

1

RF and Microwave Power Transistors

1.1 Introduction

While wireless communications standards may come and go with developments in the latest digital coding technology, or the liberation of new fragments of the electromagnetic spectrum, the common denominator among the various communications systems is the power amplifier. Over the last few decades, the transition from vacuum tubes and other forms of amplification to solid state devices has been almost complete, especially at power levels less than 1 kW. Nowadays, at the heart of the power amplifier, we find the power transistor.

In the world of RF wireless communications, the base-stations and long range transmitters use silicon LDMOS (laterally-diffused MOS) high power transistors almost exclusively. In addition to modern cellular communication systems, LDMOS devices are also used in a wide range of applications requiring radio-frequency power amplification: HF, VHF, and UHF communications systems; pulsed radar; industrial, scientific and medical (ISM) applications; avionics; and most recently in WiMAX™ communication systems. The frequency range of these applications is from a few megahertz and up to 4 GHz.

While LDMOS technology is pre-eminent in high-power RF and lower frequency microwave applications, a wide variety of compound (III–V) semiconductors are used as effective power amplification devices, especially for application frequencies above about 5 GHz, and also for lower power applications such as cellular handsets, Bluetooth™ and other wireless local area networks (WLANs), which generally demand output powers of about 1 watt or below. The most commonly used compound semiconductor material for RF and microwave applications is gallium arsenide (GaAs), which is used as the basis for field effect transistor devices such as metal-semiconductor

FETs (MESFETs), heterojunction transistors such as high electron mobility transistors (HEMTs), and heterojunction bipolar transistors (HBTs). The basics of operation of these devices will be outlined later in this chapter. Gallium arsenide FETs and pseudomorphic HEMTs (PHEMTs) are used for low-power handset power amplifiers, as well as in some high-power cellular base-station applications, and they are also used for wide bandwidth power transistors up to the millimeter-wave regime. Recent compound semiconductor technology developments have led to the introduction of gallium nitride (GaN)-based HEMT devices, which boast very high power densities, and, depending on the substrate material used, can also have very low thermal resistance, making these devices well suited for high-power amplification. Gallium nitride FETs also have high transition frequencies, similar to some GaAs FET technologies, and so they also have the potential for high-power amplification while operating at microwave and millimeter-wave frequencies.

Even as the complexity of wireless systems continues to increase, the pressure to reduce the design cycle time and the time-to-market is challenging the designer to evaluate and develop alternative design techniques. Traditional empirical design methods based on experience and measurements are being replaced with computer-aided design (CAD) approaches. In a CAD-based design flow, the need for accurate and validated models that are properly implemented in the design tools is of paramount importance.

In this book, we shall focus on the device modeling issues peculiar to high power RF applications. Here we should state that we shall use the term 'RF' to include the RF, microwave and millimeter-wave frequency bands, in which the field effect transistor is widely used for power amplification. The use of transistors at high powers brings a unique set of problems that must be overcome in the development and deployment of a transistor model that can be used in the successful design of high power RF amplifiers. The power transistors are of considerable physical size, as can be readily appreciated from Fig. 1.1. This figure shows a transistor capable of delivering 140 watts of power at 1-dB compression, at 2.1 GHz. The complexity of the transistor in terms of the number of constituent components is a major factor to be considered in how the device is modeled. This complexity is manifest in a number of issues that must be addressed in a successful realization of a model, including electromagnetic interactions between matching networks, packaging considerations, thermal management, and the self-consistent integration of the thermal model with the electrical model of the device. The power transistor has very large total gate-width, generally much too large

