Monolithic GaAs VLSI Optoelectronic Neuron Arrays

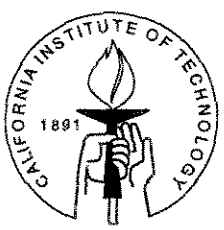
Thesis by

Jiafu Luo

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy



California Institute of Technology

Pasadena, California

1996

(Submitted September 27, 1995)

Acknowledgements

First I would like to thank my advisor, Professor Demetri Psaltis, for all his effort in making this thesis possible. It has been my privilege to be a member of his highly motivated and talented group. His superb guidance, support, and constant encouragement has established a very enjoyable working environment. I am deeply grateful for his kindness, understanding, and the large amount of time and capital he has spent on me. Without any of these elements, this thesis would not have been possible.

I owe many thanks to Dr. Steven Lin, who taught me a lot on GaAs processing skills and MESFET circuit design when I first joined this group. His initial work on GaAs neuron circuits led to the completion of this thesis. I would also like to thank Dr. Annette Grot, with whom I worked on the optical FET detectors (OPFET's) discussed in Chapter 4 and many other areas of this thesis. I appreciated her technical support.

Many thanks also go to Professor Rutledge and his student, Jung-chih Chiao, for allowing me to use their optical step camera to make photolithography masks. I would like to express my appreciation to Professor Scherer for providing the quantum well LED's described in Chapter 3 and the anti-reflection coating on modulators described in Chapter 5. A major part of this thesis would not have been possible without the help from Professor Fonstad and his students, Krishna Shenoy, Hao Wang, Royter Yakov, and Joseph Ahadian at MIT. Hao Wang and Krishna Shenoy did the MBE regrowth of multiple quantum well structure on MOSIS GaAs chips described in Chapter 5 and 6. My thanks also go to Professor Tai for providing some of the device processing needs.

Jean-Jacques Drolet was very helpful to me during my stay at Caltech. He installed many softwares on the workstations in our group, including MAGIC for integrated circuit layout design and HSPICE for circuit simulation. I would also like to

thank Ms. Yayun Liu for her support in maintaining the smooth operation of all the GaAs processing equipment. I appreciate their help.

I would like to thank all other friends and colleagues in the Optical Information Processing Group, past and present: Chuanyi Ji, Subrata Rakshit, Yong Qiao, Sidney Li, Kevin Curtis, Robert Denkewalter, David Marx, Geoffrey Burr, Xin An, Ernest Chuang, Michael Levene, Allen Pu, George Barbastathis, Greg Billock, Chuan Xie, Xu Wang, and Ali Adibi. I am also grateful to Lucinda Acosta, Linda Dozsa, and Helen Carrier for their administrative support.

Outside our group, I would like to acknowledge the help from my long-time friend, Kezhong Hu. I appreciated his valuable suggestions on the quantum well modulator discussed in Chapter 5. Together with Yibin Cao and Peng Lian, he helped me settle down at Caltech during my first year here.

My sincere gratitude goes to my grandmother, You-ying Guo, in my small hometown in southeastern Chinese countryside. Without her unconditional support and sacrifice, nothing else would have been possible. I would like to thank my parents, Shu-Jin and Jin-Ding Luo for their love and encouragement.

Finally, I would like to thank my wife, Xiaobin, for taking care of me during the last four years. Her sweet smile brings a lot of sunshine and happiness into my life. Without her patience and consideration, this thesis would not have been completed.

Abstract

This thesis focuses on the design and fabrication of GaAs monolithic optoelectronic integrated circuits (OEIC's) for use in optical neural networks. The basic circuit in a neuron array consists of GaAs MESFET (Metal-Semiconductor Field Effect Transistor) circuits and optical input/output (I/O) devices. By implementing the I/O process optically, we can greatly increase the neuron density for a 2-dimensional array and thus achieve highly parallel computation.

Because of the high loss involved in optical interconnections, high density neuron arrays require high gain photodetectors and high efficiency output devices. With responsivities up to 10^4A/W and structure compatibility with MESFET circuits, optical FET detectors (OPFET's) are an excellent choice as photodetectors. Several techniques have been investigated in order to fabricate high efficiency LED's (light-emitting-diodes) at low current levels. Low power consumption neurons based on OPFET's and GaAs/AlGaAs double-Zn-diffusion double-heterojunction LED's are fabricated using in-house facilities.

