

Pseudomorphic HEMT Technology and Applications

NATO ASI Series

Advanced Science Institutes Series

A Series presenting the results of activities sponsored by the NATO Science Committee, which aims at the dissemination of advanced scientific and technological knowledge, with a view to strengthening links between scientific communities.

The Series is published by an international board of publishers in conjunction with the NATO Scientific Affairs Division

A	Life Sciences	Plenum Publishing Corporation
B	Physics	London and New York
C	Mathematical and Physical Sciences	Kluwer Academic Publishers
D	Behavioural and Social Sciences	Dordrecht, Boston and London
E	Applied Sciences	
F	Computer and Systems Sciences	Springer-Verlag
G	Ecological Sciences	Berlin, Heidelberg, New York, London,
H	Cell Biology	Paris and Tokyo
I	Global Environmental Change	

PARTNERSHIP SUB-SERIES

1. Disarmament Technologies	Kluwer Academic Publishers
2. Environment	Springer-Verlag / Kluwer Academic Publishers
3. High Technology	Kluwer Academic Publishers
4. Science and Technology Policy	Kluwer Academic Publishers
5. Computer Networking	Kluwer Academic Publishers

The Partnership Sub-Series incorporates activities undertaken in collaboration with NATO's Cooperation Partners, the countries of the CIS and Central and Eastern Europe, in Priority Areas of concern to those countries.

NATO-PCO-DATA BASE

The electronic index to the NATO ASI Series provides full bibliographical references (with keywords and/or abstracts) to more than 50000 contributions from international scientists published in all sections of the NATO ASI Series.

Access to the NATO-PCO-DATA BASE is possible in two ways:

- via online FILE 128 (NATO-PCO-DATA BASE) hosted by ESRIN,
Via Galileo Galilei, I-00044 Frascati, Italy.
 - via CD-ROM “NATO-PCO-DATA BASE” with user-friendly retrieval software in English, French
and German (© WTV GmbH and DATAWARE Technologies Inc. 1989).

The CD-ROM can be ordered through any member of the Board of Publishers or through NATO-PCO, Overise, Belgium.



Series E: Applied Sciences - Vol. 309

Pseudomorphic HEMT Technology and Applications

edited by

R. Lee Ross

U.S. Army Research Laboratory,
Fort Monmouth, NJ, U.S.A.

Stefan P. Svensson

Martin Marietta Laboratories,
Baltimore, MD, U.S.A.

and

Paolo Lugli

Dipartimento di Ingegneria Elettronica,
Università degli Studi di Roma "Tor Vergata",
Rome, Italy



Kluwer Academic Publishers

Dordrecht / Boston / London

Published in cooperation with NATO Scientific Affairs Division

Proceedings of the NATO Advanced Study Institute on
Pseudomorphic HEMT Technology and Applications
Erice, Sicily, Italy
July 14–25, 1994

A C.I.P. Catalogue record for this book is available from the Library of Congress

ISBN-13:978-94-010-7228-1
DOI:10.1007/978-94-009-1630-2

e-ISBN-13:978-94-009-1630-2

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Kluwer Academic Publishers incorporates the publishing programmes of
D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Press.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

All Rights Reserved
© 1996 Kluwer Academic Publishers
Softcover reprint of the hardcover 1st edition 1996

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner.

CONTENTS

PREFACE

vii

INTRODUCTION

Chapter 1	Introduction to PHEMTs J.V. DiLorenzo, B.D. Lauterwasser and M. Zaitlin	1
-----------	--	---

PHEMET LAYER STRUCTURE DESIGN AND FABRICATION

Chapter 2	Pseudomorphic HEMTs: Device Physics and Materials Layer Design T. Grave	23
Chapter 3	Pseudomorphic HEMT: Materials Growth and Characterization D. C. Streit	71

PHEMET DEVICE FABRICATION

Chapter 4	Gate Formation Technologies P.C. Chao	93
Chapter 5	Vias, Backside Thinning and Passive Elements J.A. Turner	109

MODELING, DESIGN, SIMULATION AND TEST

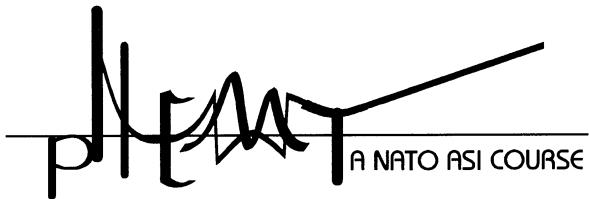
Chapter 6	Field-Effect Transistor Models and Microwave CAD R. J. Trew	125
Chapter 7	HEMT Models and Simulations P. Lugli, M. Paciotti, E. Calleja, E. Munoz, J.L. Sanchez-Rojas, F. Desenne, R. Fauquemergue, J.L. Thobel and G. Zandler	141
Chapter 8	MMIC Circuit Design S. Weinreb	165
Chapter 9	Accurate Active Device Models for Computer Aided Design of MMICs R. Tayrani	181
Chapter 10	Advanced CAD Models G. Ghione, F. Bonani, and M. Pirolo	203
Chapter 11	Test and Reliability B. R. Allen	251

RELATED TECHNOLOGIES

Chapter 12	InP-Based Power HEMTs M. Matloubian, and L. E. Larson	275
------------	--	-----

APPLICATIONS

Chapter 13	Space Applications of P-HEMT Devices G. Gatti	301
Chapter 14	European Applications of Pseudomorphic HEMTs - Military and Commercial J. Favre	329
Chapter 15	U.S. Applications of HEMT Technology in Military and Commercial Systems E. I. Sobolewski	335



PREFACE

Pseudomorphic HEMT (PHEMT) devices and their incorporation into advanced monolithic integrated circuits is the enabling technology for modern microwave/millimeter-wave system applications. Various microwave/millimeter-wave monolithic integrated circuit (MIMIC) initiatives are presently devoting their major attention to the development of a viable manufacturing/production base incorporating PHEMT technology. This technology is expected to continue its major role in proposed follow-on programs continuing into 1998 and beyond. Although still in its infancy, PHEMT MIMIC technology is already finding applications in both military and commercial systems, including radar, communication, and automotive technologies.

Solid-state scientists create and develop devices which are used as oscillators, low noise and power amplifiers, phase shifters, converters, switches and mixers which are incorporated into a broad spectrum of microwave and millimeter-wave systems. Yet the researcher's focus is usually limited in scope to a confined area of involvement and interest. Consequently, an opportunity is needed to improve the understanding of the applications of advanced device technology to overall system design and performance. Conversely, it is the circuit designer who develops the device specifications which drive the need for advanced device RF performance. The materials/device physicist must be cognizant of these specifications in order to be relevant and responsive to these needs. The system engineer, in turn, must access a technology for solutions to problems of performance and reliability, while the technical manager is concerned with cost and time-to-market considerations in the development of a new application in the competitive market place. The successful team in a globally competitive market is one in which all four individuals are cognizant of those considerations and requirements which influence each other's function. The NATO Advanced Study Institute *Pseudomorphic HEMT Technology and Applications* attempted to achieve this realization through its comprehensive format of topics and discussions, as well as the broadened scope of its participants and lecturers.

The objective of this NATO Advanced Study Institute was to continue the tradition of sharing new knowledge among solid-state researchers, engineers and technical managers in a setting isolated from distractions. In contrast to previous Advanced Study Institutes whose objectives were to treat a specific subject in depth, the aim of this program was to widen the scope of the study. The benefit to be realized is the broadened scope necessary for today's researcher and manager who must be cognizant and educated to remain relevant in an ever expanding technology base.

The purpose and scope of the Advanced Study Institute was to provide participants with a comprehensive overview of pseudomorphic HEMT technology including materials, fabrication and processing, device physics, CAD tools and modeling, monolithic integrated circuit technology and applications. Organization of these proceedings follows the format of the ASI. A major goal of the Institute was to present a realistic assessment of the current status and future potential of pseudomorphic HEMT technology. Lectures covered basic concepts in a tutorial manner as well as practical issues and challenges. Participants with specific areas of expertise were welcomed to lead group discussions of practical issues. When appropriate, comparisons were made with competing technologies in the various subject areas.

The editors wish to acknowledge Professor Minko Balkanski, Director of the International School of Material Science and Technology, for his guidance, support and enthusiasm. This ASI was the 29th course in his school. We would like to thank the Scientific Affairs Division of NATO, Dr. L. Veiga da Cunha, Director, for their kind and generous financial support of this undertaking. Gratitude is also expressed to the staff and organization of the Ettore Majorana Centre for Scientific Culture, with particular thanks to Dr. S.A. Gabriele and Miss Pinola Saralli for their help, dedication and patience. To the people of Erice and to Carmello, for their hospitality and warm kindness extended to all of us.

A special thankfulness is expressed to the 19 lecturers who gave of their time and energy to come to Erice and share their knowledge and friendship with the 80 some odd participants who attended this ASI. To the employers of these lectures who allowed them to take precious time away from their employment and who supported them, and this ASI, financially so as to make this undertaking a realization. To the Esprit Working Group ELTRASIN, Office of Naval Research European Office, U.S. Army Research Development Standardization Group U.K. and the National Science Foundation whose financial support provided grants allowing several participants to attend who otherwise would have been unable to do so. A special thank you is extended to Mr. Robin Beard, Assistant Secretary General of NATO for his wit, humor and poignant banquet address on *The Future of NATO*.

Finally, it was in July of 1981 at a NATO ASI in Erice, that two of us (RLR and SPS) met each other for the first time, RLR from America and SPS from Sweden. A friendship was born which has become ever lasting. As our careers developed over the ensuing years and as we infrequently met each other during brief occurrences, we would reminisce of the wonderful time in Erice in 1981. It was our fondest desire to return to Erice and to bring to others that which we were so fortunate to experience on that occasion. That occasion was the 3rd course in Prof. Minko Balkanski's International School of Materials Science and Technology. That desire was realized by this offering of the 29th course. It is hoped that this offering generated an environment to perpetuate this desire in others. Thank you to all that made it possible, for it truly was a labor of love.

R. Lee Ross, US Army Research Laboratory, USA, ASI Director

Stefan P. Svensson, Martin Marietta Laboratories, USA, US Technical Director

Paolo Lugli, Universita Degli Studi di Roma "Tor Vergata", Italy, European Technical Director

INTRODUCTION TO PHEMTS

J.V. DILORENZO, B.D. LAUTERWASSER, M. ZAITLIN

Raytheon Company

Advanced Device Center

362 Lowell Street

Andover, MA 01810

1. Introduction

The development of the pseudomorphic high electron mobility transistor (PHEMT) was a result of the convergence of several factors. These factors included an interest in the behavior of quantum well structures (including superlattices) and development of the concept of modulation doping, the development of a crystal growth technique (molecular beam epitaxy) capable of producing heterojunction III-V compound semiconductor materials with atomically abrupt transitions in composition and doping, and lastly, the application of both the MBE growth technique and the knowledge of modulation doped superlattice behavior to the development of field effect transistors. The superior microwave performance capabilities of PHEMT devices compared to conventional GaAs MESFETs provided the impetus for early application of PHEMT technology to advanced military systems, where ultimate performance historically outweighed cost as a system driver. The subsequent investment in PHEMT material, design, and manufacturing capabilities by the defense community, through programs such as the MIMIC program in the U.S., enabled the evolution of a cost effective PHEMT technology base. As a result, PHEMT technology is being selected for a number of emerging commercial applications including global communications, cellular phones, wireless local area networks, high efficiency lighting, and a host of others.

Table 1 shows a timeline of some of the significant milestones in PHEMT development, from the first proposal for a heterojunction device at the time that the transistor was invented to the first reported PHEMT-based MMIC. Each decade since the 1950's appears to have had its own focus with respect to PHEMT development. The 1960's saw the parallel development of the underlying theory of mobility enhancement in heterojunction devices and the practical means (MBE) of building such devices. The following decade witnessed a flurry of activity centered around the characterization of both optical and electrical behavior of quantum wells and heterojunctions, culminating in the first demonstration of a HEMT device. With that first demonstration of device performance in hand, the 1980's saw a continuous stream of device refinements and improvements, from the HEMT to the PHEMT to the pulse doped PHEMT to a fully functional PHEMT-based MMIC. The 1990's are likely to be remembered as the decade when PHEMT-based products finally entered the marketplace.

Table 1. Milestones in PHEMT Development

1951	Heterojunction device proposed (Patent #2,569,347)	Shockley	BTL
1969	Mobility enhancement in superlattice heterojunction predicted for GaAs/AlGaAs system	Esaki & Tsu	IBM
1969	Molecular Beam Epitaxy demonstrated	Cho & Arthur	BTL
1978	Mobility enhancement in GaAs/AlGaAs demonstrated	Dingle, et. al.	BTL
1978	U.S. patent for HEMT (Patent #4,163,237)	Dingle, et. al.	BTL
1980	First demonstration of HEMT device	Morkoc, et. al.	U. of III. Rockwell
1985	Pseudomorphic HEMT introduced	Mimura, et. al. Zipperian, et. al.	Fujitsu U. of III.
1986	Present InGaAs/AlGaAs PHEMT structure introduced	Rosenberg, et. al. Ketterson, et. al.	IBM U. of III.
1987	Pulse-doped PHEMT demonstrated	Moll, et. al.	HP
1988	First comprehensive (numerical) model of HEMT	Foisy	Cornell
1989	First PHEMT-based MMIC reported	Aust, et. al.	TRW

2. Physics & Characterization of Heterojunctions

As is apparent from the HEMT acronym, the original motivation for the development of AlGaAs/GaAs quantum well heterostructures was to obtain higher electron mobility. At room temperature, electron mobility is limited by scattering from optical phonons and ionized impurities. As the doping level increases, the deleterious effect of impurity scattering becomes more pronounced (for typical device bulk doping levels in GaAs, the mobility is less than half that of semi-insulating GaAs). Thus, attempts to increase the performance of MESFETs by increasing the sheet density in the channel are ultimately self-limiting.

The HEMT (and PHEMT) structure decouples the mobile carriers from their dominant scattering mechanism by physically separating them from the dopant ions. This is accomplished by juxtaposing a doped, wide bandgap semiconductor with an undoped, narrower bandgap semiconductor (see Figure 1). The potential energy of the conduction band in the narrow bandgap material is lower than that in the (doped) wide bandgap semiconductor, causing electrons to transfer to the lower energy region. This charge transfer is opposed, however, by the electric field between the, now separated, electrons and donor ions. This electric field alters the band potential, confining the carriers to a triangular quantum well region in the narrow bandgap material immediately adjacent to the wide bandgap material. This quantum well is sufficiently thin that the free carriers form a

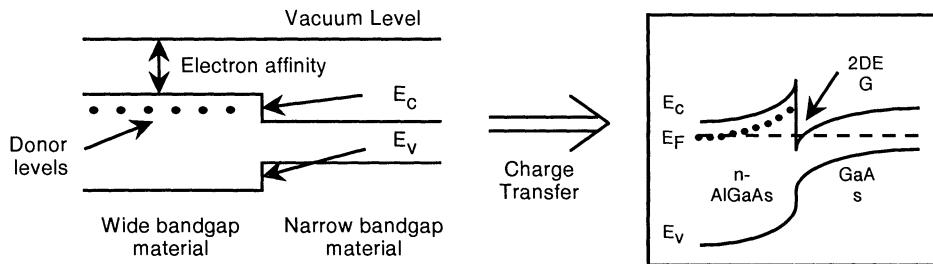


Figure 1. The effect on energy levels of joining two semiconductors with different bandgap energies. The Fermi level remains constant across the interface. Note the formation of a thin energy well in the conduction band of the narrow bandgap material.

2-dimensional electron gas (2DEG). The expected mobility enhancement resulting from the separation of electrons and donor ions was observed in 1978 by Dingle, et al, who used an AlGaAs/GaAs modulation-doped structure grown by molecular beam epitaxy. Figure 2 shows the substantial mobility improvement they obtained.

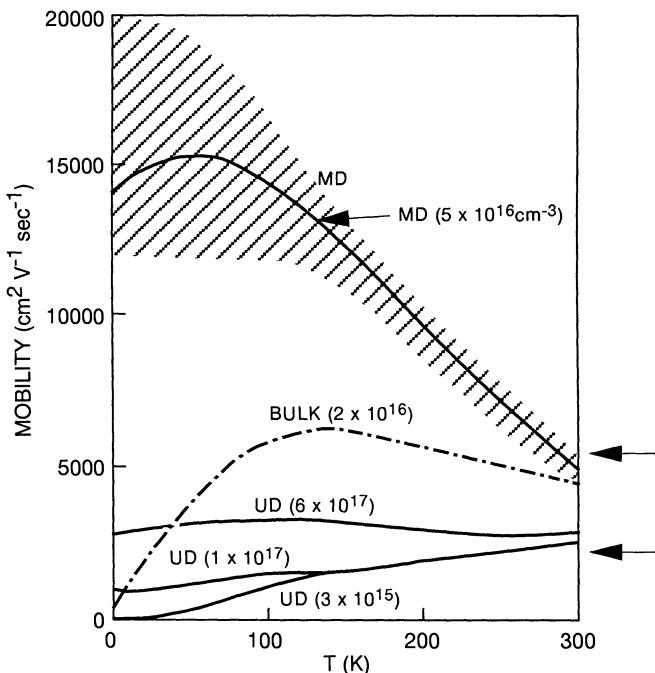


Figure 2. Mobility enhancement observed by Dingle, et. al. in modulation doped heterostructures compared to uniformly doped heterostructures.

Although conceptually simple, the epitaxial growth of a heterojunction quantum well such as that described above requires a careful choice of materials: the two constituent semiconductors must have the same lattice type, they must have similar lattice constants, and there must be sufficient difference between their bandgaps to achieve a useful degree of charge transfer. From among the III-V semiconductors (zincblend structure) the initial work on HEMTs was done using the $\text{GaAs}/\text{Al}_x\text{Ga}_{1-x}\text{As}$ system because those two materials are very closely lattice matched for all values of x . While this system was well suited to studies of the fundamental concepts of modulation-doped heterostructures, it was not the ideal material system for real devices. As efforts were made to use higher values of x in the doped barrier material (to increase the bandgap difference and thus the charge transfer to the quantum well) a fundamental difficulty with AlGaAs became apparent.

The culprit is a donor-defect complex in AlGaAs , known as the DX center. This center is linked to a higher indirect valley in the conduction band whose energy above the direct conduction band valley decreases as x increases (for $x \geq 0.45$ the indirect valleys are lower in energy than the direct valley, and AlGaAs becomes an indirect semiconductor). Because of this changing offset between the different conduction band minima, the DX center emerges as a trap level in the bandgap for $x \geq 0.25$. This level traps electrons, significantly decreasing the concentration of free carriers available for transfer to the quantum well. Thus, there is a practical limit to the bandgap discontinuity which can be achieved in the $\text{AlGaAs}/\text{GaAs}$ system.

If it is impractical to increase the AlGaAs barrier height, the only other option for obtaining a greater conduction band offset is to reduce the bandgap of the quantum well material by growing an alloy of GaAs with a narrower bandgap III-V semiconducting material (e.g., InAs). All of the narrower bandgap III-V compounds have substantially larger lattice constants than GaAs (and, therefore, AlGaAs) raising the question of the growth of lattice mismatched semiconductors. Within certain limits, lattice mismatched heterostructures can, in fact, be grown. As the growth of the new layer begins, the atoms initially align themselves to the lattice of the underlying layer. Subsequent layers continue the same atomic registration. The result is the growth of a “strained layer” in which the lattice constant in the plane of the growth is constrained to match that of the underlying material, which also alters the lattice constant in the direction perpendicular to the growth as the strained layer attempts to maintain roughly the same volume as if it were unstrained. In the case of the growth of InGaAs on GaAs (or AlGaAs), the lattice constant of the InGaAs in the growth plane is less than its bulk value, whereas the lattice constant perpendicular to the growth plane is greater than the bulk value (see Figure 3).

The total energy due to the strain in such a layer is proportional to the lattice mismatch times the thickness of the layer. At some point, the strain energy becomes sufficiently large that the layer “relaxes” by forming dislocations, which can seriously degrade device

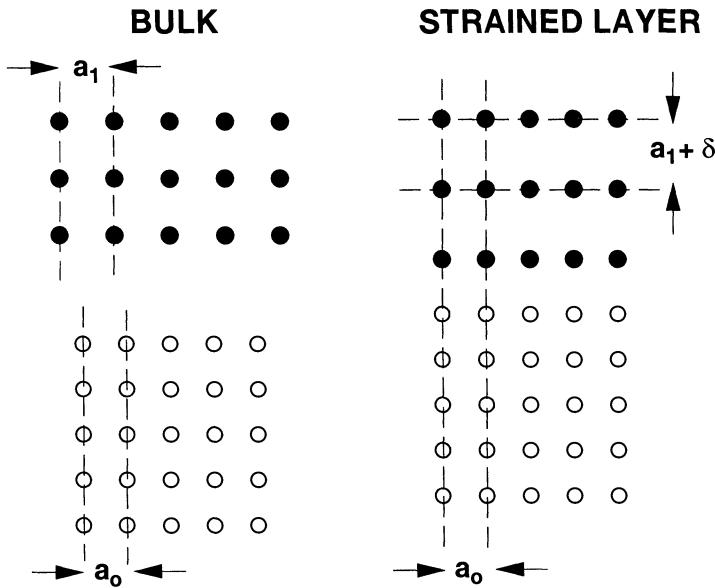


Figure 3. Effect of strained layer growth on the lattice constants of the strained layer.

performance. This relaxation occurs at the “critical thickness,” which is a function of lattice mismatch and growth conditions. In the $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{AlGaAs}$ system, since the lattice mismatch is proportional to the In fraction, the strain relaxation sets an upper limit to the InGaAs layer thickness for a given value of x . For $x = 0.20$, this limit is in the vicinity of 160 Å, the precise value again being dependent upon the particular growth conditions. Real PHEMT device structures are typically grown with InGaAs quantum wells much thinner than the critical thickness in order to provide a substantial margin of safety. To complete the quantum well structure, another barrier layer (e.g., AlGaAs) is grown on top of the strained InGaAs, resulting in a structure which resembles a square well.

The performance of PHEMT devices is particularly dependent upon the underlying epitaxial heterostructure, in terms of both the quality of the individual layers and the correctness of the overall layer structure (compositions, thicknesses, etc.). It is important, therefore, to have a means of evaluating the quality of the epitaxial material before any device processing commences. This evaluation is primarily, although not exclusively, focussed on the quality of the quantum well structure.

The most fundamental properties of the quantum well are its transport characteristics: the sheet density of the electrons contained in the quantum well and their mobility. This

measurement is complicated by the fact that there can be several doped layers in a PHEMT epitaxial structure in addition to the quantum well in which, it is hoped, most of the electrons have collected. If the Hall measurements are done at low temperature (e.g., 77 K), the carriers in the quantum well have much higher mobility (due to the much reduced impurity scattering) than those remaining in the doped regions. Virtually all of the conduction in the structure will therefore be in the quantum well and the measured mobility and sheet density will be essentially those of the carriers in the quantum well. Figure 4 shows the trend with temperature in the sheet density and mobility for a quantum well with low carrier concentration. The constant sheet density for temperatures below about 120 K indicates that there are no other paths of parallel conduction in this structure. Typical PHEMTs have quantum well sheet densities about an order of magnitude larger than that shown in Figure 4, with 77 K mobilities around $20,000 \text{ cm}^2/\text{V}\cdot\text{s}$.

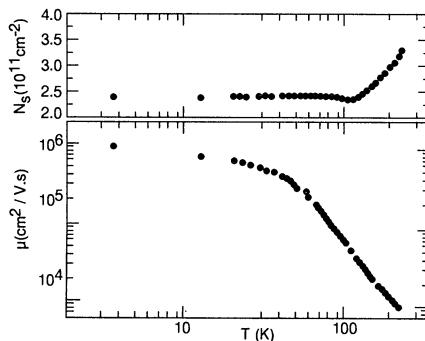


Figure 4. Trends in sheet density (top) and mobility (bottom) with temperature for a quantum well with low carrier concentration.

Although the transport properties convey critical information about the quality of the quantum well, they provide no knowledge of the structure or shape of the quantum well, factors which can also affect device performance. In addition, Hall measurements are destructive: a special sample must be prepared rendering the remainder of the epitaxial wafer useless for subsequent device processing. For these reasons, it is desirable to have a non-destructive evaluation technique which provides more detailed information about the electronic and physical structure of the quantum well region. Photoluminescence (PL) is just such a technique. Traditionally, PL has been done at very low temperatures (e.g., 4 K) where it has the ability to resolve features separated by only hundredths of a meV and it has been useful in answering questions about the interface quality of quantum wells. To be truly non-destructive, however, the characterization technique must be capable of providing useful information when applied to a full epitaxial wafer at room temperature. In the past few years, room temperature PL has been shown to meet these criteria.

Figure 5 is a room temperature PL spectrum of a typical PHEMT structure. The two steep rises in intensity occur at the two lowest electron sub-bands in the conduction band

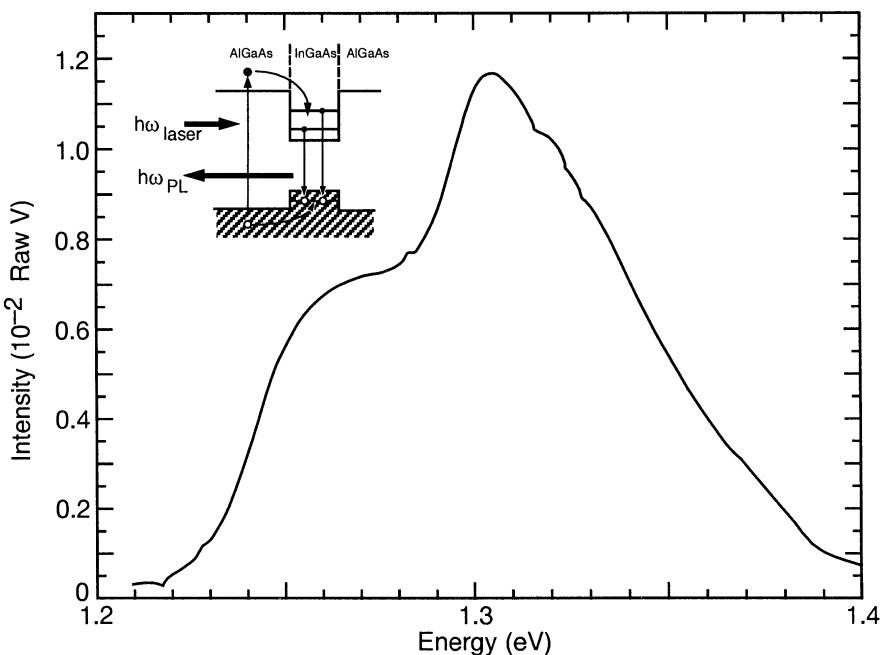
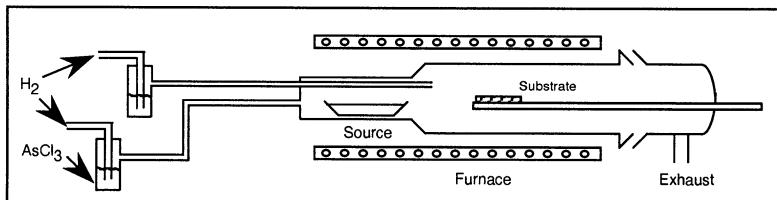


Figure 5. Room temperature photoluminescence spectrum of a typical PHEMT structure.

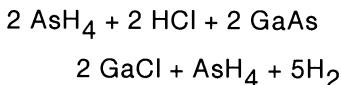
quantum well and the first heavy-hole sub-band in the valence band quantum well (see the figure inset). The energies of these transitions are related to the composition and width of the quantum well. The relative amplitudes are indicative of the underlying symmetry (or lack thereof) of the quantum well. The high energy roll-off is governed by the Fermi level, allowing one to determine the sheet density contained in the well. These, and other parameters, can be extracted by fitting the spectrum with a line shape model.

3. Evolution of Materials Technology

While the development of PHEMTs would not have been possible without a fundamental understanding of the physics of quantum well behavior, the actual demonstration of such behavior and ultimately the ability to make heterojunction devices would not have been possible without significant advances in materials technology. Prior to the early 1970's, III-V compound semiconductor epitaxial layers were grown using either vapor phase or liquid phase epitaxial growth techniques. While these techniques were capable of producing high quality epitaxial layers, they were limited in their ability to produce heterojunction materials of sufficiently high quality to demonstrate the optical and electrical properties of quantum wells which had been theorized. A brief description of each of these growth techniques will serve to underscore their limitations for PHEMT growth.



Hydride Process



Halide Process



Figure 6. Schematic diagram of a chloride based vapor phase epitaxial growth system.

Vapor phase epitaxy is based on chemical reactions between gaseous and solid source materials in thermodynamic equilibrium and on vapor phase transport of chemical reactants from a high temperature zone to a lower temperature zone. Figure 6 schematically illustrates a typical VPE reactor configuration. Two basic processes are available for the growth of III-V compound semiconductors such as gallium arsenide. In the hydride process, arsine and hydrogen chloride gases are introduced into the reactor where the HCl interacts with a liquid or solid phase source of GaAs to form GaCl, which is then transported downstream to the deposition zone. Deposition occurs on a GaAs substrate through the reverse chemical reaction, driven in the desired direction by the reduced temperature of the substrate. In the halide process, shown in Figure 6, both HCl and arsenic are produced from the decomposition of arsenic trichloride in a hydrogen ambient. Reactions in the source and substrate regions are very similar to those which occur in the hydride process.

Liquid phase epitaxy relies upon controlled solidification from supersaturated melts to produce high quality crystals. By using multiple melts of differing composition and a slider arrangement which allows the starting substrate crystal to be moved sequentially from one melt to another, heterojunction structures can be produced. LPE was used, in fact, for some of the very early work on heterojunction devices.

Both VPE and LPE are capable of producing very high quality layers of GaAs. However, both VPE and LPE techniques have inherent limitations which make them unsuitable for growing high quality PHEMT structures. Their relatively high growth rates of 10-60 $\mu\text{m}/\text{hr}$ make it very difficult to grow the extremely thin layers (15-100 \AA) required for optimum device performance. Aluminum containing compounds, such as AlGaAs, are problematic for VPE because aluminum chloride compounds react with the quartz walls of the reaction chamber. Good uniformity in layer thickness and doping con-

centration are difficult to achieve over large areas using VPE or LPE due to gas phase or liquid phase convection which results from thermal gradients. Finally, the constraints of growing under conditions of chemical and thermodynamic equilibrium in VPE and the use of a limited number of sequentially arranged melts in LPE limit the capability to grow complex structures of varying material composition.

Beginning in the late 1960's, two growth techniques were developed which appear to overcome the limitations of VPE and LPE for growth of highly complex heterostructures. Both metal-organic chemical vapor deposition (MOCVD) and molecular beam epitaxy (MBE) have been used to produce PHEMT structures, although MBE is favored for this application for reasons which are discussed below. MOCVD, as its name implies, makes use of highly volatile metal-organic compounds to introduce the group III elements of interest (Ga, Al, In) into the reaction chamber. Hydrides (arsine, phosphine) are typically used to supply the group V elements. The reaction zone is limited in volume by using rf induction heating of a graphite susceptor. In a typical MOCVD system, shown schematically in Figure 7, the substrate sits on the susceptor at a shallow angle with respect to the gas flow direction. The combination of the substrate angle and rotation of the substrate on the susceptor produces very uniform deposition conditions over relatively large areas. The use of precision mass flow controllers on the gas lines, rapid switching pneumatic valves, minimum dead volumes, and a high gas flow rate in the reactor (which may be at atmospheric or reduced pressure) results in both a very low and very controlled growth rate and the ability to switch gas stream composition (and thus material composition and doping) very abruptly. Advantages of the MOCVD process include relatively simple system construction (e.g., no high vacuum components are required), the ability to easily produce phosphorous containing compounds, and the capability to produce large volumes of wafers in a single growth run using a so-called barrel reactor configuration. Drawbacks of the technique include the use of moderately to extremely toxic materials, and the generally lower purity of the starting materials compared to those used in MBE growth.

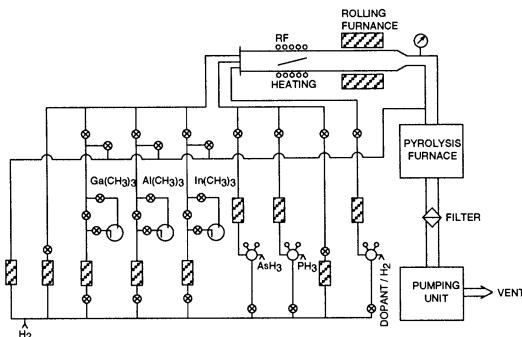
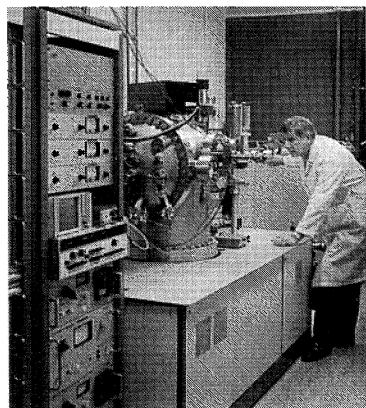
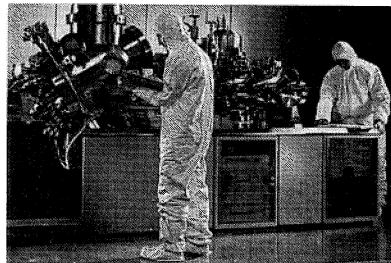


Figure 7. Schematic diagram of a typical MOCVD growth system.

Molecular beam epitaxy has been the growth technique of choice for first HEMT and then PHEMT structures since its introduction in the late 1960's. The technique is based on evaporation of high purity elements under ultra-high vacuum conditions and subsequent deposition onto a heated substrate. Modern MBE reactors incorporate design features such as tapered source cells and optimized source to substrate distances to produce extremely uniform deposition over large areas. A multi-wafer MBE system can produce up to three 4" diameter wafers (five 3" wafers) in a single growth run using a 12" diameter platen with variations in layer thickness, doping concentration, and alloy composition of less than 1% across a wafer. The growth rate in MBE systems is typically 1 $\mu\text{m}/\text{hr}$ which is equivalent to approximately one atomic layer per second. At such a low growth rate, rapid shuttering (0.1 sec) of evaporant cells produces atomic layer control of thickness, doping, and composition. The highly controlled growth environment of the MBE system, in combination with computer control of all system parameters, allows for growth of complex structures with excellent run to run reproducibility, making MBE a suitable technology for high volume production. The development of gas sources and specialized cracking furnaces in recent years has made it possible to grow ternary and quaternary compounds which incorporate phosphorous and other high vapor pressure elements. Table 2 lists some of the significant milestones in the evolution of MBE technology.

Table 2. Evolution of MBE Technology

- 1969 MBE technique invented (Cho)
- 1975 First commercial system sold
- 1980 Worldwide MBE sales = 13 systems
- 1986 First gas source MBE system
- 1986 Worldwide MBE sales = 82 systems
- 1988 First commercial multi-wafer system (3 x 2")
- 1992 First multi-wafer 4" machine
- 1993 Worldwide MBE sales = 31 systems



4. PHEMT Device Development

With the demonstration of enhanced mobility in modulation doped quantum well structures, the race was on to produce a high performance transistor based on the phenomenon. Within a relatively short span of time, several research laboratories had produced their own version of a high mobility transistor, and each laboratory produced its own unique name for the device. Thus Fujitsu developed a high electron mobility transistor (HEMT) while the University of Illinois and Rockwell produced a modulation doped field effect transistor (MODFET). Bell Laboratories touted the performance of its selectively doped heterojunction transistor (SDHT) while Thomson CSF offered its version called the two-dimensional electron gas FET (TEGFET), and with only slight modification NEC proclaimed the 2DEGFET. In the end HEMT emerged as the acronym most universally adopted, although it was not until the pseudomorphic HEMT (PHEMT) came along that all sides appeared to agree on terminology (one never sees reference, for example, to pseudomorphic TEGFETs). There are a number of reasons why the PHEMT offers performance advantages for both low noise and power applications over the conventional GaAs/AlGaAs HEMT. First, the larger bandgap discontinuity between AlGaAs and InGaAs compared to that between AlGaAs and GaAs produces higher sheet carrier density (more charge transfer) and thus higher device current in the PHEMT compared to the HEMT. Secondly, the higher saturated electron velocity of InGaAs compared to GaAs results in higher gain and better high frequency performance. Third, the improved carrier confinement in the InGaAs channel resulting from the larger bandgap discontinuity produces higher output conductance in the PHEMT. Finally, because the PHEMT can carry higher charge density in the channel than the HEMT, the AlGaAs layer above the channel can be doped to a higher level without introducing excessive parasitic current. This higher doping level reduces the access resistance to the channel. The net result of these several advantages is that PHEMTs produce higher gain with similar noise figure, and higher current for more power compared to HEMTs.

The construction of the PHEMT device, from the point of view of the underlying material structure, has evolved over time as various researchers have worked toward optimizing specific device parameters. Figure 8 shows a schematic cross-section of a typical low noise PHEMT device. The heart of the device, of course, is the InGaAs channel in which the two dimensional electron gas is confined. The channel composition is typically between 15% and 20% indium and is $110\text{\AA} - 150\text{\AA}$ thick. A superlattice buffer layer (alternating thin layers of GaAs and AlAs) beneath the channel enhances charge confinement in the channel to produce sharp pinch-off characteristics and allow for high values of transconductance at gate voltages very near pinch-off. The channel charge originates in a very narrow, very highly doped, doping “pulse” in the AlGaAs layer immediately above the InGaAs channel. The use of a doping pulse instead of a uniformly doped AlGaAs layer enhances charge transfer efficiency while reducing parasitic current in the AlGaAs layer. Power PHEMT devices frequently make use of a “double pulse doped”

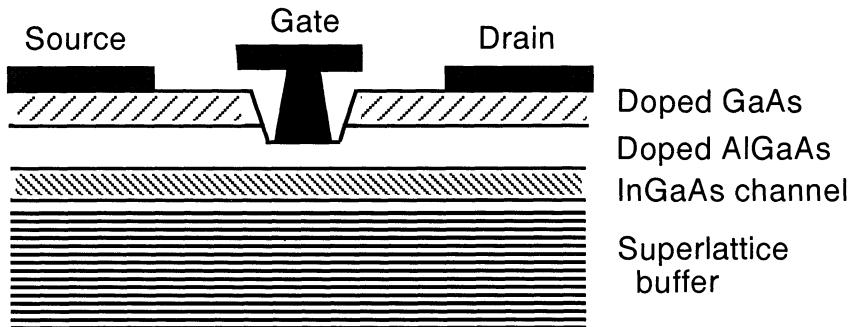


Figure 8. Schematic cross-section of a high electron mobility transistor.

structure in which doping pulses are placed both above and below the channel in order to increase the total amount of mobile charge in the channel for maximum power capability. Figure 8 also indicates the presence of a doped GaAs cap layer on top of the material stack. This layer, typically doped at $2-4 \text{ E}18 \text{ cm}^{-2}$, provides for good ohmic contact to the source and drain metal pads of the device. The FET geometry depicted in Figure 8 is a single recess T-gate geometry. The gate is recessed to produce the desired value of operating current (I_{dss}) in the device. The T-gate structure, as discussed later on, is used to reduce parasitic gate resistance.

Different approaches to material and device design are used to optimize PHEMTs for either low noise or high power operation. Much of the early work on PHEMTs was based on the advantages of those devices over HEMTs for low noise operation, especially at high frequencies. Figure 9 shows a comparison of current gain cutoff frequency (f_t) as a function of gate length between GaAs MESFETs, HEMTs, and PHEMTs (denoted InGaAs HEMT in the figure). The value of f_t varies inversely with gate length for all device types, and is directly proportional to electron velocity, which is what differentiates the curves for the three device types. Noise figure is inversely related to f_t , which is why PHEMTs have the advantage. A number of PHEMT material parameters affect device mobility and carrier concentration, and thus noise performance. These parameters include the thickness of the spacer layer between the InGaAs channel and the doping pulse in the AlGaAs layer, the thickness and composition of the InGaAs channel layer, and the thickness and doping level of the AlGaAs layer. As improvements have been made in material parameters and in extrinsic device parameters (such as reduced gate length and gate resistance), noise performance has steadily improved. Figure 10 shows the dramatic improvement made in f_t over the course of the past decade.

PHEMTs also possess a number of characteristics which make them well suited as microwave and millimeter-wave power devices. These characteristics include higher I_{max} for a given gate voltage swing (i.e., higher g_m), lower I_{min} due to sharper pinch-off

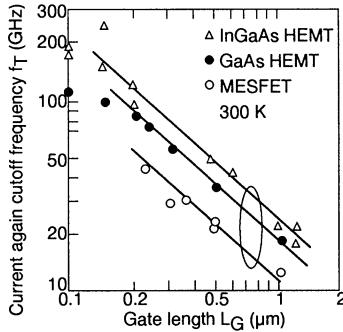


Figure 9. Variation of current gain cut-off frequency with gate length for MESFET, HEMT, and PHEMT devices.

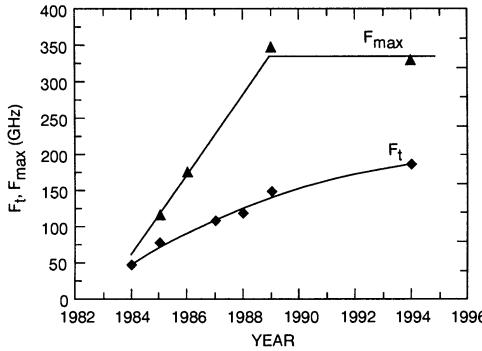
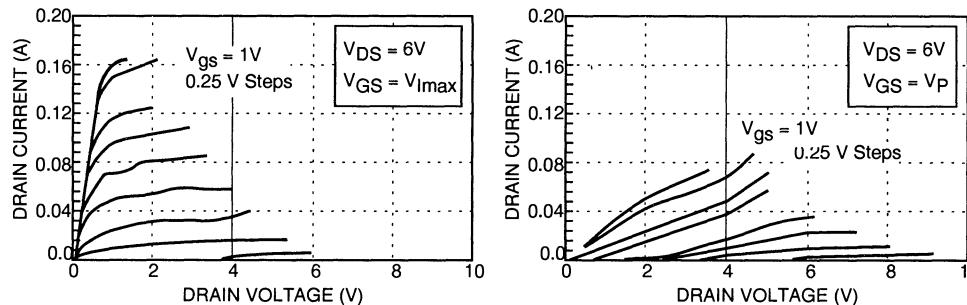


Figure 10. Improvement over time in device noise performance.

characteristic, lower knee voltage, and higher gain at high power levels all compared to MESFET devices. All of these characteristics tend to increase the device current swing for a given voltage swing which means higher output power. The evolution in PHEMT power capability over the past ten years has been dramatic. Early devices suffered from low values of I_{max} and breakdown voltage, both of which are critical to good power performance. This fact explains why most of the initial focus on PHEMTs was on low noise devices. State-of-the-art power performance for a PHEMT discrete device in 1985, measured at 15 GHz, was 140 mW output power with 8 dB gain and 37% power added efficiency. By 1993 PHEMTs were producing in excess of 1 W output power with 7 dB gain and over 50% power added efficiency at 18 GHz.

Several factors were instrumental in the development of improved PHEMT power devices. In order to increase maximum open channel current, it is necessary to increase the number of charge carriers in the device channel. Since HEMTs rely on charge transfer from the larger bandgap material to the smaller bandgap material, a method was

needed to increase the amount of charge transfer. Two methods were, in fact, developed. First, doped layers were located on both sides of the channel rather than simply in the upper AlGaAs layer as had traditionally been done. Secondly, uniform doping in the AlGaAs layers was replaced with very narrow very concentrated doping “pulses” positioned 30Å - 100Å away from the InGaAs channel. The very high doping density in these doping pulses results in increased charge transfer efficiency. A typical value of sheet carrier concentration in a double pulse doped PHEMT device might be 3.2E12 cm⁻² compared to 1.5E12 cm⁻² in a single side uniformly doped device, an improvement in channel charge of more than a factor of two. As noted above, poor breakdown voltage was a characteristic of early PHEMT devices. The introduction of a wide gate recess geometry improved device breakdown voltage by modifying the electric field profile on the drain side of the gate. Unfortunately, traps at the AlGaAs surface next to the gate can severely degrade device performance. The effect of traps is evident in the pulsed I-V characteristics of poor PHEMT power devices, as illustrated in Figure 11. By utilizing a double recess gate geometry high breakdown voltage can be maintained while removing the surface charge (traps) from the immediate vicinity of the gate. The combination of the double pulse doped material profile and the double recess gate geometry proved to be the key



The effect of surface charge on a pulse doped PHEMT is seen in the pulsed I-V characteristics of a 400 μm device with 0.33 μm gate length and a wide recess geometry. Pulse duration is 220 ns for the gate and 170 ns for the drain.

Figure 11. DC (left) and pulsed (right) I-V curves for a 400 μm PHEMT with 0.33 μm gate length and a wide recess geometry showing the effect of surface charge related traps. Pulse duration of 220 nS for the gate and 170 nS for the drain.

to unlocking the true power performance capability of the PHEMT device. Single device output power in excess of 10 watts at 2.4 GHz and power added efficiency of 40% at 44 GHz have been demonstrated (Figure 12).

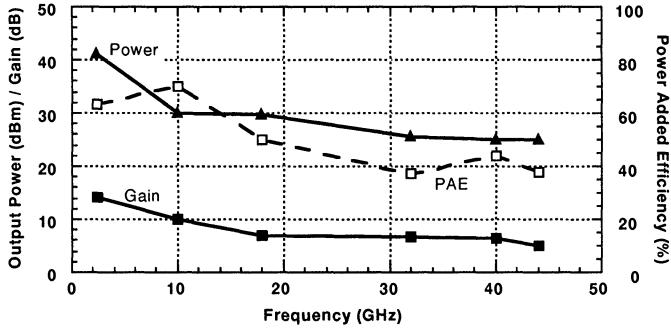


Figure 12. RF performance capability of discrete PHEMT devices as a function of frequency.

5. Designing & Fabricating PHEMT MMICs

Obtaining optimum MMIC performance using state-of-the-art PHEMT devices requires advanced design and fabrication techniques. For design purposes, accurate noise and large signal models are essential. In particular, large signal models should accurately describe pulsed I-V characteristics of the devices since, as shown in section 4, pulsed I-V characteristics may be strikingly different than dc characteristics, and it is the pulsed characteristics of the device which reflect its behavior under rf drive conditions. One of the most important design considerations when dealing with PHEMTs is stability analysis. Because PHEMTs typically have much more gain than MESFETs, circuit stability over a wide range of frequency, bias, or temperature is much more difficult to achieve in PHEMT based MMICs. Traditional K-factor analysis (where $K>1$ ensures stability) does not always accurately predict actual circuit stability. It is essential that a stability analysis methodology be employed which produces rigorous results, particularly when circuit topologies become complex. Such a methodology does exist. The technique uses a network determinant function (NDF) to calculate the number of encirclements at infinite frequency. If the number of encirclements is zero, the circuit is stable.

Fabrication of PHEMT-based MMICs follows the same sequence of steps, and involves the same process operations, as does fabrication of MESFET-based circuits. However, since PHEMT circuits are generally designed for optimum performance, advanced fabrication techniques are usually required to minimize the effects of parasitic elements such as gate resistance and source inductance. Gate resistance is particularly problematic since optimum performance often requires the use of sub-half micron gate-lengths (especially true for frequencies above 10 GHz). Due to the trapezoidal shape of conventional gate fingers, it is not possible to deposit enough gold in a quarter micron gate-length structure to produce acceptably low values of gate resistance. The standard

solution to this problem is to create a T-gate geometry in which a large (0.5 - 1.0 μm) cap structure is placed on an underlying gate stripe of desired length (0.1 - 0.25 μm). Most wafer fabs have adopted a variation on one of two basic T-gate processes. The tri-layer resist process utilizes a three layer e-beam sensitive resist stack generally consisting of a bottom layer of PMMA, a middle layer which is composed of a co-polymer of PMMA and PMAA, and a thin top layer of PMMA. A single e-beam exposure is used and the T profile is created by differential sensitivity of the resist layers to the developer chemistry, the middle co-polymer layer being most sensitive and therefore developing a wider cross section. The thin top layer of resist serves to create a lift-off lip while the bottom most layer of PMMA defines the actual gatelength. The principle advantage of the tri-layer process is the fact that only one pass through the e-beam system is required (e-beam writing can be very time consuming, depending upon the type of machine used). Drawbacks to the tri-layer process include the sensitive dependence of gate shape on resist/developer chemistry and the fact that the gate foot is exposed through a thick resist stack which limits the degree of control over gatelength. An alternative approach is to use a bi-layer resist process in which a very thin layer of PMMA is used for the gate foot exposure (thereby providing a high level of control over gatelength) followed by a second application of resist and a second e-beam exposure to define the cap of the T-gate structure. Care must be taken to avoid intermixing of the resist layers and inadvertent exposure of the lower resist layer during processing of the upper layer. The main drawback to the bi-layer approach is the need for two e-beam exposures. Advantages include the use of a very thin resist for definition of the gate foot and the fact that the gate geometry is directly controlled by the e-beam system rather than by differential development of the resist layers.

Parasitic source inductance can be a problem in large gate periphery power circuits due to the use of airbridged source metal to connect parallel source pads in the typical interdigitated FET structure. Most conventional MMIC layouts specify a 4 mil chip thickness with via hole connections between the backside ground plane metal and frontside source pads at the ends of each interdigitated FET. Interior source pads are then connected to the ground pads using air bridges. Parasitic source inductance due to the airbridge structures can severely degrade circuit performance, particularly at millimeter-wave frequencies. Using advanced processing techniques, it is possible to reduce the chip thickness (typically to 2 mils) and to create individual source vias, that is, direct via hole connections between the backside ground plane and each individual source pad of the FET. Such a process relies on a highly anisotropic via etch chemistry which can place stringent demands on the resist which is used for via patterning. Figure 13 shows schematically the differences between the conventional FET layout and the individual source via layout.

One critical difference between the behavior of MESFET wafers and PHEMT wafers during circuit fabrication is in the relationship between gate recess depth and device saturated current. This relationship is critical in a conventional recessed gate device because

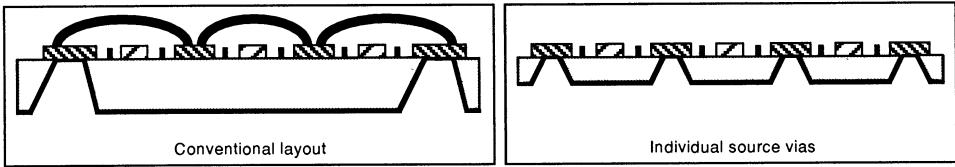


Figure 13. Schematic cross-sections of a typical interdigitated PHEMT showing the conventional layout (left) with via holes at each end of the device, and a non-conventional layout (right) with individual source vias to reduce parasitic inductance and improve rf performance.

the ability to etch the gate recess to the correct depth (as monitored by ungated saturated current) determines how reproducibly FET parameters (I_{dss} , I_{max} , G_m , etc.) can be defined from wafer to wafer, and from process lot to process lot. Because the mobile charge in a MESFET channel is distributed throughout the channel, ungated current falls off in a reasonably predictable manner with increasing gate recess depth. A trained etch operator can reliably etch a MESFET wafer to within $\pm 2\%$ of the desired target value for current without difficulty. The situation is considerably different in PHEMT wafers, however. Because so much charge is contained in such a confined region of the device channel (the so-called two dimensional electron gas), current falls off very slowly until the depletion region generated by the etched free surface begins to modulate the channel charge, at which point the fall off in current with increasing etch depth is extremely rapid. This behavior makes it very difficult to accurately hit an etch target wafer after wafer, and thus I_{dss} values in a PHEMT process tend to show far more variation than in a MESFET process. On the other hand I_{max} and G_m , which are primarily functions of the material structure, tend to be quite well controlled in a good PHEMT process. One technique which is often used to overcome problems with etch control and uniformity is to utilize a dry (plasma) etch process in conjunction with a built-in etch stop layer. Various chlorine based plasma etchants show very good selectivity in etch rate between GaAs and AlGaAs. While this technique is viable for the wide recess in a double recess power process, problems with residual etch damage and control of ungated channel dimension make it more problematic for use in the actual gate recess process.

Because widespread use of PHEMTs is a relatively recent phenomenon and because their structure is fundamentally different from that of GaAs MESFETs, concerns have arisen over the long term reliability of PHEMT devices and circuits. While the amount of PHEMT reliability data which currently exist is somewhat limited, the data that do exist suggest that PHEMT reliability is quite similar to that found for MESFETs. In particular, the measured activation energies are the same for MESFETs and PHEMTs. Figure 14 shows an Arrhenius curve based on temperature accelerated dc life testing of discrete PHEMT devices. Three life tests were conducted on 1.2 mm power PHEMT devices at channel temperatures of 250°C, 275°C, and 330°C. Over 88,000 cumulative device hours

were accumulated. The data show a very good fit to the 1.5 eV activation energy used for GaAs devices with a projected median life at 125°C channel temperature of 2E8 hours. In a separate rf driven life test at a channel temperature of 95°C, reliable operation beyond 10,000 hours has been verified.

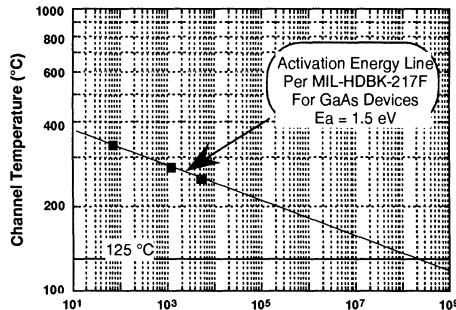


Figure 14. DC life test data measured on 1.2 mm power PHEMT devices at three channel temperatures showing an activation energy of 1.5 eV and a median life of 2×10^8 hours at 125°C channel.

6. Examples of PHEMT MMIC Performance

The state-of-the-art in PHEMT performance is continuously evolving as device and circuit technologies continue to improve. New performance standards are announced on an almost monthly basis. However, it is possible to give a snapshot view of the state-of-the-art in PHEMT MMIC performance as of mid- to late-1994.

The excellent power and low noise performance characteristics of discrete PHEMT devices which were highlighted in section 4 have translated into equally impressive monolithic integrated circuit performance. Tables 3 and 4 give a few examples of recent PHEMT MMIC results for low noise and power circuits respectively.

Figures 15 and 16 show wideband low noise performance and X-, Ku-band power performance respectively for two selected PHEMT MMIC amplifiers. Figure 15 compares the low noise performance of the PHEMT amplifier to that of the same amplifier fabri-

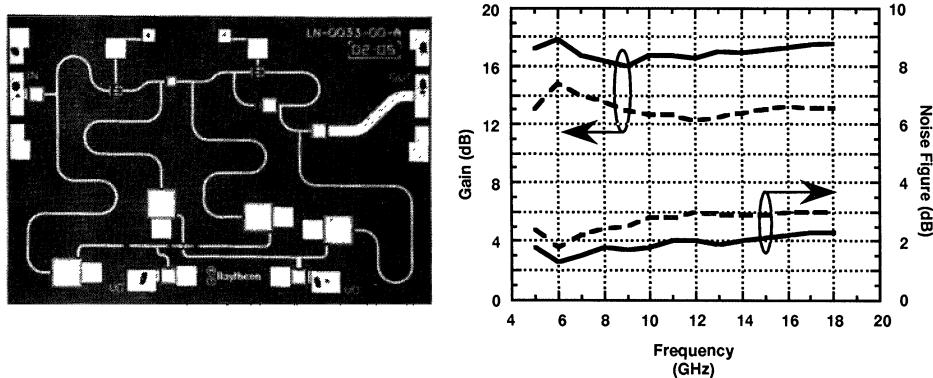
Table 3. Low Noise PHEMT MMIC Amplifier Results

Frequency	Noise Figure	Gain
6-18 GHz	2.25 dB	18 dB
9-11 GHz	1.5 dB	32 dB
43.5-45.5 GHz	3.2 dB	20 dB
60 GHz	4.8 dB	15 dB

Table 4. Power PHEMT MMIC Amplifier Results

Frequency	Power Out	Gain	PAE
900 MHz	1.5 W	27 dB	59%
2.4 GHz	2.0 W	20 dB	48%
4.3-5.0 GHz	8.0 W	24 dB	50%-60%
6-18 GHz	2.5 W	20 dB	18%
8-13 GHz	3.2 W	16 dB	40%
18-20 GHz	3.5 W	13 dB	35%
42-46 GHz	630 mW	15 dB	17%
60 GHz	335 mW	11 dB	10%

cated using an optimized quarter micron ion implant low noise process (the circuits were optimized for each process separately). The PHEMT process produces 4 dB higher gain with 0.75 dB lower noise figure across the entire band compared to the MESFET process. The low noise figures and very high power added efficiencies which are evident in the data of Tables 3 and 4 are what underlie the drive to insert PHEMT technology into both military and commercial systems, as discussed in the next section.

*Figure 15. Photograph and rf noise performance of a wideband low noise amplifier processed using both 0.25 μ m MESFET and 0.25 μ m PHEMT technologies.*

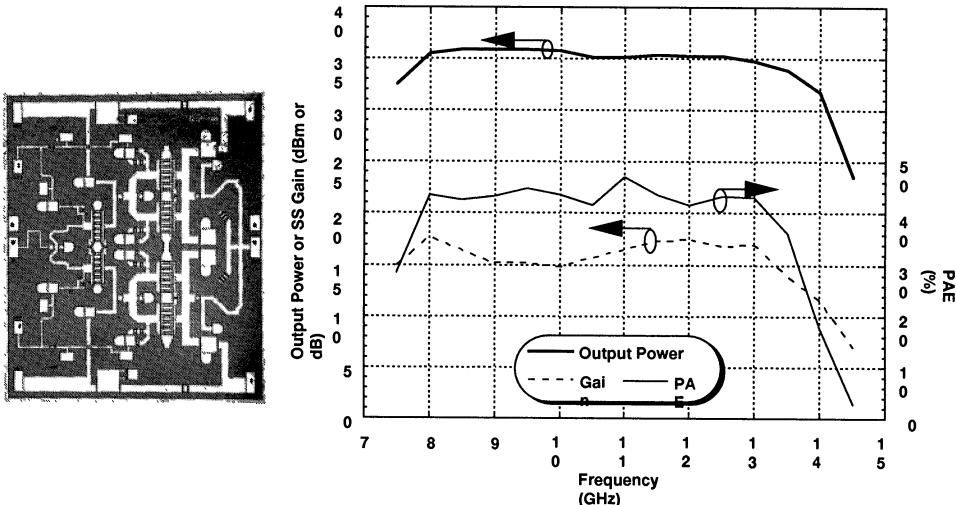


Figure 16. State-of-the-art power performance for a medium band PHEMT power amplifier.

Applications of PHEMT Technology

As noted in the Introduction, the development of a manufacturable PHEMT technology was originally driven by the need for ultimate performance in advanced military systems. Improvements in output power, gain, efficiency, and noise figure at the device level translate directly into increased range, sensitivity, and accuracy in radar-based weapons, communications, and electronic warfare systems. The U.S. department of defense has begun to field MMIC-based systems, with PHEMT devices planned for future insertions. For example, Raytheon is currently working to insert PHEMT low noise amplifiers into two existing missile systems for contracted future buys. The use of PHEMT LNA's in the advanced medium range air-to-air missile (AMRAAM) and the Patriot surface to air missile will result in improved guidance characteristics for both missiles. These insertions are planned for the 1995-1996 timeframe. An excellent example of the use of PHEMT technology in military radar systems is the advanced anti-air warfare (advanced AAW) ship-board defense system. The use of PHEMT power, driver, and low noise amplifiers in the building block transmit/receive module for this system provides extended range and resolution capabilities compared to a MESFET based system.

The benefits of PHEMT technology are being recognized in the commercial arena also, and the focus on low cost PHEMT manufacturing which has been supported by government programs targeted at military systems insertions has allowed commercial appli-

cations engineers and designers to begin to build PHEMT technology into their systems. For example, the very high power added efficiency of PHEMT devices along with their excellent output power capability make them superior devices for satellite based global communications systems such as the IridiumTM and Globalstar systems currently under development. These systems are planned to be operational by 1998. High power added efficiency at low power consumption levels make PHEMTs attractive for applications requiring battery powered operation, such as cellular phones. In addition, a number of automotive radar systems utilizing millimeter-wave devices are currently being developed, including collision avoidance, blind spot detection, vehicle position location (with GPS), and true ground speed detection systems. Within ten years, some or all of these systems should be as common as airbags and anti-lock brakes are today. Driven by these types of high volume commercial applications PHEMT technology will continue to mature rapidly in the next five to ten years, with higher yields and lower costs creating further opportunities for the technology.

TMIridium is a trademark and service mark of Iridium Incorporated.

Acknowledgements

The authors would like to thank Drs. S. Shanfield and S. Brierley of Raytheon, Dr. R. Pucel of RCP Consultants, Dr. L. Nguyen of Hughes Research Laboratories, and Professor L. Eastman of Cornell University for their contributions to this paper.

- [1] R. Dingle, H.L. Stormer, A.C. Gossard, and W. Wiegmann, *Appl. Phys. Lett.* 33, 665 (1978).
- [2] E.E. Mendez, P.J. Price, and M. Heiblum, *Appl. Phys. Lett.*, 43, 294 (1984).
- [3] S.K. Brierley, *J. Appl. Phys.*, 74, 2760 (1993).

PSEUDOMORPHIC HEMTS: DEVICE PHYSICS AND MATERIALS LAYER DESIGN

THOMAS GRAVE

Siemens AG

Corporate Research and Development

D-81730 Munich

Germany

ABSTRACT: In this paper, pseudomorphic HEMTs on GaAs substrate and both lattice-matched and pseudomorphic HEMTs on InP substrate are treated. First, the basic requirements are considered that must be met by any HEMT epitaxial layer sequence. They are compared to the fundamental possibilities for layer growth that exist within the system of III-V compound semiconductors, and the five most common HEMT layer structures are discussed. An elementary introduction into HEMT device physics is given with special emphasis on the influence of specific layer schemes on charge control. The small-signal equivalent network is used to point out differences between intrinsic and extrinsic device due to the presence of parasitics. Guidelines for HEMT design with respect to analog low-noise and power applications are given. Most optimization considerations are illustrated by means of GaAs-based HEMT examples. Properties of InP-based devices are treated more concisely. Overviews over the state-of-the-art concerning high-frequency, low-noise and power performance are presented for both types of HEMTs, and a selection of future concepts for further improvements is outlined.

1. Introduction

Present Situation. Today, high electron mobility transistors (HEMTs) are the most important analog electronic devices for the frequency range between 10 and 100 GHz. They achieved this position by their enormous versatility: HEMTs are used in low-noise as well as power circuits, offering outstanding performance in both applications.

Particularly as low-noise devices, HEMTs are rivaled by no other technology. For power purposes, two additional types of III-V compound transistors might be considered, but they are potential competitors for the HEMT only in the lower part of its frequency range. The first alternative is the well established GaAs MESFET, usable up to approximately 30 GHz, but with inferior power gain. Recently, heterobipolar transistor (HBT) technology has established a second possibility. It is presently

maturing rapidly, with prospective applications up to, say, 50 GHz. For higher frequencies, however, HEMTs will stay the only available solid-state power device for the foreseeable future.

Consequently, HEMTs are the only choice if a *universal* technology for microwave and millimeter wave ICs is required. This situation is reflected by the widespread use of HEMTs - particularly GaAs-based pseudomorphic HEMTs - in analog communication and information systems by the major companies in this field worldwide. How did this prominent role in high-frequency electronics evolve historically?

HEMT History. The first papers on the pseudomorphic HEMT's predecessor, the lattice-matched AlGaAs/GaAs HEMT, were published in 1980 by Mimura *et al.* [1] and Delagebeaudeuf *et al.* [2]. The AlGaAs/GaAs HEMT makes use of the fact that a two-dimensional electron gas (2DEG) of extremely high carrier mobility forms at the heterointerface between AlGaAs and GaAs if the AlGaAs is appropriately *n*-doped (Dingle *et al.* [3], Störmer *et al.* [4]). Due to the small discontinuity of the conduction band energy at the heterointerface and the missing barrier between 2DEG and substrate, the current that can be achieved with these devices is relatively low, and the confinement of the carriers in the channel is poor by today's standards. With sufficiently short gates, however, low noise figures and high current-gain cutoff frequencies can be obtained so that these HEMTs became rapidly acknowledged to be well suited for analog receivers and digital ICs.

It was soon recognized that an energy barrier between channel and substrate would significantly improve the HEMT concept. Such a barrier can be achieved by the introduction of a pseudomorphic InGaAs channel between the GaAs buffer and the supply layer, transforming the HEMT into a PHEMT. First PHEMT results were published in 1985 by Rosenberg *et al.* [5] who used GaAs also for the supply layer, and Ketterson *et al.* [6] who introduced the standard layer sequence of today, i.e. GaAs buffer, InGaAs channel and AlGaAs supply. The new device was soon characterized in more detail (Ketterson *et al.* [7]), and its high potential for microwave and millimeter wave applications was demonstrated (Henderson *et al.* [8]).

However, the PHEMT's performance not only as a low-noise but also as a power transistor could be fully exploited only after the so-called double-heterojunction (DH) layer sequence had been developed. In this case, there is an *n*-AlGaAs supply layer not only on top of the channel but also below it. Power DH-HEMTs with a lattice-matched GaAs channel were reported by Hikosaka *et al.* [9] in 1985. The presently most common power HEMT structure was created in 1988 and 1989 when Smith *et al.* [10,11] sandwiched a pseudomorphic InGaAs channel between the two *n*-AlGaAs supplies.

All HEMT variants discussed so far are based on layer sequences grown on GaAs substrate. The development of the other important family of HEMTs, i.e. HEMTs grown on InP, began about 1982. In these transistors, InAlAs barrier layers lattice-matched to the substrate are grown below and above the InGaAs channel. Their advantage over GaAs-based devices is twofold: the conduction band discontinuity at the channel/barrier interfaces is significantly larger, as well as the indium content in the

channel. This means increased electron concentration in the channel, together with improved carrier transport properties. An InGaAs channel lattice-matched to InP contains 53% indium. For pseudomorphic channels, even higher In mole fractions are possible. For comparison: in GaAs-based HEMTs, lattice-matched channels do not contain any indium at all, and in the pseudomorphic case, the content is usually limited to about 25%.

The first report on the successful fabrication of InP-based lattice-matched HEMTs (LMHEMTs) was made by Pearsall *et al.* early in 1983 [12], but due to the difficulties with the growth of the material, it took some more years until the devices showed the expected performance (Hirose *et al.* 1985 [13], Peng *et al.* 1987 [14]).

First results with pseudomorphic channels on InP were reported in 1986 (Kuo *et al.* [15]), and in 1988, InP PHEMTs became the first devices to exhibit a current-gain cutoff frequency f_T in excess of 200 GHz (Mishra *et al.* [16]). The result was achieved with an indium content of 62% in the channel. However, on InP substrate, the advantages of PHEMTs over LMHEMTs are not as pronounced as in the GaAs system.

The domain of HEMTs grown on InP are ultra-high frequency low-noise applications. At high drain voltages, gate leakage currents and impact ionization in the channel are hard to suppress so that these devices have for a long time not been regarded to be suitable for power application. Recently, attempts have been made to overcome these problems. They will be discussed at the end of this paper when advanced future concepts are presented.

Scope of this Article. In the present paper, we will treat HEMTs only with analog applications in mind. Within the family of HEMTs grown on GaAs, LMHEMTs and PHEMTs differ with respect to layer sequences and device physics. Since the AlGaAs/GaAs LMHEMT has become outdated for analog applications, we will not discuss it here. For HEMTs grown on InP, the situation is substantially different: LMHEMTs and PHEMTs share the same layer sequences and are often very similar in performance. There is only a difference in indium content in the channel, in many cases very small. For these reasons, *we shall discuss both LMHEMTs and PHEMTs on InP.* Thus, the subject of this review could be more precisely described by *HEMTs with InGaAs quantum well channels.*

In principle, we intend to avoid lengthy theoretical discussions, but rather want to look at HEMT device physics and epitaxial layer design from a practical point of view. This means: the present article tries to clarify the problems usually faced by a device designer when a HEMT layout for analog low-noise or power applications has to be developed. Therefore, a basic understanding of

- the possibilities and limitations of epitaxial layer design,
- the basic current-voltage characteristics of HEMTs,
- the physics of charge carrier control in HEMT channels,
- and the performance requirements for a specific application

is tried to be given in the following text.

2. PHEMT Epitaxial Layer Structure

2.1. EPITAXIAL GROWTH

The binary III-V compound semiconductors are all of the same zinc-blende crystallographic structure. Solid solutions can be formed among them, leading to a large number of ternary and quaternary alloys. Aside from technical problems, it is generally possible to grow these alloys on each other (by means of various epitaxial methods), provided

- the lattice constants are equal (*lattice-matched growth*),
- or the lattice constants do not differ too much, and the thickness of the layer mismatched to the growth substrate is kept below a critical limit. In this case, the strain within this layer will not relax, and the formation of crystalline dislocations is avoided (*pseudomorphic growth*).

If two semiconductors with different energy gaps are grown on each other, discontinuities of the conduction and valence band edges arise at the heterointerface. In most cases, the conduction band offset between adjacent layers ΔE_C is close to 2/3 of the total band gap difference ΔE_G .

There are basically two techniques that can be used for the growth of HEMT layers: molecular beam epitaxy (MBE) and metal-organic vapor phase epitaxy (MOVPE, also called MOCVD, metal-organic chemical vapor deposition).

Today, the majority of HEMT wafers is grown by MBE. This is caused by a number of reasons: the interfaces between layers are very smooth, changes in layer composition and doping can be made very abruptly, and the control of background impurities is comparatively easy. A disadvantage is the small wafer throughput of most MBE systems.

The high production capacity that can potentially be achieved has for a long time been judged the major advantage of MOVPE. Moreover, alloys containing phosphorus can be easily grown, contrary to MBE. On the other hand, MOVPE has a number of severe drawbacks. The control of background impurities is painstaking (especially if the layers contain aluminum) and only practicable if a large number of identical wafers has to be fabricated. Additionally, it is nearly impossible to "switch off" high doping concentrations abruptly enough if an undoped layer must be grown over a heavily doped one (which is the case in DH-PHEMTs, for instance). And finally, the MOVPE principle relies on extremely toxic gases like arsine and phosphine. Their handling necessitates extensive safety precautions and, in most countries, is prohibited close to residential areas.

2.2. BASIC LAYER SEQUENCE

PHEMTs belong to the family of field effect transistors (FETs): the current flow between two ohmic contacts - source and drain - is controlled by a third electrode, the gate. In most cases, the gate is a metal/semiconductor contact (Schottky contact).

The main difference to conventional FETs is the quantum well (QW) channel. This is a thin layer of *undoped* material where the energy gap E_G is lowest compared to the rest of the layers (see Figure 1). This means that electrons from the doped high-bandgap supply (or barrier) layer, located between the gate and the QW, are transferred into the QW because this is energetically favorable for them. Here, they are locally separated from their parent donor ions. Since the Coulomb potentials connected to the donors possess a finite range, electrons in the channel would be scattered by them if the doping would extent too close to the supply/channel interface. Therefore, to prevent mobility degradation from ionized impurity scattering, a few nanometers of the high-bandgap material directly adjacent to the channel are usually left undoped, forming the so-called spacer layer.

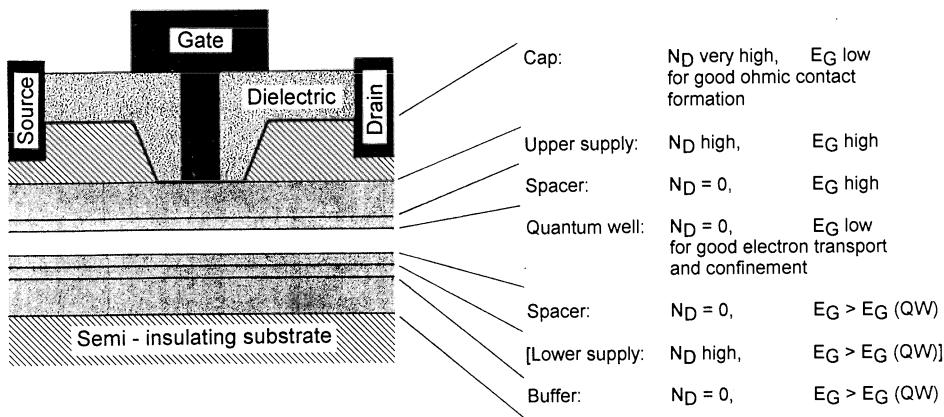


Figure 1. PHEMT layer schematic. N_D is the doping concentration.

Below the channel, there is also material with higher E_G . This creates an energy barrier for the electrons that prevents them from penetrating into the buffer where the control by the gate voltage is ineffective and transport properties are poor. In some types of PHEMTs (among them the most popular layer sequence on GaAs substrate), this lower barrier is not as high as the upper one. There are also cases in which there is additional doping in the region below the QW in order to get high electron concentration in the channel, for instance in power PHEMTs. This additional doping is usually only applied when the energy barrier below the channel is as high as the upper one (DH-PHEMTs).

Between the lower barrier layer (or the channel, if there is no separate barrier employed) and the semi-insulating substrate, the buffer layer extends, usually about one micrometer thick. This large thickness has several reasons: initial growth defects that originate from particles on the substrate surface, for instance, are reduced during the growth process, and any impurities from the growth chamber atmosphere or diffusing out of the substrate are gettered in the lower part of the buffer layer far away from the channel. Impurities in the buffer influence the device behavior even if they are several

hundred nanometers away from the channel, especially under high-voltage bias conditions, since the energy barrier between channel and buffer only reduces electron injection into the buffer but does not completely prevent it. Moreover, the number and energetical position of states in the buffer bandgap determines where the Fermi level is pinned in the gap. This influences the position of the quantum well relative to the Fermi level and, hence, its population with electrons.

The uppermost layer in the PHEMT epi structure is the cap layer. It is necessary since one of the crucial requirements for the upper supply or barrier layer is that it has a large energy gap. It is hardly possible to make good ohmic contacts to large-bandgap materials, and therefore an extra layer optimized for the formation of source and drain contacts must be introduced. The cap should have a low bandgap and a high donor density. In the region of the intrinsic transistor, i.e. below the gate electrode, it must be etched away, since here we want the current flow to take place in the quantum well. Outside the intrinsic device, the cap can contribute to a highly conductive connection to the metal contacts by forming a shunt to the channel.

2.3. MATERIALS FOR PHEMT LAYERS

We now want to discuss which of the III-V compounds can be possibly used for the individual layers of the basic PHEMT structure. Figure 2 shows the bandgap energy *vs.* the lattice constant for these compounds. Two semi-insulating substrate materials are available for epitaxial growth: GaAs and InP.

2.3.1. PHEMTs on GaAs Substrate

To GaAs, there exists no III-V compound of the same lattice constant but with smaller bandgap. Thus, for small-gap channels on this substrate, there is only the choice of *pseudomorphically* grown InGaAs. If the lattice constant is not allowed to be too different from that of GaAs, the InGaAs bandgap is not very much smaller, and the barrier between buffer and channel is low. In practice, indium contents between 20 and 25% are used.

Fortunately, there are at least a number of compounds lattice-matched to GaAs but with higher E_G . These are AlAs, $\text{In}_{0.48}\text{Al}_{0.52}\text{P}$ and $\text{In}_{0.48}\text{Ga}_{0.52}\text{P}$. The growth of materials containing phosphorus is difficult with MBE so that their use has only very recently started together with efforts to fabricate PHEMTs based on MOVPE.

The conventional choice for a high-bandgap semiconductor lattice-matched to GaAs is AlGaAs. Two basic types of PHEMTs on GaAs exist: in the first case, the InGaAs channel is grown directly on a GaAs buffer, and high-bandgap AlGaAs is used only for the upper supply layer. Until recently, this has been the standard PHEMT design on GaAs, frequently called single-heterojunction PHEMT (SH-PHEMT). In the second case, AlGaAs is also grown between GaAs substrate and InGaAs channel to create a higher energy barrier to the substrate. This design is usually called double-heterojunction PHEMT (DH-PHEMT). It is considered to be more difficult to grow because the growth front of AlGaAs tends to be rough creating an imperfect interface to the channel.

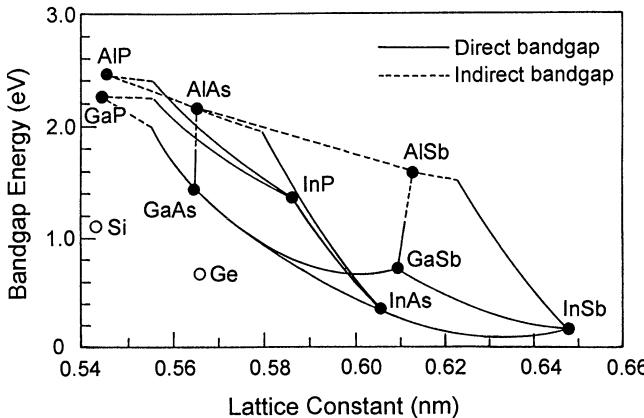


Figure 2. Bandgap energy vs. lattice constant for major III-V compounds.

The aluminum content x in $n\text{-Al}_x\text{Ga}_{1-x}\text{As}$ supply layers is limited by two effects. For $x > 0.45$, the bandgap changes from direct to indirect. The conduction band minimum is not located at the center of the Brillouin zone any more but at the three-fold degenerate X point close to the zone edge [17]. This means an enormous increase in the density of states in the lowest part of the conduction band and leads to insufficient transfer of electrons into the InGaAs channel. The second effect limiting the aluminum content in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ is the deep donor level DX which inevitably appears together with a shallow donor level when an n -type silicon doping is present (see discussion in section 5.1.). This causes ineffective donor activation and undesirable transient charge redistribution effects between these levels for x larger than about 0.23. Thus, for PHEMTs grown on GaAs substrate, the conduction band discontinuity between InGaAs quantum well and AlGaAs supply is limited to approximately 0.3 eV in most practical cases.

The uppermost layer in a PHEMT structure on GaAs substrate is usually a highly n -doped cap, also made of GaAs.

2.3.2. LMHEMTs and PHEMTs on InP Substrate

On InP, the conditions for the formation of deep quantum well channels are more favorable. Due to the aforementioned difficulties with the MBE growth of phosphides, the buffer grown on this substrate is usually not InP but lattice-matched $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ with about the same bandgap (Figure 2). On top of this buffer or lower barrier layer, a lattice-matched $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ channel is grown, followed by the upper $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ barrier. The conduction band discontinuity between channel and barriers is about 0.52 eV in this LMHEMT which is considerably larger than in the PHEMT on GaAs. As a low-bandgap cap, the same material is used as for the channel ($\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$).

The indium content in the channel can easily be increased to over 70%, creating a PHEMT. Compared to the case of GaAs, the performance advantage of a PHEMT over

an LMHEMT on InP is smaller. Sometimes, it may even be obscured by different qualities of the employed manufacturing processes.

2.4. FIVE PROTOTYPE HEMTS

The layer structures of the most popular HEMT types are compiled in Table 1. For comparison, the very first HEMT, i.e. the lattice-matched AlGaAs/GaAs HEMT, is also included. As already mentioned, this HEMT lacks an energy barrier between channel and buffer. Its triangular quantum well at the buffer/supply interface supports only relatively small drain currents and does not effectively suppress electron injection into the buffer at high drain voltages.

These deficiencies are partly removed with the second layer sequence in Table 1, the SH-PHEMT. Presently, this is the "standard" HEMT device for low-noise applications in the microwave and millimeter wave frequency range. However, the energy barrier to the buffer is quite small so that it cannot effectively confine the electrons to the channel anymore if the drain voltage approaches the order of 5 V.

TABLE 1. The five most popular HEMT layer structures

Cap	GaAs	GaAs	GaAs	$In_yGa_{1-y}As$ $y = 0.53$	$In_yGa_{1-y}As$ $y = 0.53$
Upper supply	$Al_xGa_{1-x}As$ $x \leq 0.23$	$Al_xGa_{1-x}As$ $x \leq 0.23$	$Al_xGa_{1-x}As$ $x \leq 0.23$	$In_xAl_{1-x}As$ $x = 0.52$	$In_xAl_{1-x}As$ $x = 0.52$
Cond. band discont. ΔE_c	≈ 0.20 eV	≈ 0.30 eV	≈ 0.30 eV	≈ 0.52 eV	> 0.52 eV
Channel	GaAs lattice matched	$In_yGa_{1-y}As$ $y \approx 0.20$ pseudomorphic	$In_yGa_{1-y}As$ $y \approx 0.20$ pseudomorphic	$In_yGa_{1-y}As$ $y = 0.53$ lattice matched	$In_yGa_{1-y}As$ $y \geq 0.53$ pseudomorphic
Cond. band discont. ΔE_c	0	≈ 0.12 eV	≈ 0.30 eV	≈ 0.52 eV	> 0.52 eV
Lower supply or buffer	GaAs	GaAs	$Al_xGa_{1-x}As$ $x \leq 0.23$	$In_xAl_{1-x}As$ $x = 0.52$	$In_xAl_{1-x}As$ $x = 0.52$
Substrate	GaAs	GaAs	GaAs	InP	InP
Remarks	historically first HEMT	single heterojunction PHEMT on GaAs	double heterojunction PHEMT on GaAs	standard HEMT on InP	advanced HEMT on InP

In the DH-PHEMT on GaAs (third column of Table 1), the confinement is improved. Additionally, a higher current than in the standard PHEMT can be realized. These properties enable the DH-PHEMT to be used as a power transistor. Its development meant a breakthrough for PHEMTs which thereby became a universal microwave technology applicable for receivers *and* transmitters.

The last two types of transistors in Table 1, the LMHEMT and PHEMT on InP, differ with respect to indium content of the channel and, hence, quantum well depth. InP-based devices are mostly used for purposes where lowest noise figures or extreme high-frequency performance are required. Though they are capable of even higher

drain currents than GaAs DH-PHEMTs, they are difficult to optimize for power applications (see section 9.2.).

2.5. LIMITATIONS FOR PSEUDOMORPHIC CHANNEL GROWTH

Up to now, we have only discussed the composition of the layers in a PHEMT but not their individual thicknesses. For pseudomorphic layers, these are limited depending on the amount of mismatch to the growth substrate.

2.5.1. *Pseudomorphic Growth on GaAs Substrate*

This situation was for the first time investigated by Matthews and Blakeslee [18] who used experiments with alternating layers of GaAs and GaAsP to derive a theoretical expression for the critical layer thickness L_C which cannot be exceeded if misfit dislocations must be avoided. The validity of the model for the system InGaAs/GaAs was demonstrated by Andersson *et al.* [19] by means of photoluminescence measurements. The thickness L_C up to which the mismatch of the $\text{In}_y\text{Ga}_{1-y}\text{As}$ layer can be accommodated by elastic strain is given by

$$0.07y = \frac{a(1 - \sigma/4) \left[\ln(L_C \sqrt{2}/a) + 1 \right]}{2\sqrt{2}\pi L_C(1 + \sigma)} \quad (1)$$

where y is the indium content, a is the GaAs lattice constant ($a = 0.565$ nm), and σ is Poisson's ratio (a dimensionless combination of stiffness constants which is 0.23 for GaAs). The curve defined by Equation 1 is shown in Figure 3.

For $y = 0$, the growth is lattice-matched, and the magnitude of L_C is infinite (this is nearly the case in the AlGaAs/GaAs HEMT). On the other hand, for large y , the critical thickness is only a few nanometers. Quantum wells that are too narrow contain only a small number of charge carriers, and the transport properties of these carriers are degraded due to excessive interface scattering. The best compromise with respect to transistor performance (number of carriers in the channel and carrier mobility) is found around $y = 0.2$ and a channel thickness of 12 nm. Two other commonly cited combinations are also indicated in Figure 3.

2.5.2. *Pseudomorphic Growth on InP Substrate*

Compared to the situation on GaAs, even lattice-matched quantum wells have a large ΔE_C to the barrier layers so that there is less need to drive pseudomorphic growth towards its crystallographic limits. For lattice-matched QWs, the thickness is mainly determined by device physical considerations and not by fundamental limitations. In case of pseudomorphic growth, one should expect the calculations of Matthews and Blakeslee [18] to hold also for InP.

In 1985, however, People and Bean [20] published a model for critical layer thicknesses that is not based on the mechanical equilibrium considerations of Ref. [18] but on energy balance. Recently, this new model has been experimentally verified for InGaAs/InAlAs heterostructures by Taguchi *et al.* [21]. It leads to significantly larger critical thicknesses than the older model. For $\text{In}_y\text{Ga}_{1-y}\text{As}$ with $y = 0.8$, an L_C of 20 to 25 nm is given in Ref. [21], opposed to 3.5 nm predicted by the mechanical equilibrium model. For $y = 0.7$, the values are 50 to 100 nm opposed to about 8 nm. This means that the channel dimensions of InP PHEMTs up to quite high indium percentages in the QW could be still determined by device physical considerations rather than by growth limitations. In most published InP PHEMTs, the QW thicknesses are below the maxima given by the energy balance model but larger than allowed by the theory based on mechanical equilibrium.

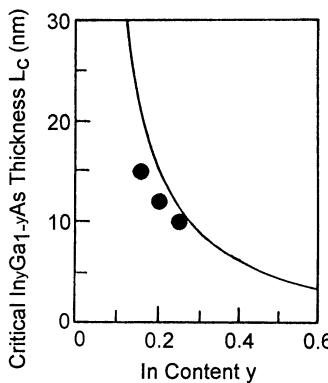


Figure 3. Critical InGaAs thickness as a function of In content after Matthews and Blakeslee [18].

Up to now, most published PHEMTs in the system InGaAs/AlGaAs on GaAs do not exceed the limitations given by the model of Matthews and Blakeslee. The performance of these devices, however, could considerably profit from deeper or wider quantum wells.

2.6. HEMT CROSS-SECTIONAL DIMENSIONS

The conduction band structures of HEMTs on GaAs and InP substrate are compared in Figure 4. The profiles are approximately to the same scale to give an impression of the larger QW cross section of the HEMTs on InP.

Apart from the mere QW sizes, several other differences can be seen. The built-in voltage $q\phi_B$ at the gate/semiconductor interface is considerably smaller for InAlAs than for AlGaAs (in practice, 400 compared to over 700 mV). This is one - but not the only - reason for the smaller thickness d of the supply layer of InP-based HEMTs since it means a smaller depletion depth under the gate. In the standard PHEMT on GaAs and the HEMTs on InP, the barrier layers below the channel are normally undoped (here, the conduction band energy E_C increases with respect to the Fermi level E_F as a

function of depth because E_F is pinned in the buffer at or below mid-gap). In the DH-PHEMT, generally both barriers are doped causing a more symmetrical shape of the QW. Here, without the doping layer below the channel, the bottom of the QW would be too high relative to the Fermi level due to the position of E_F in the AlGaAs buffer.

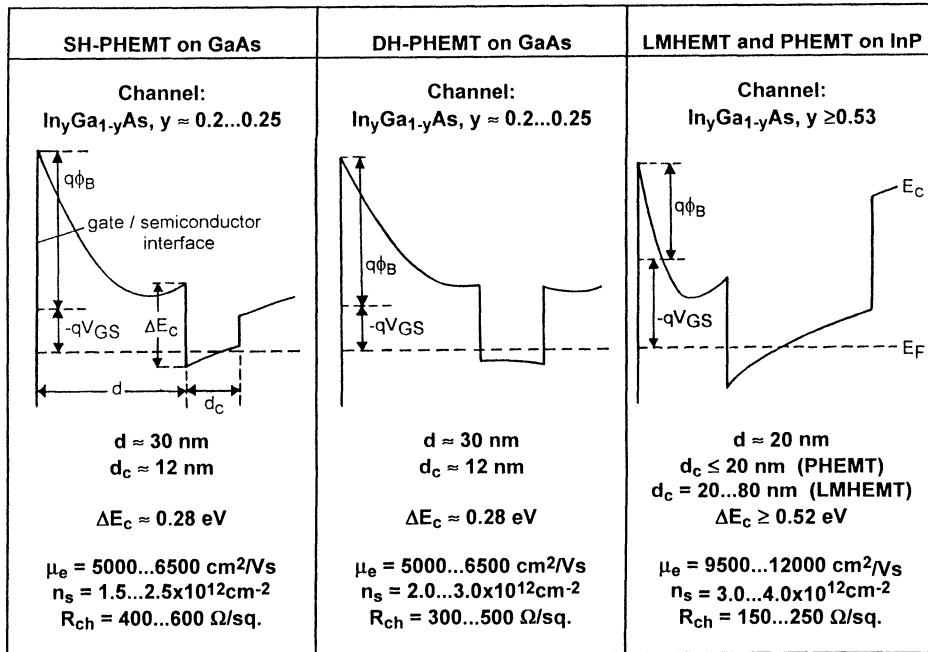


Figure 4. Conduction band structures and typical quantum well transport data for HEMTs on GaAs and InP substrate. μ_e : electron mobility, n_s : electron concentration in channel, R_{ch} : channel resistance.

3. Transport Properties of InGaAs Channels

3.1. VELOCITY-FIELD CHARACTERISTICS IN THICK LAYERS

The different indium concentrations in the QWs of GaAs- or InP-based HEMTs are one major reason for their different performances. This can be seen in Figure 5 which shows stationary electron drift velocities as a function of electric field for the compounds GaAs, $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ and InAs. The experimental data were obtained with a constant electric field in bulk crystals or thick lattice-matched epitaxial layers. In a narrow HEMT channel, the transport is not stationary, and the situation is additionally complicated by scattering of the electrons at the adjacent interfaces, by scattering at the Coulomb potentials of the dopants behind these interfaces, and by the presence of strain. Thus, Figure 5 represents an idealized situation.

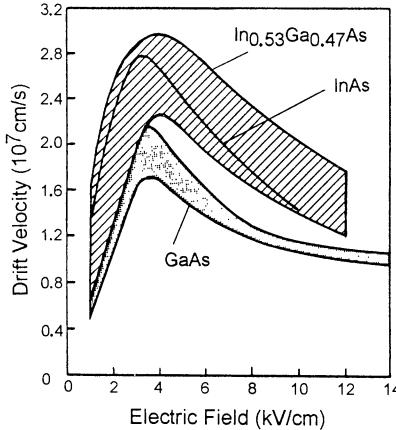


Figure 5. Electron drift velocities in GaAs, InGaAs and InAs vs. electric field.

The shaded area for GaAs comprehends the data by Ruch and Kino [22], Braslau and Hauge [23], and Houston and Evans [24]. It is composed of several distinct regimes. Up to fields E of about 3 kV/cm, the electron velocity v is determined by a constant low-field mobility μ_e (ohmic transport), where $v = \mu_e E$. Beyond the peak velocity v_p measured at fields close to 4 kV/cm, the stationary drift velocity decreases with the field until a high-field value v_{sat} is reached. The hatched area for $In_{0.53}Ga_{0.47}As$ is a compilation of data by different authors done by Littlejohn *et al.* [25]. For unstrained $In_{0.53}Ga_{0.47}As$, μ_e , v_p and v_{sat} are higher than for GaAs. The scattering of the data is considerable, probably due to different quality of the epitaxial growth, but the general superiority of the $In_{0.53}Ga_{0.47}As$ transport properties compared to GaAs is clearly visible.

The case of pure InAs is the most complicated in Figure 5. The low-field mobility is the highest of all three materials discussed here. For high fields, reliable experimental data are difficult to achieve. Thus, a theoretical curve by Brennan and Hess [26] is shown in the figure. The idealized calculation does not take impact ionization into account. However, the low bandgap of InAs (see Figure 2) causes the impact ionization rate to be considerable for the electric fields of interest here, and its neglect has a significant influence on the result of the calculation. Absence of impact ionization allows the electrons to pick up high energies from the electric field without being scattered back into states near the conduction band edge. This means that there is a high transfer rate of electrons from the central band minimum to the higher minima where the transport properties are extremely degraded. This is reflected by the curve in Figure 5 which does not indicate any superiority of the carrier peak velocity in InAs over the velocity in $In_{0.53}Ga_{0.47}As$.

In reality, however, this is not quite true because impact ionization actually plays an important role in high-field transport in InAs. It acts as a permanent cooling mechanism which prevents the electrons from attaining enough kinetic energy for transitions to the satellite valleys, thus confining them to the central band minimum. According to Brennan and Hess [18], this eliminates the decrease of the electron velocity beyond the

peak value at approximately 3 kV/cm in Figure 5 but leads to a continuous rise of v up to fields of 100 kV/cm where a new maximum of $v_{sat} = 8 \times 10^7$ cm/s is reached.

Due to the small dimensions of high-field domains in HEMTs, transport is not stationary, transient effects occur, and the relevance of the results discussed above is difficult to estimate. Nevertheless, indications remain that an increase of the indium content in PHEMT channels to values close to 100% might lead to some drawbacks. Improvement of the transport properties can be coupled to an increase of the impact ionization rate in the channel causing breakdown of the device already at low bias voltages (besides other undesirable effects like large output conductance and the creation of holes).

3.2. LOW-FIELD TRANSPORT IN THIN LAYERS

Compared to bulk material, the situation in thin layers is further complicated by the influence of the adjacent material layers, the interfaces and, often, by the presence of strain. Thus, information of general validity on thin layer transport is even more restricted because carrier velocities and mobilities are not only governed by "intrinsic" properties of the layer material but also by conditions imposed by the layer surroundings and the quality of the epitaxial growth. Additionally, the situation is obscured by the problems with high-field transport. For these reasons, mainly Hall mobility data are available for quantum well structures, and we will make only very few qualitative statements on low-field transport.

Generally, with increasing In content, the effective electron mass in InGaAs decreases, and the 300 K mobility is expected to go up. Strain, on the other hand, increases the effective mass (Liu *et al.* [27]). When the In content y is raised from 0 (unstrained case on GaAs substrate) to 0.3 (strained case), increasing strain and alloy scattering cause a decrease of the mobility at 77 K by about a factor of two (Nguyen *et al.* [28]). At 300 K, the decreasing effective mass by far dominates the counteracting effects, and the mobility improves with In concentration for InGaAs grown on both InP and GaAs substrates (Bhattacharya [29]).

3.3. TRANSPORT IN PRACTICAL HEMT CHANNELS

In this complex situation, we prefer to restrict ourselves to the statement of typical low-field mobilities for the three types of HEMT quantum well channels of Figure 4. In practice, the physical influences discussed above are mostly obscured by effects of the individual epitaxial growth quality and scattering at the Coulomb potentials of the dopant atoms in the supply layers. The proper optimization of the spacers that separate these dopants from the quantum well is important if high mobilities μ_e shall be combined with maximum electron sheet densities n_s . Both n_s and μ_e determine the channel sheet resistance R_{ch} . The quantities given in Figure 4 are typical numbers that can be obtained by Hall effect measurements at HEMT quantum well structures. They do not represent record data.

The HEMTs based on InP feature electron mobilities far superior to those on GaAs. Additionally, their larger conduction band discontinuity at the channel-barrier interface leads to substantially higher n_s . In combination, this means that InP based HEMTs possess significantly smaller sheet resistances than GaAs based devices. Before we discuss how these 2DEG properties translate into HEMT performance, we briefly sum up some basic current-voltage characteristics that the HEMT shares with all field-effect transistors.

4. Current-Voltage (I-V) Characteristics of PHEMTs

4.1. OUTPUT CHARACTERISTICS

For a PHEMT, the dependence of the drain current I_D on the drain-source voltage V_{DS} (output characteristics) is similar to that for a conventional III-V MESFET (MEtal-Semiconductor Field-Effect Transistor, i.e. FET with Schottky gate electrode). As an example, Figure 6 shows an output characteristics of a "standard" SH-PHEMT on GaAs substrate.

In the saturated current regime, I_D is not really constant for a fixed gate-source voltage V_{GS} but still increasing with V_{DS} which means that there is a finite output conductance g_{ds} given by

$$g_{ds} = (\partial I_D / \partial V_{DS}) \quad \Big|_{V_{GS} = \text{const.}} \quad (2)$$

Additionally, the curves I_D (V_{DS}) are not equidistant for the different gate-source voltages. This indicates that I_D is not a linear function of V_{GS} , i.e. that the transconductance g_m is not constant with V_{GS} . The transconductance is defined by

$$g_m = (\partial I_D / \partial V_{GS}) \quad \Big|_{V_{DS} = \text{const.}} \quad (3)$$

PHEMTs are distinguished from other FETs by their high g_m . In the example of Figure 6, the peak value is approximately $g_m / W_G = 525$ mS/mm (W_G : gate width). In combination with $g_{ds} / W_G = 25$ mS/mm, this results in a voltage gain $g_m / g_{ds} = 21$.

4.2. TRANSFER CHARACTERISTICS

The nonlinearity of the dependence of I_D on V_{GS} can be better seen in the transfer characteristics and the transconductance characteristics (Figure 7). The transconductance g_m (V_{GS}) exhibits a pronounced maximum for V_{GS} close to 0.25 V. It is one of the aims of this paper to give some insight into the underlying mechanisms so that measures to influence the shape of this curve in a purposeful way can be identified.

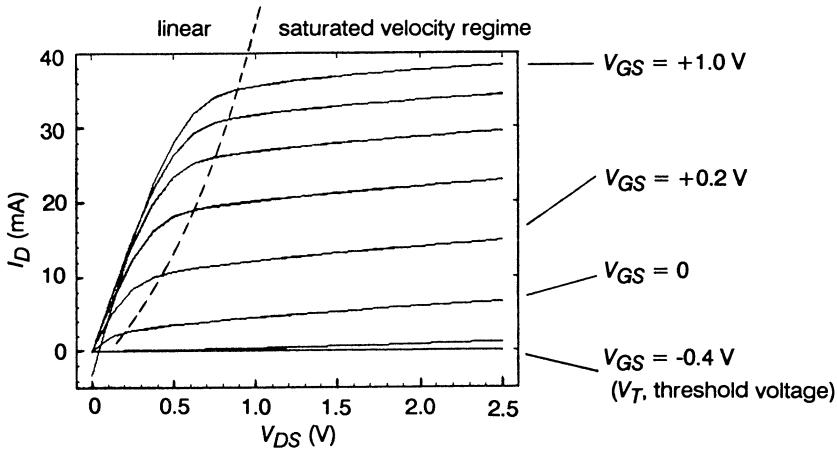


Figure 6. Output characteristics I_D (V_{DS}) of an SH-PHEMT on GaAs substrate with $W_G = 80 \mu\text{m}$ and $L_G = 0.25 \mu\text{m}$.

4.3. TRANSCONDUCTANCE OF AN IDEALIZED PHEMT

In an idealized PHEMT, all charge carriers are confined in the QW in a sheet of negligible thickness. This 2DEG is assumed to be located at $z = d + \Delta d$ (see Figure 8) where d is the thickness of the barrier layer between gate and channel, and Δd takes into account that the effective center of the electron distribution is not directly at the barrier-channel interface but somewhere in the QW.

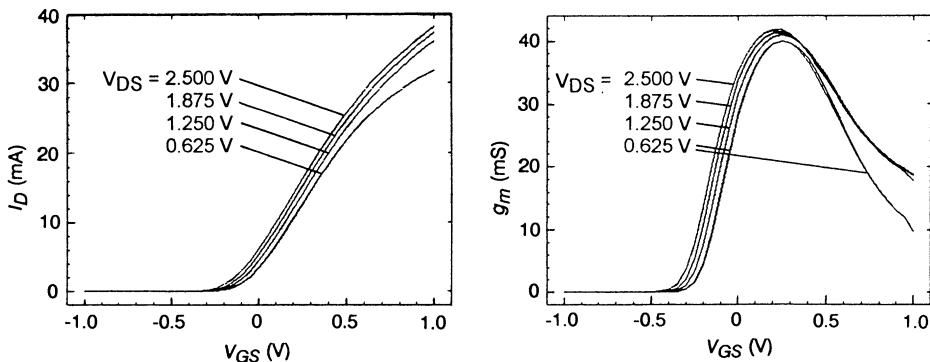


Figure 7. Transfer characteristics I_D (V_{GS}) and transconductance characteristics g_m (V_{GS}) of an SH-PHEMT on GaAs substrate with $W_G = 80 \mu\text{m}$ and $L_G = 0.25 \mu\text{m}$.

The sheet charge of the 2DEG is $Q_s = q n_s L_G W_G$ (L_G : gate length), and the sheet capacitance under the gate C_s is given by

$$C_s = \partial Q_s / \partial V_{GSi} = [\epsilon / (d + \Delta d)] L_G W_G \quad (4)$$

where ϵ is the dielectric constant of the barrier layer material. The gate voltage V_{GSi} of the *intrinsic* transistor must be distinguished from the externally applied V_{GS} because both quantities differ due to the presence of parasitic elements. We will discuss this in chapter 6.

If we assume that the charge carriers move with uniform velocity through all parts of the transistor, and that this velocity is v_{sat} , the drain current is

$$I_D = Q_s v_{sat} / L_G \quad (5)$$

which leads to the following expression for the transconductance of the intrinsic device, g_{mi} (which must also be distinguished from the externally measured g_m):

$$g_{mi} = \partial I_D / \partial V_{GSi} = [\epsilon v_{sat} / (d + \Delta d)] W_G = C_s v_{sat} / L_G . \quad (6)$$

An important consequence of Equation 6 is: for maximum transconductance, the distance d between gate and QW must be minimized. Unfortunately, this also results in a maximum sheet capacitance which is directly proportional to g_{mi} and can be reduced only by a shortening of the gate length.

Additionally, a second problem arises with the reduction of d . At the interface between gate and barrier layer, the built-in voltage $q\phi_B$ causes the conduction band edge E_C to bend upwards with respect to the Fermi energy E_F thus establishing a charge depletion layer. If d is too small, the QW will be lifted relative to E_F , and n_s will be reduced. In extreme cases, it might be possible that the QW cannot be adequately filled with electrons even if maximum positive gate bias is applied to counteract the surface depletion.

Therefore, it is favorable if the doping concentration N_D in the barrier layer is high. High doping has two consequences: it leads to a high n_s (presumed, the charge transfer into the QW is effective), and it causes a more rapid decrease of the surface potential since the depth d_d of the depletion zone is given by

$$d_d \approx (2 \epsilon \phi_B / q N_D)^{1/2} . \quad (7)$$

The common barrier material for GaAs-based PHEMTs, AlGaAs, can be practically doped up to levels of about $N_D = 3...5 \times 10^{18} \text{ cm}^{-3}$. Higher concentrations are useless because they do not translate into higher carrier densities due to decreasing donor activation and increasing electron trapping in deep levels (DX centers).

InP-based HEMTs usually possess InAlAs barrier layers. Better activation compared to AlGaAs allows silicon dopings up to $N_D = 1.5 \times 10^{19} \text{ cm}^{-3}$ so that very small barrier thicknesses d can be employed. Together with the large v_{sat} of $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, this a

key contribution to the realization of ultra-high transconductances with InP-based HEMTs.

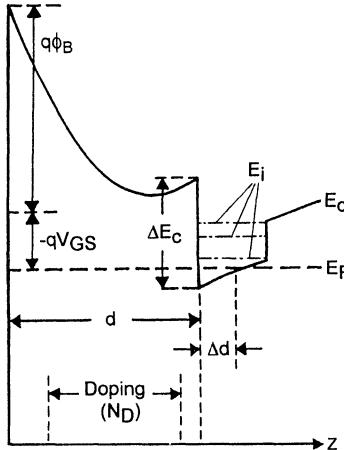


Figure 8. Conduction band diagram of a PHEMT with homogeneously doped supply layer.

However, there are two major reasons for which a high doping cannot be applied throughout the whole thickness d of the barrier layer. The first one was already mentioned in section 2.2.: since the Coulomb potentials of the donor ions have a finite range, the doping must be separated from the QW interface by a spacer typically of the order of 2 nm.

The second reason is the bad quality of Schottky contacts to highly doped semiconductors. Such contacts are characterized by low barriers, high leakage currents under reverse bias conditions, and low reverse breakdown voltages. Therefore, an undoped layer is also appropriate at the gate/barrier interface.

Extremely small barrier thicknesses d can be achieved by the concept of *delta doping* as discussed by Schubert [30] (also called *planar doping* or *pulse doping*). Here, the supply donors are concentrated to a few atomic monolayers. This can be realized by MBE growth interruption of the host crystal during deposition of the impurities. In AlGaAs, planar donor concentrations up to $n_{\text{delta}} = 7 \times 10^{12} \text{ cm}^{-2}$ are reasonable. Analogous to the case of bulk material doping, even higher concentrations can be used in InAlAs due to the better activation efficiency. Figure 9 shows a conduction band diagram for a delta-doped PHEMT which can be identified by the linear potential drop in undoped material.

The preceding discussion provided some insight how the peak value of g_m can be optimized but did not explain the particular shape of the curve $g_m(V_{GS})$. To understand this, charge control in a PHEMT channel must be discussed in more detail.

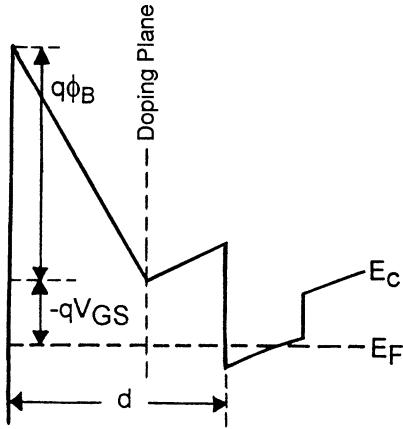


Figure 9. Conduction band diagram of a delta-doped PHEMT.

