Physics of high-frequency operation in Silicon MOSFETs

by

Richard T. Chang

Submitted to the Department of Electrical Engineering and Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Engineering in Electrical Engineering and Computer Science at the Massachusetts Institute of Technology

May 20, 1998

© Copyright 1998 Richard T. Chang. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and distribute publicly paper and electronic copies of this thesis and to grant others the right to do so.

Author

Department of Electrical Engineering and Computer Science

May 20, 1998

Certified by

Professor
Jesús del Alamo
Thesis Supervisor

Accepted by

Arthur C. Smith
Chairman, Department Committee on Graduate Theses
Physics of high frequency operation in silicon MOSFETs

by
Richard Chang

Submitted to the
Department of Electrical Engineering and Computer Science

May 19, 1998

In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

As technology continues to advance, the values of the current-gain cut-off frequency, $f_T$, in silicon MOSFETs are improving. However, the values of the unilateral power gain cut-off frequency, $f_{\text{max}}$, remain relatively low. This is problematic because $f_{\text{max}}$ is important for radio/microwave frequency applications. This thesis describes a Computer-Aided-Design (CAD) environment which assesses $f_{\text{max}}$ in MOS transistors and identifies which parasitics are hindering $f_{\text{max}}$. We have found that, for a state-of-the-art digital n-MOSFET device with a channel length of 0.1 microns, the most influential parasitic component is the gate resistance. Of next critical importance are the output resistance and overlap capacitance. We have also found that silicon-on-insulator (SOI) MOSFETs exhibit degraded high frequency potential due to their floating body. Their lower junction capacitance has a minimal positive impact on $f_{\text{max}}$.

Thesis Supervisor: Jesús del Alamo
Title: Professor
Acknowledgments

This work was performed at the IBM Thomas J. Watson Research Laboratory in Yorktown Heights, NY under the auspices of the VI-A Program. The author would like to acknowledge Dr. Clement Wann for his support and guidance in this project. Dr. Keith Jenkins was instrumental in helping with the high-frequency measurements, and Dr. MeiKei Ieong of the IBM East Fishkill Semiconductor Research and Development Center (SRDC) was extremely helpful with the FIELDAY simulations. Furthermore, the author would like to thank Dr. Yuan Taur for his managerial support, and Dr. Hon-Sum Philip Wong for his continual encouragement.

At MIT, Prof. Jesús del Alamo was instrumental in helping the author present a clear and comprehensible analysis of his work. Jim Fiorenza also helped with some of the initial MATLAB routines.

Finally, the author thanks his family and friends for basically everything else.
Table of Contents

Chapter 1: Introduction ............................................ 10

Chapter 2: The Fundamentals of $f_T$ and $f_{\text{max}}$ in silicon MOSFETs ........ 12
  2.1 Definition of $f_T$ and $f_{\text{max}}$ ........................................ 12
  2.2 High-frequency Measurement ........................................ 14
  2.3 Analytical approximations for $f_T$ and $f_{\text{max}}$ ....................... 15
  2.4 Conclusion ........................................................... 17

Chapter 3: Computer Aided Design (CAD) Framework ....................... 21
  3.1 Two-dimensional high-frequency device simulation .................... 21
  3.2 Extension from 2-D to 3-D ........................................... 22
  3.3 Extraction to a small-signal circuit model ............................ 23
  3.4 Verification of models with respect to frequency dependency .......... 24
  3.5 Verification of models with respect to drain current dependency ......... 27
  3.6 Conclusion ............................................................. 29

Chapter 4: Determining factors of $f_{\text{max}}$ in bulk MOSFETs........ 30
  4.1 Scaling trends of $f_T$ and $f_{\text{max}}$ .................................. 30
  4.2 Description of fabricated device technology ............................ 32
  4.3 Effects of gate resistance ............................................ 33
  4.4 Effects of $R_{\text{out}}$ and $C_{\text{gd}}$ ..................................... 37
  4.5 Verification of the predictive power of analytical expression for $f_{\text{max}}$ .... 41
  4.6 Drain engineering .................................................... 43
  4.7 Conclusion ............................................................. 45

Chapter 5: $f_{\text{max}}$ in bulk and SOI devices ................................. 47
  5.1 Background ............................................................ 47
5.2 Experimental observations ........................................ 47
5.3 Qualitative impact of junction capacitance on $f_{max}$ .................. 48
5.4 Quantitative impact of $C_{db}$ ........................................ 50
5.5 Impact of junction capacitance on output impedance ................... 52
5.6 Experimental Y-parameters ....................................... 54
5.7 Introduction of simulation to explain $Y_{22}$ disparity .................. 57
5.8 The impact of the floating body effect ................................ 58
5.9 Conclusions regarding SOI devices ................................ 63

Chapter 6: Conclusion ....................................................... 63

Chapter 7: References ....................................................... 65
List of Figures

1  2-port representation of device ................................... 12
2  Graphical depiction of current gain of a 2-port network .......... 13
3  Dependence of current gain and unilateral power gain on frequency of a fabricated device .............................................. 15
4  Equivalent circuit model for derivation of $f_T$ and $f_{max}$ formulas ........ 16
5  Block diagram denoting fundamental aspects of $f_T$ and $f_{max}$ ......... 17
6  Block diagram denoting detailed simulation covered in Chapter 3 ........ 20
7  Experimental and corresponding simulated FIELDAY IV characteristics ... 21
8  Modification of Y-parameter matrix to include distributed gate resistance with $n$ partitions ..................................................... 22
9  $f_{max}$ versus number of gate partitions in distributed model ........ 23
10 Small-signal equivalent circuit model with distributed gate resistance containing $n$ partitions ..................................................... 24
11 Simulated FIELDAY and corresponding small-signal equivalent circuit model Y-parameters ..................................................... 25
12 $f_T$ and $f_{max}$ for devices depicted in Fig. 11 ........................................... 25
13 Experimental and corresponding simulated FIELDAY and MATLAB model Y-parameters ..................................................... 26
14 $f_T$ and $f_{max}$ for devices depicted in Fig. 13 ........................................... 27
15 Comparision of experimental and simulated $f_T$ dependence on drain current ..................................................... 28
16 Comparision of experimental and simulated $f_{max}$ dependence on drain current ..................................................... 28
17 Experimentally reported number for $f_T$ for silicon MOSFETs, since 1991 ... 31
18 Experimentally reported number for $f_{max}$ for silicon MOSFETs, since 1991 . 31
19 Simulated $f_T$ and $f_{max}$ dependence on $L_g$ ........................................... 32
20 Cross-sectional view of NMOS device ........................................... 33
21 Experimental $I_{ds}$ vs. $V_{ds}$ ........................................... 33
22 Simulated dependence of $f_T$ and $f_{\text{max}}$ on gate resistance. ................ 35
23 Simulated dependence of $f_T$ and $f_{\text{max}}$ on width. ......................... 36
24 Simulated $f_{\text{max}}$ dependence on $R_{\text{out}}$. ............................... 38
25 Simulated $f_{\text{max}}/f_T$ dependence on $R_{\text{out}}$. ................................. 39
26 Simulated $f_{\text{max}}$ dependence on $C_{gd}$. .................................. 39
27 Simulated $f_{\text{max}}/f_T$ dependence on $C_{gd}$ .................................. 40
28 Simulated trade-off between $R_{\text{out}}$ and $C_{gd}$ for constant $f_{\text{max}}$ using equivalent circuit model (Fig. 10) ........................................... 40
29 Simulated trade-off between $R_{\text{out}}$ and $C_{gd}$ for constant $f_T$ using equivalent circuit model (Fig. 10) ................................................................. 41
30 Simulated trade-off between $R_{\text{out}}$ and $C_{gd}$ for constant $f_{\text{max}}/f_T$ using equivalent circuit model (Fig. 10) .................................................. 42
31 Side-by-side comparisons for $f_T$ and $f_{\text{max}}$ dependence on $R_g$. .......... 42
32 Side-by-side comparisons for $f_T$ and $f_{\text{max}}$ dependence on $R_{\text{out}}$ with $C_{gd}$ as a parameter ................................................................. 43
33 Simulated comparison of $f_T$ for bulk, LDD, and DMOS structures, normalized by drain current ................................................................. 44
34 Simulated comparison of $f_{\text{max}}$ for bulk, LDD, and DMOS structures, normalized by drain current ................................................................. 44
35 Cross-sectional view of DMOS structure (not to scale). .............................. 45
36 Simulated comparison of $f_{\text{max}}$ vs. $f_T$ for bulk, LDD, and DMOS structures, normalized by drain current ................................................................. 46
37 Comparison of experimental $f_T$ and $f_{\text{max}}$ of bulk and SOI devices with respect to drain current ................................................................. 48
38 Circuit model which includes parasitics source, gate, and drain resistances... 49
39 Final circuit model with lumped gate resistance. ...................................... 50
40 $f_T$ and $f_{\text{max}}$ dependence on $C_{db}$, with SOI and bulk values labelled. ..... 51
41 $f_T$ and $f_{\text{max}}$ dependence on $R_{\text{sub}}$, with SOI and bulk values labelled .... 51
42 $Y_a$ and $Y_b$ partitions ................................................................. 52
43 Experimental device Y-parameters. ...................................................... 54
44 Comparison of bulk experimental and extracted circuit model Y-parameters.. 55
Comparison of SOI experimental and extracted circuit model Y-parameters.  . 56
Comparison of Y-parameters from extracted circuit models for Bulk and SOI devices. ........................................................ 56
Comparison of simulated FIELDAY Bulk and SOI model Y-parameters. . . . . 58
Comparison of simulated $f_T$ of bulk and SOI devices with respect to drain current ................................................................. 59
Comparison of simulated $f_{max}$ of bulk and SOI devices with respect to drain current ................................................................. 59
2-port representation of an SOI device with a ground contact just above the buried oxide layer. ..................................................... 60
2-port representation of a floating-body SOI device. .......................... 60
2-port representation of a SOI device with a forward-biased body. .............. 61
A closer look at the drain region of an SOI device. .............................. 61
The $Y_{22}$ parameters for various FIELDAY simulations ....................... 63
List of Tables

