isolated and active areas, there is no longer a requirement to taper the gate metal before it comes out of the active region (as in figure 3.4). Eliminating the taper also removes the small part of the transistor where the source and drain do not lie parallel, and which is effectively a long-channel transistor. This promises to improve performance by some 10-15%. During the redesign that this heralded, all the other layer-to-layer tolerances were tightened, to reduce the size of the source and drain regions, again assisting reduction of the parasitic capacitances. It was also possible to reduce the separation between gate, source and drain as a part of this general evolution.

Thus a continual improvement in the understanding of both circuit and device operation was used to update the processing technology throughout the period of this study. Figure 3.35 shows a schematic cross section through the process at this final stage of the evolution. The latter sequence of changes were not fully incorporated until the design of the 8:1 multiplexer circuit whose description forms the basis of Chapter 5. However, the performance was then stretched to beyond 2.6 GHz and since completion of this

Figure 3.35 Detailed cross section through GaAs IC process used in this work.



work, others have continued this process of refinement to extend the performance well beyond 3 GHz [342].

3.3 Functional blocks

In principle, all digital circuits can be fabricated using only the logic building blocks described in Section 3.2. However, in most cases it is desirable to design a few of the more common elements at the transistor rather than the gate level. In complex (VLSI) systems, the driving force behind such customisation is for optimisation of the silicon area, as exemplified by the storage cells and sense amplifier cells of the random access memory (RAM) [343].

However, in this work, the concern is for optimised switching speed rather than for minimum circuit area. Here the prime target for optimisation is the latch, which holds the key to all high speed functions. Section 3.3.1 deals with this aspect of design. As switching speeds become higher, both the buffering and distribution of the clock signals assume renewed importance, and attention is given to these in the following Section. The final part (3.3.3) of this Section deals with input and output switching at high speed.

3.3.1 Latch design

There is a number of different circuits referred to as latches [344], for each of which, there is a further variation in ways of possible implementation. The latch may be synchronous if the output only changes upon application of a clock or strobe signal, or it may be asynchronous if the output directly responds to a suitable change at the input. The following latch types are recognised: D-type, T-type, R-S and J-K.

A D-type latch is necessarily synchronous, with the output being determined by the state of a single D(ata) input at the moment of the clock transition. Some designs may require both true and complementary data signals to be present, but this is still treated as only a single input, because the two signals both carry the same information.

In the ideal latch, the output state is independent of the input data except at the moment the clock line is strobed. At this point, the current input data is memorised and the present state is stored until the next strobe signal appears. However, the simplest D-latch circuit is transparent during one half of the clock cycle (i.e. when the clock is held either high or low, depending upon the design). Thus, the input data is memorised on (say) a falling edge, and held until the next rising edge, whence the output again follows changes occurring at the input, pending the next falling clock edge. To use such a circuit, it is necessary to guarantee that the input data remains constant during the transparent phase. In some limited applications, this may be an acceptable restriction, but when the ultimate high speed is demanded, it is usually unwelcome either to impose such constraints, or to modify the circuitry to ensure that this condition is met.

It is usually preferable therefore to move closer to the ideal, and employ a latch which is not transparent. This may comprise a pair of the simple latches cascaded, arranged such that the transparent phase of the second coincides with the latched phase of the first. Thus, the data at the input of the second does not change whilst it is transparent, and the data appearing at the final output is only allowed to change once every full clock cycle. It should be stressed that in this design, the data appearing at the output immediately following a strobe signal is not that present at the input immediately prior to the strobe, but that occurring half a cycle earlier. This is illustrated in figure 3.36 which shows a schematic timing diagram for this so-called Master-Slave (M-S) configuration [345]. In practice, the two latches may be identical, but simply activated off the opposite edges of the clock.

There is an alternative design of opaque D-latch. In this, the so-called edge-triggered latch [346], the clock and the data are pre-conditioned, such that the actual storage element is enabled only during a transition of the clock, rather than by one of the logic levels. A logic circuit for this has already been shown in figure 2.10.

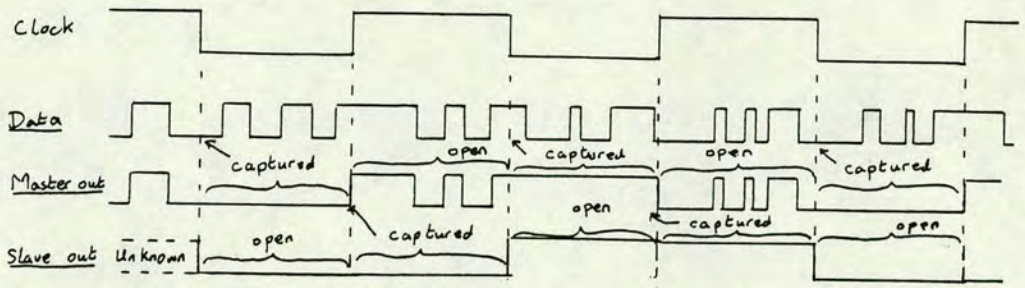


Figure 3.36 Schematic timing diagram for a master-slave latch.

In a T(oggle) type of latch, only the clock signal is required, and the latch state changes on each clock cycle, thus dividing the frequency of the clock signal by two. In an extension of the T-latch, a control signal may be provided, whereby the dividing action of the latch may be disabled (e.g. the circuit divides when there is a logic-1 present at the control input, but remains unaltered when there is a logic-0 at the input). The former (uncontrolled) version of the T-latch may be constructed by inverting the output from a D-latch and returning it to the data input. Clearly, if oscillation is to be avoided a fully opaque latch must be used.

In the R(eset)-S(et) latch, two signals are required. The latch responds to the presence of either a set or a reset signal, by changing respectively to an on or an off state. In the absence of a signal, the latch remains unaltered, but if both set and reset are present, the output state is unpredictable. When the R and S inputs are made complementary in order to prevent both signals being present together, the latch reverts to the D-type configuration already described. R-S latches may be constructed in either positive or negative logic, such that 'the presence of a signal' may mean either a logic-0 or a logic-1, according to the particular design. Both synchronous and asynchronous varieties also exist.

In the final variation of latch, the J-K, two inputs are still required. The performance is similar to that of the R-S latch, except that the unfortunate problem of the indeterminate state has

been eliminated. The J and K inputs correspond to the S and R inputs. However, when both are present, the latch behaves like a T-latch. The J-K latch thus has a complete truth table, with all possible output responses. Table 3.1 lists this truth table, assuming positive logic.

Table 3.1 Truth table for a J-K latch

J	K	Q(before clock)	Q(after clock)
0	0	0	0
0	0	1	1
0	1	X	0
1	0	X	1
1	1	0	1
1	1	1	0

where X represents don't care, and Q the latched output.

In high speed logic, timing is invariably critical. A complex system may comprise several serially latched events. Each stage in an asynchronous system responds immediately, thus introducing the minimum possible delay, and the signal may therefore propagate more quickly from input to output in an asynchronous system than in a fully synchronous environment. However, it is very difficult either to predict this delay accurately, or to effect adequate control of the delay on individual chips within a large production run. If each latch is synchronous, although several clock cycles may be required before a given input change is reflected at the output, the variation in the total delay time will be very small. Furthermore, other events could have taken place during the intervening clock periods (pipelining [347]), therefore increasing the effective throughput. Thus the synchronous system generally gives a higher rate of operation after allowing suitable design margins, and is invariably preferred for high speed operation.

The key to maximising the circuit speed lies in minimising the propagation delay through the latch itself. Depending on the application, it may also be necessary to minimise the logic being performed at the input of each latch. The choice of latch variety is usually further limited at high speed. As great care is required to

ensure non-overlap of the R and S signals in an R-S latch, it is usually impractical to use these, with the J-K variety being preferred. However, the additional circuitry required to perform the J-K function renders it slower than the D-type latch. When a toggling function is required, a dedicated T-latch is usually much faster than a modified D-latch. Generally therefore, the D-latch and the T-latch are the only varieties used at high speed.

The T-latch is the easier of the two circuits to analyse, as there is only the single input variable: operation ceases when the clock frequency exceeds the critical maximum. In the D-latch however, failure occurs either because the clock is too fast, or because the phase relationship between clock and data is in error. A detailed, non-linear analogue circuit simulation is required to predict the maximum operating frequency. However, at this stage a qualitative assessment of peak performance is required in order to limit the number of such costly simulations which are necessary. The simulation will only be used to optimise the performance of a given circuit, once the best circuit has been identified.

The simplest analysis treats the latch as a combinatorial logic element, with the feedback lines broken and each feedback connection treated as an input. The circuit is set into one of its stable states, and then a change is introduced at the clock. The corresponding new output condition is calculated after exactly one propagation delay. A new set of input vectors is generated from the revised output and this in turn is propagated to the output. This procedure is followed until a stable condition is found, whence a new change is applied to the clock. The logic diagram of the simple divider used on the initial test chip is reproduced in figure 3.37, together with the revised circuit used in this procedure. Table 3.2 shows the analysis. After twelve propagation delays, the complete divide-by-two sequence is completed. Since this takes two clock cycles, it is apparent that the minimum clock period is 6 times the propagation delay of each constituent gate.

This information may be displayed pictorially on a Karnaugh map [348]. Because of the symmetry of the circuit, the number of

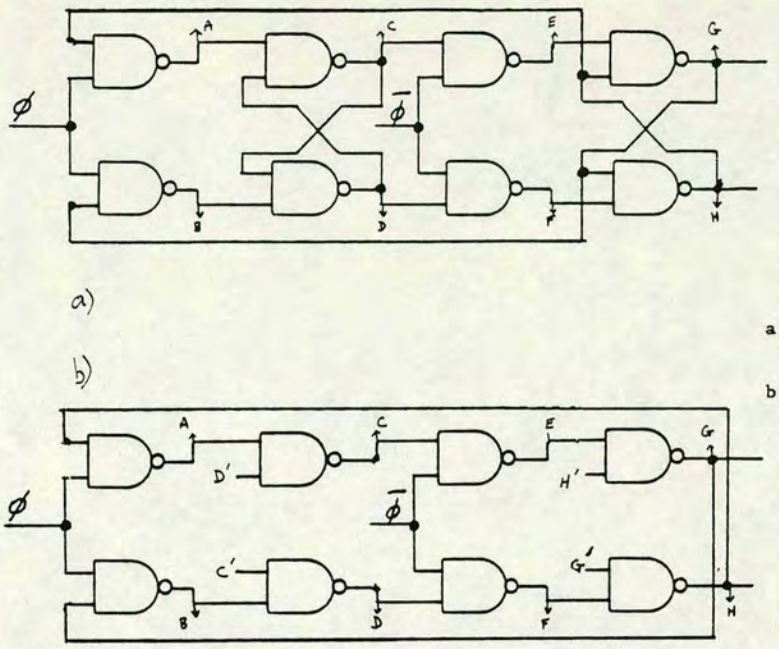


Figure 3.37 8-NAND T-latch used in the analysis:
a) complete latch, identifying nodes
b) feedback nodes "broken" and re-labelled.

allowed states is fortunately fewer than the 512 suggested for a full mapping (8 outputs and a clock). The clock information may be omitted from separate columns, provided that the analysis is only required for the case when ϕ and $\overline{\phi}$ are exactly complementary. If the analysis is restricted to sequences derived from an initial stable state, the portion of the Karnaugh map which is actually required is remarkably small. Figure 3.38 shows this selected section of the Karnaugh map. The two states of the clock are denoted by the shading, and the four stable states are clearly marked. In order to move from a stable state, the clock must be altered.

Continuing the analysis further, it is instructive to see what happens if the clock is toggled faster than allowed by the above sequence. For convenience the graphical technique will be used, although a table similar to the above is required to generate the map. The clock is changed exactly one propagation delay (t) early, and continues to change every 2t. The map shows that the circuit goes into a previously unused state after one period (t), but that after 2t, the circuit is back "on course". This pattern is followed all around the map, and the circuit still behaves as a T-latch.

Table 3.2 Analysis of the circuit in figure 3.37

ϕ	C'	D'	G'	H'	A	B	C	D	E	F	G	H	Comment
1	-	-	-	-	0	1	1	0	1	1	0	1	Stable
0	1	0	0	1	1	1	1	0	0	1	0	1	
0	1	0	0	1	1	1	1	0	0	1	1	1	
0	1	0	1	1	1	1	1	0	0	1	1	0	Stable
1	1	0	1	0	1	0	1	0	1	1	1	0	
1	1	0	1	0	1	0	1	1	1	1	1	0	
1	1	1	1	0	1	0	0	1	1	1	1	0	Stable
0	0	1	1	0	1	1	0	1	1	0	1	0	
0	0	1	1	0	1	1	0	1	1	0	1	1	
0	0	1	1	1	1	1	0	1	1	0	0	1	Stable
1	0	1	0	1	0	1	0	1	1	1	0	1	
1	0	1	0	1	0	1	1	1	1	1	0	1	
1	1	1	0	1	0	1	1	0	1	1	0	1	Stable

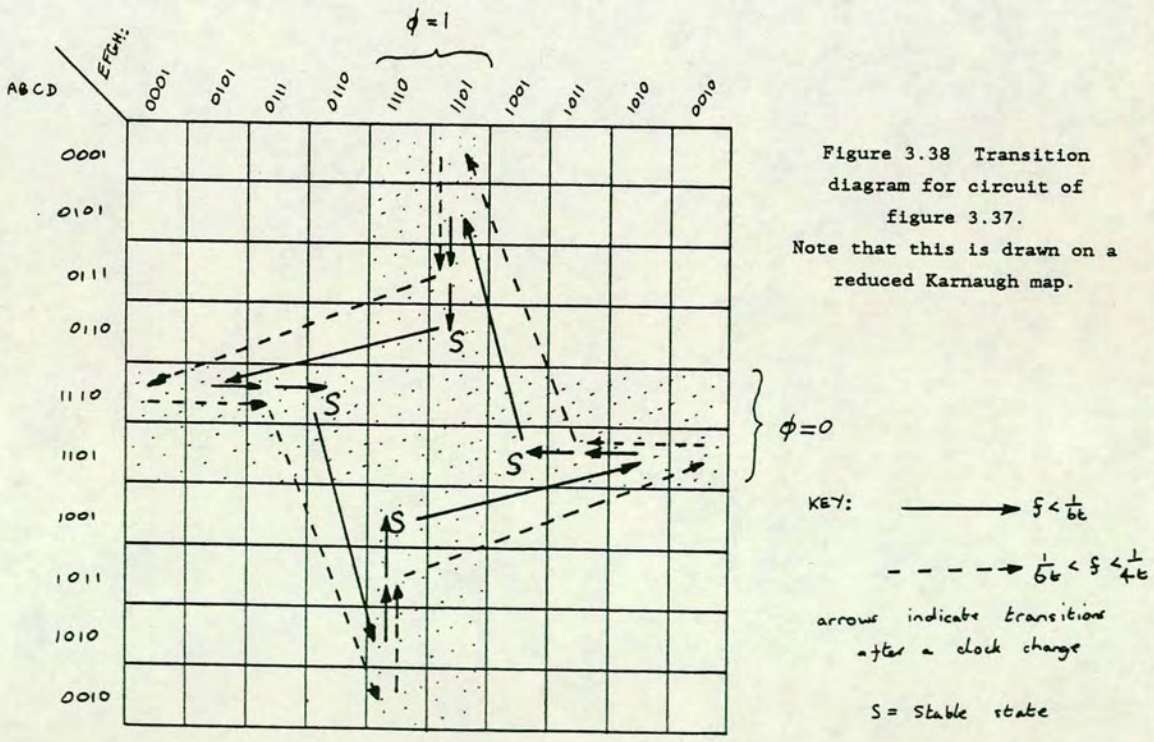


Figure 3.38 Transition diagram for circuit of figure 3.37.
Note that this is drawn on a reduced Karnaugh map.

Although 12t are required to complete the stable loop, the circuit is clearly capable of operating properly in 8t. The similarity of the two loops on the Karnaugh map leads to the suggestion (borne out in practice) that the circuit will operate properly up to a clock frequency (f) of ($1/4t$). If the clock is changed after every period, t, a similar analysis shows that the circuit will soon degenerate to an unpredictable oscillation.

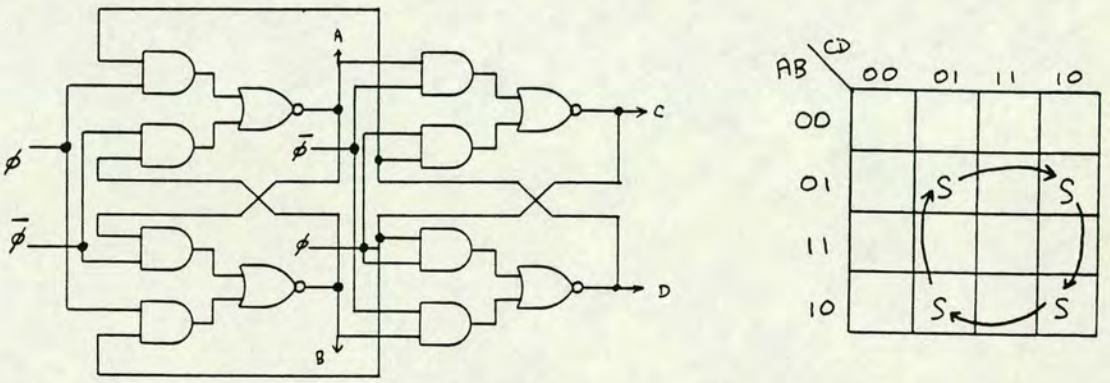


Figure 3.39 AND-NOR divider
a) logic diagram, b) Karnaugh transition map.

Applying the same technique to the alternative circuit comprising 4 AND-NOR gates (fig 3.39a) yields the map shown in figure 3.39b. It is important to note that the AND-NOR gate requires just a single propagation delay to perform both the AND and the NOR, as there is only one current being switched. Comparing the two maps, we see that the AND-NOR circuit has a theoretical upper frequency of $1/(2t')$ compared to the $1/(4t)$ of the 8 NAND circuit. Note however that the propagation delays are not equal for the two cases. Although the switching time of the NAND gate is smaller than that of the AND-NOR, the difference between the two is insufficient to make up for the factor of two, and the AND-NOR is faster than the NAND implementation. Furthermore, at or near the high frequency, the output from the latter has a 5:3 mark space ratio, where the former has a 1:1 ratio. This may prove critical in system designing. As the edge-triggered latch introduced earlier is not practical in CCL, and is much slower, it will not be considered here.

The analysis presented is in terms of a purely digital signal. A step-change is effected at the input, which appears at the output, still as a step-change, but delayed by a fixed period, the so-called propagation delay. In a real circuit, the transition is analogue, and the circuit will begin to respond very much earlier in the cycle than indicated. Further, not every gate in the circuit will share the same value of propagation delay, and even a single gate may be different in the response time for the 1-to-0 and the 0-to-1 transitions. The foregoing analysis should therefore be used as a guide, rather than as the whole truth.

The pure symmetry of the AND-NOR configuration suggests that the analysis can be applied here with more confidence than to the 8-NAND circuit. However, even here the symmetry is broken by the requirement to obtain an output. One (or two if complementary outputs are required) of the four gates will suffer a larger load capacitance than the others. Clearly, the same degree of symmetry is not present in the 8-NAND circuit. This asymmetry may be used to advantage to scale the sizes of the individual gates. Referring to figure 3.37, gate G is the most heavily loaded, and should therefore be the largest. Similarly, A is the most lightly loaded, and should be the smallest gate (both A and E have only a single NAND load, but we have established that gate G is larger than gate C). If the effective load of a gate is directly related to its size (very nearly true), the appropriate sizes for equalising the propagation delays in the eight gates may be calculated from the following equations, in which f is the effective fan-out of each gate, A,C,E and G are the widths of the pairs of gates, and M is the width of the output buffer:

$$C = A \cdot f$$
$$C + E = C \cdot f$$
$$G = E \cdot f$$
$$G + A + M = G \cdot f$$

Setting the gates of both A and M to a unit size, but treating the loading effect of the output buffer (M) as a half unit because of the smaller input capacitance of an inverter, the fan-out on each gate is 1.7, and the respective widths of the gates are:

A : C : E : G : M = 1 : 1.7 : 1.2 : 2.1 : 1

Again, the example shown is highly stylistic, but well illustrates the point. In a more detailed analysis, this calculation would be used to obtain starting sizes which would be used for a trial layout, before evaluating the actual load capacitance on each node including the effects of the parasitic elements arising from the physical layout. The whole exercise would be repeated and the calculation re-performed, using these new coefficients. The help of a circuit simulator would probably be enlisted at this point. Note that this circuit now offers a much closer performance to that of the AND-NOR gate. Although it requires twice as many propagation delays, the value of t to be used is that for an average fan-out of 1.7 instead of some 2.2.

It now becomes essential to view the latch more specifically as an element for use in a CCL circuit. Again the T-latch will be used for the study, although the analysis covers the general case of the D-latch. The crucial question is that of initialisation, as raised in Section 3.2.3. There, the general unsuitability of NOR gates within feedback loops was discussed, and these guidelines have already been followed. However, it is wise to ensure that the remaining circuits are actually acceptable.

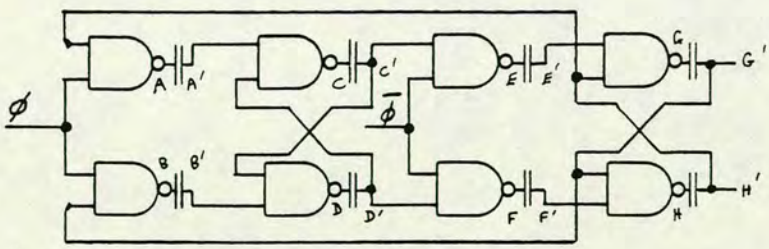


Figure 3.40 8-NAND T-latch showing capacitor placements.

The 8-NAND circuit of figure 3.37, is re-drawn (figure 3.40) to emphasise both the capacitors and the notation to be used. Signals are then propagated from the input until each node is forced into a pre-determined state. Table 3.3 charts the charging of the capacitance on each node during the first few applied clock pulses

Table 3.3 Charging of the capacitors in an 8-NAND CCL T-latch

ϕ	$\bar{\phi}$	ϕ'	$\bar{\phi}'$	A'	B'	C'	D'	E'	F'	G'	H'	M'	N'
0	1	z	1	z	z	z	z	z	z	z	z	z	z
1	0	1	0	z	z	z	z	1	1	z	z	z	z
0	1	0	1	1	1	z	z	0	0	1	1	z	z
1	0	1	0	0	0	1	1	1	1	y	y	w	w
0	1	0	1	1	1	y	y	w	w	y	y	w	w

Key: z = uncharged;
y = charged but in critical race condition i.e. one of the pair at logic-1, other at logic-0,
w = charged, but state depends on outcome of critical race.

showing that the latch starts automatically upon the application of a clock, albeit after two and a half clock cycles rather than immediately.

Repetition of this exercise for the AND-NOR divider, shows that the circuit will not pre-charge by such a simple expedient, but will sit with all capacitors uncharged. Experimentally this is generally found to be the case, although such circuit configurations do occasionally self-charge, probably through noise and imbalance in the power line distribution. This is by no means reliable however.

The other attractive features of the AND-NOR divider (viz. its superior speed) make it worthwhile pursuing. As discussed in Section 3.2.3, the problem of pre-charging the nodes boils down to one of gaining access to each logic element. The circuit of the AND-NOR divider (figure 3.39), by virtue of its symmetrical clock inputs, does already possess an external signal on each logic element. If all the clock lines (both ϕ and $\bar{\phi}$) were simultaneously switched to logic zero, then every capacitor within the latch would pre-charge, whence the circuit does behave as a true divider. Thus if the clock lines are used as dual purpose signals, both to precharge the circuit prior to operation and to control the normal sequence of operation thereafter, the higher speed potential of this circuit can be exploited to the full. It should be noted that this concept of preconditioning a circuit before operation is not new, but has been widely used in memory control for many years [349].

Before exploring the practical consequences of this approach, it is worth studying its wider implications. If a moderate to large system were to be developed using CCL, it would be necessary to explore every possible starting state to ensure that the "false" patterns encountered during the charging phase did not have disastrous consequences, such as the permanent latching into a single state or group of states. As each starting sequence may need to be propagated for very many clock cycles until the full operation is achieved, the task of design checking (and the subsequent tasks of circuit verification and chip testing) would be very costly. The alternative and more attractive approach is to use a predefined, short, control sequence which forces the circuit into a known, functional state. Thus, the apparent inconvenience of requiring a control sequence to start the circuit may in fact be subsumed by a positive desire to apply a control sequence in order to start the circuit in a well defined manner.

Turning to the practical implementation of the control circuit. It is now necessary to define the whole structure of the latch more accurately than hitherto, and this is done for one configuration in figure 3.41. Figure 3.41a shows a block diagram of the salient features, with the control circuit being inserted in the clock distribution, between the complementary clock generator and the clock buffers. In figure 3.41b, a single stage quasi-complementary inverter buffer is assumed as the clock line driver, with the control gates comprising a pair of NAND gates in each line. The first pair is used to disable the internal clock, with the other pair used to apply the external control sequence. Figure 3.41c shows the required control signals, annotated with the function of each element of the pulse train. It is apparent that a more complex configuration (e.g. a two stage buffer) would simply require extra pulses on the second control line.

Using either of the two basic T-latch circuits discussed above, still more care has to be taken before a final implementation. In this analysis presented for the charging sequence of each capacitor, only the "quasi-digital" state of the nodes has been considered. In the actual analogue circuit concerned, the two inputs of the NAND gate do

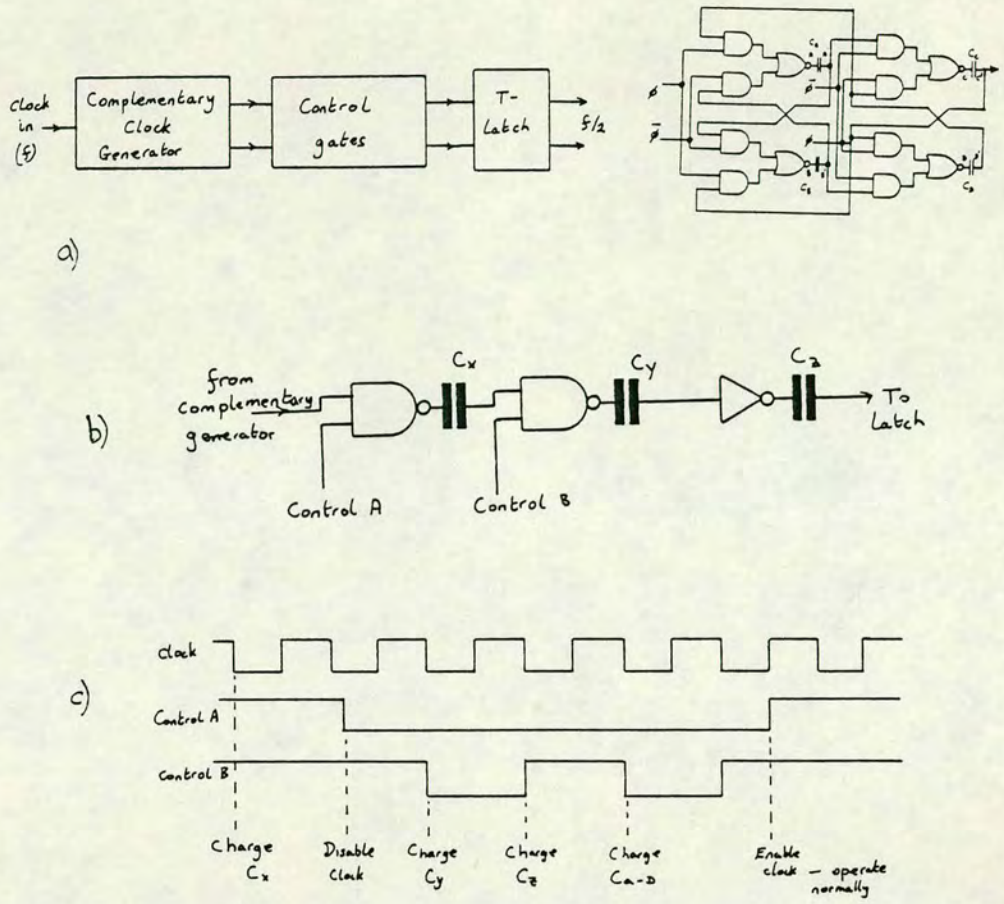


Figure 3.41 Operation of AND-NOR latch in CCL
a) overall diagram (inset = T-latch circuit)
b) control gates in one clock channel
c) charging sequence.

not behave identically. In practice, the upper and lower inputs behave quite distinctly, particularly when considering charging paths. The capacitor on the lower input always has an open charging path (through the source of the transistor), but that on the upper gate (through the source of the upper transistor and the channel of the lower one) is only open whilst the lower transistor is switched on. Looking at the symmetry and charging configuration of the AND-NOR gate, it is clear that the capacitors in the feedback branch of the loop will only be suitably biased whilst the clock is turned off. Thus it is imperative that the clock input is applied to the upper gate of the dual transistor. This condition may be relaxed if a diode is included from the upper input to ground, as shown earlier in figure 3.20.

On the grounds of flexibility, the recommendation has already been made (Chapter 3.2) that the capacitor should be associated with each input rather than with each output. However, in a customised logic element such as the latch, it is quite possible to tailor the capacitors to the exact requirements, and it is thus possible to use one capacitor for both outputs taken from each node in the AND-NOR latch. If this policy is adopted, then it is only mandatory to have one of these loads directly to the lower input, although it remains preferable that both inputs should be to the lower gates. Closer inspection of this particular circuit reveals that the second input of each of the dual gate transistors which share a single capacitor is also common. Thus a neat layout can be achieved by clustering the two dual gate transistors with just two double-sized capacitors as shown in figure 3.42. In this configuration, both clock inputs are naturally to the upper gate, with the signal tracks to the lower.

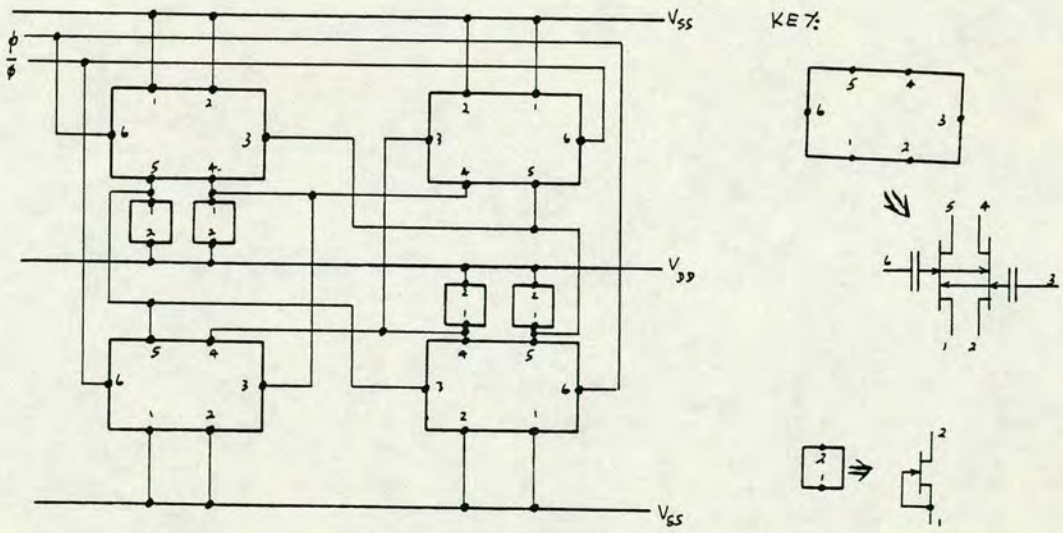


Figure 3.42 Physical layout of AND NOR latch.

Even without the constraints imposed by pre-charging, it is preferable to connect the clocks to the upper inputs, and the switching signals to the lower, in order to effect the fastest switching. Because the clock buffer may be increased in size as necessary, the loading on the clock lines is not critical, whereas that placed on the switching nodes is. If the switching nodes are returned to the lower input, they benefit from the lower input

impedance associated with the cascode stage, to which the circuit then approximates.

Thus far, this chapter has only considered the principles involved in the design of T-latches. The other important variety, the D-latch can be derived from both latches considered by removing the feedback linking the second half latch back to the first. Complementary data are required for both circuit types discussed. The basic switching performance is identical to that presented, except that now the characteristics of the data input circuit must also be considered. If the data are already available in complementary form (e.g. as in the mid-elements of a shift register), then no further input circuit is required. The normal measures of D-latch performance, the set-up and hold times, are then related to the number of gate-delays required in the toggle rate calculations. However, specification of exact figures depends upon a knowledge of the clock buffering, and the delays associated with that. If complementary data are not available, the set-up and hold times are further modified by the circuitry which conditions the data at the inputs. Although both of these factors modify the absolute set-up and hold times, the effect is simply to introduce a relative shift of clock with respect to data. The data clocking rate is still very similar to that of the T-latch.

3.3.2 Complementary clock generator

The previous Section has shown how the exact choice of latch can have a major bearing on the system performance. In this Section, the influence of the clock distribution and driver circuits will be studied. The initial investigation will concentrate on the effects of imbalance in the complementary clock circuit, insofar as it affects the maximum clocking rates of the latch. The potential sources of this imbalance will be reviewed in the latter part of the chapter.

In the analysis of latch cycle times, the two clock phases were assumed to bear an exact 180° phase relationship to each other. In practice, such an ideal is unlikely to occur, and it is instructive to see the exact impact of any deviation from this.

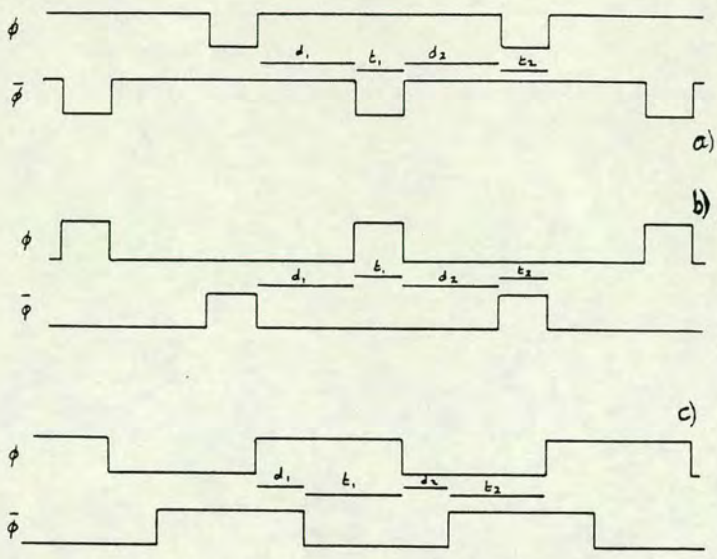


Figure 3.43 Non-ideal phasing of the complementary clock signals:
a) overlap in logic-1,
b) overlap in logic-0
c) phase skew.

In this connection, the two divider circuits are analysed under varying conditions of mis-alignment of the two clock phases. Three cases are of interest, and the appropriate timing diagrams are shown in figure 3.43. In 3.43(a and b), the mark:space ratio of the two clock signals deviates widely from 1:1, and there is necessarily overlap in one of the logic states (logic-1 in a and logic-0 in b). In 3.43c, the mark:space ratio of the clocks is correct, but there is a clock skew (or phase offset) between the two signals. Similar analysis to that used in the previous Chapter shows that there are limits on the allowed errors in phase (d_1 and d_2), and also on the minimum pulse widths (t_1 and t_2). The results of this analysis are reproduced in table 3.4.

In deriving table 3.4, the output logic level during a transition was assumed to change a half propagation delay after the causal change at the input. Taking a 1-0 transition as an example, in practice the node voltage will fall monotonically throughout the whole transition until the new level is reached after a time t. However, any input connected to the node will begin to respond to the change when the output level has crossed through some switching threshold, which is not necessarily the full logic voltage. If a second change takes

Table 3.4 Effect of clock mismatch on circuit performance.

Parameter	AND-NOR	8-NAND	Comment
Configuration a:			
Max d_1, d_2 for recovery	2.5t	2.5t	
Min t_1, t_2 for switching	1t	2t-d	for d<t
" "	1t	1t	for d>t
Min clock period (t_1=t_2=t)	2t+d_1+d_2	4t	for d<t
" "	2t+d_1+d_2	2t+d_1+d_2	for d>t
Configuration b:			
Max d_1, d_2 for recovery	0.5t	indef	
Min t_1, t_2 for switching	1t	2t	
Min clock period (t_1=t_2=t)	2t+d_1+d_2	4t+d_1+d_2	
Configuration c:			
Max d_1 for recovery	2.5t	2.5t	
Max d_2 for recovery	0.5t	indef	
Min t_1, t_2 for switching	1t	2t	
Min clock period (t_1=t_2=t)	2t+d_1+d_2	4t+d_2	for d_1<t
" "	2t+d_1+d_2	3t+d_1+d_2	for d_1>t

place at the input which causes the output to revert to its original state before that threshold is reached, the following stage will not see any change. In this assumption, that threshold voltage is assumed to be crossed after 0.5t. Thus if the clock changes during the first 0.5t no change will be propagated, but a zero will eventually (after the full time t) be recorded at the subsequent input if the clock is cycled after 0.5t. This consideration is especially important for a circuit such as a latch which passes through unstable states, some of which trigger other events.

Comparing the data in Table 3.4 for the two circuits, it is clear that the maximum frequency of the AND-NOR latch is directly affected by any deviation from complementarity of the two clock phases, but that the 8-NAND circuit is not always slowed by such a phase-error. Perhaps more important to note is the sensitivity of the AND-NOR circuit to the situation where both clock signals are simultaneously at logic-0. Where there is doubt about the ability to generate truly complementary signals, the clock drivers should be designed to err towards overlap in the logic-1 state to avoid this problem.

The full analysis clearly shows the presence of a regular oscillation when both clock phases are at logic-1. For both latch types the oscillation period is exactly 8t and this can be a very helpful diagnostic tool, permitting accurate measurement of the value of t. This measurement is much more definite and reproducible than any result obtained by any other means.

Table 3.4 clearly shows the relevance of deriving accurately phased clock signals in order to obtain the maximum circuit operating frequency. There is a definite requirement to avoid clock-overlap of greater than 50% of t when both clocks are at logic-0 when an AND-NOR configuration is being used. This sets a suitable worst case target for all configurations, as a clock skew of 0.5t on both edges causes the AND-NOR divider to suffer a 30% reduction in maximum clock frequency. Such degradation can only be tolerated under extreme circumstances. For the work performed in this study, in which a toggle frequency of 3 GHz was required for an AND-NOR divider, the clock skew had to be limited to a maximum of 60 ps.

It is essential to note that the allowed skew is determined by the maximum speed of the technology and not by the maximum required circuit speed. In slow parts of the circuit, it may be appropriate to choose the 8-NAND divider, simply to gain an extra safety margin for the clock driver circuitry. This is contrary to the option selected purely on the basis of minimising power dissipation in the low speed elements. Perhaps a better alternative is deliberately to slow the transition time of gates within a divider designed to function only at low speed. In practice this may be accomplished by using very narrow transistors, thereby simultaneously conserving power dissipation.

The earliest reported work on high speed GaAs circuits relied upon external phasing of the two clock signals [350]. As a demonstration exercise this was adequate, because suitable analogue techniques were available using tuned delay lines. For any usable system, this is cumbersome and inappropriate, and it is necessary to generate the antiphase signals on-chip. The first reported single clocked dividers used edge-triggered D-types in a NOR gate technology [351],

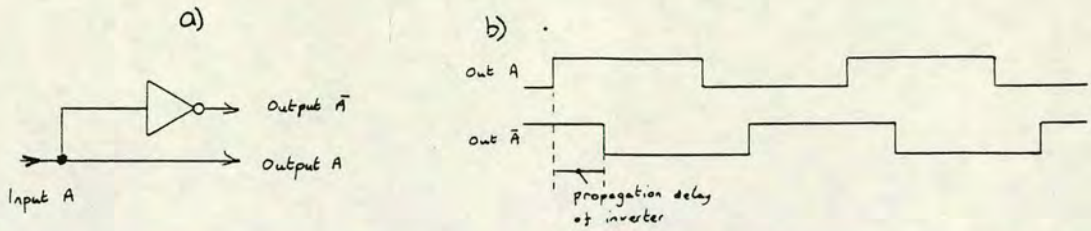


Figure 3.44 Simplest complementary signal generator
a) logic diagram, b) timing diagram.

incurring significant speed penalties from the choice of circuit. However, when compared with the result from a dual-clocked divider operating from internally generated clocks [352], this apparently poor performance proved quite good. It is thus important to trace and minimise the source of errors in the clock phasing, in order to benefit from the greater potential of the dual clocked circuits.

The simplest way of generating anti-phase signals is to pass the input signal through an inverter, taking both the input and output signals into separate buffers, as shown in figure 3.44. The two signals thus generated should have the type of skew suggested by figure 3.43c, with an error equal to the single gate delay. However, using a circuit exactly as drawn, the delay from the single gate could be quite large, because the buffer presents a significant load. The actual skew on the clocks is therefore the value appropriate to an inverter with high fan-out. During the initial phase of this development, this would have incurred a penalty of some 200 ps or so (admittedly, the divider would then only clock at 1 GHz) [353]. Even later, after refinement of the process, achieving less than 60 ps would have been marginal.

In addition to being marginal in the value of skew, the need to design for satisfactory operation over a wide spread in threshold voltage adds to the difficulties. The ratio of the switching currents available from the switch and load transistors of the inverter varies with threshold voltage (V_t). Therefore, the inverter can only be designed with ideal performance for a particular value of V_t. When operated with a threshold voltage different from the design value, the inverter becomes ill-matched, and when driven with a square wave but loaded by a perfect one, the output of the second

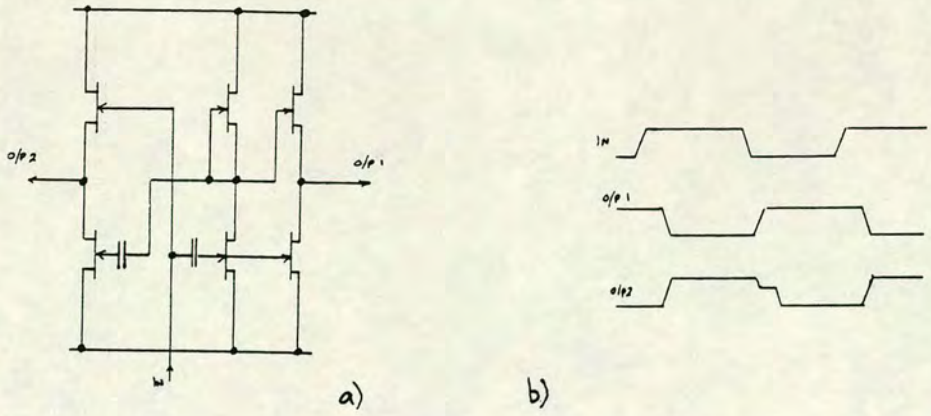


Figure 3.45 Improved complementary signal generator
a) circuit diagram, b) schematic timing diagram.

stage will display a disparity between the widths of mark and space. Adding this extra phase error to that already predicted for the heavily loaded inverter of figure 3.44 simply compounds the problems.

If the circuit has to cope with a wide range in V_t, this is an inevitable consequence, and a better system of clock generation is therefore required. Where phase is important, the effective fan out of the crucial gates must be minimised.

In attempting to improve this situation, the circuit of figure 3.45a was derived. This makes use of the fact that the clock buffers are probably fabricated from quasi-complementary drivers (see Section 3.2), in which two skewed antiphase signals are used to act as active pull-up, and active pull-down. By generating just a single pair of such signals, and feeding these in the opposite configuration into two separate push-pull stages, one may expect to improve on the phasing compared with figure 3.44. The reality is even worse. Schematic switching voltages are also shown in figure 3.45b, and these show clearly that the FET push-pull stage works well when the upper input is delayed slightly with respect to the lower, but not when vice versa. The rising output edge is controlled by both the emitter follower and the common-source stage, but the falling edge is controlled by the common-source stage. The emitter follower only aids this transition once switching has begun. Thus output-1 is

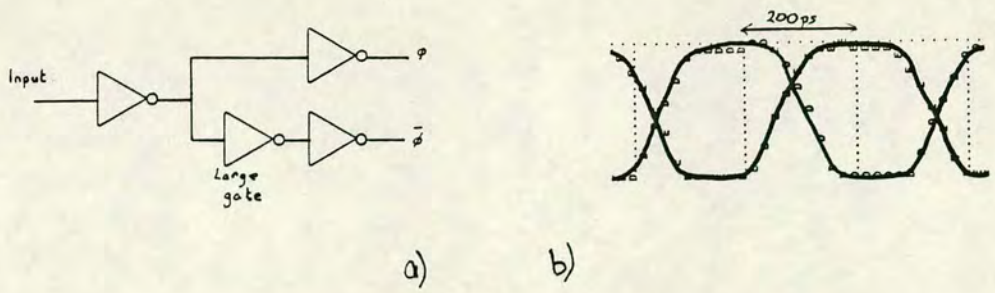


Figure 3.46 Final complementary clock generator
a) logic diagram, b) simulated timing diagram.

activated on both edges by the first input, but output-2 responds to the first edge on the rising output, and the second edge on the falling output. Output-2 thus always suffers from a non-symmetrical response.

An improved design reverts to the use of an inverter to generate the second signal. However, the careful selection of transistor sizes, coupled with an increased amount of buffering between the extra inverter and the clock drivers allows the circuit to be tuned to considerably less than the delay associated with a single, unity fan-out inverter. This configuration is shown in figure 3.46a, and the results of circuit simulation are shown in figure 3.46b. Because the skew can be guaranteed to be less than the 0.5t of the AND-NOR gates, there can be no problem of unwanted oscillation of the divider. The only remaining problem is then to identify how slow the divider is compared with its theoretical upper frequency.

Whilst the latter circuit is generally adequate for the generation of complementary signals suitable for clock driving, complementary signals are also needed for the data inputs of high speed D-type latches. Here, the large absolute delay between the input and output of the complementary clock generator may be intolerable. For this case, the simplest solution of using a large single inverter to produce the second phase appears to be the best, even though this necessarily degrades the performance of the latch.

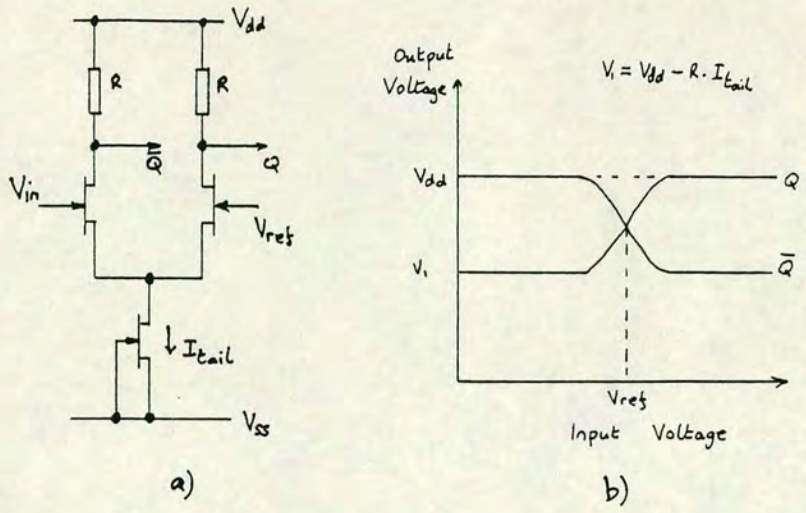


Figure 3.47 Long-tailed pair
a) circuit diagram, b) transfer characteristic.

A further circuit, namely the long-tailed pair (LTP) [354], needs to be considered. This circuit (figure 3.47) lies at the heart of the success of ECL circuits in silicon, and is also the key to SCFL in GaAs [355]. The circuit may be used either in an entirely complementary fashion, or it may be driven with one of the inputs held at a reference level. In the latter mode, a single input can be converted into two anti-phase signals, provided that the reference voltage is held midway between the two logic levels appearing at the single input. The FET version of the long-tailed pair is not as efficient at switching as the bipolar version, but it still works well. However, in a technology relying on saturated current loads rather than resistor loads, the circuit is very sensitive to imbalance [356]. It is also very sensitive to overload, particularly as the internal logic of CCL operates nearly rail to rail and, furthermore, the output is not directly compatible with the succeeding stages. In most applications then, the LTP is impractical. However, there is one particular exception where it may be considered useful: as an interface circuit onto the chip, where the voltages are already incompatible with those required internally. This, along with other interface circuits, will be considered in the following Section.

3.3.3 Input and output buffers

In the design of most integrated circuit families, one of the most difficult aspects is that of interfacing to the IC at the correct, well-defined signal levels. For high speed operation, any new components must guarantee to inter-work with ECL, the existing standard.

This standard is defined around the capabilities of bipolar transistors, most notably the extremely well controlled switching threshold, particularly when used in a long-tailed pair configuration. In conventional operation, ECL uses a 5.2 V power supply, with V_{cc}, the positive rail, as the grounded signal. The logic levels are then defined approximately as -0.9 V and -1.7 V, although the exact levels are temperature dependent. Under a rigorous definition of the interface, the output must be capable of driving directly into a 50 Ohm impedance, terminated to a -2 V power rail. Strictly, then, one could argue that ECL requires two power supplies, this terminating voltage adding the second power rail. This leads to an alternative mode of operation, whereby the output termination is taken to ground (0 V), and the IC is operated between +2 and -3.2 V, the new logic levels being 0.3 to 1.1 V.

In order to interface with ECL, either of these two modes of operation can be targetted at the outset, although a longer term solution would require operation with ECL in its standard configuration. Whichever mode is chosen, the challenge is to convert between the very small amplitude ECL signal, and the large amplitude CCL signal.

Taking a pragmatic approach, in order to demonstrate operation of a CCL circuit at high speed, it is absolutely essential to drive the circuit with all the necessary input signals, but the minimum requirement at the output is simply to observe the signals. This can be done using an oscilloscope, a spectrum analyser or a logic analyser although a high speed version of this latter was not commercially available when the project began. Both of the first two instruments can be used with any signal level above a few mV. Thus, correct interfacing at the output is not an essential objective only

a desirable one. All design effort was therefore concentrated on designing the input amplifier where correct interfacing was seen as essential. However, it still remains to explore the few options available for the output buffer.

Treating the output buffer first. If given the freedom to terminate the output load into any desired DC voltage, the problem is simply one of identifying a means of driving sufficient current into the required output impedance. This impedance must be low, say 50-100 Ohm, otherwise the time constant formed between the load capacitance and the load resistance will preclude observing fast waveforms. Furthermore, matched impedance connectors and transmission media are only available in this range. Without using good "RF practice", and matching the source and load impedances to the transmission line, all data signals will be distorted and corrupted by the pulses reflected from each mismatched node.

In translating this requirement into reality, the practical solution is to use a very wide FET as the output transistor. With the BTRL process, this demands a 150 to 200 μm transistor, offering between 20 and 40 mA of drive current, depending upon the exact threshold voltage of the transistor. This can actually comprise several narrower transistors placed in parallel.

Either an open drain or an open source configuration is acceptable, with the load resistor being provided off-chip. However, whichever circuit is being used, this transistor carries large switching currents and it is essential to separate its common voltage from the equivalent internal rail. Failure to do this has resulted in spurious oscillations in some output buffers used on test structures because of the very severe noise spikes fed back to the internal power bus.

The choice between the two possibilities depends upon the terminating voltage and which of the ECL power arrangements is being used. The internal driving of the output transistor is different in the two cases, and should first be considered.

If an open source is to be used, the gate of the FET can be connected directly to a standard inverter output, although significant gate current may be drawn if the terminating voltage is too low. In this arrangement, the output behaves like a standard source follower.

If an open drain is to be used, either the source of the transistor must be taken significantly positive of the V_{ss} rail (possibly to the 0 or -2 V ECL terminating voltage), or the gate must be capacitively coupled, as in a standard inverter.

Putting all these options together, a compromise solution was adopted, whereby the output transistor was capacitively coupled to the drive stage, and both source and drain were padded out separately (although there is actually no physical distinction between the source and drain connections). In the source follower configuration, the capacitor protects the gate from drawing excessive currents, but still allows a good drive capability at high speed.

Because the output transistor is very wide, it presents a large load capacitance to its driver, which itself needs to have a high drive capacity. A quasi-complementary buffer is therefore recommended as the driver, with the sizes of the two stages of this buffer selected to distribute the load uniformly between the internal circuitry and the output stage. Even this internal buffer presents a fan-out loading of about three on the previous stage.

Turning now to the more severe problem of the input buffer. A number of alternatives must be considered. The amplifier mentioned in Chapter 3.1.3 proved inadequate to the task, having been designed using poor models, in particular neglecting the (at that time) very large source resistance component. In attempting to improve noise immunity by adding hysteresis to the switching threshold, the circuit had become much too dependent upon both the exact threshold voltage and the transistor parasitic resistances.

In considering a new design, the very wide range of threshold voltage likely to be encountered had to be considered. It seemed apparent that the solution to this problem was to use an external reference

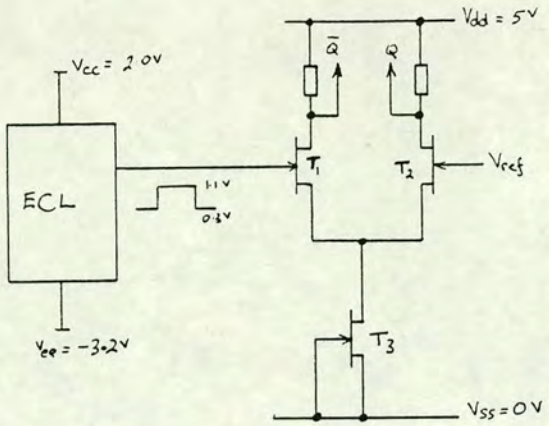


Figure 3.48 Long tailed pair input amplifier interfacing with ECL.

voltage to tune the amplifier threshold, depending upon the exact properties of the device in question. Given the comparison with ECL, this immediately suggests the long-tailed pair as the input stage, as has now been demonstrated in SCFL (although the buffer was designed before its invention).

If this is adopted as the input amplifier, it is necessary to operate the ECL circuit between +2 and -3.2 V, whilst operating the CCL from 0 to +5 V, as shown in figure 3.48. The input voltage then lies between 0.3 and 1.1 V, giving a good chance of operation over a wide range of threshold voltage. If the FET model presented earlier in this chapter is adopted, the common emitter node should sit above V_{ss}-V_t, as this is the saturation voltage of the current source. By appropriate choice of the reference voltage, the input transistor T1 can be switched between fully-off and partially-on states. Unfortunately, this model predicts that the avilable output amplitude reduces as the threshold voltage becomes more negative, directly the opposite of the requirement. With the hind-sight of the new FET model (Chapter 4), which suggests a constant saturation voltage, whose value is much reduced over that of the earlier model, this problem does not arise.

Even though the new model implies that the circuit is quite viable, two further problems arise. Firstly, the switching transistor is operating in its lowest transconductance region, giving relatively

poor rise and fall times. Some compensation can be made by using very large transistors, but clearly this also increases the parasitic capacitances. The second problem is more severe. In the BTRL process, the resistor is not available, but is instead replaced by an FET with gate and source connected together. If figure 3.48 is modified accordingly, both the load and the tail transistors are "constant current" sources, and the circuit operation then becomes critically dependent upon the output conductance of the FETs. Since this is probably the least well controlled or understood of the FET parameters, this situation cannot be allowed to obtain.

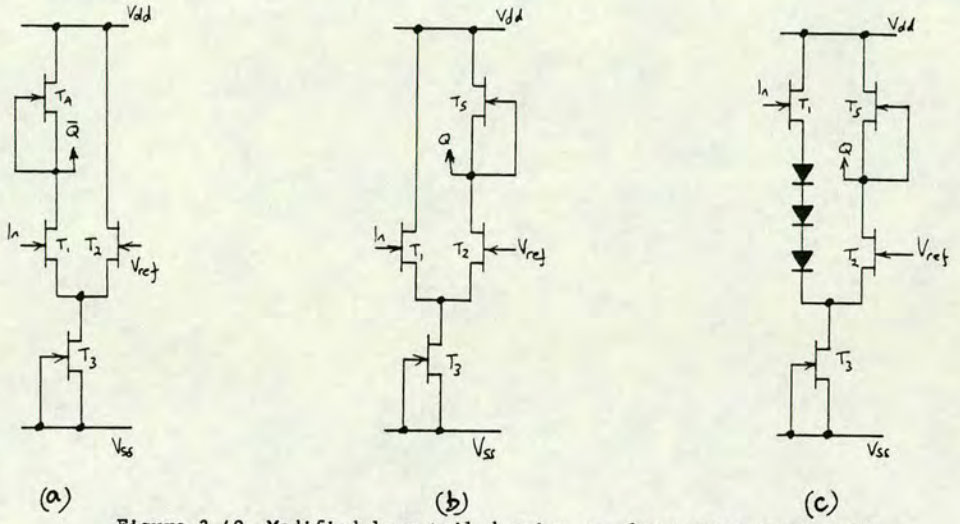


Figure 3.49 Modified long-tailed pairs as alternative input buffers.

A modified long-tailed pair, as shown in figure 3.49 goes part way towards alleviating the second of these problems. The output-low voltage is related to the common-emitter node voltage. Since this voltage is lower when T1 is the conducting transistor, option (a) offers a larger output amplitude than does option (b). As the essence of the input stage is to obtain a large amplitude, option (a) is to be preferred. This is somewhat unfortunate, because option (b) can readily be converted into the attractive circuit of figure 3.49(c), in which the input voltage-level shifting offers the potential for direct coupling to standard mode ECL.

This final revision of the long-tailed pair leads directly into the alternative input configuration. The chain comprising T1, the diodes and T3 is none other than a source follower level-shifter as used in

BFL. If this is removed, then T2 and T5 must form a viable amplifier in its own right, and by re-drawing in the more familiar configuration (figure 3.50), it is recognisable as a common-gate amplifier.

Taking into account the practical objections to the long-tailed pair circuit in the available process technology, this simplest of all configurations is the wisest choice for the input buffer. The shear simplicity of it offers a good insurance policy against failure. This is most important, because failure of any of the input nodes, particularly of the clock input, would prevent the extraction of any useful information about circuit performance.

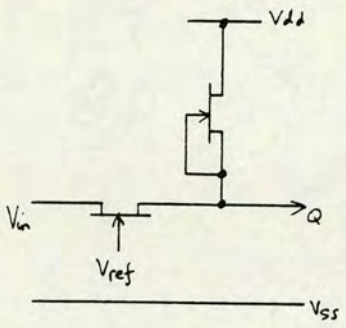


Figure 3.50 Common gate input amplifier.

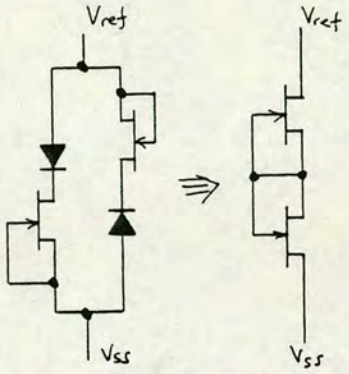


Figure 3.51 Input protection circuit.

A final nicety of interface design is to provide automatic bias and overload protection. This is most appropriate for the common gate input buffer, where an excessive voltage placed on the gate could cause damage to the transistor. The circuit of figure 3.51 has been added to all reference nodes to give some measure of over-voltage protection. Because the transistors' source and drain electrodes are interchangeable, the input signal is clamped between +/- V_{bi}. This should help to prevent damage arising from electrostatic discharge (ESD), as well as providing a lower impedance path to any "overdrive" current than is available through the input transistor. Unfortunately, this circuit presents a relatively high impedance when there is no voltage across it. The degree of auto-biasing is thus somewhat limited.

Chapter 4. Modelling

Earlier chapters have introduced GaAs and have described the operation of simple circuits in qualitative terms. In addition, the elementary models available at the outset of this research programme have been introduced. In order to develop beyond this level of understanding and to try and predict actual performance it is necessary to introduce more sophisticated and representative models. This chapter details the developments in modelling undertaken as a part of this study.

The modelling falls into three categories: diodes, MESFETs and interconnections. These conveniently form the topics of the three sections, although no one of them can be treated in complete isolation from the others. Thus, for example, the final transistor models rely on those of both diodes and interconnections.

4.1 Schottky diode modelling

There are two distinct applications for Schottky diodes in this work, depending upon the applied bias. Under forward bias, the diodes are used as elements for voltage level shifting, and under reverse bias they are used as capacitors. Thus the modelling falls into these two categories, but before embarking in detail, it is useful to discuss the physical operation of the junction.

4.1.1 Physics of Schottky junctions

It is beyond the scope of this work to delve deeply into the fundamental physics of the operation of Schottky diodes. However, it is important to grasp the basic elements of operation in order to understand the relevance of the derived models. Only n-type semiconductors are appropriate to this study, so most effort will be concentrated here.

The Schottky diode is formed from a metal and semiconductor in intimate contact [357]. To understand the behaviour of the interface it is first necessary to establish some definitions. Referring to figure 4.1, the metal is characterised by a work-function, which is a

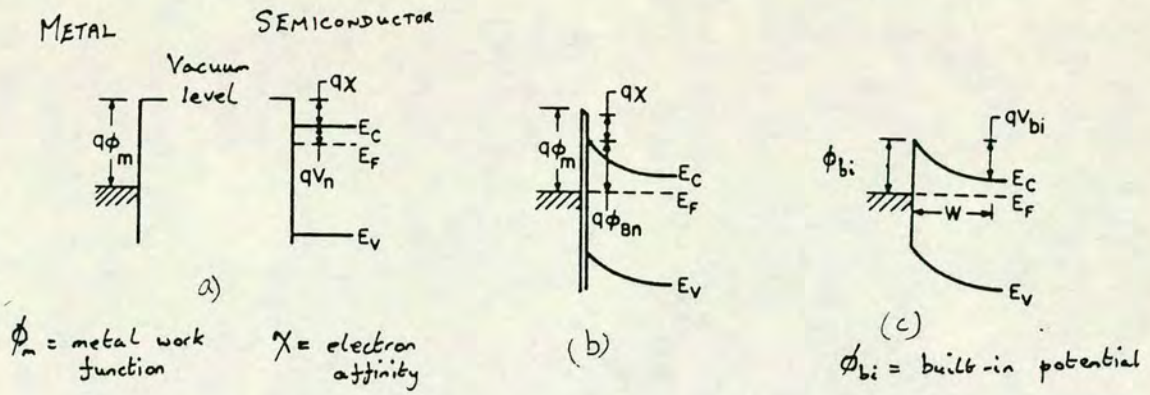


Figure 4.1 Band structure and definitions for a metal and semiconductor interface a) isolated, b) close proximity, c) in contact.

measure of the energy required to transfer an electron from within the metal out to infinity [358]. The Fermi-level [359] is a measure of the highest energy of the electrons in the material at equilibrium. For a metal, this is coincident with the work-function. In a semiconductor, the electron affinity [360] is the equivalent measure to the work-function, but it measures the energy between the bottom of the conduction band and the vacuum level. In a semiconductor at room temperature, the Fermi-level lies close to the energy of the dopant [361]. Thus, for an n-type semiconductor, it lies just beneath the donor level, for a p-type, just above the acceptor level, whereas in intrinsic material it lies at the centre of the band-gap.

As the two materials are brought into proximity (figs 4.1b and 4.1c), the conduction and valence bands in the semiconductor bend to take up the difference between the Fermi-levels of the two materials. The total amount of band-bending is related to the work-function, the electron affinity and the depth of the semiconductor Fermi-level. The amount of bending is variously termed the barrier height or the built-in potential [362]. Figure 4.1b shows the two materials close together, with contact actually established in figure 4.1c. It is clear that the built-in potential for the two cases is different. In reality, a junction is unlikely to be as perfect as that implied in figure 4.1c, because there may be some interfacial dielectric or

charge which helps to absorb some of the total voltage drop, giving a lower final built-in potential, similar to that of figure 4.1b [363].

If a voltage is applied across the interface, the bending of the bands in the semiconductor produces a barrier, inhibiting free electron flow across the junction (see figure 4.2) [364]. When the metal is made sufficiently positive, electrons can be injected from the semiconductor into the metal, forming a current flow, in the so-called forward direction. In contrast, as the metal is made negative, the electrons within the metal see a significant barrier to flow into the semiconductor, and there is no current. This is the reverse-biased mode. As the reverse-bias increases, the barrier grows in height, but effectively reduces in width. Eventually, under sufficient bias, one of two mechanisms will come into play to contribute to an increased current. The barrier may become sufficiently narrow for the electrons to tunnel through [365], or the electric field may be sufficiently large that a few energetic carriers cross the barrier and collide with the atoms in the semiconductor with sufficient force to form an electron-hole pair, which in turn contribute further carriers to form an avalanche and a rapidly increasing current in the breakdown condition [366].

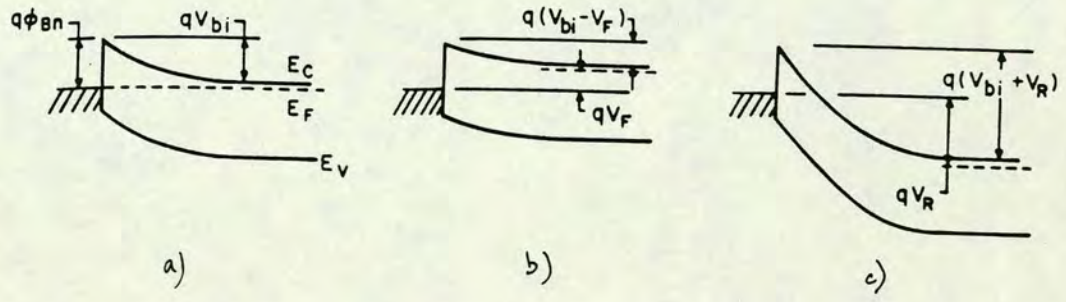


Figure 4.2 Schottky diode under bias
a) unbiased, b) forward biased, c) reverse biased.

Under forward-biased operation, there are competing mechanisms leading to the electron flow. Thermally excited electrons may be emitted over the top of the barrier according to the thermionic emission theory [367]. This theory assumes that the height of the

barrier alone is sufficient to control the current flow, with each point within the semiconductor in local equilibrium. According to the diffusion theory [368], the current is limited by the movement of electrons within the depletion layer, and so the exact shape of the doping profile is also important. These two effects are combined in the thermionic emission-diffusion theory [369]. In addition, tunneling through the barrier in the forward direction may contribute significantly to the total current [370], either when the doping level is high (narrow barrier), or the temperature low (diffusion and thermal excitation both reduced).

All these mechanisms give rise to different mathematical expressions relating the current (I) and voltage (V) and it is difficult to balance them all in a single equation. This is especially so when considering non-ideal interfaces and other detailed effects (e.g. image-force lowering [371]). However, from a practical viewpoint, the I-V behaviour is very well matched by the equation [372]:

$$I = I_o \{ \exp \frac{q.V}{nkT} - 1 \}$$

where n, the ideality factor, is an arbitrary fitting parameter close to unity. Often, the (-1) term is ignored because it is negligible, except at very low voltage.

This expression is valid, because the most dominant mechanisms lead to the exponential dependence, small perturbations from which are accounted for by the ideality factor. Generally n should be greater than unity, typically 1 to 1.1, with a value greater than 1.3 indicating a very poor device, above which the single expression ceases to be adequately representative. However, as such poor devices will not be used successfully in real circuits, a valid model is not required outside the usable bounds of the above equation.

This expression also defines the current under reverse bias, although there is generally an alternative leakage mechanism which dominates, for example by recombination at interface states [373].

Under reverse bias, the current is generally only of academic interest, but the capacitance is of prime importance. As in a p-n junction the capacitance arises because of the charge stored at the depleted interface. The depletion region is entirely supported in the semiconductor because the metal acts like a very heavily doped material, with the change in electric field supported over a very short range. The capacitance is thus identical to that expected from a one sided abrupt p-n junction [374], except that there is only a single kT/q voltage correction to take account of the majority carrier distribution [375], instead of two. Thus:

$$C = C_o \cdot (V_{bi} - V - kT/q)^{0.5}$$

where C_o is related to the doping level and the usual physical constants. Generally, the effect of the thermal voltage (kT/q) is sufficiently small to be neglected.

4.1.2 Diode current

The theory presented in the previous section is perfectly adequate for describing the current voltage behaviour of a practical diode provided that a resistance is added in series with the ideal model. In addition, one must recognise that the saturation current I_o is dependent upon the diode area and also has a square-law temperature dependence [376] The ideality factor is assumed to be independent of temperature, which is generally true whilst the value remains within the bounds already specified ($1 < n < 1.1$).

As indicated in Chapter 3, the other major consideration in the development of a usable diode model is the stability of numerical convergence [377]. Because of the very rapid change in current with voltage, the diode equation does not lend itself to stability, and it is necessary to "cheat" the model by imposing a limitation on the diode slope characteristic. The details presented in Chapter 3 fully describe the necessary alterations.

Also of interest is the speed of computation of the necessary equations, because a circuit will contain very many diodes, and each equation will be recalculated probably many thousands of times in a simulation. As conditional tests and exponentiation are both very

time consuming, it is useful to structure the expressions for the diode current very carefully. It is also necessary to trap errors which might arise (e.g. a value whose exponent is required should be checked to prevent over or underflow of the result). When using the circuit simulator, ASTAP [378], only the I-V equation needs to be specified, with the derivative being calculated numerically. In an alternative simulator, ASTEC [379], the first derivative must be expressly specified. The optimum coding of the diode equations is thus different for the two packages, as the ASTEC version should be designed for minimum overall effort in the generation of both current and slope, not just for the greatest efficiency in calculating the current alone.

Beyond this very simple level of modelling (necessarily simple to obtain an expression suitable for very quick calculation), there is advantage in studying more complex effects. This is particularly so for optimising the physical structure of diodes used for voltage-level shifting elements, where a large voltage drop must be combined with high speed operation [380]. There is clearly a trade-off between the extra voltage dropped in series resistance, and the longer time constant of such a device. Detailed studies lead to the understanding which helps in determining a suitable geometry, even when the better models are too complex to be used in large circuit simulations.

In seeking this greater physical understanding, the fundamental problem is therefore to ascertain what happens to a Schottky diode under high injection. In the p-n junction, there is a fundamental change in the current equations, simply because effects ignored at low currents now become important [381]. The most significant of these effects is that the numbers of minority carriers injected across the junction now become of similar magnitude to those of the majority carriers and their effects must be included. The net effect of this is that, after extracting the effect of resistance, the slope of the log(I) vs V curve reduces with voltage, settling at half its original value. In the Schottky diode, there is no such mechanism, as the device is purely a majority carrier device. However, the flow of a large current through the "resistive" layer of the diode may

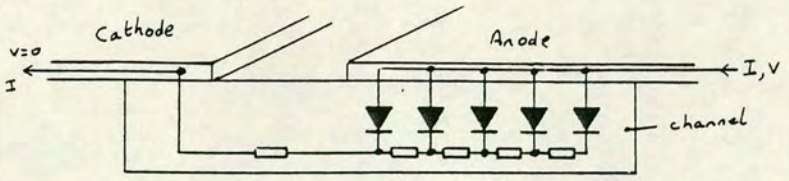


Figure 4.3 Distributed model of a diode.

result in conductivity modulation [382], one of the other second order effects. If this is neglected, there is still an interesting deviation from the simple model presented earlier.

The multi-diode model of figure 4.3 should be used to study the high current case. As the current density through the diode increases, the voltage dropped across each resistor element also increases, debiasing those diodes remote from the contact. Qualitatively therefore, the current will be concentrated more and more to the leading edge of the diode as the voltage increases. Since the effective length of the diode is decreased, the effective series resistance will also be reduced, and a large area diode will behave quite differently from expectation. There is an analytic solution to this model, if the effect of conductivity modulation is ignored, and this is given in Appendix 1. The solution is expressed as a pair of simultaneous transcendental equations in which ψ is the arbitrary parameter to eliminate:

$$I = 2w/a \; . \; nkT/q \; . \; 1/R_o \; . \; \psi \; . \; \tan(\psi)$$

$$\cos(\psi) = \frac{\psi}{a} \left\{ \frac{nkT}{q.R_o.I_o} \right\}^{0.5} . \exp -\{ \frac{q.V}{2.nkT} \}$$

and numerical simulation is required to plot the results. However, limiting cases can be established analytically. When ψ is small, the equations simplify to:

$$I = I_o \; . \; w.a \; . \exp\{ \frac{q.V}{nkT} \}$$

which is the normal (simplified) diode equation. If the resistivity is 500 Ohm/sq, and the diode has a width and length of 10 and 5 μm respectively, this criterion holds for voltages below 0.5 V, a much lower break point than expected. Using the same parameters, an upper limit can be established in which the I-V relationship changes to:

$$I = \left\{ \frac{I_o \cdot 2\,nkT}{R_o \cdot q} \right\}^{0.5} \cdot w \cdot \exp\left(\frac{q.V}{2\,nkT} \right)$$

Using the same parameters as before, this holds for voltages in excess of 0.8 V. Note that the effect on the log(I) vs V slope (reduced to half its original value) is the same as high injection in a p-n diode, even though the mechanism is totally different.
Figure 4.4 compares the current from this distributed model with that from various values of fixed resistor, after adding in the effect of lateral separation between the anode and cathode. The distributed diode may be replaced by an ideal diode with a non-linear resistor in series. The effective series resistance of one such diode is plotted in figure 4.5. Unfortunately, attempts to reproduce this curve experimentally were hampered by the excessive heating of the diode during measurement, as no pulse measurement technique was available. However other workers have observed similar behaviour at high current, without linking it to the cause [383].