5. Charge Control in PHEMTs

5.1. SCHRÖDINGER'S AND POISSON'S EQUATIONS

The local distribution of the charge carriers in the direction z perpendicular to the semiconductor-gate interface is described by two equations which have to be solved simultaneously: Schrödinger's and Poisson's equations. A discussion of these equations can be found in the review by Nguyen *et al.* [28] and in the article by Ando and Itoh [31].

Schrödinger's equation is given by

$$[-(\hbar^2 / 2m^*) (\partial^2 / \partial z^2) - qV(z) + \Delta E_C(z)] \phi_i(z) = E_i(z) \phi_i(z) \quad (8)$$

where \hbar denotes Planck's constant, m^* is the effective mass of the electrons in the conduction band, $V(z) = E_C(z) / (-q)$ is the electrostatic potential, $\Delta E_C(z)$ is a function that describes the conduction band discontinuities at the heterojunctions that enclose the quantum well, $\phi_i(z)$ is the envelope electron wavefunction, and E_i are the Eigenvalues of the electron energy in the quantum well. The envelope wavefunction is connected to the complete wavefunction $F_i(z, \mathbf{r})$ by

$$F_i(z, \mathbf{r}) = \phi_i(z) \exp(i\mathbf{q}\mathbf{r}) \quad (9)$$

where the vectors \mathbf{q} and \mathbf{r} are in a plane parallel to the heterojunction.

Poisson's equation is

$$\partial^2 V(z) / \partial z^2 = [q / \epsilon(z)] [N_D^+(z) - N_A^-(z) + N_{DX}^+(z) - n(z)] \quad (10)$$

where the dielectric constant ϵ is different for the particular layers and therefore a function of z . $N_D^+(z)$ is the concentration of ionized shallow donors, and $N_A^-(z)$ the concentration of ionized shallow acceptors. $N_{DX}^+(z)$ is a term which is important for GaAs-based PHEMTs with AlGaAs barrier layers but not for InP-based HEMTs with InAlAs barriers. It denotes the concentration of occupied "DX" type deep donor levels in doped AlGaAs. $n(z)$ is the three-dimensional electron concentration.

The quantity $N_{DX}^+(z)$ depends on the Al content of the AlGaAs layers and on their (silicon) doping concentration. Two different models that describe this dependency are discussed in the papers by Chand *et al.* [32] and Schubert and Ploog [33]. Both papers agree in the observation that for Al contents exceeding $x \approx 0.2$ the introduction of silicon does not only lead to the formation of shallow donor levels but simultaneously to a concentration of deep donors that has to be taken into account.

Schubert and Ploog postulate that the energetical position of the deep levels is independent of the Al content. However, an increase of the Al concentration causes the deep-state density to rise relative to the desired shallow-state density. This results in ineffective doping activation since an increasing part of the donor electrons will not be transferred into the quantum well or the supply layer conduction band but will be trapped in the DX levels. Here, it does not contribute to current transport but only may participate in slow trap charge redistribution processes which interfere with the regular high-frequency transport mechanisms.

In the model by Chand *et al.*, it is not the density of the DX centers that depends on the Al concentration, but their energetical position. It is assumed that the DX energy is separated from the energy of the X conduction band minimum by a constant difference which means that the DX energy rises with Al content, together with the X valley. This means that the DX occupation probability rises with Al content.

Both DX models lead to the conclusion that Al concentrations substantially exceeding, say, 23% are problematic if trap charge redistribution effects shall be avoided (especially at temperatures below 300 K), i.e. effects having to do with voltage and time dependent variations of the term $N_{DX}^+(z)$ in Equation 10.

The local electron concentration $n(z)$ is the sum of the local electron concentrations in the subbands connected to the individual energies E_i , with the sum over all envelope wavefunctions normalized to one:

$$n(z) = \sum_i |\phi_i(z)|^2 n_i. \quad (11)$$

The electron concentrations n_i in subbands i are given by

$$n_i = (kTm^* / \pi\hbar^2) \ln [1 + \exp ((E_F - E_i) / kT)] \quad (12)$$

where k is Boltzmann's constant and T the temperature. Please note that not only $n(z)$ but also all other terms in the bracket of the right-hand side of Equation 10 include occupation probabilities which depend on the position of E_F .

A simultaneous solution to Equations 8 and 10 is subject to two major boundary conditions. The first one is the potential at the semiconductor-gate interface which is given by $V(z = 0) = V_{GS} - \phi_B$, and the second one is the continuity of the electric displacement $\epsilon(z) \partial V(z) / \partial z$ at all heterojunction interfaces. When looking for a first guess for $V(z)$ to start with the iterative solution process, one might assume that $V(z)$ is a straight line in regions where no doping is introduced, for instance in the layer between delta-doping and gate, but has to keep in mind that mobile charge $n(z)$ might be transferred to parts of such regions so that $\partial^2 V(z) / \partial z^2$ is not equal to zero any more.

We shall now discuss the numerical solution of Schrödinger's and Poisson's equations for two specific cases.

5.2. CHARGE CONTROL IN GaAs-BASED SH-PHEMTS

The first example is an SH-PHEMT on GaAs substrate with a gate-to-channel separation of about 35 nm and a homogeneously doped $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$ supply layer (2 nm undoped spacer to the channel). The channel is 12 nm $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}$. Electrostatic potential and electron wavefunction are shown in Figures 10 and 11 for two different gate voltages. In Figure 10, V_{GS} is negative and the number of electrons in the quantum well is small, i.e. the device is close to pinch-off (under the assumption $v_{sat} = 1.5 \times 10^7 \text{ cm/s}$, $n_s = 2 \times 10^{11} \text{ cm}^{-2}$ roughly corresponds to a drain current of 50 mA/mm). The squares of the envelope wavefunctions show that close to 99% of the electrons reside in the ground state or the first excited state in the quantum well.

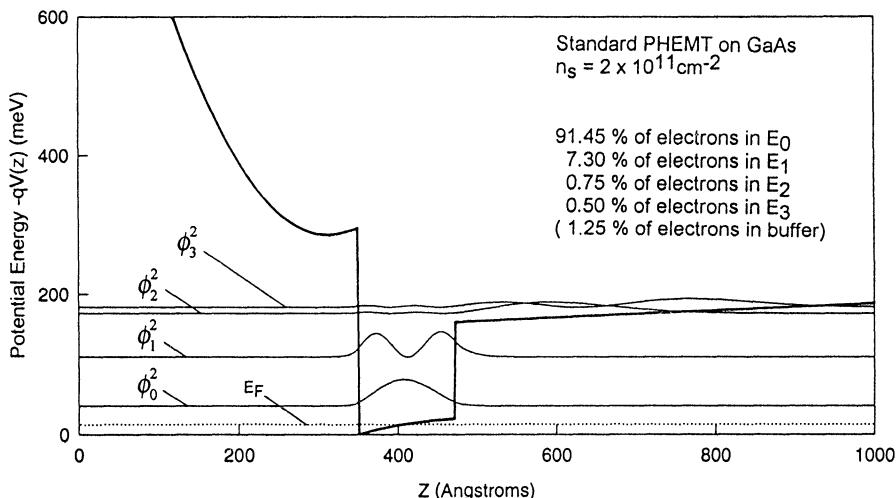


Figure 10. Conduction band diagram and electron wavefunction in an SH-PHEMT on GaAs for $V_{GS} = -0.39 \text{ V}$. Program PSIGaAs by A. Zrenner. Results courtesy H. Riechert.

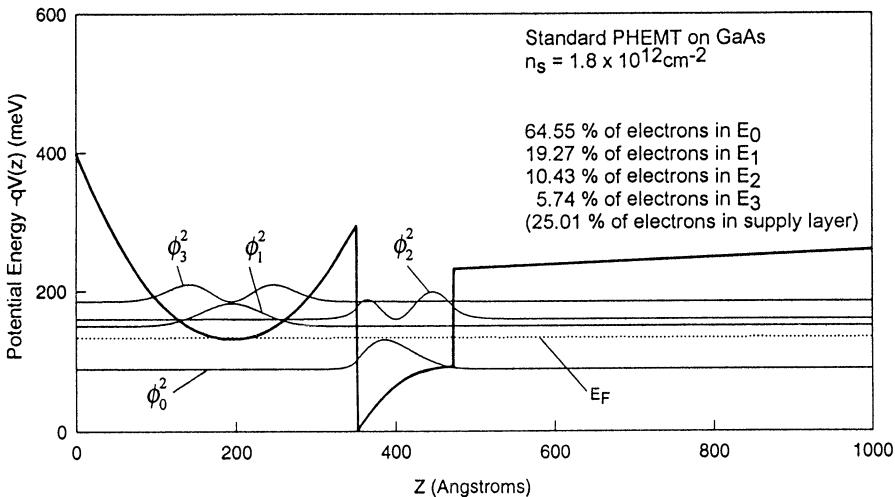


Figure 11. Conduction band diagram and electron wavefunction in an SH-PHEMT on GaAs for $V_{GS} = + 0.43$ V. Program PSIGaAs by A. Zrenner. Results courtesy H. Riechert.

However, there is a small number of electrons (1.25%) populating the excited levels E_2 and E_3 . The wavefunctions belonging to these levels are not centered in the quantum well but extend far into the GaAs buffer below it. This is a consequence of the very small energy barrier between channel and buffer.

In Figure 11, the gate voltage is positive, and the number of carriers in the channel is high ($n_s = 1.8 \times 10^{12} \text{ cm}^{-2}$ corresponds to about 430 mA/mm if the same v_{sat} is used as above). Only the carriers belonging to energy states E_0 and E_2 are confined to the quantum well. The remaining 25% (states E_1 and E_3) sit in a "parasitic" channel which has developed in the AlGaAs supply. In contrast to the case of Figure 10 where the ground state electrons are located precisely in the center of the channel, the ground state electrons (i.e. the majority of carriers) now form a large sheet charge concentrated near the gate-supply heterojunction. According to Poisson's equation, this gives rise to a local electric field. Due to the boundary condition of field continuity at all heterojunctions, this field is also present in the supply region adjacent to the channel bending the conduction band down to the Fermi level. Of course, the electrons in this parasitic channel suffer from degraded transport properties.

5.3. CHARGE CONTROL IN GaAs-BASED DH-PHEMTS

The second example is a DH-PHEMT on GaAs substrate with a gate-to-channel separation of about 30 nm and a homogeneously doped $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$ upper supply layer, also with 2 nm undoped spacer. As previously, the channel is 12 nm $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}$. Below the channel, there is a second $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$ supply layer: the doped region of 6 nm thickness is separated from the channel by a 4 nm spacer. The undoped material below this supply is again $\text{Al}_{0.2}\text{Ga}_{0.8}\text{As}$. In Figure 12, the carrier

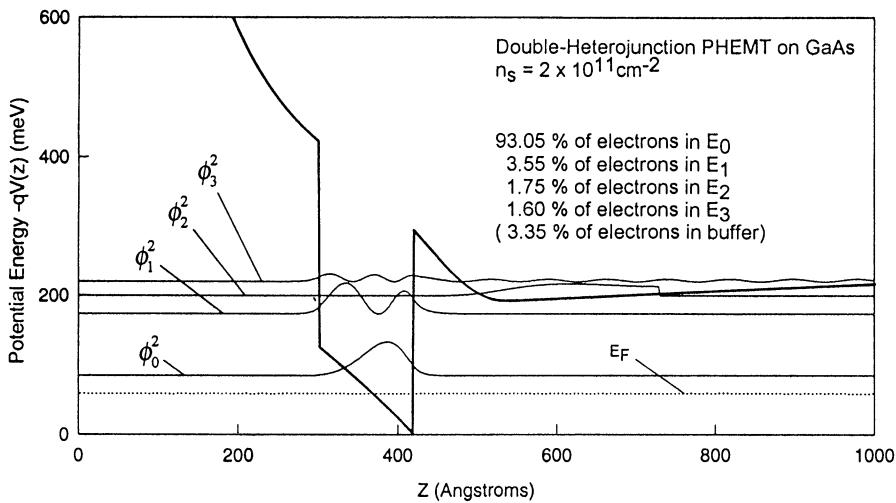


Figure 12. Conduction band diagram and electron wavefunction in a DH-PHEMT on GaAs for $V_{GS} = -0.64$ V. Program PSIGaAs by A. Zrenner. Results courtesy H. Riechert.

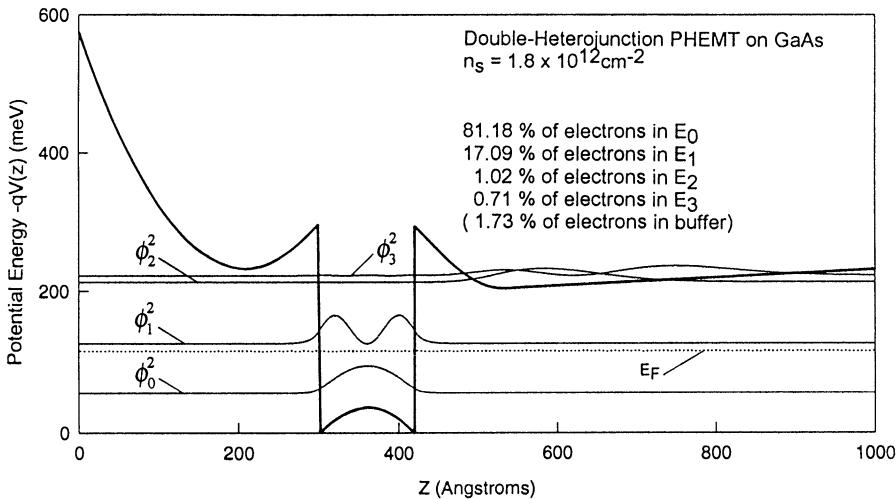


Figure 13. Conduction band diagram and electron wavefunction in a DH-PHEMT on GaAs for $V_{GS} = +0.26$ V. Program PSIGaAs by A. Zrenner. Results courtesy H. Riechert.

density is the same as in Figure 10, i.e. $2 \times 10^{11} \text{ cm}^{-2}$. As a consequence of the doping below the QW, the gate voltage necessary to deplete the channel this far is more negative than in Figure 10 though the gate-to-channel separation is smaller.

In the DH-PHEMT, the percentage of electrons in states E_2 and E_3 (3.35%) is higher than in the SH-PHEMT by a factor of 2.7. Although not all of these electrons are located in the lower supply (Figure 12 shows a finite magnitude of ϕ_3^2 inside the quantum well), it is quite clear that parasitic conduction below the channel is larger in the DH-PHEMT. This is caused by the lower doping which bends the conduction band

down. The barrier below the channel is higher than in the SH-PHEMT but its upper part is thin so that electrons in the excited states can tunnel through it. The majority of the electrons, however, is located inside the quantum well, close to the lower heterojunction.

The conduction band diagram of a DH-PHEMT with positive gate bias is shown in Figure 13. The carrier density of $1.8 \times 10^{12} \text{ cm}^{-2}$ is achieved with a smaller gate voltage compared to the SH-PHEMT. But more important, there are no parasitic electrons in the upper supply. This means that all electrons are "fast", and the same number of carriers supports a larger drain current than in the SH-PHEMT. The ground state wavefunction is centered in the channel (contrary to the SH-PHEMT where it is close to the upper heterojunction). Surprisingly, there are still a few electrons found below the quantum well.

5.4. CHARGE CONTROL AND TRANSFER CHARACTERISTICS

Transfer and transconductance characteristics of an SH-PHEMT have already been shown in Figure 7. Figures 14 presents the corresponding characteristics of a DH-PHEMT.

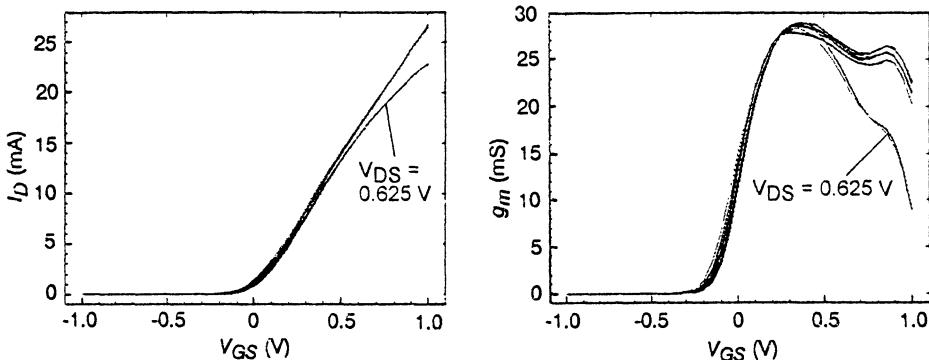


Figure 14. Transfer characteristics I_D (V_{GS}) and transconductance characteristics g_m (V_{GS}) of a DH-PHEMT on GaAs substrate with $W_G = 60 \mu\text{m}$ and $L_G = 0.25 \mu\text{m}$.

In Figure 14, it can be seen that the dependence of the drain current I_D on the gate voltage V_{GS} is more linear compared to Figure 7. This is reflected in the transconductance $g_m(V_{GS})$ which features a distinct plateau for V_{GS} between + 0.2 and + 0.9 V. Such a plateau cannot be found in the corresponding plot for the SH-PHEMT (Figure 7).

These differences in the transconductance characteristics of SH-PHEMT and DH-PHEMT can be explained by their different charge control behavior. In an SH-PHEMT, when V_{GS} increases from the pinch-off voltage towards more positive values, the 2DEG is generated in the middle of the quantum well and moves (while n_s grows) towards the upper heterojunction. The maximum g_m is obtained when V_{GS} is such that the 2DEG is

close to the heterojunction (i.e. the distance to the gate is minimized) but the conduction band in the supply is still well above E_F . For even more positive V_{GS} , the 2DEG density n_s will not grow anymore, but the supply will be increasingly populated with (slow) carriers, and g_m drops.

In the DH-PHEMT at threshold, the 2DEG does not develop in the center of the channel but shifted towards the lower heterojunction. With increasing V_{GS} , the peak of the electron distribution also moves upwards. In contrast to the SH-PHEMT, the maximum n_s is already reached when the wavefunction sits approximately symmetrically in the quantum well. This means that high carrier concentration in the channel can be achieved without creating an extremely high electric field at the upper heterojunction that forces the conduction band in the supply down to E_F . This happens only for even higher V_{GS} so that the onset of parasitic conduction in the supply is delayed. Consequently, in a DH-PHEMT, carriers confined to the quantum well are modulated without simultaneous modulation of slower carriers during a fraction of the total gate voltage swing that is significantly larger than in an SH-PHEMT. However, a transfer characteristics as linear as shown in Figure 14 can be achieved only with small threshold voltage, i.e. with comparatively thin supply layer.

Unfortunately, the favorable behavior of the DH-PHEMT is combined with a drawback: for a given supply thickness, the effective separation of the 2DEG from the gate is some nanometers larger in a DH-PHEMT compared to an SH-PHEMT. Therefore, the SH-PHEMT can be better optimized for high *peak* transconductances when this separation has to be minimized.

5. 5. MODULATION EFFICIENCY

The notion of "Modulation Efficiency" was introduced by Foisy *et al.* [34]. It is an attempt to extend the validity of the simple formulae that describe the *ideal* HEMT (in which all carriers are completely confined to an infinitely thin channel and move with uniform velocity v_{sat}) in a phenomenological way so that the more complex transport that takes place in a *real* HEMT is included.

Because of the absence of slow carriers inside or outside the channel, the ideal HEMT is the device in which the modulation efficiency η is at its theoretical limit. This means that *maximum* current modulation is achieved by *minimum* charge modulation, and $\eta = 1$.

In the *real* HEMT, only a part of the charge carriers travel in the quantum well with a velocity v_{sat} . Other parts of the electrons

- are also confined to the channel but have not yet reached high speed ("gradual channel" part),
- move in the supply layer at even lower speed,
- may be trapped in the supply so that changes in V_{GS} only redistribute trapped charge,
- or may have penetrated deep into the buffer.

This has the consequence that far more charge has to be modulated for a given current change than in a device that has only fast carriers in the channel, and $\eta < 1$.

The expressions that yield the intrinsic transconductance of the idealized device

(Equations 4 to 6) may now be appropriately modified. For the real device, the channel charge Q_s in Equation 5 must now be replaced by the total charge Q_G under the gate, moving with the smaller effective velocity $v_{sat} \eta$, and Equation 5 now reads

$$I_D = Q_G v_{sat} \eta / L_G . \quad (13)$$

The intrinsic transconductance is thus given by

$$g_{mi} = (\partial Q_G / \partial V_{GSi}) (v_{sat} \eta / L_G) = C_{GSi} v_{sat} \eta / L_G \quad (14)$$

where C_{GSi} is the gate-source capacitance of the intrinsic HEMT.

Finally, we want to derive an expression for a major figure of merit for HEMTs, the current-gain cutoff frequency or transit frequency f_T . It is often represented as the inverse of the effective transit time of the carriers through the intrinsic region: $f_T = (2\pi\tau)^{-1}$. For the real device,

$$f_T = v_{sat} \eta / (2\pi L_G) . \quad (15)$$

This equation shows that good modulation efficiency results not only in high transconductance but also in high cutoff frequency. For $\eta = 1$, f_T is at its theoretical limit.

Comparison of Equations (14) and (15) yields

$$f_T = g_{mi} / (2\pi C_{GSi}) . \quad (16)$$

In Equation 16, the time-of-flight across the intrinsic transistor is considered to be the only time constant determining f_T . However, a more detailed discussion in section 7.1. will show that this is a simplification and additional contributions to f_T must be taken into account.

6. Small-Signal Equivalent Network

6.1. INTRINSIC AND EXTRINSIC NETWORK ELEMENTS

Up to now, we have only dealt with the *intrinsic* HEMT which is essentially the physical part of the device directly below the gate contact. In practice, this intrinsic HEMT cannot be connected to its electrical environment without any losses. These losses determine which percentage of the device's intrinsic performance can be actually measured *extrinsically*. The elements which determine the amount of losses are called parasitics.

Figure 15 shows a small-signal equivalent network which includes the elements of the intrinsic HEMT and the parasitics that connect the inner device to the three terminals that are accessible to the measurement (source, gate and drain pads). In the

figure, it was tried to arrange the network elements in a way that their physical origin is reflected. For a complex device like the HEMT, the choice of a specific equivalent network never can be unique. Therefore, approaches slightly different from that in Figure 15 are also possible. These ambiguities can result in doubts if a specific element is of intrinsic or extrinsic character. A good discussion of HEMT network modeling is given in the article by Das [35].

In the network in Figure 15, elements of the intrinsic device are the current source g_{mi} , the gate-source capacitance C_{GSi} , an intrinsic series resistance to this capacitance, R_i , sometimes referred to as the "channel" resistance, the gate-drain capacitance C_{GDi} , and the output conductance g_{ds} . Extrinsic elements are the source, drain and gate resistances R_S , R_D and R_G , the gate-source and gate-drain parasitic capacitances C_{GSp} and C_{GDP} , and the drain-source capacitance C_{DS} .

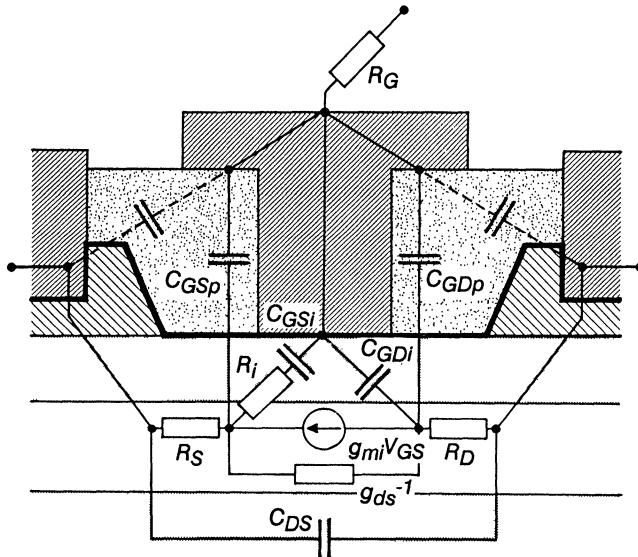


Figure 15. HEMT equivalent network.

One of the most important consequences of the existence of a finite R_S is that the extrinsically measured transconductance g_m is reduced compared to the intrinsic value g_{mi} :

$$g_m = g_{mi} / (1 + R_S g_{mi}) . \quad (17)$$

The presence of parasitic capacitances cannot be seen in the dc properties of the HEMT but has a very important effect on its high-frequency performance. The total gate-source and gate-drain capacitances are given by

$$C_{GS} = C_{GSi} + C_{GSp} , \quad (18)$$

$$C_{GD} = C_{GDi} + C_{GDP}, \quad (19)$$

and Equation 16 has to be modified to

$$f_T = g_{mi} / [2\pi (C_{GS} + C_{GD})]. \quad (20)$$

There is yet another cutoff frequency which depends on a larger number of equivalent network elements. As a figure of merit, it may therefore be valued higher than f_T . It is the maximum frequency of oscillation or power-gain cutoff frequency f_{max} given by [36, 37]

$$f_{max} = f_T [4 g_{ds} (R_S + R_i + R_G) + 2 (C_{GD} / C_{GS}) ((C_{GD} / C_{GS}) + g_{mi} (R_S + R_i))]^{-1/2}. \quad (21)$$

An optimization of f_T and f_{max} can be achieved by minimizing all network elements, with the exception of g_{mi} which must be as large as possible.

6.2. OPTIMIZATION OF NETWORK ELEMENTS

A high g_{mi} implies a small gate-to-channel separation and, hence, is coupled to a high C_{GSi} , according to Equation 14. This conflict can only be solved by realizing a short gate length. This is a major reason why L_G is such a key issue for a fast device.

R_S can be minimized by observing a couple of rules: the ohmic source and drain contacts should be placed close to the inner device, and the remaining distance should be bridged by a highly conductive cap as the topmost layer in the epi sequence. Highly doped caps with small band gaps also assist in the formation of ohmic contacts with low resistivities. The cap is intersected by the recess etch trough. Below this trough, current flow is intended in the quantum well only. Under the sole aspect of R_S reduction the recess length should be as short as possible, but we will see later that a tradeoff is necessary for the sake of other device properties, i.e. that the recess length must be larger than L_G and some finite length of "ungated channel" at both sides of the gate must be allowed. The resistance of the ungated channel depends on the sheet resistance of the 2DEG, i.e. on the electron concentration n_s and mobility μ_e in the quantum well, and on the threshold voltage V_T . If V_T is zero or even positive, the surface potential in the ungated channel will deplete the quantum well from carriers so that R_S can achieve undesirably high values if the ungated channel is too long.

The second resistance that has to be minimized is R_G . The requirement to realize short gates leads to high end-to-end resistances of the gate stripes especially if the gate widths are large as it is the case in analog devices (compared to transistors for digital IC use). This situation results in the concept to attach a gate overlayer with a large cross section and, hence, low resistance to a short-length gate stem forming a "T-gate" or "mushroom gate" as sketched in Figure 15.

The magnitude of the output conductance g_{ds} reflects the extent to which the drain current is controlled not only by V_{GS} but also by V_{DS} , a generally unwanted effect. It

can be suppressed by choosing a high "aspect ratio" $L_G / (d + \Delta d)$ which means placing the gate close to the channel so that the control of the gate potential over the charge carriers is good. Usually, aspect ratios between 5 and 10 are judged to be reasonable. Additionally, a large ΔE_C between the quantum well and the lower barrier helps to prevent the electrons from penetrating deep into the buffer where they can escape from the control by V_{GS} . This is the reason that - among the GaAs-based PHEMTs - the DH-PHEMT is superior to the SH-PHEMT with respect to g_{ds} .

A high g_{ds} cannot only be caused by electrons in the buffer below the channel but also by real-space transfer of carriers into the upper supply or the highly conductive cap between gate and drain. This can be reduced by a long ungated channel at the drain side of the gate. For reasons of process technology, the recess trough is usually symmetrical with respect to the gate center so that the ungated channel regions at the source and drain sides of the gate are of equal length. Consequently, the decision for a particular recess length for a given L_G always represents a compromise: short recess is in favor of small R_S , long recess is in favor of small g_{ds} .

The minimization of the parasitic capacitances C_{GSp} and C_{GDp} has an impact on both f_T and f_{max} . Since C_{GDp} is generally much smaller than C_{GSp} , the addition of parasitics is, of course, unfavorable with respect to C_{GSp} , but even worse for C_{GD} . In other words, large parasitics degrade f_T , but they degrade f_{max} even more. The most common reason for the existence of large parasitics is the presence of a thick dielectric layer covering the HEMT surface for the purpose of device passivation. Often, the gate overlayers are supported by these passivating layers (as shown in Figure 15) causing a capacitive coupling between overlayer and substrate (or even between overlayer and ohmic contacts, as also indicated in the figure). This undesirable effect can be avoided by choosing a thin dielectric - preferably with a small dielectric constant - which does not completely fill the space between overlayer and substrate. Even if a passivation is completely omitted, a small part of the parasitic capacitances never can be avoided. This part is the so-called fringe capacitance which arises at the edges of any metal strip on top of a semiconductor simply as a consequence of Poisson's law [38, 39].

7. PHEMT High-Frequency Properties

7.1. BIAS DEPENDENCE

In section 5.5., f_T was treated as the inverse of the time-of-flight of the carriers across the region under the gate. We now want to identify additional components which contribute to the total delay τ because this enables us to understand the dependence of the high-frequency properties on V_{DS} and I_D . The discussion is mainly based on the treatments by Moll *et al.* [40] and Greiling and Nguyen [41].

The components of the total delay divide into *intrinsic* and *extrinsic* delays, analogous to the intrinsic and extrinsic network components. The most important part of the *intrinsic* contribution is the time-of-flight across the inner HEMT (in section 5.5., this was considered the only existing time constant):

$$\tau_i = L_G / v_{sat} \quad \eta = C_{GSi} / g_{mi}. \quad (22)$$

At the drain end of the gate, there exists a dipole zone or drain depletion region the extension of which depends on device geometry (length of ungated channel at drain side) but also on drain voltage. Therefore, the transit time τ_d through this domain, the drain delay, is proportional to V_{DS} .

The third intrinsic part of the total delay is the channel charging time τ_c . It is the inverse RC constant for changing the charge state of the channel and is inversely proportional to the number of carriers in the channel, n_s , or to the drain current, I_D .

The *extrinsic* contributions to the total delay are less dependent on bias conditions. They are essentially the fringe and parasitic capacitance charging time τ_{par} and the pad capacitance charging time τ_{pad} . Since pads must have a minimum size, the relative contribution of pad capacitances is large for devices with small gate periphery, and the best current-gain cutoff frequencies can be measured at HEMTs with wide gates.

In summary, f_T is given by

$$f_T = (2\pi \tau)^{-1} = [2\pi (\tau_i + \tau_d + \tau_c + \tau_{par} + \tau_{pad})]^{-1}. \quad (23)$$

7.2. PROPERTIES OF DIFFERENT EPI LAYER DESIGNS

Current-gain cutoff frequency and maximum frequency of oscillation for three different types of GaAs-based PHEMTs are shown as a function of drain current in Figure 16. The devices in the figure are all of the same gate width and have comparable gate lengths. Pad configuration and technology process are also the same so that the influence of parasitics is identical for the three devices. However, the similarity of the f_T curves does not mean that g_m and C_{GS} are identical, too. The standard SH-PHEMT without delta-doping has the largest gate-to-channel separation and, hence, the smallest C_{GS} , but also the smallest g_m . This results in f_T values comparable to those of the two delta-doped PHEMTs with smaller d and higher g_m . The dependence of f_T on I_D reflects the fact that for high I_D , g_m stays highest for the DH-PHEMT, as already demonstrated in Figure 14, and that the the smallest channel charging times τ_c are also realized in the DH-PHEMT.

The f_{max} data reveal additional information. The large d in the standard SH-PHEMT leads to a large g_{ds} being the main reason for the low f_{max} . In the delta-doped SH-PHEMT, the gate is close to the channel, the aspect ratio is high, g_{ds} low, and f_{max} is significantly improved. The delta-doped DH-PHEMT has a comparable d , but contrary to the SH-PHEMTs, the electrons are all confined to the quantum well even for high I_D (see Figure 13) so that modulation efficiency and f_{max} are high for a very large current interval.

Figure 17 illustrates the decrease of f_T with V_{DS} caused by the increase of drain delay. The data are taken from a DH-PHEMT optimized for K-band power applications (see also section 9.3.). The wide extension of the recess trough at the drain side of the

gate of this device causes a long drain depletion region which allows to apply V_{DS} up to 10 V. On the other hand, this is the reason for the strong increase of the drain delay with V_{DS} which can be seen in the figure.

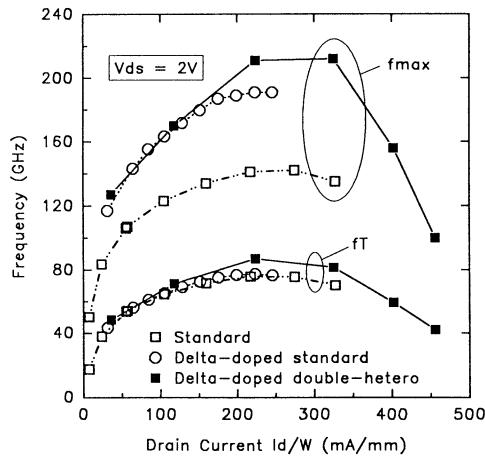


Figure 16. Cutoff frequencies f_T and f_{max} of PHEMTs on GaAs substrate. All devices passivated, $L_G = 0.18\text{--}0.20 \mu\text{m}$, $W_G = 180 \mu\text{m}$.

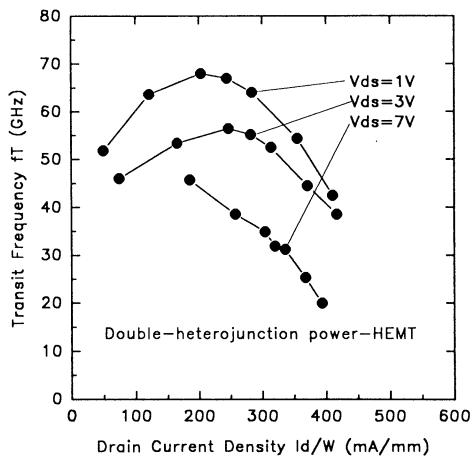


Figure 17. Current-gain cutoff or transit frequency f_T of a power DH-PHEMT on GaAs substrate. Device passivated, $L_G = 0.23 \mu\text{m}$.

7.3. CUTOFF FREQUENCIES: STATE OF THE ART

An overview over published f_T and f_{max} record results is given in Table 2.

TABLE 2. HEMT cutoff frequencies: state of the art

Device	L_G (nm)	g_m/W_G (mS/mm)	f_T (GHz)	f_{max} (GHz)	Author (Organization)
PHEMT on GaAs	150	640	100	350	Lester et al. (Cornell/GE 1988) [42]
	100	650	150	250	Nguyen et al. (Cornell 1989) [43]
	80	740...920	-	270	Chao et al. (GE 1989) [44]
	100	700	120	290	Tan et al. (TRW 1990) [45]
LMHEMT on InP	80	1150	250	220...300	Nguyen et al. (Hughes 1990) [46]
	150	1300	165	405	Chao et al. (GE 1990) [47]
	150	900	-	455	Ho et al. (GE 1991) [48]
PHEMT on InP	100	1160	210	300	Mishra et al. (Hughes 1989) [49]
	50...65	1700...1740	300...340	250...280	Nguyen et al. (Hughes 1992) [50,51]
	100	1550	305	340	Wojtowicz et al. (TRW 1994) [52]

Most of the results in Table 2 were achieved with unpassivated devices which are not handicapped by the presence of large parasitic capacitances. The collection of data clearly shows the improvement of transport properties with increasing indium content of the channel or quantum well depth, consequently leading to the highest g_m and f_T for the PHEMTs on InP with up to 80% In. Surprisingly, the superiority of the InP-based HEMTs over the GaAs-based devices with respect to f_{max} is less pronounced. This has essentially two reasons. The large energetical separation of the central conduction band valley from the satellite valleys in InGaAs with high In content allows the electrons to pick up very high energies from the electric field before they are scattered into the satellites with their degraded transport properties. These high energies facilitate real-space transfer to the barrier layers even though the quantum well is deeper compared to the GaAs-based HEMTs. The second effect is the early onset of impact ionization which leads to the presence of holes and a reduction of modulation efficiency.

8. Optimization of Low-Noise HEMTs

8.1. FUKUI FORMULA

The noise figure NF of a transistor is defined as the signal-to-noise ratio S/N of the input signal devided by the S/N of the output signal. For frequencies $f < 20$ GHz, NF can be well approximated by the semi-empirical formula given by Fukui [53]:

$$NF \text{ (dB)} = 10 \log [1 + 2\pi Kf C_{GS} ((R_S + R_G) / g_{mi})^{1/2}] , \quad (24)$$

$$NF \text{ (dB)} = 10 \log [1 + K (f/f_T) (g_{mi}/(R_S + R_G))^{1/2}] . \quad (25)$$

The quantity K is a fitting factor which takes the properties of the channel transport into account in a phenomenological way. All physical effects which make a HEMT superior to a MESFET with respect to noise performance are essentially contained in this factor. Since Fukui's equation is a simplification, it does not very well describe the noise figures at higher frequencies f especially because it does not consider the drain region as an additional noise source. Anyhow, the formula works well as a basis to identify the major criteria for a good low-noise HEMT.

8.2. LOW-NOISE BIAS CONDITIONS

First of all, the parasitics R_S , R_G and $C_{GS,p}$ must be kept as small as possible. However, the minimization of the total input capacitance C_{GS} is in conflict with the maximization of g_m which is both required by Equation 24. This makes the proper choice of the bias V_{GS} important. Figure 18 shows a one-dimensional calculation of C_{GS} (V_{GS}) by Deutschmann *et al.* [54] in comparison to measured data. In this example, the device is pinched off for $V_{GS} < -0.5$ V. The measurements were performed at a device of extremely large gate area ($L_G \times W_G = 100 \times 100 \mu\text{m}^2$), because in this case, the influence of parasitics is small compared to the intrinsic capacitance so that the behavior of the charge carriers can be studied without disturbing influences. This is the reason why C_{GS} is negligible for $V_{GS} < V_T$ in Figure 18. When V_{GS} is increased, electrons begin to populate the channel, and C_{GS} grows rapidly. For a finite range of V_{GS} around 0 V, the increase of C_{GS} with V_{GS} reduces. This is the regime of high modulation efficiency where the 2DEG in the channel is already established, and any additionally generated free electrons will also be confined to the quantum well. With V_{GS} further increasing to values significantly above 0 V, electrons begin to accumulate in the upper supply, and the increase of C_{GS} with V_{GS} becomes more rapid again.

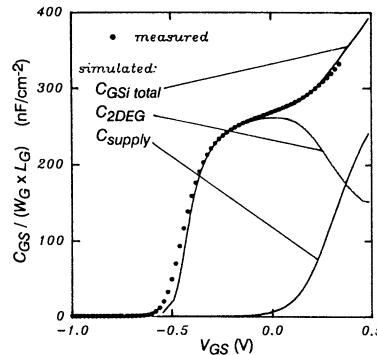


Figure 18. Gate-source capacitance C_{GS} as a function of gate voltage V_{GS} : comparison of simulation and measurement for a GaAs-based SH-PHEMT with $L_G \times W_G = 100 \times 100 \mu\text{m}^2$.

It is evident from Figure 18 that a good bias regime for low-noise operation is expected below the shoulder in the curve $C_{GS}(V_{GS})$ because C_{GS} is still quite small while g_m increases drastically with V_{GS} . In fact, for most PHEMTs, the lowest noise figures are achieved for relatively small currents $I_D / W_G = 40...100$ mA/mm, i.e. for V_{GS} not very far above pinchoff. Clearly, it must be avoided that electrons also populate the region below the channel with its inferior transport properties under this bias condition. Therefore, a device like the one shown in Figure 10 will exhibit a lower NF than the one in Figure 12.

Figure 19 shows an example how the measured NF depends on I_D in a GaAs-based PHEMT. The relatively broad minimum followed by a slow increase with I_D is a typical feature of PHEMTs. Lattice-matched AlGaAs/GaAs devices show a more pronounced increase at the right-hand side of the minimum due to the early onset of parasitic conduction in the supply.

Optimization of bias conditions with respect to NF only makes sense if a reasonable associated gain G_a can be achieved simultaneously. Therefore, the device has to meet requirements additional to those imposed by the Fukui formula. Essentially, it has to achieve a high f_{max} at low-noise bias. As we have demonstrated in Figure 16, f_{max} increases with I_D in the current regime of interest here, and so does G_a (see Figure 19). The broad minimum of the function $NF(I_D)$ exhibited by the PHEMT thus allows to shift the low-noise bias point towards higher current density in the interest of better amplification without sacrificing too much NF performance.

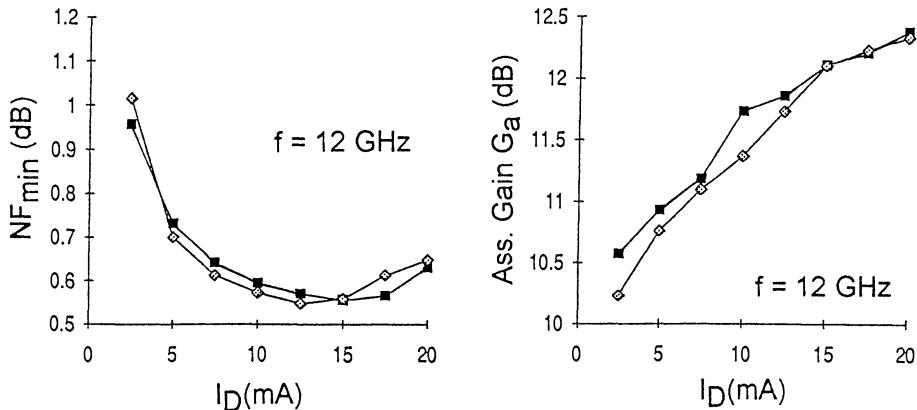


Figure 19. Noise figure NF and associated gain G_a of a standard SH-PHEMT on GaAs for $f = 12$ GHz.
Mean values from two different wafers, devices passivated and mounted in ceramic package,
 $L_G = 0.2 \mu\text{m}$, $W_G = 180 \mu\text{m}$. MBE wafer growth by Walter Schottky Institute, Garching, Germany.

8.3. NOISE FIGURE: STATE OF THE ART

Published outstanding NF data are compiled in Table 3.

TABLE 3. HEMT noise figures and associated gain: state of the art

Device	L_G (nm)	g_m/W_G (mS/mm)	f (GHz)	NF (dB)	G_a (dB)	Author (Organization)
PHEMT on GaAs	150	800	94 18	2.4 0.5	5.4 15.1	Duh et al. (GE 1990) [55]
	100	700	93.5	2.1	6.3	Tan et al. (TRW 1990) [45]
	150	680	60	1.6	6.5	Katoh et al. (Mitsubishi 1993) [56]
PHEMT (InGaP-Supply, MOVPE)	150	> 420	12 50	0.41 1.2	13.0 5.8	Takikawa et al. (Fujitsu 1993) [57]
LMHEMT on InP	200	-	63.5	0.8	8.7	Mishra et al. (Hughes 1989) [49]
	150	900	93 18	1.4 0.3	6.6 17.2	Chao et al. (GE 1990) [47]
LMHEMT on InP (MOVPE)	130	700	12	< 0.30	15	Fujita et al. (Toshiba 1993) [58]
PHEMT on InP	100	1000... 1200	95	1.3	8.2	Tan et al. (TRW 1991) [59]
	150	800	12	0.23	16	Hwang et al. (Avantek/HP 1993) [60]

The table shows a clear advantage of the InP-based HEMTs over the PHEMTs on GaAs. Analogous to the situation with f_T and f_{max} , the InP HEMT's superiority with respect to NF is more pronounced than with respect to G_a . Most of the data were measured on-wafer, i.e. not with packaged devices. Therefore, many of the cited NF values are subject to considerable uncertainty. The transconductances are mostly maximum values not measured under low-noise bias.

9. Optimization of Power HEMTs

9.1. BASIC REQUIREMENTS

During low-noise transistor operation, the high-frequency component modulating the dc gate voltage has only a very small amplitude. Thus, in such small-signal applications, the device properties are of interest only close to the particular bias point. This is completely different for power transistors: large-signal operation means that the gate voltage may sweep over the full range between maximum gate forward bias and pinchoff (or even beyond). Figure 20 schematically shows the case of class "A" power operation with an idealized, purely resistive load. In class A, the large-signal bias point is such that I_D is about half the maximum drain current of the device, I_{Dmax} , and that V_{DS} is also about half the maximum value that can be applied, V_{DSmax} .

In reality, the load of a power transistor probably is an impedance. This means that the modulation of the gate voltage does not sweep the bias point along a straight line any more but along some closed loop enclosing large areas of the output characteristics. However, the idealized class A case is sufficient for a basic discussion of the properties that are desirable for a power HEMT.

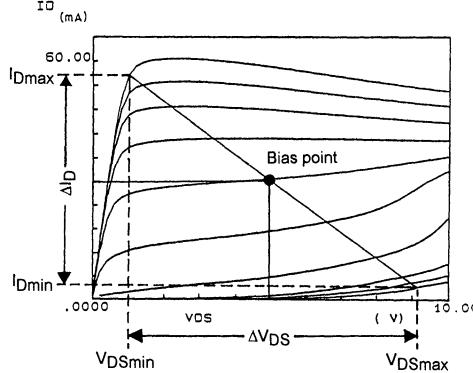


Figure 20. Power HEMT output characteristics with large-signal bias point and idealized load line.

Under these conditions, the maximum output power that can be achieved is given by

$$P_{out\ max} = \frac{1}{\sqrt{2}} \frac{\Delta V_{DS}}{2} \frac{1}{\sqrt{2}} \frac{\Delta I_D}{2} = \frac{1}{8} \Delta V_{DS} \Delta I_D. \quad (26)$$

According to this relation, the usable drain voltage and drain current swings have to be optimized for maximum output power. To the very first order, this means that a good power device should realize high values of V_{DSmax} and I_{Dmax} simultaneously. As can be seen in Figure 20, the minimum voltage V_{DSmin} along the load line is not zero due to the magnitude of R_S and R_D which determines the extent of the linear regime in the output characteristics. Additionally, I_{Dmin} is of finite size if there are leakage currents present preventing a perfect pinchoff at V_{DSmax} . Complete pinchoff at V_{DSmax} can only be achieved if the gate-drain breakdown voltage fulfills the condition $V_{GDbd} > V_{DSmax} - V_T$. Unfortunately, in the presence of high I_{Dmax} , only a low V_{GDbd} can be realized, and vice versa. This necessitates a tradeoff which is additionally influenced by considerations of frequency performance. We will return to this issue in the next section.

Another important property of a power transistor, the power added efficiency PAE , is defined by

$$PAE = (P_{out} - P_{in}) / P_{dc} = P_{out} (1 - G_a^{-1}) / P_{dc} \quad (27)$$

where P_{out} and P_{in} represent the high-frequency output and input power, $G_a = P_{out} / P_{in}$ is the power gain, and P_{dc} is the power of the dc bias. Efficiency is an important figure of merit because high PAE means small power losses and, thus, small self-heating effects and long battery operation. The theoretical limit in class A operation, $PAE = 1/2$, is reached for infinite gain G_a . The gain at a particular signal frequency is bias dependent. This imposes an additional boundary condition for the proper choice of the large signal bias.

9.2. POWER HEMT DEVICE DESIGN

A schematic cross-section of a power HEMT is shown in Figure 21. It differs from the HEMTs in Figure 1 and 15 by the introduction of an additional recess, the "wide recess" of length L_{WR} . The depth of the wide recess is usually adjusted in a way that the highly doped supply layer below the cap is depleted by the surface potential, but not the channel. Choice of the proper L_{WR} allows to find a compromise which combines reasonable R_S and R_D with sufficiently large V_{GDbd} . As discussed by Huang *et al.* [61], a long L_{WR} will improve the gate-drain breakdown but reduce f_T at a given bias V_{DS} (because the drain depletion zone is long, increasing the drain delay), whereas a short L_{WR} will cause an earlier gate-drain breakdown but a higher f_T at the same V_{DS} (because the drain delay is short).

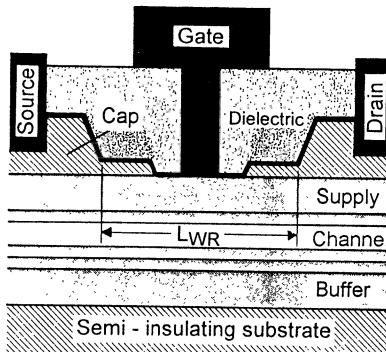


Figure 21. Power HEMT schematic cross-section.

These physical relations imply that the lateral dimensions of a power HEMT are a function of the intended frequency range: L_{WR} must decrease with increasing f , and the large signal bias V_{DS} must decrease accordingly to the reduced breakdown voltage. To realize reasonable output power again, I_{Dmax} must be sufficiently enhanced.

High I_{Dmax} can be realized by the use of epitaxial layer sequences that feature high n_s . In this respect, among the GaAs-based HEMTs, the DH-PHEMT is clearly superior to the SH-PHEMT. Moreover, the DH-PHEMT's enhanced energy barrier at the heterointerface below the channel is extremely welcome: it is the reason for a significantly improved confinement of the charge carriers to the quantum well at high V_{DS} . Thus, the DH-PHEMT not only provides the amount of current needed for power at ultra-high frequency, but also offers the best transport properties at high bias, in other words: it is the best epitaxial concept for both low and high frequency.

In principle, a compromise with respect to f_T , R_S and V_{GDbd} can also be tried by use of a single recess, i.e. by extension of the inner recess to a length of the order of L_{WR} . In most HEMT epi sequences this would expose a large area of a supply layer that contains aluminum. Huang *et al.* [62] have demonstrated that this can result in problematic high-frequency behavior due to slow surface effects. A double recess according to Figure 21 helps to avoid this. It is evident that an unsymmetrical shape of

the double recess is desirable (shorter recess at the source side) because it does not degrade R_S if V_{GDBd} is enhanced. In practice, this cannot be easily realized with the small lateral dimensions typical for HEMTs since lithographical alignment of the gate to the wide recess is difficult.

Up to now, we have not mentioned power HEMTs on InP substrate. This is because InP-based layer sequences of the type introduced so far (see Table 1) are not particularly well suited for power application. Problems arise due to the small bandgap of InGaAs with In contents of 53% and above (early onset of impact ionization), due to locally high fields (early breakdown of the gate-drain diode), and due to deconfinement of the carriers by real space transfer into the supply or buffer layer. Recently, concepts have been developed to tackle these problems. One of these will be discussed in the last section about advanced HEMT concepts.

9.3. EXAMPLE: A POWER PHEMT FOR K-BAND

In this section, we use a PHEMT designed for application in 15 to 30 GHz medium-power MMICs as an example to which result the numerous optimization tradeoffs can lead. Figure 22 shows the output characteristics of this device.

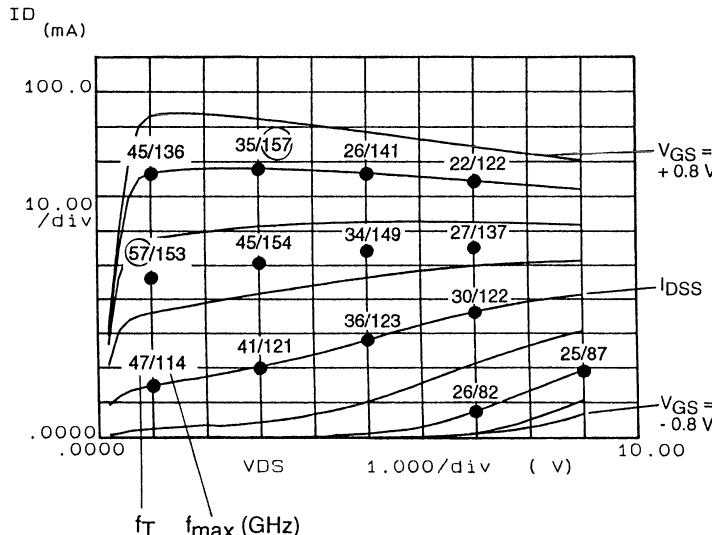


Figure 22. Output characteristics $I_D(V_{DS})$ and cutoff frequencies f_T/f_{max} of a power DH-PHEMT on GaAs substrate. $L_G = 0.25 \mu\text{m}$, $W_G = 240 \mu\text{m}$, $L_{WR} = 1.30 \mu\text{m}$.

The epitaxial layer structure used is essentially the one discussed in Figures 12 and 13. Hall measurements at a control sample resulted in $n_s = 2.6 \times 10^{12} \text{ cm}^{-2}$. In the MMICs, the large signal bias point is $V_{DS} = 5 \text{ V}$, $I_D = 48 \text{ mA}$ where f_T is about 35 GHz. In most parts of the prospective load line, f_T is well above the intended application frequency.

Figure 23 presents the result of an on-wafer load-pull measurement of the output power of such a device at $f = 10$ GHz. The power gain G_a is 17.5 dB, and the output power at 1 dB gain compression, P_{-1dB} , is 24.5 dBm (corresponding to 780 mW/mm). The figure demonstrates that G_a is constant for a large range of P_{in} . This good linearity is a consequence of the fact that g_m varies only slightly over a large fraction of the load line. The small PAE at low P_{in} is caused by P_{dc} being large compared to P_{out} under these conditions. According to Equation 27, the peak PAE of about 55% cannot be achieved in class A operation. Therefore, the measurement result is an indication that the PHEMT is already driven into so-called class AB conditions for the higher values of P_{in} , i.e. that the gate bias reaches values more negative than V_T .

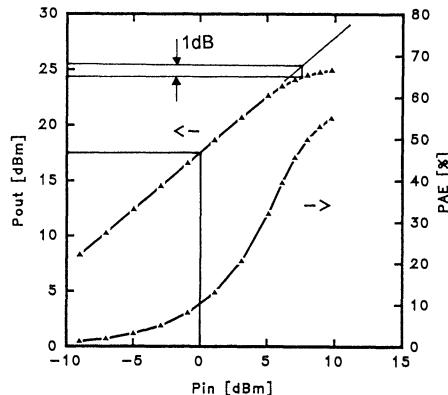


Figure 23. P_{out} vs. P_{in} at $f = 10$ GHz for a DH-PHEMT like the one in Figure 22, but with $W_G = 360 \mu\text{m}$. $V_{DS} = 7$ V. MBE growth by H. Riechert, Siemens.

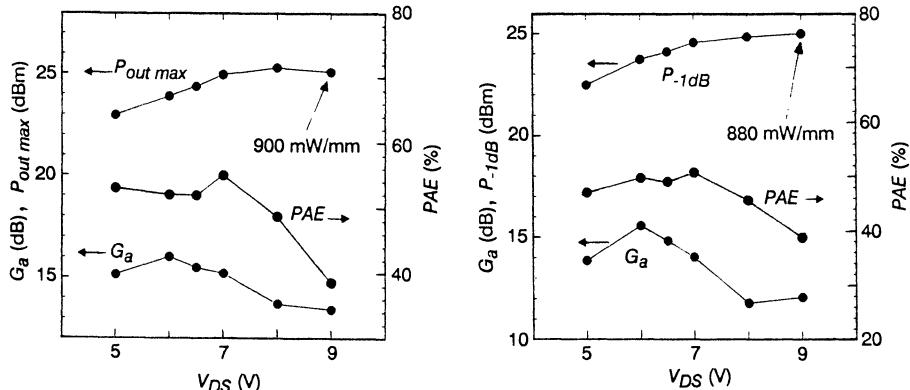


Figure 24. $P_{out\ max}$ and P_{-1dB} vs. V_{DS} at $f = 10$ GHz for the DH-PHEMT of Figure 23. MBE growth by H. Riechert, Siemens.

Figure 24 shows P_{out} , G_a and PAE as a function of drain bias. The power saturation of the device is very abrupt (as could already be seen in Figure 23) which results in P_{-1dB} being only marginally smaller than $P_{out\ max}$. Beyond about $V_{DS} = 7$ V, PAE and G_a decrease due to reduced modulation efficiency. Additional results achieved with this device are contained in Table 4.

9.4. OUTPUT POWER: STATE OF THE ART

A compilation of outstanding power HEMT results is given in Table 4.

TABLE 4. HEMT output power, gain and efficiency: state of the art

Device	f (GHz)	W_G (μm)	V_{DS} (V)	P_{out}/W_G (mW/mm)	P_{-1dB}/W_G (mW/mm)	G_a (dB)	PAE(%)	Author (Organization)
DH-PHEMT on GaAs	4.5	400	8 14	830 1260	-	17.2	63	Kao et al.(GE 1992) [63]
	10	1200	≤ 8.5	935	-	8.9	56	Huang et al. (Raytheon 1991) [64]
	18	400		870		7.4	47	
	10	360	7 9 10 10		790 880 860	14 12 13.6 16.1	51 39 50 42	Grave, Schleicher and Riechert (Siemens 1994)
doped channel DH-PHEMT on GaAs	18	75	-	1000	-	6.8	50	Kim et al. (TI 1989) [65]
"inverted" PHEMT on GaAs	25	120	5	810	680	10.9 12.9	59 45	Zhou et al. (HP 1989) [66]
DH-PHEMT on GaAs	32	400 1200	7	922 619	-	6.7 4.1	37 22	Huang et al. (Raytheon 1993) [67]
DH-PHEMT on GaAs	44	900	5.4	766	-	3.0	24	Smith et al. (GE 1991) [68]
PHEMT on GaAs, 2 doped channels	60	50	3.95	1000	~ 900	2.7	25.5	Saunier, Tserng (TI 1989) [69]
doped channel PHEMT on GaAs	60	75	\downarrow	1000		4.4	36	Smith et al. (GE 1990) [70]
DH-PHEMT on GaAs	60	400	5.0	550		4.5	25.4	Lai et al. (TRW 1993) [71]
doped channel DH-PHEMT on GaAs	94	160	3.4	392		4.0	13.2	Streit et al. (TRW 1991) [72]
LMHEMT on InP	12	300	4.0 ?	960		-	40	Matloubian et al. (Hughes 1991) [73]
	20	800	4.0	700		≈ 7	47.1	Matloubian et al. (Hughes 1993) [74]
doped channel LMHEMT on InP	57	450	3.5 3.5	444	333	3.6 2.6	17 20	Matloubian et al. (Hughes 1993) [75]

For a proper understanding of the data, some additional remarks are necessary. If PAE exceeds 50%, this is an indication that the mode of operation was not class A but class AB or B. It is technically very difficult to maintain a specific output power P_{out} / W_G measured with small W_G also for very large W_G (i.e. for a large number of gate

fingers in parallel) and to achieve high *absolute* P_{out} . Thus, values of P_{out} / W_G realized with large W_G deserve the most regard in Table 4. Additionally, it is more demanding to develop a device with high linear P_{-1dB} than with high P_{out} at unspecified gain compression.

As expected from the preceding discussion, a clear tendency can be seen in Table 4 that the large signal bias voltage V_{DS} goes down with increasing frequency. This implies that power is achieved by a large current swing. To increase drain current, some of the transistors listed in Table 4 are "doped-channel" HEMTs. In these devices, not all but part of the total doping sheet concentration is applied directly in the quantum well and not in the supply layers. Strictly speaking, this is a violation of the original HEMT principle which is based on a local separation of the impurities from the current path. Ironically, this violation is made in HEMTs intended for the highest frequencies where best transport properties are demanded.

With the doped-channel modification, it is possible to maintain a P_{out} / W_G of about 1 W/mm up to a frequency of 60 GHz. Due to the aforementioned difficulties with high fields in InP-based HEMTs, these have been inferior to GaAs-based power HEMTs up to now. Recently, ideas have been developed to overcome this deficiency. An example will be discussed in the last section.

10. Outlook: Advanced HEMT Concepts

10.1. MOTIVATION FOR ADVANCED HEMTS

In this paper, it has been discussed that InP-based HEMTs profit from their deep quantum wells with high n_s mostly with respect to g_m and f_T but that their superiority over GaAs-based HEMTs is weaker when it comes to f_{max} and is completely gone for power applications. This situation suggests two main directions for improved concepts.

The first one is the search for ideas how to improve the GaAs-based PHEMT with respect to f_T and g_m . The manufacturability of devices on GaAs substrate is so much better that efforts to delay the introduction of InP devices to regular production lines as long as possible are considered worthwhile.

The second direction is the aim to remove the InP-based HEMT's deficiencies as a power HEMT. This would create a transistor family of universally superior performance.

These two goals are certainly of practical importance. Independent from the consideration if there is an immediate technical demand, the ambition exists to drive transistor speed to its ultimate limits in the some-hundred gigahertz range. This defines a third direction for advanced concepts.

Necessarily, the following outlook must remain incomplete since the amount of new ideas is vast and constantly growing.

10.2. ADVANCED HEMTS ON GaAs SUBSTRATE

10.2.1. *Metamorphic Growth*

The basic idea is to grow the typical InP HEMT $\text{In}_{0.52}\text{Al}_{0.48}\text{As}/\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ barrier/quantum well system also on GaAs substrate (see, for instance, Wang *et al.* [76], Masato *et al.* [77]). Since this system is extremely mismatched to GaAs, it is necessary to grow a thick graded $\text{In}_x\text{Al}_{1-x}\text{As}$ buffer on the substrate where x slowly increases from 0 to 0.52. This buffer contains dislocations and microcracks. These are intended to be buried deep below the active layers of the HEMT but actually are difficult to keep restricted to these regions. Superlattice buffers made of InAlAs/InGaAs layer sequences can also be used as dislocation filters.

10.2.2. *MOVPE Growth Instead of MBE*

As already mentioned in section 2.1., MOVPE is a growth technique with the advantage that layers with phosphorus can be grown with relative ease. This property can be used to realize supply layers that do not contain aluminum, for instance $\text{In}_{0.48}\text{Ga}_{0.52}\text{P}$ (Takikawa and Joshi [57]). In this material, deep DX centers do not exist which results in better donor activation than in AlGaAs. Additionally, the surface depletion region is shallower due to a lower Schottky barrier. The advantage of both effects is the small thickness of the supply layer which can be realized. This leads to excellent suppression of short-channel effects and, hence, excellent gain. Among all GaAs-based devices, these HEMTs came closest to InP HEMT performance with respect to noise figure (see Table 3. The noise figure NF at 12 GHz was measured in package!).

A natural extension of the concept is to use quaternary AlGaInP as a supply instead of InGaP (Bachem *et al.* [78]). On the one hand, the advantage of the supply without Al is lost, but on the other hand, the conduction band offset is noticeably larger.

The general problems with MOVPE growth, for instance the difficult control of background impurities, have already been mentioned in section 2.1. For HEMTs with phosphide supplies, the additional complication of proper growth of arsenide/phosphide interfaces arises.

10.3. ADVANCED HEMTS ON InP SUBSTRATE

10.3.1. *InAs Channel HEMTs*

These HEMTs are the natural completion to all attempts to grow PHEMTs on InP with increasing indium content in the quantum well, having ever higher f_T and g_m records in mind. The idea is to create a device with maximum conduction band offset between channel and supply. The InAs channel is, of course, extremely mismatched to InP (see Figure 2) and can therefore only be very thin. In the case of Yang *et al.* [79], a 5 nm InAs channel was combined with a lattice-matched InGaAs pre-channel and a conventional InAlAs supply. The high-frequency performance did not yet exceed the results achieved with InP PHEMTs with less In in the channel possibly due to growth difficulties with the InAs layer.

Bolognesi *et al.* [80] have grown a 10 nm InAs channel in combination with AlSb barrier layers. This combination offers a conduction band offset of 1.3 eV. In view of $L_G = 0.5 \mu\text{m}$, the measured f_T of 93 GHz was promising, but the device suffered considerably from problems due to holes created by impact ionization, for instance parasitic bipolar action.

10.3.2. InP Channel HEMTs

The purpose of this development is to find a HEMT with better power performance than the GaAs-based power PHEMT. InP is considered a suitable channel material because it has a high v_{sat} but the velocity peak is only weak and delayed to higher fields compared to InGaAs. This results in high breakdown fields. The bandgap of InP is similar to lattice-matched InAlAs (see Figure 2). However, there exists a conduction band offset of about 0.39 eV between both materials. This is possible because the valence band offset is in the same direction. Aina *et al.* [81] realized HEMTs with InP channel and InAlAs barrier layers by MOVPE and achieved a very high P_{out} of 1.45 W/mm at $V_{DS} = 7 \text{ V}$ ($f = 30 \text{ GHz}$). A major advantage of InP is the very slow decrease of f_T with V_{DS} , a consequence of the moderate dependence of the drain delay time on the field. But there is a drawback: the low μ_e renders the realization of small parasitic resistances very difficult. R_S and R_D , however, determine the extension of the linear regime in the output characteristics. They are therefore parameters with an important influence on the usable voltage swing and, hence, PAE. An attempt to overcome the R_S handicap is the growth of double-layer channels composed of InP and lattice-matched InGaAs (Jelloian *et al.* [82]).

11. Conclusion

We have discussed basic properties of the two major groups of HEMTs: devices based on GaAs substrate, and devices based on InP substrate. Both have the InGaAs channel in common, but in case of the InP-based HEMTs, the indium content in the quantum well is significantly higher, and the conduction band offset between the barrier layers and the channel is larger yielding a deeper quantum well.

The two families of devices possess specific advantages and drawbacks which can mostly be explained in terms of these physical differences.

The GaAs-based PHEMT is the more mature device. Its substrate material is available in 4" size and less subject to wafer breakage than InP. This makes GaAs PHEMT processing less expensive, and manufacturing is possible on existing MESFET technology lines.

GaAs PHEMTs are inferior to InP HEMTs with respect to f_T , g_m and NF . The main reasons for this are the lower electron velocity v_{sat} (due to the smaller indium mole fraction in the quantum well) and the smaller 2DEG concentration n_s (due to the smaller conduction band offset). However, the performance which can be realized with GaAs PHEMTs is in many cases sufficient for circuits with frequencies close to 100 GHz. Additionally, at any frequency, good large-signal performance is - at the moment

- easier to achieve with GaAs PHEMTs. The reason for this is the very same responsible for the smaller v_{sat} : in high fields, the electrons cannot attain such high conduction band energies as they do in channels with higher indium percentage. In combination with the larger bandgap, this shifts the onset of avalanche breakdown towards higher drain voltages.

Though PHEMTs are now a production technology (at least when grown on GaAs) there are still dynamic developments going on. They follow several directions in parallel: it is tried to expand the range of GaAs-based PHEMTs to higher frequencies by the growth of channels more similar to InP HEMTs (concept of "metamorphic" HEMTs), remove the deficiencies of InP HEMTs with respect to power applications to create a HEMT family of universal superiority, and finally, explore how far the frequency range of HEMTs can be extended into the multi-hundred gigahertz range by use of small-bandgap quantum wells like InAs and barriers like AlSbAs.

Acknowledgments

This paper became only possible due to the support of the following colleagues: H. Riechert (MBE wafer growth and one-dimensional device simulation), L. Schleicher (HEMT process development), G. Jäger and F. Raisch (process control), A. Morav and P.W. von Basse (dc characterization), H.J. Siweris, M. Kärner and H. Tischer (HF characterization), J.E. Müller (on-wafer load-pull power measurements), and R. Deutschmann (one- and two-dimensional device simulation). The author is indebted to A. Mesquida Küsters, H. Kniepkamp and W. Kellner for fruitful discussions and critical reading of the manuscript. He further wants to express his gratitude to all other people involved in HEMT processing at both the Siemens Corporate Research and Development and the Siemens Semiconductor Production Group who are not explicitly named here. Part of the low-noise HEMTs discussed in this paper were MBE-grown by the Walter Schottky Institute (Technical University of Munich).

References

1. Mimura, T., Hiyamizu, S., Fujii, T., and Nanbu, K. (1980) A new field-effect transistor with selectively doped $\text{GaAs}/n\text{-Al}_x\text{Ga}_{1-x}\text{As}$ heterojunctions, *Jap. J. Appl. Phys.* **19**, L225-L227
2. Delagebeaudeuf, D., Delescluse, P., Etienne, P., Laviron, M., Chaplart, J., and Linh, N.T. (1980) Two-dimensional electron gas MESFET structure, *Electron. Lett.* **16**, 667-668
3. Dingle, R., Störmer, H.L., Gossard, A.C., and Wiegmann, W. (1978) Electron mobilities in modulation-doped semiconductor heterojunction superlattices, *Appl. Phys. Lett.* **33**, 665-667
4. Störmer, H.L., Dingle, R., Gossard, A.C., Wiegmann, W., and Sturge, M.D. (1979) Two-dimensional electron gas at a semiconductor-semiconductor interface, *Solid-State Comm.* **29**, 705-709
5. Rosenberg, J.J., Benlamri, M., Kirchner, P.D., Woodall, J.M., and Pettit, G.D. (1985) An $\text{In}_{0.15}\text{Ga}_{0.85}\text{As}/\text{GaAs}$ pseudomorphic single quantum well HEMT, *IEEE Electron Device Lett.* **EDL-6**, 491-493
6. Ketterson, A., Moloney, M., Masselink, W.T., Peng, C.K., Klem, J., Fischer, R., Kopp, W., and Morkoc, H. (1985) High transconductance $\text{InGaAs}/\text{AlGaAs}$ pseudomorphic modulation-doped field-effect transistors, *IEEE Electron Device Lett.* **EDL-6**, 628-630

7. Ketterson, A.A., Masselink, W.T., Gedymin, J.S., Klem, J., Peng, C.-K., Kopp, W.F., Morkoc, H., and Gleason, K.R. (1986) Characterization of InGaAs/AlGaAs pseudomorphic modulation-doped field-effect transistors, *IEEE Trans. Electron Devices* **ED-33**, 564-571
8. Henderson, T., Aksun, M.I., Peng, C.K., Morkoc, H., Chao, P.C., Smith, P.M., Duh, K.-H.G., and Lester, L.F. (1986) Microwave performance of a quarter-micrometer gate low-noise pseudomorphic InGaAs/AlGaAs modulation-doped field effect transistor, *IEEE Electron Device Lett.* **EDL-7**, 649-651
9. Hikosaka, K., Hirachi, Y., Mimura, T., and Abe, M. (1985) A microwave power double-heterojunction high electron mobility transistor, *IEEE Electron Device Lett.* **EDL-6**, 341-343
10. Smith, P.M., Chao, P.C., Lester, L.F., Smith, R.P., Lee, B.R., Ferguson, D.W., Jabra, A.A., Ballingall, J.M., and Duh, K.H.G. (1988) InGaAs pseudomorphic HEMTs for millimeter wave power applications, *IEEE MTT-S Tech. Digest*, 927-930
11. Smith, P.M., Kao, M.Y., Ho, P., Chao, P.C., Duh, K.H.G., Jabra, A.A., Smith, R.P., and Ballingall, J.M. (1989) A 0.15 μ m gate-length pseudomorphic HEMT, *IEEE MTT-S Tech. Digest*, 983-986
12. Pearsall, T.P., Hendel, R., O'Connor, P., Alavi, K., and Cho, A.Y. (1983) Selectively-doped $\text{Al}_{0.48}\text{In}_{0.52}\text{As}/\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ heterostructure field effect transistor, *IEEE Electron Device Lett.* **EDL-4**, 5-8
13. Hirose, K., Ohata, K., Mizutani, T., Itoh, T., and Ogawa, M. (1985) 700 mS/mm 2DEGFETs fabricated from high mobility MBE-grown *n*-AlInAs/GaInAs heterostructures, *Gallium Arsenide and Related Compounds 1985, Inst. Phys. Conf. Ser. No. 79*, 529-534
14. Peng, C.K., Aksun, M.I., Ketterson, A.A., Morkoc, H., and Gleason, K.R. (1987) Microwave performance of InAlAs/InGaAs/InP MODFET's, *IEEE Electron Device Lett.* **EDL-8**, 24-26
15. Kuo, J.M., Lalovic, B., and Chang, T.Y. (1986) New pseudomorphic MODFETs utilizing $\text{Ga}_{47-\mu}\text{In}_{53+\mu}\text{As}/\text{Al}_{48+\mu}\text{In}_{52-\mu}\text{As}$ heterostructures, *IEDM Tech. Digest*, 460-463
16. Mishra, U.K., Brown, A.S., and Rosenbaum, S.E. (1988) DC and RF performance of 0.1 μ m gate length $\text{Al}_{48}\text{In}_{52}\text{As}-\text{Ga}_{38}\text{In}_{62}\text{As}$ pseudomorphic HEMTs, *IEDM Tech. Digest*, 180-183
17. Pollak, F.H. (1993) Energy gaps of AlGaAs, in S. Adachi (ed), *Properties of Aluminium Gallium Arsenide*, INSPEC, the Institution of Electrical Engineers, London, p. 54
18. Matthews, J.W., and Blakeslee, A.E. (1974) Defects in epitaxial multilayers, *J. Crystal Growth* **27**, 118-125
19. Andersson, T.G., Chen, Z.G., Kulakovskii, V.D., Uddin, A., and Vallin, J.T. (1987) Variation of the critical layer thickness with In content in strained $\text{In}_x\text{Ga}_{1-x}\text{As}$ -GaAs quantum wells grown by molecular beam epitaxy, *Appl. Phys. Lett.* **51**, 752-754
20. People, R., and Bean, J.C. (1985) Calculation of critical layer thickness versus lattice mismatch for $\text{Ge}_x\text{Si}_{1-x}/\text{Si}$ strained-layer heterostructures, *Appl. Phys. Lett.* **47**, 322-324
21. Taguchi, T., Takeuchi, Y., Matugatani, K., Ueno, Y., Hattori, T., Sugiyama, Y., and Tacano, M. (1993) Critical layer thickness of $\text{In}_{0.80}\text{Ga}_{0.20}\text{As}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ heterostructures, *J. Crystal Growth* **134**, 147-150
22. Ruch, J.G., and Kino, G.S. (1968) Transport properties of GaAs, *Phys. Rev.* **174**, 921-931
23. Braslavau, N., and Hauge, P.S. (1970) Microwave measurement of the velocity-field characteristic of GaAs, *IEEE Trans. Electron Devices* **ED-17**, 616-622
24. Houston, P.A., and Evans, A.G.R. (1977) Electron drift velocity in *n*-GaAs at high electric fields, *Solid-State Electr.* **20**, 197-204
25. Littlejohn, M.A., Kim, K.W., and Tian, H. (1993) High-field transport in InGaAs and related heterostructures, in P. Bhattacharya (ed), *Properties of Lattice-Matched and Strained Indium Gallium Arsenide*, INSPEC, the Institution of Electrical Engineers, London, p. 107-116
26. Brennan, K., and Hess, K. (1984) High field transport in GaAs, InP and InAs, *Solid-State Electr.* **27**, 347-357
27. Liu, C.T., Lin, S.Y., and Tsui, D.C. (1988) Cyclotron resonance measurements of the electron effective mass in strained AlGaAs/InGaAs/GaAs pseudomorphic structures, *Appl. Phys. Lett.* **53**, 2510-2512
28. Nguyen, L.D., Larson, L.E., and Mishra, U.K. (1992) Ultra-high-speed modulation-doped field-effect transistors: a tutorial review, *Proc. of the IEEE* **80**, 494-518

29. Bhattacharya, P.K. (1993) Low- and high-field transport in pseudomorphic InGaAs based heterostructures, in P. Bhattacharya (ed), *Properties of Lattice-Matched and Strained Indium Gallium Arsenide*, INSPEC, the Institution of Electrical Engineers, London, p. 117-126
30. Schubert, E.F. (1990) Delta doping of III-V semiconductors: fundamentals and device applications, *J. Vac. Sci. Technol. A8*(3), 2980-2996
31. Ando, Y., and Itoh, T. (1988) Analysis of charge control in pseudomorphic two-dimensional electron gas field-effect transistors, *IEEE Trans. Electron Devices* **35**, 2295-2301
32. Chand, N., Henderson, T., Klem, J., Masselink, W.T., Fischer, R., Chang, Y.-C., and Morkoc, H. (1984) comprehensive analysis of Si-doped $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ($x = 0$ to 1): theory and experiments, *Phys. Rev. B* **30**, 4481-4492
33. Schubert, E.F., and Ploog, K. (1984) Shallow and deep donors in direct-gap n -type $\text{Al}_x\text{Ga}_{1-x}\text{As:Si}$ grown by molecular-beam epitaxy, *Phys. Rev. B* **30**, 7021-7029
34. Foisy, M.C., Tasker, P.J., Hughes, B., and Eastman, L.F. (1988) The role of inefficient charge modulation in limiting the current-gain cutoff frequency of the MODFET, *IEEE Trans. Electron Devices* **35**, 871-878
35. Das, M.B. (1991) HEMT device physics and models, in F. Ali and A. Gupta (eds), *HEMTs and HBTs: Devices, Fabrication, and Circuits*, Artech House, Boston, London, pp. 11-75
36. Wolf, P. (1970) Microwave properties of Schottky-barrier field effect transistors, *IBM J. Res. Dev.* **14**, 125-141
37. Das, M.B. (1985) A high aspect ratio design approach to millimeter-wave HEMT structures, *IEEE Trans. Electron Devices* **ED-32**, 11-17
38. Wasserstrom, E., and McKenna, J. (1970) The potential due to a charged metallic strip on a semiconductor surface, *Bell Syst. Tech. J.* **49**, 853-877
39. Anholt, R. (1991) Dependence of GaAs MESFET fringe capacitances on fabrication technologies, *Solid-State Electr.* **34**, 515-520
40. Moll, N., Hueschen, M.R., and Fischer-Colbrie, A. (1988) Pulse-doped AlGaAs/InGaAs pseudomorphic MODFET's, *IEEE Trans. Electron Devices* **35**, 879-886
41. Greiling, P.T., and Nguyen, L. (1991) Ultrahigh-frequency InP-based HEMTs for millimeter wave applications, *SPIE Vol. 1475*, 34-41
42. Lester, L.F., Smith, P.M., Ho, P., Chao, P.C., Tiberio, R.C., Duh, K.H.G., and Wolf, E.D. (1988) 0.15 μm gate-length double recess pseudomorphic HEMT with f_{max} of 350 GHz, *IEDM Tech. Digest*, 172-175
43. Nguyen, L.D., Tasker, P.J., Radulescu, D.C., and Eastman, L.F. (1989) Characterization of ultra-high-speed pseudomorphic AlGaAs/InGaAs (on GaAs) MODFET's, *IEEE Trans. Electron Devices* **36**, 2243-2248
44. Chao, P.C., Shur, M.S., Tiberio, R.C., Duh, K.H.G., Smith, P.M., Ballingall, J.M., Ho, P., and Jabra, A.A. (1989) DC and microwave characteristics of sub-0.1- μm gate-length planar-doped pseudomorphic HEMT's, *IEEE Trans. Electron Devices* **36**, 461-473
45. Tan, K.L., Dia, R.M., Streit, D.C., Lin, T., Trinh, T.Q., Han, A.C., Liu, P.H., Chow, P.-M.D., and Yen, H.C. (1990) 94-GHz 0.1- μm T-gate low-noise pseudomorphic InGaAs HEMT's, *IEEE Electron Device Lett.* **11**, 585-587
46. Nguyen, L.D., Jelloian, L.M., Thompson, M., and Lui, M. (1990) Fabrication of a 80 nm self-aligned T-gate AlInAs/GaInAs HEMT, *IEDM Tech. Digest*, 499-502
47. Chao, P.C., Tessmer, A.J., Duh, K.-H.G., Ho, P., Kao, M.-Y., Smith, P.M., Ballingall, J.M., Liu, S.-M.J., and Jabra, A.A. (1990) W-band low-noise InAlAs/InGaAs lattice-matched HEMT's, *IEEE Electron Device Lett.* **11**, 59-62
48. Ho, P., Kao, M.Y., Chao, P.C., Duh, K.H.G., Ballingall, J.M., Allen, S.T., Tessmer, A.J., and Smith, P.M. (1991) Extremely high gain 0.15 μm gate-length InAlAs/InGaAs/InP HEMTs, *Electron. Lett.* **27**, 325-326
49. Mishra, U.K., Brown, A.S., Delaney, M.J., Greiling, P.T., and Krumm, C.F. (1989) The AlInAs-GaInAs HEMT for microwave and millimeter-wave applications, *IEEE Trans. Microw. Theo. Tech.* **37**, 1279-1285

50. Nguyen, L.D., Brown, A.S., Thompson, M.A., Jelloian, L.M., Larson, L.E., and Matloubian, M. (1992) 650-Angstrom self-aligned-gate pseudomorphic $\text{Al}_{0.48}\text{In}_{0.52}\text{As}/\text{Ga}_{0.20}\text{In}_{0.80}\text{As}$ high electron mobility transistors, *IEEE Electron Device Lett.* **13**, 143-145
51. Nguyen, L.D., Brown, A.S., Thompson, M.A., and Jelloian, L.M. (1992) 50-nm self-aligned-gate pseudomorphic $\text{AlInAs}/\text{GaInAs}$ high electron mobility transistors, *IEEE Trans. Electron Devices* **39**, 2007-2014
52. Wojtowicz, M., Lai, R., Streit, D.C., Ng, G.I., Block, T.R., Tan, K.L., Liu, P.H., Freudenthal, A.K., and Dia, M.R. (1994) 0.10 μm graded InGaAs channel InP HEMT with 305 GHz f_T and 340 GHz f_{max} *IEEE Electron Device Lett.* **15**, 477-479
53. ukui, H. (1979) Optimal noise figure of microwave GaAs MESFET's, *IEEE Trans. Electron Devices* **ED-26**, 1032-1037
54. Deutschmann, R., Fischer, C., Sala, S., and Selberherr, S. (1993) Evaluation of effective device parameters by comparison of measured and simulated $C-V$ characteristics for conventional and pseudomorphic HEMTs, in S. Selberherr, H. Stippel and E. Strasser (eds), *Simulation of Semiconductor Devices and Processes Vol. 5*, Springer-Verlag, Wien, pp. 461-464
55. Duh, K.H.G., Chao, P.C., Ho, P., Tessmer, A., Liu, S.M.J., Kao, M.Y., Smith, P.M., and Ballingall, J.M. (1990) W-band InGaAs HEMT low noise Amplifiers, *IEEE MTT-S Tech. Digest*, 595-598
56. Katoh, T., Yoshida, N., Minami, H., Kashiwa, T., and Orisaka, S. (1993) A 60 GHz-band ultra low noise planar-doped HEMT, *IEEE MTT-S Tech. Digest*, 337-340
57. Takikawa, M., and Joshin, K. (1993) Pseudomorphic n -InGaP/InGaAs/GaAs high electron mobility transistors for low-noise amplifiers, *IEEE Electron Device Lett.* **14**, 406-408
58. Fujita, S., Noda, T., Wagai, A., Ashizawa, Y., and Hosoi, S. (1993) Extremely low noise InGaAs/InAlAs HEMT grown by MOCVD, *Electron. Lett.* **29**, 1557-1558
59. Tan, K.L., Streit, D.C., Chow, P.D., Dia, R.M., Han, A.C., Liu, P.H., Garske, D., and Lai, R. (1991) 140 GHz 0.1 μm gate-length pseudomorphic $\text{In}_{0.52}\text{Al}_{0.48}\text{As}/\text{In}_{0.60}\text{Ga}_{0.40}\text{As}/\text{InP}$ HEMT, *IEDM Tech. Digest*, 239-242
60. Hwang, T., Chye, P., and Gregory, P. (1993) Super low noise pseudomorphic InGaAs channel InP HEMTs, *Electron. Lett.* **29**, 10-11
61. Huang, J.C., Boulais, W., Platzker, A., Kazior, T., Aucoin, L., Shanfield, S., Bertrand, A., Vafiades, M., and Niedzwiecki, M. (1993) The effect of channel dimensions on the millimeter-wave power performance of a pseudomorphic HEMT, *GaAs IC Symp. Tech. Digest*, 177-180
62. Huang, J.C., Jackson, G.S., Shanfield, S., Platzker, A., Saledas, P.K., and Weichert, C. (1993) An AlGaAs/InGaAs pseudomorphic high electron mobility transistor with improved breakdown voltage for X- and Ku-band power applications, *IEEE Trans. Microw. Theo. Tech.* **41**, 752-759
63. Kao, M.-Y., Fu, S.-T., Ho, P., Smith, P.M., Chao, P.C., Nordheden, K.J., and Wang, S. (1992) Very high voltage AlGaAs/InGaAs pseudomorphic power HEMTs, *IEDM Tech. Digest*, 319-321
64. Huang, J.C., Jackson, G., Shanfield, S., Hoke, W., Lyman, P., Atwood, D., Saledas, P., Schindler, M., Tajima, Y., Platzker, A., Masse, D., and Statz, H. (1991) An AlGaAs/InGaAs pseudomorphic high electron mobility transistor (PHEMT) for X- and Ku-band power applications, *IEEE MTT-S Tech. Digest*, 713-716
65. Kim, B., Matyi, R.J., Wurtele, M., Bradshaw, K., Khatibzadeh, M.A., and Tserng, H.Q. (1989) Millimeter-wave power operation of an AlGaAs/InGaAs/GaAs quantum well MISFET, *IEEE Trans. Electron Devices* **36**, 2236-2242
66. Zhou, G.-G., Chan, K.T., Hughes, B., Mierzwinski, M., and Kondo, H. (1989) A pseudomorphic MODFET structure with excellent linear power performance at mm-wave range, *IEDM Tech. Digest*, 109-112
67. Huang, J.C., Saledas, P., Wendler, J., Platzker, A., Boulais, W., Shanfield, S., Hoke, W., Lyman, P., Aucoin, L., Miquelarena, A., Bedard, C., and Atwood, D. (1993) A double-recessed $\text{Al}_{0.24}\text{GaAs}/\text{In}_{0.16}\text{GaAs}$ pseudomorphic HEMT for Ka- and Q-band power applications, *IEEE Electron Device Lett.* **14**, 456-458
68. Smith, P.M., Ferguson, D.W., Kopp, W.F., Chao, P.C., Hu, W., Ho, P., and J.M. Ballingall (1991) A high power, high efficiency millimeter-wave pseudomorphic HEMT, *IEEE MTT-S Tech. Digest*, 717-720

69. Saunier, p., and Tserng, H.Q. (1989) AlGaAs/InGaAs heterostructures with doped channels for discrete devices and monolithic amplifiers, *IEEE Trans. Electron Devices* **36**, 2231-2235
70. Smith, P.M., Chao, P.C., Ballingall, J.M., and Swanson, A.W. (1990) Microwave and mm- wave power amplification using pseudomorphic HEMTs, *Microwave J.* May 1990, 71-86
71. Lai, R., Wojtowicz, M., Chen, C.H., Biedenbender, M., Yen, H.C., Streit, D.C., Tan, K.L., and Liu, P.H. (1993) High-power 0.15- μm V-band pseudomorphic InGaAs-AlGaAs-GaAs HEMT, *IEEE Microw. Guided Wave Lett.* **3**, 363-365
72. Streit, D.C., Tan, K.L., Dia, R.M., Liu, J.K., Han, A.C., Velebir, J.R., Wang, S.K., Trinh, T.Q., Chow, P.-M.D., Liu, P.H., and Yen, H.C. (1991) High-gain W-band pseudomorphic InGaAs power HEMT's, *IEEE Electron Device Lett.* **12**, 149-150
73. Matloubian, M., Nguyen, L.D., Brown, A.S., Larson, L.E., Melendes, M.A., and Thompson, M.A. (1991) High power and high efficiency AlInAs/GaInAs on InP HEMTs, *IEEE MTT-S Tech. Digest*, 721-724
74. Matloubian, M., Brown, A.S., Nguyen, L.D., Melendes, M.A., Larson, L.E., Delaney, M.J., Thompson, M.A., Rhodes, R.A., and Pence, J.E. (1993) 20-GHz high-efficiency AlInAs-GaInAs on InP power HEMT, *IEEE Microw. Guided Wave Lett.* **3**, 142-144
75. Matloubian, M., Brown, A.S., Nguyen, L.D., Melendes, M.A., Larson, L.E., Delaney, M.J., Pence, J.E., Rhodes, R.A., Thompson, M.A., and Henige, J.A. (1993) High-power V-band AlInAs/GaInAs on InP HEMT's, *IEEE Electron Device Lett.* **14**, 188-189
76. Wang, G.-W., Chen, Y.-K., Schaff, W.J., and Eastman, L.F. (1988) A 0.1- μm gate $\text{Al}_{0.5}\text{In}_{0.5}\text{As}/\text{Ga}_{0.5}\text{In}_{0.5}\text{As}$ MODFET fabricated on GaAs substrates, *IEEE Trans. Electron Devices* **35**, 818-23
77. Masato, H., Matsuno, T., and Inoue, K. (1991) $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}/\text{InAlAs}$ modulation-doped field effect transistors on GaAs substrates grown by low-temperature molecular beam epitaxy, *Jap. J. Appl. Phys.* **30**, 3850-3852
78. Bachem, K.H., Pletschen, W., Winkler, K., Fleissner, J., Hoffmann, C., and Tasker, P.J. (1993) $\text{AlGaInP}/\text{GaInAs}/\text{GaAs}$ -MODFETs with carbon doped p^+ -GaAs gate structure, a novel device concept, its implementation and device properties, *Proc. Int. Symp. GaAs and Related Compounds, Inst. Phys. Conf. Ser. No.* **136**, 35-40
79. Yang, D., Chen, Y.C., Brock, T., and Bhattacharya, K. (1992) DC and microwave performance of a 0.1- μm gate $\text{InAs}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ MODFET, *IEEE Electron Device Lett.* **13**, 350-352
80. Bolognesi, C.R., Caine, E.J., and Kroemer, H. (1994) Improved charge control and frequency performance in InAs/AlSb-based heterostructure field-effect transistors, *IEEE Electron Device Lett.* **15**, 16-18
81. Aina, O., Burgess, M., Mattingly, M., Meerschaert, A., O'Connor, J.M., Tong, M., Ketterson, A., and Adesida, I. (1992) A 1.45-W/mm, 30-GHz InP-channel power HEMT, *IEEE Electron Device Lett.* **13**, 300-302
82. Jelloian, L.M., Matloubian, M., Liu, T., Lui, M., and Thompson, M.A. (1994) InP-based HEMT's with $\text{Al}_{0.48}\text{In}_{0.52}\text{As}_x\text{P}_{1-x}$ Schottky layers, *IEEE Electron Device Lett.* **15**, 172-174

PSEUDOMORPHIC HEMT

MATERIALS GROWTH AND CHARACTERIZATION

DWIGHT C. STREIT

*TRW Electronics and Technology Division
Redondo Beach, California, USA*

1. Introduction

All other things being equal, the performance of field effect transistors is driven by the electron velocity in the channel. For large gate devices this may be the electron saturation velocity. For short gate devices this may be an effective overshoot velocity. In any case, one of the device engineer's principal tasks is to optimize the device structure in order to maximize the electron velocity in the channel, and thus maximize device performance.

The most obvious technique to increase the channel electron velocity is to use semiconductor material with a smaller electron effective mass. Gallium arsenide works better than silicon for microwave applications primarily because its effective mass is so much smaller. GaAs field effect transistors have been the workhorse of the microwave community for nearly 20 years. But in general GaAs MESFETs are effective only for applications through K-band, or about 25 GHz.

In the early 1980s GaAs-AlGaAs high electron mobility transistors, also called modulation doped field effect transistors, came into general use. The dopant atoms were in an AlGaAs donor layer that was physically removed from the GaAs channel where electron transport occurs. This physical separation of donor atoms and their associated electrons has several benefits, including increased overshoot velocity and lower noise due to fewer scattering centers being present in the channel. However, the device performance is still limited to the transport characteristics of the GaAs channel. In general, GaAs-AlGaAs HEMTs are useful through V-band, or about 60 GHz.

In the late 1980s pseudomorphic InGaAs-GaAs-AlGaAs HEMTs came into general use. The addition of InAs to the GaAs channel resulted in an

immediate improvement in the performance characteristics of these devices. The electron effective mass was reduced to an intermediate value between that of GaAs and InAs, depending on the amount of InAs added to the channel. On GaAs substrates the amount of InAs is limited to about 30%, making GaAs-InGaAs HEMTs useful for applications through W-band, or about 100 GHz.

The performance of HEMT devices is being pushed to even higher levels in the 1990s through the use of InP substrates. The use of InP substrates allows InGaAs channels of up to about 80% InAs. These pseudomorphic InGaAs-InAlAs-InP HEMTs have an electron effective mass in the channel approaching that of InAs, resulting in very high overshoot velocities. These devices are now being used at 140 GHz D-band frequencies, and will in the future be useful for applications above 200 GHz.

This chapter is structured somewhat along the lines of HEMT development. Most HEMTs are fabricated by molecular beam epitaxy, and this technology will be discussed as it applies to the growth of these devices. The material characterization tools that we use for the evaluation of HEMT material will also be discussed. Both InGaAs-AlGaAs-GaAs and InGaAs-InAlAs-InP pseudomorphic HEMT material growth and characterization will be covered from a practical point of view.

2. Molecular Beam Epitaxy

The technology of molecular beam epitaxy is especially suited for the production of the epitaxial layers necessary for pseudomorphic HEMTs. The cartoon shown below in Figure 1 illustrates the utility of MBE.

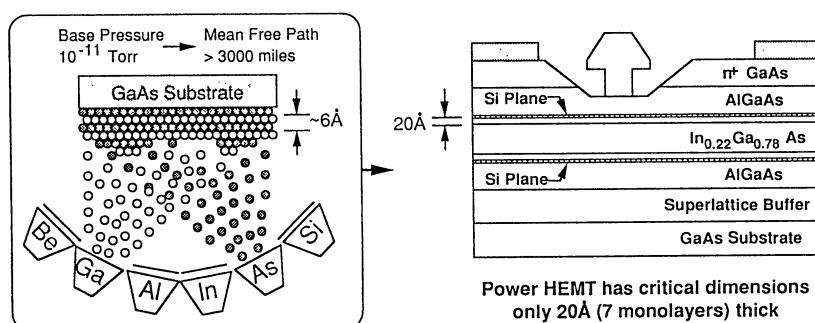


Figure 1. A schematic cartoon of the MBE process, with an InGaAs-GaAs power HEMT profile for comparison.

We are very fortunate because solid-source MBE agrees very well with HEMT production. No toxic gases or exotic growth procedures are required to produce these devices. The constituent elements necessary for HEMT growth are easily available in very high purity ingots, have reasonable vapor pressures at the temperatures available in normal MBE furnaces, and are solids at or near room temperature. Indium, aluminum, arsenic, and silicon are easily handled. Gallium melts at 29.8°C, and so a little reasonable care is required to ensure it remains solid until loaded into the MBE system, but it is not a problem.

One of the major advantages of MBE for HEMT production is the excellent uniformity that can be obtained. Figure 2 compares resistivity plots of relatively thick, heavily doped epitaxial films grown by MBE and MOCVD on 3-inch GaAs (100) substrates. The standard deviation of the resistivity measured on the MBE film is only 0.4%, whereas the MOCVD film has a standard deviation of 2.7%.

The uniformity of layer thickness, dopant concentration, and alloy composition are especially important for HEMTs because of the methods commonly used to process HEMTs. The device threshold voltage depends on the relationship between the gate position after gate etch and the distance to the InGaAs channel. Variations in thickness and alloy composition will change this spatial relationship and thus change the device threshold voltage. In addition, changes in the resistivity of the channel across the wafer will alter the channel current used to monitor the gate etch procedure, and add an additional variation to the threshold voltage.

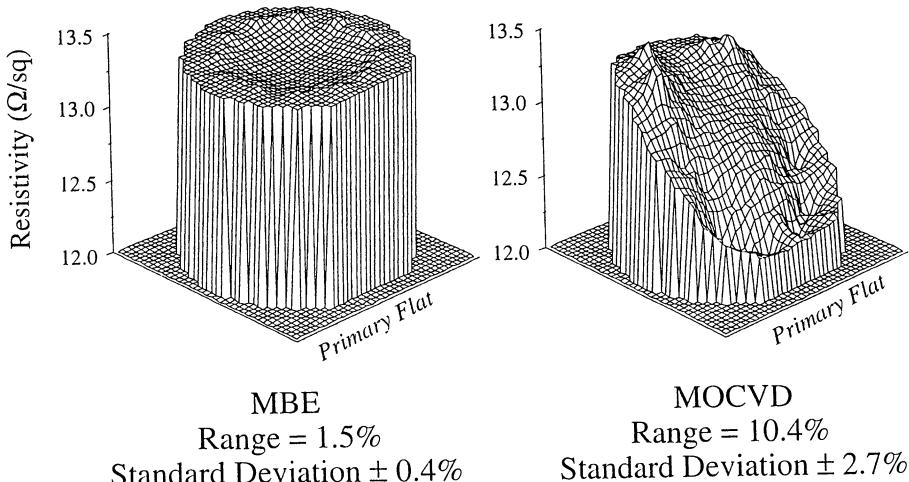


Figure 2. Non-contact eddy-current resistivity maps of heavily-doped multilayer MBE and MOCVD epitaxial films on 3-inch GaAs substrates.

3. MBE System Calibration

The growth techniques necessary to obtain good quality, high purity, GaAs films by solid-source MBE are well understood, and have been in use for nearly 20 years. There are a variety of commercially available MBE systems, and with around a thousand MBE systems built over the past two decades, almost every university and industrial compound semiconductor research or development laboratory has an MBE system available.

The ability to grow high-quality GaAs films is an absolute prerequisite to growing high-quality HEMTs. That means the MBE system is clean and well-cared for, has good liquid nitrogen shielding of all stainless parts, and has good control systems for reproducible layer timing, substrate and furnace temperatures, and temperature ramps. A clean MBE system is capable of growing epitaxial GaAs with background doping of Na-Nd $<10^{14}$ cm $^{-3}$ p-type films for As₄ sources, and Nd-Na $<10^{15}$ cm $^{-3}$ n-type films for cracked As₂ sources. We have used both types of arsenic sources for the growth of GaAs and InGaAs layers with equal success.

The calibration of the system is relatively straight forward, thanks to the well-behaved Arrhenius-type relationship of flux and temperature for each element, and their unity or reproducible sticking coefficients. The flux from each dopant or Group III effusion cell is described by $F = F_0 e^{-E_a/kt}$ cm $^{-2}$ s $^{-1}$ over the temperature range of interest. Thus when the log of the dopant concentration or flux from each cell is plotted versus 1/T, a calibration curve is obtained that can be used to determine the times and temperatures necessary for a given doping level or film thickness. Plots are shown below in Figure 3 for silicon dopant concentration in GaAs films grown at a constant 0.25 nm/s and aluminum flux as measured by an ion gauge.

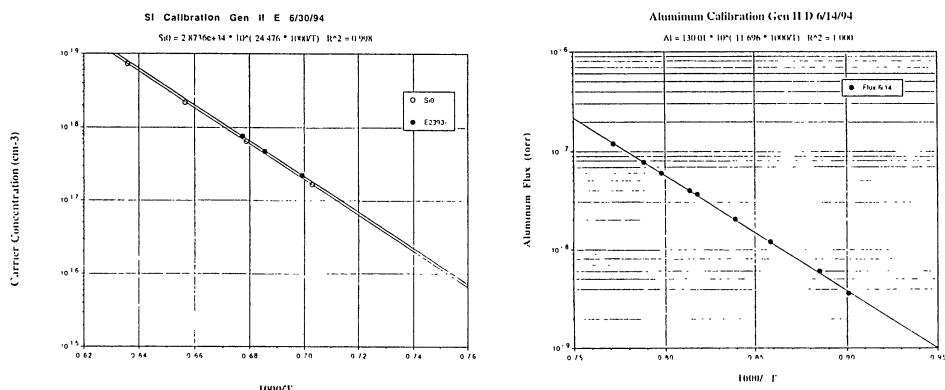


Figure 3. Measured Nd-Na and flux vs temperature for Si and Al cells.

Silicon is a well-behaved n-type dopant for (100) GaAs to concentrations near 10^{19} cm^{-3} . Its amphoteric behavior begins to show up at about $8 \times 10^{18} \text{ cm}^{-3}$, depending on the substrate temperature, arsenic flux, and growth rate. For this reason we limit our Si doping in GaAs or AlGaAs films to $6 \times 10^{18} \text{ cm}^{-3}$ in order to ensure an adequate safety margin.

Aluminum and gallium both have unity sticking coefficients below the maximum temperatures used for HEMT growth, or about 620°C . Thus once a growth rate versus ion-gauge flux is obtained at one flux, the curve as shown in Figure 3 can be used to obtain the growth rate at any other ion-gauge flux or cell temperature. These curves can then also be used to determine cell temperatures required for any AlGaAs composition at any desired growth rate in the range of interest.

Indium flux calibration is more subtle because its sticking coefficient is less than unity in the temperature range of interest, and only thin films may be grown before reaching the critical thickness and associated film relaxation. The absolute InGaAs channel composition can thus depend on the channel growth temperature. In addition, indium tends to segregate on the surface of the growing film, further complicating the issue. Methods to determine the absolute InGaAs composition in a thin strained InGaAs layer will be discussed later in this chapter.

Reflection high energy electron diffraction or RHEED oscillations are routinely used for growth rate calibrations. Typical GaAs and AlGaAs RHEED oscillations are shown in Figure 4. As the epitaxial film grows, the

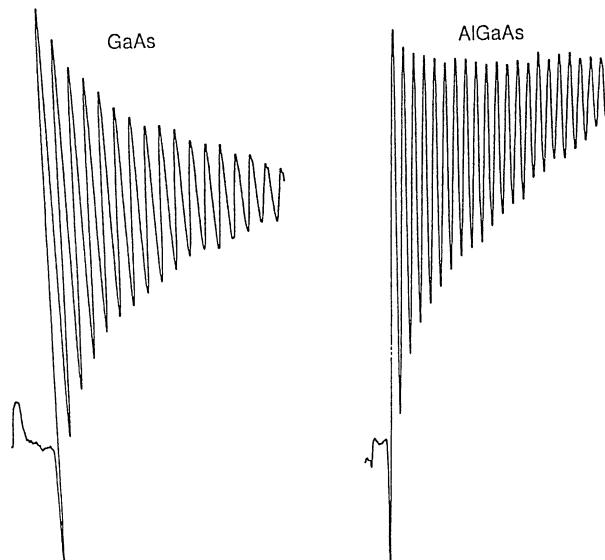


Figure 4. RHEED oscillations of GaAs and AlGaAs layers.

diffraction pattern oscillates between a minimum and maximum intensity due to the relative diffraction conditions as the surface layer fills. By simply plotting the diffraction intensity versus time on a chart recorder the growth rate at the beam spot can be determined. This technique is very useful, and there is a lot of interesting physics in what is happening to create the oscillations. However, the growth rate as measured can also be misleading. Unless sophisticated techniques are used to track a rotating substrate, the growth rate as measured in one spot on a stationary substrate may not be representative of the whole film. If this relationship is understood, RHEED can be used for routine growth rate calibrations.

X-ray and optical measurements on grown films provide good absolute calibration checks. Figure 5 shows a double-crystal X-ray diffraction spectra and a room-temperature photoreflectance spectra from an AlGaAs film with a nominal 30% AlAs content. The optical results yield 30.4% AlAs while the X-ray results using the latest AlAs parameters yields 30.6%. The difference between the values is less than the uncertainty associated with the PR and X-ray measurement techniques.

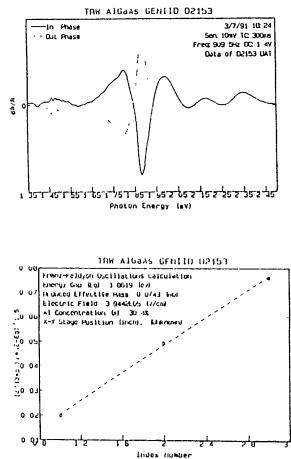
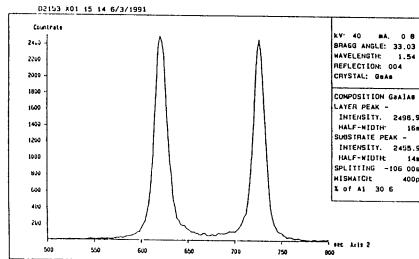
PR: $x = 0.304$ X-ray: $x = 0.306$ 

Figure 5. Photoreflectance and X-ray diffraction spectra of an AlGaAs film.

4. GaAs Substrate Preparation

The actual growth conditions used for the deposition of the different layers required for a multilayer HEMT structure may be different depending on the actual device profile and depending on where the film is being grown.

The growth conditions have been optimized here based on the overall quality of the device structure as measured by material characterization techniques and by the resulting device performance. Several conditions are recognized industry wide, however.

Substrate cleaning seems to always be an issue in the MBE community. There are many different chemical cleaning techniques available before the substrate is loaded into the system, there are different oxidation and ozone-cleaning steps that can be used, and there are many different in-situ techniques available for oxide desorption and surface cleaning. Epi-ready GaAs substrates are popular because the substrate cleaning and etching are done by the vendor, so less effort is required by the MBE operator.

Our experience has been that freshly cleaned and etched substrates using simple $\text{NH}_4\text{OH}-\text{H}_2\text{O}_2$ based solutions, and prepared immediately before loading, present reproducible surface conditions. Oxide desorption at 620°C leaves a clean well-ordered surface with carbon contamination of <less than 10^{12} cm^{-2} . We have determined that about 10% of the carbon at the substrate-epi interface is electrically active. Ozone cleaning is very effective at removing the interface carbon, but ozone oxides are difficult to desorb and therefore require higher substrate temperatures during the desorption step. A substrate prepared using ozone oxidation is more likely to have oxygen interface contamination, which in our experience is more deleterious to the epitaxial film than carbon interface contamination.