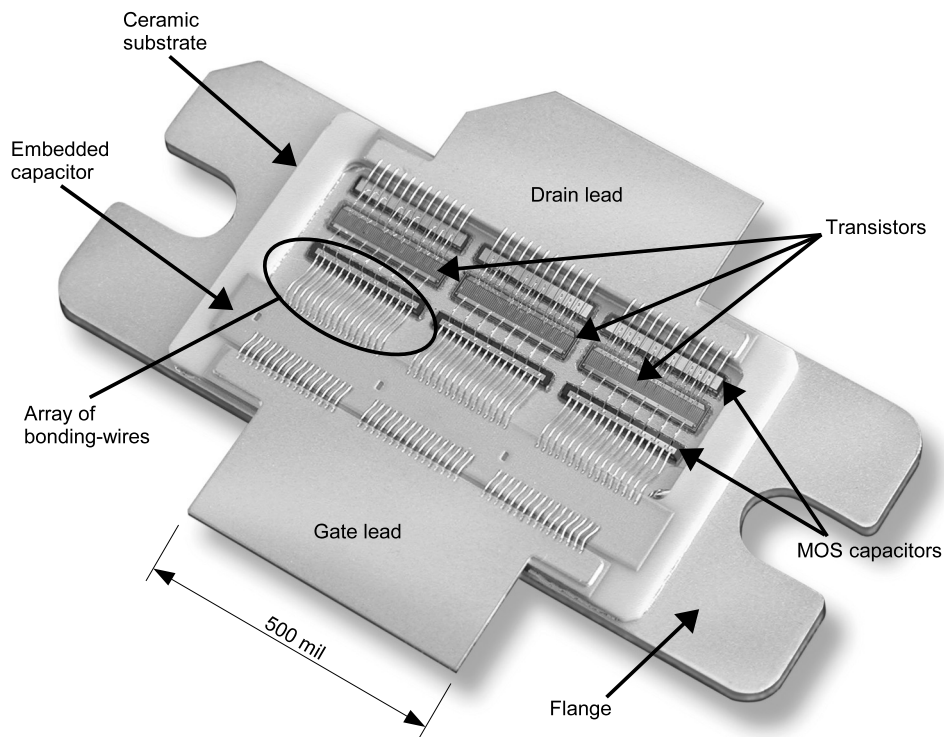


Fig. 1.1. A view of a high-power LDMOS transistor, with the lid removed to show the complexity of the internal matching networks and the LDMOS die [1]. © 2006 IEEE. Reprinted with permission.

to measure and characterize directly, so the question of scaling must be addressed. How these problems are overcome depends upon the overall device modeling strategy: the interaction between measurement techniques and function approximations, and between accuracy and implementation in the CAD tool. Our goal is to describe how these compromises can be addressed successfully.

Although the focus of this book is primarily on the high-power LDMOS FET, as this is the current market-leading technology, we shall develop, describe, and apply many measurement, analysis, and model synthesis techniques that can be more widely applied to FET modeling in general. For example, the nonlinear analysis and modeling techniques can be applied to other material technologies, such as GaAs or GaN power devices, which

may be used in different applications, such as much higher frequencies and bandwidths.

In building the transistor model, we should bear in mind the following *Laws of Simulation and Modeling*[†]:

- (i) A simulation is only as accurate as the models it is based on;
- (ii) A model is (mostly) useless unless it is embedded in a simulator;
- (iii) Models are, by definition, inaccurate; it's just a matter of degree;
- (iv) Models generally trade off complexity (simulation time) for accuracy.

1.2 Outline of the Transistor Modeling Process

In this section we present the overall processes we follow for model development and provide some insight into the tasks required. It is these tasks that are to be presented in detail in subsequent chapters. In general the process of developing and extracting a model is shown in Fig. 1.2. Not only is this typical of the process but it also serves as an overview to the book.

