NOISE IN HIGH ELECTRIC FIELD TRANSPORT AND LOW NOISE FIELD EFFECT TRANSISTOR DESIGN: THE ERGODIC METHOD

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Luiz M. Franca-Neto
May 1999

Abstract

Noise, meaning *fluctuations* in carrier transport, is a topic of research in its own right. The literature is vast on methods and approaches to noise calculations. These methods and approaches in general fall into two major categories: (1) Methods based on a broad semi-axiomatic approach to noise, where noise sources are random variables described by their statistical characteristics, and (2) Methods tailored for the analysis of a specific form of fluctuations, produced under special conditions, like transport at very low temperatures or mesoscopic dimensions.

Unfortunately, the methods above are *descriptive methods* in nature only, and other than trivial inferences, they are entirely disconnected or only indirectly connected to fundamental semiconductor device design issues.

In the work of this thesis, the goal is to develop an approach to noise phenomena in macroscopic semiconductor-devices under a framework of carrier transport formalism, but highly connected to device design. The new approach to noise should also be able to lend itself to high electric field transport. This will hopefully provide non-obvious new and general guidelines for the design of new devices specifically for low noise operation.

High electric field transport brings a new perspective to noise analysis. Under equilibrium conditions all conductors and semiconductors develop thermal noise according to the Nyquist theorem. Thus, the thermal noise generated is only a function of the resistance and the temperature of the conductor, and the bandwidth of the measurement. The thermal noise is independent of the material from which the conductor is made. Contrary to equilibrium conditions, conductors and semiconductors develop *additional noise* when a DC current flows through them and this *additional*

noise does depend on the conducting material system.

With the augment of electric field applied to the semiconductor, the *additional noise* will increasingly depart from the Nyquist theorem. An assertion made and proven in this work is that different semiconductor materials or systems will develop *additional noise* differently, but in a *predictable* and *controllable* way, hence allowing device design optimization for low noise operation.

The theoretical framework for noise analysis is established via the introduction of the *Ergodic Method* created in this work. The name stems from an *ergodic assumption* made to allow calculation of noise power from the carrier velocity distribution.

The *Ergodic Method* uses the Boltzmann Transport Equation (BTE) for the calculation of velocity distributions, however other tools, such as Monte Carlo calculation can be used to assess the carrier velocity distribution and they would be equally functional for the *Ergodic Method*. The advantage of the BTE is to present in a compact form, all the main device features that influence the carrier velocity distribution and are under the control of the device designer.

The *Ergodic Method* recovers the thermal noise expression as a special case of the current noise power when the device is at equilibrium (no DC current). The *Ergodic Method* also lends itself to high electric field transport and can thus be extended to tackle deep sub-micron devices.

Extensions of the *Ergodic Method* to be used in noise analysis of mesoscopic devices are also discussed.

Experiments are performed on MOSFETs to demonstrate the predictions brought out by the *Ergodic Method*. In this thesis, the theoretical concepts and axioms of the *Ergodic Method* are presented first, followed by justification by the correspondence between their consequent predictions and actual experimental results.

MOSFETs are the semiconductor-devices used in the experiments because MOSFETs, being three terminal devices, they allow independent change in carrier availability for transport and in the applied electric fields. The discussed experimental results are for silicon FETs, but the concepts they prove right are readily extendible to other semiconductor systems. Moreover, these theoretical concepts establish the guidelines for the design of new low noise FETs.

Acknowledgments

As I wrap things up at Stanford University, and finish the work that led me to the Ph.D. degree in Electrical Engineering, I look back on my years at Stanford and find a great number of people who strongly influenced and supported my way up here. Some of them are always-present influence from well before Stanford, others are newer influences but equally strong ones. It is my pleasure to try now to put in words my deep appreciation for them.

I should start from home, where my parents, Ayrton and Janete Franca, constructed a wonderful environment of love and care. They made their kids understand the joy of intellectual achievements, and I should say they planned all this I am experiencing now well in advance. Thank you, Mom and Dad!

Then come my Professors at ITA - Instituto Tecnológico de Aeronáutica (*Aeronautical Institute of Technology*), CTA - Centro Técnico Aeroespacial (*Aerospace Technical Center*), who trained me in Math, Science and Engineering with unyielding dedication and attention, and encouraged me to pursue graduate studies. Special thanks should go to both Prof. Darwin Bassi and Prof. Fernando Walter, who were assigned to be my program advisors at ITA only, but went well beyond the call of duty. I am also grateful to Prof. Fernando Sakane, who strongly influenced my interest in electronics and signal processing in his classes, and kindly took the time to discuss with me the good and bad aspects of the carrier paths I was planning to pursue after ITA.

Once at Stanford, I was very lucky to be adopted by Prof. James Harris, who truly deserves his nickname "Coach"! I should stress "adopted" because Prof. Harris is like a second father to all the Ph.D. students in his group and provides the intellectual

freedom in his group which is at the same time a privilege and also a responsibility for us. The open research atmosphere he created for his group is very likely the best tailored environment to train and challenge a young researcher. I am truly grateful to Prof Harris for the opportunity he gave me to pursue my own research topic and also grateful for his extreme attention and patience to always find some time in his ultra-busy schedule to discuss some research ideas and encourage me to always keep moving.

I would like to thank a number of people without whom my Ph.D. work would never had reached the level it reached. I should thank Dr. Judy Hoyt and Ken Rim, from Stanford University; Dr. Dieter Vook and Andrew Grzegorek, from Hewlett-Packard Laboratories; Dr. Francis Rotella and Dr. Francois Herbert, from Spectrian, Inc.; Dr. Paul de La Houssaye and Dr. Waddad Dubbelday, from SPAWAR; and Matt Wetzel and Prof. Peter Asbeck, from University of California at San Diego. They provided me all the advanced devices I needed for the experimental part of my research. Dr. Judy Hoyt and Ken Rim went even further and engaged in research collaboration and discussion of technical ideas with me. I truly benefited from their experience on the fabrication of Si-on-SiGe FETs and analysis of their experimental data.

I am grateful to Gail Chun-Creech and Lu Miao for their great administrative skills, which made everything run smoothly inside the Harris Group. I should then thank all the past and present Harris Group students with whom I interacted. The present ones made the life really enjoyable in the Solid State Electronics Laboratory, Center for Integrated Systems, and the past ones helped keep me motivated by their examples. Deserving mention is also Pauline Prather, who always agreed kindly to come to Stanford campus at unplanned nights to wire bond my devices when I needed experimental results for the following day. Equally, I must thank HP's Jeff Raggio who lent his expertise to help me on dealing with noise measurement equipments and HP's Dennis Hayes, who promptly offered his laser tools to blow undesired connections in my chips.

My coming to Stanford University meant meeting very warm reception by Mr. and Mrs. Anderson, who hosted my wife Marta and I for our first days in the USA,

and became our first American friends. Later, we would come to meet Mr. and Mrs. Sadlewicz, who also provided for the hospitality to make us feel at home all the time we lived our lives at Stanford. They were always most friendly and we are grateful for that.

I must thank my wife, Marta, for her love and patience all these years of hectic activities at Stanford University. She definitely provided the emotional support I needed and I am truly indebted to her.

Finally, the work for this dissertation was possible only thanks to a Doctoral Fellowship awarded to me by the Brazilian Council for Scientific and Technological Development - CNPq, Brasilia, Brazil, and to a Research Assistantship awarded to me by the Stanford University's Center for Integrated Systems and supported by Ericsson.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years, the development of deep sub-micron CMOS processes for digital circuit applications with usable analog performance in the microwave frequency range, has prompted a massive re-birth of analog and radio-frequency (RF) research both in universities and industry. This re-birth of RF research however, came with the new challenge of putting digital and analog circuitry together on the same chip without great penalties in performance degradation for either of these parts. Mixed signal Integrated Circuits (ICs) of this sort are receiving great attention since these developments coincided with an enormous popularization of wireless communication systems, cellular and PCS phones.

Figure 1.1 shows the *cutoff frequency*, f_t, for several device technologies. Note the ascent of MOSFETs to the operating frequencies of other technologies more frequently used in the past for RF applications. One of the great advantages of MOSFETs is the huge available infrastructure of fabrication facilities compared to all other technologies shown. This advantage makes MOSFETs also present in most of the Digital Signal Processing (DSP) and microprocessors chips, and their ascent to RF applications brings strong impact to the technology scenario, as they are capable of lowering the cost of manufacturing.

The re-birth of RF prompted a resurgence of concerns on *noise phenomena*. *Circuit level* solutions to cope with noise [1, 2], as well as investigations on how noise couples through the substrate from various subsystems in the same chip, [3, 4] have

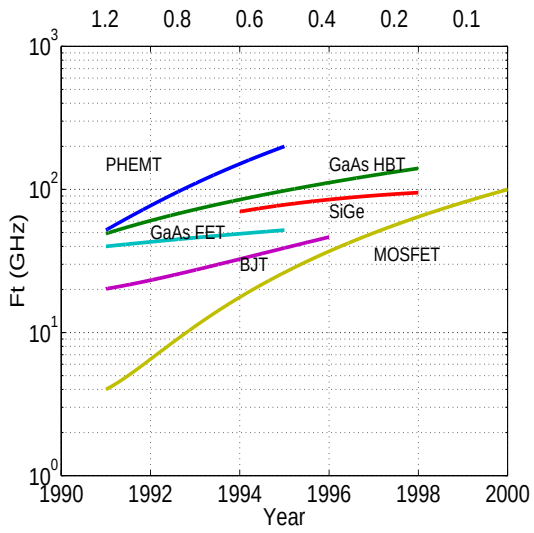

Figure 1.1: *Cutoff frequency*, f_t, for several Radio Frequency transistor technologies. The years and gate lengths for MOSFETs are displayed on the bottom and top of the plot respectively.

recently been reported with increasing volume.

However, very few results were produced at the *device level* for noise minimization. This thesis work focuses on the device level aspects of noise and develops an understanding of noise phenomena through a new approach that guides semiconductor-device design for optimum noise performance. [5, 6]

It was already known that under thermal equilibrium [7], conductors and semiconductors develop thermal noise due to the Brownian motion of the carriers scattering inside the crystal, and this thermal noise is independent of the material used in the conductor or semiconductor. Thus, the thermal noise power is only a function of the resistance of the conductor or semiconductor, the lattice temperature and the bandwidth of the noise measurement, as shown in figure 1.2. It was also known [7] that when a DC current flows through the conductor or semiconductor, *additional noise* is generated.

The important assertion made in this introduction and proven in this work is that when a DC current flows through the material, *additional noise* is generated in a *predictable* and *controllable* way, thus allowing device optimization for low noise operation.

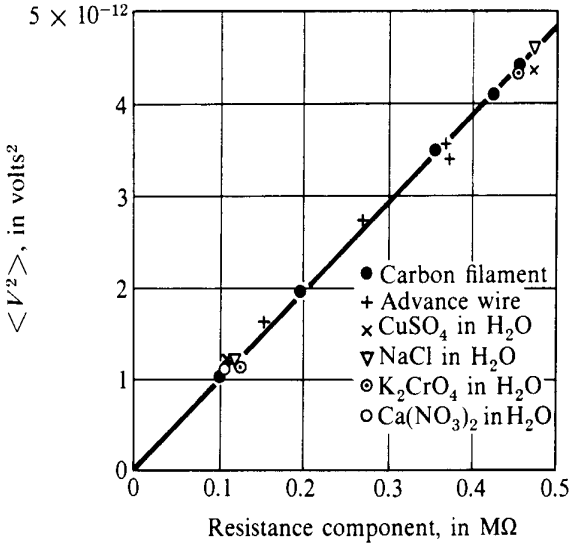

Figure 1.2: Voltage thermal noise for different conductors. The voltage thermal noise power is independent of the material that makes up the conductor and is only function of the DC-resistance value. (From "Thermal Physics", C.Kittel and H.Kroemer)

It is the goal of this thesis work to develop a new framework where the *noise phenomena* is approached from a transport problem perspective and where the noise characteristics of semiconductor devices are tied to design issues under the control of the device designer. This new framework for noise analysis is set by the introduction of the *Ergodic Method* and the calculation of noise power from the *variance* in carrier velocity distribution. It will be important to notice, as it will become clear later, that the current noise power generated in a semiconductor material is *not* the variance in carrier velocity distribution, but is directly related to it.

Standard sub-micron silicon MOSFETs, Si-on-SiGe MOSFETs, and Graded Channel MOSFETs are used in our experiments to demonstrate several predictions on noise performance from the *Ergodic Method* on the effects of *carrier effective masses* and *laterally graded doping profile* in noise performance of transistors under high electric field transport. Simpler device structures could be used for the demonstration of these effects on noise in high field transport, but being three-terminal devices, MOSFETs allow for variations in electric field and carrier population along the channel

independently, and thus provide great flexibility for transport and noise analysis. Furthermore, all the enhancements on noise performance on these devices are of great impact technologically and commercially, since these devices are the basis of most of todays semiconductor industry. Most importantly however, all the theoretical concepts developed in this thesis and proved right by these silicon FETs are readily extendible to devices based on other semiconductor systems.

1.1 Noise Terms

There are several noise terms in the literature and different noise sources in Nature.[8, 9, 10, 11, 14, 15] In this section, some of the most common types of noise are discussed and the type of noise we are going to focus upon in this work is defined.

The current noise power expressions are for equivalent shunt current sources and the voltage noise power expressions are for equivalent series voltage sources, as depicted in figure 1.3

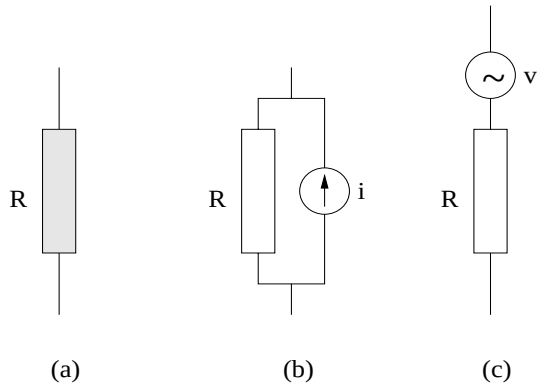

Figure 1.3: Noise sources: the noisy resistor or conductor (a) can be represented by a noiseless resistor or conductor of same resistance with a shunt noise current source (b), or the same noiseless resistance with a series voltage noise source (c).

1.1.1 Shot noise, quantum noise

In a conceptual description,[8, 11, 12, 13] shot noise is generated when there is a DC current flowing along a conductor or semiconductor in which the carriers pass through

a narrow constriction. The carriers are originally in a huge reservoir, with randomly distributed momentum. Only those carriers with proper forward momentum are allowed to proceed through the transport media, as depicted in figure 1.4.

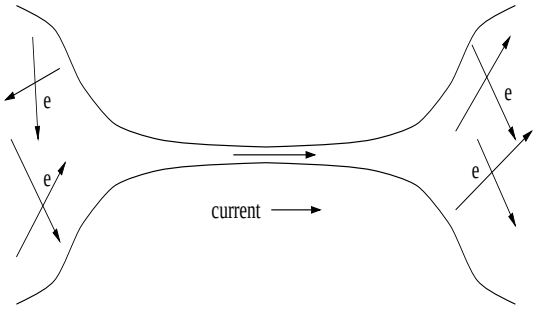

Figure 1.4: Shot noise: developed when DC current flows and carriers are strongly impaired from progressing along the transport media.

The conceptual description above is readily applied to p-n junctions or n-doped semiconductors at very low temperatures (close to 0K). In p-n junctions, under forward bias for instance, the electrons from the n-side of the junction (reservoir) must surmount a potential barrier (narrow channel concept) in order to reach the other side of the junction. In n-doped semiconductors at very low temperatures, the Fermi-level is very close to the conduction band. At these temperatures the electrons are frozen out onto their donors (reservoir) and it is with very low probability (narrowing channel) that they can get phonon-scattered into the conduction band to contribute to the DC current.

Shot noise is thus presented as evidence of the corpuscular nature of the carriers and it has a white noise spectrum. The current noise power is expressed as $\overline{i^2} = 2qI\Delta f$, where $\overline{i^2}$ is the current noise power, q, I and Δf are the electronic charge, the DC current through the device and the bandwidth of noise measurement, respectively. From the statistical perspective, the expression of shot noise is remarkable. It relates a *mean measurement*, the DC-current, to a *variance measurement*, the current noise power, and its form is a consequence of the assumption that it is very unlikely for a carrier inside the reservoir to acquire, via random scattering, a momentum appropriate for going through the channel. Because of that, shot noise is

normally modeled as a *Poisson random process*[1]. [16, 17, 18]

Moreover, it will be noted further on in the text that despite being a totally different random process, similar to shot noise, thermal noise will also have an expression that relates a *mean measurement* to a *variance measurement*, indicating a deep parallelism between the expressions of these two conceptually different noise sources.

1.1.2 Flicker noise, 1/f noise, pink noise

Flicker noise [19, 20, 21, 22] has no unique fundamental source and it is normally thought of in association with low frequency traps at interfaces or other capturing-releasing of carriers in semiconductor devices. The name stems from its spectral characteristics and in MOSFETs is normally related with surface states. This noise is normally buried by the thermal noise at frequencies of tens or a few hundred of kilohertz in MOSFETs and even lower frequencies in Bipolar Junction Transistors (BJT).

1/f noise is of particular concern in oscillators (phase noise) and circuits which up-convert or down-convert frequencies.

1.1.3 Popcorn noise, burst noise, bistable noise, Random Telegraph Signals noise (RTS noise)

Popcorn noise [23, 24]is the noise type where the noise level switches randomly between two levels, as depicted in figure 1.6.

RTS noise is normally associated with contamination, but is hardly predictable since it is not well understood.

[1]Poisson distribution can be written as the limit of a Binomial distribution when there is a great number of Bernoulli trials and very small probability of success. In the Poisson distribution the *variance* is equal to the *mean* and for the shot noise case this translates into a spectrum of noise with constant value equal to *2qI*.

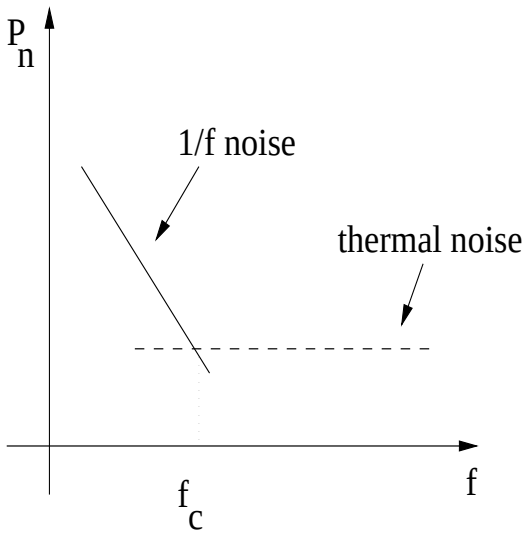

Figure 1.5: Flicker noise, or 1/f noise, gets its name from the its power spectrum. 1/f noise in MOSFETs usually gets buried by the thermal noise of the device at frequencies of tens to a few hundreds of kilohertz.

1.1.4 Thermal noise, Johnson noise, Nyquist noise, Johnson-Nyquist noise

Thermal noise [25, 26, 27, 28] is the noise generated by the thermal fluctuations in carrier velocity, which makes the carriers wander along the conductor or semiconductor with Brownian motion.

The expressions for thermal current and voltage noise sources are:

$$\overline{i^2} = 4k_B T_L G \Delta f \text{ for current noise source} \tag{1.1}$$

$$\overline{v^2} = 4k_B T_L R \Delta f \text{ for voltage noise source} \tag{1.2}$$

This is the type of noise of focus in this work, as the dominant noise source in MOSFETs is the thermal noise developed in the channel of these transistors.

Note that the current and voltage noise expressions were deduced for thermal equilibrium conditions and they are *only* guaranteed for that condition. These expressions relate (as was also the case of shot noise) a *mean measurement*, resistance

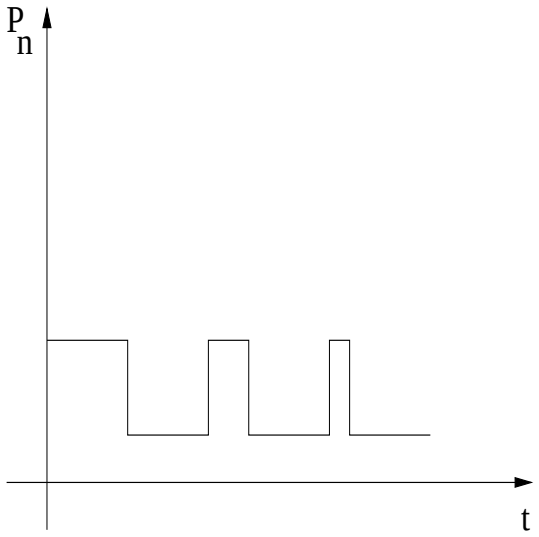

Figure 1.6: Popcorn noise or RTS noise: noise level switches randomly between two levels.

or conductance, to a *variance measurement*, the current or voltage noise power.

Because of the fact that these expressions were developed for thermal equilibrium, some questions arise for the extension of these equations when the conductor is not in thermal equilibrium anymore, and a DC current flows through it. Should the *"resistance"* used in the equations above be the *"DC resistance"* or the *"AC resistance"* when there is a current flowing? The answer to that comes from a deep appreciation of the meaning of noise, DC and AC resistance. The detailed discussions and their non-obvious but unambiguous conclusions are presented in chapter 3, where experimental evidence is provided to support the theoretical discussions.

Also in chapter 3, the concept of *excess noise* in MOSFET will be introduced. This concept will be tied to any noise level developed by the semiconductor above thermal noise. This will be an instrumental concept for the case when the device is biased and has a DC current flowing through it. A new expression, derived from the thermal noise expression to quantify *excess noise* is:

$$\overline{i^2} = 4k_B T_L G \Delta f \gamma_f \tag{1.3}$$

γ_f is called the *excess noise factor* and will be used to compare different technologies for MOSFETs and their relative performance for low noise operation under high electric fields. A definition of the conductance, G, is postponed to the discussions in chapter 3.

1.2 Extrinsic and Intrinsic Noise Sources

Noise analysis of Field Effect Transistors, FETs, is based on the characterization of the different noise sources contributing to the total noise power exhibited by the FET.

These noise sources can be divided into two classes: a) *Extrinsic noise sources*, which are those associated with the device connections through which the transistor action is externalized; and b) *Intrinsic noise sources*, which are those due to the inherent transport phenomena inside the FET.

The detrimental effect of extrinsic noise sources can be minimized by appropriate design of connections. These design decisions are normally in the general direction of making connections as large as possible and/or as short as possible, with the intent of minimizing the resistance of these connections. Similarly, connections can be made multi-fingered, with the same intent of making them less resistive. Although these are design issues of difficult implementation at times, they are standard procedures, with all the necessary knowledge-basis already established.

The intrinsic noise sources on the other hand are inherent to the device operation and are the focus of this work when addressing low noise FET design. These intrinsic noise sources, in the case of FETs, are represented by the *drain current noise* developed by the channel and the *induced gate noise*, which is also brought about by the channel under high frequency operation. [11]

The physics behind the generation of an induced gate noise from a drain current noise is that carriers taking part in the transport along the channel will experience scattering events which will make them move towards the gate oxide. At low frequencies, these fluctuations towards the gate oxide are balanced out by the sum of all similar scattering events throughout the extension from source to drain. At high

frequencies, however, the summation along the extension of the oxide-channel interface does not balance out and this imbalance will develop a net movement of the carriers in the channel towards the gate. This net movement induces a voltage at the gate node via the capacitance divider made by the gate oxide capacitance in series with the capacitance from the gate node to the AC ground. Because the focus of this thesis work is excess noise developed inside the FET channel, the frequencies used for the noise measurements reported later in this work were not high enough to make the induced gate noise noticeable.

1.3 Macroscopic, Mesoscopic and Microscopic Dimensions

The approach to noise phenomena in this work is made within a framework of transport theory and it is important to revisit the techniques and concepts in transport theory as device dimensions change from macroscopic to microscopic.[29, 30]

Microscopic dimensions are the easiest to define. This refers to dimensions of the order of the size of atoms and is a definition not related to any transport characteristics. Macroscopic and mesoscopic dimensions however, have special meanings in transport theory and are related to transport characteristics.

Macroscopic dimensions are those to which the conductors behave *'ohmically'*. That is, the DC conductance of a rectangular conductor can be written as:

$$G = \sigma W/L \tag{1.4}$$

where G is the conductance of the rectangular conductor, W and L are the width and length of the rectangular conductor, and σ is the electrical conductivity of the material.

Mesoscopic dimensions are intermediary dimensions between macroscopic and microscopic ones. Mesoscopic dimensions are of the order of any of the following dimensions: 1) the *de Broglie wavelength*, 2) the *carrier mean free path* between collisions, or 3) the *carrier phase relaxation length*. These three dimensions refer to the average

distance over which kinetic energy, momentum, or phase of the carriers is randomized by the scattering mechanisms, and thus depend strongly on temperature and the material used.

Figure 1.7 provides an idea of the typical dimensions involved in the mesoscopic domain.

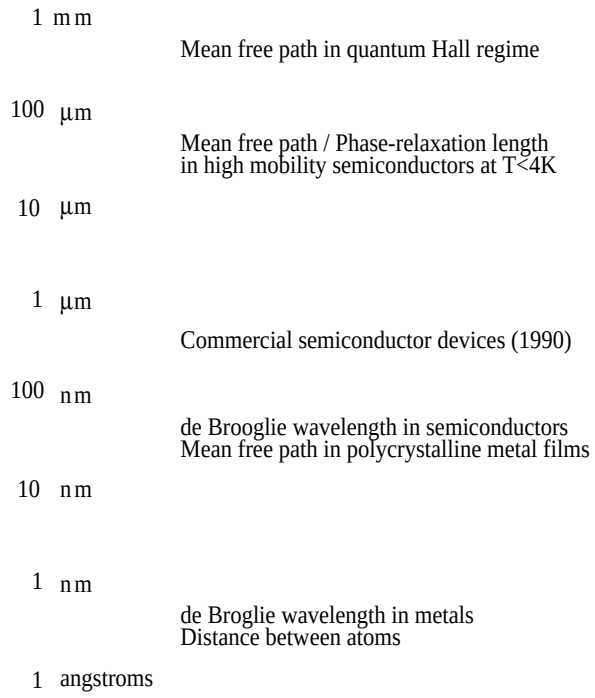

Figure 1.7: Typical dimensions involved by the mesoscopic denomination. $1\mu m = 10^{-6}m$ and $1nm = 10^{-9}m = 10\text{Å}$.