5. InGaAs-GaAs HEMT Growth

The overall growth and calibration scheme and the device profile for a typical low-noise InGaAs-GaAs HEMT used at TRW is shown in Figure 6. The device itself consists of an AlGaAs-GaAs buffer layer, an InGaAs channel, an AlGaAs planar-doped donor layer, and an n^+ GaAs contact layer. This overall HEMT profile is a staple in the microwave industry, with minor variations to the buffer layer design, alloy compositions, layer thickness, and doping levels. The modulation doping used for GaAs-AlGaAs HEMTs as described in the introduction is still present, although for certain power applications some profiles use a doped InGaAs channel.

The design of MBE buffer layers seems to be an issue that yet to be fully resolved in the MBE community. There are probably as many buffer-layer designs as there are MBE operators. The purpose of the buffer is to confine electron transport to the InGaAs channel. Any transport in the buffer layer, or electron injection through the buffer and into the substrate, provides a parallel conductance path and will degrade the device performance.

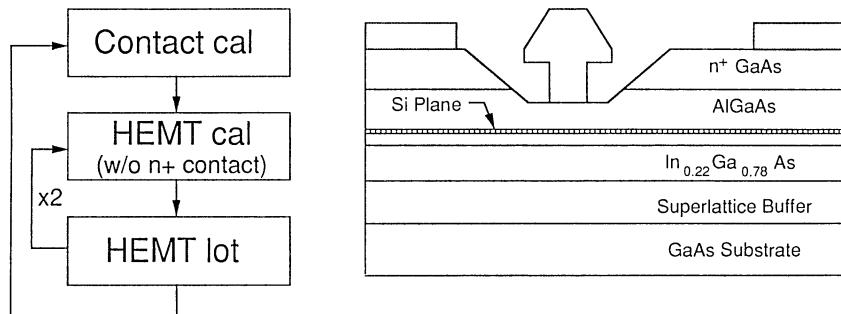


Figure 6. Calibration procedure and device profile for low-noise HEMTs.

The growth conditions used for the buffer layer will determine the quality of the buffer. Most HEMT buffers have AlGaAs in them to provide an increase in the bandgap. Superlattices were originally used to inhibit outdiffusion of contaminants and defects from the substrate, and are still incorporated in many designs. The use of As₄ usually results in a p-type buffer. A p-type buffer is probably preferred to the n-type buffer resulting from the use of As₂ in an unintentionally doped film. GaAs buffers grown at low temperature have been reported for nearly a decade with mixed results, and are not yet used as a standard buffer due to performance reproducibility and reliability concerns.

The InGaAs channel is typically grown at reduced substrate temperature for best film quality due to the tendency of indium to surface segregate at higher temperatures. The preferred temperature window for As₄ is narrower than that for As₂, which somewhat complicates the situation. The less than unity sticking coefficient for indium also adds an additional degree of freedom to the InGaAs channel composition, but with accurate substrate temperature reproducibility this effect is negligible. The absolute InGaAs composition is important for performance and model reproducibility. There are performance advantages to using higher InAs mole fractions, but these can be offset by increased reproducibility and reliability concerns, as will be discussed later.

6. The Pseudomorphic InGaAs Channel

The addition of InAs to the GaAs channel results in an alloy with reduced electron effective mass. It also results in an alloy with a larger lattice parameter. The bandgap versus lattice constant, and effective mass versus

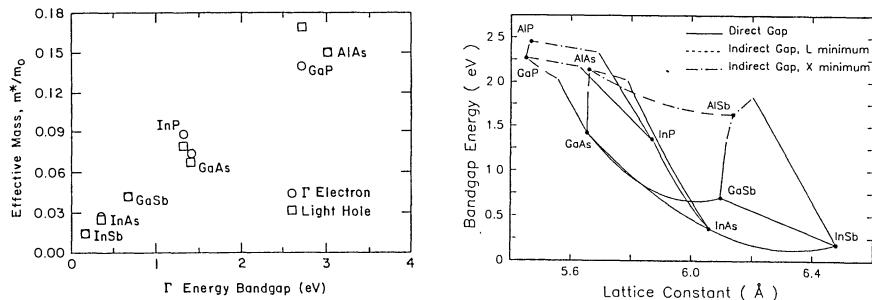


Figure 7. Effective mass, bandgap, and lattice parameter for III-V alloys.

bandgap are shown in Figure 7. In general, the larger, heavier alloys have smaller bandgaps and smaller effective masses. But they also have larger lattice parameters that complicate the growth of these films.

The technique used to produce HEMTs with the desired InGaAs channel composition is simply to grow a lattice-mismatched InGaAs film. This is shown schematically in Figure 8. The GaAs substrate and GaAs/AlGaAs buffer have lattice parameter $a_0 = 5.653\text{\AA}$. An InGaAs film with about 28% InAs mole fraction has a nominal lattice parameter about 2% larger than GaAs, or $1.02 a_0$. This difference in lattice parameter can be tolerated during the film growth by distorting the InGaAs crystal structure. The channel is put into compression in the directions parallel to the surface, and is put into tension in the direction parallel to the epitaxial growth.

The change in the InGaAs lattice parameter in the growth direction is related to the change due to the compression in the plane by Poisson's ratio, which for III-V semiconductors is about 0.3. The energy required to achieve this distortion in the InGaAs lattice accumulates, and is eventually relieved through dislocations when the critical thickness is reached.

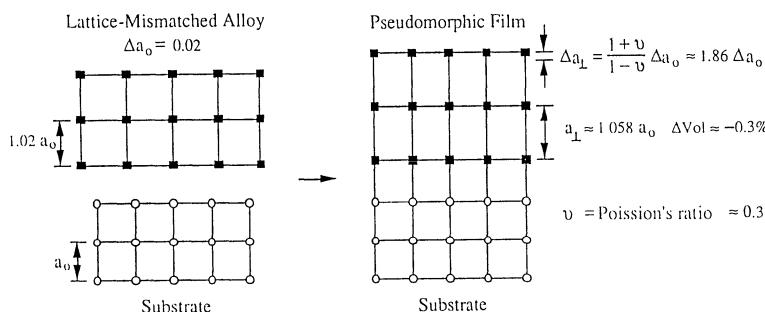


Figure 8. Schematic of pseudomorphic epitaxial InGaAs channel growth.

7. Statistical Process Control

We use statistical process control of the critical HEMT parameters in order to ensure reproducibility of the epitaxial growth process. Parameters that are tracked include InGaAs and AlGaAs composition, sheet resistivity, as well as channel sheet charge and mobility,

The AlGaAs donor layer composition and net dopant concentration are especially important. The AlGaAs composition and thickness affects the device breakdown and gate recess etching characteristics. The net dopant concentration likewise affects breakdown and gate etching, as well as channel current and threshold voltage. Control of the InGaAs channel and AlGaAs donor layer compositions, and control of the net sheet charge in the channel are critical to control of the device process and performance.

The procedure used for calibration of the HEMT profiles was shown earlier in Figure 6. A special HEMT calibration sample is grown with each six-wafer lot. The calibration sample is exactly the same as the normal device, except a thin undoped GaAs layer is substituted for the n⁺ GaAs contact layer. This allows unequivocal measurement of the channel sheet charge and mobility without the uncertainties associated with the presence of the n⁺ contact layer.

Figure 9 shows the SPC charts of channel doping for the low-noise profile of Figure 6. The target sheet charge for this profile is $2.6 \times 10^{12} \text{ cm}^{-2}$. The average sheet charge \pm one standard deviation measured over 228 wafers was 2.57 ± 0.09 , and in a separate system measured over 102 wafers the average sheet charge was $2.58 \pm 0.08 \times 10^{12} \text{ cm}^{-2}$.

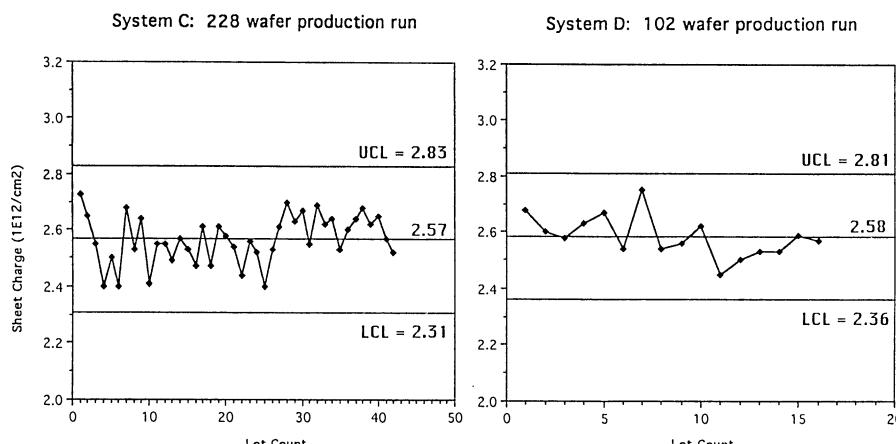


Figure 9. SPC control charts of channel sheet charge in the same HEMT profile produced by two different MBE systems at different times.

8. InGaAs Channel Composition

The performance of InGaAs-GaAs HEMTs is known to improve with increasing InAs content in the InGaAs channel. The absolute composition of the channel is limited when grown on GaAs (100) substrates to about $In_{0.3}Ga_{0.7}As$ due to the previously mentioned surface segregation tendency of indium. At higher compositions the InGaAs film tends to decompose into segregated InAs and GaAs regions, which of course completely destroys the transport properties of the channel.

Most device designs call for channel compositions ranging from 15% to 25% InAs. Why not use 30% InAs if this is achievable? It is clear that the channel properties improve with increasing InAs content, as shown in Figure 10. The room temperature Hall mobility increases from $5830 \text{ cm}^2/\text{Vs}$ for a 15% InAs channel to $6730 \text{ cm}^2/\text{Vs}$ for a 28% InAs channel. Likewise the cutoff frequency of the device increases, indicating improved electron velocity in the channel.

The most likely reason for specifying lower InGaAs compositions than can be achieved is for producibility and reliability reasons. The absolute composition control typically achieved is on the order of 0.02 mole fraction. In order to maintain high yield in the MBE process the channel composition must be specified lower than the absolute maximum.

We have conducted extensive reliability tests on InGaAs-GaAs HEMT material. We have found that for $In_{0.22}Ga_{0.78}As$ channels the material is extremely robust. However, HEMT material with $In_{0.28}Ga_{0.72}As$ channels are not as robust as that with $In_{0.22}Ga_{0.78}As$ channels. A comparison of sheet charge and mobility results at 295K and 77K for HEMTs annealed at different temperatures is shown in Figure 11. The $In_{0.22}Ga_{0.78}As$ device material showed no change in sheet charge with annealing up to 700°C , whereas the $In_{0.28}Ga_{0.72}As$ device material showed reduction in sheet charge as low as 500°C . For this reason our baseline material is $In_{0.22}Ga_{0.78}As$.

Wafer No.	Indium Comp. (x)	Channel Width (Å)	300K		77K	
			Ns (cm^{-2})	Mobility (cm^2/Vs)	Ns (cm^{-2})	Mobility (cm^2/Vs)
D1306	0.15	190	2.3E12	5830	2.2E12	22,200
D1299	0.20	150	2.5E12	6520	2.5E12	18,800
D1304	0.28	120	2.7E12	6730	2.6E12	14,600

Figure 10. Hall sheet charge and mobility for varying InGaAs compositions.

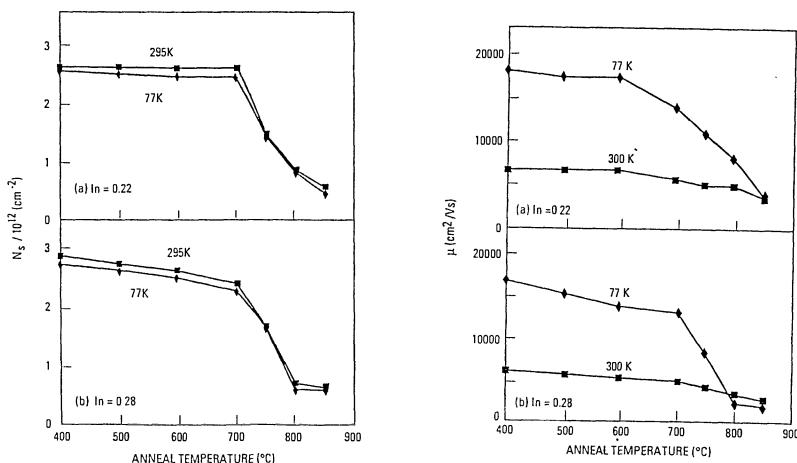


Figure 11. Hall sheet charge and mobility as a function of annealing temperature for $\text{In}_{0.22}\text{Ga}_{0.78}\text{As}$ and $\text{In}_{0.28}\text{Ga}_{0.72}\text{As}$ channel HEMTs.

9. Photoluminescence Characterization of InGaAs-GaAs HEMTs

The calibration sample grown with each HEMT lot has many uses. Because it is dedicated to characterization, the wafer can be cleaved and measured using destructive techniques. In addition to the Hall measurements made on every characterization wafer, photoluminescence measurements are made using a computerized liquid helium immersion PL system.

Figure 12 shows the spectra predicted for a $\text{In}_{0.22}\text{Ga}_{0.78}\text{As}$ HEMT with a 14 nm channel. The high energy peak is associated with transitions from the second conduction-band level to the first heavy-hole level in the valence

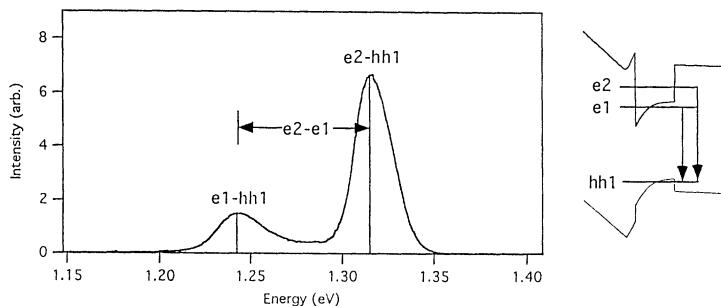


Figure 12. Photoluminescence model for $\text{In}_{0.22}\text{Ga}_{0.78}\text{As}$ HEMT profile.

band. The low energy peak is associated with transitions from the first conduction band level to the heavy hole level. The peak separation is associated with the well width. Monitoring the 4K photoluminescence can thus give information on absolute InGaAs composition as well as absolute InGaAs channel thickness.

Figure 13 shows the modeled results for low-noise InGaAs HEMT profiles with varying InGaAs channel composition. The e1-hh1 transition energy can be used to determine InGaAs channel composition. The well widths versus peak energy separation assumes an $\text{In}_{0.205}\text{Ga}_{0.795}\text{As}$ channel.

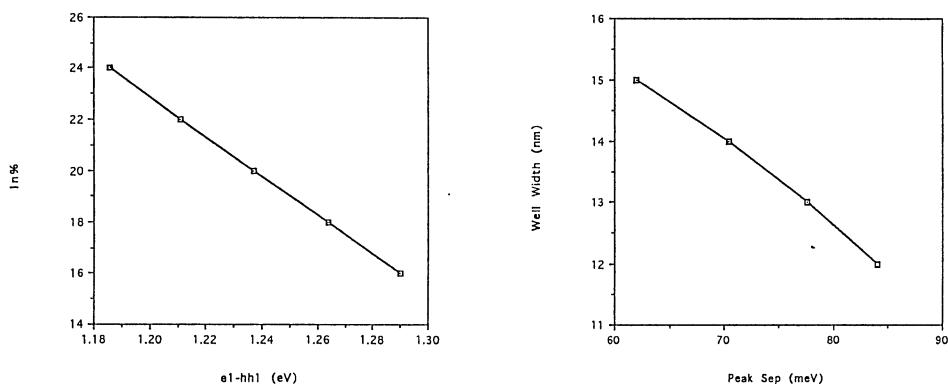


Figure 13. Predicted InGaAs channel composition versus transition energy and well widths versus peak energy separation for $\text{In}_{0.205}\text{Ga}_{0.795}\text{As}$.

The beauty of this 4K photoluminescence technique is that several important parameters can be monitored through one measurement. The peak shape and intensity gives an indication of the overall material quality. Low intensity would indicate unusual recombination, most likely in the AlGaAs or InGaAs regions. The transition energies allow both InGaAs composition and well width to be monitored and tracked by SPC.

Figure 14 shows the 4K PL spectra for two different HEMT wafers grown with the same nominal profile over one year apart in different MBE systems. The similarity in the photoluminescence spectra intensity and profile shape indicates the similarity in the InGaAs channel profile. The difference in peak energy separation indicates that the channel thicknesses are not exactly the same, but about 5 Å different. The important thing is to be able to analyze the profiles and to understand the differences between material grown to the same specifications. Figure 15 shows the SPC chart derived from spectra such as that in Figure 14 for the same 228 wafer production run as that shown in the Figure 9 sheet charge SPC chart.

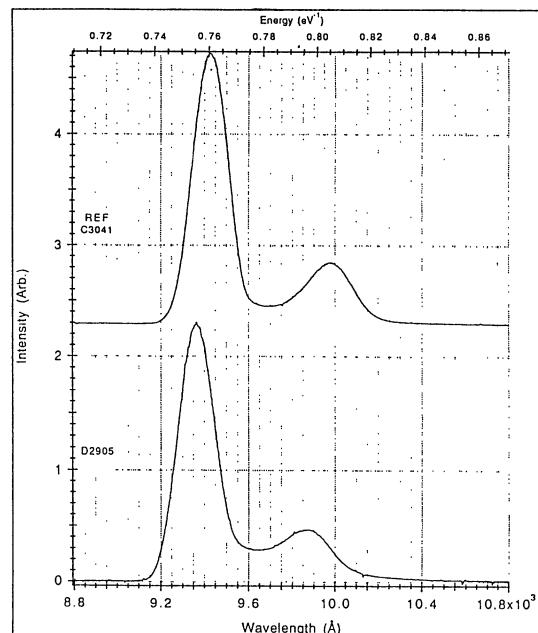


Figure 14. 4K photoluminescence from two different HEMT wafers with the same nominal profile, grown in two different systems one year apart.

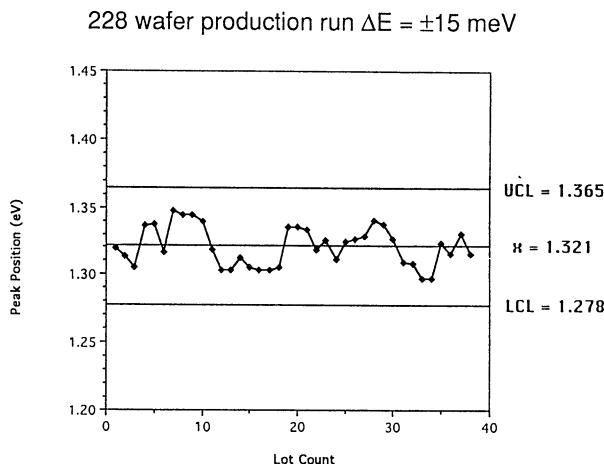


Figure 15. SPC chart of channel composition as deduced from the e2-hh1 transition using 4K photoluminescence. This is the same 228 wafer production run shown in Figure 9.

10. X-Ray Characterization of InGaAs-GaAs HEMTs

Recent advances in the sensitivity and resolution of multi-crystal X-ray diffraction systems have allowed the comparison of the measured and modeled results for very sophisticated epitaxial profiles. Figure 16 shows the measured versus modeled results for the same profile used in the photoluminescence results of the last section. Figure 17 shows the difference in the model with a 5Å change in the well thickness. Verification of the profile by multiple analysis techniques adds confidence to their validity.

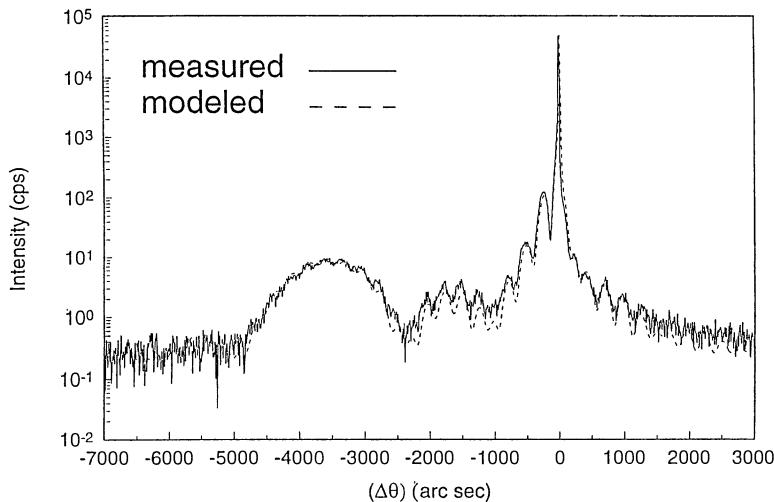


Figure 16. Three-crystal X-ray spectra and model of InGaAs-GaAs HEMT.

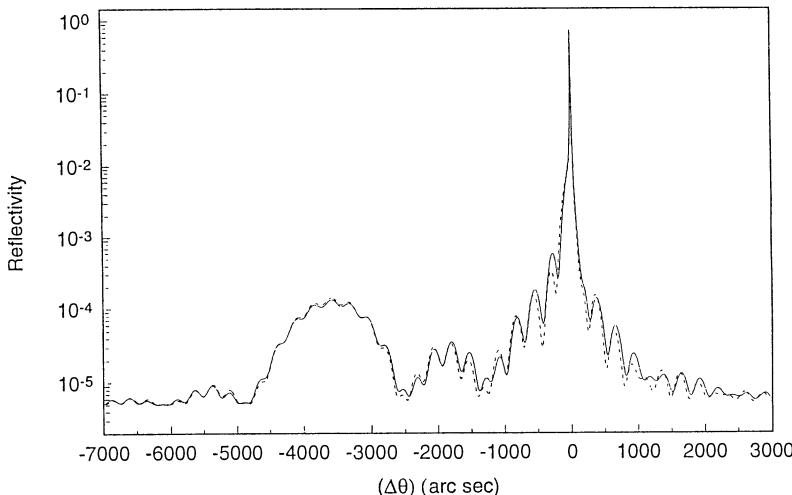


Figure 17. X-ray spectra of HEMT with 5Å change in well thickness..

11. Photoreflectance Characterization of InGaAs-GaAs HEMTs

Photoreflectance and electroreflectance are room temperature modulation spectroscopy techniques that are very useful for HEMT productoin because they are noncontact, nondestructive techniques.

Just as many HEMT features can be deduced from the 4K photoluminescence spectra, many HEMT features can be deduced from the photoreflectance spectra. The AlGaAs transistions yield the AlGaAs bandgap and thus composition. The InGaAs transistions labeled A,B, and C allow the channel composition to be determined. The MQW transistion allows the GaAs growth rate to be determined because the AlGaAs composition is known thus the peak location is uniquely determined from its energy.

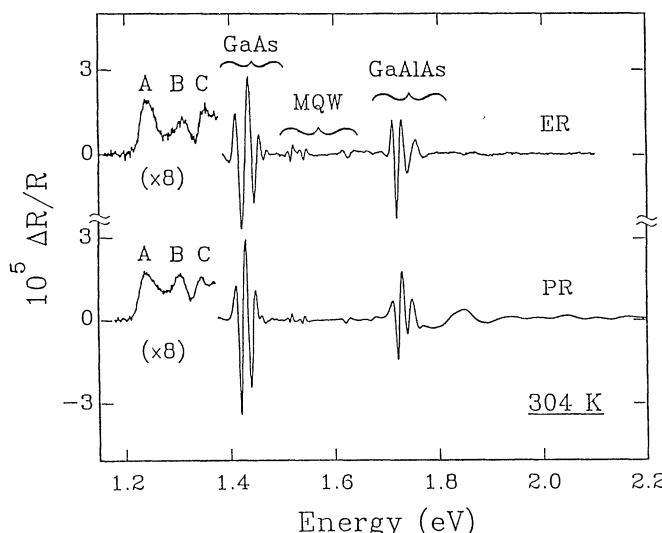


Figure 17. Photoreflectance and electroreflectance spectra of a power HEMT profile with a nominal $\text{In}_{0.22}\text{Ga}_{0.78}\text{As}$ channel. The peaks labeled A,B, and C are associated with the 21H, 32H, and 42H transistions. (with F. Pollak).

	Sheet Charge 10^{12} cm^{-2}	Field (kV/cm) Modeled	Field (kV/cm) Measured
D2561	3.3	138	137
D2569	3.6	142	145
D2576	4.0	146	151

Figure 18. Measured and modeled electric field in the donor layer of a power HEMT with different sheet charge density in the front plane.

12. Material Profiles for Increased Performance

Pseudomorphic InGaAs-GaAs HEMTs have added significantly increased performance capability to microwave system design than was available with MESFETs or GaAs-AlGaAs HEMTs. For many applications, however, front-end noise figure is so important that the search for noise figure reduction is never ending. How do we improve the performance even more?

The device equivalent circuit is the place to start in order to understand what material and profile characteristics are important for low noise, high gain devices. Fukui's empirical noise equation says that the minimum noise figure is proportional to the gate-source capacitance, and proportional to the square root of the source resistance and gate resistance, and inversely proportional to the transconductance. $NF = 1 + K_w C_{gs} [(R_s + R_g) / g_m]^{1/2}$.

We can look at Figure 19 and get some ideas about what to optimize in the device profile to improve the performance. We want high g_m , low $C_{gs} / g_m^{1/2}$, and low R_s and R_g .

We can improve g_m by adding InAs to the channel and thereby reducing the electron effective mass and improving the electron velocity. To get above $In_0.3Ga_0.7As$ we will use an InP substrate. Looking back to Figure 7 we can see that an InP substrate will allow us to obtain lattice-matched $In_0.53Ga_0.47As$. Using the same pseudomorphic techniques here we can go to 70% or 80% InAs with reasonable channel thickness.

Another technique to increase g_m is to minimize $d + \partial d$, the distance from the gate to the centroid of the electron sheet charge in the channel. If we make the InAlAs donor layer with as small a spacer layer as possible and still get good channel transport, that will help. If we etch the channel recess so the gate is close to the channel, that will help too. But we can't get too close, because that would make the device more enhancement mode, which can

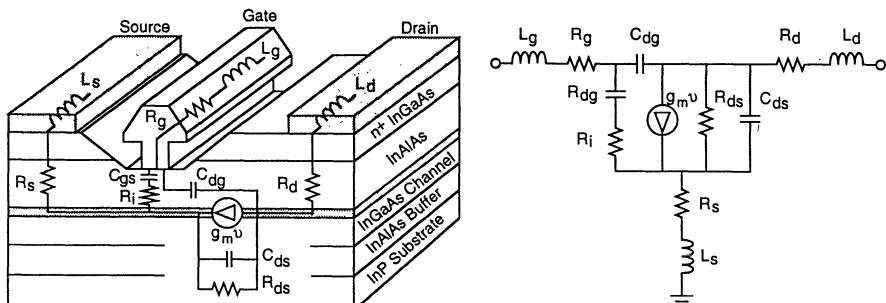


Figure 19. HEMT device equivalent circuit.

increase the gate leakage, which would hurt the noise figure. We want to keep g_m peak just about at $V_g=0$ V.

In order to minimize the source resistance we need to use a highly doped InGaAs contact. This will reduce the device breakdown, but that is acceptable since most low noise devices will only operate at one volt V_d . We can maximize R_{ds} by using an InAlAs buffer layer, which presents a very large barrier to electron injection into the buffer layer. So when we're all done, the device profile will look something like Figure 20.

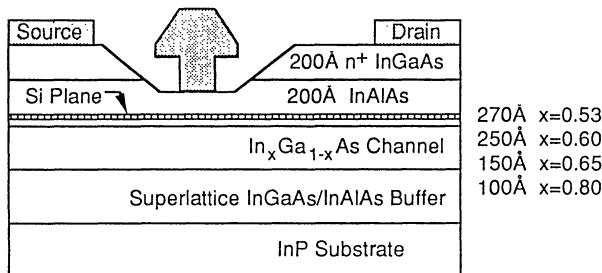


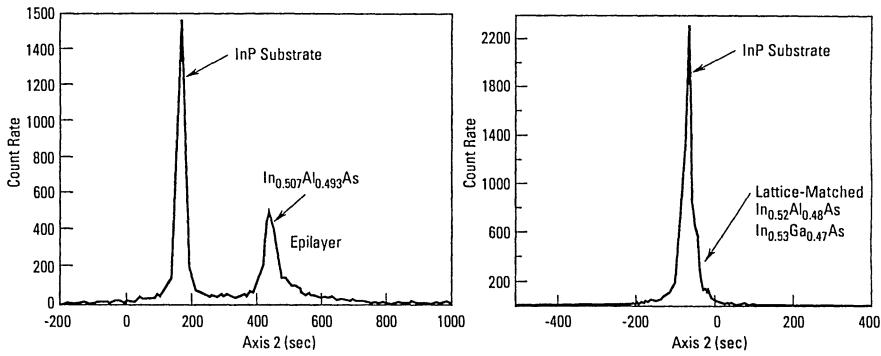
Figure 20. InGaAs-InAlAs-InP HEMT profiles with increasing InAs content.

13. InGaAs-InAlAs-InP Pseudomorphic HEMT

Pseudomorphic HEMTs build using InP substrates are inherently more difficult to grow than those on GaAs substrates. With GaAs, the buffer layer is naturally lattice-matched, and calibration procedures are relatively straight forward.

The difficulty in growing lattice-matched buffers can be seen in Figure 21. An InAlAs buffer layer grown on an InP substrate is 1.3% off in composition. This corresponds to less than 1°C in the aluminum furnace temperature. The temperature of the cell is adjusted, and the next film is nearly perfectly lattice-matched. Some tolerance in alloy composition is allowed since these films are typically pretty thin, on the order of 0.5 μm . But the fact that the epitaxial layers are not normally lattice matched to the substrate does require that the MBE system be nearly perfectly calibrated.

The benefits of increasing InAs composition can be clearly seen in Figure 22. The room temperature mobility increases from 9,750 cm^2/Vs for lattice matched $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ to 12,210 cm^2/Vs for $\text{In}_{0.65}\text{Ga}_{0.35}\text{As}$. This increase in mobility, and more importantly channel sheet resistance, translates directly to improved device performance.



Grow and measure calibration
X-ray shows InAs 1.3% low

Adjust MBE parameters
Next run has perfect composition

Figure 21. Double-crystal X-ray analysis of InAlAs calibration films.

Channel $In_xGa_{1-x}As$	$300K N_s$ (10^{12} cm^{-2})	$300K \mu$ ($\text{cm}^2/\text{V}\cdot\text{s}$)	$77K N_s$ (10^{12} cm^{-2})	$77K \mu$ ($\text{cm}^2/\text{V}\cdot\text{s}$)
0.53	3.04	9,750	2.92	32,300
0.60	3.15	10,200	3.01	40,700
0.65	2.71	12,210	2.54	46,300
0.80	3.61	10,750	3.36	33,700
Graded 0.80	3.94	12,200	3.82	50,400

Figure 22. Hall results for increasing InAs content in the InGaAs channel.

14. Benefits of Increased InAs Mole Fraction in InGaAs Channels

The ultimate test of all of the profile changes recommended here is in the device performance. Figure 23 shows the relationship between InGaAs composition and channel sheet charge and device cutoff frequency. It is clear that the benefits are real until about 80% InAs. At the point the channel thickness is so thin the the energy levels in the well are pushed up and confinement is not as good. The 0% InAs shown in Figure 23 is for a GaAs-AlGaAs HEMT, the 22% is for an InGaAs-GaAs HEMT, while the remainder are for InGaAs-InAlAs-InP HEMTs.

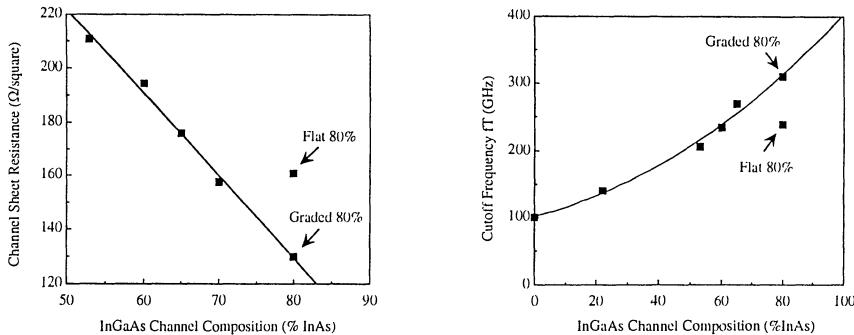


Figure 23. Relationship of channel sheet resistance and device cutoff frequency to InGaAs channel composition.

The design of the InGaAs channel is also important. There are several different designs, with flat profiles, step profiles, and graded composition profiles in the literature. Figures 24 and 25 show the difference in electron energy levels for two different profiles. Figure 24 is a 150Å In_{0.65}Ga_{0.35}As flat profile, while Figure 25 is a step profile with a flat 200Å In_{0.53}Ga_{0.47}As region and a 75Å In_{0.8}Ga_{0.2}As near the gate.

We have found a graded-composition channel to be most effective in achieving the goal of high cutoff frequency, high f_{max}, and low noise figure. However, our baseline profile is a flat channel design for the best reproducibility. Integrated circuits built using these devices are only as good as the model. The best circuit performance comes from a reproducible profile that yields a reproducible model.

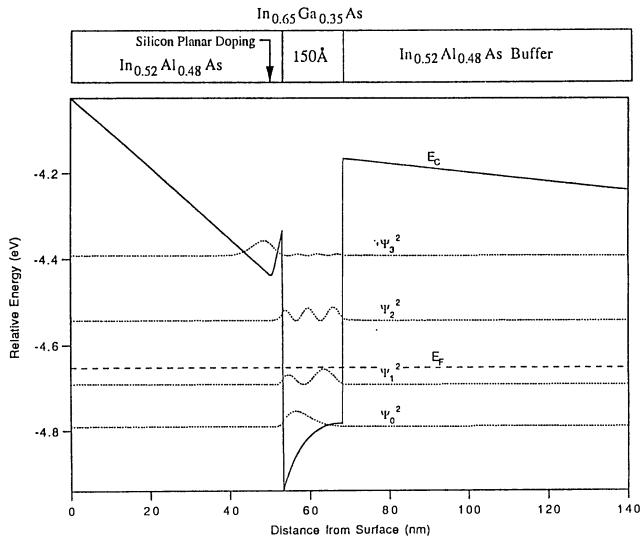


Figure 24. Modeled profile of flat channel InGaAs-InAlAs-InP HEMT.

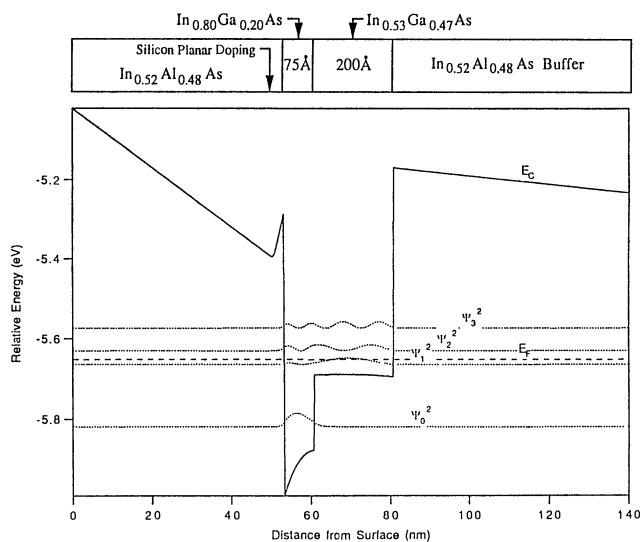


Figure 25. Modeled profile of step channel InGaAs-InAlAs-InP HEMT.

15. Conclusion

Pseudomorphic InGaAs-AlGaAs-GaAs HEMTs are now a production mainstay of the microwave industry. They are used for low-noise and power applications through 94 GHz. Production of the epitaxial material for these devices has become a commercial venture, with numerous vendors offering a variety of profiles.

The next generation HEMT devices will be based on InGaAs-InAlAs-InP material. These advanced profiles offer even lower noise and higher gain than their InGaAs-GaAs cousins. Development of power devices for W-band and higher frequencies using InP-based HEMTs is underway.

All of these devices have so far been based on molecular beam epitaxy material. With higher volumes and commercial applications, alternative growth methods are being developed. However, with proper production techniques it is possible to produce epitaxial material for HEMT devices that costs about the same as the substrate material. The total cost for substrate and epitaxial material then becomes almost a negligible part of the total cost of device wafer production.

Acknowledgements

The author gratefully acknowledges his colleagues at TRW for their contributions to this work, including Kin Tan and Rich Lai for HEMT results, Mike Wojtowicz for photoluminescence and modeling results, Tom Block for MBE, as well as Mark Goorsky at UCLA for X-ray results and Fred Pollak at Brooklyn College for photoreflectance results and modeling.

Gate Formation Technologies

P.C. CHAO

*Electronics Laboratory, Lockheed Martin Co.,
Syracuse, New York 13221*

1. Introduction

Since its first MMW demonstration in 1986, the GaAs-based pseudomorphic HEMT (PHEMT) has stimulated great interest for high-speed and high frequency low-noise and power applications. The Schottky gate, a technology involving reliable metallization, low parasitic resistance, small size, and recess control with precision, is one of the most important elements in determining the performance, yield, and reliability of the device. This note describes PHEMT gate process and its impact on the device performance. In the gate process, both commonly used self-aligned and etched channel gate technologies are compared. The gate recess technique and the low-resistance T-gate formation process are also described. Finally, the gate related failure mechanism is discussed.

2. Gate Lithography Approach

Figure 2-1 shows the cross-sectional view of a GaAs PHEMT. Figure 2-2 shows a typical PHEMT process flow chart. The gate process includes lithography, recess, metallization and liftoff.

In a PHEMT, to achieve high speed, low-noise figure and high efficiency at high frequencies, the device requires a very short gate length. Fabricating very short gates requires well-developed lithography and pattern transfer techniques. I-line stepper or deep UV optical lithography has been used to produce short gates. It has the advantage of

small capital investment and high throughput. However, the gate length over the entire wafer and from wafer-to-wafer may not be as consistent as what can be achieved using direct-write electron-beam

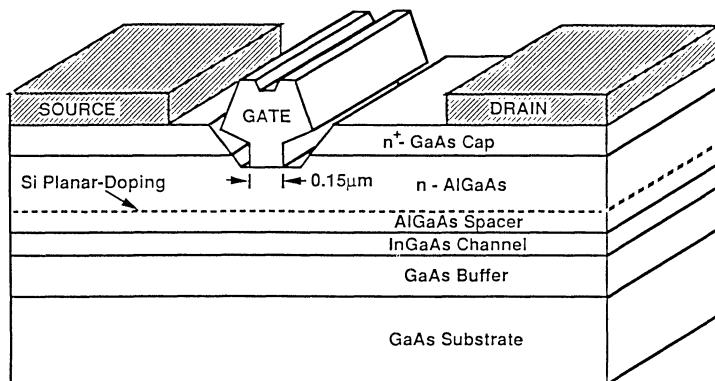


Figure 2-1 Cross-sectional view of a GaAs/AlGaAs/InGaAs PHEMT.

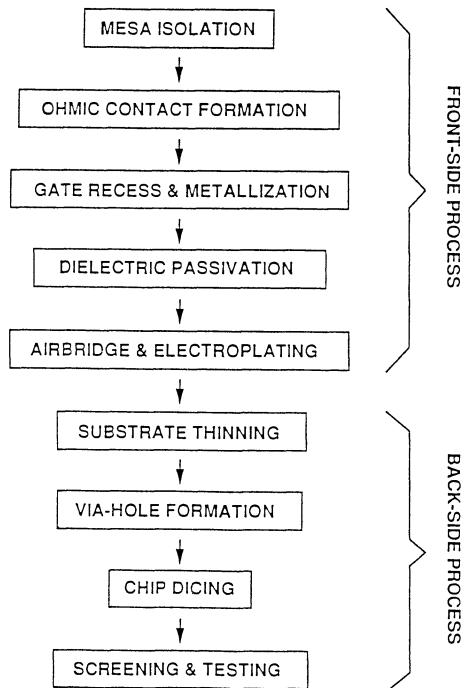


Figure 2-2 Typical PHEMT process flow chart.

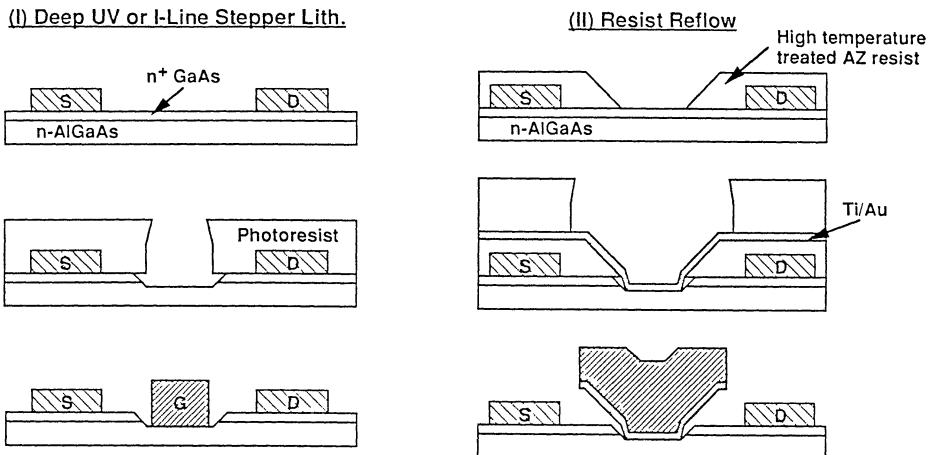


Figure 2-3 Examples of short-gate fabrication using optical lithography.

- Optical lithography

- Proximity, contact, or step-and-repeat projection printing.
- Typically used to define $0.5\text{-}1.0\mu\text{m}$ gates with high throughput.
- With deep UV exposure or special trick techniques (e.g., angle-evaporation), gates shorter than $0.5\mu\text{m}$ have been demonstrated.

- E-beam lithography

- Direct write, no hard mask required, fast turn-around.
- Low throughput.
- Typically used to produce $0.1\text{-}0.5\mu\text{m}$ gates with excellent uniformity and reproducibility.

Table 2-1 Comparison of gate lithography approaches.

(E-beam) lithography. Figure 2-3 illustrates two short-gate fabrication techniques using optical lithography.

E-beam lithography has widely been used to produce gates with dimension of $0.5\mu\text{m}$ and smaller. This E-beam approach facilitates accurate definition and alignment of very fine geometry and, most importantly, provides flexibility and fast turn around in design alterations.

One of the shortcomings of E-beam lithography is the low exposure throughput. To improve the wafer exposure throughput, an optical / E-beam “mixed” or “hybrid” lithography is typically used. In the hybrid lithography, a direct-write E-beam defines only the very short gate length for the PHEMT. Optical lithography, using a high throughput wafer stepper, for example, is then used for the remaining coarse features such as gate pad and interconnect. The hybrid lithography has the advantage of both the high throughput of optical lithography and the high resolution and good registration accuracy of E-beam lithography, and is the industry standard for the production of PHEMTs and MMICs requiring submicrometer gates. The gate lithography approach is summarized in Table 2-1.

3. Effect of Gate on PHEMT RF Performance

The cutoff frequency f_T is a function of gate length. The maximum frequency of operation f_{\max} , on the other hand, is strongly affected by gate length as well as gate resistance. For noise applications, the minimum noise figure of a PHEMT can be expressed as,

$$F_{\min} (\text{dB}) = 10 \log (1 + kfC_{gs} ((R_g + R_s) / g_m)^{1/2}) \quad (3.1)$$

where F_{\min} is the minimum noise figure, k is the Fukui constant, f is frequency, C_{gs} is input gate capacitance, g_m is transconductance, and R_g and R_s are the gate and source resistances, respectively. The above noise equation can also be expressed as

$$F_{\min} (\text{dB}) = 10 \log (1 + cf) \quad (3.2)$$

where c is a constant. For good low-noise performance, the device must have low parasitic gate resistance and capacitance. Figure 3-1 compares the best reported noise performance of $0.25\mu\text{m}$ and $0.15\mu\text{m}$ gate-length PHEMTs [1]. The $0.15\mu\text{m}$ HEMTs have consistently demonstrated lower noise figures than the $0.25\mu\text{m}$ ones over the microwave and mm-wave frequency range. A significant improvement in the device associated gain is also observed with reducing the gate-length. It is, however, very important to note that simply reducing the gate-length of a device does not guarantee better noise performance, since the gate resistance and other parasitic elements can dominate.

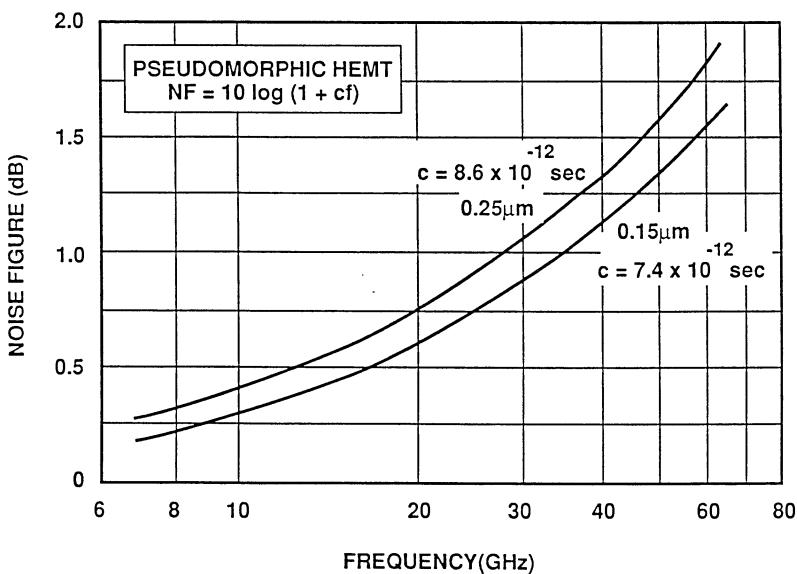


Figure 3-1 Best reported noise performance of 0.15 and $0.25\mu\text{m}$ PHEMTs.

For power applications, in order to obtain higher power from PHEMTs, shorter gate-length and larger gate periphery devices composed of many inter-digitated fingers are required (see Figure 3-2). However, as gate unit finger width increases, efficiency and gain are degraded. This is largely as a result of two parasitics -- the source inductance and the gate resistance. The effect of source inductance increases with frequency and, depending on the device layout, may also increase with gate width. The device gate resistance, a function of gate length, linearly increases with gate width. With proper device design to minimize the effect of gate resistance (e.g., Larger T-gate metal cross-section, shorter unit finger width, and airbridge gate structure), power PHEMTs should produce high output power with high efficiency and gain.

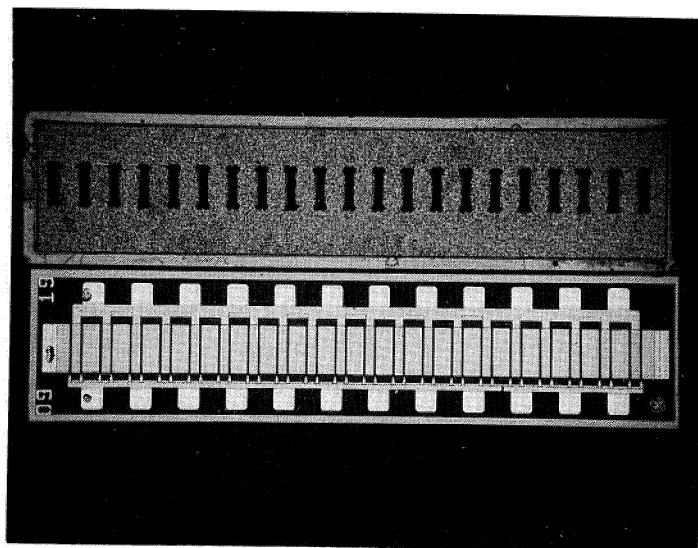


Figure 3-2 Inter-digitated gate power PHEMT (total gate width $40 \times 100\mu\text{m}$).

4. Gate Fabrication Technologies

The key elements in processing a high performance PHEMT is illustrated in Figure 4-1. The gate process is most critical among these elements. The requirements of the gate processing steps are detailed in Table 4-1.

Gates can be formed using either a self-aligned gate technology (Figure 4-2) or an etched channel gate technology (Figure 4-3). The former one is typically adopted to fabricate digital ICs where a uniform threshold voltage is required. The etched (recess) channel gate technology, on the other hand, provides an ability to tweak channel current for low feedback capacitance, source resistance, and high breakdown voltage, and is most suitable for microwave applications. In this technology, after the gate lithography and resist development, the exposed PHEMT channel area is chemically recessed to achieve the desired channel current and threshold voltage prior to the gate metallization.

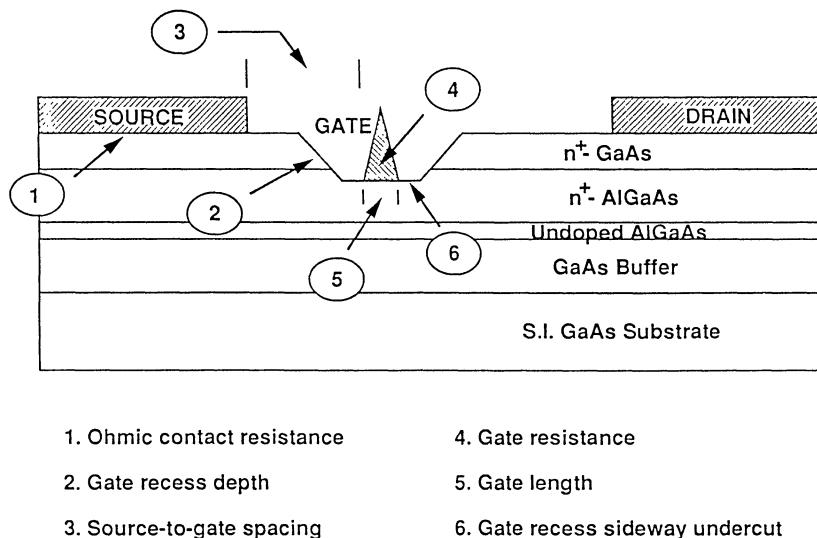


Figure 4-1 Key elements in processing a high performance PHEMT.

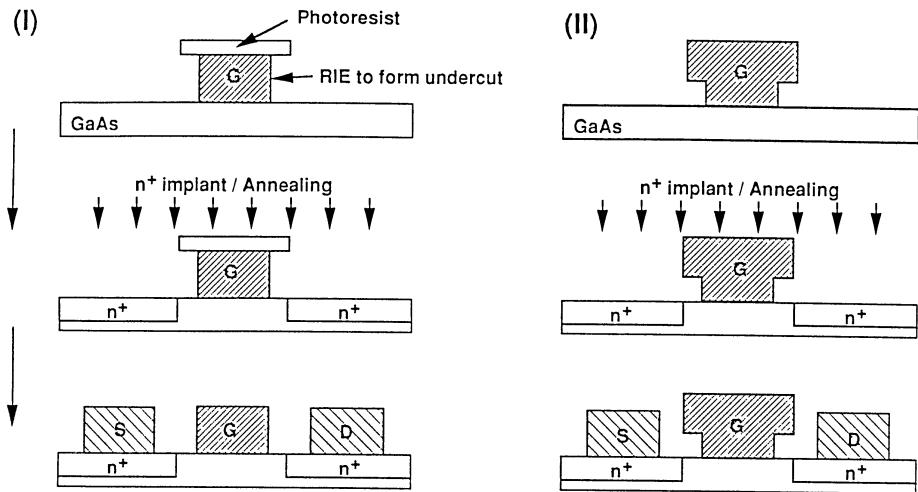


Figure 4-2 Self-aligned gate technology. Gate is defined first and used as a mask to define n^+ ohmic regions.

Gate should be capable of surviving high anneal temperature process.

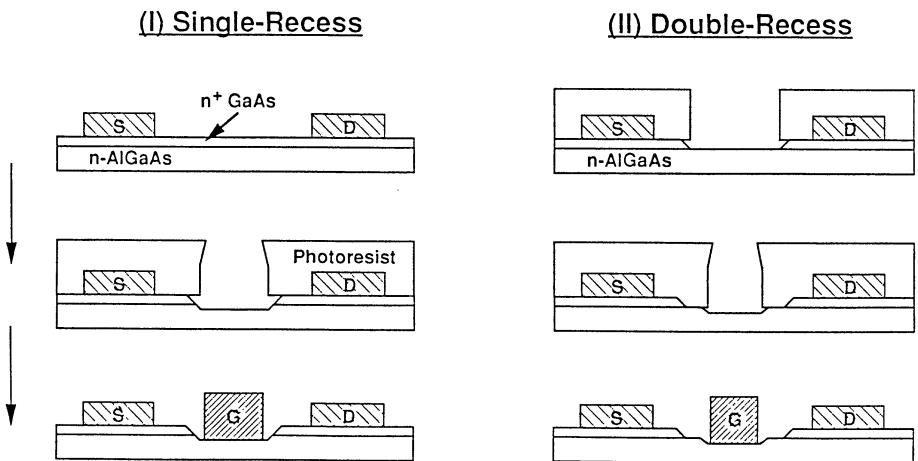


Figure 4-3 Etched channel (recessed) gate technology. The single-recess technology provides low source resistance, high g_m and is best for low-noise and MMW power applications.

- Gate recess
 - Uniform recess depth
 - Controlled recess undercut
 - Smooth etch surface
- Gate formation
 - Short gate-to-source spacing
 - Short and uniform gate length
 - Low parasitic gate resistance

Table 4-1 Requirements of gate process steps. Gate recess is the most critical process step in fabricating PHEMTs.

5. High Yield Gate Recess Process

In the process of a PHEMT, the n⁺-GaAs cap layer needs to be etched off to allow a Schottky to be made on the n-AlGaAs gate layer for a high breakdown voltage. The etchant typically consists of an acid and an oxidizing agent such as H₂O₂. It attacks GaAs because the oxidizing agent it contains removes electrons from the surface of the GaAs and thereby destroys the bonds which link the atoms of the solid. The gate recess profile affects device source resistance, feedback capacitance, output conductance, and breakdown voltage. For PHEMTs, the gate recess is shallow (typically 300-500Å) and the recess profile is less critical when compared to MESFETs.

It should be noted that a deeper gate recess is needed to maintain a high gate aspect ratio [2]:

$$\text{Gate Aspect Ratio} = \text{Gate Length} / \text{Gate-to-2DEG Spacing} > 3 \quad (5.1)$$

A high aspect ratio reduces short channel effect and ensures a high device voltage gain factor and a reasonable value of stable gain at very high frequencies. For PHEMTs, a deeper gate recess is needed to maintain a high aspect-ratio for a very short gate. For a given pinchoff voltage, a minimum value of gate-to-2DEG spacing can only be obtained by employing a highest possible doping concentration in the AlGaAs donor layer. The high doping level in the AlGaAs gate layer, however, reduces the device breakdown voltage. To solve this problem, a "planar-doped" or "pulsed-doped" layer structure can be adopted.

After the recess, the wafer is then metallized and the liftoff process is performed to form the metal gates. Aluminum or gold with a barrier layer (i.e., metal layer which functions as an inter-diffusion barrier and also promotes the adhesion between gate metal and semiconductor channel) such as Ti, Pt, TiW, Mo, or Ni are commonly used for the gate metallization in PHEMTs. Since the gates are very small, a scanning electron microscope (SEM) is typically used before and after gate metallization to measure the gate dimension and inspect for any defects produced during the gate formation process.

6. Formation of Low-Resistance T-Gate

In addition to the short gate-to-source spacing, and the short and uniform gate length, a small gate resistance is essential with PHEMTs for high gain, low noise and high power applications. One approach for obtaining low gate resistance is the adoption of a multi-finger air-bridged gate structure. The air-bridged gate structure, however, increases the number of processing steps and also the parasitic capacitance of the device. Another approach is the use of a T-shaped or mushroom-shaped gate structure. In this

As mentioned above, since the PHEMT cap and donor layers are very thin and heavily-doped, the PHEMT gate recess depth is significantly more difficult to control than with the GaAs MESFET. There are two ways

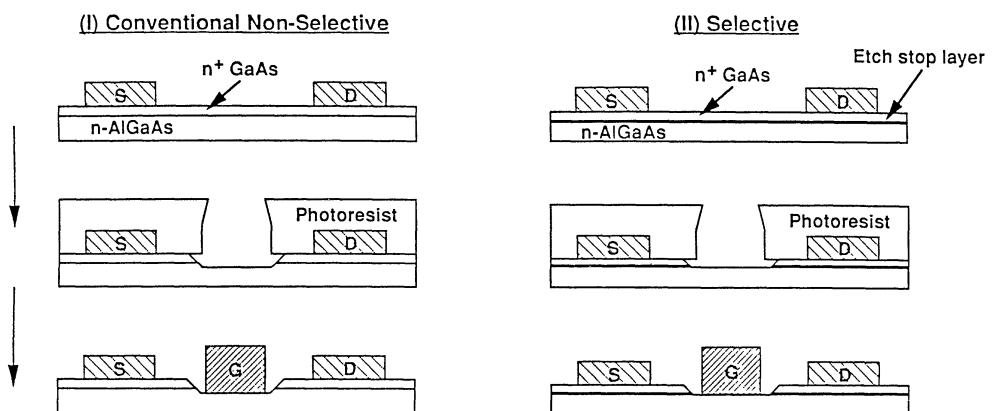


Figure 5-1 Selective gate recess stops at the GaAs/AlGaAs interface, providing excellent uniformity and yield. The AlGaAs gate layer thickness needs to be properly designed to obtain optimum channel current after recess.

to achieve better recess uniformity : a) to perform shallower gate recess, and b) to adopt a selective gate recess. As illustrated in Figure 5-1, the selective gate recess etches off only the GaAs cap layer and stops at the GaAs/AlGaAs interface, providing the uniformity of the device characteristics over the wafer. The selective gate recess, using either a wet-chemical etch or a reactive ion etching (RIE) technique, has been adopted to achieve good current or threshold voltage uniformity in PHEMTs. Compared to the non-selective one, 300-400% improvement in uniformity of device current and transconductance can be achieved using the selective gate recess technology. Furthermore, the device and MMIC yield can also increase by a factor of 2-3.

structure, the small footprint defines the length and the wide top provides a low resistance.

T-shaped gates have been fabricated using angle evaporation with optical lithography or using the double or tri-layer resist technique with E-beam lithography [3]. Figure 6-1 illustrates examples of T-gate fabrication techniques using E-beam lithography. As shown in Figure 6-2, the $0.25\mu\text{m}$ T-gates fabricated using this technique have demonstrated excellent

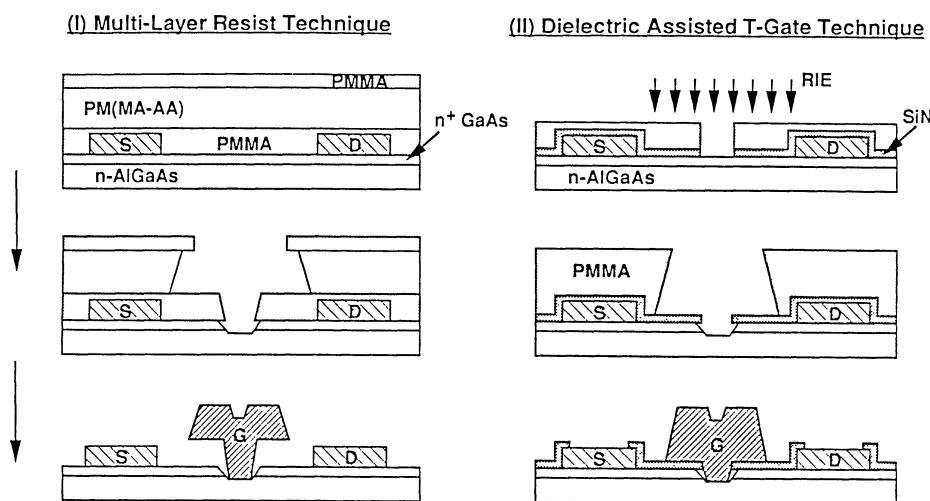


Figure 6-1 Examples of T-gate fabrication using E-beam lithography and multi-layer resist technique.

Multiple-layer resist technique is most popular in the US industry in producing very high performance T-gate PHEMTs.

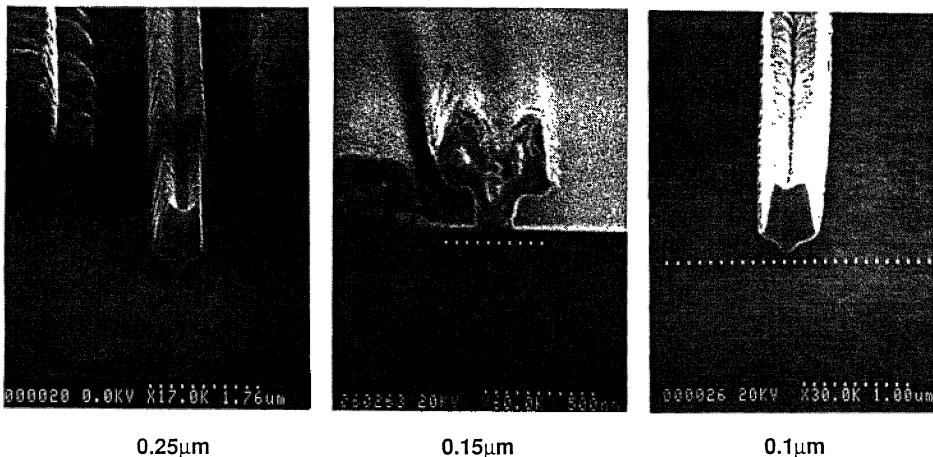


Figure 6-2 T-gates for microwave and MMW PHEMTs.

mechanical stability and also exhibit an extremely low resistance of $50\Omega/\text{mm}$, DC end-to-end. This resistance translates to only 0.4Ω in a $100\mu\text{m}$ wide PHEMT (with two of $50\mu\text{m}$ gate fingers in parallel) and is negligible when compared to the device source resistance (typically $\sim 5\Omega$). For $0.15\mu\text{m}$ long gates, however, the T-gate resistance rapidly increases to $200\Omega/\text{mm}$ and becomes significant compared to the source resistance. Therefore, with an extremely short gate length PHEMT, a tradeoff between the gate resistance (even with a T-gate structure) and the gate length needs to be made. An excellent overview of the gate formation process for FETs can be found in [4].

7. Gate Related Reliability Issues

The reliability of the PHEMT is greatly affected by the gate technology used in fabricating the device. Commonly observed failure modes associated with gate technology include :

- a) Gate sinking : gate metal and AlGaAs gate layer inter-diffusion under thermal or electrical stress, resulting in a reduced AlGaAs thickness. The deeper recessed device (i.e., more enhanced mode) is more sensitive to the gate sinking effect.
- b) Electro-migration of gate metal : between gate and drain electrodes where the electric field is the strongest.
- c) Formation of non-conductive groove next to gate metal in a high moisture environment - an electrochemical effect. The mechanism accelerates when the gate is reversed biased.
- d) Electrostatic discharge (ESD) : significant charge build-up can occur due to improper grounding. The charge is sufficient to damage the very short gate of the device during the chip or wafer handling.
- e) Degradation of characteristics under hydrogen : the device characteristics can change in a hydrogen environment.

The following describes the PHEMT hydrogen effect : To ensure that the PHEMT is reliable, Ti/Pt/Au gate is commonly used to provide thermal and electrical stability of the device. It has been found, however, that in hermetically sealed Kovar or Invar packages, the DC characteristics of PHEMTs could change within several hundred hours even with a baseplate temperature as low as 150°C. These devices can also change during the chip eutectic die attach process at 300°C under H₂ forming gas. The DC change has been explained primarily as the result of the conversion of gaseous hydrogen (released from the package) into atomic hydrogen within the Pt metallization of the gate. These hydrogen atoms subsequently diffuse into the channel region under the gate, change the Schottky barrier height and/or neutralize the Si donors, resulting in change in drain current and pinchoff voltage. The degree of current change under hydrogen is a strong function of the Pt thickness and is relatively insensitive to the Ti metal thickness of the gate. The device current becomes significantly less sensitive to hydrogen with a Pt metal

thickness of less than 50Å in the gate [5]. The search of hydrogen insensitive gate metallization is an on going activity to solve this reliability issue.

8. Summary

This note is summarized as follows:

- a) The layout and fabrication of the gate is most critical to the electrical performance of low-noise and power PHEMTs.
- b) Low resistance submicron T-gate is today's standard in fabricating high performance PHEMTs and MMICs.
- c) High gate aspect-ratio design results in less short channel effect with PHEMTs.
- d) Selective gate recess is most effective in improving PHEMT device uniformity, reproducibility and yield.
- e) Gate has a significant effect on PHEMT reliability - metal shape, position, composition, and thickness should be carefully designed for long device life time.

9. References

- 1 Chao, P.C., et. al. (1991) HEMT devices and circuit applications, in F. Ali and A. Gupta (eds), *HEMTs and HBTs*, Artech House, p. 77.
- 2 Das, M.B. (1985) A high aspect ratio design approach to MMW HEMTs, *IEEE Trans. Electron Devices*, ED-32, p. 11.
- 3 Chao, P.C., et. al. (1985) Electron-beam fabrication of GaAs low-noise MESFETs using a new trilayer resist technique, *IEEE Trans. Electron Devices*, ED-32, p. 1042.
- 4 Weitzel, C., et. al. (1986) A review of GaAs MESFET gate electrode fabrication technologies, *J. Electrochem. Soc.*, 133, p. 409C.
- 5 Chao, P.C., et. al. (1994) HEMT degradation in hydrogen gas, *IEEE Electron Device Lett.*, 15, p. 151.

VIAS, BACKSIDE THINNING AND PASSIVE ELEMENTS

J A TURNER

GEC-Marconi Materials Technology

Caswell

Towcester

Northants. NN12 8EQ

Contents

1. Introduction
2. Passive Components
 - 2.1. Capacitors
 - 2.2. Inductors
 - 2.3. Resistors
 - 2.2.1. GaAs Resistors
 - 2.2.2. Thin Film Resistors
 - 2.4. Airbridges And Dielectrically Isolated Interconnections
 - 2.5. The Integration Process
3. Backside Processing
 - 3.1. Wafer Mounting
 - 3.2. Wafer Thinning
 - 3.3. Via Hole Etching And Metallisation
 - 3.4. Bath Tub Etching And Metallisation
 - 3.5. Die Separation
4. IC Fabrication Processes
 - 4.1. Lithography
 - 4.2. Deposition Process
 - 4.2.1. Metal Deposition Techniques
 - 4.2.2. Plasma Assisted Chemical Vapour Deposition (PECVD)
 - 4.3. Etching Processes
 - 4.3.1. Ion Beam Milling
 - 4.3.2. Reactive Ion Etching (RIE) and Plasma Etching
 5. Conclusions
 6. References

1. Introduction

It is not only the gate definition process that determines the microwave and millimetre wave performance of an MMIC; the fabrication of passive elements and the processing of the backside of the wafer also have a strong effect on the capability of the circuit to perform to its highest potential.

This chapter is devoted to examining the fabrication techniques of these passive elements and also the way in which the backside processing is carried out.

The development of process technologies for high speed MMICs in compound semiconductor materials presents a number of wide ranging challenges for the process engineer. For analogue applications the diversity of frequencies from about 1GHz to 100GHz over which the PHEMT can operate means that a standard technology cannot be adopted for the whole frequency spectrum as the circuit topology required at each end of this frequency range differ widely. At low frequencies, approximately 1GHz, where the noise figure of the PHEMT holds advantage over the more conventional devices traditional distributed (microstrip) style matching elements consume too much surface area of GaAs and so RF techniques using lumped inductors, capacitors and resistors are used. It is only at frequencies in excess of 6 - 7GHz that distributed circuits are small enough to enable utilisation of this type of impedance matching.

The process engineer must first develop technologies for manufacturing all the necessary components and then devise ways of integrating them into a reproducible, reliable IC process consuming as little GaAs surface area as possible. Developing processes that minimise the fabrication temperature is key to a successful GaAs MMIC process.

The circuit manufacturing procedures can be described by the four basic processing technologies that follow.

- (a) Lithography
- (b) Etching of GaAs
- (c) Metal deposition and removal
- (d) Dielectric deposition and removal

2. Passive Components

2.1. CAPACITORS

Capacitors are used for tuning elements, for interstage isolation and for bypass. Two types are currently in use:-

- (a) Interdigitated for low capacitance (less than 1pF) [1]
- (b) Parallel plate, metal insulator metal for 1 - 20pF [2]

The interdigitated capacitor shown in Figure 1 utilises the capacitive coupling between adjacent fingers.

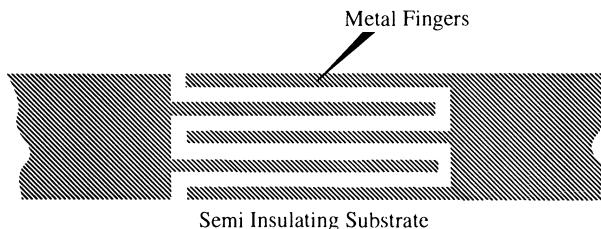


Figure 1 Interdigitated capacitor structure

To fabricate these elements, sputtered films of a gold based alloy are deposited on to the semi-insulating substrate at a rate of a few microns per hour. The most common capacitor is the metal-insulator-metal capacitor (MIM) which is formed by two metallised areas separated by a dielectric material. Figure 2 shows this type of capacitor. The dielectric used is chosen to give a reasonable capacitance/unit area, is easy to deposit, and stable in use.

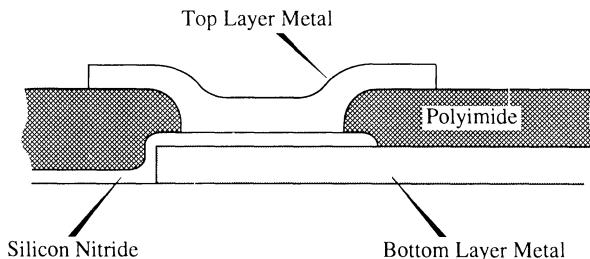


Figure 2 Metal-insulator-metal capacitor structure

Table 1 gives three dielectrics in common use, the most common being Si_3N_4 .

TABLE 1. Capacitor Dielectrics

Dielectric Material	Dielectric Constant	Typical C/unit area	Q Factor	Deposition Technique
Silicon Nitride	6.5	480	Good	Plasma assisted CVD
Silicon Dioxide	4.0	340	Good to very good	Plasma assisted CVD
Polyimide	3 - 4.5	30	Good to very good	Spun and cured film

The dielectric thicknesses used have a major outcome on the final yield of the circuit. In analogue circuits there are generally many more capacitors than MESFETs and the capacitor yields can determine the yield of the circuit. Generally dielectric thicknesses of $1000\text{-}2000\text{\AA}$ are chosen in order to minimise the number of pinholes in the film that cause DC short circuits while maintaining a reasonably high capacitance per unit area.

2.2. INDUCTORS

Inductors are used particularly in lower frequency analogue circuits as biassing and tuning elements in a MMIC. [2,3]. Figure 3 shows schematically four common configurations of inductor. They pose two major fabrication challenges - the definition of the metallisation and the connection between the centre of the spiral and the rest of the circuit.

To maintain a high circuit quality factor Q at low microwave frequencies the metal thickness must be of the order of 3 microns thick. The separation of the spirals is typically 12 microns.

An alternative technology for achieving a thick metallisation is to plate up the

spiral using a similar process to that described for the air bridge process in Section 2.4.

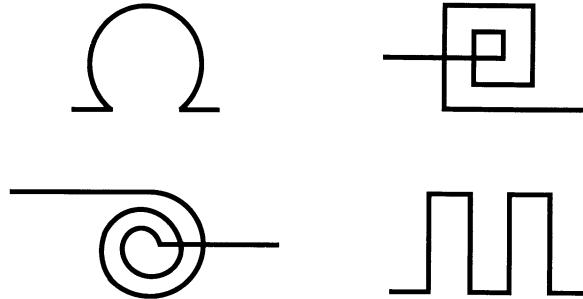


Figure 3 Four inductor configurations

One disadvantage of this process is that the edge definition of the plated spiral is somewhat worse than the milled metallisation, however it is easier to build up the metal thickness to values considerably greater than 3 microns which can be an advantage in some lower frequency circuits.

The second challenge, that of connecting the centre of the spiral to the rest of the circuit, can be met either by the use of the airbridge technology or by a metal underpass connection.

The airbridge technology offers a very good microwave solution with a low parasitic capacitance but is prone to mechanical deformation. The span is necessarily large and any downward pressure on the bridge will cause it to be distorted and to short to the spiral. The underpass connection which utilises the multilevel dielectric layers of the process is more robust but has a higher parasitic capacitance. Figure 4 shows a cross section through a completed inductor with an underpass connection.

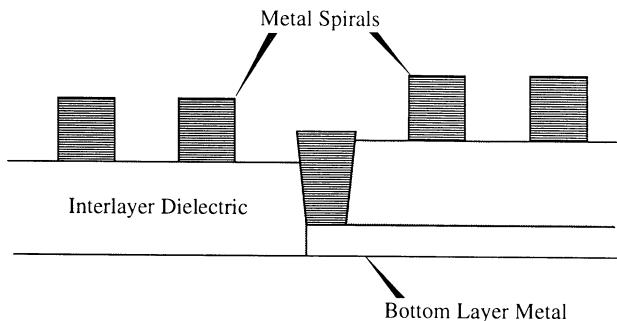


Figure 4 An inductor with underpass connection

2.3. RESISTORS

Resistors are used for determining the bias applied to the transistors, for feedback networks and for terminations in power combiners and Lange couplers (5). Two families of resistor are commonly used, those fabricated in the GaAs active layer and those produced in deposited resistive films.

2.2.1. *GaAs Resistors*

The simplest way to produce a resistor is to use the GaAs material in which the active device is fabricated. In this approach the resistor is fabricated in a similar manner to the FET itself apart from these fact that the gate electrode is omitted. Two types of GaAs resistor are shown in Figure 5. Figure 5 (a) shows a mesa resistor where isolation is achieved by etching. In Figure 5 (b) isolation is achieved by selectively implanting resistive areas in the semi-insulating GaAs substrate.

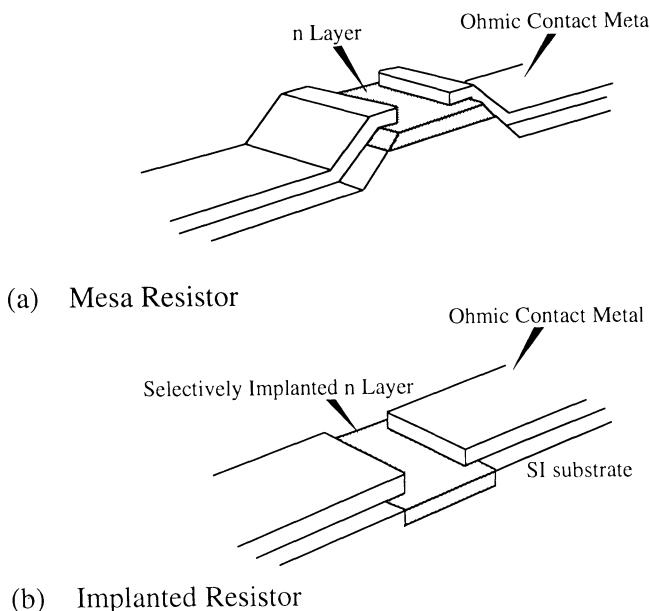


Figure 5 Two types of GaAs resistor

The use of GaAs resistors is not without its problems as an ungated resistor can act, under certain bias conditions, as a planar transverse electron oscillator (Gunn oscillator). It is therefore important to ensure that the bias voltage across the terminals of the resistor is below the threshold voltage for the onset of Gunn oscillations.

2.2.2. Thin Film Resistors

In Table 2 is listed a number of metallisations that are being used for resistors. Generally, whichever material from Table 2 is chosen, additional process steps are required in order to incorporate it into a MMIC process.

TABLE 2. Resistor Materials

Material	Resistivity ($\mu\Omega\text{-cm}$)	Temperature Coefficient of Resistance (ppm)
TaN	280	-180 to -300
NiCr	60 - 600	200
GaAs	300- 450	300

2.4. AIRBRIDGES AND DIELECTRICALLY ISOLATED INTERCONNECTIONS

In many sections of a MMIC a connection is required between-adjacent metallised areas. This can be accomplished using either airbridges or dielectric interconnections.

An airbridge is a bridge of metal running above the surface of the circuit between metal areas to be interconnected, they are formed by a photoresist and plating process outlined in Figure 6. By careful control of the resist thicknesses and the plating conditions, a very strong well defined bridge can be formed.

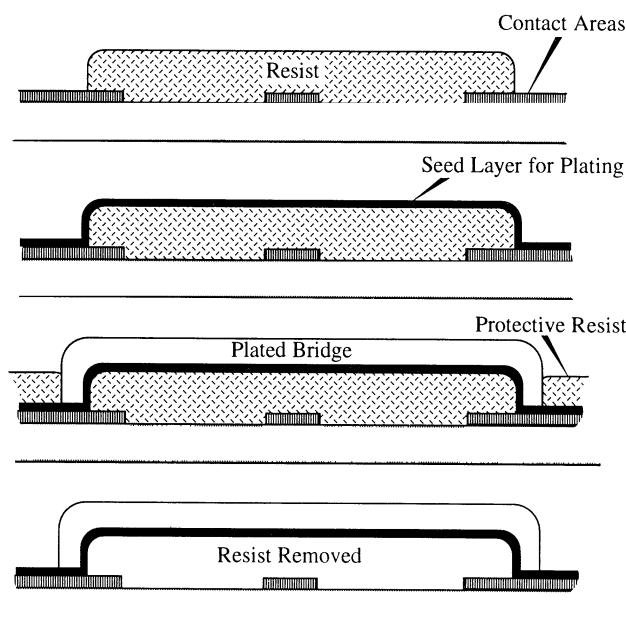


Figure 6 Airbridge fabrication process steps

Some manufacturers rely solely on this form of interconnection using it for passive component integration as well as for interconnecting the MESFET and in digital as well as analogue circuits. In digital circuits many hundreds of such interconnections are made and the process has therefore to be very reliable.

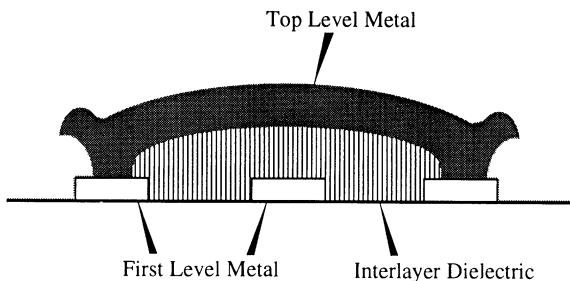


Figure 7 Underpass connection technology

An alternative interconnection process is to make use of the various levels of dielectric films in a multilevel MMIC process. This is shown in Figure 7. One advantage of this approach is the fact that it is essentially a planar process and overcomes the possibility that airbridges can be damaged during die separation and mounting. Three dielectric materials for this purpose are in common use, silicon dioxide, silicon nitride and polyimide. For fabrication simplicity the interlayer dielectric can also be the film that forms the capacitor dielectric.

In some multilevel schemes polyimide is a preferred dielectric material. As it is applied as a relatively thick liquid film (1 micron) it has the ability to 'planarise' the surface of the MMIC making subsequent metallisation over sharp underlying irregularities a relatively easy task.

2.5 THE INTEGRATION PROCESS

Defining the individual components that are described in the preceding sections of this chapter is not sufficient for the realisation of a full integrated circuit process. Its development is an iterative exercise where the requirements of the technologies to produce the components are combined with those needed to interconnect them. Generally many compromises have to be made before a manufacturable process can be finally realised involving not only aspects of the technology, but also satisfying the requirements of circuit design, assembly and packaging. The interaction necessary between circuit fabricators, circuit designers and users cannot be over stressed and all successful processes are the result of constant consultation between all interested parties.

One example of a GaAs integrated circuit process that is equally applicable to analogue and digital ICs is shown schematically in Figure 8. In this figure it is seen how the metal and dielectric layers of the circuit are built up.

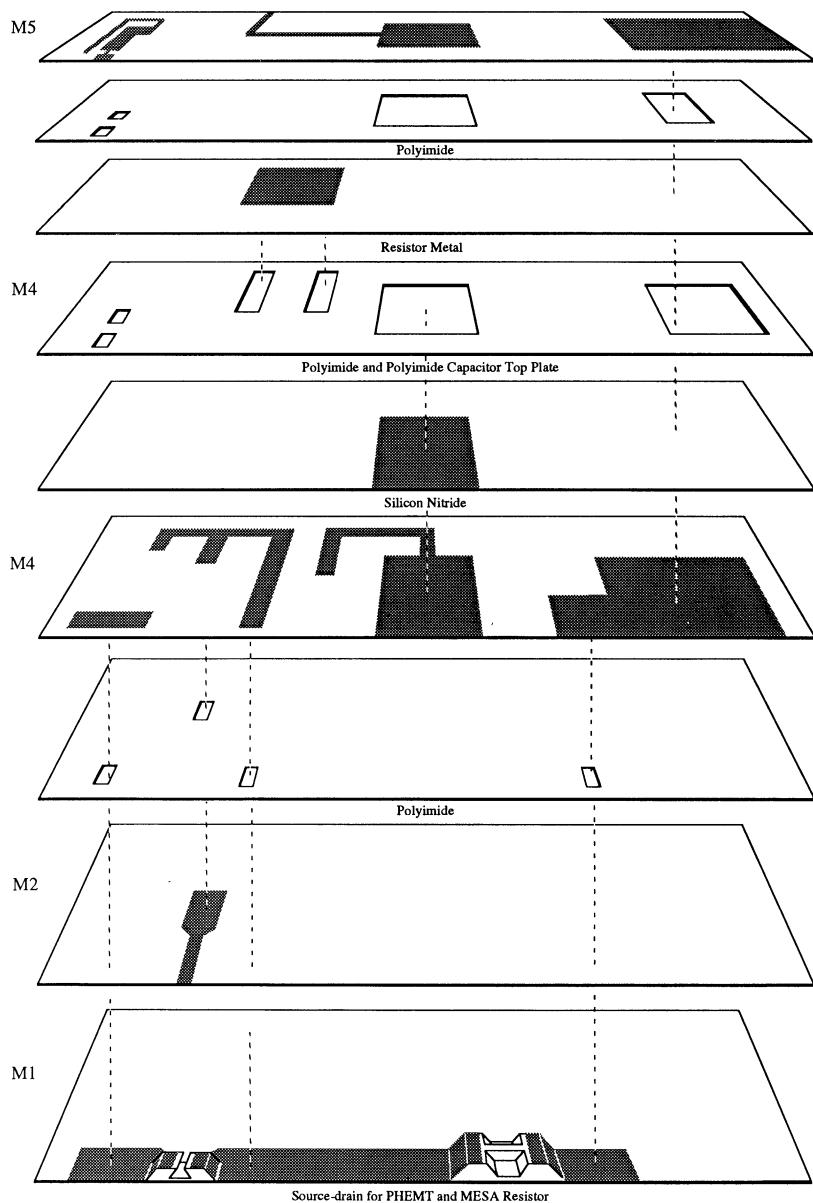


Figure 8 GaAs MMIC process schematic

3. Backside Processing

When the front side processing is complete the wafer still has to undergo many more

process steps before fabrication is complete. These are perhaps the most crucial steps, as any problems at this stage which results in the wafer being damaged means a loss of many hours front face processing time. There can be as many as five different stages to be performed on the back of the wafer before processing is complete.

- (a) Wafer mounting
- (b) Wafer thinning
- (c) Via hole etching and metallisation
- (d) Bath tub etching and metallisation
- (e) Die separation

3.1. WAFER MOUNTING

After completion of the front face processing the wafers are still at their original thickness of around 600 microns thick. The wafer will end up considerably thinner than this when backside processing is complete and must therefore be mounted on to a carrier to support it and prevent breakages. This is carried out by mounting the wafer on to a supporting plate. This plate must be very flat and must be transparent to infra red radiation. Many of the patterning techniques on the back of the wafer require alignment to features on the front face. Infra red lithography equipment is used which can 'see through' the plate and the GaAs wafer to facilitate 'back to front' alignment where necessary.

The wafer is mounted on to the plate using a high temperature fixant and during this process care must be taken to ensure the back of the wafer is parallel to the surface of the plate for reasons that become obvious in the next process step.

3.2. WAFER THINNING

In this process stage the thickness of the wafer is reduced to between 100 to 200 microns depending on the purpose for which the front face circuitry is to be used. This process must be carried out carefully and the faces of the wafers must be parallel and to within plus or minus 5 microns of their intended thickness. This is to ensure that the impedances of the transmission lines on the front face meet their design goals. Any wide deviations from the design thickness will cause impedance variations of these lines and could seriously degrade the performance of the circuits being produced.

A reduction in thickness of the wafer considerably assists in removing heat from the thermally excited areas on the front side of the wafer.

The thinning process is normally carried out using a fine abrasive powder in a slurry in a commercial lapping machine. In some cases the final 20 to 30 microns is removed using chemo-mechanical means which leaves a degree of polish on the back of the wafer.

3.3. VIA HOLE ETCHING AND METALLISATION

As the frequency of operation of the circuits increase the importance of a low grounding inductance increases. At low frequencies it is not unusual to see single or multiple bond wires connecting the grounding pads of the circuit to the package or fixture ground. At millimetre wave frequencies such a technique would give intolerably high grounding inductance. This problem has been overcome by the use of

via hole etching. In this process a hole is formed from the back of the wafer through to a metal grounding pad on the front. When filled with metal this forms a low inductance connection and, when the chip is mounted on to a conducting substrate, forms an easy contact to ground. [4]

The process requires four steps and these are shown in Figure 9. First using infra red alignment techniques the position of the vias is made to coincide with the pads to be grounded on the front face - windows in photo resist are opened up to correspond to these positions, Figure 9(a). Using the resist as a mask GaAs is removed by an etching process, RIE or wet etching techniques are commonly used, down to the back of the metal on the front face Figure 9(b). Once all the metal areas are exposed etching is terminated. The photo resist is removed and a thin metal coating is deposited over the back surface of the wafer and into the etched holes. This forms a seed layer for the final plating stage where gold is electro deposited to build up to a final thickness of 5 microns, Figure 9(d).

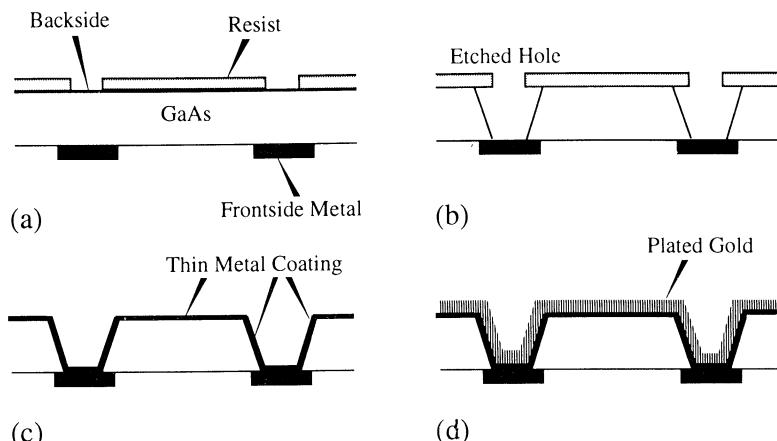


Figure 9 Through GaAs via hole process

Not only has this technique greatly improved the rf performance of the transistor and hence the MMIC it has also produced a revolution in the way MMICs are designed. Before the advent of this via technology bond wires were used to ground areas on the circuit. This necessarily meant that all grounding pads had to be on the periphery of the chip. This considerably restricted the layout of the chip and in many instances made size reduction impossible. The advent of vias has removed this layout restriction and areas deep within the circuit design can be grounded by the use of a via. This has led to more compact designs and has enabled the circuit designer to integrate many more functions together on chips of reasonable size.

3.4. BATH TUB ETCHING AND METALLISATION

Many applications of PHEMT technology can be satisfied by using the device as a

transmitter. In such areas heat sinking of the device is an important factor for efficiency and reliability considerations. It is standard practice to thin the whole of the wafer to approximately 100 microns which is a compromise between acceptable thermal and rf design. A better compromise can be reached if the wafer is selectively thinned underneath the device while maintaining a general wafer thickness of 100 microns or so. This is the so called bath tub process. In this technique the wafer is selectively thinned underneath the power transistor such that only 30 - 50 microns of GaAs remains. The metallisation on the back of the wafer therefore comes to within this distance of the transistor and its thermal characteristics are considerably improved. Figure 10 shows a cross section of the device with through GaAs vias and a bath tub.

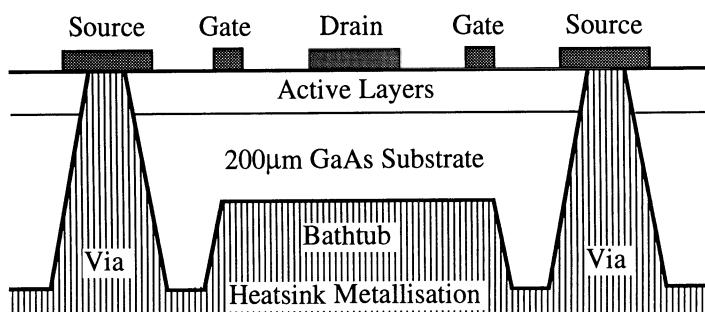


Figure 10 Bath tub heat sinking schematic

A 2 watt power transistor on a 200 micron thick substrate held at an ambient temperature of 70°C has a junction temperature of 230°C. A similar transistor on 100 micron thick substrate with a bath tub heat sink has a junction temperature of 160°C.

This technique of selectively heat sinking the transistors in the MMIC also has the benefit of reducing the wafer breakages as wafers of sub 100 micron thicknesses tend to be very prone to break during processing.

3.5 DIE SEPARATION

The final stage in the preparation of the single MMIC chips is to separate out the die from the wafer. This can be accomplished in two ways, either by sawing with a diamond tipped saw blade or by scribing with a diamond tipped tool and then breaking into individual chips with a roller. The technique chosen tends to depend on chip size and wafer thickness. Whichever technique is used saw or scribing 'streets' have to be defined in the metal on the back of the wafer - it is extremely difficult if not impossible to separate chips by either sawing or scribing if the metallisation on the back of the wafer is continuous. It is necessary either at the via hole or the bath tub metallisation stage to leave channels devoid of metal where the chips will separate. This is carried out by a selective plating process during the stage when the thick back metal is deposited.

4. IC Fabrication Processes

In this section we describe the process techniques that are used to produce the structures described in the preceding sections of this chapter.

Once the active n layer has been formed, the whole integrated circuit can be processed using three basic fabrication techniques. These are;

- a) Lithography
- b) Deposition - metal and dielectric
- c) Etching - metal, dielectric and semiconductor.

Various sequences of these three techniques are needed to allow complete circuits to be fabricated. [5]

4.1. LITHOGRAPHY

Lithography forms a vital part of any process as all process stages are preceded by a lithography step that defines the pattern for etching or for metal deposition. Basically the lithography process involves defining a pattern in a thin photo sensitive film (photoresist) by a process of optical exposure through a photographic mask and subsequent differential removal by immersion in a resist developer solution. Figure 11

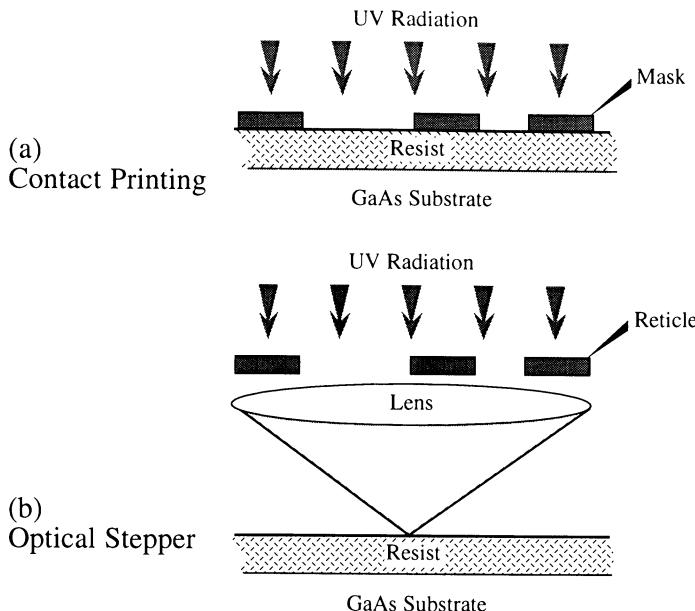


Figure 11 Contact and stepper lithography principles

shows two different methods of photolithography; by direct contact and by optical stepper.

In the direct contact method, a mask containing the pattern to be defined is placed in hard contact with the photosensitive resist film that has previously been spun on to the GaAs wafer as a uniform thickness film.

The resist not protected by the dense areas in the mask is irradiated with ultra violet light. The irradiated film of resist is placed in a developer solution and, if the resist is a positive' one, those resist areas exposed by the radiation will be removed to expose the substrate surface. If the resist is a negative one then those resist areas not exposed will be removed in the developer. The resist film left after development can then be used as an etch resistance film or as a film for the metal lift off process.

The optical stepper process is similar to the in-contact one in terms of exposure and development of the resist except that the mask is held some way away from the resist coated sample and is exposed by projecting an image of the reticle (mask) on to the surface of the GaAs slice. The feature sizes on the masks are either 5x or 10x the final required size depending on the type of stepper used.

Each process has advantages and disadvantages over the other and these are listed in Table 3 together with those for electron beam lithography for comparison purposes.

TABLE 3 Lithography Processes.

Lithography Method	Advantages	Disadvantages
Contact	Low cost equipment Fast throughput	Mask procurement difficult Mask wear
Stepper	No mask wear	Carefully controlled process conditions
Electron Beam	Mask procurement easy Resolution very good No masks Very high yield	Limited resolution High machine cost Throughput time limited

4.2. DEPOSITION PROCESS

The fabrication of a MMIC requires the deposition of both metal and dielectric films and this is achieved using some of the different techniques explained in the following sections of this chapter.

4.2.1. *Metal Deposition Techniques*

Metal films are applied to GaAs wafers either by evaporation or sputtering dependent on the required final metal thickness. In the evaporation process the metal is heated to a temperature sufficient to cause vaporisation. This vaporised metal then deposits on to the GaAs slice situated inside the vacuum chamber held at a pressure of around 10^{-7} torr. Vaporisation of the metal is achieved by either heating the metal in an electrically heated coil or by directing an electron beam on to the charge of the metal to be deposited. This method of metal deposition is a somewhat directive one and is not

recommended for metallising over steep edges but is good for lift off.

The deposition of thick films (>1 micron) and of alloys is best carried out by sputtering. In this process high energy ions are directed on to a metal target knocking metal atoms from the target which redeposit on to the GaAs slice. This is a low temperature process and relatively non directive and hence good for step coverage. Step coverage can be further enhanced by biasing the substrate. Sputtering is particularly good for alloy deposition for the nature of the process causes the deposited film to retain the composition of the target.

4.2.2. *Plasma Assisted Chemical Vapour Deposition (PECVD)*

None of the above processes can effectively be used for depositing dielectric films. The standard process for dielectric deposition is chemical vapour deposition in which the relevant gases are reacted to form the nitride or oxide which deposits onto the substrate at a high temperature. Such processes cannot however be used in GaAs technologies as the temperatures required will adversely affect prior process steps (eg ohmic contacts) and may cause the GaAs to decompose. Consequently PECVD processing is used. Here the temperature of the reaction required to form the dielectric film is lowered by the formation of a plasma in which the energetic electrons increase the reaction temperature allowing substrate temperatures below 300°C to be utilised. Silicon nitride is by far the most popular dielectric film used in GaAs processing and this is formed by reacting silane and ammonia in a closed system under a plasma. [6]

4.3. ETCHING PROCESSES

Once the metal and dielectric films have been deposited they need to be selectively removed to form the various patterns that build up the integrated circuit. This is carried out by a combination of lithography and etching. There are many etching processes available to the GaAs fabrication technologist. The major ones are listed below and are described briefly below.

- a) Ion beam milling [7]
- b) Reactive ion etching and plasma etching [8]
- c) Wet chemical etching

Of these processes only (c) is not a 'dry' process and the move away from wet chemical etching is becoming an important aspect of GaAs technology. In many instances the dry process technology is being adopted because of its controllability and cost effectiveness. In others it is proving to be the only practical method of removing many metals and dielectrics.

4.3.1. *Ion Beam Milling*

In this process an energetic ion beam is directed onto the material to be etched and metal removal is purely by the impact of these ions on to its surface. Inert gases such as argon are used and no chemical reaction takes place. There is provision within the milling equipment to rotate and tilt the substrate material and this is important as experience has shown that etch rate is a function of the angle that the substrate presents to the ion beam. Rotation of the substrate is intended to unify the etch rate across the wafer. Ion beam milling is a rather slow process with etch rates around 1 - 2 microns per hour.

4.3.2. Reactive Ion Etching (RIE) and Plasma Etching

These processes are similar to ion beam milling except that chemical compounds are introduced into the equipment to improve the etch rate. It uses relatively high energy directive ions giving only a small amount of undercutting of the masking medium. It therefore has advantages over wet etch processes and is used for instance for the etching of via holes in GaAs slice where a minimum amount of undercutting is highly desirable. For this purpose chlorine based compounds such as carbon tetrachloride are introduced into the RIE equipment. RIE is also being utilised more and more on multi-level resist processes because of its highly anisotropic etching behaviour.

The plasma etching process is very similar to reactive ion etching and they are distinguished from each other only by the operational pressures and voltages at which the equipments are operated. RIE operates at lower pressures ($\sim 10^{-2}$ torr) than plasma etching and is therefore somewhat more directional.

Figure 12 summarises these techniques with regard to the directionality of the etching process.

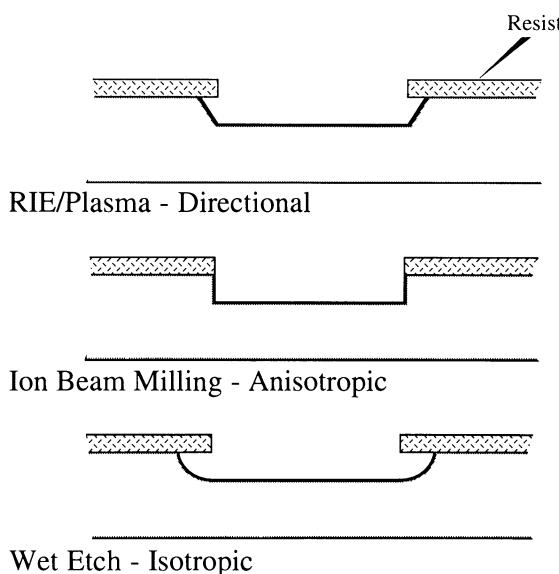


Figure 12 Etching profiles formed when using the indicated techniques

It is important to note that although wet processes are rapidly being superseded by dry ones they are not without their problems. A major concern of the GaAs technologist is the damage caused by the dry process techniques which generally are high energy processes. It appears to be relatively easy to cause damage to the surface layers of crystalline GaAs which can cause serious degradation in electrical properties of the material and hence cause changes in the circuit performance.

5. Conclusions

This section has dealt with perhaps the least glamorous aspects of the PHEMT MMIC process. However these processes and components are vital to the successful operation of the MMIC. An attempt has been made to cover as many as possible of the techniques and technologies used in today's processing lines,. These may of course change as our understanding of the technology improves and our need to move to higher frequencies and higher powers increases. As wafer diameters increase the backside processes described in section will become more difficult and an intense effort will be required to maintain wafer breakages at a low level. The move towards Indium Phosphide based PHEMTs will pose even greater problems because of the extreme brittleness of the material.

6. References

1. Esfandiari, R., Maki, D.W., Siracusa, M. (1983) "Design of interdigitated capacitors and their application to GaAs monolithic filters" IEEE Trans. Microwave Theory Tech., 31, 57.
2. Pengelly, R.S., (1982) "Microwave field effect transistors - theory, design and applications", New York: Wiley.
3. Pucel, R.A., (1981) "Design considerations for monolithic microwave circuits: IEEE Trans. Microwave Theory Tech., 29, 513.
4. D'Asaro, L.A., DiLorenzo, J.V., Fukui, H. (1978), "Improved performance of GaAs microwave field effect transistors with low inductance via-connections through the substrate" IEEE Trans. Electron Devices, 25, 1218.
5. Williams, R.E., (1984) "Gallium arsenide processing techniques" Dedham, M.A., Artech House Inc.
6. Hollahan, J.R., Rosler, R.S., (1978) "Ion-beam techniques for device fabrication" Thin Film Processes, New York: Academic Press.
7. Spencer, E.G., Schmidt, P.H., (1971) J. Vac. Sci. Technol. 8, S52.
8. Coburn, J.W., (1982) Plasma Chemistry and Plasma Proc. 2, 1.

FIELD-EFFECT TRANSISTOR MODELS AND MICROWAVE CAD

R.J. Trew

Case Western Reserve University

Mathematical models for electronic devices find wide use in the design and development of electronic systems. Since a system is composed of a large number of individual components it is necessary to have a wide range of device and component models available. A variety of systems level simulators have become available in recent years, and these simulators permit advanced design and design optimization to be performed.

Models for solid state electronic devices are the basic building blocks for these simulators, and without good models effective systems level simulators are not possible. The device models must be quantitatively accurate, yet sufficiently fast computationally to be useful in a systems level simulator where each device model may need to be solved many times in a given circuit level simulation. The computational speed consideration becomes more severe as the level of integration increases and more components need to be included.

Solid state device models can be formulated at various levels of sophistication. For example, various types of solid state device models are listed in Fig. 1, along with a chart indicating the use of the model in advanced systems level simulators. The listing indicates device models ranging from equivalent circuit type models to sophisticated and comprehensive physics based models that include quantum, non-equilibrium, and other such effects. As model complexity increases, the use of the model in systems simulators rapidly becomes more difficult since the device model itself requires significant computational effort to obtain even simple solutions. Execution time can rapidly become impractical, even for advanced computer workstations. For this reason, virtually all of the commercially available systems simulators that can be used for microwave applications include only the equivalent circuit models. These models can be efficiently implemented in the systems simulators. All of the other models have become known as "physics based" and generally require more computational resources. With only two or three exceptions, physics based solid state device models are not available in commercially available microwave systems simulators. Effort in this direction, however, is rapidly progressing and physics based device models are becoming accepted. In general, they are capable of much greater accuracy and are much easier to use since they don't require the extensive and elaborate calibration procedures required for the equivalent circuit models. Also, the physics based models permit device, as well as circuit design, to be performed and a device need not be fabricated and characterized before investigation. Significant savings result.

- A. Equivalent Circuit
- B. One-Dimensional Analytic (Linear)
- C. Analytic Large-Signal (Generally Pseudo 2-Dimensional)
- D. Two (or Three) - Dimensional Numerical (Drift-Diffusion)
- E. Two (or Three) - Dimensional Numerical (E-M Relaxation or Electron Temperature)
- F. Particle (Monte Carlo)
- G. Self-Consistent Particle (Includes Poisson's Equation)
- H. Hydrodynamic
- I. Quantum (Includes Schroedinger's Equation)

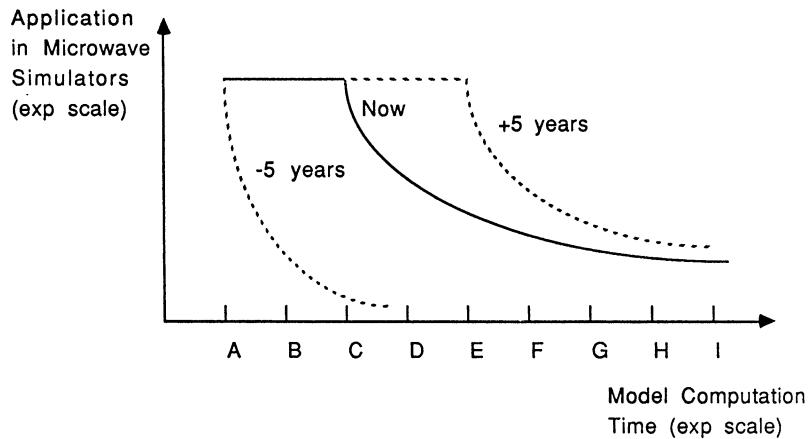


Figure 1 Types of Solid State Device Models and Their Use in Microwave Systems Simulators

Equivalent Circuit MESFET Models

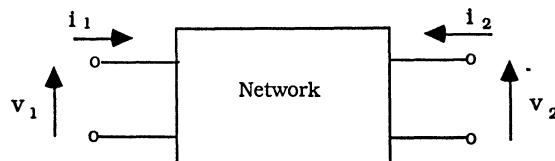
Equivalent circuits have long been used for electronic devices. The procedure is classic and widely used in electrical and electronic engineering. The basic idea is to define a circuit that accurately reproduces the experimental terminal current-voltage characteristics for a device. A topology of any combination of circuit elements that reproduce the terminal i-v characteristics is satisfactory. However, certain topologies are non-physical. For best results the equivalent circuit topology should be selected so that the elements represent the physical operation of the device.

The classic method for developing the equivalent circuit is shown in Fig. 2. Variables are defined as the input and output terminal voltages and currents. A variety of circuits can be defined, depending upon the selection of the independent and dependent variables. For example, if the voltages are selected as the dependent variables and the currents are selected as the independent variables an 'impedance' or Z parameter description for the network or device results. The Z parameters can then be implemented with some combination of circuit elements. Other descriptions for the network can be determined, depending upon the selection of the variables. It should be noted that all of the various descriptions are equivalent, in that they all describe the same device or network. In fact, all descriptions can be directly determined from any description since they all apply to the same network.