At the beginning of the modeling process it is expected that the application of the model and the ranges within which it is expected to operate are defined. For example, the requirements of model to be used in a high-power Doherty or Class F are very different from those for a low-power Class A design. Although it is our objective to generate a model with a wide range of applicability, the model must function in its intended application. From our experience, this step of defining the model is very time-consuming and is often overlooked. A clear and complete definition of the scope and deliverables of a project should be received before moving into the execution phase; this is fundamental to project management methodology. The time taken to investigate and document the model specifications is well invested, as it forces the modeling engineer and the customer to agree upon a set of objectives and validation criteria for the model. The type of model that is required is determined from these discussions. Without completing this task, which is simple in concept, but difficult to achieve in practice, there is a high probability that a model will not match the desired application or expectations of the customer.

Once the model topology has been finalized, data need to be gathered either through measurement or simulation. Typically, the frequency range, impedance ranges, and power levels are used for the specification of nonlinear models. For linear devices the frequency range and parametric variations

[†] These 'laws' have been attributed to both Colin MacAndrew and Mike Golio; we are unsure who has the precedence.

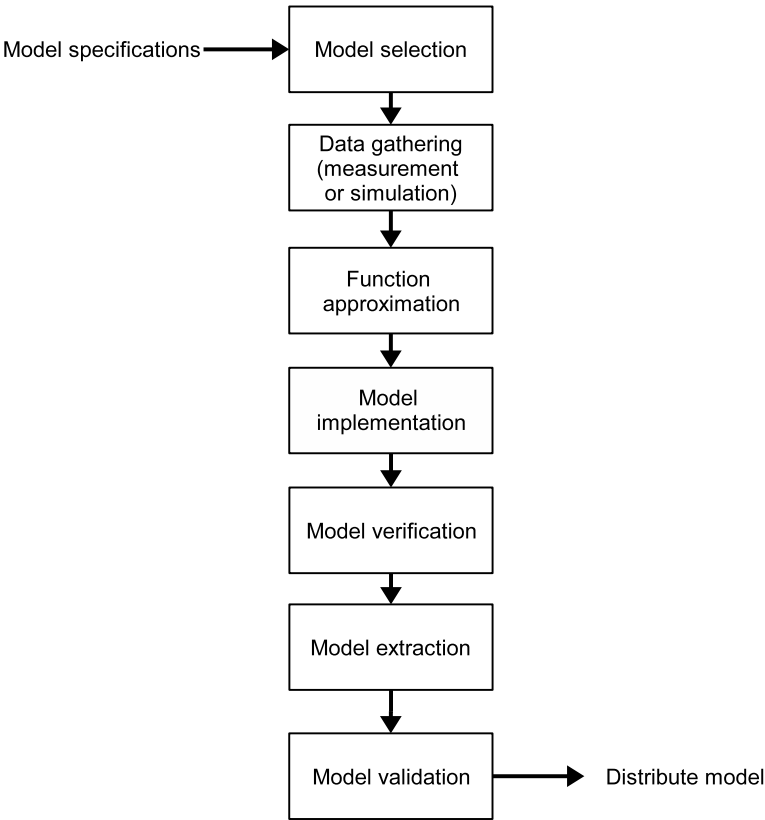


Fig. 1.2. A flow chart illustrating the distinct processes required to generate a model. Each step is covered in detail in the following chapters.

of the geometry are often specified. In Chapter 3, several measurement techniques for the extraction of linear and nonlinear models are presented. In Chapter 5, we discuss the measurements required for thermal sub-circuit generation. Electromagnetic and thermal simulation techniques are also covered in Chapters 4 and 5, respectively.

The model is generated by transforming the measured data into a format that is suitable for the implementation of the model in the chosen topology. As an example, for large-signal transistor models, the nonlinear charges and currents for the transistor model need to be extracted (see Chapter 6), then a function approximation technique is applied to generate a mathematical representation of the data that is suitable for use within a circuit simulator. Several practical approaches to function approximation are described in Chapter 7. The transistor model must then be implemented in the circuit

simulator of choice, verified, and validated prior to its release: these are the topics of Chapters 8 and 9.