In mesoscopic transport, the wave nature of the carriers becomes evident and oscillations in the device conductance with voltage preclude a description of the device by the *ohmic* equation above.

Note however, that most of the research interest in mesoscopic transport and noise focus upon very delicate transport phenomena rather than that applicable to practical FETs. [31, 32, 33]. These are phenomena that occur at very low temperatures or under highly constrained flow of carriers.

However, in real life applications, the transport characteristics need to be more robust and not change with minor environmental variations. The *Ergodic Method* to

be discussed in this work will start the noise and transport analysis under clearly macroscopic conditions at room temperature and will be extended into dimensions approaching the limits of current device technologies for FETs, around 100nm. Even these very small devices are operated in the macroscopic transport regime. Extensions of the theoretical formalism into the mesoscopic transport regimen are also discussed later in chapter 2, *"The Ergodic Method"*.

1.4 Common Noise Approaches

A number of approaches to noise phenomena in semiconductors under room temperature operation have appeared in the literature.[35, 36, 37] Some of the most commonly used approaches are briefly reviewed below to establish a starting point and help the reader to see clearly the contribution of this work.

1.4.1 Langevin Noise Source and Classical Semiconductor Equations

This approach to noise phenomena in mathematical form uses the basic equations that describe semiconductor-device operation, but adds a *Langevin Noise Source* [38, 39] to these equations. This *Langevin Noise Source* is in fact a *forcing term*, which will produce a fluctuation in the current or voltage at the device terminals.

Consider the set of equations 1.5 which can describe the operation of a p-n junction, below.[41]

$$\nabla^2\varphi = -\frac{q}{\epsilon}(p - n + N^+) \tag{1.5}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\nabla \cdot \underline{J_n} - U_n \tag{1.6}$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\nabla \cdot \underline{J_p} - U_p \tag{1.7}$$

where φ, q, ϵ, p, n, N^+, U_n and U_p are respectively the electric potential, elementary

charge, dielectric constant, hole concentration, electron concentration (in the conduction band), concentration of ionized donors (n-type semiconductor assumed), electron generation-recombination rate, and hole generation-recombination rate. The vector terms $\underline{J_n}$ and $\underline{J_p}$ are respectively the electron current and hole current densities.

In a more convenient computational representation, the above set of equations is written in matrix (vector) form as:

$$\underline{F}(\varphi, n, p, \dot{n}, \dot{p}) = 0 \tag{1.8}$$

where

$$\dot{\alpha} = \frac{\partial \alpha}{\partial t},\ \alpha = n, p \tag{1.9}$$

At this point, a vector Langevin Noise source is added to equation 1.8, which will give rise to fluctuations in the macroscopic values of potential, carrier concentration and rate of change of the carrier concentrations. This is normally written as:

$$\underline{F}(\phi + d\phi, n + dn, p + dp, \dot{n} + d\dot{n}, \dot{p} + d\dot{p}) = \underline{s} \tag{1.10}$$

where $\underline{s}$ is the Langevin Noise Source. From this point on, the job is to *adjust* $\underline{s}$ to make the fluctuations on the macroscopic values of current and/or voltage fit the experimental data. It's a legitimate approach to noise, but can hardly give any clue on what to do on the device design level to make new devices with better noise performance.

The central benefit of the *Langevin Noise source* approach is to be able to describe the fluctuation phenomena without deep knowledge of the actual mechanisms ultimately responsible for the fluctuation. It's more of a descriptive tool than a design tool.

1.4.2 Langevin Noise Source and Boltzmann Transport Equation

Similar to the previous approach to noise and bearing the same conceptual justification, this approach uses the Boltzmann Transport Equation instead of the basic

conduction equations of semiconductor-device operation. Since the basic equations of semiconductor-device operation can be derived from the Boltzmann Transport Equation formalism, this new approach had the appealing flavor of getting deeper into the semiconductor transport statistics, but comes with the highly disputable and unsupported concept of imposing a fluctuation on a statistical distribution function.

The Boltzmann-Langevin Equation appears in both the forms below: [15, 34, 35, 36, 38, 39, 40]

The first form is:

$$\frac{\partial f_c(r,k,t)}{\partial t} - \frac{qE(r,t)}{\hbar}\frac{\partial f_c(r,k,t)}{\partial k} + \frac{1}{\hbar}\frac{\partial \epsilon_c(r,k)}{\partial k}\frac{\partial f_c(r,k,t)}{\partial r} = -\frac{f_c(r,k,t) - f_o(r,k)}{\tau_c} + \gamma_c(r,k,t) \tag{1.11}$$

where $f_c(r,k,t)$ is the distribution function describing the number of electrons as a function of position, momentum and time respectively; qE, $\hbar$, $\epsilon_c(r,k)$ are the force, the Planck constant divided by 2π, and the permittivity of the medium, respectively. $\gamma_c(r,k,t)$ is the Langevin noise source, or forcing term added to the equation.

The second form is:

$$\frac{\partial n(r,k,t)}{\partial t} - \frac{F(r,t)}{\hbar}\nabla_k n(r,k,t) + v_k \nabla_r n(r,k,t) = \left(\frac{\partial n(r,k,t)}{\partial t}\right)_{coll} + \xi_c(r,k,t) \tag{1.12}$$

where $n(r,k,t)$ is the concentration of carriers as a function of position, moment and time, respectively. The correspondence to the first form is readily identified. The Langevin noise source can still be readily recognized by the addition of the term $\xi_c(r,k,t)$.

In this second form the only significant difference is the use of the *collision integral* in its full form, whereas in the first form an approximation known as the *Relaxation Time Approximation*, RTA, is used.

The *collision integral* is:

$$\left(\frac{\partial n(r,k,t)}{\partial t}\right)_{coll} = \sum_{p'} f(p')[1-f(p)]S(p',p) - \sum_{p'} f(p)[1-f(p')]S(p,p') \tag{1.13}$$

Under conditions where the distribution function, $f(r,k,t)$, is just a perturbation on the equilibrium distribution function, $f_o(r,k,t)$, the collision integral can be approximated by:

$$\left(\frac{\partial n(r,k,t)}{\partial t}\right)_{coll} = -\frac{f_c(r,k,t)-f_o(r,k)}{\tau_c} \tag{1.14}$$

The macroscopic measurable quantities to which the Langevin Noise source is adjusted can be either the total number of carriers or the total current density. The total number of carriers and the total current density are functions of the carrier distribution function, $f_c(r,k,t)$, and are written, respectively, as:

$$n(r,t) = \int \frac{d^3k}{4\pi^3} f(r,k,t) \tag{1.15}$$

$$J(r,t) = \int \frac{d^3k}{4\pi^3} v(r) f(r,k,t) \tag{1.16}$$

where $v(r)$ is the velocity of the carriers, and $n(r,t)$ and $J(r,t)$ are respectively the carrier concentration and electric current density at position, r, and time, t.

Notice however the non-mathematical concept of making $f_c(r,k,t)$, which is already a full description of a random variable, "fluctuate". [15] There is thus still the further limitation of being a descriptive noise approach with little ties to device design issues.

1.4.3 Impedance Field Method

The Impedance Field Method was introduced by William Shockley [42]. The effect of fluctuations on voltage or current by the device terminals is calculated as the addition of the contributions of many noise sources inside the device.

Small AC current sources are defined for internal positions along the device and every one of these small sources have an effect that can be mapped to the terminals as:

$$\delta V(r,f) = \nabla Z(r,f).\nabla J(r,f)d^3r \tag{1.17}$$

where $\delta V(r,f)$ is the fluctuation in voltage of frequency f at a position r of the device. $\nabla J(r,f)d^3r$ is the current fluctuation of frequency f at a position r inside the device. $\nabla Z(r,f)$ is the impedance factor relating the above voltage and current fluctuations.

The total effect of the contributions of internal current noise sources upon the terminal of the devices can be found by integration, assuming uncorrelated noise sources. The spectrum of the voltage noise at the terminals of the device can be written as :

$$S_V(f) = \int\int [\nabla Z(r,f)]^t [S_j(r,r',f)][\nabla Z^*(r',f)]d^3r d^3r' \tag{1.18}$$

where $S_j(r,r',f)$ is the spectral content of the current noise sources along the semiconductor device and $S_V(f)$ is the spectral content of the voltage noise at the device terminals.

Notice that the assumption of uncorrelation of the internal current noise sources is a consequence of the fact the device would be sliced into sections thick enough to accommodate several *mean free paths*, λ, i.e. several scattering events are expected to be experienced by a carrier traveling through each semiconductor section. The general result, however, is again a descriptive approach to noise rather than a design approach.

1.5 Preludes of the Ergodic Method

The next chapter will introduce the Ergodic Method. The method approaches noise from a perspective which brings an understanding of the mechanisms producing noise phenomena inside the semiconductor material and connects this understanding to device design issues. Therefore this will provide a tool not only for description of noise, but will be also a tool for device optimization for low noise performance.

Noise is fluctuation in carrier velocity and the current noise power is a summation over the fluctuations in velocity of each of the carriers taking part in the transport. This means noise is related to *variance* in the *time domain* statistics analysis. In order to harness the already developed tools for carrier transport, an *ergodic hypothesis* is introduced in this work so that the *ensemble* statistics of the carrier velocity distribution can be used instead of the *time statistics* of one representative carrier.

The Boltzmann Transport Equation, BTE, is used by the Ergodic Method as the tool to assess the *ensemble statistics* for the carriers taking part in the transport.

The *Ergodic Method* is shown to recover the Nyquist expression for Johnson noise (*thermal noise*) as a *special case* for equilibrium. This is achieved as the limit case for the transport under very low electric fields (or very low DC-current).

The *Ergodic Method* does *not* use any Langevin Noise Source or forcing term to retrieve noise power information from the transport equations. That is a fundamental difference from the previous methods discussed. Within the *Ergodic Method* the Scattering Rate Matrix, $S(p, p')$, used in the BTE, is the foundation of both the I-V curves of the device and its generated noise power. It is exactly because semiconductor material can be described by the $S(p, p')$ matrix that the *Ergodic Method* can make this matrix be the building block for noise analysis of new devices at *design time*.

The kernel of the *Ergodic Method* is the concept of noise directly related to *variance* in carrier velocity distribution. Therefore the path for the ultimate low noise FET is to force the carriers to travel as close to the same velocity as possible. The predictions from the *Ergodic Method* on noise performance of semiconductor-devices under high electric fields are experimentally demonstrated using state-of-the-art silicon MOSFETs.

Because of the results of this work, novel CMOS transistors can be designed for low noise operation. Three approaches have been identified and can be used simultaneously in the design of new devices: the use of *heavier carriers* whenever a hit in device speed is allowed for trade with low noise operation, the use of *laterally graded doping profile in the channel* and the use of *biaxial mechanical stress in the channel*. The first approach stems from the fact that heavier carriers are harder to have their velocity deflected by scattering mechanisms and the heating effect of the applied electric field is modulated by the mass of the carrier used in the transport also favoring heavier carriers for low noise operation. The second approach tailors the availability of carriers along the channel to maintain a tighter final velocity distribution. The third idea tailors the energy bands to promote carrier confinement and avoids having carriers with different *effective masses* taking part in the transport. This significantly reduces the *"velocity spreading"* compared to the normal case for Si where carriers of two different effective masses take part in the transport. Both type of carriers see the same accelerating applied electric field, but have different average velocities and their velocities have different distributions.

Chapter 2

The Ergodic Method

The Ergodic Method is introduced in this chapter. The objective is to develop a noise approach from a transport theory framework. This noise approach however should be closely related to device design issues in order to provide both insight into noise phenomena and device design guidelines for low noise operation.

Several assumptions on device operating conditions, which are consistent with real applications make the transport problem *macroscopic* in nature, but will still allow for device noise analysis when device dimensions enter into the deep sub-micron range (sub-micron MOSFET case).

The principle behind the presentation of the Ergodic Method is to develop the concepts in this chapter based on a semi-axiomatic method. Several starting assumptions are accepted *a priori*, and expected to be fully justified by the correspondence between the prediction of these assumptions and the actual experimental results. The experimental setup and results will be described and analyzed in the next chapters.

The extensions of the Ergodic Method are anticipated for the case of *mesoscopic* transport, but no experimental results will be discussed.

2.1 One Carrier Analysis and the Ergodic Assumption

The transport analysis for the Ergodic Method starts by assuming a non-degenerate semiconductor. That means there are many quantum states available compared to the number of carriers. Thus one carrier can scatter from its present quantum state to any other possible quantum state without finding it occupied.

Under thermal equilibrium conditions, the carriers inside this non-degenerate semiconductor will wander around, describing Brownian motion as depicted in figure 2.1(a). The momentum changes experienced by the carriers are produced by their interaction with the phonons of the semiconductor crystal. The phonons are assumed to be a huge reservoir at the temperature of the crystal lattice, thus making the carriers behave as if they were immersed in a phonon bath.

At thermal equilibrium, the carriers are in equilibrium with the phonon bath. It is then assumed that the lattice temperature is high enough to provide an abundance of phonons and to correct promptly any disturbance in the average energy of the carriers.

Once a small bias is applied to the semiconductor, the carriers will acquire an average drift velocity superimposed upon the Brownian motion, and one ordinary carrier will wander as depicted in figure 2.1(b).

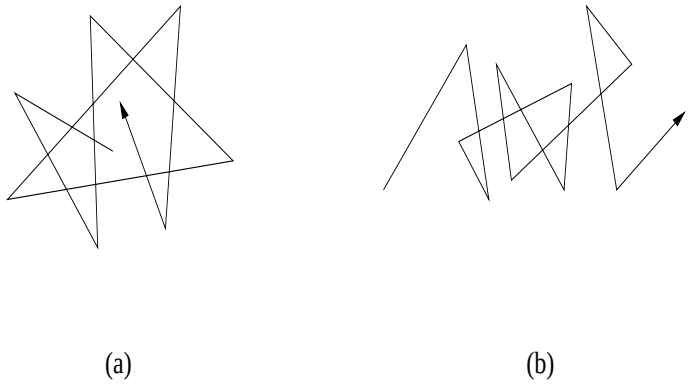

Figure 2.1: Typical carrier movement inside a non-degenerate semiconductor. (a) thermal equilibrium. (b) under DC bias.

Thus, under bias, the carriers will have and average drift velocity, v_d, plus a random velocity *fluctuation* component. The summation of the average velocity plus random velocity fluctuation will be called *total velocity* of the carrier.

Let's define $g(r, v, t)$ as the function which describes the *time* statistics of the *total velocity* of the carriers as function of position and time; so that if one ordinary carrier has its movement tracked *in time*, it will be seen that this carrier acquires different velocities due to scattering events according to the function $g(r, v, t)$.

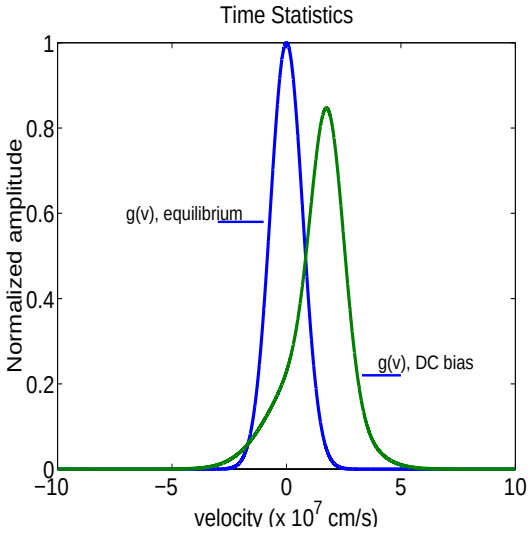

Figure 2.2: Time statistics of a single carrier at equilibrium (zero-velocity centered) and under DC bias (non-zero-velocity centered).

Assuming the random wandering of the carriers to assume velocities described by a stationary random process, $g(r, v, t)$ will be referred to as $g(r, v)$ only. $g(r, v)$ will be a Gaussian centered at *zero* velocity when the conductor is in thermal equilibrium. Under bias, $g(r, v)$ will change its mean and also its general shape. Its mean will shift from zero-velocity centered to a non-zero velocity to produce the correct average velocity for the carriers. This average velocity has to be consistent with the number of carriers and the DC current flowing through the semiconductor. Assuming a constant concentration of carriers throughout the semiconductor device, the general shape of $g(r, v)$ may change from a Gaussian shape towards a significantly skewed shape depending upon the strength of the applied electric field. At a point *"r"* inside the

semiconductor-device, figure 2.2 depicts the velocity distribution (time statistics) of a carrier at equilibrium and a possible distribution (also time statistics) for this carrier at a bias capable of producing a high applied electric field. The shift in mean and the skew on the time statistics under DC bias is exaggerated for clarity of purpose only.

From this one-carrier time statistics, an average over the total number of carriers in a thin slice of semiconductor can be written. Consider the slice of semiconductor in figure 2.3, where a DC current, I, is flowing from left to right, and inside the slice of transversal area A and length dL, there are N carriers.

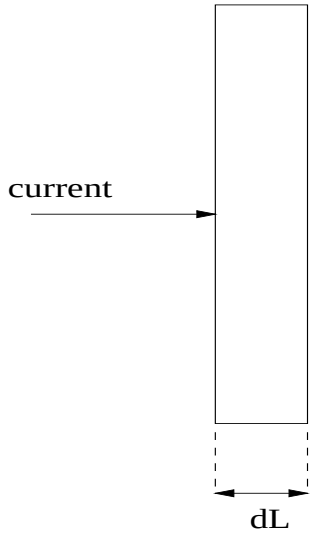

Figure 2.3: Slice of semiconductor of transversal area A and length dL supporting a DC current I. The number of carriers present in this slice is called N.

The *current noise power*, $\overline{i^2}$, developed in this slice and the *DC current* are related by:

$$\frac{\overline{i^2}}{I^2} = \frac{< (\sum_N (v - v_d))^2 >}{< (\sum_N v_d)^2 >} \tag{2.1}$$

where $< x >$ is the *expected value* for the random variable x; and v_d is the mean velocity of the carriers. If the fluctuations of carrier velocity are statistically *uncorrelated*, then:

$$\frac{<(\sum_N (v-v_d))^2>}{<(\sum_N v_d)^2>} = \frac{N<(v-v_d)^2>}{N^2 v_d^2} = \frac{<(v-v_d)^2>}{N v_d^2} = \frac{\sigma_g^2}{N\mu_g^2} \tag{2.2}$$

where σ_g^2 and μ_g are the *variance* and the *mean* of the **time statistics** distribution function *g(r,v,t)*. N is the number of carriers inside the slice of semiconductor. Therefore, from equations 2.1 and 2.2:

$$\frac{\overline{i^2}}{I^2} = \frac{\sigma_g^2}{N\mu_g^2} \tag{2.3}$$

Equation 2.3 requires information on the **time statistical** distribution function of the movement of one representative carrier for a long time in order to calculate the current noise power developed inside the semiconductor slice.

Note however that several formalisms in transport theory have developed tools to assess the **ensemble statistics** distribution function inside a semiconductor. That is, solid state physics is strongly developed on the notion of *distribution function*, meaning "number of carriers in specific energy levels", and this notion is **not** based on following "one representative carrier in time".

Nevertheless, if in some limited conditions, the *ensemble statistics* can be related to the *time statistics* above, an adequate noise calculation procedure can be devised to use the already existing set of transport analysis tools.

These two distribution functions are normally different, since the *ensemble* distribution function can describe electrons that are deep in the core of the atoms, and thus not participating in the transport. However, if we consider *only* energies above the conduction band for electrons and *only* energies below the valence band for holes, the *ensemble distribution function* will inform only the distribution of carriers with "excess energy" that will be considered as being solely *kinetic energy* of the carriers. Under these conditions, a plausible **ergodic assumption** can be made stating that the *ensemble distribution function*, $f(r, \hbar k/m, t)$, and the *time distribution function*, *g(r,v,t)*, are equal.[43] Therefore:

$$\frac{\overline{i^2}}{I^2} = \frac{\sigma_g^2}{N\mu_g^2} = \frac{\sigma_f^2}{N\mu_f^2} \tag{2.4}$$

where σ_f^2 and μ_f are the variance and mean of the (ensemble) distribution function $f(r, \hbar k/m, t)$. In simple words, the *ergodic assumption* means that if the piece of semiconductor were infinitely large and had the same (ensemble) distribution function all over and there are 100 more carriers at velocity v_1 than at velocity v_2, then if one ordinary carrier is tracked in *time*, it will occupy a state at velocity v_1 100 *units of time* greater than it will occupy a state at velocity v_2.

This equation, 2.4, will be referred throughout this thesis text as the *fundamental ergodic equation for noise*.

Given the impracticality of having an infinitely large piece of semiconductor, the assumption is taken by its plausible intuitive notion and ease of practical use. It is then expected that the experiments will prove this *ergodic assumption* correct.

Equation 2.4 points out that whenever possible, everything else being equal, low noise operation is achieved by narrowing the carrier velocity distribution inside the semiconductor-device where the carrier transport takes place.

2.2 The Boltzmann Formalism

In order to use equation 2.4, above, it is necessary to assess the carrier (ensemble) distribution function, which from now on will be referred to as simply the *distribution function*, $f(r, v, t)$, and it is to be understood this distribution has only information related to momentum/velocity or kinetic energy of the carriers.

The assessment of the carrier distribution function can be made via Monte Carlo [44, 45, 46] based simulations or via the Boltzmann Transport Equation.[47, 48, 49] In this work, the Boltzmann Transport Equation is chosen and used due to its compactness and ease to see connections between the current noise power and the factors influencing the level of this noise power.

Since the Boltzmann Transport Equation, BTE, is used, the assessment of carrier velocity distribution becomes a semi-classical procedure, where the carriers are

really seen as particles and only the scattering mechanisms are evaluated quantum mechanically. Moreover, it is assumed that consecutive scattering mechanism do not interfere with one another.

The rigorous derivation of the BTE can be found in the literature [50, 51, 52, 54, 55], and only a qualitative presentation is reproduced here.

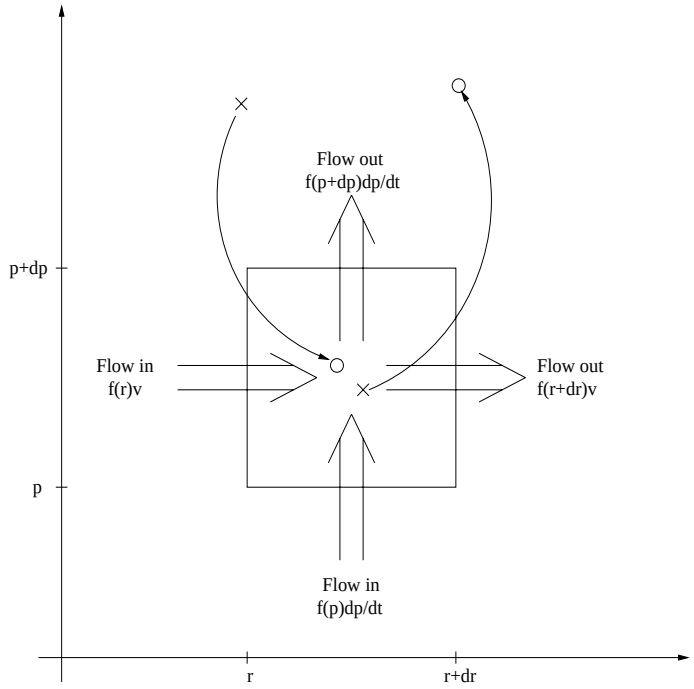

Figure 2.4: The Boltzmann Transport Equation describes the rate of change in the number of carriers in a *"box"* for each momentum, due to the effect of drift, diffusion and scattering with phonons.

As depicted in figure 2.4, the BTE extends the general concept of writing the rate of change of the number of carriers inside a box due to the effects of drift, diffusion and generation-recombination, to the concept of writing the rate of change of number of carriers inside a box for *every specific momentum*, due to the same effects of drift, diffusion *plus* the (random) effect of scattering by phonon-electron or other interactions, which may include generation-recombination of carriers..

Therefore, the BTE is written in the form:

$$\frac{\partial f}{\partial t} + \boldsymbol{v}.\boldsymbol{\nabla}_{\boldsymbol{r}} f + \boldsymbol{F}.\boldsymbol{\nabla}_{\boldsymbol{p}} f = \left.\frac{\partial f}{\partial t}\right|_{coll} \tag{2.5}$$

where $f(r,p,t)$ is the carrier velocity distribution function, which is a function of position, velocity (momentum) of the carriers in the semiconductor and time. $\boldsymbol{v}$ is the velocity of the carrier, $\boldsymbol{F}$ is the electric field force felt by the carrier, and $\frac{\partial f}{\partial t}|_{coll}$ is the collision integral. The collision integral is written as:

$$\begin{aligned}\left.\frac{\partial f}{\partial t}\right|_{coll} &= \sum_{p'} f(p')[1-f(p)]S(p',p) \\ &\quad - \sum_{p'} f(p)[1-f(p')]S(p,p')\end{aligned} \tag{2.6}$$

$S(p,p')$ is the *Transition Rate Matrix*. $S(p,p')$ is the rate of carrier transfer from momentum p to momentum p'. Accordingly, $S(p',p)$ is the rate of carrier transfer from momentum p' to momentum p.

The BTE is thus a complicated differential-integral equation where the unknown distribution function $f(r,p,t)$ appears in the differential expression in the left-hand side of the equation 2.5 and also in the summation on the right-hand side of the same equation.

The BTE in its original form gives information on both the *time evolution* and *steady state* of the velocity (momentum) distribution of the carriers. In this form, the BTE requires that the scattering events not interact with each other, in the sense that every element in the *transition rate matrix*, $S(p,p')$, has no memory of the previous scattering event. This is a prohibitive requirement for analysis of semiconductor devices in the *mesoscopic regime* and this case will be discussed later in this chapter.

The Scattering Rate Matrix, $S(p,p')$, is calculated using *Fermi's Golden Rule*,[56, 57] which assumes scattering as small perturbations to the carrier flow and proceeds representing the disturbing potential due to phonons or ionized impurities via a suitable interaction function, $U_s(z,t)$. A carrier with momentum $\hbar k_o$ at $t = 0$ interacts with the disturbing potential, $U_s(z,t)$, and emerges with momentum $\hbar k_o'$, as depicted in figure 2.5.