The equivalent circuit approach works well for linear, small-signal applications. For example, when the technique is applied to a modern GaAs MESFET (acronym for Metal-Semiconductor Field-Effect Transistor) an equivalent circuit, such as shown in Fig. 3, results. The equivalent circuit shown in Fig. 3a contains elements that accurately reflect the physical operation of the device. For this reason it could be termed a 'physics based equivalent circuit'. In practice, the element values are determined by parameter extraction from measured data. Once determined, the circuit permits circuit level design to be performed. This approach works very well for applications such as small-signal linear amplifiers, low-noise amplifiers, etc. The equivalent circuit only permits linear applications to be addressed since it does not have nonlinear capability.

In order to extend the procedure to large-signal applications, such as power amplifiers, oscillators, switches, etc. it is necessary to modify the circuit. Generally, this is accomplished by reducing the equivalent circuit and retaining only the elements of first-order importance to device operation. The circuit shown in Fig. 3b results. In many applications this circuit can be used in place of the circuit shown in Fig. 3a without loss of accuracy. However, if the circuit elements are defined from measured data it should be recognized that the element values for the two circuits will differ. This can, in fact, cause confusion if the element values are used to make judgments concerning device operation or performance.

To extend the MESFET equivalent circuit to large-signal, non-linear applications it is necessary to introduce appropriate non-linearities into the equivalent circuit. This is generally accomplished by making the current generator a non-linear element, and allowing for forward and reverse conduction of the gate electrode. The resulting large-signal equivalent circuit is shown in Fig. 4. Note that the resulting circuit is essentially the simplified circuit shown in Fig. 3b, with appropriate modifications to permit for large-signal, non-linear operation. The main non-linear element is the current generator and the mathematical expression used to describe its performance. Many non-linear 'models' have been reported, depending upon the expression used for the current generator. A listing of some of the most significant models is presented in Table 1. It should be noted that this



- Select Any Two Variables as Independent Variables and the Remaining Variables as Dependent Variables

- Many Equivalent Circuit Descriptions are Possible

e.g.

- Select Voltages as Dependent Variables and Currents as Independent Variables

Z-parameter (Impedance) Network Description

Well Suited for Tee-Network Topology

- Select Currents as Dependent Variables and Voltages as Independet Variables

Y-Parameter (Admittance) Network Description

Well Suited for Pi-Network Topology

- Select the Input Voltage and Output Current as the Dependent Variables and the Input Current and Output Voltage as the Independent Variables

H-Parameter (Hybrid) Network Description

Extensively Used for Transistors

Figure 2 Small-Signal Equivalent Circuit Formulation Procedure

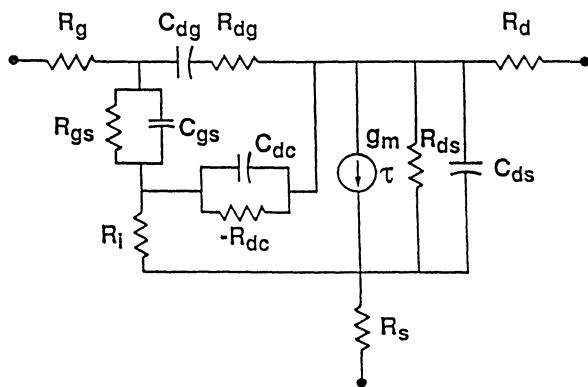


Figure 3a Physical MESFET Equivalent Circuit

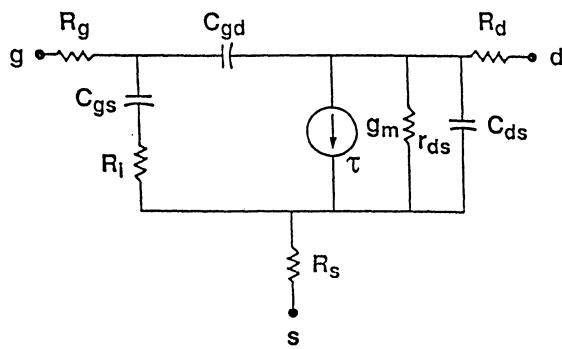
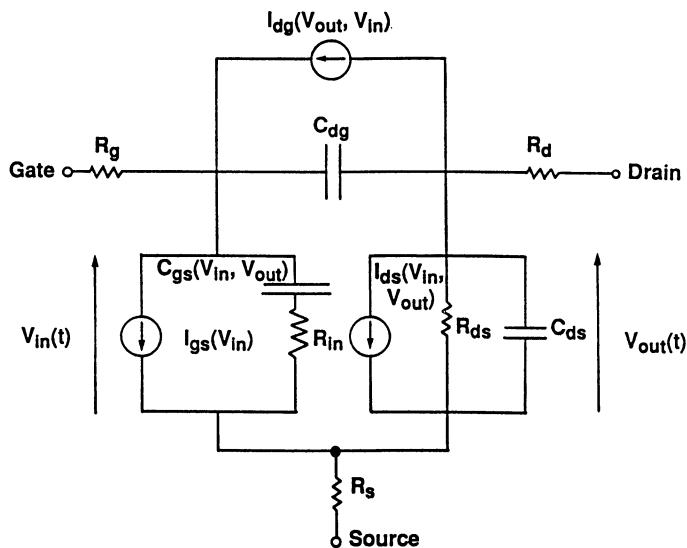


Figure 3b Simplified MESFET Equivalent Circuit



Nonlinear Circuit Elements

$I_{dg}(V_{out}, V_{in})$	Drain-gate voltage-controlled current source due to drain-gate avalanche breakdown
$I_{gs}(V_{in})$	Gate voltage-controlled current source due to forward biasing of the gate
$I_{ds}(V_{in}, V_{out})$	Drain-source voltage-controlled current source
$C_{dg}(V_{out}, V_{in})$	Drain-gate capacitance
$C_{gs}(V_{in}, V_{out})$	Gate-source capacitance
$C_{ds}(V_{out})$	Drain-source capacitance
$R_{in}(V_{in}, V_{out})$	Gate-source charging resistance
$R_{ds}(V_{in}, V_{out})$	Drain-source resistance

Figure 4 Large-Signal MESFET Equivalent Circuit

Table 1
 Large-Signal MESFET Equivalent Circuit Models
 Drain Current Generator Expressions

Shockley (1952)

$$I_{ds} = I_p \left[1 - \frac{(V_{gs} - V_{bi})}{V_p} \right]^n = \beta (V_{gs} - V_T)^2$$

SPICE2 - Shichman and Hodges (1968)

$$\begin{aligned} \text{linear region: } I_{ds} &= \beta V_{ds} [2(V_{gs} - V_p) - V_{ds}] (1 + \lambda V_{ds}) \\ \text{saturation region: } I_{ds} &= \beta (V_{gs} - V_p)^2 (1 + \lambda V_{ds}) \end{aligned}$$

Van Tuyl and Liechti (1974) / Taki (1978) / Curtice (1980)

$$I_{ds} = \beta (V_{gs} - V_p)^2 (1 + \lambda V_{ds}) \tanh(\alpha V_{ds})$$

Curtice includes a time delay: $I_{ds} = f[V_{gs}(t - \tau), V_{ds}]$

Materka and Kacprzak (1985)

$$I_{ds} = I_{dss} \left(1 - \frac{V_g}{V_p} \right)^2 \tanh\left(\frac{\alpha V_d}{V_g - V_p}\right)$$

$$V_p = V_{p0} + \gamma V_d$$

Curtice and Ettenberg (1985)

$$I_{ds} = (A_0 + A_1 V_1 + A_2 V_1^2 + A_3 V_1^3) \tanh(\alpha V_{out})$$

$$V_1 = V_{in}(t - \tau) [1 + \beta(V_{ds0} - V_{out})]$$

Statz, Smith, Pucel and Haus (1987)

$$I_{ds} = \frac{\beta (V_{gs} - V_T)^2}{1 + b(V_{gs} - V_T)} (1 + \lambda V_{ds}) \tanh(\alpha V_{ds})$$

procedure is inherently weak. That is, there is no scientific basis for the work. The non-linear model is not derived from basic principles, but is extrapolated from an established small-signal model. There is no scientific reason to expect the non-linear model to give accurate results in comparison with experiment. The technique generally ignores inter-dependencies between the circuit elements and much of the physical operation of the device. Nonetheless, the technique is widely used and good results can be obtained, depending upon the degree of non-linearity encountered. The failure of a specific model to yield good results is the driving force behind the effort to introduce new models. This effort continues to this day and new 'models' are continually being reported. However, none of the large-signal equivalent circuit based models are or can be completely adequate. It is for this reason that most commercially available microwave simulators include several device models and allow the user to select the one that best suits their application.

Physics-Based MESFET Models

Physics based device models are developed from fundamental principles according to physical describing equations. This approach is far more rigorous than the equivalent circuit approach and yields models that provide a physical understanding of device performance. These models provide significant information concerning physical phenomena involved in device operation and are, therefore, very useful in investigations of optimum device structures and operational limitations. This type of information is, in general, not available from the empirically based equivalent circuit models.

The fundamental set of semiconductor device equations provides the basis for development of a device model, either small-signal or large-signal. These equations consist of the current-density equations, the continuity equations, Poisson's equation, and Faraday's law. These equations must be solved simultaneously with the appropriate boundary conditions in order to obtain an accurate representation of the device operation. For devices such as bipolar transistors, where both electrons and holes are important to the operation of the device, two sets of equations are required, one for each type of charge carrier. These equations form the drift-diffusion approximation that is applicable to conditions where the mobile charge carriers can be considered to be in thermal equilibrium with the crystal lattice. These conditions are generally valid when the device dimensions are large compared to the wavelength of the operation frequency, or when the RF period is long compared to the charge carrier relaxation time. When these conditions are not applicable, non equilibrium (i.e., hot-electron) effects can be significant and the drift-diffusion approximation is not valid. Under these conditions alternate modeling approaches must be employed. The MESFET models discussed in this chapter will only consider conditions consistent with drift-diffusion operation.

The drift-diffusion approximation allows the current density equation to be separated into terms that describe current conduction by drift and diffusion mechanisms. For electrons the current density equation can be written as

$$\vec{J}_n = q\mu_n n \vec{E} + qD_n \nabla n \quad (1)$$

where the first term indicates that electrons drift in response to an applied electric field, with current density proportional to the electric field according to the electron mobility μ_n (units of $cm^2/V\cdot sec$). The second term indicates that electron conduction also occurs by a diffusion mechanism, where current density flows in response to a gradient in the electron

density. For the latter case the proportionality constant is the electron diffusion coefficient D_n (units of cm^2/sec). For nondegenerate semiconductors the diffusion coefficient and the mobility are related by the Einstein relation

$$D_n = \mu_n \frac{kT}{q} \quad (2)$$

The current density equation, as expressed in (1) applies to relatively low electric field operation where electron flow is linear. In this region the electron velocity is directly related to the magnitude of the electric field according to the relation

$$v = \mu_n E \quad (3)$$

For large electric fields, as are encountered under large-signal operating conditions, the electron velocity saturates and becomes a nonlinear function of electric field.

The continuity equation for electrons is

$$\frac{\partial n}{\partial t} = G_n + \frac{1}{q} \nabla \cdot \vec{J}_n \quad (4)$$

This equation states that the time rate of change of charge within a volume is equal to the rate of flow of charge out of the volume. Charge within the volume may be generated by some mechanism (G_n), such as optical excitation or avalanche ionization. Poisson's equation and Faraday's law complete the basic set of equations required for device analysis. Poisson's equation is

$$\nabla \cdot \vec{E} = \frac{q}{\epsilon} (N_d - n) \quad (5)$$

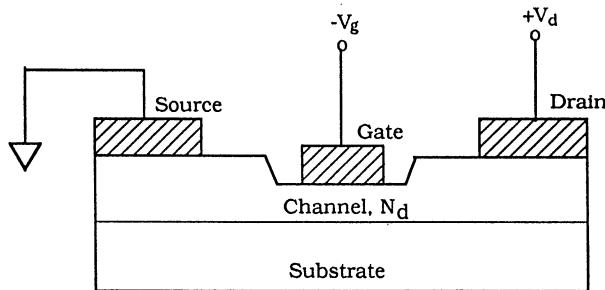
where N_d represents the density of ionized donor impurities and n is the free electron density (both have units of cm^{-3}). Faraday's law is

$$\nabla \times \vec{E} = - \frac{\partial \vec{B}}{\partial t} \quad (6)$$

Magnetic effects (B) can usually be ignored so that the right hand side of (6) is zero. This, in turn, allows the electric field to be determined from the spatial gradient of the electric potential according to the expression

$$\vec{E} = -\nabla V \quad (7)$$

In order to develop a physics based MESFET model the device equations are applied to the device geometry and solved with the appropriate boundary conditions. A typical MESFET geometry is shown in Fig. 5. Simultaneous solution of the device equations allows for a formulation for the voltage-current behavior of the device. Unfortunately, the two-dimensional device geometry produces significant complexities in the solution. The equations can be solved using numerical procedures based upon the finite difference or finite element solution techniques. Advantages of this approach are that the solution can easily



Device Equations

- *Poisson's Equation*

$$\nabla^2 \Psi = -\frac{q}{\epsilon} [N(y) - n(x, y)]$$

- *Current Density Equation*

$$\vec{J} = -qn\bar{v} + qD_n \nabla n$$

- *Continuity Equation (Conduction Current)*

$$\nabla \cdot \vec{J} = q \frac{\partial n}{\partial t}$$

- *Terminal Current (Conduction + Displacement)*

$$\vec{J}_t = \vec{J} + \epsilon \frac{\partial \vec{E}}{\partial t}$$

Where

$$\vec{E} = -\nabla \Psi$$

Figure 5 MESFET Geometry and Device Equations

include all pertinent device physics and therefore, the technique can yield very accurate results. The procedure, however, is computationally intensive and too slow to directly yield time dependent RF solutions.

A modified approach is available based upon quasi-two-dimensional techniques. In this approach a modified and simplified set of the device equations is solved. Generally, the two-dimensional Poisson's equation (eqn. 5) is solved and integrated with a one-dimensional current density equation (eqn. 1). The equations can be solved either analytically or numerically. The analytic approach produces a computationally efficient model that is easily integrated into circuit simulators. This approach has yielded excellent results in comparison with experimental data. The numerical approach is more robust and comprehensive and permits investigation of devices with any geometry. Phenomena such as surface effects are easily included. The numerical approach is more computationally intensive, but still yields a model that is easily integrated into circuit simulators.

RF Circuit Simulators, and the Harmonic Balance Method

A large-signal device model based upon the semiconductor device equations is highly nonlinear and is most efficiently solved in the time domain. Both analytic and numerical solution techniques can be applied. The analytic approach is attractive since it yields a deterministic set of analytic equations that can describe large-signal operation with good accuracy, for certain conditions. The approach requires approximations regarding device geometry and operation, which can prove limiting for certain device structures. The numerical approach is more robust and is suitable for investigation of most device structures. The technique generally requires significant execution time.

Large-signal models that have good accuracy in comparison with experimental data can quickly become complicated due to the necessity of explicitly including additional nonlinear phenomena. For example, large-signal analysis requires that many device parameters be known functions of variables, such as electric field, temperature, current density, etc. In addition, operational nonlinearities due to phenomena such as forward and reverse conduction of contact electrodes can complicate model formulation. Often it is possible to determine explicit analytic expressions for the parameter variations. However, when these expressions are included in the set of semiconductor equations the resulting device model becomes quite complicated. Although the device model can generally be solved in the time domain, it is not unusual to encounter significant model execution time. The simulation time can prove to be a major limitation to utilization of the model in frequency domain simulators.

Microwave engineers generally work in the frequency domain. The interfacing of time domain device models and frequency domain circuit simulators can be a difficult process. The harmonic balance method has been extensively developed for this purpose. The harmonic balance method is conceptually simple. In this method the circuit is divided into linear and nonlinear sub-circuits as shown in Fig. 6.

The linear sub-circuit is analyzed in the frequency domain, whereas the nonlinear circuit is characterized in the time domain. It is assumed that the excitations present in the circuit are sinusoidal and harmonically related. In advanced harmonic balance formulations it is possible to remove the latter restriction. Also, it is assumed that there are $(M+1)$ nodes between the linear and nonlinear sub-circuits including one selected as the reference (or ground) node. The time domain voltage at node m relative to the reference node can be approximated by a truncated Fourier Series

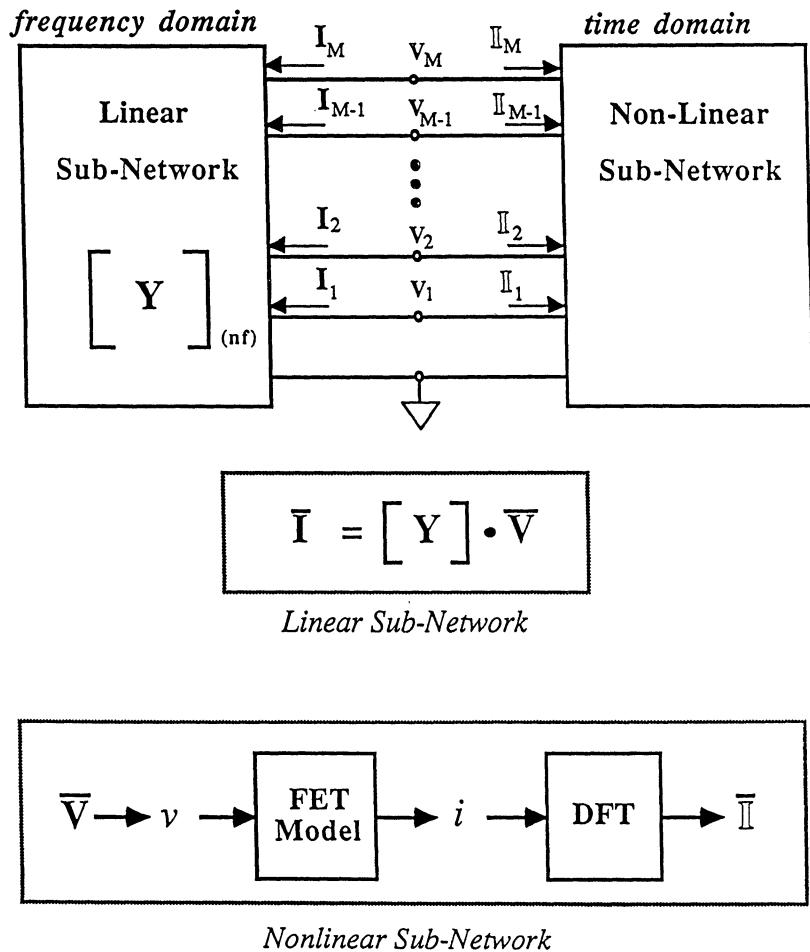


Figure 6 Harmonic Balance Technique

$$v_m(t) = \operatorname{Re} \left\{ \sum_{n=0}^N V_{mn} e^{jn\omega_1 t} \right\} \quad (8)$$

where V_{mn} is the voltage phasor at the n th harmonic and m th node. The lower index of the sum corresponds to the dc value of $v_m(t)$. Similarly, the current flowing through node m into the nonlinear sub-circuit can be expressed as

$$i_m(t) = \operatorname{Re} \left\{ \sum_{n=0}^N I_{mn} e^{jn\omega_1 t} \right\} \quad (9)$$

where I_{mn} is the corresponding current phasor at the n th harmonic and m th node. Also $i'_m(t)$ is defined as the current flowing through the m th node into the linear sub-circuit with the associated current phasors I'_{mn} . Applying Kirchoff's Current Law to the M nodes of the circuit (excluding the ground node) yields

$$i_m(t) + i'_m(t) = 0, \quad m = 0, 1, \dots, M \quad (10)$$

In practice, due to the truncation of the Fourier Series representation of the currents, equation (10) will not hold and the solution requires that the currents be adjusted until the net current flowing out of all nodes is minimized. This is equivalent to the minimization of an objective function of the form

$$E = \sum_{m=1}^M \sum_{n=0}^N |I_{mn} + I'_{mn}|^2 \quad (11)$$

in which the unknowns of the least-square minimization are the interface voltage phasors V_{mn} . Both I_{mn} and I'_{mn} are functions of V_{mn} . I'_{mn} is related to V_{mn} by the admittance matrix of the linear sub-circuit. Therefore,

$$[I'_{mn}]_{M \times 1} = [Y(n\omega_1)]_{M \times M} \cdot [V_{mn}]_{M \times 1} \quad n = 0, 1, \dots, N \quad (12)$$

where the first term on the right-hand side denotes the admittance matrix calculated at the n th harmonic. The phasor I_{mn} is a nonlinear function of V_{mn} and it is necessary to transform between the time and frequency domains using the Fourier Transform. A flow diagram for a Harmonic Balance algorithm is shown in Fig. 7.

A difficulty encountered with the harmonic balance method results from the Fourier transformation required by the algorithm. The Nyquist Sampling Theorem states that in order to recover all of the harmonics of a band-limited waveform such as $i_m(t)$ from its samples (i_{mk}), the sampling frequency $f_s = 1/T_s$ must be greater than or equal to twice the bandwidth of the original waveform. For the harmonic balance method, this translates into the condition that the number of time samples per cycle must be at least $(2N+1)$ where N is the highest harmonic number present in the circuit. Therefore,

$$N_s \geq (2N+1) \quad (13)$$

This restriction is severe in the case of inter modulation or mixing analysis where the signal bandwidths are several orders of magnitude larger than the frequency separation generally of interest.

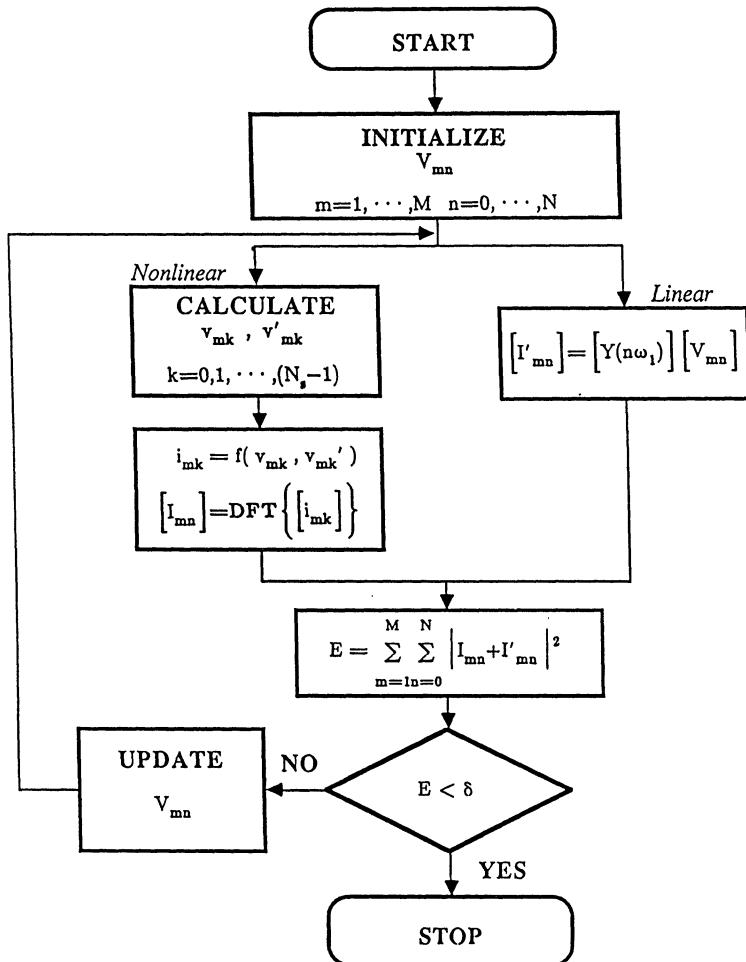


Figure 7 Harmonic Balance Simulation Algorithm

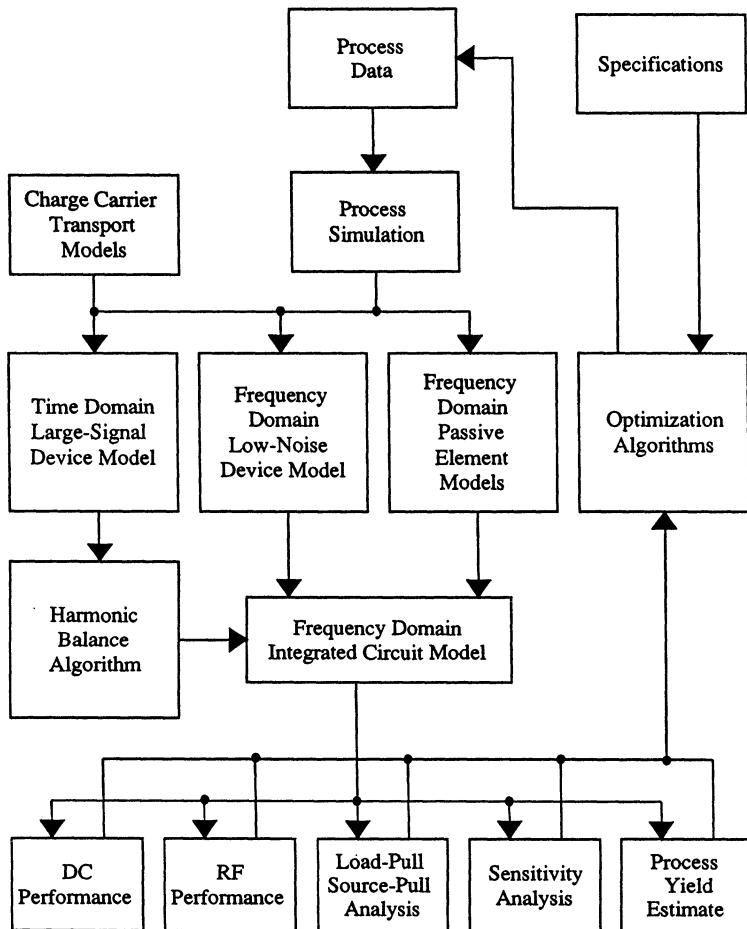


Figure 8 Comprehensive Microwave Integrated Circuit Simulator

The most significant fraction of the simulation time for the harmonic balance method is used in the error minimization. Various types of root finding algorithms have been investigated in an effort to reduce execution time, and significant progress has been achieved. Simulation time for highly nonlinear circuits, however, can still be unacceptably long. Time delays, as are often encountered in microwave devices, introduce significant complexity into Harmonic Balance algorithms, and when these delays are significant the technique often fails to converge.

Although much progress has been achieved in harmonic balance simulation procedures, it is still difficult to employ the method using highly nonlinear device models. Also, it is difficult to use numerically based physical device models in these simulators. The long execution time normally required by the device model, coupled with the harmonic balance simulation time can lead to unsatisfactory performance. For this reason, physically based models are not often used in frequency domain simulators. This has resulted in the wide application of equivalent circuit models in generally available microwave simulators, especially those with large-signal, nonlinear device models.

A block diagram for a comprehensive harmonic balance, large-signal MESFET integrated microwave simulator is shown in Fig. 8. The simulator allows the performance of a microwave integrated circuit to be determined as a function of material, process, and device design specifications. The simulator is useful for large-signal, non linear applications such as power amplifiers, oscillators, etc. The integrated simulator allows both MESFET design and embedding dc and RF circuits to be investigated. The physical MESFET model permits the dc and RF performance of a device or integrated circuit to be determined as a function of process and device design information and/or bias and RF operating conditions.

Summary

Device models are required for design and analysis of both small-signal and large-signal components in microwave systems. Small-signal components include low-noise amplifiers, detectors, etc. and large-signal components include power amplifiers, oscillators, mixers, switches, etc. Small-signal models are relatively easy to develop and are widely available in commercial systems simulators. Large-signal, non-linear device models are much more difficult to develop and only equivalent circuit based models are widely available. Large-signal conditions exist when the RF or signal voltage is significant relative to the dc operating bias. Under these conditions device operation is non linear and accurate models require descriptions of the fundamental non-linearities. Appropriate models can be developed using either equivalent circuit or physical modeling techniques. The former are generally empirically based and element values are determined from measurement data by parameter extraction. The models are widely used, are easy to implement in harmonic balance microwave simulators, but demonstrate limited accuracy in certain applications. Physical models, in contrast, are developed from the semiconductor device equations and contain, therefore, an accurate description of the physical operation of the device. The models tend to be complex and require significant computational resources to obtain solutions. The models are, in general, difficult to apply in microwave circuit simulators. Physical device models based upon analytic solution techniques provide a useful compromise between model accuracy and simulation efficiency.

HEMT MODELS AND SIMULATIONS

P. Lugli, M. Paciotti

*Dipartimento di Ingegneria Elettronica, Universita' di Roma "Tor Vergata"
Via della Ricerca Scientifica, 00133 Roma, Italy*

E. Calleja, E. Munoz, J.L. Sanchez-Rojas

*Dipartimento de Ingenieria Electronica, Universidad Politecnica de Madrid
Ciudad Universitaria, 28040 Madrid, Spain*

F. Dessenne, R. Fauquembergue, J.L. Thobel

*IEMN/DHS, Universite' des Sciences et Technologies de Lille
Avenue Poincare', 59652 Villeneuve d'Ascq, France*

G. Zandler

*Walter Schottky Institute, Technische Universität München
Am Coulombwall, 85747 Garching, Germany*

Abstract: The paper presents a series of modeling tools for pseudomorphic High Electron Mobility Transistors, discussing their physical content and presenting specific applications. The presented results will show the most peculiar features of electronic transport in such device.

1. Introduction

Pseudomorphic (PM) HEMTs are certainly among the most promising devices for microwave and millimeterwave applications [1,2]. A description of the device, covering issues such as fabrication, characterization, optimization, material choice, reliability, and applications, is provided by the other contributions to this volume. Here, we want to take a physical approach, by analyzing the various modeling tools that have been developed for PM HEMTs, looking at them in view of their underlying physical model. From this point of view, PM HEMTs constitute a quite sophisticated system. The basic structure shown in Fig. 1. The key elements are the double GaAs/InGaAs/AlGaAs heterostructure with strained InGaAs channel, the Schottky barrier at the gate contact, the ohmic contact of source and drain, the doped cap GaAs layer: a fantastic combination of materials, interfaces, dopants, a variety of processing steps, all features which need to be accounted for in some way if a realistic model of the device has to be set up.

Modeling such a system requires necessarily a series of approximations. From a geometrical point of view, a standard simplification is made by ignoring the depth of the device, thus limiting the analysis to a two dimensional section of the device (in the plane of the page in Fig. 1). As we will see later, the most complete models define the physical and transport characteristics on such 2D domain. It is nevertheless possible to further simplify the problem, by decoupling the two directions, or even discarding one of them.

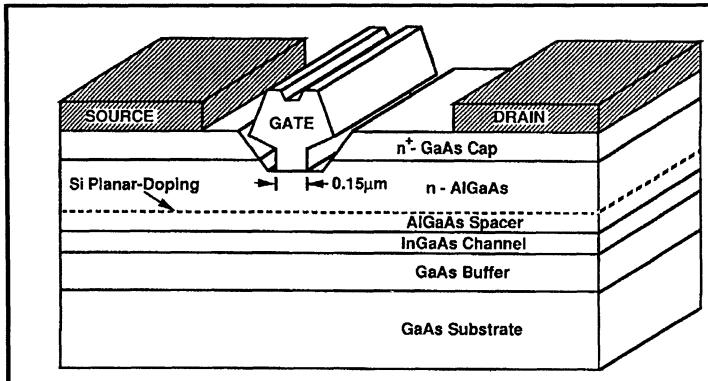


Fig. 1 Schematic 3D representation of a PM-HEMT, showing a typical submicron structure with a T-gate, cap layer, delta doping in the AlGaAs barrier.

Technological aspects such as cap layer doping, gate recess depth, ohmic contact deposition, Schottky barrier, doping profile, etc., should be included into realistic models. This is actually a very difficult task, since often very little knowledge is available on such issues, which are more a question of practice and experience rather than of consolidated theoretical background. As a result, theoretical models and simulations tend to address truly technological features at the lowest possible level, if possible by some phenomenological or ad-hoc approaches. We will see specific examples in the following.

From a physical point of view, several features need to be considered. In the first place, the presence of the double heterojunction on the sides of the InGaAs channel induces electron confinement in the channel region. If the channel width is sufficiently narrow, then the electronic motion perpendicular to the heterointerfaces is forbidden and confined discrete energy levels appear along that direction. In order to account for such effect a fully quantum mechanical approach is required. On the other side, problems related to electronic transport along the channel (such as the effect of strain and doping on mobility and drift velocity, the presence of high electric field favoring carrier heating and real space transfer, dynamical effects connected to deep traps, etc.) are better handled via semiclassical models. The basis and formal treatment of the two models, and their connection, will be dealt with in Sections 2 and 3.

A general classification that can be used for the various modeling tools is based on the hierarchy of sophistication in terms of the physical and geometrical models which are adopted. Thus we will distinguish the various approaches as:

- Analytical and semianalytical models;
- 1D numerical models
- Quasi 2D numerical models
- 2D numerical models
- 2D numerical simulations.

Examples and results will be presented in Sects. 4-6 which will show the strengths and weaknesses of each model, focusing in particular on 1D quantum mechanical model (Sect. 4), two different quasi 2D models (Sect. 5) and a full 2D Monte Carlo simulation (Sect. 6). We will always concentrate on DC phenomena, although most of the approaches we will deal with can be applied also to the AC case.

2. Quantum mechanical model

Several attempts have been made to set up a fully quantum mechanical approach for electronic transport in semiconductors (see e.g. [3] and references therein). Although some degree of success has been achieved for the general formulation of the problem, virtually no application to real devices has yet been presented. Most of the problems lie in the difficulty to combine coherent effects where the wave nature of the charge particles is dominant with the incoherent contributions brought about by collisions.

One approach which has been proven quite successful, and that does not require prohibitive computational time, is that based on a coupled description via Schrödinger's and Poisson equations. The general principles of the method will be described in this section, while applications to HEMT modeling will be presented in the following one.

2.1 SCHROEDINGER'S EQUATION

The time independent Schrödinger's equation (SE):

$$-\frac{\hbar^2}{2m_0} \nabla^2 \psi(r) + q V_C(r) \psi(r) = E \psi(r) \quad (1)$$

gives the particle wavefunction $\psi(r)$ as a function of the spatial coordinate for a given potential V_C . It is well known that if $V_C(r)$ is the periodic potential representing the crystal lattice potential, then the solution of Eq. 1 is given by a series of Bloch functions $\psi_{nk}(r) = u_{nk}(r) \exp(j \mathbf{k} \cdot \mathbf{r})$, where \mathbf{k} identifies the electron crystal momentum and $u_{nk}(r)$ are periodic functions with the same periodicity of the lattice. The eigenvalues E of Eq. 1 define the so-called dispersion relations $E = E_n(\mathbf{k})$. In the presence of external perturbations the electron wavefunctions are not pure Bloch states. It is nevertheless possible to use approximate perturbative schemes that formally allow us to keep that type of description, with the perturbative terms represented by an interaction hamiltonian and with the $E = E_n(\mathbf{k})$ relation containing all information about the band structure of the host material. In what follows, the term V_C will be the electrostatic potential seen by the electrons, which includes the external bias applied to the contacts. Thus, the solution of Eq. 1 will provide a description of the electronic states of a given structure for a given potential distribution. As we will see below, by coupling SE to Poisson equation (PE), a self consistent solution can be obtained for the equilibrium configuration in the presence of any external bias.

Schrödinger's equation could in principle be solved in three dimensions. This is though far too demanding in computational terms. It is therefore preferable to simplify the problem and to work in just one dimension. With reference to Fig. 1, the assumptions that justify such simplification are: *i*) that the potentials applied on the contact are constant with depth, *ii*) that no current goes through the gate contact, and *iii*) that charge and potential distribution only depend on the relative bias between gate and substrate. In this way, the HEMT structure can be reduced to that of Fig. 2, and the potential variation can be limited to the z direction: $V_C = V_C(z)$. This corresponds to considering a cut of the xz plane of Fig. 2 running across the gate electrode.

Under such assumptions, the electron wavefunction $\psi(r)$ can be separated as $\psi(r) = \Psi_i(z) \cdot \vartheta(x, y)$. While the solution in the xy plane is nothing else than the free

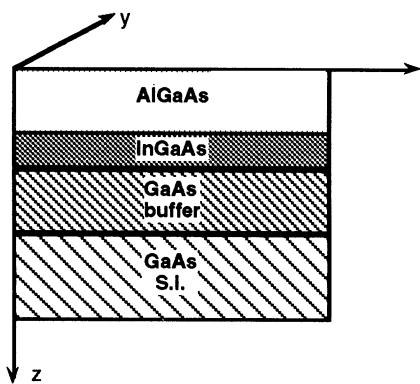


Fig. 2: Schematic representation of a HEMT structure for the solution of Schrödinger's equation

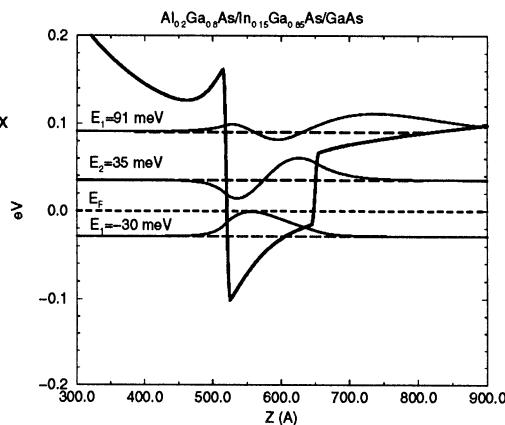


Fig. 3: Energy levels and electron wavefunction a typical HEMT structure

electron (with the appropriate effective mass), the z-component should fulfill the equation:

$$-\frac{\hbar^2}{2 \cdot m_0} \frac{d}{dz} \left(\frac{1}{m_z^*} \frac{d\Psi_i(z)}{dz} \right) + V_C(z) \cdot \Psi_i(z) = E_i \cdot \Psi_i(z) \quad (2)$$

where E_i denotes the quantized energy levels ("subbands") characteristic of the z-direction. This electron quantization is typical quantum mechanical effects, and corresponds to the impossibility for the electrons to move perpendicular to the heterointerfaces as long as their energy is smaller than the confining energy given by the conduction band discontinuity. Such quantization originates the name of "two dimensional electron gas (2DEG)" which describes electronic conduction in HEMTs [4].

Equation 2, that is the 1D SE for the z-direction, does not allow in general analytic solutions, and should be solved numerically. A typical example is given in Fig. 3 where the electronic wavefunction of the three lowest subbands are shown for a given potential, together with the potential profile (solid line) and the Fermi level (dashed line, taken a reference for the energy scale).

The knowledge of the wavefunction for all subbands allows us to calculate the 2D electron concentration in the InGaAs well, as

$$n_{2D}(z) = \frac{m(z) K_b T}{\pi \hbar^2} \sum_i \left\{ |\Psi_i(z)|^2 \ln \left[1 + \exp \left(\frac{E_f - E_i}{K_b T} \right) \right] \right\} \quad (3)$$

where the sum extends to all subbands.

2.2 POISSON EQUATION

Schroedinger's equation provides the electronic wavefunction (and therefore the concentration) for a given electrostatic potential. In the case of an HEMT structure such as that of Fig. 2, the only available information concern the applied potential between

gate and substrate. Thus, the actual potential distribution inside the structure has to be determined starting from the given boundary conditions by solving PE for the charge density ρ :

$$\frac{d}{dz} \left(\epsilon(z) \frac{dV(z)}{dz} \right) = -q \cdot \rho(z) \quad (4)$$

where $\epsilon(x)$ is the dielectric function of the material (which changes every layer) q the electronic charge, and $V(z)$ is the continuos electrostatic potential (e.g. the vacuum level), given by:

$$V_C(z) = -q \cdot V(z) + \Delta E_1 \cdot u(z - z_1) + \Delta E_2 \cdot u(z - z_2) + V_{xc}(z) \quad (5)$$

where ΔE_1 e ΔE_2 are the band edge discontinuity of the two heterojunctions, $u(z)$ is the Heaviside step function (or any function describing the transition from a material to the other across the heterointerfaces) and V_{xc} accounts for exchange and correlation effects.

The charge distribution $\rho(z)$ is given by the sum of fixed and mobile charges:

$$\rho(z) = p(z) + N_D^+(z) - n(z) - N_A^+(z)$$

where $p(z)$ e $n(z)$ are the hole and electron densities, while $N_D^+(z)$, $N_A^+(z)$ are ionized donor and acceptor concentrations, respectively. For n-type devices, as the HEMT we are considering, $\rho(z)$ simplifies to: $\rho(z) = N_D^+(z) - n(z)$, where

$$N_D^+(z) = \frac{N_D(z)}{1 + 2 \exp\left(\frac{E_f - E_D(z)}{K_b T}\right)}$$

$E_D(z)$ being the ionization energy of the donors and N_D the total donor density.

The mobile charge $n(z)$ has contributions from the different layers of the structure. For the InGaAs layer, the confined electron concentration is the one given by Eq. (3), while the charge which might populate the GaAs substrate or the AlGaAs barrier has to be calculated as

$$n_b(z) = \frac{2}{\sqrt{\pi}} N_C(z) F_{1/2}\left(\frac{E_f - V_C(z)}{K_b T}\right)$$

where $N_C(z)$ is the effective density of state for the conduction band of the material and $F_{1/2}(z)$ the Fermi integral of order 1/2. The total mobile charge is the sum of the two contributions n_{2D} and n_b .

2.3 COUPLED SCHROEDINGER AND POISSON EQUATION

We have seen in the previous section that SE requires the potential profile as an input, providing the carrier concentration (as a sum for all subbands of the integrals over the square amplitude of the wavefunctions). On the contrary, PE starts from the carrier distribution to arrive at the potential profile. Clearly, in the case of a structure such as that shown in Fig. 2, the only available information is the relative bias applied between gate and substrate. Thus, a coupled solution if the two equation is needed. This is usually

achieved by setting up an iterative solution which starts from an initial guess, for instance of the charge distribution, solves PE for the potential profile, feeds this profile into SE, and continues in an iterative manner until convergence is achieved.

The solution of each equation is usually achieved by setting up a grid over which the differential equation are discretized using conventional techniques. For instance, if a finite difference scheme is used [5], then SE takes the form:

$$-\frac{\hbar^2}{2m_0} \left[\frac{\frac{1}{m_{i+\frac{1}{2}}^*} \frac{\Psi_{i+1} - \Psi_i}{h_{i+\frac{1}{2}}} - \frac{1}{m_{i-\frac{1}{2}}^*} \frac{\Psi_i - \Psi_{i-1}}{h_{i-\frac{1}{2}}}}{\frac{h_{i-\frac{1}{2}} + h_{i+\frac{1}{2}}}{2}} \right] + \Psi_i V_{Ci} = E \Psi_i \quad (6)$$

where $\Psi_i = \Psi(z_i)$ is the wavefunction at the discrete position z_i .

Such difference equation can be rewritten as

$$a_i \cdot \Psi_{i-1} + b_i \cdot \Psi_i + c_i \cdot \Psi_{i+1} = E \cdot \Psi_i$$

where:

$$a_i = -\frac{\hbar^2}{m_0} \left(\frac{1}{m_{i-\frac{1}{2}} h_{i-\frac{1}{2}} L_i^2} \right) \quad c_i = -\frac{\hbar^2}{m_0} \left(\frac{1}{m_{i+\frac{1}{2}} h_{i+\frac{1}{2}} L_i^2} \right)$$

$$b_i = -a_i - c_i + V_{Ci} \quad L_i^2 = h_{i-\frac{1}{2}} + h_{i+\frac{1}{2}}$$

or, in matrix notation:

$$\mathbf{A} \Psi = E \Psi$$

Thus, the calculation of the energy levels E_k and electron wavefunctions $\Psi_k(z)$ is reduced to the calculation of the eigenvalues and eigenvectors of the matrix \mathbf{A} , whose elements depend on the structure material parameters.

Poisson equation can be discretized in a similar way. As pointed out above, an iterative solution is preferable with respect to a simultaneous one. In particular, it is better to force the convergence of the iterative scheme by applying "relaxation" terms that smooth the potential variation from one iteration to the next one. This can be achieved via a method introduced originally by Gummel for the solution of the coupled Poisson and current equations in Drift Diffusion algorithms (see Sect. 3.2) [6]. Within such scheme, PE takes the form:

$$\frac{d}{dz} \left(\epsilon(z) \cdot \frac{dV(z)}{dz} \right) = -q \cdot \rho(z) + q \cdot n(z) \cdot \frac{V(z) - V^0(z)}{KT} \quad (7)$$

where $V^0(z)$ is the potential profile obtained in the previous iteration. The KT term, originating from the assumption that the concentration of mobile carriers can be calculated using a Maxwellian statistics, can be seen as a convergence factor.

Equation 7 can be discretized in a similar way as SE to give:

$$\frac{\varepsilon_{i+\frac{1}{2}} \cdot \frac{V_{i+1} - V_i}{h_{i+\frac{1}{2}}} - \varepsilon_{i-\frac{1}{2}} \cdot \frac{V_i - V_{i-1}}{h_{i-\frac{1}{2}}}}{\frac{h_{i+\frac{1}{2}} + h_{i-\frac{1}{2}}}{2}} = -q \cdot p_i + q \cdot n_i \cdot \frac{V_i - V_i^0}{KT}$$

which can also be expressed in matrix notation as before. Specific examples obtained from the couple d solution will be given in Sect. 4.

3 Semiclassical transport description

While a quantum approach such as that described in the previous section provides an adequate static picture of a device structure (wavefunctions, energy levels, potential profile), it is in no way capable of describing electronic transport. This is usually achieved in terms of semiclassical approaches, where the carrier dynamics is described in terms of classical trajectories, and the collision events are taken into account quantum mechanically. The starting point for such description is Boltzmann transport equation (BTE):

$$\left(\frac{\partial}{\partial t} + \dot{k} \bullet \nabla_k + \dot{v} \bullet \nabla_r + I_k \right) f(r, k, t) = 0 \quad (8)$$

which describes the evolution of the single particle distribution function (d.f.) $f(r, k, t)$, representing the probability to find the particle at time t in a volume $dr dk$ around the phase space point r, k . The term I_k is the so-called collisional integral, describing the temporal changes in the d.f. due to collisions. In its integral form I_k is given by:

$$I_k f(r, k, t) = \frac{V}{(2\pi)^3} \int [f(r, k', t) P(k', k) (1 - f(r, k, t)) - f(r, k, t) P(k, k') (1 - f(r, k', t))] dk' \quad (9)$$

where V is the crystal volume and $P(k', k)$ the probability per unit time that a collision brings the particle from an initial state k' into a final state k . The $(1-f)$ terms account for Pauli exclusion principle. When more collision processes are present, as is usually the case, $P(k', k)$ is given by the sum of the transition probabilities of the single processes.

The knowledge of the d.f. completely determines the transport problem. In particular, the first three moments of the d.f. give directly the three main physical observable, namely:

$$- \text{carrier concentration (0th order)} \quad n(r, t) = \frac{2}{(2\pi)^3} \int f(r, k, t) dk \quad (10)$$

$$- \text{current density (1st order)} \quad J(r, t) = \frac{2q}{(2\pi)^3} \int v(k) f(r, k, t) dk \quad (11)$$

$$- \text{average energy (2nd order)} \quad \langle E(r, t) \rangle = \frac{2q}{(2\pi)^3} \int E(k) f(r, k, t) dk \quad (12)$$

The solution of BTE is a formidable task, due to the integro-differential form of its complete expression. Several approximations can be introduced which lead to numerical

solutions. From the point of view of device simulations, the most important ones are the Drift-Diffusion, the Hydrodynamical, and the Monte Carlo approaches.

3.1 THE HYDRODYNAMICAL APPROACH

We have seen above that the first moments of the d.f give directly the main physical observable quantities for the electron transport. By multiplying each term of BTE by a power of v and by integrating over v it is possible to obtain the equations that the various moments must satisfy (within the semiclassical framework of BTE the electron crystal momentum \mathbf{k} is related to the electron velocity by the relation $v = \hbar\mathbf{k} / m$; in the following we will use v as a variable instead of \mathbf{k}) [7]. Such equations are respectively:

- *the continuity equation:*

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{u}) = -G \quad (13)$$

where G is related to generation-recombination phenomena, since collisions do not contribute to the integral of the r.h.s. of BTE;

- *the momentum conservation equation:*

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) \mathbf{u} + \frac{e\mathbf{F}}{m} + \frac{1}{mn} \nabla (nK_B T_e) = \left(\frac{\partial \mathbf{u}}{\partial t} \right)_{\text{coll}} \quad (14)$$

where the term $(\partial \mathbf{u} / \partial t)_{\text{coll}}$ accounts for the effects of collision events on the average velocity; and

- *the energy conservation equation:*

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla \right) w + \nabla \cdot \mathbf{Q} + ev \cdot \mathbf{F} + \frac{1}{n} \nabla (nK_B T_e \mathbf{u}) = \left(\frac{\partial w}{\partial t} \right)_{\text{coll}} \quad (15)$$

where the vector \mathbf{Q} represents the heat flux and w is the particle average energy given by $w = mu^2/2 + 3K_B T_e/2$. There u and T_e are the average drift velocity and the average electron temperature, respectively.

The set of equations (13-15) defines the "hydrodynamical" approach [8], which is a very appealing way to solve BTE since it combines a deep physical content to a high intrinsic speed, especially when the energy equation is simplified. The solution of the hydrodynamic equations is obtained numerically via standard discretization techniques.

The collision terms in Eqs. (14) and (15) are quite complicated. A useful approximation can be obtained by using the so-called relaxation times for momentum and energy, thus having:

$$\left(\frac{\partial \mathbf{u}}{\partial t} \right)_{\text{coll}} = -\frac{\mathbf{u}}{\tau_n} \quad (16)$$

and

$$\left(\frac{\partial w}{\partial t} \right)_{\text{coll}} = -\frac{w - w_0}{\tau_e} \quad (17)$$

where w_0 is the equilibrium kinetic energy equal to $3/2K_B T_0$. The relaxation time can be taken from Monte Carlo bulk simulations.

3.2 THE "DRIFT-DIFFUSION" SCHEME

In dealing with transport phenomena it can be convenient to introduce a phenomenological expression for the current density in the form:

$$\mathbf{J}(\mathbf{r}, t) = qn(\mathbf{r}, t)\mathbf{v}_d(F) - qD(F)\nabla_r n(\mathbf{r}, t) \quad (18)$$

where the drift velocity \mathbf{v}_d and the diffusion coefficient D are taken as function of the local electric field $F=F(r,t)$. Equation (18), which describes the electron motion under the influence of an external electric field and of a concentration gradient, is at the basis of the so-called Drift-Diffusion (DD) method [9]. The drift velocity is usually defined as the product of the electron mobility and the electric field. Together with the current continuity equation:

$$\frac{1}{q}\nabla J_n = \nabla(n\mu_n F + D_n \nabla n) = \frac{\partial n}{\partial t} + R$$

equation 18 substitutes Eqs. 13-15 of the hydrodynamical model. It can actually be shown that Eq. 18 can be derived directly from BTE under the assumption that the electrons remain in thermal equilibrium with the lattice ($T_e = T_0$). Thus, the DD is the most simplified approach to the solution of BTE, and can be implemented in very fast algorithms. Its shortcomings come from the physical simplifications, that do not allow to account for effects such as velocity overshoot or carrier heating typical of short channel devices.

The DD equation can be solved either by finite difference or by finite element method. Details can be found in [10].

3.3 THE MONTE CARLO METHOD

While the methods described so far are based on the direct numerical solution of the transport differential equations, an alternative approach is provided by simulations based on Monte Carlo techniques [11]. In this case, a simulation of the particle random walk in phase space is set up based on the probabilistic rules. For the case of charge transport in semiconductors such rules are determined directly from the quantum mechanical scattering probabilities. Carriers move in a given electric field as classical particles. Their "free flight" is interrupted by collisions (with phonons, impurities, other carriers, etc.) whose occurrence is controlled by the probability distributions mentioned above. By collecting information about the state of the particle (or the ensemble of particles when several of them are considered), it is possible to calculate the main physical quantities of interest (drift velocity, average energy, etc.) and also the carrier distribution function, which can be proven to fulfill BTE. Thus, the Monte Carlo method does indeed provide a solution of BTE, although not via a direct numerical solution. When semiconductor devices are simulated, then the real space particle dynamics has to be followed, where the particles move on a given electric field distribution on the discretized simulation domain. In this case, large numbers of particles are considered in parallel, so that the average transport properties (and the carrier concentration) can be calculated in each point of the device. The Monte Carlo method has been applied to a variety of devices. For a review, see [12].

3.4 SELF CONSISTENT SCHEMES

In order to fully describe a device, the solution of BTE (either by direct methods or by simulation) has to be performed together with that of Poisson's equation, which gives the field distribution within the device. This is usually done in an iterative manner, that is by *i*) starting with an initial guess, for instance of the carrier distribution; *ii*) solving PE for the potential profile; *iii*) using such profile to solve BTE; *iv*) calculating the new charge distribution from the BTE solution; and *v*) repeating the two-step scheme (*iii*-*iv*) until convergence is achieved. For details on the various methods see [13,14].

4. One dimensional models

As a first series of results for PM-HEMTs, we analyze the outcome of the 1D model based on the coupled Schrödinger-Poisson equations described in Sect. 2. The considered structure has been plotted in Fig. 2. For an InGaAs well of 13 nm, the first two levels are confined into the well, while the third extends into the GaAs buffer (Fig. 3). In thermal equilibrium, about 77% of the electrons occupy the first level, 20 % the second, and the remaining 3% can be found in the third one. As the model gives directly the carrier concentration, we will first examine its dependence on the gate bias, that is the "Charge control" model of the device [15].

4.1 PM HEMT CHARGE CONTROL

Charge density and potential profile are plotted at two different gate voltages in Fig. 4 and 5, respectively for the PM HEMT and for a conventional GaAs/AlGaAs HEMT. As long as the gate bias is sufficiently negative, electrons are confined into the 2D channel. For higher potentials, the AlGaAs barrier layer becomes in both cases populated. This corresponds to the turning-on of a "parasitic FET" which degrades the device performance, as the transport characteristics of the doped barrier layer are far inferior to the channel ones.

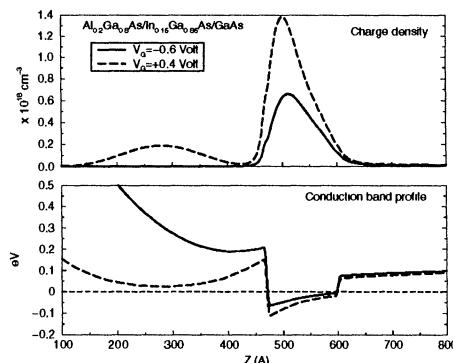


Fig. 4: Charge density and potential profile for a PM HEMT with uniform doping of $2 \times 10^{18} \text{ cm}^{-3}$ in the AlGaAs layer

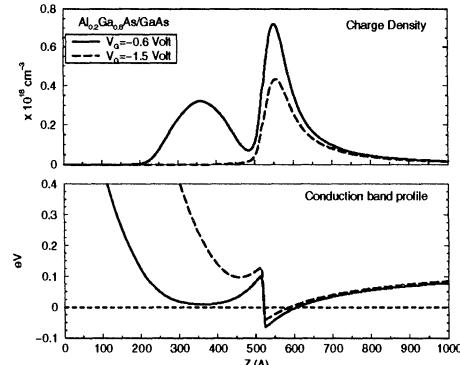


Fig. 5: Charge density and potential profile for a standard HEMT with uniform doping of $2 \times 10^{18} \text{ cm}^{-3}$ in the AlGaAs layer

The calculated dependency of the charge density on gate voltage is shown in Fig. 6. There, both the 2DEG (n_s) and the AlGaAs (n_b) densities are plotted for the PM and the

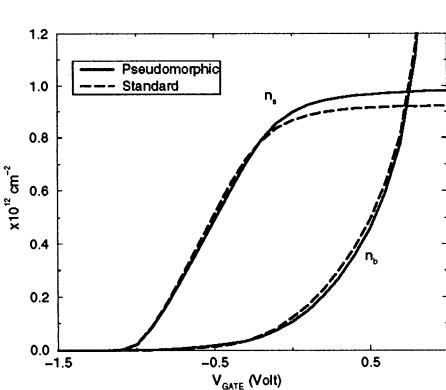


Fig. 6: Gate voltage dependence of the 2DEG density n_s and of the barrier density n_b for a PM (solid line) and a standard (dashed line) HEMT.

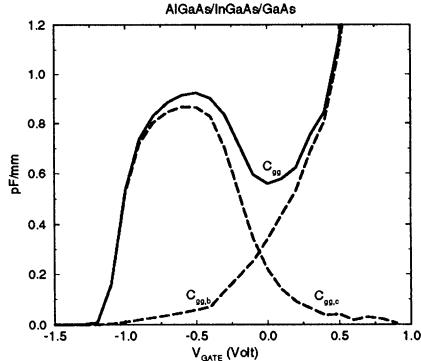


Fig. 7: Gate voltage dependence of the gate parasitic capacitance of a PM HEMT showing the contribution of the channel and of the barrier charge.

standard HEMT. Although the comparison has to be taken as qualitative, since the results depend on several factors such as the actual geometry or material composition (here the same doping and Al concentration of the AlGaAs layer has been assumed for the two HEMTs), it nevertheless shows that PM HEMT are characterized by better confinement (i.e. higher channel density with respect the total one). This is due to the higher band gap discontinuity at the InGaAs/AlGaAs interface with respect to the GaAs/AlGaAs, and to the lower electron effective mass of the InGaAs channel with respect to the GaAs one.

Another typical result provided by the 1D charge model concerns the contributions to the parasitic gate capacitance coming from the channel charge $C_{\text{gg},b}$, and from the barrier charge $C_{\text{gg},c}$, given respectively by :

$$C_{\text{gg},c} = q L_g \left. \frac{dn_s}{dV_G} \right|_{V_D=0} \quad C_{\text{gg},b} = q L_g \left. \frac{dn_b}{dV_G} \right|_{V_D=0}$$

where L_g is the gate length (taken as \$0.5 \mu\text{m}\$). The two contributions (dashed lines) and the total gate capacitance for the PM HEMT are shown in Fig. 6. Clearly, the presence of mobile carriers in the AlGaAs layer leads to high capacitance at positive gate voltages and therefore limits the high frequency performance of the device.

As a further example offered by the 1D model, Figs. 8-11 show the charge density and potential profile curves of PM HEMTs with different design of the doping distribution. In general, the use of the so called \$\delta\$-doping (that is the use of a single plane of dopants in the barrier layer) as the double benefit of a higher channel concentration, and of a better charge confinement at positive bias. The former effect is due to the possibility to put a high doping densities (up to \$5 \times 10^{12} \text{ cm}^{-2}\$) very close to the channel, the latter one to the fact that the donor plane screens the gate potential, leading to a linear potential drop from the Schottky barrier (lower graphs in Fig. 9 and 10) and consequently to a limited population of the AlGaAs layer and to better pinch-off characteristics.

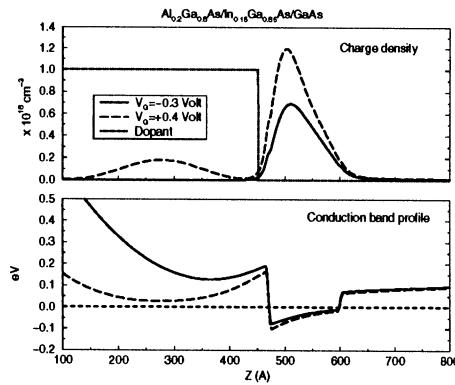


Fig. 8: Charge density (above) and potential profile (below) for a PM HEMT with uniform doping in the AlGaAs layer at two different gate voltages of -0.3 (solid) and 0.4 (dashed line) V

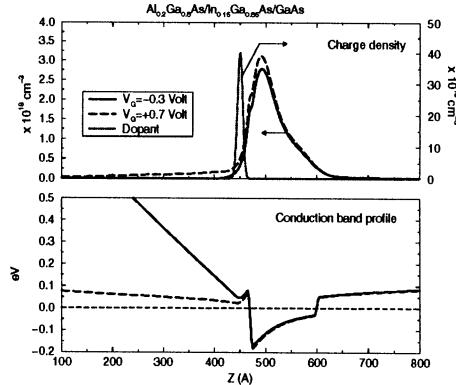


Fig. 9: Charge density (above) and potential profile (below) for a PM HEMT with δ -doping in the AlGaAs layer at two different gate voltages of -0.3 (solid) and 0.7 (dashed line) V

The comparison between Figs. 8 and 10 shows that smaller improvements (of the order of 50 % in the channel density) are achieved by using "pulse" doping , that is a step profile for the donor concentration. When high electron densities are need, for instance for power applications, then the optimal solution if provided by a double δ -doping, with one doping plane in the AlGaAs and one in the GaAs layers , as indicated by Fig. 11.

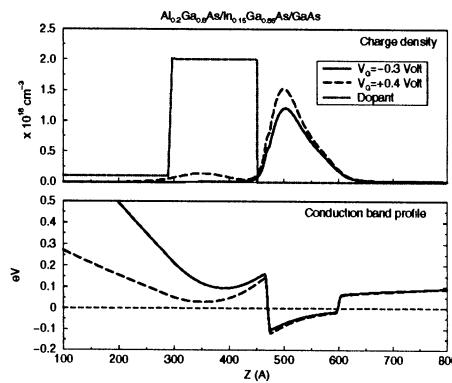


Fig. 10: Charge density (above) and potential profile (below) for a PM HEMT with step doping in the AlGaAs layer at two different gate voltages of -0.3 (solid) and 0.4 (dashed line) V.

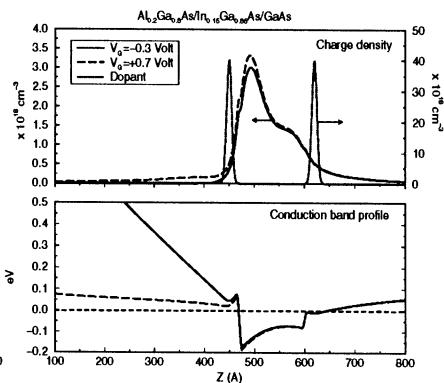


Fig. 11: Charge density (above) and potential profile (below) for a PM HEMT with double δ -doping at two different gate voltages of -0.3 (solid) and 0.7 (dashed line) V.

The implication of the previous results for HEMT performance will be discussed in Sect. 4 with the help of two different transport models.

4.2 DX CENTERS

One of the reason for adopting a δ -doping profile, is connected to the presence of DX centers [16]. Such defects are associated to Si dopant atoms in the AlGaAs layer. They act as deep traps, whose energy is associated to the L minima of the conduction band. Their presence has a series of deleterious effects such as instabilities for low temperature operations due to trapping/detrapping mechanisms, persistent photoconductivity , shift in the threshold voltage. One way to reduce the probability of occupancy of DX centers (and therefore to minimize thier effects) is to lower the Al mole fraction in the AlGaAs layer. This measure has the drawback of lowering the band edge discontinuity, thus reducing the degree of carrier confinement. A better solution is actually to resort to a design of the barrier layer that would the DX level as high as possible with respect to the position of the Fermi level. This is achieved through the use of δ -doping, as clearly indicated in Fig. 12 where the potential profile for uniformly doped structure (left) is compared to a δ -doped one. The dashed line shows the energy position of the DX center, confirming that the linear shape of the potential profile induced by the plane of dopants enhances the distance of the DX center from Fermi level.

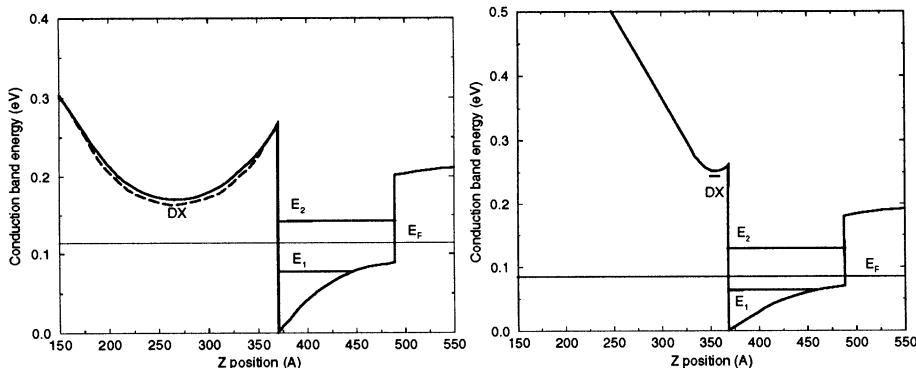


Fig. 12 Conduction band profile for a uniformly doped (left) and a δ -doped (right) HEMT structure. The quantized levels in the well are depicted by solid lines, the energy position of the DX center by dashed lines. Here, the bottom of the InGaAs conduction band is taken as reference for the energy scale.

4.3 LOW AND HIGH FIELD TRANSPORT

The results of the 1D coupled Schrödinger-Poisson model can be used to calculate the basic transport properties on the pseudomorphic structures that constitute the channel of HEMT devices. Here we will show the findings of a Monte Carlo simulation which is performed along the following lines:

1. determine conduction band profile, energy levels and electron wavefunctions for the desired structure;
2. repeat the calculation for the upper valley (L and X) by adopting the appropriate effective masses and energy separations;

3. calculate the electron scattering rates for the mechanisms of interest (here phonons, alloy scattering, ionized impurity), including scattering between different subband of the same and of different minima;

4. perform the simulation to obtain average velocity and energy, together with the population of the various subbands;

5. repeat step 1-4 to account for field induced modification in the subband population.

Non parabolic dispersion can also be included in the calculation [17].

As an application we have studied the transport properties of different heterostructures such as : $\text{Al}_y\text{Ga}_{1-y}\text{As}/\text{Ga}_{1-x}\text{In}_x\text{As}$ pseudomorphic on GaAs substrate ($x < 0.25$) or $\text{Al}_{0.48}\text{In}_{0.52}\text{As}/\text{Ga}_{1-x}\text{In}_x\text{As}$ lattice-matched ($x = 0.53$) or pseudomorphic ($x > 0.53$) on InP substrate.

When a low electric field ($E < 1\text{kV/cm}$) is applied in a direction parallel to the growth plane, the barrier layers are not of crucial importance and the structure behave as an infinite square well in which transport takes place exclusively in the $\text{Ga}_{1-x}\text{In}_x\text{As}$ layer, mainly influenced by the Indium composition and the effect of strain. Figure 13 shows the evolution of low-field mobility as a function of Indium content in the $\text{Ga}_{1-x}\text{In}_x\text{As}$ channel (solid and dotted lines) with a comparison to results obtained without taking into account the effect of strain (dashed line). We can see that strain always degrades the mobility and, for pseudomorphic structures on GaAs substrate (solid line), the mobility is very close to that of a GaAs channel. On the contrary, for pseudomorphic structures on InP substrate (dotted line), the mobility degradation due to strain is not so pronounced and mobility is an increasing function of Indium content with values always larger than those of a GaAs channel.

When the applied electric field is large enough, carriers may populate the different layers via combined intervalley transitions and real space transfer. Both because of the influence of the enhanced scattering rate, and because of the properties of the barrier layers, the mean carrier's velocity changes. Figure 14 shows the velocity-field characteristics for different heterostructures.

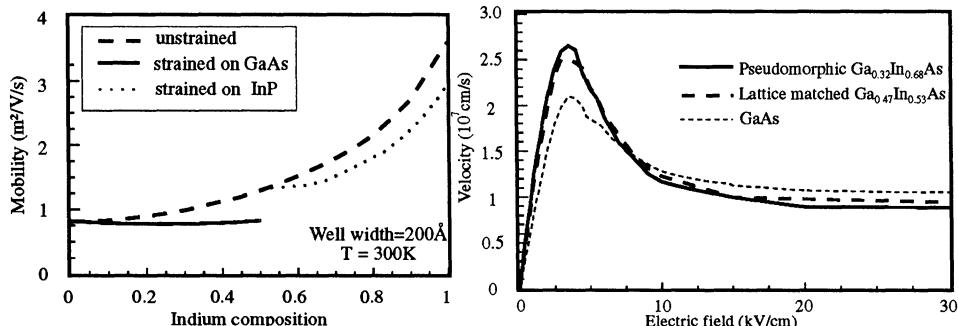


Fig. 13: Room temperature low field mobility as a function of In content for unstrained (dashed), strained on GaAs (solid) and strained on InP (dotted) structures.

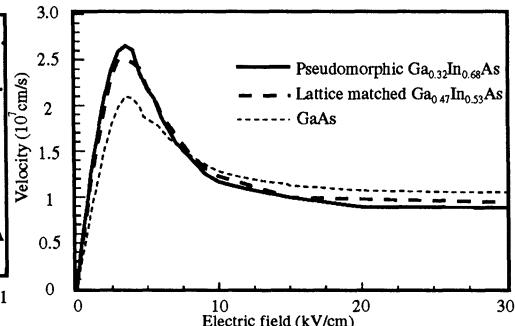


Fig. 14: Room temperature velocity field characteristics for bulk GaAs (dotted), strained (solid) and lattice matched (dashed line) on InP structures.

The pseudomorphic structure on GaAs substrate (not shown) exhibits a velocity-field characteristics quite similar to a pure GaAs layer (dotted line), whereas the results for pseudomorphic on InP structures (solid line) are very close to lattice matched to InP structures (dashed line) giving higher velocities, at least in the peak value, than the GaAs channel.

Our model allows us to determine the percentage of carriers undergoing real space transfer and escaping from the well to populate the wide gap material layers and especially the doped layer ($\text{Al}_{x}\text{Ga}_{1-x}\text{As}$ or $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ depending of the substrate). The results are presented in Fig. 15 as a function of the applied electric field. We can see that, for pseudomorphic on GaAs substrate, the percentage of transferred electrons may exceed 60%, whereas, for InP based heterostructures, this proportion is lower than 20% and decreases with increasing field resulting in a better confinement of carriers in the channel for InP based structures.

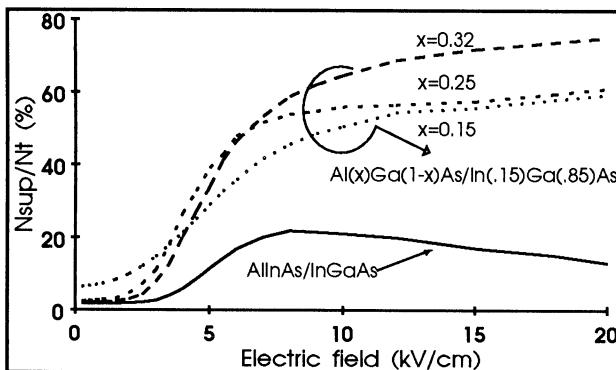


Fig. 15: Fraction of electrons in the upper valleys as a function of the electric field for PM on GaAs (dashed and dotted lines) and lattice match on InP (solid line) structures.

In summary, we can say that there is a competition between the positive effects related to increasing Indium composition of the channel and the negative effects related to strain. For GaAs-based heterostructures, these two effects compensate each other and the InGaAs channel behaves as a GaAs channel. For InP based heterostructures, the effect of strain is less important and the effect of increasing Indium content is dominant. Thus, the transport properties of a InGaAs on InP pseudomorphic structures are better than those of pseudomorphic on GaAs ones. Moreover, the confinement of carriers in the channel is largely superior in InP based structures, making these structures good candidates for high performances HEMT, as will be seen in Sect. 6.

5. Quasi 2D HEMT simulation

One dimensional models provide very useful information both on confinement effect and on transport. Yet, such model cannot account for the actual field distribution from source to drain, which requires, in principle, a fully two-dimensional approach. Unfortunately, 2D methods do not allow the full consideration of quantum features, and are usually too slow to be implemented into CAD tools. In this section we present an alternative approach, which will be defined "quasi two dimensional", that goes beyond the limitation of 1D models maintaining their speed and physical content [18].

5.1 SIMPLIFIED NUMERICAL QUASI 2D MODEL

The general HEMT structure considered for the quasi 2D model is shown in Fig. 16.

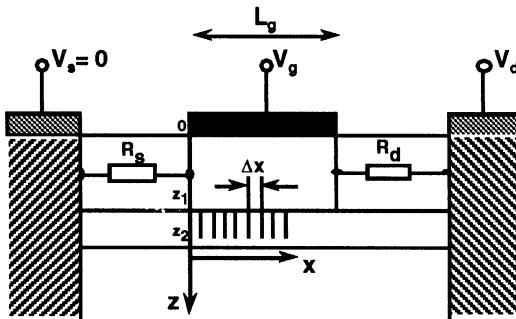


Fig. 16: Schematic representation of the PM HEMT as described by the quasi 2D model.

Following the 1D model based on the coupled Schrödinger and Poisson equation, the channel concentration as a function of the gate potential can be calculated as:

$$n_s(V_G) = \int_{z_1}^{z_2} n(V_G, z) dz \quad (19)$$

When a potential V_d is applied to the drain, then the channel potential varies with x . It is still possible to calculate the carrier density n_s of the 2DEG by dividing the channel length into sections, and considering the actual channel potential $V(x)$ for each of section. In this way $n_s(x) = n_s(V_G - V(x))$.

Our simplified model uses the drift diffusion scheme for the current definition, actually further simplified by neglecting the diffusion term:

$$I_{DS} = -q W v(x) n_s(x) \quad (20)$$

where W is the gate width. Several expressions can be used for the dependence of the velocity $v(x)$ on the longitudinal electric field F . We have use a "silicon like" velocity curve of the type:

$$v(x) = v(F(x)) = -\frac{\mu_0 F(x)}{1 - \frac{F(x)}{F_C}} = \frac{\mu_0 \frac{dV(x)}{dx}}{1 + \frac{dV(x)}{dx} \frac{1}{F_C}} \quad (21)$$

which does not include all details of the actual velocity field curve of InGaAs (in particular it does not display a negative differential resistance region), but allows the implementation of a very fast algorithm. In Eq. 21, μ_0 is the low field mobility and $F_C = v_{sat}/\mu_0$. The saturation velocity v_{sat} is actually treated as a fitting parameter, to obtain reliable values for the current. Electrons move in the channel with ohmic velocity as long as the field is lower than F_C , with saturated velocity when the field is higher.

The IV characteristics for the HEMT are calculated by assuming a value of the drain current I_{DS} , and deriving the corresponding applied voltage through the solution of the

current equation (Eq. 20) once the charge control law is given (in our case directly from the 1D Schroedinger-Poisson model). The analysis can be restricted to the channel region by defining effective potentials at the source end and at the drain end of the gate as respectively: $V_{Se} = V_S + I_{DS} R_s$ and $V_{De} = V_D - I_{DS} R_d$, where R_s and R_d are the source and drain contact resistance's. The potential V_{De} is obtained by direct integration of the electric field provided by Eqs. 20 and 21. In fact, by combining those two equations, we have:

$$I_{DS} = qW \frac{\mu_0 F(x)}{1 - \frac{F(x)}{F_C}} n_s (V_G - V(x)) \quad (22)$$

which can be discretized along the x direction of Fig. 16 as:

$$I_{DS} = qW \frac{\mu_0 F_i}{1 - \frac{F_i}{F_{i-1}}} n_s (V_G - V_{i-1} + F_i h) \quad (23)$$

with F_i and V_i giving the electric field and the potential of the i-th section. Equation 23 is solved for F_i once the V_{i-1} potential is known. Thus, starting from the boundary value V_{Se} and solving forward over all N sections, the unknown potential $V_{De} = V_N$ is found.

It is also possible to account for the parallel conduction in the AlGaAs layer when the parasitic FET is active by solving separately Eq. 23 for the channel and for the barrier layers, and summing the current contributions corresponding to equal values of V_D .

The model described above has been applied to the four different PM HEMT structures described in Sect. 4.1 (Figs. 8-11). The gate bias dependence of channel density and transconductance are shown respectively in Fig. 17 and 18 for the four different doping profiles characterizing the HEMT devices. The results confirm the indications already obtained from the investigation of the potential profile and charge density distributions.

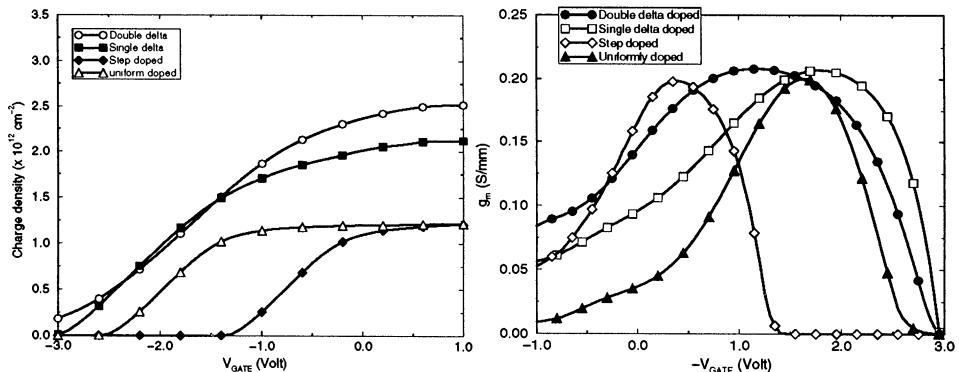


Fig. 17 2DEG charge density as a function of gate bias for 4 different HEMT structures.

Fig. 18: Transconductance as a function of gate bias for the 4 different HEMT structures of Fig. 17.

5.2 QUASI 2D HYDRODINAMICAL MODEL

In this approach the simple current equation given by the DD model (Eq. 20) is substituted by the three moment equations displayed in Sect. 3. The discretization is similar to that of Fig. 16. Poisson equation can be written on the xz plane as:

$$\nabla \bullet (\epsilon(z) F(x, z)) = q (N_D^+(x, z) - n(x, z))$$

or, by separating the two field components, as

$$\frac{d}{dx}(\epsilon(z)F_x(x, z)) + \frac{d}{dz}(\epsilon(z)F_z(x, z)) = q (N_D^+(x, z) - n(x, z)) \quad (24)$$

If we now assume that the dielectric function is the same for the various layers, and that F_x depends only on x, by integrating Eq. 24 between z_1 and z_2 we obtain

$$\frac{dF_x}{dx} \Delta_z + \Delta F_z(x) = -\frac{q}{\epsilon} N(x) \quad (25)$$

where $\Delta z = z_2 - z_1$, $\Delta F_z(x) = F_z(x, z_2) - F_z(x, z_1)$, and $N(x) = \int_{z_1}^{z_2} (n(x, z) - N_D^+(x, z)) dz$.

As for the simplified model, $\Delta F_z(x)$ is obtained from the charge control law:

$$\Delta F_z(x) = -\frac{q}{\epsilon} (n_s(V_G - V_x(x)))$$

Equation 25 is then coupled to the one dimensional (x direction) hydrodynamic equation (Eqs. 13-15).

Typical results of the quasi 2D hydrodynamical model are presented in Figs. 19 and 20 for the same 4 HEMT structures considered with the simplified model. Qualitatively, the transconductances of the various devices agree with the results of Fig. 18, although the more physical hydrodynamical model predicts a better performance of the HEMT with δ -doping. One major difference can be seen in the actual value of the peak transconductance, which is about twice as large as that predicted by the simple model.

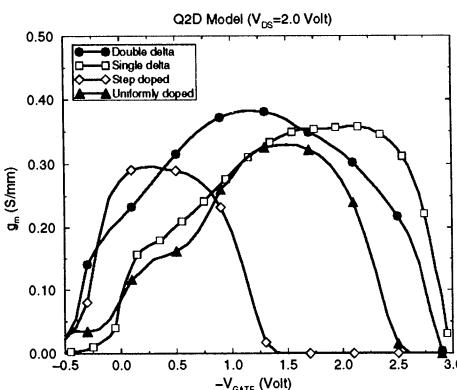


Fig. 19: Gate bias dependence of the transconductance for 4 different PM HEMT structures

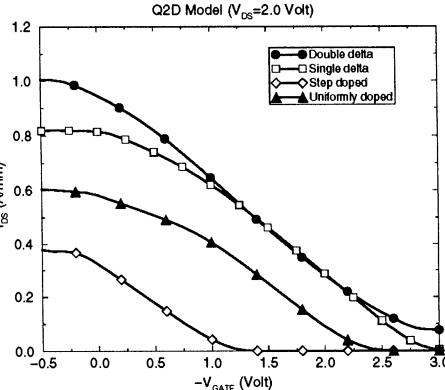


Fig. 20: Transfer characteristics of the 4 different PM HEMT structures

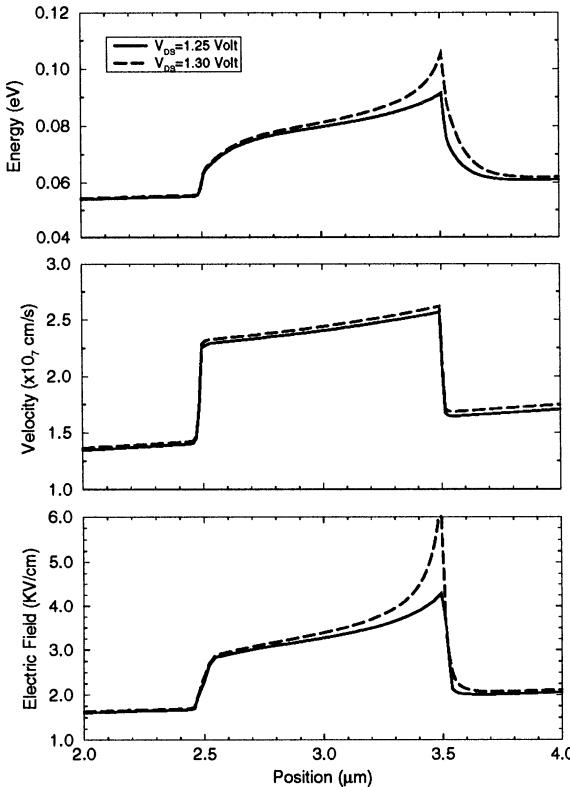


Fig. 21: 1D plot (from source to drain) of the longitudinal electric field (lower picture), average drift velocity (middle picture), and average electron energy (upper picture) for an uniformly doped PM HEMT simulated by the quasi 2D hydrodynamical model

The differences between the simplified approach (based on a reduced drift diffusion scheme) and the hydrodynamical one have been pointed out in Sect. 3. Figure 21, which shows some details of the latter model, confirms that analysis. Indeed, for the high electric field distributions in the HEMT channel at sufficiently high drain bias, the channel electrons exhibit heating and velocity overshoot. The effect would actually be much more pronounced at higher drain voltages, and for shorter gate lengths. An approach based on the DD model could take such features into account only by ad hoc unjustified modifications of the basic algorithms, while here they enter naturally since all main physical effects are included into the moment equations.

6. Two dimensional models

We have so far presented approaches that aim at a high speed by making approximation at the physical and/or at the geometrical level. This final section will illustrate the type of analysis which is possible when a fully two dimensional approach is adopted combined with a sophisticated transport model. As a case study we will take a Monte Carlo self consistent simulation. Examples of simulations based on 2D Drift Diffusion or Hydrodynamical methods can be found in [10,13,19-21]

6.1 MONTE CARLO SIMULATION OF PSEUDOMORPHIC HEMTS

The structure which has been investigated is shown in Fig. 22. It includes all features which characterize state-of-the-art PM HEMT, an indication that complex approaches such as the Monte Carlo can deal with realistic devices. Poisson equation is solved on a non uniform grid on the area shown in the figure. Two δ -doping planes are considered, the first with a donor concentration of $4 \times 10^{12} \text{ cm}^{-2}$ in the AlGaAs top layer, the second with $1 \times 10^{12} \text{ cm}^{-2}$ in a second confining AlGaAs bottom layer. The gate length is 180 nm. It should be pointed out that the simulation does not include quantization effects, that is carrier dynamics is taken as classical in the InGaAs channel. Confinement is provided by the presence of the heterointerfaces.

Typical 2D plots for a drain-to-source bias of 2 V are shown in Figs. 23-26. The potential distribution (Fig. 23) reveals that a high field region exists in the channel between the end of the gate and the beginning of the cap layer. Entering this regions electrons become extremely hot, reaching average kinetic energies of few tenths of eV, as indicated by the energy map in Fig. 24. Electron heating results in a reduced confinement within the channel, both because many of the electrons have sufficient energy to surmount the confining barrier, both because many of them transfer to the satellite L valley, which has a much smaller band edge discontinuity than the Γ one. The two effects are clearly visible in Figs. 25 and 26, which show respectively the total and the L valley electron concentrations. Because of the hot-carrier induced real space transfer, current flows both through the bottom AlGaAs layer (and partially through the GaAs substrate) as well as through the top AlGaAs barrier, which constitutes actually the access path to the drain cap region.

The actual value of the average electric field along the channel, and the corresponding average drift velocity and energy, are presented in Fig. 27. Due to the very short gate length, the field reaches a peak value of 200 kV/cm, originating the strong electron heating already shown in Fig. 24 (and confirmed in the right picture of Fig. 27), and a remarkable velocity overshoot, which is in turn responsible for the excellent microwave performance of such devices.

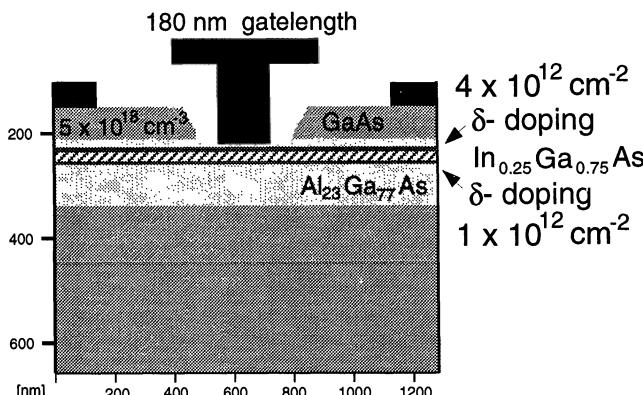


Fig. 22 Structure of the simulated PM HEMT showing the double δ -doping layers used above and below the InGaAs channel, the GaAs cap region, and the submicron T-shaped gate

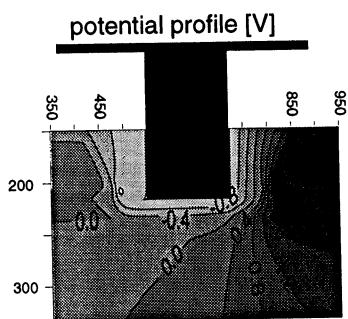


Fig. 23 Two dimensional plot of the equipotential line for the simulated PM HEMT at a drain bias of 2 V.

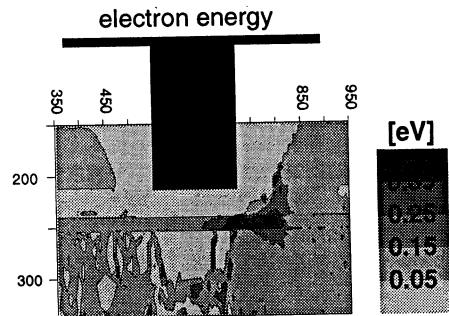


Fig. 24 Two dimensional representation of the electron average energy for the simulated PM HEMT at a drain bias of 2 V. The darkest region corresponds to an energy of 0.35 eV.

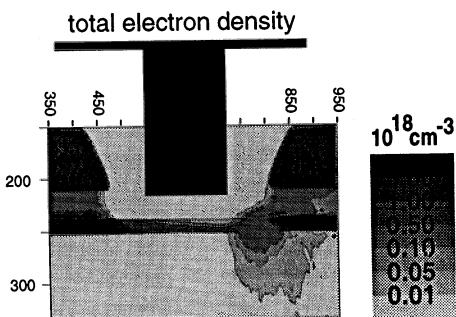


Fig. 25 Two dimensional representation of the total electron concentration for the simulated PM HEMT at a drain bias of 2 V. The darkest region corresponds to a density of $5 \times 10^{18} \text{ cm}^{-3}$.

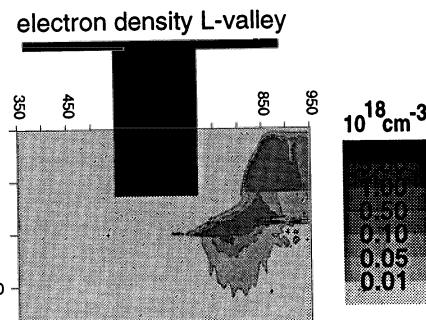


Fig. 26 Two dimensional representation of the L-valley electron concentration for the simulated PM HEMT at a drain bias of 2 V. The darkest region corresponds to a density of $5 \times 10^{18} \text{ cm}^{-3}$.

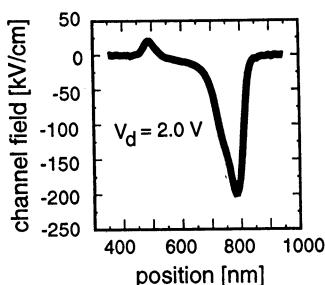


Fig. 27 One dimensional plot (along the channel direction) of the electric field (left), average drift velocity (center), and average energy (right) for two different drain voltages, respectively 0.5 V (dashed lines) and 2 V (solid lines).

The Monte Carlo simulation provide one of the best tools for predicting the performance of novel devices. As an example, we consider the InP based HEMT structure shown in Fig. 28. Two alternatives are compared: an InGaAs channel with Indium composition : of 53% for the lattice matched HEMT and one at 75% for the PM HEMT.

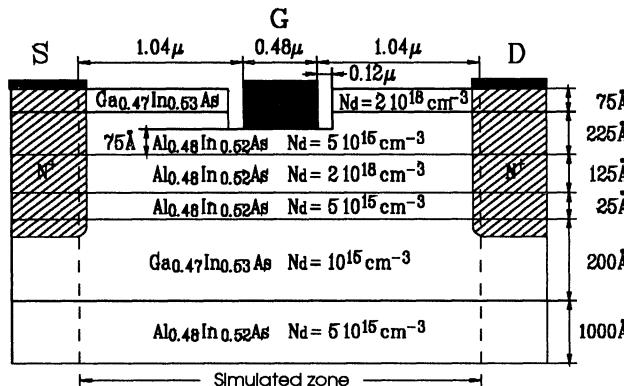


Fig. 27 Structure of the InP based simulated HEMT.

As predicted from the study of electron transport properties, one should expect for the InP based PMHEMT, a better confinement of electrons in the channel and a slight improvement of transport properties. This would lead to increase transconductance and cut-off frequency values. The evolutions of transconductance and cut-off frequency versus gate voltage are presented in Figs. 28. As expected, an improvement of almost 10% of transconductance and cut-off frequency is obtained for the PMHEMT as compared to the LMHEMT. One should notice the outstanding values predicted for InP-based HEMTs, namely around 1000 mS/mm for the transconductance and 130 GHz for the frequency.

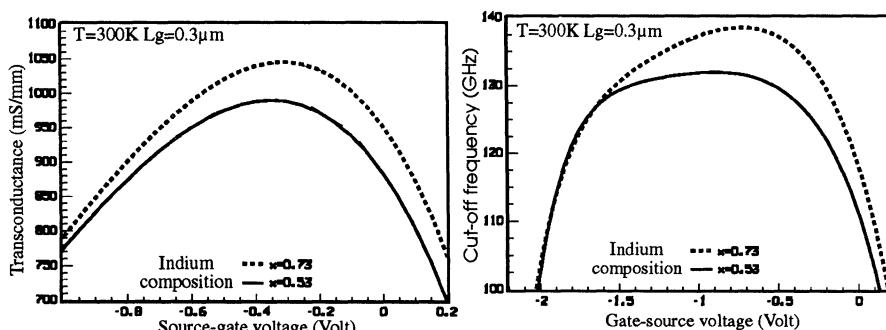


Fig. 28 Gate bias dependence of transconductance (left) and cut-off frequency (right) for a lattice matched (solid line) and a pseudomorphic (dotted line) InP-based HEMT.

Conclusions

We have presented a review of numerical tools for semiconductor device modeling. The basic features of the various methods have been discussed in relation to the physical model they implement. We have presented in detail a one dimensional model based on the solution of the coupled Schroedinger and Poisson equation, and we have shown how such approach can be extended to address the problem of transport in HEMTs. Furthermore, we have shown how sophisticated Monte Carlo simulations, coupled to a two dimensional Poisson solver, can give a very detailed description of the HEMT operation and predict the performance of novel structures.

This work was partially supported by the ESPRIT Working Group "ELTRASIN" and by the Siemens project ZFE II.

References:

- [1] H. Morkoc, H. Unlu, and G. Ji, *Principles and Technology of HEMTs*, Wiley & Sons, New York (1991)
- [2] L. D. Nguyen, L. E. Larson, and U. K. Mishra, Proceedings IEEE, **80**, 494 (1992)
- [3] D. K. Ferry and C. Jacoboni, *Quantum Transport in Semiconductors*, Plenum Press, New York (1992)
- [4] P.M. Solomon and H. Morkoc, IEEE Trans. El. Dev. **ED31**, 1015 (1984)
- [5] D. Potter, *Computational Physics*, Wiley, London (1973)
- [6] D.L. Scharfetter and H.K. Gummel, IEEE Trans. El. Dev. **16**, 64 (1969)
- [7] L. Reggiani (ed.), *Hot Electron Transport in Semiconductors*, Springer Verlag, Heilderberg (1985)
- [8] Y.K. Feng and A. Hintz, IEEE Trans. El. Dev. **ED35**, 1419 (1988)
- [9] S. Selberherr *Analysis and Simulation of Semiconductor Devices* Springer Verlag, Wien (1984)
- [10] C. M. Snowden, *Introduction to Semiconductor Device Modelling*, World Scientific, Singapore (1986)
- [11] P. Lugli, Microelectronic Engineering **19**, 275 (1992)
- [12] C. Jacoboni and P. Lugli *The Monte Carlo Method for Semiconductor Device Simulation*, Springer-Verlag , Wien (1989)
- [13] C. M. Snowden and R.E. Miles (eds.), *Compound Semiconductor Device Modelling*, Springer Verlag, London (1993)
- [14] W.L. Engl (ed.), *Process and Device Modeling*, North Holland, Amsterdam (1986)
- [15] Y. Ando and T. Itoh, IEEE Trans. El. Dev. **ED35**, 2295 (1988)
- [16] T.J. Drummond, R.J. Fisher, W.F. Kopp, H. Morkoc, K. Lee, and M. S. Shur, IEEE Trans. El. Dev. **ED30**, 1806 (1983)
- [17] J.L. Thobel, L. Baudry, F. Dessenne, M. Charef, J. Appl. Phys. **74**, 6274 (1993)
- [18] A. Cappy, A. Vanoverschelde, M. Shortgen, C. Versnayen, and G. Salmer, IEEE Trans. El. Dev. **ED32**, 2787 (1985)
- [19] W. Haensch, *The Drift Diffusion Equation and its Application in MOSFET Modeling*, Springer-Verlag , Wien (1991)
- [20] T. Shawki, G. Salmer, and O.L. El-Sayed, IEEE Trans. El. Dev. **ED37**, 21 (1990)
- [21] H. Chau, D. Pavlidis, and K. Tomizawa, IEEE Trans. El. Dev. **ED38**, 213 (1991)

MMIC CIRCUIT DESIGN

Sander Weinreb
Lockheed Martin Laboratories
1450 South Rolling Road
Baltimore, MD 21227

Abstract

The foundations for the circuit design of monolithic integrated circuits (MMICs) are briefly described in this article. It is intended for engineers with a knowledge of circuits and transistors but not necessarily microwave techniques. The topics that are introduced are: S-parameters, gain and matching, the field effect transistor (FET) equivalent circuit, transistor performance characterization, transmission lines for MMICs, matching networks, and microwave noise models. The article concludes with a description of a low-noise amplifier design.

1. S-Parameters - Definition and Examples

A starting point for MMIC circuit design is the definition and understanding of the variables used to describe a linear two-port network. At low frequencies, the commonly used variables are voltages (V_i), currents (I_k) and impedances (Z_{ik}) for ports $i,k = 1$ or 2 . At microwave frequencies the more convenient variables are the ingoing waves (A_i), outgoing waves (B_k), and scattering parameters ($S_{ik} = B_i/A_k$) of the network. The S-parameters are preferred over Z-parameters because: a) they are directly measurable with a microwave network analyzer, b) the active network tends to be unstable (it oscillates) when the open circuits required for Z-parameter measurement are applied, and c) transmission lines are more simply described with S-parameters. The defining equations for S-parameters are given in Figure 1 along with analogous expressions for Z-parameters and the relationships of voltages and currents to wave variables. Alternative matrix expressions for the variables at port 1 of a network in terms of the variables at port 2 are also given along with the expression for the power entering a port in terms of either type of variable. Equations that relate the matrix parameters can be easily derived; i.e., if the S-parameters are known, the Z-parameters can be computed. Examples of S-matrices for some common circuit elements are shown in Figure 2.

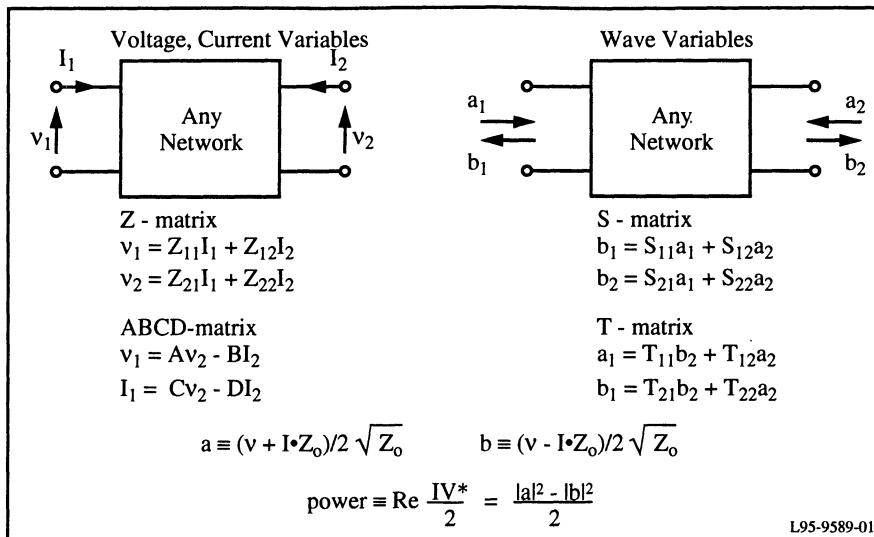


Figure 1. Key equations for description of a two-port network.

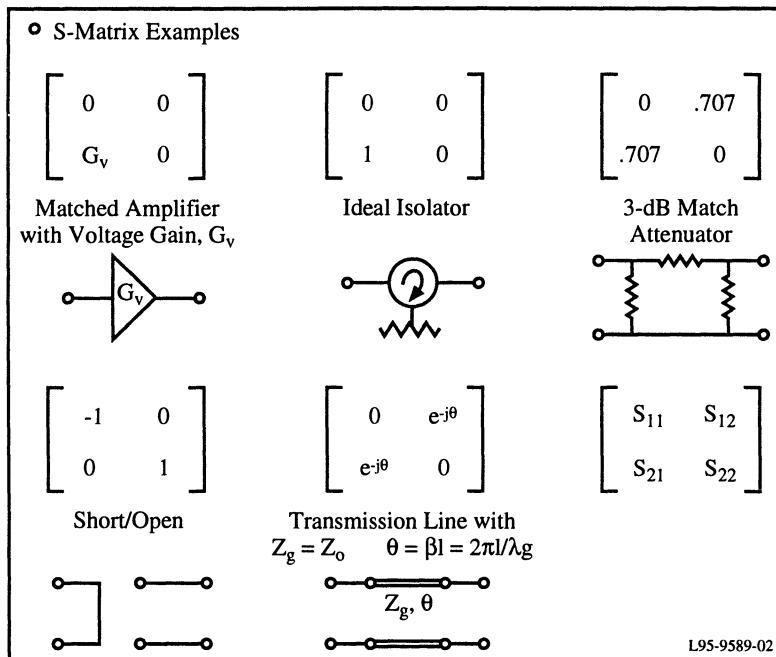


Figure 2. Examples of S-matrices for common circuit elements.

2. Gain and Matching

Important relations giving the transducer gain, G_T , (power delivered to a load divided by power available from a source) in terms of S-parameters and source and load reflection coefficients, Γ_s and Γ_L , are given in Figure 3. The network input and output reflections coefficients, Γ_{in} and Γ_{out} , are also given. Note that all of the expressions greatly simplify when Γ_s and $\Gamma_L = 0$. Since Γ_{in} is a function of Γ_L and Γ_{out} is a function of Γ_s it is not a simple matter to match input and output. These solutions for matched load and source, Γ_{ML} and Γ_{MS} , are called simultaneous bilateral match and are given in the Avantek Primer II [1] p. 5 or Gonzalez [2] p. 113.

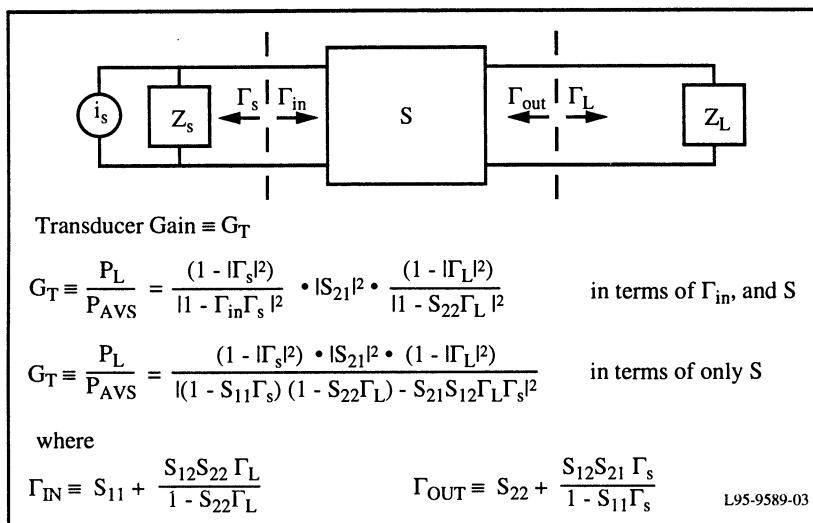


Figure 3. Power gain, G_T , of a network in terms of S parameters.

An active circuit may oscillate by producing negative resistance having magnitude > external positive resistance, as shown in Figure 4.

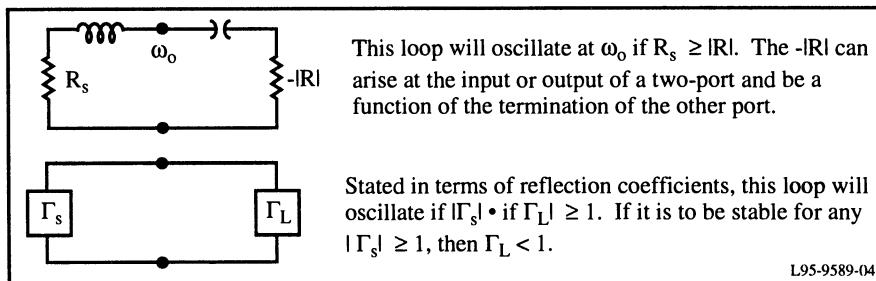


Figure 4. Active circuit oscillation.

A two-port network is stable if $|\Gamma_{in}|$ and $|\Gamma_{out}|$ are ≤ 1 for any passive load at the opposite port. Necessary and sufficient conditions for stability are:

$$\text{Stability Factor } K = \frac{|S_{11}|^2 - |S_{22}|^2 + D^2}{2|S_{12}S_{21}|} > 1 \text{ and } D < 1 \quad (1)$$

where $D = |S_{11}S_{22} - S_{12}S_{21}|$.

L95-9589-18

Some references for the topic of stability are Rollet[3] and HP Application Note #154 [4].

3. Field Effect Transistor (FET) Equivalent Circuit

A common MMIC design procedure is to determine an equivalent circuit from S-parameter measurements of the transistor at frequencies of DC to 40 GHz. The equivalent circuit elements can be identified with specific physical regions and mechanisms occurring within the transistor and thus are a help in comparing or improving transistors. For MMIC design, the equivalent circuit can be analyzed with a computer-aided design (CAD) circuit simulator along with the passive portions of the circuit. Note that the equivalent circuit allows interpolation and extrapolation with regard to frequency; the FET can be measured at one frequency and analyzed at another with generally good accuracy. An example of an equivalent circuit is shown in Figure 5.

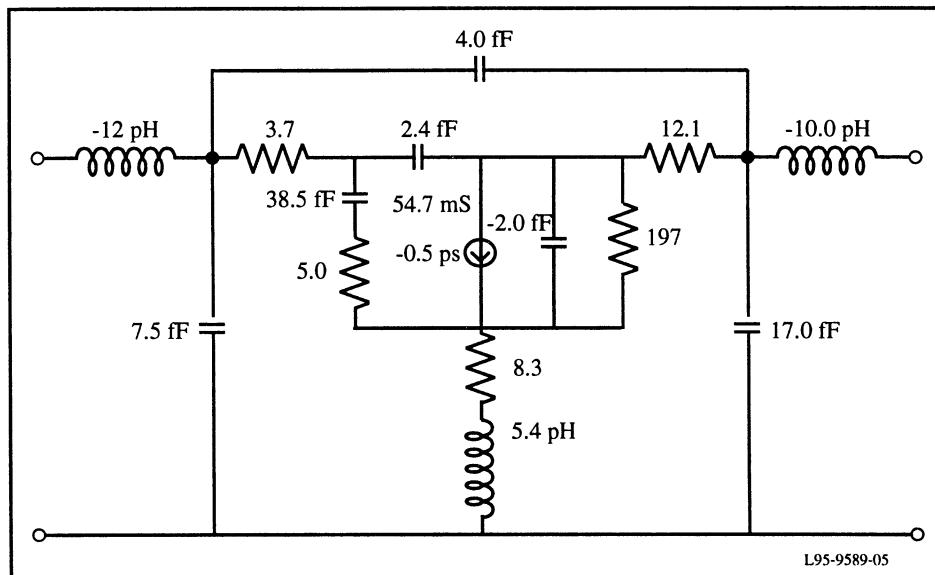


Figure 5. Equivalent circuit model of a 0.1 x 50- μm HEMT.