1 Conversion from S-parameters to Y-parameters for a 2-port network . . . . . 13
2 Summary of simulation results for similar drain currents, at $V_{ds} = 1.5$ V . . . 63
Chapter 1: Introduction

Silicon is the most popular material for today’s integrated circuits because of its low cost and wide spread use. It has traditionally been considered to be too slow for applications operating in the Gigahertz frequency range. Thus, faster, yet more expensive, substrates, such as Gallium Arsenide, have been necessary for high-frequency applications. A good example of this need is that cellular phones utilize silicon chips for the bulk of the signal processing circuitry, and use Gallium Arsenide chips for the high-speed transmission and reception of signals. However, as silicon MOSFETs become faster due to device scaling, they may be a viable solution to high-frequency applications in Gigahertz frequency range.

Device scaling has certainly improved the operating speed capability of silicon devices. One indication of this advancement is the improving unity current gain cut-off frequency, \( f_T \). This is a measure of the high-frequency current amplification of the device. This number is often accompanied by the maximum frequency of oscillation, \( f_{\text{max}} \), which is a measure of the high-frequency power gain. Together, these two figures of merit provide circuit designers with a description of the high-frequency potential of a device.

Currently, we have silicon MOSFETs with good \( f_T \) (~100 GHz), but with much lower \( f_{\text{max}} \) (~30 GHz) [1]. The high \( f_T \)'s show good promise for radio and microwave frequency applications, but circuit designers also want high \( f_{\text{max}} \) values. While \( f_T \) is largely predetermined by the limits on gate length and oxide thickness of the device, \( f_{\text{max}} \) can vary greatly depending on a greater range of factors that include layout geometry and parasitic elements. It is the goal of this project to investigate methods of improving \( f_{\text{max}} \) in silicon MOSFETs.

In this thesis, we begin, in Chapter 2, by exploring the meaning behind \( f_T \) and \( f_{\text{max}} \). The chapter also discusses how to experimentally gather these measurements. Additionally, one may estimate these numbers from the device parameters. In the following chapter, we examine the software tools that allow us to obtain the relevant high-frequency information from a variety of devices.
Using that solid foundation, we are ready to evaluate the high-frequency performance of silicon MOSFETs by examining the $f_{\text{max}}$ of a device. In Chapter 4, I will discuss the relative influence of various device parameters on $f_{\text{max}}$. In addition, I will examine the current trends in MOSFETs and suggest what changes may become necessary to enhance the $f_{\text{max}}$ of silicon device. The following chapter explores an interesting application of this $f_{\text{max}}$ analysis. We discuss the relative high frequency performances of bulk and SOI devices. In this context, the importance of the junction capacitance is evaluated. Finally, the major points of this thesis will be recapped in Chapter 6.
Chapter 2: The Fundamentals of $f_T$ and $f_{\text{max}}$ in silicon MOSFETs

$f_T$ and $f_{\text{max}}$ are two widely used figures of merit for high-frequency characterization of devices. This chapter will begin by explaining the significance of these two parameters. Then, a method of calculating these metrics from S-parameters measurements will be described. Finally, a small-signal equivalent circuit model of a MOSFET will be introduced. This will enable us to look critically at different device parameters and evaluate their impact on $f_T$ and $f_{\text{max}}$.

2.1 Definition of $f_T$ and $f_{\text{max}}$

$f_T$ and $f_{\text{max}}$ are small-signal figures of merit that provide information about the suitability of a device for high frequency operation. They can be extracted from S-parameter measurements performed using a network analyzer. This instrument measures the two-port scattering parameters (S-parameters) of a device using the gate-source terminals as the input port and the drain-source terminals as the output port (Fig. 1).

![Fig. 1: 2-port representation of device.](image-url)

The S-parameters can be converted into admittance parameters (Y-parameters) using the formulas given in Table 1, where $Y_{ij}$ is the 2-port admittance matrix parameter and $S_{ij}$ is the 2-port scattering matrix parameter, with index $i$ indicating the output port and index $j$ indicating the input port. $Y_0$ is the characteristic admittance of the wires of the measuring equipment [2]. It is often easier to think in terms of Y-parameters rather than S-paramete-
ters because Y-parameters are direct relations between the current and voltage at the ports, whereas S-parameters are mathematical manipulations that are easier to directly measure.

Table 1: Conversion from S-parameters to Y-parameters for a 2-port network

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{11}$</td>
<td>$\frac{(1 - S_{11})(1 + S_{22}) + S_{12}S_{21}}{(S_{11} + 1)(S_{22} + 1) - S_{12}S_{21}}$</td>
</tr>
<tr>
<td>$Y_{12}$</td>
<td>$\frac{-2S_{12}}{(S_{11} + 1)(S_{22} + 1) - S_{12}S_{21}}$</td>
</tr>
<tr>
<td>$Y_{21}$</td>
<td>$\frac{-2S_{21}}{(S_{11} + 1)(S_{22} + 1) - S_{12}S_{21}}$</td>
</tr>
<tr>
<td>$Y_{22}$</td>
<td>$\frac{(1 + S_{11})(1 - S_{22}) + S_{12}S_{21}}{(S_{11} + 1)(S_{22} + 1) - S_{12}S_{21}}$</td>
</tr>
</tbody>
</table>

Otherwise, the current gain with the output port shorted to ground can be calculated by using the following formula, which is in terms of S-parameters:

$$h_{21} = \frac{-2S_{21}}{(1 - S_{11})(1 + S_{22}) + S_{12}S_{21}}$$

This is graphically depicted in Figure 2. In MOSFETs, at high frequencies and when the gain is greater than one, the current gain is inversely proportional to frequency. $f_T$ is defined as the frequency at which the current gain is unity or 0 dB.

![Fig. 2: Graphical depiction of current gain of a 2-port network.](image)

In a similar fashion, the unilateral power gain is defined with the input and output ports conjugately matched and the feedback parameter neutralized to zero. It may be computed by using the following formula [2]:

$$U = \frac{\frac{1}{2}|S_{21}/S_{12} - 1|^2}{k|S_{21}/S_{12}| - Re\{S_{21}/S_{12}\}}$$

where $k$ is the stability factor, defined as:
Much like the current gain, in MOSFETs, at high frequencies, this gain is also inversely proportional to frequency, and $f_{\text{max}}$ is defined as the frequency where the unilateral power gain becomes unity.

These are ideal circumstances; devices rarely operate under these conditions when implemented into a circuit. However, these benchmarks establish the high-frequency potential of a device and are useful in giving circuit designers an idea of the maximum frequency capability of devices.

2.2 High-frequency Measurement

Measurement of $f_T$ and $f_{\text{max}}$ is by no means trivial. At high frequencies, the parasitics of the experimental setup are large enough to become significant. Only by using specialized high-frequency probes and calibration techniques can the influence of these parasitics be minimized.

To measure the S-parameters of a single device, a probe station is equipped with microwave probes that are connected to a network analyzer (HP8720) and a semiconductor parametric analyzer (HP4145B). The 4145B provides the bias voltages, while the 8720 measures the S-parameters over a wide range of frequencies (100 MHz to 20 GHz). Then, a single device, outfitted with contact pads, can be individually tested. The measurements are then de-embedded from the parasitic effects of the probes and contact pads, by means of calibration, so that information about the intrinsic device may be obtained. This involves measuring data for an open test site in which the metal pads are present, but there is no actual device [3]. Even still, if these parasitics are too large, then the cancellation process becomes ineffectual and the subsequent measurements tend to be very noisy. It is best to minimize these parasitics when initially designing the device layout.