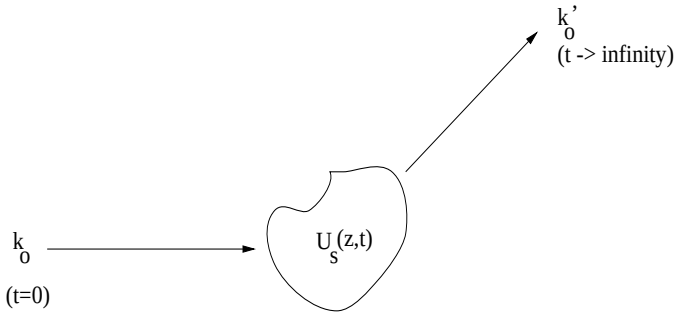

Figure 2.5: Representation of carrier being scattered by a disturbing potential $U_s(z,t)$. The carrier with initial momentum $\hbar k_o$ is scattered to momentum $\hbar k'_o$.

Thus, considering the incoming wave function, $\Psi_k(z)$, and the scattered wave function, $\Psi^*_{k'}(z)$, the Matrix Element, $H_{k'k}(t)$, is defined as:

$$H_{k'k}(t) \equiv \int_{-\infty}^{+\infty} \Psi^*_{k'}(z) U_s(z,t) \Psi_k(z) dz \tag{2.7}$$

The Scattering Rate Matrix, $S(p,p')$, is then:[50, 52]

$$S(k,k') = \frac{2\pi}{\hbar}|H^a_{k'k}|^2 \delta[E(k') - E(k) - \hbar w] + \frac{2\pi}{\hbar}|H^e_{k'k}|^2 \delta[E(k') - E(k) + \hbar w] \tag{2.8}$$

where there is the possibility of having one photon absorbed or one photon emitted. $E(k')$ and $E(k)$ are the energies of the carrier with momentum $\hbar k'$ and $\hbar k$, respectively.

The calculation of $S(k,k')$ is the only point when the quantum mechanical concepts of the carrier are considered in this semi-classical approach for transport.

It's important to notice that for transport analysis via the BTE, it is necessary to have the $S(p,p')$ matrix fully characterized for the material where the transport is to take place. Such a matrix can be either built from first principles alone or be based upon first principles and experimental data fitting. The second option (first principles plus experimental data fitting) is more attractive and feasible, and means that the structure of $S(p,p')$ is already known for the scattering mechanisms present in a given transport media, but several coefficients in $S(p,p')$ still need to be adjusted. These coefficient adjustments are done based on experimental results already taken

from the same transport media. The fact that the structure of $S(p, p')$ is physically based guarantees that after the coefficient adjustments, this matrix can be used in calculations to predict noise performance for other devices based on similar transport media at different bias points.

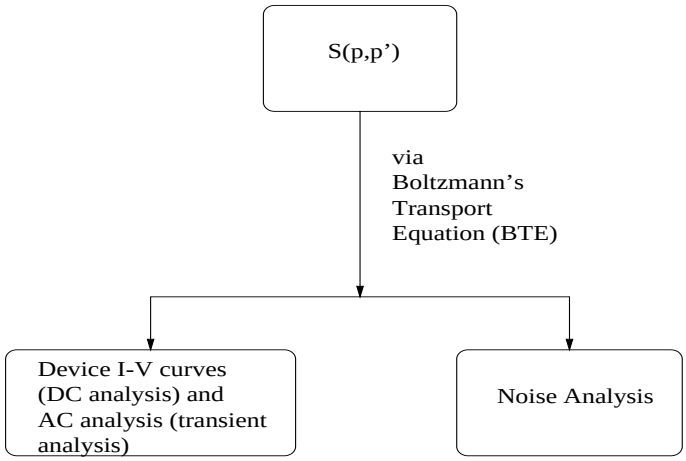

Figure 2.6: The *Ergodic Method* uses the Scattering Rate Matrix, $S(p, p')$, at the center of transport analysis. Thus, both the I-V/AC analysis and noise analysis of new devices become connected to the *average* and *variance* measures of the carrier velocity distribution, respectively.

Figure 2.6 stresses the conceptual difference between the *Ergodic Method* and other methods for noise analysis. Note that there is *no* Langevin Source in this method. The matrix $S(p, p')$ is at the heart of device operation and describes completely the random component of carrier transport inside the semiconductor material. Therefore, $S(p, p')$ will define the constant velocity carriers acquire under bias and also their fluctuation in velocity. Consequently, $S(p, p')$ will also define the I-V curves of the device, its transient behavior and its noise performance. In practice, since the main scattering mechanisms in a given semiconductor have known formulae, adjusting coefficients can be added to make the predictions of the BTE (using $S(p, p')$) match the experimental results for both I-V/AC analysis and noise analysis of the material/device. In the next step, $S(p, p')$ can be used as the core knowledge for the simulation/prediction of noise performance of new devices.

Through the *Ergodic Method*, the I-V/AC analysis of a device is connected to the *average* quantities of carrier velocities, and the noise analysis is connected to the

variance of carrier velocities in a unified approach. Note also that all of the scattering events are regarded as momentum changes. Even fluctuations in the number of carriers can be treated as momentum changes because a carrier that is trapped, can be seen as changing its momentum from a finite value to *zero momentum*. Conversely, a carrier which is released from a trap can be treated as changing its momentum from *zero momentum* to a finite momentum.

2.3 Recovering Thermal Noise

The Nyquist Theorem is recovered for current noise under equilibrium condition. The current noise for this special case is calculated by the limit of equation 2.4

$$\overline{i^2}|_{equilibrium} = \lim_{I \to 0} I^2 \frac{\overline{i^2}}{I^2} \tag{2.9}$$

This limit is investigated with the use of a *long piece of semiconductor* under bias. Consider a piece of semiconductor under electric field, $\boldsymbol{\mathcal{E}}$. The semiconductor is assumed to be in steady state and is also long enough such that the carrier distribution function, $f(r,p)$, does not change with position. In this case, the applied electric field is a minor perturbation of the equilibrium condition and the Relaxation Time Approximation (RTA) holds.[50] Hence, the BTE can be written as:

$$(\text{-q})\boldsymbol{\mathcal{E}}.\boldsymbol{\nabla}_{\boldsymbol{p}} f = -\frac{f_A}{\tau_o} \tag{2.10}$$

where τ_o is a time constant independent of the momentum of the carriers. [50, 58]

Hence, the collision integral is reduced to a simple form via RTA, and f is of the form $f = f_o + f_A$, where f_o is the equilibrium distribution and f_A is a small perturbation in the equilibrium distribution due to the applied electric field. Further simplifications assumed are that $f \approx f_o$, and that $\boldsymbol{\nabla}_{\boldsymbol{p}} f \approx \boldsymbol{\nabla}_{\boldsymbol{p}} f_o$. This then yields

$$f = f_0 + q\tau_o \boldsymbol{\mathcal{E}}.\boldsymbol{\nabla}_{\boldsymbol{p}} f_o \tag{2.11}$$

Assuming $\boldsymbol{\mathcal{E}}$ is directed along $\hat{z}$, equation 2.11 becomes $f = f_o + q\tau_o \mathcal{E}_z \frac{\partial f_o}{\partial p_z}$; which

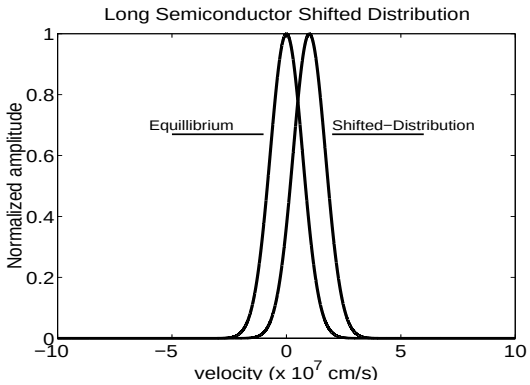

Figure 2.7: Distribution Function for a long piece of semiconductor in *equilibrium* and with an applied electric field. The distribution function f under applied electric field relates to the equilibrium distribution function f_o as $f(p_x, p_y, p_z) = f_o(p_x, p_y, p_z + q\tau_o\mathcal{E})$, where $\mathcal{E}$ is the applied electric field in the z-direction, in this example.

leads to $f(p_x, p_y, p_z) = f_o(p_x, p_y, p_z + q\tau_o\mathcal{E}_z)$. Thus, the final distribution function in a long semiconductor is a shifted-Maxwellian distribution and is depicted in figure 2.7. (an exaggerated shift is shown only for the sake of illustration.)

Then,

$$f = Ae^{[E_f - E_{CO}(r) - (p-p_o)^2/2m^*]/K_B T_L} \tag{2.12}$$

where $E_f - E_{CO}$ is the potential energy (referred to the Fermi Level) of the carrier in the conduction band and $(p - p_o)^2/2m^*$ is the kinetic energy. m^* is the carrier effective mass (which depends on the semiconductor system), and A is a normalization constant.

It is then clear that equation 2.12 is the expression of a Gaussian distribution function multiplied by the number of carriers participating in transport. This form of f in this development is totally due to the fact that it is defined to include only information on electrons with energies above the conduction band level, or holes with energies below the valence band. That is, equation 2.12 can be written in the form:

$$f = N\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(p-p_o)^2/2\sigma^2} \tag{2.13}$$

where N is the number of carrier taking part in the transport. Concentrating only in

the normalized Gaussian form, by inspection:

$$\sigma^2 = \overline{(v - v_d)^2} = \frac{K_B T_L}{m^*} \tag{2.14}$$

The reader might be more used to see the above equation written as $(1/2)\overline{(v - v_d)^2} = 3K_B T_L/2$. The more familiar expression is a 3-Dimensional result for the electron gas and equation 2.14 can be also reached if *equipartition* of energy is assumed for the electron gas. The above result along the electron drift is useful for the 1-Dimensional analysis in this section. There is a division by 3 (three) in going from the 3-D to 1-D analysis which then renders equation 2.14, above.

Let $< \tau_m >$ be the mean time between collisions. It is well known from the literature[59] that the average drift velocity can be written as:

$$v_d = q\mathcal{E} < \tau_m > /m^* \tag{2.15}$$

The microscopic relation for conductivity through a slice of semiconductor of transversal area A and thickness d, is:

$$G = \sigma \frac{A}{d} = q \frac{q < \tau_m > n}{m^*} \frac{A}{d} \tag{2.16}$$

The number of electrons inside the slice of semiconductor is related to the concentration, $N = nAd$. The electric field and voltage are related via $\mathcal{E} = V/d$, and the macroscopic description of conductance as $I = GV$.

Thus, using the fundamental ergodic noise equation, equation 2.4, we reach:

$$< i^2 >< \tau_m >= GK_B T_L \tag{2.17}$$

However, in order to actually measure this current fluctuation (noise), the idealized measurement set described in figure 2.8 must be considered.

As can be seen, the noise measurement set has an input impedance matched to the real resistor, R, but presents much less noise power and is represented by a hypothetical noiseless input impedance. Under this circumstance, the voltage fluctuation developed in the input impedance of the measuring equipment is clearly

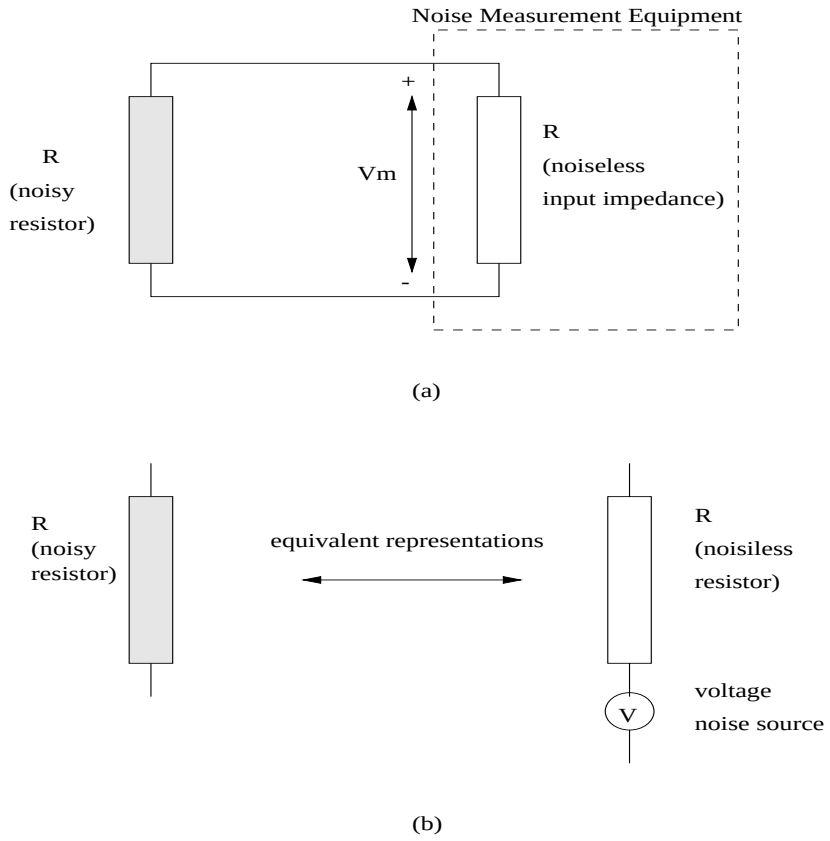

Figure 2.8: Noise measurement in an ideal instrument set. The slice of real resistor, R, has current fluctuation, $\overline{i^2}$, and the measuring equipment has its impute impedance matched to R, but hypothetically assumed to be *noiseless*. (a) The real resistor develops a voltage on the input impedance of the measurement equipment equal to $< v_m^2 >= RK_BT_L/ < \tau_m >$. (b) The same real resistor, R, can be represented by a noiseless resistor in series with a voltage noise source, $< v_m^2 >= {}_4RK_BT_L/ < \tau_m >$, in order to produce the same voltage as seen by the measurement equipment.

$< v_m^2 >= RK_BT_L/ < \tau_m >$. The same voltage on the terminals of the measurement set would be developed if the resistor, R, in figure 2.8(a), were replaced by a noiseless resistor in series with a voltage noise source as depicted in figure 2.8(b). This open-circuit noise voltage would have a value of:

$$< v^2 >< \tau_m >= \frac{(R+R)^2}{R^2} RK_BT_L = 4RK_BT_L \tag{2.18}$$

in order to account for the voltage divider represented by R and the input impedance of the measurement set.

Thus, if white noise is assumed for the current noise spectrum up to frequencies of $\frac{1}{<\tau_m>}$ and zero from this frequency up, all frequencies of practical interest would be encompassed and the above expression evolves to

$$S_v(f) = \frac{1}{\Delta f} < v_m^2 >= 4RK_BT_L; \text{ for } f < \frac{1}{<\tau_m>} \tag{2.19}$$

which is the familiar macroscopic expression for the thermal noise power spectrum, proving equation 2.4 does lead to the Nyquist Theorem under equilibrium. This is a special case reached in the limit of very low applied electric fields (or very low DC-current).

2.4 General Noise Calculation and Extensions Towards Sub-micron Dimensions and Mesoscopic Transport Regimen

The general noise calculation for a slice of semiconductor, which is a building block inside a device is straightforward if we assume the slices of the semiconductor are thick enough to accommodate several mean free paths, λ. In this case, the procedure follows the recipe described in figure 2.9, where the assessment of the carrier velocity distribution is made for every slice via BTE and the current noise source for every slice, according to the *Ergodic Method*, is calculated via the fundamental ergodic equation for noise, reproduced below.

$$\frac{\overline{i^2}}{I^2} = \frac{\sigma_f^2}{N\mu_f^2} \tag{2.20}$$

In the next step, the current noise sources of every slice are converted to voltage noise sources, using the value of the slice resistance.

These voltage noise sources are assumed *statistically uncorrelated* due to the thickness of the slices of several mean free paths, λ (extension to sub-micron devices is discussed in section 2.4.3). The sum of all the individual slice contributions give the total voltage noise at the device terminals.

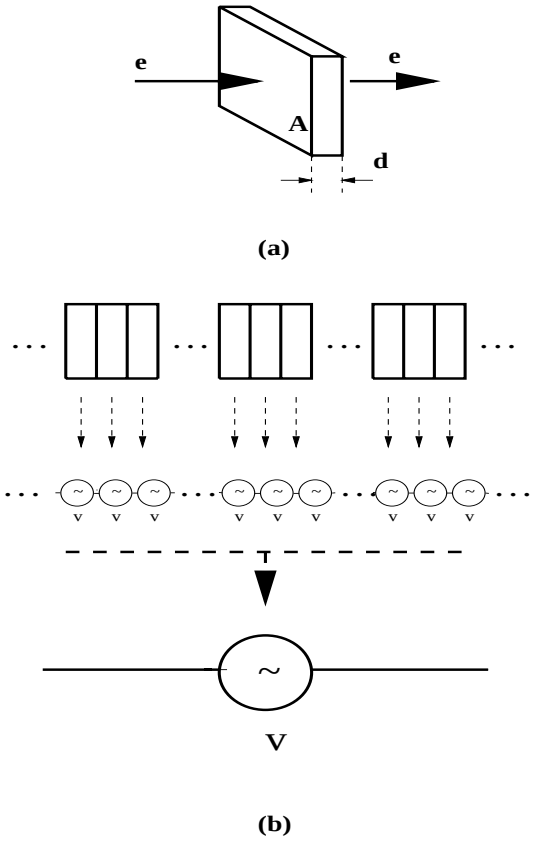

Figure 2.9: The noise analysis in carrier transport in semiconductor devices is based upon slicing the device in fine blocks along the carrier transport path. The transversal area, A, faces the carrier flow. The volume Ad contains a number of carriers referred to as N in the noise computation.(a) Shows the slab of semiconductor and (b) shows the current noise power for each slab being converted in voltage noise power, and being added to produce the total voltage noise power of the device. Each voltage noise source is considered uncorrelated to the other voltage sources and their effects are added in *variance*.

Underlying this procedure is the idea of noise sources as localized sources in each slice, as depicted in figure 2.10. It's then important that the carrier traveling through the transport media be subjected to scattering events in any slice.

This is an important observation because the Ergodic Method is easily extendible to vacuum devices as depicted in figure 2.11. But this extension needs to be done mindfully. Throughout the vacuum, it would be senseless to think about slices of transport media, since no scattering event can occur there. Therefore, it is an important assumption behind the Ergodic Method that slices and distribution functions

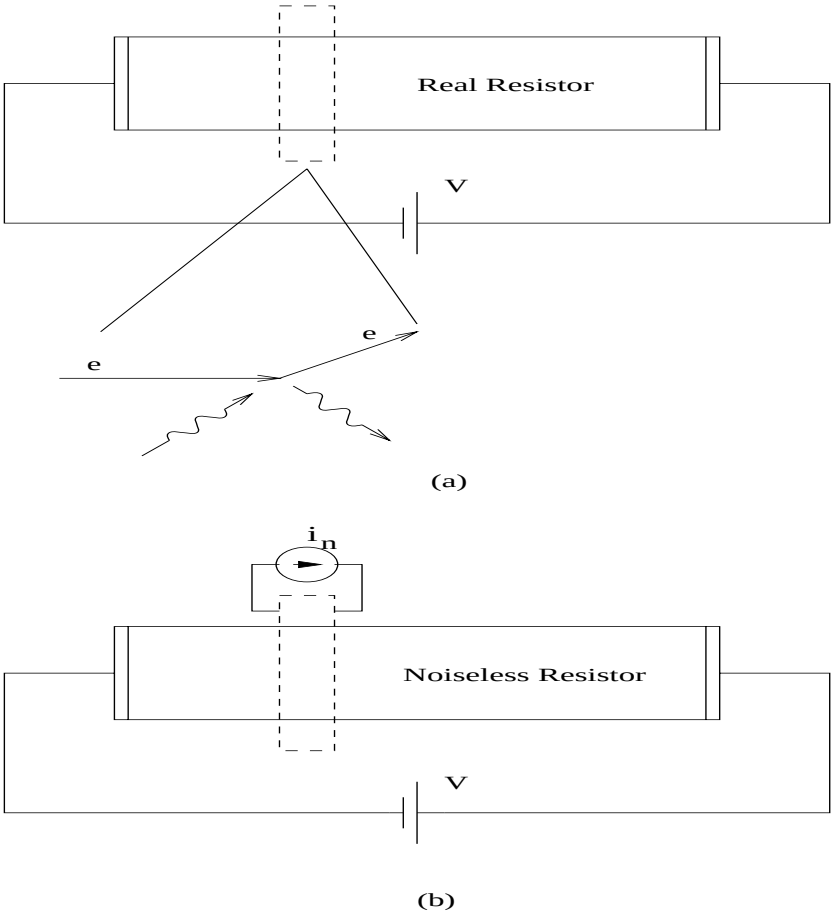

Figure 2.10: Slices are made for noise calculations only where carriers can be scattered. (a) shows a slice made in a *real resistor* and depicts a phonon scattering event. (b) shows the equivalent effect of imagining a *noiseless resistor* and adding a current noise source in parallel to each slice of the transport media.

have only meaning for noise calculations **if** a carrier can suffer a scatter event inside those slices. As an special case, the slices on a vacuum diode noise analysis only appear at the energy barrier from the cathode metal contact to the vacuum.

2.4.1 Computational Procedure for BTE

The calculation of the carrier velocity distribution function throughout the channel, iteratively using the scattering matrix and BTE tends to give computational convergence difficulties. This problem is particularly enhanced in this work because the

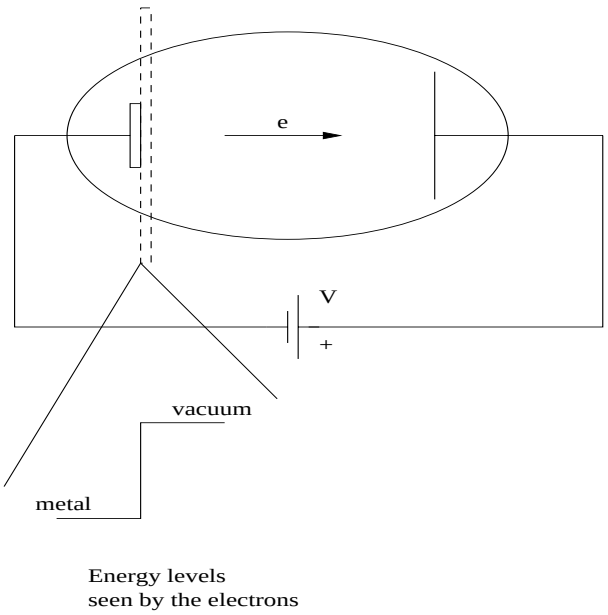

Figure 2.11: In a vacuum diode, slices are only used and analyzed at the transition from metal to vacuum. It is assumed that once set in motion in vacuum, the electrons do not experience any scattering events on their transit to the anode. No slice is used or analyzed in plain vacuum.

emerging distribution function for one slice is used as the initial distribution in the subsequent slice.

However, a more robust approach is developed if the solution of the BTE is calculated by minimizing a residue function, $R(r,p)$, to be defined later, with the constraint of trying only *physically meaningful* profiles for the distribution function, $f(r,p)$. These *physically meaningful* profiles are Gaussian due to the large number of carriers and quantum states involved in each scattering mechanism present in the transport media. This also leads to a reduction in computational load since entire carrier velocity distributions can be defined by a few numbers describing these Gaussian profiles.

Figure 2.12 shows a typical carrier velocity distribution function for a piece of pure silicon under high electric field.[50] Note the highly non-Maxwellian profile. Figure 2.13 shows that the same non-Maxwellian profile of figure 2.12 could be reproduced with three Gaussian profiles, arguably recognized as coming from three different scattering mechanisms.

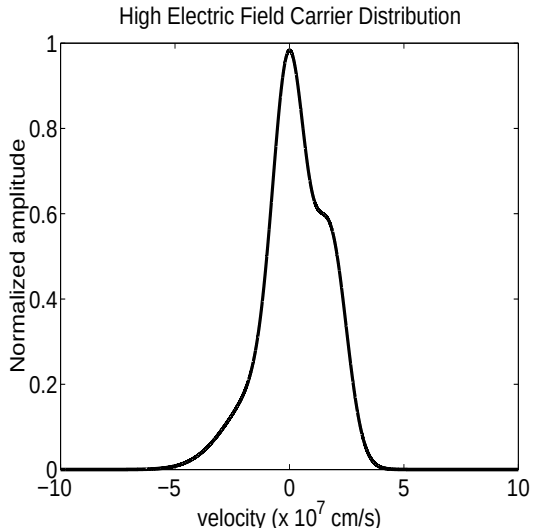

Figure 2.12: Typical distribution function for electron velocities in pure silicon under high electric field (F=100kV/cm).

Therefore, instead of trying to solve the BTE for each momentum directly, a solution for the entire distribution function will be calculated by first *guessing* the final distribution function, $f(r,p)$, should be a summation of multi-Gaussian profiles. Every Gaussian profile can then be associated with either specific scattering mechanisms or specific effective masses of carriers taking part in the transport.

It's important to mention that the summation of several Gaussian shapes as described above may produce fine structure in the envelop of the distribution function that is not present in an *"exact"* solution of the BTE. This is of secondary importance since only the *variance* and *mean* of the distribution function are relevant for the noise analysis.

Another *a priori* assumption made here is that no matter how many Gaussian profiles are defined for appropriate analysis of the transport, a carrier can scatter from one Gaussian profile to another with equal probability. Moreover, it will be the total summation of Gaussian profiles that will describe the probabilities of a carrier acquiring specific velocities (momentums). This is an assumption with some difficulties under low field transport as the scattering within a Gaussian profile may mean intra-valley scattering, and scattering from one Gaussian profile to another Gaussian

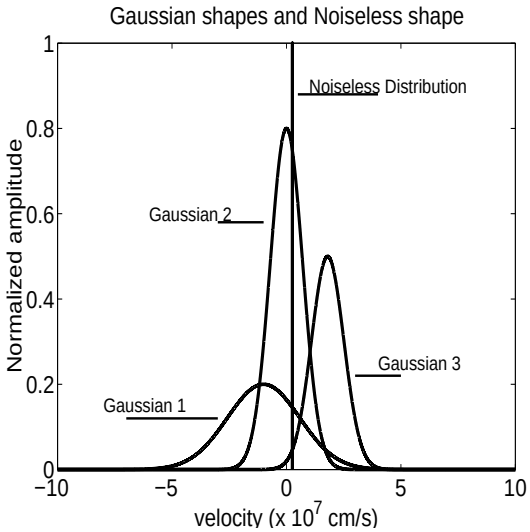

Figure 2.13: The same distribution of electron velocities of figure 2.12 is represented here as the sum of three Gaussian profiles. The hypothetical *noiseless* distribution is also depicted.

profile may mean inter-valley scattering. However, as the carriers enter the *high field regime*, the difference in probability between intra- and inter-valley diminishes and the above approximation becomes increasingly acceptable approximation. The experimental investigations described in the next chapter will demonstrate that the above *a priori* assumption is highly adequate.