4. Transistor Performance Characterization

Once a transistor equivalent circuit has been determined, many variables important to the MMIC design can be computed. A noise model (to be discussed later) allows the noise properties to be computed.

The key gain, stability, and noise parameters of a typical $0.1 \times 50\text{-}\mu\text{m}$ gate transistor are plotted as a function of frequency in Figure 6. At frequencies where the transistor is unstable, with stability factor $K < 1$, the maximum available gain is infinite and a quantity known as the maximum stable gain (MSG) is plotted. At frequencies where $K > 1$, the maximum available gain (MAG) is plotted. The minimum noise figure (NF), its associated gain, and the NF of an infinite cascade of the circuit are also plotted.

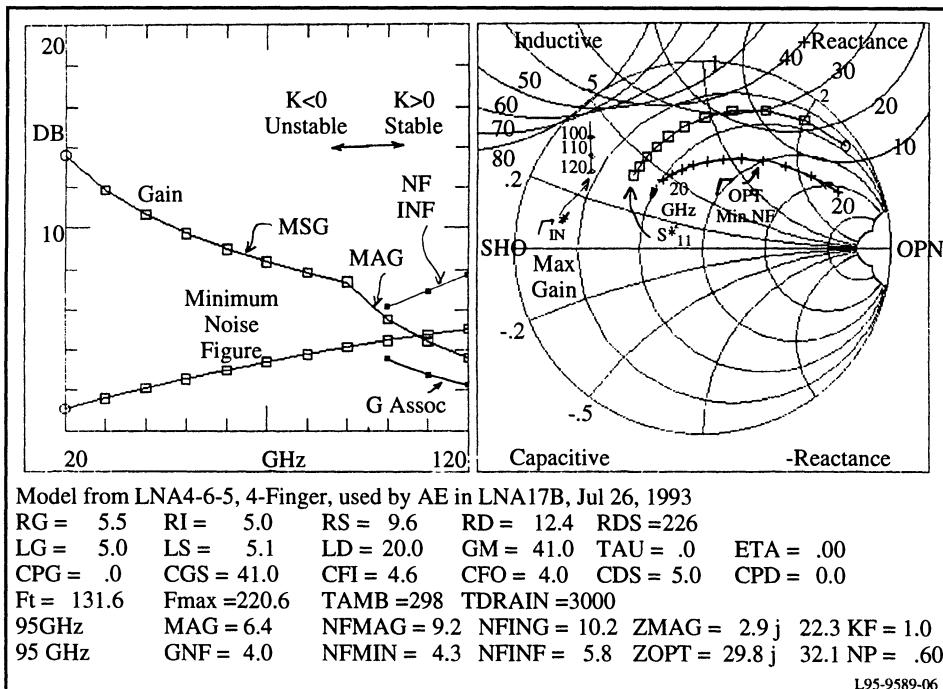


Figure 6. Performance parameters (gain, stability, NF, optimum source and load impedances) computed from equivalent circuit model.

The Smith chart plot on the right in Figure 6 gives the following information:

- (a) The conjugate match source reflection coefficient, Γ_{in} , which delivers the MAG at frequencies where $K > 1$

- (b) The source reflection coefficient, S_{11}^* , which would match the transistor terminated with a 50-ohm output impedance
- (c) The source reflection coefficient Γ_{OPT} , which produces the minimum noise for the transistor
- (d) Stability circles computed at 10-GHz intervals from 10 to 80 GHz. A source reflection coefficient within a circle will produce a negative resistance at the output of the transistors, and hence oscillation could occur with some values of load impedance. For source reflection coefficients outside the circles, oscillation cannot occur for any values of passive load impedance.

5. Transmission Lines for MMICs

Microstrip and coplanar waveguide (CPW) transmission lines are described in Figure 7. Most MMICs utilize microstrip, but grounded coplanar waveguide has been utilized for millimeter-waves because lower inductance grounds can be realized. The key parameters of a transmission line are the line width as a function of the desired characteristic impedance, the effective dielectric constant or propagation velocity, and the loss. Typical values for microstrip and grounded CPW are given in Figures 8 and 9, respectively.

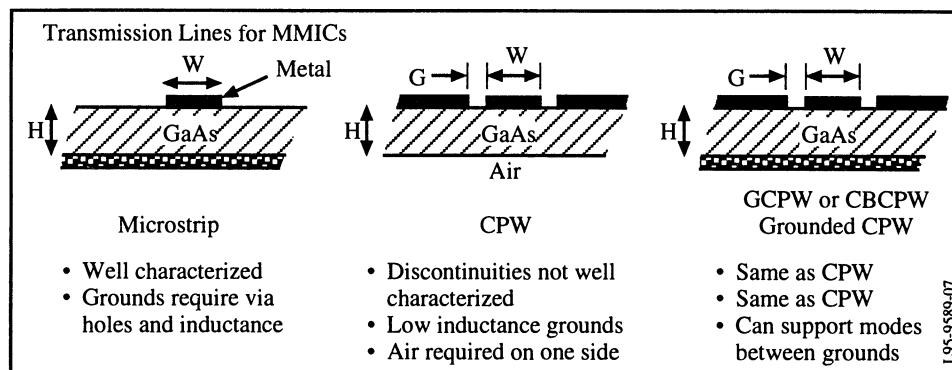


Figure 7. Transmission lines for MMICs

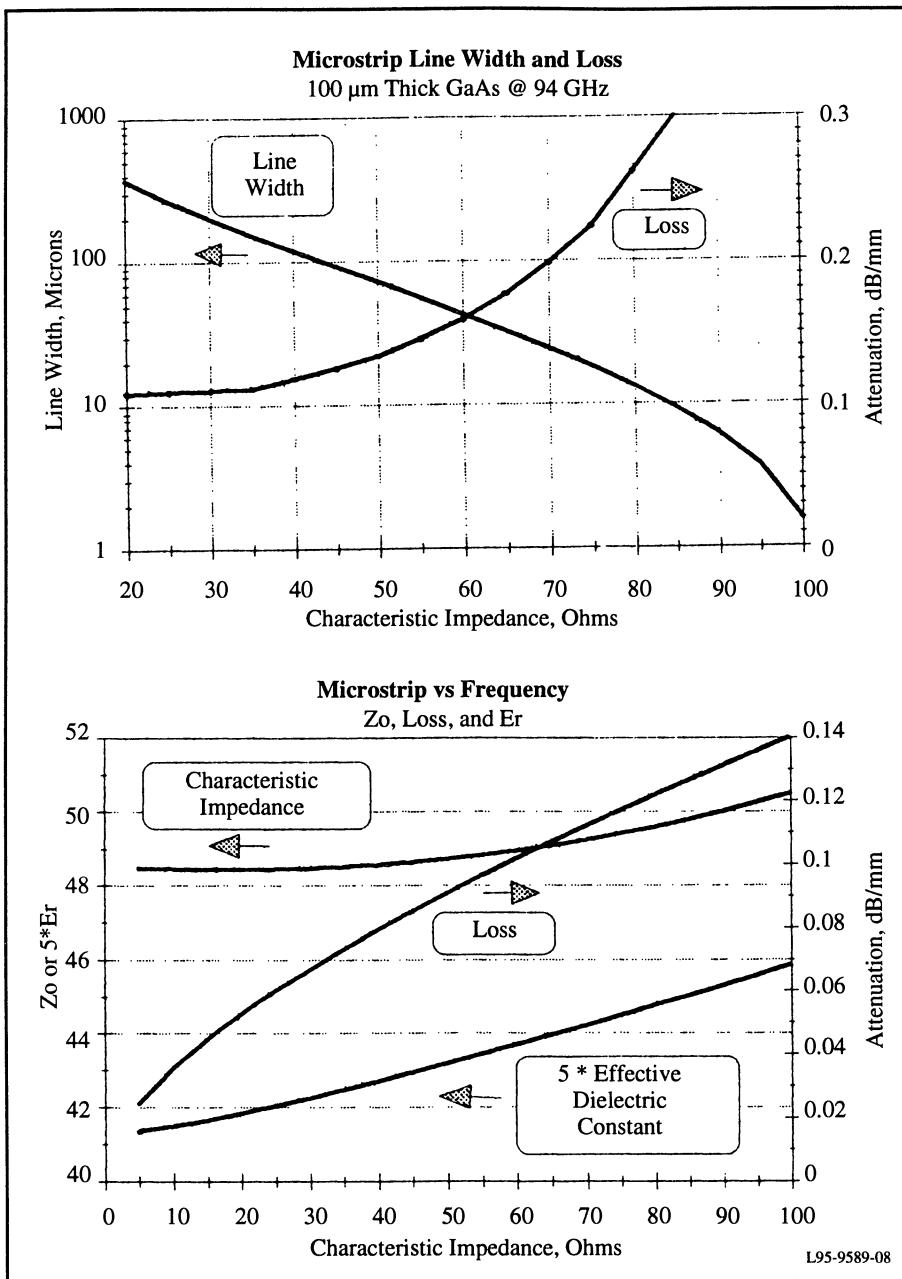


Figure 8. Key parameters of a microstrip transmission line on 0.1-mm-thick GaAs.

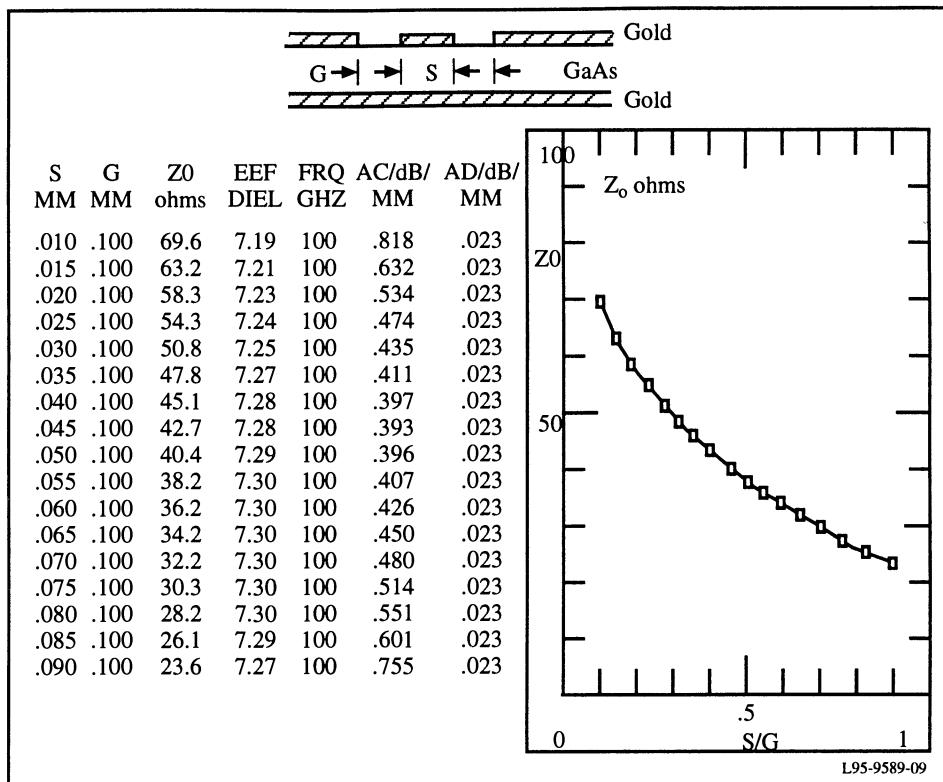


Figure 9. Characteristics of GCPW on 0.1-mm-thick GaAs. See References [5] and [6].

Transmission lines can be utilized for impedance transformation. A one-quarter wavelength line with characteristic impedance, Z_1 , transforms a load impedance, Z_L , to Z_1^2/Z_L . These concepts and examples are shown in Figure 10.

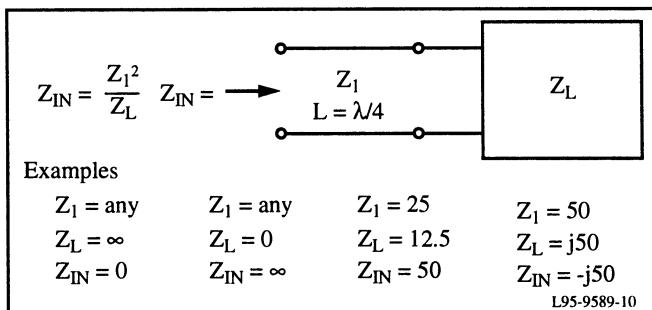


Figure 10. Quarter-wave transformers.

6. Matching Networks

A major aspect of MMIC design is the realization of input, interstage and output impedance matching networks. The concept is illustrated in Figure 11. Some key properties of these networks are:

- (a) The bandwidth is limited by load $Q = X/R$ but the network can make it narrower. High Q also increases network loss.
- (b) The resistive transformation can, in principle, be performed over a wide bandwidth (for example, with a tapered transmission line) but the circuit size and loss increase. It is difficult to match resistances < 5 ohms.
- (c) If the time delay is not specified, there are just two parameters (i.e., N and X) of the matching network and there are many two-element networks that can be utilized (i.e., a two-stub tuner).

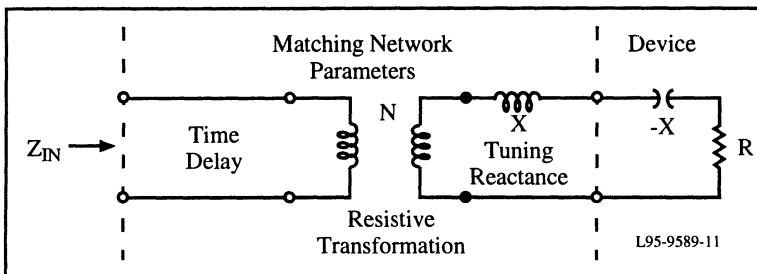


Figure 11. Simplified model of an impedance matching network used to transform a device input impedance, $R - jX$, to a desired input impedance, Z_{in} .

CAD programs analyze and optimize matching networks, but the designer must choose the circuit topology. Some examples of network topologies are shown in Figure 12.

Factors to consider in the selection of matching network topology are:

- (a) Size, shape, cost per unit area
- (b) Bandwidth
- (c) Low-frequency stability
- (d) DC bias integration
- (e) Range of realizable elements
- (f) Accuracy of discontinuity models
- (g) Loss

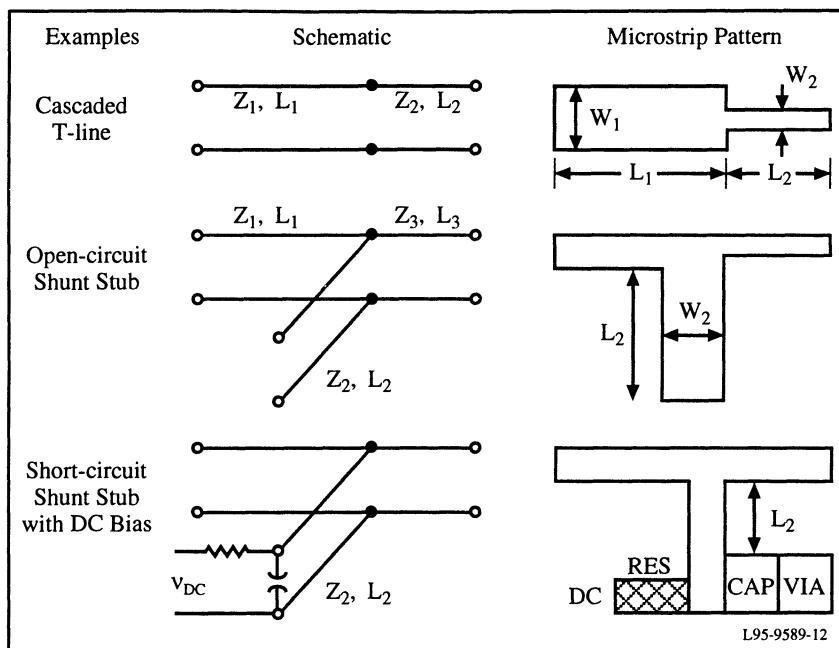


Figure 12. Schematic and microstrip patterns for three common examples of matching networks.

7.0 Microwave Noise Model and Parameters

The noise in a microwave component is usually modeled by attributing all of the noise in the device to a noise temperature, T_n , of the generator impedance, Z_s , as illustrated in Figure 13. The noise in the network is represented by T_n , which is the physical temperature of source resistance that produces thermal noise equal to the network noise.

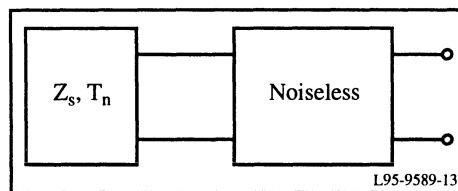


Figure 13. Noise model.

There is a value of generator impedance, Z_s , which maximizes the signal-to-noise ratio of the network. This optimum value of Z_s is Z_{op} . T_n has the following dependence on Z_s and four network noise parameters, T_{min} , $Z_{op} = R_{op} + jX_{op}$, and N ; in terms of impedance, admittance, and reflective coefficient, respectively,

$$T_m = T_{min} + N \cdot T_o \frac{|Z_s - Z_{op}|^2}{R_s R_{op}}, \quad (2)$$

$$T_m = T_{min} + N \cdot T_o \frac{|Y_s - Y_{op}|^2}{G_s G_{op}} \quad Y_s = Z_s^{-1} \quad Y_{op} = Z_{op}^{-1}, \quad (3)$$

$$T_m = T_{min} + 4 \cdot N \cdot \frac{T_o |\Gamma_s - \Gamma_{op}|^2}{(1 - |\Gamma_s|^2)(1 - |\Gamma_{op}|^2)} \quad \Gamma_{op} = \frac{Z_{op} - Z_o}{Z_{op} + Z_o}. \quad (4)$$

L95-9589-19

The noise parameters have the important properties illustrated in Figure 14. T_{min} and N are invariant to a lossless transformation and $T'_{min} \leq 4 N T_o$. The relations to other commonly used noise parameters are: *noise resistance* = $R_n = N/G_{op}$, and *noise conductance* - $G_n = N/R_{op}$.

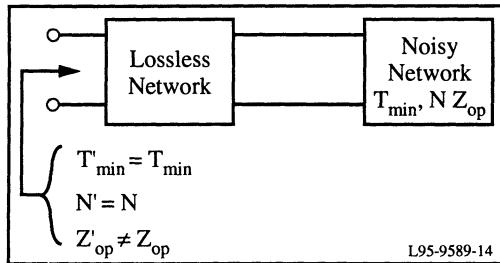


Figure 14. Transformation of noise parameters by a lossless network. See Reference [7].

A simplified noise model of an FET (Pospieszalski [] and Gupta []) allows all of the noise parameters to be computed as a function of frequency and ambient temperature given the small signal equivalent circuit (from S-parameter measurements) and one noise figure measurement at any frequency. The model assumes that all resistances in the equivalent circuit produce thermal noise determined by ambient temperature except the drain resistance, R_{DS} , which has noise temperature, T_{DRAIN} . The model is illustrated in Figure 15. T_{DRAIN} is strongly bias dependent, is independent of frequency, is of the order of 1500K at optimum noise bias, and can be determined from one noise measurement.

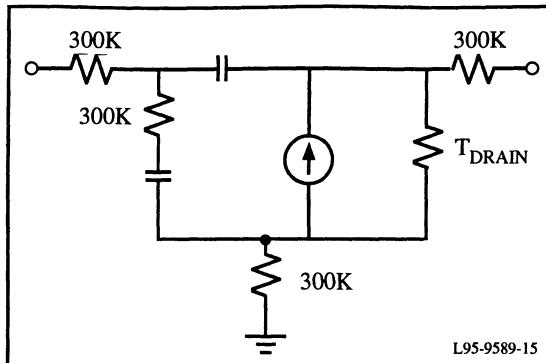


Figure 15. Simplified noise model of a FET.

8. Low-Noise-Amplifier (LNA) Design

The input matching network of an LNA transforms the generator impedance (usually 50 ohms) to a source resistance, R_s , which can be chosen either for minimum noise, $R_s = R_{op}$, or maximum gain, $R_s = R_{in}$, as illustrated in Figure 16.

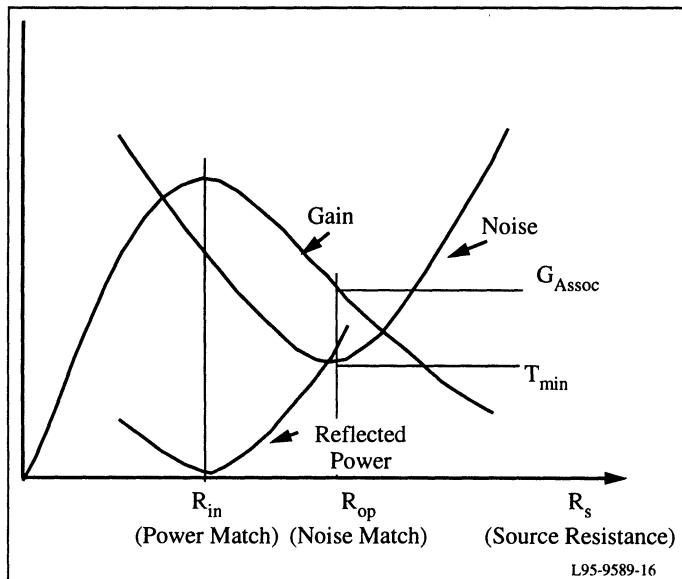


Figure 16. Behavior of noise figure, gain, and reflected input power of an amplifier as a function of impedance driving the amplifier.

The noise match and gain or power match usually occur at different values of source resistance. Three methods of making these impedances coincide are: a) input isolation, b) balanced amplification, and c) feedback (such as increased FET source inductance).

A complete four-stage MMIC LNA is shown in Figure 17. The MMIC utilizes grounded CPW transmission lines, thin-film resistors and capacitors, and via-holes to suppress dielectric waveguide modes in the GaAs substrate. This amplifier has a noise figure of approximately 5 dB and a gain of 23 dB at 94 GHz.

MMIC technology can be utilized to provide many functions as shown in Table 1, which lists the functions and key specifications of millimeter-wave MMICs designed at Lockheed Martin Laboratories.

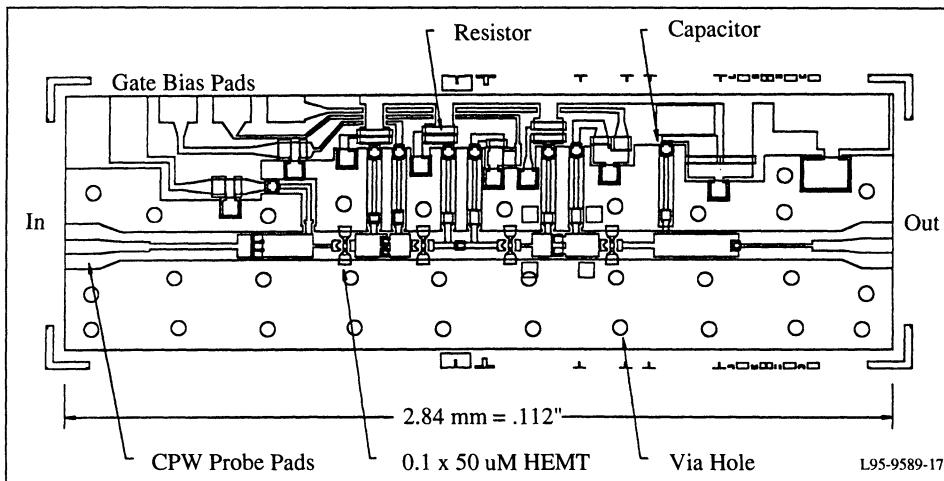


Figure 17. Example of a complete four-stage MMIC low-noise amplifier.

TABLE 1. Lockheed Martin Laboratories' Portfolio of Millimeter-wave MMIC Designs

Design	Frequency	Function	Key Specifications
LNAA17-1	75-95	LNA	20-30 dB gain, 4-6 dB NF @ 94 GHz
LNAA17-1B	90-98	LNA	15-20 dB gain, 4.5-5.5 dB NF@ 94 GHz
94PA1	90-97	PA	+16 dBm and 10 dB gain @ 94 GHz
60PA2	60	PA	+20 dBm and 15 dB gain @ 66 GHz
PALNA1	60	PA and LNA	PA, 10 dBm & 12 dB gain @ 60 GHz LNA, 5 dB NF & 20 dB gain @ 60 GHz
30PA1	27-35	PA	+20 dBm and 12 dB gain @ 31 GHz
30PA2	27-35	PA	+27 dBm and 7 dB gain @ 31 GHz
94MXRH1	80-100	Balanced mixer	10 dB conversion loss @ 94 GHz
94MXR4	92-96	Image reject mixer	13 dB conversion loss @ 94 GHz, 10 dB image rejection
W2MXR1	80-100	2nd harmonic mixer	10 dB conversion loss @ 94 GHz
200MXR1	200-210	2nd harmonic mixer	13 dB conversion loss @ 205 GHz
W3MXR1	80-100	3rd harmonic mixer	20 dB conversion loss @ 94 GHz
94X320 109-275	92-94	Tripler	+15 dBm out with +27 dBm in @ 93 GHz
WDET1	80-100	Detector	4 μ A/ μ W sensitivity @ 94 GHz
WNG1	80-100	Noise generator	ENR=17 dB, 15,000K @ 94 GHz
31FAS1	26-40	Phase-shifter	100° phase-shift range
94PS4A	75-110	Phase-shifter	360° phase-shift range, 6-11 dB insertion loss

9. References

1. "High Frequency Transistor Primer, Part II, Noise and S-parameter Characterization," Document ATP-1047/R-10-86, Avantek (Hewlett Packard) Company Technical Report.
2. G. Gonzalez, *Microwave Transistor Amplifiers*, 1984.
3. J. Rollet, "Stability and Power Gain Invariants of Linear Two-Ports," IEEE PGCT-9, March, 1962, p. 29.
4. "S-Parameter Design," Hewlett Packard Application Note 154, Document 5952-1087, April, 1972.
5. Ghione and Naldi, "Microstrip ...," IEEE MTT35-3, March, 1987, pp. 260-267.
6. Gopinath, "Microstrip ...," IEEE MTT30-7, July, 1982, pp. 1101-1104.
7. Lange, "Noise," IEEE JSSC-2, June, 1967, p. 37.

Accurate Active Device Models for Computer Aided Design of MMICs

Techniques for Direct Parameter Extraction of pHEMTs and MESFETs

Reza Tayrani
Compact Software
Paterson, NJ, USA

Introduction

State-of-the art computer-aided design methods for active microwave circuits rely heavily on the models of active devices. In the traditional approach in obtaining the equivalent circuit parameters (ECPs), the model parameters are optimized to closely fit the measured small signal scattering (S) parameters. The process of optimization is rather tedious and generally provide erroneous ECP values which lack physical significance of the circuit elements.

In this review article, several enhanced techniques for direct parameter extraction of MESFETs and HEMTs, for accurate generation of device equivalent circuit parameters (ECPs) up to millimeter-wave frequencies are explained. Such extraction tools are necessary for obtaining bias and temperature dependent ECPs leading to verifications and improvements of existing non-linear models. Figure 1 shows an example of our non-linear extractor design methodology for a MESFET/HEMT.

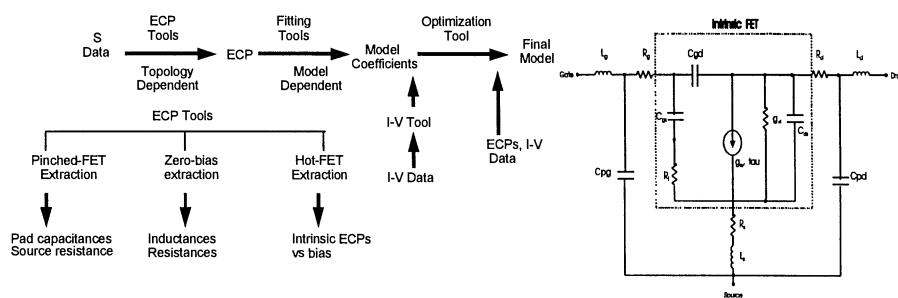


Figure 1. Non-linear extractor methodology and device model topology used for extraction.

In Section 1.1 we explain the development of a linear extractor program (EXT) for MESFETs and HEMTs. Also discussed is a novel and accurate analytical direct extraction method for the determination of MESFET and HEMT parasitic elements, which differs from the commonly used “cold-FET” technique by avoiding the forward biasing of the Schottky gate junction. In section 1.2, the methodology used for non-linear parameter extraction of MESFETs and HEMTs is described. Section 1.3 discusses the comparative study done on the performance of the three popular non-linear models (TOM, Materka, and Angelov).

1.1. MESFET and HEMT Extraction

The extraction methodology for MESFETs and HEMTs is based on the equivalent circuit shown in Figure 2 from which, the device equivalent circuit elements can be divided into two categories:

- Intrinsic bias-dependent elements C_{gs} , R_i , C_{gd} , g_m , R_{ds} , C_{ds}
- Extrinsic(Parasitic) linear elements L_g , R_g , L_s , R_s , R_d , L_d , C_{pg} , C_{pd}

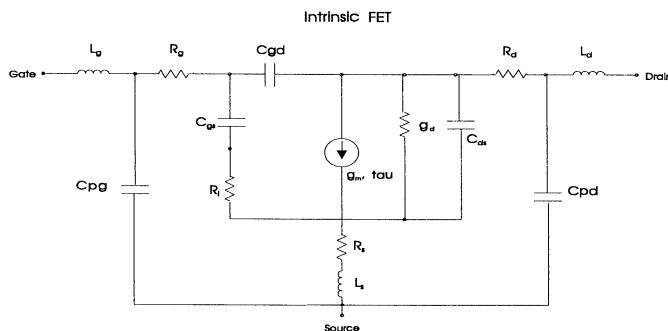


Figure 2: Small Signal equivalent circuit of a MESFET/HEMT

The extraction of ECPs are performed in successive steps in order to avoid ambiguous solutions. The extrinsic elements are extracted first, followed by the extraction of the intrinsic elements. We have studied two methods in extracting the extrinsic elements:

- Cold FET Technique, which extracts the extrinsic resistances and inductances(R_s , R_d , R_g , L_s , L_d , L_g) from forward gate-bias($V_{ds}=0$, $V_{gs} > 0.8$) measurement and the pad capacitances C_{pd} , C_{pg} are extracted from pinched measurement($V_{ds}=0$, $V_{gs} < V_p$)
- Unbiased Technique, which extracts the extrinsic resistances, inductances and capacitances (R_s , R_d , R_g , L_e , L_d , L_g , C_{pg} , C_{pd}) from zero bias($V_{ds}=0$, $V_{gs}=0$) measurement and pinch measurement($V_{ds}=0$, $V_{gs} < V_p$)

1.1a. Cold FET Technique

The cold FET technique [1,2] utilizes a forward-gate bias and pinched bias in extracting the extrinsic elements. The pinched-FET allows us to extract the parasitic capacitances associated with the gate and drain pads. The pinched-FET topology is shown in Figure 3 together with its Y matrix equations. This topology is accurate for $V_{gs} < V_{pinchoff}$ and $V_{ds} = 0$.

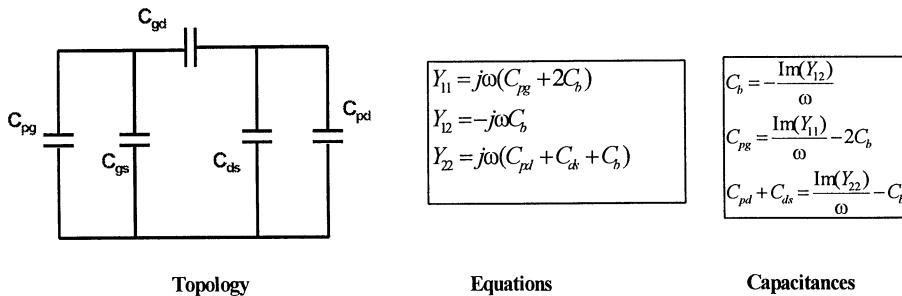


Figure 3: Pinched-FET topology

where the capacitances are:

C_{pg}	Gate-source pad capacitance
C_{pd}	Drain-source pad capacitance
C_b	Depletion layer capacitance, $C_{gs}=C_{gd}=C_b$
C_{ds}	Source-drain interelectrode capacitance

The capacitances are computed at each frequency and each pinched bias point. They are then averaged to determine a final value. The simplification of the topology is valid only at low frequencies (about 8 GHz) where the reactances of the wire bonds of the chip device are small. For wafer-probed devices, the validity of the topology

holds to higher frequencies. Since it is not possible to determine the fraction of Cpd that represents Cds, a fixed portion, namely 75%, is attributed to Cds.

By forward biasing the gate Schottky diode at two or more bias points and taking RF measurements at each bias point, we can determine the parasitic resistances and inductances analytically. This measurement uses a cold-FET, $V_{ds} = 0$, and V_{gs} set positive so that I_g is greater than a few millamps. The equivalent circuit topology for this bias is shown in Figure 4 with the Z-parameter matrix equations in (1).

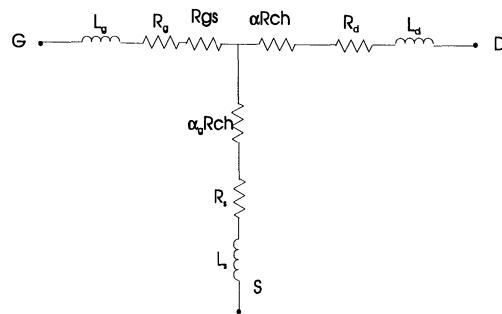


Figure 4: Equivalent circuit for the forward-biased gate FET

Forward-gate Z-parameter equations:

$$\begin{aligned} Z_{11} &= R_s + R_g + R_{gs} + \alpha_g R_{ch} + j\omega(L_g + L_s) \\ Z_{12} &= Z_{21} = R_s + \alpha R_{ch} + j\omega L_s \\ Z_{22} &= R_s + R_d + 2\alpha R_{ch} + j\omega(L_s + L_d) \end{aligned} \quad (1)$$

where:

- R_{gs} = (hVt/Ig) is the resistance of the Schottky junction
- R_s is the source resistance
- R_g is the gate metal resistance
- R_d is the drain resistance
- R_{ch} is the channel resistance
- L_g is the gate inductance
- L_s is the source inductance
- L_d is the drain inductance
- α, α_g are current factors as determined by Vogel [2], whose argument is (R_{ch}/R_{gs})

The inductances are readily determined from the imaginary parts of the Z-parameters. Ls is extracted from the slope of $\text{Im}\{Z_{12}\}$ with frequency, then Lg and Ld are determined from $\text{Im}\{Z_{11}\}$ and $\text{Im}\{Z_{22}\}$ respectively. The inductances are computed for each forward-gate bias point and averaged to determine a final value.

The ideality factor is determined by using the method reported by Vogel [2] where two forward-biased gate points are used at high gate currents. Then, α_g is virtually constant so the only bias dependent parameter in $\text{Re}\{Z_{11}\} = R_{11}$ is R_{gs} . The difference between the two bias points is the written as:

$$R_{11}^{(2)} - R_{11}^{(1)} = R_{gs}^{(2)} - R_{gs}^{(1)} \quad (2)$$

where the R_{11} 's are the average value over frequency at that bias point. We solve for the ideality factor, h in Rgs as:

$$\eta = \frac{\frac{R_{11}^{(2)} - R_{11}^{(1)}}{V_t}}{\frac{I_g^2}{I_g^1}} \quad (3)$$

The ideality factors are computed at each pair of increasing forward-gate bias points so we obtain $N-1$ ideality factors (where N is the number of bias points). These factors are then averaged to obtain a final value.

The resistances are more difficult to distinguish because there are four unknown resistances and only three equations. Therefore, we needed a method to consistently determine the resistances. Noting that $\text{Re}\{Z_{12}\} = R_{12}$ is the simplest relation of the three, and that R_s and R_{ch} are constant over bias, we can write the following at two bias points:

$$R_s = R_{12}^{(1)} - \alpha^{(1)} R_{ch} = R_{12}^{(2)} - \alpha^{(2)} R_{ch} \quad (4)$$

where the numbers in curly braces denote the bias points. Rewrite into:

$$0 = R_{ch}(\alpha^{(1)} - \alpha^{(2)}) - (R_{12}^{(1)} - R_{12}^{(2)}) \quad (5)$$

We then use a root-finding algorithm to find R_{ch} at the solution equation (5). We call this a zero-crossing method of determining R_{ch} and consequently R_s . A typical plot of (5) is shown in Figure 5. Note that the R_{12} values are averaged over frequency at each bias point.

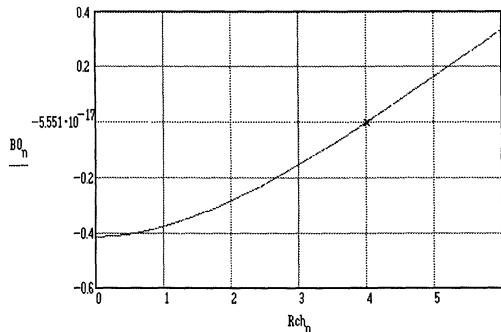


Figure 5. Equation (5) for a typical MESFET

The value of R_{ch} obtained at the two lowest bias points is used to determine R_s , R_g , and R_d at all other forward-biased points. These resistances are then averaged to obtain final parasitic resistance values.

At this point, we have obtained all parasitics relevant to the equivalent circuit model and now can compute the intrinsic equivalent circuit parameters of actively biased devices.

1.1b. Unbiased Technique

A new extrinsic parameter extraction method [3] was developed which offers the following important advantages over the cold-FET approach:

- The extraction method does not rely on forward-biasing the gate Schottky junction, eliminating any gate degradation caused by large gate currents.
- All parasitic resistances are directly obtained from the unbiased and pinched-FET measurements, allowing observation of the frequency dependence of these resistances and eliminating ambiguities that may be found in the cold-FET resistance extraction.

Figure 1.5a shows the equivalent circuit topology for the unbiased-FET. The Z-parameters are:

$$Z_{11} = R_s + 0.5R_{ch} + R_g + j\left(\omega(L_g + L_s) - \frac{1}{\omega C_g}\right) \quad (6)$$

$$Z_{12} = Z_{21} = R_s + 0.5R_{ch} + j\omega L_s \quad (7)$$

$$Z_{22} = R_s + R_{ch} + R_d + j\omega(L_d + L_s) \quad (8)$$

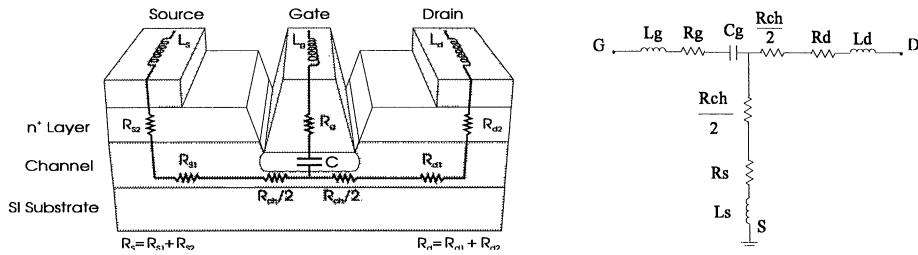


Figure 6. Cross sectional view and equivalent circuit of the Unbiased FET.

The inductances are determined from the imaginary parts of the unbiased-FET Z-parameters. L_s is extracted at each frequency point from $\text{Im}\{Z_{12}\}$ in eq. 7. L_d can then be determined from $\text{Im}\{Z_{22}\}$ in eq. 8. We solve for L_g and C_g in eq. 6 using $\text{Im}\{Z_{11}\}$ at two frequency points. Frequencies below $\sim 5\text{GHz}$ are neglected since the reactance of the inductances is small and cannot be consistently determined.

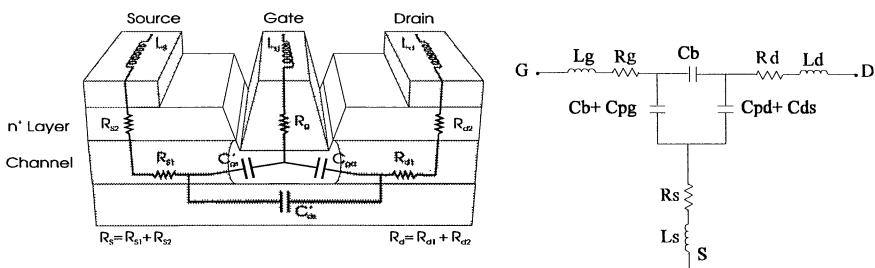


Figure 7. Equivalent Circuit of the Pinched FET.

The pinched-FET equivalent circuit topology is as shown in Figure 7. As can be noted from the above figure, we have included the effect of the parasitic resistances and inductances which the cold-FET method for the pinched FET, have avoided. This complete topology for the pinched FET enables us to obtain more accurate values for the pad capacitances. Here, we have combined the pad capacitances, C_{pg} and C_{pd} , with the intrinsic capacitances C_b and C_{ds} without additional error. Examining the real parts of the Z-parameters for both the unbiased FET of Figure 6 and the pinched-FET of Figure 7, we can write:

For the unbiased FET:

$$\text{Re}\{Z_{11}\} = R_s + 0.5R_{ch} + R_g \quad (9)$$

$$\text{Re}\{Z_{12}\} = R_s + 0.5R_{ch} \quad (10)$$

$$\text{Re}\{Z_{22}\} = R_s + R_{ch} + R_d \quad (11)$$

For the pinched-FET:

$$\text{Re}\{Z_{11}\} = R_s + R_g \quad (12)$$

$$\text{Re}\{Z_{12}\} = R_s \quad (13)$$

Eq. 13 yields R_s directly, although we have only found this to be accurate for small devices (gate width < 100mm) when using TRL calibration. Otherwise we use eq. (12) and eq. (9) to determine R_{ch} and then eq. (12) to find R_s . Once R_s is found, eq. (11) is used to compute R_g and eq. (13) is used for R_d . We have found that the value of R_s is sensitive to the accuracy of the measurement system calibration. Note that this method of extracting the resistances (and inductances) does not require the forward biasing of the gate, thus eliminating metal migration and reliability issues. Also, the ambiguity of the "Alpha" factors used in [1,2] is avoided.

The parasitic gate and drain pad capacitances, C_{pg} and C_{pd} , are obtained by performing S-Z parameter conversions from the pinched-FET measurement data. The branch impedances of the parasitic inductances and resistances, which are already known, are subtracted from the Z matrix. The new Z matrix is converted to Y parameters and C_{pg} , C_{pd} , C_b , and C_{ds} can be determined from:

$$C_b = -\frac{\text{Im}\{Y_{12}\}}{\omega} \quad C_{pg} = \frac{\text{Im}\{Y_{11}\}}{\omega} - 2C_b \quad C_{ds} + C_{pd} = \frac{\text{Im}\{Y_{22}\}}{\omega} - C_b \quad (14)$$

Since C_{ds} and C_{pd} cannot be separated, we use the assumption that $C_{pd} = 0.25*C_{ds}$.

This parasitic extraction method was used to determine the parasitics of a 0.25umX100um HEMT. The comparison of measured and modeled results for the unbiased and pinched FET are as shown in Figures 1.5c and 1.5d. The excellent agreement verifies the model topologies and extraction method used for the unbiased and pinched device.

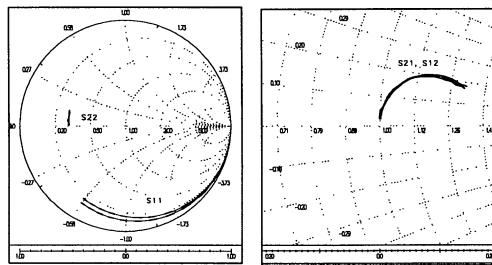


Figure 8: Measured and Modeled results for the Unbiased HEMT(0.25umX100um)

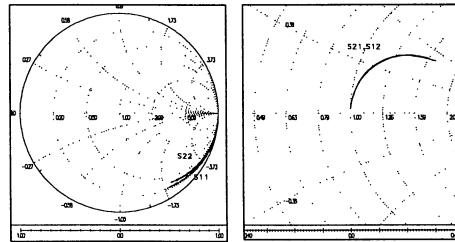


Figure 9: Measured and Modeled results for the Pinched HEMT(0.25um X100um)

At this point we have obtained all parasitics relevant to the equivalent circuit model and can compute the intrinsic equivalent circuit parameters of actively biased devices.

1.1c. Hot FET Technique

After obtaining the extrinsic parameters from either of the above two approaches, the next step involved is to extract the intrinsic parameters. De-embedding the extrinsic parameters from the equivalent circuit shown in Figure 2, we end up with a π -equivalent circuit comprising of C_{gs} , C_{gd} , R_i , G_m , G_d , Tau and C_{ds} . These intrinsic parameters are extracted using the equations published by M. Berroth and R. Bosch [4].

$$\begin{aligned}
 C_{gd} &= -\frac{\text{Im}\{Y_{12}\}}{\omega} \\
 C_{gs} &= \frac{\text{Im}\{Y_{11}\}}{\omega} - C_{gd} \\
 R_i &= \frac{\text{Re}\{Y_{11}\}}{\omega^2 C_{gs}^2} \\
 g_m &= \text{Re}\{Y_{21}\} \\
 \tau &= -\left(\frac{\text{Im}\{Y_{21}\}}{\omega} - C_{gd}\right) \frac{1}{g_m} - R_i C_{gs} \\
 g_d &= \text{Re}\{Y_{22}\} \\
 C_{ds} &= \frac{\text{Im}\{Y_{22}\}}{\omega} - C_{gd}
 \end{aligned} \tag{15}$$

1.1d Experimental Results

Based on the above described techniques, extraction were performed for several MESFETs and HEMTs. We present here the results of a HEMT(0.25um X 100um) and MESFET(0.5um X 200um), both fabricated at Foundry C. Figures 10 and 11 show the excellent agreement obtained for the modeled and measured HEMT and MESFET S-parameters biased at $V_{gs}=0V$, $V_{ds}=4V$ and $V_{gs}=-0.5V$ and $V_{ds}=5V$ respectively. It is to be noted in Figure 10, there are three traces for each of the S-parameters, which are (i)Measured (ii) Modeled(using Cold FET technique) (iii) Modeled (using Unbiased FET technique). We note from this figure that the proposed extraction(Unbiased FET) method provides as good results as the cold FET method. In Figure 11, the extraction of the extrinsic parameters for the MESFET were calculated using the Unbiased FET method. Similar results were also obtained for a Foundry B MESFET and Foundry A HEMT.

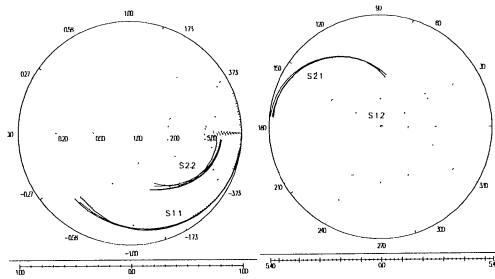


Figure 10: Measured and Modeled S-parameters of Foundry C HEMT at $V_{ds}=4V$ and $V_{gs}=0V$

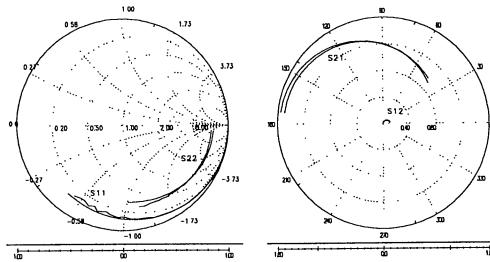


Figure 11: Measured and Modeled S-parameters of Foundry C MESFET at $V_{ds}=5V$ and $V_{gs}=-0.5V$

1.2. Nonlinear MESFET and HEMT Model Parameter Extraction Program

This section discusses the nonlinear model parameter extraction program for MESFETs and HEMTs. The ECP data over bias, obtained from the linear extractor (discussed in section 1.1) is read by the nonlinear extractor and optimizes the model coefficients to minimize the difference between computed and extracted ECPs over all the bias. The following models are implemented in our nonlinear extractor program:

1. Modified Triquint MESFET drain current model (TOM)
2. Angelov MESFET/HEMT drain current model
3. Modified Materka-Kazprazak drain current model
4. TOM capacitance equations
5. Angelov capacitance equations

6. Materka capacitance equations
7. Materka channel resistance model

The channel current determines the DC I-V curves (Id), the transconductance (Gm), and the output conductance (Gd) over bias. The FIT program uses the measured I-V curves, and the extracted Gm and Gd from the ECPs. We have had good success fitting Id and Gm , but not Gd . Shown in Figure 12 are comparisons between extracted and modeled Id , Gm , and Gd when only the Id data is used for the fitting. The device used here is a $0.5 \times 150\mu m$ power FET from Foundry B and modeled using Materka model. Note that Id fits very well, Gm is over estimated and Gd fits poorly.

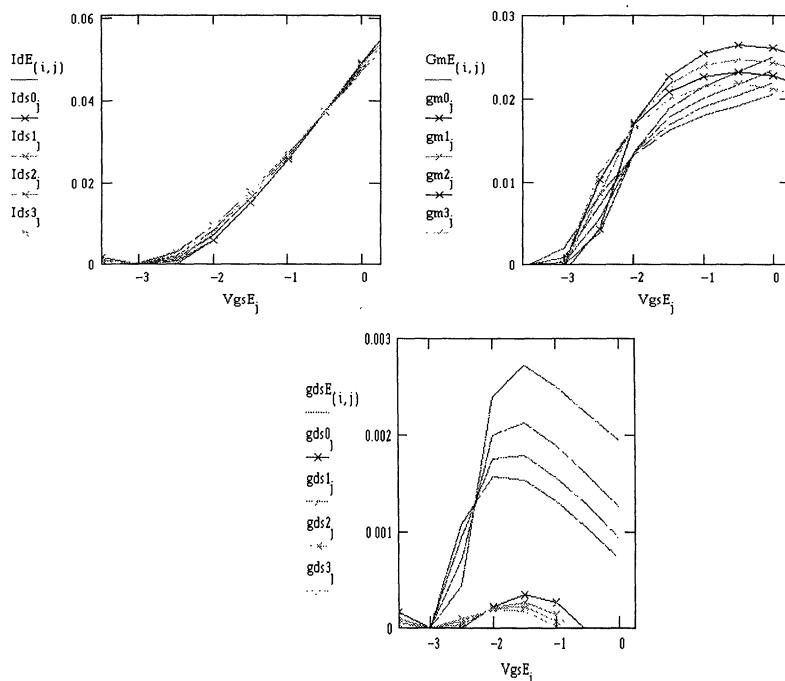


Figure 12. Fit of Id , Gm , and Gd using the Id data set only. Measured data (solid), modeled data (x's).

Figure 13 shows the comparisons when only the Gm data is used for the fitting. Here Gm fits very well, Id is under estimated and Gd has a slightly better fit than with the Id data. The reason that Id and Gm pull opposite each other at high Vgs is the

combined effects of low frequency dispersion and the thermal heating in the device at high bias levels.

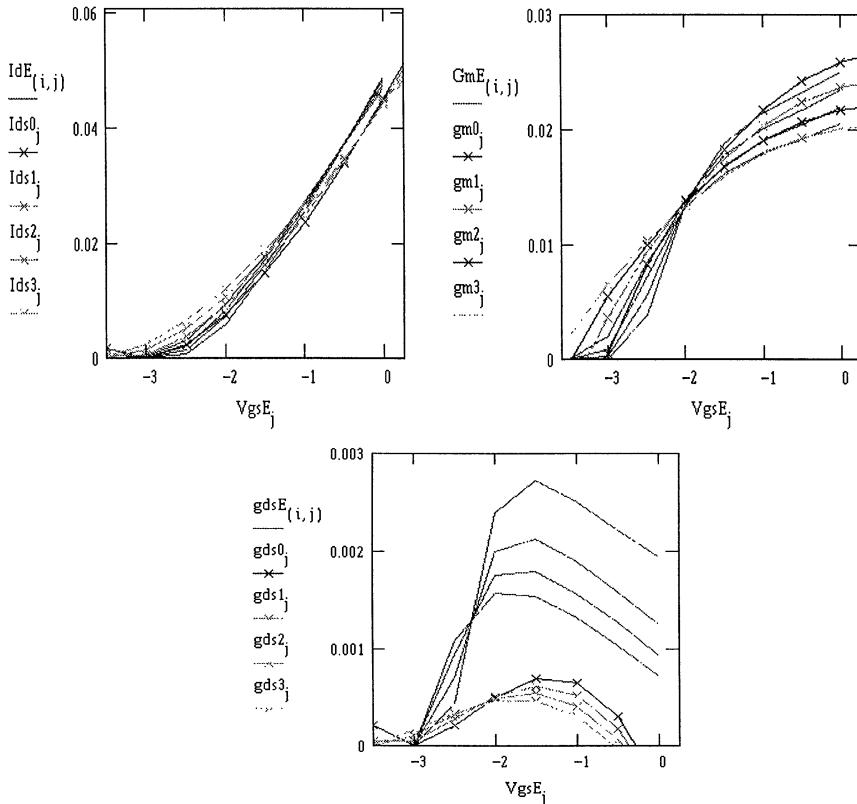


Figure 13. Fit of Id , Gm , and Gd using the Gm data set only. Measured data (solid), modeled data (x's).

Since Id and Gm do not seem to be able to match simultaneously, there will be a trade-off between matching one or the other. We can weight this trade-off, depending on our circuit application and bias point, or we can select a "typical" set of weighting factors that can be revised later. Figure 14 illustrates when Id and Gm are used in the fitting procedure. Both data sets were weighted evenly.

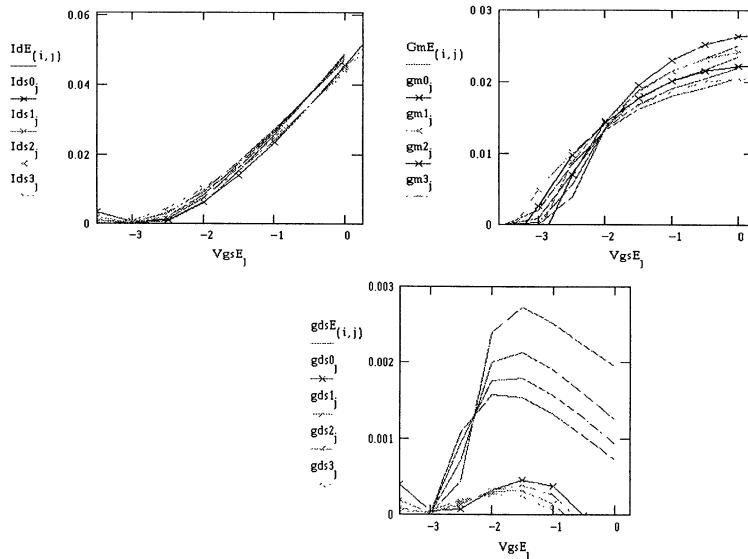


Figure 14. Fit of Id, Gm, and Gd using the Id and Gm data sets.

If we include the Gd data set in our fitting process, the data in Figure 15 results. Again, the fit for Id, Gm, and Gd is a tradeoff. The Id and Gm fits are not as good as Figure 14 where the Gd data was not included, but the Gd fit here is much better.

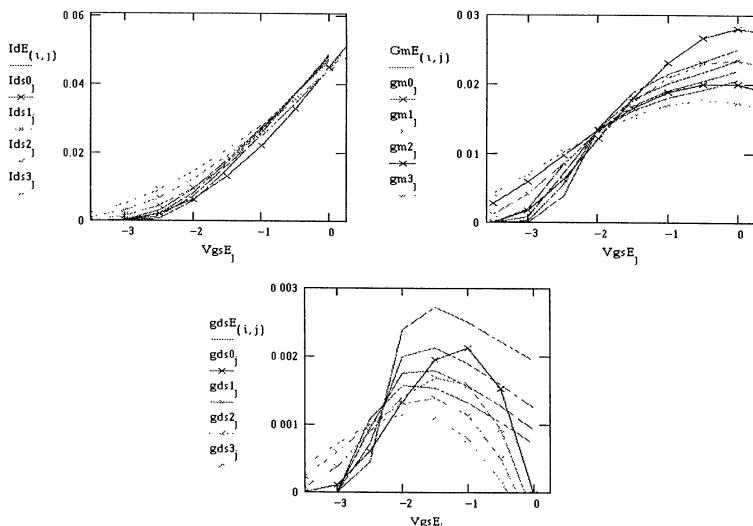


Figure 15. Fit of Id, Gm, and Gd using all the data set only. Measured data (solid), modeled data (x's).

From our study it was found that the TOM and Materka models were primarily used to model MESFETs. These models failed to predict the nonlinear characteristics of the HEMT transconductance behavior. Angelov et. al [5,6] proposed a model which rectified this discrepancy. Their model showed good fits for different technology devices such as, MESFET, pHEMT, double delta-doped pHEMT, and a lattice matched InP HEMT. The model equations for the channel current and capacitances are given as [5,6].

Figures 16 -19 illustrate the comparison between Angelov model and measured data for a Foundry C pHEMT (0.25um X 100um). that Id and Gm closely match with extracted results (The weights used were WId=1, WGm=1 and WGd=0.1), however, the Gd fit, as experienced in the case of MESFET, is not as good (Figure 18). Figure 19 compares the measured and modeled capacitances(Cgs and Cgd).

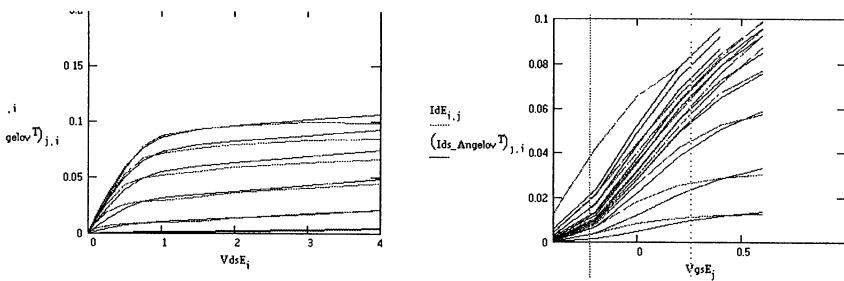


Figure 16. Modeled and Measured (Angelov) Id with respect to Vds and Vgs

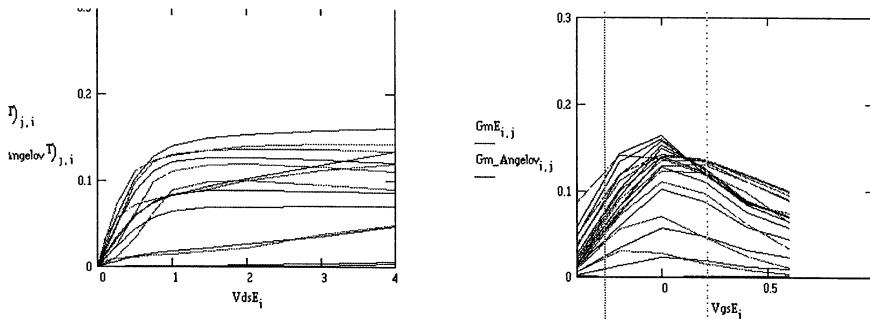


Figure 17. Modeled and Measured Gm with respect to Vds and Vgs

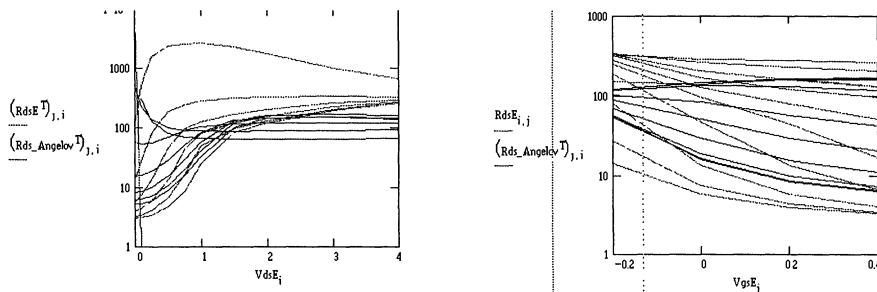


Figure 18. Modeled and Measured R_{ds} with respect to V_{ds} and V_{gs}

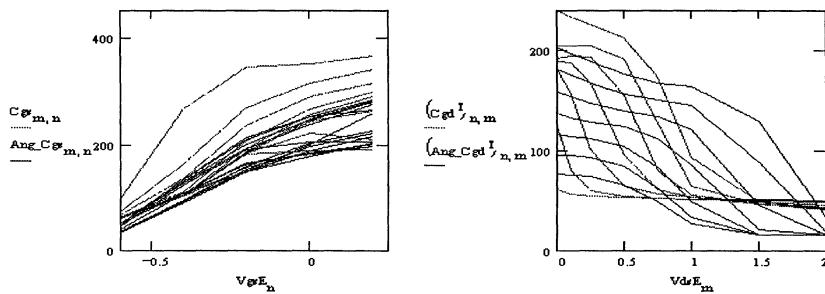


Figure 19. Measured and Modeled capacitances(C_{gs} and C_{gd})

Section 1.3. Comparison of advanced non-linear models

This section discusses the comparative study done on the three nonlinear models (TOM, Materka and Angelov) for MESFETs and HEMTs.

1.3a. Comparison study of the Nonlinear Models for MESFET

Figure 20 depicts our nonlinear extraction methodology for obtaining the device nonlinear model parameters. Depending on the choice of the model (TOM, Materka or Angelov), the equivalent circuit parameters (ECPs) versus bias obtained from the linear extractor (EXT) is provided as input to the optimizer. The optimizer then calculates the model coefficients based on these ECP input data. In the present study, we have looked into TOM, Materka and Angelov models. The TOM model is a slight modification of the Raytheon-Statz model. Therefore, we have not shown any results of the Raytheon-Statz model in the comparative study.

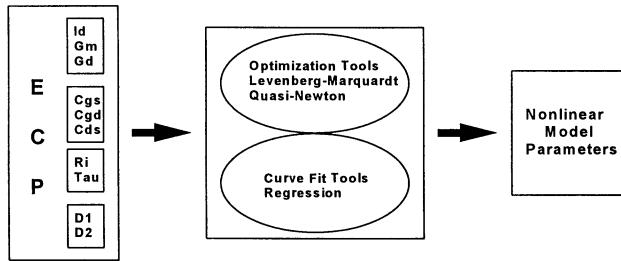


Figure 20. Methodology in extracting the nonlinear parameter

1.3b. Channel Current Model

Figure 21 shows the Foundry C MESFET fit for Id and Gm at Vds = 2.0 V and Rds at Vgs=0. The weighting factors used here are WId = 1.0, WGm = 1.0, and WGd = 0.2. The fit of Id and Gm for all three models is very good, but as generally observed by other workers, it is difficult to simultaneously obtain a good fit for all the three parameters (I-V, Gm, and Rds). Therefore, as it can be seen, the fit for Rds is not as good as the other two parameters.

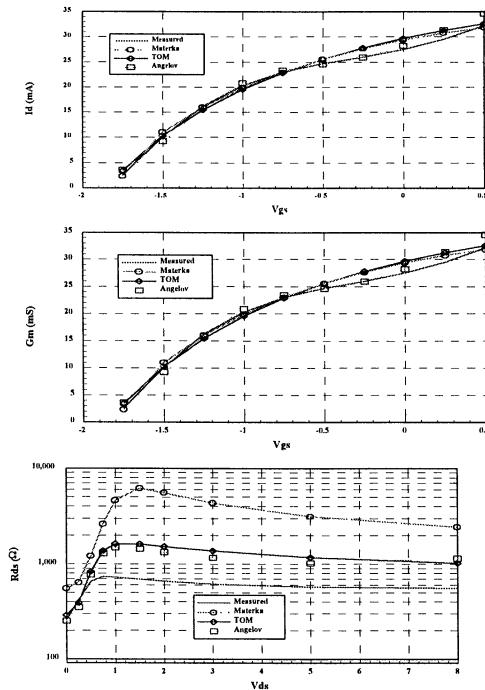


Figure 21. Plot of Id, Gm versus Vgs(at Vds=2) and Rds versus Vds(at Vgs=0)

From these comparisons of bias dependent ECPs, it is not possible to say with confidence that one MESFET channel current model is better than the others. They all show similar problems modeling Rds and are almost equally good at modeling Id and Gm. We could formulate a single error measurement unit, but we don't feel that it would be an absolute figure-of-merit indicator. It seems that depending how a particular device will be used in an application, that is, where it will be biased and where the dynamic load locus will be, will lead to using one model over another. A better indication would be to simulate a practical circuit using the parameter extraction methods developed here and compare first- and higher-order responses.

1.3c. Capacitance Model

We can compare the capacitance models for the Materka, TOM, and Angelov models for the foundry (C) MESFET. If we fix one of the bias axes and compare the models in Figure 22, we can clearly see that the Angelov model matches Cgs very well and both the TOM and Angelov models match Cgd. However, neither model can predict the "turn over" of Cgd at Vds = 0 V. It is to be noted based on the capacitance equations for the Materka model, there is no variation of Cgs with Vds. But the TOM and Angelov models predict the Cgs dependency on Vds.

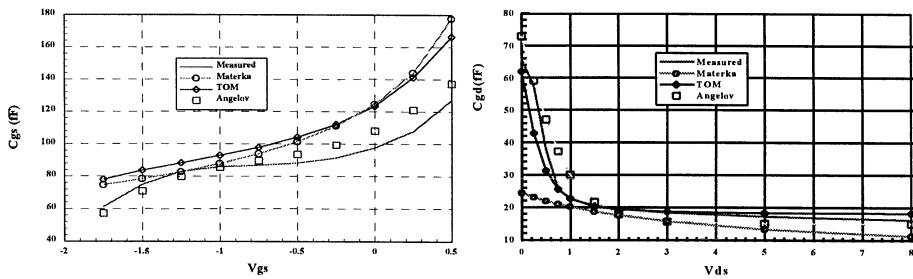


Figure 22. Comparison study of Cgs and Cgd

Therefore we can conclude that the Angelov capacitance model shows the best fit to measured data, while the TOM model would produce an acceptable fit, and the Materka model does not fit well except for Cgd in the saturated region.

The comparison study of the Nonlinear Models for HEMT are described in the following sections.

1.3d. Channel Current Model

Figure 23 compares the results of I_d , G_m and R_{ds} with respect to V_{gs} ($V_{ds}=2V$). The weighting parameters used for the three models are $W_{Id}=1$, $W_{gm}=1$ $WGd=0.1$. The Angelov Model fits very well for both I_d and G_m . Notice that the TOM model and Materka model doesn't predict the "bell shape" of the transconductance curve. For the modeling of R_{ds} , better prediction is obtained from Angelov model.

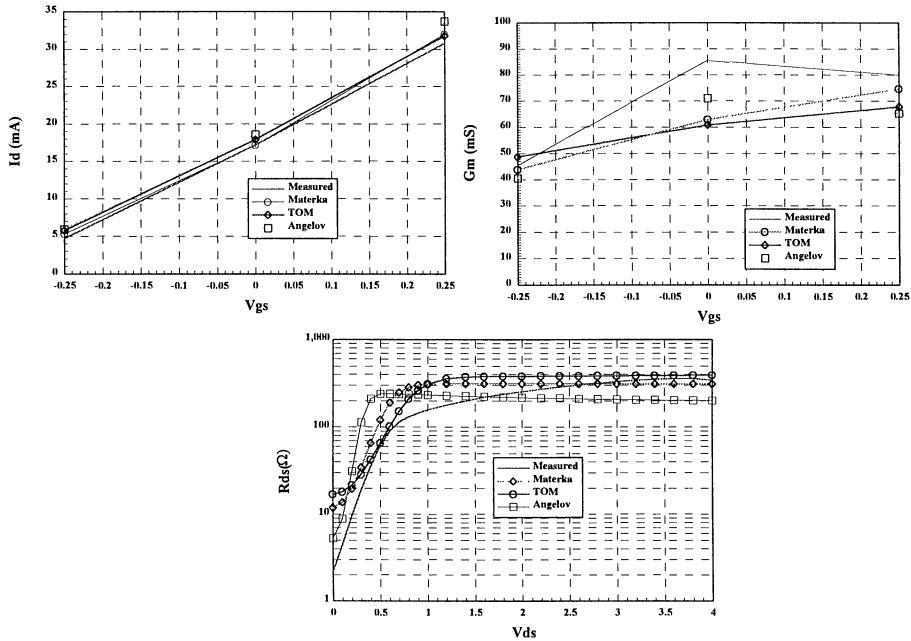


Figure 23. Comparison of the models for I_d , G_m and R_{ds} at $V_{ds}=2V$.

1.3e. Capacitance Model

Results of modeling the capacitances are shown below in Figure 24. For C_{gs} , we obtain better results using the Angelov model than the TOM or the Materka models. However, for C_{gd} we see in the linear region, the Angelov model is not as accurate as the measured while the other two models predict closely to the measured. Towards the saturation region, we see the Angelov model predicting closer to the measured.

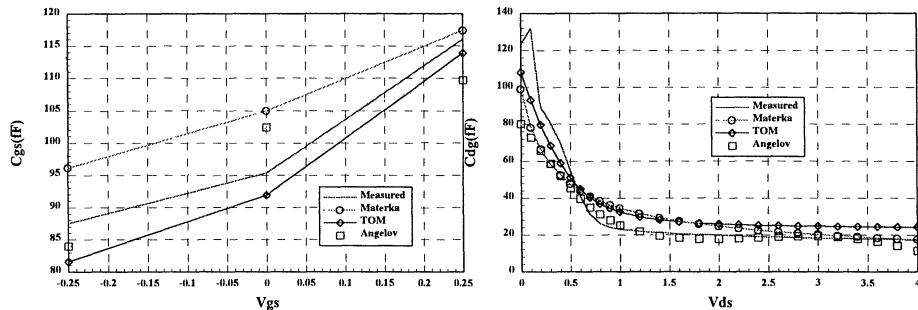


Figure 24. Comparison of the nonlinear models for $C_{gs}(V_{ds}=2V)$ and $C_{gd}(V_{gs}=0V)$

From our study on the models, we have found out that the models are more accurate in the saturation region than at the linear region. A table of the comparative study of the three models are presented below in Table

1.

ECP	Materka	Raytheon	Triquint	Angelov
Gm	■	■	■	▲
Rds	●	■	▲	■
Cgs	■	▲	▲	▲
Cgd	■	▲	▲	▲
Ri	▲	NA	NA	NA

▲ Good ■ Fair ● Bad

Table 1 Comparative study of the nonlinear models.

1.4. Summary

For the linear extraction of MESFETs and HEMTs, we have developed a novel method(Unbiased Technique) for extracting the parasitic elements of MESFETs and HEMTs. This extraction method does not rely on forward-biasing the gate Schottky junction, eliminating any gate degradation caused by large gate currents. Also, all parasitic resistances are directly obtained from the unbiased and pinched-FET measurements, allowing observation of the frequency dependence of these resistances and eliminating ambiguities that may be found in the cold-FET resistance extraction. Experimental results prove the validity of the method. From our study on the

nonlinear models for MESFETs and HEMTs, we have found the Angelov model more accurate(HEMTs) in predicting the ECPs when compared with the other models such as, TOM and Materka. In particular, the Angelov model predicts accurately the “bell-shape” transconductance curve of HEMTs, which the other models fail to do so.

Acknowledgments

This work was supported by ARPA MIMIC Phase III program. We appreciate the excellent support of the MIMIC program office led by Eliot Cohen and Elissa Sobolewskiof of ARPA, R. T. Kemerley, Mark Calcaterra, and Steve Kiss of the Wright Laboratory as well as Army and Navy personnel.

References

- [1] G. Dambrine, A. Cappy, F. Helidore, and E. Playez, "A New Method for Determining the FET Small-Signal Equivalent Circuit," IEEE Trans. MTT, vol. 36, no. 7, 1988
- [2] R. Vogel, "The Application of RF Wafer Probing to MESFET Modeling," Microwave Journal, pp. 153-162, Nov., 1988
- [3] R. Tayrani, J.E. Gerber, T. Daniel, R.S. Pengelly, U. L. Rhode , " A new and reliable direct parasitic extraction method for MESFETs and HEMTs," 23rd European Microwave Conference, Madrid, Spain, 1993.
- [4] M. Berroth and R. Bosch, "Broad-Band Determination of the FET Small-Signal Equivalent Circuit, " IEEE Trans. MTT, July, 1990, pp. 891-895.
- [5] I. Angelov, H. Zirath, and N. Rorsman, "A New Empirical Nonlinear Model for HEMT Devices," IEEE MTT Symp. Digest, pp. 1583-1586, 1992.
- [6] I. Angelov, H. Zirath, and N. Rorsman, "A New Empirical Nonlinear Model for HEMT and MESFET Devices," IEEE Trans. on MTT, vol. 40, no. 12, pp. 2258-2266, 1992.

ADVANCED CAD MODELS

G.GHIONE, F.BONANI AND M.PIROLA

*Politecnico di Torino, Dipartimento di Elettronica,
Corso Duca degli Abruzzi 24, 10129 Torino, Italy*

1. Advanced large-signal models

1.1. INTRODUCTION

Much effort has been devoted during the last few years to the development of accurate and efficient nonlinear FET models for analog circuit design. Although MESFETs, HEMTs, pseudomorphic HEMTs each have their own peculiarities, a large amount of the work carried out on MESFETs can be readily applied to PHEMT modelling. Despite the variety of models and modelling approaches suggested during the last few years, and of the significant success achieved in large-signal circuit design, a completely satisfactory solution to large-signal modelling from the standpoint of accuracy, flexibility, ease of experimental characterization and range of operating conditions still has to be found. This is due to the complexity of nonlinear modelling, but above all to the high accuracy required by analog circuit design.

In the following subsections, an overview will be provided on some of the current problems in advanced measurement-based large-signal modelling. Particular stress is laid on black-box modelling approaches, which have recently emerged as a solution towards the development of a fully automated, flexible and device-independent large-signal characterization strategy. Part of this section is devoted to the specific topic of equivalent-circuit large-signal PHEMT modelling, still a popular design tool owing to the straightforward implementation in most circuit simulators.

1.2. TRENDS IN NONLINEAR DEVICES MODELLING

Since non-linear device modelling can be seen as an extension and generalization of linear modelling, it will be helpful to recall that linear FET modelling follows two basic approaches. In the *black box approach*, small-

signal measurements are carried out on a suitable frequency band and the resulting small-signal parameters (for instance, the four scattering parameters) are directly (or through suitable interpolation techniques) inserted into a linear simulator. In the *equivalent circuit approach*, a suitable circuit topology, derived from the device physics, is assumed and the component parameters (capacitance, resistance, transconductance values) are derived through direct or optimization techniques so as to reproduce, as accurately as possible, the measured small-signal parameters on a prescribed frequency band. In linear operation, the black-box approach is exact, since linear devices, connected to any circuit, are completely characterized by a frequency-domain set of linear two-port parameters. The approximate equivalent circuit approach, although not strictly needed for circuit simulation, is still popular since it enables to derive frequency-independent figures of merit like the transconductance and cutoff frequency.

In the nonlinear case, both the black-box and the equivalent circuit approach have been pursued during the last few years. Exact black-box models for nonlinear circuits or systems are provided by the so-called *Volterra series* method, which can be interpreted as an extension, to the nonlinear case, of the linear transfer function or pulse response concepts [1]. Unfortunately, the Volterra approach is cumbersome not only from the standpoint of model implementation, but above all as far as the experimental characterization is concerned [2]. Moreover, the Volterra method is effective only when the device operates in the quasi-linear region, for instance in the distortion analysis of amplifiers. Thus, it can be said that *no exact method* for nonlinear device modelling is practically available.

Several attempts have been therefore made to develop, on the basis of some physical knowledge of FET operation, approximate black box models whose implementation and characterization only requires DC and small-signal parameter measurements carried out in different working points. Examples of such approaches, which will be described in more detail further on, are state-function based representations like the Root model [3, 4] and the Nonlinear Integral Model (NIM) proposed, on the basis of a Volterra series approach, by Filicori *et al.* [5]. At present, the only black-box model available in commercial simulators [6] is the Root model.

Equivalent-circuit nonlinear models were derived, since the pioneering work by Rauscher and Willing [7, 8], by taking a topology derived from small-signal lumped-parameter models and letting the resistive and capacitive elements vary as a function of the working point. Practical experience and device physics suggest suitable functional forms to describe the behaviour of the circuit elements (capacitors, resistors, current sources) as a function of the instantaneous voltages applied to the device terminals. The element models are parametrized and the model parameters are

fitted, through optimization techniques, so as to reproduce at best a set of measurements carried out in several working conditions. These usually are the DC characteristics, the small-signal parameters in several working points, and sometimes large-signal characteristics like the $P_{in} - P_{out}$ curve in selected working points. Most commercially available simulators support either fixed-topology nonlinear equivalent models or user-defined models in which the user can freely assign the analytical component models. A variety of MESFET models have appeared in the literature and have been implemented both in time-domain and in frequency-domain simulators: we shall confine ourselves here to a reference list of the most significant milestones [8, 9, 10, 11, 12]. With little or no modification in the model topology, the approach was also applied to HEMTs; however, the peculiarities of this device when compared to the MESFET soon made clear that the parameter modelling suggested by MESFET models was inadequate, and that *ad hoc* analytical models, more suited to the HEMT behaviour, were to be devised for this device.

1.3. LOW-FREQUENCY DISPERSION AND THERMAL EFFECTS IN FETS

The modelling of microwave FETs would be easier if these devices had (apart from capacitive effects) the same behaviour at microwave frequencies and in DC. Unfortunately, this is not the case in most practical MESFETs and HEMTs. The voltage-current (VI) curves measured at slow sweep or in true DC conditions have a different slope from the VI curves measured with a fast sweep (or *pulsed* VI characteristics [16]), and similarly the static transconductance and output conductance are different from the small-signal values measured for frequencies in excess of a few MHz. The fact that the DC and the RF VI characteristics differ is commonly referred to as *non-quasistatic* behaviour, and similarly *non-quasistatic models* are those able to allow for this behaviour, which mainly originates from two physical mechanisms.

The first mechanism is connected to surface and substrate trap dynamics (see *e.g.* [13, 14] and references therein). In slow-sweep conditions, the trap population, which is characterized by lifetimes of the order of ms to μ s, is able to follow the applied bias, and can therefore modulate the drain current; under a fast sweep, traps are frozen at the bias point and are inactive as far as the current modulation is concerned. The effect of traps is often referred to as the *low frequency dispersion* (LFD) effect. Consistent with device physics, the high-frequency (or radio-frequency, RF) transconductance of MESFETs is lower than the DC one (for instance, 70-80%); the opposite happens in HEMTs. In both devices the DC output resistance is much larger than the RF output resistance (*e.g.*, several k Ω against 200-300

Ω). To a far lesser extent, low-frequency dispersion also affects the device capacitances.

The second mechanism arises from device self-heating [15]. Since the mobility and saturation velocity are inversely proportional to the device temperature, self-heating modulates the drain current if the bias point is varied slowly with respect to the thermal time constants of the device (in the same range as the trap lifetimes). However, if the device is driven by a fast waveform, the device temperature is frozen at a value corresponding to the average power dissipation (not necessarily equal to the DC dissipation).

In order to stress the modelling difficulties caused by LFD and thermal effects, let us develop an approximate model for the FET drain current. It may be useful to visualize low-frequency dispersion effects in terms of an additional state variable q_t ; this is a lumped approximation defining the instantaneous electrical status of the trapping centers. Similarly, the average channel temperature θ can be used as the state variable for thermal behaviour. Thus, we can express the drain current as:

$$i_{DS} = I_{DS}(v_{DS}(t), v_{GS}(t), q_t(t), \theta(t)) + \frac{dQ_D}{dt} \quad (1)$$

where $Q_D = Q_D(v_{DS}(t), v_{GS}(t))$ is the drain charge, while the slow trap and thermal dynamics suggest the following relaxation-time empirical law:

$$\left(1 + \tau_t \frac{d}{dt}\right) q_t = Q_t^{DC}(v_{DS}(t), v_{GS}(t)) \quad (2)$$

where τ_t is the average trap lifetime, Q_t^{DC} is the DC trap occupancy. Similarly, the following equation can be postulated for the channel temperature:

$$\left(1 + \tau_\theta \frac{d}{dt}\right) \theta = T_0 + R_\theta p(t) \quad (3)$$

where τ_θ is the thermal time constant, R_θ the thermal resistance, T_0 the heat sink temperature, and $p(t) = v_{DS}(t)i_{DS}(t)$ is the instantaneous dissipated power.

Let us now evaluate the small-signal parameters associated to Eq. (1). For the sake of simplicity, we shall suppose that the drain current does not depend on temperature, and consider only the effect of traps. Suppose that the driving voltages and the trapped charge can be split into DC and small-signal components as $V_{DS} + v_{DS}(t)$, $V_{GS} + v_{GS}(t)$, $Q_t + q_t(t)$. Then, the small-signal equations read, in the frequency domain:

$$i_{DS} = \frac{\partial I_{DS}}{\partial v_{DS}} v_{DS} + \frac{\partial I_{DS}}{\partial v_{GS}} v_{GS} + \frac{\partial I_{DS}}{\partial q_t} q_t + j\omega C_{DS} v_{DS} + j\omega C_{GS} v_{GS} \quad (4)$$

where all partial derivatives are evaluated at the DC working point, and: $C_{DS} = \partial Q_{DS}/\partial v_{DS}$, $C_{GS} = \partial Q_{GS}/\partial v_{GS}$. Finally, the dynamic equation for the trap density becomes:

$$(1 + j\omega\tau_t)q_t = \frac{\partial Q_t^{DC}}{\partial v_{DS}}v_{DS} + \frac{\partial Q_t^{DC}}{\partial v_{GS}}v_{GS}. \quad (5)$$

On substituting Eq. (5) into Eq. (4) one obtains:

$$\begin{aligned} i_{DS} &= \frac{\partial I_{DS}}{\partial v_{DS}}v_{DS} + \frac{\partial I_{DS}}{\partial v_{GS}}v_{GS} + \frac{1}{1 + j\omega\tau_t} \times \\ &\times \frac{\partial I_{DS}}{\partial q_t} \left(\frac{\partial Q_t^{DC}}{\partial v_{DS}}v_{DS} + \frac{\partial Q_t^{DC}}{\partial v_{GS}}v_{GS} \right) + j\omega C_{DS}v_{DS} + j\omega C_{GS}v_{GS} \end{aligned} \quad (6)$$

Taking into account that for $\omega \rightarrow 0$ one has:

$$i_{DS} \rightarrow \left(\frac{\partial I_{DS}}{\partial v_{DS}} + \frac{\partial I_{DS}}{\partial q_t} \frac{\partial Q_t^{DC}}{\partial v_{DS}} \right) v_{DS} + \left(\frac{\partial I_{DS}}{\partial v_{GS}} + \frac{\partial I_{DS}}{\partial q_t} \frac{\partial Q_t^{DC}}{\partial v_{GS}} \right) v_{GS}, \quad (7)$$

that is:

$$i_{DS} \rightarrow g_{DS}^{DC}v_{DS} + g_m^{DC}v_{GS}, \quad (8)$$

while, in the limit $\omega\tau_t \gg 1$:

$$i_{DS} \approx \frac{\partial I_{DS}}{\partial v_{DS}}v_{DS} + \frac{\partial I_{DS}}{\partial v_{GS}}v_{GS} + j\omega C_{DS}v_{DS} + j\omega C_{GS}v_{GS} \quad (9)$$

that is:

$$i_{DS} \approx (g_{DS}^{RF} + j\omega C_{DS})v_{DS} + (g_m^{RF} + j\omega C_{GS})v_{GS}, \quad (10)$$

we can write, for arbitrary frequency, the admittance parameters of the device as:

$$Y_{DS}(\omega) = \frac{1}{1 + j\omega\tau_t}g_{DS}^{DC} + \frac{j\omega\tau_t}{1 + j\omega\tau_t}g_{DS}^{RF} + j\omega C_{DS} \quad (11)$$

$$Y_{GS}(\omega) = \frac{1}{1 + j\omega\tau_t}g_m^{DC} + \frac{j\omega\tau_t}{1 + j\omega\tau_t}g_m^{RF} + j\omega C_{GS}. \quad (12)$$

An example of the frequency behaviour of Y_{DS} is shown in Fig. 1.

The above analysis points out that, owing to LFD effects, the small-signal parameters are different at DC and RF; similar remarks hold for thermal effects, which can in turn be interpreted as a form of low-frequency dispersion. Furthermore, the dynamic slope of the RF characteristics under a large-signal sweep is generally different from both the DC and the RF small-signal slope. In fact, let us consider the static VI curves and the

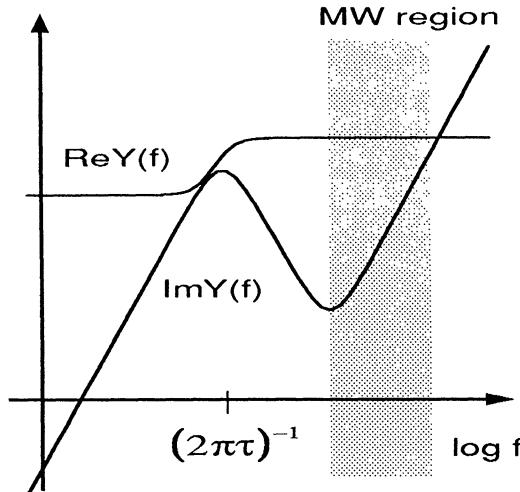


Figure 1. Frequency behaviour of FET output admittance with low-frequency dispersion effects.

pulsed VI characteristics measured starting from the DC working point V_{DS0}, V_{GS0} . The static VI curves are:

$$i_{DS}^{DC} = I_{DS}(v_{DS}, v_{GS}, Q_t^{DC}(v_{DS}, v_{GS}), \theta^{DC}(v_{DS}, v_{GS})) \quad (13)$$

where θ^{DC} is the DC channel temperature corresponding to the working point v_{DS}, v_{GS} . On the other hand the dynamic (pulsed) characteristics are:

$$i_{DS}^{RF} \approx I_{DS}(v_{DS}, v_{GS}, Q_t^{DC}(V_{DS0}, V_{GS0}), \theta^{DC}(V_{DS0}, V_{GS0})), \quad (14)$$

since under a fast sweep the trap occupancy and channel temperature are frozen at the starting point V_{DS0}, V_{GS0} . Owing to the different trap occupancy and channel temperature, the two functions are different and so are their partial derivatives with respect to the driving voltages. Therefore, the dynamic slope of the IV characteristics is different from the one measured in small-signal conditions.

The modelling of LFD and heating effects directly based on the model of Eq. (1) is not straightforward, since the trap occupancy cannot be independently controlled, while the channel temperature can only go up to a certain extent (by changing the heat sink temperature). In particular, the look-up table identification of the trap- and temperature-dependent model of Eq. (1) requires not only DC and small-signal measurements, but also pulsed DC measurements carried out starting from different working points

and at different heat sink temperatures. On the other hand, suitable approximations based on device physics can help in devising models for LFD and heating effects, which can be implemented into black-box or equivalent circuit models for microwave FETs. An example of implementation will be discussed further on, when dealing with PHEMT equivalent circuit modelling approaches.

1.4. ADVANCES IN BLACK-BOX FET MODELLING

Black-box measurement-based nonlinear FET models directly exploit measured DC or small-signal device characteristics according to a look-up table approach, thus avoiding any equation fitting process. Since no direct equation-based modelling of lumped equivalent circuit elements is carried out, the model is flexible and can accommodate, without any equation tuning, geometrical or technological changes. Nevertheless, look-up based models must be coupled with proper automated measurement techniques able to explore the DC and RF device characteristics. Besides, look-up table based models do not have the possibility of extrapolating, in a smooth way, the model representation beyond the measured area; thus, an automated measurement system must be available to explore the device in the whole significant region of operation and to make a denser sampling whenever the device characteristics exhibit a rapidly varying behaviour. Last but not least, such a system must handle the huge amount of data (DC samples, small-signal frequency domain samples) needed to carry out a complete characterization on a wide frequency band.

A final remark concerns the measurements needed to identify the black-box model. Since DC and small-signal measurements are comparatively easy to perform and less demanding, from the standpoint of the equipment needed, when compared to other measurements (like pulsed DC or load-pull characterizations) most black-box models (not directly based on the load-pull approach) try to make use of “standard” DC and SS measurements. Unfortunately, as mentioned in the previous section, those are not enough to *exactly* characterize a non-quasistatic model; on the other hand, the ability to perform exotic measurements is wasted if those are not correctly exploited to identify a model. In conclusion, the models described in what follows are based on DC and SS measurements only, but this limitation is not to be considered as an absolute one.

1.4.1. *State-function based models*

Let us suppose that the FET has been stripped of all external parasitic elements and that the DC and SS parameters of the *intrinsic device* are thus made available. This has two advantages: first, the intrinsic device is

closer to a voltage-controlled two port than the full device, and its black-box representation is made easier. Secondly, the removal of device parasitics is an important step if a *scalable* model is sought, *i.e.* a model able to simulate devices with different gate widths. This requirement is very important in integrated device modelling and in MMIC CAD, and has to account for the fact that parasitic elements and the parameters of the intrinsic device have different scaling rules.

Device physics suggest for the quasi-intrinsic FET a *voltage controlled* representation, the controlling voltages being the gate and drain voltages with respect to the source. With some inaccuracy, we can define v_{GS} and v_{DS} as the *state variables* of the model. The principle of state-function models, first proposed by Root [3] is that gate and drain currents can be approximated, as a function of the driving voltages, by superimposing the contributions of time-domain linear differential operators applied to *state functions* defined as nonlinear functions of the state variables (the driving voltages). According to this ansatz, the Root model postulates [4]:

$$i_{DS} = \sum_{k=0}^{\infty} h_{DS}^{(k)}(d/dt) X_{DS}^{(k)}(t) \quad (15)$$

and similarly:

$$i_{GS} = \sum_{k=0}^{\infty} h_{GS}^{(k)}(d/dt) X_{GS}^{(k)}(t) \quad (16)$$

where the state functions $X_{DS}^{(k)}$, $X_{GS}^{(k)}$ depend on the instantaneous values of the driving voltages:

$$X_{DS}^{(k)} = X_{DS}^{(k)}(v_{GS}(t), v_{DS}(t)) \quad (17)$$

$$X_{GS}^{(k)} = X_{GS}^{(k)}(v_{GS}(t), v_{DS}(t)) \quad (18)$$

and the dynamic operators $h_{DS}^{(k)}(d/dt)$, $h_{GS}^{(k)}(d/dt)$ can be approximated as rational functions of d/dt . A possible approximation suggested by Root [4] is to truncate the series of the drain current to the third term, setting:

$$h_{DS}^{(0)} = \frac{1}{1 + \tau \frac{d}{dt}} \quad (19)$$

$$h_{DS}^{(1)} = \frac{d}{dt} \quad (20)$$

$$h_{DS}^{(2)} = \frac{\tau \frac{d}{dt}}{1 + \tau \frac{d}{dt}} \quad (21)$$

while, for the gate current, the series is truncated to the second term, with:

$$h_{GS}^{(0)} = 1 \quad (22)$$

$$h_{GS}^{(1)} = \frac{d}{dt}. \quad (23)$$

Concerning the meaning of the state variables, $X_{DS}^{(0)}$ coincides with the DC drain current; $X_{DS}^{(1)}$ is the drain charge Q_{DS} , and $X_{DS}^{(2)}$ is the drain current measured at high frequency, *i.e.* beyond the cutoff frequencies of trapping and thermal effects. One can easily notice that the first state variable does not contribute for fast variations of the driving voltages, while the third state variable only contributes at high frequency; the second term is capacitive. $X_{GS}^{(0)}$ is the DC gate current (negligible in the active region, not necessarily so in breakdown or direct gate conduction), while $X_{GS}^{(1)}$ is the gate charge Q_{GS} .

The DC characterization of the device directly provides the DC drain and gate currents. The gate and drain charges as a function of the driving voltages can be recovered by integrating the small-signal capacitances, obtained from the imaginary part of the small-signal admittances as a function of frequency. Finally, the high-frequency drain current is obtained from the measured RF transconductance and drain conductance. However, measurements only provide the *partial derivatives* of the gate and drain charge and of the high-frequency drain current. Calling $X = X(v_{GS}, v_{DS})$ one of these state functions, measurements yield the gradient of X with respect to the driving voltages:

$$\nabla X = \frac{\partial X}{\partial v_{GS}} \hat{v}_{GS} + \frac{\partial X}{\partial v_{DS}} \hat{v}_{DS}. \quad (24)$$

It is well known, however, from the theory of vector fields, that if X is a scalar field, then its gradient must be conservative, *i.e.* the line integral of the gradient along a closed path must vanish. This condition is equivalent to the requirement that the second-order mixed partial derivative of X must be independent of the derivation order. If the gradient is conservative, X can be recovered by integrating ∇X with respect to any direction, and starting from an arbitrary point P_1 in the (v_{GS}, v_{DS}) plane, see Fig. 2.

Because of the already mentioned dependence of the device current on LFD and thermal effects, measured FET small-signal parameters often do not yield exactly conservative data [17], above all for the high-frequency drain current i_{DS}^{RF} . In this case, i_{DS}^{RF} cannot be uniquely defined through integration of the small-signal parameters, see Fig. 3. If LFD and thermal effects are not dramatic, the measured data can be slightly renormalized so

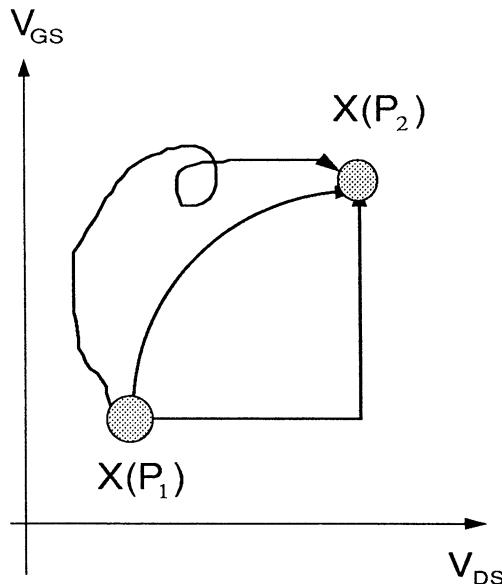


Figure 2. State function identification by integration of its gradient along an arbitrary path, conservative case.

as to make them ideally conservative. It should be stressed that, since conservation of the high-frequency current is implicit in the model structure, non-conservative data inserted into the model lead to spurious solutions, since non-conservative components driven along a cycle in phase space do not reach the same condition cycle after cycle; this behaviour is somewhat similar to the behaviour of charge-pumped MOSFET devices modelled with nonconservative charge models.

The Root model is non-quasistatic, since the low- and high-frequency behaviour of the VI characteristics are not necessarily the same. Other non-quasistatic models akin to the Root model have been recently presented; an example is the relaxation time model proposed in [18, 19]. While, in its simplified form [4], the Root model postulates an instantaneous dependence of the terminal charge on the driving voltages, thus neglecting, for the gate charge, the effect of the intrinsic or channel resistance, the model in [18] recovers this effect by postulating a model where the gate and drain charges are subject to a relaxation equation:

$$\left(1 + \tau_G \frac{d}{dt}\right) Q_G = Q_G^{ss}(v_{GS}, v_{DS}) \quad (25)$$

$$\left(1 + \tau_D \frac{d}{dt}\right) Q_D = Q_D^{ss}(v_{GS}, v_{DS}). \quad (26)$$

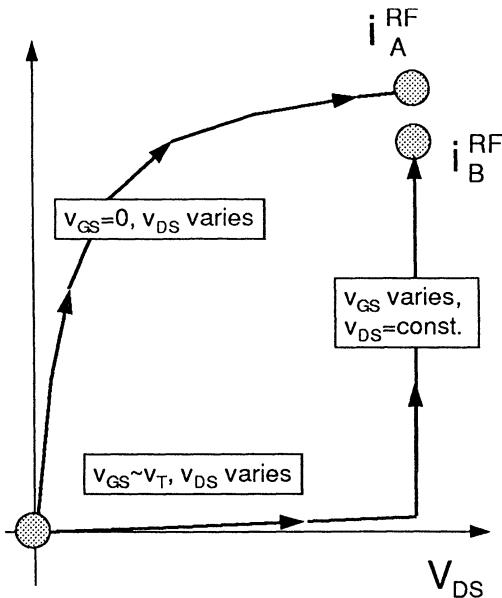


Figure 3. Non-unique state-function identification by integration of its gradient along an arbitrary path, non-conservative case.

The relaxation times are in general a function of the instantaneous working point; they can be evaluated with reference to the high-frequency behaviour of the imaginary part of the small-signal parameters.

In conclusion, state-function models, coupled to an automated DC and small-signal measurement system, have shown excellent potentialities in the device-independent FET characterization. Their only limitation seems to be, at present, connected to the built-in conservative properties of the model, which make some compromise indispensable in the treatment of low-frequency dispersion effects and thermal effects.

1.4.2. The nonlinear integral model

The Volterra series approach has been recently revisited, in an approximate form, by the so-called Nonlinear Integral Model (NIM) proposed by Filicori *et al.* [5]. The NIM is a technology-independent, look-up table based model which can be directly derived from conventional measurements (DC and small signal) carried out in several working points. The basic idea of the NIM is to derive, from the Volterra series model, a nonlinear model which is exact in the two limits: (a) linear device with infinite bandwidth, or, as it is often said, infinite memory; (b) nonlinear, memoriless device. In every intermediate case, the model is approximate only. Thus, the DC characteristics are exactly recovered, and so are the small-signal bias-dependent

parameters. In order to introduce the NIM, let us consider, for the sake of simplicity, a one-port device; the treatment of multiports is fully analogous. Let us suppose that the device is voltage controlled; thus, algebraic manipulations carried out on the conventional Volterra series lead to the following integral series:

$$\begin{aligned} i(t) &= F_{DC}(v(t)) + \\ &+ \int_{t-\tau_m}^t G^{(1)}(v(t), t-\tau_1)(v(\tau_1)-v(t))d\tau_1 + \\ &+ \iint_{t-\tau_m}^t G^{(2)}(v(t), t-\tau_1, t-\tau_2)(v(\tau_1)-v(t))(v(\tau_2)-v(t))d\tau_1 d\tau_2 + \\ &+ \iiint_{t-\tau_m}^t G^{(3)}(v(t), t-\tau_1, t-\tau_2, t-\tau_3) \times \\ &\quad \times (v(\tau_1)-v(t))(v(\tau_2)-v(t))(v(\tau_3)-v(t))d\tau_1 d\tau_2 d\tau_3 + \dots \end{aligned} \quad (27)$$

where $i(t)$ and $v(t)$ are the instantaneous voltage and current of the device. Owing to the presence, in Eq. (27), within the convolution integrals, of voltage-dependent kernels $G^{(n)}(v(t), t-\tau_1, \dots, t-\tau_n)$ and terms $v(\tau)-v(t)$, the convergence of the series is fast not only when the nonlinearity is mild (as in the classical Volterra series case) but also when the nonlinearity is strong, provided that this is characterized by a short memory with respect to the operating bandwidth. In fact, in such conditions the terms $v(\tau)-v(t)$ are small even when the voltage signal v has large amplitude. To be practically effective, the NIM series must be truncated to the second term, thus yielding the simple expression:

$$i(t) \approx i_{DC}(v(t)) + \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{Y}(v(t), \omega) V(\omega) \exp(j\omega t) d\omega \quad (28)$$

with:

$$\tilde{Y}(v(t), \omega) = Y(v(t), \omega) - Y(v(t), 0). \quad (29)$$

$V(\omega)$ is the Fourier transform of the driving voltage. Clearly, the above formulation can be also symbolically expressed under the form:

$$i(t) = i_{DC}(v(t)) + \tilde{Y}(v(t), d/dt)v(t). \quad (30)$$

Since $\tilde{Y}(v(t), 0) = 0$ by definition, i_{DC} simply are the DC device characteristics. The dynamic admittance parameters can be directly measured at different working points so as to yield a look-up table representation. In order to correctly appreciate the meaning of Eq. (30), it should be noticed that the d/dt operator in $\tilde{Y}(v(t), d/dt)v(t)$ acts only on the $v(t)$ which follows the operator, or, if we write for the sake of clarity $\tilde{Y}(v(t), d/dt)v(t) =$

$\tilde{Y}(v_a(t), d/dt)v_b(t)$ with $v \equiv v_a \equiv v_b$, d/dt only acts on $v_b(t)$. Thus, the NIM is a *parametric* model in v_a .

The truncated expression of the NIM for a two-port device is formally similar to the expression given above for a one port. Owing to its parametric nature, the NIM is not affected by the lack of data conservativeness; as an example of this, let us consider a simple NIM model for the drain current (capacitive effects excluded):

$$i_{DS} = i_{DC}(\underline{v}) + [\underline{g}^{RF}(\underline{v}) - \underline{g}^{DC}(\underline{v})] \cdot \frac{\tau \frac{d}{dt}}{1 + \tau \frac{d}{dt}} \underline{v} \quad (31)$$

where $\underline{v} = (v_{GS}, v_{DS})$ and the corresponding small-signal parameters $\underline{Y} = (Y_{DG}, Y_{DD})$ are:

$$\underline{Y}(\underline{v}, \omega) = \frac{j\omega\tau}{1 + j\omega\tau} \underline{g}^{RF}(\underline{v}) + \frac{1}{1 + j\omega\tau} \underline{g}^{DC}(\underline{v}). \quad (32)$$

In the RF limit one has:

$$i_{DS} = i_{DC}(\underline{v}(t)) + [\underline{g}^{RF}(\underline{v}(t)) - \underline{g}^{DC}(\underline{v}(t))] \cdot \underline{v}(t) \quad (33)$$

which is *always* an instantaneous function of the state variables \underline{v} , independent of the conservation properties of \underline{g}^{RF} .

Although the NIM is a non-quasistatic model, the treatment of LFD and thermal effects is approximate only. As far as LFD effects are concerned, it can be shown [20] that the NIM is exact if (see Eq.(1)) i_{DS} is a *linear* function of q_t , and Q_t^{DC} is a *linear* function of v_{GS} and v_{DS} . Since the linear approximation is inappropriate for thermal effects (because the dissipated power, and therefore the temperature, is a strongly nonlinear function of the driving voltages), the NIM is not expected to provide an accurate solution unless the temperature or the dissipated power are assumed as a controlling variable and the model is modified accordingly. A possible technique for this is suggested in [20]. Simulations and experimental results [5, 20] have confirmed the good accuracy of this non-linear integral model in the large-signal performance prediction of GaAs FETs.

1.5. EQUIVALENT CIRCUIT PHEMT MODELS

This section is devoted to the equivalent-circuit large-signal modelling of PHEMTs. Despite the progress made in black-box approaches, it should be considered that the equivalent-circuit approach still is the most popular choice in large-signal modelling, owing to the widespread availability of

such models in all circuit simulators. During the last few years, predefined models (whose equations cannot be changed by the user) have been gradually replaced, in most circuit simulators, by user-defined models in which, on the basis of a standard topology, arbitrary equations can be entered in symbolic form to define the behaviour of the lumped elements of the equivalent circuit with respect to the working point.

When compared to the black-box approach, the equivalent-circuit approach relies more heavily on the experience of the user and on the effectiveness of the optimization techniques needed to perform the parameter fitting on a suitable set of measured characteristics. In a way, the literature on equivalent circuit models does not yield a direct and unique solution to model development, but rather resembles to a collection of recipes from which the experienced user can sort out the best solution to his own modelling problem. The increased flexibility offered by user-defined models and powerful optimization tools [21] allows today the circuit designer to obtain, in a reasonable time, a set of models tailored to his specific application. Moreover, facilities in data handling offered by automated measurement set-ups make it possible to base the model fitting on large sets of DC, small signal and even large-signal parameters. Despite this, equivalent-circuit modelling still is more like an art than a straightforward and exact method, and much is left to the insight and ingenuity of the model designer.

A typical, though unwelcome, feature of the equivalent circuit approach is its being, at least up to a certain extent, device-dependent. Since the analytical approximations which describe the behaviour of the circuit elements with respect to the bias point cannot be arbitrary, any feature not included into the model equations from the start cannot be accurately simulated. Thus, the first step in the modelling of a new device is to understand its peculiarities, also with respect to the technologically closest devices already known. For the HEMT, the obvious next-of-kin is the MESFET.

Despite their similarity, the MESFET and the HEMT differ under a few respects. The main point of discrepancy concerns the different behaviour of the transconductance *vs.* the gate voltage. While in MESFETs the transconductance increases as a function of v_{GS} from pinchoff to direct gate conduction, the HEMT transconductance has a maximum in correspondence of the onset of parallel conduction in the supply layer, and drops beyond this point, see Fig. 4. Gain compression cannot be neglected, since PHEMTs usually operate near the maximum gain bias point. This occurs because, owing to the larger built-in voltage of the gate Schottky junction and to the reduction of pinch-off voltage in recent devices, the operating point for maximum power turns out to be close to the maximum gain operating point.