An example of these measurements, shown in Fig. 3, is plotted, in general, on a semi-log scale. For each metric, the data is extrapolated to 0 dB, since the equipment is not capable of reaching the frequency where the gain is unity. This extrapolated line should have a slope close to -20 dB per decade of frequency. At the selected bias point, this particular device has an $f_T$ of approximately 120 GHz and an $f_{\text{max}}$ of approximately...
70 GHz.

\[
\begin{align*}
V_{gs} &= 0.7 \text{ V} \\
V_{ds} &= 1.5 \text{ V} \\
L_g &= 0.09 \mu\text{m} \\
W_g &= 5 \mu\text{m} \times 20 \text{ fingers}
\end{align*}
\]

Fig. 3: Dependence of current gain and unilateral power gain on frequency of a fabricated device.

2.3 Analytical approximations for \( f_T \) and \( f_{\text{max}} \)

It is useful to extract an AC equivalent circuit model when characterizing a device technology. Then, one can derive an analytical expression for \( f_T \) and \( f_{\text{max}} \) by writing the two-port S-parameters in terms of circuit elements. This is accomplished by applying the S-parameter formulas (Eqs. 1 and 2) to the circuit model and solving for the case when the gain is equal to unity. After simplifying the equations by making suitable approximations, one arrives at a useful analytical first-order approximation for \( f_T \) and \( f_{\text{max}} \). Figure 4 shows such an equivalent circuit model that incorporates circuit elements and may be used as the basis for deriving an approximate analytical expression.

These estimates provide valuable insight into this figure of merit. In terms of equivalent small-signal circuit model elements, the unity current gain is given by [4]:

Chang 15
\[ f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})}, \]  

where \( g_m \) is the transconductance, \( C_{gs} \) is the gate-source capacitance, and \( C_{gd} \) is the drain-gate capacitance. Similarly, to the first order, \( f_{\text{max}} \) is given analytically by [4-7]:

\[ f_{\text{max}} \approx \frac{f_T}{2 \sqrt{\frac{R_g + R_s + R_i}{R_{out}}} + 2\pi f_T C_{gd} R_g}, \]

where \( R_{out} \) is the output resistance, and \( R_g, R_s, \) and \( R_i \) are the gate, source, and channel resistances, respectively.

Fig. 4: Equivalent circuit model for derivation of \( f_T \) and \( f_{\text{max}} \) formulas.

These formulas also prove to be very useful in examining more complex circuit models of a MOS device, as we shall see in succeeding chapters. In the process of evaluating the relative importance of each parameter, we will then also assess the accuracy of these formulas.

Figure 5 shows a block diagram of the basic strategy of how one can start with a device and characterize it by measuring \( f_T \) and \( f_{\text{max}} \). Next, one can extract an equivalent circuit model and obtain \( f_T \) and \( f_{\text{max}} \) as a function of various device parasitic circuit ele-
ments. These analytical equations can then be used as aids in device characterization and design. For example, one may be able to pinpoint the strengths and weaknesses of silicon MOSFETs in high-frequency applications to specific device attributes, such as output resistance or junction capacitance, rather than more obscure, less intuitive, S-parameters.

\[ f_{T}, f_{\text{max}} = f(R_g, C_{gd}, R_{out}, \ldots) \]

Fig. 5: Block diagram denoting fundamental aspects of \( f_T \) and \( f_{\text{max}} \).

2.4 Conclusion

In this chapter, we have established some basic ideas regarding \( f_T \) and \( f_{\text{max}} \). First, the S-parameters are measured using high-frequency equipment and calibration techniques. From S-parameters, the unity current gain and the unilateral power gain may be extracted, so that \( f_T \) and \( f_{\text{max}} \) can then be extrapolated. In addition, based on these S-parameters, an equivalent circuit model of a device may be extracted. This model helps to distinguish how different aspects of a device affect \( f_{\text{max}} \).

In the next chapter, I will introduce a simulation environment that allows for esti-
mation of $f_T$ and $f_{\text{max}}$ based on a physical description of the MOSFET. In addition, I will expound upon the circuit extraction process.
Chapter 3: Computer Aided Design (CAD) Framework

In order to determine which characteristics of a device are key to obtaining a high $f_{\text{max}}$, it would be ideal to have tens of thousands of fabricated test devices with varying widths, channel lengths, and other relevant device parameters. However, fabricating such a wide array of individual test devices is expensive. Additionally, the turnaround time between design and fabrication of a chip is on the order of months. A quicker, more practical, and more economical method of analyzing and designing devices is to use computer simulation tools. This chapter provides a detailed illustration of the tools used in this evaluation. Since three-dimensional device level simulations are difficult and computationally very expensive, the modeling is broken into two parts: first, the two-dimensional cross-sectional behavior of the device is simulated, and second, the key three-dimensional parasitic effects are added. Afterwards, a small-signal equivalent circuit is extracted. This simulation framework will be validated by comparison with experimental data.

There are many other compelling reasons to use a Computer-Aided-Design (CAD) environment. It provides a convenient way to examine devices without many of the complications of dealing with actual hardware. It also eliminates doubt and ambiguity due to processing variations; the doping profile can be known with certainty, the gate oxide thickness is uniform and without defects, and the channel length can always be ascertained. These traits are even more beneficial when making comparisons between devices. For example, one can evaluate devices with different channel lengths in a controlled environment.

However, the simulator cannot replace measurements of actual fabricated devices. After all, the models are far from perfect. The mobility modeling is especially suspect; in my experience, it often does not have the proper vertical electric-field dependency. Finally, the device is destined for a circuit, where processing variations are inevitable.

This chapter fills in the details of the process flow that was outlined in the last chapter (Fig. 5). Of noteworthy importance is the introduction of CAD tools that allow a device to be described by its cross-sectional physical characteristics. Then, its electrical
Fig. 6: Block diagram denoting detailed simulation covered in Chapter 3.

Characteristics can be simulated. Additionally, the critical parasitics of the third dimension can be extracted.
sion are added into the model. These virtual devices may be seemlessly substituted in the place of fabricated devices in our examination (Fig. 6). Following that, from these results, an equivalent circuit model can be extracted so that the trends of $f_{\text{max}}$ may be estimated with respect to the various device parameters.

3.1 Two-dimensional high-frequency device simulation

To aid in this analysis of $f_T$ and $f_{\text{max}}$, I will be using “FIELDAY,” a simulator that solves Poisson’s equation and preserves current continuity to model the two-dimensional electron transport within a device. This basically involves using a drift-diffusion approximation. This powerful software tool allows for a specification of different device structures with unique doping profiles. FIELDAY can perform DC simulations to obtain I-V characteristics, as well as perform AC simulations to generate small-signal data, such as the conductance and capacitance characteristics for each terminal. These results may then be compiled to generate the two-port admittance parameters.

In general, we will be modeling n-type MOSFETs with gate lengths of 0.1 $\mu$m, of which we have experimental data to compare against. Figure 7 shows FIELDAY simulation current-voltage characteristics plotted along with those of a experimental device. The two match fairly well. The fabricated devices, upon which the models are based, will be described in greater detail in the next chapter.

![Fig. 7: Experimental and corresponding simulated FIELDAY IV characteristics.](image-url)
3.2 Extension from 2-D to 3-D

FIELDAY produces two-port small-signal data based on a two-dimensional cross-sectional view of a device. For evaluating the operation of the device under DC conditions, the results can just be multiplied by the appropriate width. For analyzing the device under AC conditions, we need to include the gate resistance in the model. To do this, MATLAB, a software package used for mathematical manipulation, is utilized to extend the device into the width direction with the proper gate resistance.

First, the Y-parameter matrix is scaled according to the width of the device. Then, the gate resistance can be lumped into a discrete element and added to the two-port network obtained from the device simulator, by the method described in [6]. A better approach is to more accurately model the device by distributing the gate resistance, as shown in Fig. 8 [7]. Then, new two-port parameters may be taken from this 3-D device to calculate $f_T$ and $f_{\text{max}}$.

![Diagram of Y-parameter matrix modification with gate resistance distributed](image)

**Fig. 8:** Modification of Y-parameter matrix to include distributed gate resistance with $n$ partitions.

Figure 9 shows the impact of distributing the gate resistance on $f_{\text{max}}$ and $f_T$. At ten gate partitions, $f_{\text{max}}$ reaches approximately 95% of its final value. On the other hand, $f_T$ is unchanged with respect to the number of gate partitions since it is independent of the gate
resistance. All simulation results reported here use a distributed gate resistance model, when appropriate.

![Graph showing f_max versus number of gate partitions in distributed model.]

Fig. 9: f_max versus number of gate partitions in distributed model.

3.3 Extraction to a small-signal circuit model

To aid in device design, a small signal equivalent circuit model can be created from a set of device Y-parameters (Fig. 10). EESOF Series IV Communications Suite, a software program, is used to compute elemental values of the circuit so that its Y-parameters closely match those of either a FIELDAY simulated device or an experimental device. This optimization is performed in the range of 1 to 20 GHz and must be individually computed for each desired bias point. The model then may be subjected to a sensitivity analysis to determine which parameters dominate the behavior of f_max and f_T. This will be useful in assessing what changes to the device, if any, are necessary for high-frequency operation.