As was explained in previous section, the carrier transport media of the semiconductor device under analysis is sliced into fine slabs. For each slab, a distribution function is calculated from the BTE. This is done iteratively, finding the best Gaussian profile for $f(r,p)$, that minimizes a residue function, $R(r,p)$. $R(r,p)$ is defined as:

$$R(r,p) = \frac{\partial f}{\partial t} + \boldsymbol{v}.\boldsymbol{\nabla_r} f + \boldsymbol{F}.\boldsymbol{\nabla_p} f - \left.\frac{\partial f}{\partial t}\right|_{coll} \tag{2.21}$$

The first try for $f(r,p)$ adjusts the amplitude, variance, and mean of the first Gaussian profile in order to minimize $R(r,p)$. Once the best fit is found, second Gaussian profile is added to first Gaussian profile in order to further decrease $R(r,p)$. This process is repeated for as many Gaussian profiles as it is necessary to fully describe the scattering mechanisms or the different effective masses of carriers involved

in the transport.

It should be noted that the integral from $-\infty$ to $+\infty$ for each Gaussian profile should be consistent with either the different scattering rates; or with the different number of carriers of specific effective mass.

The total mean of the final profile for the slab (slice of semiconductor device) is referred to the *global mean* and is the mean of the summation of all the Gaussian profiles. This *global mean* is marked in figure 2.13 as the *"noiseless distribution"*, meant to represent a distribution from a hypothetical *noiseless crystal* that produces the same DC-current of the actual multi-Gaussian distribution, but produces no noise since its variance is *zero*.

The total variance, which is related to the current noise power produced in the slice, is written as:

$$\sigma_T^2 = \sum_k [A_k(\sigma_k^2 + d_k^2)] \tag{2.22}$$

where σ_T^2 is the total variance of the multi Gaussian profile, σ_k^2 is the variance of the Gaussian profile k, and d_k^2 is the square of its distance to the global mean. A_k is a normalization factor dependent on the number of carriers under each Gaussian profile, k. Therefore, the process of fitting the Gaussian profiles produces directly the variance values and distances necessary for the total variance.

The last steps to the noise power calculation use this total variance, total current and total number of carrier in the slab, in accordance with equation 2.20, to calculate $\overline{i^2}$, and then $\overline{v^2}$ for the slab. The conductance of the slab is proportional to the number of carriers and this information is used to convert a current noise source ($\overline{i^2}$) into a voltage noise source ($\overline{v^2}$).

2.4.2 Discretization of the BTE

In order to use the BTE in simulations, the following procedure was introduced for discretization:

The transport channel is first sliced, and the BTE is solved starting from a known end of the device. In the case of FETs, the velocity distribution of carriers at the

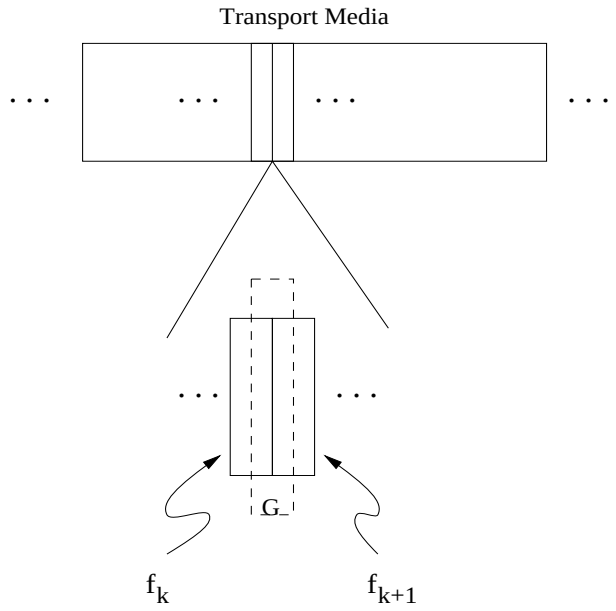

Figure 2.14: Boltzmann Transport Equation is discretized taking into account the slicing of the transport media. Two consecutive slices have their velocity distribution functions referred to as $f_k(r, v, t)$ and $f_{k+1}(r, v, t)$, and the volume described by G is used as reference for continuity when calculating carriers entering and leaving this volume, according to their velocity and position.

source end is assumed to be Gaussian and from that, the velocity distribution for the next slice in the channel is calculated.

The process of slicing the transport media, and finding the velocity distribution function for the carriers from left to right is illustrated in figure 2.14 and shows how the BTE is discretized.

Accordingly, the Residue function, $R(r, p)$, introduced in the previous section becomes in the discretized version:

$$\begin{aligned} R_{k+1}(v) &= v(N_{k+1}f_{k+1}(v) - N_k f_k(v))/dL \\ &+ N_{k+1}\frac{qE}{m_o^*}\frac{f_{k+1}(v+dv) - f_{k+1}(v)}{dv} \\ &- \sum_{v'}[N_{k+1}(1 - f_{k+1}(v))f_{k+1}(v')S(v', v)] \\ &+ \sum_{v'}[N_{k+1}f_{k+1}(v)(1 - f_{k+1}(v'))S(v, v')] \end{aligned} \tag{2.23}$$

where steady state was assumed (there is thus no dependence on time). The Residue function, $R(r, p)$, the distribution function, $f(r, p)$, and the Scattering Rate Matrix, $S(p, p')$ were written in their discrete form as functions of *velocity* and not of *momentum*. N_{k+1} and N_k are the number of carriers inside slice k and slice $k+1$, respectively.

2.4.3 Extension Towards Sub-micron Dimensions

This procedure of converting the current noise into a voltage noise in each slice and subsequent adding the voltage noise sources assuming no correlation can be carried well into the sub-micron dimension regime provided some cautions are taken.

As is clear from the fundamental ergodic equation for noise below:

$$\frac{\overline{i^2}}{I^2} = \frac{\sigma_f^2}{N\mu_f^2} \tag{2.24}$$

there is no information on the spectrum of the noise, only its total power. As stressed in recovering the expression for thermal noise in a previous section, additional information on the spectrum of the noise must be provided for meaningful noise calculation. In the earlier section, this additional information was provided by assuming a *flat noise spectrum* up to a frequency, $f = 1/ < \tau_m >$, i.e. up to a frequency of the inverse of the *mean time* between collisions and *zero* beyond this frequency..

The mean time between collisions for silicon or GaAs at room temperature establishes these frequencies around 100GHz,[50] with some variations depending on doping levels. Since for most applications of silicon FETs, the frequencies of interest are above the region dominated by 1/f noise and below 10GHz, the noise spectrum

can be considered flat for all practical purposes.

Once it is established that the noise spectrum in the frequency range of interest is *flat*, hypothetically making the mean time between collisions go to *zero* only change the roll-off frequency, which was already around 100GHz to $+\infty$, but does not change what is happening at those frequencies which were already in the *flat* region of the spectrum.

Therefore, from the perspective of the frequencies which were already in the flat region, having a meantime between collisions going *"artificially"* to zero makes no essential change, but makes a strong practical effect in device simulation. That's because a completely flat spectrum allows us to slice the FET channel as thin as we want, and allow us to continue assuming no correlation between voltage noise source from neighbor slices.

Assuming complete flat spectrum for the noise, or equivalently, assuming the mean time between collisions is *zero* is only a computational artifact, which allows for the analysis of deep sub-micron FETs, and the simulation results will always match the actual noise measurements provided they refer only to frequencies where the *real world* noise spectrum is actually flat.

Besides the flat spectrum requirement, the noise analysis of deep sub-micron FET will benefit from this computational artifact while the dominant noise source is the device channel. When the channel becomes so small that the dominant noise contributors is the interface between *source* and channel, the noise will start to change its behavior from thermal-like to shot-noise-like behavior, and there will be no need any more to consider channel slicing.

2.4.4 Extension Toward The Mesoscopic Regime

The first aspect to be emphasized is that a formalism for transport has to include both *dynamics* and *dissipative functions*. The dynamics in the BTE are represented by the *reversible* left member of the equation, which includes terms related to drift and diffusion of carriers. The *dissipative functions* are represented by the right member of the equation where the *Scattering Rate Matrix*, $S(p, p')$, describe the interaction of

carriers and the phonon bath (reservoir) inside the crystal.

In this work, mesoscopic transport is not treated, however the required adjustments in the *Ergodic Method* can be pointed out for the appropriate treatment for this kind of transport.

First, mesoscopic transport is achieved when the conductor dimensions and temperature of operation are such that the *wave nature* of the carriers are prevalent in the transport characteristics. The carriers will propagate in *modes* and will be affected by the presence of the other carriers. The effect of the other carriers is felt through the Pauli exclusion principle or Coulombic repulsion between carriers. Other than these interactions, mesoscopic transport is almost 100% ballistic.

Therefore, the *Ergodic Method* can retain its interpretation of noise phenomena as a fluctuation in carrier velocity and is thus related to variance in carrier velocity. However, the semi-classic Boltzmann formalism must be changed to a formalism which is based upon the wave nature of the carriers.[51, 52, 53]

2.5 Noise in sub-micron CMOS

Noise for the *long channel* case in CMOS transistors was treated by Van der Ziel [60, 61]. Quite different results were reported by Abidi [62] and Jindal [63] for the *short channel* case. The discrepancy between the results for long and short channel CMOS can be readily explained from the development of the *Ergodic Method*.

For the long channel case, the formal assumes enough *gate* bias to turn on the channel. At $V_{DS} = 0V$, the current thermal noise developed in the channel appears at the drain-source terminals as:

$$\overline{i^2} = 4k_B T_L G_{do} \Delta f \tag{2.25}$$

where $\overline{i^2}$ is the current noise power and G_{do} is the conductance of the channel at V_{DS} =$0V$. k_B, T_L and Δf are the Boltzmann constant, the lattice temperature and the bandwidth of the noise measurement, respectively.

Assuming an NMOS transistor, when V_{DS} is increased, the channel conductance

diminishes. Therefore, the current noise developed by the channel at $V_{DS} \neq 0V$ should be lower than that described by equation 2.25 above. Not surprisingly, a correcting factor, γ, should be introduced if the conductance G_{do} is to be kept in the noise equation. The adjusted equation is:

$$\overline{i^2} = 4k_B T_L G_{do} \Delta f \gamma \tag{2.26}$$

As expected, the correction factor γ was calculated and found to be 2/3, at the onset of saturation. It is also noticed to be constant to some extent for higher values of V_{DS}.

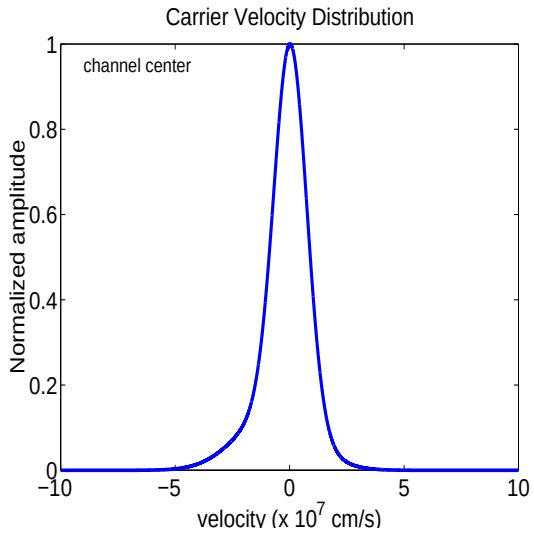

Figure 2.15: Typical distribution of electron velocities in the center of the channel for a long channel MOSFET ($L = 5\mu m$)

.

It should be pointed out, however, that $\gamma = 2/3$ was calculated above with the assumption that the carriers are always in equilibrium with the lattice temperature. It was assumed that *no carrier heating* was taking place inside the channel. This is a reasonable assumption for long channel CMOS, but for short channel CMOS, carriers *do* get substantially *hotter* than the lattice.

Carrier heating in short channel CMOS devices comes from the presence of a high lateral electric field which accelerates the carriers to velocities not readily thermalized

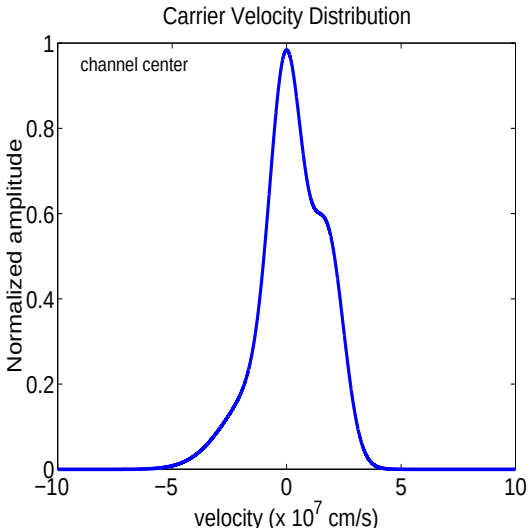

Figure 2.16: Typical distribution of electron velocities in the center of the channel for a short channel MOSFET ($L = 0.5\mu m$)

.

by the phonon bath inside the channel. Therefore, γ factors well in excess of 5 have been reported in the literature for the sub-micron CMOS transistors.[60] - [65].

Typical results for long channel and short channel carrier velocity distribution are shown in figures 2.15 and 2.16. This provides deep insight into the noise phenomena in CMOS transistors. Notice that the short channel device is poised (according to the Ergodic Method) to develop far more noise than the long channel device because the distribution of carrier velocity has more *variance* in the short channel device than in the long channel device. This result clearly suggests that only making smaller devices may not be the best path for low noise operation. Shorter channel devices may have some carriers experiencing *ballistic* transport, which means transport without scattering. However, having only some carriers going ballistically along the channel while having others traveling sluggishly behind greatly increases the variance in carrier velocity distribution and thus increases the current noise power. The key for low noise operation is then to make the carriers move along the channel with the narrowest possible velocity distribution.

2.6 The Ergodic Method Predictions

Up to this point, the *Ergodic Method* was used to give deeper insight into noise phenomena, including the *excess noise* in deep sub-micron CMOS transistors. This constitutes the descriptive characteristics of the method and is its basis for CAD tools capable of predicting noise performance of semiconductor-devices at *design time*.

Another asset of the *Ergodic Method*, which is valuable for device designers is its guidance to the design of new low noise devices. The objective in stressing this asset is to make clear that the *Ergodic Method* provides a vantage point to the appreciation of noise phenomena in semiconductor-devices from which new and non-obvious low noise device designs can be devised. The next chapter is focused on experimentally demonstrating the predictions of noise performance that will be drawn in this section. The experimental demonstration of the *Ergodic Method's* is also used to prove the method's correctness and adequacy of its *a priori* assumptions.

First, the fundamental equation of the *Ergodic Method* and the Boltzmann Transport Equation are recast below:

Fundamental Equation:

$$\frac{\overline{i^2}}{I^2} = \frac{\sigma_f^2}{N\mu_f^2} \tag{2.27}$$

BTE:

$$\frac{\partial f}{\partial t} + \boldsymbol{v}.\boldsymbol{\nabla_r} f + \frac{q}{m^*}\boldsymbol{E}.\boldsymbol{\nabla_p} f = \left.\frac{\partial f}{\partial t}\right|_{coll} \tag{2.28}$$

$$\left.\frac{\partial f}{\partial t}\right|_{coll} = \sum_{p'} f(p')[1-f(p)]S(p',p) - \sum_{p'} f(p)[1-f(p')]S(p,p') \tag{2.29}$$

Previously established guidelines for the design of low noise devices suggest the device should be designed so that (1) carrier transport is away from interfaces, i.e.

buried channel devices and (2) carriers should travel in a high-concentration pack. Both guidelines are inspired by the right-hand side of the BTE above. In (1), the objective is to diminish the scattering rates ($S(p,p')$) by providing transport away from interfaces where there is a much higher concentration of crystal defects than in the bulk. In (2), the objective is to take advantage of the fact that if a quantum state is already occupied by a carrier, another carrier can't scatter there ($(1-f(p))$). Both elements (1) and (2) are remarkably accomplished in the design of HEMTs (High Electric Mobility Transistors).

Because of the *Ergodic Method*, it is clear that the left-hand side of the BTE can also be explored for the design of low noise devices.

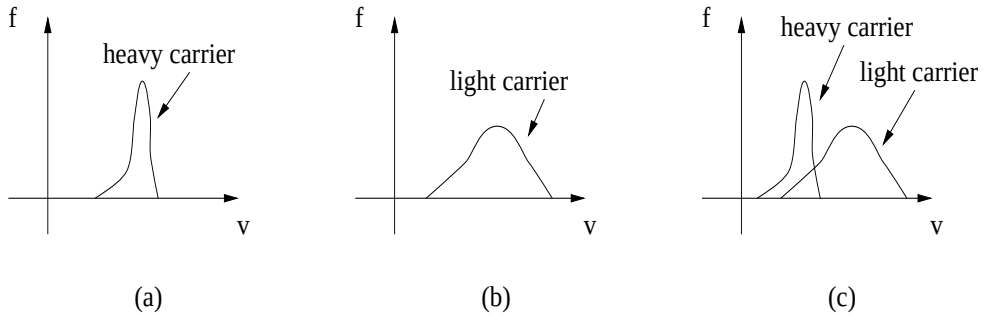

Figure 2.17: Carriers velocity distribution under high electric field transport. In (a) only heavy-effective-mass carriers take part in the transport, in (b) only light-effective-mass carrier take part in the transport and in (c) light and heavy effective mass carriers take part in the transport.

Note that the left-hand side of the BTE works as a driving force to the right-hand side. If it were not for the left-hand side, the carrier velocity distribution would always look Maxwellian as in thermal equilibrium. The effect of the electric field makes the carrier velocity distribution non-Maxwellian and in the case of short channel devices, *highly* non-Maxwellian.

Note that the effect of the electric field on the velocity distribution is modulated by the *effective mass* of the carriers, as can be seen in the BTE above. Thus, a *heavier carrier* should produce a device less prone to develop *excess noise*. This is a surprising result. Unfortunately, the drawback to using a heavier carrier is its lower mobility and resultant lower frequency device. In figure 2.17(a) and (b) the pictures convey the fact that heavier carriers will have less average velocity and narrower distribution

of velocity than light carriers.

Consistent with the above, if the semiconductor material has the carriers traveling at two possible effective masses, as depicted in figure 2.17(c), the BTE should be solved twice, one time for each carrier. The total variance in carrier velocity in high fields will be much broader than if only one type of carrier were allowed to participate in the transport. n-type silicon, for instance, conducting electric current in the $< 100 >$ direction, will have some carriers transported with the *transverse effective mass* and others transported with *longitudinal effective mass*. An appropriate mechanical stress in Si can change the energy band diagram and force the carriers to travel under practically only one effective mass.[66] - [80] This leads to narrower velocity distribution for the carriers, hence lower noise device operation. Note however that this one-type-carrier transport will hold into the high field regime till extreme high electric fields end up heating carriers so excessively that they might access other valleys (i.e. other transport modes), making the carrier transport return to a multi effective mass mode..

If the carrier type of the low noise device has already been fixed, a final path for low noise design is still suggested in the left-hand side of the BTE. The *carrier concentration* can be varied along the transport channel so that the diffusion term on the left-hand side of BTE counter-balances the effect of the electric field which drives the distribution function to be non-Maxwellian.[81]

These three approaches for the design of low noise operation predicted by the *Ergodic Method* will be experimentally demonstrated using silicon FETs in the next chapter. The effect of heavier carrier on noise in high electric fields is experimentally investigated with comparison of *state-of-the-art sub-micron NMOS and PMOS transistors*. The effect of only one versus more than one effective mass carrier will be investigated using mechanical stress on *Si-on-SiGe FETs*. The effect of varying carrier concentration along the transport channel is investigated with *laterally graded channel MOSFETs*.

Chapter 3

Experimental Developments

The Ergodic Method and the Boltzmann Transport Equation produced three approaches for controlling and minimizing the generation of *excess noise* in high electric field transport. The first approach is to use heavier effective-mass carriers in the transport; the second approach is to force the carriers to transport with only one effective-mass and the third approach is to tailor the doping profile to minimize the *variance* of the carrier velocity distribution.

These approaches for controlling *excess noise* generation are demonstrated in this chapter through experiments aimed at testing each of the three approaches for controlling *excess noise* in high field transport. 1) The effect of different effective masses in noise performance is checked via comparison between state-of-the-art sub-micron NMOS and PMOS transistors; 2) the effect of more than one effective-mass carrier taking part in the transport is checked with Si-on-SiGe FETs with different degrees of tensile stress; and 3) the effect of lateral graded doping profiles in noise performance is checked via comparison of uniformly doped and graded doped n-type MOSFETs.

In this chapter the *Excess Noise Factor*, γ_f, is introduced to allow comparison of excess noise generation among different materials and technologies.

3.1 Experimental Setup

A schematic diagram of the experimental setup is shown in figure 3.1. The output of the spectrum analyzer gives the current noise power per hertz.

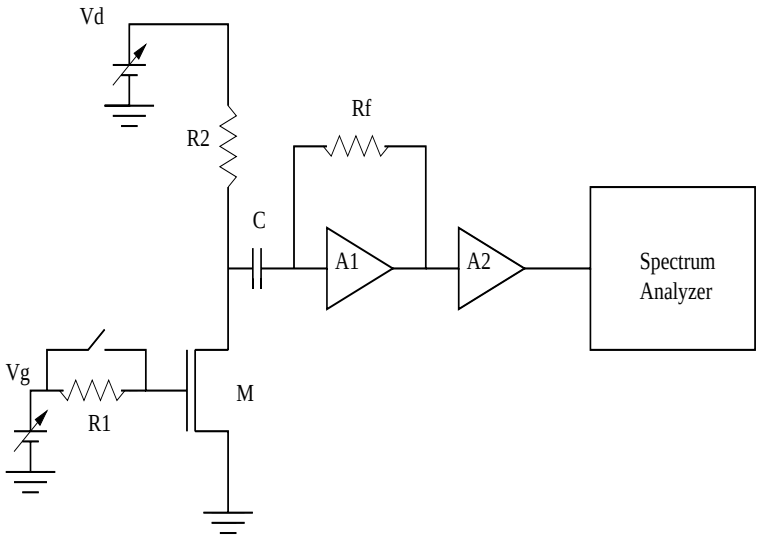

Figure 3.1: Schematic of the system designed for measurement of drain current noise power of FETs.

The drain current noise power of the FET is amplified by the low noise transimpedance amplifier and pos-amplifier to levels above the noise floor of the spectrum analyzer, which in our system was at -89dBm/Hz.

The amplifier, A_1, is a low noise transimpedance amplifier. Its *input referred current noise* and *input referred voltage noise* sources need to be low enough to make sure the contribution of current noise power coming from the *drain* of the transistor under test is the dominant noise component at the input of this stage. The *input referred current noise* of this amplifier, A_1, is an issue for measuring the noise power of a low drain current noise power transistor; and the *input referred voltage noise* of A_1 is an issue for low output impedance transistors (normally occurring for large transistor biased in the linear region).

In our setup, A_1 has an *input referred noise current* of $2.5pA/\sqrt{Hz}$ and *input referred voltage noise* of $1.25nV/\sqrt{Hz}$. Resistor R_2 is chosen to be high enough to produce much less current noise than the targeted transistors to be tested. R_f was

chosen high enough to provide enough gain to insure that the noise contribution of the following stages did not dominate the noise power level measured. Resistor R_1 was used to check the contribution of the gate resistance noise of the transistor under test. The dominant noise source from the transistor was always the thermal noise generated in the transistor channel.

Because of the high gain involved and the assumption that *thermal* or *excess noise* have a flat spectrum over frequencies up to 100GHz, the experiments only tried to stay away from the 1/f noise corner. Therefore, measurements were made and reported in this work for f=800kHz. This frequency proved to be far enough away from the 1/f noise corner and within the capabilities of the amplifiers used.

The setup was then tested with *known values of resistors* in place of the transistor in figure 3.1. Carbon film resistor were used. This was then a **low electric field** current noise measurement. The experimental results and linear curve fitting are reproduced in figure 3.2. Note that the current noise power of these resistors follows the value predicted from the Thermal Noise expression $< i^2 > /\Delta f = GK_BT_L$. The experimental setup is capable of correctly measuring current noise contributions from a resistor with resistance values between 10Ω and $6k\Omega$.

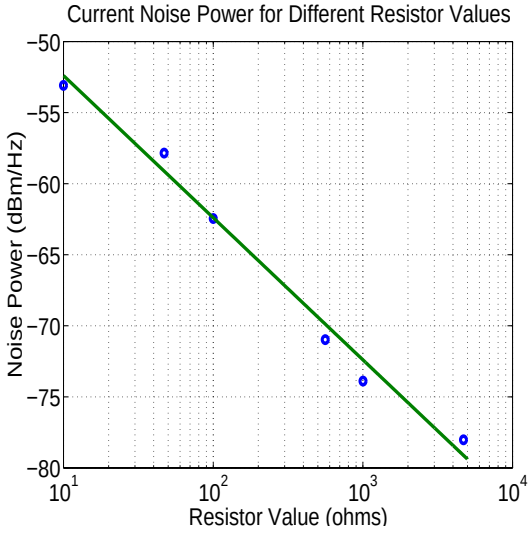

Figure 3.2: Experimental results for *low electric field* current noise power of resistors of known resistance values.