1.3 A Review of the Commercial Applications of High-Power Transistors

Since the beginnings of radar development during the Second World War [2], there has been a need for components that can generate large amounts of power at RF and microwave frequencies. Initially, pulsed radar systems were designed to use klystrons, traveling wave tubes (TWT), and cavity resonators (magnetrons): vacuum tube microwave devices. In the 1960s the first semiconductor devices, GaAs transferred electron devices and silicon and GaAs IMPATT (impact ionization transit time) diodes, began to be used in radar systems as the main RF power amplification component. Compared with vacuum tubes, these first solid-state devices generated relatively low power, and were only used for short-range applications. Nowadays, semiconductor devices are widely used in pulsed radar applications. These applications present another interesting collection of design challenges: the pulsed nature of the signals used in many radar systems means that the devices are not as thermally stressed as they are in CW applications, and their transient performance can be optimized for these pulse signals.

The commercial development of HF, UHF, and VHF communication and broadcast systems drove the growth and technology innovations of high power transistors, power amplifiers and communication systems further during the 1980s. Paging and cellular communication systems then brought a new wave of products and technologies operating at higher frequencies. The first generation of cellular systems, based on analogue modulation, were introduced in the 1980s and were limited to use within the 900 MHz frequency bands. During the 1990s, the second-generation ('2G') cellular systems revolutionized the communication industry by integrating voice and data through the introduction of digital modulation. These second-generation systems also shifted to higher frequencies in their quest for available spectrum and higher data rates (900 MHz, 1.9 GHz, and 2.1 GHz). In the early years of this century, third-generation ('3G') cellular systems were becoming available; these more complex systems are distinguished by even higher data rates, resulting in more stringent requirements on the efficiency and linearity of the power amplifier.

A broad category of applications that requires the generation and amplification of high powers is the industrial, scientific, and medical (ISM) market. Examples of such applications are plasma generators for etching, surface

finishing, coating, magnetic resonance imaging (MRI), industrial lighting, microwave heating and drying, and so forth. In this market, there is a growing need for higher power levels, and for some applications many devices are assembled in parallel to obtain power amplifiers that can deliver several kilowatts. A new trend is to offer high performance RF transistors that are designed to operate at 50 V, thereby increasing the power density per device, and hence reducing the number of devices per power amplifier. A further benefit of using the higher voltage transistors is that the high power levels can be obtained with higher terminal impedances [3], making it easier to design amplifiers of broad frequency bandwidth.

A relatively new application in which RF high-power semiconductor devices are being used instead of tube RF amplification devices is digital television or video broadcasting (DVB). This application spectrum covers 480 MHz to 880 MHz, and because of its large bandwidth, presents a different set of challenges to the RF transistor designer. In this application, many transistors are used in a large combining network to achieve the required amount of RF power. To maximize the amount of power output per device, the devices used in digital broadcast applications operate at 32 V DC instead of the more common 28 V DC that is used in other RF power applications, such as wireless infrastructure base-stations.

Another recent application area for high-power RF transistors is in WiMAX communication systems. These systems represent a new opportunity for LDMOS power transistors as they are pushing the frequency of operation higher than is found in the traditional wireless communication systems. WiMAX systems are being designed and built to operate at several frequency bands for different regions of the world, with the majority of the development centered around 2.7 GHz and the 3.4 to 3.8 GHz bands. These higher frequencies also bring a new set of challenges to designers, as more distributed phenomena must be considered for optimum transistor and circuit design. In addition, these systems are designed with bandwidths that are much wider than traditional wireless communication systems, further increasing the complexity of the design.