Industrial foundries provide the most convenient answer to the challenge of fabricating high density 2-D neuron arrays. Two approaches will be described. The first approach utilizes the FET-SEED (self-electrooptic effect device) process from AT&T Bell labs. It provides monolithically integrated circuits with optical I/O devices and depletion mode FET's. In the other approach, GaAs/AlGaAs multiple quantum well modulators are grown on MOSIS GaAs MESFET circuits by MBE regrowth. It is found that in both approaches, the FET's can be used as high gain photodetectors even though the mechanisms are different, thus making it possible to achieve low power consumption high density neuron arrays. Various kinds of complex optoelectronic circuits can be fabricated through these two approaches.

Contents

List of Tables

List of Figures

Chapter 1 Introduction

With the continuing development of the processing techniques for semiconductor integrated circuits, both the speed and integrated density of IC's have greatly increased. However, interconnection delays and packaging of input/output (I/O) pins gradually become the limiting factor of the performance of integrated processors [1, 2, 3]. Optoelectronic integrated circuits (OEIC's) provide one possible solution for these bottleneck problems [4]. With optical signals as inputs, an OEIC performs the required computation using electronic integrated circuits, and the result is presented as another optical signal which can be interconnected optically through free space or optical elements to other processing units. Because optical signals do not interfere with each other in the same way as electrical signals, we can achieve parallel computation and eliminate electrical interconnection delays and I/O problems. As a result, the performance of an integrated processor is greatly enhanced [5, 6].

1.1 OEIC's for Optical Neural Networks

Typically, OEIC's are arranged into two-dimensional arrays, and the third dimension is used for interconnections by using lenses or holograms [4]. The computation power of such a system depends on three factors: the computing speed of the system, the number of I/O elements, and the functionality, or smartness, of each I/O as represented by the number of IC elements per I/O [7]. In this sense, there are two approaches in which OEIC's can be applied to achieve enhanced performance as compared to electronic processors. For conventional processors, the electronic circuit has very complex functionality and performs most of the computation, i.e., the number of IC elements per I/O is very large. By using the I/O and interconnection optically, we can bypass interconnection delays and electrical I/O packaging problems. As a result, both the speed of the system and the number of I/O elements can be greatly

increased, so does its computation power. One example is the free-space switching network [8].

On the other hand, the number of I/O elements can be increased to get more computation power. Neural networks, which usually have many computation elements (neurons) while each element is a simple processing unit [9], is such an example. The computation power of such systems comes from its large number of interconnections and the resulting parallel processing. Such a system can be found in the human brain which consists of about 10^{11} neurons [10]. The implementation of neural networks has many applications in many areas, such as pattern recognition, where animals outdo current electronic computers. Unfortunately, the number of neurons can be limited in a purely electrical implementation [11] due to some bottleneck problems such as the requirement of massive interconnections among all the neurons. Furthermore, the dynamic modification of interconnections required by error-driving algorithm [12] is difficult to implement electrically. On the other hand, optics is well suited for this kind of system because the interconnections can be implemented via the third dimension by utilizing the parallelism of optical signals [13, 14].

A schematic of the optical neural network architecture is shown in Figure 1.1. Neurons are arranged as 2-dimensional arrays; each array corresponds to one neuron layer in the network. While there are always one input layer and one output layer, the number of hidden layers can be changed according to the complexity of the task. Because the interconnections between different neuron layers are implemented by optical signals, each pixel must be able to detect incoming optical inputs and produce optical outputs. Therefore, each neuron is generally comprised of three components: optical receivers, nonlinear electronic elements, and optical output transmitters. The most commonly used nonlinear function in neural networks is the threshold function [9]. Other neuron functions include radial-basis functions, "bump" functions as the derivative of threshold functions, etc. [15]. Input signals detected by the optical receivers are mapped to the optical output transmitters through nonlinear electronic