Figure 4.6 shows the predicted I-V characteristics for a number of diodes of increasing anode length. Because of the current crowding, the voltage drop at high current is relatively insensitive to the length. This is of tremendous benefit in patterning circuits in which the diode drop is quite critical (e.g. in CDFL circuits as described in Chapter 6.2), because the inevitable variations in alignment between the Schottky metal and the underlying cathode have little effect on performance. Figure 4.7 shows two diode configurations, one whose area is independent of alignment, and the other which is registration dependent. The latter is physically much smaller, and it is useful to be able to use the latter knowing that the consequence is small.

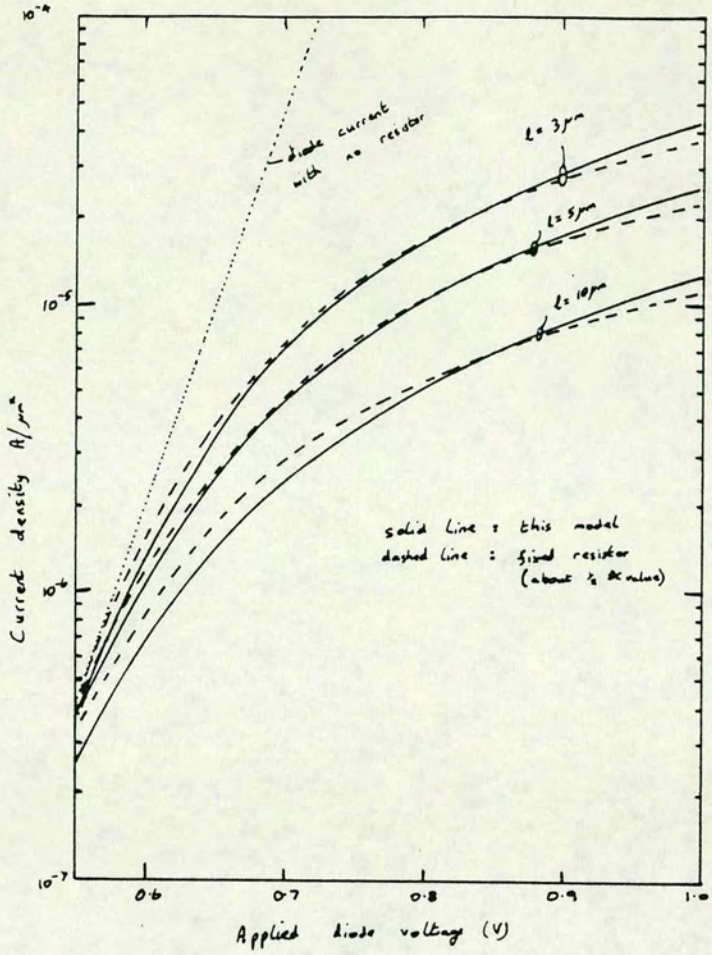


Figure 4.4 I-V characteristics from a distributed diode model.

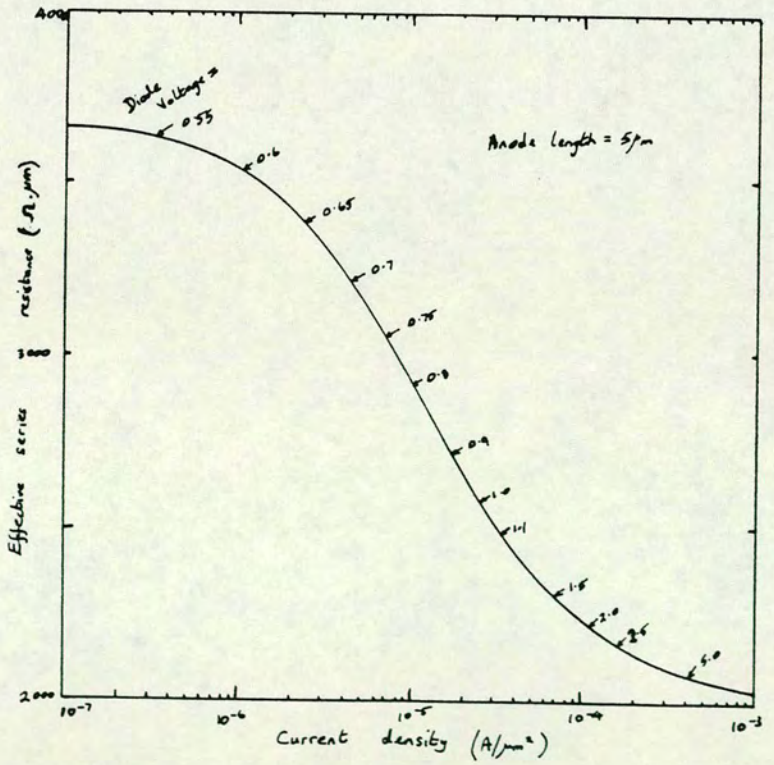


Figure 4.5 Effective series resistance of a distributed diode.

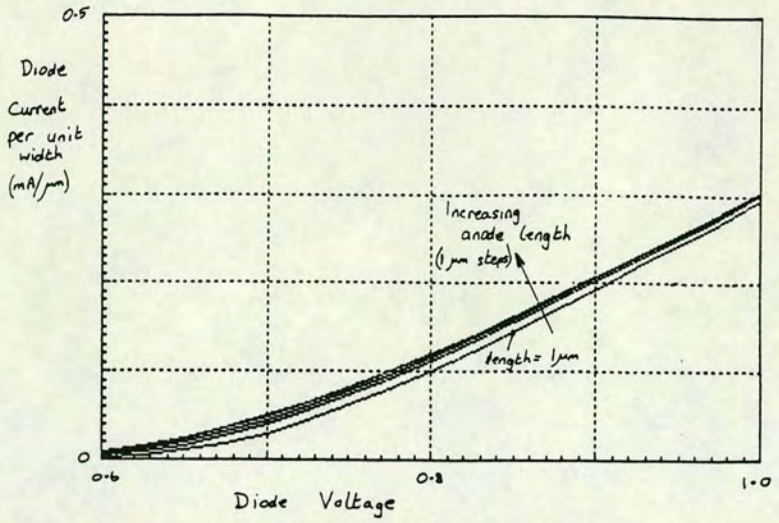


Figure 4.6 I-V relationship of a distributed diode operated at high current density.

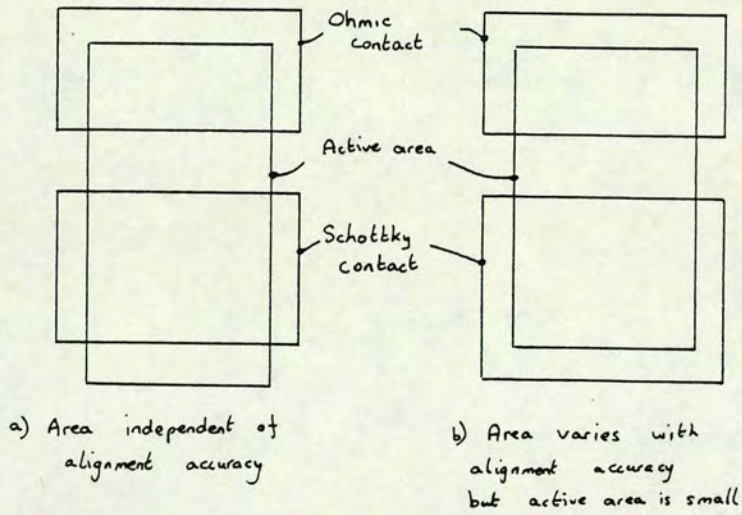


Figure 4.7 Physical construction of Schottky diodes.

As with most detailed modelling, the insight gained is of much greater value than the more accurate simulations made possible. The new model is not simple to encode, and is thus not suitable for circuit studies. The shape of figure 4.5 might suggest a slightly improved model using a resistor whose value falls proportional to the increasing current, provided that the diode is only used over a small current range. A more detailed model has also been suggested [384]. However, if this modelling work is to be used quantitatively, further experimental work is necessary. Diodes with different geometry should be measured carefully using short voltage pulses to avoid heating effects. Test equipment is now available specifically for this type of measurement [385].

4.1.3 Diode capacitance

The modelling of the diode capacitance falls into two distinct phases. In the early work, large periphery digitated diodes were used to increase the coupling efficiency of the diode under high applied reverse voltage. This was to try to overcome the full depletion of the shallow doped layer designed primarily for transistors (see Section 3.2.4). When this proved an inappropriate technique, the technology was altered to allow deep diodes to be fabricated, and the modelling emphasis changed somewhat.

Despite this ordering of events, it is more appropriate to introduce the two phases of modelling in reverse order, as the very deep diode is actually easier to present, with the shallow diode being a modification of this model.

As with the equations governing diode current, the capacitance of the diode is adequately modelled over a wide range of applied voltage by the standard expression:

$$C = C_o \cdot (V_{bi} - V)^{-0.5}$$

Strictly, this expression is only valid for a region of constant carrier concentration, and the factor C_o is a function of this level. Figure 4.8 shows the voltage dependence of the capacitance for a range of doping levels. The theoretical capacitance for the deep implanted diodes is shown in figure 4.9: the implant being performed through a nitride layer, and the Schottky diode being subsequently recessed to the approximate depth shown, leaving the lower part of the profile in the diode. Figure 4.8 also compares the theoretical capacitance profile of this region with that measured on a large area diode. Note that the measured device has a significant leakage current above about 4.5 V, and that the corresponding phase angle of the device prevents measurement beyond this voltage.

Despite the non-uniform doping profile, the diode capacitance is matched well if a constant doping level is assumed, albeit that this level is double that expected. The reason for this has not been established, despite the fact that the measured value was

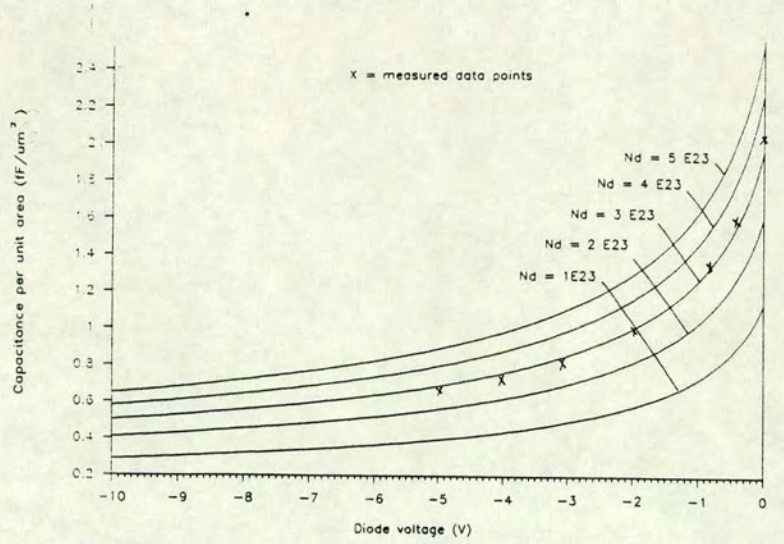


Figure 4.8 C-V relationship for plane layers of constant doping.

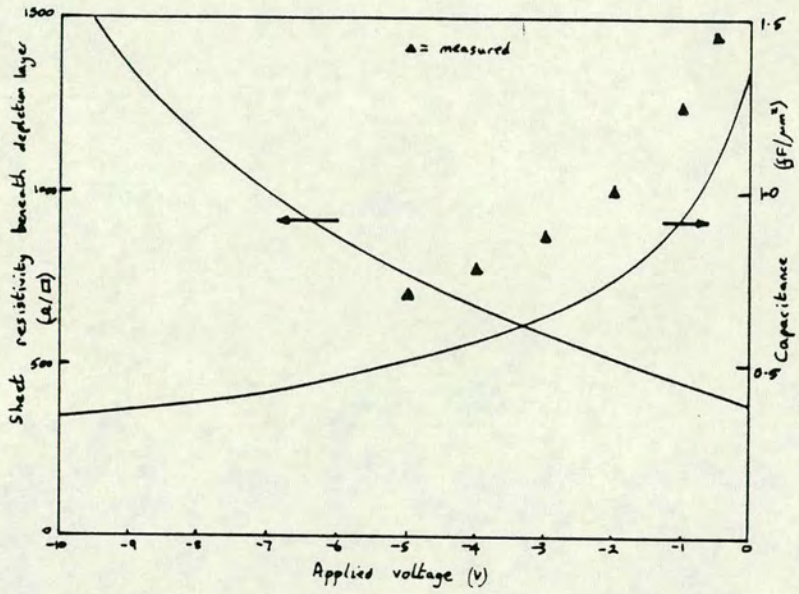


Figure 4.9 Theoretical capacitance and layer resistivity for a deep implant into semi-insulating GaAs.

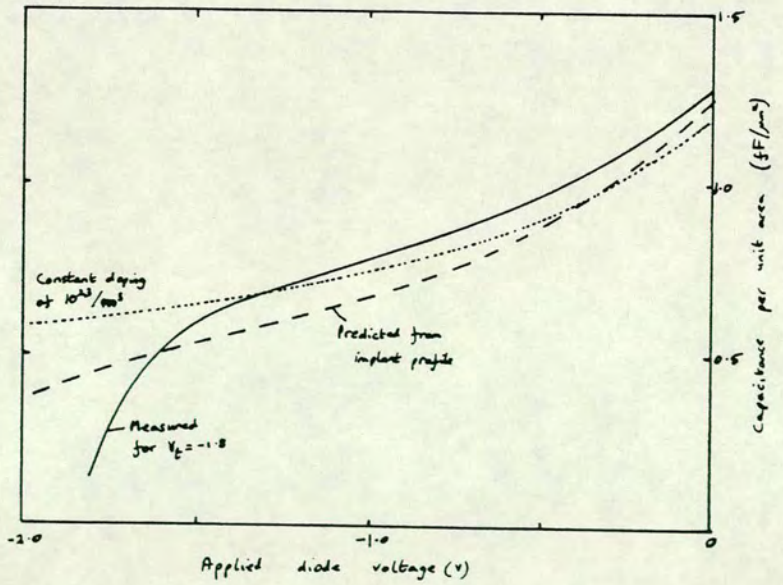


Figure 4.10 C-V of a shallow layer.

consistently high. It is possible that there is some additional charge storage at the interface, although this would have to be at a very high level. In view of the consistency of this measured result, and the model simplicity that it allows, the high constant doping figure has been used in the design work. The first priority is to obtain models which represent the real world, and which can be used in circuit design. Understanding the origin of the models, whilst helpful, must remain secondary.

There are two further complicating issues: what happens to the capacitance once the layer punches through, and what happens under forward bias?

Turning firstly to the case of punchthrough. This arises because the doped region at the surface is of finite thickness, and at a certain voltage (the threshold voltage of the layer) the depletion region extends through the layer into the underlying substrate. Clearly, in a detailed model which takes accurate account of the doping level variation with depth, punchthrough is automatically treated when the doping level falls rapidly to zero. In the more straightforward model adopted here, the punchthrough condition needs to be forced. Figure 4.10 shows the measured capacitance of a shallow layer, showing the dramatic fall in capacitance near the threshold voltage of the layer. The capacitance of a fixed doping level is included. Comparison of the two plots shows that the capacitance roll-off occurs over the last 10% of the applied voltage range. By applying an exponential tail to the standard capacitance model, the desired effect is achieved. Thus:

$$C = \frac{C_o}{(V_{bi} - V)^{0.5}} \cdot \{ 1 - \exp\left(\frac{k_1 \cdot (V_t - V)}{V_{bi} - V_t} \right) \}$$

The value of k_1 is obtained by assuming that the roll-off factor has introduced a 10% change in capacitance when 10% of the voltage remains. Thus a value of 22 should be used for k_1. To prevent the capacitance becoming negative, the exponential cutoff term should only be applied for voltages above the threshold.

The most positive limit of the capacitor equation is in fact more problematic than the negative limit. Clearly the equation has a pole when the input voltage equals the built-in potential. Like the forward biased current, this causes numerical problems. In addition though, there is a further apparent anomaly, because a real diode can actually be forced hard into forward bias, such that the voltage drop exceeds the built-in potential. Accurate modelling of this is then much more than applying the simple numerical confidence trick used with the diode current equations.

The numerical problem is solved quite simply, as before, by attaching a tangent to the curve and using this for the forward bias condition. The actual position of attachment is then open to discussion. For simplicity, it could be added at the zero cross-over, simplifying the equations, or it could be added at any other voltage. Figure 4.11 shows the effect of choosing various break points. Apparently a very large error can arise from an injudicious choice, but this is a very misleading picture as will be demonstrated.

The question of how to treat capacitance under forward conduction only arises as a problem when ignoring the effect of the diode series resistance coupled with the onset of current flow. The additional voltage drop caused by the current flow through the resistor has been included in figure 4.12, and the apparent capacitance error is much reduced, and larger voltages are now allowed. Even figure 4.12 exaggerates the effect of the capacitance, by ignoring the charge injected due to current flow. Takada [386] has shown that there is an overall effective capacitance which may be defined from the derivative of the instantaneous charge with respect to voltage. This method has been adopted in the derivation of the curves of figure 4.13, which show the effective capacitance of the diode, using the break point of the forward-biased equation as a parameter. It can be seen that, once all the parameters are accounted for, the error in simulation caused by selecting a low break-point is much less than expected.

Note that the techniques described to correct the capacitance both beyond threshold and under forward bias are purely arbitrary. Since

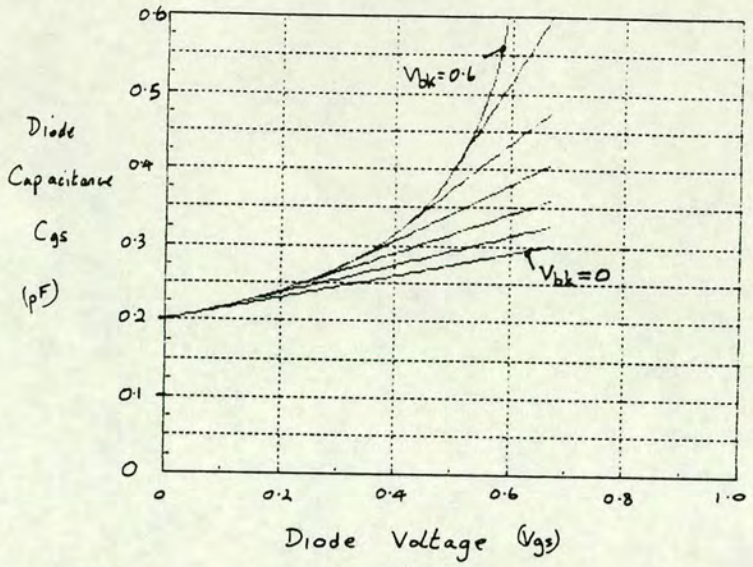


Figure 4.11 Capacitance of a Schottky junction under forward bias. Area = 100 um^2. Tangent applied at V=0 to V=0.6 V, in 0.1 V increments.

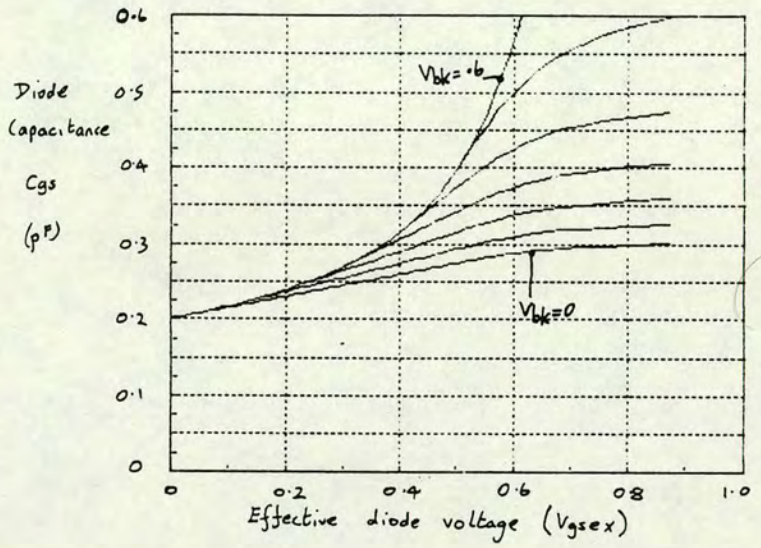


Figure 4.12 Figure 4.11 replotted using the effective diode voltage, allowing for the current flowing in the extrinsic resistor.

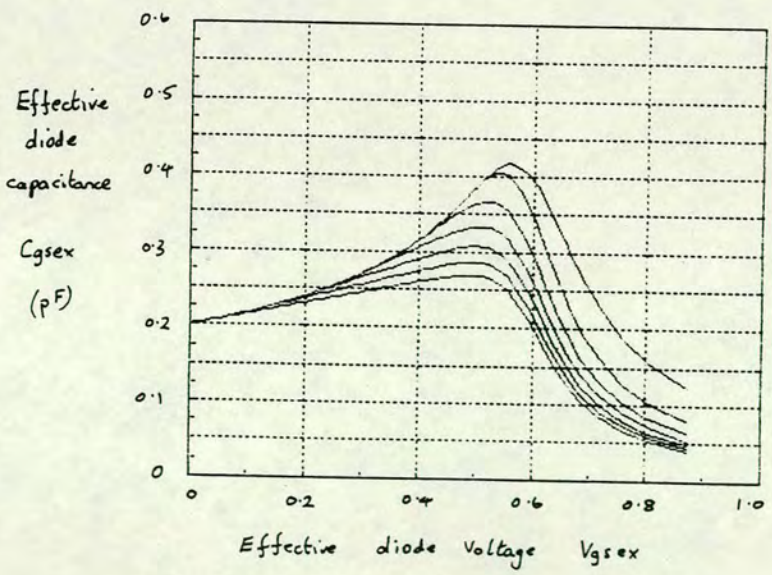


Figure 4.13 Data of figure 4.12 replotted using the effective capacitance, derived from dQ/dV.

the transient circuit simulation in which these models are used employs an integral charge calculation, the validity of the capacitor model can be verified by integrating the total charge in the depletion region. The theoretical value can then be contrasted with the expected value obtained by integrating the total donor concentration within the layer. This is a very useful technique, especially as the particular case of interest is often the Schottky capacitance of the transistor whilst it switches from fully-off to fully-on. In this switching action, the entire charge from the depletion region is indeed swept out.

Figure 4.14 shows that the effect of the forward-biased correction is to produce an offset in charge, but since only the change in charge is used, this offset is irrelevant. In figure 4.15, the percentage error in the change in charge is plotted against the external voltage, showing that the error of using this model should be less than 10%. Since a real device does not have a constant doping level, it is pointless trying to improve the fit to theory still further, when this would be at the expense of numerical ease.

Thus far, only a plane depletion region has been considered. Below threshold, the depletion region extends beyond the doped layer, and additional capacitance only comes from the periphery of the device. Consideration of the peripheral effects are thus important.

When the diode is not punched through, the depletion layer extends sideways at the periphery, as shown diagrammatically in figure 4.16. Conventional approaches suggest that the depletion layer be treated as a quadrant emanating from the metal corner [387]. The predicted capacitance is then independent of voltage because the extra charge stored grows in direct proportion to the increasing separation of that new charge from the metal. The value is given by:

$$C_p = 0.5\pi . \epsilon\epsilon_o \qquad \text{per unit length,}$$

where ϵ is the dielectric constant of the GaAs, and ϵ_o the permittivity of free space.

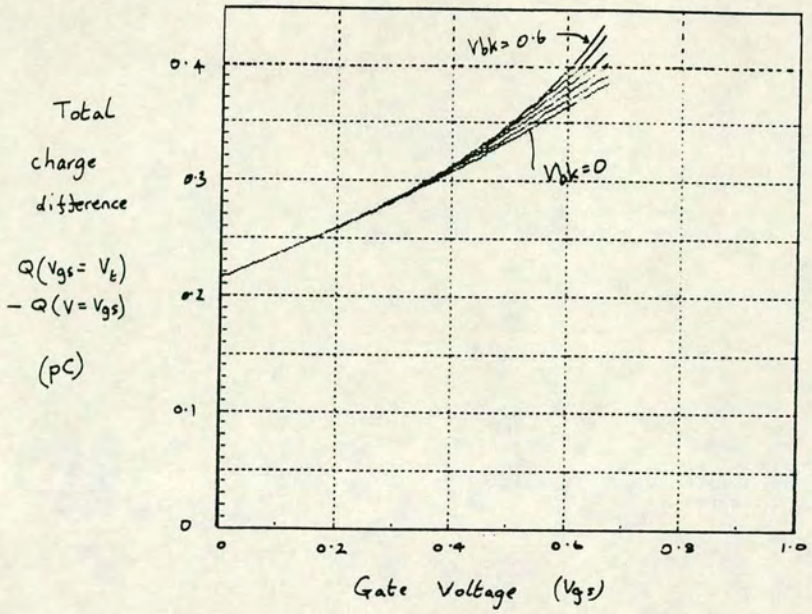


Figure 4.14 Effect of breakpoint used in model of capacitor on total charge switched from on to off.

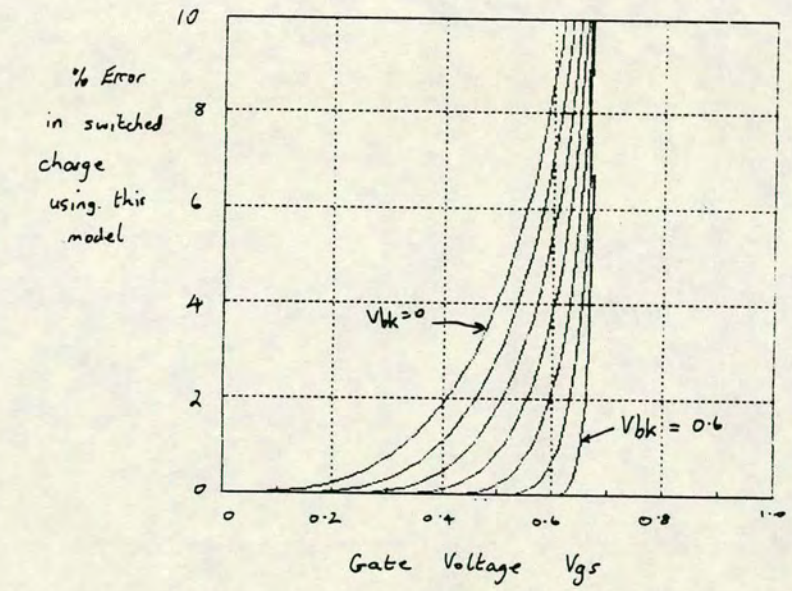


Figure 4.15 Effective error in the charge required to switch a transistor off according to this model of capacitance.

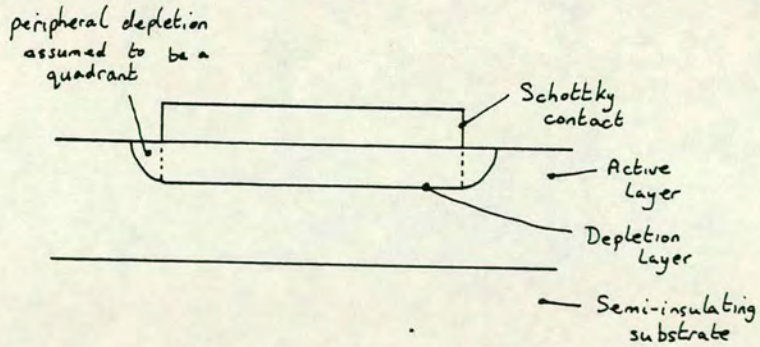


Figure 4.16 Geometry conventionally assumed for depletion layer periphery.

Although this model is widely applied, it is inconsistent. The flux across the interface between the dielectric and the semiconductor must be continuous [388], yet this shape of depletion layer implies a surface discontinuity in flux. Numerical simulations [389] indicate that the correct shape is more appropriately as shown in figure 4.17a. This can be conveniently approximated by two quadrants (4.17b). The net area within the periphery is now reduced from $\pi/4\ r^2$ to $1/2\ r^2$, and the charge therefore shows a similar reduction. Since the voltage dependence of the radius is unaffected by the geometry, the capacitance is reduced in the same ratio, now being given simply by $\epsilon\epsilon_o$ per unit length, a factor of 1.57 lower than conventional estimates.

Once the reverse voltage applied to the diode exceeds the threshold voltage, the depletion region extends through the layer and the lateral depletion region is modified accordingly. Based on the foregoing, there is a number of alternative shapes which could be assumed for the depletion layer edge under these circumstances, as shown in figure 4.18. The exact formulae for the appropriate capacitances of these three shapes are all complex, and very much against the spirit of adopting simple equations, especially as the shape used is already an approximation. Figure 4.19 indicates the change in capacitance with voltage for these "exact" solutions, and shows that the approximation:

$$C = \epsilon . \epsilon_o . \left(\frac{V_{bi} - V_t}{V_{bi} - V} \right)^{0.5}$$

may be used as an adequate representation of the capacitance per unit length.

The complete model for the capacitance can thus be expressed as:
linear with voltage when under forward bias,
continuous slope at zero bias,
inverse square root law under reverse bias,
an exponential tail-off near threshold,
periphery only capacitance below threshold.

Despite all these contingencies built into the model, this is only a starting point for the successful simulation of diode capacitance. At the beginning of the programme, digitated diodes with a very large

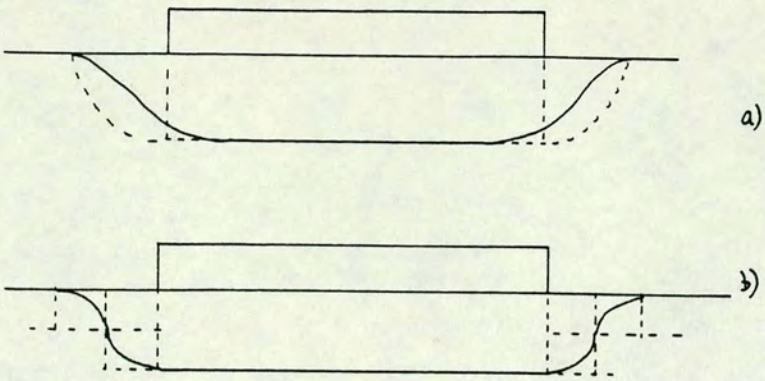


Figure 4.17 Improved modelling of shape of depletion periphery: a) given by [389], b) approximated to (a).

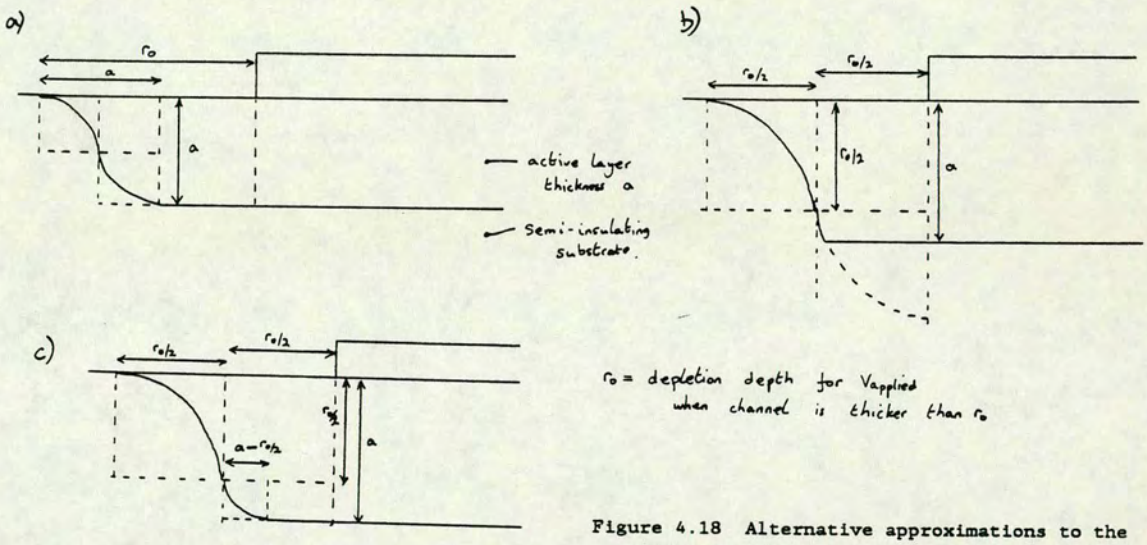


Figure 4.18 Alternative approximations to the peripheral depletion shapes for a punched-through layer ($V<V_t$).

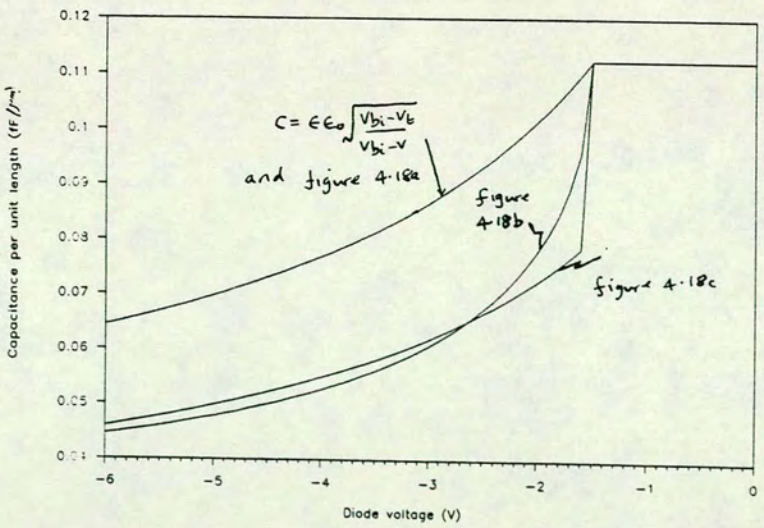


Figure 4.19 Peripheral capacitance approximations beyond threshold.

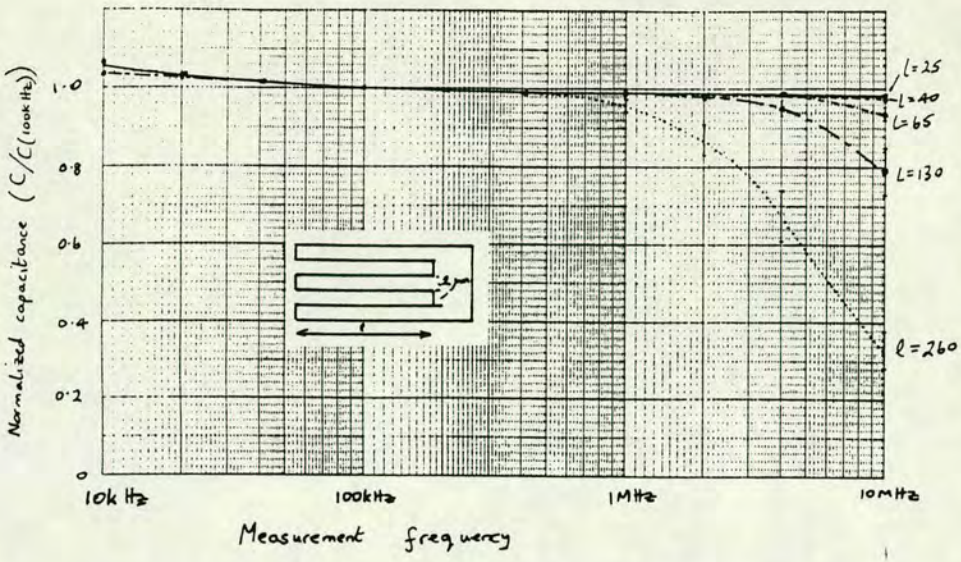


Figure 4.20 Roll-off in capacitance of digital diodes.

periphery were used, and the initial model for these related the total capacitance to the area and periphery dependent terms. It was soon realised that these diodes were not behaving as expected at high frequency, especially when a thin, low threshold-voltage active layer was used. Measurements performed at different frequencies on large diodes showed that there was a significant drop-off in capacitance with frequency, as shown in figure 4.20.

A physical quasi-3D model was derived for these digitated diodes, splitting the diode into physically similar regions, each having several capacitor and resistor elements to represent the network. The capacitance of each element was selected from the equations already presented, linked with the elemental area. Each element also had a series resistance associated with it, the value being dictated by the thickness of the sub-depletion region (and hence both the applied voltage and threshold voltage of the layer), together with the geometry of the element. The gaps between the metal fingers were modelled simply as a series of resistors whose values were chosen from the doping-level and the initial layer thickness. Figure 4.21 shows the results of such a simulation, indicating the close correlation between the measured and simulated result.

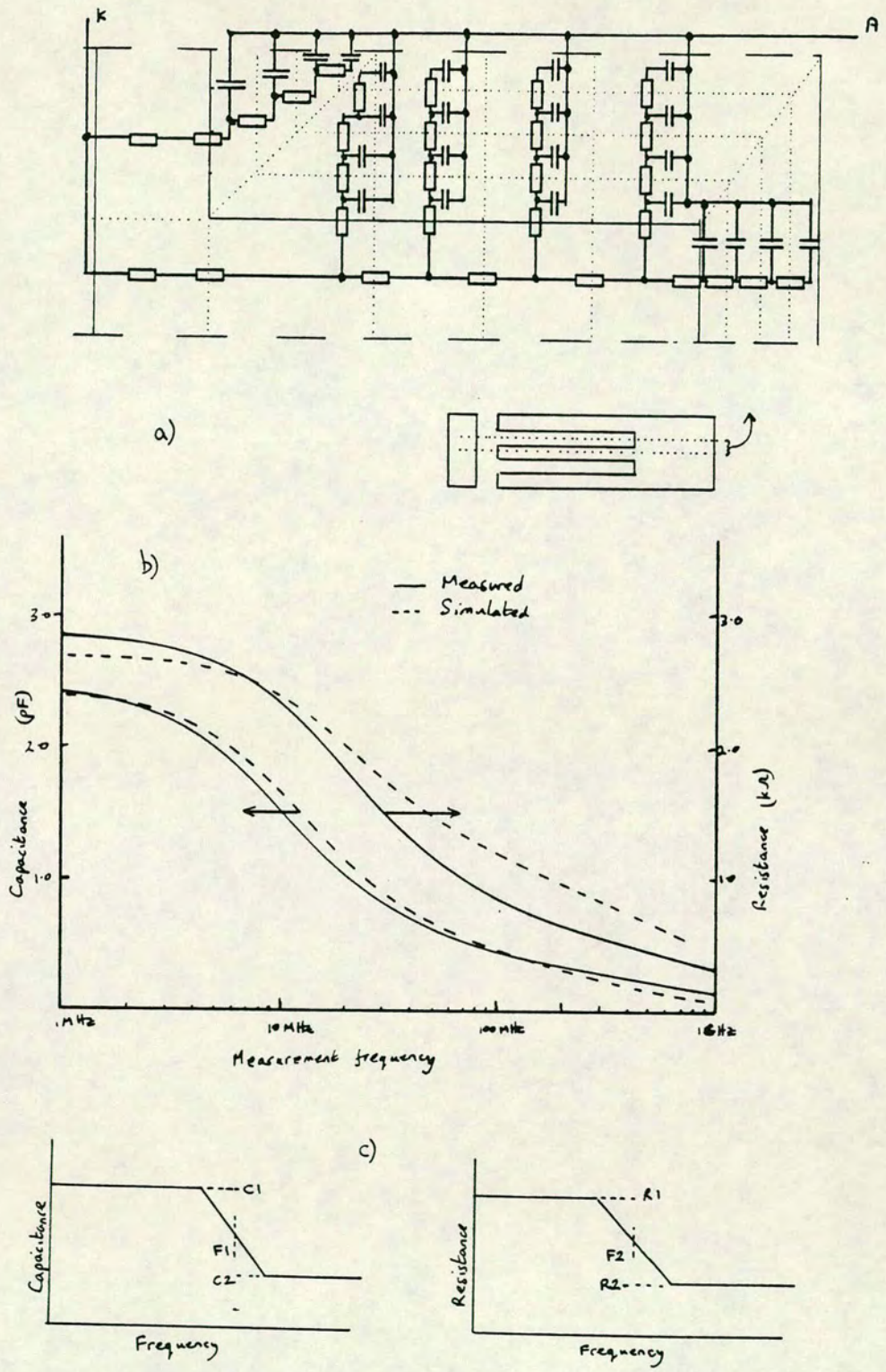


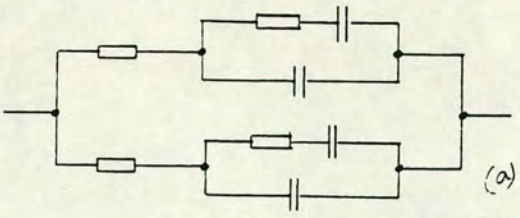
Figure 4.21 Modelling high frequency effects in digitated diodes:
a) derivation of model
b) comparison of measured and simulated data for a large diode
c) schematic diagram showing 6 required degrees of freedom in model.

Note that this simulation technique requires solely an understanding of the operation of this type of diode, coupled with an intelligent splitting into elements. A standard circuit simulator, in this case ASTAP, was then used to derive the frequency response of the network. Alternatively, a complete solution could have been obtained by solving Poisson's equation within the boundaries of this structure. The latter would have demanded the writing of a computer programme to solve the equation in 3D, but would have been more accurate. The ease with which the solution was obtained using the lumped element circuit model was a vital part of improving the component design.

This physical representation of the digitated diode is much too complex to include in a circuit simulation. However, once the predicted frequency response of the diode has been derived using the physical model, curve-fitting can be used to obtain a representation of the diode behaviour. By keeping the model used for the curve fitting very simple, this second model can be used successfully in circuit simulations. More recently, an identical technique has been adopted by IBM to obtain equivalent circuit models from 3D device models of bipolar transistors. [390]

Figure 4.21c shows that there are at least six degrees of freedom required to represent the frequency response of the capacitor network, even without attempting to control the slope of the decay. In order to allow more scope in the curve fitting, an eight element network was chosen to represent the diode. Values for the eight elements were then obtained by minimising the error in the fit to the two curves of figure 4.21b. A good correlation was obtained from this model, as depicted in figure 4.22.

The frequency plots were obtained by selecting appropriate values of both sheet capacitance and sheet resistivity, assuming a single applied voltage and therefore no localised debiasing of the network. Unfortunately, when this was repeated for a number of voltages, the eight extracted values showed no systematic variation with voltage. This was probably due to a very broad minimum of the error function between the required curves and the eight element model. It was therefore not possible to make the final model voltage dependent. As



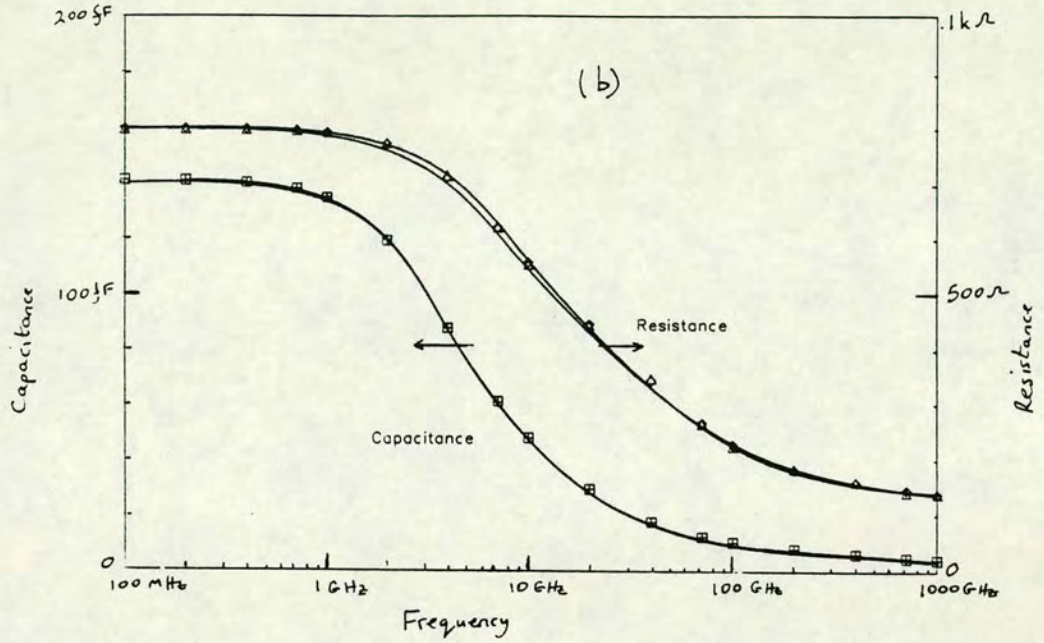


Figure 4.22 Fit between multi-component model of capacitor and simple 8-element model: a) 8-element model, b) simulation results.

this was already a curve fit to an approximation, the extra error arising from the small voltage changes during circuit operation was deemed small enough to ignore.

Despite reservations about the final accuracy of this model, its main application was to identify the acceptability of a diode construction. Clearly, the original design using a low threshold layer was unsatisfactory, entirely relying as it did on the periphery capacitance. When the fabrication process was modified to allow a deep active layer of high threshold voltage, a significant improvement was noticed, but the digitated diode was still a source of performance degradation.

The model was able to point the way to a plane diode construction, as sketched in figure 4.23, and was able to indicate the range of acceptable dimensions. The maximum usable anode length proved quite short at 5 μm, even with a very deep active layer of low resistivity.

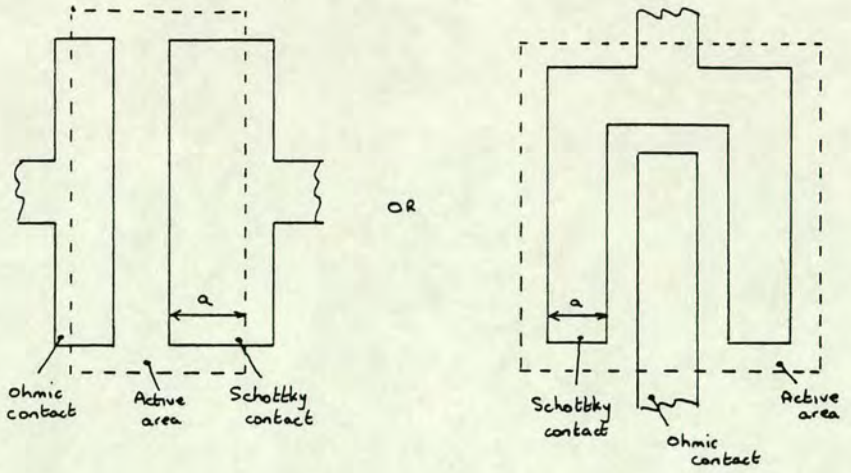


Figure 4.23 Modified diode construction for high frequency usage.
Note: a<5 μm for BTRL process.

Having now adopted both a simple geometry and a deep active layer, the complex model described above became obsolete. A single R.C pair is sufficient to represent the behaviour of this diode, and is easily incorporated into simulations of even the most complex circuits. The correct value of capacitance to use is that expected for the given area, and the resistance is simply the sum of the extrinsic and intrinsic components: the extrinsic value being directly derived from the geometry, and the intrinsic value being that of one third the "expected" value. The factor of one third arises because of the distributed nature of the resistor and capacitor [391]. Since the diode representation is once again straightforward, both these elements can be made voltage dependent in a rational manner.

Provided that care is exercised in the diode construction, this simplest of models is perfectly acceptable for circuit simulation. The physical modelling described above has enabled sensible bounds to be placed on the dimensions of the diode for the particular process adopted. The insight gained from this work will allow complete confidence to be placed on the derivation of new new design rules required for any future change in the overall processing.

4.2 MESFET Modelling

As with the Schottky diode, some physical understanding of the device behaviour must precede the development of detailed models. Again,

these models can be divided into those elements describing the DC behaviour and those impacting the AC or transient behaviour. This Section is divided into three, individually treating the physics and both types of model.

4.2.1 Physics of MESFET behaviour.

As discussed in Chapter 2, and in the introduction to the elementary modelling in Chapter 3, the MESFET is a variant on a JFET, but in which the p-n junction has been replaced by a Schottky junction. Thus, in modelling the transistor behaviour, we can call upon the diode modelling described in the previous Section, coupled with the basic JFET physics [392].

The conventional understanding of JFET behaviour is well treated in standard texts and it is not appropriate to regurgitate this material in other than very brief form. The standard model assumes that there is a depletion region which is associated with the junction beneath the gate electrode. This region is devoid of mobile charge, and the conductivity of the remaining channel between source and drain is therefore altered by its size. Since the width of the depletion region is a function of the applied voltage, the gate bias may be used to change the flow of drain current. Here in essence is the transistor action.

Because the channel is resistive, the voltage across the depleted layer will change along the length of the channel, and the depletion region will therefore adopt a shape similar to that shown in figure 3.10. Using the conventional assumption that the depletion layer is an abrupt boundary (the so-called "depletion approximation" [393]), the width of the depletion zone can be expressed in terms of this voltage. Since the voltage, and hence electric field, are related to the charge distribution by Poisson's equation [394], an integral expression can be derived relating the current flow to the applied voltages and the physical dimensions [395].

The debate over the exact form of the model to use depends upon the assumptions used to solve this integral. However, the transistor's

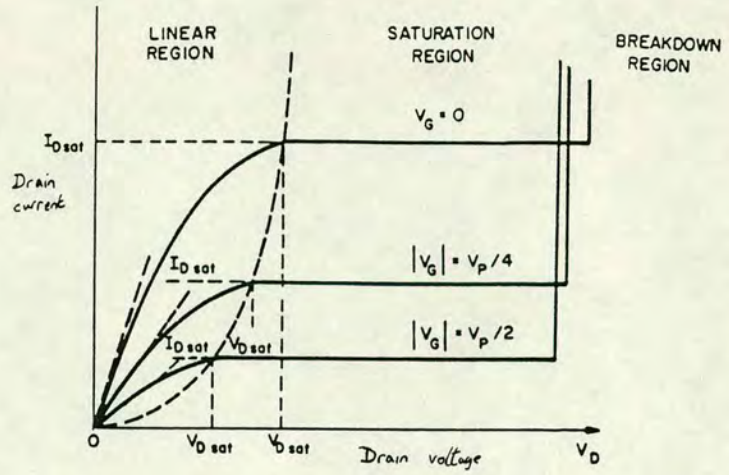


Figure 4.24 Generalised MESFET output characteristics.

output characteristics are of the form shown in figure 4.24 [396] irrespective of these assumptions. The curves are divided into two regions, the so-called linear and saturated regions (not to be confused with the equivalent designations in bipolar device physics, where the same terms take on completely different meanings [397]).

In the linear region, a small voltage applied to the drain will cause a current to flow. Initially this current will be linear with voltage, but soon the drain voltage distorts the shape of the depletion zone, reducing the channel thickness, and causing the current flow to fall. Eventually, the peak voltage between the gate and channel equals the so-called pinchoff voltage (equal to the built-in potential minus the threshold voltage) [398] and, according to the above model, the depletion region will completely close the channel. Some faith in the mathematics is required at this point, as the equations predict that a current still flows, despite the "pinchoff point" having a region of zero channel thickness, albeit for a zero length!

At larger voltages, the transistor is in the saturation regime, whereby the current is assumed to be voltage independent. Nowhere is the physics of this region adequately explained on a mathematical basis. Several factors mitigate in defining the boundary between the linear and saturation regions. Firstly, adherence to the assumptions of the depletion approximation is no longer acceptable. The second effect is that a space-charge region must build up on either side of

the pinchoff point [399], changing the voltage distribution in the channel. The third factor, particularly in transistors with active channels formed directly into a semi-insulating substrate, is that the field beneath the pinchoff point is quite intense, forcing carriers down into the substrate [400]. All these factors are predicted from Monte-Carlo particle simulations of devices [401]. They combine to give a non-zero slope in the output characteristics [402], usually, but not exclusively [403] positive. This leads to the so-called output conductance parameter.

4.2.2 DC Models

The model available at the outset of this study, described in Chapter 3, is obtained by assuming that the electron mobility is independent of electric field. This is a good approximation in silicon for devices with channels a few microns and longer [404]. It may also be applied to GaAs for even longer channels (say in excess of 5 μm), but for all state-of-the-art devices this is an invalid model. In GaAs, deviation from this model at all but very long channels arises because of the non-monotonic nature of the velocity versus field characteristics of the electrons, described in Chapter 2.

The first major attempt to solve the problem for field dependent mobility used a simple expression [405] to account for velocity saturation characteristics. The velocity has been fitted with:

$$v = \mu.E \cdot \frac{v_s}{v_s + \mu.E}$$

such that at low field, the velocity is given by the mobility times the field, and at high field, the velocity is constant. Figure 4.25 sketches this effect. Clearly, this expression is good for silicon, where the velocity saturates due to lattice interactions, but it does not fit GaAs where the major velocity saturation is due to inter-valley scattering (compare figure 4.25 with 2.2).

It is still useful to identify the consequences of this change, as similar effects would be expected in the GaAs device, even if the quantitative fit is wrong. In their analysis of this model, Lehovec

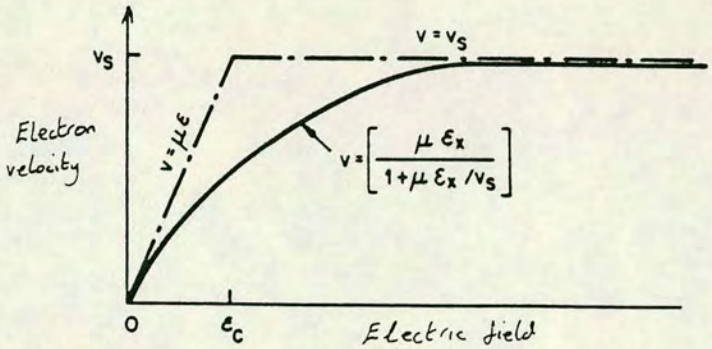


Figure 4.25 Approximation to velocity-field characteristic.

and Zuleeg [406] identify a "scaling" parameter z which is given by the expected velocity in the channel at pinchoff for the constant mobility model ratioed against that according to to the saturated velocity model. Thus, z increases as the channel is shortened, and the fields in the device become higher. When contrasted with the constant mobility model, as z becomes larger, this model predicts a lowering of both the saturation voltage and the saturation current and a much more linear transfer characteristic. Figures 4.26-4.28 reproduce the curves showing these trends.

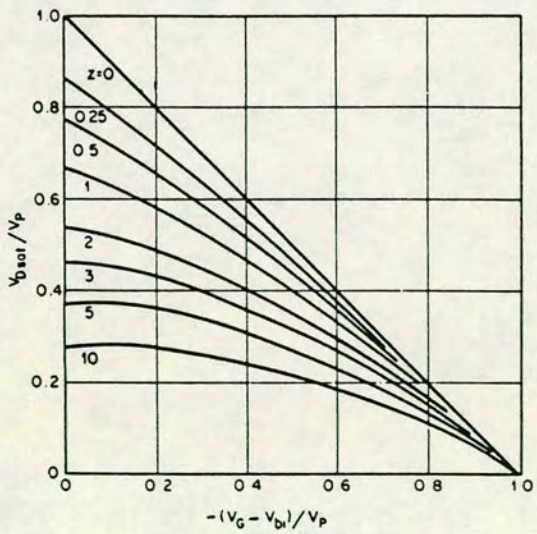


Figure 4.26 Saturation voltage according to [406] (as z increases Vdsat becomes constant).

In a further extension, Pucel [407] has split the channel into two regions, one with constant mobility and one with constant velocity. The boundary separating the two regions is moved according to operating voltage. Essentially this model fits a piecewise linear curve to the velocity field distribution of figure 4.25. Because

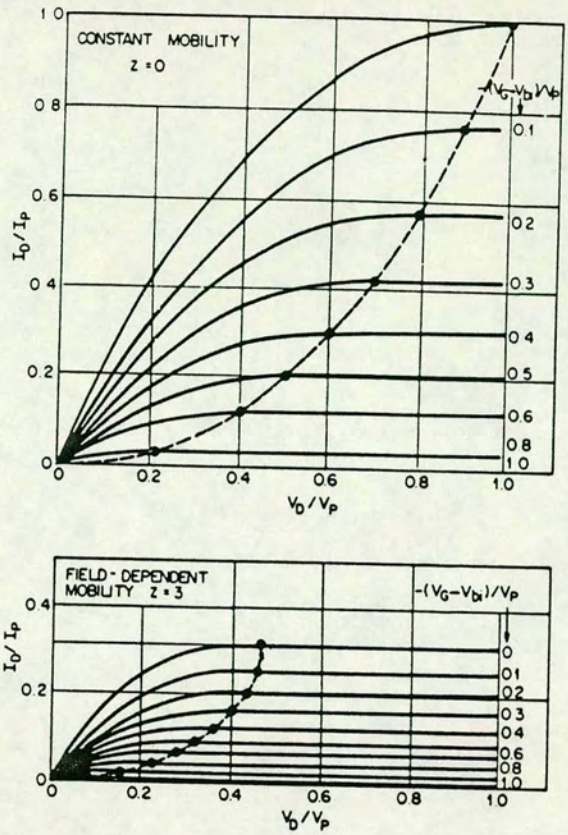


Figure 4.27 Linearising of saturation current according to [406] (I_p=maximum channel current when z=0).

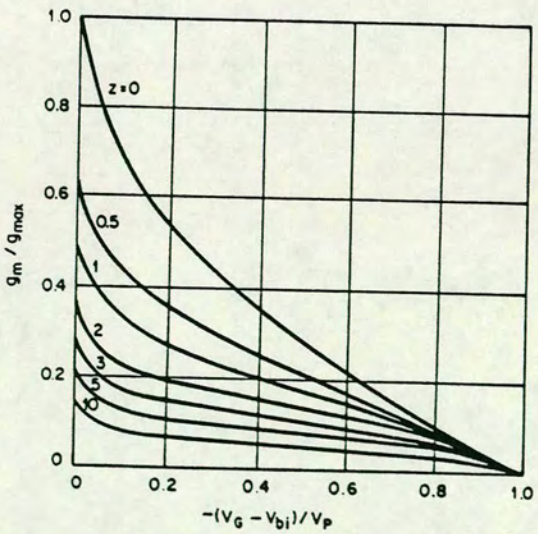


Figure 4.28 Transconductance according to [406] (as z increases, g_m becomes more linear).

this gives a sharper transition between the two regions than the previous model, it is much more representative of GaAs, in which the mobility remains high right up to the critical field, when the velocity suddenly falls.

Even here, the negative mobility region is not however modelled. If the effect were to be included, the current flow would have to be modified by the formation of the space-charge layer at the drain end of the gate [408]. Detailed modelling of any FET device shows this phenomenon, but in GaAs, the sudden drastic slowing of carriers when the electrons change bands exaggerates this.

Although these two models come closer to expressing the device physics, even if they were completely accurate, the equations would be much too complex to be used in circuit simulations. Here, one must again draw the distinction between predictive and representative models. The former are required to explore devices of new, generally smaller dimensions, whilst the latter are required to exploit to the full existing, measured devices. In this work, a representative model is required for maximising circuit performance by optimising the operating conditions of each transistor, and an empirical fit to the experimental data is therefore adequate. The importance of the physical models should not however be underestimated, because any justification for new, empirical models comes from studying the physical trends described above.

Several transistors were measured, covering a wide range of threshold voltages and saturation currents. The effects of the source and drain resistances were eliminated from the characteristics by first using a Kelvin measuring technique to identify their values [409], and then by increasing the gate and drain biases to compensate for the predicted Ohmic voltage drops across these parasitics. In the resulting curves, the common observation was that all devices followed two of the three trends identified from Lehovec's model; namely a very low saturation voltage, and a much more linear transfer characteristic than expected. The third observation (of reduced saturation current) was probably present, but was masked by the unpredictability of the factor linking threshold voltage with

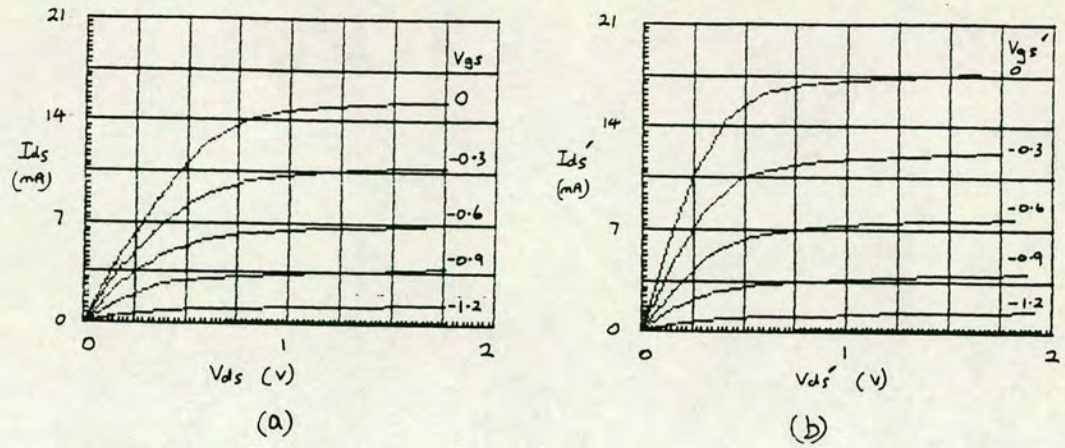


Figure 4.29 Output characteristic of a MESFET (V_t=-1.5):
a) measured directly,
b) applied voltages modified to eliminate R_s=R_d=8.3

saturation current. This was the main manifestation of the major source of unrepeatability in early GaAs processing, arising through variations in mobility caused by uncontrolled implant activation and compensation. Figure 4.29 compares the transistor output characteristics before and after de-embedding these resistances.

Following all these observations, and justified by the theoretical trends already described, the first step in introducing a new, usable model was to establish that the saturation voltage itself approaches a constant value as the gate voltage increases. In most devices, this is much smaller than expected. In this new model, the saturation voltage is defined as a key parameter which can be calculated from:

$$V_{sat} = \left\{ \frac{1}{(V_{gs} - V_t)^m} + \frac{1}{(V_{sat0})^m} \right\}^{-1/m}$$

where V_{sat0} is set equal to the constant saturation voltage, typically 0.75 V. The parameter m determines how quickly the saturation turns over from the conventional expression at low voltage (given by the first term in the equation), to the saturated value (the second term) at high voltage. For numerical convenience, m=2 is a good choice, as the square root function may be used to avoid recourse to logarithms.

In the standard model, the saturation current is given by [410]:

$$I_{sat} = I_{dss} \cdot \left\{ \frac{V_{gs} - V_t}{V_t} \right\}^2$$

In these experimental devices, the saturation current is modelled quite well by a similar square law when the voltage is well away from threshold, but an offset voltage, V_{of} must be added to the expression:

$$I_{sat1} = I_{dss} \cdot \left\{ \frac{V_{gs} - V_t + V_{of}}{V_t + V_{of}} \right\}^2$$

This new expression now exaggerates the current near threshold, so it is necessary to modify the current for low gate voltages. The appropriate factor should reduce the above expression to zero at threshold but multiply by unity at large gate voltage. This factor has already been introduced, because the equation to calculate the saturation voltage satisfies both conditions when suitably scaled. The new empirical expression for saturation current is now:

$$I_{sat} = I_{sat1} \cdot \frac{V_{sat}(V_{gs})}{V_{sat}(0)}$$

where $V_{sat}(0)$ is the saturation voltage when the gate voltage is zero.

Thus far, the I-V locus of the boundary between the saturation and linear regions has been defined. The model is completed in the linear region by fitting the peak of an inverted parabola to this boundary, in such a way that the parabola passes through the origin. Thus:

$$I_{ds} = I_{sat} \cdot \left\{ \frac{2.V_{sat}(V_{gs}) - V_{ds}}{V_{sat}^2(V_{gs})} \right\} \cdot V_{ds}$$

An output conductance is readily added to this equation.

Figure 4.30 shows the accurate fit which is obtained between this model and the experimental data for two completely different wafers, with widely spread threshold voltages.

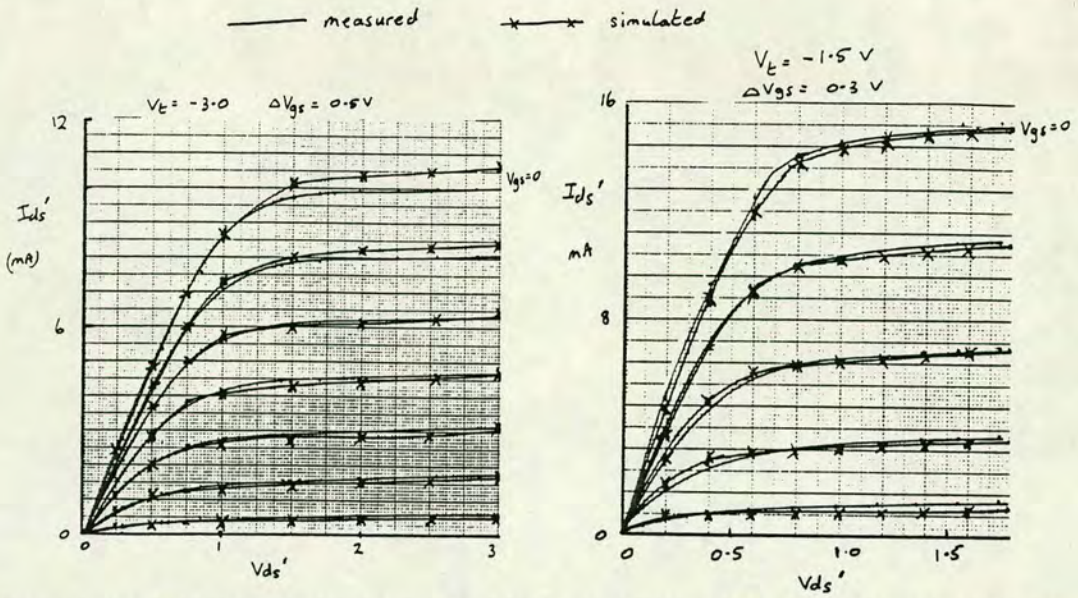


Figure 4.30 Fit between measured and simulated output characteristics for widely different transistors, using new model.

Using this model, the saturation current and the threshold voltage, although loosely linked by processing, must be specified individually. The only questionable parameter is the arbitrary V_{of}, and the inflexibility of the model lies in its inability to fit more than a narrow range of slopes of transfer characteristic, although admittedly only a narrow range has actually been encountered. The value of V_{of} lies very close to both the peak saturation voltage and the barrier height of the junction. Intuitively there are tempting reasons to link to either of these, but there is no compelling theoretical connection.

This model has been used in all the circuit design work described in the succeeding Chapters of this thesis. The appropriate values of I_{dss} and V_t were originally taken from a few early measurements, but as the data bank of device measurements has been increased, the link between these two key parameters has been established much more definitely. I am indebted to colleagues who have amassed these data in scattergram form, as shown in figure 4.31. From this graph, (and according to any theoretical model) only one of these two parameters is needed. The equation of best fit to the experimental data of figure 4.31 can be included as part of the device model.

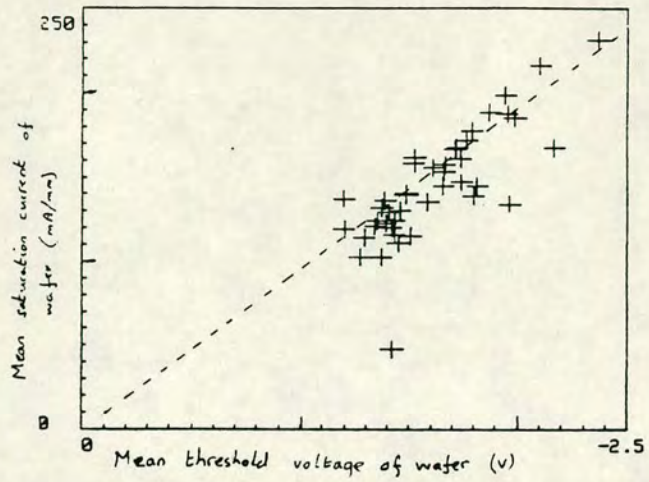


Figure 4.31 Correlation between threshold voltage and saturation current of FETs.

The transistor model has also been taken one stage further, primarily as an attempt to ease the extraction of the modelling parameters from measured characteristics, but also to remove the arbitrary nature of the model in the saturation region. Instead of describing the behaviour in this region using the offset voltage V_{of} coupled with an arbitrary roll-off at low-current, the power law describing the transfer characteristic is changed. Thus:

$$I_{sat} = I_{dss} \cdot \left\{ \frac{V_t - V_{gs}}{V_t} \right\}^m$$

where m is now the only fitting parameter. If m = 2 then the model is similar to the conventional long-channel JFET model, and if m = 1, the model describes a device with constant carrier velocity [411]. Thus, the value is allowed to range between 1 and 2 with some justification, and the correct choice of m is now readily obtained from the ratio of the forward transconductance under small and large signal conditions. In moving to this new model, the saturation voltage has been fixed at V_{sat0}, independent of the gate voltage. This is a minor simplification which is again justified to save on computation, but which could not have been applied to the earlier model, where the variable value was also required in the drain current equations.

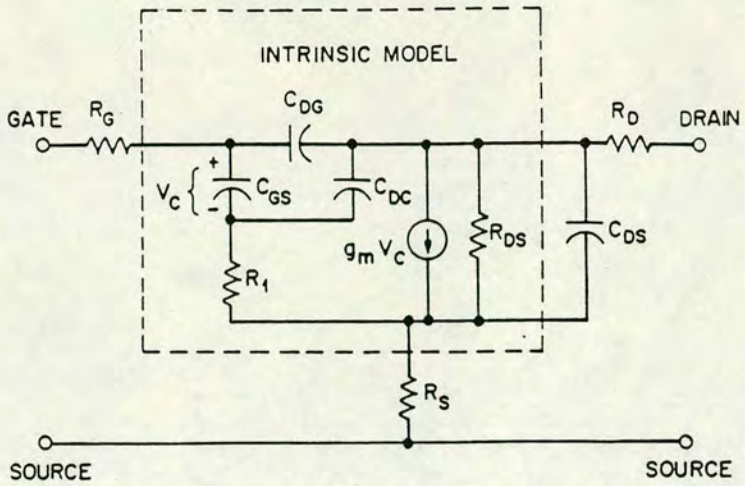


Figure 4.32 Typical small signal MESFET model.

This is a much more practical model. Since the value of m may be obtained from a simple measurement, it is possible to monitor it routinely. Through the empirical link with the theoretical models described earlier, any variations in this parameter can be linked directly to variations in carrier behaviour, probably arising from differences in the internal field distribution (through different dopant activation, varying ratio between length and depth of channel, different gate recess profile, etc.).

Note that throughout this modelling work, the effects of the source and drain resistances have first been eliminated, to obtain a so-called "intrinsic" transistor. This is the real key to obtaining good models and physical understanding, but is apparently often neglected. The importance of the source resistor in particular can be observed because the effective transconductance is directly affected by this parasitic element [412]. It is therefore important to obtain a low value, and a significant effort is expended in process design simply to minimise this component, not forgetting that in a dual gate transistor, the inter-gate region forms a part of the source resistance of the upper device.

4.2.3 AC Models

Most commonly used AC models for transistors are small-signal, as required in the design of amplifiers and similar analogue circuits. An equivalent circuit similar to that in figure 4.32 is usually used to describe the transistor behaviour.

The parameters are obtained directly from an s-parameter [413] or other high frequency measurement technique. Using the scattering parameters, measurements of transmission and reflection coefficients are made for both ports (input and output) at different frequencies, and the values extracted from curve-fitting. Separate values have to be obtained for each operating condition, as this model is a linear representation of the non-linear device.

The value of the current source in this equivalent circuit can also be obtained directly from the DC model for the particular operating point of interest. Quite commonly the two values are in considerable disagreement, probably because of the varying shunting effect of the capacitor representing the charge stored in the stationary Gunn domain, discussed earlier.

This small signal model is of little importance in digital circuit design, where the voltage swings are invariably large, such that the non-linearities must be considered. If it is difficult to rationalise the parameters in the DC model with those in the small signal equivalent circuit model of figure 4.32, it is near impossible to obtain sensible values for a large signal model. Nevertheless some attempt must be made.

As discussed in Chapter 3, the simplest large signal AC model is obtained by the addition of non-linear capacitors to the DC model. Figure 4.33 shows a suitable model in more detail than that presented earlier. Capacitors are added to the intrinsic device to represent the active junction and the space-charge region, and the inter-electrode capacitors are added to the extrinsic model. Only the junction capacitors were used in the earlier model.

Because the gate is formed from a Schottky diode, the equations developed in Section 4.1 are appropriate for the junction capacitors, although the respective areas of the source and drain components need to be determined. In the inherited model, the source capacitance was calculated assuming that half the diode area was responsible for this component. The drain capacitance was calculated from the other half area, (but with the appropriate voltage applied), and then 10% of

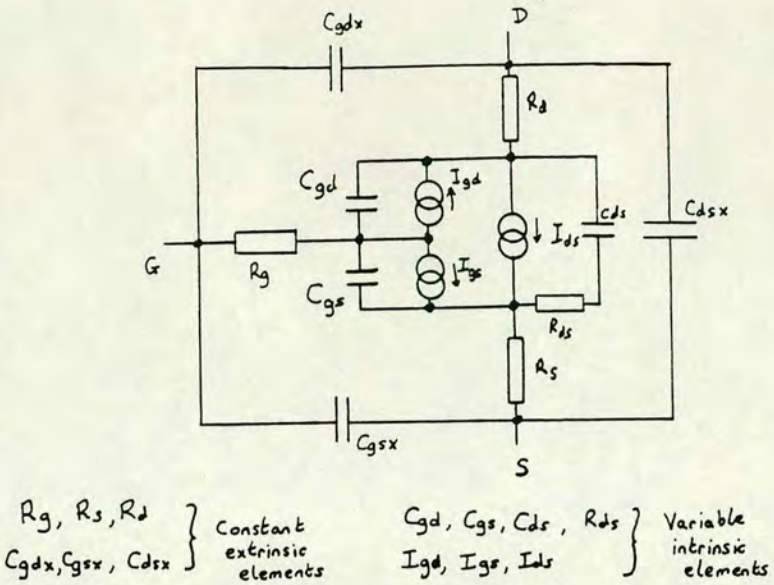


Figure 4.33 Large signal model of a MESFET. Note: R_{ds} and C_{ds} have been omitted here.

this value was used. This would generally be fitting for most small signal models because of the typical bias voltages applied, but is quite inappropriate over the full voltage range used here.