Inserting *transistors* into the measurement setup, quite different results are observed for **high electric field** current noise measurements. In figure 3.3, the experimental results for a NMOS transistor having $W = 10\mu m$ and $L = 0.48\mu m$ are depicted. Notice that for every value of V_{GS}, the sweep from V_{DS} =0V to $V_{DS} = 3V$ gives noise power that departs from the straight line in the plot. The straight line in the plot indicates the level of thermal noise associated with the *equivalent linear resistor* in thermal equilibrium which would have a resistance value equal to the DC-resistance, V_{DS}/I_{DS}. This straight line is instrumental for current noise power comparison among different technologies. The separation between this straight line and the actual noise measurements for the transistor will be defined as the *excess noise factor*, γ_f, and it will be recorded for every bias point[1].

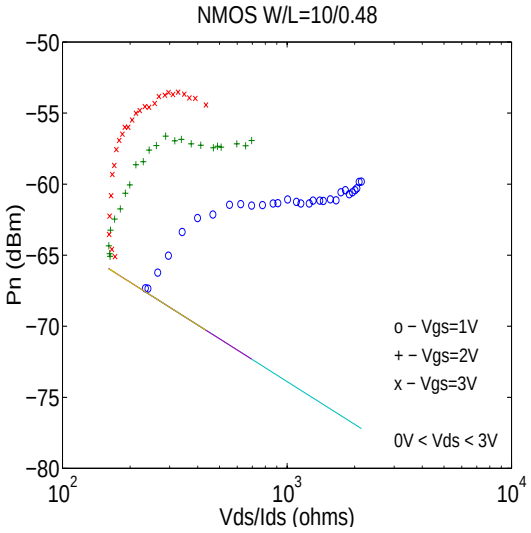

Figure 3.3: Experimental results for *high electric field* current noise power of an NMOS transistor with dimensions $W = 10\mu m$ and $L_{eff} = 0.48\mu m$. At the highest values of V_{DS}, the transistor develops much more noise than would be generated by a linear resistor in thermal equilibrium with resistance value equal to V_{DS}/I_{DS}. These linear resistors would be developing noise levels which would follow the straight line indicated in the plot.

The interpretation of the graph shown in figure 3.3 will be clear after the introduction of the excess noise factor, γ_f in the next section; but for now it is important

[1]This definition of γ_f is different from the definition of γ by other authors, and they differ numerically by one order of magnitude normally.[82, 83]

to note that the noise power measured and shown in figure 3.3 is **not** coming from the gate resistance amplified by the g_m of the devices used. This is proven by post processing of the experimental data, calculation of g_m from the I-V curves of every device and calculation of the noise power at the output port from the product of gate resistance thermal noise multiplied by g_m. In the transistor depicted in figure 3.3, the gate resistance noise power referred to the output provided a noise power level below -70dBm/Hz. This result is also supported if we focus our attention on the flat parts of the three curves. Neither the gate resistance nor the value of g_m changes significantly at those values of bias to provide such a difference in measured noise power. Therefore, the noise power measured and depicted in figure 3.3 comes entirely from the channel of the device.

The *excess noise* behavior of figure 3.3 is not particular to the transistor used for these measurements. This is rather typical behavior of transistors operating under *high fields*, meaning horizontal electric fields in excess of $10^5 V/cm$.

Before moving on to the experimental results of the transistors, several concepts and a figure of merit for comparison among different FET technologies are introduced in the next section. Upon the knowledge basis developed in this section, the presentation and interpretation of the experimental results follow naturally.

3.1.1 Excess Noise and Excess Noise Factor, γ_f

As already discussed, the thermal noise expression, $< i^2 > /\Delta f = 4GK_BT_L$, is only valid in thermal equilibrium, or when **no** DC current is flowing. [84, 85] If a DC current flows, additional noise develops and this additional noise is called *excess noise*.

Because of the small dimensions and high electric fields involved in the drain current noise measurements, MOSFETs used in this work exhibit typical curves of current noise power described in figure 3.4(a) and (b). Figure 3.4(c) depicts the case for a long channel MOSFET. In this figure, the drain current noise power is plotted as function of *DC resistance* V_{DS}/I_{DS}. The decision why to use *DC resistance* instead of *AC resistance* will be discussed later in this section, but for now it suffices stating

the *qualitative* appearance of the graphs plotted in figure 3.4 would be the same for either the DC or the AC resistance used for the x-axis.

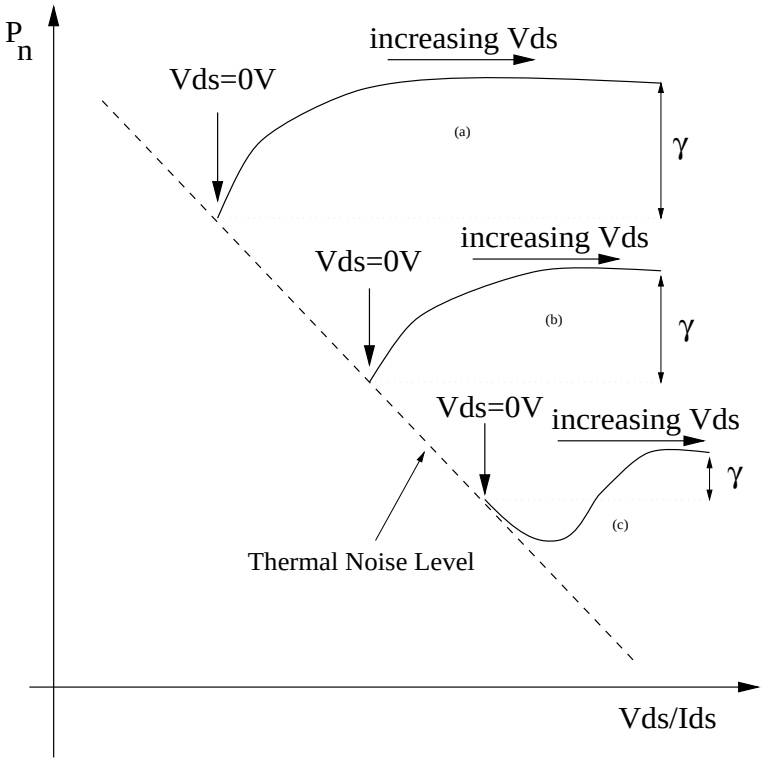

Figure 3.4: Typical curves for drain current noise power as a function of V_{DS}/I_{DS}. The dashed line indicates the *thermal noise level* if the noise were developed by an equivalent linear resistor with resistance equal to V_{DS}/I_{DS}. The plots (a) and (b) show results for short-channel MOSFETs, which produce thermal noise only at V_{DS} =0V and develop increasing *excess noise* above V_{DS} =0V. (c) shows the case for a long channel MOSFET, which also develops increasing *excess noise* with V_{DS}, but stays below the noise developed under V_{DS} =0V for a significant range of V_{DS} values. The gamma factor, γ, introduced by other authors, is defined as the difference in noise power from the measured noise to the noise developed at V_{DS} =0V.

It's important to note in figure 3.4 that the dashed line indicates the current noise power that would be measured if an *equivalent linear resistor* of resistance value V_{DS}/I_{DS} were placed in the measurement setup instead of the actual transistor.

The typical curves for *short channel MOSFETs* in figure 3.4(a) and (b) show transistors which develop increasing *excess noise* with V_{DS} and always develop more noise than at V_{DS} =0V. Figure 3.4(c) shows a typical case for *long channel MOSFETs*, which also develop increasing *excess noise* with V_{DS}. However, they develop less noise

than when at V_{DS} =0V for a substantial range of V_{DS} values. Eventually, the long channel MOSFETs develop more noise than when at V_{DS} =0V.

As mentioned in the previous chapter, a gamma factor, γ, is already used in the technical literature to cope with the noise of MOSFETs at biases of V_{DS} $\neq$0V. This is defined by the separation between the transistor noise power at a certain bias and the noise power developed by the same transistor at V_{DS} =0V. Therefore it's common in the literature to write the drain current noise power as:

$$< i^2 > /\Delta f = 4G_{do}K_BT_L\gamma \tag{3.1}$$

$< i^2 > /\Delta f$ is the current noise power per hertz. K_BT_L is the product of the Boltzmann constant and the lattice temperature. Note that the conductance value used is G_{do}, which is the conductance of the channel at V_{DS} =0V. γ is the multiplicative correction to make the expression produce numerical values equal to the experimental values. Sometimes, G_{do}, the DC conductance of the channel is substituted by g_{do}, the AC conductance of the channel at V_{DS} =0V, as they are numerically equal, and in some cases, even g_m, the transistor transimpedance is used when its value is numerically equal to g_{do}. As can be seen from A. Van der Ziel [82], the gamma value for a long channel MOSFET in saturation is 2/3. This value is calculated considering only the value of the channel conductivity in saturation and comparing it to the channel conductivity at V_{DS} =0V. This does not consider any carrier heating by the electric field, as can be seen from the arrows indicating the definition of γ in figure 3.4.

Since the drain current noise power seems to *"saturate"* at some value at high values of V_{DS}, this equation is very appealing and convenient. Unfortunately this same equation obscures the physical action that takes place at high values of V_{DS}, in particular, as it is seen from figure 3.4, the transistor continues to develop increasing values of *excess noise*, diverging even more from the thermal noise level (dashed line) and shows no sign of *"saturation"*.

In figure 3.5, a **new gamma factor** called the **excess noise factor** (note the subscript *"f"*), γ_f, is introduced. The drain current noise level developed by the transistor is compared to the noise developed by a linear resistor of resistance value

equal to V_{DS}/I_{DS}. The resistor is assumed to develop thermal noise and its current noise level is marked by the dashed line in figure 3.5.

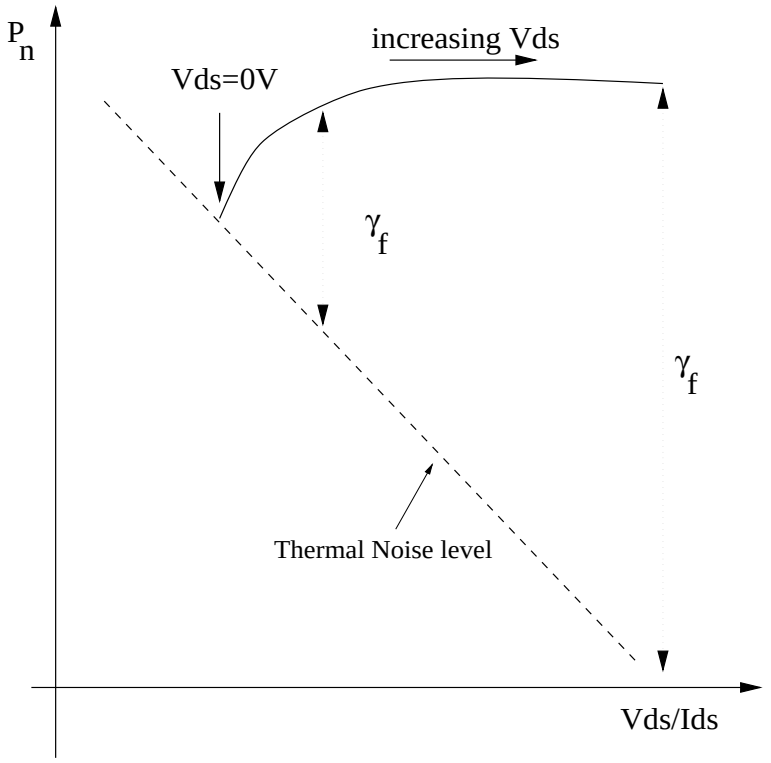

Figure 3.5: Typical curves for drain current noise power as a function of V_{DS}/I_{DS}. The dashed line indicates the *thermal noise level* if the noise were developed by an equivalent linear resistor with resistance equal to V_{DS}/I_{DS}. The *excess noise factor*, γ_f, is defined as the difference in noise power from the measured noise to the noise developed by the equivalent linear resistor of resistance value equal to V_{DS}/I_{DS}.

Therefore, the drain current noise is now expressed as:

$$< i^2 > /\Delta f = 4GK_BT_L\gamma_f \tag{3.2}$$

$< i^2 > /\Delta f$ is the current noise power per hertz. K_BT_L is the product of the Boltzmann constant and the lattice temperature. The new aspect of this expression is the introduction of the conductance G, which tracks the bias as $G = I_{DS}/V_{DS}$. The excess noise factor, γ_f, is now clearly a figure of merit indicating how different materials develop excess noise. This factor will be different for different materials

and carriers involved in the transport. Furthermore, since G absorbs the difference in biases and sizes of the transistor, γ_f is only dependent on electric field and carrier profile. This will be clear from the experiments.

It is important to pause here for some words on the choice of a DC resistance (large signal resistance) versus AC resistance (small signal resistance) for computing γ_f.

3.1.2 DC Resistance versus AC Resistance

In figure 3.6, linear resistors develop thermal current noise. Resistors depicted in figure 3.6 (a),(b),(c) and (d) develop thermal current noise powers expressed by the equations below in the same order:

$$\overline{i^2}/\Delta f = 4(1/R_1)K_B T_L; \text{ for resistance } R_1 \quad (3.3)$$

$$\overline{i^2}/\Delta f = 4(1/(R_1+R_2))K_B T_L; \text{ for resistance } R_1+R_2 \quad (3.4)$$

$$\overline{i^2}/\Delta f = 4(1/(R_1+R_2+R_3))K_B T_L; \text{ for resistance } R_1+R_2+R_3 \quad (3.5)$$

$$\overline{i^2}/\Delta f = 4(1/(R_1+R_2+R_3+R_4))K_B T_L; \text{ for resistance } R_1+R_2+R_3+R_4 \quad (3.6)$$

Consider now figure 3.7. In figure 3.7(a) the transistor is "on" and in the *linear region*. In figure 3.7(b), the transistor is also "on", but in the *saturation region*. As depicted in this figure, the channel in the linear region can be seen as equivalent to a resistor of value R. Any uniform slicing of the channel would produce slices of equal resistance. In the saturation region (b), however, the channel is disturbed by the non-zero voltage applied at the drain and a uniform slicing of the channel will **not** produce slices of equal resistance. As depicted in figure 3.7(b), the channel (DC) resistance can be approximated by a stack of resistances of different values for equally spaced slices of the channel.

Therefore, as V_{DS} changes, the channel resistance changes and describes either one of the typical curves shown in figure 3.8. In figure 3.8(a) the I-V curve depicts a normal transistor and in figure 3.8(b), the I-V curve depicts a transistor suffering

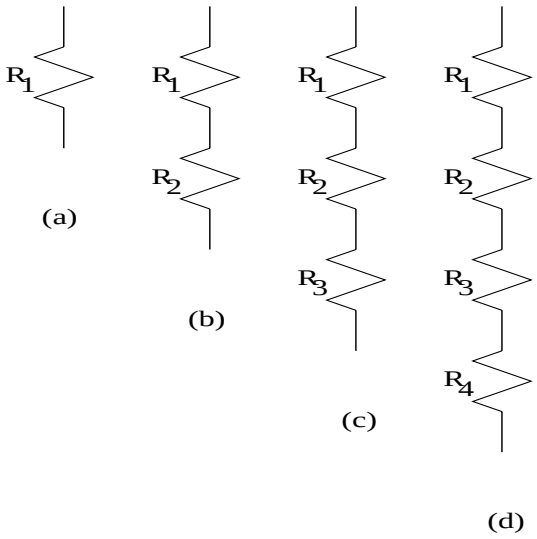

Figure 3.6: Linear resistors developing thermal noise.

self-heating effects (explained later in section 3.3, "Modes of Transport and Noise"). At bias points shown in figure 3.8(a) and (b), the DC conductance I_{DS}/V_{DS} is a positive value. However, the AC conductance $g_o = \partial I_{DS}/\partial V_{DS}$ is a *positive* value in figure 3.8(a), but a *negative* value in figure 3.8(b).

As seen in the experimental data, either transistor depicted by the I-V curves in figure 3.8(a) and (b) develop *excess noise* in the *very same way*. There is therefore *no* correlation between noise behavior and the change in signal of the AC conductance g_o. Note also that the DC conductance, $G = I_{DS}/V_{DS}$, does not change signal from one transistor to the other.

At this point, it is clear that one should use the DC conductance instead of the AC conductance in the drain current noise of equation 3.2. However, more fundamental insight is gained if it is recalled that the Ergodic Method fundamentally calculates current noise power by probing the transport media by following a single ordinary carrier and tracking the statistics of its velocity in time.

Consider a transistor biased in saturation. If a small variation is applied to the value of V_{DS}, a small variation in I_{DS} will appear, such that $\partial I_{DS}/\partial V_{DS} = g_o$. This is however a **collective** response of the channel, no matter how small the driving variation in V_{DS}.

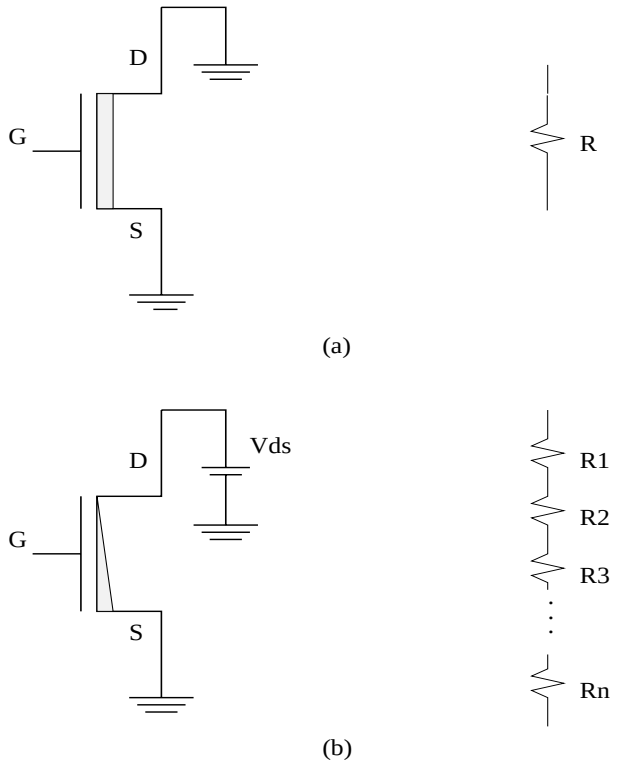

Figure 3.7: FETs' channel equivalent DC resistance. (a) In linear region the channel is an uniform resistor. (b) In the saturation region the channel varies its resistance from source to drain.

However, the resistance experienced by *one ordinary carrier* while traveling through the same channel is very different. If one ordinary carrier is tracked along the channel of this same transistor, it will experience the DC resistance established by the bias along the channel. It is clear this single carrier will **not** be able to change the channel resistance in any collective fashion.

Therefore, it is the DC conductance that matters for the analysis of thermal and excess noise in the channel of FETs. As a matter of fact, it should come as no surprise that the AC conductance is only a linear representation of the *collective* response of a device or network for small signals, but bears little fundamental physics. What an ordinary carrier sees while traveling through the channel is the DC resistance of the channel.

This brings a strong similarity in the structure for the noise equations for *shot noise*

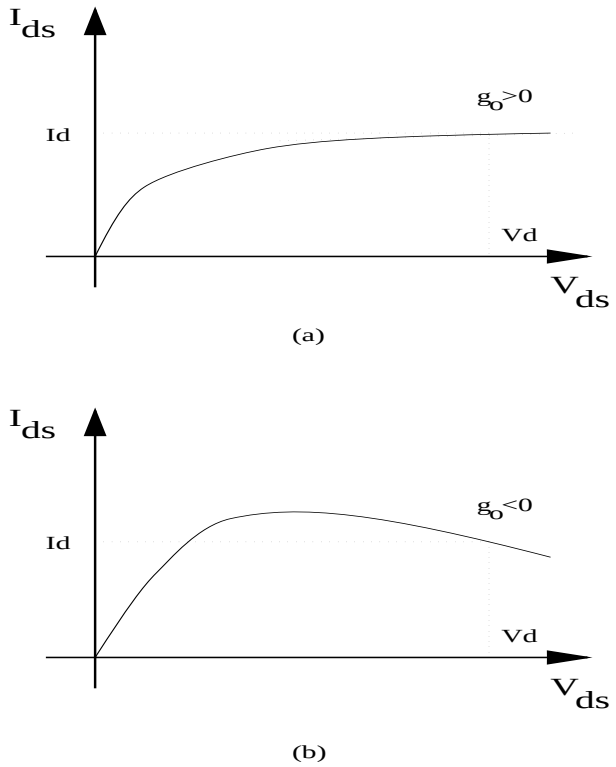

Figure 3.8: I-V curves of FETs. In (a), the transistor has an ordinary curve, with $g_o > 0$. In (b), probably due to *self heating* effects, the transistor develops a curve with $g_o < 0$. Both transistors develop excess noise in similar ways and this fact provides evidence that g_o bears no fundamental connection to thermal noise generation. Only the DC conductance, $G = I_{DS}/V_{DS}$, relates to noise.

and *thermal noise*. The shot noise expression $< i^2 > /\Delta f = 2qI$, connects a small signal $< i^2 >$ to a DC value I. In our excess noise expression, $< i^2 > /\Delta f = 4GK_BT_L\gamma_f$, accordingly, a small signal $< i^2 >$ is connected to a DC value of conductance, G.

3.1.3 The Excess Noise Factor as a function of Electric Field

Once the previous section established the appropriate use of the DC channel Resistance for the excess noise expression, it is of interest to plot the dependence of the excess noise factor, γ_f, with the applied lateral electric field in the channel.

Sticking with the definition of excess noise associated with the DC conductance of the channel, $< i^2 > /\Delta f = 4GK_BT_L\gamma_f$, the plot of the excess noise factor, γ_f has the general behavior depicted in figure 3.9.

Actually, the x-axis is the *average electric field* rather than a single value of electric field. That's because the voltage $V_{DS} \neq 0V$ applied from drain to source of the FET disturbs the the carrier concentration along the channel such that normally there is *no* constant electric field throughout the channel. In practical plots, the x-axis will mark values retrieved from the division V_{DS}/L_{eff}, where L_{eff} is the effective length of the channel.

Note that from the development above, the plot of γ_f as function of electric field should always start at $\gamma_f = 1$ for the case of low electric field, i.e. $V_{DS} = 0V$. This is consistent with the fact that at $V_{DS} = 0V$, there is no heating effect from the field in the channel and the channel should be developing only thermal noise. As the field increases (by augmenting V_{DS}), excess noise is developed and γ_f becomes much bigger than 1, as can be seen in figure 3.9.

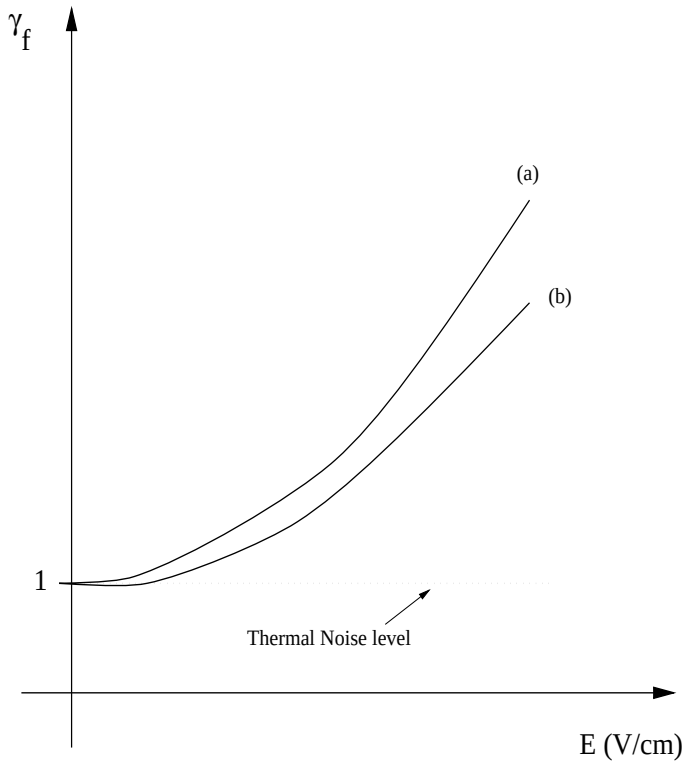

Figure 3.9: Excess noise factor, γ_f, as a function of electric field. $\gamma_f = 1$ when no electric field is applied and increases as the applied electric field increases. γ_f is used to compare the noise of different technologies for field effect transistors (FETs). (a) describes a FET noisier than the FET described by (b).

The graph of γ_f, as shown in figure 3.9, illustrates how the excess noise factor is used as a figure of merit to compare different technologies of field effect transistors. As can be seen in 3.9, curve (a) describes an hypothetic FET that develops excess noise much faster than a FET described by (b). Therefore, the device described by (b) has better noise performance than device (a). In fact, if sized accordingly, device (b) can provide the same DC Resistance as device (a), but still develop less current noise under high fields.

In the next subsections, it experimental results will demonstrate the predictions of the Ergodic Method developed in previous chapters via comparison of the excess noise factor, γ_f.

It should be noted that the actual values of excess noise factor, γ_f, for deep submicron transistors produce numerical values one or two orders of magnitude larger than those values of *"ordinary"* γ reported in the literature. This is first because they are based upon different definitions and calculated differently and second, because γ_f is reported in this work for channel lengths much shorter than those used in the literature. In any case, the numerical values *per se* are of secondary importance. Of primary importance is the relative values of the excess noise factor γ_f between different technologies.

3.2 Different Effective Masses and Noise

In order to test the Ergodic Method prediction that heavier mass carriers will develop less excess noise than lighter ones, drain current measurements were done on NMOS and PMOS devices.

Identical silicon NMOS and PMOS transistors were used in this experiment with the following dimensions in *microns*:

W/L:
40/0.4, 10/1.38, 10/0.78, 10/0.48, 10/0.23, 10/0.18

where W is the lithographically defined *width* and L is measured *effective length* of the channel.

There are thus transistors with same defined width $10\mu m$, but *effective* lengths varying by one order of magnitude, from $1.38\mu m$ to $0.18\mu m$. In addition, a transistor dimension of width $40\mu m$ and length $0.4\mu m$ was added to check for any effects of variation in width on noise performance. These dimensions were used for both NMOS and PMOS transistors.