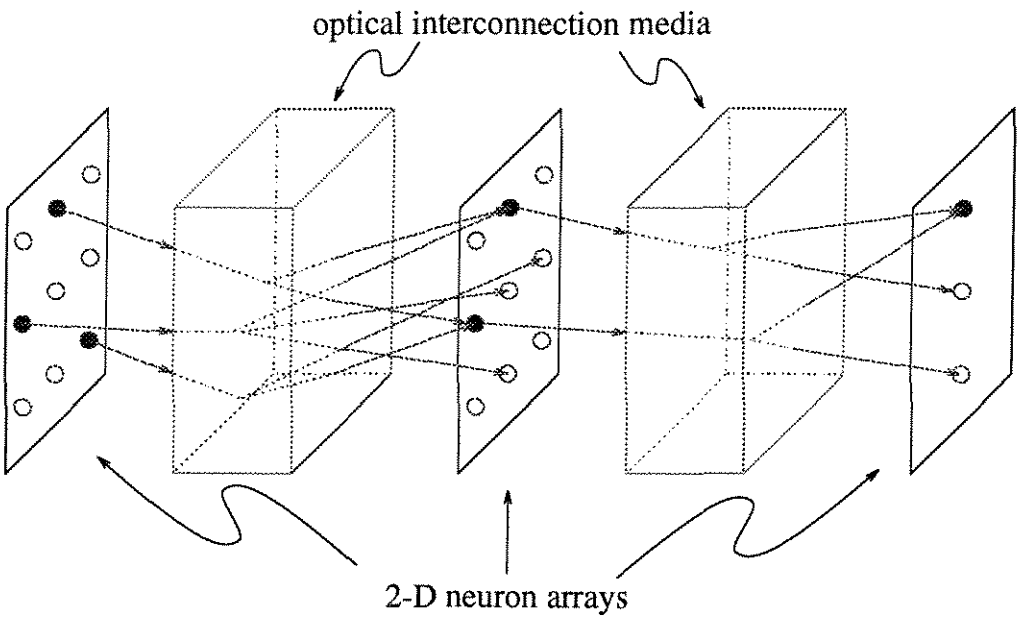


Figure 1.1: Schematic architecture of a multiple-layer optical neural network. Each circle represents a neuron while a filled circle means that the neuron is on the "ON" state.

elements as:

$$y_j = G\sigma(\sum_{i=1}^{N} w_{ij}x_i) \tag{1.1}$$

where y_j is the optical output from the jth neuron in current layer, G is the optoelectronic gain, $\sigma(x)$ is the non-linear neuron response function, N is the total number of input neurons in the previous layer, w_{ij} represents the interconnection weights between the two neuron layers, and x_i is the input from ith neuron in the previous layer. A neural net can be trained to solve a specific task by adjusting the interconnection weights using an algorithm, such as the back propagation learning algorithm [12].

Besides the easy implementation of optical I/O's, an optical neural network has two significant advantages compared to its electrical counterpart. First, in such a densely interconnected network, the weight w_{ij} represent a large database which can be easily implemented in the form of holographic memory [16]. Second, dynamic holograms recorded in photorefractive crystals have allowed the interconnection weights to be adjusted in real-time [17, 18, 19]. Therefore, the training of an optical neural

network can be done relatively easily by updating w_{ij} according to the comparison result between the output of the neural net and the desired response.

1.2 Fabrication Methods for OEIC Neuron Arrays

The fabrication of OEIC neuron arrays can be done either monolithically or in a hybrid process. Several hybrid integration approaches have been investigated, including flip-chip bonding [20, 21, 22], liquid crystal spatial light modulators on silicon CMOS circuits [23, 24], epitaxial lift-off [25, 26], etc. All these techniques have the advantage of a relatively mature Si VLSI integrated circuit technology to provide the necessary nonlinear functionality. Unfortunately, because silicon is an indirect bandgap material, it is very difficult to fabricate high efficiency optical output devices on silicon substrates even though there are some progress in this area [27, 28]. Furthermore, due to the build-in lattice mismatch between Si and direct bandgap III-V materials, especially GaAs [29], direct epitaxial growth of III-V materials on Si [30, 31] is normally subjected to strain, which may cause defects and thus degrade the device performance. Hence, optical output elements are typically built separately and then physically transported to silicon circuits. For example, the flip-chip bonding technique uses indium solder bumps to electrically and physically connect two different substrates, one with optical output devices on it, and the other, typically silicon, with electronic circuits and photodetectors.

The advantage of hybrid integration approaches is that individual components of the OEIC can be optimized before they are finally integrated. However, there are some disadvantages. First, the complicated mechanical assembly process may result in reliability and stability problems [32]. Second, such techniques are typically limited to single chip processing and therefore difficult to extend to large wafer fabrication due to assembly and packaging problems.