A better empirical approach is to acknowledge that the space-charge layer isolates the drain from the junction under most operating voltages and, therefore, the whole gate area should be used for the source component, with only the peripheral capacitance at the drain end being used in the gate-drain capacitor. This is a much better model, but still under-estimates the critical gate-drain capacitance (critical because of Miller effect). Perhaps a better value still is given from the Pucel model [414], but at the expense of considerable complication.

The problems in gate-drain capacitance only arise if the drain-source voltage is very small, when the two gate capacitors should be near equal. None of the models cater adequately for this, yet it is this component which dominates the performance of BFL circuits during a portion of the switching transient [415]. In the work of this thesis, devices were not operated in this regime, and the inadequacy of the simple (revised) model was tolerated. Great emphasis was placed on the calculation of the interelectrode parasitics, as described in Section 4.3, and since a major stray component arises in parallel with the gate-drain capacitor (especially in CCL with one

capacitor per switch transistor), minor errors in the actual value of gate-drain capacitor were once again quite tolerable.

The other element of particular interest is the space-charge capacitor. There is no suitable model for this value, and it was omitted from these simulations. In the absence of physical models, it is impossible to assess the magnitude of the errors arising from its omission. The effect is to ignore one component of the phase change through the devices, which although critical in analogue circuits, is probably less important in digital circuits. Improved models now becoming available in commercial packages [416] begin to tackle this element. One of the further problems, even with new models, is that real transistors tend to show some charge storage or parameter drift effects which are time dependent, so no single value of capacitor would be accurate [417].

The three electrode capacitors have all been assessed in detail for a variety of transistor geometries, using the models developed in the Chapter 4.3. Rather than derive the individual values of each transistor, complete capacitor matrices were calculated for the commonly occurring elements such as inverters or NAND gates. This is a much more realistic situation than assuming that two transistors placed together behave as if the same transistors were patterned far apart. This approach is particularly necessary because of the enhanced inter-electrode matrix values associated with the very thick dielectric of the semi-insulating substrate.

4.3 Interconnection modelling

4.3.1 Introduction

Historically, the modelling of on-chip interconnection for ICs has been a trivial problem. This notion is based on silicon ICs, in which the tracks are of large dimension compared with the underlying dielectric. Each track therefore closely approximates to a parallel plate capacitor, with the underlying semiconductor forming one plate, and the track the other.

Even as chips have increased in complexity, this simple model has been retained. The average and maximum lengths of track used

internally have been directly related to the chip size using statistical data. Problems were first encountered in MOS memory ICs, where polysilicon was used as a convenient interconnection [418], rather than metal. As track lengths increased, the relatively high resistivity of this material resulted in poor time constants [419]. The answer to this was to advance to multiple levels of metal for interconnection [420], when once again the capacitance (obtainable from simple calculations) dominated the interconnect modelling. More recently, the track dimensions have shrunk to become similar to the dielectric thickness and a much more complex modelling technique is required [421]. The calculation of capacitance between tracks on different levels is also non-trivial.

In bipolar circuits, the generally much higher current levels dictated an earlier move to multi-level metal, avoiding the resistance problem. However, the demand for much higher speed, especially from ECL, has called for more complete calculation of track capacitance than is obtained statistically. Also, the use of much smaller signal voltages [422] imposes a tighter constraint on signal cross-coupling between tracks. The advent of sophisticated CAD, has introduced back-annotation of the finished layout, in which the complete circuit including parasitic elements is reconstructed from the final geometry. This has enabled the study of the effects of interconnect capacitances, particularly those lying on the critical timing paths of the circuit. This advance has come during the span of this project.

In GaAs, most of these problems were encountered immediately. The semi-insulating nature of the substrate removes the back-plane to the back of the wafer such that the track dimensions (and most importantly spacings) are always much smaller than the dielectric thickness. Thus, cross talk is potentially high, capacitance calculations are difficult, and the capacitance is critical because circuits are intended for high speed operation.

4.3.2 Track capacitances

There is a number of methods for the calculation of capacitance of a single track on GaAs, depending upon the required accuracy and

available computation power. In modelling a single track, the assumption of infinite length is invariably made, but practically, this only means that the track is long enough to ignore end effects. As such, the simplest techniques are strictly only two dimensional.

The first of these methods is to recognise that a single track placed directly on GaAs represents a microstrip line [423], with the back of the substrate forming the return path. Conformal transform techniques [424] allow closed form solutions to be derived for the calculation of such lines, and these are well documented, although even one author quotes different levels of approximation [425]. Because of the need to make approximations in obtaining these closed form solutions, they are generally restricted to a narrow range of linewidths for a given substrate permeability and thickness. In the case in question, the limits of validity of the equations invariably lie outside the region of interest. All the tracks being used are too narrow to fall within the documented tolerance of these equations. The adoption of this technique must therefore be open to some interpretation.

The Green's function technique is one of the iterative computation techniques available for solving the problem [426]. In this method, there is a known relationship between the potential at any point on the track of interest and a unit charge placed anywhere else. By dividing the track into filaments and assuming a point charge at the centre of each filament, but with differing amounts of charge on adjacent filaments, the potential on each filament can be found. This potential may now be used to calculate new values of charge. By iteration, a self-consistent charge distribution is eventually obtained, with the capacitance being derived quite simply from this.

This technique has been applied to the single track on GaAs, both for the wide and narrow tracks. Figure 4.34 shows the results for different numbers of filaments, and compares the results with the analytic expressions. Clearly, the more filaments, the more accurate the method and the figure thus agrees well with the conformal transform solution, even for the narrow stripes where the latter is questionable.

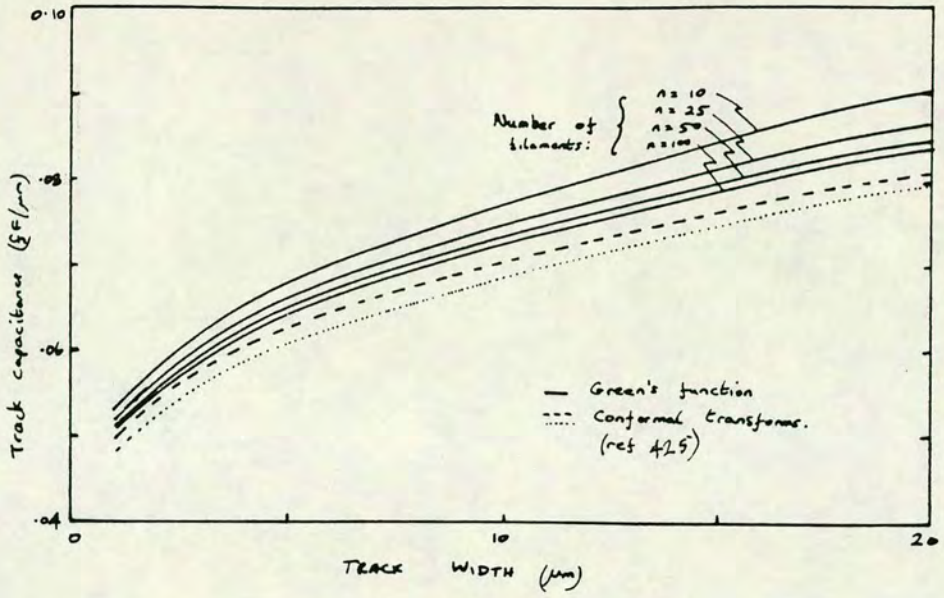


Figure 4.34 Capacitance of a single track an GaAs.

Any other technique for solving Poisson's equation may be adopted [427], but this particular technique has been used here because of its simplicity and accuracy compared with the accepted method (used in its valid regime).

The next trial is to move to two equal sized strips placed close together. By analogy, this may be solved using the coupled microstrip approximations [428], or by using Green's functions. This time the gap between the two tracks must also be filled with filaments, but each bearing no charge. Because of the limitation on the size of matrix which may be numerically inverted, this restricts the accuracy when the strip widths differ significantly from the gap width. Figure 4.35 shows the result of these computations. The implication is that the Green's function technique approaches the coupled microstrip solution for coarse structures, but only if sufficient filaments can be used to represent the potential adequately, both in the gap and on the tracks. The second conclusion is that the Green's function method becomes more accurate as the gap and track are of similar size. Thus, it should predict accurately for all typical track geometries. At small dimensions, the coupled

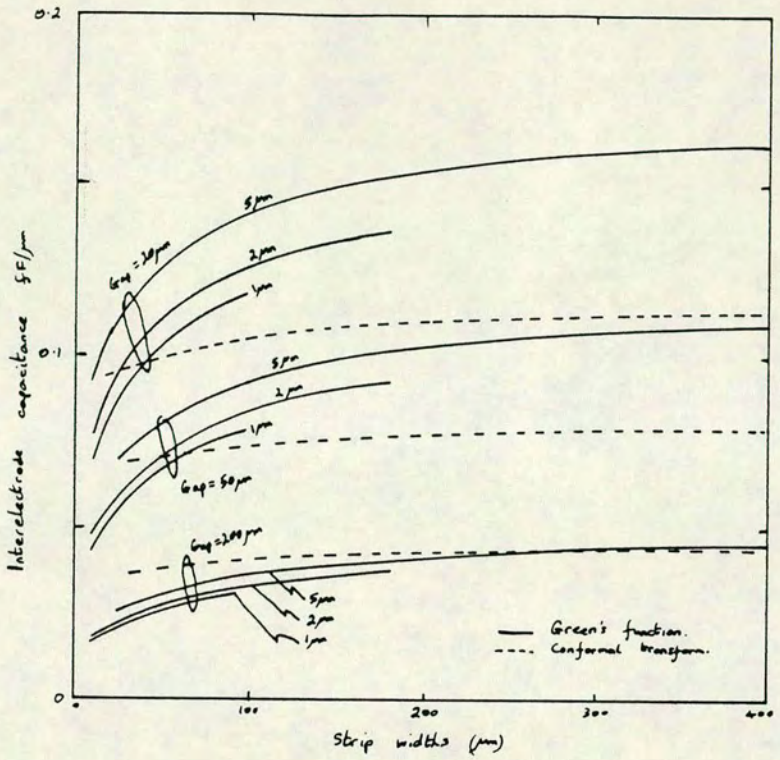


Figure 4.35 Capacitance between a pair of equal sized tracks on GaAs.

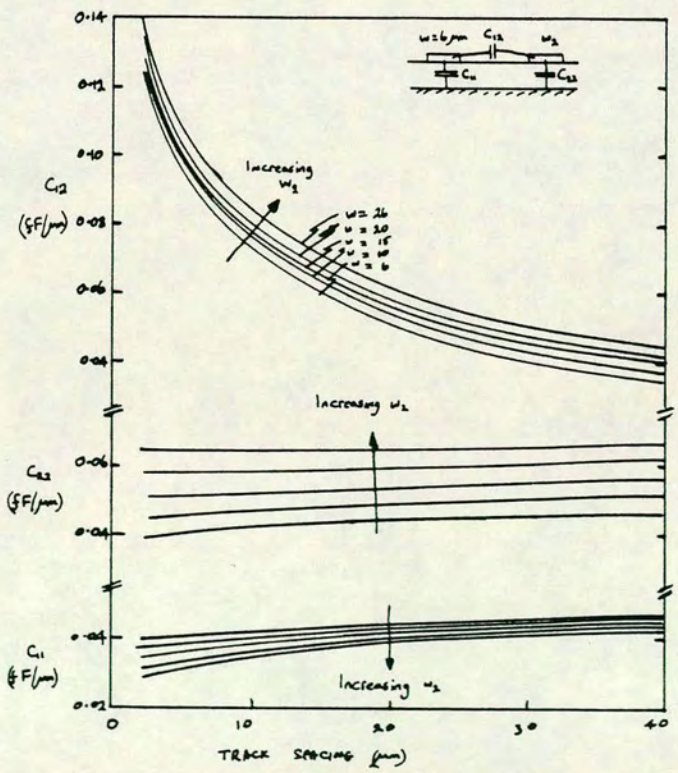


Figure 4.36 Capacitance matrix for unequal tracks on GaAs.

microstrip calculation appears to fail. This is unlike the calculation for the single microstrip where accuracy was apparently maintained even outside the specified bounds.

Thus, for more than one track, the existing closed form solutions appear inappropriate. However, the result predicted from the Green's function method may over-estimate capacitance by some 10% or more, even when a large number of filaments is used.

The charge distribution on the tracks is such that the filaments should ideally be concentrated close together near the edge of the track. In its simplest implementation, a fixed filament size has been adopted, but in a refinement, the filaments could be grouped into the regions where the charge changes most rapidly (i.e. at the edge). By careful choice, a much more accurate solution would be obtained. Even greater accuracy could come from an adaptive technique, in which filaments are introduced only where the charge is changing rapidly and removed where there is only a small change. However, to retain user flexibility and simplicity, this has not been implemented. Instead, the computation time for an accurate solution has been reduced by obtaining an initial solution for a small number of filaments, then reducing the filament size using interpolation to estimate the charge distribution on the extra filaments, and resimulating. The second solution is much more accurate than the first and converges in fewer iterations than required for a similar accuracy solution obtained directly. This process can be repeated to the desired level of accuracy.

As already described, these hitherto simple, symmetrical examples could be calculated with varying degrees of accuracy using established models. Generally, the cases of interest are more complex, including asymmetric coupled microstrips (two tracks of differing widths), and larger numbers of tracks. These lie outside the realms of the established techniques, but are made tractable by the Green's function method.

Figure 4.36 summarises the trends observed in the elements of the capacitance matrix as the track sizes and separation are altered.

Again, it should be stressed that these curves pertain to infinitely long parallel tracks. However, they may be used with discretion to calculate the effects of adjacent shapes of finite length.

Such models are usually considered only for metal tracks or interconnects. However, the same rules apply to the calculation of the stray capacitance between a track and an active area, (e.g. in most CCL gate cells, the coupling capacitor sits with its long edge parallel to the V_{ss} power rail). Strictly, this is only true for tracks directly on the GaAs surface (first-level metal). However, whilst the lower dielectric constant of the polyimide used to separate the metal levels should lead to a smaller capacitance to the upper layer, the effect of the height separation will offset this with an increased capacitance. For convenience therefore, the coupling between track and nearby active area has been assumed to be independent of metal level. This is especially useful as all tracks, independent of level, have been kept at a minimum distance from the active area. Furthermore, the circuit layout rules adopted for this work have deliberately excluded the crossing of active areas by tracks on the second level metal, so each active device may have a single value of stray capacitance calculated for it, the value being obtained by assuming that the device is flanked on all sides by tracks at the minimum separation.

Since the FET capacitance is particularly important, it is treated as a special case. As indicated in Section 4.2, the parasitic capacitances of the FET must be calculated by treating it as an array of parallel tracks. Figure 4.37 shows how this is achieved, splitting the transistor into three different regions.

Model 1 is applicable outside the active areas of the transistor, and represents the full capacitance of this area, both in the GaAs substrate, and in the air (or polyimide) above. The width of this region is artificially increased by a "phantom" width, to compensate for the edge effect [429].

In the active areas, the GaAs beneath the gate is conductive, and there is thus no free dielectric field in this region. The electrode

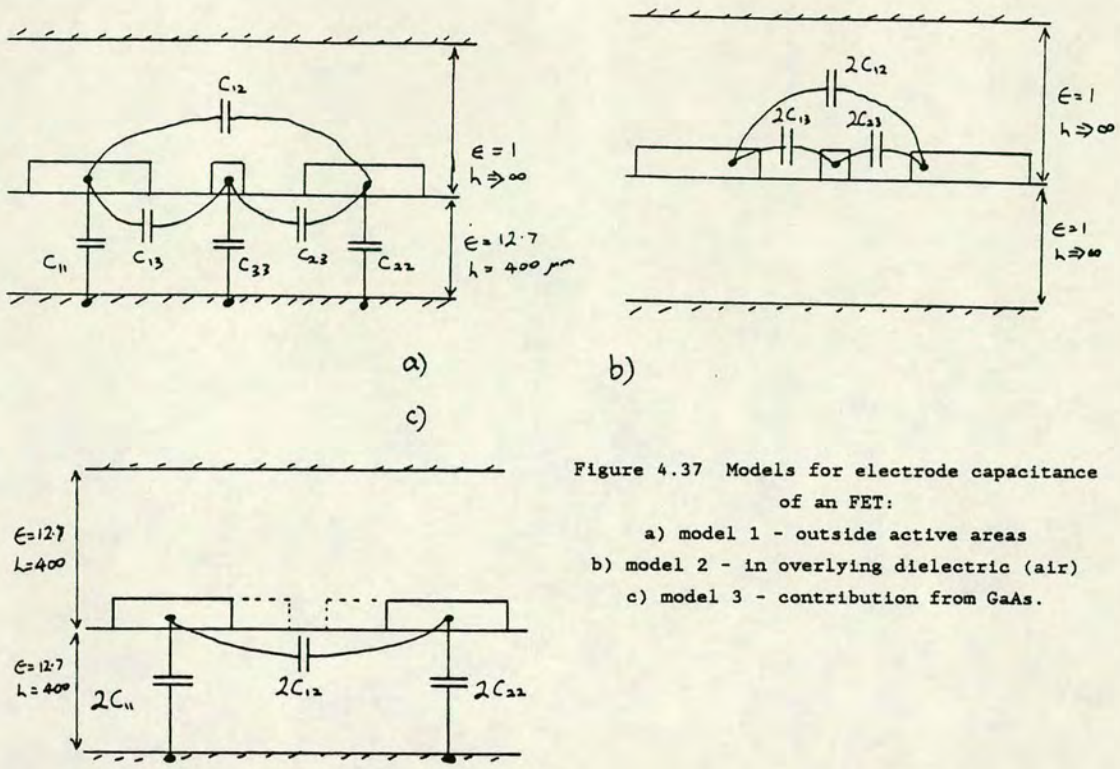


Figure 4.37 Models for electrode capacitance of an FET:
a) model 1 - outside active areas
b) model 2 - in overlying dielectric (air)
c) model 3 - contribution from GaAs.

capacitance of the gate therefore only arises in the air above the GaAs. Making the assumption that the equipotentials lie normal to the GaAs surface, the contribution to the capacitance may be calculated from the perfectly symmetrical model 2.

Model 3 is the equivalent model for calculating the substrate component of the capacitance. Because the conducting channel must be accounted for, the electrodes in model 3 are set equal in size to the metal contacts plus the appropriate spacing between that electrode and the gate electrode. This is effectively assuming that all the voltage is dropped across the channel immediately beneath the gate. In the saturated regime this is not a bad assumption, and in any case gives the largest (worst case) capacitance value. As with model 2, the capacitance value is obtained by halving the value obtained from the symmetrical situation.

The capacitance values obtained from each of the three models are added together to give the final values. Having calculated the

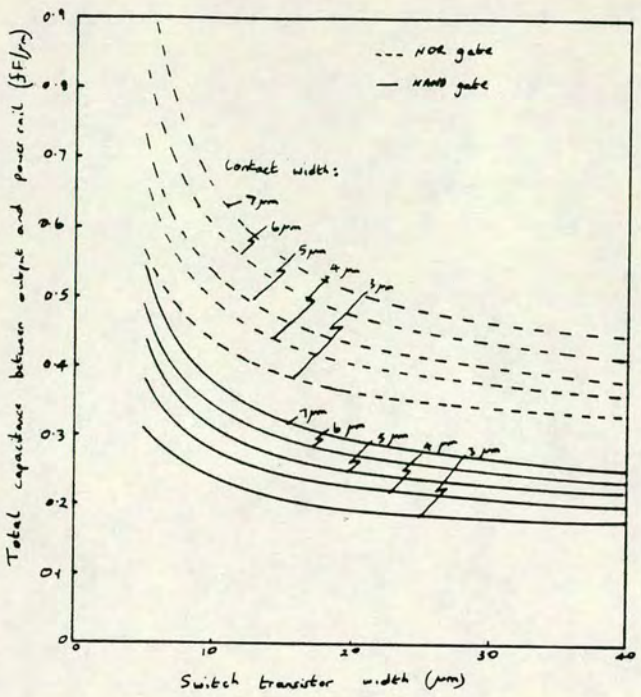


Figure 4.38 Output capacitance vs switch width for NOR and NAND gates.

inter-electrode capacitances for the transistor, similar models may be used for any combination of transistors and tracks. Generally only a small number of specific geometries is worth reviewing independently. Figure 4.38 shows the simulated effects of varying the contact sizes for some of the most commonly used physical structures.

4.3.3 Cross-over capacitances

The difference between the effective capacitance of tracks on the first and second metal levels is really quite small as already suggested. The only significant extra contribution to consider is the inter-track capacitance at a crossover. This is unfortunate, because it is also one of the hardest structures to model well. For this reason two different approaches have been taken.

In the first approach, the crossover shown in figure 4.39 can be split into two sections as shown. In the XX' section, it behaves as a single microstrip line with the polyimide as the dielectric, and the first level metal as a groundplane. In the second case (YY'),

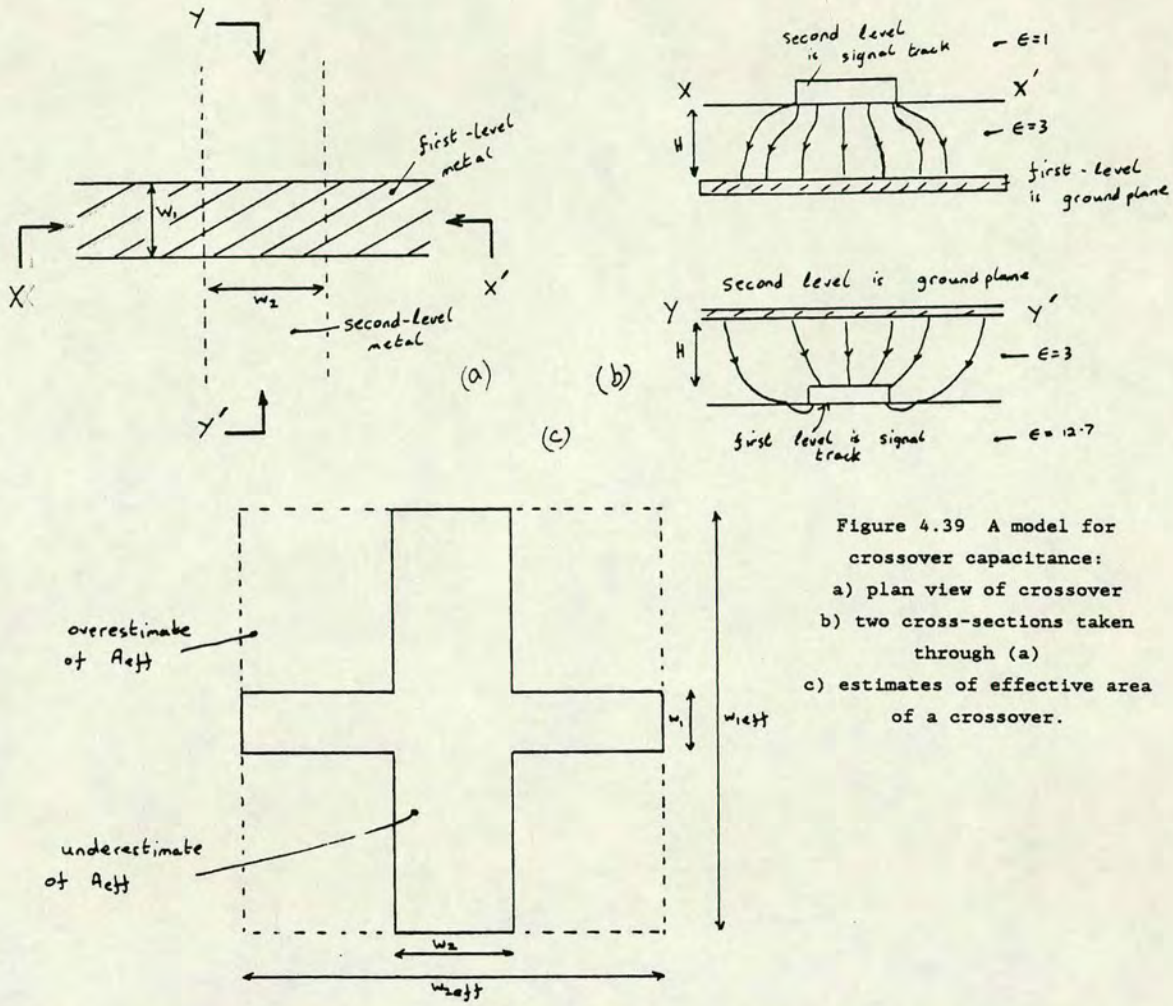


Figure 4.39 A model for crossover capacitance:
a) plan view of crossover
b) two cross-sections taken through (a)
c) estimates of effective area of a crossover.

the groundplane is on top, and the microstrip is buried at the interface between two dielectrics, the polyimide and the GaAs. Assuming that this model is sufficiently representative, despite the very short perpendicular lengths, effective widths, w_{eff}, can be defined for each of the two tracks. Here, a parallel plate capacitor of width w_{eff} has the same capacitance per unit length as does the microstrip line. Since the tracks are much wider than the inter-layer dielectric thickness, the standard microstrip capacitance formulae may be used to calculate the effective widths. Intuitively, from figure 4.39, the effective area of the crossover must obey:

$$w_2 \cdot w_{1eff} + w_1 \cdot w_{2eff} - w_1 \cdot w_2 < A_{eff} < w_{1eff} \cdot w_{2eff}$$

For a typical crossover comprising equal track widths of 6 um with a 1 um thick dielectric, these inequalities imply that the capacitance lies between 2.7 and 3.0 fF. This very tight range suggests that the

optimum value is obtained by simple averaging of the two values. Thus this model predicts that the crossover capacitance is given by:

$$A_{eff} = 0.5 \; (\; w_2 \; . \; w_{1eff} + w_1 \; . \; w_{2eff} - w_1 \; . \; w_2 + w_{1eff} \; . \; w_{2eff} \;)$$

The second model is also applied using various assumptions. Firstly, the lower track is replaced by an infinite metal sheet, and the upper track by a circular disc, to give capacitance C_U. The process is then reversed, with the disk on the lower level, and the groundplane on the upper, to give C_L. The two results are then averaged to give an effective crossover capacitance. Again standard formulae may be applied [430]. The problem lies in assigning an area to the disc, compared to the actual area of overlap. For equal sized tracks, four regimes are worth considering. These are shown in table 4.1 for 6 μm tracks with a 1 μm dielectric.

Table 4.1 Crossover capacitance modelled as a circular disc:

method	C_U	C_L	C_{eff}
radius = semi side	1.33	3.17	2.25 fF
radius = semi diagonal	2.43	5.18	3.81 fF
equal areas	1.64	3.76	2.70 fF
equal periphery	2.02	4.45	3.24 fF

Again, there is justification in using either the diagonal diameter or the equal periphery results as upper bounds, with either of the other two for the lower bound. If all four values are averaged, the result is very close to that obtained from the alternative method. However, when dissimilar track widths are used on the two layers, only the area and periphery related values are calculated easily. Under these circumstances, the equal periphery value would be used for the upper bound, with the equal area for the lower bound. Because of the assumptions, this would only be a suitable model for tracks of near-equal width.

Since the two completely different methods predict similar values we may conclude that either of these models may be adopted with confidence. However, the first method is adopted in preference,

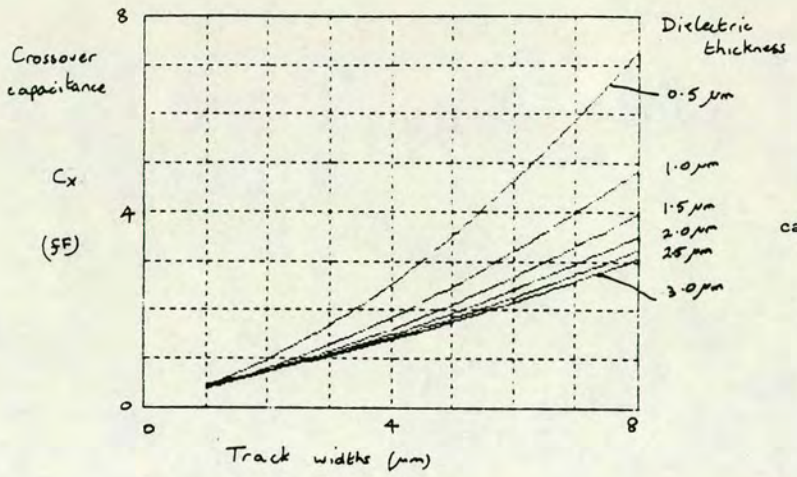


Figure 4.40 Crossover capacitances for equal sized tracks.

because unequal track widths are more readily accommodated. Figure 4.40 shows the computed crossover capacitance for various combinations of track widths and dielectric thickness.

Formulae are available to modify the equations for a finite thickness of metal [431]. Allowing for this, up to a further 10% capacitance may be added for 0.2 um to 0.5 um metal thickness. Figure 4.41 shows the effect of metal thickness of the two levels, indicating that only the lower thickness is important (because of the underlying

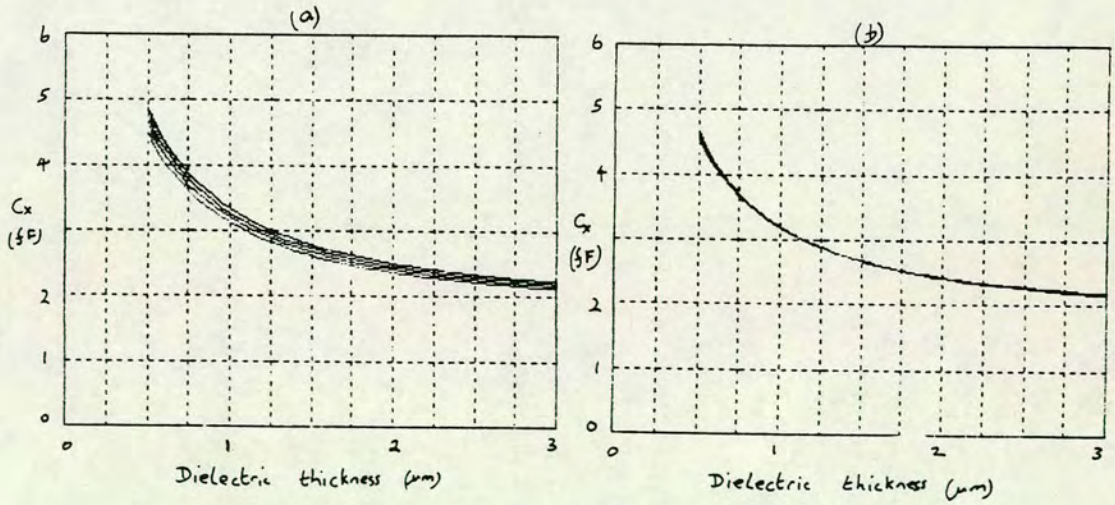


Figure 4.41 Effect of metal thickness (in 0.1 um steps) on crossover capacitance of 6 um tracks:
a) first level = 0.2 to 0.5 um, second level = 0.5 um
b) first level = 0.3 um, second level = 0.4 to 0.8 um.

dielectric constant). Thinning of the dielectric due to planarisation has not been included in this figure.

One important design factor is made clear from these models. If the tracks are grossly dissimilar in widths, the capacitance of the crossover differs depending upon which track is the wider. If the wider track is on the lower level, the capacitance is the lower of the two combinations. This arises from the very high dielectric constant of the GaAs [432] compared with the upper insulating layer. If the narrow track is on the GaAs, it will have a very wide effective width, because the field between the tracks spreads out in the GaAs. When the large track is on the lower level, the proportional increase in effective width is much less. Figure 4.42 shows the capacitances for these two cases, in which the smaller track is constant at 6 um.

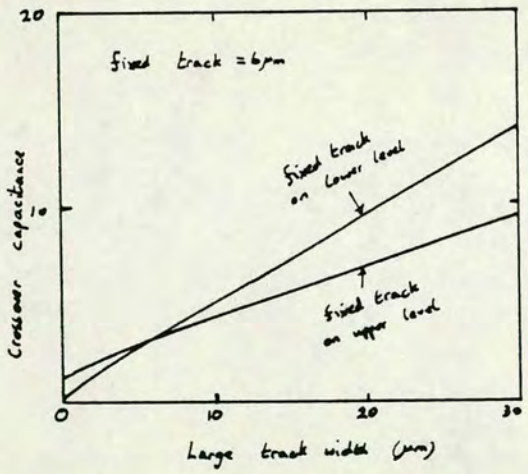


Figure 4.42 Crossover capacitance for dissimilar tracks.

4.4 Discussion

The modelling described in this chapter can neither be described as the most sophisticated nor the most complete work. In particular, the capacitance modelling of the MESFET is limited, especially with regard to the gate-drain component. Furthermore, all of the capacitance modelling would have benefited tremendously from a greater abundance of experimental verification. Some of this work

could have been achieved simply, in other aspects only with much more difficulty. However, the main thrust of the work was directed towards obtaining working ICs, for which this modelling was only one step along the way. The work described here was indeed sufficient to meet these objectives, (at least in a first iteration) as indicated in Chapter 6, and the final results were very good.

Chapter 5. The design of an MSI Circuit in CCL

From the very outset of this research work, the goal was to build a suitable circuit at MSI complexity, to prove both the principles of capacitor coupled logic, and to demonstrate the feasibility of building useful digital circuits in GaAs. This Chapter details the initial specification of the chosen circuit, its design and simulation, finishing with discussion of the measured performance of the fabricated circuit. As such, this Chapter forms the climax of the thesis.

5.1 Specification

An 8:1 multiplexer was chosen for this demonstration primarily because of its usefulness in a variety of telecommunication applications. A circuit of this nature is a key element in achieving high data rate transmission [433], and the parallel-in, serial-out circuit also has significant potential as a building block for local area networks (LANs) [434].

When the project began, a data-rate of 140 Mbit/s was being introduced into the European public switched telephone networks (PSTN) to supplement the widely used 34 Mbit/s systems [435]. Following the conventional CCITT recommended hierarchy adopted in Europe [436], the new data-rate represented a quadrupling of capacity after allowing for the control and retiming functions necessary in a plesio-synchronous system [437]. Also following this convention, systems operating at 565 Mbit/s were just beginning development [438]. The latter represented the limit of what was thought feasible using silicon technology, and it was felt that a further quadrupling to 2.4 Gbit/s was unlikely: instead, a mere doubling of the data rate to 1.2 Gbit/s would be probable. The 8:1 multiplexer thus represented one of the building blocks for a 1.2 Gbit/s prototype system. Because of the performance limitations of printed circuit boards, the merging of eight 140 Mbit/s streams was preferred over the alternative 2:1 interleaving of two 565 Mbit/s data streams.

Taking this specific application, the absence in CCL of low frequency operation is not considered a limitation, as data are deliberately line-encoded before transmission [439], to avoid any problems of DC drift or large bandwidth. Again the 8:1 circuit represents an excellent choice of demonstrator.

With this background in mind, a limited specification for the circuit could be defined. The low-speed inputs were required to be compatible with ECL 10K logic levels [440], so that these could be driven directly from the outputs of standard 140 Mbit/s systems without any interface circuits. Similarly, the frame synchronisation output was required to drive directly into ECL logic, as the control circuit would be implemented in this standard technology. Furthermore, the circuit must work from only a single master clock to avoid problems with phase errors between complementary signals. In the absence of any other standard for high data rate pulses, again ECL 10K logic levels were chosen for both the clock input and the single data output, although it was recognised that this voluntary specification was "open to interpretation" as necessary. The final requirement was that the chip should be powered from a single 5.2 V supply to avoid the need to generate a separate, non compatible power rail.

Although compatibility with ECL 10K has been described, the full specification of this logic family includes a detailed description of the voltage tracking of the inputs and outputs with variations in temperature [441]. At this early stage in the development of a GaAs technology, no account of this could be taken, and only the room temperature behaviour has been assumed for the specification. Clearly, here is an avenue for future development before full incorporation into system designs may be contemplated.

The original specification, as justified here, was for operation at 1.2 Gbit/s. However, the development pace of a CCL technology was slower than anticipated; the first MSI design was flawed and other SSI components were developed to prove the fabrication process before a second design was undertaken. These intermediate components had proven the possibility of achieving a much higher data rate than

1.2 Gbit/s, with other technologies also showing similar advancements during the period. Taking account of this, the most likely data rate for the next generation of transmission systems in Europe is now set at 2.4 Gbit/s - once again falling into the traditional fourfold increase. Although the 8:1 multiplexer function is no longer directly relevant to this application, the circuit was retained for demonstrating feasibility, but the specification on speed was increased to 2.5-3 Gbit/s, thereby widening the possible application of the technology.

5.2 Design

This Section discusses the design of the multiplexer circuit. A wide range of options is available to implement the chosen function, but not all lend themselves to very high speed operation. The first division of this section addresses the design options, indicating why the particular solution was finally selected from this range. The second part of this discussion centres on the translation of this optimum design into a finished layout on GaAs.

5.2.1 Options

There is a range of designs available for implementing a multiplexer function, and the first task is to choose the most appropriate of these for obtaining peak operating frequency.

The first of the choices is between using a parallel-in, serial-out shift register, and using a counter and decoder addressing technique. Figure 5.1 shows the relevant block diagrams.

In the former, all eight inputs are loaded simultaneously into the register, and these are then clocked through at the output data rate. Data appear automatically at the output in the correct sequence, and fully synchronised against the clock, so further retiming is not required. Similarly, the simultaneous loading of the input channels is an advantage in adjusting their relative phases. The shift register can act as a self-latching circuit to memorise the input data, allowing the maximum tolerance on the phase-matching.

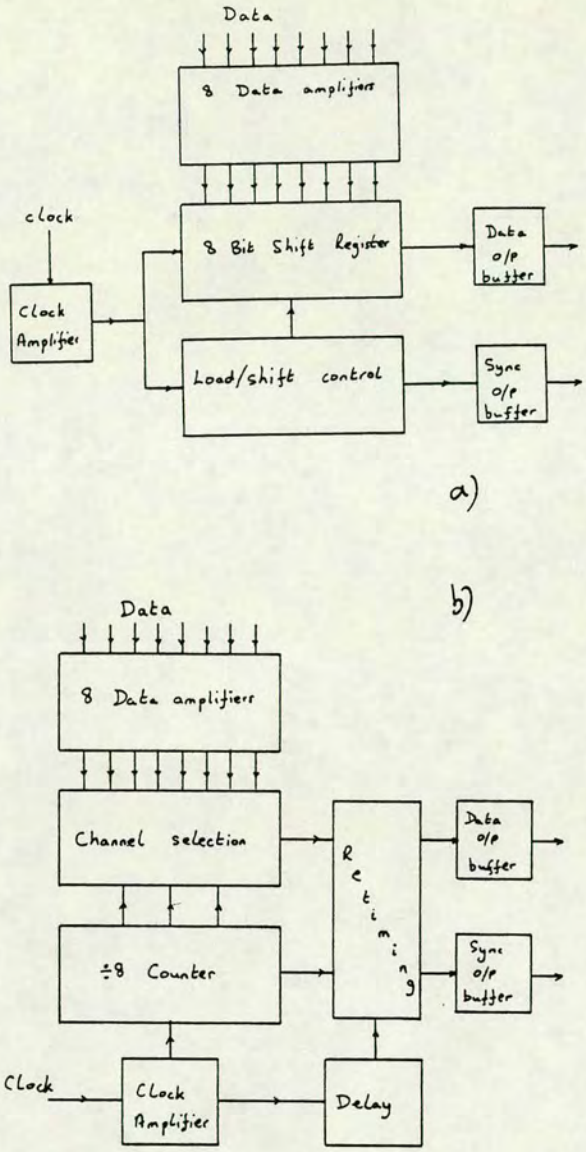


Figure 5.1 Block diagrams of alternative 8-channel multiplexers:
a) using a shift register
b) using a counter and encoder.

The most significant problem is the organisation of the control circuit which switches the register elements from parallel-load into serial-shift. This problem is most readily overcome by using a seven stage register, to load all but the least significant bit (the first out). This bit is fed directly to the output whilst the other data are being loaded, and its input is then disconnected when the serial shift is implemented. Since this single bit must be re-timed to coincide with the other data, it must be passed through a D-latch at the output. The control signal, which will also serve as the frame-sync pulse, is then a single pulse every eight clock periods. Figure 5.2 shows this new configuration in more detail. However, by

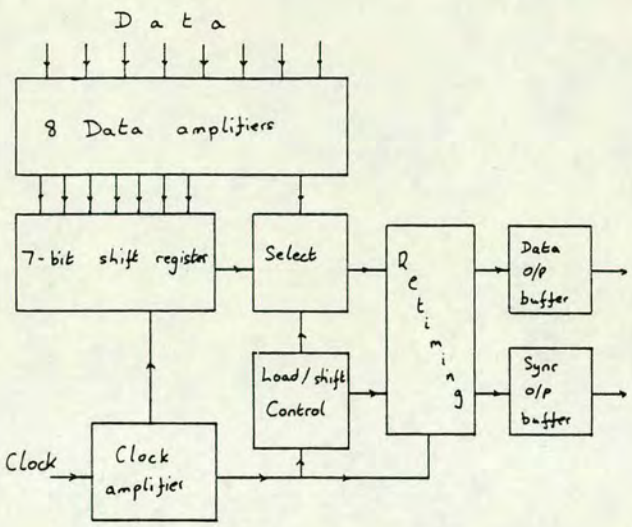


Figure 5.2 Regrouped shift-register multiplexer.

simple element regrouping, it transpires that the output latch duplicates the function of the eighth register bit of the original design. The net result is a circuit identical to the initially conceived eight-channel shift register.

Each element in the shift register is a D-type latch, and superficially two options are available. Either the latch must be transparent (see Section 3.2), or the strobe pulse must be very narrow, even at low operating frequency. The former is preferable, and in CCL, this demands that one of the master-slave configurations be used, although no distinction will be made between the two at this juncture.

If it were feasible to use multi-input gates at the input to the D-latch, both the serial-input with its enable, and the parallel-load with its controlling signal could be merged with latch element. A suitable logic diagram is shown in figure 5.3. This function, [442] although implemented in rival Si ECL technology, and possible in GaAs BFL technology, is not available in CCL, and a much more complex equivalent must instead be used. Complexity is not the key issue however; of much more concern is the additional propagation delay of the latter. Since the combined delay of the latch and the switching logic must still lie within a single clock period, the technology required for this configuration must be much faster than

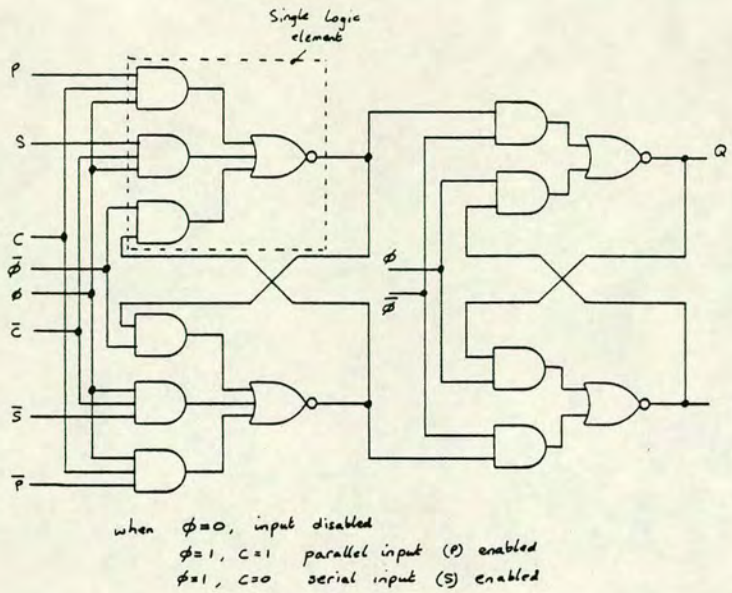


Figure 5.3 Multi-input gate used in a shift register.

that required simply to operate a latch at the clock frequency. For this reason, this construction appears unsuitable for attempting to reach a clock rate of 2.5 GHz.

In the second option, the incoming clock frequency is divided by eight, with the three outputs of the divider used to encode the binary address of each of the eight channels. The address information is combined with the relevant incoming data, and the eight signals merged together in an eight-input gate. Since this eight-input gate is a part of a combinatorial rather than a sequential network, it may be patterned in either a NAND or NOR logic. In the latter, eight inputs can indeed be made in one element, but for reasons of conservatism, the "8-input" gate is likely to be fabricated from several cascaded gates using Boolean combinations. Because it does not lie within a clocked structure (i.e. this gate does not lie within the feedback path), its absolute delay is irrelevant, and the use of multiple, cascaded gates may readily be tolerated.

However, the output stream is not synchronous with the clock because of the various, unmatched internal propagation delays. It should therefore be re-timed in a known relationship with the clock signal

by passing through a D-latch. Although the absolute delay through the address encoding block is irrelevant, the delay relative to the re-timing clock is most important. By artificially delaying the re-timing clock, the pertinent relative delay may be minimised. Consequently, the relationship between the timing of the output signal and the master clock, although regular, may be undetermined. It is therefore necessary to provide an output coinciding with the delayed clock. A further requirement is to provide a signal which unambiguously identifies each channel. In practice, a frame-sync signal is produced for every eight complete clock cycles. Generally this marks the period when channel zero is addressed, and it may be obtained directly from the data-select line for this channel. In an ideal situation, this frame-sync pulse would also be re-timed against the delayed clock, but this is not essential.

In this design, the input data are not latched, all eight inputs being addressed at different times. It is thus quite difficult to obtain the correct input phase for each channel which ideally should be separated by 45^{o} of the master clock, to give the largest immunity to errors.

It is clear from the foregoing that the second method is to be preferred when speed of operation is paramount, although the former method has other significant advantages. Table 5.1 lists the merits of the two approaches.

Table 5.1 Shift register versus direct addressing.

	Shift register	Direct address
Latched input	YES	NO
Fully synchronous	YES	NO
High speed	NO	YES
Gate count	HIGH	MEDIUM

Before dismissing the shift register approach entirely, it is worth considering the block diagram of figure 5.4. Instead of a direct conversion to the full data rate, two separate streams are converted to half speed and then interleaved. As drawn, this design is

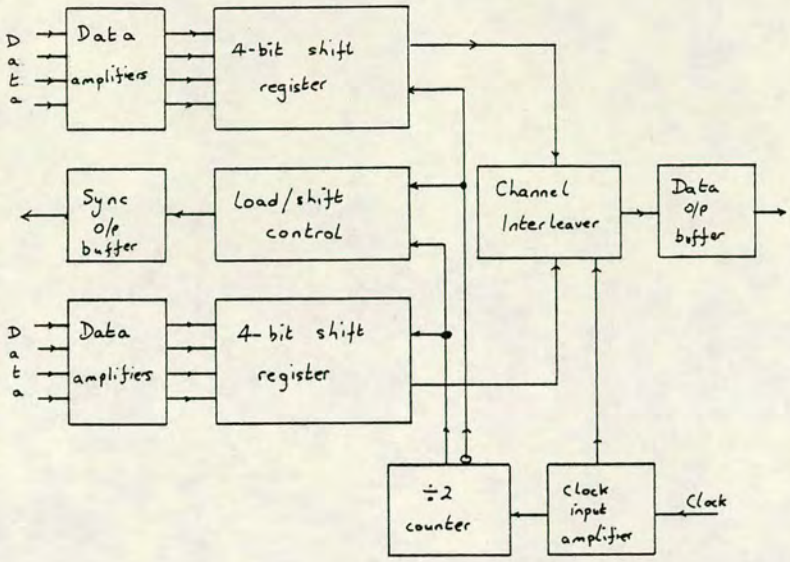


Figure 5.4 The use of two half-speed multiplexers to reduce the speed requirements.

sufficiently general to be implemented in either of the methods. The major advantage is that most of the circuit is only operating at half the output frequency, with just the final stage and the clock divider operating at full speed. In the case of the shift register circuit this advantage is bought for only a marginal increase in the overall complexity, but when direct addressing is used, the extra circuitry is not a trivial overhead. This technique is thus more readily justified if the shift register is used. In practice, this method should be considered seriously when the input latching and synchronous operation are thought worth the premium, but not otherwise.

Having selected direct addressing as the most appropriate technique, it is necessary to choose a type of divider. Three options spring to mind, a ripple counter, a synchronous divider and a ring counter. Schematics of all three are shown in figure 5.5. Within each of these options, it is important to remember the two available latch circuits, and their differing properties. The important characteristic of the divider is that eight, clean, non-overlapping select pulses should be produced. (Strictly, these pulses may

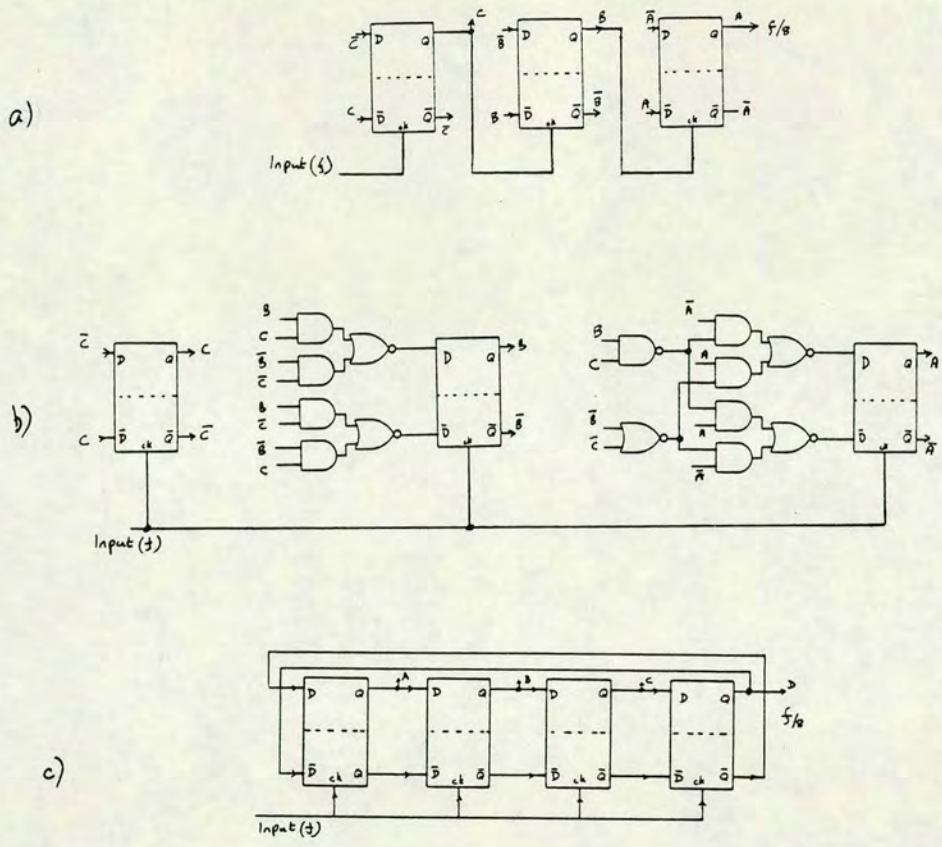


Figure 5.5 Various modulo-8 counters: a) ripple cpunter, b) synchronous counter, c) ring counter (Johnson counter).

overlap if the timing on the final latch is correctly set-up, but non-overlapping pulses allow a greater tolerance in the timing.)

This specification clearly rules out the simplest circuit - the ripple counter. Here, the disparity of 2t (twice the propagation delay) between the timing of the three binary bits is a very significant fraction of the maximum clock period (2 or 4t depending which circuit is used), and the divider must operate well below its theoretical maximum speed.

The conventional circuit of a synchronous divider is similarly problematic. Here, the outputs change in unison, so there is no problem in obtaining the clean selection pulses, but once again the circuit is prevented from operating near the maximum toggle frequency of the D-latch because of the amount of logic required to generate

the input to the lowest significant bit (LSB). It is ironic that the synchronous approach suffers from the delays associated with the slowest element!

The circuit shown in figure 5.5 for the synchronous divider has been obtained by assuming that the counter follows a true binary sequence. There is no justification for this, other than it being the most obvious first choice. By changing the count sequence from binary, the counter construction can apparently be simplified by reducing the amount of logic required, as shown in figure 5.6. In figure 5.7, this count sequence is contrasted on the Karnaugh map with that of the binary counter. One is tempted to argue that the simplification derives from the reduction in number of times the counter jumps to a non-adjacent cell on the map. Following this argument to the extreme, perhaps the Gray code [443], also included in figure 5.7, is the ideal count sequence. This code is translated into a logic diagram in figure 5.8. Indeed, the symmetry of the code translates directly into a symmetry of logic, and although figure 5.8 may appear more complex than figure 5.6, the performance of this counter will certainly be the best of all the three synchronous counters. More

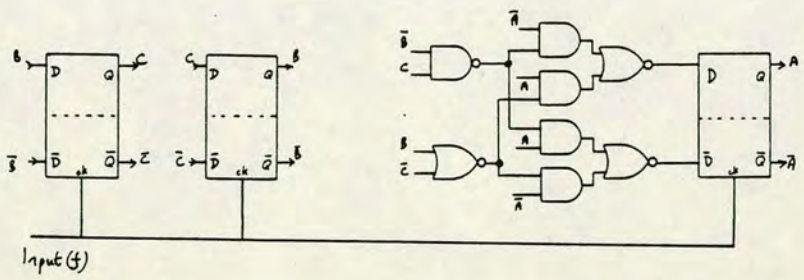


Figure 5.6 Synchronous modulo-8 counter (non-binary).

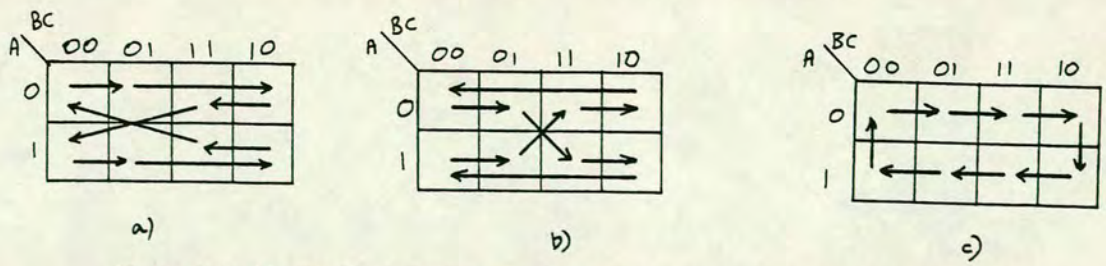


Figure 5.7 Karnaugh transition map of modulo-8 counters: a) binary counter b) modified count sequence (figure 5.6), c) Gray counter (simplest count sequence).

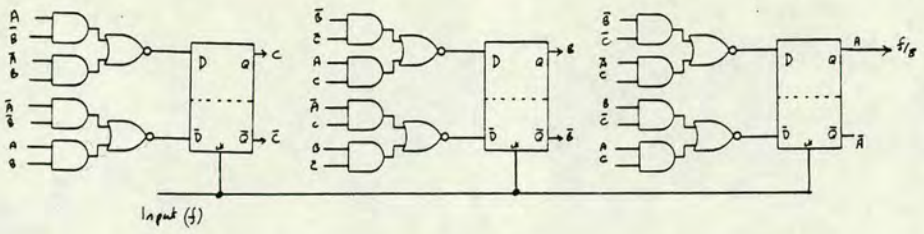


Figure 5.8 Logic diagram of Gray code modulo-8 counter.

logic is performed, but because this is evenly distributed on the three latches, the worst case path is significantly improved (all paths are in fact equal, so the worst case is also the best!).

Despite obtaining a significant improvement by challenging the assumption of a binary count sequence, the synchronous counter is still unable to perform close to the potential operating speed of a single latch. Instead of the clock period being fixed at the 2t (or 4t) of the latch (depending upon design), at least a further 1 or 2t must be added, (probably nearer to 2t if the clock is expected to have equal mark:space ratio).

The final circuit choice is the ring counter. The most natural choice for a divide by eight ring counter is the four stage twisted ring (or Johnson counter [444]) as depicted in figure 5.5. In this circuit, four (rather than three) latches are needed but no additional logic is required. Clearly, therefore, this counter can operate at the maximum speed of the individual latches. With four outputs rather than three, this counter does not operate in a conventional binary manner, even after allowing for arbitrary count sequences, as suggested earlier. Decoding into the eight individual select pulses is therefore different. It is in fact easier because only a dual input gate is required for each of the eight counts. Figure 5.9 shows the count sequence and its decoding.

The redundancy introduced by the extra latch means that only a partial sequence is executed. It is therefore important to ascertain what happens if the shift register enters one of the unused states. In fact, this four stage counter has two legal sequences, each of

A	B	C	D	Decoding
0	0	0	0	$\bar{A}\ \bar{D}$
1	0	0	0	$A\ \bar{B}$
1	1	0	0	$B\ \bar{C}$
1	1	1	0	$C\ \bar{D}$
1	1	1	1	$A\ D$
0	1	1	1	$\bar{A}\ B$
0	0	1	1	$\bar{B}\ C$
0	0	0	1	$\bar{C}\ D$
0	0	0	0	Repeat

Figure 5.9 Count sequence of 4-stage Johnson counter.

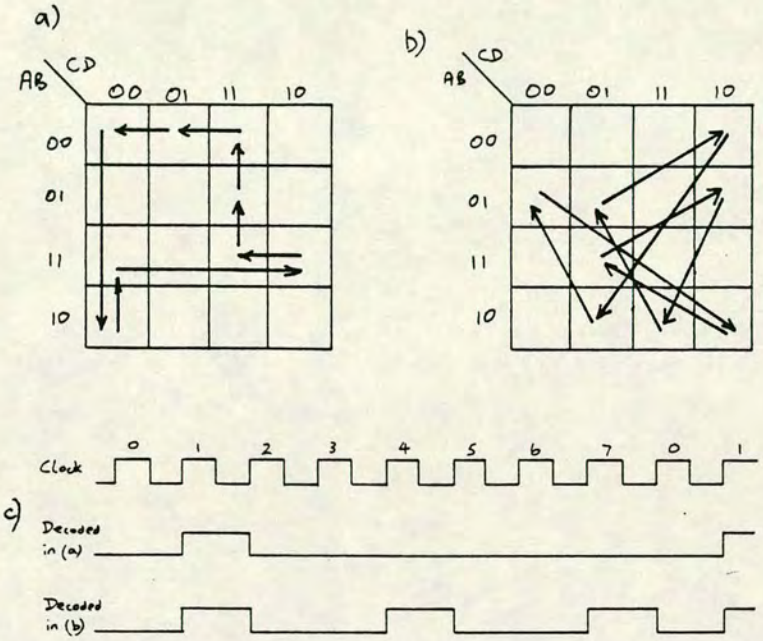


Figure 5.10 Johnson counter sequences:
a) correct sequence
b) incorrect sequence
c) the result of using a pair of inputs to decode the eight outputs from a Johnson counter.

eight cycles. Unfortunately, if the counter gets into the wrong sequence, then it will remain there. The NAND-gate decoding used to generate the eight select lines from the valid sequence will actually select each line three times out of every eight if the incorrect count sequence is in force. Figure 5.10 shows the incorrect count sequence, and the erroneous results of the decoding.

There are three possible solutions to this problem. The first is to use one of the eight decoded outputs as a reset line, and then to force the shift register into the subsequent state, instead of allowing it to free count. Thus one might choose the state decoded by ($D.\bar{C}$), and use this to force the state 0000 on the next clock cycle. In this case, a reset or clear input to each latch would be required. Unfortunately, this is not available within the self-imposed CCL design rules.

An alternative is to force the A input to zero either when D=1 or when (B=0 and C=1). In the correct count sequence, this would have no effect, but in the incorrect sequence, the false count 0010 would be succeeded by 0001, a valid count. This measure would require some additional logic within the ring, which would have a deleterious effect on speed.

The third, and most efficacious measure is to change the decoding. It is possible to decode eight individual channels correctly from either of the count sequences using a three input gate rather than the two input gate. Figure 5.11 shows these configurations. Now there is no requirement to designate a correct and an incorrect sequence - either are correct. If an error occurs which transfers the counter from one sequence to another, then it is possible temporarily to lose data. However, in the application envisaged, errors can occur for this or many other reasons and there is an established procedure for recovery from occasional errors [445]. This error detection and system recovery is accomplished using the lower frequency demultiplexed signals.

Correct sequence				Incorrect sequence				
A	B	C	D	A	B	C	D	Decoding
0	0	0	0	0	1	0	0	$\bar{A}\,\bar{C}\,\bar{D}$
1	0	0	0	1	0	1	0	$A\,\bar{B}\,\bar{D}$
1	1	0	0	1	1	0	1	$A\,B\,\bar{C}$
1	1	1	0	0	1	1	0	$B\,C\,\bar{D}$
1	1	1	1	1	0	1	1	$A\,C\,D$
0	1	1	1	0	1	0	1	$\bar{A}\,B\,D$
0	0	1	1	0	0	1	0	$\bar{A}\,\bar{B}\,C$
0	0	0	1	1	0	0	1	$\bar{B}\,\bar{C}\,D$
0	0	0	0	0	1	0	0	Repeat

Figure 5.11 Improved decoding from a 4-stage Johnson counter.

It is interesting to note that the ring counter will always start in the same fixed state if the faster of the two CCL latches is used with a pre-starting sequence. Whilst this erstwhile disadvantage has once again become an advantage, it should not be considered sufficient to assume that the counter will always remain in the correct sequence, as there is always the possibility of a noise induced error shifting operation into the other state. The circuit should always be designed to recover from such an error.

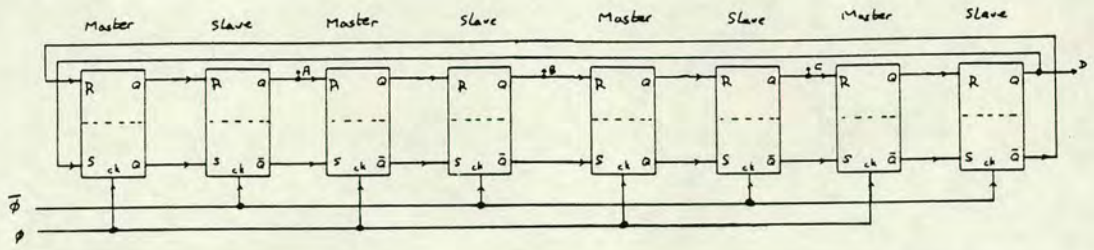


Figure 5.12 4-stage Johnson counter with master-slave gates expanded.

Thus, it appears that the ring counter, suitably decoded, is the only counter capable of fulfilling the high speed requirements of this application. However, there is one further avenue still to explore and this will in fact prove the most useful of all. The counters discussed to date have all operated only on one half of the clock cycle. When trying to optimise performance, it is worth exploring circuits which respond to both the rising and falling edges (or the logic-0 and logic-1 levels) of the clock, allowing the clock speed to be halved. This is related, though not identical, to the splitting of the 8-bit shift register into two four bit registers discussed earlier.

In fact, master-slave D-types have been assumed throughout these discussions, and these do respond on both clock edges, but the internal nodes between master and slave are usually forgotten. If the ring counter is expanded to include both master and slave, when the 8 pairs of nodes are analysed on both clock edges, 16 individual states can be identified. The circuit, as expanded in figure 5.12, is actually an 8-stage twisted ring counter, not simply a 4 stage counter. Similarly, if any of the synchronous binary counters is fully expanded, 16 states can be identified.

This argument suggests that either a four stage twisted ring-counter, or a two bit binary counter will suffice to produce the 8 selection line signals. If we follow the guidelines already laid down, then the Gray code counter will be the most efficient of the sequence counters. Figure 5.13 shows this counter (note that the circuit of figure 5.6 used this for the two LSBs). As it turns out to be

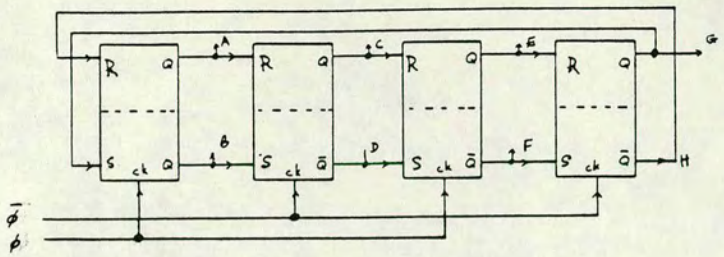


Figure 5.13 Final choice of counter.

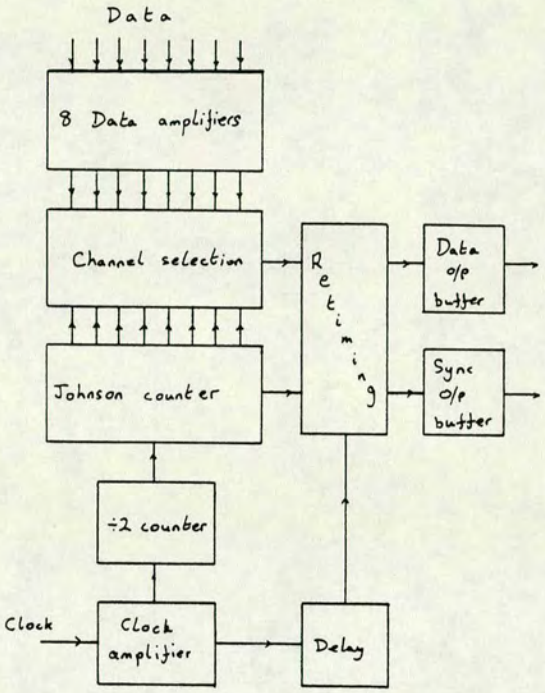


Figure 5.14 Revised multiplexer block diagram.

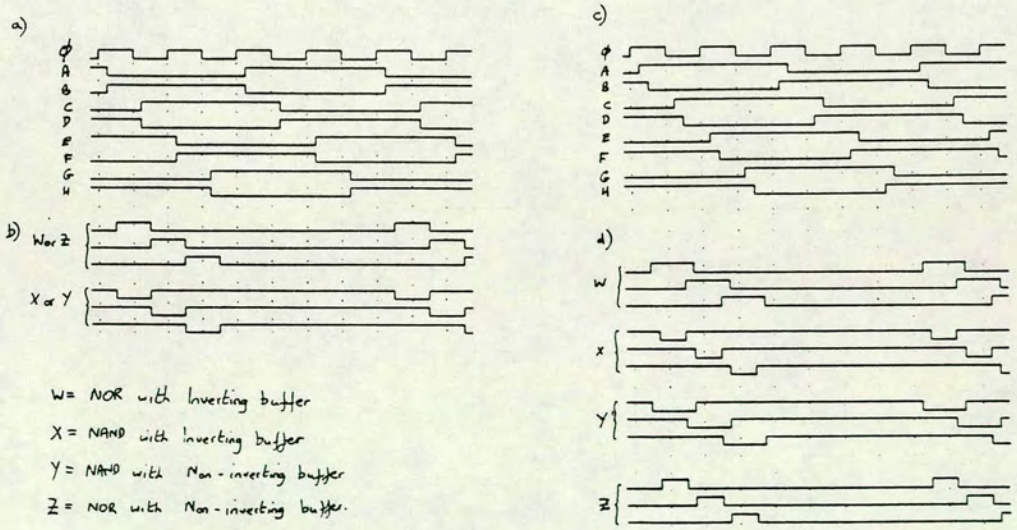


Figure 5.15 Schematic waveforms from Johnson counter:
a) AND-NOR counter, b) NAND counter
c) encoded waveforms from (a), d) encoded waveforms from (b).

identical to the four stage twisted ring counter, there is no comparison to make. It is useful to have approached this circuit from both angles, as further analysis is simplified. Because it is a ring counter, decoding can be performed using just dual-input gates, but because it is a full binary-type counter, there are no redundant states and no illegal sequences. Thus, this circuit apparently has the best of all worlds: it is only clocked at half speed, has the capability of the highest possible speeds available with the technology, has simplified decoding, but does not have illegal count sequences. This then **must** be the correct choice of counter.

If this half-speed counter is used at the core of the multiplexer, additional elements are required to make up the full 8:1 function. Figure 5.14 shows a block structure of the new circuit. The incoming master clock follows two paths: it is halved in a divide-by-2 prescalar before being passed into the frequency divider, but also drives the output retiming circuitry via a delay line whose delay is chosen to match the propagation path through the divider, selector and encoder circuitry to that through the output latch.

The output sequence of the Johnson counter is shown in figure 5.15 for implementation in both AND-NOR and NAND logic. It is preferable to buffer the outputs from the counter, to avoid excess fan-out loading, and this may be with either a source-follower (as a non-inverting buffer) or a conventional inverter stages. The extra drive capability of the quasi-complementary stage is not required. Because of the symmetrical nature of the counters, either a NAND or a NOR function may be used to derive the select lines. The effect of the inverting buffer following the NAND is the same as the non-inverting buffer with NOR, and vice versa. Figure 5.15 also shows schematically the quality of the gating signal produced from the two counters for the two logical implementations. This figure was derived in the absence of clock skew, and the clear advantage of the AND-NOR configuration can be seen. The decision between the two types of latch is thus made on the quality of the gating signal alone.

The choice of the AND-NOR configuration of latch also gives a higher speed, but does demand the extra control circuitry required for the pre-charge sequence. The adoption of this circuit also ensures that there is a means of starting the circuit in a known state, and a means of diagnosing performance, through the self-oscillation mode of the divider. Because there are three control circuits, associated with the main divider, the pre-scalar and the retiming, all parts of the circuit can be isolated for individual diagnosis in the event of circuit failure. This is a most important consideration for the combination of new design and new technology. In the early (failed) design, the control lines for each of these circuits were indeed separately accessible, but this was deemed an unnecessary luxury on the final circuit where pin-out constraints were applied more vigorously. Instead, some flexibility was retained by allowing the output latch to be clocked externally via a separate input gate which could be disabled.

Having established the main timing path through the multiplexer, there is yet another choice to be made, to decide how to pattern the data selection and encoding. The eight signals available from the divider are shown in figure 5.15. If the input data are labelled a-h, and the eight divider phases A-H, the Boolean function required from the encoder is:

$$\text{output} = aBD + bCE + cFH + dBG + eAC + fDF + gEG + hAH.$$

This can be implemented in many ways, depending upon the amount of logic performed in each individual gate and upon the allowed combinations. As the circuit is to be implemented in CCL, thought must also be given to the unused inputs. In operation, the requirement is that all channels should operate above some minimum frequency. However, it is helpful to be able to test the circuit by stimulating it with a reduced number of inputs. A measurement requiring just a single input is by far the simplest to configure. When NOR gates are included in the encoding chain, all inputs to each NOR gate must be exercised, but where NAND gates are used, it is permissible to "float" some inputs. It may therefore be more practical to expand a NOR logic function in its NAND equivalent.

There is a compromise between encoder complexity and the minimum number of input vectors required during testing to ensure correct operation.

Since truly complementary data are available from the divider, there is theoretically no preference of either NAND or NOR logic at the input to the encoder chain. However, if some skew is present between the two clocks, the output of the AND-NOR counter will no longer show perfect complementarity, instead degrading so that the outputs become much more like those available from the NAND type of latch. Intuitively, wider but overlapping gate pulses appear more attractive than very narrow, non-overlapping signals, so the decoder would be chosen with this as a preference. As the inverter buffer is to be preferred over the source follower because of its better switching characteristics, this would imply that NOR logic be used to generate the select pulse. One should also decide whether or not to generate the individual select pulses before or during the merging with the channel data.

The other parameters of interest in the overall design are the power dissipation (related to the total gate count), the potential yield hazard (related to the total FET gate width), and the effective fan-out loading presented to the ring-counter.

In total, some 17 different encoder configurations were studied and table 5.2 compares the parameters associated with these designs. Not all possible designs were studied: the greater the minimum number of channels required, the fewer the options considered. Figure 5.16 shows some of them.

The final decision was made after comparing the results from table 5.2 and answering the following questions:

i. should the gating pulse be generated before merging with data?
ii. if so should it be NAND or NOR?
iii. is single channel operation essential?

Although there are some neat configurations using untried logic gates, particularly circuit 13 which uses a 2-wide, 3-input OR-NAND,

Table 5.2 Comparison of encoding techniques

Design:	Gating:	Fan-out:	Channels:	Current:	FET width:	Verified:
1	NAND	4	1	29	152	YES
2	NOR	2	1	29	132	YES
3	NAND	4	1	17	116	YES
4	NOR	2	1	17	96	YES
5	NAND	4	2	25	113	YES
6	NOR	2	2	25	93	YES
7	NAND	4	1	37	153	YES
8	NOR	2	1	37	133	YES
9	none	2	1	29	115	YES
10	none	6	1	21	162	NO
11	none	2	2	17	75	YES
12	none	6	2	25	151	NO
13	none	4	1	13	66	NO
14	none	6	2	9	117	NO
15	NAND	4	2	23	133	YES
16	NOR	2	2	23	113	YES
17	none	2	4	11	56	YES

Key: gating = method of pre-generating the gating pulse
channels = minimum inputs required during testing
current = number of unit width load transistors
FET width = number of unit width transistor gates
verified = YES if all logic elements used are already tested

these were deemed too risky, even though this particular example does not break the ad hoc rule to limit NAND inputs to two. Preparation of the gating signal prior to mixing with the input was thought to be highly desirable, noting also that the greatest advantage in NAND vs NOR derives from reduction in both loading and total gate width in the NOR option. Option 4 was thus chosen as the best overall compromise. It should be emphasised that the circuits considered seriously were not restricted simply to those offering single channel operation, but that the one selected is clearly better than the alternatives in most respects.