Several approaches have been proposed to incorporate transconductance

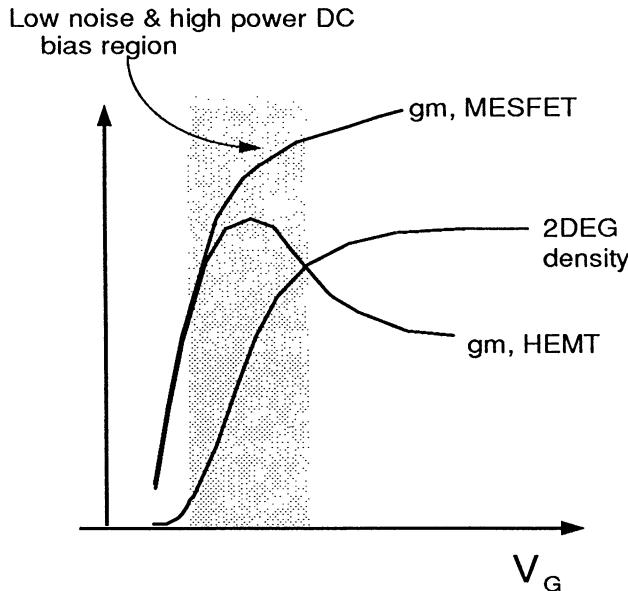


Figure 4. Behaviour of the transconductance *vs.* bias in MESFETs and HEMTs

compression into the output current generator of the lumped device model. Golio *et al.* [22, 23] have proposed modification of existing MESFET models based on two-region approaches. In these models, two different analytical expressions are introduced, one for gate voltages below the maximum transconductance, one for gate voltages above it. Although this approach leads to fairly accurate results when properly fitted, two-region approximations lead to difficulties when implemented into general-purpose simulators as user-defined models. A different approach was proposed by Angelov *et al.* [24], whereby a unique expression incorporates the behaviour below and above the compression point.

Concerning capacitance modelling, most authors exploit for the HEMT the same analytical approximations which are currently used for the MESFET. This assumption seems to be reasonable, since the experimental capacitance behaviour of PHEMTs is close enough to the one of long-gate MESFETs. Special expressions were proposed by Angelov [24], which, however, do not lead to far better results than the ordinary diode-like expressions included in MESFET models such as the Curtice [9] or the Curtice-Ettenberg [10] model.

Finally, the impact of LFD and thermal effects on HEMTs is somewhat controversial. Since LFD phenomena are induced by the dynamics of trapping centers located in the supply layer, such effects are expected

to be negligible for gate voltages below the saturation of the channel electron charge and the corresponding transconductance maximum. Since the typical low-noise operating bias points fall within this region, HEMTs for low-noise application do not experience dramatic LFD effects, nor thermal effects, owing to the low power level. In devices for medium- or high-power applications, on the other hand, the operating region above the maximum g_m is interested and the dissipated power grows, thus leading to significant thermal effects. Although circuit techniques based on the implementation of low-pass RC networks (see e.g. [14]) have been proposed to derive equivalent circuits with non-quasistatic properties, this is, in practice, a waste of effort. In fact, the lowest harmonics relevant to microwave circuit operation (at least under single-tone excitation) are the DC and the signal frequency, well below or above, respectively, the inverse of the relaxation times typical of trap or temperature dynamics. Thus, to all practical effects, a smooth transition between the DC and RF behaviours is immaterial, and the model can be made depend on two sets of variables, the instantaneous (RF) values of voltages and currents, and the bias values. In many circuit simulators the DC value is directly available; at any rate, it can be recovered, also in the case in which strong AC-DC conversion is present, through a low-pass filter with very low cutoff frequency. Low-pass filtering, coupled to a careful use of controlled generators, enables to recover, as a control variable, the average power dissipated by the device, thus allowing a simple thermal model to be incorporated into the equations of equivalent circuit models.

1.6. A NON-QUASISTATIC PHEMT MODEL

As an example of equivalent circuit modelling, we consider the large-signal modelling of PHEMTs for Ka-band medium-power amplifiers developed, manufactured and characterized by SIEMENS [25]. The model, whose structure is shown in Fig. 5, was implemented as a user-defined equation set within the framework of the HARPE/OSA [21] simulator, and the model structure was derived, starting from Angelov's equations, through work on three successive generations of PHEMTs which we shall denote hereafter as PHEMT1, PHEMT2 and PHEMT3. This adaptive approach to equivalent circuit modelling is fairly typical in the presence of a rapidly evolving technology. Later, the model was also exported to other circuit simulators (MDS). With respect to the original equations, many significant effects were included into the model: LFD effects, soft breakdown (or DC kink), current compression due to thermal effects. The resulting model yields very good agreement of the DC and small-signal parameters in a wide range of bias points; preliminary large-signal measurements have also shown good

agreement. The model for the drain current reads:

$$i_{DS} = \frac{i_1(v_{DS}, v_{GS})}{1 + (K_\theta \langle v_{DS} i_1 \rangle)^\chi} + G_D^{RF}[v_{DS} - V_{DS}] + G_m^{RF}[v_{GS} - V_{GS}] \quad (34)$$

where the DC current is described by:

$$i_1 = I_{pk}[1 + \tanh(\Psi)] \tanh(\alpha v_{DS})(1 + a_{sb}(v_{DS} - v_{GS})^3] \quad (35)$$

$$\begin{aligned} \Psi = & P_1(v_{GS} - V_{pk}) + P_2(v_{GS} - V_{pk})^2 + P_3(v_{GS} - V_{pk})^3 + \\ & + P_4(v_{GS} - V_{pk})^4 + P_5(v_{GS} - V_{pk})^5 \end{aligned} \quad (36)$$

and the parameters for the non-quasistatic model are:

$$G_D^{RF} = R_0 + R_1 V_{GS} + R_2 V_{GS}^2 \quad (37)$$

$$G_m^{RF} = R_3 V_{DS} i_1(V_{DS}, V_{GS}). \quad (38)$$

V_{pk} is the voltage where the output transconductance reaches its maximum; the coefficient a_{sb} models the soft breakdown effect. With $a_{sb} = 0$, i_1 is the original Angelov model. The model includes thermal effects; lowercase letters stand for instantaneous values, while uppercase variables are DC values. The term $\langle v_{DS} i_1 \rangle$ (the symbol $\langle \cdot \rangle$ stands for low-pass averaging obtained through a proper symbolic equivalent circuit) is an approximation of the average dissipated power, K_θ is an empirical coefficient (with dimension W^{-1}) describing the sensitivity of the drain current to the dissipated power, while χ is an additional fitting parameter whose value is close to 1.

Finally, diode-like current generators were introduced into the model according to standard approaches [10] in order to account for strong breakdown effects and direct gate conduction. Although these generators are often neglected as trivial in the discussion of large-signal equivalent circuits, it may be recalled that gain compression in the large-signal regime ultimately depends on the value of the breakdown voltage and on the range where direct gate conduction comes into play.

The charge model is:

$$Q_{GS} = C_{GS}^0[1 + \tanh(k_2 v_{DS})] \left\{ v_{GS} + \frac{\log[\cosh(k_1 v_{GS})]}{k_1} \right\}, \quad (39)$$

$$Q_{DG} = C_{GD}^0[1 + \tanh(k_3 v_{GS})] \left\{ v_{DG} - \frac{\log[\cosh(k_4 v_{DS} + k_5 v_{DS} v_{GS})]}{k_4 + k_5 v_{GS}} \right\} \quad (40)$$

from which the following capacitance model is obtained:

$$C_{GS} = C_{GS}^0[1 + \tanh(k_1 v_{GS})][1 + \tanh(k_2 v_{DS})] \quad (41)$$

$$C_{DG} = C_{GD}^0[1 + \tanh(k_3 v_{GS})][1 - \tanh(k_4 v_{DS} + k_5 v_{DS} v_{GS})]. \quad (42)$$

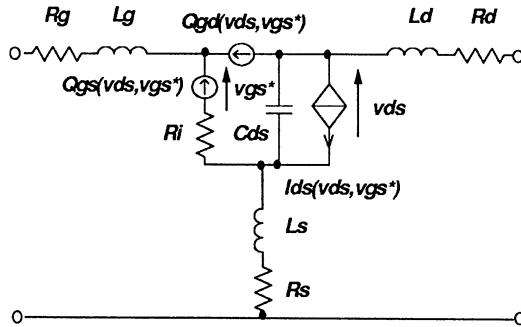


Figure 5. Structure of large-signal HEMT model.

The HARPE/OSA simulator directly exploits the charge equations; based on its derivation, the capacitance model is conservative by definition.

Some results are presented concerning the second- and third-generation devices PHEMT2 and PHEMT3. The large-signal model was optimized through HARPE on the DC and scattering parameters measured in several working points on the frequency band 1-26 GHz (PHEMT2) and 1-50 GHz (PHEMT3). For PHEMT2, an example of the results obtained is shown in Fig. 6 for the measured and simulated DC curves and in Fig. 7 for the measured and simulated scattering parameters in the class A operating point; good agreement is also obtained for other bias points.

The modelling of PHEMT3 devices posed some challenge, since the device shows a fairly large amount of DC heating that partly compensates for the soft breakdown behaviour, thus introducing considerable low-frequency dispersion in the transconductance and output resistance. In order to carry out the model fitting on the DC and scattering parameters, the model parameters were divided into two classes: those mainly influencing the DC behaviour, and those only influencing the AC behaviour. A series of “orthogonal” optimizations was then performed, also taking into account the different parameter sensitivity with respect to the objective function, which can be evaluated through HARPE. Fig. 8 shows the measured and simulated DC curves for the PHEMT3 device; the agreement obtained is excellent over the whole measured range. The device was to be exploited for a medium-power class A amplifier with bias point around $V_{GS} = 0$ V, $V_{DS} = 5$ V; for this working point, the measured and simulated scattering parameters on the frequency band 1-50 GHz are compared in Fig. 7. The agreement obtained is satisfactory also for different values of the working point.

Fig. 9 shows the measured and simulated $P_{in} - P_{out}$ curve for the fundamental harmonic. The source and load impedances are $Z_S = 4.66 + j16.04$

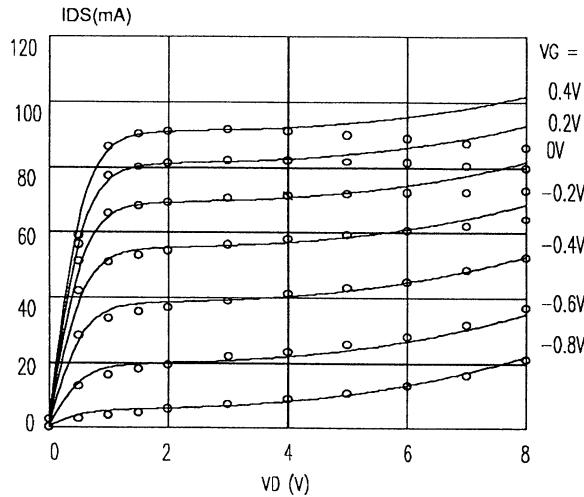


Figure 6. Measured and simulated DC curves of PHEMT2 device.

Ω , $Z_L = 50.27 \Omega$, the operating frequency is 10 GHz and the bias point is $V_{DS} = 5$ V, $I_{DS} \approx 70$ mA. The measured data were obtained from a device belonging to the same wafer as the one on which the DC and small-signal measurements were carried out; thus, a small difference in the small-signal gain can be detected, although the measured and simulated devices were polarized so as to obtain the same output current. The error bars in the measured data were extrapolated by means of a MonteCarlo scatter plot based on available data on DC and small-signal parameter uniformity, and refer to the estimated standard deviation which turns out to be of the order of 0.5 dBm.

2. Noise models

2.1. INTRODUCTION

Conventional and pseudomorphic HEMTs are typical *low noise* devices. This section is devoted to a review of the noise parameters of two-ports and to the noise performances of microwave FETs. A short overview on physics-based noise modelling techniques is also provided.

The physical origin of noise in semiconductor devices lies in carrier transport mechanisms [26, 27, 28]. During their history, carriers undergo a sequence of *free flights* and *scattering events* (*i.e.*, interactions with impurities, phonons, other carriers, etc.). Since the individual carrier veloci-

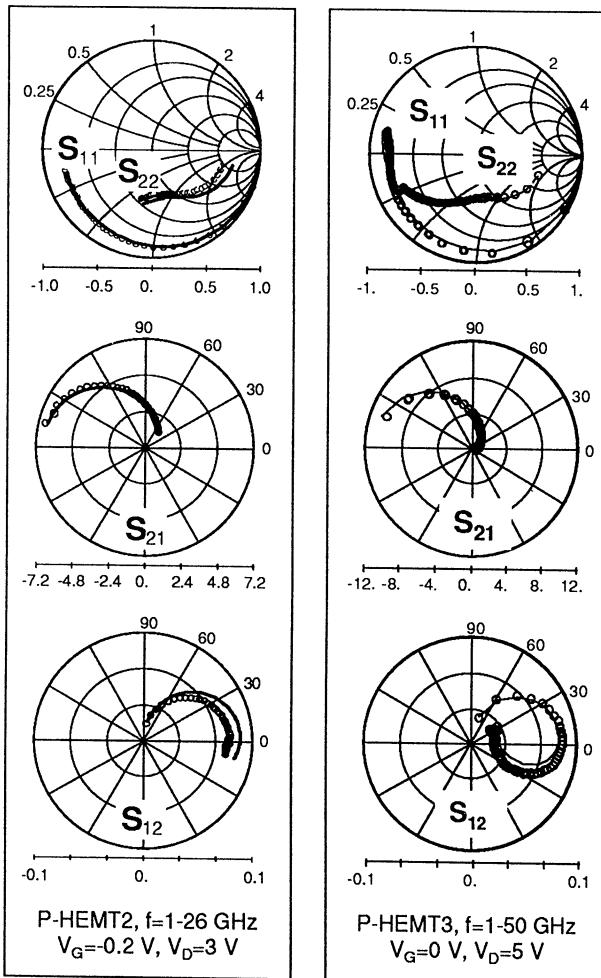


Figure 7. Examples of measured and simulated scattering parameters of PHEMT2 and PHEMT3 devices.

ties experience wild fluctuations mainly induced by scattering events, also the average velocity of a carrier ensemble *fluctuates*, although, as statistics suggests [29], ensemble fluctuations are much smaller than individual fluctuations [29].

In low-field conditions the semiconductor is near thermodynamic equilibrium and velocity fluctuations yield *thermal noise*, meaning that the power spectrum of velocity fluctuations is proportional to the equilibrium temperature of the system. With increasing electric field, high-field trans-

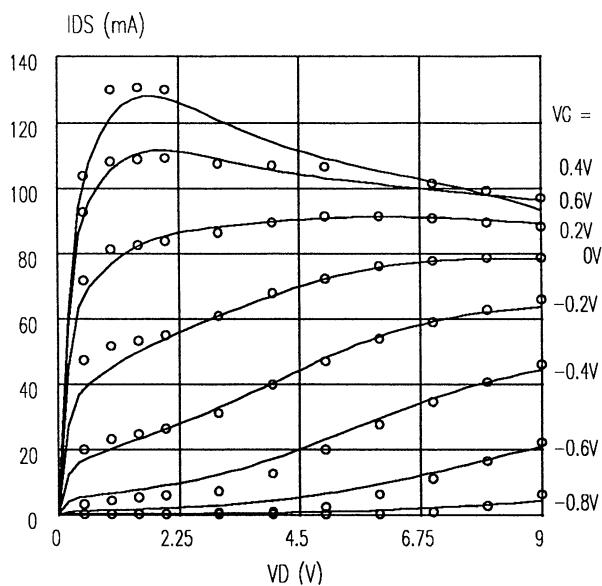


Figure 8. Simulated and measured DC curves of PHEMT3 device.

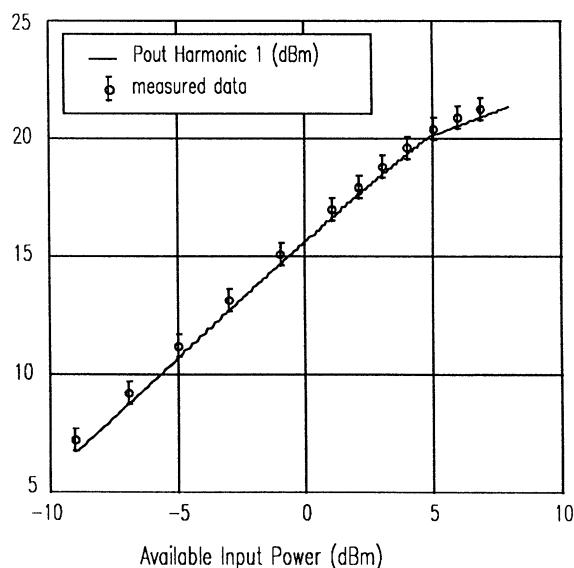


Figure 9. Measured and simulated power sweep of PHEMT3.

port prevails, and the carrier fluctuations can be characterized in terms of the more general *diffusion noise* [27], whereby the carrier velocity fluctuation spectrum is proportional to the carrier diffusivity. Also for diffusion noise, a generalized interpretation in terms of thermal noise is possible, provided that the carriers are assigned a proper thermodynamic temperature, which is the *electron temperature* for an electron ensemble. Diffusion noise is *white*, i.e. frequency independent, up to frequencies of the order of the inverse of the mean time between scattering events, well above the microwave range.

Another kind of semiconductor noise originates from population fluctuations induced by trapping and detrapping phenomena. Such fluctuations typically have low-pass power spectra, whose cutoff frequency is related to the inverse of the trap lifetime. The collective effect of traps having different lifetimes yields a low-frequency fluctuation spectrum that decreases with increasing frequency. While the interpretation of low-frequency noise still is somewhat controversial, the experimental behaviour of the power spectrum can be approximated as $1/f^\alpha$, with α around 1. The $1/f$ noise, as it is commonly called, is present in all microwave devices. Since both velocity fluctuations and population fluctuations ultimately lead to *current density fluctuations*, the power spectrum of the current density fluctuations in a bulk semiconductor subject to diffusion and $1/f$ noise typically takes the appearance shown in Fig. 10.

The crossing between the $1/f$ noise and the white (diffusion noise) plateau takes place between 100 MHz and 1 GHz in MESFETs and HEMTs. Since the operating frequency of most microwave circuits is above this limit, $1/f$ noise could appear as irrelevant to microwave operation. This is indeed true for linear circuits such as low-noise amplifiers; however, nonlinear circuits (for instance oscillators) up-convert low frequency noise through amplitude modulation. For instance, oscillator noise around the carrier frequency is mainly up-converted $1/f$ noise. In this context, however, we shall neglect $1/f$ noise and only consider the white part of the fluctuation spectrum, relevant to linear microwave operation.

The basic source of electrical noise in field effect transistors are the fluctuations of the channel current, which directly cause noise in the short-circuit drain current [30]. Capacitive coupling to the gate also yields induced fluctuations in the short-circuit gate current. Short-circuit current fluctuations can be interpreted as random current generators at the input and output of the device, see Fig. 11. Since the ultimate microscopic origin of input and output fluctuations is the same, statistical correlation may exist between the two generators. From the standpoint of the external circuit, then, the noise behaviour of a FET (and, generally, of a two-port device) can be described by the power spectra of the short-circuit current fluctu-

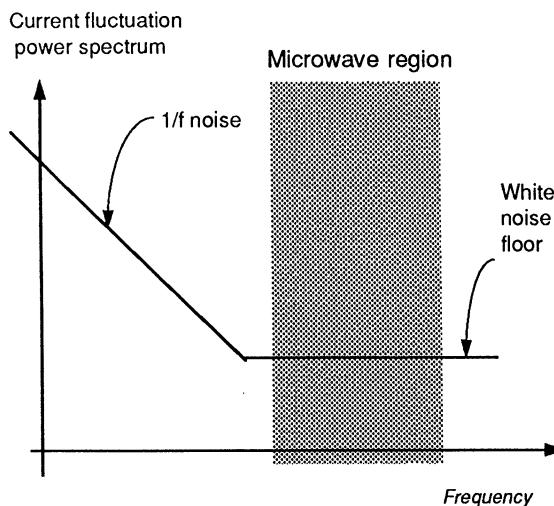


Figure 10. Power spectrum of current fluctuations in a bulk semiconductor.

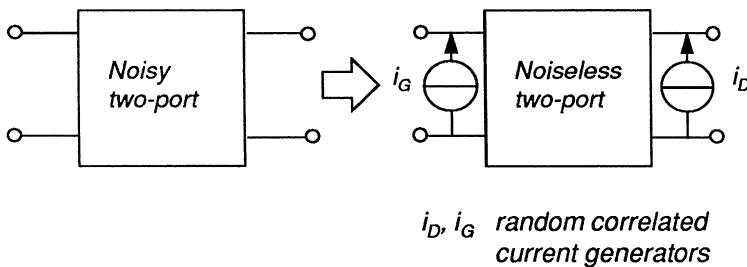
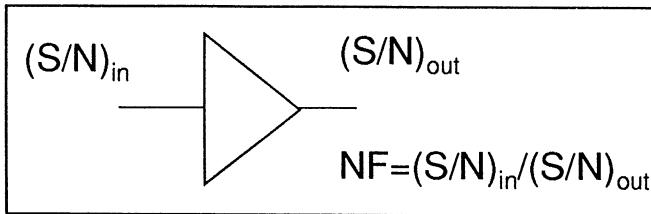


Figure 11. Representation of a noisy two-port in terms of short-circuit current fluctuations.

ations and by their correlation spectrum; power spectra are real functions of frequency, while the correlation spectrum is a complex function of frequency. Alternative representations are possible in terms of open-circuit voltage fluctuations or mixed combinations between short-circuit current and open-circuit voltage fluctuations.

The noise characterization in terms of current or voltage fluctuation spectra and their correlation is often replaced, at a circuit or system level [26, 31], by overall noise parameters that provide information on the maximum signal-to-noise ratio achievable at the device output, the optimum way to connect the device so as to minimize its noise, and the sensitivity of noise to a change with respect to the optimum conditions. Having defined

- Minimum NF (noise figure) F_{\min}



- Noise resistance R_N
- Optimum source impedance
 $Z_{opt} = R_{opt} + jX_{opt}$

Figure 12. System-oriented noise parameters.

the *noise figure* of a two-port as the ratio between the input and output signal-to-noise ratios (Fig. 12), the relevant system noise parameters are the *minimum noise figure*, defined as the minimum ratio which can be achieved, at a given working point, between the input and output signal-to-noise ratios; the *optimum source impedance* $Z_{s,opt}$, i.e., the generator impedance that minimizes the noise figure; and the *noise resistance*, a measure of the noise figure sensitivity to changes of the source impedance with respect to the optimum value. It may be noticed that the system noise parameters amount to four real functions of frequency, exactly as the intrinsic noise parameters (power and correlation spectra).

Experimental and theoretical considerations suggest that the minimum noise figure increases with frequency. A few experimental data (from [32] and references therein) on the noise behaviour of microwave FETs are reported in Fig. 13. Besides the noise figure, another important figure of merit for a low-noise device is the gain in the optimum noise conditions, also called the *associated gain* G_{ass} . Since the bias and loading conditions for minimum noise and maximum gain do not coincide, the associated gain can be significantly lower than the maximum device gain. The associated gain of the microwave FETs of Fig. 13 is shown in Fig. 14. The superiority of PHEMTs over MESFETs as low-noise devices is not only due to the lower noise figure, but also to the higher associated gain of these devices.

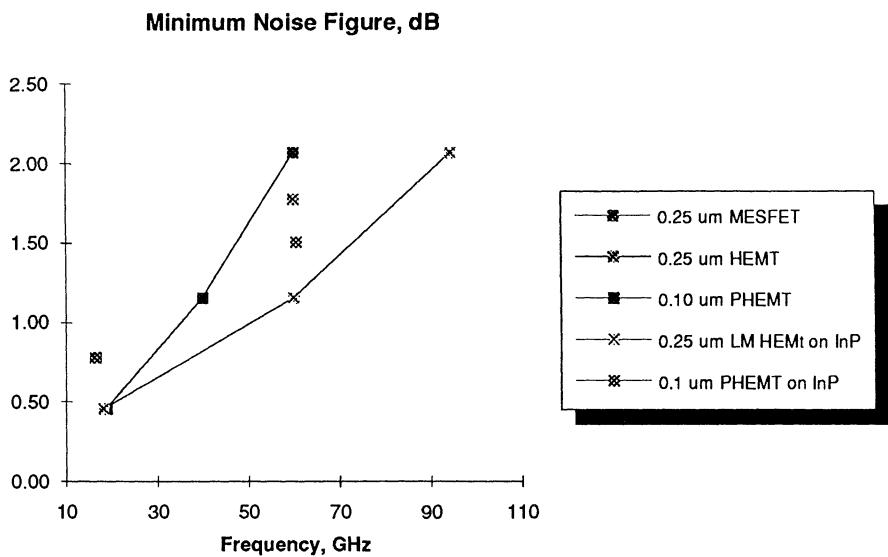


Figure 13. Minimum noise figure of microwave FETs [32].

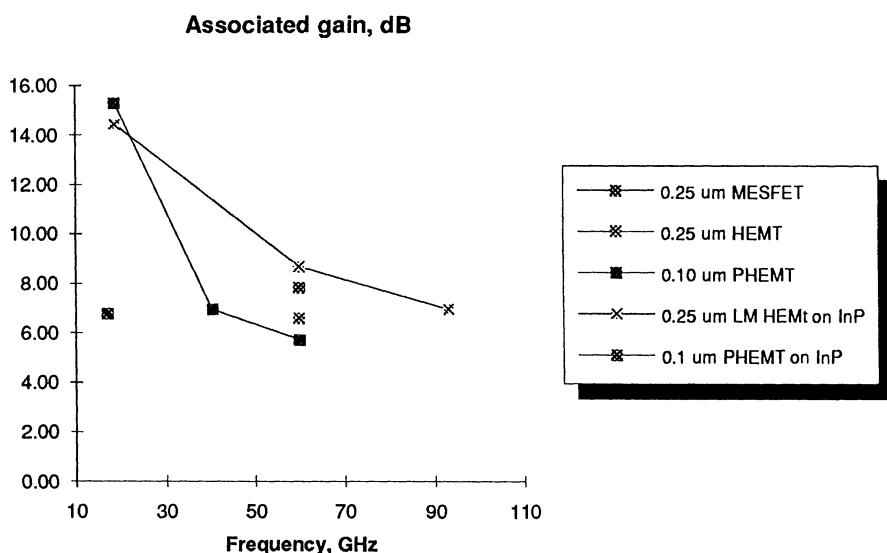


Figure 14. Associated gain of microwave FETs [32].

2.2. TWO-PORT NOISE CHARACTERIZATION

It is well known from basic circuit theory that a linear and autonomous one-port can be represented by a Norton and/or Thevenin equivalent circuit, *i.e.* it is equivalent to a frequency-dependent admittance and/or impedance. The equivalent circuit of a noisy one-port also includes a short circuit current generator i_n (Norton) or an open circuit voltage source e_n (Thevenin). These random generators are statistically characterized by their unilateral power spectra $S_{e_n e_n}(f)$ and $S_{i_n i_n}(f)$, related to the quadratic averages $\overline{|e_n|^2}$ and $\overline{|i_n|^2}$ by:

$$\overline{|x_n|^2} = \int_0^{+\infty} S_{x_n x_n}(f) df \quad (43)$$

where $x \equiv e$ or $x \equiv i$. For *white noise* one has $\overline{|e_n|^2} = S_{e_n e_n} \Delta f$, $\overline{|i_n|^2} = S_{i_n i_n} \Delta f$ where Δf is the system bandwidth. Finally, thermal noise is characterized, according to the Nyquist theorem, as $\overline{|e_n|^2} = 4kTR \Delta f$, $\overline{|i_n|^2} = 4kTG \Delta f$, where $R = \text{Re}\{Z\}$ and $G = \text{Re}\{Y\}$.

The representation of a noisy multiport is similar. A noisy two-port can be completely identified by the small-signal parameters and the power and correlation spectra of the noise generators [31]. According to the small-signal representation chosen, the noise generators can be short-circuit noise current generators (admittance characterization), open-circuit noise voltage generators (impedance characterization) or hybrid generators (H matrix).

Thus, the noise characterization of a two-port amounts to the knowledge of two power spectra and one correlation spectrum, *i.e.*, $S_{e_{n1} e_{n1}}$, $S_{e_{n2} e_{n2}}$, $S_{e_{n1} e_{n2}}$ for the series representation and $S_{i_{n1} i_{n1}}$, $S_{i_{n2} i_{n2}}$, $S_{i_{n1} i_{n2}}$ for the parallel representation. As already recalled, power spectra are real, while correlation spectra are in general complex; therefore, all the noise information amounts, for a two-port, to *four* real functions of frequency. The normalized correlation spectra, commonly called *correlation coefficients*, are often used instead of the simple correlation spectra:

$$C_{x_{n1} x_{n2}}(f) = \frac{S_{x_{n1} x_{n2}}(f)}{\sqrt{S_{x_{n1} x_{n1}}(f) S_{x_{n2} x_{n2}}(f)}} \quad (44)$$

where $x \equiv e$ or $x \equiv i$.

From a system standpoint, it is often convenient to transfer all noise generators to the input of the two-port, as shown in Fig. 15. To perform this task, it is helpful to take into account a few properties of power and correlation spectra of random processes that allow to exploit a sort of operational calculus for their manipulation. The starting point is the following property: consider two linear systems with outputs y_1 , y_2 and inputs x_1 ,

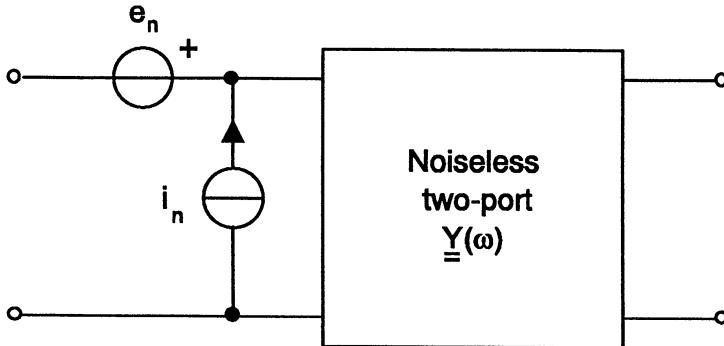


Figure 15. Representation with input correlated generators.

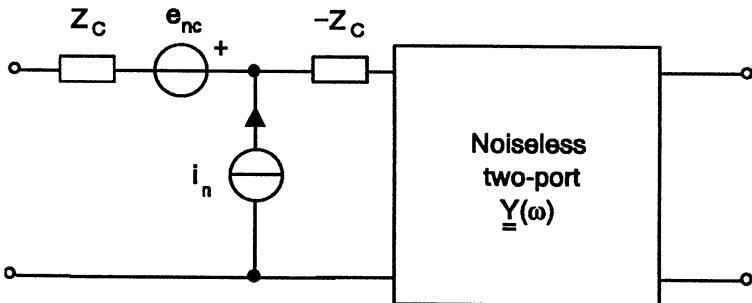


Figure 16. Series representation with input uncorrelated generators.

x_2 , characterized, in the frequency domain, by the transfer functions H_1 , H_2 as:

$$Y_1(\omega) = H_1(\omega)X_1(\omega) \quad (45)$$

$$Y_2(\omega) = H_2(\omega)X_2(\omega). \quad (46)$$

Then, the following relationships hold [29] between the input and output power and cross-correlation spectra:

$$S_{y_1 y_1} = |H_1|^2 S_{x_1 x_1} = H_1 H_1^* S_{x_1 x_1} \quad (47)$$

$$S_{y_2 y_2} = |H_2|^2 S_{x_2 x_2} = H_2 H_2^* S_{x_2 x_2} \quad (48)$$

$$S_{y_1 y_2} = H_1 H_2^* S_{x_1 x_2}. \quad (49)$$

Taking into account the above discussion, the two input generators can be easily related, e.g., to the open-circuit noise voltage generators as:

$$e_n = e_{n1} - e_{n2} Z_{11}/Z_{21} \quad (50)$$

$$i_n = -e_{n2}/Z_{21}. \quad (51)$$

The power and correlation spectra of the open-circuit noise voltage generators are explicitly expressed as a function of the corresponding parameters for the short-circuit noise current generators as:

$$S_{e_{n1}e_{n1}} = |Z_{11}|^2 S_{i_{n1}i_{n1}} + |Z_{12}|^2 S_{i_{n2}i_{n2}} + 2\operatorname{Re}\{Z_{11}Z_{12}^*\} S_{i_{n1}i_{n2}} \quad (52)$$

$$S_{e_{n2}e_{n2}} = |Z_{22}|^2 S_{i_{n2}i_{n2}} + |Z_{21}|^2 S_{i_{n1}i_{n1}} + 2\operatorname{Re}\{Z_{22}^*Z_{21}\} S_{i_{n1}i_{n2}} \quad (53)$$

$$\begin{aligned} S_{e_{n1}e_{n2}} = & Z_{11}Z_{21}^* S_{i_{n1}i_{n1}} + Z_{12}Z_{22}^* S_{i_{n2}i_{n2}} + \\ & + Z_{11}Z_{22}^* S_{i_{n1}i_{n2}} + Z_{12}Z_{21}^* S_{i_{n1}i_{n2}}^*. \end{aligned} \quad (54)$$

The above representation is still somewhat unsatisfactory, however, since the two generators are correlated. To obtain two uncorrelated generators, the following procedure can be exploited. Let us decompose the Fourier-transformed process e_n into the sum of two processes:

$$e_n = e_{nc} + e_c \quad (55)$$

where e_{nc} is uncorrelated from i_n , while e_c is completely correlated to i_n , hence:

$$e_c = Z_c i_n \quad (56)$$

where Z_c is a complex function of frequency called the *correlation impedance*. The correlation impedance is evaluated to implement the uncorrelation condition; from a circuit standpoint, Eq. (55) is equivalent to the circuit shown in Fig. 16. In terms of the power spectra and correlation coefficient of the open-circuit voltage fluctuations one has:

$$Z_c = Z_{11} - Z_{21} C_{e_{n1}e_{n2}} \sqrt{\frac{S_{e_{n1}e_{n1}}}{S_{e_{n2}e_{n2}}}} \quad (57)$$

$$S_{e_{nc}e_{nc}} = S_{e_{n1}e_{n1}} \left(1 - |C_{e_{n1}e_{n2}}|^2\right) \quad (58)$$

$$S_{i_ni_n} = S_{e_{n2}e_{n2}} \frac{1}{|Z_{21}|^2}. \quad (59)$$

To express the noise generators in a more manageable form, current noise sources are commonly associated to *noise conductances*, while voltage noise sources to *noise resistances*, with the help of a formal Nyquist relationship. Therefore we can define:

$$S_{e_{nc}e_{nc}} = 4kT r_n \quad (60)$$

$$S_{i_ni_n} = 4kT g_n \quad (61)$$

where T is the ambient temperature. As a conclusion, an equivalent set of noise parameters can be taken as the set g_n, r_n, R_c, X_c , where the two last parameters are the correlation resistance and reactance.

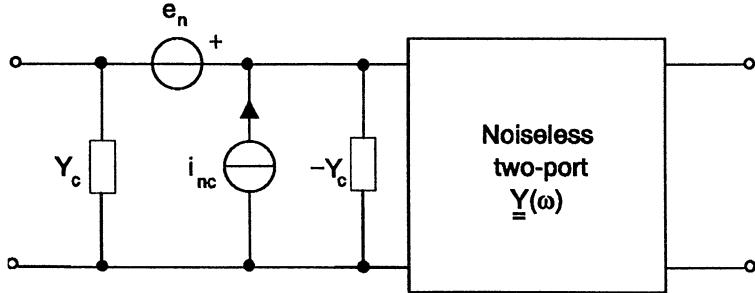


Figure 17. Parallel representation with input uncorrelated generators.

An equivalent (parallel) representation is obtained from the following decomposition:

$$i_n = i_{nc} + i_c \quad (62)$$

where i_{nc} is uncorrelated from e_n , and i_c is completely correlated to e_n . Thus:

$$i_n = i_{nc} + Y_c e_n \quad (63)$$

where Y_c is the *correlation admittance*. It should be noted that Y_c is *not* the reciprocal of Z_c . Eq. (63) represents the circuit of Fig. 17, which is the parallel counterpart of Fig. 16.

It can be shown that [31]:

$$Y_c = Y_{11} + Y_{21} C_{i_{n1} i_{n2}} \sqrt{\frac{S_{i_{n1} i_{n1}}}{S_{i_{n2} i_{n2}}}} \quad (64)$$

$$S_{i_{nc} i_{nc}} = S_{i_{n1} i_{n1}} \left(1 - |C_{i_{n1} i_{n2}}|^2\right) \quad (65)$$

$$S_{e_n e_n} = \frac{S_{i_{n2} i_{n2}}}{|Y_{21}|^2} \quad (66)$$

where $S_{i_{nc} i_{nc}}$ and $S_{e_n e_n}$ can be associated to a *noise conductance* G_n and a *noise resistance* R_n such that:

$$S_{i_{nc} i_{nc}} = 4kT G_n, \quad (67)$$

$$S_{e_n e_n} = 4kT R_n. \quad (68)$$

The series and parallel representations are connected by the relations [31]:

$$Y_c = \frac{Z_c^*}{|Z_c|^2 + r_n/g_n}, \quad (69)$$

$$G_n = \frac{r_n}{|Z_c|^2 + r_n/g_n}, \quad (70)$$

$$R_n = r_n + g_n |Z_c|^2. \quad (71)$$

2.2.1. Noise figure

Suppose to connect the input of the two-port to a source of internal impedance $Z_s = R_s + jX_s$. The noise figure of the two-port can be expressed as a function of the series circuit parameters as [26]:

$$F = 1 + \frac{r_n}{R_s} + \frac{g_n}{R_s} \left[(R_s + R_c)^2 + (X_s + X_c)^2 \right] \quad (72)$$

where $Z_c = R_c + jX_c$. In the above expression the noise figure is in natural units. The noise figure can be minimized by properly choosing the internal impedance of the generator. The minimum noise figure is attained for $Z_s = Z_{s,\text{opt}}$ where

$$R_{s,\text{opt}} = \sqrt{\frac{r_n}{g_n} + R_c^2} \quad (73)$$

$$X_{s,\text{opt}} = -X_c. \quad (74)$$

The resulting minimum noise figure is:

$$F_{\min} = 1 + 2g_n(R_c + R_{s,\text{opt}}). \quad (75)$$

By combining Eq. (72) and Eq. (75) one can express the noise figure as:

$$F = F_{\min} + \frac{g_n}{R_s} \left[(R_s - R_{s,\text{opt}})^2 + (X_s - X_{s,\text{opt}})^2 \right]. \quad (76)$$

Thus, the noise conductance g_n is a measure of the sensitivity of the noise figure with respect to the minimum for variations from the optimum source condition.

Alternatively, by using the parallel representation and considering a generator of internal admittance $Y_s = G_s + jB_s$, one would write:

$$G_{s,\text{opt}} = \sqrt{\frac{G_n}{R_n} + G_c^2} \quad (77)$$

$$B_{s,\text{opt}} = -B_c. \quad (78)$$

where $Y_c = G_c + jB_c$, and:

$$F_{\min} = 1 + 2R_n(G_c + G_{s,\text{opt}}). \quad (79)$$

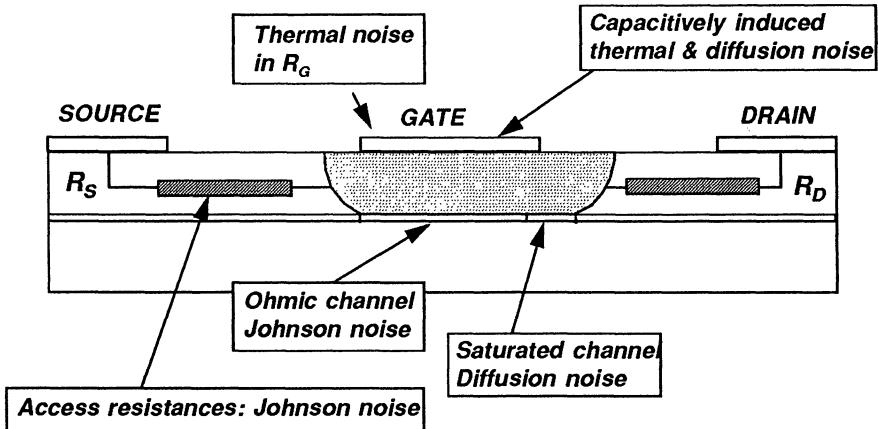


Figure 18. Distribution of noise sources in a PHEMT device.

Finally, one can write:

$$F = F_{\min} + \frac{R_n}{G_s} \left[(G_s - G_{s,\text{opt}})^2 + (B_s - B_{s,\text{opt}})^2 \right]. \quad (80)$$

Again, the noise resistance R_n is a sensitivity measure for the noise figure with respect to its minimum value.

2.3. FREQUENCY AND BIAS BEHAVIOUR OF PHEMT NOISE

The noise sources in a HEMT are usually divided into two classes: intrinsic and extrinsic noise sources. Extrinsic noise is thermal noise originated by the access resistances to the channel and to the gate resistance (Fig. 18). Intrinsic noise is connected to the channel noise, which in turn includes thermal noise originating in the ohmic part of the channel, and diffusion noise coming from the velocity saturated region.

In order to derive a simple and approximate picture of the bias dependence of the noise generated by the device [34], let us neglect the contribution of the saturated part of the channel and evaluate the device noise at the end of the linear region. In this case, and using a simple two-zone model for the velocity-field relationship of the material, we can approximate the channel resistance R_i as:

$$R_i \approx \mathcal{E}_{\text{th}} L_G / I_D \quad (81)$$

since at the onset of velocity saturation the voltage drop on the channel is approximately given by the threshold field \mathcal{E}_{th} multiplied by the channel length L_G . Since $\mathcal{E}_{\text{th}} \approx v_{\text{sat}}/\mu_0$, where v_{sat} is the saturation velocity and μ_0

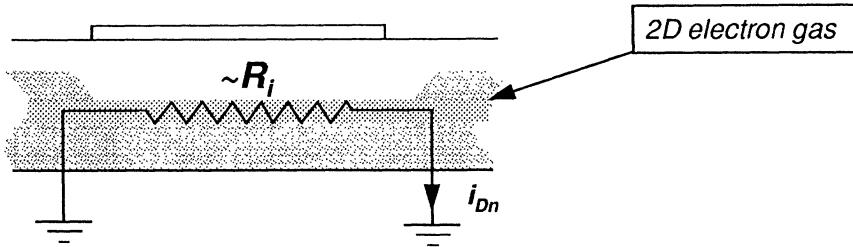


Figure 19. Simplified model for short-circuit drain current fluctuation spectrum.

the initial mobility, one finally obtains for the spectrum of the short-circuit drain current fluctuations the approximation:

$$S_{i_{Dn}i_{Dn}} \approx \frac{4kT}{R_i} \approx \frac{4kT\mu_0}{v_s L_G} I_D. \quad (82)$$

Thus, the spectrum is white and increases with I_D .

Channel current fluctuations induce, through capacitive coupling to the gate, fluctuations in the short-circuit gate current. Again, the simplified model of Fig. 20 can be exploited to evaluate such fluctuations. In fact one has, from the equivalent circuit shown and in the low-frequency limit:

$$S_{i_{Gn}i_{Gn}} \approx 4kTR_i\omega^2C_{GS}^2 \approx 4kT\omega^2C_{GS}^2 \frac{v_s L_G}{\mu_0 I_D}. \quad (83)$$

Thus, short-circuit gate current fluctuations are proportional to f^2 and their behaviour as a function of bias depend on the ratio C_{GS}^2/I_D . Since in HEMTs the gate-source capacitance is approximately constant far from pinchoff and decreases at pinchoff owing to the displacement of the channel charge away from the gate, the resulting behaviour of the short-circuit gate current fluctuation spectrum is a decreasing function of current, as confirmed by more accurate numerical simulations [35]. In the simplified model introduced, the correlation between i_{Dn} and i_{Gn} is $C \approx j$, i.e. fluctuations are fully correlated.

An alternative representation of the noise parameters is obtained by replacing the fluctuation spectra and their correlation (approximately assumed as purely imaginary, jC) with the adimensional parameters R and P [30] defined as:

$$S_{i_{Dn}i_{Dn}} = 4kTg_m P \quad (84)$$

$$S_{i_{Gn}i_{Gn}} = 4kT \frac{\omega^2 C_{GS}^2}{g_m} R \quad (85)$$

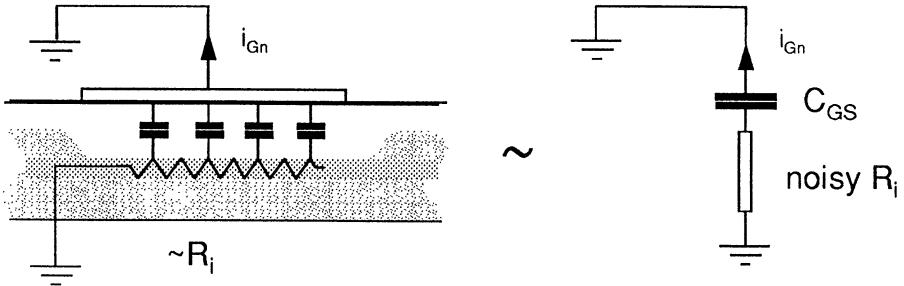


Figure 20. Simplified model for short-circuit gate current fluctuation spectrum.

where T is the device temperature, g_m the transconductance, C_{GS} the gate-source capacitance. Note that the P , R , C representation is not exact, even if these coefficients are assumed as functions of frequency, since the correlation coefficient is purely imaginary rather than complex. With the above approximations one has $P \approx I_D/g_m \mathcal{E}_{th} L_G$, $R = 1/P$. By taking into consideration the gate and source parasitic resistances and expanding the minimum noise figure in powers of frequency, one obtains, to the first order:

$$F_m \approx 1 + 2 \frac{f}{f_T} \sqrt{g_m(R_s + R_g)(P + R - 2C\sqrt{PR}) + (1 - C^2)PR}, \quad (86)$$

where $f_T = g_m/2\pi C_{GS}$ is the cutoff frequency. Eq. (86) is similar to the so-called Fukui expression [36], which ultimately postulates a linear frequency dependence of the minimum noise figure, well verified in practical devices. Taking into account that the transconductance is approximately constant for high drain current, while it decreases towards pinchoff, where $g_m \approx \sqrt{I_D}$, one sees that for high current the minimum noise figure increases approximately like $\sqrt{I_D}$, while for low currents it again increases due to the increase of R and to the decrease of f_T . This ultimately leads to the bell-shaped behaviour in Fig. 21.

By neglecting the gate noise and assuming unit correlation, one finally obtains a version of the Fukui formula [34]:

$$F_m = 1 + 2 \sqrt{\frac{I_{DS}}{E_c L_g g_m}} \frac{f}{f_T} \sqrt{R_s + R_g}. \quad (87)$$

The minimum noise figure increases linearly as a function of frequency and is directly proportional to the drain current, inversely proportional

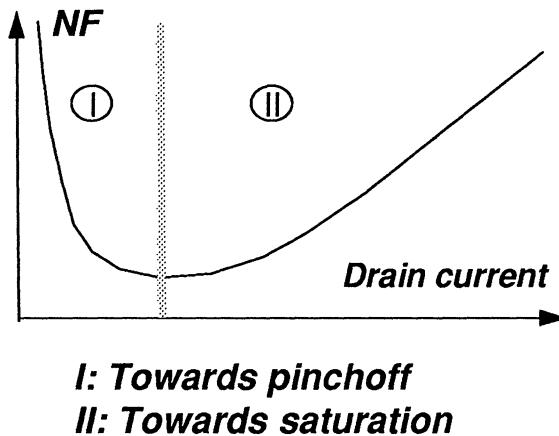


Figure 21. Qualitative behaviour of FET noise figure versus the drain current.

to the cutoff frequency and the transconductance. Thus, the superior performances of PHEMTs with respect to other FETs partly derive from the higher cutoff frequency and specific transconductance achieved by these devices. These qualitative conclusions are confirmed by the experiment and also by more complex expressions accounting for the gate noise.

Although the qualitative behaviour of the noise figure in MESFETs and PHEMTs is similar, it may be noticed that pseudomorphic HEMTs often exhibit a nearly flat minimum noise figure for a wide enough bias range to include the bias points yielding maximum gain and also maximum power, see Fig. 22. Thus, not only PHEMTs have superior noise properties, but also their associated gain in low-noise bias is considerably better than for MESFETs.

2.4. MEASUREMENT-BASED NOISE MODELS

As discussed, the exact characterization of the noise behaviour of a two-port requires four real parameters at each frequency. The $P\ R\ C$ model assumes that the correlation is purely imaginary; moreover, the noise coefficients are often approximately taken as frequency independent. This remark is supported by the experiment, which shows that the current fluctuation spectra are either white or ω^2 , thus depending on a single parameter. Recent theoretical investigations [37] have stressed that, under certain assumptions, the frequency-independent noise parameters can be reduced to two, since it can be shown that the correlation coefficient is approximately $C \approx \sqrt{P/R}$.

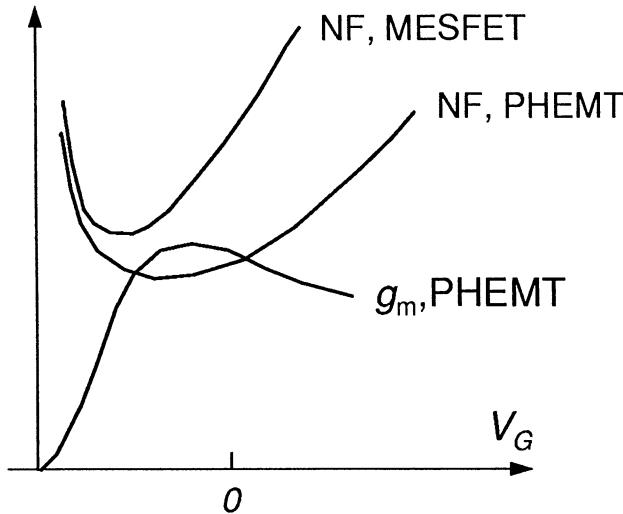


Figure 22. Qualitative behaviour of MESFET and PHEMT noise figure and gain versus the drain current.

Thus, it may be concluded that the frequency-dependent noise characterization of a microwave FET reduces, at least approximately, to the identification of two frequency independent noise coefficients. This conclusion had already reached by some experimentalists; in particular, Pospieszalski [38] proposed a noise model whose noise parameters are two noise temperatures assigned to the intrinsic (channel) resistance and to the output conductance, see Fig. 23. By introducing the noise temperatures T_G and T_D , one obtains:

$$S_{i_{Dn}i_{Dn}} = 4kT_D g_{DS} + 4kT_G \frac{g_m^2 R_i}{1 + \omega^2 C_{GS}^2 R_i^2} \quad (88)$$

$$S_{i_{Gn}i_{Gn}} = 4kT_G \frac{R_i \omega^2 C_{GS}^2}{1 + \omega^2 C_{GS}^2 R_i^2} \quad (89)$$

$$S_{i_{Gn}i_{Dn}} = 4kT_G \frac{j\omega g_m C_{GS} R_i}{1 + \omega^2 C_{GS}^2 R_i^2}. \quad (90)$$

Measurements point out that T_G is close to the ambient temperature while T_D takes values in excess of several thousands K. Despite the controversial correlation with the electron temperature in the hot-carrier regime, the two-temperature model provides a simple way to identify by means of a frequency-independent parameter set the overall device noise behaviour for a single operating point.

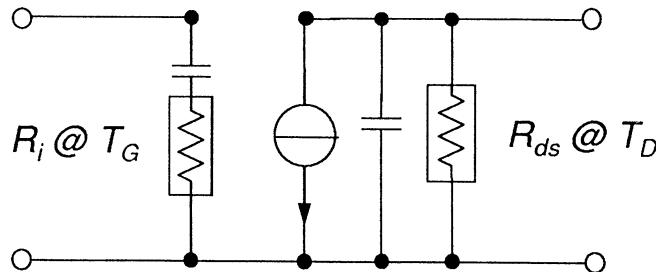


Figure 23. Two-temperature Pospieszalski model.

2.5. AN OVERVIEW ON PHYSICS-BASED PHEMT NOISE ANALYSIS

Physics-based semiconductor noise modelling techniques based on drift-diffusion or hydrodynamic transport equations [39] are classically derived by considering that, owing to the small amplitude of noise fluctuations, the device response to microscopic noise sources can be estimated by means of a linearized model [40], once that the statistical characterization of the microscopic noise sources is known from semiconductor physics. This is the basis of the *impedance field method* [40] for noise analysis.

As already recalled, the motion of carriers in a semiconductor is a random sequence of free flights and scattering events. Consider a sample of N electrons having random velocities $\underline{v}_i(t)$; the mean velocity of the sample $\underline{v}(t) = (1/N) \sum_i \underline{v}_i(t)$ is itself a random process [29] whose average value $\langle \underline{v} \rangle$ is the carrier velocity (in the usual sense) and whose variance is inversely proportional to the sample population N . Since N is large but finite, the variance of the mean velocity will be small but not zero. We define as fluctuation (here, velocity fluctuation $\delta\underline{v}(t)$) the small-amplitude random process $\delta\underline{v}(t) = \underline{v}(t) - \langle \underline{v} \rangle$ having zero average but non-zero mean square value. Carrier velocity fluctuations cause the current density to fluctuate, yielding diffusion noise, or, in low-field, thermal noise [26]. Current density fluctuations can also be caused by population fluctuations arising from G-R phenomena (G-R noise and $1/f$ noise) or intervalley scattering (intervalley scattering noise), as in III-V semiconductors [26, 27, 28].

Small-amplitude microscopic current density fluctuations induce voltage or current fluctuations at the device terminals. These can be (somewhat artificially) interpreted as the small-signal response of the noiseless device to an impressed, distributed current source modelling microscopic fluctuations [40]. Thus, a direct relationship exists between small-signal and noise modelling: the small-signal response is the device response to a small-amplitude *external* excitation; noise is the device response to a small-amplitude, dis-

tributed, *internal* random current excitation given by the current density fluctuation $\delta\underline{J}(\underline{r})$.

The aim of physics-based noise analysis is therefore to evaluate the statistical properties of the macroscopic fluctuations (*e.g.*, the power spectra of the open circuit noise voltages or short-circuit noise currents at the device ports) given the statistical properties of the microscopic ones. The statistical characterization of the impressed current density is derived from semiconductor physics [26] according to the noise mechanism considered, and is usually expressed in the form of the correlation spectrum:

$$\mathbf{S}_{\delta\underline{J}(\underline{r}_1)\delta\underline{J}(\underline{r}_2)}(\underline{r}_1, \underline{r}_2, \omega) = \overline{\delta\underline{J}(\underline{r}_1, \omega)\delta\underline{J}^*(\underline{r}_2, \omega)}, \quad (91)$$

where the bar sign denotes ensemble average. Since the correlation length of microscopic fluctuations is of the order of the mean free path, these are often considered as spatially uncorrelated, *i.e.*:

$$\mathbf{S}_{\delta\underline{J}(\underline{r}_1)\delta\underline{J}(\underline{r}_2)}(\underline{r}_1, \underline{r}_2, \omega) = \mathbf{K}_{\delta\underline{J}\delta\underline{J}}(\underline{r}_1, \omega)\delta(\underline{r}_1 - \underline{r}_2), \quad (92)$$

where \mathbf{K} is usually referred to as *local noise source* [27]. For uncorrelated diffusion noise [40, 26] the local noise source is:

$$\mathbf{K}_{\delta\underline{J}\delta\underline{J}}(\underline{r}, \omega) = 4q^2\mathbf{D}_0n_0(\underline{r}) \quad (93)$$

where \mathbf{D}_0 and n_0 are the working point diffusivity matrix and electron density, respectively.

The small-signal potential fluctuations induced by $\delta\underline{J}(\underline{r})$ can be generally expressed by means of the Green's function of the problem and of a superposition integral extended to the device volume Ω . This is the principle of Shockley's impedance-field method (IFM) [40], in which the Green's function (the *vector impedance field*) is a vector $\underline{Z}(\underline{r}_1, \underline{r}_2, \omega)$ such that

$$\delta\phi(\underline{r}_1, \omega) = \int_{\Omega} \underline{Z}(\underline{r}_1, \underline{r}_2, \omega) \cdot \delta\underline{J}(\underline{r}_2, \omega) d\underline{r}_2, \quad (94)$$

where $\delta\phi(\underline{r}_1, \omega)$ is the induced potential fluctuation. In turn, $\underline{Z}(\underline{r}_1, \underline{r}_2, \omega)$ can be obtained as $\nabla_{\underline{r}_2} Z(\underline{r}_1, \underline{r}_2, \omega)$ where Z is the *scalar impedance field*, *i.e.*, the response to a spatially impulsive scalar current source impressed in point \underline{r}_2 [40].

From the definition of the vector impedance field, and considering spatially uncorrelated sources, the power ($i = j$) or cross ($i \neq j$) spectra of the potential fluctuations δe_i induced on electrode i, j , take the form:

$$\mathcal{S}_{\delta e_i \delta e_j}(\omega) = \int_{\Omega} \nabla_{\underline{r}} Z(\underline{r}_i, \underline{r}, \omega) \cdot \mathbf{K}_{\delta\underline{J}\delta\underline{J}}(\underline{r}, \omega) \cdot \nabla_{\underline{r}} Z^*(\underline{r}_j, \underline{r}, \omega) d\underline{r}. \quad (95)$$

The physics-based noise modelling of microwave FETs has been carried out in the past through simplified numerical or analytical implementations of the impedance-field method, in which the drain voltage fluctuations induced by the channel current density fluctuations are evaluated through a 1D version of the IFM applied to the one-dimensional channel transport model, while the gate voltage capacitively induced by the channel current fluctuations is derived according to a quasi-static charge control model. A representative example of numerical, quasi-2D model is the MESFET model proposed in 1981 by Carnez, Cappy *et al.* [41], based on the quasi-2D non-stationary model [42] and later extended to HEMTs [33]. Analytical noise models based on the two-region channel approximation were proposed for epitaxial MESFETs by Baechtold [43] and later by Statz, Haus and Pucel [30]. Statz's model was recently applied to the HEMT by Brookes [44] and Ando [45]; although the results are in satisfactory agreement with the experiment, some basic difficulties arise in characterizing the fluctuations of the 2DEG current [46]. Concerning HEMTs and PHEMTs, the use of an IFM implementation based on a full 2D model (see [47]) for device characterization is perhaps less favourable than in MESFETs, since 2D models usually approximate in a classical way the charge control relationship yielding the 2DEG population as a function of the external bias. A good compromise between accuracy and computational efficiency was obtained through the so-called active line method [48] which is also amenable to numerical implementation. A representative example of quasi-2D PHEMT simulator with noise modelling capabilities is the HELENA simulator developed at the Lille University [49].

3. EM models for (M)MICs

3.1. INTRODUCTION

Hybrid and Monolithic Microwave Integrated Circuits (MICs and MMICs, see [50, 51]) include passive elements, whose design through electromagnetic (EM) analysis techniques is fundamental. Passive elements are either distributed (like transmission lines [52, 53] and related elements, such as hybrid couplers) or concentrated (resistors, capacitors, inductors [50, 51, 54]). Moreover, the circuit is always enclosed in a package [51, Ch.10] and linked to the outside world through electrical connectors; neither the package nor the connectors are ideal, *i.e.*, electrically transparent (Fig. 24).

The traditional approach towards the EM design of microwave circuits based on planar transmission lines and components (such as microstrip and coplanar lines) is based on the partition of the circuit into non-interacting elements. These can be lumped elements, such as resistors, capacitors or inductors, see Fig. 25, or distributed elements, such as simple or coupled

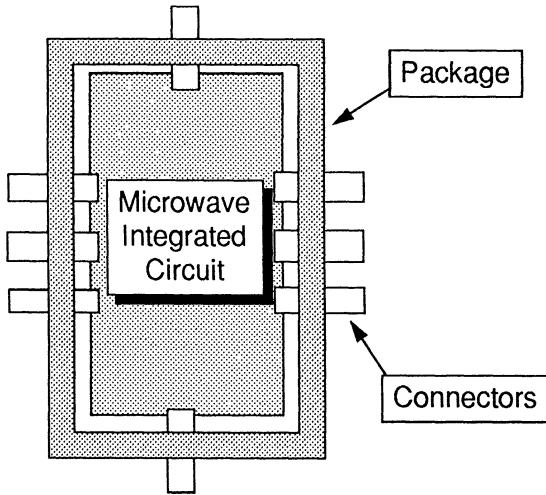


Figure 24. Microwave integrated circuit, package and connectors.

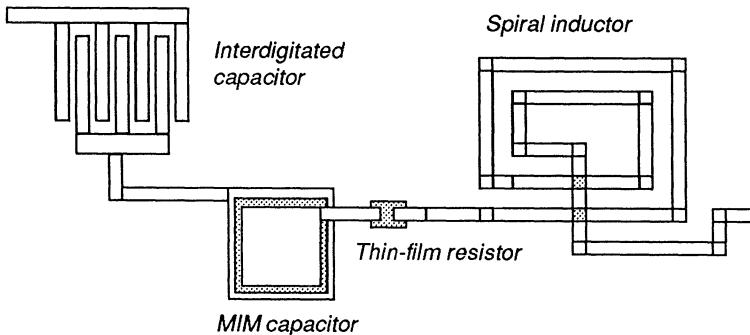


Figure 25. MMIC lumped elements.

transmission lines, see Fig. 26. In the traditional circuit approach, elements are characterized in isolation (either experimentally, or through EM analysis) and their equivalent circuit is established in terms of ideal lumped or distributed elements. Thus, for instance, an interdigitated capacitor is modelled by a lumped ideal capacitor connected to some parasitic elements (inductors and resistors), and a microstrip line is modelled as an ideal transmission line having characteristic impedance Z_0 , phase velocity v_f and attenuation α .

The above design strategy neglects spurious interactions between ele-

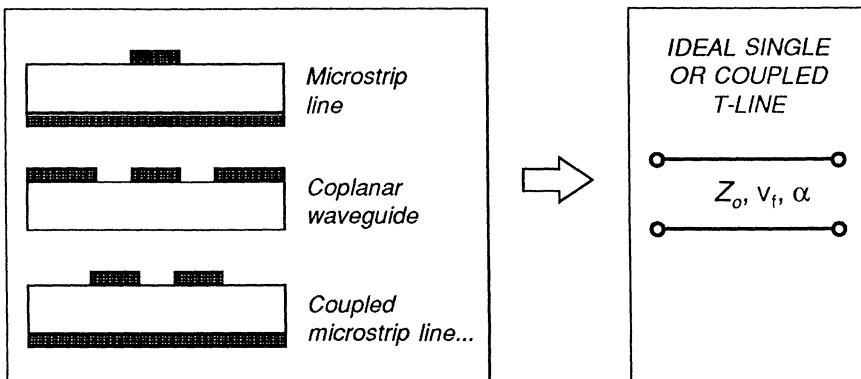


Figure 26. MMIC distributed components.

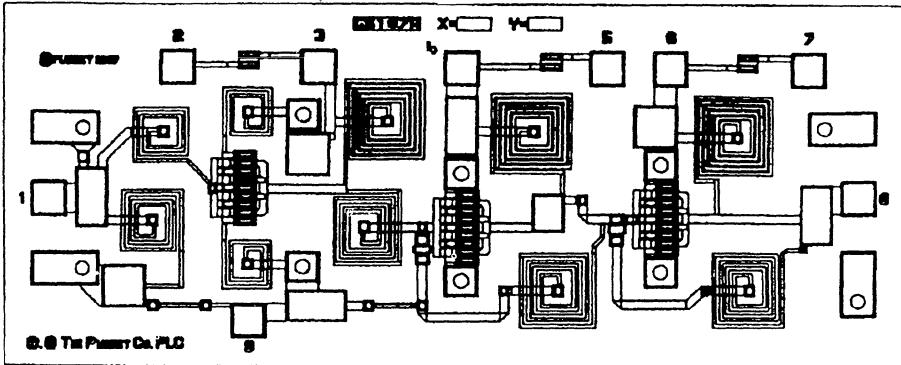


Figure 27. MMIC example (3-stage low-noise amplifier), from [51].

ments. Although in most practical circuits on comparatively low-frequency bands (*e.g.* below 10 GHz) spurious couplings actually turn out to be negligible, MMIC downsizing (related to the increasing operating frequency and to the minimization of the chip area) requires to keep under control the EM interaction between elements. This means that layout rules must be established to minimize spurious couplings. Since those may not be immediately evident, the global EM analysis of the circuit would ultimately allow the overall optimization of the circuit layout to be performed without any simplifying assumption on the nature of element interaction.

Unfortunately, microwave circuit design based on overall EM analysis still is far beyond the possibility of today's computational resources. Indeed, several tools for the three-dimensional EM analysis of microwave circuits have been made available during the last few years. However, as the com-

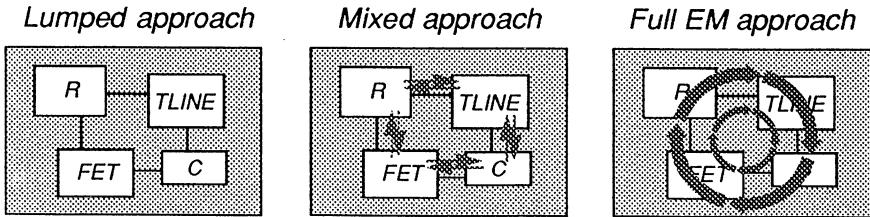


Figure 28. EM modelling approaches for MMICs.

plexity of an even simple example of MMIC would suggest (Fig. 27), owing to the memory and speed limits of today's workstations such EM tools are practically unable to allow for the EM analysis (let alone design and optimization) of complete circuits.

Because of these limitations, a more realistic view of the EM design of MMICs can be described as shown in Fig. 28. A first design approach could be defined as *lumped*, in the sense that element interaction is considered as second-order only. In the lumped approach elements are characterized in isolation and empirical layout rules are established on the basis of the experiment or of EM models, with the aim of making the interaction between elements to be actually negligible. This design approach is somewhat similar to the one followed in large-scale digital integrated circuit design and is easily amenable to layout automation, since the placement of elements is ruled by simple, one-to-one rules on element separation. In the *mixed approach* some parts of the circuit are singled out as critical (for instance, after that a tentative initial layout has been made) and their full EM analysis is carried out to control local spurious interactions. The mixed approach is within the possibilities of the currently available EM simulators; moreover, these can be profitably applied to the characterization of single elements such as line discontinuities, spiral inductors, interdigitated capacitors and so forth, whose frequency limits are practically dictated by the full-wave behaviour. Finally, in the *full EM approach* a global EM analysis is performed on the whole circuit, so as to automatically include all possible interactions. With the commonly available computational resources, only extremely simple circuits, or components with prevailing EM non-static behaviour (like radiating elements), can be treated this way.

Last but not least, the full EM analysis of packaging structures has become a matter of interest, since the packaging usually deteriorates the electrical performances of the *in chip* circuit. Three-dimensional electromag-

netic design tools can greatly help in optimizing package structures both from the standpoint of in-out transitions and from the one of parasitic resonances.

3.2. ADVANCED EM ANALYSIS TECHNIQUES FOR MMICS

The analysis and design of passive elements in hybrid and monolithic microwave integrated circuits requires the solution of the electromagnetic model provided by the Maxwell equations. Full-wave techniques directly address the solution of Maxwell equations in time or frequency domain, while quasi-static approaches exactly or approximately reduce the problem to a static formulation, *i.e.* to the time-independent Laplace or Poisson equations. For a review on the problem formulation and available solution techniques, see *e.g.* [55, 56, 57] and references therein.

The analysis of lumped elements can be, at least in the lowest frequency range, carried out through the solution of quasi-static models [54]. Capacitors typically pose a potential problem, whose formulation in terms of the Laplace equation is straightforward; the solution provides the structure capacitance, and therefore a viable model in the frequency range where parasitic effects can be neglected. In a similar way, the analysis of resistors can be reduced to the solution of Laplace equation in a conducting medium and the analysis of inductors amounts to the solution of Poisson equation with source term proportional to the current density flowing in the inductor.

A quasi-static treatment often suffices also to the characterization of transmission lines encountered in MMICs, such as the microstrip and the coplanar line. Although such lines are, strictly speaking, dispersive (*i.e.* characterized by frequency-dependent parameters), in the low-frequency range (which, for MMIC substrates, can extend up to the millimeter wave range) planar lines can be analyzed through the well known per-unit-length capacitance and inductance concepts, whose evaluation amounts again to the solution of a static problem on the line cross section. Line discontinuity problems can also be formulated, at least in the low-frequency range, in terms of parasitic capacitances or inductances, thus requiring a quasi-static EM model.

Despite its simplicity with respect to the full EM problem, the numerical treatment of the Laplace or Poisson equations in a 2D or 3D system is not straightforward. Apart from conformal mapping techniques, which have been extensively used for the solution of 2D problems and therefore for the line characterization, two basic approaches are available for the solution of Laplace equation.

According to the first approach [58], the differential problem is reduced to an integral equation, the unknown being the charge or current distribu-

tion of a metallic strip lying in a multilayered dielectric medium. The kernel of the integral equation is the Green's function of the problem, which can be efficiently computed in many practically significant cases (like in homogeneous and planar stratified dielectric media). Since the Green's function has a more straightforward representation as a Fourier integral or series, integral equation based techniques are often termed as (static) spectral domain Green's function techniques [59]. One of the advantages of the integral formulation is the fact that it reduces by one the dimension of the problem. For instance, a two-dimensional problem, like the evaluation of a line capacitance, is reduced to an integral equation whose unknown is the charge density on the line cross section, which is one-dimensional. Three dimensional problems involving the charge or current density supported by planar conductors can be reduced to integral equations whose domain is two-dimensional, *i.e.* to a set of currents or charges defined in the dielectric interface whereon metallic strips lie.

The second approach is a brute-force discretization of the differential problem in the whole domain. This can be carried out through techniques like the Finite Difference method or the Finite Element method (FEM) [60]. Brute-force approaches have high computational cost but no geometrical limitations on the structure analyzed. Besides, several computer codes are now commercially available to conveniently set up the problem according to the FEM or finite-differences on medium-power workstations. Integral equation based techniques require a larger amount of analytical preprocessing and are less flexible, but often more efficient on specific problems.

The full solution of Maxwell equations was first undertaken in the analysis of planar transmission lines. As already mentioned, these structures are only approximately TEM, and can be exactly modelled only by solving, in the frequency domain, the Maxwell equations or an equivalent set on the line cross-section. Owing to the planar nature of most (M)MIC transmission lines, and to the previous experience already gained in the solution of electrodynamic problems on multilayered dielectric substrates by means of Green's function techniques, the spectral-domain Green's function method obtained, during the late seventies and early eighties, great popularity in the solution of transmission line problems (see [61, 62] and references therein). Many general-purpose computer codes were developed on the basis of this method for the analysis of several classes of planar lines (microstrip, coplanar line, slot line).

The full-wave analysis was extended to line discontinuities by means of the so-called equivalent waveguide approach [63]. According to this technique, microstrip lines can be replaced by an equivalent closed plane-parallel line having suitable equivalent dimensions; microstrip transitions can be approximately modelled as transitions between equivalent closed waveguides.

This technique brought to immediate use a wealth of analysis methods inherited from the analysis of discontinuities in closed waveguides, such as the mode matching method [55, 56]. Accurate characterizations in closed or quasi-closed form were obtained for large classes of microstrip discontinuities.

Limitations of the spectral-domain Green's function exist in the treatment of lines with thick metallizations (like the ones arising in MMICs) or non-planar dielectrics; moreover, the extension of the method to three-dimensional components is possible [61], but not entirely satisfactory in terms of generality and structure flexibility. The closed-waveguide approach to discontinuity modelling, on the other hand, does not allow for many microstrip and all coplanar discontinuities. In view of such limitations, the growing computational efficiency and memory resources of today's workstations make possible to exploit brute-force approaches also for the solution of the full-wave problem. New techniques, like the Transmission Line Method (TLM) (see [64, 65] and references therein) and the Finite Differences Time Domain (FDTD) method (see [66, 67]) have been shown to effectively handle large 3D problems with arbitrary geometries. At the same time, Finite Element techniques applied to the dynamic problem [68] have also been demonstrated as viable tools for the analysis of large 3D structures.

Computer-aided design (CAD) tools are commercially available nowadays, which support the spectral-domain method in 2D or 3D, the finite-element method, or the TLM method. From a practical standpoint, time domain techniques such as the TLM and the FDTD have not reached yet a widespread acceptance from the technical community, since data analysis is not straightforward. The use of 3D simulators for circuit analysis or design still is sporadic, and often confined to extremely advanced design. However, although most circuit design is carried out through conventional approaches, the use of 3D electromagnetic techniques is already widespread in setting up model libraries. The electromagnetic design of realistic planar components can be now performed (with the help of adequate computational resources), and will probably become a common experience for the next generation of circuit designers, equipped with more powerful workstations and more efficient and user-friendly electromagnetic CAD tools. Analysis techniques will continue their development, but a more widespread availability of computer resources will probably favour general-purpose, brute-force approaches with respect to more restricted, though efficient, techniques such as the Green's function approach. This is, at least, the trend followed during the last few years. Nevertheless, more refined approaches are not likely to become obsolete, although their application is likely to be confined to specific areas.

Acknowledgments

Part of the work described in Sec. 1 and 2 was supported by the ESPRIT projects "MANPOWER" and "COSMIC".

References

1. Maas, S.A. (1988) *Nonlinear microwave circuits*, Artech House.
2. Boyd, S., Tang, Y.S., and Chua, L.O. (1983) Measuring Volterra kernels, *IEEE Trans. Circuits and Systems* **30**, 571–577.
3. Root, D.E., Fan, S., and Meyer, J. (1991) Technology-independent large-signal non quasi-static models by direct construction from automatically characterized device data, *Proc. 21st European Microwave Conf.*, Stuttgart, 927–932.
4. Root, D.E. (1992) Foundations of measurement-based active device modeling for circuit simulation, *Proc. ICEAA*, Torino, 141–146.
5. Filicori, F., Vannini, G., and Monaco, V.A. (1992) A nonlinear integral model of electron devices for HB circuit analysis, *IEEE Trans. Microwave Theory Tech.*, 1456–1465.
6. MDS - HP 85150B Microwave and RF Design Systems. Distributed by: Hewlett Packard.
7. Willing, H.A., Rauscher, C., and De Santis, P. (1978) A technique for predicting large-signal performance of a GaAs MESFET, *IEEE Trans. Microwave Theory Tech.* **26**, 1017–1023.
8. Rauscher, C., and Willing, H.A. (1979) Simulation of nonlinear microwave FET performance using a quasi-static model, *IEEE Trans. Microwave Theory Tech.* **27**, 834–840.
9. Curtice, W.R. (1980) A MESFET model for use in the design of GaAs integrated circuits, *IEEE Trans. Microwave Theory Tech.* **28**, 448–456.
10. Curtice, W.R., and Ettenberg, M. (1985) A nonlinear GaAs FET model for use in the design of output circuits for power amplifiers, *IEEE Trans. Microwave Theory Tech.* **33**, 1383–1394.
11. Statz, H., Newman, , Smith, I.W., Pucel, R.A., and Haus, H.A. (1987) GaAs FET device and circuit simulation in SPICE, *IEEE Trans. Electron Devices* **34**, 160–169.
12. Jastrzebski, A.K. (1987) Non-linear MESFET modelling, *Proc. of 17th European Microwave Conf.*, Roma, 599–604.
13. Graffeuil, J., Hadjoub, Z., Fortea, J.P., and Pouysegur, M. (1986) Analysis of capacitance and transconductance frequency dispersions in MESFET's for surface characterization, *Solid State Electronics* **29**, 1087–1097.
14. Goyal, R., Golio, M., and Thomann, W. (1989) Low-frequency anomalies in GaAs MESFETs, in R. Goyal (ed), *Monolithic microwave integrated circuits*, Artech House, Dedham, Sec. 4.1.6.1.
15. Selmi, L., and Riccò, B. (1991) Thermal characterization of GaAs MESFETs by means of pulsed measurements, *Proc. IEDM 1991*, Washington, 255–258.
16. Vidalou, J.F., Grossier, J.F., Chaumas, M., Camiade, M., Roux, and Obregon, J. (1991) Accurate nonlinear transistor modeling using pulsed S-parameter measurements under pulsed bias conditions, *Proc. IEEE MTT-S Symposium*.
17. Root, D.E., and Fan, S. (1992) Experimental evaluation of large-signal modeling assumptions based on vector analysis of bias-dependent S-parameter data from MESFETs and HEMTs, *Proc. IEEE MTT-S symposium*, Atlanta, 255–258.
18. Daniels, R.R., Harrang, J.P., and Yang, A. (1991) A nonquasi-static, large signal FET model derived from small signal S-parameters, *Proc. of ISDRS 91*, Charlottesville, 601–604.
19. Foisy, M.C., Jeroma, E., and Martin, G.H. (1992) Large-signal relaxation-time model for HEMTs and MESFETs, *Proc. IEEE MTT-S symposium*, Atlanta, 251–

254.

20. Filicori, F., Vannini, G., Mediavilla, A., and Tazon, A. (1993) Modelling of deviations between static and dynamic drain characteristics in GaAs FETs, *Proc. 23rd European Microwave Conf.*, Madrid, 454–457.
21. OSA90/hopeTM, (1991), Optimization System Associates Inc., Dundas, Ontario, Canada.
22. Golio, J.M. (1991) *Microwave MESFETs and HEMTs*, Artech House.
23. Halchin, D., Miller, M., Golio, M., and Tehrani, S. (1994) HEMT models for large signal circuit simulation, *Proc. IEEE MTT-S Symposium*, San Diego, 985–988.
24. Angelov, I., Zirath, H., and Rorsman, N. (1992) A new empirical nonlinear model for HEMT and MESFET devices, *IEEE Trans. Microwave Theory Tech.* **40**, 2258–2266.
25. Ghione, G., Pirola, M., Dortu, J.M., and Müller, J. (1994) An improved P-HEMT large-signal model for medium-power Ka-band amplifiers, *Proc. GAAS94*, Torino, 423–426.
26. van der Ziel, A. (1970) *Noise: sources, characterization, measurement*, Prentice-Hall, Englewood Cliffs.
27. Nougier, J.P. (1980) Noise and diffusion of hot carriers, in D.K. Ferry, J.R. Barker and C. Jacoboni (eds), *Physics of nonlinear transport in semiconductors*, Plenum Press, pp. 415–465.
28. Nougier, J.P. (1987) Origine du bruit dans les dispositifs à semiconducteurs, *Revue Phys. Appl.* **22**, 803–819.
29. Papoulis, A. (1965) *Probability, random variables, and stochastic processes*, McGraw-Hill, Kogakusha.
30. Pucel, R.A., Haus, H.A., and Statz, H. (1975) Signal and noise properties of gallium arsenide microwave field-effect transistors, *Advances in Electronics and Electron Physics* **38**, Academic Press, 195–265.
31. Rothe, H., and Dahlke, W. (1955) Theory of noisy fourpoles, *Proc. IRE* **44**, 811–818.
32. Bonani, F., and Ghione, G. (1993) Noise modelling of HEMT's, *Alta Frequenza – Focus on GaAs electronics: toward zero dimensional structures* **5**, 28–36.
33. Cappy, A. (1988) Noise modeling and measurement techniques, *IEEE Trans. Microwave Theory Tech.* **36**, 1–10.
34. Delagebeaudeuf, D., Chevrier, I., Laviron, M., and Delescluse, P. (1985) A new relationship between the Fukui coefficient and optimal current value for low noise operation of field effect transistors, *IEEE Electron Dev. Lett.* **6**, 444–445.
35. Cappy, A., Vanoverschelde, A., Schortgen, M., Versnaeyen, C., and Salmer, G. (1985) Noise modeling in submicrometer-gate two-dimensional electron-gas field-effect transistors, *IEEE Trans. Electron Devices* **32**, 2787–2795.
36. Fukui, H. (1979) Optimal noise figure of microwave GaAs MESFET's, *IEEE Trans. Electron Devices* **26**, 1032–1037.
37. Danneville, F., Happy, H., Dambrine, G., Belquin, J.-M., and Cappy, A. (1994) Microscopic noise modeling and macroscopic noise models: how good a connection?, *IEEE Trans. Electron Devices* **41**, 779–786.
38. Pospieszalski, M.W. (1989) Modeling of noise parameters of MESFET's and MOD-FET's and their frequency and temperature dependence, *IEEE Trans. Microwave Theory Tech.* **37**, 1340–1350.
39. Selberherr, S. (1984) *Analysis and simulation of semiconductor devices*, Springer Verlag, Wien.
40. Shockley, W., Copeland, J.A., and James, R.P. (1966) The impedance field method of noise calculation in active semiconductor devices, in O. Lowdin (ed), *Quantum theory of atoms, molecules and solid state*, Academic Press, pp. 537–563.
41. Carnez, B., Cappy, A., Fauquenbergue, R., Constant, E., and Salmer, G. (1981) Noise modeling in submicrometer-gate FET's, *IEEE Trans. Electron Devices* **28**, 784–789.
42. Carnez, B., Cappy, A., Kaszynski, A., Constant, E., and Salmer, G. (1980) Modelling

- of a submicrometer gate field-effect transistor including effects of nonstationary electron dynamics, *J. Appl. Phys.* **51**, 784–790.
43. Baechtold, W. (1972) Noise behaviour of GaAs field-effect transistors with short gate lengths, *IEEE Trans. Electron Devices* **19**, 674–680.
 44. Brookes, T.M. (1986) The noise properties of high-electron mobility transistors, *IEEE Trans. Electron Devices* **33**, 52–57.
 45. Ando, Y., and Itoh, T. (1990) DC, small-signal and noise modelling for two-dimensional electron gas field-effect transistor based on accurate charge-control characteristics, *IEEE Trans. Electron Devices* **37**, 67–78.
 46. Bonani, F., Ghione, G., and Naldi, C.U. (1992) A CAD oriented quasi-physical HEMT noise model for device design and optimization, *Proc. of GAAS92*, Nordwijk.
 47. Ghione, G., and Filicori, F. (1993) A computationally efficient unified approach to the numerical analysis of the sensitivity and noise of semiconductor devices, *IEEE Trans. CAD* **12**, 425–438.
 48. Cappy, A., and Heinrich, W. (1989) High-frequency FET noise performances: a new approach, *IEEE Trans. Electron Devices* **36**, 403–409.
 49. Happy, H. (1992) HELENA: un logiciel convivial de simulation des composants à effet de champ, PhD Thesis, Lille University.
 50. Hoffmann, R. (1983) *Integrierte mikrowellenschaltungen*, Springer Verlag, Berlin; English translation: (1987) *Handbook of microwave integrated circuits*, Artech House, Dedham.
 51. Goyal, R. (1989) *Monolithic microwave integrated circuits: technology and design*, Artech House, Dedham.
 52. Ghione, G. (1989) Transmission lines, in R. Goyal (ed), *Monolithic microwave integrated circuits: technology and design*, Artech House, Dedham, pp. 347–382.
 53. Gupta, K.C., Garg, R., and Bahl, I.J. (1979) *Microstrip lines and slotlines*, Artech House, Dedham.
 54. Pettenpaul, E., Kapusta, H., Weisgerber, A., Mampe, H., Luginsland, J., and Wolff, I. (1988) CAD models of lumped elements on GaAs up to 18 GHz, *IEEE Trans. Microwave Theory Tech.* **36**, 294–304.
 55. Itoh, T. (ed) (1989) *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, John Wiley.
 56. Sorrentino, R. (ed) (1989) *Numerical Methods for Passive Microwave and Millimeter Wave Structures*, IEEE Press.
 57. Itoh, T. (1986) An overview on numerical techniques for modeling miniaturized passive components, *Annales des Télécommunications* **41**, 449–462.
 58. Harrington, R.F. (1967) Matrix Methods for Field Problems, *Proc. of IEEE* **55**, 136–149.
 59. Crampagne, R., Ahmadpanah, M., and Guiraud, J.L. (1978) A simple method for determining the Green's function for a large class of MIC lines having multilayered dielectric structures, *IEEE Trans. Microwave Theory Tech.* **26**, 82–87.
 60. Zienkiewicz, O.C. (1971) *The finite element method in engineering science*, McGraw-Hill, London.
 61. Jansen, R.H. (1985) The spectral-domain approach for microwave integrated circuits, *IEEE Trans. Microwave Theory Tech.* **33**, 1043–1056.
 62. Jansen, R.H., Arnold, R.G., and Eddison, I.G. (1988) A comprehensive CAD approach to the design of MMIC's up to MM-wave frequencies, *IEEE Trans. Microwave Theory Tech.* **36**, 208–219.
 63. Wolff, J., Kompa, G., and Mehran, R. (1972) Calculation method for microstrip discontinuities and T-junctions, *Electron. Lett.* **8**, 177–179.
 64. Hoefer, W.J.R. (1989) The transmission-line matrix (TLM) method, in T. Itoh (ed), *Numerical Techniques for Microwave and Millimeter-Wave Passive Structures*, John Wiley, pp. 496–591.
 65. Hoefer, W.J.R. (1985) The transmission-line matrix method - theory and applications, *IEEE Trans. Microwave Theory Tech.* **33**, 882–893.

66. Tafove, A., and Umashankar, K.R. (1989) Review of FD-TD numerical modelling of electromagnetic wave scattering and radar cross section, *Proc. of IEEE* **77**, 682–689.
67. Gwarek, W.K. (1988) Analysis of arbitrarily shaped two-dimensional microwave circuits by finite-difference time-domain method, *IEEE Trans. Microwave Theory Tech.* **36**, 738–744.
68. Arlett, L., Bahrani, A.K., and Zienkiewicz, O.C. (1968) Application of finite elements to the solution of Helmholtz's equation, *Proc. of IEE* **115**, 1762–1766.

TEST AND RELIABILITY

Barry R. Allen
TRW Space & Electronics Group
One Space Park
Redondo Beach, California USA

INTRODUCTION

This section discusses the test and reliability of pHEMT devices and circuits. A wide range of electrical tests are required to assess pHEMT device performance, provide circuit design models, and verify circuit performance. The tests discussed here are for completed devices or circuits and do not include the in-process physical and electrical tests needed to monitor pHEMT device fabrication. Reliability of pHEMT will also be addressed, emphasizing the testing needed to establish device and circuit reliability.

TEST

Although most of this discussion will be concerned with RF testing, both RF and DC testing are required to fully characterize pHEMT devices and circuits. Testing usually has one of three purposes[1]:

- Device performance evaluation -- assessment of critical performance parameters such as gain, noise figure, and power output
- Device model development -- accurate, de-embedded device measurement to provide input to empirical or physical device models
- Circuit performance measurement -- verification that discrete or monolithic circuit meets performance requirements.

Since short gate pHEMT devices achieve >6 dB gain per stage at 100 GHz, testing of pHEMT devices and circuits may require techniques capable of accurate measurement through at least 100 GHz. The devices also have good performance as both low noise amplifiers and power amplifiers[2] resulting in high dynamic range testing requirements. When coupled with monolithic integration, the wide applicability of the pHEMT further complicates testing, since complete sub-systems are now implemented on a single chip. Special test techniques for cryogenic and quasi-optical systems are also needed to cover the range of pHEMT applications.

Performance evaluation

Testing for performance evaluation is usually not as difficult as testing for accurate model development since most of the critical device performance parameters can be established without fully de-embedding the measurement. Usually the loss of the input and

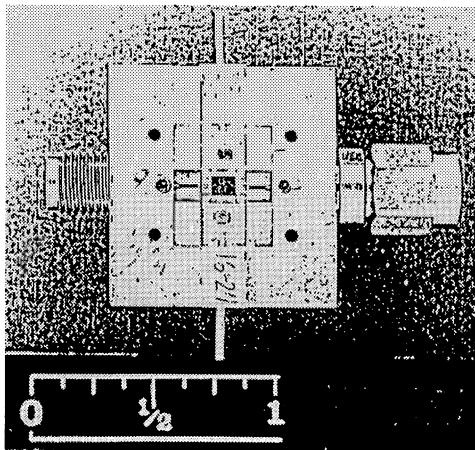


Figure 1. Split block coaxial test fixture for accurate device and MMIC characterization below 40 GHz[1].

put matching circuits of the test fixture is the only data needed to derive the device performance from scalar measurements. As long as the fixture losses are low, the accuracy is reasonably good. However as the fixture loss increases, the accuracy of the measurement will be reduced. For noise figure and output power measurement, the tuning required for optimum device performance will normally be far from the normal $50\ \Omega$ test set impedance, resulting in additional dissipation loss from circulating current in the matching networks. Typically, device performance measurements are made in a coaxial test environment below 20 GHz and in a waveguide test environment above 40 GHz. Between 20 GHz and 40 GHz both coaxial and waveguide tests are used.

A typical microstrip fixture for device performance evaluation is shown in Figure 1. This fixture has about 0.5 dB loss in each coax-to-microstrip transition. This design uses two coax-to-microstrip transitions and a separate center section to hold the device under test, to enable evaluation of both discrete devices and MMIC circuits. The three-piece design also simplifies the characterization of the transition sections for dissipation and reflection loss. Similar designs have been used for fused silica and alumina microstrip substrates with thickness of 0.25 to 0.64 mm.

For higher frequency testing waveguide interfaces are more repeatable and provide lower loss than coax. A waveguide test fixture for WR-22 waveguide is shown in Figure 2. This fixture is also split to allow different center sections to be inserted. The waveguide-to-microstrip transition is an E-plane probe on alumina, the alumina probe is 1.5 mm wide by 0.25 mm thick with a small circle printed on the end inserted into the waveguide to improve the match. The transition provides about 4 GHz bandwidth with 1.4 dB loss for

two transitions connected back to back. The design of the center section for both the coax

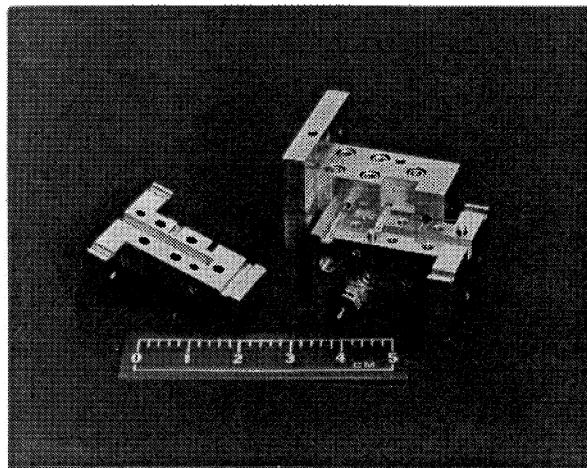


Figure 2. Split block WR-22 waveguide test fixture using alumina microstrip E-plane probe transitions, 44 GHz loss 0.7 dB per transition[1].

and WR-22 waveguide test fixture is identical, enabling test of the same device over a wide frequency range.

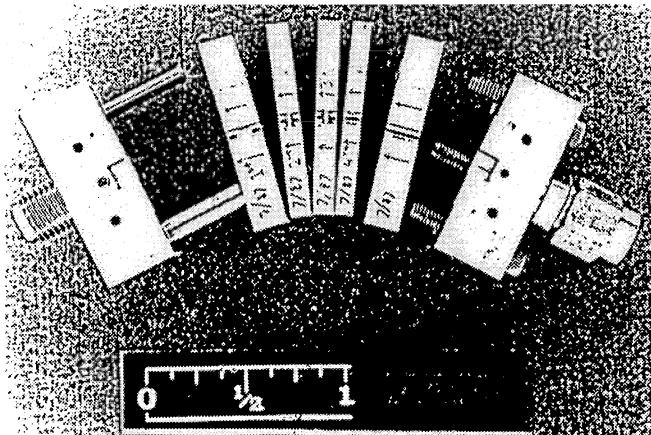


Figure 3. Calibration structures for de-embedding device performance from measurements in split block test fixtures[1].

Using test fixtures that are split with insertable center sections enables improved calibration and fixture measurement accuracy, if the interface of the insertable device is repeatable. Figure 3 shows a set of calibration structures that can be used to calibrate the coaxial and WR-22 microstrip test fixtures. The calibration structures include pairs of

shorts, opens, and loads. Microstrip through connections of two different line lengths are also provided to enable both short open load through (SOLT) or through reflect load (TRL) S-parameter calibration techniques. If the calibration structures are measured using a previously calibrated network analyzer, the measurement of the fixture calibration structures can also be used to determine both the dissipation and reflection losses of the fixture for measurements other than S-parameters.

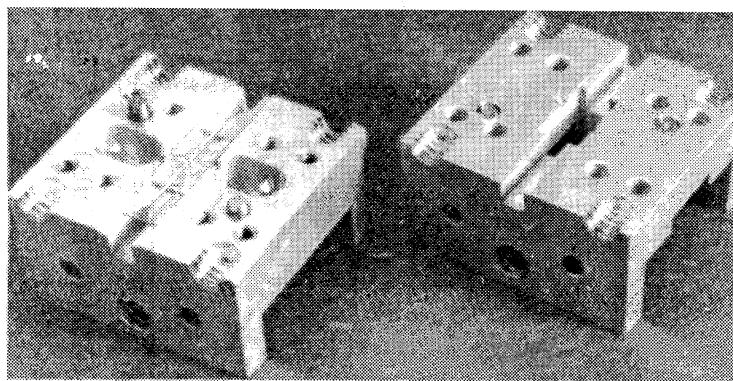


Figure 4. WR-15 waveguide test fixture using fused silica microstrip antipodal finline transitions, 60 GHz loss 0.4 dB per transition[3].

At 60 GHz and above, split block fixtures are also used, but it is much less common to use insertable calibration structures to de-embed the measurements. At the higher frequency, usually only the fixture through loss and VSWR is available and the measurement accuracy is not as good. Single piece test fixtures are common at the higher frequency because of the difficulty in providing repeatable, low VSWR with internal fixture interfaces. Figure 4 shows an example of a test fixture for device characterization at 60 GHz. This fixture uses antipodal finline waveguide-to-microstrip transitions. The transitions and matching networks are constructed on 0.13 mm thick fused silica. Similar fixtures are used through 140 GHz to evaluate discrete device and MMIC performance.

Device model development

Most device modeling tests now use on-wafer measurement. A typical on-wafer S-parameter test set is shown in Figure 5. This test set covers 40 MHz to 62 GHz. Recently on-wafer test sets that cover 40 MHz to 110 GHz with a single probe touch down were demonstrated by Wiltron. The on-wafer probe interface has better repeatability than most connectors and can be used to easily measure many device samples.

A difficulty with on-wafer measurement is verification of the measurement accuracy, since repeatability does not imply accuracy. Fully calibrated S-parameter measurement requires several on-wafer structures to de-embed the effects of the wafer probe and probe interface from the measurements. Several different methods have been used to provide the required calibration. The most accurate and most easily verified technique is a modification of the TRL S-parameter calibration technique developed by NIST[4,5]. This calibra-

tion method uses several different length transmission lines. The S-parameters were originally referred to the line impedance, but the most recent work provides the de-embedding to a fixed resistance, most commonly $50\ \Omega$. This calibration method provides accu-



Figure 5. RF wafer probe test system for 62 GHz provides accurate S-parameters for device modeling and performance assessment[1].

rate calibration, but the probes need to be physically moved to interface with the different line lengths during calibration.

A calibration method that does not require movement of one probe relative to another is desirable to automate the on-wafer calibration and measurement sequence. Figure 6

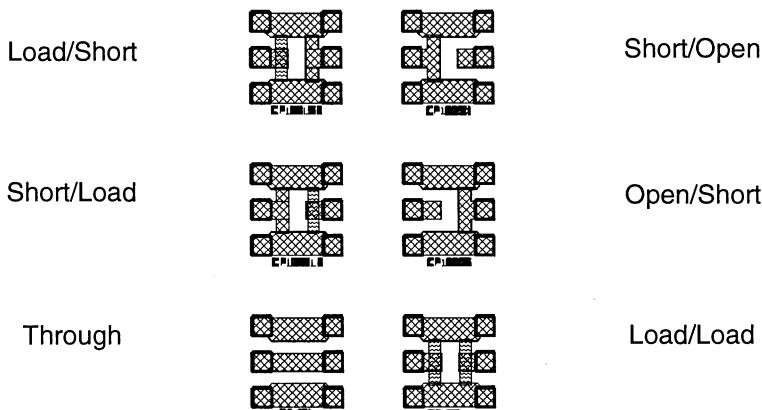


Figure 6. Coplanar wave guide on-wafer calibration structures for SOLT calibration designed with offset to insure identical probe to transmission line interface for each structure[6].

shows the SOLT on-wafer coplanar calibration structures that we use to automate on-wafer calibration and device measurement. The wafer probe interface is offset from the

open, short, and load to make the field pattern at the probe interface nearly independent of the type of structure under measurement. A critical assumption of all S-parameter calibration methods is that the interface is identical for every calibration standard. The offset is a compromise between small physical size and measurement accuracy. For improved accuracy, more offset would be desirable to insure that all evanescent waves on the coplanar line generated by the probe discontinuity were negligible. However increasing the offset may lead to other problems related to mixed microstrip and coplanar modes on thin wafers with backside metallization. The models used for the on-wafer calibration structures have been developed through numerical EM simulation. The verification of the models using the NIST coplanar calibration method is ongoing. There is some debate with NIST about the accuracy of this calibration method, but we have used it for all our device modeling for MMIC development through 140 GHz.

Almost all pHEMT device modeling uses coplanar on-wafer measurement, but microstrip is the most common design approach for pHEMT MMICs. Coplanar device

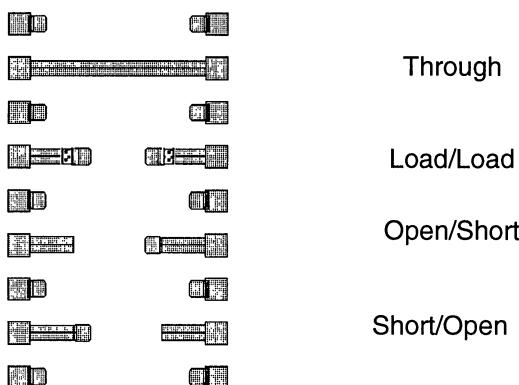


Figure 7. Microstrip on-wafer calibration structures for SOLT calibration designed with offset to insure identical probe-to-transmission line interface for each structure[6].

measurement is accurate for the internal device parameters because the device is usually small in terms of wave lengths. However, the inductive and capacitive parasitic elements are different in coplanar and microstrip environments. To provide accurate on-wafer measurement of microstrip devices and circuits, the calibration structures shown in Figure 7 were developed. Because these structures are not used on every wafer processed, the size was not as important as for the coplanar calibration structures. The open, short, and load are offset about two substrate thicknesses from the probe interface so that the probe interface is identical for each structure. Numerical EM simulation of each structure was done to establish the model used for calibration. The additional TRL microstrip lines shown in

Figure 8 were used with the NIST calibration procedures to verify the EM models.

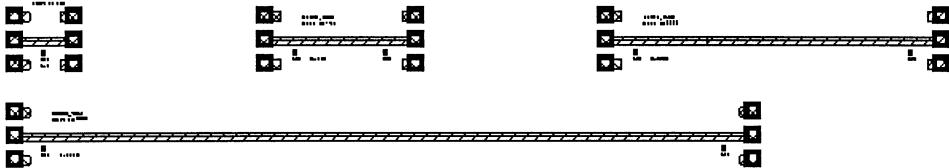


Figure 8. Microstrip on-wafer calibration structures for TRL calibration used to verify SOLT calibrations structure models[6].

Because of the high loss in the cables and wafer probes, on-wafer measurement of device parameters other than small signal S-parameters, are subject to significant errors. Fixture measurement of device and MMIC performance in some cases provide superior accuracy, but on-wafer measurement provides much larger data sets. For certain measurements much of the loss of the probe and cables can be effectively calibrated.

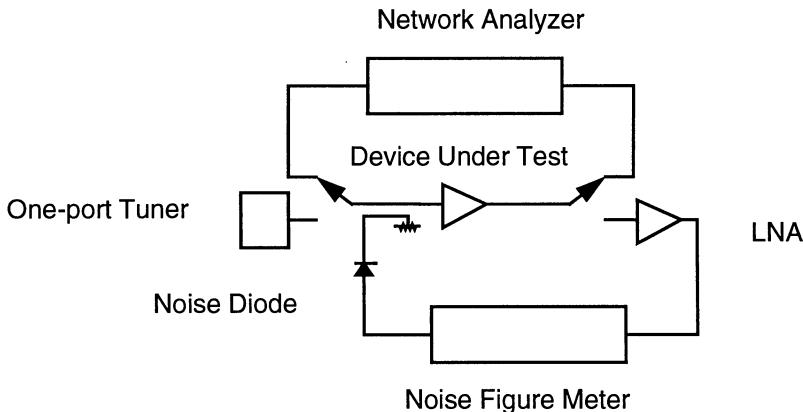


Figure 9. Typical noise parameter test set used for both fixture and on-wafer measurements

One of the significant advantages of pHEMT is the low noise figure of the device. For accurate design of low noise amplifiers, device noise parameters are required. Figure 9 shows a typical noise parameter test set. This test set measures the noise figure of the device under test with different input impedances. From this data the minimum noise figure and other device noise parameters are determined. For on-wafer test, the loss of probe and cables limit the range of input impedance available at the device input, and the cable loss requires significant measurement correction. These limitations and the very low noise figure at microwave frequency of the pHEMT result in many questionable noise parameter

measurements. Extreme care is needed to make usable noise parameter measurements.

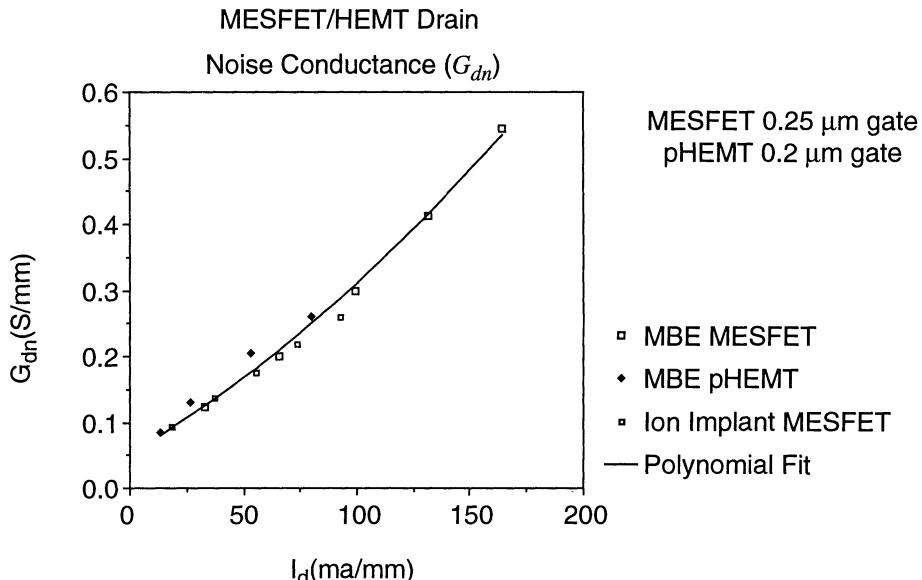


Figure 10. Comparison of measured drain noise conductance for pHEMT and 0.25 μm MESFETs.

A comparison of the measured output noise current source of pHEMT and MESFET is shown in Figure 10. Much of the device noise is due to drain current fluctuations, which appear to be nearly the same in MESFET or pHEMT. The lower noise figure of the pHEMT is due higher gain, especially at lower drain current density, compared to the MESFET.

Measurements needed for device modeling of pHEMT become more difficult as the nonlinearity of the device is modeled. The measurement needed to develop pHEMT nonlinear models depend on the form of the nonlinear model. There is reasonable agreement on the form of the device model for small signal S-parameters. For noise parameters, several different model forms are used, but for nonlinear modeling, even the form of the model has not been well established. For power amplifiers only a few db into compression, models such as the Curtice model give reasonable results, but at high gain compression the Curtice model is not effective. Other model approaches are also needed for accurate intermodulation prediction.

An important modeling task is the accurate prediction of the pHEMT nonlinearities that produce intermodulation. To check device models that model intermodulation, the test set shown in Figure 11 is used. The test set provides the level harmonic distortion as a function of bias. The different order harmonic provide the magnitude of same order non-

linearity as the harmonic order. For example, the level of the third harmonic will be pro-

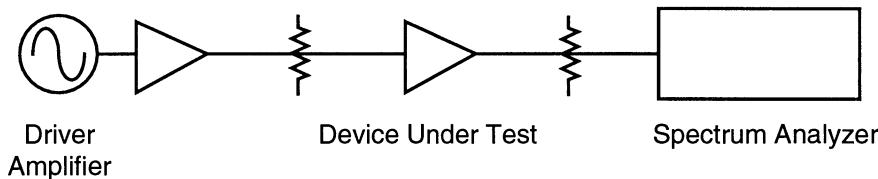


Figure 11. Harmonic generation mapping test set for distortion modeling.

portional to the level of any third order distortion product. A typical measurement of

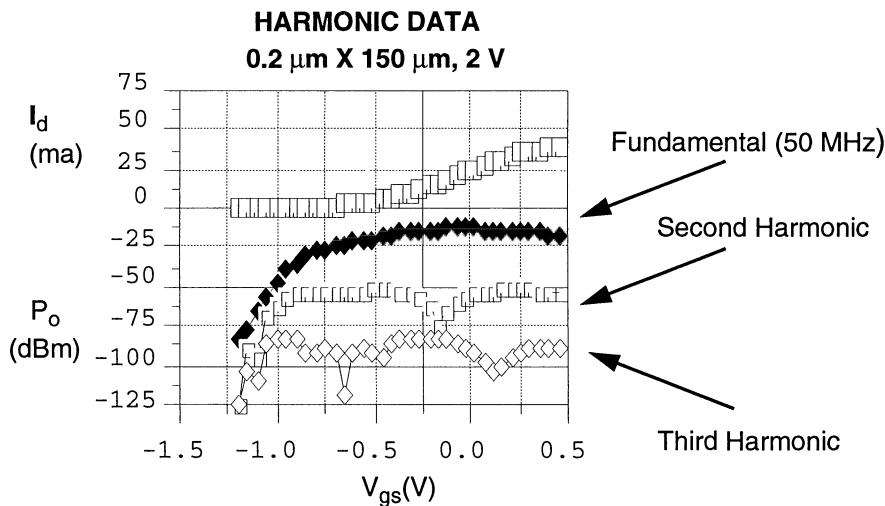


Figure 12. Measured pHEMT harmonic generation versus drain voltage, each harmonic amplitude is approximately proportional to transconductance derivative of the same order.

pHEMT distortion over bias is shown in Figure 12. The fundamental amplitude is proportional to the RF transconductance and the other harmonic orders provide accurate measurement of derivatives of the transconductance. These are the main causes of distortion in the pHEMT, only effects related to variable capacitance generated distortion would not be included in this measurement.