In matching the circuit to the simulation data, we achieve a great degree of success (Figs. 11, 12). Overall, the parameters match to within 5%. Similar results are obtained when matching the circuit Y-parameters to those from an experimental device. This will be shown in the chapter 5, when discussing the differences between bulk and SOI devices.
3.4 Verification of models with respect to frequency dependency

To verify that the FIELDAY simulations accurately reflect device characteristics, we compare its Y-parameters with those taken from an actual device (Figs. 13, 14). The fabricated device has an effective channel length of approximately 0.09 μm and consists of 10 fingers, each 10 μm long. The gate and drain terminals are both biased at 1.8 V. Given the difficulty of estimating the doping profile, the model does a reasonable job of matching to the experimental results.

The major discrepancies exist in comparing the real parts of $Y_{21}$ and $Y_{22}$ ($\text{Re}\{Y_{21}\}$ and $\text{Re}\{Y_{22}\}$, respectively). Since $\text{Re}\{Y_{21}\}$ is related to $g_m$, this difference can be attributed to differences in channel length between the model and the actual device. The simulated device has a shorter effective channel length and, therefore, a higher $g_m$. $\text{Re}\{Y_{22}\}$ is related to the output conductance ($1/R_{out}$). The simulated device has a higher output resistance, which is most likely due to the inaccurate model of the doping profile. Overall, the model yields an $f_T$ which is within 10% and an $f_{max}$ which is about 25% higher than experimentally seen.
Fig. 11: Simulated FIELDAY and corresponding small-signal equivalent circuit model Y-parameters.

Results:

- $f_{T,\text{sim}} = 84 \text{ GHz}$
- $f_{\text{max, sim}} = 87 \text{ GHz}$
- $f_{T,\text{circuit}} = 81 \text{ GHz}$
- $f_{\text{max, circuit}} = 85 \text{ GHz}$

Fig. 12: $f_T$ and $f_{\text{max}}$ for devices depicted in Fig. 11.
Fig. 13: Experimental and corresponding simulated FIELDAY and MATLAB model Y-parameters.

Fig. 14: $f_T$ and $f_{\text{max}}$ for devices depicted in Fig. 13.
This reasonable correlation of simulation to experimental results partially validates the FIELDAY model. The width and distributed gate resistance construction will be compared to experimental width dependencies in the following chapter. There, we will show that this modeling is also reasonably accurate.

3.5 Verification of models with respect to drain current dependency

Furthermore, we can examine the behavior of $f_T$ and $f_{\text{max}}$ with respect to different bias points. Figure 15 shows experimentally measured $f_T$ values as a function of increasing drain current. For this measurement, the drain potential was held at 1.5 V, while the gate was ramped from 0 V to 1.5 V. These are different devices than those shown in Figures 13 and 14.

There are a few salient features to these curves. First of all, the peak $f_T$ occurs just above the threshold voltage, where the maximum $g_m$ is. This peak occurs here despite the fact that this is a region where the gate-source capacitance is relatively high, since the device is at the onset of inversion. This indicates that $f_T$ is more sensitive to $g_m$ with changing $V_{gs}$. After this crest, $f_T$ gently roll-offs with increasing gate voltage. This is explained by noting that $g_m$ is decreasing due to mobility degradation caused by the increasing vertical electric field.

Figure 15 also shows the simulated drain current dependence of $f_T$. $f_T$ is lower than measured; however, all of the same trends are present, if not exaggerated in the simulation. By using a vertical-field dependent mobility model, we observe a peak in $f_T$. On the other hand, there is no peak in $f_T$ if a constant mobility model, which is independent of vertical electric field, is substituted into FIELDAY. Instead, $f_T$ monotonically increases with higher gate voltages without levelling out.

$f_{\text{max}}$ exhibits a slightly different behavior, though similar in shape (Fig. 16). Since it is directly proportional to $f_T$, this is not unexpected. The major difference is that, due to a small Early voltage, the output resistance varies over different gate biases, meaning that $R_{\text{out}}$ increases with decreasing gate voltage. Therefore, the peak $f_{\text{max}}$ should occur at low gate voltages. From the experimental results, we see that the peak occurs at a lower drain current bias than does the maximum $f_T$. Again, FIELDAY confirms this behavior; with
the constant mobility model, there is no peak $f_T$, yet the peak $f_{\text{max}}$ still occurs at low bias currents.

Fig. 15: Comparison of experimental and simulated $f_T$ dependence on drain current.

Fig. 16: Comparison of experimental and simulated $f_{\text{max}}$ dependence on drain current.
3.6 Conclusion

This chapter presented a simulation environment that begins with a structural description of a device and ends with an equivalent circuit model of the device. This encompasses a device simulation and a mathematical implementation of a distributed gate resistance. The resulting Y-parameters are then used to calibrate a circuit model. Now, we are finally in a position to investigate how individual parasitic components of a device affect $f_{\text{max}}$. This is the basis for examining ways of increasing $f_{\text{max}}$. 
Chapter 4: Determining factors of f_{max} in bulk MOS-FETs

In this chapter, we will quantify the influence of the various device parameters on \( f_{\text{max}} \). We will make use of experimental data of sub-tenth micron NMOS devices and a small-signal equivalent circuit model derived from device simulations (with estimated gate resistance).

We will begin by examining where, in terms of \( f_T \) and \( f_{\text{max}} \), device scaling has brought us and what improvements we may expect in the future. Then, we examine the width dependency of \( f_T \) and \( f_{\text{max}} \). Finally, we try to narrow the gap in values between \( f_T \) and \( f_{\text{max}} \) by adjusting the gate resistance, overlap capacitance, and output resistance of the MOSFET.

4.1 Scaling trends of \( f_T \) and \( f_{\text{max}} \)

Primarily because of device scaling, we are in a position to consider the silicon MOS transistor as a viable candidate for high-frequency operation. Its high \( f_T \)'s of over 100 GHz for the tenth-micron generation of devices certainly show promise for operation in the Gigahertz range. However, its relatively low \( f_{\text{max}} \) is a concern for 0.18 \( \mu \text{m} \) technology [8].

Will this problem solve itself as we continue to scale devices further? As shown experimentally in [9], \( f_T \) is inversely proportionally to the square of the channel length \( (L_g) \) for longer channel lengths, but this dependency begins to saturate as \( L_g \) approaches 0.1 \( \mu \text{m} \). Figures 17 and 18 show reported values of \( f_T \) and \( f_{\text{max}} \) for n-MOSFETs, dating back to 1991 [5, 8, 10-24]. As channel lengths get smaller, \( f_T \) is increasing. On the other hand, \( f_{\text{max}} \) is not improving. Clearly, \( f_T \) is not the problem; people seem to know how to make silicon devices with high \( f_T \)'s. Therefore, we will focus our attention on understanding and improving \( f_{\text{max}} \) in silicon MOSFETs.
Fig. 17: Experimentally reported numbers for $f_T$ for silicon MOSFETs, since 1991.

Fig. 18: Experimentally reported numbers for $f_{\text{max}}$ for silicon MOSFETs, since 1991.
Our device simulations show the scaling trends which correspond to what is experimentally seen. For these simulations, only the gate length is scaled; all other features, such as the gate oxide and threshold voltage, are kept constant. The general $1/L_g^2$ dependency of $f_T$ is reproduced for long gate lengths. Furthermore, the simulations indicate that the curve will deviate slightly from this trend near channel lengths of 0.1 μm. It should approach a $1/L_g$ dependency due to velocity saturation effects. Figure 19 indicates that $f_{\text{max}}$ is inversely proportional to $L_g$ at longer channel lengths and dramatically saturates at the tenth-micron regime. The dissimilar dependencies on $L_g$ indicate that as we scale devices beyond 0.18 μm, the problem of low $f_{\text{max}}$ will only continue to get worse. Therefore, we will target a 0.1 μm NMOS device and examine which parasitic elements of the transistor are primarily responsible in determining $f_{\text{max}}$.

![Figure 19: Simulated $f_T$ and $f_{\text{max}}$ dependence on $L_g$.](image)

4.2 Description of fabricated device technology

The technology available for analysis was an NMOS process with an effective gate length of approximately 0.1 μm and a gate oxide that is 27 Å thick. The gate, source, and
drain regions were silicided using a self-aligned process (Fig. 20). In addition, shallow trench isolation (STI), retrograde channel doping, and shallow junction implants are employed. A typical I-V curve for one of these devices is shown in Fig. 21.