In deriving table 5.2, the requirement for a frame-sync pulse was ignored. Having selected an option with the gating signal derived separately, one channel is already available for this pulse. The complexity of adding the frame-sync is limited to the extra gates required to pad the delay as necessary. If pre-gating had not been chosen, extra logic gates would have been required to obtain the

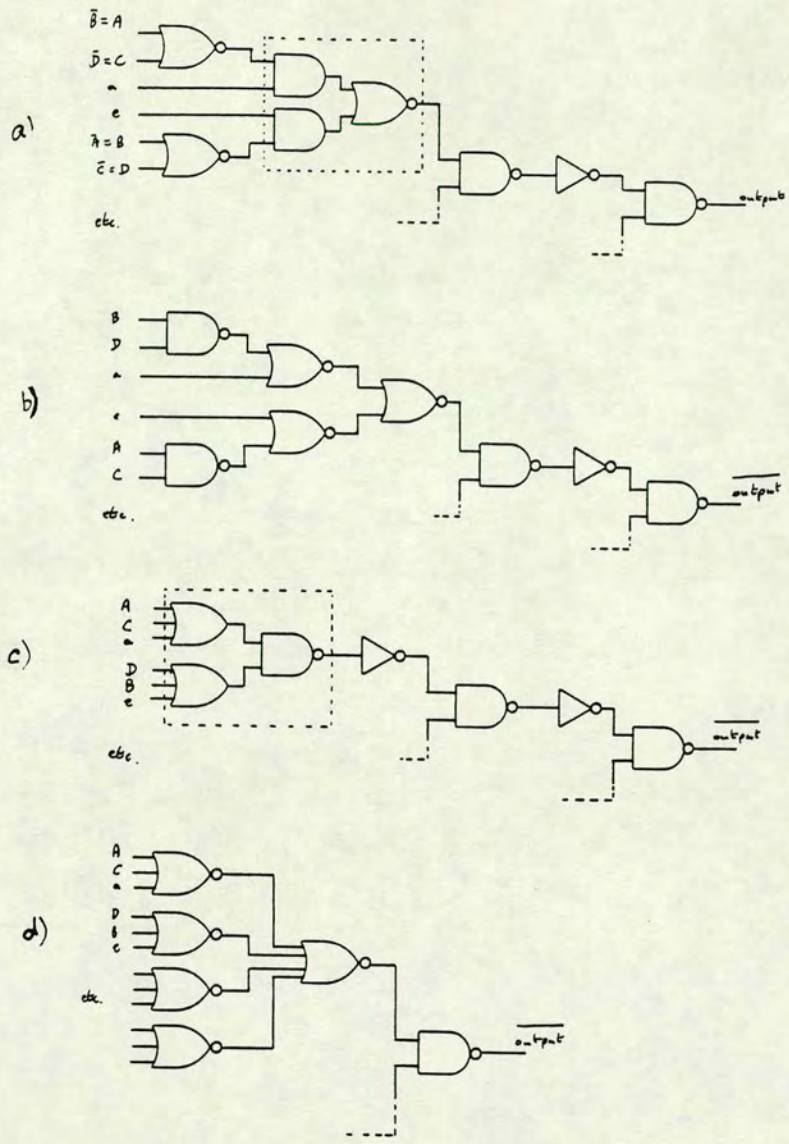


Figure 5.16 Various implementations of the channel encoding:
a) option 4,
b) option 5,
c) option 13,
d) option 17.

desired signal, so the penalty associated with demanding pre-gating is actually less than that shown in the table.

One final option remains: how to implement the output re-timing. As already hinted, having selected the AND-NOR solution for the Johnson counter, there is no further penalty to pay in adopting the same solution for the output D-latches, but there is considerable advantage (Chapter 3), and once again this type of gate is adopted. However, the final timing arrangement depends upon the clock as well as on the latch. The major design problem was that there were

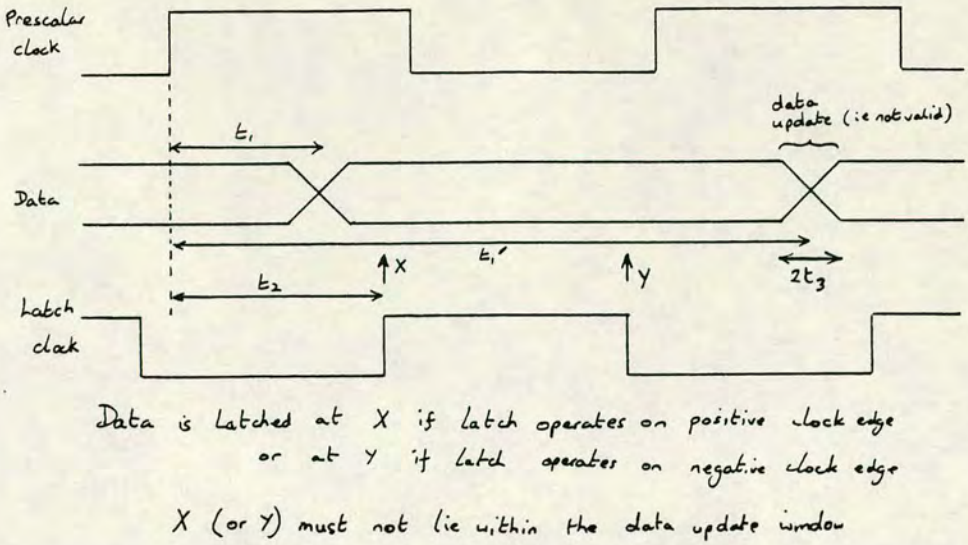


Figure 5.17 Critical timing path for correct latching of data.

insufficient data on the propagation delay characteristics of the logic elements, yet it was necessary to match the delay through the clock path quite closely to that through the data path.

Although the simulation models were much improved at this stage, the accuracy of prediction was still too low. Process variations were still quite large, and experimental data were also inadequate to the task. Furthermore, as a feasibility study, the design had to cope with a very wide margin of threshold voltage and grossly different propagation delays. An approach was needed which could tolerate these wide variations.

The problem is illustrated in figure 5.17. The clock input ϕ is defined as that to which the prescalar responds on the rising edge. The relevant data appear at the output latch after a time t_1. At low speed this lies within the same clock period, but at high speed this may be several clock periods later. The latch is activated by its clock at a time t_2 later than the prescalar. To ensure functionality, the latch must not be activated during the data uncertainty period, which is defined as $+/-t_3$ from the switching point. We may therefore define a set of bounds on the delays:

$$(t_1 - t_2 - t_3) < (n/f) < (t_1 - t_2 + t_3)$$

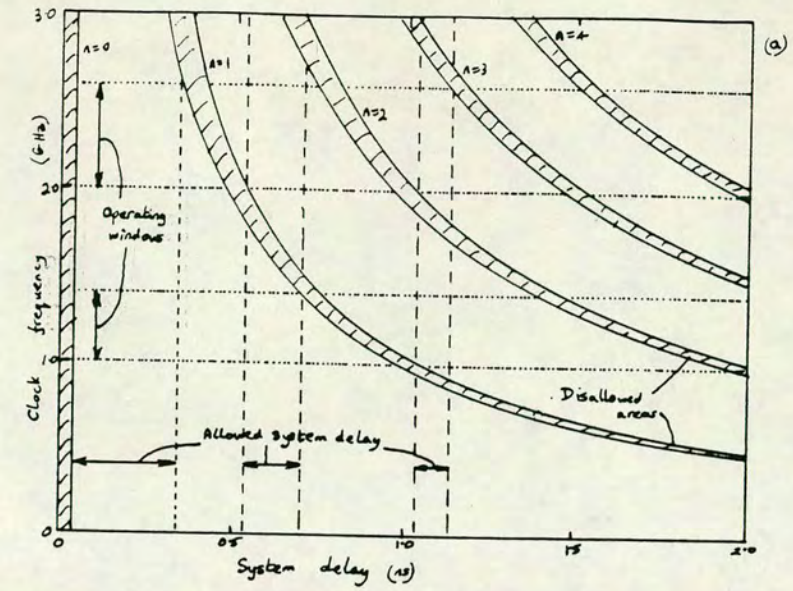
where n is an integer and f is the clock frequency. The integer n indicates the fact that the data may appear correctly latched, but several clock cycles later than it appeared at the input.

A second condition obtains if the latch is deliberately clocked on the negative going edge, rather than the positive edge. Now, (n + 0.5) replaces (n) in the above equation.

The effect of these equations is shown in figure 5.18 when clocking from either edge. A data-invalid width of +/- 30 ps was selected, and the regions of erroneous operation are marked. For the application discussed, the circuit is required to work over the region 1 to 1.4 GHz and 2 to 2.6 GHz, and these are therefore plotted on the graphs. The third highlighted region denotes the acceptable delay between the data and clock paths for the circuit to function correctly over these two frequency windows.

The ideal design would target on a delay centred on 0 ns using opposite edge latching, or centred around 0.2 ns when latching off the same edge. However, to hit either of these targets, the delay padding in the clock lines must be quite long, consequently with a large associated error. Also, such a large padding delay is adding unnecessarily to the chip complexity, thereby increasing power consumption and decreasing yield. Two alternative windows are attractive: those centred at 0.87 and 0.62 ns. The former has two non-operating regions within the total frequency spectrum, and the latter has only one.

In order to guarantee operation at the frequency of interest, it is necessary to predict the relative delays of the two paths to within about 150 ps or 80 ps depending on which of these four windows is targeted. As this is not possible, another approach has been adopted. Two different operation modes have been allowed, programmable from a single logic input, with the default mode aimed at the centre of one window requiring no external control, provided that the simulation data are correct.



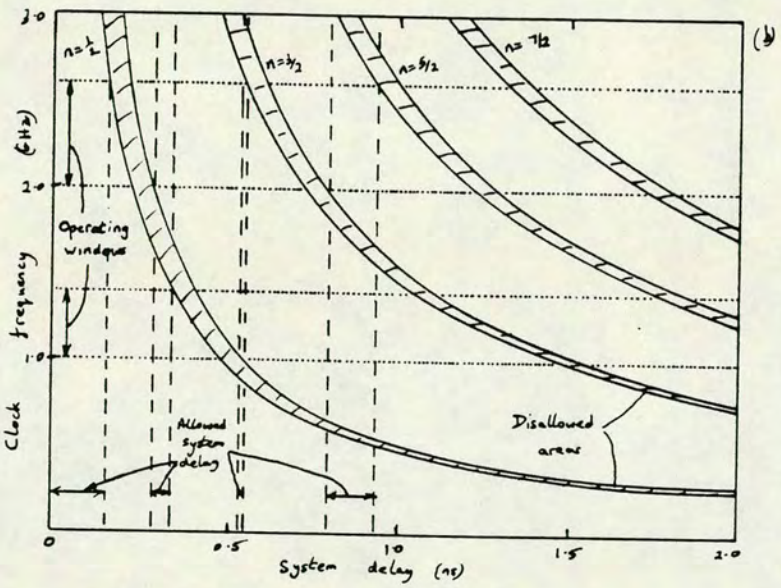


Figure 5.18 Timing analysis for multiplexer operation:
a) latch and prescalar off same edge
b) off opposite edge.

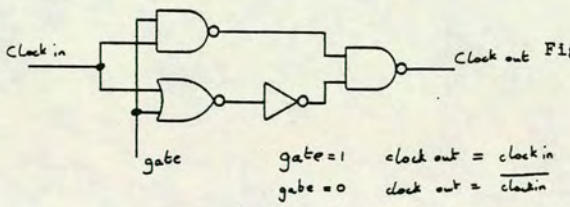


Figure 5.19 Circuit to allow switching of operation of latch by 180°.

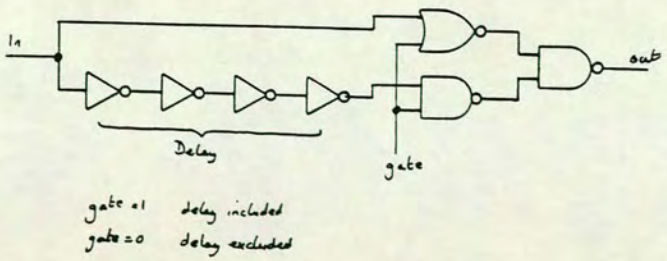


Figure 5.20 Circuit to allow variation of the delay time. (Note, as configured this will only work with CCL - see text)

Two alternative techniques are available to allow this mode switching. The first is simply to switch clock edges for data latching, thereby switching from one to the other of the two graphs in figure 5.18. The second is to retain the clock phase, but to switch the relative delay.

If the clock edges are to be switched, this can be achieved with the circuit of figure 5.19. Unfortunately, this is not simply edge switching, but also shifts the overall delay by one propagation delay. This makes the separation between the error states of the two modes of operation less than ideal.

A suitable circuit to switch the delay is shown in figure 5.20. An even number of gates should be switched out to retain the clock phase, and the quantity chosen should be selected to improve the chances of hitting one of the target windows. More pedantically, this delay should be chosen to ensure that if the default mode coincides with the erroneous operation at the frequency band of interest, then the second mode should guarantee missing the errors. Thus, the optimum delay is 200-250 ps, obtained by subtracting the time of the lower edge of the error region at 2.6 GHz from that of the upper edge at 2 GHz. This figure coincides well with the delay from four relatively slow gates.

The switched delay was adopted as the final solution, as operation at all frequencies is guaranteed. The 0.62 ns window was selected as the best compromise between circuit complexity and operating range, and this was targeted for the default, with the four inverters being switched out of the clock delay line in the secondary mode.
Figure 5.20 shows the finished circuit used to control the switching. Some comment on this is called for, as the control input is DC, despite the logic family operating only above 100 kHz. In the default mode, the control input is tied to the V_{dd} rail through an input autobias circuit. This will allow the NAND gate to operate but will disable the NOR gate by turning one of its input fully-on. When a negative bias is applied to the input, the NAND gate is cut-off even at DC, and the NOR path activated. Thus, if the non-default mode is required, the provision of an additional negative voltage is

necessary. Whilst this would be unacceptable in a finished system, it is an acceptable penalty to pay to guarantee operation, especially in a prototype.

5.2.2 Layout philosophy

The various circuit options have now been explored, and a basic overall design has been established. This section will treat the principles on which the circuit was engineered, using the top-down approach.

Table 5.3 Pad count for the 8:1 MUX

Function	Req on chip	Req in package	Note
Master-clock input	3	2/3	1,2
Data inputs	17	9/17	1-3
Sync output	2	2	4
Pre-scalar output	2	0/2	4,5
Counter output	2	0/2	4,5
Data-out latched	2/3	2/3	4,6
Data-out unlatched	2	0/2	4,5
Chip V_{ss}	1	1	
Chip V_{dd}	1	1	
Control lines	2	2	
Programmable delay	1	1	
Ext latch clock amp	4	-	1,2,7,8
Total	39/40	20/40	
Test transistor	4	-	
Ring oscillator	5	-	

Notes: 1 each input needs amplifier, reference and separate V_{dd}
2 See text for input amplifier V_{dd}
3 Data amplifiers can share a reference
4 Each output needs a separate V_{dd}
5 These outputs are optional
6 Complementary data sharing a common V_{dd} are available
7 Also requires an enable signal
8 This input is optional

A number of external constraints should be introduced at this point. Firstly, it is of no value to design a high speed circuit without any possibility of operating at the design speed. One of the factors which is often neglected, but which may seriously influence the final performance, is the package. The number of suitable choices is very restricted when operation at 2-3 GHz is expected [446]. Thus the

first external constraint is testing. To examine this area, it is necessary to define what is required for operation, and obtain a rough count of the number of pads, as listed in table 5.3

A second outside influence should be introduced at this point. A number of suitable structures for process verification and device parameter extraction already exists, and all of which are designed with a chip size of 2x2 mm^2. Furthermore, to retain complete compatibility, the lower left corner of the chip must be occupied by the test transistor.

Coupling these two factors together, it is possible to fit the required number of pads (40 plus the test transistor) onto a target chip size of 2x2 mm^2 using pads of 80 um^2 spaced at a pitch of 160 um. Of the packages suitable for high speed use, the leaded or leadless chip carriers, or the flatpacks offer the best performance characteristics [447]. The leadless varieties of chip carrier pose significant problems for non-destructive testing prior to "shipment" to customer, so the leaded varieties are to be preferred. Amongst these classes of package, two types were readily available, both with 24 pins. One of these, a 24 pin leadless chip carrier was well suited for a 2 mm^2 chip, having a cavity edge of 4.5 mm. The other, a 24 pin flat pack offered better testing prospects, but having a cavity edge of 6.3 mm, it required over-long bond wires if the die were to be centrally placed. There was no prospect of obtaining a new custom-designed package for preliminary tests, although better results might have been anticipated. Both these packages were square, with six bonding islands per side internally, although the latter package was arranged with eight and four lead-outs along the external edges.

Thus, taking all considerations into account, the target of a 2x2 mm^2 chip still seemed sensible, but with a larger die size also acceptable if necessary. Only 24 leads would be simultaneously accessible in a fully packaged version, thus loss of some of the diagnostic features after assembly would have to be tolerated, as any attempt to increase the pin count would only increase the package size, threatening yet more deterioration in performance.

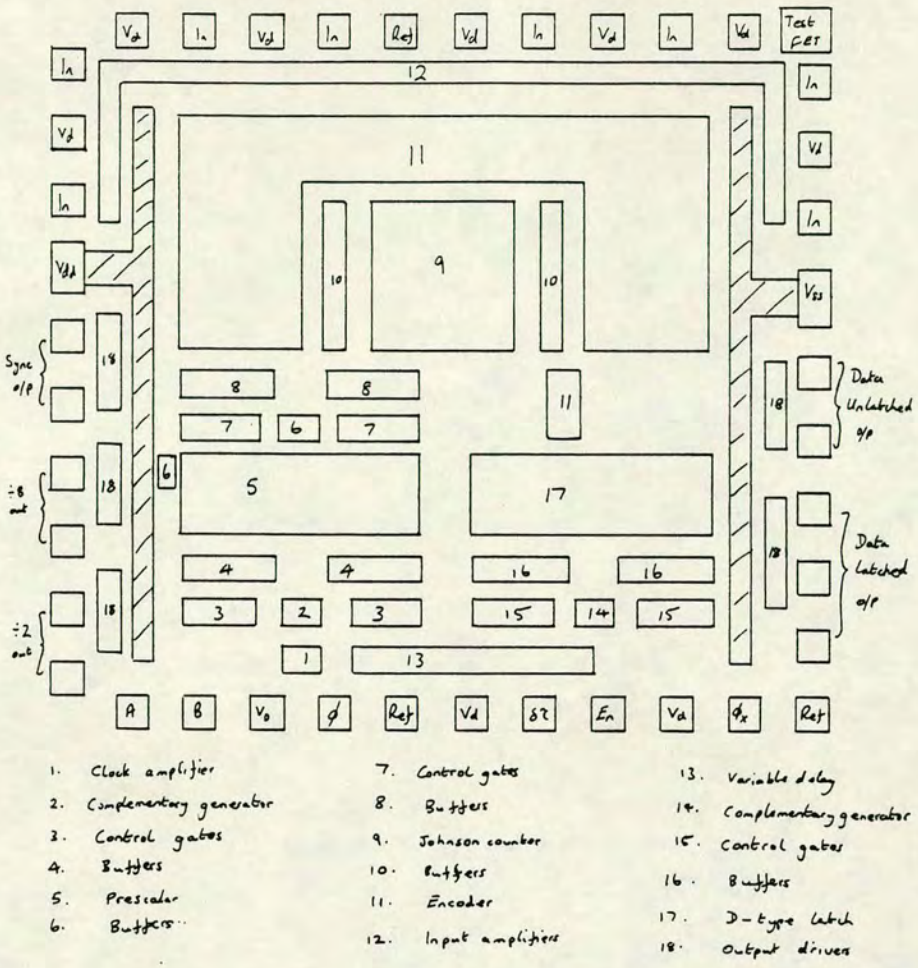


Figure 5.21 Preliminary floor plan and pad layout of the multiplexer.

In order to bond from this chip into either of these packages, the 24 signal lead-outs had to be suitably distributed around the chip to give six per side. In order to achieve this, a new block diagram was produced, physically locating each function within the square chip area. Preliminary layout of some of the more crucial elements was performed prior to this activity, in order to assess their area requirements. This was particularly important for the twisted ring counter at the heart of the multiplexing function. At this early stage, a rough attempt was made to balance the propagation delay through the tracks where this was likely to prove important. Figure 5.21 shows the area and pad allocation at this stage.

One matter of great concern in this design was the avoidance of oscillation and feedback due to cross talk. Thus, wherever possible, signals with low RF impedance to ground were interspersed between genuine RF inputs and outputs. Similarly, as already indicated in table 5.3, separate V_{dd} supply lines were used for the input, the output and the bulk of the circuit. Notice that the design of both the I/O stages makes provision automatically for the V_{ss} inputs to be separated in similar manner.

The grounding of the capacitor coupled logic circuit is interesting. Normally in logic circuits, the ground is chosen to be that supply rail which is less immune to noise. The lower impedance of the ground rail is used to minimise the uptake of noise on this input. One would initially expect the CCL circuit to be more susceptible to noise on the V_{ss} rail, because of the similarity with nMOS [448]. However, the ability of the capacitor to hold the transistor gate well below threshold suggests that small changes in V_{ss} will not affect the output state of the transistor in the vulnerable output-high state. Changes in either rail voltage will alter the output low, but only as V_{dd} decreases or V_{ss} increases. In a way, the capacitor decouples the circuit from the immediate effects of noise. Just as it was impossible to define a sensible noise margin for the logic gate, it is difficult to assess the effects of power-rail noise on the circuit, but a qualitative argument would suggest that either rail is suitable for use as ground.

The same argument does not follow either at input or output. At input, the most susceptible node is the reference node, which cannot be used as ground. Since this input amplifier is capacitively coupled to the succeeding stage, noise here is not a problem. Choosing the V_{ss} as ground would lower the effective impedance of the reference node, but would be of little other benefit. Choosing V_{dd} as ground ensures that each of the amplifiers is well decoupled from the others. Since the output consists of a capacitively coupled transistor gate, with both drain and source open, the concept of a ground cannot be applied. However, in both cases, ECL compatibility is provided only when V_{dd} is grounded together with the V_{cc} of the ECL.

Overall therefore, it is logical to ground the V_{dd}. Full advantage of this can be taken by using the die cavity as a groundplane, and bonding each of the amplifier V_{dd} lines to this backplane. Thus, each input signal bondwire is screened by a grounded wire without recourse to multiple external grounds. Clearly this is less effective than using a fully screened package, but it is still quite helpful. The block layout of figure 5.21 takes this into account in offering maximum screening between signals. Also in this initial layout, the main chip supply lines are used to separate the low and high frequency parts of the chip, and the internal power distribution is designed to isolate the amplifiers from the internal chip noise. This choice of target locations for the lead-outs also allows the chip to be offset towards the bottom right of the cavity, allowing the shortest bonding wires to the data output and the clock input lines. The lower frequency input signals, and the less important synchronisation and diagnostic outputs can tolerate somewhat longer bond wires if necessary.

Power distribution was the next consideration. The most important criterion here was that both V_{dd} and V_{ss} power rails should have low impedance paths to the package pins. This requires multiple bonding between the chip and the package to give low inductance and the power pads on the chip were therefore much enlarged. Internally, power distribution was still of concern.

One solution was to provide large rails at each side of the chip, with a horizontal bus distribution from these rails, with alternating V_{ss} and V_{dd} lines. Figure 5.22 shows this power grid. The only decoupling between the two rails is provided by the parasitic capacitances of the logic elements, and this acts as a very small distributed capacitor. Since these nodes also provide the source of the power-rail noise currents, their effectiveness in decoupling is slight, and the noise voltages are kept small by using large tracks of both low inductance and low resistance.

The alternative pattern of distribution uses a vertical first level metal grid for one power rail and a horizontal second level grid for the other rail. This gives much better decoupling of noise, because

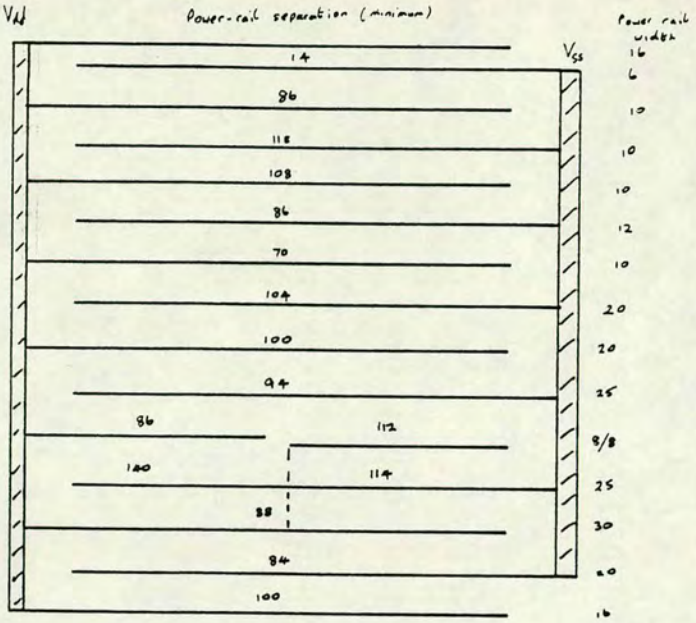


Figure 5.22 Schematic layout of power distribution showing preliminary allocation of space and track widths.

each crossing node acts as a small capacitor. However, it also gives the highest risk of catastrophic failure, because a pin-hole in the dielectric under any of the cross-over points renders the chip totally non-functional. Although this risk should still be slight, insufficient data were available on the quality of the polyimide dielectric for this complexity of circuit, and this layout was therefore cautiously avoided. The very high yield of working circuits obtained suggests that this caution was unwarranted, and one may recommend the improved decoupling of an orthogonal power mesh for future circuits.

Having obtained a crude block layout as shown in figure 5.21, together with the power grid of figure 5.22, the approximate logic gate positions were defined. Those parts of the circuit required to operate at maximum speed were allocated somewhat wider transistors, whilst the remainder of the circuit was chosen at a nominal size. The exception to this was in the delay line, where smaller than nominal devices were used deliberately to increase the propagation delay of these stages.

Having allocated a position and size to each logic element, the maximum current flowing in each branch of the power rail was calculated assuming that each gate was fully switched on. For CCL this is a sensible assumption, as each gate does draw maximum current at power-on. This calculation was also performed for the worst case transistor threshold voltage when the saturated current of each transistor would be maximum. The width of each arm of the power grid was then derived from this figure, assuming a current density of 1 μm/mA subject to a minimum track size of 6 um. This figure is tolerably below the electromigration limit for gold [449], even assuming the worst case metal thickness of 0.25 μm. Since the circuit in operation will consume less than 50% of this static power, this is a very safe margin. Because of this large operating margin, the power rail widths were not subsequently modified as the sizes of individual logic gates were refined.

Sketch layouts of the individual function blocks were used to estimate the number of routeing channels required between each pair of power rails to obtain values for the minimum separation between branches of the power grid. The minimum total chip height was therefore calculated at roughly 1.8 to 1.9 mm. Rather than reduce the chip side from the target of 2 mm, those areas of the chip where cross-talk or cross-coupling capacitance were likely to cause problems were identified, and the spacing between critical elements increased, or further screening introduced. The dimensions shown in figure 5.22 are obtained as a result of this selective relaxation of design rules. Notice that the power rail bisecting both the pre-scalar and the output D-type is split into two separate sections. By splitting so, the internal loop wiring of the latches can be completed with minimum cross-over capacitance to the power rail. This design was adopted to minimise the capacitance on the internal feedback loop in the latch, where it is most critical, but at the expense of the capacitance elsewhere in the circuit, where it is less critical.

Having converted the crude block layout into a complete logic diagram and thence into a transistor placement diagram, a certain amount of refinement was performed prior to simulation. In particular, care

was taken to balance the loading on both of the complementary clock lines for each of the latches. Whilst this was a straightforward exercise for the prescalar and Data latch, some additional balancing was necessary for the clock drivers to the Johnson counter, where layout constraints dictated that one of the clock inputs be split into two tracks, incurring significantly higher crossover capacitance than the other. The appropriate tracks, together with the extra ballasting, are highlighted in figure 5.23.

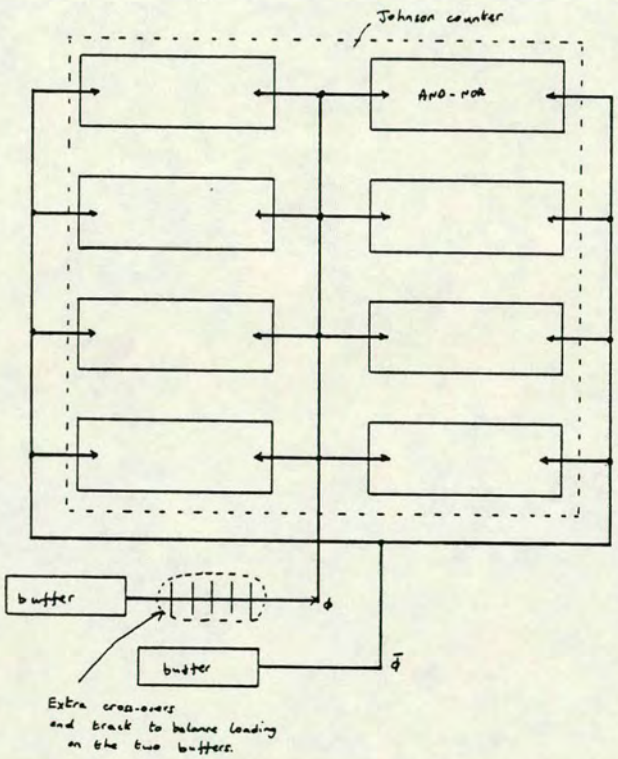


Figure 5.23 Balancing of loading on the clock distribution to the Johnson counter.

Similarly, the loading was carefully balanced for the eight outputs of the Johnson counter. Dividing the encoder into the three sections shown in figure 5.21: two 2-bit and one 4-bit section, it is clear that the line and cross-over capacitance is largely controlled by the separation between the relevant Johnson counter output and the 4-bit encoder section. By selecting the position of each logic gate within the 4-bit encoder, the different loading placed on the eight channels can be somewhat modified. A number of layouts were tried, with that offering the most uniform matching being selected. Additional spurs

of track and cross-overs were added to the lightly loaded tracks to attempt to give matched delays across all eight lines. Table 5.4 shows the results of these adjustments, assuming the load capacitance models introduced in Chapter 4. Because of possible errors in these models, both the crossover count and the line length were individually matched as far as possible.

Table 5.4 Track loading on Johnson counter outputs.

Node	------- Unmatched ------			--- Extra Ballast ----		
	X-over	Line-length	C_{tot}	X-over	Length	C_{final}
A	7	600	95	+2	+150	120
B	7	650	100	+2	+100	120
C	12	920	140	-1	-	135
D	13	820	140	-	-	140
E	4	320	50	+6	+450	120
F	8	620	100	+2	-	120
G	11	730	115	+1	-	120
H	9	750	115	-	-	115

Note 1: A rearrangement of the power distribution was used to reduce the capacitance on node C, and at the same time this contributed some of the ballast to nodes F and G.

Note 2: Part of the ballast on node E derives from the divide-by-8 output buffer.

Note 3: X-overs = simple cross-over count, Line-length is in μm, capacitance is in fF

Although there is still some discrepancy, the final capacitance values are sufficiently close to each other, given the accuracy of the available models.

The final consideration in the layout of the encoder was the need to minimise the frequency at any individual node and thereby reduce the possibility of errors arising from unwanted overlap of adjacent pulses. This was achieved by selecting the order of the inputs at the first stage of the encoder, such that channel 0 was paired with 4, 1 with 5 etc.. At the second stage, pair 0 and 4 was merged with pair 2 and 6, and similarly for the other four channels. Only at the very final stage of the encoding do any consecutive data bits become merged together. Thus, only the final NAND gate carries the full frequency of data, although the other nodes all carry signals with high frequency components arising because of the fast transitions.

To proceed further in the design process, it is necessary to turn to circuit simulation, as discussed in the following Section.

5.3 Circuit simulation

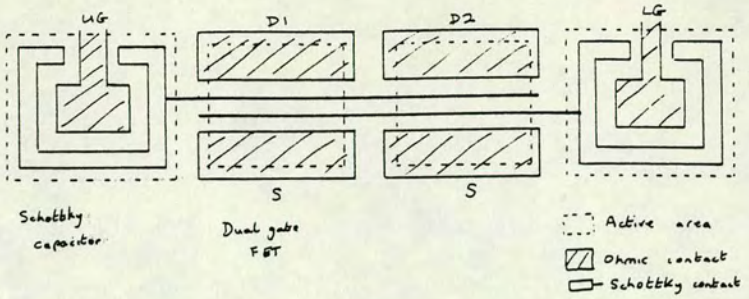
Following the basic stage of layout which included the small number of obvious refinements indicated, further improvements in design could not be made without recourse to simulation. This was also required in order to obtain confirmation of the choice of transistor widths used in key areas of the circuit, and of the exact size required for the delay line. During the circuit simulation phase, a number of key changes were introduced to eliminate a number of potential design hazards.

Detailed simulations were performed for the pre-scalar circuit, the complementary clock generator and associated driver circuitry. Less detailed studies were performed for the remainder of the circuit, although most of the function blocks were studied at some point.

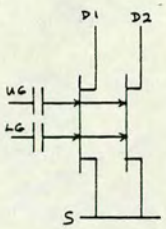
5.3.1 Ground rules for the simulations.

In all the simulations, appropriate models (as defined in Chapter 4) were used to derive the parasitic capacitance and track resistances associated with the actual layout employed. In order to obtain a good representation of a function block, it was necessary to include the preceding drive circuit, and the succeeding load. For example, the simulation of the prescalar included both clock drivers, the T-latch and the load devices.

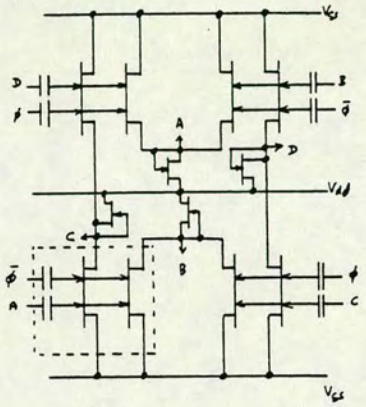
After including a large number of parasitic elements, many of the functional blocks represents a significant problem to the simulator, both in terms of element and node counts. In order to simplify the computation problem, a number of modifications were made to the "base-level" models. Instead of using transistors and diodes as the fundamental elements, much more complex basic element were used. These were usually based on the physical building blocks used in layout of the circuit. Thus, the basic element in the pre-scalar comprised a pair of dual FETs and the two associated coupling diodes. Figure 5.24 shows a sketch of the pre-scalar layout, highlighting this physical building block. Figure 5.25 compares this approach



a)

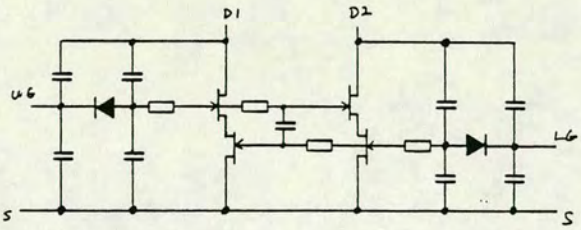


b)

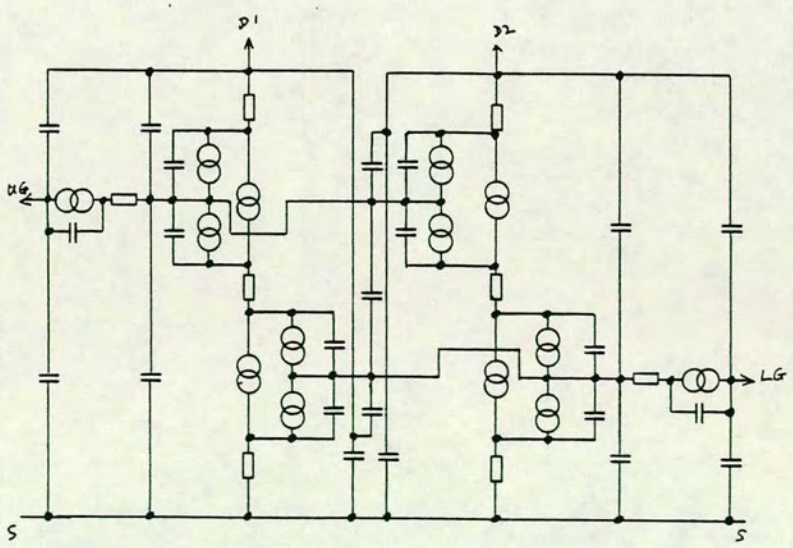


c)

Figure 5.24 Prescalar design
a) physical layout of building block
b) electrical model of building block
c) complete circuit



a)



b)

Figure 5.25 Models for the elements of figure 5.24b
a) detailed model (each transistor/diode includes parasitic elements)
b) "Simplified" equivalent circuit, after expansion of transistor/diode models.

with the more traditional method of deriving circuit models. In the conventional approach the emphasis is placed on the convenience of circuit definition which by implication eliminates accidental errors quite readily. According to this technique, the circuit comprises only 17 elements. However, each FET consists of 19 components, and each diode has 4 elements, resulting in 59 constituents in the fully exploded circuit. By combining elements, the circuit of figure 5.25b represents exactly the same function, but with only 47 components. At the same time, the number of nodes is reduced from 21 to 17. Whilst this technique requires more manual data reduction at the outset, the extra effort is recouped in faster, more reliable simulations.

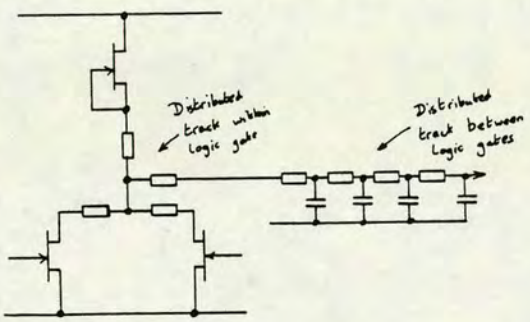


Figure 5.26 Parasitic elements within and between logic gates.

Using this much larger basic model, but with fewer repeat units, the circuit is completed by adding the track capacitance and resistance associated with the wires linking these physical blocks. In adding these extra components, one is again mindful of the need to reduce overall node and element counts. In practice, the track resistance may be split as shown in figure 5.26, and the capacitance should be distributed down the line rather than being lumped. The worst case result is obtained by lumping these elements, with the capacitor placed after the resistor (i.e. at the end of the line) and loaded to ground.

Having built up suitable circuit configurations representing the worst case of the process parameters (thinnest metal, thinnest inter-

layer dielectric etc.) it remains to select the remaining important device parameters. Most of the simulations were performed using nominal values of threshold voltage and saturation current density and a 5 V power supply, although specific results were obtained for extremes of these parameters.

The Sections which follow describe the simulation of various elements of the complete circuit. In order to maintain some structure to this Chapter, the simulations will be discussed in order of the flow of clock and data through the chip. In following this approach it will be necessary to assume the results of some simulations not yet described, because of the iterative nature of the design cycle and the complete interdependence of the functional blocks.

5.3.2 Clock generator and driver stages

The initial simulations of the clock driver circuit were performed on the complementary clock generator which was subsequently abandoned (see Chapter 3). As well as highlighting the shortcomings of the complementary clock generator, these simulations also highlighted the poor and imbalanced drive available from the NAND gates in the control loop. Figure 5.27 depicts the simulated circuit, and table 5.5 lists the rise and fall times at the nodes labelled. The table clearly shows the non-optimised choice of device dimensions because equal rise and fall times of around 75 ps are required at each node if the target speed of 2.4 GHz is to be met. The poor rise times indicate the inadequate pull-up current available from the small load transistor used in these simulations, and which was also used in the earlier experimental work.

In order to improve the drive characteristics of the clock driver circuit, it is only necessary to study the portion of the circuit

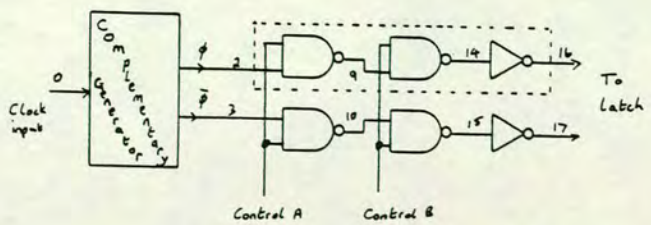


Figure 5.27 Clock generator and driver used in simulation.

Table 5.5 Switching transients for circuit of fig 5.27

Node	Rise time	Fall time
0	120	120
2	110	40
3	70	65
9	175	85
10	165	70
14	250	110
15	210	95

Notes: Simulated at 2 GHz, for V_t = -1.5; V_{dd} = 5 V
transition times represent 20-80%

Table 5.6 Switching transients for fig 5.27 with V_t = -1.5 V

Switch:load Ratio	-- Rise time -- Node 9	Node 14	-- Fall time -- Node 9	Node 14
2.5	120	120	70	85
2.0	100	90	75	85
1.6	80	75	85	90
1.3	70	75	100	105

highlighted in figure 5.27. Applying an input with rise and fall times equal to the internal target transition times at these nodes ($t_{20-80\%}$ = 75 ps), and varying the ratio of the widths of the switch and load transistors of the NAND gates, the results of table 5.6 were obtained, suggesting that the appropriate ratio should be near 1.6:1. An actual ratio of 1.67:1 was chosen to allow the convenience of using transistors in integer widths.

Because the essential reason for using CCL is to cope with the poor experimental control of threshold voltage, it is necessary to study this circuit behaviour over a wide range of values. Table 5.7 gives the results of similar studies for both -1 V and -2 V thresholds (although the power supply voltage was reduced to -4 V for the case of V_t = -1 V).

Clearly the aspect ratio of 1.67:1 is an appropriate ratio for the complete range of interest of threshold voltage, because the rise and

Table 5.7 Switching transients for fig 5.27 vs V_t

	Switch:load Ratio	-- Rise time -- Node 9	Node 14	-- Fall time -- Node 9	Node 14
V_t = -1:	2.5	200	230	95	120
	2.0	160	165	105	120
	1.6	130	130	115	120
	1.3	105	110	130	135
V_t = -2:	2.5	85	85	60	70
	2.0	70	70	70	75
	1.6	55	65	75	75
	1.3	55	65	95	95

fall times are approximately equal throughout the range. However, it is clear that the performance deteriorates quite substantially when the threshold voltage is allowed to approach -1 V. Although circuit operation can be assured, the performance specification will not be matched. One must therefore assume that there is sufficient short-range control of V_t to ensure that some circuits will contain only devices with V_t between -1.5 and -2 V, and will therefore meet the specification. The use of CCL will ensure that other circuits are functional, but only with a relaxed specification, whereas other circuit technologies would almost certainly not operate over this entire range.

Following on this thought however, even the data on the devices with nominal V_t suggest that the frequency performance will be narrowly missed if these sizes are employed. Simple calculations of load capacitance (taking account of the different effects within the quasi-complementary stage) suggest that the two NAND gates in the highlighted portion of figure 5.27 are near-equally loaded with a capacitance of around 60-70 fF (assuming 25 um width switches in the NAND, a 13 um inverter and 35 um driver transistor). This finding is borne out by the figures in table 5.6. To improve on this performance, the gate widths need to be improved, and because of this similarity, it is only necessary to simulate the effect of changing the size of one gate. The optimum size of the other can be calculated from the results. Table 5.8 shows the effects of altering

Table 5.8 Switching transients vs gate width

Switch Width	-- Rise time -- Node 9	Node 14	-- Fall time -- Node 9	Node 14
20	100	80	100	90
30	85	80	75	85
40	55	75	65	80
50	50	75	55	80

the switch transistor width of the leading gate, whilst maintaining the switch:load ratio at 1.67.

In addition to helping identify the appropriate logic gate sizes, this table indicates that even the 20-80% transition times of a stage are affected markedly by the input transition times. Even though the logic may have a fast slew rate region in mid transition, the capacitor coupling biases the effective switching region out of this high speed portion. This suggests that the circuits should be operated at the optimum supply voltage, rather than allow a large amount of "undershoot" in the voltage between diode capacitor and transistor.

The initial layout of the chip described in Chapter 5.2 indicated that the power budget of the chip may under some circumstances come close to the nominal 1 W, arbitrarily set as an upper limit at the outset (the poor thermal dissipation of GaAs may cause problems during probe test even though it has little impact once packaged). Whilst table 5.8 might suggest an ideal size of around 35 um for the switch transistor width, the more marginal size of 30 um was chosen. In order to improve the switching of the whole driver chain, gate A6 was increased from 25 to 30 um, and gate A5 was therefore increased to 35 um to compensate for this. Simulations were repeated for this improved driver chain, and table 5.9 shows that this new network offers the required performance.

The correct dimensions for the constituent gates have now been established and the circuit of the complementary clock generator revised. After allowing for the additional loading factor associated

Table 5.9 Switching transients for fig 5.27 (new sizes)

Threshold voltage	-- Rise time -- Node 9	Node 14	-- Fall time -- Node 9	Node 14
-1.0	110	115	95	105
-1.5	65	70	70	75
-2.0	50	50	65	65

with the tap-off point for the clock generator/driver for the output latch, the final circuit shown in figure 5.28 was simulated. The results of these simulations, performed at a clock frequency of 2 GHz, are also shown. Some degradation in waveform quality is seen as the signal propagates through the cascaded gates. Nevertheless, the output waveforms from the final buffer stage show only a small skew with very little deviation from a 1:1 mark:space ratio. The waveforms at gates 6 and 10 are the ones of interest because these are the final waveforms correctly simulated (those from the buffers themselves represent unloaded transitions). It is these waveforms

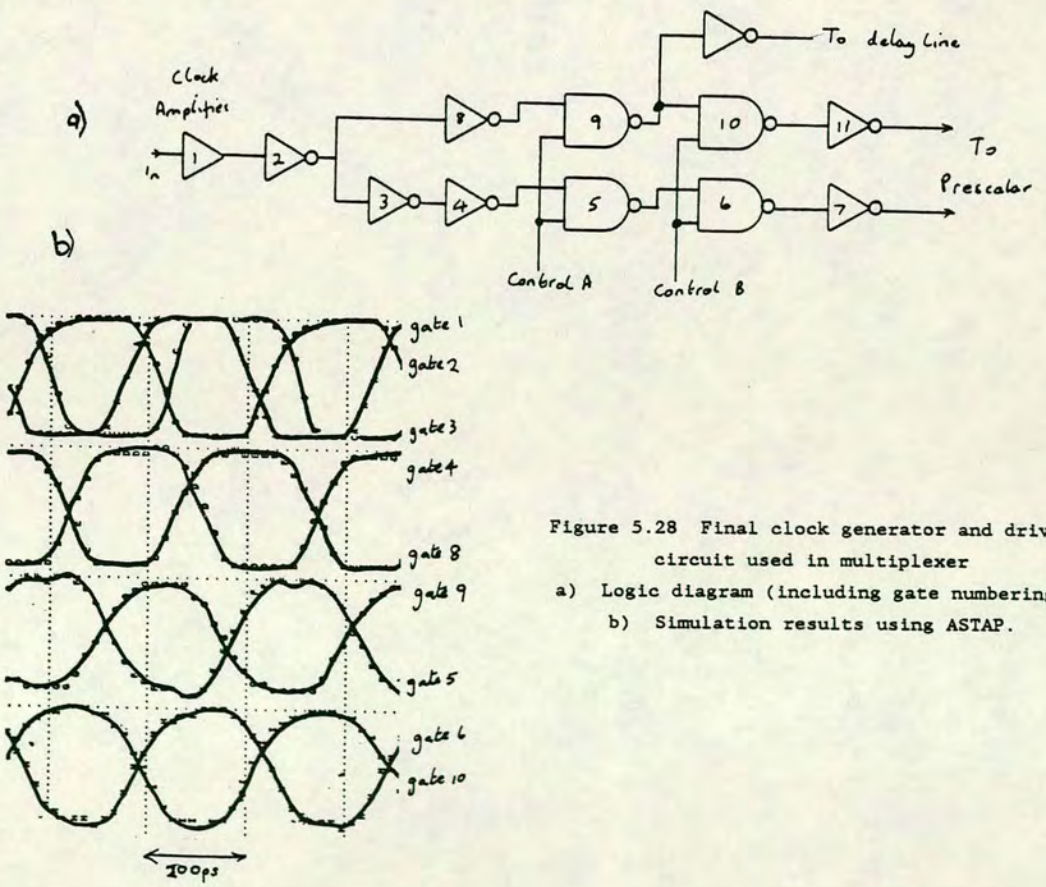


Figure 5.28 Final clock generator and driver circuit used in multiplexer
a) Logic diagram (including gate numbering)
b) Simulation results using ASTAP.

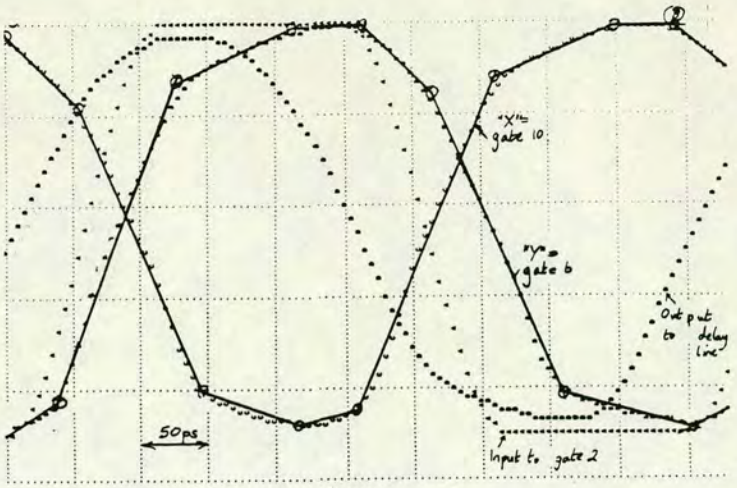


Figure 5.29 Simulated output from gates 6 and 10, showing piecewise approximation to the waveforms.

which have to be synthesised for use as input waveforms into the buffers included in the simulation of the pre-scalar circuit. Figure 5.29 shows these waveforms in detail, together with the piecewise approximation used for this synthesis.

5.3.3 Pre-scalar simulations

For the pre-scalar, all simulations were performed using 40 μm wide switch transistors with these over-large devices offering a greater speed potential than those of nominal size (25 μm). As well as the rise and fall times already studied at length, there is considerable interest in the propagation delay for most practical circuits, but unfortunately, in many circuits it is much more difficult to obtain a repeatable, unambiguous prediction for the propagation delay. However, in the toggle circuit the prediction of the overall propagation delay is very much simplified by using the self-oscillation mode. It is therefore interesting to repeat some of the studies described in the previous Section, but concentrating on the average propagation delay of the gates. One such simulation compares the propagation delay for different values of the ratio between the widths of the switch and load transistors. Table 5.10 indicates that the circuit continues to speed-up in a monotonic fashion with increased load, even though the requirement to minimise both the rise and the fall-time would indicate that an optimum ratio exists. In practice, at some (undetermined) minimum width, a catastrophic failure will occur, despite this trend of continuous performance improvement. Even though this table indicates that further improvement would be possible, the ratio of 1.67:1 (selected to

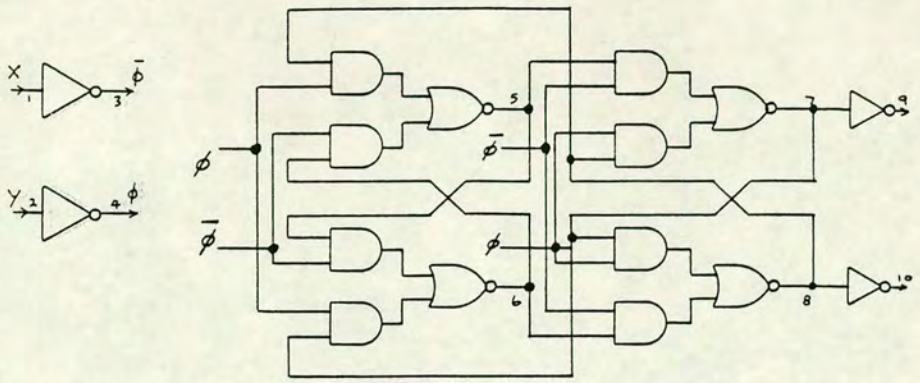


Figure 5.30 Circuit used for prescalar simulations.

Table 5.10 Predicted propagation delay of AND-NOR gate vs load width

Aspect ratio	Prop delay ps	Max clock speed GHz
2.5	195	2.6
2.0	166	3.0
1.6	150	3.4
1.3	132	3.8

Note: These are simulated results derived from the self-oscillation condition of the pre-scalar with 40 μm width switches.

maintain equal rise and fall times) has been retained as a safe, if not fully optimised solution (especially remembering the need for satisfactory operation over a wide spread in threshold voltage).

A simulation of the effect of clock skew was also performed on the divider circuit whose nodes are labelled in figure 5.30. Figure 5.31 shows the waveforms at the four principle nodes within the pre-scalar (two outputs from each of the master and the slave latches) for varying degrees of skew. Quite clearly, the effect of skew is only observed at either the master or the slave of the latch, depending upon which of the two clock lines is delayed. By ensuring that the erroneous pulse generated within the divider appears at the master rather than the slave, it will not be propagated into the outside circuitry, unless the internal pulse is sufficiently large to trigger a change in the logic state of the slave. The amplitude of this

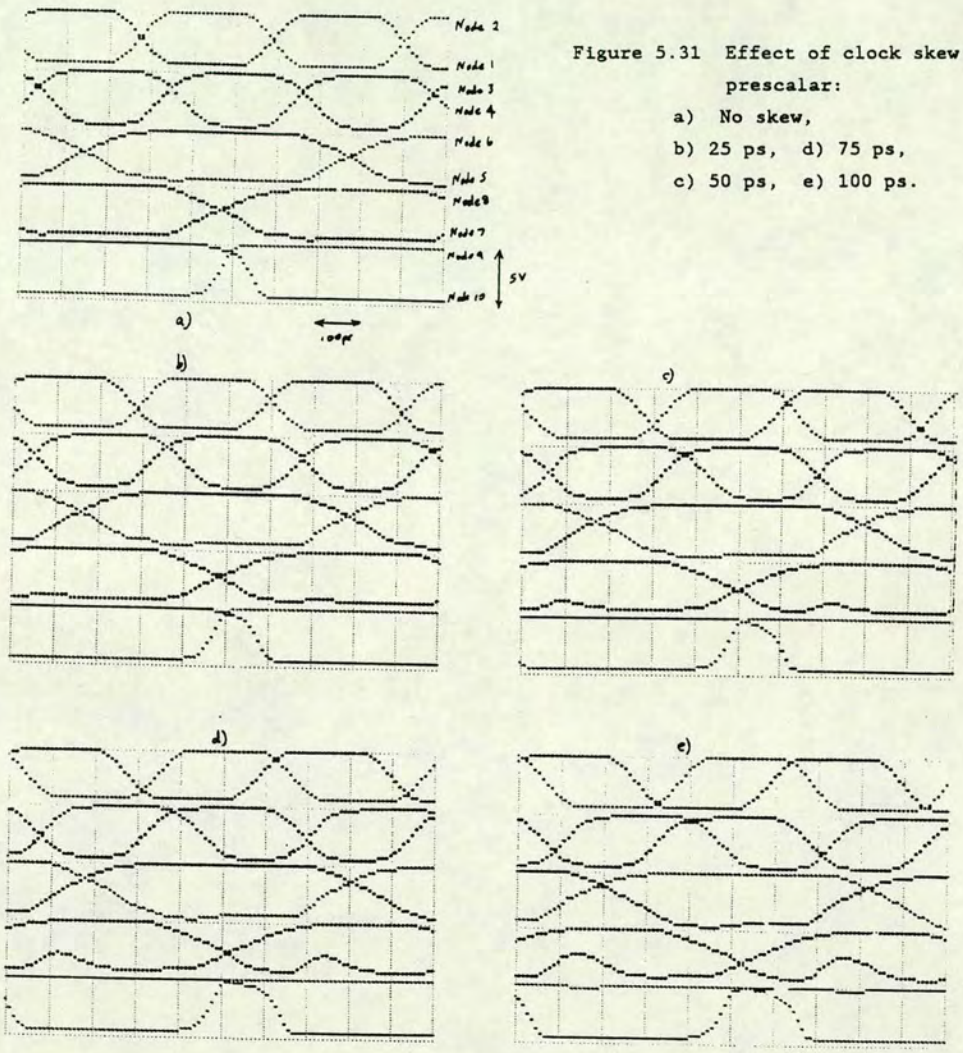


Figure 5.31 Effect of clock skew on prescalar:
a) No skew,
b) 25 ps, d) 75 ps,
c) 50 ps, e) 100 ps.

Table 5.11 Noise spike amplitude vs clock skew

Clock skew	Amplitude
0 ps	0 mV
25 ps	0 mV
50 ps	500 mV
75 ps	1300 mV
100 ps	2000 mV

pulse is listed in table 5.11 as a function of the phase skew on the clock pair. These data suggest that 50 ps of skew is the maximum tolerable. The data already presented for the clock driver circuit lie well within this limit.

This simulation work further suggests that the latch is more tolerant of clock skew than it is of a non-unity mark-space ratio.
Figure 5.31 also shows the output waveforms from inverters placed after the prescaler. The very fast transitions are once again indicative of the fact that these inverters are unloaded, driving an unrealistically low capacitance. Notwithstanding this simplification, it is clear that the latch outputs produce a significant difference in the widths of the mark and space. This is

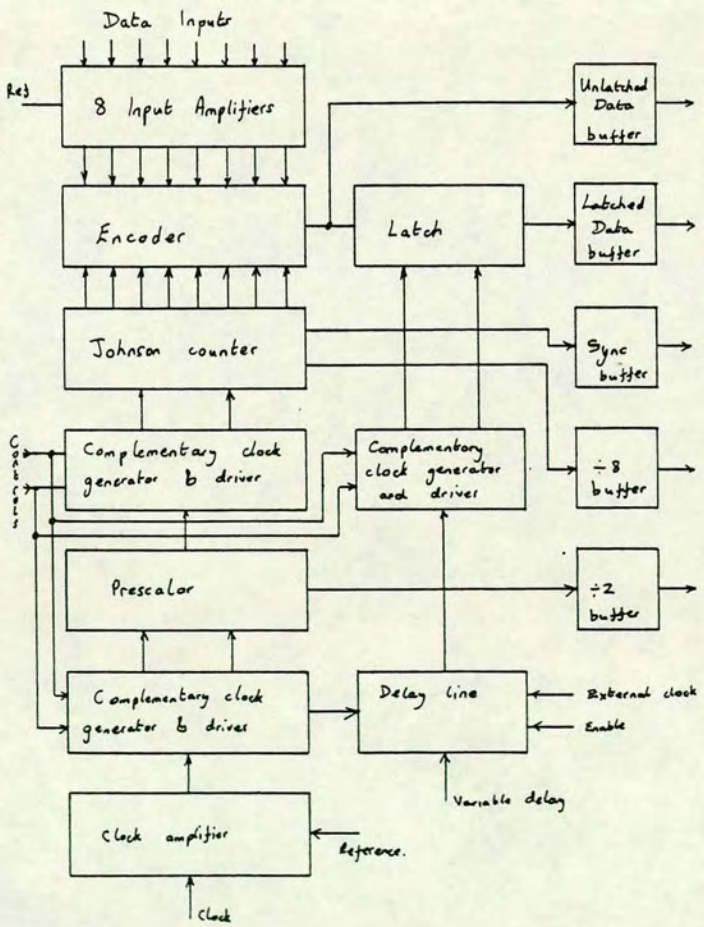


Figure 5.32 Final block diagram of multiplexer.

despite the theory (Chapter 3) which predicts an exact 1:1 ratio for this circuit element. The explanation for this lies entirely with the operation of CCL, the switching threshold for which does not lie at the centre of the voltage swing. As the toggle circuit has relatively slow edges compared with the output buffer, the close proximity of the switching threshold to the logic-1 level results in a longer time delay for the falling edge compared to the rising edge.

Thus, although the AND-NOR pre-scalar theoretically produces complementary outputs, in reality the signals are inadequate for use as complementary clock inputs to another stage. Following this

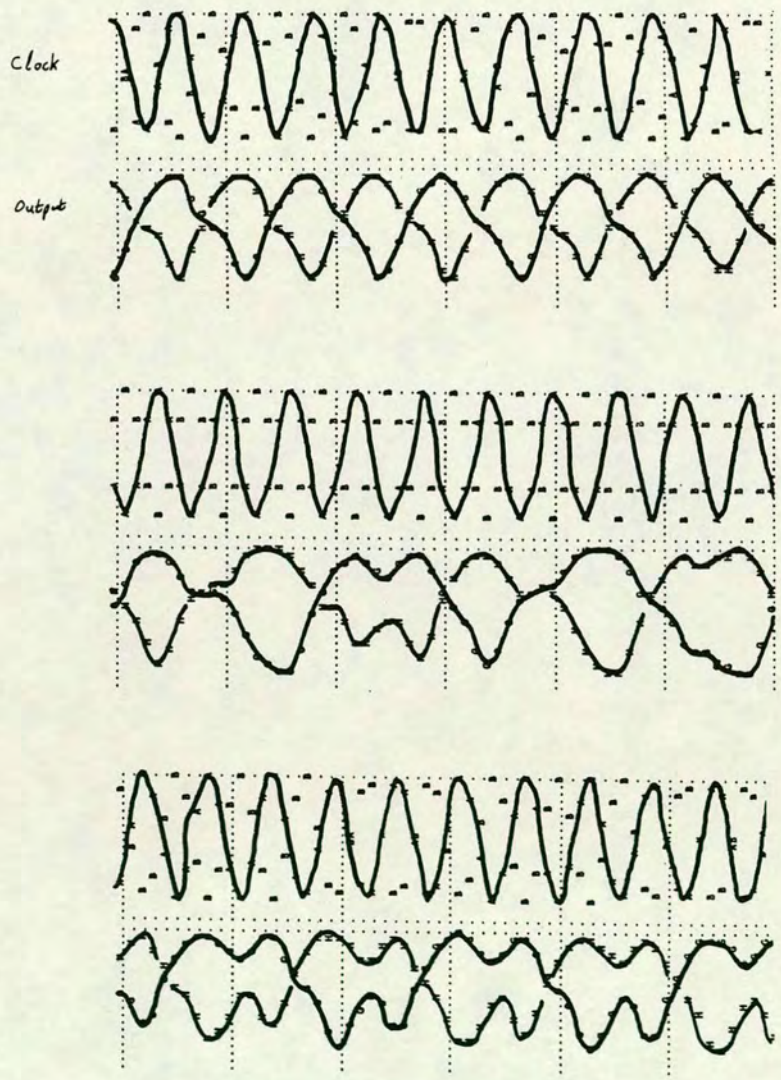


Figure 5.33 Simulated waveforms for T-latch driven at high speed:
a) f = 3.3 GHz: correct operation, b) f = 3.4 GHz: arbitrary count
c) f = 3.5 GHz: incorrect divide-by-4.

simulation therefore, the block diagram of the multiplexer was modified to take only one output from the pre-scalar. An additional complementary clock generator was added between this output and the input to the Johnson counter. Figure 5.32 shows this new block diagram.

Having reduced the loading on the pre-scalar, the circuit was re-simulated, and despite the removal of some of the symmetry, the predicted maximum circuit speed increased from 3.2 GHz to 3.3 GHz. This simulated maximum operating frequency of 3.3 GHz compares well with the 3.4 GHz maximum predicted from the self oscillation period (table 5.10). Figure 5.33 shows the simulated waveforms when the clock is driven beyond its maximum acceptable input frequency. The output degenerates into almost arbitrary wave-shapes.

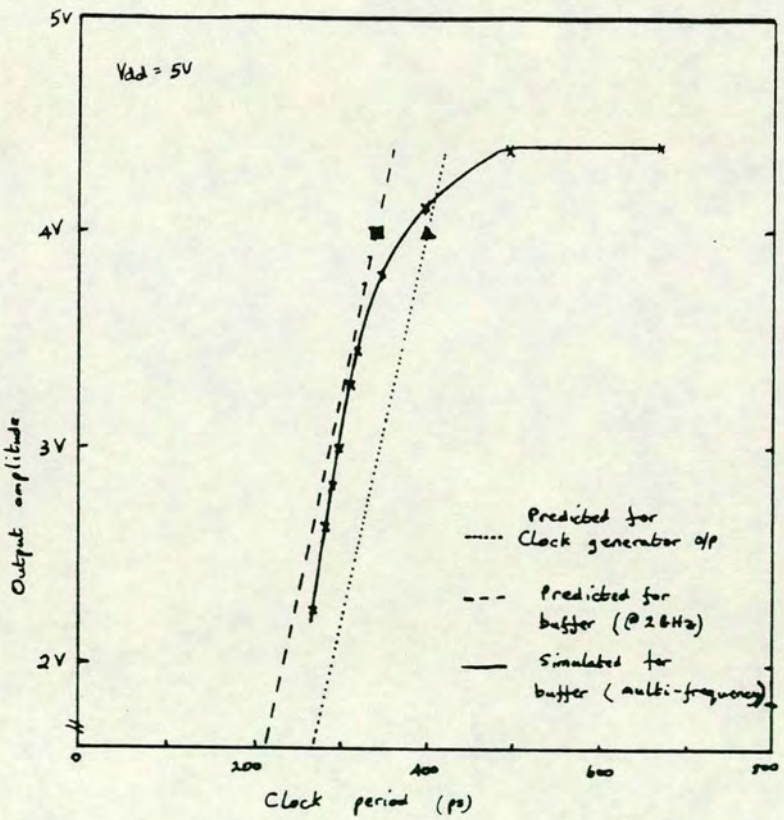


Figure 5.34 Roll-off in voltage levels for high frequency operation: comparison of prediction and simulation.

These simulations, used to determine the maximum clock frequency for correct division were also useful to study the effects of roll-off in the frequency capability of the buffer circuitry. Figure 5.34 shows the output amplitude available from the quasi-complementary buffer as the frequency is increased. Also added to this graph is a prediction

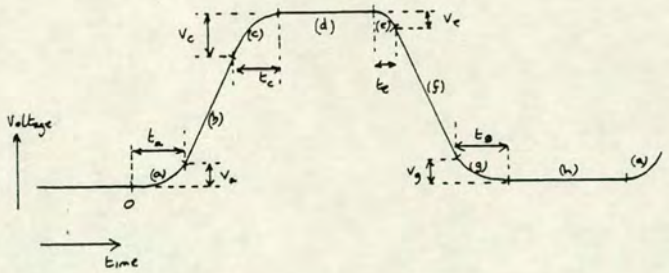


Figure 5.35 Partitioning of the switching waveform.

of this amplitude roll-off, derived from the single simulation performed at 2 GHz. This prediction was based on the assumptions indicated in figure 5.35, in which the output waveform is split into elements representing maximum slew rate for both edges (b and f), fixed elements representing the beginning and end of each transition (a,c,e,g), and static levels (d and h). If d and h are both set to zero, the clock period (t) and amplitude (A) are therefore related according to:

$$T = t_a+t_c+t_e+t_g + (A-V_a-V_c)/R_b + (A-V_e-V_g)/R_f$$

where t_x is the time, V_x the voltage and R_x is the slew rate during element(x).

By comparing this prediction with the roll-off derived from several simulations, it is apparent that the prediction is somewhat optimistic, but, provided that the amplitude remains above 3 V, it is not grossly in error. Since this voltage is probably close to that at which the circuit fails, the single 2 GHz simulation is a good indication of viability and, as each simulation of this complexity requires several minutes of CPU even on an IBM 3081, minimising the number of runs is essential.

Figure 5.34 also indicates that the final clock driver buffer is of exactly the right size, as the buffer is beginning to fail at the same time as the divider (i.e. the amplitude has fallen to 3 V at about 3.3 GHz). If the roll-off of the other gates in the clock driver chain is similarly derived from the 2 GHz simulations, the gate sizes appear to be marginal, despite having been increased. The

earlier study suggested that only the final NAND gate need be investigated, and the predicted roll-off for this gate is included as the final curve in figure 5.34. Allowing for the prediction being optimistic, it appears that this gate is only capable of operating to 2.8 GHz. Whilst this is still within the design specification, it has a smaller margin than the remainder of the circuitry, and the gate size should be modified in a future redesign. Needless to say, this study was concluded after design completion!

5.3.4 Input amplifiers

Section 3.3.3 described the basic factors influencing the design of the input amplifiers, and concluded that the common-gate amplifier would be the most appropriate choice for use in the early development of GaAs ICs, giving a good compromise between performance and simplicity. This circuit was required to fulfil two applications. Firstly, it should operate at the high speed required for the clock input. In this context, the circuit would be placed optimally next to its load, the complementary clock generator. Secondly, it should also serve as a useful amplifier for the data streams where the amplifier could only be sited remotely from its loading elements, leading to a potential large load capacitance.

Figure 5.36 shows a DC simulation indicating that the output amplitude is quite sufficient, and only marginally dependent upon the relative transstor widths. As the reference voltage is increased, the available amplitude decreases, but figure 5.37 shows that the amplifier continues to operate over a wide range of input conditions. Transient simulation at the nominal threshold voltage showed that the amplifier was capable of operating from a sine-wave input to well in excess of the 2.4 GHz required, if set at the optimum reference condition. Since the eight data amplifiers will share a single reference node, the circuit's tolerance to an incorrectly set reference voltage is of particular interest. Figure 5.38 shows the effect of load capacitance on the rise and fall times of the amplifier when driven with an 800 mV amplitude, 400 MHz sine wave. Three conditions were used; the optimum refernce voltage, and +/- 300 mV away from this value. The circuit performs satisfactorily

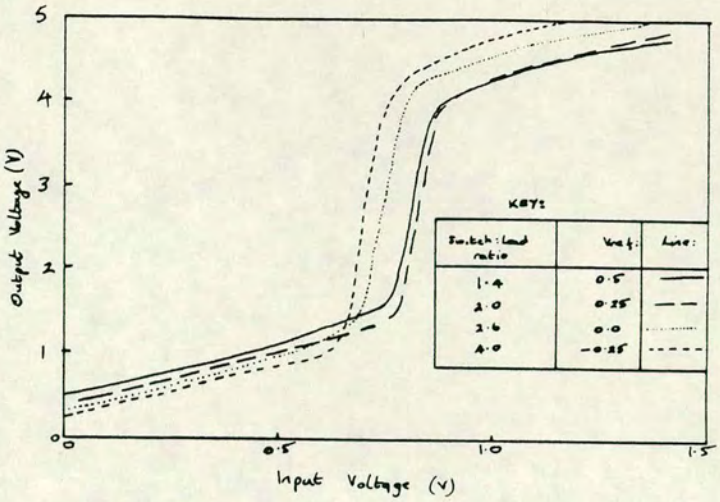


Figure 5.36 Effect of width ratio of FETs on input amplifier's performance.

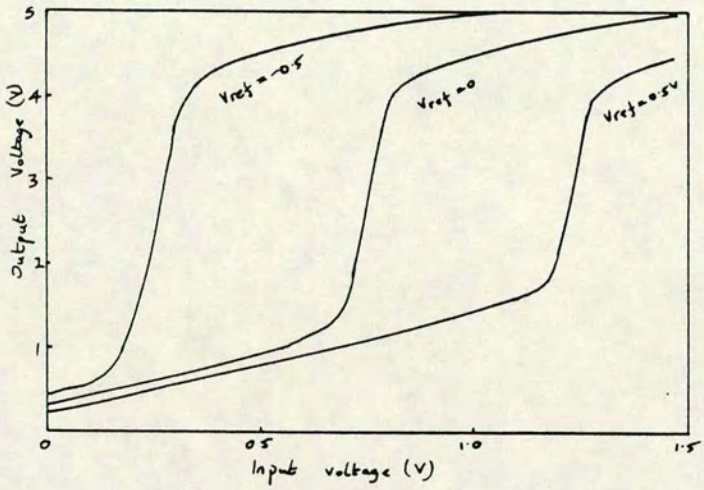


Figure 5.37 Transfer characteristic of input amplifier with varying reference voltage.

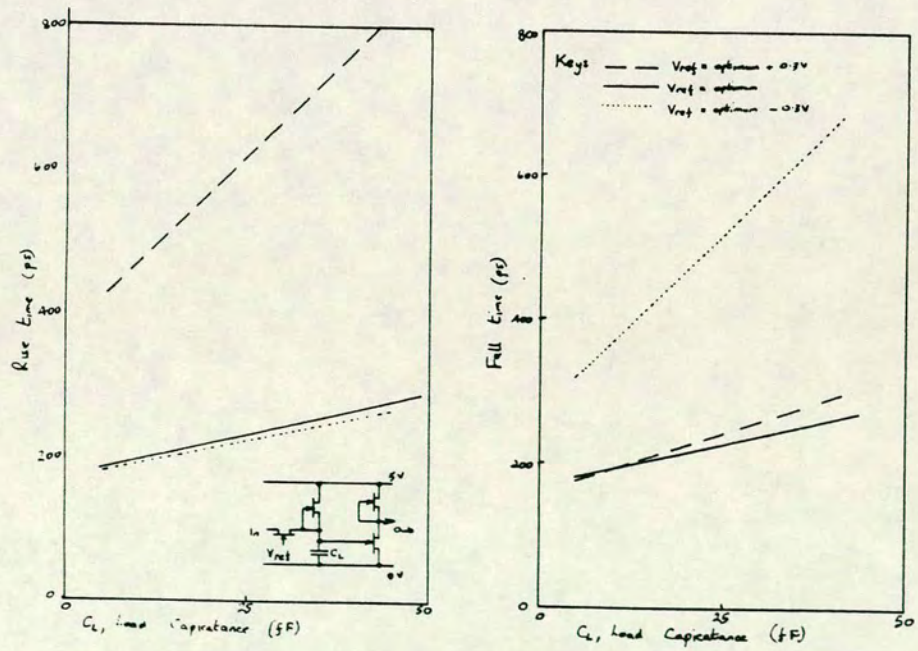


Figure 5.38 Variation in input amplifier transition times with load capacitance and reference voltage:
a) rise times, b) fall times

under all these operating conditions, implying that local buffering is not required, and that the reference voltage will not be critical.