Monolithically integrated optoelectronic circuits [33] provide an alternative solution. Using the well established GaAs integrated circuit fabrication technology, we can build large 2-dimensional OEIC arrays that consist of monolithically integrated

photodetectors, electronic elements, and light sources on a single wafer. Compared to Si circuits, GaAs can have lower power consumption and/or faster speed. Even though there exist compromises in the choice of device structures, GaAs monolithically integrated photonic circuit offers overall advantage over many of the hybrid approaches. This thesis addresses the design and fabrication issue of such high density monolithic GaAs optoelectronic neuron arrays, including the optimization of each individual components and the integration process of the whole array.

1.3 Limiting Factors on Circuit Design

The computation power of an optical neural network depends on the number of neurons on each layer and the number of interconnections an optical system can support. Given a certain optical system, the size of the neuron chip is limited due to the finite dimension of optical lenses and photorefractive hologram-recording materials. Therefore, the higher the neuron density is on an array, the more neurons we can fit into the optical system, which means a better performance for the neural net.

The requirement of high density integration and multiple layer cascade operation imposes severe limit on individual components and the design of neuron circuits. First, the optical I/O devices and electronic IC elements occupy real estate areas, which places a geometric limit on the maximum neuron density that can be achieved. The required area for signal and voltage transferring metal lines must be taken into account as well, even though some sort of vertical integration is possible due to multiple layers of interconnect metallization. For a density higher than 10^4neurons/cm^2, each pixel only has an area less than $100 \times 100 \mu$m^2. As a result, we must use the simplest circuit that performs the required neuron functionality in order to reduce the pixel size and thus increase the neuron density. Therefore, a neuron circuit can only employ a minimum number of electronic amplification stages. This will save not only the occupied area of each pixel but also the power consumption of each neuron, which we will discuss next.

Besides the foregoing geometric limit, a semiconductor wafer can only dissipate a

certain amount of generated heat. This dissipation limit can be improved by advanced packaging techniques [34]. For GaAs wafers, typically the dissipation limit is in the order of $1 \sim 10\text{W/cm}^2$. Higher power dissipation will raise the chip temperature and result in abnormal operation of the circuits, or even cause permanent damage on the overheated chip. Assuming the maximum allowed power dissipation per unit area on a wafer is P_{max} and each neuron circuit has a power consumption of P_{neuron}, then the maximum neuron density is:

$$N = \frac{P_{max}}{P_{neuron}} . \tag{1.2}$$

As a result, in order to achieve the goal of high density neuron arrays, we must reduce not only the required real estate area for IC elements but also the power consumption of each single neuron.

Unfortunately, there are limits on the power consumption of each neuron. Given a certain type of neuron circuit, there is a minimum switching current, I_{switch}, that is required to turn on and off the neuron within a certain time, τ, due to the parasitic capacitance. Such current will result in some power consumption. In addition, the minimum optical input power, P_{in}, required to switch the circuit will be:

$$P_{in} = \frac{I_{switch}}{\eta_{DEC}} \tag{1.3}$$

if we assume that the photodetector has a responsivity of η_{DEC}, which corresponds to an optical-to-electronic conversion gain of:

$$\eta = \eta_{DEC} \times \frac{h\nu}{q} , \tag{1.4}$$

where q is the charge of an electron, h is the Plank's constant, and ν is the optical frequency of incident photons. Remember that in a cascade system like the one shown in Figure 1.1, this input signal actually comes from optical outputs of neurons on previous layer, interconnected through the optical system. Hence, a minimum