![Cross-sectional view of NMOS device](image)

**Fig. 20:** Cross-sectional view of NMOS device.

![Experimental I<sub>ds</sub> vs. V<sub>ds</sub>](image)

**Fig. 21:** Experimental I<sub>ds</sub> vs. V<sub>ds</sub>.

4.3 Effects of gate resistance

Since the improvement in f<sub>T</sub> outpaces that of f<sub>max</sub> as devices get smaller, low f<sub>max</sub> will continue to be a problem at shorter lengths. We look to the analytical expression of
f_{\text{max}} (Eq. 5) for methods of improving f_{\text{max}}. First, this equation can be simplified by realizing that, oftentimes, in a MOSFET, $R_g \gg R_s + R_i$ (Eq. 6). Then, the following expression is valid:

$$\begin{align*}
\frac{f_{\text{max}}}{f_T} &\approx \left. \frac{R_g}{2\sqrt{R_{\text{out}}} + 2\pi f_T C_{\text{gd}} R_g} \right|_{R_g \gg R_s + R_i} \\
\end{align*}$$

For example, a 0.1 μm device with a width of 5 μm and titanium silicide on its gate (resistivity approximately equal to 10Ω/□) has an effective lumped gate resistance of about 200 Ohms, a source resistance of 40 Ohms, and a channel resistance of 1 Ohm. Under these conditions, Eq. 6 is a good approximation for $f_{\text{max}}$.

This equation reveals that the gate resistance plays a very important role in determining $f_{\text{max}}$. Explicitly, this means that:

$$f_{\text{max}} \propto \left. \frac{1}{\sqrt{R_g R_s + R_i}} \right|_{R_g \gg R_s + R_i}$$

Therefore, the use of different gate materials will have a direct impact on $f_{\text{max}}$. Different simulations with silicides of 2, 10, and 25Ω/□ are indicated on the graph in Fig. 22. By switching from titanium silicide (10Ω/□) to tungsten silicide (2Ω/□), we can theoretically gain a $1/\sqrt{0.2} = 2.2x$ improvement in $f_{\text{max}}$. With an even lower resistivity, a metal gate has the potential to be of even greater benefit.

Another simple method of reducing the gate resistance is to decrease the width of the device. By doing this, we will also decrease other parasitics, namely the overlap capacitance and output conductance ($1/R_{\text{out}}$), that detrimentally affect $f_{\text{max}}$. On the other hand, $f_T$ is relatively independent of width since the width dependence of $g_m$, $C_{\text{gs}}$, and $C_{\text{gd}}$ cancel one another in Eq. 4. For appreciable widths where $R_g \gg R_s + R_i$, $f_{\text{max}}$ is inversely proportional to the width (Fig. 23). This can be inferred from Eq. 6 by noticing that, at a given bias point, $R_g$ and $C_{\text{gd}}$ are proportional to width, $R_{\text{out}}$ is inversely proportional to width, and $f_T$ is independent of width. This suggests that both terms in the denominator of
Eq. 6 vary as $W_g^2$, so that overall, $f_{\text{max}}$ is inversely proportional to $W_g$.

Scaling the width will decrease the amount of drain current, which may be undesirable. Therefore, one can instead use multiple fingers, each having a shorter finger length, and conserve the total width. If we assume that each finger is identical and that the overhead wiring parasitics are negligible, then $f_{\text{max}}$ is only dependent on the parasitics of a single finger. This can be shown if we associate a two-port Y-parameter network with each finger and place the two-port networks in parallel with one another. This is equivalent to the Y-parameters of a single finger that are scaled by the number of fingers. After converting to S-parameters, this scaling factor then essentially drops out when computing $f_{\text{max}}$ using Eq. 2.

The $f_T$ and $f_{\text{max}}$ width dependencies are confirmed by experimental data comparing a device having 20 fingers, each 5 μm wide, against a device consisting of 10 fingers, each 10 μm wide (Fig. 23). Both transistors have similar $f_T$'s of roughly 120 GHz, but the

![Fig. 22: Simulated dependence of $f_T$ and $f_{\text{max}}$ on gate resistance.](image-url)
20 x 5 μm device has a $f_{\text{max}}$ of 70 GHz, while the 10 x 10 μm device has a $f_{\text{max}}$ of 35 GHz. Furthermore, by using a 3 μm wide device, simulations predict that one can obtain a device with both $f_T$ and $f_{\text{max}}$ that are approximately equal at over 100 GHz.

But, one cannot push $f_{\text{max}}$ infinitely high by using a very short width; in the limit of width reduction, $f_{\text{max}}$ levels off because $R_g$ is no longer much greater than $R_s + R_i$. According to Eq. 5, as the width decreases and the value of the gate resistance approaches that of the source resistance, the source resistance becomes the limiting factor. More explicitly,

$$
\lim_{W \to 0} f_{\text{max}} = \frac{f_T}{2} \sqrt{\frac{R_{\text{out}}'}{R_s'}}
$$

where $R_s'$ and $R_{\text{out}}'$ denote the source and output resistances per unit width, respectively.

Fig. 23: Simulated dependence of $f_T$ and $f_{\text{max}}$ on width (with experimental data points labeled).

For example, Figure 23 also demonstrates that our fabricated devices, which have an approximate gate resistivity of $10\Omega/\square$, should not show any further improvement in
f_{\text{max}} by decreasing the width below 0.5 microns. This breakpoint occurs at different widths, according to the resistivity of the gate material.

This poses somewhat of a problem; for applications that require large amounts of currents, wide devices are used; however, as we have seen, this is harmful to f_{\text{max}}. To a certain extent, one may, as mentioned before, use many parallel fingered devices. Then, the problem becomes one of creating a layout that is free of timing problems and minimizes wiring parasitics.

In addition to a practical limit on the amount of current, there is a constraint on the maximum drain voltage that may be used. A voltage placed across the gate oxide which is too large will induce breakdown. This breakdown voltage for silicon MOSFETs is low; for a 0.1 \mu m device, the maximum voltage that can be placed across the oxide without permanent damaging it is around 2 V. This will limit the power performance of the device. Bipolar and GaAs devices, with their higher breakdown voltages, have an advantage in this respect. These results suggest that silicon MOS devices may be more appropriate for low-power/low-noise applications.

In light of this, drain engineering which increases the breakdown voltage of the device may prove beneficial. For example, one may make use of a DMOS structure which increases the drain resistance to accommodate higher voltages. The advantages and disadvantages of this structure will be discussed later in this chapter.

### 4.4 Effects of R_{\text{out}} and C_{gd}

Following the gate resistance, the most important factors of f_{\text{max}} are the output resistance and the overlap capacitance. This is readily seen upon examination of Eq. 6; assuming that R_g >> R_s + R_i and then factoring out R_g leaves two distinct terms in the denominator: one is 1/R_{\text{out}} and the other includes C_{gd}. This second term may be simplified, by substitution of Eq. 4 for f_T:

\[
2\pi f_T C_{gd} = g_m \frac{C_{gd}}{C_{gs} + C_{gd}}
\]

Then, we can compare these terms to find out which is larger and will dominate the behavior of f_{\text{max}}. We can focus our attention on R_{\text{out}} and C_{gd}, since C_{gs} and g_m are largely fixed.
by the channel length.

As the output resistance increases from 0 to 2 kΩ, $f_T$ increases by about 50% and $f_{\text{max}}$ increases by about a factor of three from 20 GHz to 60 GHz (Fig. 24). In this figure, it is also apparent that even though $f_T$ is ideally independent of $R_{\text{out}}$, as the value of $R_{\text{out}}$ approaches that of the drain resistance ($R_d$), $f_T$ rolls off. This is due to the fact that $R_{\text{out}}$ begins to shunt an appreciable amount of current from the drain terminal. Though $f_{\text{max}}$ is roughly proportional to $f_T$, it is not solely because of the change in $f_T$ that $f_{\text{max}}$ is dependent on $R_{\text{out}}$. Normalizing $f_{\text{max}}$ with $f_T$ by plotting $f_{\text{max}}/f_T$ against the output resistance, we see that $f_{\text{max}}$ has its own dependency on $R_{\text{out}}$ (Fig. 25).

![Fig. 24: Simulated $f_{\text{max}}$ dependence on $R_{\text{out}}$.](image)

Turning our attention to the overlap capacitance, we see that $f_T$ and $f_{\text{max}}$ decrease with larger $C_{gd}$ (Fig. 26). Again, normalizing by $f_T$ shows that $f_{\text{max}}$ is independently affected by the overlap capacitance (Fig. 27).

The relative importance between these two parameters, $R_{\text{out}}$ and $C_{gd}$, is dependent on the values for each. In Fig. 28, we place $C_{gd}/C_{gs}$ on the x-axis and $R_{\text{out}}$ on the y-axis and plot the contours of constant $f_{\text{max}}$. This gives an idea of how much, for a given $R_{\text{out}}$
Fig. 25: Simulated $f_{\text{max}}/f_T$ dependence on $R_{\text{out}}$.

Fig. 26: Simulated $f_{\text{max}}$ dependence on $C_{\text{gd}}$.
Fig. 27: Simulated $f_{\text{max}}/f_T$ dependence on $C_{gd}$.