Together, these simulations indicate that the choice of common-gate amplifier was wise, and that it performs well over a remarkably wide range of input conditions.

5.3.5 Remaining circuitry

Although significant effort was expended in simulating the remainder of the circuitry, no new techniques were used, over and above those already described. Similar techniques were used to minimise the node count in the Johnson counter circuit, which was simulated both as an oscillator and a divider. In its self-oscillation mode, the Johnson counter showed a propagation delay of around 210 ps, indicating a maximum clocking frequency of 2.3 GHz. Whilst this is well above its intended operation, the more important aspect is the transition time of each node, because these signals are used to generate the channel gating pulses. Because of this requirement, the buffer gates on the Johnson counter outputs were also increased in size following simulation.

Another item of concern which was raised as a result of these simulations was the width of the overlap region between adjacent data pulses. Because of the large loading capacitances within the encoder circuitry, long rise and fall times are experienced, and these contribute to the need to use a wider error bar in the derivation of figure 5.18 than that employed. As this circuit is to be used primarily in a feasibility study, the consequences of this were not explored further, although it should be noted that the new error bar is still narrower than the value of delay which was externally programmable.

The remaining requirement from these simulations was to study the overall propagation delay from clock and data inputs via different routes. The results of these simulations indicated that the delay line should comprise a total of some 16 gates' delay before the complementary clock generator of the D-type. One problem is that this delay line actually carries the highest speed of the chip, yet

it is required to be low power, and to minimise the gate count, each gate should present a large delay. The circuit used is a compromise, with most of the gates having switch transistors only 8 um wide, compared with the standard size of 13 um. In order to work satisfactorily, the sizes of the remaining gates, particularly gates 13, 17-19 and 24-28 have been tailored to effect near equal loading on all the gates within the chain. The final simulated timing parameters are a loop delay difference of 600 ps between the data path to the D-type, and the clock path to the D-latch. Of this, 250 ps is programmable externally. These data represent the nominal situation of V_{dd} = 5 V and V_t = -1.5 V. Because of the large amount of simulation involved (virtually the whole circuit must be simulated, albeit split into functional blocks), no data are available on the effects of changes in threshold or supply voltage.

5.4 Multiplexer operation

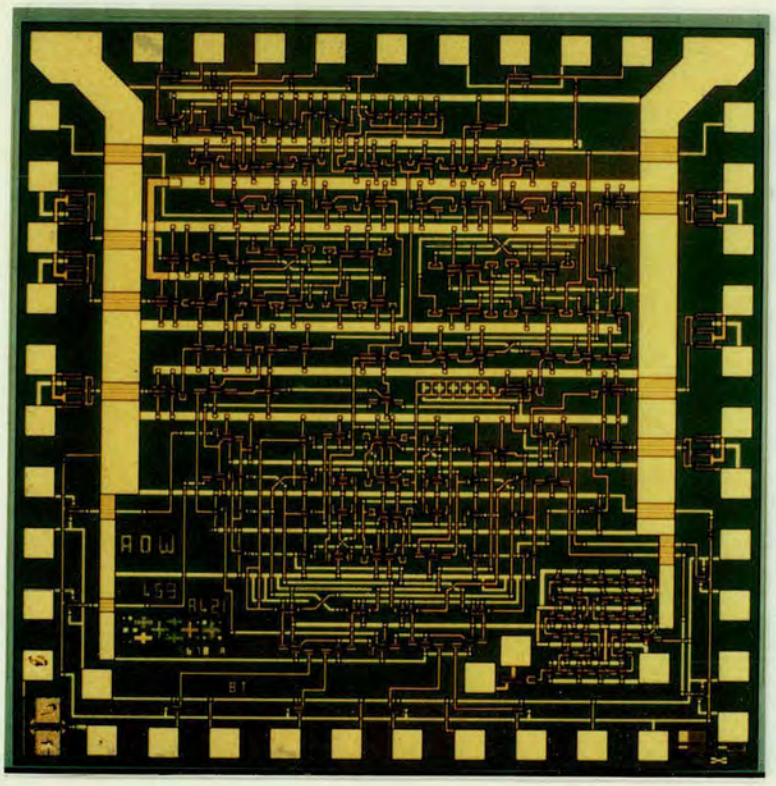
The circuit hitherto described was fabricated within the GaAs IC facility at BTRL. A chip photograph is included as figure 5.39. This section deals with the experimental results obtained from the circuit [450]. The tools available for testing are first described, followed by discussion of the actual results.

5.4.1 Test techniques

The initial testing of devices was performed using a commercial probe card consisting of standard bladeless probes mounted in an epoxy ring. This card was modified by removing all signal tracks, instead feeding the high frequency signals directly onto the probes via high quality coaxial cable. In order to maintain the desired flexibility of test setup, a metal flange carrying RF sockets leading to these coaxial leads was added to the probe card - the data input lines using SMC connectors [451], and the clock and output signals using the higher quality SMA terminations [452]. In addition, each of the stimulus lines is terminated in 50 Ohm at the probe, and each of the DC and control contacts contain RF bypass capacitors on the card.

This technique had proved successful in the testing of simpler high speed circuits such as toggle circuits, even at speeds up to 2 GHz, although detailed measurement of voltage levels were affected by

a)



b)

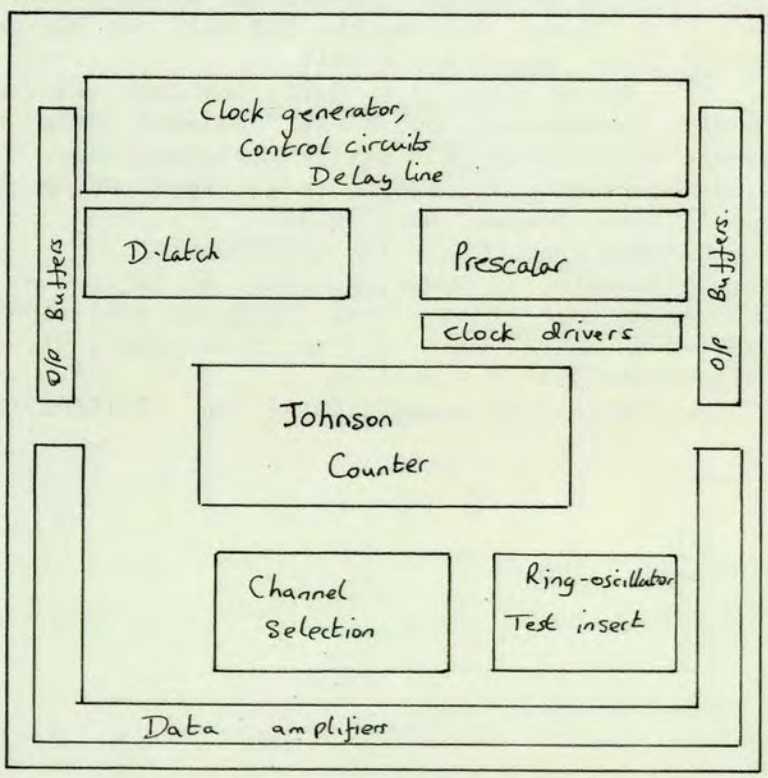


Figure 5.39 Finished chip layout: a) micrograph, b) floorplan.

degradation of the signal due to both crosstalk and mismatching. For the multiplexer, the large number of signals required exacerbated these problems, and made test results unreliable.

A number of alternative techniques was available, or under development, for the circuit testing. At the outset, the most reliable of these was to dice the wafer, and bond individual chips directly onto ceramic substrates with impedance-matched tracks. Signals were introduced onto the microstrip tracks on this substrate via SMA stripline launchers mounted in a precision jig. The substrates could be removed from the jig, to allow different samples to be measured. Figure 5.40 shows a photograph of this test jig and substrate.

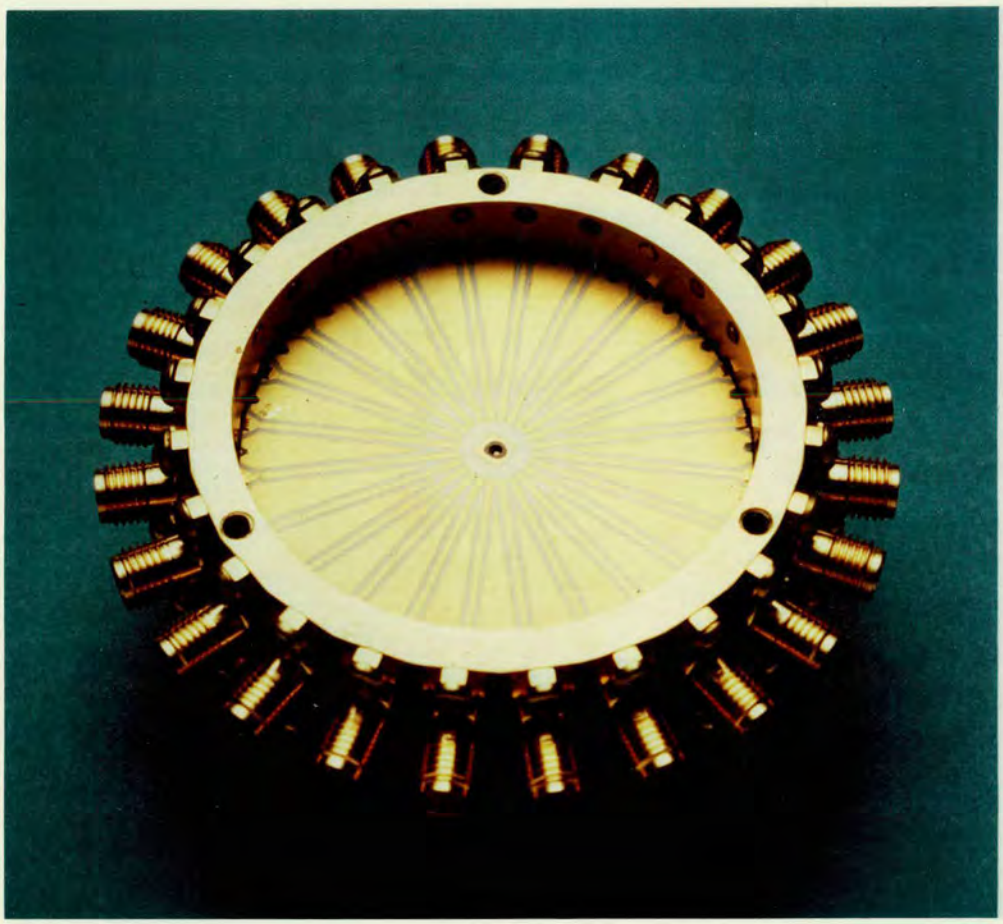


Figure 5.40 Photograph of high frequency test jig.

Although this technique was the most reliable for qualification of the circuit, it is a destructive test technique as the chip cannot subsequently be removed from the ceramic. As an intermediate technique, alternative substrate patterns were available onto which pre-packaged chips could be soldered. With care, it was possible to unsolder the chip for re-use, but again, this technique was only applicable to small sample quantities. Holders which allow rapid, and non-detrimental substitution of the packaged chips are available for both types of package employed. These holders are only designed for DC and low frequency test, and are totally unsuited to any testing above a few 10s of MHz. During the early period of circuit testing, colleagues developed a technique using ultra-thin, high dielectric constant substrates, for which the width of a 50 Ohm microstrip line exactly mated with the leads on the flatpack package. Using a rubber pressure pad to force a good contact between the package and the substrate, a high frequency non-destructive test technique was obtained.

Further improved results were obtained when a full functional wafer probe test at 10 MHz was implemented, using a programmable high-speed logic analyser (note the difference in the commercial and laboratory use of the terminology "high speed"). The functional test was coupled to wafer mapping of the speed capability of the circuits, as obtained from the self-oscillation frequency of the pre-scalar. Together, these were used to screen samples prior to packaging. Figure 5.41 shows a wafer map produced as a result of this functional test, indicating the remarkably high yield obtained from the first wafers of this complexity processed at BT [453]. This shows the resilience of the CCL logic technique.

The results presented here were obtained from all of these package techniques. In addition to the physical process of deciding on the test fixture, the source of the test signals is of considerable concern at such high speed.

At such high speeds, pulse waveforms are difficult to obtain, but a suitable clock signal can be obtained from a variety of sine wave sources, including sweep oscillators. A DC offset must be added to

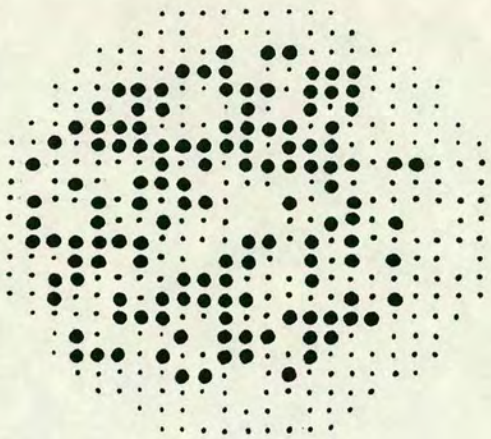


Figure 5.41 Wafer map of multiplexer yield at low-frequency functional test.

• pass
. fail

the sine wave to ensure compatibility with the input amplifiers, and this is added using a bias-T. A matched 50 Ohm environment must be adopted throughout, and the measurement or monitoring of this signal (including the added DC component) is then not trivial. In the final set-up, the clock input was passed through a broadband 20 dB coupler, with the through path providing the main signal for the circuit stimulus, and with the coupled signal used for monitoring. The correctly scaled DC offset was added separately to the two signals after the coupler.

For some measurements, the data channels can be driven with square waves bearing an exact sub-harmonic relationship to the clock signal. For these tests, the on-chip divide-by-8 signal was used to derive a lower clock speed which could lie within the range of commercial components [454]. By further dividing this frequency, a small range of different data patterns were obtained.

In addition to these simple patterns, more complex patterns were available from a purpose built pattern generator. This consisted of a number of parallel channels, each of which contained a sixteen stage shift register connected in a continuous ring. Microprocessor control was used to load parallel-words into each channel, and then to circulate this data using a master clock driven externally. Initially this was built to operate at 150 MHz, but was later improved to operate at 250 MHz. Using this generator, the circuit

could be tested at up to 2 GHz with simulated data streams, allowing any pattern sensitivity of the circuit to be detected. Such pattern dependence is unlikely to be detected either using low speed data, or with simple high speed tests using square wave patterns.

5.4.2 Results

An initial investigation of the circuit showed that it did indeed behave correctly as an 8:1 multiplexer. Figure 5.42 shows an oscillogram of a randomly selected sequence. The difficulty is to assess how well the circuit behaves, and whether it behaves correctly over the whole specified range. In practice, as this is a demonstration of process capability, the interest lies in determining the bounds of operation, almost irrespective of the adherence to a specification.

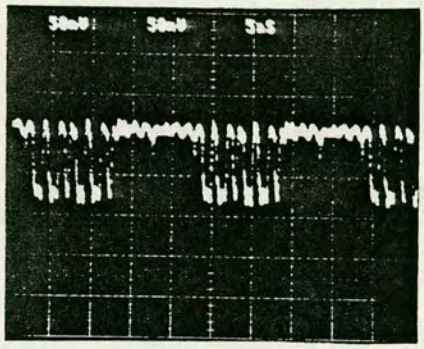


Figure 5.42 Multiplexer operating at 2.8 GHz with arbitrary data pattern.

Initially, the unlatched data outputs were studied, as these involved fewer of the internal circuit functions. Each of the eight inputs was fed in turn with a data signal at 1/16th of the clock frequency, whilst the remaining inputs were tied to ground to avoid spurious pickup. All inputs were shown to work, with the pulses appearing in the correct sequence. Figure 5.43 shows the measured waveforms under

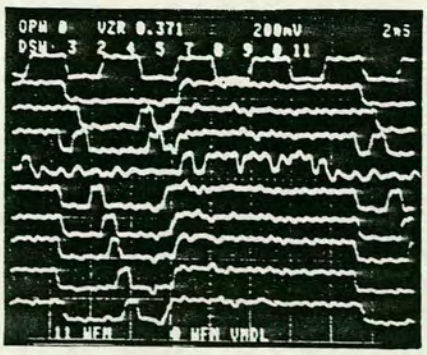


Figure 5.43 Multiplexer operating at 2.2 GHz with each channel identified in turn (channel 2 non-functional).

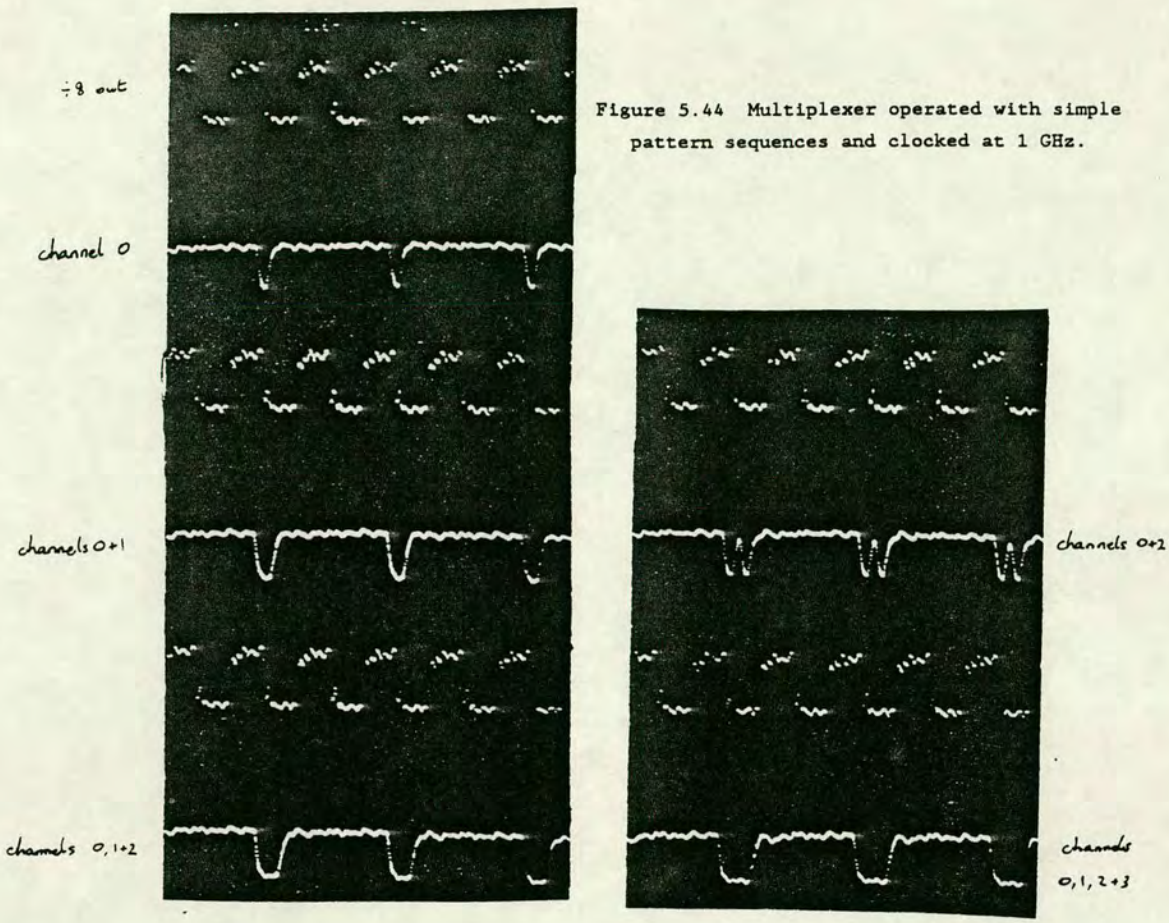


Figure 5.44 Multiplexer operated with simple pattern sequences and clocked at 1 GHz.

this test condition, as the data (second trace from the top) is applied to each input in turn (traces 3-10 from the top correspond to channels 0 to 7 respectively). For this particular sample, the first circuit measured, channel 2 is faulty, but other samples were later found which were fully functional. It should be noted that these results are measured at a clock frequency of more than 2 GHz, with the top trace showing the divide-by-eight output. This measurement was performed on a device mounted directly onto a ceramic substrate.

Other input sequences were studied to try to identify either pattern sensitivity, or timing errors between the channels. Figure 5.44 portrays the behaviour with a number of different patterns, showing that there is no gap between pairs of consecutive pulses either in the logic-0 or logic-1 state. It also shows that an alternating (i.e. 1010) sequence achieves near full height on all bits. Unlike

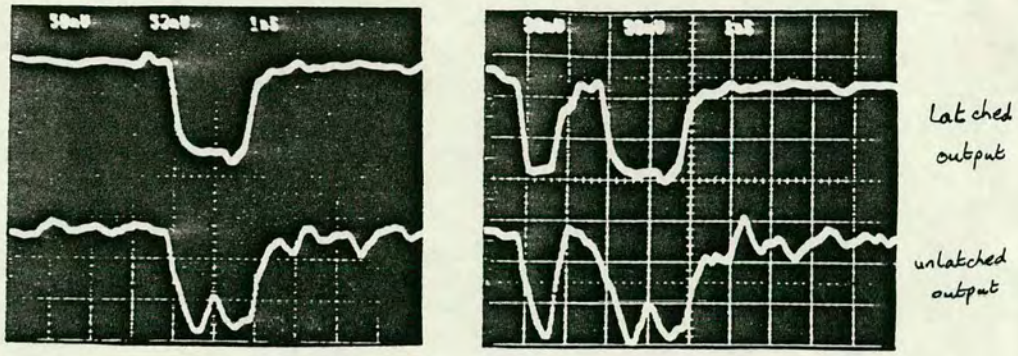


Figure 5.45 Operation of the output latch to clean-up the data pulses.

the previous figure, these oscillographs were obtained for a clock frequency of 1 GHz, but by direct probing of the wafer. The poor transition times on the pulses arise partly because of the wafer probing, but a more detailed study, including the latched outputs as well indicates that the output transitions of the latched data are much sharper, as well as cleaner (figure 5.45). This may have been coincidence, as operation at 2.8 GHz (only unlatched) has been shown for a packaged circuit tested under conditions of properly matched impedances (see figure 5.42). Further evidence of signal quality is found in the study of the eye diagram, which is shown in figure 5.46 (although this particular result has been obtained from a single D-latch, rather than from the multiplexer).

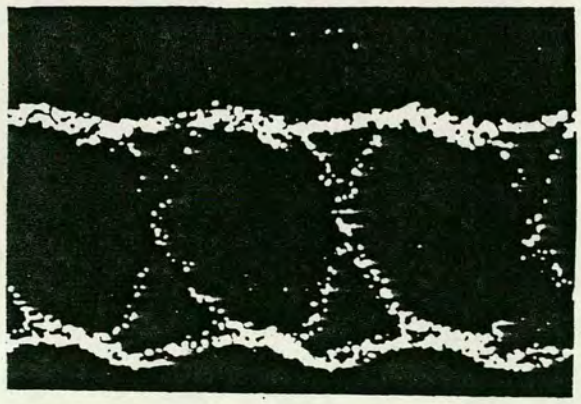


Figure 5.46 Eye diagram of a D-type latch operating at 2 GHz.

Having indicated that the circuit performs its function correctly, it is important to note that there are some restrictions on this operation. Although the circuit was designed to operate at a supply voltage of 5 or 5.2 V and a threshold voltage of -1.5 V, the

operation is very much better if the supply voltage is reduced to 3 or 4 V whilst retaining the same threshold voltage. Indeed, the operation at 2.8 GHz was obtained for a 3 V supply. Clearly the CCL design does not allow as much tolerance to excess voltage swings as would be desirable.

The self-oscillation mode of both the Johnson counter and the pre-scalar is an excellent diagnostic tool, as it eliminates uncertainty of drive conditions from the measurements. Figure 5.47 shows the effect of operating voltage on the gate delay of each AND-NOR as calculated from the oscillation period of the pre-scalar. The gate delay predicted from simulations was 150 ps for 5 V operation. Figure 5.48 shows a similar plot with the saturation current, I_{dss}, of the FETs as the variable. However, from the few multiplexer wafers processed, there was insufficient spread in the threshold, and hence in I_{dss}, and ring oscillator data were used.

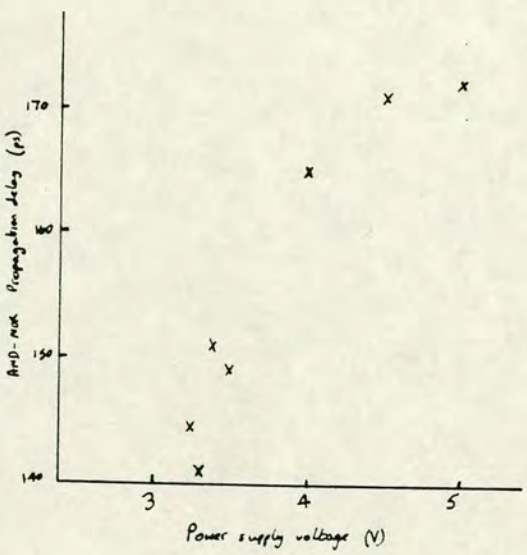


Figure 5.47 Variation of AND-NOR propagation delay with circuit operating voltage - obtained directly from multiplexer.

In practice, the chip dissipates about 600 mW at 5 V, and operates with a case temperature of 30-40°C, implying a junction temperature some 5-10°C higher. Since no temperature modelling of device parameters has been undertaken, it is not possible to verify the simulated data with confidence. However, the ratio between the oscillation speeds of the prescalar and the Johnson counter is both simulated and measured at 2.8:1, and the discrepancy between the

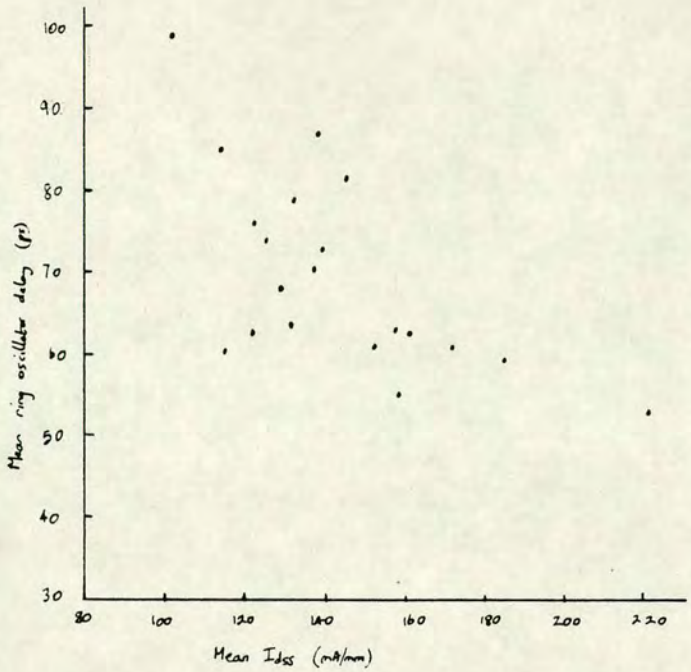


Figure 5.48 Variation of ring oscillator delay with saturation current of test transistor.

measured 170 ps and simulated 150 ps propagation delay (for the pre-scalar) is not great. These facts help to lend credence to the models, even though more work is required.

The other area which occupied much of the design effort, was the overall chip timing. In order to verify the work performed, it is necessary to compare the latched and unlatched data outputs. A difficulty arises in distinguishing between on-chip propagation delays, and measurement system delays. A variable delay line (built into the input amplifier of the sampling oscilloscope) was used to balance the delays through the two paths, such that at low frequency, the latched and unlatched signals showed a full clock period difference (figure 5.49a). As the clock frequency is increased, the relative timing of the two signals changes, as shown in the sequence of oscillographs in figure 5.49a-e. At a critical frequency (figure 5.49f), the latch is unable to store the data correctly, because its input is not constant throughout the sampling period. This waveform during this period of uncertainty appears as a series of dots, simply because a sampling oscilloscope is being used to display the result. This uncertainty spreads over two clock periods as the frequency is

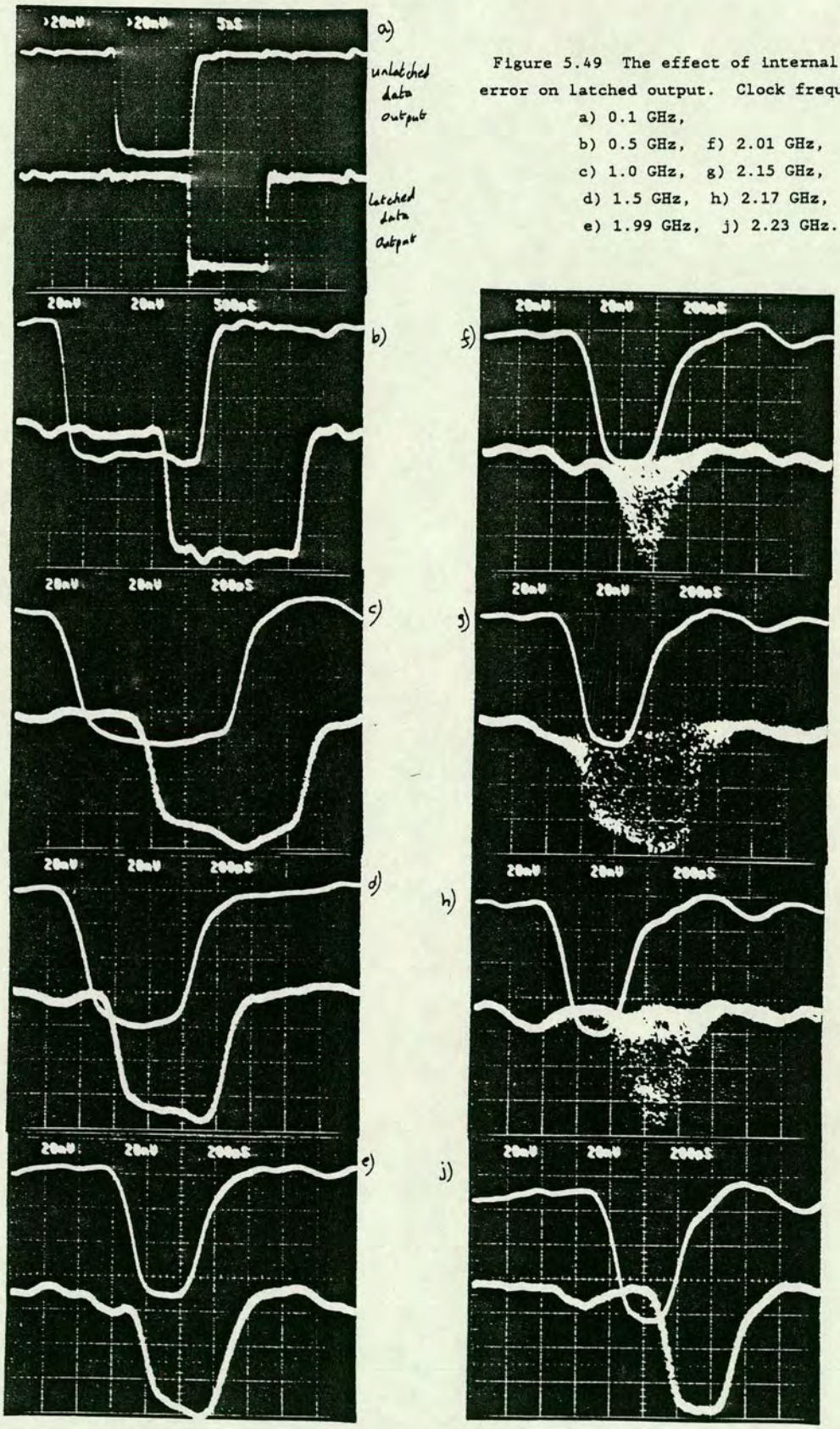


Figure 5.49 The effect of internal timing error on latched output. Clock frequency is:

a) 0.1 GHz,
b) 0.5 GHz, f) 2.01 GHz,
c) 1.0 GHz, g) 2.15 GHz,
d) 1.5 GHz, h) 2.17 GHz,
e) 1.99 GHz, j) 2.23 GHz.

increased further (figure 5.49g), because the data is shifting from one clock sampling window to another (see figure 5.18). Eventually, the data re-emerges properly latched, but one clock cycle later (fig 5.49j).

This behaviour, captured in figure 5.49, is exactly as predicted. The band of uncertain operation is quite narrow, but this is to be expected when only a single input is studied. This does not necessarily imply that the input data to the latch has an extremely fast edge, and in fact this has been shown not to be the case, as observed by the slow transitions at the unlatched data outputs. Unfortunately, it has not been possible to repeat this sequence for other operating voltages, because the frequency at which the circuit ceased to function (f_{stop}) coincided with the disappearance of the data. However, by changing the internal timing via the programmable delay, the window over which the data were in error was moved to a lower frequency, making measurement easier. Figure 5.50 shows the shift in one output edge as the controlled delay is switched in and out of circuit. In this condition, the uncertainty window was measurable at all operating voltages, but was still very narrow. The results of this analysis are presented in table 5.12.

These measurements allow some data on timing to be extracted, and these too are shown in table 5.12. Of some surprise is the very large voltage dependence of the programmable delay. This delay is caused merely by the propagation delay of the four inverters in the switchable section of the delay line. The implication is that the inherent switching time of these gates is much faster than the slew rate limited portion (see figure 5.34). This in turn suggests that these gates carry too great a fan-out. This finding is in direct agreement with the early failure of the latched output, despite the continued operation of the unlatched output. Note that this theory may be verified by driving the latch clock directly through the external input. Unfortunately, because of the restricted pin count on the existing packages, this experiment has not yet been performed.

Subsequent simulation of the behaviour of the high frequency clock propagation through a long chain of inverters, although not

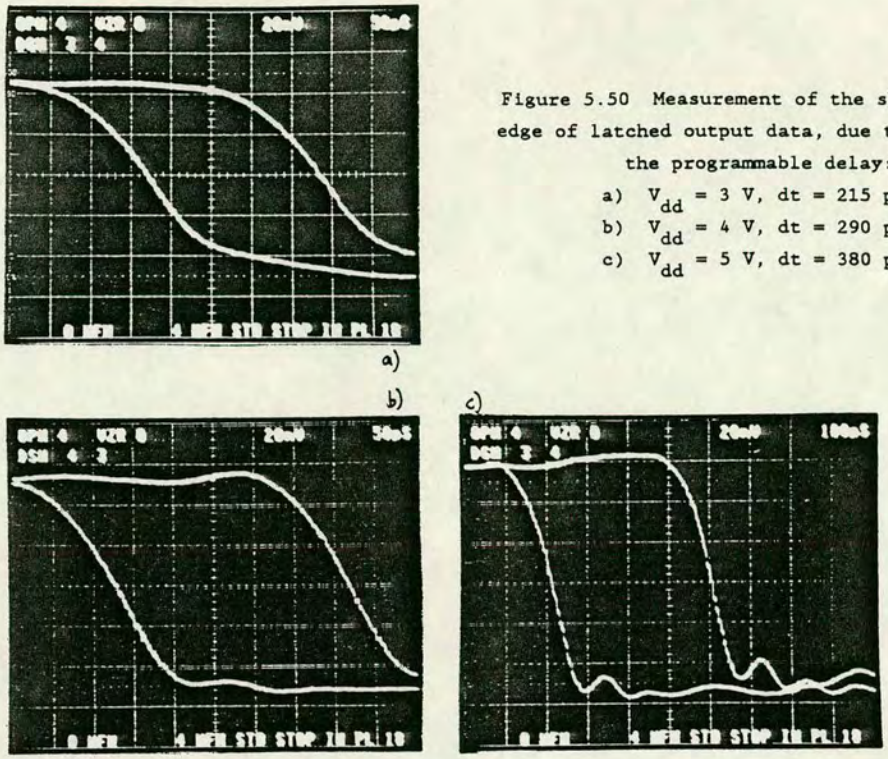


Figure 5.50 Measurement of the shift in one edge of latched output data, due to switching the programmable delay:
a) V_{dd} = 3 V, dt = 215 ps,
b) V_{dd} = 4 V, dt = 290 ps,
c) V_{dd} = 5 V, dt = 380 ps.

Table 5.12 Timing analysis of the multiplexer

Measurement	V_{dd}=3 V	V_{dd}=3.5 V	V_{dd}=4 V	V_{dd}=5 V	
Long delay:					
Data disappears	1.99	1.85	1.7	1.42	GHz
Data reappears	2.23				GHz
System delay	500	540	590	700	ps
Short delay:					
Data disappears	1.42	1.2	1.12	0.9	GHz
Data reappears	1.53	1.25	1.13	0.92	GHz
System delay	705	835	895	1110	ps
Prog. delay	205	295	305	410	ps
Latch failure		2.0	1.85	1.45	GHz

Note 1: system delay is defined as the unbalanced delay between clock and data at the D-latch. This is measured to first loss of data, because of failure to reappear. Strictly, however, it should be measured to window centre.

Note 2: Latch failure is identified by a DC level change when the circuit stops being clocked.

predicting failure at the low frequency measured, does indicate a much more rapid degradation of the signal than had been expected from lower frequency simulations. This is an area which must be addressed quite carefully for future circuits, especially those containing a large clock distribution network, and it should be noted quite carefully that the clock distribution is invariably the highest speed element on any chip. Circuits such as shift registers are particularly prone to problems. This situation should be contrasted with the timing within latches. Here, although the delays within the feedback loop are especially critical, the internal gates only operate at one half the frequency of the clock. Quite clearly, the clock distribution network requires the same degree of effort as that expended to optimise the latch design.

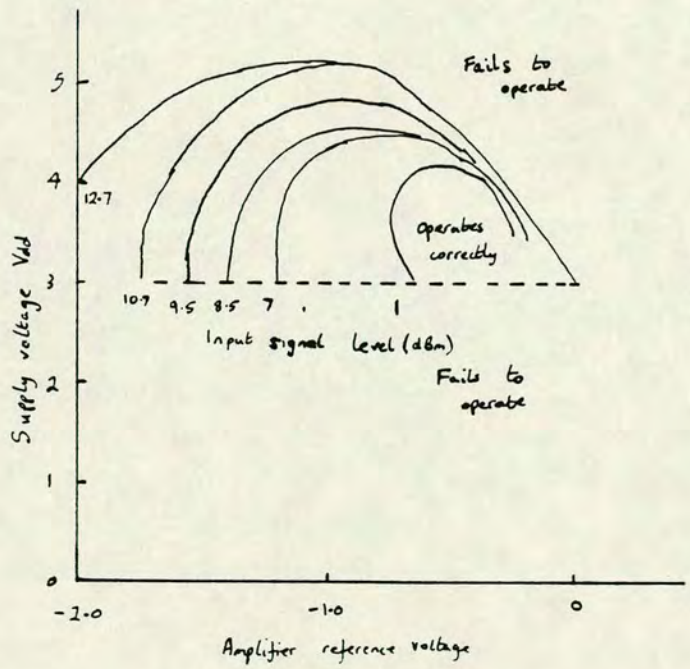


Figure 5.51 Schmoo plot of amplifier operation for input frequency of 2.3 GHz - no DC bias added to sine wave.

The final interesting aspect of circuit performance is the behaviour of the input amplifiers. These amplifiers were not used prior to this circuit, and the output was not directly available, so there is no direct correlation of performance. However, some data are available from an investigation of circuit operation versus input conditions. Figure 5.51 shows a schmoo plot of the input sensitivity

of the clock, showing that this type of amplifier forms an excellent interface between ECL and GaAs levels.

There is still room for more detailed analysis of circuit behaviour, especially if this is linked to further simulation. However, the circuit has been shown to function well, albeit with certain reservations, most notably over the clock distribution circuits. As a demonstration of the feasibility of CCL as a useful technique for MSI logic in GaAs, operating over a very wide range of threshold voltage, this circuit has been most successful. At the time that this design was performed, three design iterations would have been considered standard in order to ensure that a circuit functioned exactly to specification. By these standards, this performance is well on target for a first iteration in a completely new technology. Although no further design iteration has been performed for this circuit, a number of simpler, higher performance circuits has been built on the foundations established during this design exercise [455].

Chapter 6 Discussion and Recommendations

Thus far, this thesis has considered GaAs as an engineering material and focused towards the demonstration of capacitor coupled logic circuits built on GaAs. Along the way, consideration has been given to the limitations of the material, to alternative circuit structures, and to the necessary models and techniques which needed to be established before the chosen MSI circuit could be successfully designed. In this Chapter, the luxury of speculation is allowed in discussing the prospects for continued development of the material, the circuit techniques and the design methodology. The first Section will treat those modifications which might be made to the existing designs in the light of the circuit performance observed experimentally. The second Section will tread further afield and discuss the changes which might be made in circuit design in order to exploit the technology to the full. Thus, this will address the limitations of the capacitor coupled logic technique and suggest future modifications which will retain the low power and high speed, but which will allow operation at low as well as high frequency.

6.1 Improved performance from CCL

Until recently, the normal working practice for IC designs has demanded that 2 or 3 iterations (occasionally more) are required to obtain a circuit which functions fully according to the specification. More recently, computer aided design (CAD) tools have provided the means to verify circuit designs to a very high level of confidence before committing them to processing. In this so-called correct-by-construction technique [456], the functionality of the circuit can be guaranteed provided that the device models are accurate. These models for silicon have been built up over a number of different generations of designs, with most design changes being evolutionary in character, allowing existing models to be applied with only minor changes. Thus, IC design on silicon is now assured of a very high success rate.

6.1.1 Circuit design reworks.

With this as a backdrop, one must therefore assess this first generation IC design in GaAs. Because the same CAD tools are available, one may expect to claim the same level of confidence in the design. However, this is to ignore the revolutionary aspect of the switch from silicon to GaAs. Yes, the tools are available to ease the burden of design, but no, the models are not yet sufficiently adequate to guarantee first-time success. Thus, the nearest analogue from the history of silicon IC development must be drawn from the experimental phase of the CAD design tools. There, the degree of experience of silicon IC design was high, but the CAD tools were unsophisticated. In this case, the CAD tools are highly refined, but the models are lacking, or at least lagging behind. By analogy, one would still expect to need 2 or 3 iterations for this generation of design, but with that number reducing to 2 or perhaps even a single pass by the second generation of GaAs IC design. Thus, the performance of this first-pass design of the multiplexer circuit is creditable, and one would expect to improve significantly with a further iteration, during which the few existing weaknesses are eliminated, or improved.

The first question to address in this discussion is "where do these minor weaknesses lie?", or "how can the performance be improved?" Quite clearly, the limitation on the multiplexer circuit was in the performance of the clock buffer/ delay line combination, where the bandwidth of several cascaded stages proved inadequate compared with the remainder of the design.

As an experimental solution, it is clear that larger transistors should be used throughout the clock distribution network, as this is the only circuitry operating at the full speed, and one can afford to expend a greater fraction of the overall power budget in this area. This raises two further aspects: firstly how much should this increase be, and secondly what are the consequences of this increase?

The second question is actually the easier to answer, at least in the abstract. The obvious change is an increase in power consumption, which is accompanied by increased problems with both the distribution

of and noise currents on the voltage supply rail. The exceptional yield obtained for the multiplexer leads one to conclude that pinholes within the dielectric have formed a very low hazard, associated with short circuits between tracks on the two levels of metal. It therefore becomes feasible to consider the preferred method for power distribution, in which the positive and negative supplies are distributed in horizontal and vertical grids (or vice-versa) on the chip. It is even possible to consider the addition of MIM (metal : insulator : metal) capacitors [457] at the crossover nodes in order to help to maintain a very low impedance power supply. Both of these techniques lead to reduced cross-coupling arising from noise on the power rail. With this design modification, the additional power consumption should not prove too problematic as the earlier design limit was actually quite conservative.

Returning to the first question, by how much should the transistors be increased? Clearly, this problem should be tackled with the aid of a simulator. Further studies of the circuit have indeed shown that the clock distribution is the weak link. More detailed simulations, performed at the expected operating speed (rather than extrapolated to the operating speed) should give a basis for a decision.

However, in general, the simulations are significantly in error, particularly, for instance, when looking at the performance as a function of operating voltage. How much credence can therefore be placed on these simulations? Herein lies a significant dilemma, but one should bear in mind that the original design work, based on these inadequate simulations, yielded a circuit with very good overall performance, quite close to the predictions. Further simulation with the existing models is still therefore a worthwhile starting point.

To go beyond this stage, it is essential to improve the modelling still further. The first step in this chain is to continue to extract transistor parameters for a wide range of transistor sizes and threshold voltages. As the processing tolerances tighten, these data become more meaningful, and correlation between parameters becomes more realistic.

The vital element which is missing from the modelling thus far is to extract the temperature dependence of these parameters. When dissipating about 0.5 W, the operating temperature of the chip is probably some 10-20°C above ambient. This large temperature rise is found because no deliberate steps have been taken to maximise the thermal efficiency either in the bond between the die and package or between the package and the ambient environment. Quintessentially, the device models are only relevant if they can be used to obtain an accurate match between predicted circuit performance and the measured data: the transistor models alone are of little consequence. Thus to complete the loop accurately, temperature dependent models must be used to allow simulation of the effects of the temperature rise on the chip. By completing the loop of device measurement, device model, circuit model, circuit measurement, new device model etc, both device models and accuracy of circuit performance predictions can be improved. Although the predicted performance of the overall circuit was not grossly in error in this work, the detailed comparison between performance and prediction has been hampered by being unable to compare like with like. Only by obtaining a true comparison, which includes temperature effects, can the final confirmation be given that all is well.

In addition to the improvement arising from the introduction of temperature dependence into the device models, further confidence in the model parameters can be obtained by direct comparison with measurements using more than one technique. Thus, the transistor models should be verified at RF as well as DC. Equally, where possible, the measurement should be made using a truly representative geometry. Although the capacitance elements in the models well represent the diodes which have sufficiently large area to be measured directly, confirmation of these models in the transistor environment (small area) would be beneficial. This comparison can be obtained using s-parameter (or equivalent) measurements, coupled with a suitable small-signal device model.

Once these much improved models are available, it becomes practical to consider worst case design rather than the nominal value design which has formed the basis of most of this work. Although a major

advantage arising from the use of CCL circuit structures was the tolerance of the circuit to variations in device parameters, the experimental results have pointed to a highly significant degradation in performance as the circuit is operated away from the optimum conditions. Thus, the maximum clocking frequency of the circuit is considerably reduced as the supply voltage is increased from 3.5 to 5 V. During the course of this study, control over device parameters has been much improved. Similar improvements in the models can be anticipated and these would enable the optimum process requirements to be specified more accurately. It then becomes feasible to tune the fabrication process much more closely to the demands of the application, so that the circuits can operate near their peak potential. By demanding a much tighter control in threshold voltage, the transistor width ratios used in the circuits can be chosen with much greater care, to optimise the performance. With both tighter parameter spreads and operating specifications, the worst case analysis becomes not only practical, but also highly desirable.

6.1.2 Technology improvements.

So far, it has been assumed that the multiplexer performance can be improved by minor reworking of the design, in order to correct the unforeseen problems which arose. This assumption has been justified by comparison with a similar phase in the history of silicon IC development. Whilst this is a laudable goal, there are other, more radical ways in which circuit performance will be improved, but once again, the use of improved models is essential to differentiate fully between alternatives.

These performance improvements can be obtained by making selected changes to the process technology or to the layout philosophy. The value of accurate simulations lies in pinpointing those changes which are most beneficial, and therefore which offer the best return on the investment.

The most significant of these improvements arises from shrinking the dimensions (or design rules) of the process. The conventional understanding of such scaling is that the benefits accrue from the shrinking of the minimum feature size. As shown in Chapter 4, in the

"long channel" FET, the transconductance scales inversely with gate length whilst the input capacitance decreases with gate length. As both factors contribute to improved switching speed, the benefit is clear. Even when short channel effects begin to dominate, and these relationships are no longer accurate, there is still some continued improvement to the switching speed arising from a smaller gate length (figure 6.1).

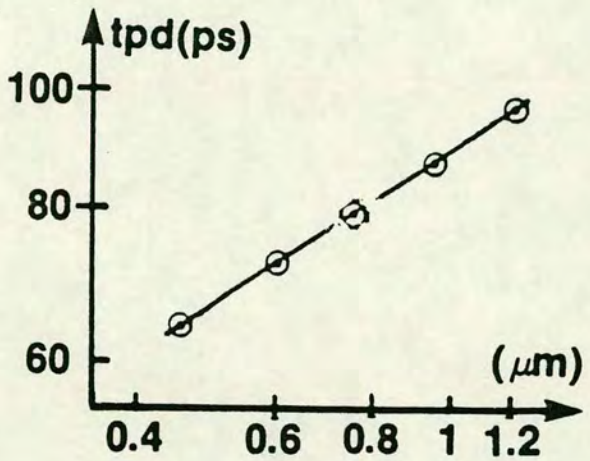


Figure 6.1 Increase in performance as gate-length is shrunk - after [458].

For a 1 um process, there is a considerable difficulty in continued scaling, because the dimensions approach the fundamental limits of resolution for optical lithography, and new, more expensive techniques are required. However, as the minimum dimension used in silicon ICs has reduced to 1 um and less, photolithography equipment capable of 0.7 um (and occasionally smaller) resolution has become available by moving to light sources of shorter wavelength [459]. Even so, the process of shrinking geometry to gain improved speed is running out of steam, and becoming more costly for each reduced improvement.

The major advantage of shrinking the minimum feature size comes when a similar shrink is applied to all dimensions. Then, the increased packing density and reduced power consumption allow greater complexity within the chip, giving an overall improvement in system

speed, by reducing the "wasted" performance associated with interface buffering. As the process used in the multiplexer uses 1 μm only for the gate stripe, whilst using (typically) 6 μm for the non-critical features, there is still a significant advantage to be gained from reducing these other dimensions.

If 2 or 3 μm were adopted for the minimum contact widths on the source and drain, as well as for the signal tracks (with a commensurate reduction in separation to 3 or 4 μm), the reduced capacitance would be reflected in improved speed. Not only would the parasitic capacitances within the transistor be reduced but, much more drastically, the loading due to the wiring capacitance would lessen. On semi-insulating material, the total effective capacitance per unit length does not vary significantly with the track width, except insofar as the lengths of track required to wire the same circuit are also reduced. Such a size reduction should not cause a significant change in yield, but would contribute usefully to a higher performance. Compared with further shrinking of the gate dimension, this improvement is obtained at a much reduced risk.

There is one remaining technique which offers further speed improvement. To date, the circuits have been designed by associating the coupling capacitor with each input. This method was chosen such that the logic element design was independent of fan-out; each input uses just the correct value of capacitor for its size. Furthermore, the node between the capacitor and the gate of the transistor is that of highest impedance, and is therefore the node most susceptible to the possibility of cross-talk. By directly associating the capacitor with the gate, this track is kept to minimum length, and the risk of crosstalk is minimised.

If the circuit was built on a conducting (junction isolated) substrate, there would be little penalty for this design choice. The capacitor parasitics would be largely area dependent, and the difference between n capacitors of area A, and one of area n.A would be slight. However, these substrates are semi-insulating and the parasitics are largely periphery dependent. Now, the use of several small capacitors is much more inefficient in reducing parasitics than

using one large one. The extra efficiency of the single large capacitor is even more apparent if the contact forming the source of the load transistor and the drain of the switching transistor is increased in size and combined with the cathode of the coupling diode capacitor, as shown in figure 6.2. The penalty for making this design change is that in many cases, the capacitor will be larger than required, unless each logic gate is individually tailored. However, the total stray capacitance on the output node of the logic gate should still be reduced.

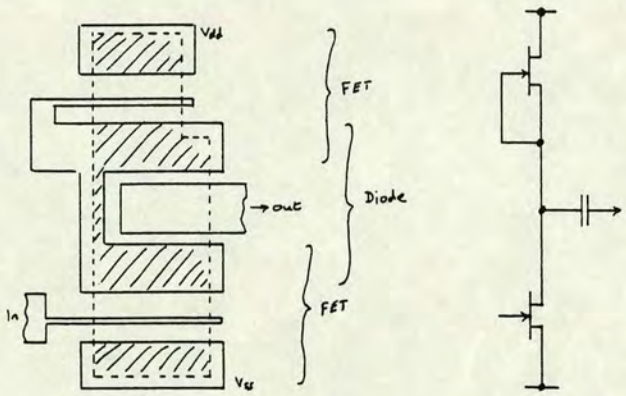


Figure 6.2 Layout of regrouped CCL inverter: a) physical construction, b) electrical circuit.

There is one remaining hidden benefit from this change. In the existing design, a significant portion of the stray capacitance of the coupling diode is directly to the drain of the succeeding transistor, and the Miller capacitance must therefore be considered in calculating the effective load presented to the previous stage. By physically shifting the capacitor to the previous output, the total effective gate-drain capacitance on any transistor will be reduced quite markedly.

In some gates, such as NAND or AND-NOR, there is considerable scope for positioning the capacitor on the drain of any of the switch transistors. Alternatively it may be distributed between the various drain contacts, as befits the optimum layout configuration.

The as-yet unquantified disadvantage of this technique lies with the increased risk of crosstalk.. In the simplest model of coupling, an

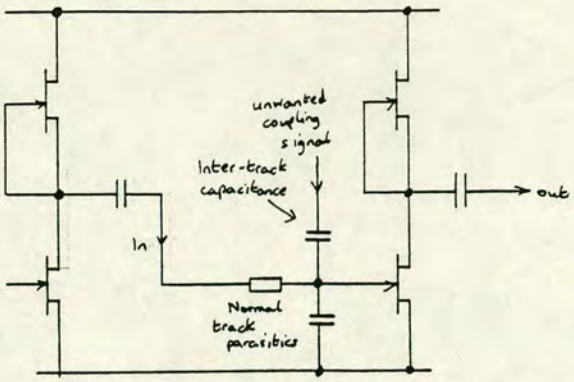


Figure 6.3 Model for crosstalk in modified CCL.

unwanted signal is attached to the sensitive node through a capacitor in the equivalent circuit of figure 6.3. However, the very nature of the coupling capacitor is that its stored charge far-exceeds the switching charge of the transistor gate capacitance. Thus, for any disturbance caused by the cross-coupled signal to be significant, the cross-coupled charge must be even larger than that holding the node and, this clearly demands an unrealistically large value for a stray coupling element. Thus, the low frequency effects of accidental coupling between lines should be small.

Looking at the dynamic performance, if the track resistance between the coupling capacitor and the transistor gate is large, even a small capacitor could couple sufficient charge to alter the state of the gate. Again this is unrealistic, because the circuit simply would not function with such a large resistor between elements.

One final possibility is that a resistive coupling path may exist. Using the new layout, this would allow the capacitor to be discharged more rapidly because of the larger area over which this leakage would occur. However, this mechanism may be dismissed even more readily than the others, as it would demand a dielectric so lossy that correct operation using the existing layout would be precluded.

We thus conclude that the problem of crosstalk is likely to be far less than perhaps envisaged, and one should therefore use this new layout in a trial circuit to verify these arguments.

A number of useful possibilities for enhancing the circuit performance, has been discussed. These arise by making some changes in both layout and design rules and also by correcting some faults from the existing design. Such improvements are possible without making severe changes to the circuit technology or to the process. Improved device, and thereby circuit modelling, is the key towards optimising circuit performance, aimed not solely at addressing the deficiencies of the multiplexer design, but also enabling future changes in process design rules to be well targeted, reaping the maximum performance benefit for the minimum risk changes.

A further benefit arises from the better understanding of device behaviour which accompanies improved device modelling. This better understanding can be used to tailor the doping profiles in the transistor to optimise the switching behaviour of the channel. As disclosed in chapter 2, many workers have started to use p-type buried layers beneath the gate to control the approach to threshold, thereby maintaining the transconductance at a high value, just when it is needed most. This should reduce the time spent on the corners of transitions (phases a and g of figure 5.35), and therefore it should give significant speed enhancements. In the fabrication process used for the multiplexer, the doping profile at the bottom of the channel was not well defined because of the recessed gate technology. A move to one of the self-aligned technologies would therefore be most beneficial, as this demands a much shallower profile which affords a sharper transition with the substrate. There would also be a small advantage in source resistance although the recessed gate already has a fairly low source resistance because of its pseudo-self aligned nature.

6.2 Improved circuit technology

The fundamental motivation for pursuing the capacitively coupled logic family was to enable circuits to be built, despite the very poor control of the transistor parameters. This had already been achieved using the BFL logic family [460], which was, at the time, a very power-hungry solution. The CCL approach has proved quite capable of coupling high speed with low power, as has been amply

demonstrated with this work, but the absence of static operation has been shown to be more restrictive than first envisaged.

A supplementary attraction of CCL is operation from a single power supply. Initially, this looked highly attractive for two reasons; to retain overall system compatibility with ECL circuits, and to simplify the on-chip routeing. Whilst the first of these still remains desirable, the performance of CCL is apparently compromised by the relatively high voltage required by ECL and, in truth the interfacing with ECL still requires additional reference lines and a non-standard power configuration. On the second point however, no net simplification to the wiring has been found in practice, because the overall design has been complicated by the limitations imposed by CCL.

It is thus highly desirable to improve the circuit technology whilst retaining the useful advantages of CCL, and this has in fact been achieved by a number of workers, primarily by marrying the CCL and BFL designs, as already described in Chapter 2 [461]. Section 6.2.1 will discuss this capacitor diode FET logic (CDFL) in more detail, and Section 6.2.2 will describe a novel, alternative approach which offers the possibilities of further improved performance, but which is based on the ideas developed for CCL. This new family of designs has been designated Capacitor Enhanced FET Logic (CEFL).

6.2.1 Capacitor Diode FET Logic

Here the BFL and CCL circuits are merged together to combine the attractions of the two techniques. Whilst the individual techniques offered high speed, only CCL offered genuine low-power with a very high tolerance to parameter variations, and only BFL the full static operation. Two basic designs are available based on the variants of BFL: with and without the source follower. Figure 6.4 shows the design of an inverter using these two approaches. The capacitor is used to give the high speed switching, whilst the diode chain and the current source are used to maintain a DC capability. The current source operates at a much reduced current compared with BFL, to give a low overall power dissipation. At first sight, the current through the pull down transistor is only required to bias the diodes at the

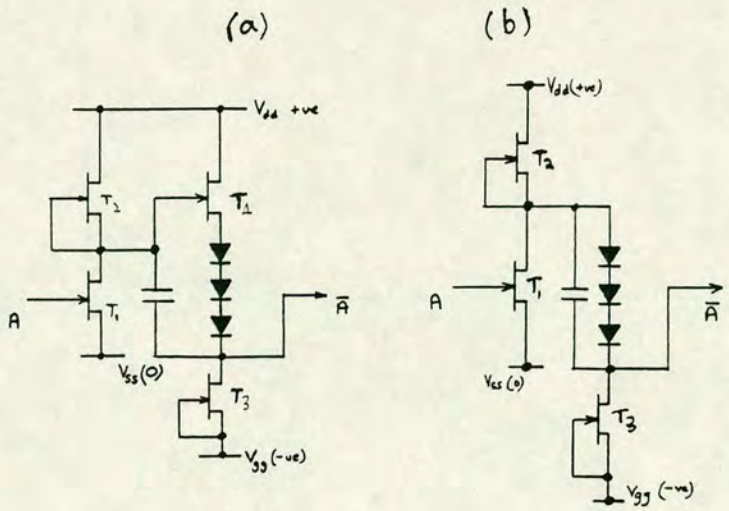


Figure 6.4 Capacitor diode FET logic inverter (CDFL)
a) with buffer,
b) without buffer.

knee voltage and to counteract the leakage currents of the capacitor. This can therefore be a very small transistor, and the V_{gg} rail does not contribute much additional power dissipation to the finished chip. However, a comparison of the two designs requires a more detailed analysis than this.

Turning firstly to the design incorporating the source follower. When the inverter is off (i.e. the output is high), the input to the level-shift branch sits virtually at the V_{dd} rail, as only the gate current of T4 flows through T2. T4 is thus turned hard-on and the lower end of the diode chain is clamped by the forward voltage on T1(b) of the succeeding stage. The current through T4 is thus divided between the gate current of T1(b) and the saturated current of T3. The coupling capacitor is thus fully charged to the voltage $(V_{dd} - V_{bi})$.

If only the capacitor's contribution to the coupling is firstly considered, then the route through T3, T4 and the diodes may be ignored. The combination is now much like CCL. As soon as T1(a) begins to switch on, the displacement current of the coupling capacitor will begin to switch T1(b) off, with the initial charge of the capacitor being partly shared with T1(b) to create the deep depletion region and cut-off the device. The parasitic capacitance on the drain of T1(a) simply slows down the slewing rate of this node, and that on the gate of T1(b) participates in the charge

sharing, effectively slowing down the transfer of charge to the desired node and reducing the efficiency.

If instead, the effect of the capacitor is neglected, then the switching of the succeeding stage begins because the gate voltage of T4 is pulled low when T1(a) switches on. The bias on T4 is such that the transistor has been forced very firmly into the linear region with a large drain current flowing and with both the gate and the drain sitting at nearly the same voltage. In this regime, the gate-drain capacitance is very large and there will be an initial switching delay because of charge-storage effects in the source follower. Following this delay, T4 turns off, because the stray capacitance on each of the nodes in the diode chain will temporarily hold the voltage on these nodes fixed. The final constituent of the switching action is controlled by the rate at which T3 can discharge these capacitors at the same time charging the depletion region under the gate of T1(b). This is the normal switching behaviour of BFL, and although not reported in the literature, the charge storage at the beginning of the transition is a major problem [462]. The stray capacitance on T1(a) drain and T1(b) gate also act to slow down the transition, in much the same way as in both BFL and CCL. The CCL circuit is thus expected to be faster on this transition than the BFL circuit.

When the combination circuit is considered, the charge storage on T4 will still be present, but once the T4 gate turns off, the fastest switching mechanism will now be via the capacitor. Once T4 has turned off, T3 will aid the capacitor in switching T1(b). During this transition, the voltage across the capacitor will fall (as it does in CCL), because it is sharing charge with T1b. The current sink action of T3 will help to balance some of this lost charge, reducing the voltage drop when compared with that occurring in CCL. The voltages on the intermediate nodes within the diode chain will not respond as quickly as the capacitor voltage, but once the diode chain comes into equilibrium, the voltage across the capacitor will be nearly fully replenished. The total voltage across the capacitor will still however be slightly smaller in this state because the

current through the chain is reduced as T1(b) is no longer drawing gate current.

Thus, for this transition, the speed of the merged circuit will lie between that of the CCL and BFL configurations.

Turning now to the transition at switch off, again the three circuit types must be considered. In CCL, T1(a) is switched off when its gate is taken low. The current flowing in the load device is therefore used to charge-up the drain capacitance of T1(a) and turn T1(b) on. The depletion charge of T1b will be transferred back into the coupling capacitor, causing the voltage across this component to rise once again. Thus, the rate of change at the gate of T1(b) must be slower than that at T1(a) drain, albeit that the total required swing is also reduced. As the end of the transition approaches, T2 comes out of saturation, reducing the switching current. Also, the instantaneous value of input capacitance of T1(b) becomes very high, reducing the coupling efficiency between the diode and T1(b). Completion of the switching transition is therefore relatively slow.

In BFL, the switch-off transition can be very efficient indeed. As the drain of T1(a) rises, the gate of T4 is made highly positive, producing a very large source current which is able to charge the intermediate nodes quickly as well as to force a very early turn-on of T1(b). The effective load placed on T2 is small, and its source therefore switches more quickly than in the CCL case. The current from T2 becomes small as it enters the linear region, but at the same time, T4 is switching from the saturated to the linear region maintaining an almost constant source current independent of the actual gate voltage. The transition in this circuit is therefore fast throughout. The switching rate of the circuit is controlled largely by the quality of the diodes - if they have a high series resistance, then they limit the switching currents available from T4, and if they have a large capacitance, more of the switching current is diverted into the capacitance. The design of these diodes is therefore important, as ideally they should offer a high voltage drop, and very low resistance and capacitance.

When the merged circuit is considered, the exact analysis is made much more difficult. Initially the capacitor will begin the transition, because the FET will be delayed by its intrinsic switching time. If one assumes that T4 then switches hard-on to dominate the later part of the cycle, then the voltage across the capacitor cannot increase because it is is no longer pulling charge out of T1(b). The capacitor must therefore act as a voltage clamp across the diode chain, preventing T4 from turning hard-on. We must then take the alternative view, in which the capacitor performs all the switching. Because now it does charge-up, its voltage must increase, forcing T4 to turn hard-on, again contradicting the assumption. The only consistent solution is that both elements play some part in the switching.

Again, the conclusion must be that the switching time of the merged circuit falls between the two cases. In switching from either of the two original circuits to the merged circuit, the balance of the compromises on switching times must therefore be altered. Overall, however, the merged circuit appears to behave very well.

The second of the merged circuits to be considered is that in which there is no source follower transistor. Again the analysis commences with the logic gate switched off. Now, the transistor T3 still draws current through the diode chain, and therefore through T2. In contrast to all the other circuits described, the drain of T1(a) no longer sits at the V_{dd} rail voltage in the logic-1 state.

When T1(a) switches on, T1(b) begins to switch off via the capacitor. As before, this causes the voltage across the capacitor to fall. Since this is also the voltage across the diodes, their voltage is reduced cutting off the flow of current. The current sink action of T3 does not change, so the current which was biasing the diodes must now be drawn from T1(b), helping to switch the input. Eventually, at equilibrium, the capacitor will again charge to the full voltage drop across the diodes, whence they will conduct as normal. This switching is very similar to that described for the version incorporating the source follower, except that there is no charge storage on the gate-drain capacitor to prevent immediate switching.

The overall switching time must thus be better than that with the source follower.

On the opposite transition, when T1(a) switches off, the current pulling T1(b) gate high must come partially from the capacitor and partially through the diode chain, exactly as described for the source follower circuit. However, the transistor T2 is not decoupled from the effects of capacitance on T1(b), and this node will switch more slowly. Because current through T2 never falls to zero, the final part of the switching cycle is more efficient than with plain CCL.

One must conclude overall that both circuits offer potential improvements over either CCL or BFL, because it is difficult to match both rise and fall times for these circuits - BFL is inherently good at pulling up, and CCL is very good at switching off. Both designs of CDFL offer a compromise in which the capacitor is used to switch-off efficiently and the diode chain to switch-on. The choice of circuit is dictated by the overall loading expected. For a lightly loaded circuit, the initial charge-storage delay of the source follower version is to be avoided, whereas the heavily laden circuit will switch faster with the help of the higher switching current of the source-follower. In principal, it would be ideal if the circuits could be used alongside each other, depending on the exact requirement. This again calls for further compromises, as it is difficult to optimise both circuits (with slightly differing voltage swings) together.

Since the aim is to produce a very low power circuit, practical considerations of transistor sizes dictate that the source follower should not be used. In either case, the size of T3 dictates the power dissipation in the level-shift chain. Since this is a constant current source (in the idealised circuit), its only effect on switching times comes from its drain capacitance, so there is no reason to use a short gate FET. A long, narrow gate transistor can therefore be used to obtain a low saturation current. However, when used, T4 does play a role in the switching time, and it must therefore be a fast device. This necessitates a short channel FET,

and it becomes impossible to balance the pair of transistors T3 and T4 whilst retaining very low power. Thus, unless very heavy loading is expected on most gates, the version without the source-follower is to be preferred.

Because the capacitor is now an integral part of the level shifting at the output, it is essential to integrate a single large capacitor (sufficient for worst-case loading) into each level-shift branch, rather than to associate one capacitor per following input. This forces the design change already recommended in Chapter 6.1.

In BFL, the design of low capacitance diodes was critical to good performance. These needed to have low series resistance and be operated at very high current density to give a large voltage drop. As this mode of operation forces the diode into both the high injection and the current crowding regimes, detailed modelling is difficult but important. In CDFL, the low resistance remains important, but the demands on low capacitance are not now as important. This allows the user more flexibility to add an extra diode into the chain, so that they may be operated at a much lower current density. Since the overall current through the chain should be kept low, this is a most useful relaxation, which allows the entire circuit to be optimised in a practical way.

One additional bonus accrues from the move to CDFL from CCL. Because the input high voltage on the gate node of each input transistor is actually fixed by the choice of supply voltages and component sizes, it is possible to bias the switch transistors into their fastest operating conditions. In practice this means forcing a large gate current in order to maximise the transconductance. This voltage clamping of each stage also ensures that a chain of inverters will operate in known conditions, even over a wide temperature range, and gives more confidence in the ability of the circuit simulator to predict the circuit behaviour (given the appropriate models). In CCL, the "floating" nature of these nodes was a potential source of unpredictability, especially at elevated temperatures.

It is therefore recommended that the CCL circuits be converted into CDFL designs without source followers. Obtaining the exact transistor sizes to be used for best results will require considerable simulation, beyond the scope of this Chapter.

Following this switch to static operation, a number of possibilities are opened up for new circuit designs and some of these are worth exploring.

The most immediate benefit arises from the removal of the constraints associated with charging the coupling capacitors, especially where the pre-charging control circuitry was required. Further advantages arise from the lifting of the constraint on NOR gate circuitry within feedback loops. The greater flexibility of fan-in of the NOR gate allows the edge triggered flip-flop to be used in places where the ultimate speed is not required. The AND-NOR latch which proved so efficient a circuit can be configured afresh, using an AND-NOR gate for the master and a (new) OR-NAND gate for the slave (the reverse is also possible). The layout of this circuit is more complex than that containing just the single type of building block, but it may offer a better isolation of noise from input to output, especially when there is skew on the clocks. An important derivative of this circuit is shown in figure 6.5. Because the clock control of the feedback paths has been removed, this circuit is not as fast as the earlier design, requiring a period of 4.t for a complete clock cycle. However, it is probably the fastest single clocked latch circuit. Since there is no possibility of slowing down the circuit by skew on the complementary pair of clock lines, this circuit may, in practice, prove nearly as fast as the AND-NOR and it can be driven from just the single input. Where circumstances allow, the adoption of this circuit can give a significant saving in the overall power budget of a chip, eliminating all control gates and half the buffer drivers.

Thus, the elimination of the non-static operation associated with CCL can be translated directly into benefits. The simplified design translates immediately into a cost advantage at the design phase, and the small power penalty of each gate can actually be regained from a reduction in total gate count. The corresponding penalty of extra

component count per gate cannot be quantified without recourse to specific design rules. Again the reduction in total number of gates allows some of the increase in area to be recouped. It is a useful exercise to study various possible layout configurations of CDFL circuits to obtain more detailed trade-offs, especially if this is performed against various sets of design rules. This is left for further studies.

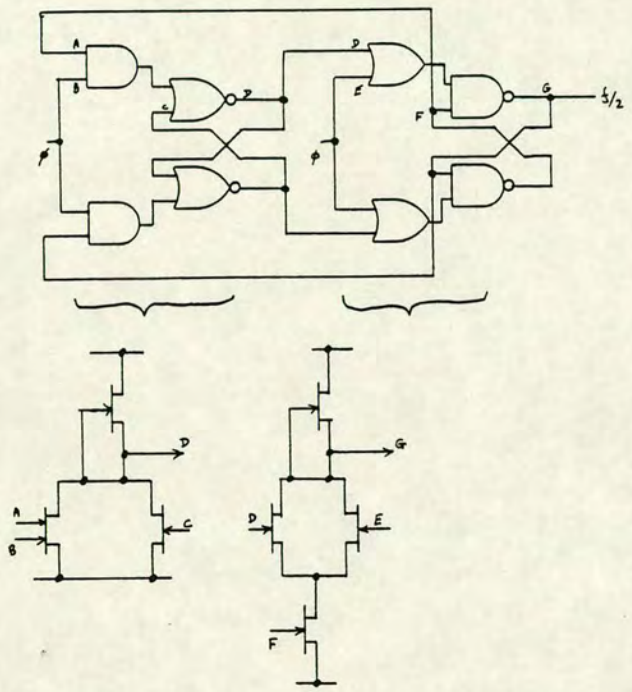


Figure 6.5 AND-NOR: OR-NAND divider and constituent gates.

6.2.2 Capacitor Enhanced FET Logic (CEFL)

So far, this chapter has investigated alternative circuit techniques based on established work, but has used the insights gained from this study to make reasoned comparisons. However, the knowledge gained from this research suggests that there is an, as yet, unexplored family of circuits which hold considerable potential. These circuits have been labelled as Capacitor Enhanced FET Logic or CEFL, because the principles of capacitor coupling are used to enhance the performance of the logic rather than to perform the switching.

Figure 6.6 shows a BFL inverter converted into a CEFL inverter, by adding the capacitor between the input and the gate of T3. In the

absence of the capacitor, when T1 switches on, T4 is cut off, thereby eliminating the current through the diode chain. The saturated current of T3 is thus available to switch the following stage. The switching speed is thus directly linked to the current and therefore the power consumption of the source follower chain. With the capacitor included, the input transition from low to high level turns T4 off as expected, but at the same time causes the pull-down current of T3 to increase by forcing its gate into forward bias. Thus, a small transistor may be used for T3 giving a low static current (and therefore low power consumption), with its current being increased automatically when a large value is required to switch off the succeeding stage.

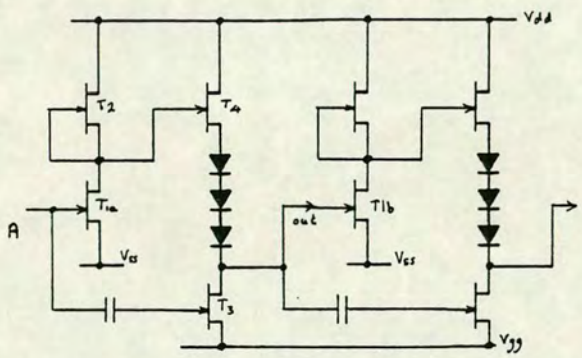


Figure 6.6 BFL Inverter converted to CEFL.

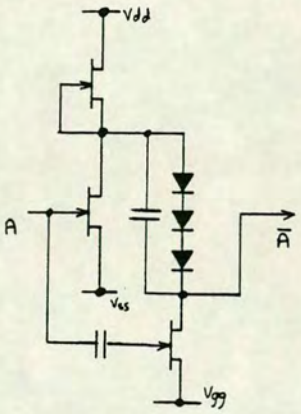


Figure 6.7 CEFL extension to CDFL inverter.