For power amplifiers, the optimum load needed for highest power or highest effi-

ciency is needed for design. The load pull test set shown in Figure 13 is used to map the

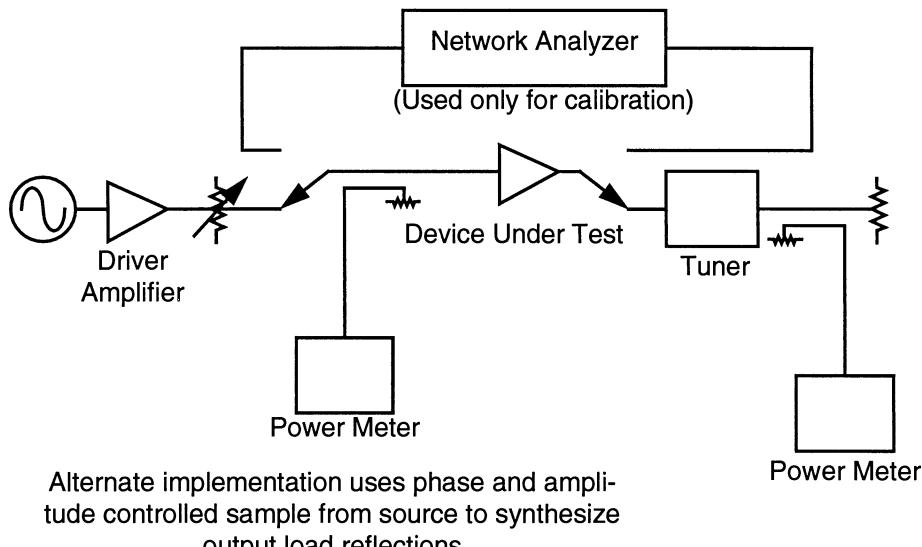


Figure 13. Load pull test set to measure optimum output load for maximum power or efficiency.

power output at different load impedance. Similarly to the noise parameter test set, loss between the tuner and the output of the device under test restricts the measurement tuning range and compromises the measurement accuracy. Unlike the noise parameter measurement, there is no generally valid mapping of measurement to theoretical contours because of the highly nonlinear nature of the problem. Although the test set only maps fundamental

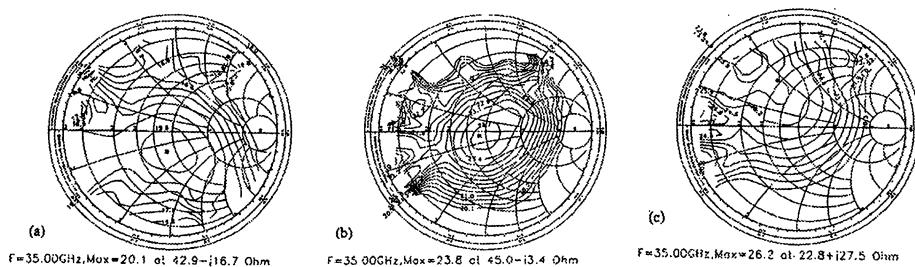


Figure 14. Measured on-wafer output power contours for Ka-band pHEMT devices with on-wafer matching[7].

tal impedances, the impedance at harmonic frequency will also effect the optimum fundamental tuning. The load pull measurement approximates the optimum load for maximum output power or efficiency. Figure 14 shows a typical set of power output contours for a matched millimeterwave pHEMT device. Because of the loss in the cable, probe, and tuner on-wafer matching circuits are usually required to make millimeter-wave load pull measurements. The matching networks are designed to be easy to model and de-embed if the device optimum load is required.

Circuit performance measurement

Measurement of circuit performance for any single performance parameter requires less accuracy than measurement for device modeling, but circuit performance measurement may include measurement of many parameters for each circuit. Many circuits require multiport measurement, which are inherently more difficult to make and calibrate than the two port measurement used for device characterization. Multiport and multifrequency measurement such as required for frequency conversion circuits are some of the most difficult measurements to make accurately. The wide frequency coverage of the pHEMT device may require circuit performance measurement to above 100 GHz[8,9]. Because of the wide range of circuit types and measurements possible, I will only cover a few measurement techniques.

The most common circuit performance measurement is two-port S-parameters. Figure 15 shows the S-parameter measurement of a three-stage pHEMT MMIC for 75-110

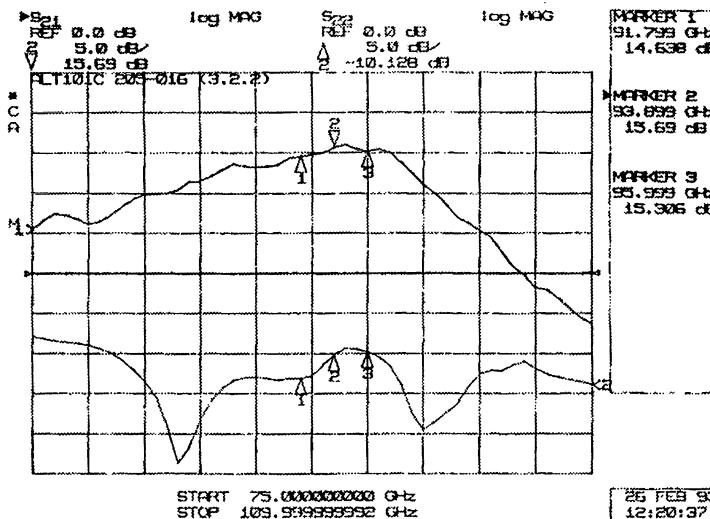


Figure 15. On-wafer measurement of 94 GHz three-stage pHEMT MMIC[9].

GHz. The techniques needed for circuit performance measurement are identical to the requirements for device S-parameters. However the required accuracy is not as difficult to

obtain for circuits as for devices, since most circuits are designed for reasonably low port VSWR. Measurement accuracy is a function of the product of the effective port reflection coefficients of the calibrated test set and device under test. The speed and ease of measurement make two-port S-parameter the most common RF performance screening measurement.

To extend S-parameter measurement to high power pHEMT amplifiers, the pulsed

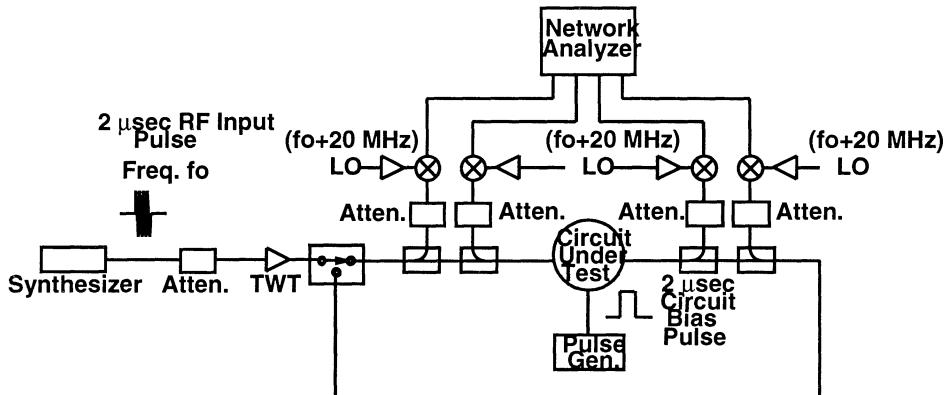


Figure 16. On-wafer millimeter-wave pulsed power test set[10].

power test set shown in Figure 16 was developed. This test set enables full power on-wafer measurement of millimeter-wave power MMICs. Both the RF drive and the DC drain

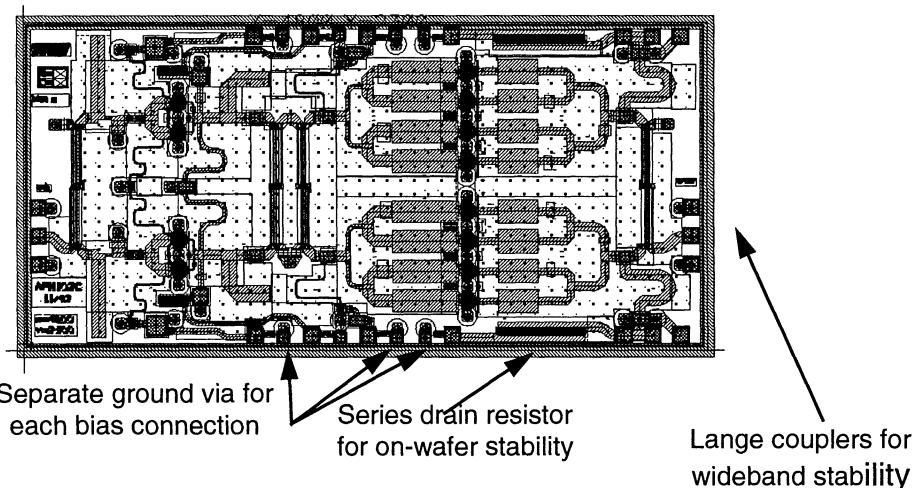


Figure 17. Ka-band chip designed for on-wafer power test[10].

voltage is pulsed to reduce the temperature of the channel during testing. The original test

set was developed as a custom test set using separate millimeter-wave mixers at each port, but similar test sets are now offered by Hewlett-Packard for millimeter-wave pulsed power measurement.

To enable full power testing of MMICs, several design features were incorporated into the chip. The chip layout and key design elements for on-wafer test are shown in Figure 17. Separate ground vias for each bias connection improves the low frequency stability and reduces the chance feedback in the bias circuit. A frequent cause of feedback during on-wafer test is the common inductance of the ground probes. When bypass capacitors are connected between ground and gate and drain supplies on the probe card, the bypass capacitors and inductance between the probe and chip ground results in unwanted feedback. Maintaining stability for on-wafer test becomes more difficult as the device periphery increases. For example the MMIC in Figure 17 has over 4 mm of gate periphery, resulting in a total transconductance of more than 2 S if the gate and drain of both stages were paralleled.

Heating Effects with 2 μ sec RF Pulse Length

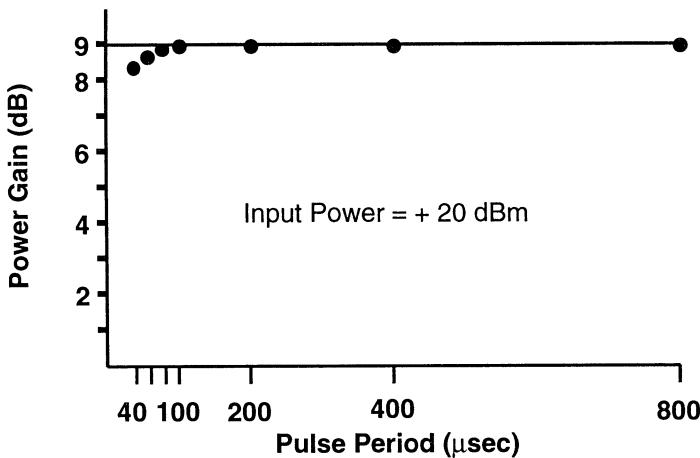


Figure 18. Ka-band MMIC test results for on-wafer pulsed power measurement showing effect of device heating on power output, and advantage of pulsed measurement[10].

The results for on-wafer pulsed power testing is shown in Figure 18. As the duty cycle increases the temperature rise in the chip decreases. Most of the testing was done at 0.5% to 1% duty cycle. At 5% duty cycle the output power has dropped by nearly 1 dB compared to 1% duty cycle. Full power on-wafer CW testing would likely destroy the part due to the poor thermal resistance. With the improved heat sinking in fixture testing, very little change in output power between CW and pulse was observed. In a test fixture with 21-22 dBm input power, the chip has >30 dBm CW output.

Testing of multiport frequency conversion circuits is difficult because each test set is usually custom designed[11,12]. The ports are at different frequencies which complicates

the calibration and test problem. The loss of the test fixture or wafer probe at each port should be independently determined. The frequency converter pump signal should be present at the local oscillator port for all measurements, since port impedances will depend on the LO drive. For good conversion loss accuracy and repeatable conversion spurious measurements, all ports need to have low VSWR. The harmonics, intermodulations products and spurious conversion products in a mixer can reflect back into the mixer ports resulting in frequency dependent ripple in measured performance. Attenuators are often used to insure wideband match at the test ports. Measurement of port reflection coefficients are also complicated by the unwanted frequencies present in a mixer. Commonly a spectrum analyzer and directional coupler are used to measure port return loss so that unwanted mixer products can be avoided. Modern network analyzers can also be used for port reflection measurements, but spurious signals degrade measurement accuracy. With care in test set design, conversion loss accuracy of a few tenths dB are achievable.

The high performance of pHEMT devices, coupled with advances in MMIC design and development, lead to some specialized testing requirements. Even for millimeterwave applications, MMICs can be designed to be small (at least in one dimension) compared to the wavelength, leading to quasi-optical system configurations that borrow from optical system design. Both transmit and receive applications benefit from quasi-optical designs. For transmit applications, spatial power combining provides the potential to efficiently combine the output power from hundreds of power amplifier MMICs. For receive systems, quasi-optical design simplifies generation of multiple receive antenna beams. Testing of quasi-optical systems requires very different test techniques from traditional coaxial or waveguide systems.

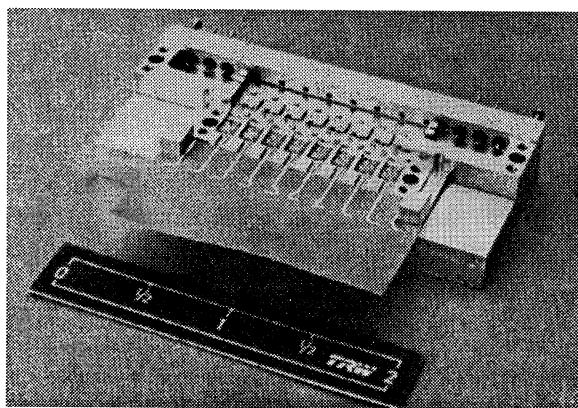


Figure 19. Direct-detection Imaging Array requiring quasi-optical testing[13].

Shown in Figure 19 is part of a quasi-optical multiple beam receiving system. This system uses a lens antenna to focus received signals on the tapered slot antennas. Each of the 8 channels has a 94 GHz LNA/detector chip to provide a DC output proportional to the input power. The system is used to provide a millimeter-wave thermal image. Testing of

this system includes both conventional and quasi-optical test techniques. Each receiver chip can be on-wafer probed to provide known good die for assembly. The sensitivity of the 8-channel assembly is tested by placing a large millimeter-wave absorbers covering the field of view of the tapered slot antennas. These absorbers are about 1 m X 1 m. By using two different temperature absorbers, the temperature calibration of the DC output voltage can be obtained. Measurement of the RMS noise output is also made to determine the system sensitivity. Similar tests are made when the 8-channel assembly is placed behind the lens. There is very little commercial test equipment designed for quasi-optical testing especially below 220 GHz, so that most test sets are custom designed.

RELIABILITY

Establishing the reliability of pHEMT is critical for system application of the technology. Expected device operating life times need to be at least several years for almost any application. For long mission space applications, mean time to failure (MTTF) should be even longer. The most desirable and accurate reliability tests would be to test a large sample of parts at the device operating conditions for the required mission life. However with some mission life requirements exceeding 10-20 years, methods that accelerate device degradation and failure are desirable. To accurately assess the reliability under accelerated aging, the physical mechanisms that result in degradation or failure need to be identified. From earlier reliability work on GaAs MESFETs, several reliability limiting mechanisms, such as gate metal diffusion, ohmic contact degradation, and metal migration, were identified that are accelerated by high temperature. There are other reliability limiting physical processes that can not be effectively accelerated by temperature, such as hot electron damage, hydrogen reactions, electrostatic discharge (ESD). However most of these mechanisms can be avoided by proper handling, packaging, and operating conditions. Reliability testing is used to establish the acceptable handling, packaging and operating limits. There was the additional concern that the more complicated structure of the pHEMT compared to the MESFET would lead to additional failure mechanisms not present in the MEFSET. Although reliability investigations to date have not identified failure mechanisms that are not shared by the MESFET, the pHEMT is more sensitive than the MESFET to several of the failure mechanisms. The increased sensitivity is a direct result of the reduced spacing between the gate and the channel, resulting in more sensitivity to changes in the local gate structure. Gradual degradation due to diffusion of gate metal toward the channel (gate sinking) observed in MESFET, is also present in the pHEMT. The effect on drain current for a given gate diffusion is greater in the pHEMT because of the more effective control of the drain current by the gate. In the pHEMT, gate sinking at first results in a shift in the gate bias voltage required for peak transconductance. The value of the peak transconductance may not change significantly, but if the gate bias voltage is fixed, the shift in threshold voltage will cause the operating transconductance to decrease. Using biasing techniques that maintain constant drain current provide protection against gain degradation due to small threshold shifts. However, even with the increased sensitivity of the pHEMT to threshold shifts, fixed gate voltage biasing provides more than acceptable reliability for > 10 year mission life.

To establish device reliability in a reasonable time, accelerated life testing is employed. Operation at high temperature accelerates the failure mechanisms that are most likely to result in long term degradation or failure of a well designed pHEMT circuit. The operating conditions of the accelerated life test should encompass the expected operating conditions expected in application. For low noise and small signal parts, d-c biasing is sufficient. However for large signal and power amplifiers, the most accurate life test should include both d-c and RF signals during the accelerated testing. In either case, RF tests should be used to set the failure conditions. During accelerated life testing, the mean time to failure at several operating temperatures are measured to establish the activation energy

of the dominant failure mechanism. The failure rate is proportional to $e^{-\frac{E_a}{kT}}$ where E_a is the activation energy of the failure mechanisms. The number of sample parts measured at each temperature is a compromise between accurate determination of the mean time to failure for each test condition and cost of running the life test.

For life test of small signal pHEMT we have developed several techniques to reduce the cost and complexity of the life test. The small signal accelerated life test uses the two-stage amplifier pHEMT MMIC shown in Figure 20. The use of a MMIC chip simplifies

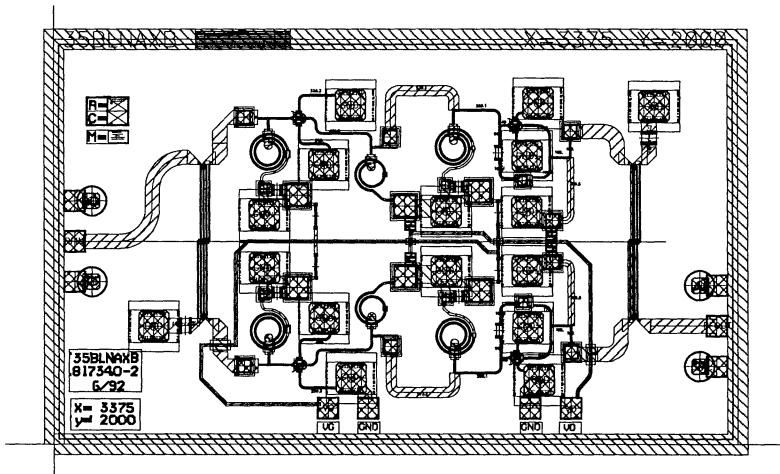


Figure 20. Two-Stage pHEMT MMIC used for small signal reliability test vehicle[14].

the biasing and device tuning to provide RF performance data. During temperature accelerated aging, the drain bias is maintained at a fixed current to maintain constant channel temperature. Periodically the devices under test are removed from the ovens and tested for RF and d-c performance at room temperature. The recorded gate voltage shift and the shape of the transconductance versus gate voltage can be used to estimate lifetime at fixed gate voltage biasing.

To establish the required operating temperatures for the life test, an initial step stress

is performed. In this test the devices under test are biased and operated for a fixed time at a given temperature. The performance of the parts is measured at room temperature after stress. The acceleration temperature is increased and the test repeated until at least half the sample fails.

For step stress test and accelerated life test, the header shown in Figure 21 was used. Two MMICs are mounted on each header. During accelerated aging, bias is applied to the MMICs through the mounting pins of the header. For room temperature performance measurement, RF probes are used contact the input and output of the mounted chip. This method eliminates the need for expensive test fixtures that can survive the high temperature of the life test. After mounting devices on the headers, burn-in at 125 C ambient for 320 hours is performed to screen out early failures. The burn-in and accelerated aging tests are performed in a automated test system using dry nitrogen purged ovens. The test sys-

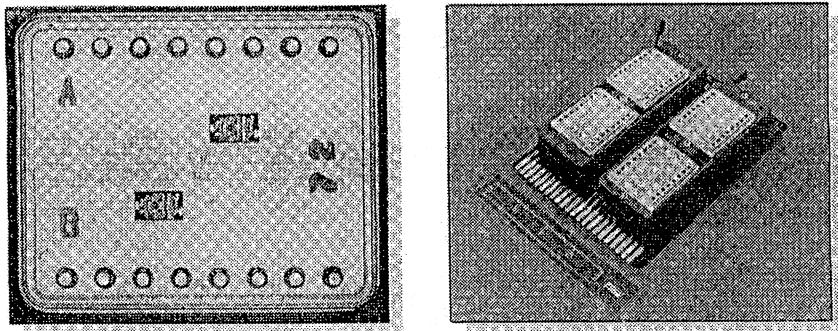


Figure 21. Low cost header and test socket for DC biased MMIC life test, room temperature
RF test data is obtained using RF probe.

tem measures and records the operating condition during the high temperature test. The automated data collection system and life test oven is shown in Figure 22. Periodically, the devices under test are brought to room temperature, removed from the ovens, and measured for RF and d-c performance. Accelerated testing continues until the 35 GHz room temperature small signal gain degrades by 1 dB or 0.5 dB per stage.

Life testing is used both for establishing MTTF for a MMIC process and to verify particular lots or wafers are reliable. To establish reliability of the pHEMT process, devices are selected from at least two lots. A complete three temperature life test is performed with 18 samples per temperature. The data is analyzed to provide expected operating life and activation energy of the major failure mechanism. To insure that particular lots or wafers

are reliable, single temperature accelerated aging test are used with the previously deter-

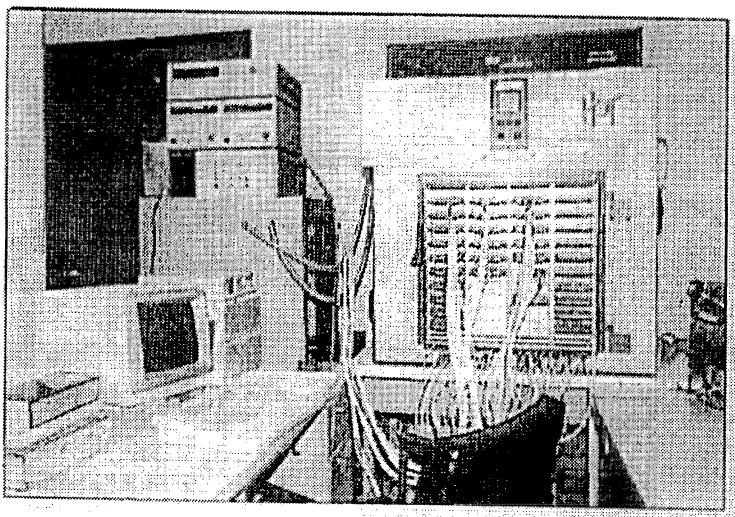


Figure 22. Reliability test ovens.

mined activation energy.

Shown in Figure 23. are typical frequency response measurements of the Ka-band

Gain vs Frequency Plot of Ka-Band LNA 230°C Ambient

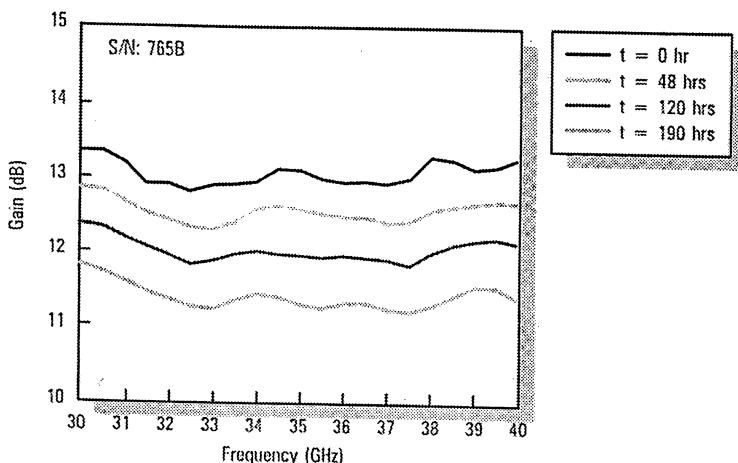


Figure 23. Typical room temperature amplifier frequency response after 230 C testing.

MMIC amplifier after testing at 230 C. The gain change is nearly identical at each frequency. All life test data analysis uses the gain at 35 GHz. The measured gain versus time and the shift in gate voltage required to produce 60 ma constant drain current is shown in

Figure 24. The MMIC amplifier has four 100 μm devices with a typical transconductance

Gain Change and Gate Voltage Shift During Life Test

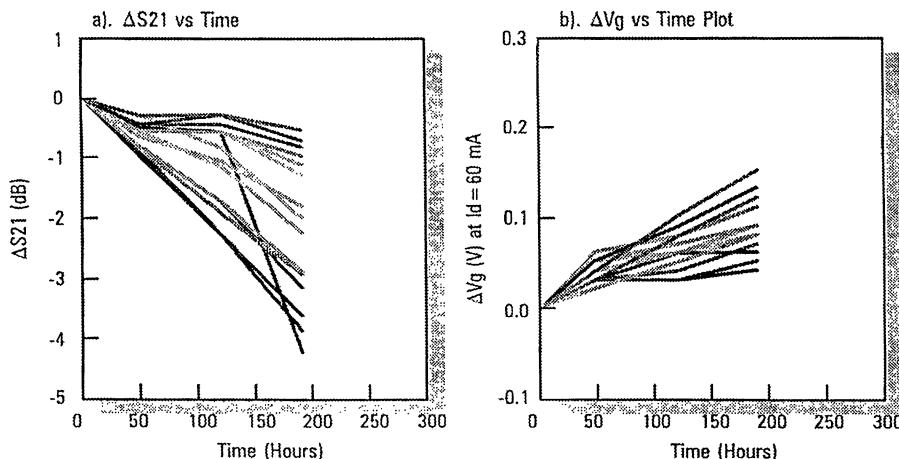


Figure 24. Measured room temperature center frequency gain and gate voltage shift for constant drain current as function of time at high temperature.

of about 400 mS/mm. With these parameters, the operating current decreases less than 25% for a 0.1 V shift in gate threshold if a constant gate bias voltage is used. The percent-

Lognormal Distribution of 35 BLNA, 3-Temp Life Test

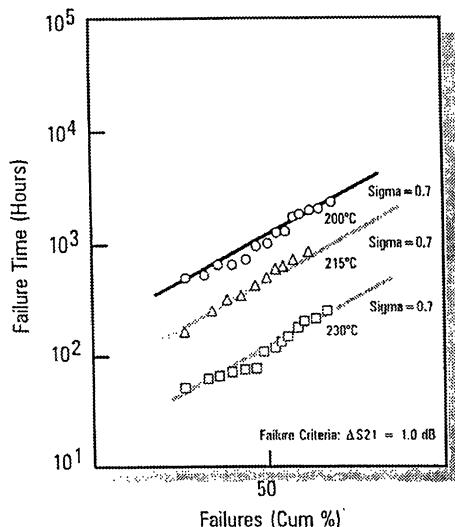


Figure 25. Log normal distribution of cumulative failures at 1 dB room temperature gain degradation.

age change in transconductance would be much smaller. By analyzing the shape of the transconductance curve an accurate determination of the effect constant gate voltage bias on life time is feasible.

The failure times and lognormal fits for devices at each of the three temperatures are plotted in Figure 25. The fits to the data are used to provide the expected time to 50% failure at each channel temperature. The Arrhenius plot to determine the activation energy is shown in Figure 26. This plot shows an extrapolated MTTF of over 10^7 hours at 125 C

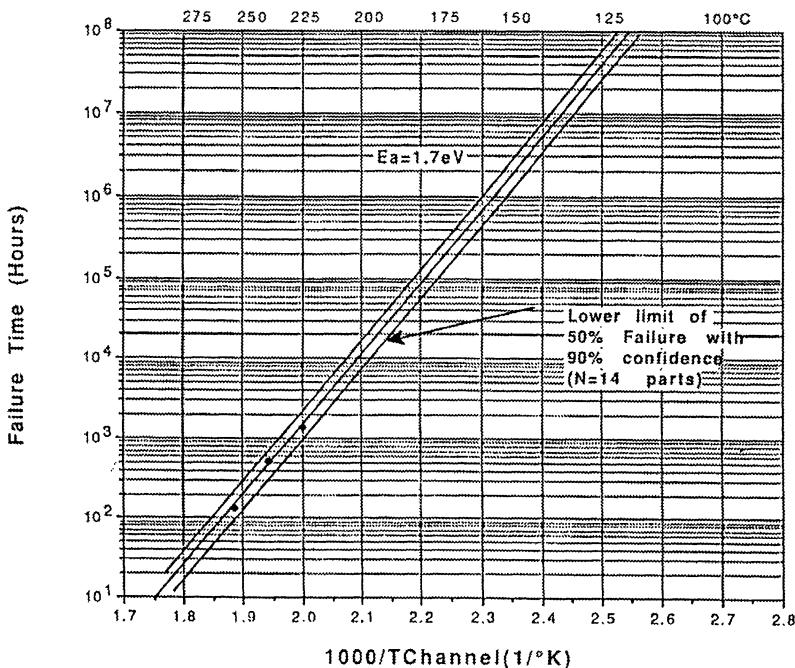


Figure 26. Projected small signal pHEMT mean time to failure.

channel temperature.

Accurate reliability assessment of power pHEMT devices and MMICs is more difficult. For small signal testing d-c biasing provides a good approximation to the device operating conditions, but power device should be tested under RF driven conditions. In addition to the complexity in the test set and fixturing to enable continuous RF drive, the decrease in device gain and output power capability at high temperature complicates matching RF stress under temperature acceleration and normal operating conditions. To date less data has been obtained for RF driven life tests than for d-c biased life tests.

Figure 27 shows the cumulative failure data and lognormal fits from a recently

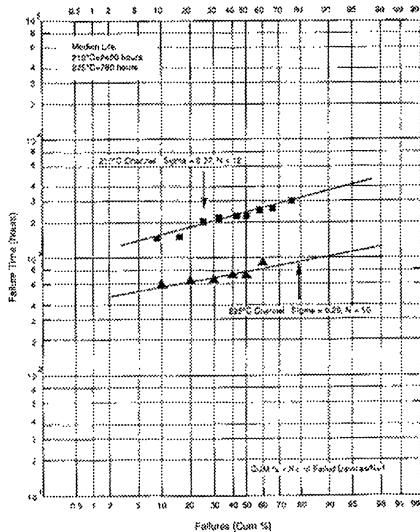


Figure 27. Log normal distribution of RF driven cumulative failures at 1 dB room temperature power output degradation[15].

reported RF driven pHEMT two temperature accelerated life test[15]. The measurements were taken on a pHEMT MMIC at 60 GHz with constant input power. The failure criterion

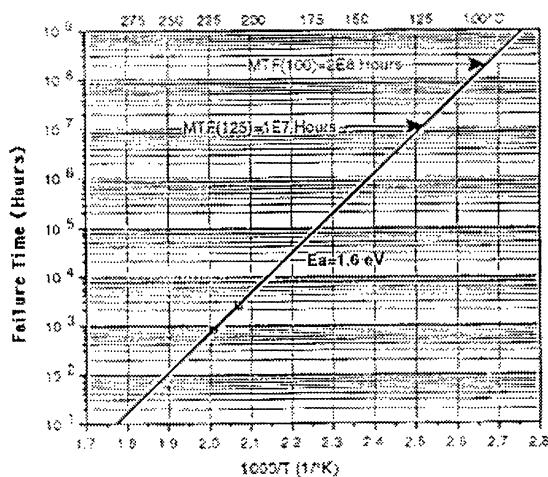


Figure 28. Projected power pHEMT mean time to failure[15].

in this case is a 1 dB drop in output power. Under RF driven conditions, the MTTF, as

shown in Figure 28, was only slightly lower than the small signal pHEMT.

CONCLUSIONS

Testing requirements for pHEMT circuits will continue to become more complex as pHEMT applications expand. The tests cover the full range of microwave and millimeter-wave applications. Reliability of pHEMT has been established through accelerated life test, but further work is needed, especially for large signal applications.

ACKNOWLEDGMENT

Many colleagues contributed to the work reported here. Although I could not list all, I would like to thank Ed Rezek for the small signal reliability data.

REFERENCES

1. Barry R. Allen, "Testability and Yield of MMICs," Joint Symposium, IEEE Antennas and Propagation Society and International Union of Radio Science (URSI), Chicago, July 1992.
2. A. Kurdoglian, C. S. Wu, W. Yau, J. Chen, M. Hu, C. Pao, and D. Bosch, "The demonstration of Ka-Band multi-functional MMIC circuits fabricated on the same PHEMT wafer with superior performance", *IEEE Microwave and Millimeter-Wave Circuits Symposium Digest*, pp. 97-98, 1993.
3. M. Sholley, B. Allen, S. Maas, A. Nichols, "HEMT Millimeter Wave Amplifiers, Mixers, and Oscillators," *Military Microwaves Conference Proceedings*, Brighton, England, June 1986, pp. 517-522.
4. R. B. Marks, "A Multiline Method of Network Analyzer Calibration," *IEEE Transactions on Microwave Theory and Techniques*, vol 39, July 1991, pp. 1205-1215.
5. R. B. Marks and D. F. Williams, "Characteristic Impedance Determination using Propagation Constant Measurement," *IEEE Microwave and Guided Wave Letters*, vol. 1, June 1991, pp. 141-143.
6. S. Chen, "Accuracy in On-Wafer Measurements; Design of Calibration Standards," *42nd Automated RF Techniques Group (ARFTG) Conference*, San Diego, CA, May 1994.
7. A. K. Sharma, G. S. Dow, M. Aust, J. Canyon, B. Allen, S. Pak, D. Yang, Y. Hwang, and K. Tan, "Ka-band power pHEMT on-wafer characterization using prematched

- structures,” *IEEE International Microwave Symposium Digest*, Atlanta, GA, June 1993, pp. 1343-1346.
8. S. M. J. Liu and G. G. Boll, “A new probe for W-band on-wafer measurements”, *IEEE MTT-S International Microwave Symposium Digest*, Atlanta, GA, June 1993, pp. 1335-1338.
 9. E. M. Godshalk, “A W-band wafer probe,” *IEEE International Microwave Symposium Digest*, Atlanta, GA, June 1993, pp. 171-174.
 10. M. V. Aust, B. Allen, G. S. Dow, R. Kasody, G. Luong, M. Biedenbender, and K. Tan, “A Ka-band HEMT MMIC 1 watt power amplifier,” *IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest*, Atlanta, GA, June 1993, pp. 45-48.
 11. H. Wang, K. W. Chang, T. N. Ton, M. Biedenbender, S. Chen, J. Lee, G. S. Dow, K. L. Tan, and B. R. Allen, “High-yield W-band monolithic HEMT low-noise amplifier and image rejection downconverter chips”, *IEEE Microwave and Guided Wave Letters*, vol. 3, no. 8, pp. 281-283, August 1993.
 12. K. W. Chang, H. Wang, T. H. Chen, K. Tan, J. Berenz, G. S. Dow, A. C. Han, D. Garske, and L. C. T. Liu, “A W-band monolithic pseudomorphic InGaAs HEMT downconverter”, *IEEE Microwave and Millimeter-Wave Circuits Symposium Digest*, pp. 55-58, 1991.
 13. D. C. W. Lo, L. Yujiri, G. S. Dow, T. N. Ton, M. Mussetto, and B. R. Allen “A W-band direct-detection radiometric imaging array,” *IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest*, San Diego, CA, May 1994, pp.41-44.
 14. J. Yonaki, R. Carandang, B. R. Allen, M. Hoppe, W. L. Jones, D. C. Yang, and C. L. Brunnenmeyer, “35 GHz InGaAs HEMT MMIC downconverter,” *IEEE Microwave and Millimeter-Wave Monolithic Circuits Symposium Digest*, Boston, MA, June 1991, pp. 47-50.
 15. C. H. Chen, Y. Saito, H. C. Yen, K. Tan, G. Onak, and J. Mancini, “Reliability study on pseudomorphic InGaAs power HEMT devices at 60 GHz,” *IEEE International Microwave Symposium Digest*, San Diego, CA, May 1994, pp. 817-820.

InP-BASED POWER HEMTs

MEHRAN MATLOUBIAN, AND LAWRENCE E. LARSON
*Hughes Research Laboratories,
Malibu, CA 90265*

Abstract

In recent years research in the development of InP-based HEMT's for microwave and millimeter wave power applications has increased dramatically. This has led to the development of InP-based power HEMT's with state-of-the-art performance. A review of the development of InP-based HEMT's for microwave and millimeter wave power applications along with some of the most recent results on InP-based power HEMT MMICs are presented in this work.

1. Introduction

Advantages of High Electron Mobility Transistors (HEMTs) over conventional GaAs MESFETs for millimeter wave applications have been well recognized [1]. Figure 1 shows the "evolution" of III-V FET structures that has led to the development of InP-based HEMT's. Performance improvement is achieved in the HEMT structure over conventional GaAs MESFETs by incorporating an AlGaAs/GaAs heterojunction which creates a band edge discontinuity that physically separates the donors in the wider bandgap AlGaAs, from the mobile electrons in the lower bandgap GaAs channel. This separation significantly reduces ionized impurity scattering of the electrons in the channel and leads to higher carrier mobility and higher f_T values. By adding indium to the GaAs

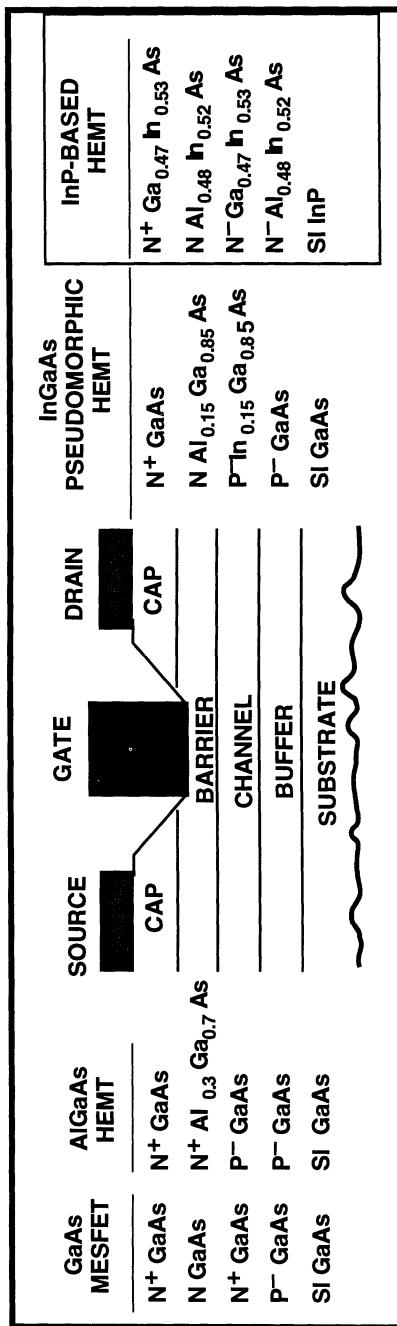


Figure 1. "Evolution" of III-V FETs from GaAs MESFETs to InP-based HEMTs.

channel to form GaInAs channels (GaAs-based pseudomorphic HEMTs), the device characteristics are further enhanced. The electron mobility, conduction band discontinuity (ΔE_C), and mobile carrier density all increase with increasing indium content; however, the addition of indium is limited by the fact that GaInAs is not lattice matched to the GaAs substrate, and the layer is increasingly strained as the indium content is increased. To avoid relaxation of the crystal structure and creation of dislocations in the channel, the indium content of AlGaAs/GaInAs/GaAs HEMT structures is typically limited to a maximum of approximately 30%.

In contrast, the InP-based HEMT incorporates an $\text{Al}_{0.48}\text{In}_{0.52}\text{As}/\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ structure that is lattice-matched to InP substrate (Figure 2). Compared with the GaAs-based pHEMTs, the $\text{Al}_{0.48}\text{In}_{0.52}\text{As}/\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ materials system exhibits higher electron sheet density, and a higher room temperature mobility (10,000 versus 6,000 $\text{cm}^2/\text{V}\cdot\text{s}$). In addition, the electron peak velocity in $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$, grown on an InP substrate, is also higher than that in pseudomorphic $\text{Ga}_{0.75}\text{In}_{0.25}\text{As}$, grown on a GaAs substrate. InP-based HEMTs exhibit a current gain cutoff frequency (f_T) of approximately 30% higher than that of the best GaAs-based HEMT or pHEMT.

The unique properties of the InP-based HEMT materials system have established this system as the leading transistor for millimeter wave low-noise applications [2-4]. They have set record performances for the fastest transistor operating at room temperature with an f_T of 343 GHz [5], and have set the world record for the fastest monolithic integrated circuit operating at a frequency of 213 GHz [6]. As mentioned earlier the characteristics that has made this materials system ideal for high frequency applications are due to several factors: very high mobility and velocity of electrons in the GaInAs channel, high conduction band discontinuity between the AlInAs and GaInAs allowing very high electron sheet concentration in the channel, and high modulation efficiencies [1]. Typical device cross-section for a low-noise AlInAs/GaInAs/InP HEMT is shown in Figure 3. All the layers are grown lattice-matched on a semi-insulating InP substrate. (To further increase the mobility of the channel, the indium content of the channel can be increased beyond 53% [5]. This results in a pseudomorphic InP-based HEMT).

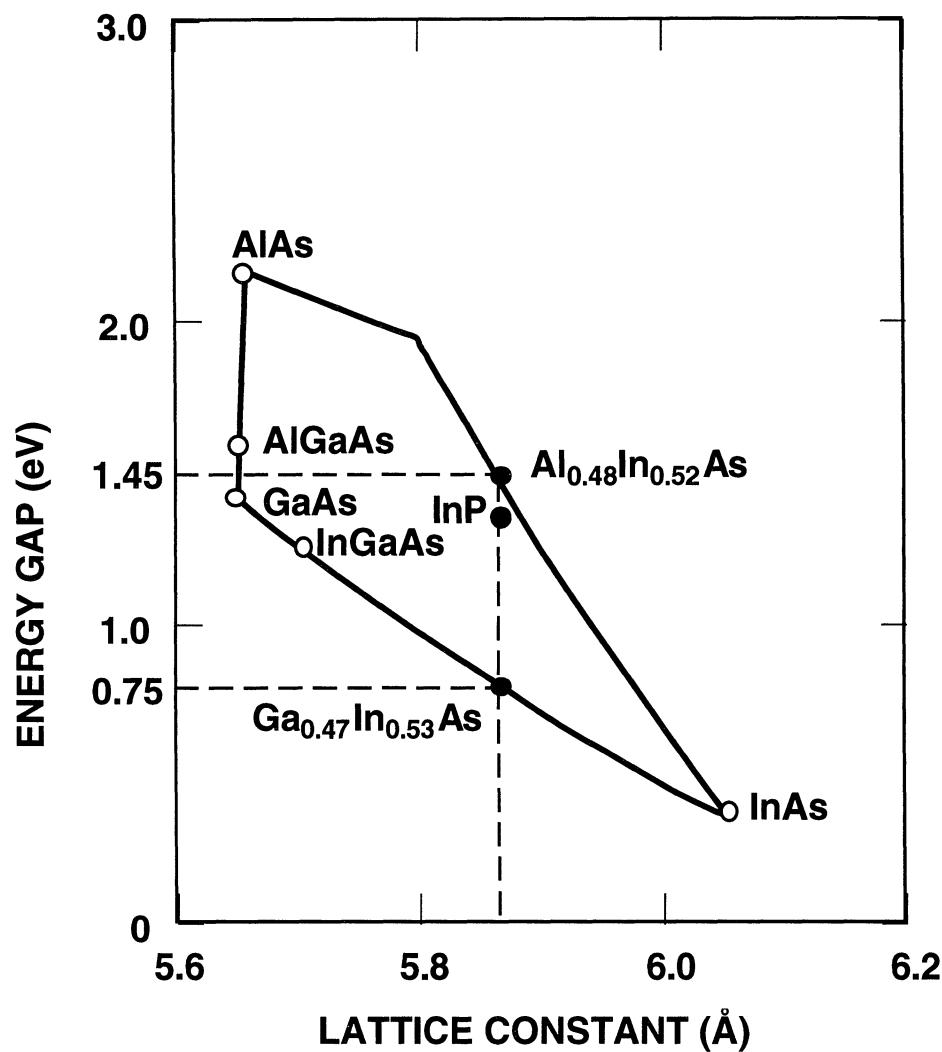


Figure 2. Band gap versus lattice constant for GaAs-based and InP-based FETs. $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ and $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ are lattice matched to InP substrate.

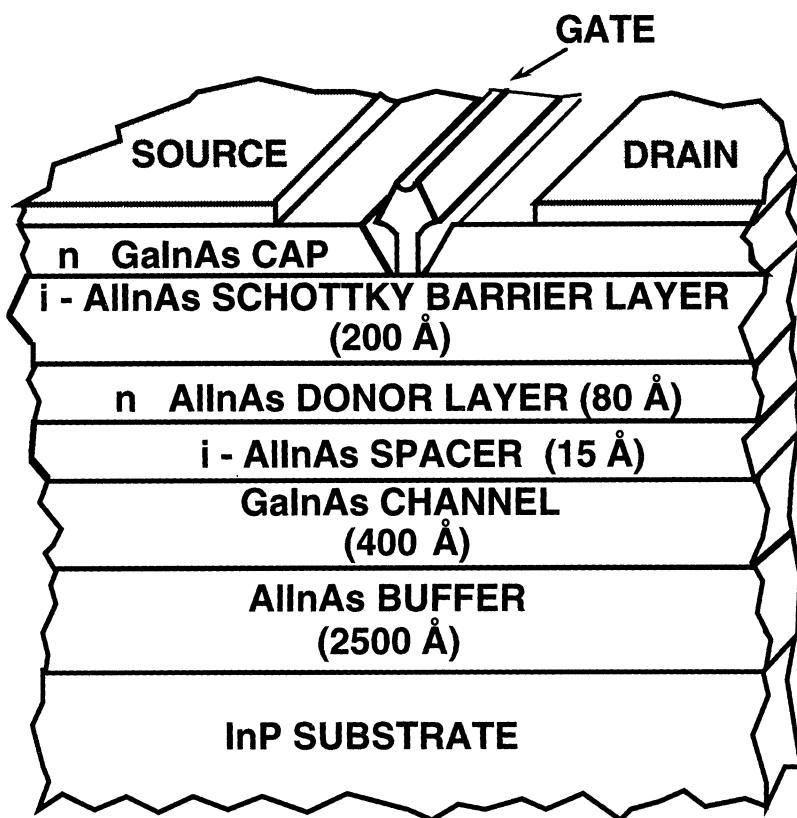


Figure 3. Typical device cross-section for a low-noise AlInAs/GaInAs/InP HEMT.

For power applications InP-based HEMT's offer a number of advantages over GaAs-based HEMT's. The thermal conductivity of InP is 40% higher than GaAs, allowing a lower operating channel temperature for the same power dissipation. In addition the larger conduction band discontinuity in the InP-based materials system leads to higher electron densities in the channel. Coupled with the higher electron velocity in the channel, higher current densities can be achieved. Despite these advantages, until recently, very little work had been done on InP-based HEMT's for power applications [7,8]. This has been primarily due to the low gate-to-drain breakdown voltage and low Schottky barrier height of low-noise InP-based HEMT's. But by proper device layer design it is possible to overcome the drawbacks of InP-based HEMT's for power applications, and achieve state-of-the-art power results. The approaches taken to optimize InP-based HEMT's for power applications will be reviewed in this work.

2. Power Transistors

Multiple interacting factors are involved to optimize the performance of power transistors. Figure 4 shows some of the major factors involved in optimizing the performance of power HEMTs. The optimization goal for a power transistor is to achieve an output power level with a particular gain and power-added efficiency. (In addition, in some applications, the reliability of the power transistor will also be very important.) The interaction between different factors can be very complex and optimization of some of these factors are sometimes in conflict resulting in more difficulty in optimization of devices for power applications as compared for low noise applications. For example to achieve higher gains both shorter gate-lengths and higher transconductances are desirable. But the gate-to-drain breakdown voltage of a transistor decreases as the gate-length is reduced. Also, to achieve a higher transconductance, a deeper gate-recess is needed, but a deeper recess increases the tunneling and thermionic emission current and results in reduction of gate-to-drain breakdown voltage. Therefore, optimization of power transistor performance requires careful optimization of a number of conflicting device parameters.

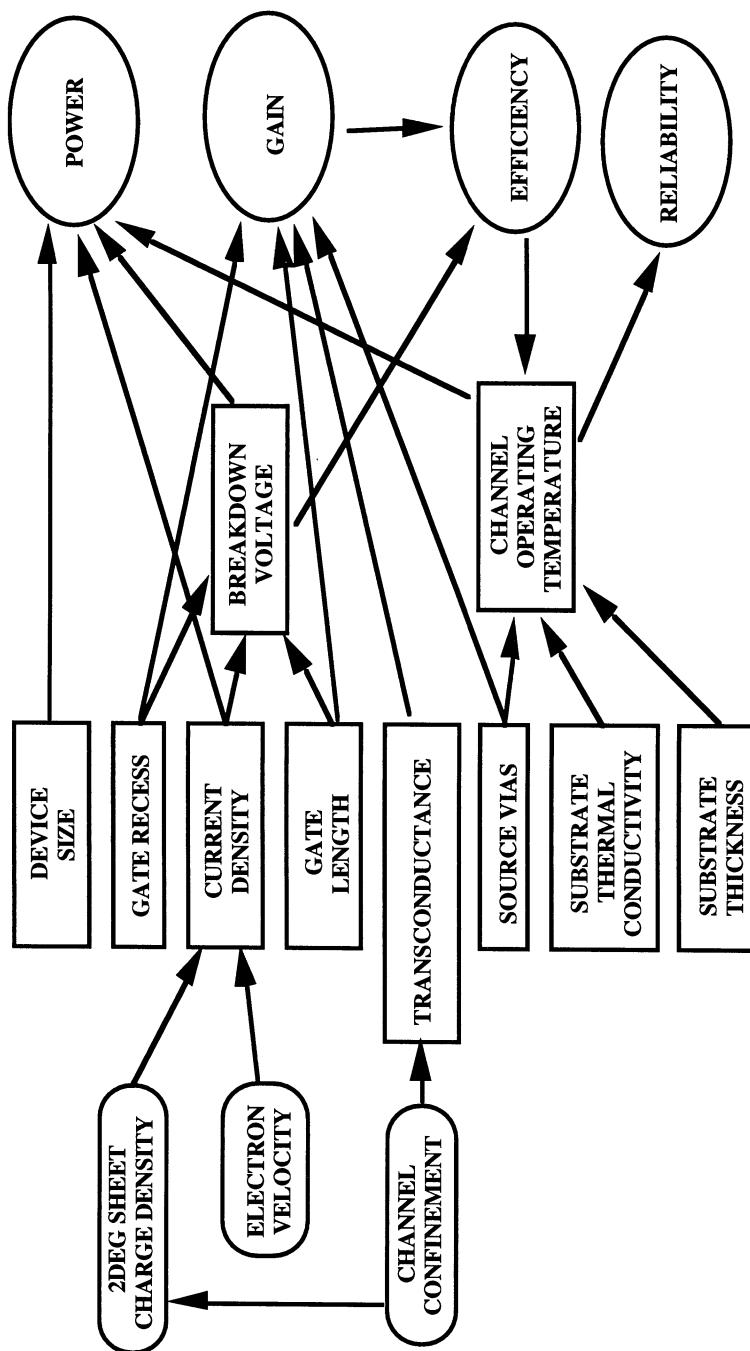


Figure 4. Multiple interacting factors are involved to optimize the performance of power HEMTs. Optimization of power transistor performance requires careful optimization of a number of conflicting device parameters.

The basic requirements for a good power transistor for microwave and millimeter wave applications are the following: low knee voltage, high gate-to-drain and drain-to-source breakdown voltage, high current density, and high gain. Depending on the frequency of operation certain device parameters become more critical. For example for high power microwave HEMTs operating below 20 GHz the drain-to-source breakdown voltage is very important because typically the operating voltage is approximately 7 to 10 V. Therefore it is desirable to optimize the material layer design and device layout to achieve breakdown voltages of more than 15 V.

But at millimeter wave frequencies the operating voltage of the device is limited to less than 5 V due to the rapid drop of f_T with the drain-to-source bias. Therefore, breakdown voltages of approximately 10 V are adequate for most applications and maximizing the power gain (without sacrificing output power) to improve the power-added efficiency typically becomes the most critical parameter.

Until recently, little work had been done on InP-based HEMT's for power applications due to the low gate-to-drain breakdown voltage and low Schottky barrier height of these HEMT's. There are two important breakdown voltages that determine the power performance of InP-based HEMTs the off-state breakdown voltage and the on-state breakdown voltage. The three-terminal off-state breakdown voltage is the drain-to-source breakdown voltage of the transistor with the device pinched-off. The off-state breakdown voltage of InP-based HEMT's has been shown to be a two step process [9]. The off-state breakdown voltage of AlInAs/GaInAs/InP HEMT's is believed to be dependent on a combination of thermionic-field emission of electrons from the gate into the channel, and impact ionization in the channel. The on-state breakdown voltage (drain-to-source breakdown with $V_{gs}=0$ V) is an important parameter because it limits the operating bias of the device for Class A or Class AB operation. The on-state breakdown voltage in these HEMT's has been shown to be dominated by impact ionization in the channel layer [10], and is typically low due to the low breakdown field of the narrow bandgap $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel. Therefore, increasing the bandgap of the channel should have an impact on improving both the off-state as well as the on-state breakdown voltages of these HEMT's.

3. Material Optimization

The efforts in improving the power performance of InP-based HEMT's can be divided into two major categories (1) Variations in the Schottky layer design, and (2) Variations in the channel layer design. Of course there are a number of other material design parameters that also impact the power performance of the device. Such as; doping and thickness of the Cap layer [11, 12], buffer layer design [13], and 2-DEG concentration in the channel as well as distribution of the charges in the donor layers [14]. In this discussion we will primarily focus on Schottky layer and channel layer design variations.

Figure 5 shows three approaches taken in modifying the Schottky layer to improve the breakdown voltage of InP-based HEMT's. The difference in layer structure between the different approaches has been highlighted. In the first approach the Al mole fraction in the AlInAs Schottky layer can be increased from 48% (the lattice matched case) to as high as 70%. The increase in the Al mole fraction of the AlInAs layer increases the bandgap of the Schottky layer resulting in improvement of the turn-on voltage as well as the gate-to-drain breakdown voltage [15, 16]. The disadvantage of such a scheme to improve the breakdown voltage of the device is the potential source of reliability problems due to the high Al content of the Schottky layer and the use of a strained Schottky layer.

In the second approach alternative material systems have been used instead of AlInAs to improve the gate-to-drain breakdown voltage. The Schottky can consist of such materials as AlInP or AlInAsP [17-19]. The advantages of these materials is that they have lower aluminum content as well as they can be grown with less strain onto InP substrates. But these novel material systems are less mature and require more study to optimize their growth conditions to achieve more consistent results.

In the third approach a pn-junction is formed instead of the Schottky gate to modulate the electrons in the channel. In the so-called JHEMT (Junction HEMT) approach a heavily doped p-type AlInAs layer is grown on top of a standard HEMT layer structure [20]. In addition to improving the breakdown voltage of the HEMT this approach has the advantage of achieving very uniform threshold voltages across the

Al-rich Schottky		AlnP/AlInAsP Schottky		JHEMT	
AlInAs	CAP	AlInAs	CAP	AlInAs	P+ - DOPED
AlInAs	SCHOTTKY	AlInAsP	SCHOTTKY	AlInAs	P-DOPED
AlInAs	CHANNEL	AlInAs	CHANNEL	AlInAs	N-DOPED
AlInAs	BUFFER	AlInAs	BUFFER	AlInAs	CHANNEL
InP	SUBSTRATE	InP	SUBSTRATE	InP	SUBSTRATE
<u>ADVANTAGES</u>		<u>ADVANTAGES</u>		<u>ADVANTAGES</u>	
Most mature InP-based power HEMT		Lower Al mole fraction		High gate-drain breakdown voltage	
High f_T and f_{max}		Potentially more reliable		Uniform threshold voltage	
<u>DISADVANTAGES</u>		<u>DISADVANTAGES</u>		<u>DISADVANTAGES</u>	
Oxidation of the Al-rich layer		Difficult to grow the Schottky layer		Higher parasitic capacitance	
Strained Schottky		Less mature device technology		Less mature device technology	

Figure 5. Variation of the Schottky layer design to improve the power performance of InP-based HEMTs.

wafer. But this technology is less mature and requires further optimization of the layer and device structure to achieve device operation at millimeter wave frequencies.

Figure 6 shows three common channel designs used for the development of InP-based power HEMTs. In the standard layer design a lattice matched $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel with a thickness of 200 to 300 Å is used [21]. This channel design, as mentioned earlier, suffers from low on-state breakdown voltage due to the low bandgap of the channel. One approach to increase the bandgap of the channel is to increase the Ga percentage in the channel (such as using a $\text{Ga}_{0.60}\text{In}_{0.40}\text{As}$ channel) [22].

Using the wider bandgap InP as the channel material also improves the drain-to-source breakdown voltage [23]. InP has a very high electron saturation velocity but InP has a lower mobility than $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ resulting in a higher sheet resistance and a higher source resistance compared to $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel HEMTs. In addition it is difficult to achieve very low contact resistances to InP without doping the AlInAs Schottky layer. This results in lower gain at millimeter wave frequencies and lower power-added efficiencies for InP channel HEMTs.

In the composite channel HEMT approach a combination of a thin layer of GaInAs and InP is used as the channel [24, 25]. This approach uses the advantages of both materials (high mobility of GaInAs at low fields, and high breakdown and saturation velocity of InP at high fields). By further reducing the thickness of the $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel it is possible to effectively increase the bandgap of the channel due to energy quantization [26]. Composite channel HEMTs with $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel thicknesses as small as 30 Å have been reported for high performance 0.15 μm InP-based power HEMTs with on-state breakdown voltages of more than 8 V [27]. Figure 7(a) shows a composite channel HEMT with a $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel with varying thicknesses. As the $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ channel thickness is decreased both the gate-to-drain breakdown voltage and the drain-to-source breakdown voltage is increased. Figure 7(b) shows that by decreasing the channel from 200 Å to 30 Å the breakdown voltage can be increased from approximately 6 V to over 10 V.

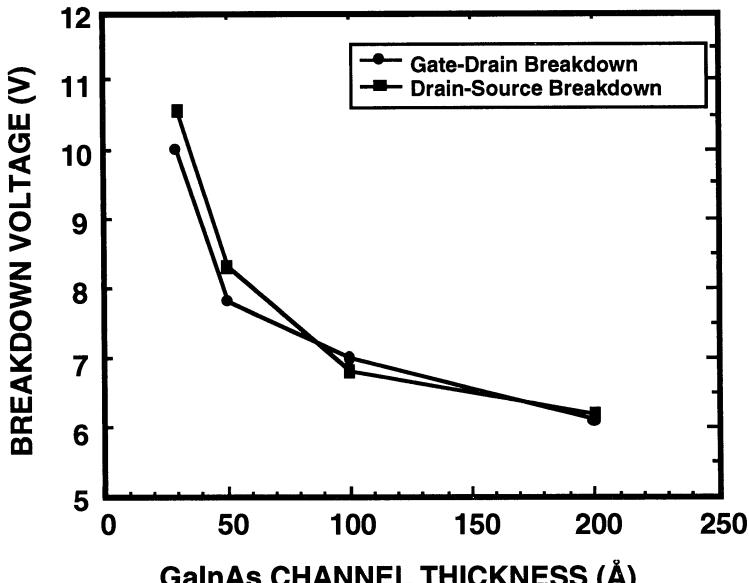
GaInAs CHANNEL		InP CHANNEL		COMPOSITE GaInAs/InP CHANNEL	
GaInAs	CAP	GaInAs	CAP	GaInAs	CAP
AllnAs	SCHOTTKY	AllnAs	SCHOTTKY	AllnAs	SCHOTTKY
GaInAs CHANNEL		InP CHANNEL		GaInAs/InP CHANNEL	
AllnAs	BUFFER	AllnAs	BUFFER	AllnAs	BUFFER
INP SUBSTRATE		INP SUBSTRATE		INP SUBSTRATE	
ADVANTAGES		ADVANTAGES		ADVANTAGES	
Most mature InP-based HEMT		Very high channel breakdown		High channel breakdown voltage	
f_T and f_{max} at low operating voltages		Higher velocity at high fields		High velocity at high fields	
DISADVANTAGES		DISADVANTAGES		DISADVANTAGES	
Low channel breakdown		Higher ohmic contact resistance		Less mature device technology	
Low electron velocity at high fields					

Figure 6. Variation of the channel layer design to improve the power performance of InP-based HEMTs.

$\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$	CAP	70 Å
$\text{Al}_{0.60}\text{In}_{0.40}\text{As}$	SCHOTTKY	250 Å
$\text{Al}_{0.48}\text{In}_{0.52}\text{As}$	Si DOPED	50 Å
$\text{Al}_{0.48}\text{In}_{0.52}\text{As}$	SPACER	15 Å
$\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$	CHANNEL	X Å
InP	Undoped CHANNEL	50 Å
InP	Doped CHANNEL	100 Å
$\text{Al}_{0.48}\text{In}_{0.52}\text{As}$	BUFFER	2500 Å

InP SUBSTRATE

(a)



(b)

Figure 7. (a) InP-based HEMT with a GaInAs/InP composite channel. The thickness of the GaInAs channel (X) is varied from 200 Å to 30 Å. (b) Dependence of the gate-to-drain and the drain-to-source breakdown voltage on the thickness of the GaInAs channel.

4. InP-Based Power HEMT's

Optimization of the material layer structure has led to the development of InP-based power HEMT's with state-of-the-art power performance at microwave and millimeter wave frequencies. Output power densities of more than 1 W/mm with very high power-added efficiencies have been reported at 4 and 12 GHz [7, 28]. Using a composite channel layer design and by optimizing the Schottky layer, gate-to-drain breakdown voltages of more than 20 V have been reported for 0.25 μm gate-length high-performance InP-based HEMTs [28]. At 60 GHz output powers as high as ~300 mW [21] and power-added efficiencies of as high as ~50% have been reported [29]. Table 1 shows a summary of the power performance of discrete InP-based HEMT's.

InP-based HEMT's are ideal for millimeter wave power applications where high gains and power-added efficiencies are desirable. Also, these HEMTs are ideal for applications requiring low-voltage (battery operated), and high-efficiency operation. For low-voltage operation of power transistors, low knee voltage, high drain current density, as well as high power gain are essential for achieving high power-added efficiencies. The efficiency of transmitters are critical in extending the life of batteries for millimeter wave portable wireless applications. Using InP-based power HEMT technology we can easily obtain electron sheet charge densities (2DEG) of more than $5 \times 10^{12} \text{ cm}^{-2}$ which has led to transistors with current densities of 1 A/mm. The high electron density combined with the high electron mobility in the channel also leads to a low source resistance and a low knee voltage. The combination of high current density, low knee voltage, and high gain make the InP-based power HEMT technology ideal for low voltage applications.

5. InP-Based Power HEMT MMICs

Development of high-performance InP-based HEMTs has also led to the development of InP-based power HEMT MMICs with state-of-the-art performance. In this section fabrication of InP-based power HEMT MMICs and some power HEMT MMIC results are presented.

Channel	Freq. (GHz)	Lg (μm)	Width (μm)	Power (mW)	PAE (%)	Gain (dB)	Ref.
δ-doped	4	0.25	2000	1100 600	50 66	14.1 13.4	31
composite	4	0.25	500	560	63	15	28
δ-doped	10	0.25	1000	450	59	11	31
composite	10	0.15	450	350 380	60 58	12 12.7	19
double-doped	12	0.22	150 300	110 288	50 40	11.1 11	7
double-doped	20	0.20	800	516	47	7.1	16
Ga _{0.47} In _{0.53} As	20	0.15	50	20.5 39	52 44	10.5 10.2	8
InP	30	0.3	100	145 120	13 23	---	23
δ-doped	60	0.15	448	155	30	4.9	
δ-doped	60	0.15	2x448	288*	20.4	3.6	
double-doped	60	0.20	448	145	24	4.2	21
composite	60	0.15	450	170	30	4	32
Ga _{0.47} In _{0.53} As	60	0.1	50	20.6 14.9	45 49	8 8.6	29
Ga _{0.33} In _{0.67} As	60	0.1	200 400	80 192	36 30	6.6 4.4	29
Ga _{0.47} In _{0.53} As	94	0.15	50	15	21	4	33
Ga _{0.33} In _{0.67} As	94	0.1	100 200	26 50	33 26	6 5	34

* Two HEMTs power combined

Table 1. Summary of the power performance of discrete InP-based HEMTs [30].

Figure 8 shows some of the key process sequences for fabrication of InP-based HEMT MMICs. The process starts with the evaporation and lift-off of the ohmic metallization (AuGe/Ni/Au) onto the InP substrate with the HEMT epi-layer grown on top. The active areas of the device are covered with photoresist to protect it from boron ion implantation used for device isolation. Using ion implantation rather than mesa isolation for device isolation has the advantage of a planar device structure. This prevents the overlap of the gate with the channel on the mesa edge which causes excessive gate leakage, and requires a separate selective etch step to undercut the channel [36].

Following the ohmic alloy step the gates are defined using e-beam in a bilayer resist process. Following the gate recess etch in a citric based etch, Ti/Pt/Au is evaporated and lifted-off to form the T-shaped gate. This step is followed by overlay metallization which in addition to covering all the ohmic patterns with a thicker Au layer also forms all the transmission lines and the matching circuits. The devices are then passivated with 500 Å of SiN_X, and for capacitors a 2000 Å layer of SiO_X is used. Using two different dielectric materials for passivation and capacitor dielectric allows the use of a thinner dielectric for device passivation to achieve higher gains at millimeter wave frequencies.

In the last top process step Au airbridges are fabricated to connect all the source fingers as well as forming the top plate of the capacitors. In the backside processing the wafers are thinned to a thickness of either 100 or 50 µm. Then the source vias are formed using a HBr/Br wet etch process. Finally the backside metal is evaporated and then plated with Au to a thickness of 5 to 8 µm.

Figure 9 shows a 20 GHz InP HEMT MMIC power amplifier [37]. The two stage amplifier consists of a 75 µm wide device in the first stage and a 250 µm wide device in the second stage. The chip size is 3.8 x 1.4 mm and was fabricated on a 100 µm thick InP substrate. The amplifier has a power gain of more than 20 dB with an output power of 120 mW and power-added efficiency of 45%. This amplifier is ideal for array applications where high-gain and high-power added efficiency along with a high thermal conductivity substrate are desirable.

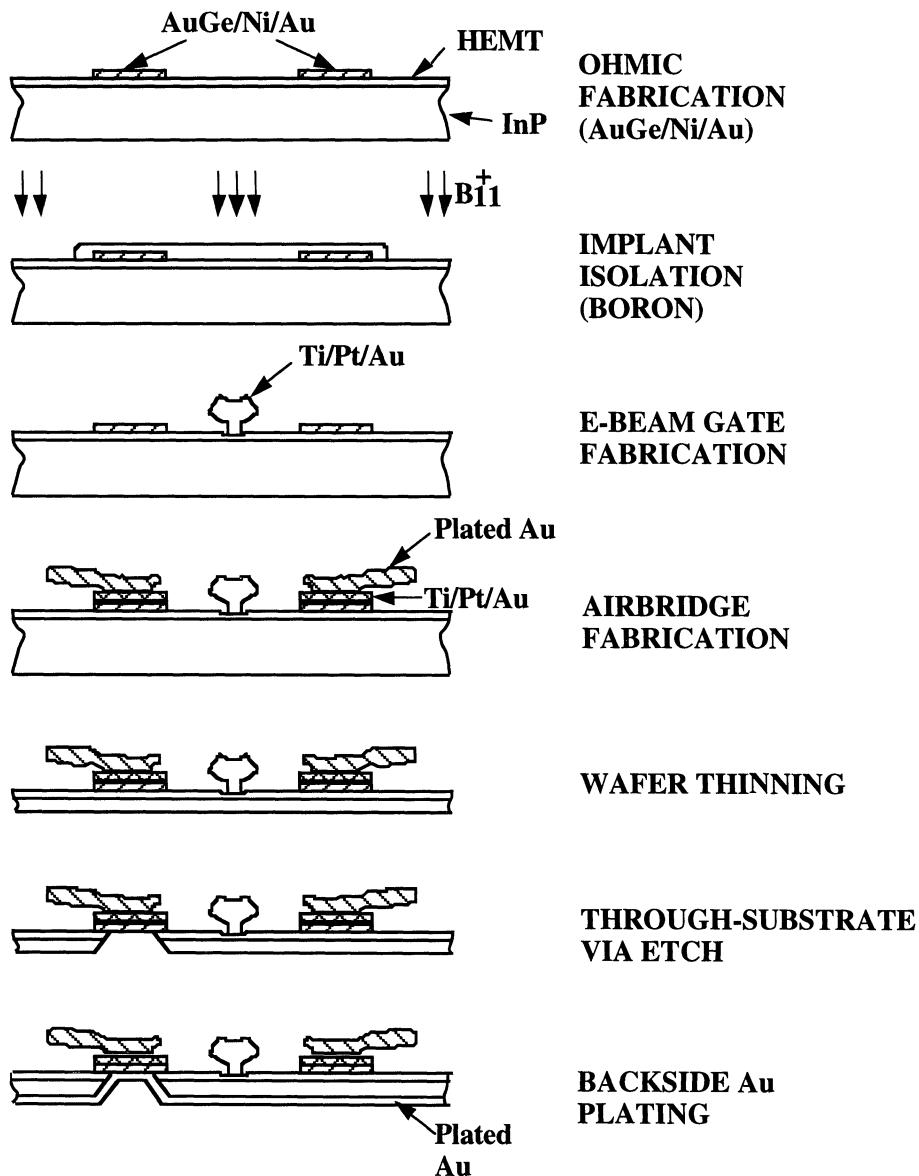


Figure 8. Key process sequences for fabrication of InP-based HEMT MMICs.

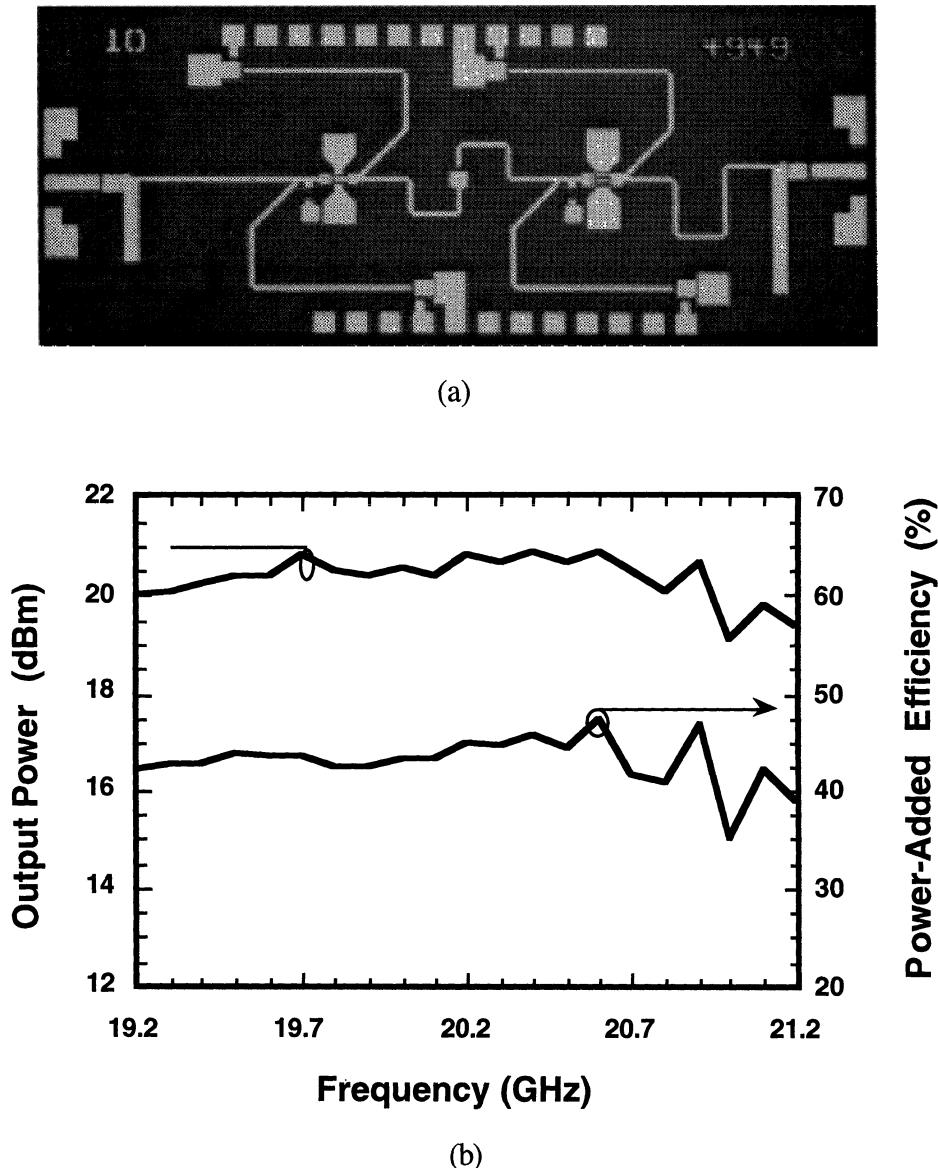


Figure 9. (a) 20 GHz InP-based HEMT MMIC power amplifier. The chip size is 3.8 x 1.4 mm with a chip thickness of 100 μm . (b) Measured power performance of the amplifier with an input power of -1 dBm [37].

In Figure 10 a 44 GHz power amplifier is shown consisting of two 600 μm wide HEMTs power combined on a 100 μm thick chip [38]. The chip size is 2 x 2.25 mm. The power amplifier has a linear gain of 9 dB and a power gain of 6.7 dB with an output power of 400 mW and 30% power-added efficiency. At 44 GHz an MMIC using a single 600 μm wide device has demonstrated an output power of 250 mW (420 mW/mm) with 33% power-added efficiency [38]. To achieve higher output powers at 44 GHz we have power combined four 450 wide HEMTs (1800 μm total gate periphery) on a single InP chip [39]. Figure 11 shows this chip which measures only 1.4 x 1.2 mm and has a thickness of 50 μm . Using a thinner chip allows a more compact circuit layout and lower thermal resistance for the chip. This amplifier has demonstrated state-of-the-art power performance with an output power of 800 mW with a power-added efficiency of 30% and 6 dB gain.

6. Summary

Recent progress in the development of InP-based power HEMTs have demonstrated that these devices are very promising for microwave and millimeter wave applications. State-of-the-art power results have been obtained from InP-based HEMT devices and MMICs. InP-based power HEMT technology is less mature than GaAs pHEMT technology but with further material, device, and process development the InP power HEMT technology will replace the GaAs pHEMT technology in some applications. In particular at W-band frequencies and above the higher gain offered by the InP-based HEMT technology will be instrumental in achieving high power-added efficiencies.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge the contributions of the staff of the Microwave and Devices Laboratory at the Hughes Research Labs., the staff of Hughes Space & Telecommunications Company, and Hughes Gallium Arsenide Operation (GAO) in Torrance.

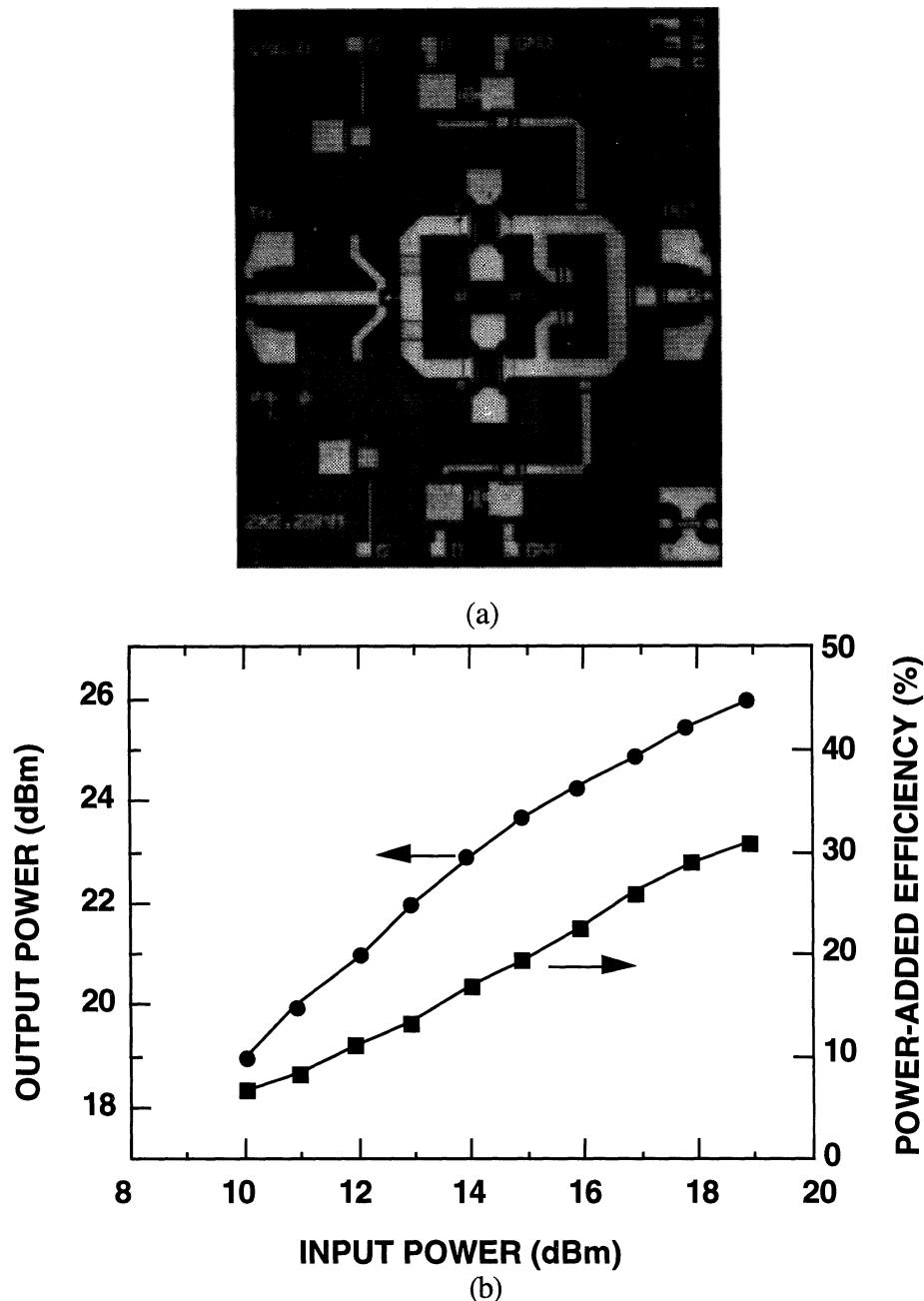
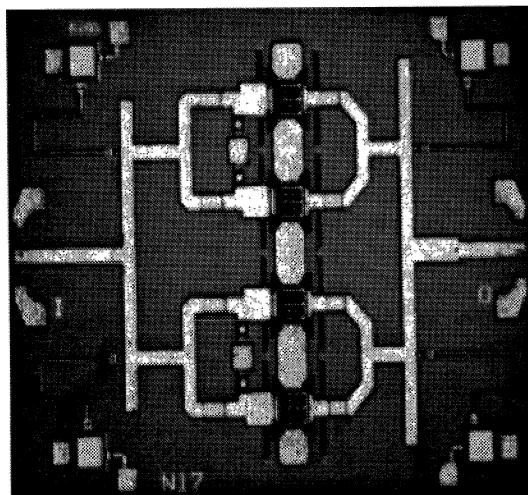
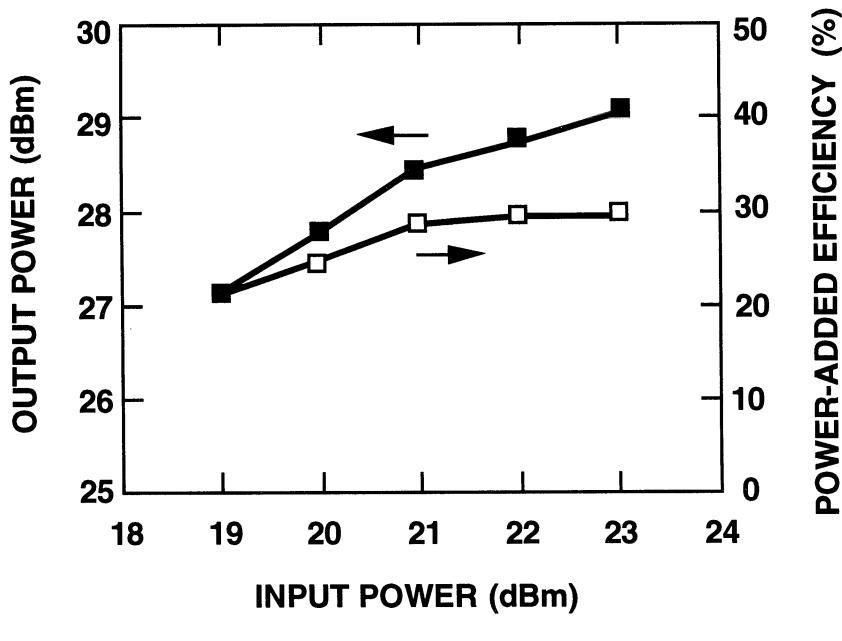


Figure 10. (a) 44 GHz InP-based HEMT MMIC power amplifier consisting of two $600 \mu\text{m}$ wide devices power combined. The chip size is 2×2.25 mm with a chip thickness of $100 \mu\text{m}$. (b) Measured output power and power-added efficiency versus input power [38].



(a)



(b)

Figure 11. (a) 44 GHz InP-based HEMT MMIC power amplifier consisting of four 450 μ m wide devices power combined. The chip size is 1.4 x 1.2 mm and has a final thickness of 50 μ m. (b) Measured output power and power-added efficiency versus input power for the amplifier [39].

REFERENCES

- [1] L.D. Nguyen, L.E. Larson, and U.K. Mishra, "Ultrahigh-speed modulation-doped field-effect transistors: A tutorial review," *Proceedings of the IEEE*, Vol. 80, No. 4, pp. 494-518, Apr. 1992.
- [2] U.K. Mishra, A.S. Brown, M.J. Delaney, P.T. Greiling, and C.F. Krumm, "The AlInAs-GaInAs HEMT for Microwave and Millimeter-Wave Applications," *IEEE Trans. Microwave Theory and Techn.*, Vol. 37, No. 9, pp. 1279-1285, Sept. 1989.
- [3] K.H.G. Duh, P.C. Chao, P. Ho, M.Y. Kao, P.M. Smith, J.M. Ballingall, and A. A. Jabra, "High performance InP-based HEMT millimeter-wave low-noise amplifiers," in 1989 IEEE MTT-S Int. Microwave Symp. Dig., pp. 805-808.
- [4] K.L. Tan, D.C. Streit, P.D. Chow, R.M. Dia, A.C. Han, P.H. Liu, D. Garske, and R. Lai, "140 GHz 0.1 μm gate length pseudomorphic In_{0.52}Al_{0.48}As/In_{0.60}Ga_{0.40}As/InP HEMT," in IEDM Tech. Dig., pp. 239-242, Dec. 1991.
- [5] L.D. Nguyen, A.S. Brown, M.A. Thompson, and L.M. Jelloian, "50-nm self-aligned-gate pseudomorphic AlInAs/GaInAs high electron mobility transistors," *IEEE Trans. Electron Dev.*, Vol. 39, No. 9, pp. 2007-2014, Sept. 1992.
- [6] S.E. Rosenbaum, L.M. Jelloian, A.S. Brown, M.A. Thompson, M. Matloubian, L.E. Larson, R.F. Lohr, B.K. Kormanyos, G.M. Reibez, and L.P.B. Katehi, "A 213 GHz AlInAs/GaInAs/InP HEMT MMIC Oscillator," in IEEE IEDM Tech. Dig., pp. 924-926, Dec. 1993.
- [7] M. Matloubian, L.D. Nguyen, A.S. Brown, L.E. Larson, M.A. Melendes, and M.A. Thompson, "High power and high Efficiency AlInAs/GaInAs on InP HEMTs," in 1991 IEEE MTT-S Digest, pp. 721-724.
- [8] M.Y. Kao, P.M. Smith, P.C. Chao, and P. Ho, "Millimeter wave performance of InAlAs/InGaAs/InP HEMTs," in Proc. IEEE/Cornell Conf. on Advance Concepts in High Speed Semiconductor Devices and Circuits, 1991, pp. 469-477.
- [9] S.R. Bahl, J.A. del Alamo, J. Dickmann, and S. Schildberg, "Off-state breakdown in InAlAs/InGaAs MODFET's," *IEEE Transactions on Electron Devices*, Vol. 42, No. 1, pp. 15-22, Jan. 1995.
- [10] J. Dickmann, S. Schildberg, A. Geyer, B.E. Maile, A. Schurr, S. Heuthe, and P. Narozny, "Breakdown mechanisms in the on-state mode of operation of InAlAs/In_xGa_{1-x}As pseudomorphic HEMTs," Proc. 6th Int. Conf. on InP and Related Materials, Santa Barbara, CA, pp. 335-338, 1994.
- [11] Y.C. Pao, C. Nishimoto, M. Riaziat, R. Majidi-Ahy, N.G. Bechtel, and J.S. Harris, "Impact of surface layer on In_{0.52}Al_{0.48}As/In_{0.53}Ga_{0.47}As/InP high electron mobility transistors," *IEEE Electron Dev. Lett.*, Vol. 11, No. 7, pp. 312-314, July 1990.

- [12] J. Dickmann, H. Dambkes, H. Nickel, R. Losch, W. Schlapp, J. Böttcher, and H. Künzel, "Influence of surface layers on the RF-performance of AlInAs-GaInAs HFET's," IEEE Microwave and Guided Wave Lett., Vol. 2, No. 11, pp. 472-474, Nov. 1992.
- [13] U. Auer, R. Reuter, C. Heedt, H. Künzel, W. Prost, and F.J. Tegude, "InAlAs/InGaAs HFET with extremely high device breakdown using an optimized buffer layer structure," in 1994 Proceedings of Int. Conf. on Indium Phosphide and Related Materials, pp. 443-446.
- [14] M. Matloubian, unpublished results.
- [15] C.L. Lin, P. Chu, A.L. Kellner, H.H. Weider, and E.A. Rezek, "Composition dependence of Au/In_xAl_{1-x}As Schottky barrier heights," Appl. Phys. Lett., Vol. 49, pp. 1593-1595, Dec. 8, 1986.
- [16] M. Matloubian, A.S. Brown, L.D. Nguyen, M.A. Melendes, L.E. Larson, M.J. Delaney, M.A. Thompson, R.A. Rhodes, and J.E. Pence, "20-GHz high-efficiency AlInAs-GaInAs on InP power HEMT," IEEE Microwave and Guided Wave Lett., Vol. 3, No. 5, May 1993.
- [17] K.B. Chough, C. Caneau, W.P. Hong, and J.I. Song, "Al_{0.25}In_{0.75}P/Al_{0.48}In_{0.52}As/ Ga_{0.35}In_{0.65}As graded channel pseudomorphic HEMT's with high channel-breakdown voltage," IEEE Electron Device Lett., Vol. 15, No. 1, pp. 33-35, Jan. 1994.
- [18] J.J. Brown, M. Matloubian, T.K. Liu, L.M. Jelloian, A.E. Schmitz, R.G. Wilson, M. Lui, L.E. Larson, M.A. Melendes, and M.A. Thompson, "InP-based HEMTs with Al_xIn_{1-x}P Schottky barrier layers grown by gas-source MBE," in 1994 Proceedings of Int. Conf. on Indium Phosphide and Related Materials, pp. 419-422.
- [19] L.M. Jelloian, M. Matloubian, T. Liu, M. Lui, and M.A. Thompson, "InP-Based HEMT's with Al_{0.48}In_{0.52}As_xP_{1-x} Schottky Layers," IEEE Electron Dev. Lett., Vol. 15, No. 5, May 1994.
- [20] J.B. Shealy, M.M. Hashemi, K. Kiziloglu, S.P. DenBaars, U.K. Mishra, T. Liu, J.J. Brown, and M. Lui, "High-breakdown-voltage AlInAs/GaInAs junction-modulated HEMT's (JHEMT's) with regrown ohmic contacts by MOCVD," IEEE Electron Device Lett., Vol. 14, No. 12, pp. 545-547, Dec. 1993.
- [21] M. Matloubian, L.M. Jelloian, A.S. Brown, L.D. Nguyen, L.E. Larson, M.J. Delaney, M.A. Thompson, R.A. Rhodes, and J.E. Pence, "V-Band High-Efficiency High-Power AlInAs/GaInAs/InP HEMT's," IEEE Trans. MTT, Dec. 1993.
- [22] M. Matloubian, et. al., "InP-based HEMTs with Ga_{0.60}In_{0.40}As channels," to be submitted.
- [23] O. Aina, M. Burgess, M. Mattingly, A. Meerschaert, J.M. O'Connor, M. Tong, A. Ketterson, and I. Adesida, "A 1.45-W/mm, 30 GHz InP-Channel Power HEMT," IEEE Electron Dev. Lett., Vol. 13, No. 4, May 1992.

- [24] T. Enoki, K. Arai, A. Kohzen, and Y. Ishii, "InGaAs/InP double channel HEMT on InP," in 1992 Proceedings of Int. Conf. on Indium Phosphide and Related Materials, pp. 14-17.
- [25] M. Matloubian, L.M. Jelloian, M. Lui, T. Liu, L.E. Larson, L.D. Nguyen, and M.V. Le, "GaInAs/InP composite channel HEMT's," 51st Device Research Conf., Santa Barbara, CA, June 1993.
- [26] S.R. Bahl, and J.A. del Alamo, "Breakdown voltage enhancement from channel quantization in InAlAs/n⁺-InGaAs HFET's," IEEE Electron Device Lett., Vol. 13, No. 2, pp. 123-125, Feb. 1992.
- [27] M. Matloubian, T. Liu, L.M. Jelloian, M.A. Thompson, and R.A. Rhodes, "K-band GaInAs/InP channel power HEMTs," IEE Electronics Lett., Vol. 31, No. 9, pp. 761-762, Apr. 1995.
- [28] M. Matloubian, L.M. Jelloian, M. Lui, T. Liu, and M. Thompson, "Ultra-High Breakdown High-Performance AlInAs/GaInAs/InP Power HEMTs," in 1993 IEEE IEDM Digest, pp. 915-917.
- [29] P. Ho, P.M. Smith, K.C. Hwang, S.C. Wang, M.Y. Kao, P.C. Chao, and S.M.J. Liu, "60 GHz Power Performance of 0.1 μm Gate-Length InAlAs/InGaAs HEMTs," in 1994 Proceedings of Int. Conf. on Indium Phosphide and Related Materials, pp. 411-414.
- [30] U.K. Mishra, and J.B. Shealy, "InP-based HEMTs: Status and Potential," in 1994 Proceedings of Int. Conf. on Indium Phosphide and Related Materials, pp. 14-17.
- [31] J.J. Brown, A.S. Brown, S.E. Rosenbaum, A.S. Schmitz, M. Matloubian, L.E. Larson, M.A. Melendes, and M.A. Thompson, "Study of the Dependence of Ga_{0.47}In_{0.53}As/Al_xIn_{1-x}As Power HEMT Breakdown Voltage on Schottky Layer Design and Device Layout," in 1993 Device Research Conference Digest.
- [32] M. Matloubian, L.M. Jelloian, M. Lui, T. Liu, L.E. Larson, M. Le, D. Jang, R.A. Rhodes, "GaInAs/InP Composite Channel HEMTs for Millimeter Wave Power Applications," in 1993 Proceedings of Int. Conf. on Millimeter and Sub-Millimeter Waves and Applications, pp. 579-580.
- [33] K.C. Hwang, P. Ho, M.Y. Kao, S.T. Fu, J. Liu, P.C. Chao, P.M. Smith, and A.W. Swanson, "W-Band High Power Passivated 0.15 μm InAlAs/InGaAs HEMT Device," in 1994 Proceedings of Int. Conf. on Indium Phosphide and Related Materials, pp. 18-20.
- [34] S.C. Wang, M.Y. Kao, S.M.J. Liu, P. Ho, and K.G. Duh, "High Performance W-band Pseudomorphic InAlAs/InGaAs/InP Power HEMTs," in 1994 Device Research Conference Digest.
- [35] L.E. Larson, M. Matloubian, J.J. Brown, A.S. Brown, R. Rhodes, D. Crampton, and M. Thompson, "AlInAs/GaInAs on InP HEMTs for low power supply voltage operation of high power-added efficiency microwave amplifiers," IEE Electronics Lett., Vol. 29, No. 15, pp. 1324-1326, July 1993.

- [36] S.R. Bahl, and J.A. del Alamo, "Elimination of mesa-sidewall gate leakage in InAlAs/InGaAs heterostructures by selective sidewall recessing," *IEEE Electron Device Lett.*, Vol. 13, No. 4, pp. 195-197, Apr. 1992.
- [37] M. Case, S. Rosenbaum, C. Ngo, M. Matloubian, A. Brown, J. Henige, M. Thompson, and L. Larson, "High efficiency, high gain K- and Ku-band InP-based HEMT MMIC amplifiers for array applications," in *1994 MTT-S European Topical Congress*, pp. 49-54.
- [38] A. Kurdoghlian, W. Lam, C Chou, L. Jelloian, A. Igawa, M. Matloubian, L. Larson, A. Brown, M. Thompson, and C. Ngo, "High-efficiency InP-based HEMT MMIC power amplifier," in *1993 IEEE GaAs IC Symposium Digest*, pp. 375-377.
- [39] W. Lam, M. Matloubian, A. Igawa, C. Chou, A. Kurdoghlian, C. Ngo, L. Jelloian, A. Brown, M. Thompson, and L. Larson, "44-GHz high-efficiency InP-HEMT MMIC power amplifier," *IEEE Microwave and Guided Wave Lett.*, Vol 4, No. 8, pp. 277-278, Aug. 1994.

SPACE APPLICATIONS OF P-HEMT DEVICES

GIULIANO GATTI

European Space Research and Technology Centre

European Space Agency

P.O.Box 299, 2200 AG, Noordwijk, The Netherlands

1. Introduction

Pseudomorphic High Electron Mobility Transistors (P-HEMTs) are relatively new devices for space applications. The advantages and disadvantages of their application will be critically discussed. For this purpose the paper will first describe the most important drivers for the development of microwave equipments for space applications. Then a short description will be given of the main type of payloads used in satellites, and review the basic link equations and the consequent trade-off factors. This will then lead to the definition of the major constraints on applications of new technologies, such as P-HEMTs. Next the main characteristic of P-HEMT that make them interesting in spacecraft equipments will be examined and some examples given. A review of technologies that compete with P-HEMTs in space applications will also be given for comparison. As P-HEMTs can be used in the form of discrete devices or integrated technologies (Monolithic Microwave Integrated Circuits, MMICs), this will be discussed giving the rationale for a choice. After that the basic concepts of reliability for space hardware will be presented, and the relevant consequences on the application of P-HEMTs will be described. After a short reference to ground segment applications, the paper will conclude with some prospective views for P-HEMTs, indicating the areas where improvements shall be sought in the future.

2. Main drivers for microwave equipments for space applications

Microwave equipments for space applications are very expensive: this is caused by the fact that space is a very unfriendly environment (vacuum, radiations, temperatures). Equipments shall be designed and manufactured to survive in this environment providing excellent performance as resources on-board (mass and power) are very limited. On top of that, spacecrafts shall be launched and this implies that the equipments shall also sustain without damage high levels of vibrations and shocks. The above constraints put a high price tag on any piece of hardware that shall go in space, including all sorts of components used. Typically a microwave component for space applications costs from ten to hundred times more than for commercial applications, depending on quantities and specifications.

One important contribution to the high cost of space applications is the launch cost, typically between 50 to 70 million US dollars. This makes the cost of launching 1 Kg of hardware into orbit to be of the order of 30000 dollars, in average. So it is not surprising that one of the major drivers for the development of equipments for space applications is mass reduction, if we also consider that, in any case, launch capability is limited!

But making an equipment less heavy is not enough. If we exclude passive equipments, the other types of equipments consume DC (direct current) power. This power is produced on board by solar panels and by batteries (needed when the spacecraft does not see the Sun during the eclipses). The bigger the solar panels and the batteries, the higher the fuel required for keeping the satellite in its orbit (station keeping). Hence the power consumption can be considered equivalent to additional mass in the satellite. For a geostationary satellite this equivalence is of the order of 200 g per watt of power consumption. So, in practise, an equipment of 1 Kg of mass that consumes 10 W is actually a 3 Kg equipment (90000 \$ launch cost)! Therefore space equipments shall be designed also to minimize their power consumption.

After mass and power consumption reduction another important point for space equipments is maximization of reliability. In the majority of cases, once launched, space hardware can not be repaired. Therefore all efforts and precautions shall be taken during the design, the selection of components, the assembly and integration. This is to make sure that the equipment will be perfectly functioning during the mission life time that, presently, for geostationary telecommunication satellites, is required to be between 10 to 15 years.

Hence, the applicability of a new technology in space is also strictly dependent on its reliability in the space environment, and, very often, the applicability domain is much more restricted (e.g. in terms of biasing voltage range, level of gain compression, temperature range, etc.) than in military, professional or commercial applications.

In summary, before a new technology is used the advantages and the risks shall always be properly traded-off carefully. Eventually, the decision of using it, and the extent of use, will be related to advantages in at least one of the key drivers (mass reduction, power consumption reduction, reliability increase) without, or with minimum, disadvantages.

In addition, during the development phase, equipments for on-board space applications go through various phases (breadboard model, BM, engineering model, EM, engineering qualification model, EQM, proto-flight model, PFM, and, at the end, flight model, FM). This development phase can last from a minimum of two-three years to even five, six years, depending on the program. Additionally it is quite typical to see orders for these equipments for a first spacecraft, and some years after for another spacecraft. This long development phase and the need to maintain over a certain number of years the production capabilities, put strong requirements on stability and consistency on the technologies used. For example, microwave components procured for the FMs shall have very similar performance to those used in the BM. This means that well established technologies are preferred as they are very stable and can produce devices with reproducible performance. For relatively new technologies any evolution shall be avoided during the equipment manufacturing life, or any improvement shall not affect the circuit performance, therefore impacting on schedule and cost.

3. On board microwave payloads

In this paragraph a very simple summary will be given of the most common on-board microwave payloads, highlighting the basic relationships between various system performance.

Microwave equipments are present in all satellites that belong to the following main categories:

- telecommunication
- remote sensing of the Earth
- scientific application.

The number of microwave equipment used is very high in telecommunication satellite payloads, and in microwave instruments of remote sensing satellites. For scientific satellites the microwave equipments are, very often, only present in the telemetry, telecommand and control (TT&C) sub-system.

3.1 TELECOMMUNICATION SATELLITES

Telecommunications satellites can contain various type of payloads:

- fixed satellite system (FSS) payloads
- direct broadcasting system (DBS) payloads
- mobile satellite system (MSS) payloads
- data relay satellite system (DRS) payloads
- navigation payloads
- search & rescue payloads.

FSS payloads are used to relay the signals between ground stations located in fixed positions, MSS satellites between mobile ground stations (e.g. ships, airplanes, cars, trucks or individual persons) and DRS payloads between another satellite and a control centre on the ground. DBS payloads typically broadcast television channels that can be received on the ground by millions of very small receivers. Navigation payloads transmits signals that, when received by a ground user, allow to determine its position accurately. Search and rescue payloads pick up signals from beacons of distressed airplanes, ships, etc. and relay them to control centres.

Independently from the satellite purpose, generally on a telecommunication satellite we have payloads similar to the one sketched in figure 1. In this payload the receiver (RX) antenna receives the signal transmitted by ground (or by another satellite), often called the up-link signal. The signal is amplified by a low noise amplifier (LNA) and then converted to another frequency by a down converted (D/C), using the signal generated by a local oscillator (LO). Normally the signal received by ground is in-fact made up of several carriers, at slightly different frequencies. These carriers are split using a set of narrow band filters (input-multiplexer, IMUX) and separately amplified first by a channel amplifier (CAMP) and then by a power amplifier (PA). Eventually

all the amplifier carriers are combined using another set of narrow band filters (output multiplexer, OMUX) and transmitted to ground (or to another satellite) by a transmit (TX) antenna. This other link is normally called the down-link.

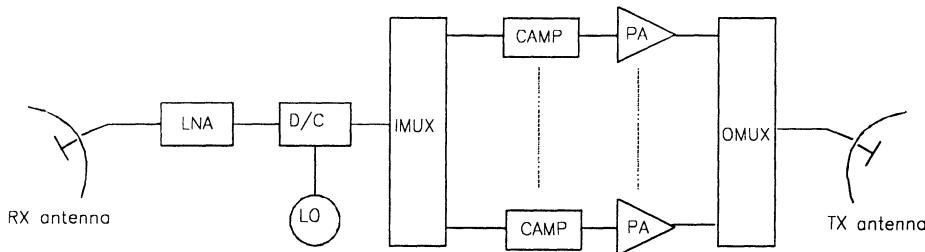


Figure 1. Typical telecommunication payload

This simple payload in practise can be much more complex: for example an intermediate frequency can be used after the D/C and the signal has to be up-converted before being transmitted. Or the antenna configuration is more complex (e.g. phased array, see paragraph 6), or the signals are demodulated and the re-modulated on-board (regenerative transponder) using digital circuits. It is beyond the scope of this lecture to go into details of these various type of payloads and the reader should consult specific references (e.g. [1] to [3]).

The link between a transmitter and a receiver can be related by this simple equation:

$$\left(\frac{C}{N}\right)_{dB} = 10\log P_r G_r - 20\log \frac{4\pi R}{\lambda} + 10\log \frac{G_t}{T} - 10\log L_A - 10\log k - 10\log B_{IR} \quad (1)$$

C	received carrier power
N	received noise power
P _t	transmitted power
G _t	transmit antenna gain
R	range
λ	wave-length
G _r	receive antenna gain
T	total noise temperature at receiver input ($^{\circ}$ K)
L _A	additional losses (e.g. pointing, polarization, atmospheric losses), >1
k	Boltzmann constant ($1.38 \cdot 10^{-23} \text{ J K}^{-1}$)

B_{IF}	intermediate frequency bandwidth
$P_T G_T$	EIRP (Effective Isotropically Radiated Power)
G_R/T	very often indicated as G/T of receiving station (but in dB/K)

This equation gives the ratio between the carrier power and the noise power, which determines the quality of the link. And it introduces the inter-dependence of the various variables. For example the noise temperature of the receiving system referenced at receiver input (T) is strictly related to the transmitted power (P_T). If the noise temperature is higher then more power has to be transmitted to get the same link quality (same C/N).

It is useful to visualize on a graph this dependency, as in figure 2. From this figure, for example, if the system noise figure from 2 dB (2 dB on the y-axis) becomes 0.2 dB higher (0.2 dB on the x-axis) the overall link budget is degraded of 0.5 dB. This can be recovered only by transmitting 0.5 dB higher power.

Effect of NF degradation on link budget

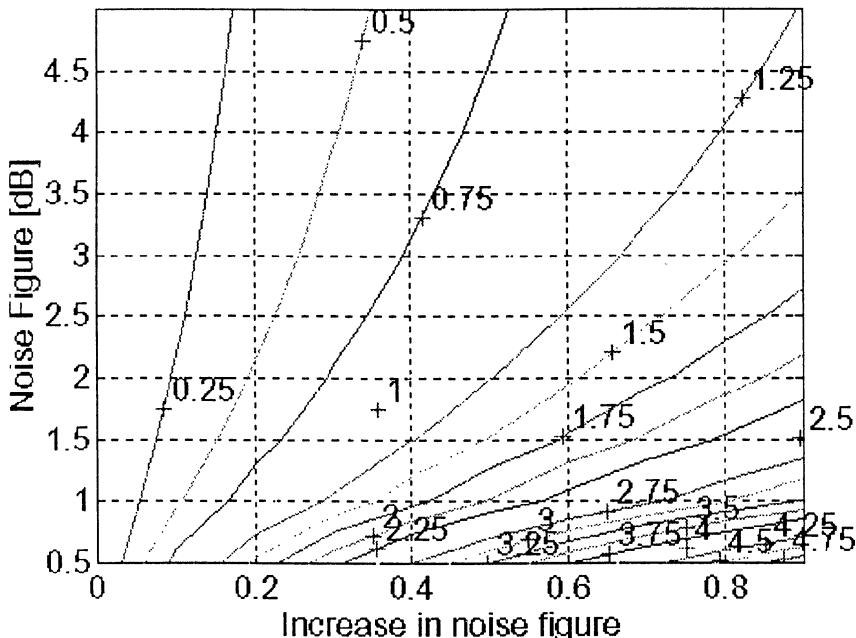


Figure 2

In the previous equations the noise temperature is linked to the noise figure, more commonly used in equipment design, by the following equation:

$$(NF)_{dB} = 10 \log\left(1 + \frac{(T)_K}{290}\right) \quad (2)$$

NF	noise figure receiver
T	receiver noise temperature

From the above considerations it is clear that it would be useful to decrease as much as possible the system noise temperature to reduce the transmitted power. This does not completely correspond to reduce the noise temperature of the receiver. In fact the system noise temperature is related to the receiver noise temperature by the following equation:

$$T = \frac{T_A}{L} + T_L \left(1 - \frac{1}{L}\right) + T_R \quad (3)$$

T	total noise temperature at receiver input
T_A	antenna noise temperature
T_L	physical temperature of lossy elements between antenna and receiver
L	loss between antenna and receiver, >1
T_R	receiver noise temperature

Also in this case it is useful to see this relation graphically in figure 3.