Fig. 28: Simulated trade-off between $R_{out}$ and $C_{gd}$ for constant $f_{\text{max}}$ using equivalent circuit model (Fig. 10), where $V_{gs} = 0.7$ V, $V_{ds} = 1.5$ V, $L_g = 0.1 \mu$m, $W_g = 5 \mu$m.
and \( C_{gd} \), we must change one parameter in order to remain at the same \( f_{\text{max}} \) due to a small change in the other. At low values of \( R_{\text{out}} \), the contours of constant \( f_{\text{max}} \) are almost horizontal, implying that a small change in \( R_{\text{out}} \) requires a drastic change in \( C_{gd} \) to maintain the same level of performance. Here, \( R_{\text{out}} \) clearly dominates the behavior of \( f_{\text{max}} \). For higher values of \( R_{\text{out}} \), the contours are much steeper, and \( C_{gd} \) has a stronger effect. Our experimental devices lay somewhere in the intermediate region, so minimizing \( C_{gd} \) will have a similar performance boost as improving \( R_{\text{out}} \) for the 0.1 \( \mu \)m generation.

The plot for constant \( f_T \) is shown in Fig. 29. Here, \( R_{\text{out}} \) only factors in when its value approaches that of the drain resistance.

![Simulated trade-off between \( R_{\text{out}} \) and \( C_{gd} \) for constant \( f_T \) using equivalent circuit model (Fig. 10), where \( V_{gs} = 0.7 \) V, \( V_{ds} = 1.5 \) V, \( L_g = 0.1 \) \( \mu \)m, \( W_g = 5 \) \( \mu \)m.](image)

After normalizing to \( f_T \), the plot for constant \( f_{\text{max}}/f_T \) is shown in Fig. 30. Clearly, both the output resistance and the overlap capacitance are important in \( f_{\text{max}} \).

4.5 Verification of the predictive power of analytical expression for \( f_{\text{max}} \)

At this point, we can pause to check if the \( f_{\text{max}} \) formulation (Eq. 5) predicts the observed behavioral dependencies. Instead of varying the element values inside of the
Fig. 30: Simulated trade-off between $R_{\text{out}}$ and $C_{gd}$ for constant $f_{\text{max}}/f_T$ using equivalent circuit model (Fig. 10), where $V_{gs} = 0.7 \ \text{V}$, $V_{ds} = 1.5 \ \text{V}$, $L_g = 0.1 \ \mu\text{m}$, $W_g = 5 \ \mu\text{m}$.

equivalent circuit model schematic, measuring the S-parameters, and extrapolating to find $f_{\text{max}}$, we can take the values from one circuit, plug them into the formula, and directly compute $f_{\text{max}}$.

Fig. 31: Side-by-side comparisons for $f_T$ and $f_{\text{max}}$ dependence on $R_g$. Equation 5 is used to generate the plot on the left, and the equivalent circuit model is used to generate the plot on the right.
A comparison between the dependence of $f_{\text{max}}$ and $f_T$ on gate resistance is shown for both the analytical equation and the circuit model in Fig. 31. The plot generated from the equation shows very close correlation to that of the circuit.

In Fig. 32, the predictions of the analytical equation are plotted alongside the results of the circuit model for the $f_{\text{max}}$ dependence on the output resistance for different overlap capacitances. The results are remarkably accurate despite the fact that the equation omits any dependency on drain resistance or a distributed gate resistance. Instead, one must estimate an effective lumped gate resistance. The end results show that, given the correct values of the circuit elements, the analytical equation can be a powerful tool for making predictions regarding $f_{\text{max}}$.

![Graph showing $f_{\text{max}}$ vs $R_{\text{out}}$ with different overlap capacitances.](image)

**Fig. 32:** Side-by-side comparisons for $f_{\text{max}}$ dependence on $R_{\text{out}}$, with $C_{gd}$ as a parameter. Equation 5 is used to generate the plot on the left, and the equivalent circuit model is used to generate the plot on the right.

### 4.6 Drain engineering

One method of improving $f_{\text{max}}$ is to use drain engineering. A reduced drain extension doping concentration -- Lightly Doped Drain (LDD) -- will improve $R_{\text{out}}$ by reducing the channel length modulation. It will also improve the overlap capacitance, since the drain depletion width is larger for similar $V_{gd}$. Indeed, a circuit extraction of a FIELDAY simulation of device with an asymmetric LDD structure shows that $R_{\text{out}}$ improves by a factor of 2, while the overlap capacitance is reduced by 33%. A side effect is that $g_m$ also decreases by 45% as a result of the longer effective channel length. Once these values are
Fig. 33: Simulated comparison of $f_T$ for bulk, LDD, and DMOS structures, normalized by drain current.

Fig. 34: Simulated comparison of $f_{max}$ for bulk, LDD, and DMOS structures, normalized by drain current.

normalized to their drain currents, $f_T$ and $f_{max}$ both show roughly a 15% increase in value.
over conventional devices (Figs. 33 and 34).

Alternatively, one can use a DMOS structure -- a device with a very large drain area (Fig. 35). This structure increases the breakdown voltage of a device by causing a voltage drop across the drain resistance rather than across the gate oxide. As mentioned before, neither $f_T$ nor $f_{\text{max}}$ give much information regarding breakdown voltage of a device. After normalizing to the drain current, the DMOS device offers an increase of 9% in both $f_T$ and $f_{\text{max}}$ over the conventional bulk device (Figs. 33 and 34).

![Cross-sectional view of DMOS structure](image)

**Fig. 35:** Cross-sectional view of DMOS structure (not to scale).

One may argue that the increase in $f_{\text{max}}$ is due to the increase in $f_T$, since the percent increase in each is the same. However, as Fig. 36 demonstrates, for comparable $f_T$'s, the LDD and DMOS show modest improvement in $f_{\text{max}}$ over conventional devices. These structures help increase the $f_{\text{max}}/f_T$ ratio (dotted line on Fig. 36). The immediate goal is to create devices that lie on or above this line, and as far to the upper right as possible. However, alternate criteria, such as breakdown voltage, must be carefully weighed before making a final judgment on the worth of these devices.

### 4.7 Conclusion

As channel lengths of MOS devices continue to shrink, $f_T$ will continue to increase, but, at a reduced pace. The slope of $f_{\text{max}}$ with $1/L_g$, which is less than that of $f_T$ to begin with, is hurt even more as devices are scaled. To improve $f_{\text{max}}$, one must, most importantly, minimize the gate resistance. Next, one can decrease the overlap capacitance and increase the output resistance, the impacts of which depend on the relative values for each. For the 0.1 $\mu$m generation of devices, $C_{gd}$ and $R_{out}$ are equally significant in influ-
Fig. 36: Simulated comparison of $f_{\text{max}}$ vs. $f_T$ of bulk, LDD, and DMOS structures.

One way of addressing the low $f_{\text{max}}$ issue is to use drain engineering to improve the output resistance and overlap capacitance. A lightly-doped-drain structure best achieves this. In the next chapter, we will alter our MOSFET even more by using a buried substrate. Then, we can evaluate whether SOI and its lower junction capacitance offers any high-frequency performance assistance.
Chapter 5: $f_{\text{max}}$ in bulk and SOI devices

So far, we have analyzed the impact of parameters such as output resistance, overlap capacitance, and gate resistance on $f_{\text{max}}$. Now, we apply this knowledge about the trends of $f_{\text{max}}$ to evaluate the relative high-frequency performance of two different device structures, a conventional bulk and a silicon-on-insulator (SOI) device. In this context, the relevance of junction capacitance (the drain-to-bulk capacitance, $C_{db}$) to $f_{\text{max}}$ is evaluated.

5.1 Background

Recently, the role of silicon MOSFETs in the field of microwave applications has sparked interest for their enticing potential of offering high performance at lower cost [8]. At the same time, SOI devices have been touted for offering faster switching speeds over conventional bulk devices [25]. These two parallel developments naturally arouse inquiry into whether silicon-on-insulator (SOI) can provide improved high frequency performance. With our understanding of $f_{\text{max}}$, we are in a good position to perform a detailed comparison between bulk and SOI devices for microwave applications.

5.2 Experimental observations

To address this issue, we have fabricated an SOI and a bulk device using identical processing steps. The channel lengths for these devices, which were extracted using a shift-and-ratio method, show close correspondence: the bulk device has an effective channel length of 0.0922 $\mu$m, while the SOI device has an effective channel length of 0.0919 $\mu$m. In addition, both devices have 20 fingers, each 5 $\mu$m long. At a drain voltage of 1.5 V, these devices have similar peak $f_T$'s, 141 GHz for the bulk device and 137 GHz for the SOI one, but SOI has a lower peak $f_{\text{max}}$ of 63 GHz compared to the 80 GHz $f_{\text{max}}$ of the bulk counterpart (Fig. 37). The difference in $f_T$ is too minor to account for the 20% lower $f_{\text{max}}$ of the SOI device.
There are two possible sources for this difference in $f_{\text{max}}$: the lower junction capacitance and the floating body effect. Intuitively, we would guess that lower junction capacitance would, if anything, help $f_{\text{max}}$. In the next section we will analytically examine the impact of $C_{\text{db}}$ on $f_{\text{max}}$.