Similarly on the other transition, in the conventional BFL circuit, the current through T3 subtracts from the total current available to switch T1(b). At this point, it would be convenient to reduce the current through T3, and in CEFL, this does indeed happen.

The concept of CEFL is not restricted to the BFL circuit, but may indeed be applied to any of the level-shifted FET logic families, excluding CCL in its original form. Figure 6.7 shows the application to CDFL. The extension to further variants is clear.

The intention of the CEFL modification is to reduce the power delay product associated with the logic gate, either by reducing power at a

fixed speed, or by increasing speed for a given power consumption. As with any new circuit, their are some penalties to pay for these attractive properties. The most obvious is the increase in both complexity and chip area occupied by the gate. The capacitor must have a larger value than the gate-source capacitance of T3, but since the transistor is small, this area penalty may be quite small. Note however, that the long gate transistors, as suggested for CDFL in Section 6.2.1, may not be suitable in CEFL because of the large capacitance and poor switching characteristics.

Other problems arise too. The input capacitance of the logic gate is increased, with this capacitor now being governed by the combined gate areas of T1 and T3, plus the additional stray capacitance of the capacitor itself. This increase will actually slow down the previous gate. However, this effect is anticipated to be much smaller than the potential gain arising from the use of the new circuit.

There is, however, one serious problem. This arises when a logic gate more complex than an inverter is patterned. If a NOR gate is required, then each input to the NOR gate should be capable of enhancing the current through T3. One solution is to wire together a capacitor from each of the inputs, as shown in figure 6.8a. This will cause major problems with cross-talk to other nodes, as a change at input A will cause a noise spike at input B etc. A second alternative is to split T3 into two parallel transistors for the two input gate, with each half responding to the appropriate input (figure 6.8b). Each input can now only enhance the current by half the total amount. Clearly, the larger the fan-in, the more complex the circuit, and the smaller the enhancement available. Furthermore, the switching speed is now pattern dependent, as multiple inputs switching together give a faster transient than that of a single input change. Whilst the NOR configuration is possible but untidy, the NAND combination becomes even more problematic.

At first glance therefore, the CEFL circuit families are only of use for inverters. This does not detract from the value, because almost invariably, buffer circuits use a single input, and the CEFL

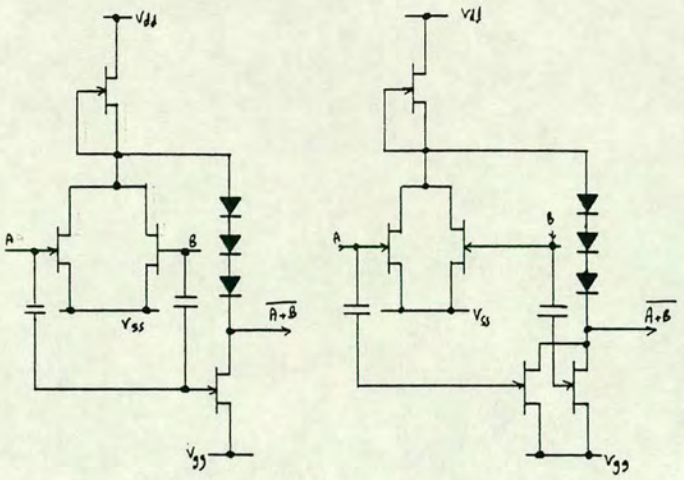


Figure 6.8 Extending CEFL to NOR gates.

connection could be viewed as a method of power reduction in large buffer circuits used for line driving.

There is a second category of circuits which would benefit, perhaps unexpectedly. Table 6.1 gives the switching sequence of the familiar AND-NOR T-latch.

Table 6.1 AND-NOR T-latch sequence

ϕ	A	B	C	D
0	1	0	0	1
1	0	1	0	1
0	0	1	1	0
1	1	0	1	0
0	1	0	0	1

In this circuit, the gate giving output A has four separate inputs (ϕ , $\bar{\phi}$, B and D), and all four would need combining to give the correct enhancement at the pull down transistor on output A. However, the transient from node B alone can be used to enhance the switching speed, as output A is always its inverse. Similarly, the output of A can be used to enhance B, and also the pair of outputs, C and D, can enhance each other. Once the switching has started to take effect, because of the logic conditions on all four inputs, the

capacitor enhancement from just the single input will speed-up that transition.

Since the speed performance of cross-coupled gates (as in the T-latch just described) and of clock buffer lines have proved to be two of the most critical areas in determining the overall performance, it is well worth noting that these are the areas most suited to the application of capacitor enhancement.

6.3 Concluding remarks.

This thesis has described a wide-ranging research project targetted specifically at the demonstration of the ability to design and manufacture MSI complexity integrated circuits in gallium arsenide. The work began in the infant days of IC processing with GaAs, and this thesis has therefore covered the development of a number of techniques which would now be taken for granted. The author has had to battle with unforeseen problems related to poor materials quality in the early days, as well as proceeding with those aspects more naturally related to IC design. New, and practically useful models have been developed along the way, aimed at helping with the understanding of component behaviour in GaAs. Furthermore, the design studies have included considerations of the extra complexities of component layout and packaging for use at Gbit/s data rates. Despite the profound depth of some of these individual side-topics, and the considerable temptation to study some at greater depth, none of them has been allowed to detract from the main goal of the exercise - the successful fabrication of the 8:1 multiplexer for 2.4 Gbit/s operation; the chosen MSI complexity IC.

I am pleased to report that this quest was triumphant, and that the IC was fabricated successfully and worked close to the specification. Perhaps of greater relevance, is that other workers have been able to use the knowledge and techniques gained along the way, not only to repeat the exercise and build other GaAs components, but also to act upon some of the recommendations of this Chapter, in order to extend the performance to 4 Gbit/s. Interestingly, the capacitor coupled logic technique pioneered in this work has also been adopted by many

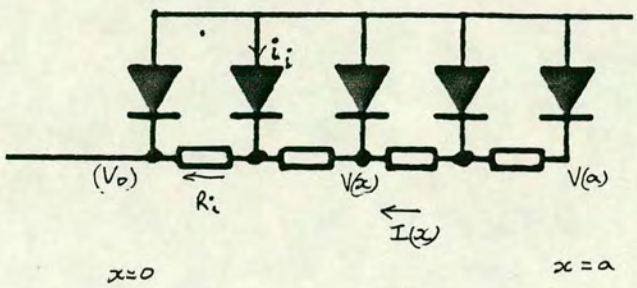
of the commercial companies, in their desire to improve both manufacturability and operating margins, whilst keeping low power operation [463].

Beyond this, this Chapter has discussed some of the issues associated with improving the performance of GaAs integrated circuits. The first Section has suggested ways of improving the types of circuit which have formed the motivation of the work for this thesis. Section 6.2 has concentrated on ways of improving performance by combining the advantages of capacitor coupling with other techniques, both established and novel. Neither of these Sections has deviated very far from the technology pattern set by the earlier Chapters of this work. This is a deliberate decision, as one of the most profound truisms which I have learnt during the period of this study is that the best solution to a problem is to be found by taking the smallest possible steps from the current position. Nowhere can this be more true than in semiconductor research.

If one were to have the freedom to start afresh in 1988, I should aim either for a silicon bipolar or a heterostructure bipolar solution. If I were constrained to work on GaAs MESFETs, I might just contemplate working with normally-off devices, which have held promise for so long. However, there is no doubt that, for the present, and despite all the opinion to the contrary, normally-on GaAs MESFETs still hold the edge in both usefulness and availability.

Appendix 1. Current in a distributed diode.

The aim is to obtain the current-voltage relationship of the distributed diode model shown in figure A1.



We know that the current flowing in a diode filament and the resistance beneath that filament is given by:

$$i_i = I_o.w \; . \; \{ \exp \left(\frac{q[V_o - V(x)]}{n.kT} \right) - 1 \} \; . \; dx \qquad -1$$

$$R_i = \frac{R_o}{w} \; . \; \{ \frac{[V_{bi} - V_t]^{0.5}}{[V_{bi} - V_t]^{0.5} - [V_{bi} - V_o + V(x)]^{0.5}} \} \; . \; dx \qquad -2$$

where I_o is the saturation current density and R_o is the sheet resistivity of the complete layer, which is assumed to have a constant doping level. The -1 term in equation 1 is only relevant at extremely low current, under which conditions, the whole model could be replaced by a non-distributed model. This term can therefore be ignored under high injection, when it is appropriate to adopt the distributed model. We can therefore formulate the integral equations relating the diode voltage and current at a position x:

$$I(x) = \int_{y=x}^{y=a} I_o.w. \exp \left(\frac{q[V_o - V(x)]}{n.kT} \right) .dy \qquad -3$$

$$V(x) = \int_{y=0}^{y=x} \frac{I(y).R_o}{w} \; . \; \{ \frac{[V_{bi} - V_t]^{0.5}}{[V_{bi} - V_t]^{0.5} - [V_{bi} - V_o + V(y)]^{0.5}} \} .dy \qquad -4$$

from which we may derive a second order differential equation:

$$\frac{d^2V}{dx^2} \cdot \left\{ \frac{[V_{bi}-V_t]^{0.5} - [V_{bi}-V_o+V(y)]^{0.5}}{[V_{bi}-V_t]^{0.5}} \right\} \qquad \}-5$$

$$- \left\{\frac{dV}{dx}\right\}^2 \cdot \frac{1}{2\{[V_{bi}-V_t].[V_{bi}-V_o+V(y)]\}^{0.5}} + R_o.I_o.\exp\left\{\frac{q[V_o-V(x)]}{n.kT}\right\} = 0$$

The relationship between the current and voltage of the diode may then be derived by solving this equation subject to the pair of boundary conditions:

$$\text{at } x=0,\ V=0\ ;\quad \text{at } x=a,\ \frac{dV}{dx} = 0 \qquad -6$$

This equation can only be solved in the general case using numerical techniques. However, under high injection, we may assume that the resistivity of the sub-layer is nearly constant, because the depletion is very small. Equation 5 now reduces to the form in (7), and by using substitutions (8), this may be rewritten as equation (9) and integrated to give (10).

$$\frac{d^2V}{dx^2} = - R_o.I_o.\exp\left\{\frac{q[V_o-V(x)]}{n.kT}\right\} \qquad -7$$

$$\text{Put } \phi = \frac{q.V(x)}{n.kT} \quad \text{and} \quad U = \frac{q.R_oI_o}{n.kT} \cdot \exp\left\{\frac{q.V(x)}{n.kT}\right\} \qquad -8$$

$$\left\{\frac{d\phi}{dx}\right\}^2 = -2U.\exp\{-\phi\}.d\phi \qquad -9$$

$$\frac{d\phi}{dx} = +/- \{ 2U [\exp(-\phi) + c']\}^{0.5} \qquad -10$$

Now we know that the current is proportional to the left hand side of this equation, so only the positive root is required. Applying the boundary condition at x=a, we can see that, following back substitution, the constant of integration is given by (11) which always forces a real solution to equation 10.

$$c' = - \exp\left\{- \frac{q.V(a)}{n.kT}\right\} \qquad -11$$

We may therefore make a further substitution into equation 10, to enable the second integration to be performed. Using (12) as this new substitution, (10) is transformed into (13), which is readily integrated to give equation 14.

$$z^2 = \exp(-\phi) - C^2 \quad \text{and} \quad C = \exp\left\{-\frac{q.V(a)}{2.n.kT}\right\} = \exp\{-\phi_o\} \qquad -12$$

$$\frac{1}{z^2 + C^2} \cdot \frac{dz}{dx} = \sqrt{\frac{U}{2}} \qquad -13$$

$$\frac{1}{C} \cdot \tan^{-1}\left\{\frac{z}{C}\right\} = -\sqrt{\frac{U}{2}} \cdot x + B \qquad -14$$

Re-applying the boundary condition at x=a, we note that z=0, from which we can obtain the new integration constant. Equation 14 can now be rewritten (15) and recombined with (12), to give equation (16).

$$z = C.\tan\left\{C\sqrt{\frac{U}{2}}.(a-x)\right\} \qquad -15$$

$$\exp\{-\phi\} = \exp\{-\phi_o\} \cdot \sec^2\left\{\exp\left[-\frac{\phi_o}{2}\right].\sqrt{\frac{U}{2}} \cdot (a-x)\right\} \qquad -16$$

Equation 16 may be substituted into (3) to give (17), which can again be integrated into (18).

$$I = I_o.w.\exp\left\{\frac{q.V_o}{n.kT} - \phi_o\right\}.\int_{x=0}^{x=a} \sec^2\left\{\exp(-\phi_o).\sqrt{\frac{U}{2}} \cdot (a-x)\right\} \qquad -17$$

$$I = I_o.w.\sqrt{\frac{2}{U}}.\exp\left\{\frac{q.V_o}{n.kT} - \frac{\phi_o}{2}\right\} \cdot \tan\left\{\exp\left(-\frac{\phi_o}{2}\right).\sqrt{\frac{U}{2}}.a\right\} \qquad -18$$

A final change of variable using (19) allows the current to be expressed in terms of a single unknown, ψ, as in (20). Equation 16 may be rewritten using (19) to give equation 21. By applying the boundary condition at x=0 to this, and using (8), ϕ_o may also be written (22) as a function of this single parameter, ψ.

$$\psi = \exp\left\{ -\frac{\phi_o}{2} \right\} \cdot \sqrt{\frac{U}{2}} \cdot a \qquad -19$$

$$I = 2 \cdot \frac{w}{a} \cdot \frac{n.kT}{q} \cdot \frac{1}{R_o} \cdot \psi \tan(\psi) \qquad -20$$

$$\exp\{-\phi\} = \exp\{-\phi_o\} \cdot \sec^2\left\{ \psi \cdot \left(1 - \frac{x}{a} \right) \right\} \qquad -21$$

$$\exp\left\{ -\frac{\phi_o}{2} \right\} = \cos(\psi) \qquad -22$$

Finally, by substituting (22) and (8) into (19), this arbitrary parameter may be expressed solely in terms of the known parameters (23).

$$\cos(\psi) = \frac{\psi}{a} \cdot \left\{ \frac{2.n.kT}{q.I_o.R_o} \right\}^{0.5} \cdot \exp\left\{ -\frac{q.V}{2.n.kT} \right\} \qquad -23$$

Equations 20 and 23 form the pair of equations which must be solved numerically in order to obtain a solution. However, when ψ is small, $\cos(\psi)$ may be set to unity, and $\tan(\psi)$ may be set equal to ψ, in which case the current is given by the standard diode equation (24). Note that this vindicates the omission of the (-1) term from equation 1, as we have shown that at low currents the resistor has no effect.

$$I = I_o.w.a. \exp\left(\frac{q.V_o}{n.kT} \right) \qquad -24$$

In the other extreme, when V is very large, ψ must tend towards $\pi/2$, whence, $\tan(\psi)$ becomes equal to $\sec(\psi)$, and equation 23 can then be directly substituted into 20 to give (25).

$$I = \left\{ \frac{I_o.2.n.kT}{R_o \cdot q} \right\}^{0.5} \cdot w \cdot \exp\left(\frac{q.V_o}{2.n.kT} \right) \qquad -25$$

By substituting typical values and an anode length of 5 μm, the low current case is given for V<0.5 V and the high current case for V>0.8 V, surprisingly low values.

REFERENCES

1 C. Hilsum, Semiconducting III-V compounds, C. Hilsum, A.C. Rose-Innes, pub Pergamon, Oxford, (1961)

2 M.J. Howes, D.V. Morgan (eds), GaAs materials, devices and circuits, pub Wiley, Chichester, (1985)

3 M. Rocchi, Tech dig. ESSDERC, (1984)

4 D. McKie, C. McKie, Crystalline solids, pub Nelson, London (1974)

5 L.C. Kimmerling, J.M. Parsey (eds), Defects in semiconductors, pub Metallurgical soc., New York, 1985

6 J. Lagowski, H.C. Gatos, J.M. Parsey, K. Woda, M. Kamuriska, W. Walakeiwicz, Appl. Phys. Lett., 40, 342 (1982)

7 R. A. Smith, Semiconductors, pub Cambridge University press, Cambridge, (1978)

8 J.R. Chelikowski, M.L Cohen, Phys. Rev., B14, 556, (1976)

9 R.A. Smith, Wave mechanics of crystalline solids, pub Chapman and Hall, London (1969)

10 C. Kittel, Introduction to solid state physics, pub Wiley, New York (1976)

11 D. Long, Energy bands in semiconductors, pub Wiley, New York, (1968)

12 ibid.

13 H. Kressel and J.K. Butler, Semiconductor lasers and heterojunction LEDs, pub Academic press, New York, (1977)

14 H.C. Casey, M.B. Panish, Heterostructure lasers parts A and B, pub Academic press, New York, (1978)

15 J.L. Moll, Physics of semiconductors, pub McGraw Hill, New York, (1964)

16 Properties of GaAs, pub Inspec, London (1986)

17 A.D. Welbourn, IEE Proc., 129I 157 (1982)

18 P. Smith, M. Inoue, J. Frey, Appl. Phys. Lett., 37, 797, (1980)

19 C.S. Chang, H.R. Fetterman, Solid State Electron., 29, 1295, (1986)

20 H.M. Rosenberg, The solid state, pub. Clarendon Press, Oxford, (1978)

21 A. Animalu, Intermediate theory of crystalline solids, pub Prentice Hall, New Jersey (1977)

22 R.J. Elliot, A.F. Gibson, An introduction to solid state physics, pub Macmillan, London (1974)

23 J.M. Ziman, Principles of the theory of solids, pub Cambridge University Press, Cambridge, (1972)

24 K. Hess, IEEE Trans. Electron Dev., ED-28, 937, (1981)

25 Y.C. Wang, Y.T. Hseih, Int. J. Electronics, 47, 49, (1979)

26 J.G. Ruch, W Fawcett, J. Appl. Phys., 41, 3843, (1970)

27 C.D. Thurmond, J. Electrochem. Soc., 122, 1133, (1975)

28 A.S. Grove, Physics and technology of semiconductor devices, pub Wiley, New York, (1967)

29 C. Hyun, C. Vogelsang, J. Notthoff, GaAs IC Symp. Tech. dig., 119, (1986)

30 C.T. Foxon, J.A. Harvey, B.A. Joyce, J. Phys. Chem. Sol., 34, 1693, (1973)

31 N. Newman, W.G. Petro, T. Kendelewicz, S.H. Pan, S.J. Eglash, W.E. Spicer, Jap. J. Appl. Phys., 57, 1247, (1985)

32 Properties of GaAs, pub Inspec, London (1986)

33 W. T. Kleinman, J.C. Phillips, Phys. Rev., 118, 1164, (1960)

34 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)

35 R. Zuleeg, J.K. Notthoff, G.L. Troeger, IEEE Electron Dev. Lett., EDL-5, 21, (1984)
36 D.E. Holmes, R.T. Chen, K.R. Elliot, C.G. Kirkpatrick, Appl. Phys. Lett., 40, 46, (1980)
37 C.Y. Ho, R.W. Powell, P.E. Lilley, Thermal conductivity of the elements - a comprehensive review, pub Am. chem. soc., New York (1975)
38 B. Bosch,IEE Proc., 127I, 254, (1980)
39 B. Bosch, Proc. IEEE, 67, 340, (1979)
40 C.C. Lee, A.L. Palisoc, GaAs IC Symp. Tech. dig., 115, (1986)
41 Properties of GaAs, pub Inspec, London (1986)
42 D.F. Horne, Microcircuit production technology, pub Adam Hilger, Bristol, (1986)
43 S.K. Gandhi, The theory and practice of microelectronics, pub Wiley, New York, (1968)
44 P.M. Asbeck, C.P. Lee, M.F. Chang, IEEE Trans. Electron Dev., ED-31, 1377, (1984)
45 M. Hasegawa, T. Sawada, T. Kitagawa, GaAs IC Symp. Tech. dig., 144, (1983)
46 M. Simons, GaAs IC Symp. Tech. dig., 124, (1983)
47 C.F. Gibbon, Avionics panel 49th symp., (AGARG-CP-380), 8, 1, (1985)
48 R. Loudon, Quantum theory of light, pub. Clarendon Press, Oxford, (1979)
49 H.C. Casey, M.B. Panish, Heterostructure lasers part A, pub Academic press, New York, (1978)
50 R.D. Martin-Royle, Br. Telecom Eng. J., 1, 190, (1983)
51 H.C. Casey, M.B. Panish, Heterostructure lasers part B, pub Academic press, New York, (1978)
52 F. Capasso, Margaritondo, Heterojunction band discontinuities: physics and device applications, pub North Holland, Amsterdam, (1987)
53 M.A. Alframowitz, Solid State Commun., 15, 59 (1974)
54 F. Capasso, Margaritondo, Heterojunction band discontinuities: physics and device applications, pub North Holland, Amsterdam, (1987)
55 G.P. Agrawal, N.K. Ditta, Long-wavelength semiconductor lasers, pub. van Nostrand Rheinhold, (1986)
56 ibid.
57 N. Williams, Electron Aust., 45, 28, (1983)
58 S. Gallinger, Opt. Info. systems, 7, 391, (1987)
59 B. Bosch, Proc. IEEE, 67, 340, (1979)
60 R. L. van Tuyl, IEEE Trans. Electron Dev., ED-29, 1031, (1982)
61 P.M. Solomon, Proc. IEEE, 70, 489, (1982)
62 S. Seki, H. Hasegawa, GaAs IC Symp. Tech. dig., 119, (1982)
63 P. Wolf, IBM J. Res. Dev., 14, 125, (1970)
64 R.S. Pengelly, Microwave field effect transistors - theory, design and applications, pub Wiley, Chichester, (1982)
65 A.D. Welbourn, IEE Proc., 129I 157 (1982)
66 D.T. Young, J.C. Irwin, Proc. IEEE, 53, 2130, (1965)
67 C. Hilsum, Proc. IRE, 50, 185, (1962)
68 J.B. Gunn, Solid State Commun., 1, 88 (1963)
69 H. Kroemer, Proc. IEEE, 52, 1736, (1964)
70 J. Mun, Electron. Lett., 13,275, (1977)
71 A.A. Salles, J.R. Forrest, Proc. 11th European Mic. conf. tech. dig., 81 (1981)

72 A.D. Welbourn, IEE Colloquium on ICs above 1 GHz, (1985)
73 P.M. Solomon, Proc. IEEE, 70, 489, (1982)
74 S. Evans, J. Delaney, C. Fuller< D. Boone, C. Dubberley, J. Hoff, J. Stidham, B. delaTorre, M. Vernon, M. Wdowik, GaAs IC Symp. Tech. dig., 109, (1987)
75 H. Kroemer, Proc. IEEE, 70, 13, (1982)
76 E.H. Nicollian, J.R. Brews, MOS physics and technology, pub Wiley, New York, (1982)
77 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)
78 K. Yamaguchi, S. Takahashi, IEEE Trans. Electron Dev., ED-28, 581, (1981)
79 W.E. Spicer, P.W. Chye, P.R. Skeath, C.Y. Su, I. Landau, J. Vac. Sci. Technol., 16, 1422, (1979)
80 D. Frohman, Bentchowsky, Proc. IEEE, 58, 1207, (1970)
81 M. Rocchi, Tech dig. ESSDERC, (1984)
82 F.L. Schuermeyer, IEEE Trans. Electron Dev., ED-28,541, (1981)
83 N. Yokoyama, T. Mimura, M. Fukuta, IEEE Trans. Electron Dev., ED-27, 1124 (1980)
84 F.L. Schuermeyer, H.P. Singh, R.L. Scheres, D.L. Mays, IEDM Tech. dig., 441 (1980)
85 P.J.T. Mellor, A.W. Livingstone, Electron. Lett., 16, 749, (1980)
86 R. Zuleeg, J.K. Notthoff, G.L. Troeger, EASCON tech. dig., 146, (1980)
87 R.L. van Tuyl, C.A. Liechti, R. Lee, E. Gowen, ISSCC tech. dig., 198, (1977)
88 C.A. Liechti, E. Gowen, J. Cohen, ISSCC tech. dig., 158, (1972)
89 A.D. Welbourn, IEE Colloquium on ICs above 1 GHz, (1985)
90 H.L. Grubin, IEEE Trans. Mic. Theory Tech., MTT-28, 442, (1980)
91 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)
92 R.R. Bockemuehl, IEEE Trans. Electron Dev., ED-10, 31, (1963)
93 E.H. Nicollian, J.R. Brews, MOS physics and technology, pub Wiley, New York, (1982)
94 Y. Kato, M. Dohsen, J. Kasahara, K. Taira, N. Watanabe, Electron. Lett., 17, 951 (1981)
95 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)
96 C. A. Mead, Proc. IEEE, 54,307, (1966)
97 E. H. Rhoderick, Metal semiconductor contacts, pub. Clarendon Press, Oxford, (1978)
98 R. Zuleeg, J.K. Notthoff, G.L. Troeger, Gallium Arsenide technology vol II
99 M. Abe, T. Mimura, K. Nishiuchi, A. Shibatomi, M. Kobayashi, GaAs IC Symp. Tech. dig., 150, (1983)
100 L. Esaki, R. Tsu, IBM Res report RC 2418, (1969)
101 C. Lee, D. Hou, S. Lee, D. Miller, R. Anderson, GaAs IC Symp. Tech. dig., 162, (1983)
102 M. Laviron, D. Delagebeaudeuf, J.F. Rochette, P.R. Jay, P. deLescluse, J. Cherrier, N.T. Linh, Physica, 129b, 376, (1985)
103 T. Mimura, S. Hiyamazu, K. Joshin, K. Hikosaka, Jap. J. Appl. Phys., 20, L317, (1981)
104 D. Delagebeaudeuf, M. Laviron, P. Delescluse, P.N. Tung, J. Chaplat, N.T. Linh, Electron. Lett., 18, 103, (1982)

105 R. A. Kiehl, M.D. Fever, R.H. Hender, J.C.M. Hwang, V.G. Keramidas, C.L. Allyn, R. Dingle, IEEE Electron Dev. Lett., EDL-4, (1983)
106 H. Morkoc, IEEE Electron Dev. Lett., EDL-2, 260, (1982)
107 C.P. Lee, S.J. Lee, B.M. Welch, IEEE Electron Dev. Lett., EDL-3, 97, (1983)
108 S. Makram-Ebeid, P. Minondo, GaAs IC Symp. Tech. dig., 142, (1983)
109 A.D. Welbourn, A.W. Livingstone, G.L. Blau, IEE Colloquium on High Speed Logic (1982)
110 A.D. Milne, MOS devices: design and manufacture, pub Edinburgh Univ. press, (1982)
111 N. Yokoyama, A. Shibatomi, S. Okhara, M. Fukuta, H. Ishikawa, GaAs and related compounds conf., pub Institute of Physics conference series 33b, (1976)
112 J. V. diLorenzo, Proc. 6th Cornell elec. eng. conference, 1, (1977)
113 ibid.
114 J.F. Gibbons, Proc. IEEE, 56, 1062, (1972)
115 C.P. Lee, Proc III-V Mat. conf., 324 (1982)
116 N.F. Mott, E.A. Davis, Electronic processes in non-crystalline materials, pub. Clarendon Press, Oxford, (1978)
117 C.P. Lee, S.J. Lee, B.M. Welch, IEEE Electron Dev. Lett., EDL-3, 97, (1983)
118 C. Kocot, C.A. Stolte, IEEE Trans. Mic. Theory Tech., MTT-30, 963, (1982)
119 Y. Kawakami, K. Tanaka, M. Tsunotani, H. Nakamura, Y. Sano, GaAs IC Symp. Tech. dig., 107, (1984)
120 E. H. Perea, F. Damaykavala, G. Nuzillat, C. Arnodo, IEEE J. Solid State Circ., SC-18, 365, (1983)
121 M.S. Birritella, W.C. Seelbach, H. Goronkin, IEEE Trans. Electron Dev., ED-29, 1134, (1982)
122 Personal communication with E. Strid
123 D. d'Avanzo, IEEE Trans. Mic. Theory Tech., MTT-30, 955, (1982)
124 M. Berth, M. Cathelin, G. Durand, IEDM Tech. dig., 201, (1977)
125 W.M. Paulson, M.S. Birrittella, T.H. Miers, K.L. McLaughlin, GaAs IC Symp. Tech. dig., 166, (1982)
126 C.P. Lee, R. Vahrenkamp, S.J. Lee, Y.D. Shen, B.M. Welch, GaAs IC Symp. Tech. dig., 169, (1982)
127 J.B. Clegg, G.B. Scott, J.P. Hollins, A. Mireau-Roussel, J. Appl. Phys., 52, 1110, (1981)
128 H.M. Cox, J.V. diLorenzo, GaAs and related compounds conf., pub Institute of Physics conference series 33b, 11, (1976)
129 C.P. Lee, B.M. Welch, IEEE Electron Dev. Lett., EDL-3, 200, (1982)
130 Y. Umemoto, S. Takahashi, N. Matsunga, M. Nakamura, K. Ueyanagi, GaAs IC Symp. Tech. dig., 173, (1982)
131 S. Makram-Ebeid, P. Minondo, GaAs IC Symp. Tech. dig., 142, (1983)
132 A.W. Livingstone and A.D. Welbourn, 14th Conference on Solid State Devices, Tokyo,1982. [also Jap. J. Apply Phys 22 Supplement 22-1, pp393-6, (1983)]
133 S. Nishine, N. Kito, K. Fujita, M. Sekinobu, O. Shikatani, S. Akai, GaAs IC Symp. Tech. dig., 58, (1982)
134 J.B. Mullin, R.J. Heritage, C.H. Holliday, B.W. Straughan, J. Cryst. growth, 34, 281, (1968)

135 L.R. Weisberg, J. Blanc, E.J. Stofko, J. Electrochem. Soc., 109, 642, (1962)

136 R.D. Fairman, F.J. Morin, J.R. Oliver, Inst. of Physics conf. series, 45, 134, (1979)

137 O. Mizumo, S. Kikuchi, Y. Seki, Jap. J. Appl. Phys., 10, 208, (1971)

138 H. Yamazaki, Y. Nanishi, S. Miyazawa, Y. Ishii, GaAs IC Symp. Tech. dig., 30, (1983)

139 P.M. Asbeck, J. Tandon, E. Babcock, B. Welch, C. Evans, V. Deline, IEEE Electron Dev. Lett., EDL-1, 35, (1980)

140 J.B. Clegg, G.B. Scott, J.P. Hollins, A. Mireau-Roussel, J. Appl. Phys., 52, 1110, (1981)

141 N. Matsunaga, S. Takahashi, N. Hashimoto, GaAs IC Symp. Tech. dig., 159, (1984)

142 J. V. diLorenzo, D.D. Khandelwal (eds), GaAs FET principles and technology, pub Artech, Dedham Mass., (1982)

143 D.E. Holmes, R.T. Chen, K.R. Elliot, C.G. Kirkpatrick, Appl. Phys. Lett., 40, 46, (1980)

144 J. Lagowski, H.C. Gatos, J.M. Parsey, K. Woda, M. Kamuriska, W. Walakeiwicz, Appl. Phys. Lett., 40, 342 (1982)

145 M.R. Brozel, J.B. Clegg, R.C. Newman, J. Phys. D, 11, 1331, (1978)

146 A.S. Jordan, R. Carcuso, A.R. van Neida, Bell Syst. Tech. J., 59, 593, (1980)

147 Y. Nanishi, S. Isida, S. Miyazawa, Jap. J. Appl. Phys., 22, L54, (1983)

148 I.R. Grant, IEE coll. on ICs above 1 GHz, tech dig 35/86, paper 1, (1986)

149 D. Rumsby, R.M. Ware, B. Smith, M. Tyjberg, M.R. Brozel, E.J. Foulkes, GaAs IC Symp. Tech. dig., 34, (1983)

150 H.V. Winston, A.T. Hunter, H.M. Olsen, R.P. Bryan, R.E. Lee, Proc. 3rd III-V Semi-insulating materials conf., (1984)

151 K. Tada, S. Murai, S. Akai, T. Suzuki, GaAs IC Symp. Tech. dig., 49, (1984)

152 U.K. Mishra, A.S. Brown, L.M. Jelloian, L.H. Hackett, M.J. Delaney, IEEE Electron Dev. Lett., EDL-9, 41, (1988)

153 K. Terashima, J. Nishio, S. Washizuka, M. Watanabe, J. Cryst. growth, 84, 247, (1987)

154 R.S. Pengelley, Microwave FET theory, design and applications, pub Wiley, Chichester, (1982)

155 ibid.

156 R.S. Butlin, D. Parier, I. Crossley, J. Turner, GaAs and related compounds conf., pub Institute of Physics conference series 33b, 237, (1976)

157 G. Dearnaley, J.H. Freeman, R.S. Nelson, J. Stephen, Ion implantation, pub North Holland, Amsterdam, (1978)

158 P.N. Favennec, J. Appl. Phys., 47, 2532, (1976)

159 J.A. Higgins, B.M. Welch, F.H. Eisen, G.D. Robinson, Electron. Lett., 12, 17, (1976)

160 T. Inada, H. Miwa, S. Kato, E. Kobayashi, T. Hara, M. Mihara, J. Appl. Phys., 49, 4571, (1978)

161 F.H. Eisen, B.M. Welch, H. Muler, K. Gamo, T. Inada, J.W. Mayer, Solid State Electron., 20,219, (1977)

162 T.J. Harris, B.J. Sealy, R.K. Surridge, Electron. Lett., 12, 664, (1976)

163 R.C. Eden, B.M. Welch, R. Zucca, IEEE J. Solid State Circ., SC-13, 419, (1978)

164 J.P. Donnelly, GaAs and related compounds conf., pub Institute of Physics conference series 33b, (1976)

165 ibid.

166 J.L. Tandon, M.A. Nicolet, F.H. Eisen, Appl. Phys. Lett., 34, 165, (1979)

167 R. Zuleeg, J.K. Notthoff, G.L. Troeger, EASCON tech. dig., 146, (1980)

168 Y. Umemoto, S. Takahashi, N. Matsunga, M. Nakamura, K. Ueyanagi, GaAs IC Symp. Tech. dig., 173, (1982)

169 Y. Kato, K. Gonoi, K. Taira, J. Kasahara, M. Dohsen, M. Arai, N. Watanabe, GaAs IC Symp. Tech. dig., 187, (1982)

170 R.G. Huisperger, R.G. Wilson, D.M. Janha, J. Appl. Phys., 43, 1318, (1972)

171 S. Nozama, Y. Kawasaki, Jap. J. Appl. Phys., 17, 1845, (1978)

172 M. Sittig, Handbook of toxic and hazardous chemicals, pub Noyes, New Jersey, (1981)

173 R.K. Surridge, B.J. Sealy, A.D.E. d'Cruz, K.G. Stephens, GaAs and related compounds conf., pub Institute of Physics conference series 33b, 16, (1976)

174 B.J. Sealy, Microelectron. J., 13, 21, (1982)

175 R.L. Mozzi, W. Fabian, F.J. Piekarski, Appl. Phys. Lett., 35, 337, (1979)

176 B.J. Sealy, S.S. Kular, K.G. Stephens, R. Croft, A. Palmer, Electron. Lett., , 14, 720, (1978)

177 D. Kahng (ed) Silicon ICs part C, pub Academic press, Orlando, (1985)

178 P.D. Scovell, ESSDERC tech dig., 280, (1985)

179 C.T. Foxon, J.A. Harvey, B.A. Joyce, J. Phys. Chem. Sol., 34, 1693, (1973)

180 unpublished results obtained at BTRL

181 M. Feng, V. Eu, H. Kanber, W.B. Henderson, J. Electron. Mat., 10, 973, (1981)

182 J. V. diLorenzo, D.D. Khandelwal (eds), GaAs FET principles and technology, pub Artech, Dedham Mass., (1982)

183 A.G. Foyt, J.P. Donnelly, W.T. Lindley, Appl. Phys. Lett., 14, 372, (1969)

184 J.P. Donnelly, W.T. Lindley, C.E. Hurwitz, Appl. Phys. Lett., 27, 41, (1975)

185 H. Nishii, S.O. Kamura, T. Inada, H. Hashimoto, GaAs and related compounds conf., pub Institute of Physics conference series 63, 365, (1982)

186 D.E. Holmes, R.T. Chen, K.R. Elliot, C.G. Kirkpatrick, Appl. Phys. Lett., 40, 46, (1980)

187 B.M. Welch, Y.D. Shen, R. Zucca, R.C. Eden, S.I. Long, IEEE Trans. Electron Dev., ED-27, 1116, (1980)

188 W.M. Paulson, M.S. Birrittella, T.H. Miers, K.L. McLaughlin, GaAs IC Symp. Tech. dig., 166, (1982)

189 M. Berth, M. Cathelin, G. Durand, IEDM Tech. dig., 201, (1977)

190 D. d'Avanzo, IEEE Trans. Mic. Theory Tech., MTT-30, 955, (1982)

191 C. Kocot, C.A. Stolte, IEEE Trans. Mic. Theory Tech., MTT-30, 963, (1982)

192 W. Shockley, Bell Syst. Tech. J., 28, 435, (1949)

193 E. H. Rhoderick, Metal semiconductor contacts, pub. Clarendon Press, Oxford, (1978) .

194 M. Rocchi, Tech dig. ESSDERC, (1984)
195 A.M. Cowley, S.M. Sze, J. Appl. Phys., 36, 3212, (1965)
196 ibid.
197 D.J. Elliott, Integrated Circuit fabrication technology, pub McGraw Hill, New York, (1982)
198 ibid.
199 D. A. Allan, P.J. O'sullivan, IEE Coll. on ICs above 1 GHz tech dig 35/86, paper 9, (1986)
200 M.J. Howes, D.V. Morgan (eds), Reliability and degradation, pub Wiley, Chichester, (1981)
201 ibid.
202 D. A. Allan, P.J. O'sullivan, IEE Coll. on ICs above 1 GHz tech dig 35/86, paper 9, (1986)
203 E. Kohn, IEDM Tech. dig., 469, (1979)
204 N. Yokoyama, T. Ohnishi, H. Onedera, T. Shinoki, A. Shibatomi, H. Ishikawa, ISSCC tech. dig., 44, (1983)
205 R.A. Sadler, L.F. Eastman, IEEE Electron Dev. Lett., EDL-4, 215, (1983)
206 N. Yokoyama, T. Ohnishi, K. Odari, H. Onedera, M. Abe, IEDM tech. dig., 80, (1981)
207 M.P. Lepselter, J.M. Andrews, Ohmic contacts to semiconductors, pub Electrochem. Soc., New York, (1969)
208 R. A. Smith, Semiconductors, pub Cambridge University press, Cambridge, (1978)
209 M.P. Lepselter, J.M. Andrews, Ohmic contacts to semiconductors, pub Electrochem. Soc., New York, (1969)
210 M.J. Howes, D.V. Morgan (eds), GaAs Materials Devices and Circuits, pub Wiley, Chichester, (1985)
211 ibid.
212 K. Chino, Y. Wada, Jap. J. Appl. Phys., 16, 1823, (1977)
213 C.P. Lee, B.M. Welch, W.P. Fleming, Electron. Lett., 17, 407, (1981)
214 D. A. Allan, P.J. O'sullivan, IEE Coll. on ICs above 1 GHz tech dig 35/86, paper 9, (1986)
215 M. Ino, M. Hirayama, K. Ohwada, K. Kurumada, GaAs IC Symp. Tech. dig., 2, (1982)
216 D.J. Hamilton, W.G. Howard, Basic integrated circuit engineering, pub McGraw Hill, New York, (1975)
217 K.C. Saraswat, F. Mahammadi, IEEE Trans. Electron Dev., ED-29, 645, (1982)
218 ibid.
219 Wolf, Semiconductor data book, pub Wiley, New York
220 Bergeron, Kent and Morrett, Int. rel. phys. symp. tech. dig., 229, (1984)
221 Manufacturers data sheets.
222 Bergeron, Kent and Morrett, Int. rel. phys. symp. tech. dig., 229, (1984)
223 R.A. Pucel, IEEE Trans. Mic. Theory Tech., MTT-16, 342, (1968)
224 R.S. Pengelly, D.C. Rickard, 7th European Mic. conf. tech. dig., 460, (1977)
225 P.R. Shepherd, IEEE Trans. Mic. Theory Tech., MTT-34, (1986)
226 J.R. Anderson, M. Omori, H.F. Cooke, IEDM Tech. dig., 133, (1978)
227 G. McCormack, A. Rode, E. Strid, GaAs IC Symp. Tech. dig., 25, (1982)

228 A. Rode, T. Flagel, G. LaRue, GaAs IC Symp. Tech. dig., 178, (1983)
229 R.C. Eden, B.M. Welch, IEEE Trans. Electron Dev., ED-24, 1209, (1977)
230 T. Mizutani, N. Kato, M. Ida, M. Ohmori, IEEE Trans. Mic. Theory Tech., MTT-28, 479, (1980)
231 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)
232 S.M. Sze, Physics of semiconductor devices, pub Wiley, New York, 1981
233 G. Nuzillat, F. Damay-Kavala, G. Bert. C. Arnodo, IEE Proc., 127I, 287, (1980)
234 C.A. Liechti, IEEE Trans. Mic. Theory Tech., MTT-24, 279, (1976)
235 H. Hasegawa, T. Sawada, T. Kitagawa, GaAs IC Symp. Tech. dig., 145, (1983)
236 R.R. Bockemuehl, IEEE Trans. Electron Dev., ED-10, 31, (1963)
237 S.H. Wemple, M.L. Steinberger, W.O. Schlosser, Electron. Lett., 16, 459, (1980)
238 R. Zuleeg, J.K. Notthoff, A.F. Behle, Int mic symp dig., MTT-S, 512, (1979)
239 C.H. Vogelsang, J.L. Hogin, J.K. Notthoff, GaAs IC Symp. Tech. dig., 149, (1983)
240 H. Fukui, IEEE Trans. Mic. Theory Tech., MTT-27. 643, (1979)
241 P. Frey, B. Gabillard, M. Rocchi, GaAs IC Symp. Tech. dig., 62, (1983)
242 R.L. van Tuyl, C.A. Liechti, IEEE J. Solid State Circ., SC-9, 269, (1974)
243 Integrated device and connection technology, ed Bell telephone labs.,pub Prentice Hall, New Jersey, (1971)
244 B.M. Welch, Y.D. Shen, R. Zucca, R.C. Eden, S.I. Long, IEEE Trans. Electron Dev., ED-27, 1116, (1980)
245 R.L. van Tuyl, C.A. Liechti, R. Lee, E. Gowen, ISSCC tech. dig., 198, (1977)
246 A. Barna, C.A. Liechti, IEEE J. Solid State Circ., SC-14, 708, (1979)
247 R. Yamamoto, A. Higashisaki, S. Asai, T. Tsugi, Y. Takayama, S. Yano, ISSCC tech. dig., 403 (1983)
248 S.H. Bennett, S. Hollock, P.J. Ward, ESSCIRC tech. dig., 34, (1975)
249 R.C. Eden, B.M. Welch, R. Zucca, S.I. Long, IEEE J. Solid State Circ., SC-14, 221, (1979)
250 Y. Yamauchi, K. Nagagt, O. Nakajima< H.Ito, T. Nittono, T. Ishibashi, Electron. Lett., 23, 881, (1987)
251 T. Sakai, S. Konaka, Y. Yamamotot, M. Suzuki, IEDM Tech. dig., 18, (1985)
252 N.G. Einspruch, R.K. Watts (eds), VLSI electronics & microstructure science, 16, (1987)
253 S.K. Wiedmann, IEEE J. Solid State Circ., SC-19, 282, (1984)
254 A.W. Wieder, Siemens Forsch. - u Entwickl-Ber. Bd, 13, 246, (1984)
255 G.L. Patton, J.C. Brauman, J.D. Plummer, IEEE Trans. Electron Dev., ED-33, 1754, (1986)
256 H.D. Barber, Can. J. Phys., 63, 683, (1985)
257 W. Kellner, Compound semiconductor devicces and ICs workshop, Germisch, May 1987

258 R.C. Eden, B.M. Welch, IEEE Trans. Electron Dev., ED-24, 1209, (1977)

259 B.M. Welch, Y.D. Shen, R. Zucca, R.C. Eden, S.I. Long, IEEE Trans. Electron Dev., ED-27, 1116, (1980)

260 R.L. van Tuyl, C.A. Liechti, R. Lee, E. Gowen, ISSCC tech. dig., 198, (1977)

261 R. Zucca, B.M. Welch, R.C. Eden, S.I. Long, IEEE Trans. Electron Dev., ED-27, 1109, (1980)

262 G. Lee, T. Steeves, R. Nelson, T. Vu, B. Hanzal, P. Roberts, M. Helix, S. Jamison, S. Hanka, J. Brown, GaAs IC Symp. Tech. dig., 174, (1983)

263 M.J. Helix,S.A. Jamison, S.A. Hanka, R.P. Vidano, P. Ng, C. Chao, GaAs IC Symp. Tech. dig., 108, 1982

264 L.J. Herbst, Monolithic IC techniques and capabilities, pub. Clarendon Press, Oxford, (1985)

265 A. Barna, C.A. Liechti, IEEE J. Solid State Circ., SC-14, 708, (1979)

266 Private discussions with P. Greiling

267 E. H. Perea, F. Karda, G. Nuzillat, C. Arnodo, ESSCIRC tech. dig., 126, (1982)

268 P.J.T. Mellor, A.W. Livingstone, Electron. Lett., 16, 749, (1980)

269 A.W. Livingstone, A.D. Welbourn, G.L. Blau, IEEE Electron Dev. Lett., EDL-3, 284, (1982)

270 A.D. Welbourn, A.W. Livingstone, G.L. Blau, IEEE J. Solid State Circ., SC-18, 359, (1983)

271 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)

272 J.M. Calvert, M.A.H. McCausland, Electronics, pub Wiley, Chichester, (1978)

273 R.C. Eden, GaAs IC Symp. Tech. dig., 11, (1984)

274 M.R. Namordi, W.A. White, GaAs IC Symp. Tech. dig., 21, (1982)

275 H.P. Singh, R.A. Sadler, A.E. Geissberger, D.G. Fisher, J. Irvine, G. E. Gorder, GaAs IC Symp. Tech. dig., 11, (1986)

276 J. K. Notthoff, C.H. Vogelsang, GaAs IC Symp. Tech. dig., (1979)

277 A.D. Welbourn, J. Mavor, IEE Colloquium on a Comparison of GaAs and Silicon, (1986).

278 S.M. Sze, Physics of semiconductor devices, pub Wiley, New York, (1981)

279 R.R. Bockemuehl, IEEE Trans. Electron Dev., ED-10, 31, (1963)

280 S.M. Sze, G. Gibbons, Appl. Phys. Lett., 8, 111, (1966)

281 J.R. Brews, Solid State Electron., 21, 345, (1978)

282 D. Delagebeaudeuf, N.T. Linh, IEEE Trans. Electron Dev., ED-29, 995, (1982)

283 R.A. Warriner, IEE J. Solid State Electron. Dev., 1, 105, (1977)

284 A.D. Welbourn, IEE Proc., 129I, 157, (1982)

285 T. Takada, Y. Yokoyama, M. Ida, T. Sudo, IEEE Trans. Mic. Theory Tech., MTT-30, 719, (1982)

286 S.S. Haykin, R. Barrett, Transistor circuits in electronics, pub Butterworth, London, (1971)

287 K.C. Gupta, R. Garg, R. Chadha, Computer-aided design of microwave circuits, pub Artech, Mass., (1981)

288 M. Idda, T. Takada, T. Sudo, IEEE Trans. Mic. Theory Tech., MTT-32, 5, (1984)

289 Y. Mitsui, K. Nishitani, O. Ishihara, M. Nakatani, GaAs IC Symp. Tech. dig., 115, (1982)
290 Y. Ishii, M. Ino, M. Idda, M. Hirayama, M. Ohmori, GaAs IC Symp. Tech. dig., 121, (1984)
291 A.W. Livingstone, A.D. Welbourn, G.L. Blau, IEEE Electron Dev. Lett., EDL-3, 284, (1982)
292 R.A. Sadler, L.F. Eastman, DRC. tech. dig., (1983)
293 P.M. Asbeck, C.P. Lee, M.F. Chang, IEEE Trans. Electron Dev., ED-31, 1377, (1984)
294 H. Nakamura, Y. Sano, T. Nonaka, T. Ishida, K. Kaminishi, GaAs IC Symp. Tech. dig., 134, (1983)
295 M. Ohmori, ESSDERC tech dig, (1985)
296 W.E. Spicer, P.W. Chye, P.R. Skeath, C.Y. Su, I. Landau, J. Vac. Sci. Technol., 16, 1422, (1979)
297 I. Ohta, Y. Oishi, M. Nishiuma, M. Hagio, M. Kazumura, G. Kano, I. Teramoto, GaAs IC Symp. Tech. dig., 55, (1986)
298 L.J. Herbst, Monolithic IC techniques and capabilities, pub. Clarendon Press, Oxford, (1985)
299 T.Terada, Y. Kitaura, T. Mizoguchi, M. Mochizuki, N. Toyada, A.Hojo, GaAs IC Symp. Tech. dig., 138, (1983)
300 R.E. Lee, H.M. Levy, D.S. Matthews, GaAs IC Symp. Tech. dig., 177, (1982)
301 T. Furutsuka, T. Tsuji, F. Hasegawa, IEEE Trans. Electron Dev., ED-25, 563, (1978)
302 R.E. Lee, H.M. Levy, D.S. Matthews, GaAs IC Symp. Tech. dig., 177, (1982)
303 T. Ohnishi, T. Onodera, N. Yokoyama, H. Nishi, IEEE Electron Dev. Lett., EDL-6, 172, (1985)
304 T. Onodera, T. Ohnishi, N. Yokoyama, H. Nishi, IEEE Trans. Electron Dev., ED-32, 2314, (1985)
305 H. Nakamura, Y. Sano, T. Nonaka, T. Ishida, K. Kaminishi, GaAs IC Symp. Tech. dig., 134, (1983)
306 Y. Yamasaki, K. Asai, K. Kurumada, IEEE Trans. Electron Dev., ED-29, 1772, (1982)
307 M. Ino, M. Hirayama, K. Ohwada, K. Kurumada, GaAs IC Symp. Tech. dig., 2, (1982)
308 K. Asai, K. Kurumada, M. Hirayama, M. Ohmori, ISSCC tech. dig., 46, (1983)
309 R. Zuleeg, J.K. Notthoff, K. Lehovec, IEEE Trans. Electron Dev., ED-25, 628, (1978)
310 Personal comunication with R. Zuleeg
311 R. Zuleeg, J.K. Notthoff, K. Lehovec, IEEE Trans. Electron Dev., ED-25, 628, (1978)
312 P.M. Solomon, Proc. IEEE, 70, 489, (1982)
313 G. Nuzillat, G. Bert, T.P. Ngu, M. Gloanec, IEEE Trans. Electron Dev., ED-27, 1102, (1980)
314 G. Nuzillat, F. Damay-Kavala, G. Bert, C. Arnodo, IEE Proc., 127I, 287, (1980)
315 A. Barna, High speed pulse and digital techniques, pub Wiley, New York, (1980)
316 S. Katsu, S. Nambu, A. Shimano, G. Kano, IEEE Electron Dev. Lett., EDL-3, 197, (1982)
317 Personal deduction from simulation work.
318 M. Idda, T. Takada, T. Sudo, IEEE Trans. Mic. Theory Tech., MTT-32, 5, (1984)

319 T. Takada, N. Kato, M. Ida, IEEE Electron Dev. Lett., EDL-7, 47, (1986)
320 K.M. Chan, S.Y. Oh, D. Chin, J.L. Moll, Computer aided design and VLSI development, pub Kluwer Acadamic, Massachusets, (1986)
321 R.W. Knepper, S.P. Gaur, T.Y. Cheng, G.R. Srinivassan, IBM J. Res. Dev., 29, 218, (1985)
322 S.M. Sze, Physics of semiconductor devices, pub Wiley, New York, (1981)
323 J. Vlach, K. Singhal, Computer methods for circuit analysis and design, pub. van Nostrand Rheinhold, New York, (1983)
324 E. H. Rhoderick, Metal semiconductor contacts, pub. Clarendon Press, Oxford, (1978)
325 ibid.
326 G. Dearnaley, J.H. Freeman, R.S. Nelson, J. Stephen, Ion implantation, pub North Holland, Amsterdam, (1978)
327 K.C. Gupta, R. Garg, R. Chadha, Computer aided design of microwave circuits, pub Artech, New York, (1981)
328 MECL 10k data book, Motorola
329 P.J.T. Mellor, A.W. Livingstone, Electron. Lett., 16, 749, (1980)
330 M.J. Howes, D.V. Morgan (eds), GaAs materials, devices and circuits, pub Wiley, Chichester, (1985)
331 ibid.
332 J.M. Calvert, M.A.H. McCausland, Electronics, pub Wiley, Chichester, (1978)
333 D.G. Ong, Modern MOS technology: processes devices and design, pub McGraw Hill, New York, (1984)
334 R.R. Bockemuehl, IEEE Trans. Electron Dev., ED-10, 31, (1963)
335 J. Watson, Analogue and switching circuit design, pub Adam Hilger, Bristol, (1984)
336 P.M. Chirlian, Analysis and design of integrated electronic circuits vol 2: digital electronics, pub Harper and Row, London, (1982)
337 M. Rocchi, Tech dig. ESSDERC, (1984)
338 D.J. Walter, Integrated circuit systems, pub Butterworth, London, (1971)
339 H. Fujiwara, Logic testing and design for testability, pub MIT press, Boston, (1985)
340 A.W. Livingstone and A.D. Welbourn, 14th Conference on Solid State Devices, Tokyo,1982. [also Jap. J. Apply Phys 22 Supplement 22-1, pp393-6, (1983)]
341 S.M. Sze, G. Gibbons, Appl. Phys. Lett., 8, 111, (1966)
342 R.P. Merrett, IEE Colloquium on a Comparison of GaAs and Silicon, (1986).
343 A. Cardon, L.J.L. Franson, Dynamic semiconductor RAM structures, pub Pergamon, Oxford, (1984)
344 J.J. Sparkes, Transistor switching and sequential circuits, pub Pergamon, Oxford, (1969)
345 S. Okugawa, N. Inoue, Electron. Lett., 23, 909, (1987)
346 ibid.
347 M. Foster, H.T. Kung, IEEE Computer, 13, 26, (1980)
348 P.M. Chirlian, Analysis and design of integrated electronic circuits vol 2: digital electronics, pub Harper and Row, London, (1982)
349 A. Cardon, L.J.L. Franson, Dynamic semiconductor RAM structures, pub Pergamon, Oxford, (1984)

350 R.L. van Tuyl, C.A. Liechti, R. Lee, E. Gowen, ISSCC tech. dig., 198, (1977)
351 R.C. Eden, B.M. Welch, IEEE Trans. Electron Dev., ED-24, 1209, (1977)
352 M. Gloanec, J. Jarry, G. Nuzillat, Electron. Lett., 17, 763, (1981)
353 A.D. Welbourn, A.W. Livingstone, G.L. Blau, IEEE J. Solid State Circuits, SC-18, 359, (1983)
354 H. Ahmed, P.J. Spreadbury, Analogue and digital electronics for engineers, pub Cambridge University Press, Cambridge, (1984)
355 S. Katsu, S. Nambu, A. Shimano, G. Kano, IEEE Electron Dev. Lett., EDL-3, 197, (1982)
356 Personal observation from detailed design studies.
357 E. H. Rhoderick, Metal semiconductor contacts, pub. Clarendon Press, Oxford, (1978)
358 J.M. Ziman, Principles of the theory of solids, pub Cambridge University Press, Cambridge, (1972)
359 ibid.
360 S.M. Sze, Physics of semiconductor devices, pub Wiley, New York, 1981
361 R.A. Smith, Semiconductors, pub Cambridge University Press, Cambridge, (1978)
362 H.K. Henisch, Rectifying semiconductro contacts, pub. Clarendon Press, Oxford, (1957)
363 A.M. Cowley, S.M. Sze, J. Appl. Phys., 36, 3212, (1965)
364 E. H. Rhoderick, Metal semiconductor contacts, pub. Clarendon Press, Oxford, (1978)
365 S.M. Sze, Physics of semiconductor devices, pub Wiley, New York, 1981
366 S.H. Wemple, M.L. Steinberger, W.O. Schlosser, Electron. Lett., 16, 459, (1980)
367 H.A. Bethe, MIT Radiation Lab report, 43-12 (1942)
368 W. Schottky, Naturwissenschaften, 26, 843, (1938)
369 C.R. Crowell, S.M. Sze, Solid State Electron., 9, 1035, (1965)
370 C. Y. Chang, S.M. Sze, Solid State Electron., 13, 727, (1970)
371 C.R. Crowell, S.M. Sze, J. Appl. Phys., 37, 2685, (1966)
372 W. Shockley, Electrons and holes in semiconductors, pub van Nostrand, Princeton, (1950)
373 J. M. Andrews and M.P. Lepselter, Solid State Electron., 13, 1011, (1970)
374 S.M. Sze, Physics of semiconductor devices, pub Wiley, New York, 1981one sided abrupt p-n junction capacitance
375 C.G.B. Garrett, W.H. Brattain, Phys. Rev., 99, 376, (1955)
376 A.G. Milnes, Semiconductor devices and integrated electronics, pub. van Nostrand Rheinhold, Princeton, (1980)
377 J. Vlach, K. Singhal, Computer methods for circuit analysis and design, pub. van Nostrand Rheinhold, New York, (1983)
378 R. W. Knepper, S.P. Gaur, T.Y. Cheng, G.R. Srinivassan, IBM J. Res. dev., 29, 218, (1985)
379 P.A. Witting, Computer aided eng. J. 2, 186, (1985)
380 D.B. Estreich, IEEE Trans. Computer Aided design, CAD-2, 106, (1983)
381 H.K. Gummel, Solid State Electron., 10, 209, (1967)
382 R.M. Burger and R.P. Donovan (eds), Fundamentals of silicon integrated device technology vol. 2, pub Prentice Hall, New Jersey, (1968)

383 D.B. Estreich, IEEE Trans. Computer Aided design, CAD-2, 106, (1983)
384 ibid.
385 see Hewlett Packard HP4142 data sheet
386 T. Takada, Y. Yokoyama, M. Ida, T. Sudo, IEEE Trans. Mic. Theory Tech., MTT-30, 719, (1982)
387 J.A. Copeland, IEEE Trans. Electron Dev., ED-17, 404, (1970)
388 J.D. Jackson, Classical electrodynamics, pub Wiley, New York, (1975)
389 J.A. Copeland, IEEE Trans. Electron Dev., ED-17, 404, (1970)
390 R.W. Knepper, IEEE Bipolar Circuits and tech. Meeting tech. dig., 1, (1987)
391 I.E. Getreu, Modelling the bipolar transistor, pub Elsevier, New York, (1978)
392 R.M. Burger and R.P. Donovan (eds), Fundamentals of silicon integrated device technology vol. 2, pub Prentice Hall, New Jersey, (1968)
393 E.H. Nicollian, J.R. Brews, MOS physics and technology, pub Wiley, New York, (1982)
394 Morse and Feshbach, Methods of theoretcial physics, pub McGraw Hill, New York, (1953)
395 R.R. Bockemuehl, IEEE Trans. Electron Dev., ED-10, 31, (1963)
396 S.M. Sze, Semiconductor devices: physics and technology, pub Wiley, New York, (1985)
397 the "linear" region is strictly defined as the "forward active" region.
398 W. Gosling, W.G. Townsend, J. Watson, Field effect electronics, pub. Butterworths, London, (1971)
399 C. A. Liechti, IEEE Trans. Mic. Theory Tech., MTT-24, 279, (1976)
400 R.A. Warriner, IEE J. Solid state electron dev., 1, 105, (1977)
401 J.M. Hammersley, D.C. Handscomb, Monte-Carlo methods, pub. Methuen, London, (1964)
402 M.S. Shur, L.F. Eastman, IEEE Trans. Electron Dev., ED-25, 606, (1978)
403 T. Wada, J. Frey, IEEE Trans. Electron Dev., ED-23, 519, (1976)
404 R.R. Bockemuehl, IEEE Trans. Electron Dev., ED-10, 31, (1963)
405 K. Lehovec, R. Zuleeg, Solid State Electron., 13, 1415, (1970)
406 ibid.
407 R.A. Pucel, H.A. Haus, H. Statz, Advances in Electronics and electron physics, 38, pub Academic press, New York, (1975)
408 C. A. Liechti, IEEE Trans. Mic. Theory Tech., MTT-24, 279, (1976)
409 S.C. Thorp, IEE Coll. on GaAs ICs, (1983)
410 R.D. Middlebrook, I. Richer, Solid State Electron., 6, 542, (1963)
411 R.E. Williams, D.W. Shaw, Electron. Lett., 13, 408, (1977)
412 K. Yamaguchi, H. Kodera, IEEE Trans. Electron Dev., ED-24, 1061, (1977)
413 Hewlett Packard application note 154, (1972)
414 R.A. Pucel, H.A. Haus, H. Statz, Advances in Electronics and electron physics, 38, pub Academic press, New York, (1975)
415 Personal observation confirmed in discussion with Paul Greiling.
416 H. Statz, P. Newman, I.R. Smith, R.A. Pucel, H.A. Haus, IEEE Trans. Electron Dev., ED-34, 160, (1987)

417 M. Rocchi, ESSDERC Tech. dig., (1984)
418 K.C. Saraswat, F. Mahammadi, IEEE Trans. Electron Dev., ED-29, 645, (1982)
419 ibid.
420 L. . Herbst, Monolithic IC techniques and capabilities, pub. Clarendon Press, Oxford, (1985)
421 F.M. Klaassen, Inst. Phys. conf. series, 82, 113, (1986)
422 P.H. Saul, IEE coll. on high speed logic, tech dig, 70/1982, (1982)
423 M.V. Schneider, Bell Syst. Tech. J., 48, 1421, (1969)
424 H.A. Wheeler, IEEE Trans. Mic. Theory Tech., MTT-25, 631, (1977)
425 K.C. Gupta, R.Garg, R. Chadha, Computer aided design of microwave circuits, pub Artech House, (1981)
426 J.A. Weiss, T.G. Bryant, Electron Lett.,6, 462, (1970)
427 J. Vlach, K. Singhal, Computer methods for circuit analysis and design, pub van Nostrand Rheinhold, New York, (1983)
428 R. Garg, Int. J. Electron., 47, 587, (1979)
429 R.L. van Tuyl, C.A. Liechti, R.E. Lee, E. Gowen, IEEE J. Solid State Circ., SC-12, 485, (1977)
430 H. A. Wheeler, IEEE Trans. Mic. Theory Tech., MTT-30, 2050, (1982)
431 I.J. Bahl, R. Garg, Proc. IEEE, 64, 1611, (1977)
432 Properties of GaAs, pub Inspec, (1987)
433 R.D. Martin-Royle, Br. Telecom Eng. J., 1, 190, (1983)
434 M.G. Rowlands, Br. Telecom Eng. J., 2, 6, (1984)
435 R.D. Martin-Royle, Br. Telecom Eng. J., 1, 190, (1983)
436 CCITT (consultative committee for international telephone and telegraphs) is an organisation for developing international communications standards.
437 H.R. Schemmel, Fernmelde Praxis, 63, 343, (1986)
438 R.M. Brooks, G.P. Coombs, R. Rowley, V.W. Stanbrook, A. Stevenson, S. Whitt, Br. Telecom Tech. J., 3, 46, (1985)
439 R.M. Brooks, A. Jessop, Int J Electron, 55, 81, (1983)
440 MECL 10k data book, pub Motorola
441 ibid.
442 P.H. Saul, IEE Coll on ICs above 1 GHz, (1985)
443 D.J. Walter, Integrated circuit systems, pub Iliffe, London, (1971)
444 A. Barna, High speed pulse and digital techniques, pub Wiley, New York, (1980)
445 A.J. Sharland, R.M. Brooks, G.P. Coombs, S. Whitt, Br. Telecom Tech. J., 1, 31, (1983)
446 D. Wilson, N. Frick, J. Kwiat, S. Lo, J. Churchill, J. Barrera, GaAs IC Symp. Tech. dig., 13, (1982)
447 ibid.
448 E.H. Nicollian, J.R. Brews, MOS Physics and technology, pub Wiley, New York, (1982)
449 Personal discussion with S.J. Wright
450 A.D. Welbourn, R.P. Merrett, G.L. Blau, IEE coll. on design and fab of GaAs ICs, (1984)
451 SMC is a widely available standard for subminiature microwave connectors.
452 SMA is a widely available standard for miniature microwave connectors.

453 R.P. Merrett, IEE Coll. on ICs above 1 GHz, tech dig 26/85, paper 4 (1985)
454 commercially available ECL components from Motorola, Fairchild Plessey etc. operated at around 250-300 MHz at this time.
455 R.P. Merrett, S.C. Thorp, P.J. Smith, IEE Coll. on ICs above 1 GHz tech dig 36/86, paper 16, (1986)
456 K.M. Charn, S.Y. Oh, D. Chin, J.L. Moll, Computer aided design and VLSI device development, pub Kluwer Academic, Mass. (1986)
457 GaAs component data sheets from Harris semiconductors.
458 R. Soares, J. Graffeuil, J. Obregon (eds), Applications of GaAs MESFETs, pub Artech, Dedham, Mass. (1983)
459 C. deProst, ESSDERC tech. dig., 23, (1985)
460 C. Liechti, G. Baldwin, E. Gowen, R. Joly, M. Naziro, A. Podell, IEEE Trans. Mic. Theory Tech., MTT-30, 998, (1982)
461 M.R. Namordi, W.A. White, GaAs IC Symp. Tech. dig., 21, (1982)
462 Personal discussion with P. Greiling.
463 component data sheets from Gigabit Logic Inc., California.

Personal work published during this research.

The following papers have been presented or published in connection with this work. The journal publications marked with an asterisk are included in this volume.

* A.W. Livingstone and A.D. Welbourn, "Design Considerations of Coupling Capacitors in GaAs Integrated Circuits", 14th Conference on Solid State Devices, Tokyo,1982. and Jap. J. Apply Phys 22 Supplement 22-1, pp393-6, (1983).

A.D. Welbourn, A.W. Livingstone, G.L. Blau, "Back Bias Effect in GaAs Logic Circuits", IEE Colloquium on High Speed Logic (1982)

* A.D. Welbourn, A.W. Livingstone, G.L. Blau, "A High Speed 8-bit Multiplexer Using Capacitor Coupled Logic", ESSCIRC, Brussels (1982) and IEEE J. Solid State Circuits SC-18 pp359-64 (1983).

* A.W. Livingstone, A.D. Welbourn, G.L. Blau, "Manufacturing Tolerance of Capacitor Coupled Logic", IEEE Electron Dev. Lett. EDL-3 pp284-5 (1982).

* A.D. Welbourn, "Gigabit Logic - A Review", IEE Proc 129 I pp157-72 (1982).

A.D. Welbourn, G.L. Blau, "The Determination of Criteria for the Design of GaAs Digital ICs", 1st European GaAs IC Workshop, Nice (1984)

A.D. Welbourn, R.P. Merrett,G.L. Blau, "Design, Simulation and Testing of a 2.4 GBit/s 8-Channel Multiplexer", IEE Coll on Design and Fab of GaAs ICs (1984)

A.D. Welbourn, "A Physical Comparison of GaAs and Silicon for High Speed Digital Integrated Circuits", IEE Colloquium on ICs above 1 GHz, (1985)

A.D. Welbourn, chapter entitled "Material Properties of GaAs" to be published in "Subnanosecond Digital Electronics", edited by L.J. Herbst, pub Prentice Hall.

A.D.Welbourn, lectures on "Material Properties of GaAs" ,"Device and Circuit Models for GaAs ICs" and "HEMT device Physics", [illegible] ITG Advanced Course on Subnanosecond Digital Electr[illegible]ics, Teesside Polytechnic, (1984).

A.D. Welbourn, T.J. Chandler, P.E. Holmes, "A N[illegible]vel Device Structure for High Speed Digital ICs", ESSDERC, Aachen 1985)

T.J. Chandler, A.D. Welbourn, P.E. Holmes, "Design and Assessment of Base Doping for High Speed Bipolar Proce[illegible]s", IEE Colloquium on ICs above 1 GHz, (1986).

A.D. Welbourn, J. Mavor, "A Comparati[illegible]e Overview of the Development of GaAs and Silicon for High Speed Digi[illegible]al ICs", IEE Colloquium on a Comparison of GaAs and Silicon, (1986).

A.D. Welbourn, T.J. Chandler, P.E. Holmes, "M[illegible]delling of the base-emitter breakdown voltage of a bipolar transiste[illegible]", ESSDERC 1986, Cambridge, September 1986.

A.D. Welbourn, "Process development using BIPOLE", Alv[illegible]y Process and device modelling meeting, Edinburgh, October 1986.

C. Smith, A.D. Welbourn, "Prospects for a heterostructure bi[illegible]olar transistor using a Silicon Germanium Alloy", presented at the IEEE Bipolar Circuits and Technology meeting, Minneapolis (1987)

Proceedings of the 14th Conference (1982 International) on Solid State Devices, Tokyo, 1982;
Japanese Journal of Applied Physics, Volume 22 (1983) Supplement 22-1, pp. 393–396

Design Considerations of Coupling Capacitors in GaAs Integrated Circuits

Alec LIVINGSTONE and David WELBOURN

British Telecommunications Research Laboratories,
Martlesham Heath, Ipswich, Suffolk, England, IP5 7RE

Capacitor coupling of GaAs depletion mode FETs is being pursued as a method of achieving low power integrated circuits which are tolerant of widely varying process parameters. The performance of these capacitors has been examined in three particular areas; the effect of back-bias applied to an adjacent contact, the effect of geometry on the frequency dependence of the capacitance and alternative capacitor structures. The results have significance beyond capacitor coupling because they reveal the interaction of active device layers with other circuit elements, results which become increasingly important as high packing densities are approached on GaAs wafers.

§1. Introduction

Logic stages comprising GaAs depletion mode FETs are fast and simpler to fabricate than those using enhancement mode transistors. However, their interconnection requires level shifting circuitry which may use a large area of wafer, be intolerant of process parameter variations, consume power and require two power supplies. Passive interconnection of logic stages by capacitors has been proposed[1,2] to provide some improvement in all of these areas.[1,2]

The basis of capacitor coupling is shown in Fig. 1 in which two depletion mode FET inverters are connected. When node A is high, node B is clamped by the forward bias gate current, and the capacitor is charged through Q2 to about 4.5 V with a 5 V supply. When node A moves to its low logic level a portion of the capacitor voltage is retained, taking node B to a negative voltage, beyond pinch-off, turning off the second inverter. The voltage across the capacitor in both states is always greater than the FET pinch-off voltage.

Several techniques could be chosen to fabricate the coupling capacitors. One of the simplest, and the one we have chosen to implement, is to use reverse biased Schottky diodes which can be formed with the FET gates, thus simplifying manufacture. This method of interconnection fulfils all the requirements of the capacitor coupling technique and simple dividers and latches have been made operating at frequencies up to 1 GHz and over a very wide range of FET pinch-off voltages.

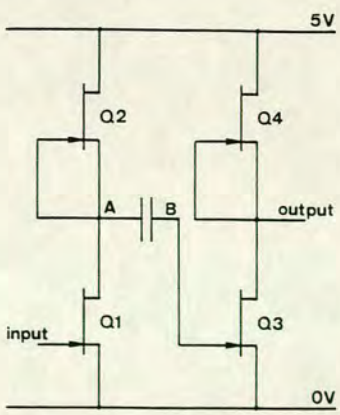


Fig. 1. Two capacitor coupled inverters consisting of depletion mode FETs.

§2. Capacitor Structures

Curve A in Fig. 2 shows the C-V plot of a

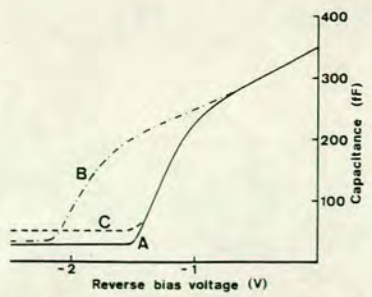


Fig. 2. Capacitance-voltage plots for Schottky diodes of (A) low periphery, (B) high pinch-off voltage and (C) high periphery.

typical Schottky diode on implanted GaAs. The disadvantage of such a diode as the coupling capacitor on the same active layer as the FETs is that there is very little useful capacitance beyond the FET pinch-off voltage. There are two ways in which the situation can be remedied; one is to fabricate the capacitor on material of higher pinch-off voltage than the FET, either by using two implants or by selectively etching the transistor channels, as shown in curve B. An alternative approach, which does not involve additional fabrication steps, is to use the same active layer for both capacitor and FET but increase the contribution from the capacitor periphery as shown in curve C.

To establish an accurate model for the C-V characteristic of high-periphery capacitors a test mask has been prepared containing structures as shown in Fig. 3. Design variables include the finger width, spacing and length, the back-bias contact material and its separation from the capacitor. In addition, inverters with capacitor inputs were included to confirm the effects on circuit operation.

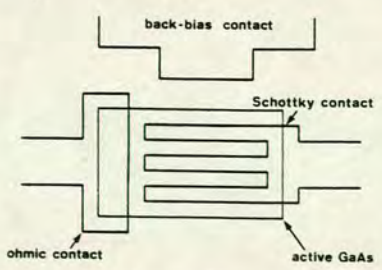


Fig. 3. Test structure used to evaluate back-bias effect.

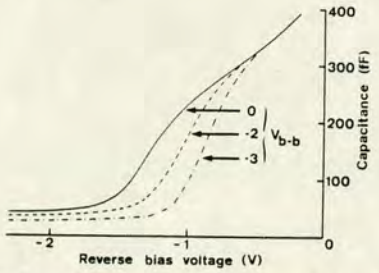


Fig. 4. Effect of back-bias on the Schottky diode C-V characteristic.