Looking at figure 3 if the noise figure is of 1 dB and the antenna temperature is 300 °K the total system temperature, referenced at the receiver input, is of about 3.6 dB. Whereas if the antenna temperature is 30 °K the total temperature is about 1.7 dB. The first example is very much similar to the common situation of a receiving antenna on the satellite where the antenna is looking to the Earth that is at an average temperature of the order of 300 °K. The second example is more applicable to a ground station where the antenna gives a noise contribution of the order of less than 50 °K for elevations higher than 20-30°. From figure 3 it is evident that depending on the situation (ground station or on-board receiving antenna) there is a different amount of interest in reducing the noise figure of a receiver below a certain value. This also depends on frequency and antenna design.

3.2 REMOTE SENSING SATELLITES

These satellites include various type of microwave instruments (the equivalent of "payloads" in telecommunication satellites) that are used to detect specific characteristics of the ground or of the atmosphere. Some of the instruments are simply passive receivers (for example radiometers) that measure the emission from ground sources (different terrains, gases, etc.). Others are active instruments that operate as radars (for example synthetic aperture radars, SAR, radar-altimeters, cloud radars, rain radars, etc.). The typical simplified block diagram of the front-end of a radar instrument is shown in figure 4.

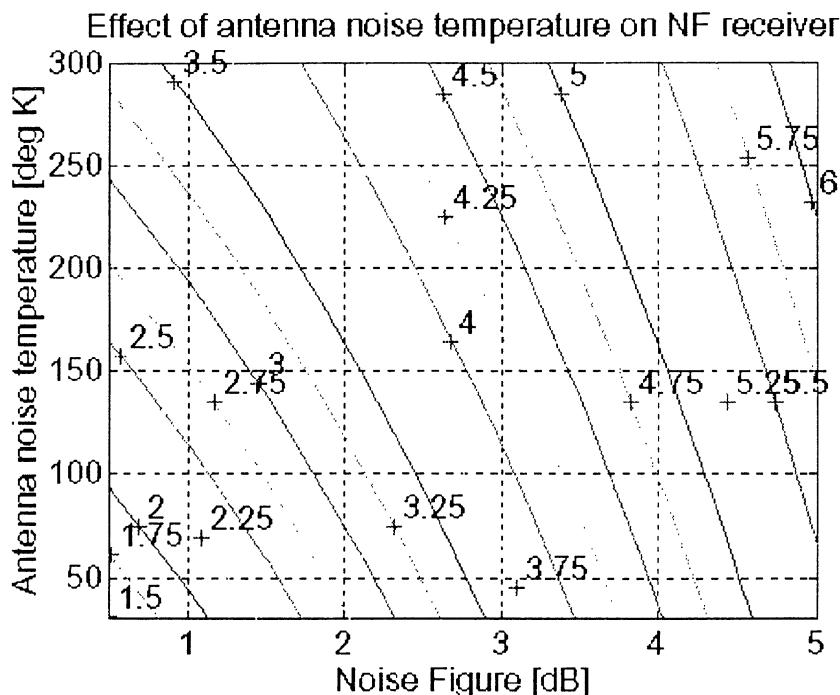


Figure 3: assuming $L=1$ dB and $T_L=290^\circ\text{K}$

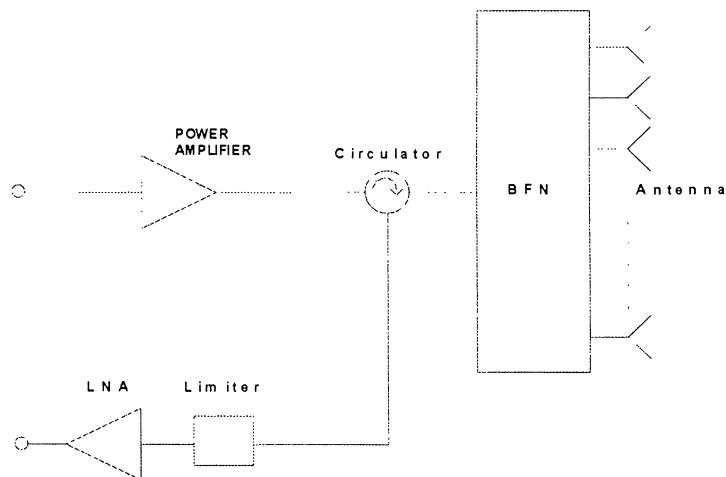


Figure 4: Simplified radar front-end block diagram

The power amplifier generates a high power pulsed signal that is directed, through a circulator, to an antenna. This antenna is quite often an array of elements, driven by a beam forming network (BFN). The signal transmitted by the antenna reaches a target (e.g. the surface of the Earth) and is reflected back. The reflected signal depends on the reflectivity and on the dimensions of the target. The reflected signal is then received back from the antenna, routed by the circulator to a low noise amplifier (LNA) and then processed. A limiter is often inserted in front of the LNA to protect it against the leakage of the very high power during the transmission pulses.

Even though the mode of operation of radars differs quite significantly compared to a microwave telecommunication payloads the main system performance can be discussed in a very similar way, at least at link-budget level, as shown in equation 4. This very generalized equation includes terms exactly identical to those of equation 1. The basic differences are the fact that the target illuminated by the transmitted radar pulse reflect a certain part of this signal back to the radar antenna (proportionally to its radar cross section). Secondly the range R is squared in equation 4 to take in account the round trip of the signal (from antenna to ground and back to the antenna) and that the same antenna is used in reception and transmission. Additionally an equivalent bandwidth is included to generalize the equation avoiding too specific discussions, for example on integration time, for which the reader is again recommended to refer to specialized text-books (e.g. [4] to [9]).

$$\left(\frac{C}{N}\right)_{dB} = 10 \log P_r G_{TR} - 20 \log \frac{4\pi R^2}{\lambda} + 10 \log \frac{G_{TR}}{T} - 10 \log L_A - 10 \log k - 10 \log B_{eq} + 10 \log \frac{\sigma}{4\pi} \quad (4)$$

σ	target radar cross-section
G_{TR}	transmit/receive antenna gain
B_{eq}	equivalent bandwidth

Looking to equation 4, similar relationships can be drawn as for equation 1, highlighting also for radars the close link between the various system parameters, including noise temperature and transmitted power.

The importance of system temperature is also fundamental for radiometers, whose typical block diagram is shown in figure 5. In this case the higher the system temperature the worse is the radiometric sensitivity or the longer shall be the integration time. This relationship is given by the following equation:

$$\Delta T = T_{sys} \sqrt{\frac{1}{B\tau} + \left(\frac{\Delta G_s}{G_s}\right)^2} \quad (5)$$

ΔT	radiometric resolution
T_{sys}	system noise temperature
B	bandwidth (IF)
τ	integration time
G_s	system gain
ΔG_s	rms system gain variation

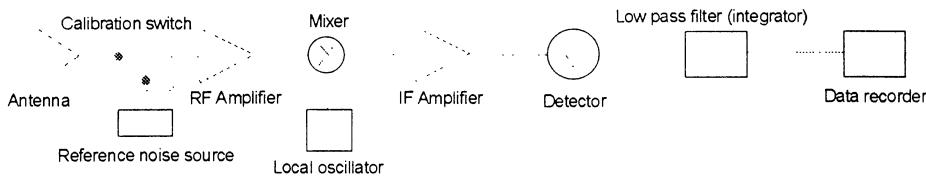


Figure 5: Total power radiometer block diagram

4. P-HEMTS, Considerations on their Applicability in Space Equipments

After having reviewed the basic equations and considerations that concern the most important microwave applications on-board satellites, let us concentrate in trying to better define the role that P-HEMTs can have in developing microwave equipments for space applications. Few details will be given in this section on P-HEMT devices as these are very well described in the other sections of this NATO/ASI volume and only the application point of view will be dealt with.

Main advantages of P-HEMTS are certainly:

- high cut-off frequency (F_T)
- higher gain per stage
- better efficiency
- higher power density per mm of gate periphery
- break-down voltage similar (potentially) to MESFETs.

The above advantages are combined with reliability figures comparable to those of the well established MESFET (metal semiconductor field effect transistors) technology. And recalling paragraph 2 considerations, this is a fundamental requirement for a technology to be applied in on-board space applications.

Let us examine how these advantages impact on microwave payloads.

A higher cut-off frequency enables the realization of circuits using two port active devices up to very high mm-wave frequencies. At the present state of the art useful gain and noise figure has been demonstrated already above 100 GHz. Millimeter wave applications in space are numerous: from inter-satellite links at 50-60 GHz, to fixed and mobile communications with users on ground at 40-50 GHz, to radiometer applications. These last applications operate on a wide spectrum to observe various characteristic radiations of surfaces, gases, clouds, etc., on ground. Radiometer applications start as low as at VHF (very high frequency) to go as high as at 1 THz! The possibility of having "transistor like" circuits at mm-waves allow to simplify the realization of the above type of instruments, improving performance and reliability, and reducing mass and size.

An higher gain per stage reduces power consumption. If, for example, we want to obtain a gain of about 20 dB at Ku-band most likely we can get it with a two stage amplifier using P-HEMT devices, whereas we should use three devices (wasting also some gain) with conventional HEMTs. If this amplifier is used in an active antenna with, say, 500 elements we can reach a reduction of power consumption of the order of 10 W and a weight saving of 7 Kg (10 g less per unit plus the equivalent mass reduction of 10 W)! Additionally less components are used, improving reliability, and the assembly is less time-consuming, reducing cost.

Moreover a higher gain per stage allows to reduce the noise-figure of a multi-stage amplifier. In fact in this amplifier the total noise figure is given by the equation:

$$NF_{amp} = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \dots \quad (6)$$

NF_{amp} total noise figure of a three stage amplifier (in number)
 NF_x noise figure of stage x (in number)
 G_x gain of stage x (in number)

For example let us use the characteristics of a P-HEMT and of a HEMT device that are space qualified. These are the type CFY-66 and CFY-67 from Siemens (Germany), qualified by ESA under the SCC (space component coordination group) detailed specification 5613/002 and 5613/004.

Both devices have the same dimension (4 fingers, 45 micron each) but the CFY-66 is a standard AlGaAs/GaAs HEMT whereas the CFY-67 is a AlGaAs/AlInAs P-HEMT. At 18 GHz they have approximately the same noise figure of less than 1 dB. However the associated gain is 9 dB for the HEMT and 10 dB for the P-HEMT. Using the previous equation, and neglecting other contributions (circuit losses, bandwidth, etc.) a two stage amplifier followed by a mixer with 8 dB noise figure, would have a noise figure of 1.3 dB using two P-HEMTs and of 1.5 dB using two HEMTs. Using figure 2 this difference corresponds to a 0.75 dB channel degradation!

The higher gain of P-HEMTs eases the utilization of these devices in more non linear class of operations (e.g. class B) where the gain is normally degraded compared to class A operation.

Better efficiency of P-HEMTs allow to realize solid state power amplifiers (SSPAs) with reduced power consumption and reduced heat dissipation. This efficiency is normally intended as power added efficiency (PAE) and takes into account not only the intrinsic capability of a device to produce RF (radio frequency) output power using limited DC power, but also the gain of the device. In this respect one of the reasons for P-HEMT to provide better efficiency is their capability to give higher gain than other devices. Power added efficiency is defined using the following equation:

$$PAE = \frac{P_{output} - P_{input}}{P_{DC}} = \frac{P_{input}}{P_{DC}}(G-1) \quad (7)$$

PAE	power added efficiency
G	gain power amplifier
P _{input}	input power
P _{DC}	DC power

Taking the previous example of the active antenna with 500 elements, each transmitting 1 W, a power P-HEMT at Ku-band would give an advantage, over a conventional MESFET, of about 5 to 10 % in PAE, achieving PAEs of the order of 45-50 %. This would reduce the power consumption of the transmitter by more than 130 W. Additionally the devices would run cooler with better gain, output power and improved reliability.

Combined with the high efficiency, the higher power obtainable per mm of gate periphery from a single P-HEMT device, and at higher frequencies, allows to simplify the design of SSPAs, reducing the number of devices to be combined to obtained a required output power. This reduces the mass and size of the amplifiers and the manufacturing time and cost, and enables an even higher efficiency, removing undesired combination losses.

P-HEMTs can provide break-down voltages similar to those of MESFETs (15-20 V), at least potentially (published data are still scarce in this respect). This is also very important for space applications, and an even higher voltage would be desirable. First of all the amplifiers using these P-HEMTs could be biased with higher voltage and less current. This would be very beneficial as power supplies and DC distribution harness can provide the DC power required by the SSPA with a higher efficiency (as the losses in DC circuits are proportional to the square of the current, multiplied by the circuit resistance). Secondly the operation of the P-HEMT could be considered in more non-linear amplification classes (such as class B), yielding even better power added efficiency, without affecting reliability. Additionally a higher break-down voltage would allow a higher ratio between pulsed power and continuous wave (CW) power, easing and improving the design of radar instruments.

Do P-HEMT have any disadvantage? Until now it seems that the only real disadvantage is their performance in terms of phase noise: in fact they typically give a phase-noise that is worse than MESFET, and much worse than HBTs (heterojunction bipolar transistors). For example in [10] the residual phase noise of an amplifier has been measured showing at 10 KHz from the carrier -160 dBc/Hz for AlGaAs/GaAs HBTs, -155 dec/Hz for a GaAs MESFET and -150 dec/Hz for a AlGaAs/InGaAs P-HEMT amplifier, for a carrier frequency around 2 GHz. Therefore for oscillators at lower microwaves frequencies (below K-band) better performance can be obtained using HBTs, MESFETs or even Si-bipolar transistors (now available with cutoff frequencies up to 20 GHz). However for higher frequencies P-HEMT can offer the capability to have signal generation without the need of multipliers (that would degrade phase noise inherently). Possibly, even though it has still to be verified experimentally, better phase-noise can be obtained by using medium-power P-HEMTs as low power oscillator devices. This has been proven to be successful already with MESFETs and it should be successful for P-HEMTs as well. Intuitively a bigger device is more linear than a small one and, therefore, it up-converts less efficiently noise at very low frequencies (0 to 100 KHz) to a carrier at microwave frequencies.

Another disadvantage is the still limited maturity of power P-HEMT devices: manufacturers in all the world are still trying to optimize the device structures and semiconductor recipes. Even though excellent results have been presented there is not any commercial device available that could be considered for an evaluation program for verifying its applicability to space.

Paradoxically, one of the advantages of P-HEMTs, the increased gain, could constitute a problem when considering stability issues: for example a device with 10 dB gain at Ku-band could give more than 20 dB at C-band and as high as 40 dB at VHF frequencies! Therefore the use of P-HEMTs could be problematic if proper, and complex, stabilization techniques are not adequately used. In some cases you can end up losing all the gain and noise figure advantages of P-HEMTs due to the introduction of lossy stabilization networks, and you would be better off using HEMTs or even conventional MESFETs.

5. Integrated versus Discrete P-HEMT Technologies

P-HEMT technologies are today available in both discrete devices and MMICs. The situation in Europe (including other GaAs technologies, beside P-HEMTs) is summarized in Table 1 (from [11]).

When discussing new space applications it therefore important to assess the advantages of using one or another approach. It is clear that a MMIC implementation is very advantageous for mm-wave applications (due to the removal of parasitic reactances), and also for other applications (in particular those based on active phased-arrays) for improved reliability (at least for low signal applications, due to the reduction of the number of assembly steps), for better repeatability of performance (at least among devices coming from the same wafer), and for cost reduction (even though this point could be less important for space applications than it is for the commercial field, where high volume

production could be the driving force for MMIC use). However GaAs substrates are more lossy than, for example, Alumina substrates. For example at 12 GHz, a $\lambda/4$, 50Ω , microstrip line would loose about 0.1 dB more on GaAs than the same line on Alumina. This directly affects noise figure (in a low noise amplifier) and efficiency (in a power amplifier). Recalling the equation that gives the degradation of noise figure of an amplifier when a lossy element is put in front of it:

$$NF'_{dB} = 10 \log L + NF_{dB} \quad (8)$$

NF'	resulting noise figure
L	loss at the input, >1
NF	initial noise figure

it is clear that any losses in front of a P-HEMT device degrades noise figure of the same amount. Therefore when the application (e.g. front-end LNAs) requires the last tenth of dB the use of a discrete device could be more effective than a MMIC. On the other side MMICs could be used after one or two stages of discrete amplifiers.

Similar consideration applies to power amplifiers. If we express the resulting power added efficiency of a device when losses are included at its output, we get the following relationship:

$$PAE' = \frac{PAE}{L_{out}} + \left(\frac{1}{L_{out}} - 1 \right) \frac{P_{input}}{P_{dc}} \approx \frac{PAE}{L_{out}} \quad (9)$$

PAE'	resulting efficiency
PAE	initial efficiency
L_{out}	loss output amplifier, >1
P_{input}	input power
P_{dc}	DC power

where the approximation is normally valid for today available microwave power devices. This equation can be shown on the graph in figure 5. Looking at figure 5 an amplifier with 45% power added efficiency would have a total efficiency of only 35-36% if 1 dB additional losses are inserted at the output (this gives a good rule of thumb of 1 % loss in efficiency for any 0.1 dB additional loss at the output). Therefore also in the case of power amplifiers the use of discrete P-HEMTs would maximize performance. As in the case of low noise applications, the use of MMICs would be acceptable in the lower signal part of the power amplifier chain (e.g. up to, and including, the pre-driver amplifier).

Table 1: GaAs MMIC Processes in Europe

COMPANY	LN/GP MESFET	LN HEMT	LN P-HEMT	Power MESFET	Power P-HEMT	HBT	Diode process	E/D-Digital process	Foundry Service	Discrete MESFETs or HEMTs	Discrete low noise P-HEMT Devices
ALCATEL TELETTRA	I2P	---	Future	I2P I3P	Future	---	---	---	Restricted	Yes	---
ALENIA	Yes	---	Yes	Yes	Future	Future	---	---	Restricted	Yes	---
DAIMLER BENZ	E05 F025	H1025	---	---	H2025	Future	B025	---	Yes	Yes	Yes
IAF	---	Yes	Future	---	---	---	---	Yes	Restricted	---	---
GMMT	F20	---	F40	F20	---	Future	---	---	Yes	Yes	---
IMEC	---	Yes (InP)	---	Yes	---	---	---	---	---	---	---
PML	D07A/M D05A/L	D05AH	D02AH	D07M	Future	---	---	ER07AD ED02AH	Yes	Yes	---
SIEMENS	DIOM 15/20 LN	HEMT 40	HEMT 50	DIOM 15 SWMP DIOM 15/20 HP DIOM 15 HPM	Future	Future	---	---	Restricted	Yes	Yes
TCS	LN05	VLN02	Future MM015	HP07 HP05	Future MM015	Future	---	H-GaAsII	Yes	Yes	Yes

LN: Low Noise

GP: General Purpose

E/D: Enhancement/Depletion

To partially reduce the problem of insertion losses at the output and to maintain, at least partially, the advantages of MMIC technologies some developers have utilized a MMIC power amplifier where the output matching networks are not implemented on GaAs but on an external Alumina substrates.

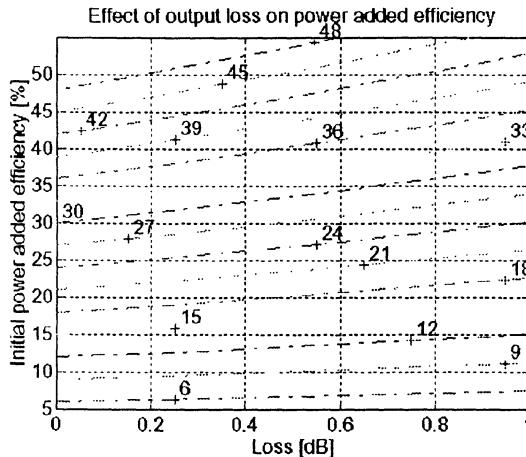


Figure 6

There is another very good reason to use discrete devices for the output stages of SSPAs and it is thermal. GaAs does not have a good thermal conductivity (46 W/m°C compared to 118 W/m°C of Si, 156 W/m°C for Aluminum and 2200 W/m°C for Diamond). Therefore it is very important to have the GaAs power devices built on very thin substrates. Values of 30 to 50 μm are feasible for discrete devices. However 100 μm is typically the minimum for power MMICs if reasonable manufacturing yields have to be maintained. This means that power P-HEMTs on MMICs typically operate at a higher temperature than the same (in terms of gate periphery) device in discrete form. This means worse performance (e.g. lower output power, gain and efficiency) and reliability (that is directly related to the operating temperature).

The thermal characteristics of MMICs could be improved by, for example, using local reduction of the GaAs substrate thickness only beneath the active devices and plating-in gold to implement a sort of heat-sink. As P-HEMT devices have better efficiency, this helps in reducing this thermal problem as the dissipated power in the device, generating the heat, is inversely proportional to the device efficiency (for the same output power):

$$P_{diss} = (P_{DC} + P_{input}) - P_{output} = P_{DC}(1 - PAE) \quad (10)$$

P_{diss} dissipated power in the device (heat)
 P_{DC} DC power consumed by the device

P_{input}	input power
P_{output}	output power
PAE	power added efficiency

One additional point to be considered when comparing integrated versus non-integrated technologies is that typically, when implementing MMIC devices, to guarantee a reasonable overall yield designs are normally wider-band than what is actually required for space applications (where typically the wider bandwidth used is less than 10-15% of the centre frequency). This has the consequence that all performance (including noise figure and power added efficiency) are degraded compared to a narrow-band design that is normally possible (because of tuning capability) using discrete components.

6. Competing Technologies for Space Applications

Having discussed the applicability of P-HEMTs it is important to briefly discuss competing technologies in space applications.

Talking about low noise applications, it is clear that already now P-HEMTs are used for low noise amplifiers from S-band to mm-wave frequencies, due to the better noise figure and gain achievable. However their applicability to the lower end of the microwave spectrum, and below, is questionable because of the stability problems mentioned before, and because reasonably good performance is obtainable from Si bipolar transistors (for lower frequency application), MESFETs and conventional HEMTs.

Competition to P-HEMT technologies comes also in space applications such as radiometers, at high mm-wave frequencies and sub-mm wave frequencies. Here the present applications utilize either planar Schottky diodes (60-100 GHz) or whisker-contact Schottky diodes in waveguide structures (above 100 GHz), both used in direct conversion receiver configurations (without RF amplifier).

For solid state power amplifiers space applications are still dominated (almost 100 %) by MESFETs. These components are well known for many years and a significant amount of life-test results and in-orbit performance parameters are available to demonstrate their reliability. Additionally, MESFET output power capability is very high, at least below X-band, with reasonable power added efficiency (10-20 W Pout, CW, with 45-50 % PAE is now typical around C-band).

Competition to MESFET at frequencies below X-band does not come from P-HEMTs: Si power bipolar transistors can provide useful performance up to 1.5-2 GHz for CW applications, and up to 3.5-4 GHz for pulsed applications. One of the biggest advantages, in this last case, is that Si bipolar junction transistors (BJTs), because of their very high voltage break-down characteristics, can provide extremely high pulsed power per device (of the order of 60-70 W per device). Additionally they use high collector voltage (35-40 V), and they do not need a pulse modulator to pulse the collector voltage (as FETs do to pulse the drain voltage) as the devices can be reliably operated in Class B or C.

For example in figure 7 the results of a trade-off are shown between Si transistors and MESFETs for a 60 W pulsed power amplifier at 3.3 GHz. From this example it is clear that the solution using MESFETs (and, for extrapolation, using P-HEMTs, except if these could be made with much higher break-down voltage and/or power capability) would be not recommendable.

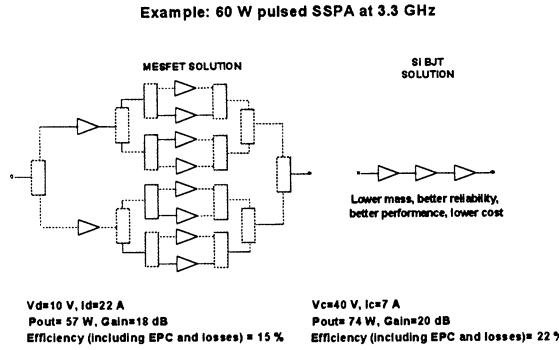


Figure 7

At higher frequencies (above 30-40 GHz) Gunn and Impatt diodes are still used in space applications either for reflection type power amplifiers, or in oscillators.

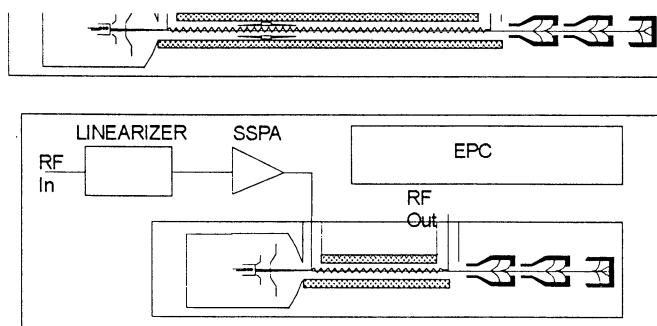
Another technology that often is considered competitive with P-HEMTs is the HBT [12]. However, until now, HBTs have shown competitive electrical capabilities only at frequencies below X-band. Therefore being P-HEMTs more adequate for higher frequency applications we can consider the two technologies to be relatively complementary. It is clear that HBT has many advantages compared to P-HEMT such as higher power density per mm of emitter periphery, better linearity, better phase noise performance and better capabilities in pulse amplifiers. However HBTs have also the big disadvantage, due to their vertical structure, of running quite hot, which significantly affects their reliability and will make their application to space hardware quite limited, at least for power applications, at the present state of the art.

However the technology that represents the strongest competitor to P-HEMTs for microwave power amplifiers (and for all other possible semiconductor based power amplifiers) is the Travelling Wave Tube Technology (TWT). Just to give an idea presently tubes are routinely delivered for telecommunication satellites at Ku-band with output power capabilities between 80 to 110 W and power added efficiency of more than 60 %. They are able to operate with base-plate temperatures up to 100 °C, and over very wide frequency bands. Additionally TWTs have accumulated many millions of hours in orbit, which has demonstrated reliability figures close to those obtainable with SSPAs. Other disadvantages of TWTs, bigger mass and size, worse linearity, have been leveraged out, in most of the cases, by their extremely efficient power capabilities. In the next foreseeable future TWTs will continue to dominate the area of satellite power amplifiers, at least for telecommunication satellites at Ku-band [13].

At lower frequencies (L-S band) SSPAs are still advantageous but in most of the cases the requirements for space applications can be covered by MESFETs or Si bipolar transistors.

One of the most "commercial" space application of SSPAs is in the area of C-band telecommunication transponders. Until now this has been dominated by SSPAs based on MESFETs, but in the future P-HEMTs could be introduced to improve performance, [14]. Also at higher frequencies there are some specific (niche) application areas (e.g. telemetry transmitters at Ku/Ka band, down-link transmitter for mobile satellite feeder links at Ku-band) where SSPAs are used and where better performance using P-HEMT could be interesting for future developments.

However, paradoxically, the evolution of the TWT open new space for semiconductor technologies, and in particular P-HEMTs. In fact to reduce TWT mass and size a new approach has been taken first in the military field and soon will be taken for space applications: the so-called Microwave Power Module (MPM). This module consists of a short (therefore smaller) TWT combined with a MMIC solid state driver (with output power of the order of few hundred milliwatts) and a miniaturized power supply, integrated in a single, compact housing (see a simple block diagram of a MPM in figure 8), [15]. By combining these technologies significant reduction in mass and size are achieved while maintaining very high efficiency, power capability and maximizing integration of functions (before carried out by separate equipments such as channel amplifier, driver amplifier, linearizer, equalizer, etc.). The MPM concept brings also to cost advantages for volume production (e.f. for ground based radar and electronic-counter-measure jammers) but also for space applications as a single unit shall be procured and integrated instead of three or more separate equipments. But for the MPM to be effective the solid state part shall be manufactured with very good performance especially in terms of power added efficiency and wide-band operation: that is where P-HEMT technologies, and most likely P-HEMT MMICs, will be required.



Microwave Power Module (MPM)

Figure 8

In addition there are other type of space applications where P-HEMT will play a dominant role in the future, those based on active phased array. In simple words an active phased array is constituted by an antenna made by many radiating elements (or feeds), each connected to small solid state power amplifiers (for a transmitting array), or to low noise amplifiers (for a receiving array), or to transmit/receive modules (in a radar). All the modules are driven by a beam forming network (BFN) that provides the proper phase and amplitude settings to the antenna radiating elements so that the desired antenna pattern is obtained. A schematic of an active phased-array radar front-end is shown in figure 9. Such payloads are normally required in remote sensing missions where the antenna pattern shall be modified to cover different areas on ground, introducing flexibility and re-configurability in the instrument capabilities. Active phased-arrays are also required in telecommunications missions. In some cases very high antenna gain on-board is required such as the communication link is feasible with very simple and small ground terminals (e.g with very small aperture terminals (VSATs) or mobile terminals). High gain from an antenna, however, automatically implies that only a small area on ground can be covered by the antenna beam (due to the reduction of the beam width). To recover from this situation multi-beams shall be generated at the same time and one way to do so is through the use of active phased-arrays.

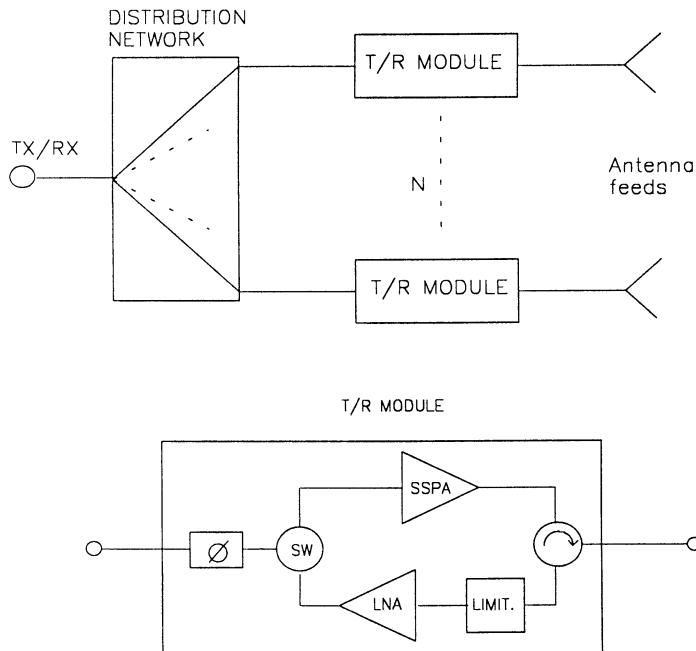


Figure 9: Active (distributed) phased-array

An active phased-array (sometimes called distributed phased-array) is alternative to a conventional phased-array, whose block diagram is shown in figure 10. This approach has some advantages compared to a active phased-array but also many disadvantages. In particular

the centralized solution can use a much more efficient power amplifier (TWT) and uses only one receiver. However it needs all units to be made redundant (to minimize the risk of loss of the instrument functions) and all parts operate at very high power level (with the risk of discharges and multipaction phenomena). High power TWTs and power supplies are very bulky, the losses in the BFN are very critical (and this means that bulky wave-guide components have to be used, and techniques such as amplitude shaping, to reduce side-lobe levels, are not practical) and the limiter in front of the LNA is very critical (due to the high power levels involved).

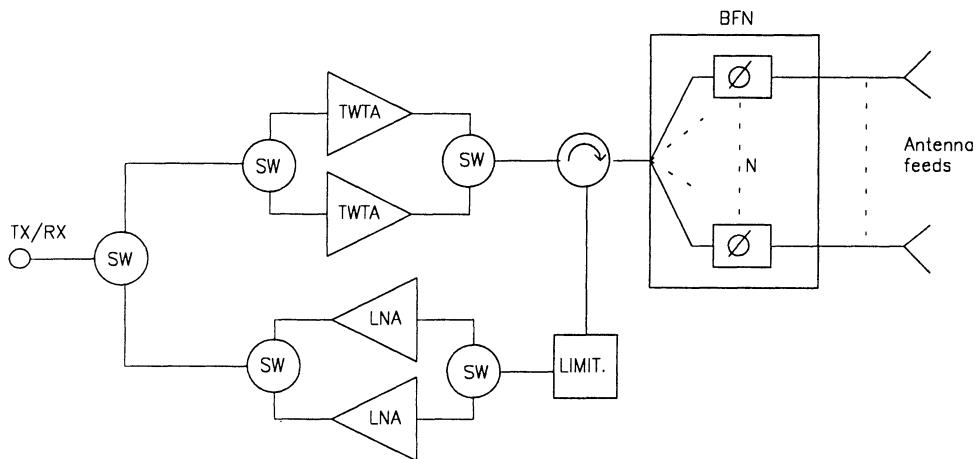


Figure 10: Centralized phased-array

On the opposite side, an active phased-array can be realized using miniaturized technologies in all functions, including the BFN (as its losses in this configuration are not critical any more). It is intrinsically reliable because the failure of few T/R modules will only slightly impact on the antenna pattern, and amplitude shaping is feasible, which provides better lower side-lobes from the antenna. Main disadvantages of an active phased-array are mainly the high complexity, the tracking requirements between modules, the cost and the need for high performing solid state technologies. This is exactly where P-HEMT technologies, and in particular P-HEMT MMICs (for reasons of miniaturization, reliability, cost and reproducibility of performance) will dominate in the future [16].

Another possible applications of P-HEMT technologies would be in the area of extremely fast digital circuits. As the technologies available in this domain are still very limited it is premature to discuss on their applicability to space. When maturity is reached, these technologies could be very useful in space equipments such a digital direct synthesizers (DDS), auto-correlation spectrometers, digital signal processing, etc.

7. Reliability Constraints on P-HEMT Operation

As explained at the beginning reliability is a fundamental requirement for on-board space equipments and, consequently, for any semiconductor technology used. Reliability is an intrinsic characteristic of a technology (and in this respect it has already been mentioned that, until now, P-HEMTs have shown characteristics very similar to MESFETs) but also depending on the way a particular technology is used in a specific application.

As far as space applications the limits in the applicability of components such as P-HEMTs come from the known failure mechanisms of these devices than, until now, have been shown to be similar to those of MESFETs (and that are well described in other sections of this NATO/ASI book). In summary limitations shall be applied to:

- maximum channel operating temperature
- maximum current density in the device metallization layers
- maximum level of overdrive
- stability requirements.

In general space applications requirements set the maximum channel temperature (or junction temperature, in case of bipolar devices) to 110 °C for any semiconductor device. This is normally a very tough requirement especially in power amplifiers, which affects significantly the electrical, mechanical and thermal design and manufacturing of equipments as the equipment shall be operated over extended temperature ranges where the hottest point is typically between + 55 to +75 °C. The situation can be improved clearly by having devices, such as P-HEMTs, with better power added efficiency and by devices that can operate in more efficient classes of operation (such as class AB or even class B).

This last remark is also very important in most space applications where the signal in a power amplifier could vary over a wide dynamic range and when in some cases, the power amplifier is powered, but no signals are present at its input. This represents for a class A amplifier the worse thermal, and therefore reliability, condition as all the DC power is dissipated into the device as heat, and it is not converted to RF power. By operating the device in class AB or, even better, in class B the drain current decreases for lower signal power and, overall, the power dissipated can be reduced. These classes of operation are, however, less linear than class A and therefore their compatibility with the type of signal to be amplified shall also be verified.

The maximum current in the device metallization layers shall be limited (typically below $5 \cdot 10^5$ A/cm² for Al metallization) to avoid long-term degradation known as electro-migration. This phenomena causes degradation of the metallization (and consequent performance degradation) and, eventually, catastrophic failure of the device. Most of the problems are caused by the gate finger metallization as these, for obtaining good microwave performance, are very tiny. A typical limit applied, in this case, is a maximum average gate current below 1 to 2 mA per mm of gate periphery.

Gate current is normally generated in a power device because the input signal drives the device

in gate-drain breakdown and in the gate-source forward conduction (and for very high level of signal even to gate-drain forward conduction), as schematically shown in fig.11.

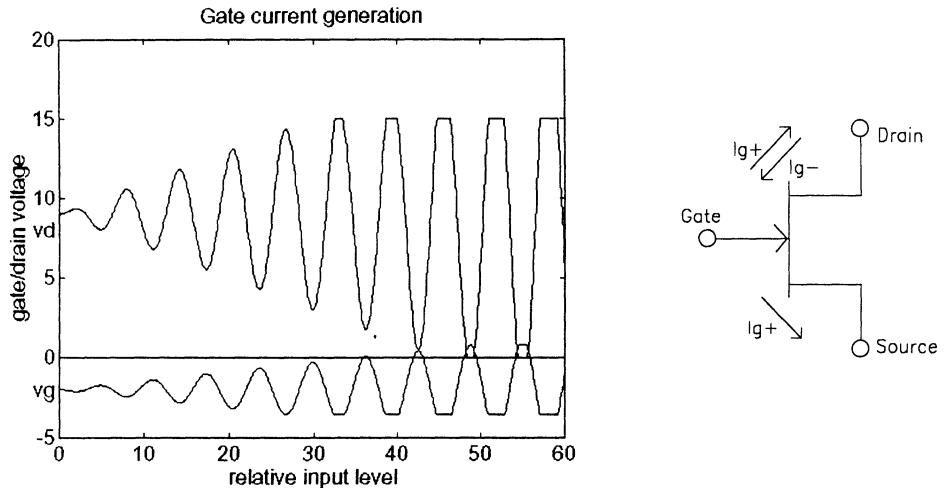


Figure 11

The higher the input signal (drive level) the higher the gate current: therefore the maximum input power shall be limited. This corresponds to a limit in the degree of gain compression at which the device is operated. Limitation of the average (breakdown plus forward conduction currents) gate current can be obtained by inserting a series resistor of a proper value between the voltage source supplying the gate voltage and the gate itself. Care shall be taken to avoid an excessive value of the resistor as this would degrade the power characteristics of the device (power, efficiency, linearity).

The value of the gate-drain break-down gate current (normally called negative gate current) is higher when a device is biased closer to pinch-off (therefore in class AB or class B) due to the higher DC voltage difference between gate and drain. In this respect, therefore, it would be better to operate the device in class A. This is in contradiction with reducing the device channel temperature and, therefore, a compromise shall be achieved.

Apart from generating excessive gate current a high level of input drive is also producing other degradations in the device, caused, for example, by damages in the device surface (e.g. due to impact ionization). These degradations become apparent, for example, as an increased power consumption of the device combined with increased gate current, [17]. On the long term these degradations, therefore, can also accelerate the phenomena described before. Also in this case to avoid such degradations the device shall not be operated above a certain gain compression point.

High input drive levels not only do affect power amplifiers but also devices used in LNAs: for example in radar applications the LNA shall be protected against the transmitter pulse power leaking to the LNA, by means of proper limiters. Additionally, any space equipment shall show robustness when, by mistake (for example during the integration of the equipment into a payload), excessive input power is applied to the equipment itself. Typical specification for space equipments call for no damages caused by input power as high as 6-10 dB above the nominal level. These design constraints are imposed to avoid that very expensive equipments can be damaged easily causing big impacts in schedule and cost of a space program, and often generate the need to insert proper protecting circuits (e.g. variable attenuator controlled by a detector sensing the input power) in the space equipments.

Care shall be also taken when using devices in a oscillator. Also in this case the operation is very non-linear and precautions shall be taken to ensure that the device is always operating in a reliable condition.

Further limitations in space applications come from the fact that, normally, power amplifiers are not operated in the so-called single carrier, CW (continuous wave) condition. This is also the condition where life-tests have been carried out by a certain semiconductor manufacturer to prove the reliable utilization of a device in a space environment. Especially in some telecommunication payloads (for example for communication with many mobile users on ground) the power amplifier has to amplify complex signals made up by, for example, thousand of carriers.

In other cases the amplifier has to amplify a phase modulated signal (e.g. filtered QPSK (quadrature phase shift keying)). In both situations these signals cause the voltage swing on the active device to be very different compared to that of a single carrier signal. In other words the device remains, for example, for more time in a hazardous region such as the avalanche region, and this is evidenced by an higher gate-current and an higher stress level than for a situation where only a single carrier, CW signal is applied, with the same average input power level. Therefore the operation point of a device shall be further limited compared to the life-test conditions to prevent degradations or failures due to complex signal operation, [18].

Also stability is an important reliability requirement for space applications. Microwave equipments for space applications are normally required to be unconditionally stable in all operating conditions. Instabilities could cause unwanted signals, which would oblige, for example, an operator to switch-off a transponder. Moreover, the insurgence of RF oscillations can cause, especially in power amplifiers, a device to fail. Stability has to be implemented starting from the design of each circuits and thoroughly verified at equipment level. In this respect the already mentioned potential drawback of P-HEMTs at lower microwave frequencies shall be watched carefully. Additionally unconditional stability is normally assumed if the stability factor K is greater than 1 for a reasonably wide frequency range. This is not enough as internal loops in circuits can also cause oscillations, and this potential problem is not detected by just looking at the K-factor. One typical example is in balanced structures in MMIC amplifiers where a potential oscillation loop can be present causing the so-called odd-mode oscillations. Typically, as these last type of instabilities cannot be detected by external

measurements, provisions shall be taken at design level to avoid this type of problems.

But how reliable shall an equipment be? The allocation of a certain reliability figure to an equipment is usually coming from the reliability required by the satellite costumer for a certain mission. This is a very stringent number (reliability figures close to 0.99 are normal). From the satellite figure allocations are made at payload level and then at equipments level. And for an equipment the total reliability figure shall be demonstrated by summing up the predicted reliability figures of the various components, starting form the typical life-tests results of each of them. This is done using the MTBF (medium time before failure) defined for N units submitted to the life-test by:

$$\text{MTBF} = \frac{\sum t_i}{N} \quad (11)$$

t_i time to failure of i-unit
 N number of units

From the MTBF the failure rate is defined:

$$\lambda = \frac{1}{\text{MTBF}} \quad (12)$$

λ failure rate

In space applications the failure rate is usually expressed in "fits" (for $\text{MTBF} = 1 * 10^9$ hours, $\lambda = 1 * 10^{-9}$, failure rate defined as 1 fit). For example a 20 W power MESFET has normally 20-100 fits. Then the number of fits for each component are summed up to get the total fit count for an equipment. A 30 W SSPA can have a fit count between 500 to 700 fits, including the power supply. By knowing the fit number and the expected mission life the reliability of the equipment is predicted by the formula:

$$R = e^{-\lambda T} \quad (13)$$

R reliability (probability of mission success)
 T mission time (hrs)

assuming (typically) an exponential distribution for the mission. What about if this computed number is more than the expected allocation for that equipment? Then the equipment has to be made redundant by duplicating, triplicating or even quadruplicating it, so that if one unit fails it can be replaced, by properly operating redundancy switches, by an identical unit. In this way the reliability figure can be improved for the function carried out by that particular equipment,

as shown in figure 11. In this figure if an equipment with reliability equal to 0.5 is made redundant the overall reliability increases to 0.75.

Of course ensuring a high reliability of space equipments goes much beyond a simple fit count analysis. This is implemented by well defined procedures and documents that go in very deep detail into all possible reliability areas including, for example, selection of components, proper operating conditions, worse-case analysis, etc.

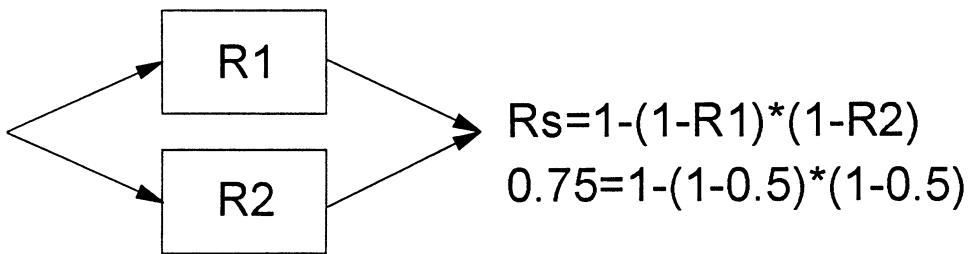


Figure 12: Concept of redundancy

8. Ground Segment Applications

Most of the information given until now are referring to on-board space applications. But of course most of the on-board payloads would be useless without an on-ground counterpart. This is normally called the ground segment and in this paragraph we will briefly discuss the applicability of P-HEMT technologies to this domain. The ground segment includes all ground stations, control stations, network interfaces with ground communications systems, etc. Limiting the discussion to ground stations we can have, for different category of users, for example:

- large, fixed, transmit/receive stations (telephone trunking, control, etc.) at S/C/Ku band
- VSATs at C/Ku/Ka-band
- small, receive only, DBS station (C and Ku-band)
- small, portable or in a car, transmit/receive mobile terminal (L/S band, for future projects for world-wide communication via satellite such as IRIDIUM, GLOBASTAR and INMARSAT P21)
- small, portable, receive only navigation receivers (L-band)
- small, portable or installed in aircraft/ships, search & rescue, beacon transmitters (L-band)
- small, receive only, weather images receivers (VHF or L-band).

Most of the above applications are at L-band where, as indicated before, the use of P-HEMTs will not be so common. The area where P-HEMTs could find a very wide range application is in low noise amplifiers for DBS receivers (outdoor units), especially at Ku-band. Here, however, P-HEMTs shall be very cost effective to be used, considering that a kit (including the microwave building blocks, such as low noise amplifier chip, D/C chip, etc.) for a DBS receiver outdoor unit shall cost around 10-20 \$. This means, for example, that for either discrete P-HEMTs or P-HEMT MMICs plastic packages, for surface mounting, shall be used. Moreover, the price shall be competitive to standard MESFET/HEMT devices with the same functions.

P-HEMTs will find a reasonable use also in VSAT terminals, both in the low noise amplifier and in the transmitter SSPA (that typically requires from 5 to 20 W output power). Also in this case the cost driver will be one of the most important: the price for a complete transmit/receiver unit for a VSAT terminal is targeted to be of the order of less than 5000 \$!

For longer term applications, some preliminary developments are being carried out, especially in Japan, for personal communication systems (PCS) at mm-wave frequencies (40/50 GHz) aiming at having wrist-watch type terminals! If this idea progresses, a new large volume opportunity for P-HEMTs will appear. Here the cost factor will have to be traded-off with the type of P-HEMT technology to be used to provide sufficient performance.

P-HEMTs are also found in equipments for high performing large ground stations, especially in the LNAs (sometimes even cooled to get better noise figure: remember the discussion on contribution of antenna noise in paragraph 3.1). In these large stations the transmitter normally includes a tube amplifier (TWT or Klystron) as power involved can be of the order of hundreds of watts. However these tubes have to be substituted often, with high maintenance cost. One way to solve the problem is to obtain the required output power with a combination of several (twenty to forty) SSPAs. In case of failure of one module the output power is slightly reduced but the station can be kept in operation, and the module substituted without interrupting the service. This new approach has started with ground station at S-band (for TT&C) using MESFETs, and it will progressively used for stations at higher frequencies, where P-HEMTs will be used.

9. Conclusions

P-HEMT technologies have been already used in several space applications as they give advantages that impact positively on the overall system performances. A further expansion of their application will depend on their capability to demonstrate significant advantages in comparison with already well established alternative technologies (MESFETs, standard HEMTs and Si bipolar transistors, in particular).

What would space equipment manufacturers like to see from P-HEMTs in the future? There are several areas where efforts should be concentrated, and these are mainly linked to operation in power amplifiers. One area is related to higher voltage capability. P-HEMTs in the best cases have shown break-down voltages close to those obtained with MESFETs. This

characteristic shall be improved so that better overall efficiency of the system is achieved (due to reduced current requirements for the same output power). Additionally, higher break-down voltages would allow operation in more efficient class (e.g class B) and an effective use of harmonic tuning techniques, allowing further efficiency improvement, while maintaining high reliability.

The second area is related to the availability of larger gate periphery devices: targets could be a power HEMT with 10-20 W output power at S/C-band, and 55-60 % efficiency, 8-10 W at Ku-band with 50-55 %, and 3-6 W at K-band with 45-50% efficiency. Clearly all these devices will have to be manufactured with a stable process that can be submitted to space qualification.

The thermal problem of GaAs based devices would be clearly another area where improvement shall be sought. Of course devices with better efficiency help in this respect. Some more advantages could be obtained by using techniques such as integral heat-sinks, or the use of an higher number of source grounding via holes. In a more futuristic way we could also expect improvements made by growing GaAs structures on diamond substrates.

Another area to be expanded is the realization of low noise P-HEMTs with shorter and shorter gate length (less than 0.1 μm) for better performance at mm and sub-mm frequencies. Here, however, reliability constrains shall be taken very early into consideration, due to the very fine metallization required.

And of course new semiconductor materials, such as InP, will mature, give improved performance, and demonstrate reliable operation.

What next? Using the fantasy one could hope in seeing in the future P-HEMTs manufactured on the material that has all the properties (saturation velocity, break-down voltage and thermal conduction) to be the ideal semiconductor material for future space (and non-space) applications: the diamond!

10. References

1. G.Maral, M.Bosquet, "Satellite communication systems", J.Wiley&Sons, ISBN 0-471-90220-9
2. E.Henakis,(1984), "Satellite communications", Mc Graw-Hill, ISBN 0-07-022594-X
3. R.L.Freeman, (1987), "Radio system design for telecommunications (1-100 GHz)", J.Wiley&Sons, ISBN 0-471-81236-6
4. C.Elachi,(1990), "Spaceborne radars remote sensing: applications", IEEE Press, ISBN 0-87942-241-6
5. D.K.Barton, C.E.Cook, P.Hamilton, (1991), "Radar evaluation handbook", Artech House, ISBN 0-89006-488-1
6. M.I.Skolnik, (1962), "Introduction to radar systems", Mc Graw-Hill, ISBN 0-07-057909-1
7. L.J.Cantafio, (1989), "Space-based radar systems", Artech House, ISBN 0-89006-281-1
8. F.T.Ulaby, R.K.Moore, A.K.Fung, "Microwave remote sensing", vol.I, Artech House, ISBN 0-89006-190-4
9. R.C.Dorf, (1993), "The electrical engineering handbook", CRC press, ISBN 08493-0185-8
10. D.Costa, A.Khatibzadeh, (1994), "A wide-band AlGaAs/GaAs HBT amplifier optimized for low near carrier noise applications up to 18 GHz", 1994 MMT-S Proceedings, paper TH3E-5
11. G.Gatti, (1994), "European GaAs processes", ESA, Preparing for the Future, vol.4, no.1
12. K.Fricke, G.Gatti, H.L.Hartnagel, V.Krozer, J.Würfl, (1992), "Performance capabilities of HBT devices and circuits for satellite communication", IEEE MTT. Trans, vol.40, no.6
13. G.Gatti, D.Perring(1992) "Microwave power amplifiers for space applications", Proceedings of 1994 ISSSE,

Paris

14. J.Shu et alt., (1994), "C-band high performance IMFETs and SUPERIMFETs using MESFET and P-HEMT technology for satcom applications", 1994 MMT-S Proceedings, paper WE1A-2
15. Proceedings of the 1994 Space TWTAs Workshop, ESTEC, May 1994
16. G.Gatti, P.Mancini, (1993), "Critical technologies for advanced spaceborne remote sensing missions", Proceedings of GAAS'93, Milan
17. W.Bösch, F.Garat, (1994), "Impact of multi-carrier RF signals on the reliability of power GaAs-FETs for space applications", Proceedings of GAAS'94, Turin
18. XRM Technical Note, Vol.III, N.5, ESTEC, May 1994
19. R.Coirault, S.Feltham, G.Gatti, M.Guglielmi, D.Perring, (1992), "Overview of microwave components activities at the European Space Agency", IEEE MTT. Trans., vol.40, no.6
20. G.Gatti, R.Coirault, (1994), "Microwave components and technologies: trends and prospective for the future", ESA TD(94)1, section 3.8.4, pp.3.86-3.88
21. Proceedings of IEEE, July 1990, Special issue on satellite communications
22. ANT, (1992), "Millimetre wave local oscillator development", Final report ESTEC contract 9245/90
23. W.Houston, R.D.Rey, (1994), "Military communication satellite development with international cooperation", Proceeding 1994 AIAA Conference, AIAA-94-1129CP
24. H.Krauss, C.Bostion, (1980), "Solid state radio engineering", John Wiley & Sons
25. Chi-Jung Li, "Avalanche breakdown and surface deep-level trap effects in GaAs MESFETs", IEEE Trans. ED-40, no.4, Apr.1993
26. S.H.Wenple, "Control of gate drain avalanche in GaAs MESFETs", IEEE Trans. ED-27, no.6, Jun.1980
27. S.Toyoda, "High efficiency single and push-pull power amplifiers", Proceedings IEEE 1993 MTT-S, paper OF-1-17
28. ANT, (1990), "Study and development of a Ku-band linearizer", Final report ESTEC contract 7338/87
29. MBB, (1990), "L-band SSPA linearization bread-boarding", Final report ESTEC contract 8094/88
30. Workshop on microwave HBTs and HEMTs, Circuits and Reliability, Atlanta, June 1993

EUROPEAN APPLICATIONS OF PSEUDOMORPHIC HEMTS: *Military and Commercial.*

Jacques Favre,
Thomson-CSF Semiconducteurs Spécifiques,
R.D. 128, B.P. 46,
91401 Orsay, France

1. Introduction

Since the mid 1980s, pseudomorphic HEMT technology has been developed world wide and has been exploited in many demonstrators, driven mainly by military requirements. Consequently, the process is now quite mature and several producers claim to offer foundry services. Nevertheless, behind the label *pseudomorphic*, various trade-offs are hidden, such as cost versus performances or large power versus high current gain cut-off frequency. Depending on the major topics addressed (X-band power generation, W-band very low noise receiver, high-speed low power digital circuits, etc.), these foundries have focused on different device topologies. This has led to an initial sharing of the market, which is now expanding dramatically in the area of civil applications.

2. Overview of the Gallium Arsenide Industry in Europe

For any application, the most advanced technology is always available as an option. However, the incapability of the older processes to accommodate the foreseen system objectives always drives the user to trade off his initial need with the capability of the available devices.

Table 1 defines the major categories of devices that can call for pseudomorphic HEMT technology.

TABLE 1. Major categories of opportunities for pseudomorphic HEMT technology

	Requirements	
	Commercial	Military
High-speed, low-power-consumption digital circuits	xxx	x
Low-bias power generation (cellular phones)	xxx	
X-band large power generation	x	x
Ka- and Q-band transmitter/receiver	xx	x
V- and W-band transmitter/receiver	xxx	x

Note: Xs indicate relative levels of foreseen quantities.

To cover all these different applications, at least two different processes are necessary. One with a very narrow recess to offer very high cut-off frequencies for mm-wave use and another with a large breakdown voltage to provide a large output power. This point is discussed in more detail below.

Seven gallium arsenide foundries are located in Europe:

- Alenia (Italy)
- Alcatel Telettra (Italy)
- Daimler-Benz (Germany)
- GEC Marconi (Great Britain)
- Philips Microwave (France)
- Siemens (Germany)
- Thomson-CSF (France).

Some have fabrication lines capable of large production. All have shown a deep interest in pseudomorphic HEMTs. But, depending on their internal priority and the technology they own or develop, each has distinct technology preferences. Currently, the greatest interest is in millimeter-wave. Thus, single, rather narrow recess processes with either 0.25- or 0.15- μm -long gates are commonly available. It is this availability that defines the domains of the applications, which either focus on very low noise, even up to W-band, or tend to medium power delivery. No double recess processing aiming at X-band power MMIC design and fabrication is presently available in these foundries.

Two other factors contribute to reluctance in Europe to use pseudomorphic HEMTs in lower-end applications.

First, the design-to-cost ratio is a key parameter for any commercial application. Apart from any professional preference for performance, the final cost of the chip appears to be the major decision parameter. Thus, the use of epitaxial substrates instead of an implantation process adds expenses that currently are not compensated by any clearly identified market request.

Second, for military applications, MESFET technologies have proven their reliability; so the performance improvements of the pseudomorphic material is less important. The delay in the availability of the technology and the associated lower experience level also influence decisions. Therefore, for applications up to X-band, the temptation to skip the pseudomorphic HEMT technology and directly jump to HBT is strong.

The decision becomes easier when the frequency is increased beyond 20 GHz. On one hand, HBTs still seem far from actual operation in this frequency range, except for frequency generation. On the other hand, the performances of MESFETs are heavily degraded as the gate length reduction cannot compensate the low gain. This comes from the aspect ratio: gate length/gate to channel distance, which was proven to be an accurate physical demonstrator of the capability of FETs. Thus, in order to keep a sufficient margin in gain, or output

power, the pseudomorphic substrate appears as the most mature solution; its only competitor is the more advanced, and much less mature, InP based transistor (AlInAs/GaInAs HEMT).

2.1. SYSTEM NEEDS AND P-HEMT USE

In a very basic way, two families of systems require MMICs: communication networks and radars. Figures 1 and 2 are typical synoptic diagrams of such systems. The system requirements are different in each case. For instance, radars need large, saturated output power while communication networks need lower output power but higher linearity. However, the needs are similar in terms of processing, as devices are very linear when they are far from saturation.

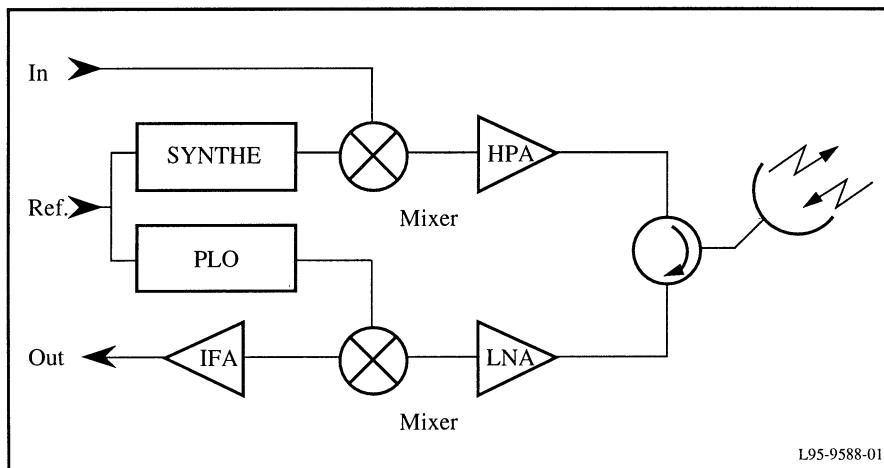


Figure 1. Typical sketched synoptic diagram of a node of a communication network.

In any case, basic functions such as up and down converters, low noise amplifiers, power amplifiers, and oscillators are required. While trade-offs in the design-to-cost ratio can eliminate the low noise amplifier, reduce the necessary output power, or modify any other parameter, the need for a large set of different chips still exists.

The higher the frequency the larger the impact of bondings will be on the chip performance. While it may be possible to fabricate the necessary prototypes with standard wire interconnections, the capability of such a technique to support large quantity low cost production is not proven. Thus, the integration of functions is clearly demanded, which transfers the cost of critical interconnections to MMIC integration. Therefore, a larger number of different functions can be realized in a single technology, as we exclude the use of much more complex - and expensive - processes that gather different devices such as HBTs and FETs. As long as the objective is low cost but efficient technology, trade-offs will be necessary. Such a technology is qualified as multi-function,

and the single recess P-HEMT is the ideal candidate due to its very high cut-off frequency, medium breakdown voltage and high current density. The success of such a procedure depends only on the capability to achieve high enough fabrication yields with 0.1- μm technologies.

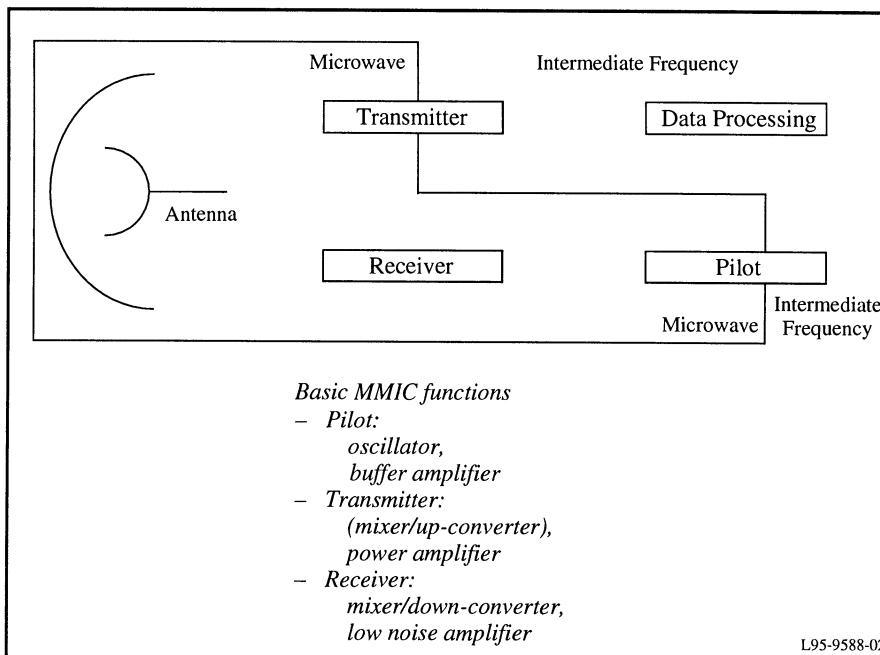


Figure 2. Typical sketched synoptic diagram of a radar transmitter/receiver module.

2.2. APPLICATIONS REQUIRING P-HEMTS

Finally, in the large number of applications where P-HEMT is likely to be used, only a few, rather short term programs can be identified, all of them dealing with millimeter-wave frequencies. (See Table 2).

TABLE 2. European applications of P-HEMT

	Status	Comments
High-speed, low-power-consumption digital circuits	Implanted MESFET	Cost (epitaxy)
Low-bias power generation (cellular phones)	Implanted MESFET	Cost (epitaxy)
X-band, large power generation	MESFET HBT	Reliability proven Next generation
Ka- and Q-band transmitter/receiver	P-HEMT	Availability
V- and W-band transmitter/receiver	P-HEMT	Availability

Considering the categories of applications in Table 2 as a point of reference, one can assume that P-HEMTs are viable for the programs listed in Table 3, which are underway in Europe.

An application common to several frequency bands is commercial communications. Present bands are over-crowded and several bands are being allocated, thus producing a need for a new channel with sufficient bandwidth for high-rate transmission.

TABLE 3. European programs with potential applications for
P-HEMT technology

Ka-band:
Commercial communication networks
Military communication networks
Friend-or-Foe Identification
Q-band:
Military ground-to-satellite communications
Commercial communication networks
V-band:
Commercial communication networks
Military communication networks
Automotive (beacon-to-mobile) communications
W-band:
Obstacle detection systems for cars
Proximity guiding radars.

Another application leading to mass production is the automotive market. Beacon-to-mobile communications require the expensive installation of a wide infrastructure, which will delay this program beyond year 2000. But, the obstacle detection systems are standalone products and should be incorporated into automobiles very soon (Figure 3). Prototypes have been tested for months in Germany and France. The major difficulty is related to the high frequency (77 GHz), which was chosen as the standard, and the very low target price. Thus, lorries, trucks and upper class cars should be the initial markets.

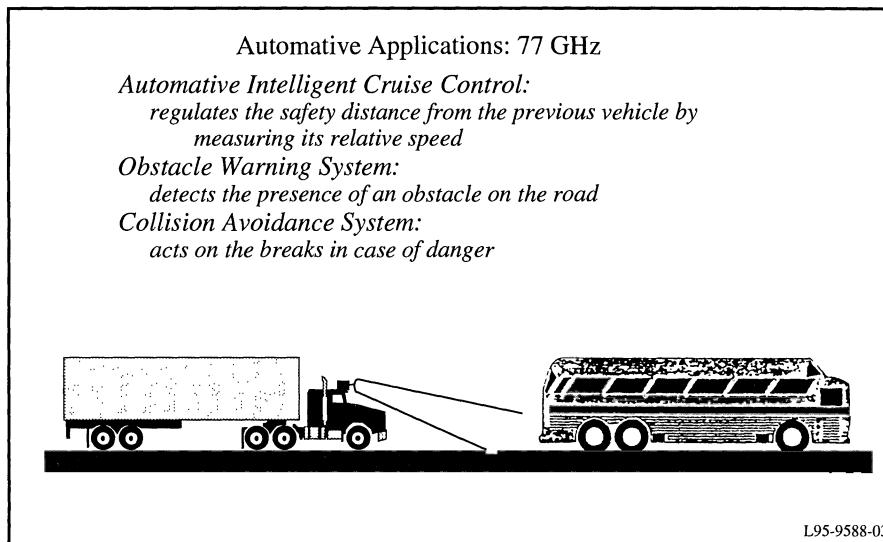


Figure 3. Obstacle detection systems.

3. Conclusion

In this paper, we have briefly described the gallium-arsenide industry in Europe and the market for P-HEMTs. The competition with other technologies, most of which are several years older, is very strong. Because users are conservative, often preferring reliability to performance, P-HEMTs seem to be excluded from low frequency (<20 GHz) applications. Nevertheless, when the frequency is increased beyond the standard capability of MESFETs, P-HEMTs become a viable option. Its relative youth, which is a drawback compared to MESFET, becomes an advantage against the brand new HBT or InP FETs. The millimeter-wave range is the ideal application area for this very promising technology. However, each program has specific needs and trade-offs, so the competition between the different processes will continue for a long time.

U.S. APPLICATIONS OF HEMT TECHNOLOGY IN MILITARY AND COMMERCIAL SYSTEMS

ELISSA I. SOBOLEWSKI
*U. S. Department of Defense
Advanced Research Projects Agency
Arlington, VA, U.S.A.*

1. Introduction

Discrete gallium arsenide (GaAs) devices have been used in an increasing number of microwave systems for about 15 years. The emergence of monolithic GaAs integrated circuits from the laboratory to use in fielded systems has taken place only within the past five years. Almost all of the current systems applications make use of ion-implanted metal-semiconductor field effect transistors (MESFETs). The improvements in basic materials, computer aided design (CAD) accuracy, patterning resolution, processing control, test speed, and packaging efficiency that have made the application of MESFET technology possible are, with very little further delay, being applied to the design and manufacture of high electron mobility transistor (HEMT) integrated circuits (ICs). Many applications being targeted for HEMT technology are in the millimeter wave range from 35 GHz to 94 GHz where MESFET performance begins to fall off. However, even in the frequency range from about 5 GHz to 35 GHz, particularly for those applications that require the lowest noise or highest efficiency circuits, HEMT technology has advantages. The particular tradeoff between cost and performance required by a system will determine which technology is chosen. As the cost and performance of available devices change with time, the system program manager's options increase and the choice may change.

As a result of the progress that has been made in the manufacture of HEMT components, they are being designed for use in a large number of specific systems in the United States for both military and commercial applications — sometimes as a cost-effective replacement for MESFETs that have already been designed into the system, and sometimes as original, enabling system components.

In order to merit consideration for insertion into the hardware design of any system, whether military or commercial, each technology under consideration must meet a number of criteria:

- *Performance:* It must meet the form, fit, and functional performance requirements of the system with adequate margin to ensure acceptable operation in the field.

- *Availability:* Circuits must be available in assured quantities from reputable vendors, preferably with second-source availability.
- *Reliability:* A level of reliability must be demonstrated which is compatible with the operational mission of the system.
- *Affordability:* Components must be affordable within the context of the system use.

The most prominent parameters of interest and measures of performance are the noise figure, the power output, and the efficiency as a function of frequency. In all three factors, HEMT ICs are competitive with other GaAs technologies below 35 GHz and generally superior above that frequency. At all frequencies, HEMT technology currently provides the lowest noise figure of any GaAs device.

The availability of HEMT IC components from merchant foundries within the past two years is largely a result of the stimulus provided by the U.S. Microwave and Millimeter Wave Monolithic Integrated Circuits (MIMIC) Program. It also is the result of the prospects for a rapidly developing market in commercial systems both at microwave and millimeter wave frequencies that require performance characteristics that can only be achieved readily through the use of HEMT monolithic microwave/millimeter wave ICs (MMICs). A number of foundries are now offering HEMT fabrication and/or design capabilities.

Some of the foundries offering HEMT capabilities include:

- Hughes
- Raytheon
- TRW
- Texas Instruments
- Martin Marietta
- Avantek/Hewlett-Packard
- Litton
- Alpha
- MiSig
- Westinghouse

The typical services provided by a GaAs HEMT foundry and the typical turn-around times are as follows:

Activity	Time to Complete
Deliver design and model data to customer	2 weeks
On-site design rule review	1 to 3 days
Customer design with consultation	1 to 6 months
Error rule checks, design rule checks, and error reporting	2 weeks
Reticle layout and final review	2 weeks
Mask tooling/procurement and verification	3 weeks
Wafer frontside fabrication, standard PCM test and yield analysis	7 weeks
Optional RF circuit test	1 week
Backside processing	2 weeks
Dice, chip sort, package, and shipment	1 week

An example of foundry yield results for 1993 are shown below. The data indicate the relative maturity of the HEMT manufacturing process at the present time.

	0.2- μ m HEMT	HBT	0.5- μ m MESFET
Line Yield (%)	75	92	80
RF Yield (%)	71	52	61
Cycle Time (weeks)	8	8	8

Data on the reliability of HEMT devices are just becoming available as the fabrication processes and recipes stabilize. The data indicate that HEMT MMICs are being produced with mean time to failure (MTTF) in the range of 10^6 to 10^7 hours at 150°C channel temperature. Testing under both dc and RF bias was performed.

The cost of all GaAs MMICs, especially HEMTs, is in a state of rapid change at the present time. Cost reductions depend strongly on the efficiency of the manufacturing process and on how fully the manufacturing lines are loaded. Using U.S. dollars per square millimeter as a measure, IC cost has decreased from \$10 to \$20/mm² about five years ago to projections of less than \$1/mm² after 1994. These projections assume that the number of chips manufactured of a given type is in the range of 100,000 or more and that a continuous level of throughput is maintained in the foundry.

There are expectations that the military and commercial markets for microwave systems in the 1 GHz to 100 GHz frequency range will grow substantially during the next two decades. There is little doubt that commercial demand will quickly exceed and then dominate the total market both in units produced and in total sales revenue. What is uncertain is the magnitude of the overall market as a function of time and, for the outlook of HEMT technology, the relative market share it captures. One view of the GaAs MMIC market, provided by BIS Strategic Decisions, Inc., is shown in Figure 1, which depicts the GaAs MMIC merchant market by process technology type for the years 1992 and 1997.

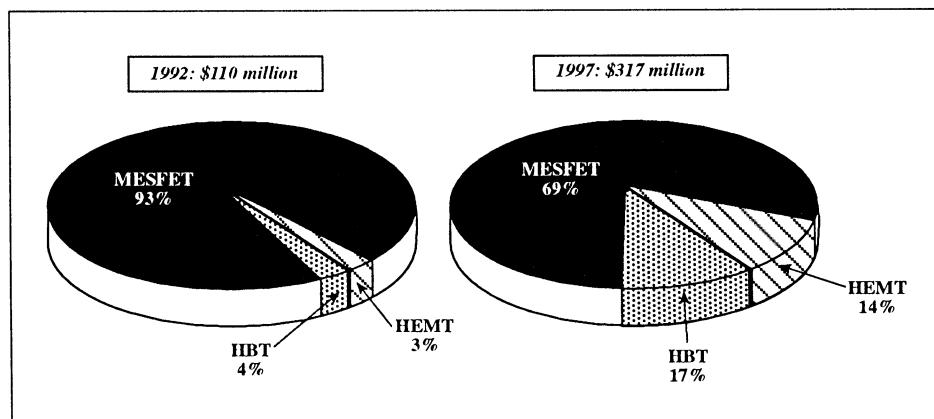


Figure 1. GaAs MMIC merchant market by process technology: 1992 and 1997

The status of the GaAs HEMT technology was described in the preceding paragraphs. It is now feasible to consider HEMT technology for insertion on a wide scale into all military microwave application areas. For commercial systems, the availability of affordable MESFET and HEMT GaAs technology has opened many areas of application that have not, until now, been technically feasible or have been prohibitively expensive. Some of these application areas are as follows:

Military		Commercial
Radar	Communications	Instrumentation
Electronic warfare	<ul style="list-style-type: none"> - Wireless phones 	<ul style="list-style-type: none"> - Remote sensing and data transmission
Smart weapons	<ul style="list-style-type: none"> - Cellular phones 	Direct broadcast satellites
Missiles	<ul style="list-style-type: none"> - Personal data/phone systems 	
Fuzes	<ul style="list-style-type: none"> - Wireless LAN/WAN 	<ul style="list-style-type: none"> - Satellite
Communications	Automotive	<ul style="list-style-type: none"> - Ground terminals
Imaging	<ul style="list-style-type: none"> - Vehicular obstacle detection/ warning 	<ul style="list-style-type: none"> - TVTs
Identification	<ul style="list-style-type: none"> - Highway surveillance/tracking - Toll collection - Traffic control - Vehicle identification 	<ul style="list-style-type: none"> Domestic products Navigation/GPS

The key benefits of HEMT insertions into systems are: reduced cost, lower noise, higher efficiency, higher power at millimeter wave frequencies, higher levels of integration, and reduced size and weight. These benefits lead to greater system capability in range, bandwidth, resolution, accuracy, and security.

2. The U.S. Department of Defense Microwave/Millimeter Wave Monolithic Integrated Circuits (MIMIC) Program

During the past five years in the United States, much of the progress made in applying GaAs MMIC technology to systems has been accomplished through the MIMIC Program sponsored by the U.S. Department of Defense. This program has provided the funding resources and management focus, from both the Government and industry, needed to bring this high-performance technology to an affordable level for insertion into fielded systems. The objective of the MIMIC Program is to develop and demonstrate the applicability, affordability, and sustained availability of MMIC technology and products through the development of chips, modules, and subsystems for military applications. The MIMIC Program also places special emphasis on bringing HEMT technology to a more robust manufacturing status for application to microwave and millimeter wave systems.

Phase 1 of the MIMIC Program began in 1988 and was completed in 1991. Four contractor teams comprising 27 companies participated in this phase. After successful completion of Phase 1, Phase 2 contracts were awarded to three contractor teams in August 1991. Phase 2 of the MIMIC Program will be completed in

fiscal year (FY) 1995. The MIMIC Phase 2 prime contractors and team members are listed below.

- **Hughes:** Alliant Techsystems, Cascade, EEsorf, ITT, Litton, M/A-COM, Martin Marietta (formerly GE)
- **Raytheon/Texas Instruments:** Aerojet, Airtron, Consilium, General Dynamics, Hittite, Lockheed-Sanders, Teledyne
- **TRW:** Alliant Techsystems, Hercules, Hughes Missile Systems, MiSig (formerly COMSAT), Northrop Grumman, Westinghouse

The examples of military applications of HEMT technology described in this paper are a result of the tasks carried out under the MIMIC Phase 2 Program.

3. HEMT Technology and U.S. Military System Applications

The MIMIC Phase 2 contractor teams are each performing a brassboard development and demonstration task. For this task, they have selected applications for which the advantages of MMIC technology use can clearly be demonstrated with brassboards or, in some cases, demonstration modules that incorporate one or more MMICs. Benefits accruing from use of MMIC technology include system cost savings, improved performance, improved reliability, and reduced size and weight. In many cases, HEMT technology is the only means for achieving the desired system capabilities.

The Hughes team is developing MMIC chips for use in four selected applications in radar, smart weapon, and communication systems (see Figure 2). Of the 18 chip types developed for the four applications, 13 employ pseudomorphic-based (PHEMT) MMIC technologies. These advanced technologies are being validated during MIMIC Phase 2 at the Hughes and Litton foundries.

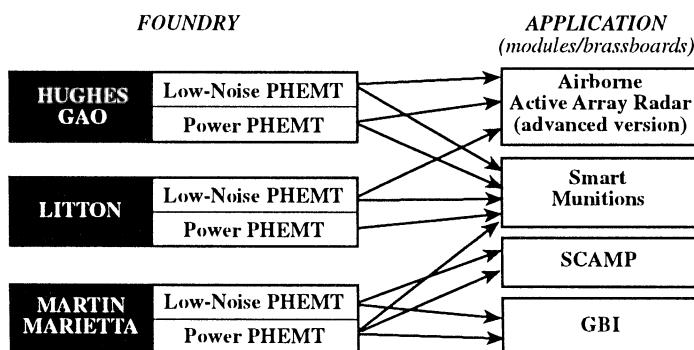


Figure 2. Hughes team HEMT military application

The Raytheon/Texas Instruments joint venture is developing four brassboards, 10 demonstrators, and 27 chips that directly apply MMIC technology improvements to a broad range of military systems (see Figure 3). The four brassboards and five of the 10 demonstrators use HEMT technology. Of the 27 chips, four are being fabricated in HEMT technology; nine others can be fabricated using MESFET and HEMT technologies, with the choice to be made at the end of the program.

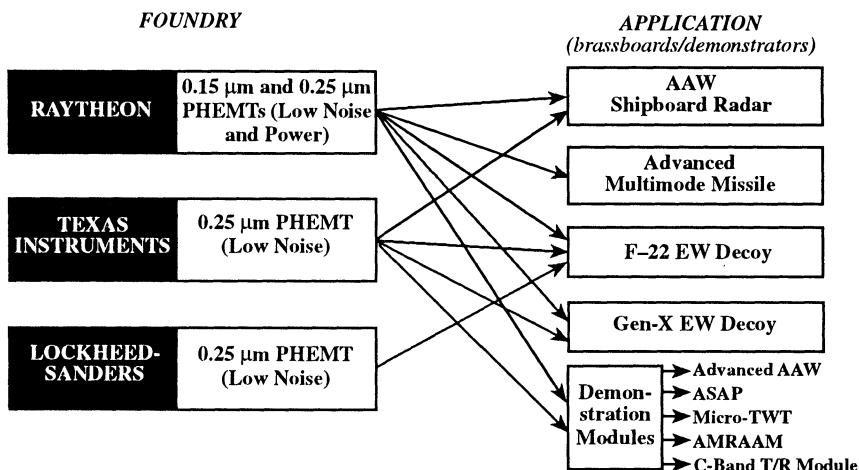


Figure 3. Raytheon/Texas Instruments team HEMT military applications

The TRW team is developing seven brassboards and three demonstrators for shipboard and airborne radars, electronic warfare, EHF communication links, imaging, and smart munitions/missiles (see Figures 4 and 5). Of the 29 chips being developed under the MIMIC Program, 19 use HEMTs, five use heterojunction bipolar transistors (HBTs), and the remainder use MESFETs. Particular emphasis has been placed on the development of PHEMT MMICs using a 0.1 μm process for two W-band (94 GHz) systems (MLRS/TGW and X-Rod). In both cases, size, weight, and power consumption are at a premium. The benefits gained from using HEMT technology allow the missiles to carry more payload to longer ranges with higher probability of hitting the target.

3.1. EXAMPLES OF HEMT TECHNOLOGY INSERTION IN RADAR SYSTEMS

The airborne active array radar brassboard being developed by the Radar Systems Group of Hughes Aerospace and Defense Sector has the potential of becoming one of the highest volume military uses of MMICs. Although the brassboard under development specifically addresses airborne radar, its essential features and capabilities are equally applicable to surface-based radars. The single-channel

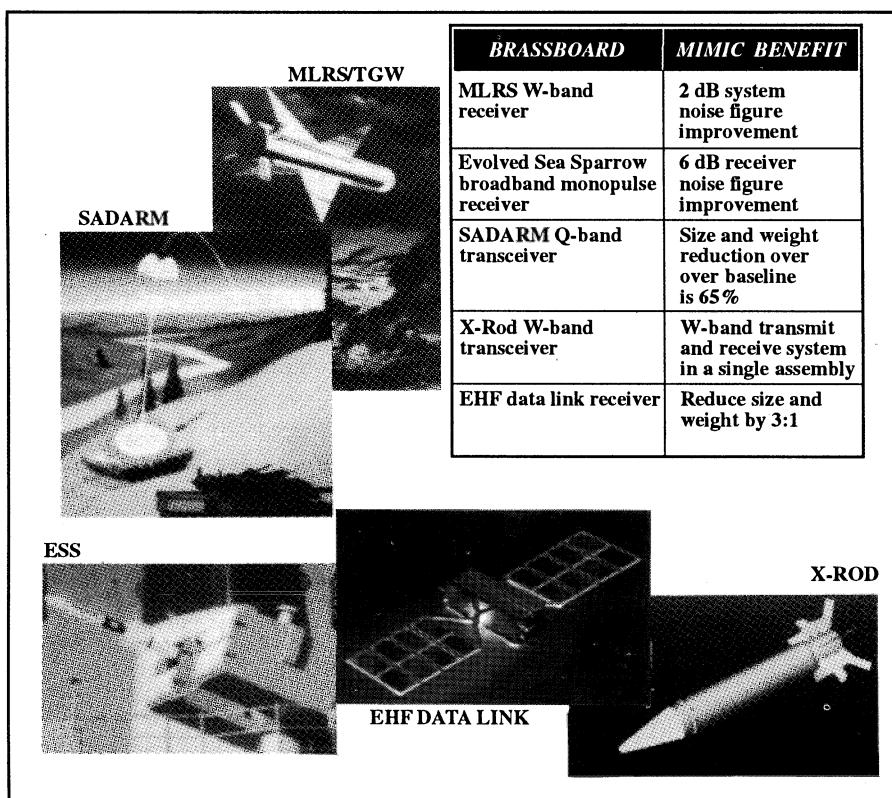


Figure 4. TRW team HEMT military applications

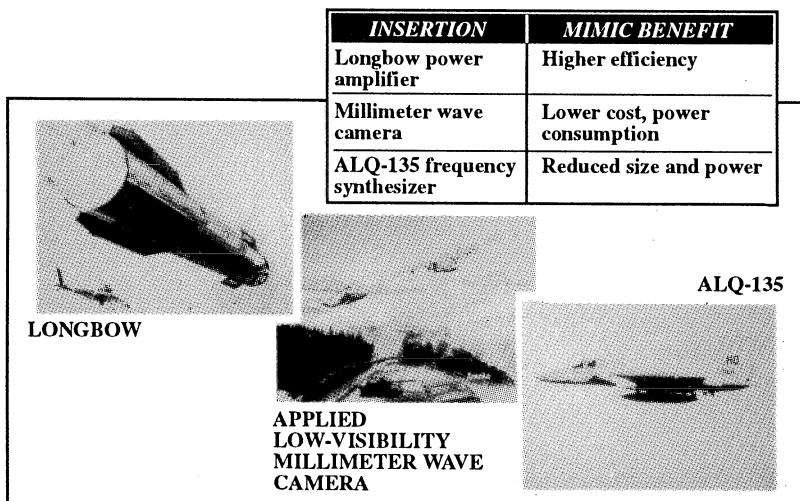


Figure 5. TRW team HEMT military applications

transmit/receive (T/R) modules used in the brassboard contain both microwave and digital circuits, and are the basic repetitive building blocks of an active array antenna. MMICs are essential for active arrays to meet required size and weight constraints and to achieve high reliability. Among the major performance parameters being improved are transmit power, efficiency, and noise figure. Performance increases are being achieved through improved chip technology, including the use of HEMT devices. The airborne active array brassboard will consist of an 8-watt T/R module and integral test equipment, including an interface box and software for automated testing. Two different RF module designs will be evaluated for use in the airborne radar brassboard: a baseline module design will use MESFET-based chips to meet requirements, and an advanced module design will attempt to achieve higher performance by using a PHEMT multifunction low noise amplifier (LNA) that has less than 2.5 dB noise figure and has been compacted to 30% of the size of the original MESFET design.

Raytheon's Anti-Air Warfare (AAW) brassboard is a four-element T/R module subarray building block for a large multifunction phased array. Both T/R module and subarray design approaches are based upon radar system specifications. The characteristics assumed for this radar are typical of those required for an AAW ship defense weapon system. The phased array system addressed by the AAW brassboard is multifaced, with several thousand active elements per face. The brassboard consists of a 4 x 1 element subarray. It will demonstrate the 1.028 inch horizontal element spacing and 0.115 inch maximum module height appropriate for use in a full-size AAW array. A T/R module architecture was chosen in which each housing contains the complete circuitry for a single T/R channel. The T/R module contains 11 GaAs MMIC chips (seven chip types) to accomplish both transmit and receive functions. Five of the chip types are processed using Raytheon's 0.25 μm PHEMT power and low noise processes. Four power amplifiers each provide a nominal 3.6 watts of output power and 12 dB of gain with a design goal of 50% peak transmit efficiency. The LNA has 30 dB of gain with a specified 1.8 dB noise figure. Achievement of 30% transmit module efficiency and 3.9 dB module receive noise figure is expected to require three chip design iterations and fully optimized PHEMT processes.

Raytheon's advanced AAW demonstrator provides a fully form-factored, specification-compliant T/R module demonstration. The module will be capable of supporting tactical AAW system requirements. Future AAW systems will require a dedicated radar capable of performing search, track, and target illumination functions. The system performance requirements for successful execution of these functions demand T/R module performance and producibility capabilities that can only be met with advanced MMIC technologies. An advanced 0.15 μm power PHEMT process will be used to improve module power efficiency from 25 to 30%. The 0.25 μm LNA PHEMT process will be used to meet the 3.9 dB module noise figure specification.

Texas Instruments' airborne shared aperture (ASAP) module demonstrator is a T/R subassembly that demonstrates the advantages of MIMIC Phase 2 technology in wideband T/R modules suitable for use in multifunction airborne arrays. A feature of the demonstrator is the application of a high-efficiency,

wideband 0.25 μm PHEMT power amplifier. One of the main requirements that must be met by modules intended for use in an airborne phased array is to achieve the desired power output while minimizing power dissipation. This new power amplifier will allow the ASAP module efficiency to increase from the 2–4% range achieved during the previous phase of ASAP to over 10% at a 45% transmit duty cycle in the present module. The power amplifier will also have more than 3 dB greater power output than the previous ASAP module.

3.2. EXAMPLES OF HEMT TECHNOLOGY INSERTION IN ELECTRONIC WARFARE SYSTEMS

Texas Instruments' generic decoy brassboard (Gen-X) consists of a full decoy system that will integrate upgrades of Modules 2 and 3 within the Gen-X production design. The Module 3 design approach incorporates new wideband power amplifier (0.25 μm PHEMT) chips and new wideband LNA (0.25 μm MESFET) chips to reduce overall assembly complexity while maintaining module performance. The Module 2 upgrade consists of integrating a wideband downconverting mixer, two single-pole/double-throw switches, and a low-gain amplifier (0.5 μm MESFET) on a single chip. These upgrades will reduce the total chip count of Modules 2 and 3 by 50% versus the Gen-X full-scale engineering development design. The modules will be form, fit, and function compliant with the Gen-X production configuration to enable direct mating with complementary hardware and, ultimately, to allow for production insertion.

The F-22 aircraft technology insertion effort, led by the team of Lockheed/Boeing/Pratt & Whitney, uses an integrated system approach for avionics functions. This approach includes electronic support measures (ESM); communication, navigation, and identification (CNI); fire control; electronic countermeasures (ECMs); and pilot awareness (cockpit displays). Lockheed-Sanders has the responsibility for the development of the ESM and ECM subsystems. Part of the subsystem consists of several electronic steerable arrays located throughout the platform to provide both azimuth and elevation coverage. The F-22 EW brassboard being developed on the MIMIC Phase 2 Program is an integral part of the arrays. The core of the brassboard is a single-channel low noise receive module. The channel receives signals from a vertically polarized aperture, uses phase and amplitude control for beam steering and tapering, and has built-in test circuits for in-flight diagnostics. The single-channel receive module will make use of advanced MMIC technology (0.25 μm PHEMT LNA) and state-of-the-art packaging technologies. The impact of the 0.25 μm PHEMT LNA on the modules is a reduction of the noise figure to 4.0 dB.

Teledyne's Micro-Traveling Wave Tube (Micro-TWT) driver module serves as the driver amplifier for a Microwave Power Module (MPM) now being developed under separate ARPA/tri-Service contracts. The MPM is intended to be a generic power source capable of being used in a variety of electronic warfare and other applications, including onboard jammers, expendable decoys, and phased array transmitters. Five companies (Hughes, Litton, Raytheon, Teledyne, and Varian) completed development of Phase 1A MPM modules in

September 1992; three companies were selected to continue development of a 50 to 100 watt, 6–18 GHz amplifier consisting of the Micro-TWT driver module, a TWT power amplifier, and an electronic power conditioner that provides power for both the TWT and driver modules. As part of the MIMIC Phase 2 Program, Teledyne plans to demonstrate a common solid-state driver module that uses wideband power chip sets designed under the MIMIC Program. To meet the more stringent Phase 1B MPM power and temperature requirements, the final Micro-TWT driver design will make use of 0.25 μm PHEMT power chips.

Northrop Grumman's MIMIC Multi-Macro Synthesizer chip is the primary component necessary for realizing many different wideband synthesized sources covering 0.1 to 26 GHz. The synthesizer chip was initially designed to operate from 2 to 20 GHz. The chip was designed and tested by Northrop and integrates 35 functions on a single chip. Included are amplifiers, mixers, oscillators, filters, switches, attenuators, and power dividers. The chip size is 3.7 mm x 4.1 mm. The chip is fabricated using TRW's 0.15 μm HEMT process. HEMT technology contributes to reduced size and power. The synthesizer chip was developed for insertion into systems such as the Advanced Technology Radar Jammer (ATRJ), ALQ-135 Electronic Warfare System Advanced Repeater Tuning Unit (ARTU), ALQ-135 Digitally Controlled Oscillator, local oscillator source for ALQ-162, and Virtual Instrument System Analyzer. There are commercial applications for this chip as well (e.g., use in a spectrum analyzer).

3.3. EXAMPLES OF HEMT TECHNOLOGY INSERTION IN SMART MUNITIONS/MISSILE SYSTEMS

Smart weapon platforms require the size, weight and cost reductions, and higher functional integration benefits that HEMT technology provides. Without all of these benefits, many systems under development are either not feasible or are severely limited in mission capability.

Alliant Techsystems is developing a PHEMT-based brassboard for use in Army smart munitions applications. The brassboard consists of two millimeter wave (Ka-band) radar sensors (one forward-looking transceiver and one downward-looking transceiver) and planar circuit antennas. Each sensor is based on a frequency-modulated continuous-wave (FMCW) radar architecture with LNA input receivers. Since considerable commonality exists between the two radars, only five unique chip designs are needed to implement all of the brassboard functions. The five chip types are forward VCO (FVCO), downward VCO (DVCO), balanced power amplifier, mixer, and LNA. All chips utilize PHEMT technology. A total of 12 single-function MMIC chips will be inserted into each T/R module pair, eight in the forward module and four in the downward module.

The Advanced Multimode Missile (AMMM) receiver brassboard being developed by Raytheon's Missile Systems Division is a complete four-channel microwave receiver for use in advanced multimode missiles. The primary objective of the brassboard is to demonstrate specified performance in a volume compatible with missile requirements, approximately 50 in³. The multimode receiver comprises three types of subassemblies: Ka-band, X-band, and C-band. Four Ka-band

subassemblies, one X-band subassembly, and one C-band subassembly are integrated, enabling Ka-band active (monopulse), Ka-band passive (interferometer), and X-band semiactive (monopulse) receive functions. The Ka-band subassembly provides frequency downconversion from Ka-band to C-band. Five MMIC chip types are being developed for use in the subassemblies. A total of 20 of these chips are required for a receiver brassboard. The chips include a Ka-band LNA (0.25 μm PHEMT) that is used in four places within the receiver.

Raytheon's Advanced Medium Range Air-to-Air Missile (AMRAAM) demonstrator module involves developing two MMIC chips for system insertion. The chips are the X-band LNA (PHEMT) and the mixer/intermediate-frequency (IF) amplifier. Raytheon is targeting the use of these chips in the RF processor (a near-term MMIC insertion opportunity) and the data link processor. Two versions of the data link processor have been developed, one based on the 0.25 μm PHEMT LNA and one on the 0.5 μm MESFET LNA. The version that provides the maximum benefit in terms of cost and performance will be implemented into production.

The Multiple Launch Rocket System/Terminal Guidance Warhead (MLRS/TGW) is a multinational program to develop an advanced "fire and forget" anti-armor weapon, including a terminally guided submunition (TGSM) that can operate under adverse weather and battlefield conditions. The W-band millimeter wave seeker-sensor is the crucial element in providing all-weather targeting under battlefield conditions. The seeker-sensor for this weapon autonomously performs target search, detection, acquisition, tracking, and terminal homing. There is a critical need to reduce the cost of this seeker to meet MLRS/TGW program unit-production cost goals. The U.S. Army Missile Command (MICOM) is supporting a Manufacturing Methods and Technology (MMT) program to achieve this cost reduction. In parallel with the MMT effort, the TRW MIMIC program includes the development of a 94 GHz (W-band) brassboard receiver. TRW's brassboard uses 0.1 μm pseudomorphic indium GaAs (InGaAs) HEMT technology chips that are designed to be form, fit, and function replacements for the existing MMT LNA/mixer circuits. TRW is planning to insert the W-band brassboard as a replacement for the high production cost items in the MMT receiver.

Hughes Missile System's broadband, monopulse, four-channel receiver brassboard is targeted for use in the Evolved Sea Sparrow Missile. Each channel performs low noise amplification of incoming signals and converts the signal to the IF determined by the system's local oscillator. A change in performance requirements for "U.S. Only" non-Aegis class ships requires a 6 dB receiver noise figure improvement. HEMT technology is the likely solution. Two MMIC chip types are being developed for this brassboard, a 9- to 14-GHz LNA and an image reject mixer frequency downconverter. Each uses TRW's 0.2 μm pseudomorphic InGaAs HEMT technology.

Alliant Techsystems' Q-band transceiver brassboard is based on the U.S. Army's sense-and-destroy armor (SADARM) radar-radiometer integrated front end (IFE). The current SADARM baseline IFE consists of a single Q-band planar antenna with separate transmitter and receiver modules. The IFE design is

similar for both the SADARM MLRS and the 155 mm configurations. The 155 mm configuration is the MIMIC Phase 2 brassboard demonstrator baseline. The dual-mode interim modules use a mixer and PHEMT LNA for the receive circuit and a VCO and PHEMT X2 multiplier/power amplifier for the transmit circuit. The final brassboard will use a single-module approach that integrates the two separate T/R circuits onto a single carrier and housing to handle all RF functions. The X2 multiplier/power amplifier will use PHEMT technology. MMIC insertion into SADARM offers significant performance, reliability, and cost advantages compared to previous hybrid prototype designs. The key impact of HEMT usage on SADARM is improved design performance margin through optimum selection of chip set topology and material technology. This will result in increased subsystem yield and lower submunition costs.

Hercules' W-band transceiver brassboard is targeted for insertion into the U.S. Army's X-Rod smart munition. The brassboard transceiver is an FMCW W-band millimeter wave single-channel transmitter combined with a dual-channel monopulse receiver. The receiver RF circuits and transmitter channel are implemented using 0.1 μm pseudomorphic InGaAs LNA, image reject mixer frequency downconverter chips, and 0.2 μm pseudomorphic InGaAs HEMT power amplifier chips. The MMIC insertion into the X-Rod smart munition provides a more cost-effective technology alternative for this future weapon with a potential total cost savings of $\sim \$1$ million (classified quantity). HEMT MMICs provide significant weight and volume reduction as well as reduced receiver noise and increased power output.

The Longbow Hellfire missile is a fire-and-forget system designed for the Army AH-64 Apache attack helicopter. The Longbow missile homes in on targets using millimeter wave radar technology instead of laser guidance. The missile radar transceiver requires solid-state technology (as opposed to TWT amplifiers) in order to meet system size, weight, and operational constraints. Under MIMIC Phase 2, TRW developed a 1 watt Ka-band power amplifier breadboard to demonstrate the suitability of the new higher output power MMIC chip for use in Longbow. The Ka-band power amplifier module contains four power amplifiers—a driver amplifier MMIC followed by a three-way power combiner circuit summing the outputs of three MMIC power amplifier chips. Each power amplifier chip is 4.8 mm x 2.3 mm. The chips were processed using TRW's high-reliability passivated 0.15 μm pseudomorphic InGaAs power HEMT technology.

3.4. EXAMPLES OF HEMT TECHNOLOGY INSERTION IN COMMUNICATIONS SYSTEMS

Many communications applications require very small, lightweight and low-cost receivers and transmitters. A number of defense space programs that are either in procurement or under consideration by the U.S. Government have communications systems with similar architectural needs. For example, the systems generally require uplinks (ground to space) that use the 44 GHz band and downlinks (space to ground) that use the 20 GHz band. For protection against ground-based jammers and detectors, cross-links (between space-based platforms)

are required to be in the 60 GHz band to take advantage of the band's high atmospheric attenuation characteristics.

Martin Marietta's Government Communications Systems Division will demonstrate the application of MMIC to 20 and 44 GHz communications modules for the Single Channel Anti-Jam Man Portable (SCAMP) terminal. MMIC technology provides small size, lightweight, and low cost, while meeting MILSTAR needs for reliable anti-jam data and voice communications. MMICs developed for SCAMP are being implemented into the LNA/converter module of the T/R assembly. The plan is to use MMICs in both production phases of the program. The chips being developed are a 20 GHz (K-band) LNA, a 44 GHz (Q-band) gain block amplifier, and a 44 GHz intermediate power amplifier. The chips use 0.25 μm PHEMT technology.

Martin Marietta's Astro Space Divison is developing insertion modules for application to a new EHF communications subsystem designed for widespread use in missiles and spacecraft. During MIMIC Phase 2, Martin Marietta is concentrating on achieving initial MMIC insertion into the Ground-Based Interceptor (GBI) program. The GBI communications system must provide a reliable, high-performance EHF link to the system ground segment employing MILSTAR frequencies. Low noise receivers and high power transmitters are required for maintaining the link in the presence of atmospheric propagation effects, ECM, and vehicle maneuvers. The components selected for MMIC insertion into GBI are the 44 GHz single-chip LNA, 5 watt module for the uplink receiver and a 20 GHz transmitter amplifier for the missile downlink. Both the 44 GHz LNA chip and the 20-GHz power amplifier chip being developed for the modules use 0.25 μm PHEMT technology. Although these devices are being developed for the GBI program, the resulting MMIC technology will have broad application to a wide range of missiles and spacecraft.

TRW's EHF data link transceiver brassboard is intended to demonstrate the ability of MMIC technology to provide lighter, smaller, and higher performance communications subsystems. The brassboard contains five component sections or slices: a 60 GHz (V-band) receiver slice; a 60 GHz (V-band) transmitter slice; a 44 GHz (Q-band) receiver slice; a 20 GHz (K-band) transmitter slice; and a frequency source slice. The entire brassboard is contained within 127 cubic inches and will weigh approximately 4.03 pounds, including the housing, which weighs 2.52 pounds. Of the 13 chips being developed for the brassboard, 11 are fabricated using TRW's HEMT technology. Implemented using 0.2 μm pseudomorphic InGaAs HEMT technology are: the Q-band downconverter, V-band downconverter, K-band upconverter, V-band upconverter, V-band 2-stage LNA, and V-band 4-stage LNA. Implemented using 0.15 μm pseudomorphic InGaAs HEMT technology are: the I-band/J-band mixer, V-band X4 multiplier, K-band X2 multiplier, and K-band medium power amplifier. The V-band 2-stage balanced amplifier is fabricated using TRW's 0.1 μm InGaAs HEMT technology. MMIC insertion offers a threefold weight and volume reduction, a reduction in parts count, and a 4:1 increase in functionality and reliability. These advantages are multiplied by the large number of satellites or interceptors planned for deployment.

An example of the desirability of HEMT technology at lower microwave frequencies is the C-band T/R module demonstrator being developed by Raytheon. This module is being designed for a U.S. Navy airborne communication active array application. The specifications for the module are derived from system-level requirements as specified to the Raytheon/Texas Instruments Joint Venture by the Navy. The T/R module contains five GaAs MMIC chip types to accomplish both the transmit and the receive functions. Three of the chip types are processed using Raytheon's 0.25 μm PHEMT power and low noise processes. Two power amplifiers each provide approximately 10 watts of output power and 22 dB gain with a design goal of 55% peak transmit efficiency. (The equivalent design using MESFET technology is estimated to provide 8 watts of power, 18 dB gain, and ~30% power added efficiency. This corresponds to an 18% efficient module.) The LNA has 20 dB of gain with a specified 1.25 dB noise figure. (MESFET technology would produce approximately 1 dB higher noise figure at both the chip and module level.) Achievement of 32% overall module efficiency and 4.0 dB noise figure is expected to require two chip design iterations. Eighteen pilot production modules will be built using the first design iteration in late 1994. Five hundred production modules will be built using the second design iteration chips during the first half of 1995.

3.5. EXAMPLES OF HEMT TECHNOLOGY INSERTION IN IMAGING AND IDENTIFICATION SYSTEMS

TRW has demonstrated a MMIC-based single-pixel W-band direct detection receiver. This receiver uses two MMIC W-band three-stage LNAs and a preamplified detector MMIC, all fabricated using 0.1 μm pseudomorphic InGaAs HEMT technology. Combined, the amplifier and video detector perform a complete direct detection receiver function on a single chip. MMIC direct detection eliminates the necessity for the local oscillator and mixer used in the conventional heterodyne (hybrid) receiver approach (17.5 dB system noise figure) for reduced complexity, power, and cost, and provides a better noise figure (an 8 dB improvement). The single chip was integrated into a W-band 1 x 8 focal plane array for a passive MMW camera application. The receiver was connected to a TRW Passive MMW Imaging System.

TRW has demonstrated a Battlefield Combat Identification System (BCIS) using existing MMIC products. BCIS is a millimeter wave (Ka-band) query-and-answer system and will help save combat soldiers' lives. Prototype units were built in three months. The baseline transceiver uses five existing MMIC chip types per system. HEMT technology insertion offers performance benefits and helps reduce unit production cost.

4. HEMT Technology and U. S. Commercial System Applications

Although not specifically addressing commercial products, MIMIC Program successes have spun off into a wide variety of commercial applications. HEMT

technology is being applied to commercial products that require high power and linearity combined with low power consumption at voltages as low as 3 volts. The competitive price and performance advantages of HEMTs provide an opportunity for their use at operating frequencies ranging from L-band to millimeter wave. The classes of applications where HEMTs have a particular advantage over the more conventional MESFET technology are described below.

Ground Receiver Stations: HEMT front-end amplifiers provide high gain and low noise figure for improved sensitivity of ground receiver stations. System applications include Direct Broadcast Satellite (DBS) receivers, which continue to grow in use in Europe and the Pacific Rim countries, and Global Positioning System (GPS) receivers.

Wireless Communications: Insertion of HEMT technology into wireless commercial transceivers is necessary to achieve high linearity and low power consumption at 3 volt operation. Systems being produced include wireless local area networks and cellular telephones. Also included within the category of wireless communications are the wireless city data networks known as MANs (metropolitan area networks) and WANs (wide area networks) that enable data communications between buildings and/or between cities. High power requirements and high data rates that necessitate higher operating frequencies give HEMT technology the edge over MESFET technology in these applications.

Satellite Communications: HEMT technology is key for providing high power, high efficiency, and low noise figure T/R modules for use in phased-array space satellite systems. The Iridium^{TM/SM} system, which consists of 66 low altitude orbiting satellites for data and voice communications, is based on HEMT technology. The Very Small Aperture Terminal (VSAT) system, used for data transmission, requires the high power and efficiency that HEMT technology provides.

Automotive Applications: A number of automotive products are being developed at millimeter wave frequencies (e.g., collision warning systems, collision avoidance systems, autonomous cruise control). A radar lane-changing aid has been demonstrated at 35 GHz. At W-band (94 GHz), TRW is adapting a transceiver chip, developed under the MIMIC Program, for use in potential automotive radar applications. The transceiver is a single-chip FMCW design using 0.1 μ m HEMT technology. The unit has a noise figure of less than 6 dB, The final unit dimensions will be less than 3" x 4" x 4".

Other Commercial Applications: Other novel applications include the use of HEMT technology in microwave stimulated lighting for industrial and automotive use.

5. Conclusion

HEMT technology has the potential for widespread use in both U.S. military and commercial systems. Military applications most likely to use HEMT technology include radar, electronic warfare, smart munition/missile, communications, and imaging/identification systems. Potential commercial applications include

ground receiver stations, wireless communications, satellite communications, and automotive systems.

It is expected that both military and commercial applications for HEMT technology in the microwave and millimeter wave frequency ranges will increase rapidly during the next five to 10 years. Estimates show HEMT technology capturing an increased share of the expanding MMIC market.

Acknowledgement

The author wishes to thank her colleagues throughout the Department of Defense and industry for their many helpful discussions about HEMT technology and its use in both military and commercial systems. She also wishes to thank Mr. Albert Brodzinsky for his time, patience, and many valuable suggestions. A special thanks goes to System Planning Corporation for their graphical support.