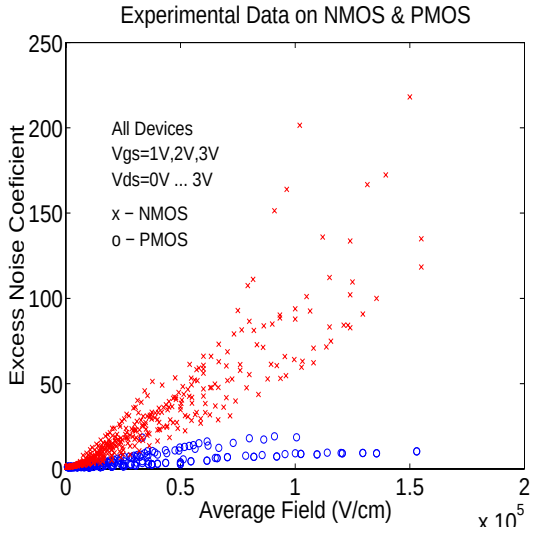

Figure 3.10: Excess noise factor, γ_f, for silicon NMOS and PMOS. All the NMOS devices ("x") develop more excess noise than PMOS devices ("o") at the same electric field.

Figure 3.10 shows the results of the excess noise factor, γ_f, for all of these devices. The measurements were performed on every transistor, using $V_{GS} = 1V$, $2V$, and $3V$, while V_{DS} was swept from $0V$ to $3V$ for NMOS transistors and $0V$ to $-3V$ for PMOS transistors.

As can be seen from figure 3.10, all NMOS transistors develop more *excess noise* than PMOS transistors. In silicon, electrons are significantly lighter than holes, thus the experimental results in figure 3.10 supports the prediction of the *Ergodic Method* that devices based on heavier carrier are less noisy than those based on lighter carrier.

3.3 Modes of Transport and Noise

The theoretical development of the Ergodic Method predicted that materials where carriers are transported by more than one *mode* should develop more excess noise than a material where carriers transported by a single *mode*.

The word *mode* is used here in the sense of having carriers traveling via a *specific effective mass*. Silicon, for instance, if n-type and supporting an electric current in the $< 100 >$ direction, has electrons traveling with both *longitudinal* and *transversal effective mass*. On the other hand, n-type GaAs, will have all its electron traveling with the same effective mass.

Effective masses are defined by the energy-band diagram of the semiconductor material. The energy-band diagram is a function of the periodicity of the atomic potentials inside the semiconductor lattice. Semiconductor materials have different energy band structures which may be quite complex, as it is depicted in figure 3.11 for the case of germanium, silicon and GaAs.

The *effective mass* of the carriers is defined by the periodicity of the atomic potentials inside the lattice. This potential periodicity and thus the effective mass of the carriers, can be varied if the lattice spacing is altered via *mechanical stress*.

In order to focus in the essential point of this section, a hypothetical semiconductor with an energy-band structure conceptually described in figure 3.12 will be used. In figure 3.12(a) the semiconductor is in equilibrium or nearly so under very low field. The electrons occupy only the central bottom of the conduction band. In 3.12(b) the application of a strong electric field makes current flow and significantly alters the carrier occupancy of the central band. Notice however that in (b) the carriers are forced to stay in this central valley and their momentum variance is constrained to the energy curves of this valley. In figure 3.12(c), the material has secondary satellite valleys, here assumed to imply different effective mass for carriers which occupy it. The transport described in (b) is called *single mode transport* and the transport described in (c) is called *two mode transport*.

In fact, different valleys may or may not have carriers with different effective

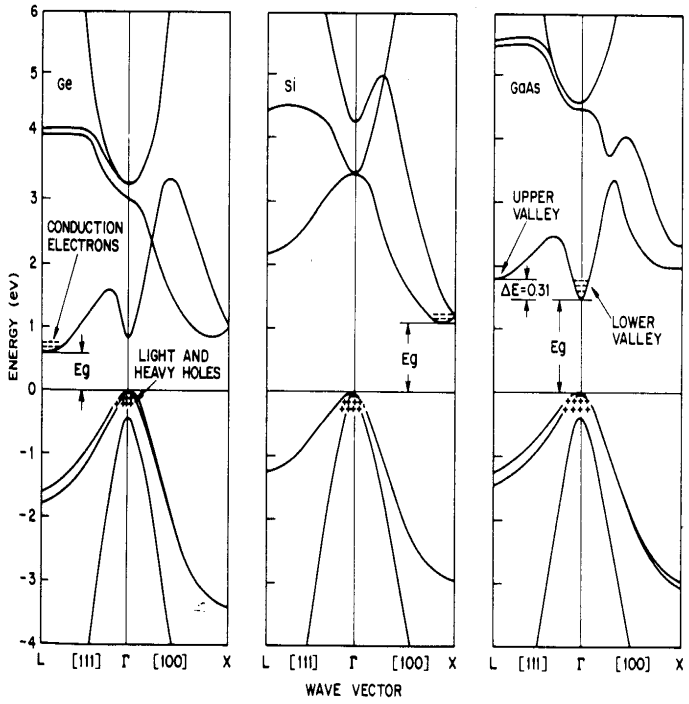

Figure 3.11: Band-energy structure for (a) germanium, (b) silicon and (c) GaAs. (From "Physics of Semiconductor Devices", S. Sze)

masses. Therefore, more accurately, the names single mode, two mode or even multi-mode transport will refer to the presence of one, two or several effective mass carriers taking place in the transport. This will become clear in the use of silicon on silicon-germanium (Si-on-SiGe).

The controlled experiments on *single mode* and *multi-mode* transport were done with Si-on-SiGe FETs. The SiGe alloy has a larger lattice constant than the Si lattice. This difference can be adjusted depending on the content of Ge in the SiGe alloy. The Si-on-SiGe FETs are built by depositing a very thin layer of Si on a thick lattice relaxed layer of SiGe. To provide some background for the rationale of using Si-on-SiGe in the experiments of this section, consider the effects of tensile and compressive stress on a thin film of semiconductor shown in figure 3.13.

Starting with a thick and lattice relaxed layer, the deposited thin layer of a second semiconductor will grow epitaxially adapting its lattice constant to the lattice

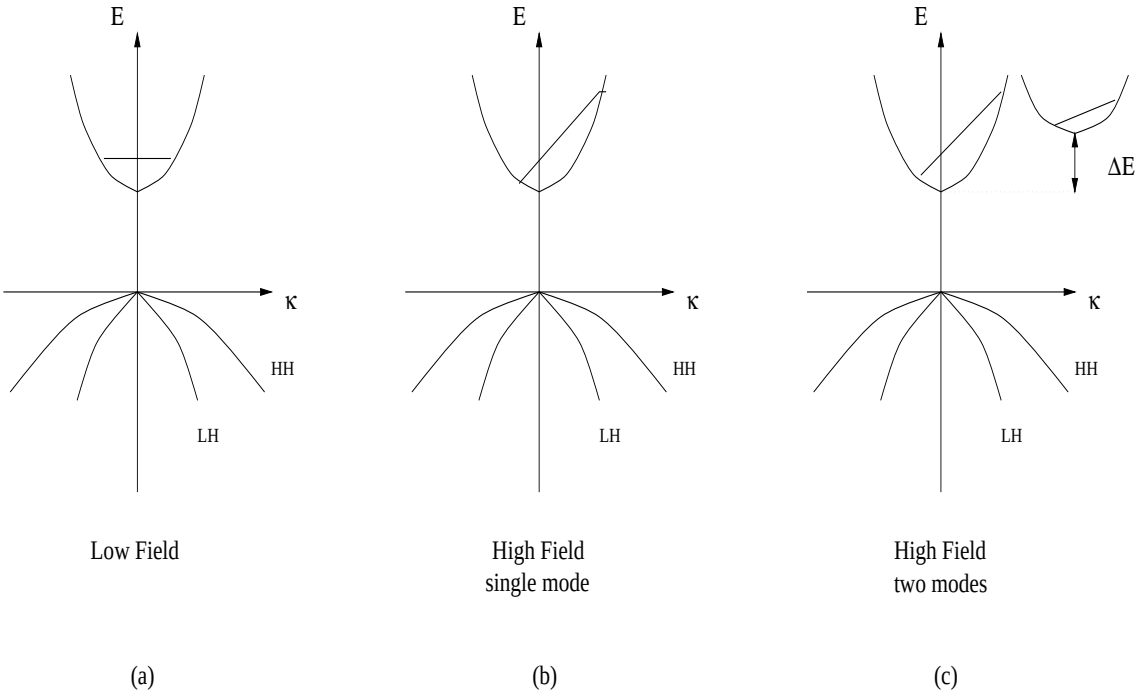

Figure 3.12: Electrons and occupancy of bands with electric field. (a) At low electric field, electrons occupy the central bottom of the conduction band. (b) High field disturbs the electron occupancy of the central band, but carriers can occupy only one minima. (c) High field and the presence of another valley allows electrons to transport with two different "modes", one through the central valley and another through the satellite valley.

constant of the first lattice relaxed layer. In figure 3.13(a) a material with a larger lattice constant is deposited on the substrate and accordingly, the film lattice constant becomes compressed in the x-y plane and larger in the z direction. In figure 3.13(b) the material deposited has a smaller lattice constant and the film becomes tensely stressed.

Recall that the energy band structure of the materials and the consequent effective masses of the carriers are produced by the periodicity of the lattice. Thus mechanical stress (tensile or compressive) will have a strong effect on both the band structure and effective mass of the carriers.

Figure 3.14 shows the effect of tensile stress on silicon. In figure 3.14(a), the constant-energy surfaces are equally populated and for a current flowing in the $< 010 >$ direction, the carriers in the ellipsoids oriented on the z-axis and x-axis travel with

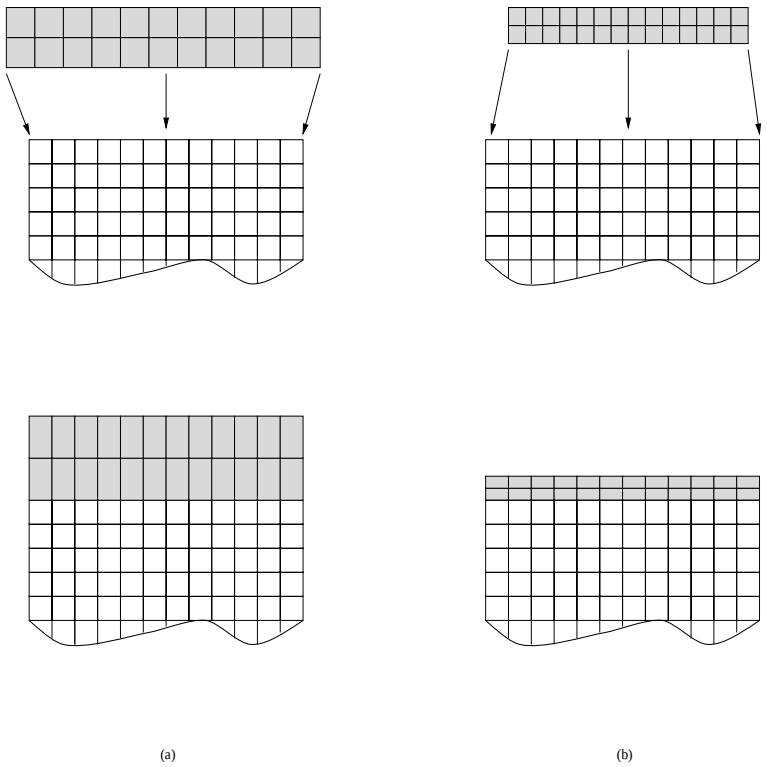

Figure 3.13: Effect of compressive and tensile stress. The strain is produced by epitaxy growth of a film layer of material with bulk lattice constant (a) larger or (b) smaller than the substrate. The film in (a) becomes compressed in the horizontal plane and elongated in the vertical direction. The film in (b) is tensile stressed in the horizontal plane and compressed in the vertical plane.

transverse effective mass. The carriers in the ellipsoids oriented in the y-axis, however, travel with longitudinal effective mass.

Once biaxial tensile stress on the x-y plane is applied to silicon, the constant-energy surfaces on the horizontal axis and those on the vertical axis separate in energy. The darkened vertical ellipsoids of constant-energy in figure 3.14(b) are of lower energy than the horizontal ellipsoids. Therefore, carriers tend to populate these vertical ellipsoids preferentially compared to the horizontal ellipsoids. Therefore, for a current flowing in the $< 010 >$, under tensile stress, the carriers will travel only with transverse effective mass.

The tensile stress on silicon can be changed depending only on the content of germanium in the underlying SiGe relaxed layer. The offset between the vertical ellipsoids and horizontal ellipsoids is given by equation 3.7

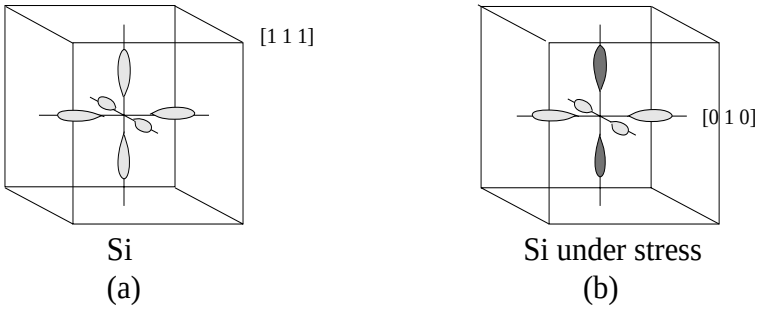

Figure 3.14: Constant-energy ellipsoids for (a) relaxed bulk silicon and (b) biaxially tensile stressed silicon. The tensile stress in the horizontal plane makes the vertical constant-energy ellipsoid (blackened) of less energy than the horizontal ellipsoids. In the case of tensile stressed silicon, a current flowing on the horizontal plane will have carriers traveling with only transverse effective mass.

$$\Delta E = 0.62x(eV), \text{ for } Si_{1-x}Ge_x \tag{3.7}$$

The Si-on-SiGe FETs are then built in accordance to the description in figure 3.15.

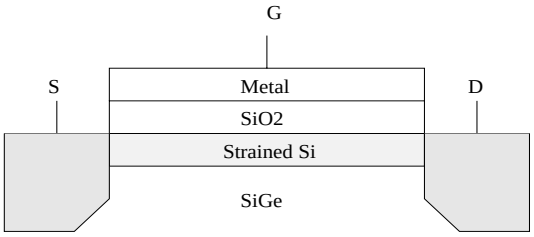

Figure 3.15: Si-on-SiGe FET. A thick relaxed layer of SiGe is grown and a thin layer of Si is epitaxially grown on top. The content of Ge in the SiGe relaxed layer defines the amount of stress imposed on the thin Si layer which is grown on top. The channel is created totally inside the Si layer.

In our experiments, there were two group of Si-on-SiGe FETs. One group used a SiGe layer with 10% germanium and the other group used a SiGe layer with 30% germanium. Therefore the offset between the vertical and horizontal ellipsoids was 62meV and 186meV, respectively.

Therefore, the 30% germanium $Si - on - Si_{0.7}Ge_{0.3}$ FETs will tend to keep the electrons traveling along the channel under a *single mode* more strongly under higher fields than the 10% germanium $Si - on - Si_{0.9}Ge_{0.1}$ FETs.

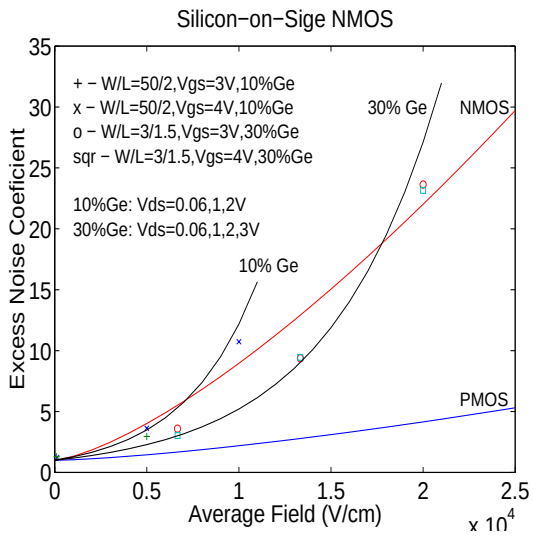

Figure 3.16: Excess noise factor for **long** channel Si-on-SiGe FETs. The 30% germanium content FETs develops much lower levels of excess noise than the 10% germanium content FETs.

In figure 3.16 and 3.17, the 30% germanium FETs show excess noise factor values diverging from the thermal value of *1* (one) much slower than the 10% germanium FETs, as expected from the theory.

In both figures 3.16 and 3.17, the *"typical results"* for standard sub-micron transistors of approximately the same *effective gate length* are plotted to provide some benchmark. These *"typical results"* for NMOS and PMOS are indicated by the names "NMOS" and "PMOS" in the figure. Moreover, the experimental results for the Si-on-SiGe FETs are not fitting curves, but only aid put in to the eye.

Notice that the germanium content makes the SiGe layer underneath these FETs be increasingly less thermally conductive. At higher levels of current, the 30% germanium FETs suffer *self-heating* much more strongly than the 10% germanium FETs. The benefit for low noise operation of the more tensely stressed FETs is washed out by the *lattice heating* and they may become equally noisy or noisier than the 10% germanium FETs. The effect of *self-heating* can be clearly seen from the I-V curves for the Si-on-SiGe FETs depicted in figure 3.18. The *negative* inclination of the I-V curves in saturation is more intense in the 30% Ge FET than in the 10% Ge FET. Note however that if the poorer thermal conductivity of the 30% Ge FET started to

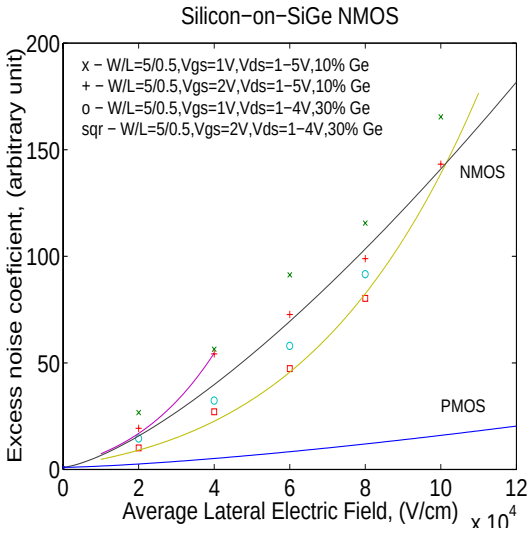

Figure 3.17: Excess noise factor for **short** channel Si-on-SiGe FETs. The 30% germanium content FETs develops much lower levels of excess noise than the 10% germanium content FETs.

play a role in noise performance at low current this would be a detrimental effect to this transistor, and the experimental fact that it develops *less* excess noise than the 10% Ge FET, is *per se* a strong evidence in favor of the interpretation that assigns to the *single-mode* carrier transport in the 30% Ge FET the reason for its lower noise operation.

Moreover, as mentioned before, this *negative* inclination made it clear that the DC-Resistance was the correct resistance value to use in the excess noise expressions for the FETs instead of the AC resistance.

Strained Si and *relaxed SiGe* layers form a heterojunction with lower conduction band energy on the Si side, and higher valence band energy on the SiGe side. As illustrated in figure 3.19, *only* NMOS transistors can be built because a PMOS transistor would have an extra undesired channel for holes in the SiGe layer.

However, a *relaxed Si* and *compressed SiGe* layer can be used as the basis for a *buried layer* PMOS, where holes travel through a channel created in the compressed SiGe layer, as illustrated in figure 3.20. [86, 87].

In this work, only NMOS structures depicted in figure 3.19 were used.

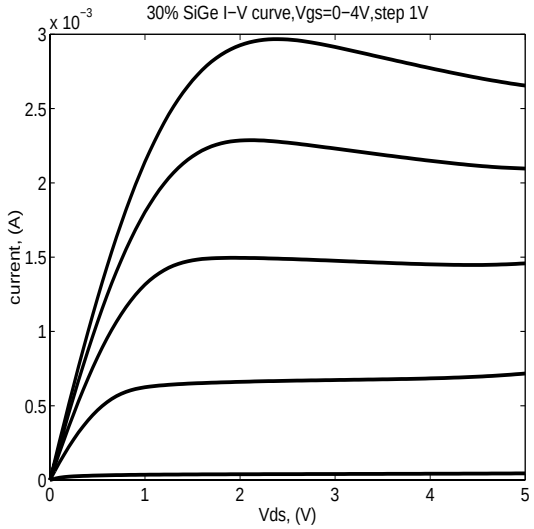

Figure 3.18: Drain current vs. drain-source voltage for a 30% Si-on-SiGe FET, W=5 μm, L=0.5 μm. At high drain current, self-heating in the channel causes the drain current to drop as V_{DS} increases.

3.4 Graded Doping Profile and Noise

The effect of the carrier doping profile in the generation of excess noise under high electric field transport is examined with a comparison between graded channel MOSFETs and standard (uniformly doped) MOSFETs.

According to the *Ergodic Method*, the concentration of carrier along the channel can be used to counter-effect the heating effect of the high lateral electric field. This counter-effect can be realized by doping the channel such that more carriers are available for conduction from source to drain.

The rationale is explained with the help of figure 3.21. Figure 3.21(a) shows that such a device is harder to invert close to the source than it is close to the drain. The channel close to the drain may even be always *"on"*, and the part of the channel close to the drain requires gate voltages different from zero to produce a channel. Thus, the amount of carriers available for conduction augments from source to drain. Such a device with an increasing carrier density from source to drain is accomplished by using a NMOS transistor built with double diffusion.

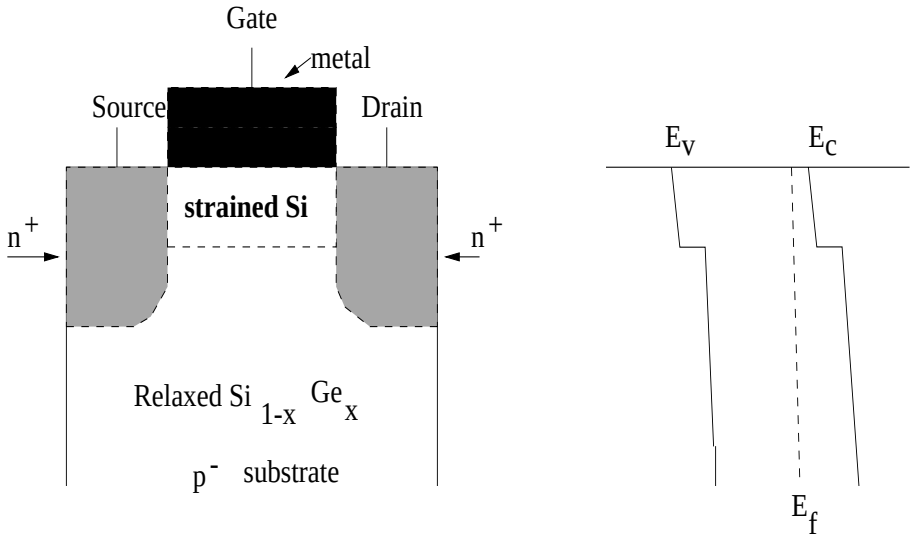

Figure 3.19: Device structure and band diagram. The $Si_{1-x}Ge_x$ layer is very thick and relaxed to the bulk lattice constant. The thin Si layer, ($\approx$ 200 Å) is tensile stressed to a larger lattice constant than its bulk constant.

Graded transistors of this type were first designed to produce higher speed operation. Consider figure 3.21(b). In a uniformly doped channel, the electric field peaks close to the drain (dashed line) far more strongly than for a graded doped channel (continuous line) when the transistor is biased into saturation. This is the normal operation regime for amplifiers (where the drain - output - current is intended to be only controlled by the gate - input - voltage and not by the drain - output - voltage).

Therefore, the electric field in graded channel MOSFETs tends to be more distributed than in uniformly doped MOSFETs. If the graded channel MOSFET has the electric field above the critical filed, E_c, (straight line in the plot) then the carriers will travel at the uniform *saturation velocity* for a greater fraction of the channel length than the uniformly doped MOSFET and this will make the graded channel MOSFET a faster device.

On the noise performance issue, the graded channel MOSFET is equally a superior device. In both uniform and graded channel MOSFET, at the source, there are many carriers and the velocity distribution is Maxwellian. The distribution is almost zero-centered since a small non-zero average velocity for the carriers will be enough to support a huge current through the channel. Because the electric current

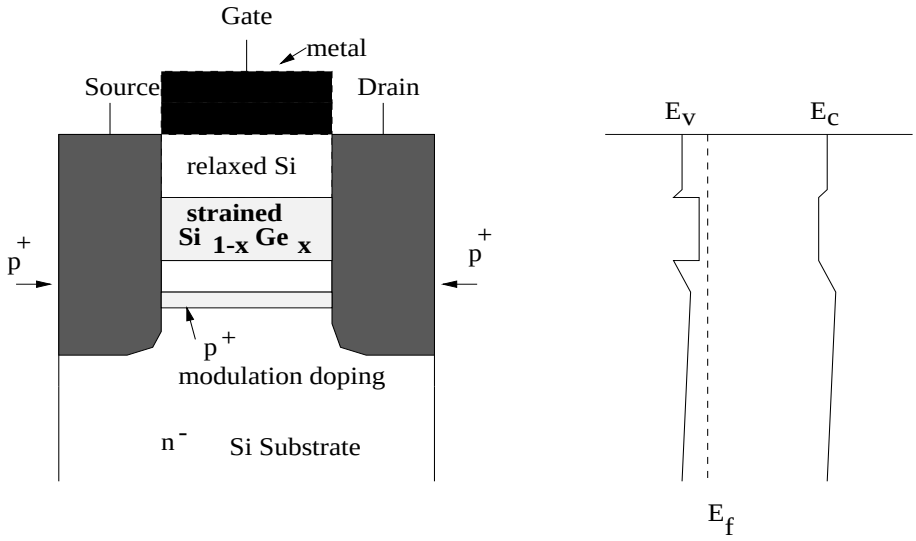

Figure 3.20: Modulation-doped p-MOSFET. Holes are supplied by the thin ($\approx$ 50Å) p-type doping spike (p^+) and these holes travel by the *buried* channel provided by the (compressively) stressed $Si_{1-x}Ge_x$ layer.

is proportional to the product of carriers density and average velocity, the velocity distribution inside the channel must be centered at a significantly non-zero value to produce the same electric current as the source. In figure 3.21(c), the distribution of velocities in the channel is exaggerated to illustrate this point.

Therefore, as the carriers enter the channel, they are *"pushed"* towards higher velocities, but this *push* is considerably softer in the graded than the uniform channel MOSFET because more carriers are available as one progresses down the channel, and the distribution does not need anymore to continuously shift it center as abruptly as in the case of the uniformly doped channel MOSFET.