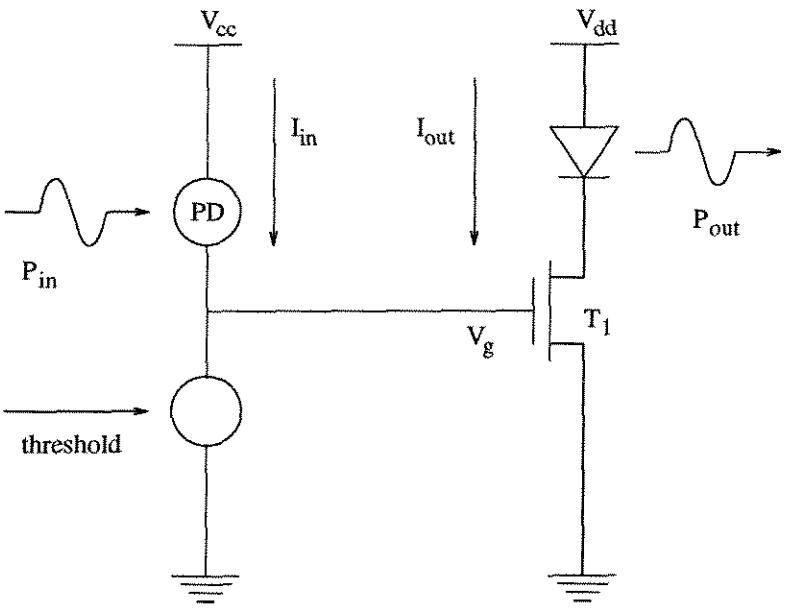


Figure 1.2: A simple threshold neuron circuit with light-emitting diode (LED) output. PD is a photodetector and the threshold can be set electrically by a MESFET or optically by another photodetector.

optical output power of:

$$P_{out} = P_{in} \times G = \frac{I_{swtich}}{\eta_{DEC}} \times G \tag{1.5}$$

is needed, where G is the required optoelectronic gain in order to compensate the high loss involved in the optical interconnection system. As we will see in the next section, the consumed power in order to generate P_{out} places a significant limit on the maximum neuron density that can be achieved.

1.4 Limitation on Neuron Density

In order to estimate the maximum neuron density, we look at the simple threshold neuron circuit shown in Figure 1.2. This circuit consists of two branches: an input branch is used to compare the optical input signal, P_{in}, detected by a photodetector, PD, with a threshold value, P_T, which can be set either electrically or optically; the

compared result is fed into the output branch as the gate voltage, V_g, of the driving MESFET, T_1. When $P_{in} > P_T$, an excess current is generated by the photodetector, which then charges up the gate capacitor of T_1, thus increasing V_g up to near V_{cc}. As a result, transistor T_1 is turned on and we have a high current to drive the on-chip LED source. The neuron is on the "ON" state because its optical output is high. The reverse case occurs when $P_{in} < P_T$ and the neuron is on the "OFF" state. Now we want to figure out what is the maximum neuron density for such a simple design.

First, there is a geometry limit due to lithography minimum features and the real estate requirement. Assuming the following area for the four devices: 5μm$\times 5\mu$m for the LED, same area for the photodetector PD and maybe the threshold unit if it is set optically (electrical threshold unit may require less space), T_1 is about 3μm$\times 5\mu$m, and some area for the metal connection lines. The total area is about $10 \times 10\mu\text{m}^2 = 10^{-6}\text{cm}^2$. As a result, the geometry limit on the maximum neuron density is about 10^6neurons/cm^2.

Assuming the gate capacitance of T_1 is C_g, the minimum current required to switch the neuron within time τ is:

$$I_{switch} = \frac{\Delta Q}{\tau} = \frac{C_g \Delta V_g}{\tau} \simeq \frac{C_g V_{cc}}{\tau} \ . \tag{1.6}$$

This means the photodetector must generate a photocurrent of:

$$I_{in} = I_{switch} + I_{threshold} \ . \tag{1.7}$$

The minimum power consumption case happens when $I_{threshold} = 0$, i.e., the threshold is set at zero level. In this case, from Eq. 1.3 and Eq. 1.5, the minimum output is:

$$P_{out} = \frac{I_{switch}}{\eta_{DEC}} \times G = \frac{G V_{cc} C_g}{\tau \eta_{DEC}} \ . \tag{1.8}$$

So, the driving MESFET T_1 must allow an output current of:

$$I_{out} = \frac{P_{out}}{\eta_{LED}} = \frac{G V_{cc} C_g}{\tau \eta_{DEC} \eta_{LED}} \tag{1.9}$$

where η_{LED} is the external current efficiency of the LED and we assume a linear relationship between the optical output of the LED and its driving current. The effect of carrier lifetime on the output power of LED is also neglected, which is true when the lifetime is much shorter than τ [35]. Putting everything together, the power consumption of this neuron is:

$$\begin{aligned} P_{neuron} &= \frac{1}{2}V_{cc}I_{in} + V_{dd}I_{out} \\ &= \left(\frac{V_{cc}}{2} + \frac{V_{dd}G}{\eta_{DEC}\eta_{LED}}\right)\frac{V_{cc}C_g}{\tau} \end{aligned} \tag{1.10}$$

where the factor of $1/2$ comes from the charge of capacitor C_g. Combining Eq. 1.10 and Eq. 1.2, we have the maximum neuron density as:

$$N = \frac{P_{max}\tau}{V_{cc}C_g} \cdot \frac{\eta_{DEC}\eta_{LED}}{\frac{1}{2}V_{cc}\eta_{DEC}\eta_{LED} + V_{dd}G} . \tag{1.11}$$