5.3 Qualitative impact of junction capacitance on $f_{\text{max}}$

Though reduced junction capacitance is important in digital applications, does it bring increased performance at high-frequency? At DC, the electrical characteristics of the bulk and SOI devices are basically identical; both exhibit the same current-to-voltage relationships. To compare their AC characteristics, we must delve into the details of $f_{\text{T}}$ and $f_{\text{max}}$.

The analytical formulas for $f_{\text{T}}$ and $f_{\text{max}}$ (Eq. 4, 5) offer no hint of the impact of the junction capacitance since these equations have no explicit dependency on $C_{\text{db}}$. This may be because these formulas are based on a small-signal equivalent circuit model which is...
too simplistic (Fig. 4).

This model suggests that $f_T$ should be independent of drain-bulk capacitance, since the drain terminal is shorted to ground (the source) when calculating $f_T$. Thus, both $R_{\text{out}}$ and $C_{db}$ are shorted out. Likewise, in $f_{\text{max}}$, $C_{db}$ is negated since the output is conjugately matched.

![Circuit model which includes parasitics source, gate, and drain resistances.](image)

**Fig. 38:** Circuit model which includes parasitics source, gate, and drain resistances.

After using a more complete model (Fig. 38) which includes the drain resistance, this is no longer the case. The drain resistance creates a current divider when the drain terminal is shorted to ground. This does not appreciably affect $f_T$ when the output impedance of $R_{\text{out}}$ and $C_{db}$ is sufficiently larger than $R_d$. However, this effect may become more pronounced at shorter channel lengths, since the $R_{\text{out}}$-to-$R_d$ ratio likely suffers, due to inadequate scaling of the drain extension depth and doping concentration.

Similarly, $C_{db}$ will shunt current from the drain terminal and lower $f_T$. This parasitic causes the current gain-versus-frequency slope to deviate from an ideal slope of 20 dB/decade. An additional breakpoint is introduced when $1/(\omega C_{db}) \approx R_d$. Since $C_{db}$ is usually only a few femtofarads and $R_d$ is a few tens of ohms, this pole will be in the tens of Terahertz, which is well outside the frequency regime which we are concerned.
5.4 Quantitative impact of $C_{db}$

To quantify the impact of $C_{db}$ on $f_{max}$, an even more complicated equivalent circuit is constructed. The presence of a drain-to-bulk resistance ($R_{sub}$) is key to increasing the accuracy of the circuit model in relation to experimental data. The final circuit is depicted in Fig. 10 and is used as the basis for extracting parameter values. The circuit in Fig. 39 is similar, but, for simplicity, the gate resistance is lumped into a single element. This schematic is employed to ease explanation and hand analysis.

For device dimensions of $0.1 \, \mu m$ length by $5 \, \mu m$ width, two distinct circuit models are created to correspond to SOI and bulk experimental devices. The bulk device has a $C_{db}$ of $2.7 \, fF$, and the grounded SOI device has a $C_{db}$ of nearly $0 \, fF$. This, by itself, translates into a difference in $f_{max}$ of about $10\%$, as seen from figure 40, which explicitly shows the relationship between the two metrics, $f_T$ and $f_{max}$, and $C_{db}$, based on the circuit model.

This $f_{max}$ dependence on junction capacitance is weaker than either the output resistance or overlap capacitance dependence. Furthermore, the junction capacitance of a bulk device is small enough that the move to SOI does not offer much improvement. Additionally, the substrate resistance has a negligible impact on $f_T$ and $f_{max}$, as indicated by Fig. 41.
Fig. 40: $f_T$ and $f_{\text{max}}$ dependence on $C_{\text{db}}$, with SOI and bulk values labelled.

Fig. 41: $f_T$ and $f_{\text{max}}$ dependence on $R_{\text{sub}}$, with SOI and bulk values labelled.
5.5 Impact of junction capacitance on output impedance

Since the output port is conjugately matched as one of the conditions for $f_{\text{max}}$, one would think that the effect of the junction capacitance would be completely neutralized. This is not true because the junction capacitance also contributes a non-imaginary component to the $Y$-parameters. More specifically, the junction capacitance manifests itself in the real part of $Y_{22}$ ($\text{Re}\{Y_{22}\}$).

Returning to our simplistic circuit model (Fig. 4), we see that:

$$Y_{22} = \frac{1}{R_{\text{out}}} + j\omega(C_{gd} + C_{db})$$

(10)

and that, in this case, $\text{Re}\{Y_{22}\}$ is inversely proportional to only the output resistance. Once more components are added to the circuit, $\text{Re}\{Y_{22}\}$ becomes a function of these other parameters, and the analytical equation must be updated to reflect this change. For purposes of understanding the repercussions of these new elements, the $f_{\text{max}}$ formula may be modified by substituting the circuit element $R_{\text{out}}$ with this new $1/\text{Re}\{Y_{22}\}$, in essence, an effective output resistance. Then, we can probe into $\text{Re}\{Y_{22}\}$ to see how $C_{db}$ ultimately affects $f_{\text{max}}$.

Fig. 42: $Y_a$ and $Y_b$ partitions.

A simplified view of the schematic in Figure 39 may be used to focus on the junction capacitance. Partitioning the circuit model into $Y_a$, $Y_b$, and $R_d$ (Fig. 42) yields:
\[ Y_{22} = \frac{1}{R_d} \| (Y_a + Y_b) \]  
\[ = \frac{Y_a + Y_b}{1 + R_d(Y_a + Y_b)} \]  

(11)  

Now, Re\{Y_{22}\} is no longer simply a function of \( R_{\text{out}} \). Then, considering the drain resistance is approximately 40 ohms, and overall, \( Y_{22} \) is a few tenths of a milliSiemen (in the frequency range which we are looking at), this suggests that the denominator is close to unity and that:

\[ Y_{22} = Y_a + Y_b \]  

(13)

Then, we can write \( Y_b \) may be in terms of \( C_{db} \):

\[ Y_b = \frac{1}{R_{\text{sub}}} \| j\omega C_{db} \]  

(14)

Thus,

\[ \text{Re}\{Y_b\} = \frac{\omega^2 C_{db}^2 R_{\text{sub}}}{1 + \omega^2 C_{db}^2 R_{\text{sub}}^2} \]  

(15)

\( C_{db} \) is approximately 2.5 fF, and \( R_{\text{sub}} \) is about 500 ohms, so at a frequency of 20 GHz,

\[ \omega^2 C_{db}^2 R_{\text{sub}}^2 = (0.15)^2 \]  

(16)

This leaves the denominator of Re\{Y_{22}\} close to 1, and

\[ \text{Re}\{Y_b\} = \omega^2 C_{db}^2 R_{\text{sub}} \]  

(17)

Finally, we arrive at the following expression:

\[ \text{Re}\{Y_{22}\} \approx \text{Re}\{Y_a\} + \omega^2 C_{db}^2 R_{\text{sub}} \]  

(18)

Therefore, \( C_{db} \) does have an effect on Re\{Y_{22}\}. This is indeed the case; increasing the value of \( R_{\text{sub}} \) or \( C_{db} \) will increase Re\{Y_{22}\}, which, in turn, decreases the effective output resistance and reduces \( f_{\text{max}} \). Again, this is only a second order effect, since \( Y_a \) dominates over \( Y_b \). The magnitude difference in \( f_{\text{max}} \) due to a change in \( C_{db} \) from 2.5 fF to 0.1 fF is
about 5 GHz.

Up to this point, we have shown the minimal role that the junction capacitance plays. We have yet to determine the cause of the lower $f_{\text{max}}$ in SOI devices. In the next section, we will take a close look at the Y-parameters to see if we can determine the cause.

5.6 Experimental Y-parameters

Next, we can examine the experimental Y-parameters and see if we can pinpoint the reason for the disparity in $f_{\text{max}}$ between bulk and SOI devices (Fig. 43). These results show close correspondence between the bulk and SOI in terms of $Y_{11}$, $Y_{21}$, and $Y_{12}$. On the other hand, distinct differences appear in both the real and imaginary parts of $Y_{22}$. The $\text{Re}\{Y_{22}\}$ of SOI is 50% larger than that of bulk, while the $\text{Im}\{Y_{22}\}$ is one-fourth that of...
the bulk device.

Then, we can extract equivalent circuit models for both bulk and SOI devices (Figs. 44 - 46). These circuits show two major differences: SOI has lower junction capacitance, as expected, and lower output resistance.

The lower junction capacitance, as demonstrated previously, should help $f_{\text{max}}$, if anything. Thus, we must explain why the floating body effect causes lower output resistance.