Note that the interface source-channel can be equally noise prone for both uniform and graded channel MOSFET, but the noise performance of both transistors is defined by the development of the carrier transport along the channel and *not* at neither source-channel nor channel-drain interfaces.

Thus, according to the Ergodic Method, the velocity distribution function along the channel is narrower in graded channel than in the uniform channel MOSFET, as depicted in figure 3.21(c).

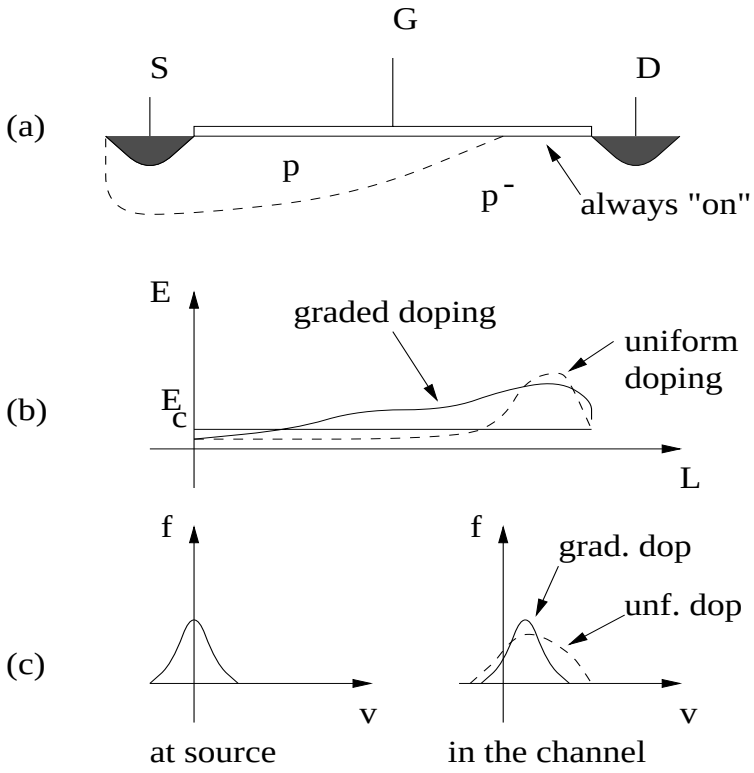

Figure 3.21: Graded channel n-MOSFET. (a) structure, (b) electric field better distributed than uniform doped FET, and (c) less variance in velocity distribution, hence less noise in **high electric fields**.

Figure 3.22 shows the results of excess noise factor for two graded channel MOSFETs. One has a gate width of $W = 480\mu m$ and the other has a gate width of $W = 360\mu m$. The lithographically defined gate length of both devices is $L_{drawn} = 1\mu m$, but the *effective* gate length is around $L_{effective} \approx 0.65\mu m$, since part of the channel close to the drain is by design always *"on"*.

Also, the devices in figure 3.22 were power transistors and their threshold voltages, V_{th}, were around $5V$. Therefore the gate voltages, $V_{GS} = 7, 8, 9V$, represent *gate overdrives* equivalent to the gate overdrive of the previous uniformly doped submicron NMOS and PMOS used in the experiments of this chapter, whose *"typical"* values for γ_f for the effective gate length of $0.65\mu m$ are plot on the same figure and referred to as the lines marked *"NMOS"* and *"PMOS"*. Note how close the excess noise factor for the graded channel NMOS FET gets to the PMOS values, which

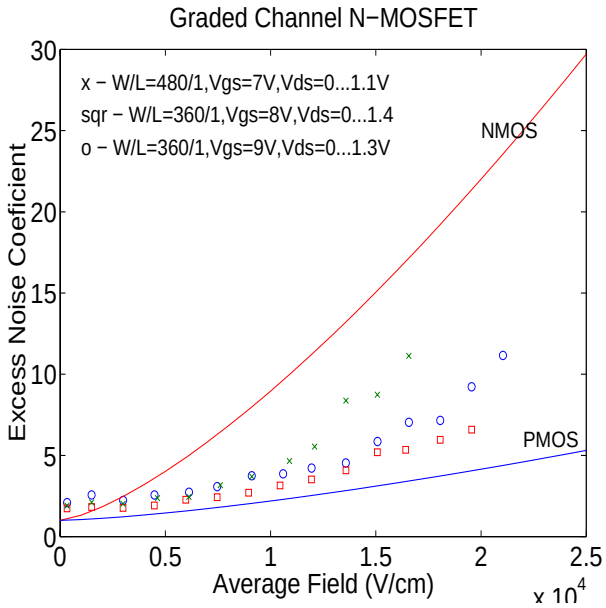

Figure 3.22: Excess noise factor for graded channel MOSFET. The graded channel MOSFET develops much lower excess noise than standard NMOS transistors of equivalent length and almost perform as good as PMOS transistors, which use much heavier carriers (holes).

are due to much heavier carriers (holes). This suggests the graded channel NMOS devices could be a strong contender in low noise device class, by virtue of their noise performance and simple manufacturing process.

Chapter 4

Data Analysis

This chapter connects the theoretical developments of the *Ergodic Method* of chapter 2 and the experimental results of chapter 3.

Inside the *Ergodic Method*, noise phenomena is related to the *variance* in carrier velocity distribution and it was predicted that under high electric fields the *excess noise* level is a function of the carrier effective mass, field, and carrier concentration along the transport media.

The goal in this chapter is to produce fitting curves based on the above concepts of the *Ergodic Method* that can approximate the experimental results presented in the previous chapter. These data fitting equations will explain the experimental results for noise level, reinforce the correctness of the predictions of the *Ergodic Method* and justify the assumptions accepted *a priori* during the theoretical development of the method.[89]

The values for effective masses of electrons and holes are taken from the low field regime and they are used in the high field regime as the *"first guesses"*. The high field effective masses should be found from an average of the second derivative of the energy bands at the energy of the carriers. It's assumed that these high field effective masses can be approximated *to first order* as a multiplicative constant times the the low field effective masses of carriers in the same material. This multiplicative constant is found from the fitting of the experimental data.

4.1 Sub-micron NMOS and PMOS Transistors, and Graded Channel n-MOSFET

The general behavior of the Excess Noise Factor, γ_f, in standard sub-micron NMOS and PMOS devices was already presented, but is reproduced here again in figure 4.1 for discussion. It is clear that the NMOS transistors develop more *excess noise* than PMOS transistors for the same *average lateral electric field*.

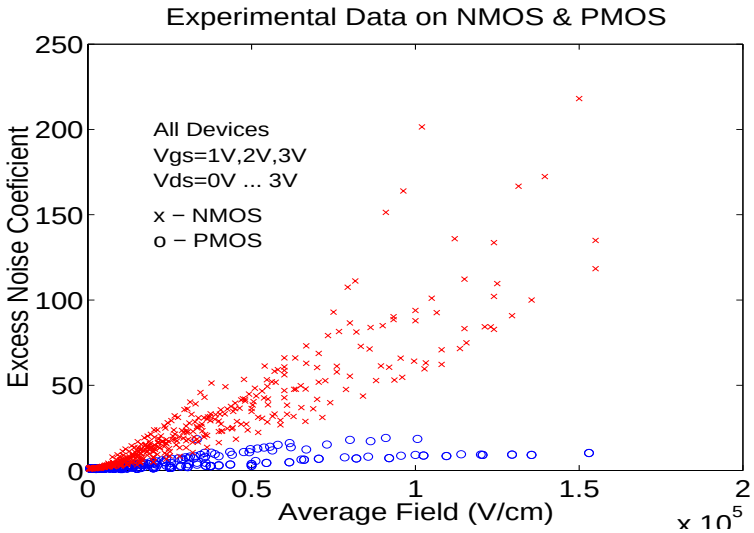

Figure 4.1: Excess noise factor, γ_f, for silicon NMOS and PMOS. All the NMOS devices ("x") develop more excess noise than PMOS devices ("o") at the same electric field.

Another important feature of γ_f is that it did *not* show any dramatic change in behavior for gate lengths varying from as large as $L_{eff} = 1.4\mu m$ to as small as $L_{eff} = 0.28\mu m$. Figure 4.2 shows the smooth evolution of γ_f for NMOS with $W = 10\mu m$ and $V_{GS} = 3V$.

Some spreading on γ_f does occur however for L_{eff} below $0.28\mu m$, as depicted in figure 4.3. This spreading is believed to be caused by mismatch on doping levels and defects not averaged out by the very small dimensions involved. Therefore, this spreading is not likely function of any fundamental physics behavior associated with our analysis, and the description of its statistical characteristics will be deferred in this present work.

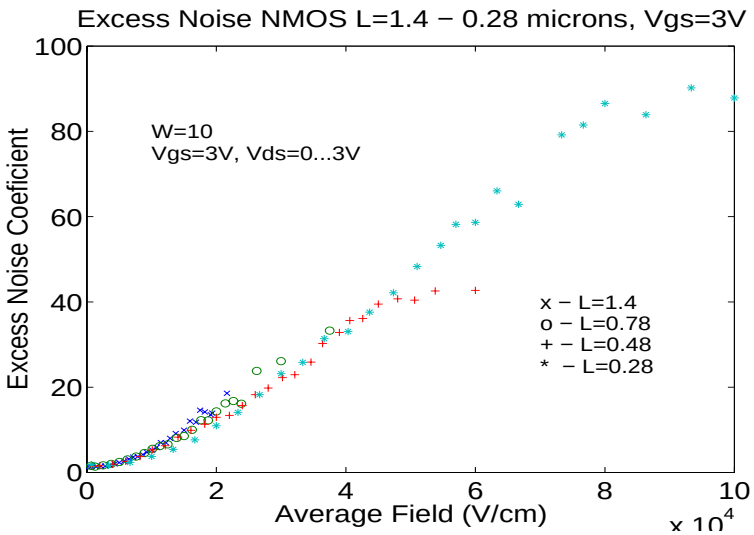

Figure 4.2: γ_f *vs.* average electric field for several transistor effective lengths. $W = 10\mu m$, $V_{GS} = 3V$, and V_{DS} varies from $0V$ to $3V$. As L_{eff} diminishes, transport is made under increasingly higher electric fields, and γ_f shows smooth evolution throughout the range of electric fields applied.

The analysis of the data from the standard NMOS and PMOS devices is done via the *Excess Noise Factor*, γ_f. The goal is to find the best fit for the experimental data from the *Ergodic Method* perspective. Thus, γ_f for these devices will be written as a function only of electric field, carrier concentration profile and effective mass of the carriers. Therefore, once these three variables are able to capture the physics and successfully makes the calculated γ_f follow the experimental γ_f, the correctness of the Ergodic Method and its *a priori* assumptions are reinforced.

Once a MOS transistor changes its transport condition with bias, the attempt to data fitting is made in **two** steps. The *first step* tries to fit data of γ_f for bias in the linear region. In this first step, γ_f is related to the *average electric field* along the channel, V_{DS}/L_{eff}. In the *second step*, γ_f is calculated for every slice of the channel, thus allowing the analysis progress from an *average electric field* to actual electric field, carrier concentration profile and effective mass of the carriers for every slice of the transistor channel. In this second step, the information on electric field, carrier concentration and thickness of the slices of the channel are found via the commercial device simulator ATLAS/SILVACO.[92]

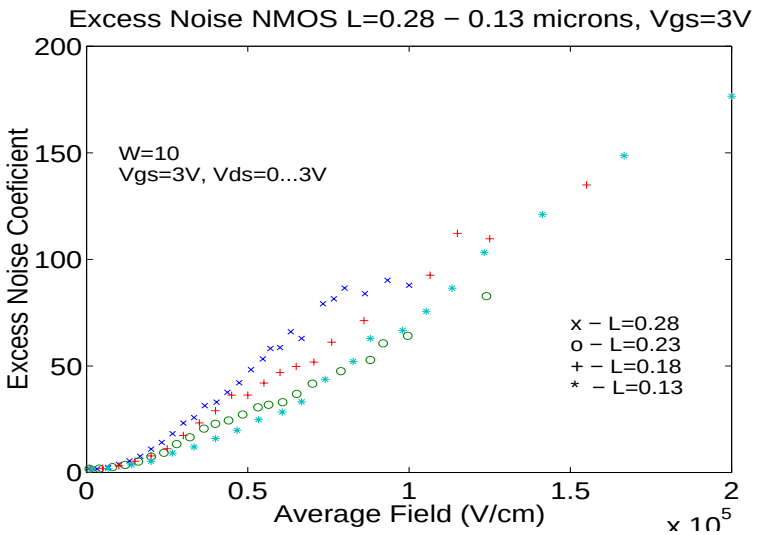

Figure 4.3: γ_f *vs.* average electric field. $W = 10\mu m$, $V_{GS} = 3V$, and V_{DS} varies from $0V$ to $3V$. The spreading in γ_f values is very likely function of mismatches on doping levels and defects on the silicon not averaged out by the small dimensions of the transistors used.

4.1.1 The General Behavior $\gamma_f = 1 + \alpha E^2$

Considering standard NMOS and PMOS, deep sub-micron transistors have γ_f *vs.* average electric field as depicted in figures 4.4 and 4.5. In both of these figures, the transistors have $L_{eff} = 0.18\mu m$, and are biased as close to the linear region as possible in order to avoid strong skewing of the carrier concentration profile along the channel. This allows the difference in effective mass of the carriers to be the dominant distinction in γ_f between NMOS and PMOS transistors.

In figures 4.4 and 4.5, the very good data fitting is made with $\gamma_f = 1+\alpha E^2$, where E is the average electric field, V_{DS}/L_{eff}, and α is a constant.

In figure 4.6, the data fitting for several device lengths is depicted. All the curves have the form $\gamma_f = 1+\alpha E^2$, and the values of α for every length and type of transistor are indicated in table 4.1.

An exponential fit for the data was also tried but with much less success than the polynomial form $\gamma_f = 1+\alpha E^2$. The coefficient α varies with transistor effective length, L_{eff}, and it will be shown to be function of both carrier effective mass and carrier concentration profile. Because of the definition of γ_f, this excess noise factor becomes

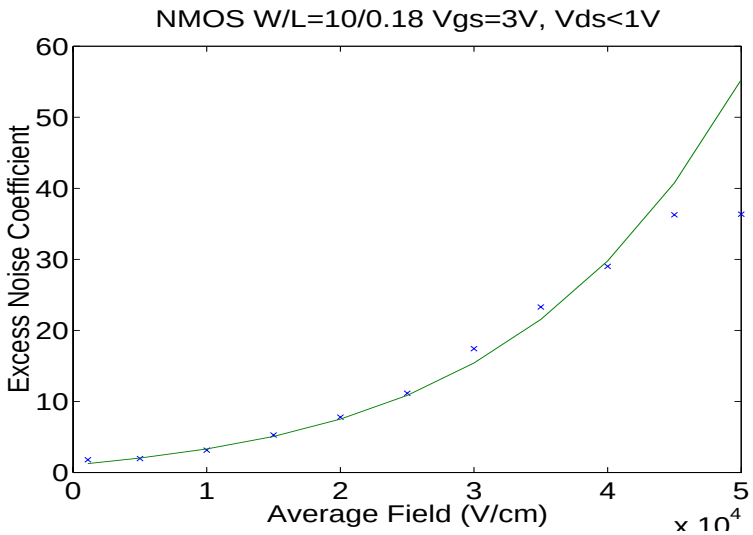

Figure 4.4: γ_f *vs.* average electric field for NMOS transistor of $L_{eff} = 0.18\mu m$. The device is biased such that most of the data is in the linear regime of transistor operation. Data is for $V_{GS} = 3V$.

closely related to the *excess temperature* that carriers acquire under high electric field transport. This is a reinforcement of the correctness of the polynomial data fitting above that other authors have published that the ratio of *excess temperature* of the carriers to the lattice temperature of the semiconductor material can be written as $T_e/T_{lattice} = 1 + cE^2$, where "$c$" is a function of carrier mobility (function of effective mass and dominant carrier scattering mechanism) among other terms. [90, 91].

From this point on, the data analysis concentrates on the coefficient α.

Table 4.1: **Coefficient α values for** $\gamma_f = 1 + \alpha E^2$

Effective Length (μm)	NMOS α (cm^2/V^2)	PMOS α (cm^2/V^2)
0.78	$5.2\text{x}10^{-8}$	$1.9\text{x}10^{-8}$
0.48	$3.6\text{x}10^{-8}$	$9.0\text{x}10^{-9}$
0.28	$2.4\text{x}10^{-8}$	$6.5\text{x}10^{-9}$
0.23	$1.6\text{x}10^{-8}$	$2.5\text{x}10^{-9}$
0.18	$1.7\text{x}10^{-8}$	$1.2\text{x}10^{-9}$

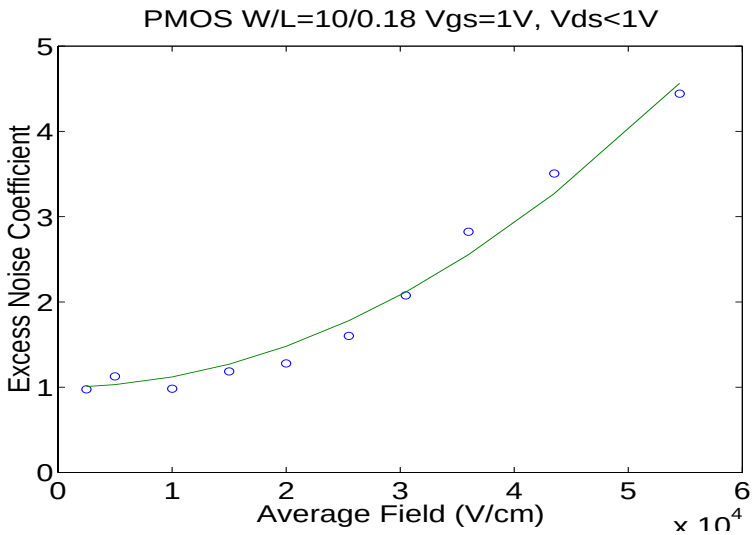

Figure 4.5: γ_f *vs.* average electric field for PMOS transistor of $L_{eff} = 0.18\mu m$. The device is biased such that most of the data is in the linear regime of transistor operation. Data is for $V_{GS} = 1V$.

4.1.2 The Coefficient α

Data analysis on the results which led to figure 4.6 showed that the dramatic change in inclination in the γ_f from PMOS to NMOS transistors is strongly correlated with the *effective mass* of the carriers (electrons for NMOS and holes for PMOS). Within the NMOS and PMOS however, the spreading of the curves for either one of these transistor types correlates strongly with *effective gate length*, but correlates with the *wrong* signal (negative coefficient in front of term $1/L_{eff}$).

The *wrong* signal in the correlation of the coefficient γ_f is an *artifact* caused by the use of the *average* electric field, $E = V_{DS}/L_{eff}$, in the plot of γ_f. Because of that, in figure 4.6, experimental data from *long channel* CMOS deeply in saturation aligns with experimental data from *short channel* CMOS deeply in the linear regime.

The problem with the plot in figure 4.6 is that a *long channel* CMOS, deep in saturation, have peak of electric field close to the drain significantly higher than the *average field* along the channel, and this pick can also be significantly higher than for a *short channel* CMOS deep in the linear region, with the same *average electric field*. In order to avoid this ambiguity, the search for the true dependence of the coefficient

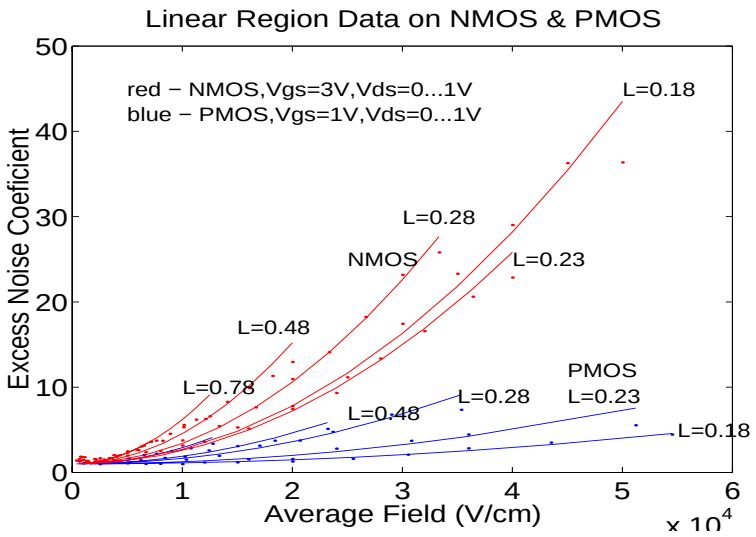

Figure 4.6: γ_f *vs.* average electric field for NMOS and PMOS transistors of various values of effective channel length, L_{eff}. All the curves follow the form $\gamma_f = 1 + \alpha E^2$, where E is the average electric field, V_{DS}/L_{eff}, and α is a constant.

Table 4.2: **BTE transformations**

BTE Term	standard representation	representation for γ_f
drift	$\frac{qE}{m^*}$	$\frac{qE}{m^*}$
diffusion	$v.\frac{\partial f}{\partial x}$	$\mu E \frac{\partial n}{\partial x}$

γ_f, or equivalently the coefficient α, has to be found for thin slices along the channel, and the summed over to fit the measured value of γ_f at the device terminals.

An excess noise factor, γ_{f_i}, can be defined for each slice *"i"*. The excess noise factor, γ_f is then:

$$\gamma_f = G \sum R_i \gamma_{f_i} \tag{4.1}$$

where G is the conductance of the entire transport media, measured from the external terminals (*source* and *drain* in our case), γ_{f_i} is the excess noise factor at each slice.

If the form $\gamma_{f_i} = 1 + \alpha_i E^2$ still applies for each slice, the general form for the coefficient α_i can be inferred by the similarity present in the Boltzmann Transport Equation, BTE, as described in table 4.2.

The expression for the excess noise factor, γ_{f_i}, can be thought as coming from the relations:

$$\gamma_{f_i} = 1 + \left(c_1 \frac{q}{m^*} E + c_2 \mu \frac{\partial n}{\partial x} E \right)^2 = 1 + \alpha_i E^2 \tag{4.2}$$

Therefore,

$$\alpha_i = \left(c_1 \frac{q}{m^*} + c_2 \mu \frac{\partial n}{\partial x} \right)^2 \tag{4.3}$$

or,

$$\alpha_i = \left(C_1 \frac{1}{m^*} + C_2 \frac{1}{(m^*)^{5/2} + (m^*)^{1/2} N_I} \frac{\partial n}{\partial x} \right)^2 \tag{4.4}$$

where C_1 and C_2 are constants and will be calculated by fitting experimental data. The mobility, μ, was written as function of effective mass, m^*, including acoustic phonon and ionized impurity scattering. N_I is the concentration of ionized impurities.

The data analysis for finding C_1 and C_2 included data from all the standard submicron transistors and graded-channel transistors as well. The *carrier concentration*, *lateral electric field*, and *length of slice* along the channel came from the commercial device simulator ATLAS/SILVACO.[92]

The biases, transistor sizes, and experimental values of γ_f used to find C_1 and C_2 is reported in tables 4.3 and 4.4. The last column 'γ_f (calc.)' contains the calculated results from the data fitting.

The average error between experimental values of γ_f and calculated value of γ_f is 30%. Despite the fact that the agreement can be made considerably with addition of other terms to the fitting, it is important to stress is that only first order effects provided by equations 4.2 to 4.4 were used, and therefore only two coefficients, C_1 and C_2 (equation 4.4) were adjusted during the data fitting. Moreover, as shown in figure 4.7, where the experimental and calculated γ_f's are plotted, it is clear that the data fitting tracks very closely the variations in the experimental γ_f. The values of γ_f were taken from the last two columns of tables 4.3 and 4.4. The plot in figure 4.7 thus provides evidence that the first order expressions for the dependency of γ_f on

Table 4.3: γ **and Uniform Channel FET**

L_{eff} (μm)	V_{GS} (V)	V_{DS} (V)	γ (exp.)	γ_f (calc.)
1.38	1	1	4.0	2.97
1.38	1	2	9.0	9.71
1.38	1	3	14	19.5
1.38	2	1	4.0	4.50
1.38	2	2	13	9.59
1.38	2	3	20	23.3
1.38	3	1	3.9	5.85
1.38	3	2	9.5	13.4
1.38	3	3	19	24.9
0.78	1	1	12	2.55
0.78	1	2	28	13.2
0.78	1	3	50	29.8
0.78	2	1	7.0	2.41
0.78	2	2	20	11.6
0.78	2	3	32	26.3
0.78	3	1	7.0	2.42
0.78	3	2	20	11.0
0.78	3	3	34	24.2
0.48	1	1	15	8.86
0.48	1	2	31	34.7
0.48	1	3	60	62.3
0.48	2	1	11	8.25
0.48	2	2	23	31.9
0.48	2	3	38	59.4
0.48	3	1	12	7.72
0.48	3	2	35	30.4
0.48	3	3	43	54.4

Table 4.4: γ **and Uniform Channel FET**, continuation

L_{eff} (μm)	V_{GS} (V)	V_{DS} (V)	γ (exp.)	γ_f (calc.)
0.28	1	1	35	16.2
0.28	1	2	70	61.2
0.28	1	3	190	132.1
0.28	2	1	20	14.2
0.28	2	2	55	48.5
0.28	2	3	95	106.4
0.28	3	1	25	12.4
0.28	3	2	70	44.5
0.28	3	3	90	88.4
0.18	1	1	45	27.0
0.18	1	2	100	91.4
0.18	1	3	200	-
0.18	2	1	35	24.2
0.18	2	2	70	79.6
0.18	2	3	115	151.0
0.18	3	1	40	22.2
0.18	3	2	85	69.7
0.18	3	3	135	130.8

electric field and carrier concentration profile do capture most of the basic physics involved.