From Eq. 1.11, it is clear that we want to increase P_{max} and reduce V_{cc} and V_{dd} in order to increase the maximum neuron density. Unfortunately, these parameters are more or less fixed. P_{max} is determined by the technique used to package the GaAs neuron chip. V_{cc} is the switching voltage of MESFET T_1, so it is unlikely that V_{cc} can be less than 0.5V. V_{dd} is the power supply for the LED, so it must be larger than the sum of the LED turn-on voltage and the saturation voltage of MESFET T_1. However, we can decrease the gate capacitance, C_g, to get better neuron density. From semiconductor physics [36],

$$C_g = \frac{\epsilon A}{d} = \frac{\epsilon W \times L}{d} \tag{1.12}$$

where ϵ is the dielectric permittivity of GaAs, W and L is the gate width and length of T_1, respectively, and d is the depth of the depletion region underneath the MESFET gate, which is determined by the doping level of the MESFET channel and the applied voltage. By changing the parameters in our circuit design to reduce this parasitic capacitance, we can increase the maximum neuron density.

Eq. 1.11 also imposes a requirement on the optical input and output devices. First, for the LED light source, a higher current efficiency, η_{LED}, will result in higher density, N. It is very important to notice that these LED's are operating at low current levels, typically less than 1mA, when the neuron density is increased to more than 10^3neurons/cm^2. This problem will be discussed in Chapter 3. Second, it is essential to use high gain photodetectors in order to get high neuron density. If η_{DEC} is low, in order to generate the same amount of switching current, P_{in} must be higher, so a much higher P_{out} is needed. The end result is an increased power consumption and a lower neuron density. The discussion on high gain photodetectors will be presented in Chapter 4.

To get a feeling for the magnitude of N, let's assume the following parameters: $P_{max} = 5\text{W/cm}^2, \tau = 10\text{nsec}$, $V_{cc} = 0.5\text{V}$, $V_{dd} = 2\text{V}$, $\eta_{LED} = 0.006\text{W/A}$, $\eta_{DEC} = 200\text{A/W}$, $G = 100$, and $C_g = 12\text{fF}$, which represents a gate area of $1.0\mu\text{m}\times 10\mu\text{m}$ with about $1000\mathring{A}$ depth of depletion region. From Eq. 1.11, the maximum neuron density that can be achieved is 5.0×10^4neurons/cm^2.

Eq. 1.11 gives us the maximum neuron density that can be achieved by using LED as optical output devices in the threshold circuit shown in Figure 1.2. From this equation, we have $N \propto \tau$, which results from the requirement that these circuits are cascaded. The longer the switching time can be, the smaller the switching current is needed for charging up the same capacitor. Therefore, P_{in}, and thus P_{out}, is allowed to be weaker and a smaller driving current is required for the LED. A lower power consumption and higher neuron density is the final result. So, we can trade the speed of the system for the maximum neuron density, or vice versa. However, Eq. 1.11 is only partially true: the switching current can not be reduced indefinitely because of the existence of gate-source leakage current of MESFET T_1. The exact modeling of MESFET gate-source leakage current is very complex [37]. From the metal-semiconductor junction theory [36], a simplified result is given as:

$$I_{gs} = W \times L \cdot J_{sat}(e^{qV/nkT} - 1) \tag{1.13}$$

where J_{sat} is the reverse-biased saturation current density depending on the Schottky barrier height, n is the ideality factor and ranges between 1 and 2, k is the Boltzmann constant, and T is the environment temperature. When the neuron is on the "OFF" state, V_g is close to 0, so the leakage current is very small. When the neuron is turned on, however, V_g can rise up to near V_{cc} and the leakage current is much higher than I_{sat}. In order to support this leakage current and maintain the "ON" state, a minimum P_{in} is required. As a result,

$$P_{out,min} = \frac{I_{gs}(V_g = V_{cc})}{\eta_{DEC}} \times G \ . \tag{1.14}$$