1.4 Silicon Device Technology Development

Silicon metal-oxide-semiconductor field effect transistors (MOSFETs) are the most common devices in use today in the high-power RF market. They offer several advantages over bipolar devices in RF power applications as they tend to exhibit lower intermodulation distortion, and do not suffer from thermal runaway [4]. With the requirement for ever more linear transistors,

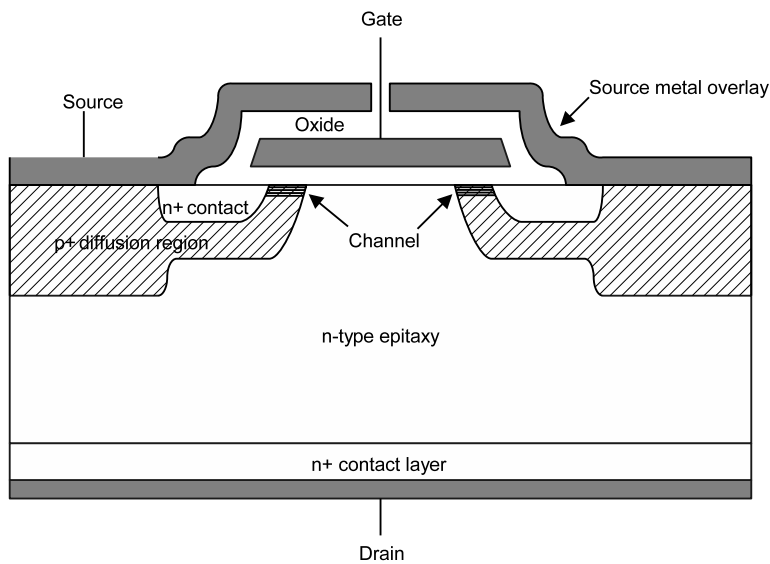


Fig. 1.3. Cross-section of a VDMOS transistor.

the BJT was relegated in the early 1990s. The cost and performance trade-off of the MOS devices also provides a clear advantage over many III–V technologies at frequencies below 4 GHz. The silicon LDMOS FET is the predominant technology among all the silicon high power FET transistors. In this section we will review the technology progression that led to the development of the silicon LDMOS transistor.

The development of the power MOSFET was marked by the introduction of a structure that allowed the current to flow through the substrate vertically. Figure 1.3 shows a cross-section of a vertically-diffused MOS (VDMOS) transistor, which is characterized by a structure that has the source and drain terminals on opposite sides of the silicon wafer. The VDMOS FET allows a more efficient use of the space on the wafer since it enables a higher concentration of the active cells. The VDMOS structure is an enhancement-mode device, which requires only a single bias polarity.

The vast majority of RF transistors are used in the common source configuration and, therefore, the VDMOS device structure has a major disadvantage in that its drain electrode is located in the back side of the wafer. As a consequence the source of the transistor, which is at the top surface of the silicon wafer, needs to be connected to the grounded terminal of the package. This results in a package construction that is very complex, requiring the use of an insulating material to isolate the back

side of the wafer (the drain terminal of the transistor) from the grounded metal carrier of the package. Beryllium oxide (BeO) was commonly used as the insulating material in the VDMOS RF packages, which was expensive and highly toxic. Wire bonds were used to provide the source connection from the top of the wafer to the grounded part of the package, introducing an inductive path to ground, severely limiting the gain and the frequency response of the transistors. As a result the commercial success of VDMOS transistor was limited to applications less than 1 GHz.

Another high-power transistor technology that enjoyed commercial acceptance in the 1980s and early 1990s is the high-power bipolar junction transistor (BJT). The state-of-the-art sub-micron, high-power RF BJT technology achieved power levels around 60 W in single-ended common emitter Class AB devices at 2 GHz [5]. The high-power RF BJT suffered from the same drawback as the VDMOS device as the back side of the wafer is the collector terminal, which has to be insulated from the grounded package, and the emitter contact is at the top of the silicon wafer, requiring bondwires from the top of the wafer to the grounding bar of the package. These bondwires introduced a high feedback inductance limiting the gain of the devices. In addition, the mutual coupling between the collector and emitter bondwire arrays made the design of the matching network more challenging. The ruggedness of this sub-micron BJT was quite limited, and special care was needed in the design of bias network to ensure the thermal stability of the device, and prevent thermal runaway.