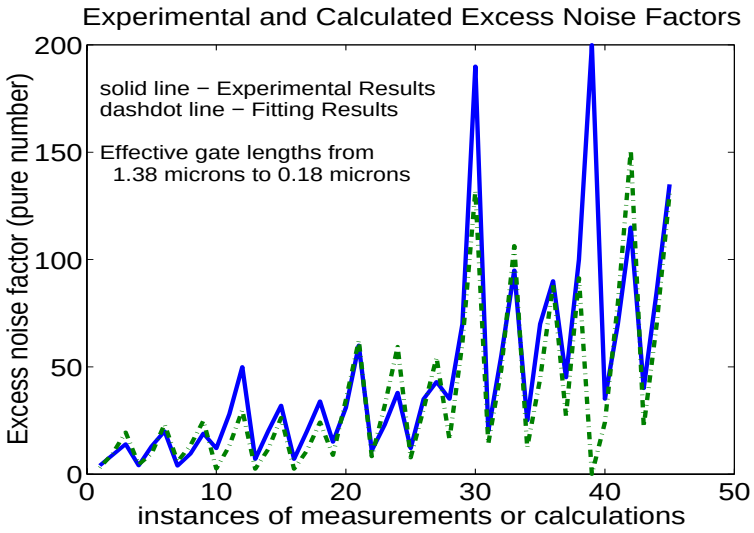

Figure 4.7: γ_f at the transistor terminals calculated by weighted summation of γ_{f_i} from every slice along the transport media. γ_f from experimental results variations with bias and transistor gate length are closely closely tracked by γ_f from data fitting with only the coefficients C_1 and C_2 introduced in equation 4.4. The horizontal axis indicates the lines on tables 4.3 and 4.4 from which the γ_f were taken.

The experimental results for the graded channel MOSFET however were significantly better than those calculated via equations 4.2-4.4. In order to accommodate the results for the graded channel MOSFET, other terms needed to be added to the expression of equations 4.2-4.4. These terms were only function of E and $\partial n/\partial x$ to higher powers, and therefore despite the difficulties to provide a physical interpretation for the mechanism which required such terms, it is still clear that the only variables necessary to accommodate all the experimental results for both uniform and graded channel MOSFET excess noise are the electric field and the carrier concentration profile along the channel, reinforcing the prediction from the Ergodic Method.[93, 94, 95]

4.2 Single Mode and Multimode MOSFET

The experimental results for both multimode and single mode transport are analyzed in the same framework of data fitting developed in the previous section for the standard and graded-channel MOSFET.

The transistors used for comparison between multimode and single mode transport are Si-on-SiGe FETs. Data for the first order analysis comes from ATLAS, the same way as in the previous section. The only difference is that in the previous section, *two* different effective mass carriers took part in the the transport, and this was masked by the data fitting used, which analyzed the data as it were all coming from a sort of "equivalent" effective mass carrier.

In the Si-on-SiGe FET case, however, the carriers transport predominantly as one single effective mass (*transverse effective mass*). At high fields, however, the 10% Ge $Si-on-Si_{0.9}Ge_{0.1}$ FET tends to have slightly more carriers going into the *longitudinal effective mass* transport mode than the 30% Ge $Si-on-Si_{0.7}Ge_{0.3}$ FET.

It is than straightforward to adapt the Si-on-SiGe FET results into the data fitting developed in the previous section, and the procedure for that is outlined below:

For Si-on-SiGe FETs, the offset of vertical and horizontal constant-energy ellipsoids is given by:

$$\Delta E_s = 0.67x \text{ eV} \tag{4.5}$$

where *"x"* is the fraction of germanium in the SiGe relaxed layer. Therefore, the carrier population will be split according to:

$$n_t = \frac{n}{1 + e^{-\Delta E_s/V_{th}}} \tag{4.6}$$

$$n_l = \frac{ne^{-\Delta E_s/V_{th}}}{1 + e^{-\Delta E_s/V_{th}}} \tag{4.7}$$

where n, n_t, and n_l are respectively the total carrier concentration, carrier concentration with transverse effective mass and carrier concentration with longitudinal effective mass at each mesh point along the transistor transport media. $V_{th} = K_B T_L/q$.

Table 4.5: **Concentration of Carriers and Effective Mass - Low Field**

	10% Ge Si-on-SiGe FET	30% Ge Si-on-SiGe FET
Trans-Eff-Mass Carrier	$0.93n$	$1n$
Long-Eff-Mass Carrier	$0.07n$	0

Table 4.6: **Concentration of Carriers and Effective Mass - High Field**

	10% Ge Si-on-SiGe FET	30% Ge Si-on-SiGe FET
Tran-Eff-Mass Carrier	$0.7n$	$0.9n$
Long-Eff-Mass Carrier	$0.3n$	$0.1n$

For the case of the 10% and 30% germanium content Si-on-SiGe FETs used in this work, equations 4.5 - 4.7 give the results in table 4.5 for low electric field transport.

Under high electric field transport, a larger number of carriers will go into the *longitudinal-effective-mass* transport mode. Thus, the data fitting starts with the numbers from table 4.5 as *"good guesses"* and the number of carriers in each mode is adjusted until the best fit is found. The coefficients C_1 and C_2 are the same as those in the previous section for standard and graded-channel MOSFETs.

For our particular case, a faster calculation which gives acceptable results takes advantage of the fact that C_1 made its term in equation 4.4 dominate the term with C_2 for most of the data in the high electric field transport. Then the ratio between the γ_f results for 10% Ge and 30% Ge Si-on-SiGe FETs is the square of the inverse proportion of the ratios of their carrier's "equivalent" effective mass. "Equivalent" effective mass here meaning the weighted average of the transverse and longitudinal effective masses presented in the transport; the weights being the concentration of carriers with each one of these effective masses.

Straightforward calculation leads to the conclusion that the population of carriers transported with transverse and longitudinal effective mass for both 10% Ge and 30% Ge FETs under high field transport should be altered to the values in table 4.6, to accommodate the experimental results for γ_f found for both type of transistors.

These results reinforce the *Ergodic Method* interpretation of the difference in excess noise development in Si-on-SiGe FETs as coming from differences in *transport modes*.

Chapter 5

Summary and Conclusion

This research work started with two *known facts* and one *assertion*. The contribution of this thesis is the development of both a theoretical framework and experimental results to prove this *assertion*.

The first known fact was that at *equilibrium*, all conductors and semiconductors develop *thermal noise* the same way, and that the *thermal noise* power is only a function of the resistance and temperature of the conductor or semiconductor. The second known fact was that when a DC-current flows, conductors and semiconductors develop *additional* noise, usually called *excess noise*.

The assertion which launched this research effort was that while different materials develop this *additional* noise differently, they develop it in a **predictable** and **controllable** way, thus enabling device optimization for low noise operation.

The theoretical structure was built with the development of the *Ergodic Method* and the use of the Boltzmann Transport Equation formalism for noise analysis. This work had two main objectives: 1) the development of device simulation tools capable of *predicting* the noise performance of new semiconductor devices at design time; and 2) the design of truly low noise semiconductor devices based on new design guidelines. This required that noise phenomena be understood from a phenomenological perspective which is linked to device parameters under the control of the device designer.

In this work, both objectives were tackled and their basis established. The main strategy was to start with the basic idea of following the movement of a single carrier

inside a semiconductor material. This work establishes that the *mean* and *variance* of the velocity distribution in the random movement of this single carrier define both the DC-current and current-noise produced by this carrier. This result is averaged over the entire number of carriers contributing to the electric current and present in a slice of semiconductor. Note however, that these are *time domain* measures and do not lend themselves to the available transport analysis framework for semiconductors.

The enabling step was made by the *ergodic assumption*, which allowed this noise calculation to be made based *not* on the analysis of the movement of a single carrier *in time*, but to be based on the *ensemble* velocity distribution of the carriers contributing to the current at a *given time*. Then a path to use the semiconductor carrier velocity distribution function in the noise calculations is open, provided this distribution function only refers to electrons above the conduction band minimum or holes below the valence band maximum (non-degenerate).

The next step was to establish how a device simulator should work, recover known results for thermal noise (*Nyquist Theorem*), and most importantly, using the vantage point provided by this method, anticipate several *new* and *non-obvious* effects of carrier effective mass and concentration profiles on noise performance. The prediction of new and non-obvious effects on noise performance is the key aspect of this work. The prediction of new and non-obvious effects on noise makes it clear that the *Ergodic Method* was providing a deeper and more physically based understanding of noise phenomena, principally (excess) noise in high electric field transport.

The correctness of the Ergodic Method was verified by experimentally testing structures with predicted lower noise properties. Experimental measurements were performed to check the effect of carrier mass and concentration profile on *excess noise*. At the onset of this work, an *excess noise factor*, γ_f, was introduced as a figure of merit in comparing noise performance of devices of differing dimensions and technologies. These γ_f results were the numerical test for the *Ergodic Method* predictions. One of the most important results of this analysis and formalism is that it is valid for the *high electric field transport* regime in sub-micron devices.

In this regime, it is assumed that intra and inter valley scattering occur with equal probability. Carriers can thus change their effective mass, or *mode of transport*, and

there is a different distribution function for each effective mass. These different distributions are then added to form a *total* distribution function describing the probability of a carrier having any given velocity. These are the main particularity and limit of validation of the *means of computing* noise power from carrier velocity distribution taken in this work. It is important though to note that a change in this means of calculation, or even a lack of validation of the Boltzmann formalism in some transport regime, does *not* affect the principles of the *Ergodic Method* on its ground. The *Ergodic Method* only required an ergodic assumption to start working. Depending on the transport regime, the *Ergodic Method* is allowed to use the Boltzmann formalism, or any other more sophisticated one, without lack of generality, and also to use more elaborate routine to compute noise power from the distribution function if the scattering mechanisms and transport regime don't allow the treatment of adding multi Gaussian shapes as was used in this work.

The development of the Ergodic Method provided a vantage point from where remarkable results for low noise semiconductor-device operation can be appreciated. At this vantage point, remarkable results which were achieved using stress or graded-channel in a field effect transistor (FET) [96, 97, 98, 99] are now understood under a unified framework. New designs that combine all these techniques in order to obtain low noise operation by making the carriers travel with the narrowest possible velocity distribution are on the path for the ultimate low noise FET. [100].

The experiments in this work were based on bulk and strained silicon FETs with uniform and graded channel doping, but the central concepts developed are readily extendible to other semiconductor systems.

5.1 Suggestions of Future Work

As in any establishment of a new theory or method, several new perspectives emerge. New uses and tests for the theory are generated. There are three further developments that became visible from this work. The first is the investigation of other formalisms for transport analysis that extend the *Ergodic Method* to other transport regimes. The second area is the creation of a new and robust device simulator based on the *Ergodic*

Method. The third area is the design of new low noise FETs based upon the central design rule of creating devices that force the carriers to travel with the narrowest possible velocity dispersion. The first area is of great scientific interest and extends noise analysis based upon the *Ergodic Method* into the transport analysis realm. The second and third areas shares both scientific and industrial interest and will hopefully provide a path for scientific understanding of noise and great commercial impact in the rapidly expanding telecommunications world.

Bibliography

Chapter 1: Introduction

[1] A. Hajimiri and T.H. Lee, "A General Theory of Phase Noise in Oscillators," IEEE Journal of Solid-State Circuits, February 1998, pp. 179-194.

[2] D.K. Shaefer and T.H. Lee, "A 1.5V, 1.5 GHz CMOS Low Noise Amplifier," IEEE Journal of Solid-State Circuits, May 1997, pp. 745-759.

[3] T. Blalack et al., "Experimental Results and Modeling of Noise Coupling in a Lightly Doped Substrate," IEDM Tech. Dig., pp. 23.3.1-23.3.4, Dec. 1996.

[4] D.K. Su et al., "Experimental Results and Modeling Techniques for Substrate Noise in Mixed-Signal Integrated Circuits," IEEE J. Solid-State Circuits, vol. 28, pp. 420-430, Apr. 1993.

[5] L.M. Franca-Neto et al., "Low Noise FET Design for Wireless Communications," IEEE/IEDM, Digest of Technical Papers, Washington, DC, December, 1997.

[6] L.M. Franca-Neto and J.S. Harris, "Low Noise RF FET Design Using Graded Doping and Stress," IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems," Ann Arbor, Michigan, 17-18 September 1998.

[7] C. Kittel and H. Kroemer, *Thermal Physics*, W.H. Freeman and Company, 2nd ed., 1980.

[8] M.J. Buckingham, "Noise in Electronic Devices and Systems," John Wiley & Sons, 1983.

[9] A. Van der Ziel, "Fluctuation Phenomena in Semi-conductors," London, Butterworths Scientific Publications, 1959.

[10] A. Van der Ziel, "Noise," New York, Prentice-Hall, 1954.

[11] A. Van der Ziel, "Noise: Sources, Characterization, Measurement," Prentice-Hall, New Jersey, 1970.

[12] R.C. Liu, "Quantum noise in mesoscopic electron transport," Ph.D. Thesis, Stanford University, 1998.

[13] W.H. Louisell, "Radiation and Noise in Quantum Electronics," McGraw-Hill, 1964.

[14] D.A. Bell, "Noise and the Solid State," Wiley, 1985.

[15] Sh. Kogan, "Electronic Noise and Fluctuations in Solids," Cambridge University Press, 1996.

[16] A. Leon-Garcia, "Probability and Random Processes for Electrical Engineering," 2nd ed., Addison Wesley, 1994.

[17] A. Papoulis, "Probability, Random Variables, and Stochastic Processes," 3rd ed., New York, McGraw-Hill, 1991.

[18] C. Ruhla, "The Physics of Chance", Oxford University Press, 1993.

[19] A. Van Der Ziel, Sixth Quantum 1/F Noise and other Low Frequency Fluctuations in Electronic Devices Symposium, Woodbury, N.Y.,1996.

[20] D. Costa, "Microwave and Low-frequency Noise Characterization of NPN AlGaAs/GaAs Heterojunction Bipolar Transistors," Ph.D. Thesis, Stanford University, 1991.

[21] R.R. Brynsvold, "Conductance, Conductance Fluctuations, and Low-Frequency Noise in Amorphous-silicon-barrier Tunnel Junctions," Ph.D. Thesis, Stanford University, 1989.

[22] J. Rhayem et al., "1/f Noise Investigations in Small Channel Length Amorphous Silicon Thin Film Transistors," J. of Appl. Phys., **83**, 7, 1998.

[23] M.E. Pistol et al., " Random Telegraph Noise in Photoluminescence from Individual Self-assembled Quantum Dots," Phys. Rev. B, **59**, 16, 1999.

[24] H.H. Mueller and M. Schulz, "Random Telegraph Signal: An Atomic Probe of the Local Current in Field-Effect Transistors," J. of Appl. Phys., **83**, 3, 1998.

[25] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," Nature **119**, 50 (1927).

[26] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," Phys. Rev. **32**, 97 (1928).

[27] H. Nyquist, "Thermal Agitation of Electric Charge in Conductors," Phys. Rev. **32**, 110 (1928).

[28] H. B. Callen and T. A. Welton, "Irreversibility and Generalized Noise," Phys. Rev. **83**, 34 (1951).

[29] S. Datta, "Electronic Transport in Mesoscopic Systems." Cambridge University Press, 1995.

[30] C.W.J. Beenaker and H. van Houten, "Quantum Transport in Semiconductor Nanostructures," Solid State Physics, vol. 44, eds. H. Ehrenreich and D. Turnbull, Academic Press, 1991.

[31] Th. Martin and R. Landauer, "Wave-packet Approach to Noise in Multichannel Mesoscopic Systems," Phys. Rev. B, **45**, 1742 (1992).

[32] M. Buttiker, "Scattering Theory of Current and Intensity Noise Correlations in Conductors and Wave guides," Phys. Rev. B, **46**, 12485 (1992).

[33] M. Buttiker et al., "Dynamic Conductance and the Scattering Matrix of Small Conductors," Phys. Rev. Lett., **70**, 4114 (1993).

[34] C.M. van Vliet, "The Diffusion Noise Source and The Steady State Fluctuation-Dissipation Theorem for a Boltzmann Gas," Physica 133A (1985) 35-52.

[35] C.M. van Vliet, "Macroscopic and Microscopic Methods for Noise in Devices," IEEE Trans. Elec. Dev., November 1994.

[36] A. van der Ziel et al., "1/f Noise in Mobility Fluctuations and The Boltzmann Equation," Physica 121B (1983) 420-422.

[37] J.P. Nougier, "Fluctuations and Noise of Hot Carriers in Semiconductor Materials and Devices," IEEE Trans. Elec. Dev., November 1994.

[38] H.S. Min and D.Ahn, "Langevin Noise Sources for The Boltzmann Transport Equations with the Relaxation-Time Approximation in Nondegenerate Semiconductors," J.Appl.Phys. **58**(6), 15 September 1985.

[39] K.G. Moh et al., "Equivalent Noise Source for Boltzmann Transport Equation with Relaxation-Time Approximation in Nondegenerate Semiconductors," J.Appl.Phys. **74**(10), 15 November 1993.

[40] J.P. Nougier, "Fluctuations and Noise of Hot Carriers in Semiconductor Materials and Devices," IEEE Trans. Elec. Dev., November 1994.

[41] F. Bonani et al., "An Efficient Approach to Noise Analysis Through Multidimensional Physics-Based Models," IEEE Trans. Elec. Dev., January 1998.

[42] W. Shockley, in "Quantum theory of atoms, molecules, and the solid state; a tribute to John C. Slater," edited by Per-Olov Lhowdin. New York, Academic Press, 1966.

Chapter 2: The Ergodic Method

[43] R.C. Tolman, *The Principles of Statistical Mechanics*, Dover Publications, New York, NY, 1979.

[44] C.Z. Mooney, "Monte Carlo simulation," Sage Publications, 1997.

[45] G.S. Fishman, "Monte Carlo: concepts, algorithms, and applications," New York, Springer-Verlag, 1996.

[46] M. Fu and J.-Q. Hu, "Conditional Monte Carlo: gradient estimation and optimization applications," / Michael Fu, Boston, Kluwer Academic Publishers, 1997.

[47] H. Stewart, "An Introduction to the Theory of the Boltzmann Equation," New York, Holt, Rinehart and Winston, 1971.

[48] C. Cercignani, "Theory and Application of the Boltzmann Equation," Edinburgh, Scottish Academic Press, 1975.

[49] C. Cercignani, "The Boltzmann Equation and its Applications," New York, Springer-Verlag, 1988.

[50] M. Lundstrom, "Fundamentals of Carrier Transport", in *Modular Series on Solid State Devices*, Vol X, Addison-Wesley, Reading, MA, 1989.

[51] S. Datta, "Electronic Transport in Mesoscopic Systems." Cambridge University Press, 1995.

[52] S. Datta, "Quantum Phenomena," in *Modular Series on Solid State Devices*, Vol VIII, Addison-Wesley, Reading, MA, 1989.

[53] T. Ando et al., "Mesoscopic Physics and Electronics," NanoScience and Technology Series, Springer, 1998.

[54] C. Kittel and H. Kroemer, *Thermal Physics*, W.H. Freeman and Company, 2nd ed., 1980.

[55] W.A. Harrison, "Solid State Theory," New York, Dover Publications, 1979.

[56] J.J. Sakurai, "Modern Quantum Mechanics," Addison-Wesley, 1994.

[57] H. Kroemer, "Quantum Mechanics for Engineering, Materials Science, and Applied Physics," Prentice Hall series in Solid State Physical Electronics, Prentice Hall, 1994.

[58] C.M. Wolfe et al., "Physical Properties of Semiconductors", in *Prentice Hall Series in Solid State Phys. Elec.*, Prentice Hall, 1989.

[59] W. Shockley, "Electrons and Holes in Semiconductors," D. van Nostrand Company, Inc., 1950.

[60] A. Van der Ziel, "Noise: Sources, Characterization, Measurement," Prentice-Hall, New Jersey, 1970.

[61] A. Van der Ziel, "Noise in Solid State Devices and Circuits," New York, Wiley, 1986.

[62] A.A. Abidi, "High-Frequency Noise Measurement on FET's with Small Dimensions," *IEEE Trans. on Electron Devices*, vol. ED-33, no. 11, pp. 1801-1805, Nov. 1986.

[63] R.P. Jindal, "Hot-electron effects on channel thermal noise in fine-line NMOS field-effect transistors," IEEE Trans. Electron Devices, vol. ED-33, pp. 1395-1397, September 1986.

[64] S. Tedja et al., "Analytical and experimental studies of thermal noise in MOSFET's," IEEE Trans. Electron Devices, vol. 41, pp. 2069-2075, November 1994.

[65] B. Wang et al., "MOSFET thermal noise modeling for analog integrated circuits," IEEE J. Solid-State Circuits, vol. 29, pp. 833-835, July 1994.

[66] R. People, "Physics and applications of Ge_xSi_{1-x}/Si strained-layer heteroestructures," IEEE J. Quantum Electron., **QE-22**(9), p. 1696, 1986.

[67] J.R. Chelikowsky and M.L. Cohen, "Nonlocal pseudopotencial calculations for the electronic structure of eleven diamond and zinc-blend semiconductors," Phys. Rev. B, **14**(2), p. 556, 1976.

[68] R. Braunstein et al., "Intrinsic optical absorption in germanium-silicon alloys," Phys. Rev., **109**(3), p. 695, 1958.

[69] D.V. Lang et al., "Measurement of the band gap of Ge_xSi_{1-x}/Si strained layer heterostructure," Appl. Phys. Lett., **47**(12), p. 1333, 1985.

[70] J.C. Hensel and G. Feher, "Cyclotron resonance experiments in uniaxially stressed silicon: valence band inverse mass parameters and deformation potentials," Phys. Rev., **129**(3), p. 1041, 1963.

[71] T. Manku and A. Nathan, "Effective mass for strained p-type $Si_{1-x}G_x$," J. Appl. Phys., **69**(12), p. 8414, 1991.

[72] D.K. Nayak and S.K. Chun, "Low-field hole mobility of strained Si on (100) $Si_{1-x}G_x$ substrate," Appl. Phys. Lett., **64**(19), p. 2514, 1994.

[73] T. Manku and A. Nathan, "Energy-band structure for strained p-type $Si_{1-x}G_x$," Phys. Rev. B., **43**(15), p. 12634, 1991.

[74] J.M. Hinckley and J. Singh, "Hole transport theory in pseudomorphic $Si_{1-x}G_x$ alloys grown on Si(001) substrates," Phys. Rev. B, **41**(5), p. 2912, 1990.

[75] C.G. Van de Walle and R.M. Martin, "Theoretical Calculations of heterojunction discontinuities in the Si/Ge system, " Phys. Rev. B, **34**(8), p. 5621, 1986.

[76] C. Herring and E. Vogt, "Transport and deformation-potential theory for many valley semiconductors with anisotropic scattering," Phys. Rev., **101**(3), p. 944, 1956.

[77] T. Vogelsang and K.R. Hofmann, "Electron transport in strained Si layers on $Si_{1-x}G_x$ substrates," Appl. Phys. Lett., **63**(2), p.186, 1993.

[78] T. Manku and A. Nathan, "Electron drift mobility model for devices based on unstrained and coherently strained $Si_{1-x}G_x$ grown on $< 001 >$ silicon substrate," IEEE Trans. Electron Devices, Vol. 39-9, p.2082, 1992.

[79] R. People and J.C. Bean, "Band alignments of coherently strained Ge_xSi_{i-x}/Si heterostructures on $< 001 > Ge_ySi_{1-y}$ substrates," Appl. Phys. Lett., **48**(8), p. 538, 1986.

[80] T. Vogelsang and K.R. Hofmann, "Electron mobilities and high-field drift velocity in strained silicon on silicon-germanium substrates," IEEE Trans. Electron Devices, vol. 39-11, p. 2641, 1992.

[81] J. Ma et al., "Silicon RF-GCMOS IC Technology for RF Mixed-Mode Wireless Applications," IEEE Radio Frequency Intgr. Circ. Symp., Denver, CO, June 1997.

Chapter 3: Experimental Developments

[82] A. Van der Ziel, "Noise in Solid State Devices and Circuits," John Wiley & Sons, 1986.

[83] A.A. Abidi, "High-Frequency Noise Measurement on FET's with Small Dimensions," *IEEE Trans. on Electron Devices*, vol. ED-33, no. 11, pp. 1801-1805, Nov. 1986.

[84] C. Kittel and H. Kroemer, *Thermal Physics*, W.H. Freeman and Company, 2nd ed., 1980.

[85] A. Van der Ziel, "NOISE," New York, Prentice-Hall, 1954.

[86] J.J. Welser, "The Application of Strained-Silicon/Relaxed-Silicon Germanium Heterostructures to Metal-Oxide-Semiconductor Field-Effect Transistors," Ph.D. Thesis, Stanford University, December 1994.

[87] J. Welser et al., "Strain Dependence of the Performance Enhancement in Strained-Si n-MOSFETs," IEEE IEDM Tech. Dig., December 1994.

[88] J. Ma et al., "Silicon RF-GCMOS IC Technology for RF Mixed-Mode Wireless Applications," IEEE Radio Frequency Intgr. Circ. Symp., Denver, CO, June 1997.

Chapter 4: Data Analysis

[89] S. Brandt, "Data Analysis: statistical and computational methods for scientists and engineers," 3rd ed., New York, Springer, 1999.

[90] S.M. Sze, "Physics of Semiconductor Devices," 2nd. edition, John Wiley & Sons, 1981.

[91] J.L. Moll, "Physics of Semiconductors," McGraw Hill, New York, 1964.

[92] Atlas, Device Simulator, SILVACO International, Inc, 1996.

[93] T.W. Anderson and J.D. Finn, "The New Statistical Analysis of Data," New York, Springer, 1996.

[94] J. Mandel, "The Statistical Analysis of Experimental Data," New York, Interscience Publishers, 1964.

[95] "Statistical Data Analysis and Inference," edited by Yadolah Dodge. International Conference on Recent Developments in Statistical Data Analysis and Inference, Switzerland, 1989.

Chapter 5: Summary and Conclusions

[96] R.E. Williams and D.W. Shaw, "Graded Channel FET's: Improved Linearity and Noise Figure," IEEE Trans. Elec. Dev., June, 1978.

[97] J. Ma et al.,"Silicon RF-GCMOS IC Technology for RF Mixed-Mode Wireless Applications," IEEE Radio Frequency Intgr. Circ. Symp., Denver, CO, June, 1997.

[98] P.R. de La Houssaye et al., "Silicon MOSFET's with Very Low Microwave Noise,"

[99] R.A. Johnson et al., "Advanced Thin-Film Silicon-on-Sapphire Technology: Microwave Circuit Applications," IEEE Trans. Elec. Dev., (submitted)

[100] L.M. Franca-Neto and J.S. Harris, Jr.,"Low Electron Velocity Dispersion Transistor," disclosure for the Stanford University Office of Technology Licensing. In preparation.