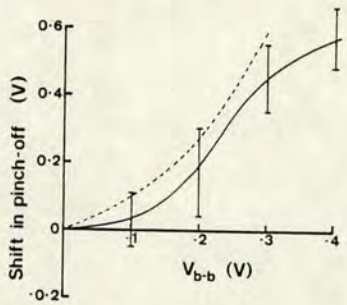


Fig. 5. Shift in pinch-off voltage averaged over 3 cm wafers with $V_p = -1.67$ V (——) and $V_p = -3.99$ V (· · · ·). The error bars show the spread over 123 capacitors.

§3. Experimental Results

3.1 *Back-bias*

The C-V characteristics of the structures of Fig. 3 have been measured at 100 kHz with back-bias applied to contacts at various distances from 5 μm upwards. The characteristics are modified by back-bias applied to a mesa 5 μm from the capacitor as shown in Fig. 4. There are two significant factors for capacitor coupled logic. Firstly, for high periphery diodes, the peripheral contribution to the capacitance beyond pinch-off is reduced by back-bias; secondly, if high pinch-off material has been used to provide the additional capacitance, then its pinch-off voltage is reduced by back-bias. Thus, the effect of back-bias voltage is detrimental, whichever scheme is employed to increase the capacitance of the coupling diode. The effect on pinch-off increases markedly when the back-bias is taken to less than -1 V as shown in Fig. 5, which shows the effect on two wafers, one with a low pinch-off voltage suitable for FET fabrication and one with a higher value which might be suitable for a capacitor or ohmic contact. Although there is a substantial increase in pinch-off shift beyond -1 V, there is not the threshold which Lee[3] has reported. The size of the back-bias effect is dependent on how the bias is applied; back-bias applied to Schottky metal deposited on semi-insulating (SI) GaAs has every much less effect than that applied to ohmic contact metal, whether on active material or the substrate.

3.2 *Geometrical effects*

When driving a large number of logic stages it is possible to use several small capacitors associated with each input gate, or to use one large capacitor to drive many gates. The latter approach might use long fingers,

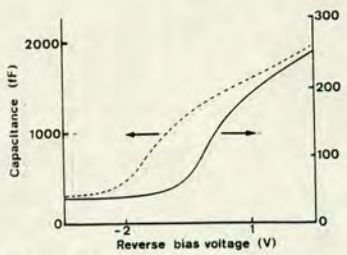


Fig. 6. Effect of geometry on capacitor pinch-off voltage. (· · · ·) 265 μm fingers, (——) 25 μm fingers.

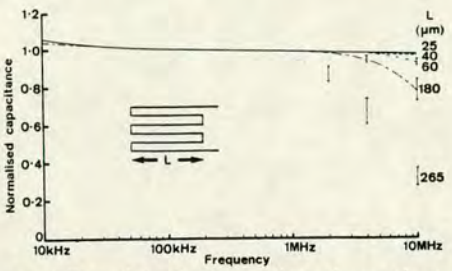


Fig. 7. Effect of the frequency dependence of capacitance on finger length.

maintaining the finger width and spacing. Figure 6 shows that such geometrical variations cause the capacitor pinch-off voltage to change. The figure shows the C-V characteristics, at 100 kHz, for capacitors with long (265 μm) and short (25 μm) finger lengths, both having a spacing and width of 2 μm. The figure has been scaled to show the pinch-off voltage increase of almost 0.5 V for the larger device.

A series of such capacitors has also been measured as a function of frequency. As Fig. 7 shows, there is a drop in capacitance at higher frequencies as the finger length increases. The curves in the figure have been normalised to their 100 kHz value to aid comparison.

§4. Discussion of Results

The shift in device characteristics due to back-bias has been attributed to modulation of the active layer at the active layer/substrate interface.[4] The shift in C-V curves shown in Fig. 4 is very similar to those of Rossel *et al.*[4] when applying the bias voltage by means of a substrate contact. Results made on our structures, using bias applied either from the substrate or adjacent surface features, are consistent if it is assumed that the substrate is a highly resistive block of material and the effective back-bias voltage is a result of a simple summation of potentials applied on the surface or elsewhere.

McIntyre[5] has calculated the pinch-off voltage of various transistors as a function of their implanted profile. The same calculation can be used to predict the pinch-off voltage of the capacitor as the active layer interface moves towards the surface under the influence of back-bias. The capacitance beyond pinch-off is purely peripheral, obtained simply from the product of periphery and effective layer depth. The device of Fig. 4 has a reduction of pinch-off voltage from -1.66 to -1.1 V, accompanied by a reduction in capacitance of 38%. The simulation using the predicted LSS implant profile gives a reduction in pinch-off voltage from -1.5 to -1.0 V with a reduction of capacitance of 36%. The C-V plots show a decrease in the peripheral capacitance as the reverse bias is increased beyond pinch-off; the detailed analysis of this effect is still in progress.

§5. Consequences for Circuit Design

In formulating design rules for the capacitors the first objective is to minimise the back-bias effect by suitable design of the logic stage. It was noted earlier that Schottky metal deposited directly on the GaAs substrate had very little effect on nearby active devices. Our present technology uses this metal, Ti/Au, for the lower level of interconnection, the ohmic metal only being used as an intermediate stage between the interconnect and the active material.

Having reduced the effect as much as possible it is desirable to select the optimum method of fabricating the diodes, either high periphery devices on the same active material as the FETs, or to use an additional implant. Figure 8 shows a simple layout (equivalent to half of

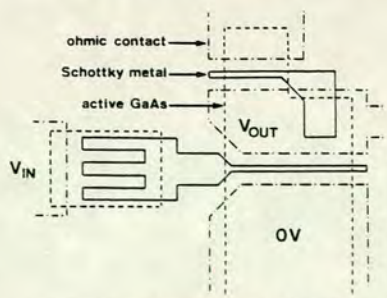


Fig. 8. Simple layout used to predict effect of back-back-bias on inverter characteristic.

Table I. Comparison of plane and high-periphery diode capacitors.

Periphery	Area	Implant		Charge
		Energy	Dose	
(μm)	(μm)^2	keV	ions·cm^{-2}	(fC)
234	304	225	2.25×10^{12}	69
70	300	350	5×10^{12}	129

Fig. 1) in which the capacitor is considered to be modified only by the voltages on the drain and source of the nearby FET; present design rules could allow the two areas of active of active material to be only 8 μm apart. When the input voltage is high, 5 V with a 5 V supply, the output voltage may be near 0.5 V. The voltage presented to the capacitor can be taken to be the mean of the output voltage and $V_{SS}(=0\text{ V})$ so the back-bias voltage on the capacitor is about -4.75 V. When the input is low, near 0.5 V, the output voltage is 5 V, the back-bias voltage is >0 V and there is negligible back-bias effect. It is possible to take two diodes of similar area and calculate the total charge available through the transition from logic high to low. The results are summarised in Table I, and show that the plane diode on material of a higher pinch-off voltage is the better choice. The total charge available from the high periphery diode is 14% less than would have been expected in the absence of back-bias, 23% less in the case of the high pinch-off diode. The choice of high pinch-off diode is more obvious when large capacitors are required because of the geometrical effects noted earlier.

§6. Conclusion

The design of integrated circuits using capacitor coupled logic need not be compromised by back-bias effects. Design rules restricting the use of some metal layers can minimise the effect and the use of a second implant for the capacitors can provide adequate margin for circuit design. Capacitor coupling eases the design problem by removing the need for a negative supply rail.

Acknowledgement

Acknowledgement is made to the Director of Research of British Telecommunications for permission to publish this paper. The devices were made and measured in the GaAs Microwave Device Section of the British Telecom Research Laboratories.

References

1) P. J. T. Mellor and A. W. Livingstone: Electron. Lett. **16** (1980) 749.
2) N. Hashizume *et al.*: Electron. Lett. **17** (1981) 553.
3) C. P. Lee: IEEE Electron Device Lett. **EDL-3** (1982) 97.
4) P. Rossel *et al.*: Revue de Physique Appliquee **13** (1978) 503.
5) N. McIntyre: Electron. Lett. **18** (1982) 208.

Manufacturing Tolerance of Capacitor Coupled GaAs FET Logic Circuits

A. W. LIVINGSTONE, A. D. WELBOURN, AND G. L. BLAU

Abstract—**Interstage coupling of depletion mode GaAs digital circuits by means of capacitors has been proposed as a simply fabricated technique combining both low power and process tolerance. This is now substantiated by results of ring-oscillators and dividers, operating at up to 1 GHz, on wafers over which the FET pinch-off voltages vary between −0.5 V and −4.0 V.**

GaAs DEPLETION mode Schnottky gate MEFSET's have been used in the manufacture of high-speed digital integrated circuits. Although the quality of active layers of GaAs has circuits. Although the quality of active layers of GaAs has improved dramatically in recent years, there is still a need for an integration technique which is tolerant of process paramter variations. Capacitor coupling has been proposed as a method of achieving passive, low power interconnection of logic stages without, necessarily, any high-speed compromise although there is a lower frequency limit [1]. The practicality of the technique has been demonstrated both in capacitor coupled logic (CCL) and a similar technique, Schottky-barrier coupled Schottky-barrier FET logic (SSFL) [2].

One of the claims made for CCL is that it is less process parameter dependent than other interconnection techniques. This paper reports the successful operation of dividers and ring-oscillators on wafers with FET pinch-off voltages varying from −0.5 V to −4.0 V.

A simple capacitor coupled inverter is shown in Fig. 1, in which the capacitor is formed from a reverse biased Schottky diode. When the input voltage is at its high level V_{IH}, the gate of FET P1 is clamped at 0.5 V by the Schottky gate, resulting in a reverse bias diode voltage of $(V_{IH} - 0.5)$ V. When the input goes to its low voltage level V_{IL}, the FET gate moves to a negative voltage

$$V_{GL} = V_{IL} - \alpha(V_{IH} - 0.5)$$
$$\cong 1.0 - \alpha(V_{DD} - 0.5)\text{V}.$$

The parameter α (the ratio of the diode voltages before and after switching) is the transfer efficiency due to the charge sharing between the coupling diode and the FET gate. In practice, V_{IH} from the preceding logic stage is equal to the supply voltage V_{DD}, and V_{IL} is normally 1.0 V, so the FET gate low voltage is −1.9 V for a typical transfer efficiency of 65% and a 5 V supply. For best performance, $V_{GL} \leqslant V_P$,

Manuscript received June 1, 1982; revised July 12, 1982.

The authors are with British Telecommunications Research Laboratories, Martlesham Heath, Ipswich, Suffolk, England, IP5 7RE.

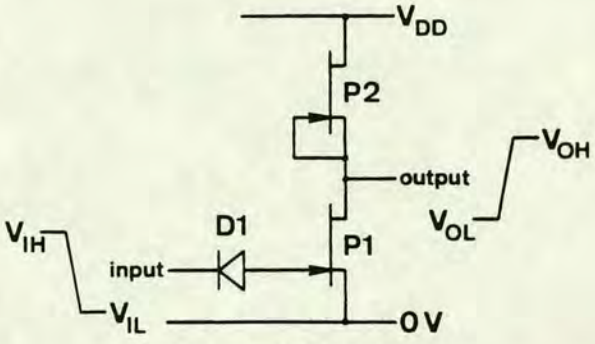


Fig. 1. Simple CCL inverter.

so the normal pinch-off voltage requirement of such a circuit would be about −1.6 V.

To achieve high transfer efficiency α, the coupling diode, must have high capacitance at a reverse bias substantially beyond pinch-off. Figure 2 shows two ways in which this might be achieved. Curve *A* shows the capacitance of a conventional plane diode which pinches-off at the transistor pinch-off voltage, V_P. Beyond V_P the capacitance is small (due only to the diode periphery) and would not result in good transfer efficiency. The solutions normally adopted are to use diodes with large periphery (as shown in curve *B*) or different impurity concentrations for the FET and coupling diode. The results reported here use large periphery diodes on high dose, deep implant active material, giving the result of curve *C* in Fig. 2. Typical high transfer efficiency diodes may occupy as much as 40% of total inverter area; α is closely related to diode area.

Having achieved a high transfer efficiency, the earlier equation for V_{GL} showed that the circuit would operate for various values of V_P if V_{DD} were modified accordingly. To test this, two wafers were processed with a Se implant of 5×10^{12} ions cm^{-2} at 400 keV, giving a pinch-off voltage near −6 V.

This active layer was used for the capacitors, but the FET gates were recessed using wet etching in order to reduce the FET pinch-off voltage. Although the majority of FET's on the two wafers were within normal limits there was a gradual variation across the wafers with a significant number of FET's throughout the range $-4 < V_P < -0.5$.

Figure 3 shows ring oscillator propagation delay measurements for the two wafers. It shows satisfactory operation over the whole pinch-off range; the inverters had $FI = FO = 1$ with 12×1.5 μm driver transistor gates. Apart from some variations at low pinch-off voltages, the propagation delay is

0193-8576/82/1000-0284$00.75

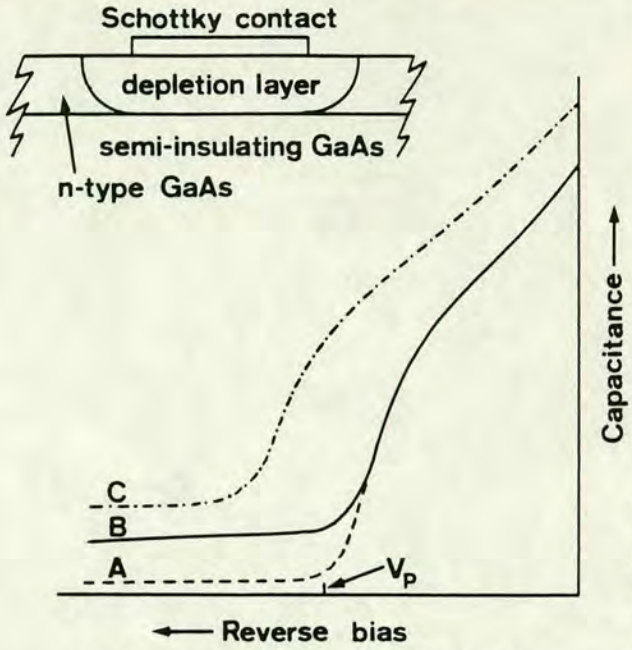


Fig. 2. *C-V* characteristics of various Schottky diodes.

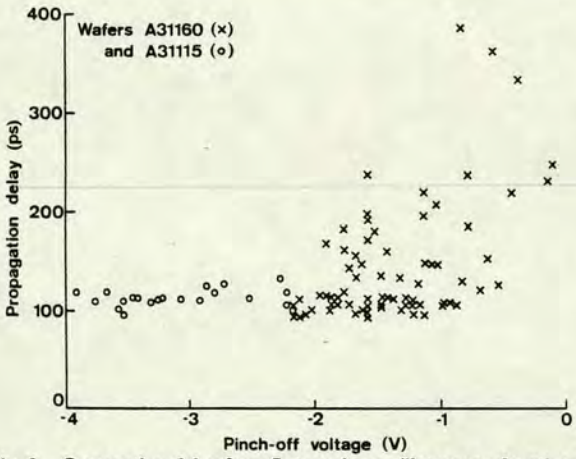


Fig. 3. Propagation delay from 7-stage ring-oscillators as a function of pinch-off voltage.

constant at about 105 ps. This delay is higher than might be achieved if ultimate speed performance was important. However, some compromise has been made in this particular design by increasing the inverter aspect ratio to match the wide tolerance to pinch-off voltage variation provided by CCL. The constant propagation delay is directly related to the transistor channel after the gate recess etch. As the pinch-off voltage is reduced, the transistor saturation current falls at the same rate as the total charge under the gate. The ratio of these last two parameters directly controls the delays in the circuit, hence the small variation over the full range of pinch-off voltages.

The processed wafers also contained D-type latches and dynamic binary dividers made using NAND gates, in a master/slave configuration, with two 20 × 1.5 μm input gates in a

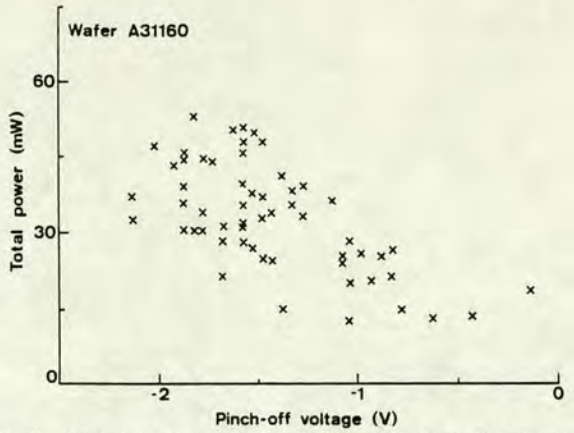


Fig. 4. Variation of power and pinch-off voltage for divider circuits.

7 μm source-drain spacing. The dividers on the test wafers also operate over a wide range of pinch-off voltages, some at speeds of up to 1 GHz. This is the highest frequency at which devices have been tested since these measurements were made on wafer probes on unpackaged circuits. Because of the capacitance coupling, CCL dividers have a low frequency limit. At room temperature the test circuits operated down to 20 kHz, which was adequate for many applications. In order to get working circuits at higher pinch-off voltages, higher supply voltages are required as indicated by the earlier equation for V_{GL}. Figure 4 shows how this results in increased dissipation. For $V_P = -1.5$ V the dissipation is typically 3 mW per gate at $V_{DD} = 5$ V.

The tolerance of CCL to process parameter variations has been demonstrated in the ring-oscillator and divider results reported here. The results in Fig. 4 represent a 40% yield of working dividers with pinch-off voltages up to −2.2 V. Dividers on the other wafer operated up to $V_P = -3.0$ V. Clearly, although working circuits can be achieved over a very wide range of pinch-off voltage, their performance depends on the margin allowed for process parameter variation; optimum performance still requires a restricted range of pinch-off voltage. For a $V_P = -1.6$ V ± 10% and with the use of NAND/NOR gates, dividers have been designed which are expected to have improved speed with less than 1 mW/gate. Although this dissipation and power/delay product are within those offered by other depletion mode MESFET techniques, they are achieved in addition to the improved process parameter tolerance of CCL.

Acknowledgment

We thank the Director of Research, British Telecommunications for permission to publish this paper. The circuits were made by members of the GaAs Microwave Devices Section of the British Telecommunications Research Laboratories.

References

[1] A. W. Livingstone and P. J. Mellor, "Capacitor coupling of GaAs depletion-mode FET's," *IEE Proc.*, vol. 127, pt. I, pp. 297–300, 1980.

[2] N. Hashizume et al., "Low-power gigabit logic by GaAs SSFL," *Electron. Lett.*, vol. 17, pp. 553–554, 1981.

A High Speed GaAs 8 Bit Multiplexer Using Capacitor-Coupled Logic

A. DAVID WELBOURN, GRAHAM L. BLAU, AND ALEC W. LIVINGSTONE

Abstract—Capacitor-coupled logic has been used to design and fabricate a GaAs eight channel multiplexer IC for use at 1.2 Gbits/s, which is fully compatible with ECL, and which offers good stability and very high tolerances to device parameters and circuit voltages. A technique has been developed to enable initial charging of all the coupling capacitors, upon application of a simple pulse sequence to control lines. Preliminary results show correct operation of the multiplexer when operated on wafer probes up to 250 MHz, the present practical limit for such measurements. Higher frequency measurements will be carried out on packaged devices, but these results are not yet available. The divide-by-two elements in the multiplexer can be programmed to self oscillate at $\frac{1}{3}$ their maximum usable frequency, allowing simple testing of high frequency performance. A very good agreement between the measured maximum usable frequencies and those predicted from the oscillation frequencies has been achieved, with over 60 percent yield for dividers. On the basis of these preliminary results, indicating operation at speeds up to about 600 MHz, it is anticipated that future wafers with 1 μm gate lengths will operate at 1.2 Gbits/s.

Manuscript received November 2, 1982; revised January 18, 1983.

The authors are with British Telecom Research Labs, Martlesham Heath, Ipswich, Suffolk, England.

I. Introduction

In GaAs logic circuits, depletion mode MESFET's offer higher speed and drive capability than enhancement mode FET's of similar dimensions, which, together with a simpler fabrication procedure makes them attractive for many applications, despite the need for voltage-level shifting between stages. Capacitor coupling [1], [2] has been implemented as a low-power level shifting technique, maintaining the inherent advantages of depletion mode MESFET's, while combining the additional advantages of operating from a single power supply, and of offering a very high degree of tolerance to variations in device parameters and circuit voltages [3] when compared with other level shifting techniques [4]–[6].

The extension of capacitor-coupled logic (CCL) to MSI complexity has been achieved with the fabrication of an 8 bit multiplexer for use in telecommunications. Similar circuits have been successfully fabricated using BFL [7] and SDFL [8], but the circuit reported here makes use of the advantages

0018-9200/83/0600-0359$01.00

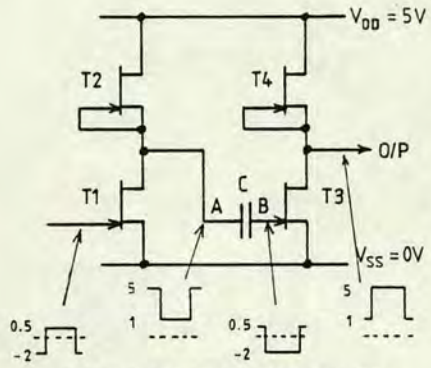


Fig. 1. Capacitor-coupled logic inverters.

of CCL, while at the same time offering a high degree of testability.

The use of a capacitor as the level shifting element implies a minimum frequency of operation of the circuits, with a low frequency cutoff F_{min} being determined by the leakage currents and relative sizes of the coupling capacitor and reverse-biased gate-source junction of the FET. Experimental results indicate that F_{min} is a few kilohertz, which is entirely acceptable for many applications, such as in telecommunications. Here, the number of consecutive ones and zeros is restricted in order to reduce the bandwidth and maintain the synchronism of data transmitted over long distances, and the minimum bit rate in high speed systems is several orders of magnitude higher than F_{min} of capacitor-coupled circuits; thus the low frequency cutoff is not a problem.

Operation of Capacitor-Coupled Logic Circuits (CCL)

Normally-on MESFET's are used for both the switch and load devices of the logic gates. Two such inverter stages are shown in Fig. 1, with a capacitor used for interstage coupling. When $T1$ is off, node A of the capacitor is charged to V_{DD} through $T2$, with node B being clamped at about 0.5 V by the forward-biased gate-source diode of $T3$. As $T1$ switches on, node A falls to the output low voltage (V_{OL}) designed to be less than 1 V, with node B following, such that the gate of $T3$ becomes negative, switching $T3$ off. Because of the charge sharing between the capacitor and the gate capacitance of $T3$, the voltage transfer efficiency between nodes A and B depends upon their relative dimensions, and is designed to be typically about 60 percent, which for V_{DD} = 5 V gives -2 V at the gate of $T3$. Once the capacitor is charged, and provided that the input frequency is sufficiently high to maintain this charge, the circuit acts as an inverter with compatible input and output logic levels defined by the power supply voltage V_{DD}, and the inverter aspect ratio, if the pinchoff voltage (V_P) is greater than -2 V. Because these voltage levels are almost independent of device parameters, a very high degree of tolerance to device parameter variations is achieved. Correspondingly, the circuits will operate over a wide range of supply voltages for a given value of V_P, allowing the system supply voltage to be chosen freely

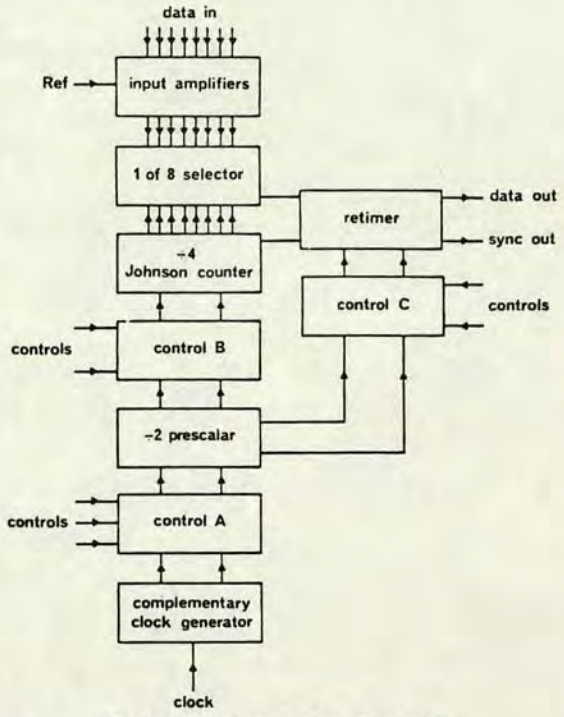


Fig. 2. Block diagram of multiplexer IC.

to suit system requirements. For convenience of fabrication the capacitors are implemented as reverse-biased Schottky diodes, occupying typically 40 percent of the gate area.

Circuit Requirements for Telecommunications Applications

When designing components for operation within a large system or subsystem, it is important that those components will interface with other elements of the system, without having individually tailored requirements. The advantages offered by CCL have been used to meet this criterion in the design of an 8 channel multiplexer for use in the continued development of experimental high speed transmission systems, capable of taking advantage of the large bandwidth offered by fiber optic cables. The multiplexer has been designed to interface with the 140 Mbit/s PCM systems already existing in the telecommunications network, requiring only a single -5.2 V supply, and capable of operating with external logic levels of -0.9 to -1.7 V, compatible with ECL voltages and signals. The high tolerance of CCL to variations in circuit voltages enables the system voltage to be selected at the -5.2 V dictated by ECL, without the CCL circuit requiring a separate, carefully chosen supply voltage. The high tolerance to variations in device parameters offers high yield, and the prospect of long term stability in a system environment.

Circuit Description

A block diagram of the multiplexer is shown in Fig. 2, with a logic diagram shown in Fig. 3. With the exception of the control circuits, the design of the multiplexer is conventional.

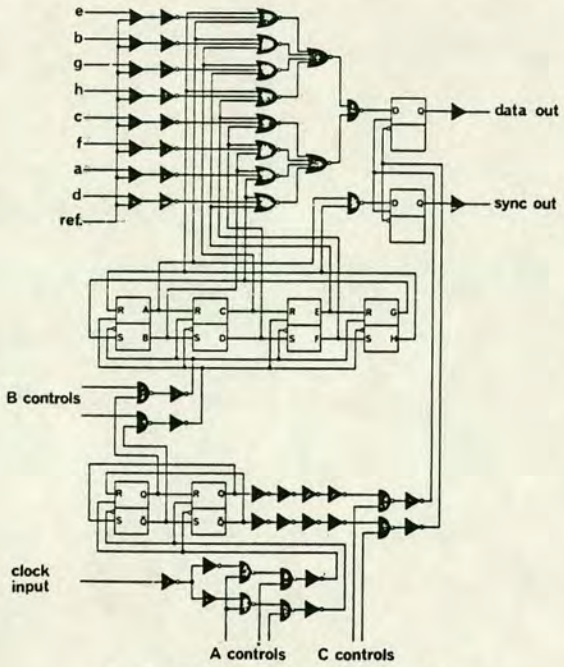


Fig. 3. Logic diagram of multiplexer IC.

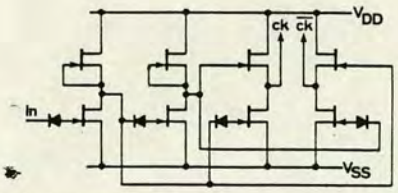


Fig. 4. Complementary clock generator used in the multiplexer.

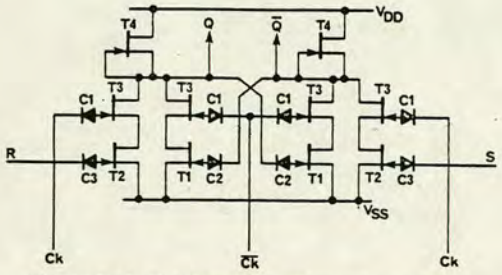


Fig. 5. *RS* latch using two complex AND–NOR gates.

The use of a Johnson counter as the main divider simplifies the decoding of the gating signals in the data selector, but requires an additional divide-by-two circuit as a prescaler.

Dual-clocked latches are used throughout the circuit because of their inherent speed advantage over the equivalent single-clocked latches. A complementary clock generator as shown in Fig. 4 is fabricated on-chip to allow operation from a single external clock. The input and output signals of an inverter are used to drive two quasi-complementary output stages,

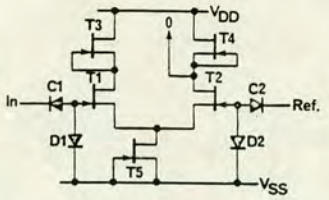


Fig. 6. Input amplifier used to convert ECL logic swings into larger CCL swings.

such that these stages produce complementary outputs. Simulations indicate that the two signals so derived are very closely complementary. Because depletion mode transistors are sued, multilevel logic gates can be constructed, enabling simple fabrication of a basic *RS* latch from two cross-coupled AND–NOR gates as shown in Fig. 5.

Input amplifiers (Fig. 6) are provided on each of the eight input channels, to convert the ECL logic levels, to the larger swing used internally by CCL. The amplifier consists of a long-tailed pair, using depletion mode FET's as the load elements. The input signal is capacitively coupled to the FET $T1$, and because of charge sharing between $C1$ and the gate capacitance of $T1$, only a small voltage appears at $T1$ gate. To maximize the gain from the amplifier, $T1$ is biased close to pinchoff using diode $D1$ and the voltage across the current source $T5$. The actual bias point of $T1$ is governed by the ratio of the forward-biased current through $D1$ and the leakage current of $C1$. As this voltage is highly dependent upon operating conditions, the reference voltage is derived in a similar manner using $C2$ and $D2$. Normally the external reference voltage is connected to V_{DD}, but for test purposes the reference voltage may be set by an external supply. Full ECL compatibility is achieved by using output buffers capable of driving ECL signal levels into a 50 load, although it is expected that the load device will be a semiconductor laser driver, the specification for which has yet to be defined.

Preliminary results have been obtained using wafers uniformly implanted with a dose of 3×10^{12} selenium ions at 400 keV, to give a large punch-through voltage for the reverse-biased coupling diodes, although selective implantation will be used on future circuits. The pinchoff voltage of the FET's is reduced by recessing the transistor gates, using a wet chemical etch. Au Ge-Ni is used as the ohmic contact metallization with TiAu used for the gate and second level of interconnect metallization, with a 1 μm thick polyimide dielectric layer between the two levels of interconnect. A lift-off technology is currently used for patterning all the metallizations.

In total, the multiplexer is an 80 gate circuit, consisting of 260 FET's and 168 Schottky diodes. A microphotograph of the 2×2 mm chip is shown in Fig. 7, with test elements including a 7 stage ring oscillator, visible along the top edge. For satisfactory operation of output buffers and logic elements V_P must lie between -1.2 and -1.8 V, although a larger spread may be tolerated with changed output amplitudes and supply requirements. With a -1.6 V pinchoff voltage, power consumption is less than 250 mW at the designed -5.2 V supply, decreasing with reduced supply voltage.

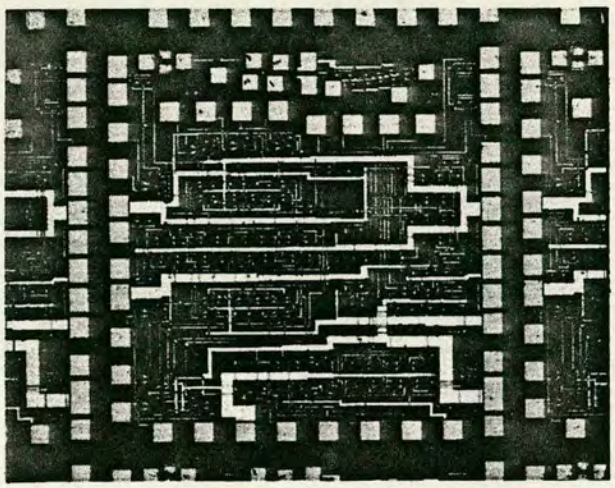


Fig. 7. Chip microphotograph of the multiplexer.

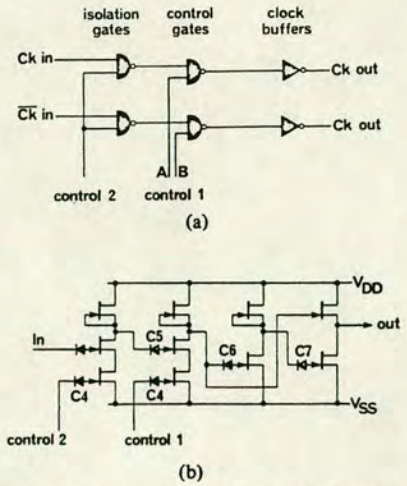


Fig. 8. Schematic of the control circuit used for charging the capacitors. (a) Logic diagram. (b) Circuit diagram of one channel.

Circuit Initialization and Testing

The correct operation of a CCL circuit requires that all of the capacitors are fully charged. In a combinatorial circuit, charging may be achieved by ripple through from the inputs, but in a sequential logic circuit, the situation may obtain where, due to the feedback loop in the flip-flops, it is possible for some capacitors to remain unchanged.

In a toggle flip-flop, two of the latches shown in Fig. 5 are cross coupled, with the R and S inputs being coupled to the Q and $\overline{Q}$ outputs of the latch. Capacitors $C2$ and $C3$ can only be charged if nodes Q and $\overline{Q}$ become high, requiring both branches of each AND-NOR gate to be turned off. $T1$ and $T2$ cannot be used to turn the gate off until those capacitors are charged, so the clock and $\overline{\text{clock}}$ inputs must be used to turn both transistors $T3$ off simultaneously. During normal operation those signals are in antiphase, and a control circuit as shown in Fig. 8 is incorporated, such that, on the application of external gating signals, these two lines may be driven in-phase, thereby charging all the capacitors $C2$ and $C3$. The pulse sequence required on the control inputs is shown in Fig. 9. Provided that the pulse widths are less than $1/2F_{\text{min}}$ the circuit will initialize, and once intialized no refresh is required while the clock frequency exceeds F_{min}. Fig. 10 shows an oscillogram of the outputs of a simple TTL circuit used for this initialization.

The provision of these control circuits amounts to an increase in complexity of the multiplexer of about 10 percent. However, by separating the control line inputs as shown in Fig. 8, the control circuits may be used in other ways, greatly increasing the testability of the circuit. The divide-by-two circuit will undergo oscillation with a period of 8 τ_{PD} (where τ_{PD} is the AND-NOR gate propagation delay for a fan-out of 3), which relates to the maximum allowable clock frequency of $1/(2\tau_{PD})$, giving a simple test of the operation of the circuit. In addition, external clocks may be applied via the control lines in order that the Johnson counter may be tested independently of the operation of the divide-by-two prescalar. In this way useful data may be obtained from only part functional circuits. The provision of several buffered test nodes also helps with the testing of such a circuit in a new technology.

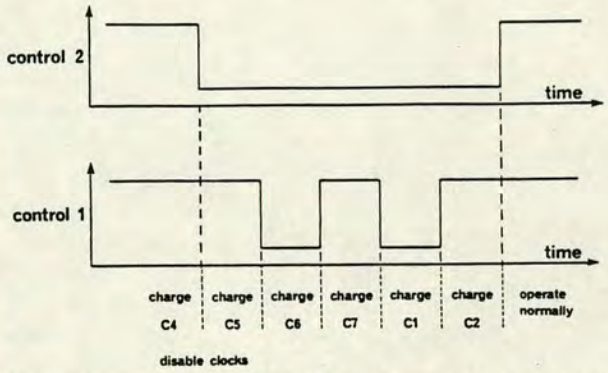


Fig. 9. The waveforms and associated actions required on the control lines of Fig. 8.

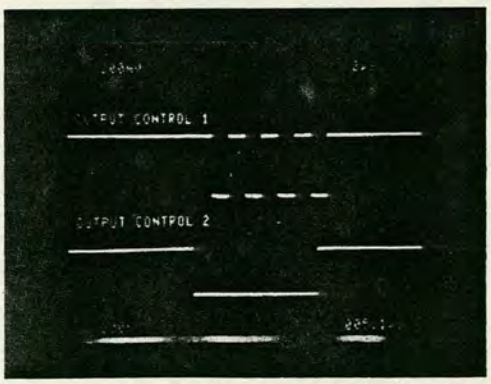


Fig. 10. Oscillogram showing the outputs of a simple TTL circuit designed to control the charging circuit.

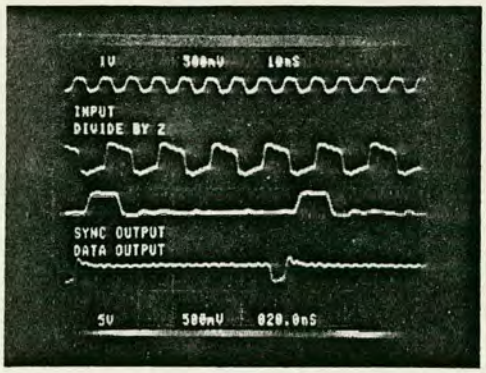


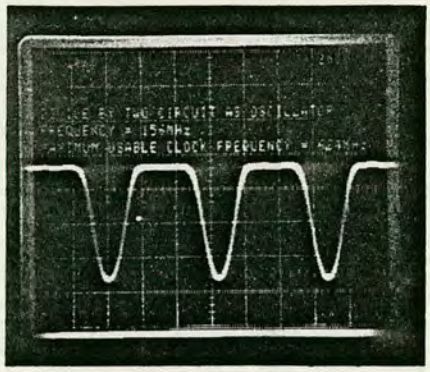
Fig. 11. Operation of the multiplexer with a single input activated.

Results

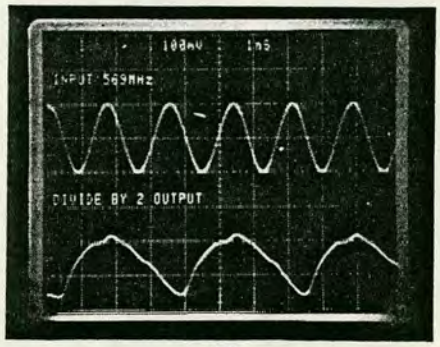
Satisfactory operation of the multiplexer has been demonstrated and preliminary results are reported here. Because of the design adopted, at least four inputs to the multiplexer must be activated before correct operation takes place. Fig. 11 shows the resultant waveforms obtained by operating one channel at high speed, while three channels were toggled with narrow pulse width, low speed, asynchronous signals, with the remaining four channels held at V_{DD}, giving a 11111110 pattern. These results were obtained from measurements on wafer probes, with the attendant problems of poor impedance matching of input lines, and cross-talk on the output. Higher frequency measurements on complete circuits can only be made on packaged devices which are not yet available. Packaged devices will be tested using a programmable 8-channel 200 MHz word generator, which is being developed in this laboratory using ECL III circuits. It is intended that automatic testing of the devices will be carried out using this word generator and a programmable sampling oscilloscope.

Predictions of the high frequency performance of the multiplexer are possible because of the design features already discussed. The operation of the divide-by-two circuits in their self-oscillating mode is at a frequency $\frac{1}{4}$ of their maximum usable clock rate offering two useful features for probe testing: the frequency is comparatively low, well within the probe card capability, and, as the clock signal is internally generated, there are no impedance matching problems. Fig. 12 shows operation of the divide-by-two circuit as an oscillator and as a divider near the maximum predicted clock frequency of 624 MHz for this particular wafer location. The inverter propagation delay corresponding to these results was 200 ps, obtained from a seven stage ring oscillator with unity fan-in and fan-out.

These results were obtained on a wafer with 1.5 μm gate-length which had been over-etched giving a V_P of -0.9 V, outside the range for which the circuit was designed. Consequently the specification of a maximum frequency >1.2 GHz has not been achieved on this wafer. It is anticipated however that further devices will meet this specification. The low frequency operation of the divider is depicted in Fig. 13, where



(a)



(b)

Fig. 12. Outputs of the divide by two circuit. (a) Operating as an oscillator at $f = 1/(8\tau_{PD})$. (b) Operating as a divider at $f_{min} = 1/(2\tau_{PD})$.

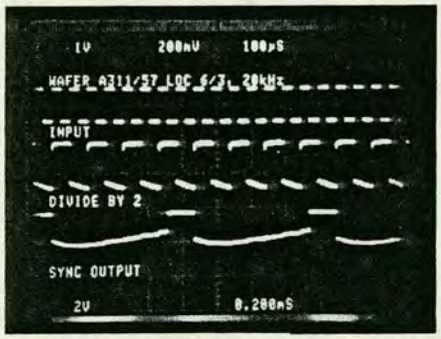


Fig. 13. Low frequency operation of the dividers, seen here operating at 20 kHz input clock frequency, corresponding to $f = 2.5$ kHz.

the input clock frequency of 20 kHz corresponds to a minimum frequency of operation for a single gate of 2.5 kHz.

Results obtained from testing the ring oscillators and dividers substantiate the claims made for CCL. A high yield, 60 percent, of dividers has been obtained, with most devices operating over the supply voltage range -3.5 V to -6.5 V (higher voltages have not been used). Similarly, a wide range of clock amplitudes is tolerated.

Conclusion

An 80 gate 8-channel multiplexer IC has been designed and fabricated using capacitor-coupled logic, and a technique has been successfully developed to charge the coupling capacitors

by means of external control signals. Successful operation of the circuit has been observed at frequencies up to 250 MHz, the limit of our test equipment suitable for making wafer-probe measurements. Higher frequency operation of packaged devices is predicted from oscillation frequencies of the divider elements on this first wafer, and the full specification for a maximum frequency of 1.2 GHz is anticipated from future wafers. The high degree of tolerance of CCL to device parameters and circuit voltages has been demonstrated at MSI complexity. This feature is considered to be of considerable importance to systems design engineers.

Acknowledgment

The authors would like to thank their colleagues in the Gallium Arsenide Microwave Devices section at British Telecom Research Labs for the fabrication of these circuits, and the Director of Research for permission to publish this paper.

References

[1] P. J. T. Mellor and A. W. Livingstone, "Capacitor-coupled logic using GaAs depletion mode FET's," *Electron Lett.*, vol. 16, pp. 749–750, Sept. 1980.

[2] N. Hashizume, H. Yamada, and K. Tomizawe, "Schottky barrier coupled, Schottky barrier gate GaAs FET logic," *Electron Lett.*, vol. 17, pp. 51–52, Jan. 1981.

[3] A. W. Livingstone, A. D. Welbourn, and G. L. Blau, "Manufacturing tolerance of capacitor coupled GaAs FET logic circuits," *Electron Devices Lett.*, vol. EDL-3, pp. 284–285, 1982.

[4] R. L. van Tuyl, C. A. Liechti, and R. E. Lee, and E. Gowen, "GaAs MESFET logic with 4 GHz clock rate," *IEEE J. Solid-State Circuits*, vol. SC-12, pp. 485–496, May 1977.

[5] R. C. Eden, B. M. Welch, and R. Zucca, "Planar GaAs IC technology: Applications for digital LSI," *IEEE J. Solid-State Circuits*, vol. SC-13, pp. 419–426, Aug. 1978.

[6] M. R. Namordi and W. A. White, "A novel low-power static GaAs MESFET logic gate," *IEEE Electron Devices Lett.*, vol. EDL-3, pp. 264–267, Sept. 1982.

[7] G. McCormack, A. Rode, and E. Strid, "A GaAs MSI 8-bit multiplexer and demultiplexer," in *Proc. GaAs IC Symp.*, 1982.

[8] S. I. Long, F. S. Lee, R. Zucca, B. M. Welch, and R. C. Eden, "MSI high speed, low power, GaAs integrated circuits using Schottky diode FET logic," *IEEE Trans. Microwave Theory*, vol. MTT-28, pp. 466–472, 1980.

A. David Welbourn was born in Brighouse, Yorkshire, England, in 1957. In 1976 he was awarded a University scholarship with British Telecom, and studied natural sciences at Trinity College, Cambridge, England, obtaining the B.A. degree in theoretical physics in 1980.

Since 1980, he has worked in the Gallium Arsenide Microwave Devices section of British Telecom Research Laboratories, where he is responsible for the circuit design and testing of GaAs digital integrated circuits. He is currently studying part-time for an external Ph.D. degree from Edinburgh University, Edinburgh, Scotland.

Graham L. Blau was born in Grimsby, England, on June 8, 1956. He received the B.Sc. degree in electronic engineering from the University of Salford, England, in 1981.

He joined British Telecom in 1973 and worked on transmission and exchange systems. Since 1981 he has worked in the research division where he is involved in the design and test of high speed GaAs digital IC's.

Alec W. Livingstone was awarded the B.Sc. degree in 1969 by St. Andrews University, Fife, Scotland, and in 1972 he received the Ph.D. degree for work carried out on ZnSe Schottky diodes in the Wolfson Institute of Luminescence at the same university.

Since 1972, he has been at the British Telecom Research Laboratory. His early interests were in the application of high speed digital test systems to design validation of LSI circuits, such as microprocessors and codecs. Since 1979, he has been leading the GaAs IC production group fabricating high speed circuits of up to 100 gates. His main interests are in circuit design, fabrication and testing.

Gigabit logic
A review

A.D. Welbourn, B.A.

Indexing terms: *Semiconductor devices and materials, Integrated circuits, Logic, Transistors, Schottky-barrier devices*

Abstract: The current state of gigabit logic is presented. Recent developments have made a significant contribution to this field, and it is the purpose of this work to compare these new developments with the more traditional approaches. The relationships between device and circuit parameters are derived in an Appendix, and these relationships are used to explore the conditions which must be satisfied by the devices if they are to offer gigabit logic operation. To meet large system requirements, it is shown that the highest levels of integration must be used: for a given complexity the requirements of individual components, e.g. power dissipation, parameter tolerances, are derived. The various circuit types, including the recent results from HEMTs and submicron silicon MOSFETs, are discussed. For the purposes of this review, only those devices which have already been integrated are studied; thus bipolar III-Vs, permeable base transistors and other similar devices are not discussed. The developments in submicron silicon present a challenge to GaAs speed superiority. This potential superiority forms the basis of the discussion in the paper, together with a projection for the future.

1 Introduction

The progress which has been seen in the electronics industry over the last decade is virtually unparalleled. This progress has been led by digital integrated circuits, with almost all of the effort being towards an increase in circuit complexity, with the empirical 'law' of the doubling of device complexity every year (Moore's law). This growth has been stimulated by three distinguishable forces:

(*a*) the desire to prove that more complex devices can be fabricated

(*b*) the expected needs of the systems engineer for more complex IC, i.e. direct sponsorship of projects

(*c*) the opening up of new fields to the systems engineer because the more complex ICs offer greatly improved system reliability, and the consequent exploitation of such devices.

Only a relatively small portion of the research effort in ICs has been devoted to the development of higher speed devices, mainly for transmission, computing and military applications, culminating in so-called 'gigabit electronics'. The same three driving forces can, however, be identified, in the field of gigabit logic, although there has been more hesitation on the part of systems engineers to explore the new fields opened up by gigabit electronics.

In the desire to prove that devices can be made faster, one of the most commonly quoted figures in the literature is the propagation delay of an inverter obtained from ring-oscillator (RO) data (see Tables 1 and 2). Such a figure is difficult to interpret, as these figures are normally obtained for the ideal case of unity fan-in and fan-out, have carefully optimised supply voltages, and often have nonsaturating excursions. The actual logic circuit operation of gates is more nearly represented by an average fan-in/fan-out of 3/3, and Bosch [1] gives a Table of the factor m, for various logic approaches, which is reproduced in Table 3, where:

$$\tau_{PD(FI/FO=3/3)} = m\tau_{PD(RO)}$$

One of the difficulties in comparing logic families and giving predictions about circuit results is in obtaining truly equivalent results. To this end, Micheel *et al.* [2] have given a rule-of-thumb method for calculating the various contributions to the propagation delays.

The major stimulus in the development of gigabit electronics has been, and still is, the increasing requirement for higher speeds in systems applications. Bosch [3] presents an extensive review of these applications and a brief review is given here in Section 2. Most of these applications fall into category (*b*) above; the subject is too immature as yet for category (*c*) to be making a great impact, but this situation is rapidly changing.

Several long-term projects have been defined to stimulate rapid progress toward gigabit systems. In the USA a programme is under way to develop very high speed integrated circuits (VHSICs) using submicron feature sizes in silicon, and in Japan a 7-year programme is being undertaken into the use of modulation-doped structures, promising extremely high electron mobilities (see Section 4. 6).

2 Applications of gigabit logic

The applictions of gigabit logic fall into five main categories:

(i) high-speed transmission systems
(ii) high-speed signal processing
(iii) satellite communications
(iv) high-speed computers
(v) test and measurement systems.

2.1 High-speed transmission systems

To meet the growing needs for telephone, televison and data links higher bit rates are being considered for use with fibre-optic transmission systems [4, 5]. A viable alternative, however, to the 1.12 and 2.24 Gbit/s PCM links being considered is to be found with the use of wavelength multiplexing at existing bit rates; such systems require the use of monomode fibres with several lasers operating in one of the longer wavelength low-loss windows of 1.3 or 1.5 μm. Integrated optics mixers would be used to combine the lower bit rates of 140, 280 and 560 Mbit/s systems which are either already in use or under development. With the fields of gigabit electronics and integrated optics both still in their infancy, it is uncertain at this stage which of the two methods will be adopted for increasing the amount of data transmitted in one link. With the use of GaAs or other III-V compound semiconductors for high-speed logic, there are also prospects for the integration of logic elements with the source and/or detector [6] in such systems.

Paper 2144 I, first received 5th May and in revised form 6th August 1982

The author is with British Telecom Research Laboratories, Martlesham Heath, Ipswich IP5 7RE, England

Table 1: Ring oscillator results — minimum propagation delay

Propagation delay	Power delay product	Dimensions	Circuit type	Comments	Reference
ps	fJ	μm			
13	00.034	2.5	Josephson	4K	106
17.1	16.4	1.7 X 33	HEMT	77K	92
18.4	16.6	0.7 X 20	HEMT (LPFL)		117
19.6	170	0.5 X 10	GaAs E MESFET	self-aligned	122
29	168	0.4	Si MOS		113
33	2	0.7 X 20	HEMT (LPFL)		117
33	37	0.4	Si MOS		113
34		0.5	GaAs BFL		42
39.5	158	0.8 X 20	GaAs E MESFET	self-aligned	62
45	126	1.3 X 50	GaAs E JFET		66
50	287	1.5 X 30	GaAs E MESFET	refractory gate self-aligned	63
51	97	0.8 X 20	GaAs E MESFET	77K	60
55	193	1.2 X 20	GaAsCCL		54
55	237	1 X 20	GaAs LPFL		73
56	26	1.7 X 33	HEMT		92
60	1200		reduced power BFL — GaAs		44
61	4000	0.75 X 40	GaAs BFL		41
66	188	1.2 X 20	GaAs E MESFET		20
72	140	1.2 X 30	GaAs MOSFET		78
72.5	2800	0.8 X 20	Si MOS		33
75	170	1 X 20	GaAs SDFL		103
77	76	0.8 X 40	GaAs E MESFET		79
85	187		Si bipolar		26
85	34	2 X 10	GaAs E JFET		66
86	3900	1 X 20	GaAs BFL	FI/FO = 2/2	15

Table 2: Ring oscillator results — minimum power delay product

Power delay product	Propagation delay	Dimension	Circuit type	Comments	Reference
fJ	ps	μm			
0.034	13	2.5	Josephson	4K	106
1.6	200	0.8 X 20	GaAs E MESFET		79
2	33	0.7 X 20	HEMT (LPFL)		117
6.9	87	1.2 X 20	GaAs E MESFET		20
11.9	173	1 X 20	GaAs LPFL		73
15.0	100	1.5 X 30	GaAs E MESFET		63
16.4	17.1	1.7 X 33	HEMT	77K	92
16.6	18.4	0.7 X 20	HEMT		117
20	385	1.5 X 100	GaAs MOS		76
27	156	1 X 5	GaAs SDFL		76
33	67	1.3 X 10	GaAs E JFET		66
34	85	2 X 10	GaAs E JFET		67
36	157	1.2 X 30	GaAs MOS		76
37	33	0.4	Si MOS		113
37	45	0.8 X 10	GaAs E MESFET		62
47	140	1 X 10	GaAs SDFL		103
75	102		Si bipolar		26

Room temperature operation except where stated. E = enhancement mode, D = depletion mode

Table 3: Factor m relating FI/FO = 3/3 propagation delays to ring oscilator delays [1]

Circuit type	m
Si bipolar	1.5
Si MOS	4.5
GaAs BFL	1.5
GaAs SDFL	1.5
GaAs E MESFET	2.5
GaAs E JFET	1.8

2.2 High-speed signal processing

This area could prove to be the major application for gigabit logic. The aim is to replace inflexible analogue processing systems with more flexible digital processing. Some of the applications are for real-time processing, e.g. of radar signals, such that adaptive beam forming and target recognition become feasible. Real-time spectrum analysis and synthesis have many applications, including meteorological studies, propagation predictions, image processing etc. Real-time data compression for low-bandwidth video links is a further

application, requiring very-high-speed computation if it is to be carried out digitally. Even low-data-rate voice links of 64 k bit/s require gigabit processing rates if the complexity of the compression algorithm is increased to offer a smaller bandwidth. Most of these applications centre around the ability to be able to process the data in real time, as it is generated or acquired, where currently this is not feasible.

2.3 *Satellite links*

With the rapidly increasing use of satellite communications, it readily becomes apparent that expansion of the facilities available is not simply a matter of increasing the number of satellites in use. The minimum angular separation of geo-synchronous satellites has recently been reduced to 4°, still permitting less than 20 satellites to serve the whole of the USA. Thus, the most efficient use of satellite links must be made. This leads to requirements for very-high-bandwidth highly compressed data links. Both compression and multiplexing require high bit rates. Further problems arise because the ground station facilities are inadequate for the full utilisation of existing satellites. Some of this ground traffic could be relieved if intersatellite links were available. These would require high-speed low-power processing facilities on board the satellites. To date the satellites only carry the necessary electronics for transversion, whereas 'intelligent switching' would be required for intersatellite communications. In addition to the requirements of high speed and low power, the processor must be 'radiation-hard'.

2.4 *High speed computers*

There is a continuing requirement for faster and more powerful computers. Modern computer architecture mixes both parallel processing, requiring high complexities, with high-speed components in critical areas to reduce bottlenecks. A projection by Eden *et al.* [7] has suggested that a single GaAs chip of about 1 cm^2 could replace the 1680 Si bipolar packages required for the CPU of a state of the art computer, offering improvements in both speed and power – by as much as two orders of magnitude each – although there is no suggestion of the feasibility of such a project. The Japanese fifth-generation computer project [113] and the IBM Josephson Signal Processor (JSP) project [114] are aimed at satisfying the demand for large, fast 'super computers', capable of more than 100 million floating-point operations per second (MIPS).

2.5 *Test and measurement systems*

To realistically test LSI and VLSI integrated circuits, testers are required which are capable of operating at speeds up to an order of magnitude higher than the device under test. With the speeds of consumer devices increasing, there is a requirement for the performance of specialist test devices to match such increases [8]. In addition to such highly specialised equipment, laboratory test equipment is beginning to utilise the higher-speed devices, with direct prescaling becoming a commonplace replacement for down-mixing in frequency counters for example.

3 Requirements for gigabit logic

Having discussed the applications for gigabit logic systems, it remains to define some of the requirements, before describing the types of devices which are available. As will be shown later, one of the hardest tasks in designing gigabit digital systems is that of interfacing the device off-chip. The speed of devices is largely controlled by the rate at which the node capacitances can be charged. It seems logical, with only a few exceptions (such as prescalers), to aim towards the highest level of integration possible, thereby eliminating the relatively large capacitances associated with IC pins and PCB tracks. An additional reason for obtaining a high level of integration is that 'microwave' problems can be reduced by keeping tracks as short as possible. Unfortunately, this inevitably means the reduction of track separations, increasing other problems. These 'microwave' problems are so often alluded to [3] but are rarely tackled in any detail. They arise from the shortening of the wavelength of the component frequencies of the waveform. Thus, for example, to achieve 50 ps rise times, the 20 GHz component must not suffer serious phase shift. As this component has a wavelength of less than 10 mm in a typical metal track, tracks must be much shorter than this, unless they are considered as transmission lines. The skin effect begins to influence the characteristics of the conductors, and radiation from tracks begins to pose serious crosstalk problems. These effects have been studied by Hasegawa [9], who suggests that very short (~10 ps) pulses will not propagate in striplines using Si/SiO^2 as a substrate, but that GaAs is a suitable substrate material. Track lengths must be kept short – less than 28 mm for 1 ns pulse widths, and 0.28 mm for 10 ps pulses. No problems have as yet been encountered, but this cannot be assumed to be the case in some of the large circuits which have been projected. Currently the largest circuits produced are a 1000-gate multiplier in GaAs, offering 5.2 ns multiply time for two 8-bit words [10], and a 16-bit multiplier in submicron silicon with a 9 ns multiply time [11].

Table 4 shows the maximum allowable dynamic switching energies at various levels of integration, and various clock frequencies. It can be seen that to approach VLSI at frequencies in excess of 1 GHz the power-delay product (dynamic switching energy) must be lower than 100 fJ. In general, the minimum clock period of a flip-flop – and hence any synchronous logic system – is at least two to four times higher than the propagation delay of the individual gates. In turn, individual gate propagation delays are upwards of twice the RO delays (see Table 3). As the scale of integration increases, inevitably the tracks become longer when compared with the dimensions of the switching device, and so increasing the load capacitance. Thus minimum RO propagation delays of around 100 ps are needed for clock frequencies in excess of 1 GHz, requiring power dissipations of less than 1 mW. A summary of the results for various of the contenders is shown in Table 1 and discussed in Section 6. This shows that many of them do, in principle, meet the requirements in terms of speed and power.

The relationships between FET parameters and the pro-

Table 4: Maximum allowable dynamic switching energies of gates against clocking frequency assuming 2W chip dissipation, at various levels of integration

Gates/chip	Clock frequency					
	0.1 MHz	1.0 MHz	10 MHz	100 MHz	1 GHz	10 GHz
ULSI = 10^5	100 pJ	10 pJ	1 pJ	0.1 pJ	10 fJ	1 fJ
VLSI = 10^4	10^3 pJ	100 pJ	10 pJ	1 pJ	0.1 pJ	10pJ
LSI = 10^3	10^4 pJ	10^3 pJ	100 pJ	10 pJ	1 pJ	0.1 pJ
MSI = 10^2	10^5 pJ	10^4 pJ	10^3 pJ	100 pJ	10 pJ	1 pJ
SSI = 10	10^6 pJ	10^5 pJ	10^4 pJ	10^3 pJ	100 pJ	10 pJ

pagation delay and dynamic switching energies are derived in Appendix 8.1, and are reproduced here:

$$\tau_{PD} \simeq \frac{8}{3} \frac{C_L l}{w \beta V_m} \quad (1)$$

$$P_D \tau_{PD} \simeq \frac{8}{3} \frac{C_L^3 l^2}{\beta^2 w^2 \tau_{PD}^2} \quad (2)$$

$$P_D \tau_{PD} \propto C_L V_m^2 \quad (3)$$

where C_L is the load capacitance which the inverter is driving, l and w the gate length and width, and β is the constant in the Schockley relationship

$$I_{DS} = \frac{\beta w}{2l} (V_{GS} - V_p)^2 \quad (4)$$

Here V_{GS} and I_{DS} have their usual meanings, and V_p is the pinch-off voltage of the FET. V_m is the amplitude of the voltage swing between high and low logic levels. For true logic operation

$$V_m \simeq |V_p - V_B| \quad (5)$$

as the device is switched off at $V_{GS} = V_p$, and the gate conducts excessive current at $V_{GS} \geqslant V_B$ (the built-in voltage or barrier height). Eqn. 1 shows that to reduce propagation delays the parameters w, V_m or β must be increased, or C_L or l reduced. As can be seen from eqns. 2 and 5, making V_p more negative will drastically increase the switching energy, and thus is not a suitable approach for VLSI. From Appendix 8.1 the theoretical value of β is given by.

$$\beta = \frac{\epsilon \mu}{a} \quad (6)$$

where a is the mean channel thickness, μ the electron mobility and ϵ the permittivity of the channel material. However, in real devices, β is defined by eqn. 4. There is a discrepancy between the two cases because the derivation of eqn. 6 neglected the FET parasitic resistances, and, in particular, ignored the gate-source resistance.

In reality, the parasitic resistances of the circuit affect this relationship, while maintaining the square law of eqn. 4. Thus, the experimental value for β may be increased by reducing the parasitic resistances, or by using a material of higher electron mobility. This latter point explains the motives behind the move to GaAs, and more recently the move to HEMTs (see Sections 4 and 5).

Eqns. 1–3 show that the load capacitance is significant in both τ_{PD} and $P_D \tau_{PD}$. This capacitance is made up of three components:

(*a*) the source-drain capacitance of the FET C_D

(*b*) the gate capacitance of the loading logic gate C_G

(*c*) the stray capacitance of the tracks due to the substrate C_S.

In general

$$C_S > C_G \gg C_D \quad (7)$$

The reduction of stray capacitances or of track lengths and widths is therefore beneficial in reducing both propagation delays and power. The former may be achieved in several ways. GaAs has a semi-insulating (SI) substrate, and Si circuits may be fabricated using silicon on sapphire (SOS), or by recrystallising polysilicon on SiO_2 or Si_3N_4 using lasers [12] electron beams [13]. Shorter gate lengths will also improve performance, both directly in eqns. 1–3, and by the reduction of C_G in eqn. 7. C_D will be increased slightly, but this effect is negligible. The effect of changing the device width is less obvious, as both C_G and C_D vary with w. If the substrate capacitance C_S of eqn. 7 were to be reduced such that C_G were the largest term, then increasing w would result in the increase of the speed-power product. However, reduction of w would also have an adverse effect on both the delay and power for two reasons:

(i) C_S would become increasingly important.

(ii) C_D is proportional to $(w + w_o)$, where w_o is an 'intrinsic width' due to fringing capacitance [14, 15], which is of the order of 3–8 μm for a 1 μm gate length. Thus C does not scale with device width

In addition to these factors, a further requirement on VLSI is that the overall logic gate area is compatible with fabricating 10 000 gates on a reasonable sized ($< 1\,cm^2$) chip. Thus each logic gate must be smaller than 10 000 μm^2. In order to pack the devices densely, the power density must be small, and fairly constant across the chip. In a move to reduce the power consumed by circuits, there has been much work carried out on GaAs enhancement-mode devices, where the optimum threshold voltages have been shown to be 0.2V [16], and so V_m is reduced to 0.5 V. This gives noise margins as low as 50 mV [17], and certainly less than 200 mV. Assuming 100% device functionality, a circuit will only function if none of the devices exhibits a threshold voltage above a given cut-off value. Long [10] points out that for a Gaussian distribution of transistor parameters on a wafer, finite yield is only obtained at VLSI levels if this cut-off value is some four or five standard deviations away from the mean threshold voltage. Current ion-implantation technology seems capable of achieving no better than 50 mV as the standard deviation of the threshold voltage [18–20], whereas epitaxial layers achieve typically 275 mV, although by an ingenious etching technique this can be improved to 75 mV [21]. Uniformity of ion implantation over small areas seems adequate to a 30 mV standard deviation of threshold [18], showing the promise of improvements in parameter control in the future. However, these figures are not promising at the moment for the VLSI prospects of enhancement-mode GaAs circuits.

As a result more complex circuits have been designed which can tolerate a 400–600 mV spread in threshold voltages [22]. Operation of GaAs at liquid nitrogen temperatures can be beneficial in several ways. The noise margins can be smaller because of the reduced thermal noise, yet the higher barrier height offers larger noise margins and logic swings. The electron mobility is greater at 77 K than at room temperature, and thus the propagation delay is reduced. These two effects combine to permit dynamic switching energies at 77 K to compare well with those at 300 K, despite the higher voltage swings [23]. Low-temperature operation of other structures (HEMTs and Josephson devices) offers dramatic improvements, but these benefits may be offset by the need for the provision of cooling plant.

4 What are the available devices?

The contenders which either are or have been in the running for gigabit logic systems are reviewed here, including discussion of the merits of the devices. Devices discussed in this Section are restricted to those which have already been integrated. Other devices are being investigated with a view to future integration – e.g. permeable-base and hetrojunction bipolar transistors – but for details on these and other devices

the reader is referred to more specialist literature (see for, example, References 119 and 120). A comparison of the results from the various logic families will be presented in Section 6, together with a discussion of future prospects. Looking first at what may be loosely termed 'conventional' transistor devices, both silicon and GaAs devices are to be found. Silicon circuits are discussed under two headings, corresponding to bipolar and MOS devices. GaAs circuits then follow, divided into three categories, namely depletion-and enhancement-mode MESFETs, and IGFETs. A discussion of the properties of the new electron-gas confinement FETs follows, before a brief summary of Josephson and transferred electron devices. The latter two are beginning to disappear from current discussions on practical gigabit logic devices.

4.1 Silicon bipolar junction transistors

Silicon bipolar technology was for a long time the leader in the world of high-speed logic, with the type of logic, termed nonthreshold logic, of which ECL is the best known example. Fig. 1 shows the basic ECL configuration, with variations offering lower power dissipation, but essentially the designs are similar. Reductions of device size, and variations of processing technology, have been employed to obtain smaller propagation delays [24–27], and projections suggest that even better results are possible [28–30]. A very ingenious design resulted in the fabrication of a versatile circuit suitable for use as a multiplexer or divider, capable of operating up to 2.8 Gbit/s [31]. The whole operation was controlled by a 'circulating travelling wave' timing unit. Most other circuits to date have been prescalers [27], but ECL is becoming increasingly available in the form of uncommited logic arrays (ULAs), or gate arrays. This technology has to date been restricted to upper MSI levels of integration because of the power consumption, but submilliwatt power levels per gate have been predicted. Being the only commercially available logic family capable of gigabit operation – dividers to 1.8 GHz are available in the Plessey series III ECL – and as it is the logic family used in present mainframe computer processors, silicon bipolar technology may be expected to be well in the running as a possible technology for larger-scale integration for future systems. An inverter using bipolar junction transistors will dissipate approximately the same power as one using MOSFETs of similar linewidths, when operated at maximum clock frequencies. However, the MOSFET power consumption can be greatly reduced in a 'standby' mode, whereas this is not true of bipolar circuits. Consequently, in an LSI circuit, for a given feature size, the power dissipation of a bipolar circuit is intrinsically higher than the equivalent MOS circuit. LSI levels of integration of ECL are nevertheless being pursued quite successfully.

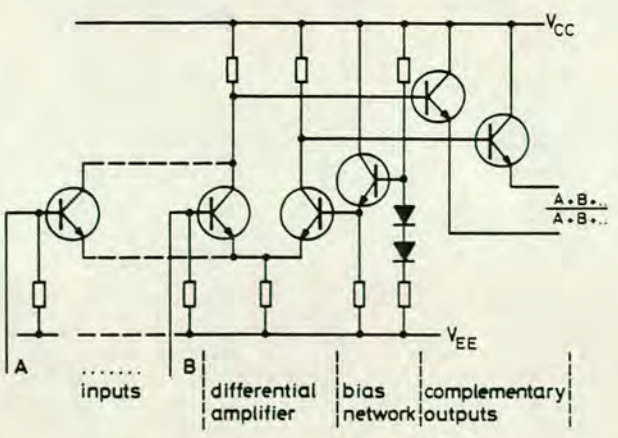


Fig. 1 *Basic ECL NOR gate circuit*

4.2 Silicon MOS

Silicon MOSFETs, already fabricated at VLSI complexities, have made a late appearance into the gigabit logic field. Nevertheless, that entry has been quite dramatic. As will be discussed later, the construction of MOS devices is such that the gate length can be quite small without having to align to very close tolerances, and thus very short source-drain spacings are possible. Recent results on MOSFETs with 0.4 μm gate lengths, obtained by workers at Bell Laboratories [113], have given propagation delays as low as 29 ps, and frequency divider operation at 2.5 GHz. Because of the relatively poor mobilities of holes in silicon it might be expected that CMOS devices would not operate at subnanosecond propagation delays. However, a recent paper [32] has reported 300 ps delays, at a power dissipation of 1 mW. This has been achieved by using a 1 μm geometry for the PMOS and 2 μm for the NMOS transistors. To achieve an increase in the carrier mobility, Nishiuchi *et al.* utilised a burried-channel MOSFET to obtain low propagation delays of 72.5 ps for fundamental gates, but with a high power dissipation of 43 mW [33]. This was achieved for 0.8 μm gate lengths. A maximum toggle speed of 1.64 GHz was reported for these devices, showing the marked loading effect found in MOS circuits.

More recently, work on scaled devices has yielded very-fast large-scale integrated circuits, with a 16-bit parallel multiplier capable of obtaining a product in 9 ns. However, there are several problems associated with the scaling of devices. It has been pointed out that for gate widths in excess of 20 μm – for 1 μm gate length – the gate must be considered as a distributed resistance/capacitance network. This is due to the relatively high resistance of polysilicon gates. The resulting time constant is similar to the switching speed of the device [33].

Table 5 shows first-order results of scaling device and circuit parameters (after Dennard [34]). Some of the more notable problems arise because the third dimension must be scaled, along with the length and width of devices, to obtain optimum results. This results in the need for a 20 nm thick oxide, and gives increased current densities and resistances in the tracks. In addition to the expected problems arising from the results of Table 5, there are short channel problems in MOS devices, arising from the dependence of threshold voltage on the back substrate bias, as discussed by Dennard [34, 35]. Also, the threshold voltage decreases markedly for a given implant dose as the source-drain spacing drops below 2 μm. For example, the standard deviation of threshold voltage is 100 mV, arising solely from a 0.3 μm variation in channel length from a nominal 1.3 μm. Such a figure is pessimistic as the value for standard deviation on a single wafer, because device lengths

Table 5: Scaling factors associated with reduction of FET dimensions

Parameter	Scaling factor
Dimensions	$1/K$
Doping concentration*	K
Voltage*	$1/K$
Current	$1/K$
Capacitance	$1/K$
Delay time	$1/K$
Power dissipation	$1/K^2$
Power delay product	$1/K^3$
Power density	1
Line resistance	K
Normalised voltage drop	K
Line response time	1
Line current density	K

*Note that the built-in voltage does not scale. Thus doping concentration is increased by a greater amount to compensate.

are likely to track with each other. However, wafer-to-wafer variations may prove intolerable. As devices are scaled more and more, further problems would be expected to arise. In a projection to 0.25 μm channel lengths Dennard points to the requirement for an oxide thickness of 6 nm. At the very high impurity levels required, low mobilities are obtained because of increased electron scattering. Yu [36] points out that the nonscaling of the subthreshold performance of MOSFETs limits the gate length to a minimum of 0.8 μm for satisfactory room-temperature operation of dynamic RAMs. The normally small subthreshold currents assume much more importance with submicron dimensions, such that the drain current is never pinched off.

One of the insurmountable problems resulting from scaling down is that the number of impurity atoms becomes quite small in a device channel. In a 0.25 x 0.25 μm FET there are only about 600 impurity atoms, giving rise to a significant statistical variation of device parameters, such that in 10^5 devices some will experience 20% shifts in the threshold voltage. Soft errors arising from radiation damage and charge trapping become more significant as the percentage shift in device characteristics due to each incidence becomes higher.

The results obtained by Boll *et al.* [113], with gates 0.4 μm long, indicate that this pessimism may be unfounded. Gate lengths as short as 0.1 μm were suggested. Silicon MOS will be a strong contender in gigabit electronics, because of the highly refined processing techniques which have been developed, the wealth of information about the material which is available to the designer, and its acceptance by system design engineers. An interesting comparison may be drawn between the development of high-speed silicon and GaAs. In GaAs development, results have been presented largely in terms of ring oscillator delays, and single-divider results. Boll's results on submicron gate length silicon devices are already at MSI level, and LSI complexity has been achieved for 1 μm gates, with 1 GHz clock frequency.

4.3 Depletion-mode GaAs MESFETs

GaAs MESFETs were first reported in integrated logic applications in 1974 [37]. These devices gave propagation delays as short as 60 ps, which, at the time, was an order of magnitude lower than anything else. The lack of a suitable native oxide, and the instability of the charge trapping at the GaAs-insulator interface, meant that earlier attempts to fabricate GaAs FET ICs had been unsuccessful. Because MESFETs (or metal Schottky FETs) utilise a Schottky-barrier gate electrode, the source-drain spacing is large compared with the gate length (see Fig. 2), unless a self-aligned technique is used. This construction may be considered as being a short-channel device with relatively high parasitic resistances.