In the late 1980s and early 1990s, researchers at Motorola's Semiconductor Product Sector, now Freescale Semiconductor, developed a laterally-diffused MOS field effect transistor, or LDMOS FET, to address the inherent limitations of the VDMOS FET and silicon BJT transistor [6–9]. Figure 1.4 shows the LDMOS FET structure, with the source, gate, and drain contacts at the top surface of the wafer. It is desirable to have the source terminal at the back of the die because it can be easily grounded to an electrically and thermally conductive heatsink. The LDMOS FET also has a very low resistance and inductance connection from the source terminal at the top of the wafer to the back side of the wafer, which simplifies the design of the package and eliminates the need for source bondwires.

The LDMOS structure has a number of important advantages over the VDMOS structure. The LDMOS structure has significantly lower parasitic capacitances, by virtue of its structure; this feature results in an extended high-frequency response compared with the vertical structure. The gate-to-source capacitance, C_{gs} , and the drain-to-source capacitance, C_{ds} , have inherently lower values compared with the VDMOS structure. Another

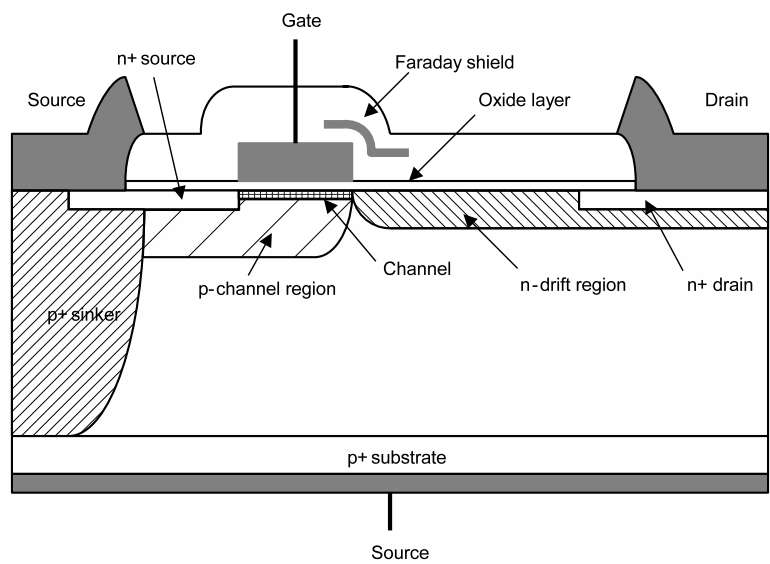


Fig. 1.4. Cross-section of an LDMOS transistor.

important advantage is that the LDMOS FETs can be used in RF integrated circuits, since they have the gate and drain terminals located on the same side of the wafer, enabling the use of microstrip transmission lines, integrated capacitors, resistors and inductors, and so forth, in the design of integrated power amplifiers and circuits.

The commercial predecessor of the LDMOS device was the sub-micron RF BJT, so it is interesting to compare the main performance differences between these two technologies. The LDMOS transistor is more reliable in the field owing to its improved ruggedness over the BJT transistor. For example, it is common for LDMOS transistors to withstand a 10:1 VSWR when driven well past their rated output power without resulting in any damage, whereas the BJT can only accept 3:1 VSWR at rated power before damage to the transistor is observed. Also, LDMOS FETs have a significant gain advantage over the BJTs. One reason for this is that the LDMOS source is connected directly to the package ground plane, avoiding the inductive feedback of the grounding bondwires.

The power saturation characteristic of an LDMOS device is quite different from that of a BJT device. The LDMOS FET has a soft saturation compared with the more abrupt power saturation of the bipolar transistor. This gives the LDMOS FET an advantage when the peak-to-average ratio (PAR) of the stimulus signal is high, avoiding the sudden clipping of the peaks in