The minimum power consumption of the neuron is:

$$\begin{aligned} P_{neuron,min} &= V_{cc} \cdot I_{gs}(V_g = V_{cc}) + V_{dd} \cdot \frac{P_{out,min}}{\eta_{LED}} \\ &= \left(V_{cc} + \frac{V_{dd} G}{\eta_{DEC} \eta_{LED}} \right) \cdot I_{gs}(V_g = V_{cc}) \ . \end{aligned} \tag{1.15}$$

So, no matter how large τ can be, the maximum neuron density that can be achieved is always less than:

$$N_{max} = \frac{P_{max}}{I_{gs}(V_g = V_{cc})} \cdot \frac{\eta_{DEC}\eta_{LED}}{V_{cc}\eta_{DEC}\eta_{LED} + V_{dd}G} \ . \tag{1.16}$$

Assuming the same parameters as we used in last section and $n = 1.12, J_{sat} = 3 \times 10^{-7}$A/cm^2 [38], which corresponds to a leakage current of 0.93μA at $V_g = 0.5$V for 1.0μm$\times$10μm gate, the maximum neuron density is 3.2×10^4neurons/cm^2.

Combining Eq. 1.11 and Eq. 1.16, we have the final result for the maximum neuron density:

$$N = \begin{cases} \frac{P_{max}\tau}{V_{cc}C_g} \cdot \frac{\eta_{DEC}\eta_{LED}}{\frac{1}{2}V_{cc}\eta_{DEC}\eta_{LED} + V_{dd}G} & \text{when } \tau < \tau_0. \\ \frac{P_{max}}{I_{gs}(V_g = V_{cc})} \cdot \frac{\eta_{DEC}\eta_{LED}}{V_{cc}\eta_{DEC}\eta_{LED} + V_{dd}G} & \text{when } \tau > \tau_0. \end{cases} \tag{1.17}$$

where τ_0 can be solved by substituting Eq. 1.16 into Eq. 1.11:

$$\tau_0 \approx \frac{V_{cc}C_g}{I_{gs}(V_g = V_{cc})} . \tag{1.18}$$

We calculate τ_0 to be 6.4 nanoseconds using the same parameters as above.

1.5 Choice of Output Light Sources

In the previous discussion, we calculated the maximum neuron density when using light-emitting diodes to provide optical outputs. The advantage of LED's is that it is an on-chip light source so the optical setup is simple compared to OEIC chips using optical modulators [39, 40]. In addition, it is relatively easy to integrate LED's into an OEIC chip. All we need to do is to define top and bottom ohmic contacts for the diode. However, there are also disadvantages for light-emitting diodes as optical outputs. First, even though the internal quantum efficiency of a double-heterojunction GaAs/AlGaAs LED can be very close to 100% [41], the fact that LED's emit light in all direction makes it very difficult to collect the optical output. Therefore, LED's typically suffer from low external efficiencies. As a result, higher driving current is needed, which results in high power consumption. Second, since the output of the LED comes from spontaneous emission, it has a wide-band spectrum and is therefore temporally incoherent. This incoherence makes the recording of holographic interconnections impossible [18] and limits the applications of LED-based neuron arrays to non-adaptive systems.

Laser diodes are the other candidate for on-chip light sources. Its bandwidth is much narrower compared to LED's and therefore has high degree of coherence. The light is emitted in one direction and therefore it has very high differential efficiency. Even though laser diodes require more complicated material preparation and processing technique, the emergence of vertical cavity surface-emitting lasers (VCSEL's) [42] has greatly simplified the fabrication of 2-dimensional laser diode arrays.

For laser diodes, the same circuits as those shown in Figure 1.2 can be used as

a threshold neuron. Hence, we can use the same equations 1.2–1.8 to calculate the power consumption and neuron density of such neurons except that we have to take into account the threshold current of the laser diode in Eq. 1.9, which then becomes:

$$I_{out} = \frac{P_{out}}{\eta_{LD}} + I_{th} = \frac{GV_{cc}C_g}{\tau\eta_{DEC}\eta_{LD}} + I_{th} \tag{1.19}$$

where η_{LD} is the differential current efficiency of the laser diode. So, the power consumption and maximum neuron density of circuits with laser diodes as optical outputs are:

$$P_{neuron} = \left(\frac{V_{cc}}{2} + \frac{V_{dd}G}{\eta_{DEC}\eta_{LD}}\right)\frac{V_{cc}C_g}{\tau} + V_{dd}I_{th} \tag{1.20}$$

$$N = \frac{P_{max}\tau}{V_{cc}C_g} \cdot \frac{1}{\frac{V_{cc}}{2} + \frac{V_{dd}G}{\eta_{DEC}\eta_{LD}} + \frac{V_{dd}I_{th}\tau}{V_{cc}C_g}} . \tag{1.21}$$

Compared with Eq. 1.11, we can see that even though we have a much better η_{LD} compared to η_{LED}, the additional term due to I_{th} will likely reduce the total neuron density.