All the early work on MESFET circuits involved depletion-mode devices, as the thin channels required for enhancement-mode FETs could not be fabricated reproducibly. Thus the circuits demonstrated by van Tuy *et al.* [37] consisted of logic gates with a level shifting buffer. The buffer requires dual power rails, in order that the output of one stage can be converted into the negative voltage levels required to drive the following stage (see Fig 3.) A significant fraction of the

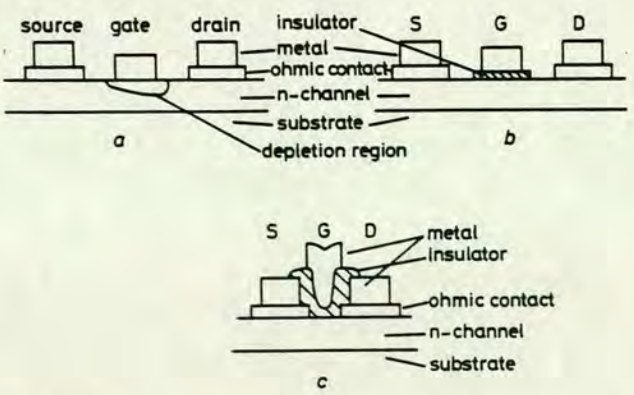


Fig. 2 *Schematic cross-sections of MOSFETs and MESFETs*
a Normally-on MESFET *b* IGFET (MOSFET) *c* Self-aligned MOSFET

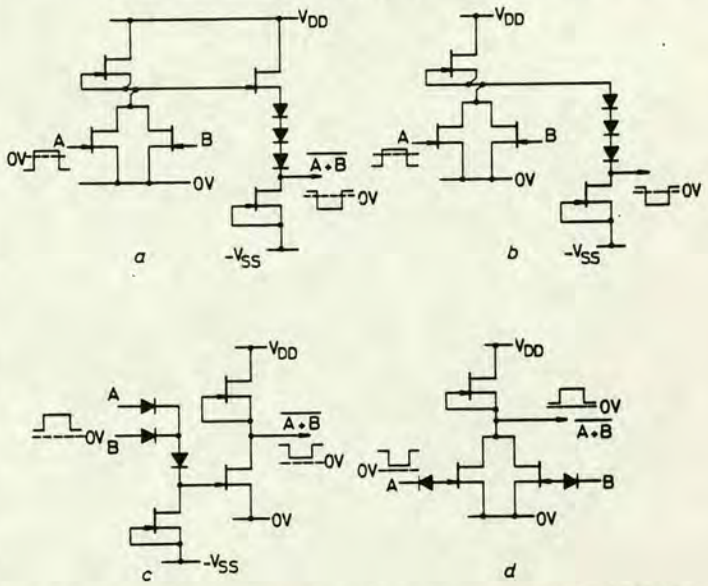


Fig. 3 *Nor gate circuit configurations using normally-on MESFETs*
a BFL *b* Low-power BFL *c* SDFL *d* CCL

power consumed by these so-called buffered-FET logic (BFL) circuits is used in the level-shifting circuit which is continuously carrying current. As a result the power consumption is some 40 mW per gate, and far from the low power dissipation required for LSI. Nevertheless, in terms of sheer speed, BFL is unmatched. The first frequency divider result offered 4.5 GHz frequency division [15, 38], and other workers have since achieved 5.5 GHz [39] and 5.7 GHz [40] division using complementary clocked dividers. Single-clocked dividers have achieved 3.5 GHz [39], and, more recently, 4.4 GHz [40]. The fastest reported D-type flip-flops have operated at 3 GHz [41]. All of these results have been achieved with gate lengths in the 0.6–1 μm region. Ring oscillator results on electron-beam patterned 0.5 μm gate lengths have yielded propagation delays as low as 34 ps [42].

Various studies have been carried out to reduce the power dissipation. Reduction of the number of level-shifting diodes, with corresponding increase (less negative) of pinch-off voltage, has yielded lower powers, but at the expense of speed. The elimination of the source follower of the buffer circuit can result in significantly reduced power dissipation [43]. Such a circuit is shown in Fig. 3*b*. As might be expected, the circuit is less tolerant to the increasing of the fan-out. Even with this reduced power circuit, the dissipation is some 12 mW per gate, with 60 ps propagation delay [44]. This limits BFL to MSI densities. Nevertheless, useful circuit results have been obtained. These include an 8:1 multiplexer circuit, designed for use in a 5 GHz pattern generator [44], and a 4-bit ALU [45]. Although IC complexity cannot be increased above this level, there appears to be a good future for BFL circuits at the very-high-speed MSI level of integration. In particular, basic BFL circuit design has been used in a wide-band amplifier, offering 26 dB gain over the frequency range 5 MHz to 3 GHz [46], showing that BFL technology is ideally

suited for the integration of combined analogue and digital components.

It was with the introduction of a totally different means of level shifting that gate powers were significantly reduced. In this circuit, shown in Fig. 3*c*, the logic is carried out by small diodes, with a pull-down transistor performing the level shifting. It is these diodes which give this logic family its name of Schottky diode FET logic (or SDFL). By utilising localised ion implantation [47] to produce shallow high-carrier-concentration channels for low-pinch-off high-transconductance transistors, and deep low-carrier-concentration active regions for the diodes, to produce low series resistance and capacitance, power dissipations as low as 0.3 mW were achieved. Propagation delays as low as 75 ps have been reported, with projections to 0.25 μm suggesting proagation delays of less than 20 ps with gate dissipation of 1.5 mW [114]. The completely planar technology achieved is a feature considered to be essential for high yields of VLSI circuits. Indeed, LSI levels of integration have already been achieved [10], but with very low yields. Among the SDFL circuits already fabricated and tested are:

Variable modulus dividers (60 gates) [48]
an 8 : 1 multiplexer (64 gates) [19]
a 3-bit multiplier (75 gates) [19]
a 5-bit multiplier (260 gates) [49]
an 8-bit multiplier using over 1000 gates [10].

However, the fabrication technology necessary for this type of circuit is especially demanding, with the fabrication of $1 \times 2\,\mu$m diodes, and only one report of work on SDFL, other than that of the Rockwell team, is known [12].

In a further attempt to reduce the power levels associated with level shifting circuitry, Livingstone *et al.* have utilised capacitor coupled logic (CCL) in which the coupling is carried out using reverse-biased Schottky diodes as capacitors [51, 52]. This type of coupling has the advantage that no power is consumed in the capacitors as, once they are charged up, the charge is merely transferred between the FET gate and the reversed diode. As this capacitor is in series with the gate capacitance, the loading capacitance is reduced. However, there is an increase in the stray capacitance associated with the relatively large area of the coupling diode. These circuits have the attraction of needing only a single supply rail. A disadvantage, however, is that they will not operate below about 20 kHz [53]. In many applications this is inconsequential – e.g. in a telecommunications environment there are already system constraints which demand a small mark/space ratio. A further advantage is the tolerance to variations in device parameters, with circuits working satisfactorily for pinch-off voltages from -0.5 to -4 V [53]. This method of coupling has been used independently as SSFL (Schottky-coupled Schottky-barrier FET logic), with propagation delays as low as 55 ps and power levels of 3.5 mW [54, 55]. GaAs IGFRTs have also been studied in terms of capacitor coupled logic [56, 57].

4.4 *Enhancement-mode GaAs FETs*

Enhancement-mode FET circuits on GaAs form a family of logic termed direct-coupled FET logic (DCFL). Normally-off or enhancement-mode operation occurs because the thickness of the FET channel is less than the thickness of the zero-bias depletion region, such that, with zero gate bias, the channel is pinched off. As the gate bias is made more positive, the current flows in the channel as the depletion region becomes narrower. Direct coupled logic operation is permitted because all the voltages in the circuit are positive with respect to the source. Two principal types of enhancement mode FETs are used, namely MESFETs and JFETs. The threshold voltage sets the lower logic level, and is typically chosen to be 0.2 V, while the onset of gate conduction sets the high logic level. For MESFETs this is typically 0.6–0.8 V, whereas JFETs have a higher built-in voltage, offering 0.9–1.1 V logic high level. Thus JFETs have the advantage of 0.9 V logic swing, compared with 0.5 V for MESFETs. However, the higher logic swing is only achieved at the expense of a more complex fabrication procedure. It was shown in Section 3 that low logic amplitudes are an advantage in terms of power dissipation, but a disadvantage in terms of noise margins. The final choice of amplitude must be a compromise between the two.

In the fabrication of normally-off MESFETs the channel thickness must be reduced to about 800 Å, but this thickness must be controlled very accurately, to give a tight control of threshold voltage. It has been shown [58] that the dense surface states in GaAs produce a depletion layer at the exposed surface, resulting in abnormally high parasitic resistances for MESFETs fabricated in the normal way on such thin channels. The reduction of the parasitic gate-source resistance poses a second problem in the fabrication of enhancement-mode devices. Several techniques have been used to overcome these two problems.

To control the thickness of the channel accurately, anodic oxidation has been used extensively to remove a known amount of material [23, 59, 60]. A more sophisticated anodic oxidation technique has been developed by Mun *et al.* [21]. Here the etching takes place such that the density of carriers per unit surface area in the channel becomes constant, independent of localised variations in doping levels or initial epitaxial material thickness. Thus threshold voltage variations of less than 75 mV are achieved, despite variations of 275 mV in the starting material. To reduce the parasitic resistance, Ido [60] has used anodic oxidation to produce a uniform 1400 Å thick channel, but then removed a further

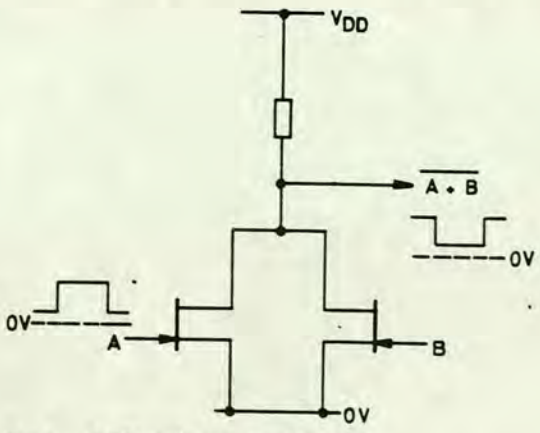


Fig. 4 *DCFL NOR gate configuration*

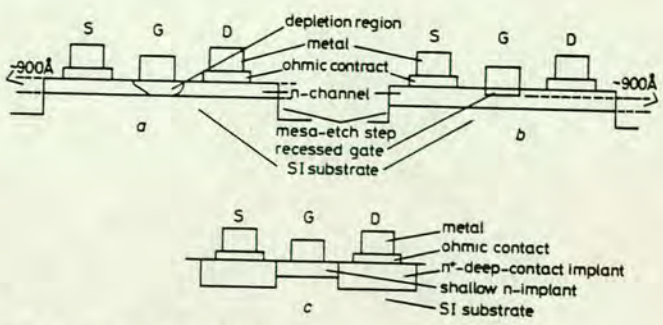


Fig. 5 *Normally-off MESFET construction*

a Mesa technology *b* Mesa technology with recessed gate
c Fully ion-implanted technology

500 Å by chemical etching. This was carried out after the gate patterning, and before the deposition of metal, resulting in thicker material between the gate and drain as compared with the channel region, and so reducing the parasitic resistance. This introduction of chemical etching after the anodic oxidation is likely to lead to a significant variation in channel thickness. (Ido reports less than 10% variation in channel thickness, which implies 150 mV variation in threshold voltage, around the nominal 0.1 V.) A 50% increase in drain current is reported as a result of using thicker starting material, and recessing the gate by 1200 Å instead of 200 Å, but no indication is given as to the effect on the variation of threshold voltage [59].

An alternative to the use of chemical etching for recessing the gate has been considered by Hojo *et al.* [61]. They have used Pt as the gate metal, which they have sintered at high temperatures, thus driving the Pt-GaAs interface into the active layer. A standard deviation of 153 mV has been achieved for a nominal threshold voltage of 0.06 V. Both of the problems mentioned above have been overcome by the use of this technique. The use of two different implant energies and doses also permits the fabrication of a thin transistor channel, while using a heavier implant underneath the contacts to reduce the contact resistance. Utilising a source-gate separation of only 0.4 μm for 8 μm gate lengths, propagation delays of 66 ps have been achieved [20]. At the same time parameter variations have been kept quite small. A self-aligned process has been reported in which the spacing has been reduced to 0.1 μm, offering RO delays of 39 ps for 0.8 μm gate lengths [62], and 20 ps for 0.5 μm gates [122]. This is a factor of four improvement on previously reported results for DCFL circuits. An earlier attempt at producing self-aligned gates resulted in 50 ps RO delays. Here the parasitic resistances were reported to be as low as 25 Ω for a 30 μm contact. This technique required the use of thermally stable gate metallisations, capable of withstanding the ($>$ 700° C) anneal temperatures [63]. Other work* on these refractory metal (W or Mo) gates suggests that yields may be very low. The new technique has obviated the need for the high-temperature anneal with the gates in position, but requires more fabrication steps, with consequently reduced yield. A new refractory gate metallisation, TiW silicide, appears to solve the problems of Schottky-barrier instability [115]. It is reported that a standard deviation of 40 mV is achieved for a nominal 0.16 V threshold, allowing successful operation of a 1 K static RAM – the largest GaAs IC fabricated to date.

Similar problems arise in the fabrication of JFETs. The channel thickness is controlled by the difference between the initial thickness of the *n*-channel and the thickness of the *p* region. In early devices, the *p*-region was formed by diffusion of Mg into the GaAs. The difficulty in controlling channel thickness is then one of controlling diffusion depth. One advantage of these devices is that the threshold can be modified by a further drive-in diffusion after the fabrication of the gates. However, this does not offer localised control. Hence poor uniformity may not be compensated for in this way. Only relatively long gates ($\sim$ 3 μm) have been fabricated by this technique [64]. Parasitic resistance is not as great a problem in JFETs as it is in MESFETs, as there is necessarily a gate recess, by the thickness of the *p*-region. However, attempts have been made to further reduce the contact resistance of a JFET, with the so-called column gate FET – see Fig. 6. Here the normal profile of the *p*-region Fig. (6*a*) has been replaced with the semicircular one of Fig. 6*b*. Ring oscillator data give a respectable 45 ps for a 2 μm gate length [65]. Difficulties of fabrication are not descussed however.

The use of ion implantation has enabled much shorter gate lengths to be produced, resulting in RO delays as low as 45 ps for a relatively wide, 1.3 μm long transistor [66]. A more realistic (in terms of LSI) gate width of 10 μm has yielded propagation delays in the 65–85 ps region [66–68]. These low figures reflect the higher logic swing available, when compared with MESFET results. No data on the variations of device parameters are as yet available for JFETs.

In recognition of some of the weaknesses of enhancement-mode FETs, various circuit configurations have been produced to counteract them. Early work used epitaxial material, and it was not easy to use active loads for the inverters (a problem not encountered with normally-on devices). The standard inverter can be characterised well in depletion-mode circuits,

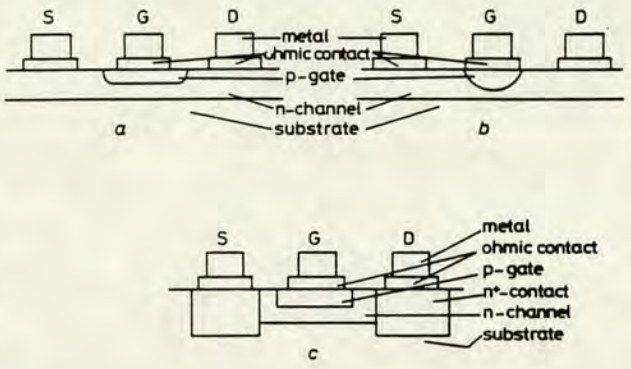


Fig. 6 *GaAs JFET construction*

a Diffused gate JFET *b* Column gate JFET *c* Fully ion-implanted JFET

because variations in device parameters are offset by the tracking effect of the load and source transistors. However, when different devices are used for switch and load, the device parameters do not track. The simplest load is a resistor, formed from channel material, but with no gate metal. Thus, both switch and load characteristics vary with channel thickness, but in different ways. The use of resistive loads gives poor rise times because of the low current available to charge the capacitance as the capacitor voltage approaches the supply voltage. Although fall times may be small, there is a high effective propagation delay. Zuleeg investigated the effects of using resistive, depletion- and enhancement-mode load devices [17, 69]. Fig. 7 shows these various approaches. The depletion load gives the fastest RO operation, but at the expense of more power than the resistive load. When the fan-out is large, the quasicomplementary circuit offers the fastest switching times, giving a significant rise time improvement. It is envisaged that in practical circuits, a variation of the designs will be used, depending on the fan-in and fan-out. The quasicomplementary circuit is impractical for fan-ins greater than two.

A more interesting approach to the solution of the problems associated with enhancement devices has been adopted by Nuzillar *et al.* [22, 70–72]. By increasing the complexity of the circuit design (Fig. 8) they have managed to shift the optimum threshold voltage from 0.2 to 0 V. In addition, the permissible variation of the threshold has been increased from 50–100 mV to 250 mV, making this family much more suited to LSI or VLSI. Fig. 8 shows six circuit configurations considered [70], although later work has been confined to the first four of these designs. As the logic high level is fixed by the onset of conduction at the gate of the loading transistor, the incorporation of the series diode enables the logic swing to be higher, as there is a voltage drop across this diode. The various designs are obtained by associating this diode with

*ALLAN, D.A: Unpublished

either the input or the output of the stage. Closer inspection reveals that four of these circuits are not new. The circuit of Fig. 8*a* is obtained by reducing the number of level shift diodes of BFL to one, while at the same time eliminating the negative supply. That of Fig. 8*b* is the power reduced BFL circuit of Reference 43. Fig. 8*d* reproduces the standard SDFL circuit with the negative bias removed, and that of Fig. 8*e* is the quasicomplementary circuit with the incorporation of a series diode.

The use of these circuits constitutes a significant achievement, as some of the power advantage of DCFL is traded for a greater tolerance to device parameter variations. They have been named low pinch-off FET logic (LPFL) because they can tolerate pinch-off voltages near zero, which are either positive or negative. The additional circuit complexity involved presents no real problems, as, like BFL and SDFL, multilevel logic can be achieved [71]. This practice is not normally adopted with either Si or GaAs normally-off devices.

In addition to ring oscillators and dividers, enhancement-mode devices are being used in larger circuits. Several workers are investigating the use of enhancement-mode circuits for static RAMs, with an immediate systems application for such circuits [74, 75].† A 256-bit static RAM has been reported, requiring only 9.4 mW, and with an access time of 50 ns [116]. A 1 K RAM has been fabricated, but no performance parameters are reported [115]. It is possible that larger (4K), and faster (1 ns) RAMs will be available by mid-1983.

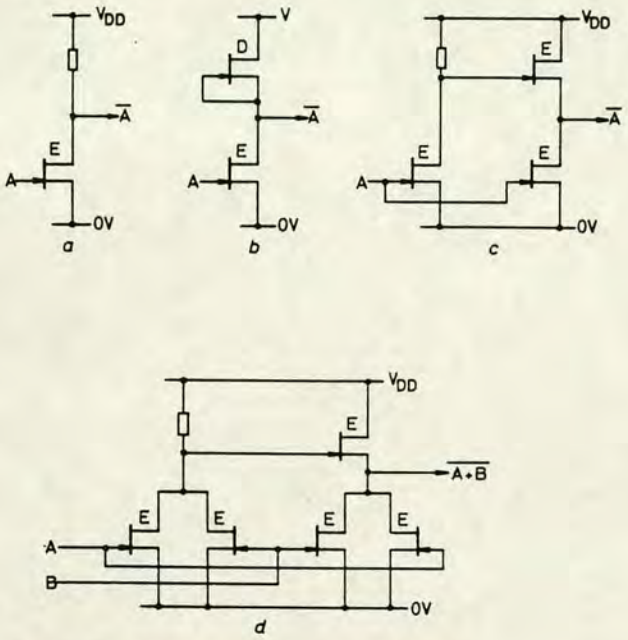


Fig. 7 *Load configurations for DCFL inverters*

a Resistor *b* Depletion-mode FET
c Enhancement-mode FET in quasicomplementary circuit
d Circuit of *c* extended to NOR gate

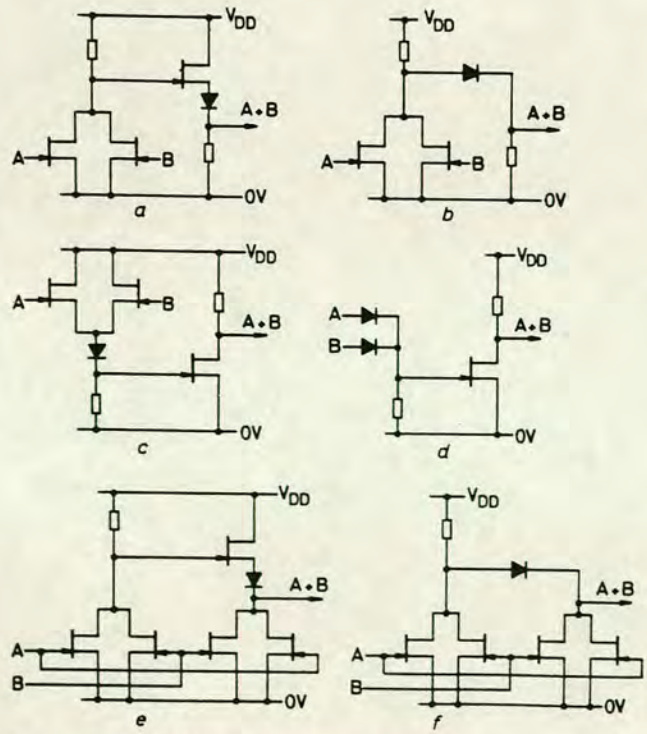


Fig. 8 *Various LPFL NOR gate configurations*

4.5 *GaAs MOSFETs (or IGFETs)*

Until recently GaAs MOSFETs would not have been included in a review of devices for gigabit logic applications. This is because of the instability of the oxide-semiconductor interface, which gives hysteresis in the FET transfer characteristics if the circuit time constants are long compared with those of the interface states. Thus DC performance of the circuits is difficult, if not impossible to predict. As has already been mentioned, DC operation of the devices is not essential, indeed is not available in CCL [51]. Of more concern with MOSFETs is the long-term stability of the interface, and the reproducibility of device characteristics, especially from wafer to wafer. Little has been reported to date on these phenomena. Schuermeyer has utilised the hysteresis of the interface states to produce 'electrically settable' IGFETs, which he analyses in terms of capacitor coupled MESFETs [56, 57].

Yokoyama [76] and Yamaguchi [77] independently point out the theoretical advantages of MOSFETs over MESFETs in GaAs. MOSFETs have a higher f_T and g_m, promising smaller propagation delays. In addition there are the advantages of much higher logic swings, and simpler fabrication – the gate needs not be formed within a gap between source and drain, but can overlap both. This assumes that the insulator can be easily formed or deposited. MOS logic gates have been succesfully fabricated in GaAs using a low-temperature plasma oxidation process. These devices do indeed offer the promise of high speed, with 72 ps delays reported [76, 78]. Unfortunately, no low-frequency studies have been reported on these devices as yet. Insulators other than native oxides are being considered with regard to integration of IGFETs on GaAs and other related materials. Compound semiconductors other than GaAs have been used for the fabrication of IGFETs [73, 80–85], or logic gates [85]. In these other materials – to date InP and GaInAs – there are less problems associated with interface states than there are in GaAs, but they are less stable than silicon MOSFETs, leading to DC instability. However, as these devices are very far from anything other than SSI, they will be discussed in relation to their material properties in Section 5.

4.6 *HEMTs*

A new device has recently emerged, termed variously a high-electron-mobility transistor (HEMT), a two-dimensional electron-gas FET (TEGFET), a modulation-doped FET, or a heterojunction FET (HJFET). Both enhancement- [86] and depletion- [87] mode devices can be fabricated. These devices directly replace other FETs in logic circuits. A separate Section has been devoted to them simply because they possess very important, unique properties.

† ZULEEG, R.: Private communication

Various geometries have been suggested for HEMTs, the basic concept being the fabrication of a very thin layer of doped semiconductor adjacent to a thin layer of undoped semiconductor of differing bandgap (see Fig. 9). Because of the relative ease of fabrication of lattice-matched heterojunctions between GaAs and GaAlAs, most devices have used this combination, although GaInAs has been used as the ternary [88]. As shown in Fig. 9*b*, the band bending at the

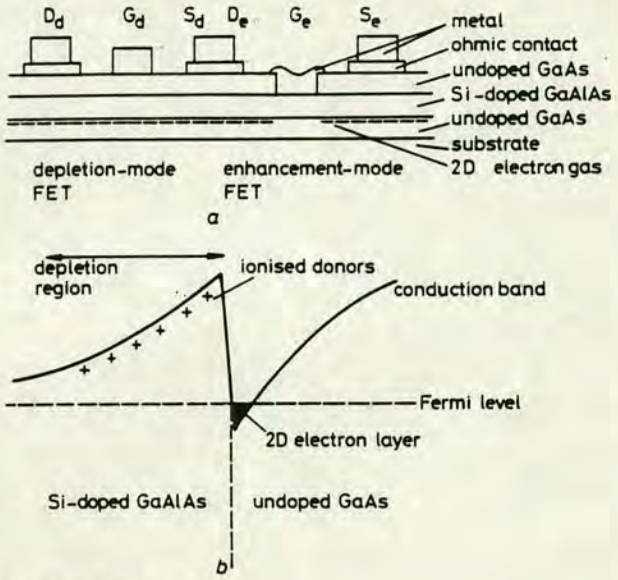


Fig. 9 *HEMTs*

a **Schematic inverter structure showing enhancement and depletion-mode FETs [92]**

b **Energy-band structure showing the origin of 2D electron layer**

heterojunction produces a well, in which the free carriers from the GaInAs become trapped. This trapping is only one-dimensional, resulting in a two-dimensional sheet of 'electron gas'. Because the electron gas is confined to move in two dimensions, electron-phonon interactions are confined to a circular shell in k-space, rather than the normal spherical shell for unconfined electrons. Consequently, interactions are relatively rare in these devices, and the resultant electron mobility is very high. Only the high-energy phonons are involved in the collisions, and therefore these can be 'frozen out' by operating at reduced temperature [89–91]. Room-temperature mobilities as high as 8400 $cm^2\ V^{-1}\ s^{-1}$ have been achieved, and 10 K mobilities as high as 210 000 $cm^2\ V^{-1}\ s^{-1}$ have been reported. A Schottky-barrier electrode can be used to change the occupation of this 2D region, and the resulting FET exhibits the standards square-law characteristics. As seen in Section 3, the electron mobility of an FET has a direct influence on device speed, and to achieve mobilities as high as these (see Table 5), when compared with those of GaAs (2000–5000 at 300 K and only 6000 at 77 K), is of great value. Indeed, when these devices were applied to logic circuits, a RO result of 17.1 ps at 77 K was achieved [92]. This is a factor of two lower than other FET circuits reported [42], but the gate length was some three times longer, at 1.7 μm.

More recent work has indicated that the use of GaAlAs Schottky barriers, as found in these HEMTS, gives further advantages over GaAs [93]. The barrier height is greater, allowing larger signal levels, and the surface potential is lower, reducing the parasitic resistance. These advantages have been incorporated into an LPFL type of circuit, yielding a staggering 19.1 ps propagation delay at room temperature, for 0.7 μm gate lengths [93], and more recently 18.4 ps, with 0.9 mW, or 30 ps at 0.06 mW [117]. Further progress seems possible, with the suggestion of a structure offering confinement into a 1D line. Electron mobilities could be as high as 3×10^8 $cm^2\ V^{-1}\ s^{-1}$ [94, 95]. Further advances are possible if quantum well structures are fabricated, because these would have useful optical properties in addition to the high electron mobilities. Although some very exciting results are appearing, only two teams have reported successful fabrication of HEMT ICs [92, 96, 117]. Much work is being carried out into the application of HEMTs, but there are as yet no available data as to the practicability in LSI. It is known that there is an extensive 7-year project under way in Japan towards this end. The promise of very fast devices with only moderate cooling – possibly none at all – is appealing. Josephson junction devices, which are the only other devices capable of operating at these speeds and power levels, require liquid helium operating temperatures. Fabrication is much the same as for other circuits, with the exception that the wafer must be grown as a modulation-doped structure. This requires molecular-beam epitaxy (MBE) or possibly metallorganic chemical vapour phase deposition (MOCVD) fabrication techniques, and it is uncertain as to the practicalities of this for production scale devices. The MBE technique ought to be capable of providing the reproducible layer thicknesses, provided that compositional variations can be eliminated. Such variations may ultimately limit the scale of integration of these devices. However, these devices are less than two years old, and they already appear to be very strong contenders.

4.7 Josephson junction devices

A brief overview of Josephson devices is presented here. For fuller treatment of device physics, circuits, systems and metallurgy, the reader is referred elsewhere (see e.g. Reference 107). Josephson devices offer very-low-power high-speed switching, and were considered ideal for high-speed computers [112] until recently, with IBM and Fujitsu being the major researchers. However, it appears that the progress in both silicon and GaAs has cast doubt on the future of Josephson circuits. One of the major problems is that they must be operated below the critical temperature at which the devices become nonsuperconducting; in practice this means operation at liquid helium temperatures. Although superconducting alloys with higher critical temperatures are available, the fabrication constraints impose restrictions on the choice of superconductor. Another major problem is that the structures must withstand thermal cycling between 4 K and 300 K. Emphasis has been placed on the use of thin-film tunnel junctions having an oxide only tens of ångströms thick, which are easily damaged by differences in thermal expansions. Van Duzer [108] points to the growth of 'hillocks' several thousands of ångströms high, caused by the release of thermally induced stresses. These will destroy the thin oxide film.

Fig. 10 shows a typical I/V characteristic for a tunnel junction. A voltage develops across the junction if a current in excess of the critical current is flowing. The transition time between the two states is about 1 ps. Two families of devices can be distinguished, those utilising an increase in current, and those using magnetic coupling as the gate signal. A simple circuit such as that of Fig. 10*b* will switch along a load line, as shown in Fig. 10*c*, and the circuit will latch when switched into the non-zero-voltage state. To switch the device off, an AC power source is used in these so-called latching circuits. As the design of Josephson circuits is unlike the design of semiconductor circuits, it is not dwelt on here. Suffice it to say that there are numerous confugurations available using both latching and nonlatching circuits. All of the standard logic elements can be fabricated, including several

flip-flop circuits, enabling LSI logic systems to be contemplated.

Logic voltage swings are small, with the gap voltage being about 2 mW, and capacitances are generally low, yielding fast very-low-power switching. A further advantage is to be found in that the transmission lines used for interstage coupling are dispersionless because of their superconducting nature.

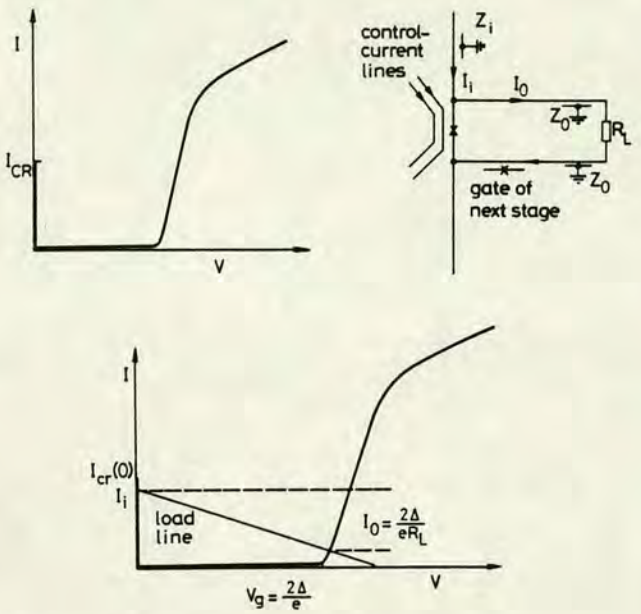


Fig. 10 *Josephson tunnel junctions*

a Typical *I/V* characteristics *b* Simple logic gate
c Switching transfer curve for *b*

Despite these advantages, it appears that the overriding factor is the need for operation in a liquid helium cryostat, which in many applications is an intolerable burden. In sheer speed, Josephson devices are not significantly faster than the latest HEMT results, operating at 77 K (or even 300 K) The differences from a systems point of view are that 77 K is easily achieved, but that 4 K is much more difficult. The extremely low power dissipation of Josephson devices must be offset against the power requirements for maintaining the low temperatures, and it appears that more conventional devices will be able to meet the immediate system requirements.

4.8 Transferred electron logic devices

Transferred electron logic devices (TELDs) have been used in logic circuits [109]. Great interest was shown in these devices in the early days of gigabit logic, because pulsewidths corresponding to the transition time of a single domain were possible. The early promise of these devices has been overshadowed by the rapid progress made in GaAs circuits, and the realisation that TELD thresholds had to be controlled within very tight margins to cascade the circuits. Additionally, TELDs for logic circuits are large, and consequently require high powers.

5 Silicon or GaAs?

Having presented a review of the types of circuit currently available for gigabit logic, it is pertinent to ask the question: 'Why put so much effort into GaAs, when silicon seems capable of meeting the requirements?' This question is a justifiable one, as the research effort in silicon has been so enormous over the last decade that it is now a well characterised material. The correspondingly small research effort in GaAs leaves very many questions still to be answered. It is perhaps easier to consider the question the other way around, and ask what GaAs offers that silicon does not.

In answer to this question, there are two parts:

(i) The electron mobility is higher in GaAs

(ii) GaAs can be easily fabricated with a semi-insulating substrate.

When work began on GaAs logic circuits, these two factors were considered to be sufficient motivation to move to a new material. It has been suggested that this move is conceptually no different to the move from PMOS to NMOS Si circuits. Here the higher carrier mobility was utilised at the expense of a more difficult fabrication procedure.

There has been much speculation as to how much is to be gained from the mobility advantage of GaAs. The low field mobility of GaAs is five or six times higher than that of silicon, but the saturated drift velocity is only a factor of two higher. Bosch [3] points to the fact that the saturated drift velocity should be used, because the devices are operated at relatively high fields. Eden, however, points to the fact that the devices operate in the low-field region for much of the time [7]. Experimental evidence seems now to point to the latter, with the temperature dependence of propagation delays being directly linked to that of low-field mobility [7, 92]. Thus the use of GaAs offers a fivefold advantage in terms of electron mobility. Bosch points out that the actual switching speed of the device is inconsequential, because the propagation delay of the circuit can be split into three components:

$$\tau_{PD} = \tau_{FET} + \tau_{LOAD} + \tau_{LINE}$$

This can be compared with eqn. 7. The intrinsic switching speed of the FET is a very small part of the whole. Such an argument conceals the fact that the factors involved in determining the actual switching speed of the device also control the transconductance of the FET, and hence the propagation delays of the load and line. Appendix 8.1 outlines the calculation used by Eden in determining the relationship between devices parameters and circuit switching speeds. The work on HEMTs has conclusively shown that high electron mobilities are important.

In a theoretical study of switching characteristics, Grubin compares the switching speeds of GaAs and Si, based on the expected electron transit times [97]. The results obtained are that at high bias levels the switching speeds are the same for normalised device dimensions, but that at low bias levels, GaAs is faster. However, the normallisation factors are different, such that the silicon FET must be less than half the length of the comparable GaAs FET. A similar study by Ino [98] reaches similar conclusions. When such studies are taken to smaller dimensions, a difficulty is reached, in that the concept of mobility in very short channels is not well defined [99], and the concept of improved transistor performance due to ballistic effects also needs to be discussed. Ballistic transport occurs where the device length is short compared with the mean free path of the carriers. In GaAs a negative differential mobility is observed, as the electrons are scattered from the central, into the six outlaying valleys. When the devices are short, this scattering will not occur, and the electron transit time corresponds to a higher electron mobility. Wada *et al.* have studied these effects in GaAs and InP. InP has a lower saturated drift velocity and a lower low-field mobility, but has a higher peak velocity [100]. Consequently, the f_T is potentially higher in InP MESFETs than in GaAs. The lower barrier height (~ 0.5 V) means that poor logic gate performance is expected. GaInAs offers higher speeds than GaAs for similar reasons (see Table 5). InP does have a stable insulator interface, and so advantage can be taken of its high f_T in MOS devices. These are being studied, as mentioned in Section 4.6.

Table 6: Properties of various semiconductors

Material	Hall mobility 300 K	77 K	Peak velocity	Schottky-barrier height
	$cm^2 V^{-1} s^{-1}$		$cm s^{-1}$	eV
p-Si	500			0.5
n-Si	1300		0.7×10^7	0.7
p-GaAs	400			
n-GaAs	4500	6000	1.8×10^7	0.8
InP	3000		2.3×10^7	0.5
InGaAs	7800		2.5×10^7	0.3
HEMT 2D Electron gas	$7-10 \times 10^3$	$30-100 \times 10^3$		

The second advantage which GaAs enjoys is that of an SI substrate. The substrate capacitance term of eqn. 7 is consequently reduced. However, there are problems still remaining in GaAs, and Si is making progress in this area. Isolation has been found to present problems at small geometries with GaAs, despite the semi-insulating substrate. Proton bombardment has been used to improve isolation [101]. There are also problems associated with conversion of a thin surface layer of the SI material during the high-temperature anneal after ion implantation. A mesa technology removes this surface layer, but results in a nonplanar technology.

Various advances in silicon processing offer an SI substrate. Silicon-on-sapphire circuits are well established, but are difficult to fabricate without lattice-matching problems. Oxide isolation is used in the now widespread ISOCMOS process. Other techniques are making rapid progress, such as the use of recrystallised polysilicon for the active areas [12, 13], or the use of neutron radiation as a means of damaging the substrate.

It was originally thought that silicon would be completely unable to compete with GaAs at gigabit rates. This has since proved to be fallacious, as silicon gas proved capable of working at high speeds, by the use of very small geometries. As silicon has proved itself capable of high speeds, are the other factors favouring silicon not sufficient to cause work on GaAs to be dropped? Bosch points to the factor of two difference in thermal conductivities, which limits GaAs to lower total power dissipations. In addition, silicon has a very stable native oxide and well characterised parameters. As indicated in Section 4, there is speculation as to whether dimensions can be reduced still further in silicon. Some workers suggest that silicon has now reached a fundamental limit, with further size reductions impossible, athough other workers maintain that this limit is far to the furure. It should be noted that comments have been made previosly concerning the end to improvements in silicon performance. If GaAs continues to be developed, it is not unreasonable to suppose that GaAs devices could be fabricated with the dimensions now being used for silicon, when, once again, the mobility advantage would come to the fore, with ballistic effects offering further advantages to GaAs. In the short term, of more significance is the fact that silicon does not have any notable optical properties. Thus monolithic integration of optical and electronic components must rely on the existence of a suitable technology in one of the direct-gap III-V compounds semiconductors. For short links, where GaAs lasers are suitable, GaAs logic would be required, and for longer transmission paths, where ternary or quaternary lasers would be used, GaInAs or InP offers the prospects of high speeds

Although it was once thought that silicon could not achieve gigaherz clock rates, this is now no longer the case, and GaAs is now being challenged. Silicon gas long had LSI capability, and scaling to smaller dimensions offers high speed LSI. GaAs has always had the speed advantage, and is only just beginning to realise LSI densities. There still seems to be the need for GaAs logic, and if sufficient effort is available to achieve larger scale integration, it will complement the silicon circuits, offering better prospects for still higher speeds.

6 Results projection and conclusion

Having so far dealt with the various approaches to gigabit logic, the results are discussed here, with a brief review of the merits of the various approaches. A projection is then produced, based on that of Bosch [1].

Tables 1 and 2 compare the best available data on ring oscillator results in terms of propagation delays and dynamic switching energies. Table 7 summarises the current data on the maximum clock rates of divide-by-two circuits. A comparison of Tables 1 and 7 shows the anomaly which is created by using RO data alone. Recent results on enhancement-mode circuits have produced some very good RO data. However, Table 3 shows how fan-in and fan-out variations affect the speed of the circuits in more realistic situations such as those represented tn Table 7. These results represent the fact that BFL circuits are not greatly slowed down by leading, whereas those of MOS and normally-off circuits are slowed much more radically. A further fact reflected in these results is the relative maturity of BFL circuits when compared with DCFL circuits.

In assessing the LSI and VLSI prospects of the various configurations, the problems outlined in Section 3 need to be considered, together with the progress which has been shown by the individual circuit types. In the near future it seems certain that silicon MOS will capitalise on the wealth of

Table 7: Maximum clock frequencies of divide-by-2 circuits

Clock frequences	Power	Dimensions	Circuit type	Reference
GHz	MW	μm		
5.7	240	0.8	GaAs – dual-clocked BFL	40
5.5	160	0.6 × 30	GaAs – dual-clocked BFL	39
4.5	160	1 × 20	GaAs – dual-clocked BFL	15
4.4	440	0.8	GaAs – single-clocked BFL	40
4.1	330	1.2 × 75	GaAs – dual-clocked BFL	102
3.5	400	0.6	GaAs – single-clocked BFL	39
3.4	30	0.6	GaAs – E MESFET	103
3.0	230	0.7 × 20	GaAs – D-type BFL	76
2.8	60	0.75 × 20	GaAs – dual-clocked LPFL	72
2.5		0.4	Si MOS – 4-stage binary	113
2.4	54	1.2 × 80	GaAs – E MESFET dual clocked	59
2.0	19	1.2 × 20	GaAs – dual-clocked SSFL	54
2.0			Si bioplar	24
1.9	12	1	GaAs – D-type SDFL	7, 103
1.8			Si bipolar	24
1.6	85		Si MOS	33
1.6			Si bipolar – circulating travelling-wave timer	31
1.4	8.3	1.2 × 20	GaAs – E MESFET dual-clocked	59

knowledge already existing, to enable it to advance further in high-speed submicron devices.

BFL, having reached MSI complexities, is unlikely to offer larger integration levels, but will capitalise on its speed at these levels. The current development is in the fabrication of analogue and digital circuits. Work is being carried out with CCL circuits to MSI level, and this looks promising. Although lower LSI would be possible in terms of power, packing density may present problems. In addition, design of the circuits is made more complex because of the limitations imposed by the need for charging the coupling capacitors before correct logic operation ensues.

Although projections (see Fig. 10) show enhancement-mode circuits reaching well into VLSI, there are problems associated with variations in device parameters, requiring improvements to be made in the areas of material quality and fabrication. LSI densities have now been reached, but until these problems associated with the materials aspects of GaAs have been solved, VLSI levels seem unlikely. It is expected that these materials problems will be solved over the next few years. The most promising way to VLSI among the remaining technologies seems to be the LPFL circuit, where high speed and low power are traded for increased circuit tolerances.The interface problems associated with MOSFET circuits in GaAs are by no means resolved, and much more work is necessary in this area.

In conclusion, it seems certain that gigabit electronics will continue to expand. The appearance of silicon MOS already at VLSI levels, and now at high speeds, will be a key factor in the future. If it continues its progress smoothly to the 0.5 μm and 0.25 μm geometries on a production basis, which is not without its problems, then the rate of expansion of GaAs devices will be adversly affected. Given the choice of functionally identical components in GaAs and silicon, the systems designer will choose the silicon component. However, if this is the case, the death of GaAs logic is not indicated, as the way forward is then fourfold:

(i) the continued development of the material and technology, with the fabrication of logic circuits of increasing complexity, but with no immediate outlet

(ii) the development of other III-V compound semiconductors along the lines of (i)

(iii) the cofabrication of logic circuits with optoelectronic components

(iv) in the longer term, GaAs devices will be able to surpass the performance of silicon circuits, taking advantage of the developments made in (i), and those now being made in the fabrication of submicron silicon.

Even without the advances in silicon technology, GaAs circuits are likely to develop along the lines of (i)-(iii). In the immediate future, there is little to choose between GaAs and silicon, for general applications. There will always be specialised applications such as (iii) which will require the special properties of GaAs and its related materials. Where smaller scales of integration are acceptable, but higher speeds are required, BFL circuits are unlikely to be matched.

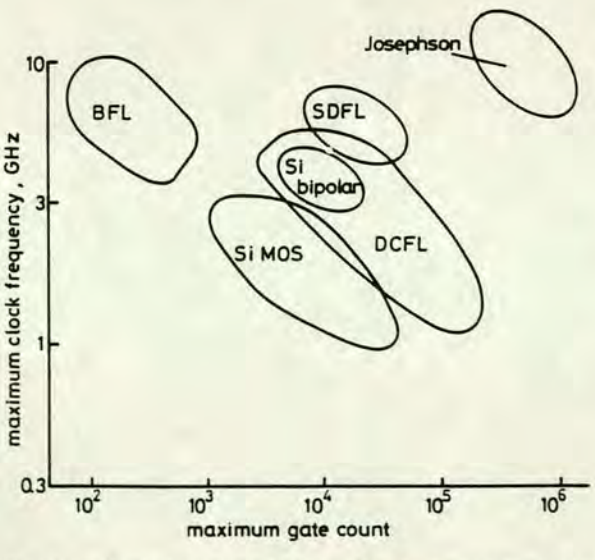


Fig. 11 *Projection of future limits of various logic families [1]*

7 Acknowledgments

Permission to publish this work has been granted by the Director of Research, British Telecom. The author acknowledges the encouragement and assistance of colleagues from the Electrical Engineering Department of Edinburgh University and from BTRL.

8 References

1 BOSCH, B.: 'Device and circuit trends in gigabit logic', *IEE Proc. I, Solid-State & Electron Devices*, 1980, **127**, (5), pp. 254–265
2 MICHEEL, L.J., and RABANUS, G.G.: 'Integrated gigabit logic – State of the art assessment and performance projections'. 1st Microwave Symposium Digest, 1978, pp. 50–53
3 BOSCH, B.: 'Gigabit electronics – A review', *Proc. IEEE*, 1979, **67**, pp. 340–379
4 *JEE*, Sept. 1975, pp. 28–29
5 ALBRECHT, W., BAACK, C., ELZE, G., ENNING, B., HEYDT, G., PETERS, K., WALF, G., and WENKE, G.: '2.24 Gbit/s optical transmission system at 0.85 μm wavelength', *Electron. Lett.*, 1981, **17**, (18), pp. 664–666
6 CARTER, A.C., FORBES, N., and GOODFELLOW, R.C.: 'Monolithic integration of optoelectronic, electronic and passive components in GaAlAs/GaAs multilayers', *ibid.*, 1982, **18**, (2), pp. 72–74
7 EDEN, R.C., WELCH, B.M., ZUCCA, R., and LONG, S.I.: 'The prospects for ultrahigh-speed VLSI GaAs digital logic', *IEEE J. Solid-State Circuits*, 1979, **SC-14**, pp. 221–239
8 WILSON, D.O., TAYLOR, D.M., and PHILLIPS, D.H.: 'High speed SOS and GaAs IC testing in the 1980s', *Solid-State Technol*, June 1981, pp. 95–98
9 HASEGAWA, H., TETSUTANI, N., and SAWADA, T.: 'Transmission properties of microstrip and coplanar strip lines for GaAs digital and analog integrated circuits'. GaAs IC symposium, 1980
10 LONG, S.I., WELCH, B.M., ZUCCA, R., and EDEN, R.C.: 'Gallium arsenide large scale integrated circuits', *J. Vac. Sci. & Technol.*, 1981, **19**, pp. 531–536
11 LEPSELTER, M.P.: 'Silicon picocircuits'. GaAs IC symposium, 1981
12 MABY, E.W., GEIS, M.W., LECOZ, Y.L., SILVERSMITH, D.J., MOUNTAIN, R.W., and ANTONIADIS, D.A.: 'MOSFET's on silicon proposed by moving melt zone recrystallization of encapsulated polycrystalline silicon on an insulating substrate', *IEEE Electron Device Lett.*, 1981, **EDL-2**, pp. 241–243
13 KAMINS, T.I., and VON HERZEN, B.P.: 'MOSFET's in electron-beam recrystallized polysilicon', *ibid.*, 1981, **EDL-2**, pp. 313–315
14 ALEXOPOULOS, N.G., MAUPIN, J.A., and GREILING, P.T.: 'Determination of the electrode capacitance matrix for GaAs FET's', *IEEE Trans.*, 1980. **MTT-28**, pp. 459–466
15 VAN TUYL, R.L., LIECHTI, C.A., LEE, R.E., and GOWEN, E.: 'GaAs MESFET logic with 4-GHz clock rate', *IEEE J. Solid-State Circuits*, 1977, **SC-12**, pp. 485–496
16 INO, M., HIRAYAMA, M., and OHMORI, M.: 'Analysis for optimum threshold voltage and load current of E-D-Type GaAs DCFL circuits', *Electron. Lett.*, 1981, **17**, (15), pp. 522–523
17 ZULEEG, R., NOTTHOFF, J.K., and BEHLE, A.F.: 'Gigabit logic prospects of GaAs E-JFET integrated circuits', IEEE MTT-S International Microwave Symposium Digest, 1979, pp. 512–515
18 ZUCCA, R., WELCH, M., EDEN, C., and LONG, S.I.: 'GaAs digital IC technology/statistical analysis of device performance', *IEEE Trans.*, 1980, **ED-27**, pp. 1109–1115
19 WELCH, B.M., SHEN, Y.D., ZUCCZ, R., EDEN, R.C., and LONG, S.I.: 'LSI processing technology for planar GaAs integrated circuits', *ibid.*, 1980, **ED-27**, pp. 1116–1124
20 FURUTSUKA, T., TSUJI, T., KATANO, F., HIGASHISAKA, A., and KURUMADA, K.: 'Ion-implanted E/d-type GaAs IC technology', *Electron. Lett.*, 1981, **17**, (25), pp. 944–945

21 MUN, J., PHILLIPS, J.A., BARRY, B.E.: 'High-yield process for GaAs enhancement-mode MESFET integrated circuits', *IEE Proc.* I, *Solid State & Electron Devices*, 1981, **128**, (4), pp. 144–147

22 NUZILLAT, G., BERT, G., DAMAY-KAVALA, F., and ARNODO, C.: 'High-speed low-power logic IC's using quasi-normally off GaAs MESFET's', *IEEE J Solid-State Circuits*, 1981, **SC-16** pp. 226–232

23 MIZUTANI, T., KATO, N., IDA, M., and OHMORI, M.: 'High speed enhancement-mode GaAs MESFET logic', *IEEE Trans.*, 1980, **MTT-28**, pp. 479–483

24 BENNETT, J.H., HOLLOCK, S., and WARD, P.J.: 'Sub-nanosecond bipolar LSI'. 1st ESSCIRC extended abstracts, 1975, pp. 34–35

25 SUDO, T., KAMOTO, T., MUKAI, H., SASAKI, H., SHIBA, H., and MORINO, A.: 'A monolithic 8 pJ/2 GHz logic family', *IEEE J. Solid-State Circuits*, 1975, **SC-10**, pp. 524

26 SAKAI, T., SUNDHARA, Y., NAKAMURA, H., and SUDO, T.: 'A 100 PS bipolar logic'. IEEE international solid state circuits conference, 1977, pp. 196–197

27 SHIMIZU, S., KOMATSU, S., and TORII, K.: 'Multi-collector ECL circuit family and its application to 1 GHz prescaler IC'. International Electron Device Meeting Technical Digest 1980, pp. 386–380

28 TANG, D.D., and SOLOMON, P.M.: 'Bipolar transistor design for optimized power-delay logic circuits', *IEEE J. Solid-State Circuits*, 1979, **SC-14**, pp. 679–684

29 WARD, P.J., and HOLLOCK, S.: 'A 200 ps silicon bipolar logic capability'. Colloquium on high speed logic, Oct. 1980

30 KNORR, S.G.: 'The potential of biopolar devices in LSI gigabit logic', *IEEE Circuits & Syst. Mag.*, 1981, **3**, pp. 2–6

31 HUGHES, J.B., COUGHLIN, J.B., HARBOTT, R.G., VAN DEN HURK, T.H.J., and VAN DEN BERGH, B.J.: 'A versatile ECL multiplexer IC for the Gbit/s range', *IEEE J. Solid-State Circuits*, 1979, **SC-14**, pp. 812–817

32 ZIMMER, G., FIEDLER, H., HOEFFLINGER, B., NEUBERT, E., and VOGT, H.: 'Performance of a scaled Si Gate *n*-well CMOS technology', *Electron. Lett.*, 1981, **17**, (18) pp. 666–667

33 NISHIUCHI, K., SHIBAYAMA, H., NAKAMURA, T., HISATSUGU, T., ISHIKAWA, H., and FUKUKAWA, Y.: 'A gigabit MOS logic circuit with buried channel MOSFET's', *IEEE J. Solid-State Circuits*, 1980, **SC-15** pp. 809–816

34 DENNARD, R.H.: 'Technology challenges for ultrasmall silicon MOSFETS', *J. Vac. Sci. & Technol.*, 1981, **19**, pp. 537–539

35 DENNARD, R., GAENSSLEN, F.H., YU, H.N., RIDEOUT, V.L., BASSOUS, E., and LeBLANC, A.R.: 'Design of ion-implanted MOSFET's with very small physical dimensions', *IEEE J. Solid-State Circuits*, 1974, **SC-9**, pp. 256–268

36 YU, H.N., REISMAN, A., OSBURN, C.M., and CRITCHLOW, D.L.: '1 μm MOSFET VLSI technology: Part 1 – An overview', *IEEE Trans.*, 1979, **ED-26**, pp. 318–324

37 VAN TUYL, R.L., and LIECHTI, C.A.: 'High-speed integrated logic with GaAs MESFET's', *IEEE J. Solid-State Circuits*, 1974, **SC-9**, pp. 269–276

38 VAN TUYL, R., LIECHTI, C., LEE, R., and GOWEN, E.: '4-GHz frequency division with GaAs MESFET ICs'. ISSCC, 1977, pp. 198–199

39 CATHELIN, M., GAVANT, M., and ROCCHI, M.: 'A 3.5 GHz self-aligned single-clocked binary frequency divider on GaAs'. *IEE Proc. I, Solid-State & Electron Devices*, 1980, **127**, (5), pp. 270–277

40 GLOANEC, M., JARRY, J., and NUZILLAT, G.: 'GaAs digital integrated circuits for very high-speed frequency division', *Electron. Lett.*, 1981, **17**, (20), pp. 763–765

41 GLOANEC, M., NUZILLAT, G., ARNODO, C., and PELTIER, M.: 'An E-beam fabricated GaAs D-type flip-flop IC', *IEEE Trans.*, 1980, **MTT-28**, pp. 472–478

43 BARNA, A., and LIECHTI, C.A.: 'Optimization of GaAs MESFET logic gates with subnanosecond propagation delays', *IEEE J. Soild-State Circuits*, 1979, **SC-14**, pp. 708–715

44 BALDWIN, G.. GOWEN, E., HENNING, F., LARRICK, R., NAMJOO, M., PODELL, A., and LIECHTI, C.: 'A GaAs MSI word generator operating at 5 Gbit/s data rate'. GaAs IC symposium, 1980

45 SUYAMA, K., KUSUKAWA, H., OKAMURA, S., YAMAMURA, S., and FUKUTA, M.: 'GaAs LSI for high speed data processing'. *Ibid.*, 1980

46 ESTREICH, D.B.: 'Microwave GaAs FET oscillators and amplifiers' 'SSCC, 1982

47 EDEN, R., WELCH, B.M., and ZUCCA, R.: 'Planar GaAs IC technology. Applications for digital LSI', *IEEE J. Solid-State Circuits*, 1978, **SC-13**, pp. 419–426

48 WALTON, E.R., LEE, F.S., SHEN, Y.K., and DIKSHIT, R.: 'High speed SDFL divider circuit'. GaAs IC symposium, 1981

49 LEE, F.S., SHEN, E., KAELIN, G.R., WELCH, B.M., EDEN, R. C., and LONG, S.I.: 'High speed LSI GaAs digital integrated circuits'. *Ibid.*, 1980

50 EDEN, R.C., and WELCH, B.M.: 'Low power depletion mode ion-implanted GaAs FET integrated circuits', *IEEE Trans.*, 1977, **ED-24**, pp. 1209–1210

51 MELLOR, P.J.T., and LIVINGSTONE, A.W.: 'Capacitor-coupled logic using GaAs depletion-mode f.e.t.s', *Electron. Lett.*, 1980, **16**, (19), pp. 749–750

52 LIVINGSTONE, A.W., and MELLOR, P.J.T.: 'Capacitor coupling of GaAs deflection-mode f.e.t.s', *IEE Proc. I, Solid-State & Electron. Devices*, 1980, **127**, (5), pp. 297–300

53 LIVINGSTONE, A.W., WELBOURN,A.D., and BLAU, G.L.: (to be published)

54 HASHIZUME, N., YANADA, H. KOJIMA, T., MATSUMOTO, K., and TOMIZAWA, K.: 'Low-power gigabit logic by GaAs SSFL', *Electron. Lett.*, 1981, **17**, (16), pp. 553–554

55 HASHIZUME, N., YANADA, H., and TOMIZAWA, K.: 'Schottky-barrier coupled Shottky-barrier gate GaAs FET logic', *ibid.*, 1981, **17**, (1), pp. 51–52

56 SCHUERMEYER, F.L., SINGH, H.P., SCHERER, R.L., and MAYS,D.L.: 'GaAs IGFET. A new device for high speed digital ICs'. IEDM Technical Digest, 1980, pp. 441–444

57 SCHUERMEYER, F.L.: 'GaAs IGFET digital integrated circuits', *IEEE Trans.*, 1981, **ED-28**, pp. 541–545

58 SPICER, W.E., CHYE, P.W., SKEATH, P.R., SU, C.Y., and LANDAU, I.: 'New and unified model for Schottky barrier and III-V insulator interface states formation', *J. Vac. Sci. & Technol.*, 1979, **16**, pp. 1422–1433

59 KATANO, F., FURUTSUKA, T., and HIGASHISAKA, A.: 'High speed normally-off GaAs MESFET integrated circuit', *Electron. Lett.*, 1981, **17**, (6), pp. 236–237

60 IDA, M., MIZUTANI, T., ASAI, K., UCHIDA, M., SHIMADA, K., and ISHIDA, S.: 'Fabrication technology for an 80-ps normally-off GaAs MESFET logic', *IEEE Trans.*, 1981, **ED-82**, pp. 489–493

61 HOJO, H., TOYODA, N., MOCH IZUKI, M., MIZOGUCHI, T., and NII, R.: 'Planar E/D type GaAs IC's by pt buried gate technology'. GaAs IC symposium 1981

62 YAMASKI, K., ASAI, K., MITZUTANI, T., and KURUMADA, K.: 'Self-align implantation for n^+-layer technology (SAINT) for high-speed GaAs ICs', *Electron. Lett.*, 1982, **18**, (3), pp. 119–121

63 YOKOYAMA, N., MIMURA, T., FUKUTA, M., and ISHIKAWA, H.: 'A self-aligned source/drain planar device for ultrahigh-speed GaAs MESFET VLSIs'. ISSCC, 1981, pp. 218–219

64 ZULEEG, R.: 'LSI prospects of GaAs FETS (logic configuration)', *Jpn. J. Appl. Phys.*, 1980, **19**, (19–1), pp. 315–318

65 ISHII, Y., ASAI, K., INO, M., KOMBAYASHI, M., and KURUMADA, K.: 'Direct-coupled logic operation of GaAs column gate FET', *IEEE Electron Device Lett.*, 1981, **EDL-2**, pp. 118–120

66 KATO, Y., DOHSEN,M., KASAHARA, J., TAIRA, K., and WATANABE, N.: 'Planar normally-off GaAs JFET for high speed logic circuits', *Electron. Lett.*, 1981, **17**, (25), pp. 951–952

67 KASAHARA, J., TAIRA, K., KATO, Y., DOHSEN, M., and WATANABE, N.: 'Fully ion-implanted GaAs ICs using normally-off JFETS', *ibid.*, 1981, **17**, (16), pp. 621–623

68 ZULEEG, R., NOTTHOFF, J.K., and TROEGER, G.L.: 'GaAs junction FET LSI'. EASCON 80, 1980, pp 146–150

69 ZULEEG, R.: 'Prospects of gigabit logic for GaAs FETS'. Institute of Physics Conference Series 53, 1980, pp. 75–81

70 NUZILLAT, G., BERT, G., NGU, T.P., and GLOANEC, M.: 'Quasi-normally-off MESFET logic for high-performance GaAs ICs', *IEEE Trans.*, 1980, **ED-27**, pp. 1102–1109

71 NUZILLAT, G., DAMAY-KAVALA, F., BERT, G., and ARNODO, C.: 'Low pinch-off voltage f.e.t. logic (l.p.f.l.): l.s.i. oriented logic approach using quasinormally off GaAs m.e.s.f.e.t.s.', *IEE Proc. I, Solid-State & Electron Devices*, 1980, **127**, (5), pp. 287–296

72 DAMAY-KAVALA, F., NUZILLAT, G., ARNODO, C.: 'High-speed frequency dividers with quasi-normally-off GaAs MESFETS', *Electron. Lett.*, 1981, **17**, pp. 968–970

73 YAMOTO, A., and UEMURA, C.: 'Anodic oxide film as gate insulator for InP MOSFETs', *Electron. Lett.*, 1982, **18**, pp. 63–64

74 KINELL, D.K., and WILSON, D.O.: 'GaAs static random access memory cell design'. GaAs IC symposium, 1981

75 BERT, G., MORIN, J.P., NUZILLAT, G., and ARNODO, C.: 'High speed GaAs static random access memory'. *Ibid.*, 1981

76 YOKOYAMA, N., MIMURA, T., and FUKUTA, M.: 'Planar GaAs MOSFET integrated logic', *IEEE Trans.*, 1980, **ED-27**, pp. 1124–1128

77 YAMAGUCHI, K., and TAKAHASHI, S.: 'Theoretical characterization and high-speed performance evaluation of GaAs IGFET's', *IEEE Trans.*, 1981, **ED-28**, pp. 581–587

78 YOKOYAMA, N., MIMURA, T., KASAKAWA, H., SUYAMA, K., and FUKUTA, M.: 'GaAs MOSFET high-speed logic', *ibid.*, 1980, **MTT-28**, pp. 483-486
79 HENRY, L., LECROSNIER, D., L'HARIDON, H., PAUGAM, J., PELOUS, G., RICHOU, F., and SALVI, M.: '*n*-channel MISFETs on semi-insulating Inp for logic applications', *Electron. Lett.*, 1982, **18**, pp. 102–103
80 LILE, D.L., and COLLINS, D.A.: 'Carrier profiling of InP', *ibid.*, 1978, **14**, (20), pp. 657–659
81 MEINERS, L.G., LILE, D.L., and COLLINS, D.A.: 'Inversion layers on InP', *J. Vac. Sci. & Technol.*, 1979, **16**, pp. 1458–1461
82 MEINERS, L.G., LILE, D.L., and COLLINS, D.A.: 'Microwave gain from an *n*-channel enhancement-mode InP MISFET', *Electron. Lett.*, 1979, **15**, (18), p. 578
83 TSENG, W.F., BARK, M.L., DIETRICH, H.B., CHRISTOU, A., HENRY, R.L., SCHNIDT, W.A., and SAKS, N.S.: 'A virtual self-aligned process for *n*-channel InP IGFET's (or MISFET's)', *IEEE Electron Device Lett.*, 1981, **EDL-2**, pp. 299–301
84 MAO, A.S.D., LEHENY, R.F., NAHORY, R.E., and DEWINTER, J.C.: 'An $In_{0.53}$ $Ga_{0.47}$ As/Si_3 N_4 *n*-channel inversion mode MISFET', *ibid.*, 1981, **EDL-2**, pp. 288–290
85 MESSICK, L.J.: 'A high-speed monolithic InP MISFET integrated logic invertor', *IEEE Trans.*, 1981, **ED-28**, pp. 218–221
86 MIMURA, T., HIYAMIZU, S., JOSHIN, K., and HIKOSAKA, K.: 'Enhancement-mode high electron mobility transistors for logic applications', *Jpn. J. Appl. Phys.*, 1981, **20**, pp. L317–L319
87 MIMURA, T., HIYAMIZU, S., FUJII, T., and NANBU, K.: 'A new field-effect transistor with selectively doped GaAs/*n*-Al_x $Ga_{1-x}As$ heterojunctions', *ibid.*, 1980, **19**, pp. L225–L227
88 OHNO, H., BARNARD, J., WOOD, C.E.C., RATHUN, L., and EASTNMAN, L.F.: 'Double heterojunction GaInAs devices by MBE'. IEDM Technical Digest, 1980, pp. 434–437
89 MORKOC. H.: 'Current transport in modulation doped (Al,Ga) As/GaAs heterostructures: Applications to high speed FET's', *IEEE Electron Device Lett.*, 1981, **EDL-2**, pp. 260–261
90 DRUMMOND, T.J., NORKOG, H., SU, S.L. FISCHER, R., and CHO, A.Y.: 'Enhanced mobility in inverted $Al_xGa_{1-x}As$/GaAs heterojunctions: Binary on top of ternary', *Electron. Lett.*, 1981, **17**, (23), pp. 870–871
91 HIYAMIZU, S., MIMURA, T., FUJII, T., NANBU, K., and HASHIMOTO, H.: 'Extremely high mobility of two-dimensional electron structures grown by MBE', *Jpn. J. Appl. Phys.*, 1981, **20**, pp. L245–L248
92 MIMURA, T., JOSHIN, K., HIYAMIZU, S., HIROSAKA, K., and ABE, M.: 'High electron mobility transistor logic', *ibid.*, 1981, **20**, pp. L598–L600
93 DELAGEBEAUDEUF, D., LAVIRON, M., DELESCLUSE, P., TUNG, P.N., CHAPLART, J., and LINH, N.T.: 'Planar enhancement mode two-dimensional electron gas FET associated with a low AlGaAs surface potential', *Electron. Lett.*, 1982, **18**, (2), pp. 103–105
94 SAKAKI, H.: 'Summary abstract: High mobility effect of electrons in ultrafine semiconductor wire structures', *J. Vac, Sci. & Technol.*, 1981, **19**, pp. 148–149
95 SAKAKI, H.: 'Electrons in ultrafine semiconductor wire structures', *Jpn. J. Appl. Phys.*, 1980, **19**, pp. L735–L738
96 TUNG, P.N., DELAGEBEAUDEUF, d., LAVIRON, M., DELESCLUSE, P., CHAPLARY, J., and LINH, N.T.: 'High-speed two-diminsional electron-Gas FET logic', *Electron. Lett.*, 1982, **18**, (2), pp. 109–110
97 GRUBIN, H.L.: 'Switching characteristics of nonlinear field effect transistors: Gallium-arsenide versus silicon', *IEEE Trans.*, 1980, **MTT-28**, pp. 442–448
98 INO, M., and OHMORI, M.: 'Intrinsic response time of normally off MESFET's of GaAs, Si and InP', *ibid.*, 1980, **MTT-28**, pp. 456–459
99 GRUBIN, H.L., and FERRY, D.K.: 'Conceptual problems in modelling submicron device physics', *J. Vac. Sci. & Technol.*, 1981, **19**, pp. 540–544
100 WADA, T., IKOMA, T., and ADACHI, Y.: 'Effects of hot-carrier phenomena on the performance semiconductor MESFET's'. Proceedings of ICCC, 1980 pp. 1–4
101 D'AVANZO, D.: 'Proton isolation for GaAs integrated circuits'. GaAs IC symposium, 1981
102 YAMAMOTO, R., and HIGASISAKA, A.: 'High-speed GaAs digital integrated circuit with clock frequency of 4.1 GHz', *Electron. Lett.*, 1981, **17**, (8), pp. 291–292
103 LONG, S.I., LEE, F.S., ZUCCA, R., WELCH, B.M. and EDEN, R.C.: 'MSI high-speed low-power GaAs integrated circuits using Schottky diode FET logic', *IEEE Trans.*, 1980, **MTT-28**, pp. 446–472
104 MITSIU, Y., OTSUBO, M., ISHIHARA, O., and NAKATANI, M.: 'Normally-off GaAs MSEFET logic', GaAs IC symposium, 1981
105 HO, V.Q., and SUGANO, T.: 'Fabrication of Sikim MOSFETS using neutron irradiated silicon as semi-insulating substrate', (to be published)
106 GHEEWALA, T.R.: 'A 30-ps Josephson current injection logic (CIL)', *IEEE J Solid-state Circuits.*, 1979, **SC-14**, pp. 787–793
107 'Special issue on Josephson devices', *IBM J. Res. & Dev.*, 1980, **24**
108 VAN DUZER, T.: 'Josephson digital devices and circuits', *IEEE Trans.*, 1980, **MTT-28**, pp. 490–500
109 UPADHYAYULA, C.L., SMITH, R.E., WILHELM, J.F., JOLLY, S.T., and PACZKOWSKI, J.P.: 'Transferred electron logic devices for gigabit-rate signal processing'. IEEE MTT-S microwave symposium, 1976, pp. 164–165
110 LEHOVEC, K.: 'GaAs enhancement mode FET-tunnel diode ultra-fast low power inverter and memory cell', *IEEE J. Solid-State Circuits*, 1979, **SC-14**, pp. 797
111 ABE, M., MIMURA, T., YOKOYAMA, N., and ISHIDALA.: 'New technology towards GaAs LSI/VLSI for computer applications'. GaAs IC symposium, 1981
112 TSUI, F.F.: 'JSP: A research signal processor in Josephson technology', *IBM J. Res. & Dev.*, 1980, **24**, pp. 243–252
113 BOLL, H.J., LEVINSTEIN, H.J., and SMITH, G.E.: 'Silicon MOSFETS attain high speed via submicron dimensions', *MSM Microwave Syst. News*, Feb. 1982, pp. 75–89
114 SHUR, M.S., and LONG, D.: 'Performance prediction for submicron GaAs SDFL logic', *IEEE Electron Device Lett.*, 1982, **EDL-3**, pp. 124–127
115 YOKOYAMA, N., OHNISHI, T., ODANA, K., ONODERA, H., and ABE, M.: 'Ti/w silicide technology for silf aligned GaAs MESFET VLSIs'. IEDM, 1981, pp. 80–83
116 OHWADA, K., INO, M., MIZURANI, T., and ASAI, K.: 'GaAs 256 bit static RAM', *Electron. Lett.*, 1982, **18**, (7), pp. 299–300
117 TUNG, P.N., DELESCLUSE, P., DELAGEBEAU DEU, D., LAVIRON, M., CHAPLANT, J., and LINH, N.T.: 'High speed low power DCFL using planar 2D electron gas FET technology', *ibid.*, 1982, **18**, (12), pp. 517–519
118 EDEN, R.C.: 'Comparison of GaAs device approaches for ultra-high speed VLSI', *Proc. IEEE*, 1982, **70**, pp. 5–12
119 BOZLER, C.O., and ALLEY, G.D.: 'The permeable base transistor and its application to logic circuits', *ibid.*, 1982 **70**, pp. 46–52
120 KROEMER, H.: 'Heterostructure biploar transistors and IC's', *ibid.*, 1982, **70**, pp. 13–25
121 VERMA, K.B., NODE, A.G., and BEUTHER, R.R.: 'GaAs planar digital IC structures'. GaAs IC symposium 1980
124 YAMASAKI, K., JAMANE, Y., and KURUMADA, K.: 'Below 20ps/gate operation with GaAs Saint FETs at room temperature', *Electron. Lett.*, 1982, **18**, (14), pp. 592–593

9 Appendix

To consider the practicalities of the various approaches to gigabit logic, a crude, but nevertheless quantitative, theory is reproduced here, to relate the various device parameters to the propagation delays, and power-delay products. FET circuits are considered because of their simplicity, and the fact that they show better prospects for VLSI than do bipolar circuits. The argument follows Eden [7] for the GaAs MESFET, but both MOSFETs and JFETs are similar. The drain current in a channel of width w is given by

$$I_{DS} = w Q_c v_d \tag{8}$$

for Q_c the charge per unit area of the channel, and v_d the drift velocity. The average field $\bar{\epsilon}$ in the channel for the drain-source voltage to be saturated (i.e. greater than $V_{GS} - V_p$, where V_{GS} is the gate-source voltage and V_p the pinch-off voltage) is

$$\bar{\epsilon} \sim \frac{(V_{GS} - V_p)}{l} \tag{9}$$

If the charge density is approximated to the product of the average gate-channel voltage and the gate to channel capacitance per unit area, then

$$Q_C \sim \epsilon \frac{V_{GS} - V_p}{2a} \tag{10}$$

where a is the average gate-channel separation, and ϵ the semiconductor permittivity.

This approximation will give the required relationship between the variables, with the omission of the integration constants. Further, assuming the Schokley model for the FET, the drift velocity is simply the electron mobility times the average field:

$$v_d = \mu \bar{e} \tag{11}$$

This assumes no velocity saturation, as discussed in Section 5. From eqn. 8

$$I_{DS} = \frac{w\beta}{2l} (V_{GS} - V_p)^2 \tag{12}$$

where

$$\beta = \frac{\epsilon \mu}{a} \tag{13}$$

For an FET circuit as shown in Fig. 12, if the load transistor is assumed a constant current source, providing half the maximum drain current of the switching FET, the pull-up and pull-down will be equal. For a maximum voltage amplitude V_m, the transition times are given approximately by

$$\Delta T \sim \frac{2C_L V_m}{I_{DS}} \tag{14}$$

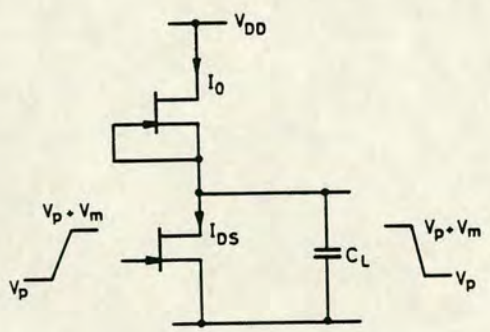


Fig. 12 *Inverter model used in switching analysis*

The propagation delay of the circuit is proportional to ΔI with a constant of proportionality of 2/3 according to Eden:

$$\tau_{PD} \sim \frac{4}{3} \frac{C_L V_m}{I_{DS}} \tag{15}$$

At the maximum switching speed, the load transistor is acting as a constant current source, either charging the capacitor, or passing current through the switch FET. At lower speeds, the power could be halved. The power and dynamic switching energy are given by

$$P_D = V_{DD} I_o = \frac{1}{2} V_{DD} I_{DS} \tag{16}$$

$$P_D \tau_{PD} = \alpha C_L V_{DD} V_m \tag{17}$$

for d $\simeq$ 2/3.

As the maximum output voltage swing will equal the input swing, eqn. 11 gives

$$I_{DS} = \frac{w\beta}{2l} V_m^2 \tag{18}$$

$$\tau_{PD} \sim \frac{8}{3} \frac{C_L l}{w\beta V_m} \tag{19}$$

Assuming that the minimum supply voltage is proportional to V_m, then

$$P_D \tau_{PD} \propto C_L V_m^2 \tag{20}$$

$$P_D \tau_{PD} \propto \frac{C_L^3 l^2}{\beta^2 w^2 \tau_{PD}^2} \tag{21}$$

with the constant of proportionality being approximately 6 to 8 for a typical relationship between V_{DD} and V_m. It should be emphasised that this is a worst-case value for the dynamic switching energy, ignoring the power contributed by the level-shifting technique. The level shifting will add to the power consumed, but varies with the circuit configuration. Switching speed is assumed to be independent of the level-shifting technique employed, for the purposes of this calculation.