![Comparison of bulk experimental and extracted circuit model Y-parameters.](image)

Fig. 44: Comparison of bulk experimental and extracted circuit model Y-parameters.
Fig. 45: Comparison of SOI experimental and extracted circuit model Y-parameters.

Fig. 46: Comparison of Y-parameters from extracted circuit models for Bulk and SOI devices.
5.7 Introduction of simulation to explain $Y_{22}$ disparity.

FIELDAY can be programmed to simulate an SOI device that is ensured to differ from the bulk device only by the presence of a buried oxide layer; the remainder of the device structure is identical. This inherent advantage of simulation permits a comparison of bulk and SOI free from discrepancies due to processing variations in channel length and doping profile. This is a major problem when trying to compare actual experimental devices; one cannot easily decouple the performance differences due to processing variations from those caused by the intentional structural changes. It should also be noted that, in these particular simulations, impact ionization is not modeled.

FIELDAY simulations faithfully reproduce the relevant trends seen in the $Y$-parameters of the experimental bulk and SOI devices (Fig. 47). This re-assures us that these trends are not due to processing variations and are indeed indicative of a genuine distinction between bulk and SOI devices. In addition, the degree of correspondence gives us confidence that we have an adequately modeled the device. The major shortcoming of this modeling is that the absolute value of the Re{$Y_{22}$} is lower than experimentally seen. A modification to the doping profile to decrease the output resistance would alleviate this problem. However, for our evaluation purposes, the model does an adequate job of reproducing all the relevant trends.

In terms of our figures of merit, SOI has a slightly higher peak $f_T$ with respect to the drain current, and a lower peak $f_{max}$ (Figs. 48, 49). The experimental devices may show that the SOI has a lower $f_T$ due to self-heating or slight differences in channel length. In any case, the simulations indicate that $f_T$ is approximately the same, while $f_{max}$ is lower in SOI devices, as is observed experimentally.

With this credible simulation environment, we are in a favorable position to delve into the crux of the matter: explaining the difference in $Y_{22}$ between bulk and SOI and the influence of this disparity on $f_{max}$.

First, we will take advantage of another useful property of the simulation. We can observe and control the potential in different areas of the device that would otherwise be inaccessible in a real device. For example, in the first device under scrutiny, a contact is placed just above the buried oxide so that the body of the device is firmly grounded (Fig.
Fig. 47: Comparison of simulated FIELDAY Bulk and SOI model Y-parameters.

50). This allows the two characteristics of an SOI device to be decoupled; the first comparison will demonstrate the effects of the lower drain-bulk capacitance, while the second comparison will assess the impact of the floating body effect.

5.8 The impact of the floating body effect

Next, the ground contact in the body of the SOI device was removed to allow the body to float, as is the case in real devices (Fig. 51). However, it is important to note that no impact ionization is being simulated.
Fig. 48: Comparison of simulated $f_T$ of bulk and SOI devices with respect to drain current.

Fig. 49: Comparison of simulated $f_{\text{max}}$ of bulk and SOI devices with respect to drain current.
The DC simulation show similar characteristics to those of the grounded SOI. Examination of the Y-parameters shows that the major difference lies in the real and imaginary parts of $Y_{22}$ (Fig. 47). The discrepancy in $\text{Im}\{Y_{22}\}$, caused by the difference in junction capacitance, is not of large concern since it will be canceled due to conjugate matching of the output in calculating $f_{\text{max}}$. The $\text{Re}\{Y_{22}\}$ does not cause any shift in $f_T$, but does produce a 10% lower $f_{\text{max}}$ of 95 GHz in SOI. The equivalent circuit model is useful in illuminating the reasons for this disparity.

The extracted circuit for a floating-body SOI device has an output resistance which is almost two times smaller than that of a bulk device. An initial conjecture at explaining this observation may be to blame the floating body effect which forward-biases the body. However, a body raised to a higher potential would produce a higher output resistance, due to the smaller depletion width near the drain. This is confirmed by simulating an SOI device with its body potential raised to 0.5 V (Fig. 52). This back-biased device has a higher $R_{\text{out}}$ than the grounded-body SOI device.
The decrease in output resistance appears to be a result of the floating body's inability to quickly drain away charge. When a small signal voltage is applied to the drain, the depletion width increases across the drain-bulk junction (Fig. 53). On the body side, the majority carriers are pushed away and move into the body. The time it takes to be collected in the grounded source is on the order of microseconds, far slower than the time allowed in the radio frequency regime. Therefore, the charge remains in the body and slightly raises its potential. This decreases the threshold voltage and results in higher drain current. Thus, an overall decrease in output resistance is observed. This explanation also applies to $g_m$; a small signal voltage applied to the gate will repel positive carriers into the body, inducing a decrease in the threshold voltage. Again, the drain current increases. Simulation shows that $g_m$ of an SOI device is less than 5% higher than that of a bulk device.

Since impact ionization is not simulated, the floating body picture is not complete. With this phenomenon, the body of a device will be further forward-biased. The impact on the small signal parameters should be minimal since the conduction current should not generate impact ionization current. In this case, the DC characteristics are modified; the body would be forward-biased. However, as the back-biased device simulation shows, $R_{out}$ only increases by 5% for a 0.5 V body potential.
Fig. 53: A closer look at the drain region of an SOI device (circled in the upper cross-section). The dashed line represents the depletion width edge, which modulates due to the applied small-signal voltage source, $v_{ds}$.

The results of the set of FIELDAY simulations are summarized in Figure 54 and Table 2. All measurements are taken at similar drain currents and a drain-to-source potential of 1.5 V. The floating-body SOI device has a higher $\text{Re}\{Y_{22}\}$ than the bulk, grounded SOI, and back-biased SOI devices. In addition, the bulk device shows a higher $\text{Im}\{Y_{22}\}$ than the other three SOI devices.
Fig. 54: The $Y_{22}$ parameters for various FIELDAY simulations.

Table 2: Summary of simulation results for similar drain currents, at $V_{ds} = 1.5$ V.

<table>
<thead>
<tr>
<th></th>
<th>$f_T$</th>
<th>$f_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bulk</td>
<td>83 GHz</td>
<td>105 GHz</td>
</tr>
<tr>
<td>grounded SOI</td>
<td>83 GHz</td>
<td>103 GHz</td>
</tr>
<tr>
<td>back-biased SOI, $V_{bs} = 0.5$ V</td>
<td>83 GHz</td>
<td>104 GHz</td>
</tr>
<tr>
<td>floating-body SOI</td>
<td>84 GHz</td>
<td>95 GHz</td>
</tr>
</tbody>
</table>

5.9 Conclusions regarding SOI devices

In conclusion, SOI devices have similar $f_T$'s to those of bulk devices for devices of comparable channel length. On the other hand, SOI has a lower $f_{\text{max}}$ due to its floating body effect. This is mainly due to a decrease in output resistance. Additionally, the drain-to-bulk capacitance is only a second order effect in $f_{\text{max}}$. Bulk devices have small enough junction capacitance such that its impact on $f_T$ and $f_{\text{max}}$ is minor.
Chapter 6: Conclusion

As silicon MOSFET technology that is designed for digital applications continues to evolve, the prospects for using MOSFETs in the more demanding niche of radio/microwave frequency applications improves. The high-frequency suitability of a device can be examined by means of $f_T$ and $f_{\text{max}}$, two widely regarded indicators of good high-frequency performance. Already, there have been numerous reports of silicon MOSFETs with $f_T$'s reaching above 150 GHz. In contrast, $f_{\text{max}}$ in silicon MOSFETs is relatively low. To address this problem, this work aims to increase understanding of $f_{\text{max}}$ in silicon MOSFETs.

In this thesis, I have developed and validated a CAD framework that helps understand which parasitics are of highest importance in influencing the $f_{\text{max}}$ of MOSFETs. This framework consists of a combination of a two-dimensional device simulation modified to include a distributed gate resistance and a small-signal equivalent circuit model. These tools indicate that the gate resistance is of primary importance. The overlap capacitance and the output resistance have the next greatest impact on $f_{\text{max}}$.

This work suggests that there are several different methods of improving $f_{\text{max}}$, ranging from simple layout modifications to complicated changes in the fabrication process. For example, the gate resistance may be decreased by using smaller finger lengths or by utilizing different gate materials that have lower resistivity. This CAD framework allows these and other design strategies to be examined and evaluated before device fabrication.

This framework can also be used to identify strengths and weaknesses of different device structures. One such example, Silicon-on-insulator (SOI), is discussed in detail. It is found that junction capacitance has little impact on $f_{\text{max}}$. However, the floating body of an SOI device decreases the output resistance, and consequently adversely affects the device's $f_{\text{max}}$. Experimentally, an SOI device has a 20% lower $f_{\text{max}}$ than a comparable bulk device that has undergone identical processing.

The work presented in this thesis provides a deeper understanding of the salient
issues involving $f_{\text{max}}$ in silicon MOSFETs. These ideas should aid in designing silicon MOSFETs with high $f_{\text{max}}$. 
Chapter 7: References