There is currently active research on how to reduce the threshold current of laser diodes and also on the so-called "indistinct threshold" lasers [43]. However, most of the low-threshold laser diodes reported today still have threshold currents in the order of hundred microamps. To the best of my knowledge, for cw operation of laser diodes at room temperature, the lowest reported threshold current was 70μA [44] for a VCSEL with a native oxide confinement structure. However, the efficiency of such a laser is very low. The efficiency achieved in Ref. [44] is only 0.02W/A for devices with $I_{th} = 70\mu$A. It increases to 0.08W/A for $I_{th} = 0.75$mA devices. Recently, threshold currents as low as 9μA has been reported [45].

To calculate the maximum neuron density with the laser diode as optical outputs, we use two different sets of parameters for the laser diode: $I_{th1} = 0.6$mA, $\eta_{LD1} = 0.5$W/A, and $I_{th2} = 70\mu$A, $\eta_{LD2} = 0.1$W/A. Other parameters are the same as those in the LED discussion. From Eq. 1.21, we can calculate $N_1 = 4.2 \times 10^3$neurons/cm^2

and $N_2 = 3.4 \times 10^4$neurons/cm^2 from the two threshold currents.

The same leakage-current-limit discussed in Eq. 1.13 to 1.16 is still valid since we have the same circuit. With the addition of I_{th}, Eq. 1.15 and 1.16 become:

$$\begin{aligned} P_{neuron,min} &= V_{cc} \cdot I_{gs}(V_g = V_{cc}) + V_{dd} \cdot \left(\frac{P_{out,min}}{\eta_{LD}} + I_{th} \right) \\ &= \left(V_{cc} + \frac{V_{dd} G}{\eta_{DEC} \eta_{LD}} + V_{dd} \cdot \frac{I_{th}}{I_{gs}(V_g = V_{cc})} \right) \cdot I_{gs}(V_g = V_{cc}) \quad (1.22) \\ N_{max} &= \frac{P_{max}}{I_{gs}(V_g = V_{cc})} \cdot \frac{1}{V_{cc} + \frac{V_{dd} G}{\eta_{DEC} \eta_{LD}} + V_{dd} \cdot \frac{I_{th}}{I_{gs}(V_g = V_{cc})}} . \quad (1.23) \end{aligned}$$

For the same parameters, Eq. 1.23 gives $N_{max1} = 4.2 \times 10^3$neurons/cm^2 and $N_{max2} = 3.3 \times 10^4$neurons/cm^2. The reason that Eq. 1.23 and Eq. 1.21 give almost the same result is because in both cases, the neuron circuit density is basically limited by the threshold current of the laser diode rather than anything else. It is not until very high speed operation that the result given by Eq. 1.21 will be significantly different from the result of Eq. 1.23. The point when we begin to get noticeable difference is at:

$$\tau_0 = \frac{10 V_{cc} C_g}{V_{dd} I_{th}} \cdot \left(V_{cc} + \frac{V_{dd} G}{\eta_{DEC} \eta_{LD}} \right) . \quad (1.24)$$

We calculated τ_0 to be 0.125nsec for $I_{th1} = 0.6$mA and 1.07nsec for $I_{th2} = 70\mu$A.

Equations 1.17, 1.21, and 1.23 are plotted in Figure 1.3 for both the high threshold-current and the low threshold-current laser diodes. We can see that LED's are preferred over laser diodes when the laser diode has a high threshold current. However, when the threshold current of the laser diode is reduced, the performance of laser-diode-based neuron arrays becomes comparable to that of LED-based arrays. Further reduction in the threshold current will make the laser diode a preferred choice. However, one major problem for OEIC arrays with laser diodes is the difficulty in monolithic integration because of the high requirement on the material quality and the fabrication technique. For very high speed cases, e.g., when $\tau < 0.83$nsec, laser diodes perform better because they have much higher current efficiencies. For such high speeds, however, our result is not applicable because we did not consider the

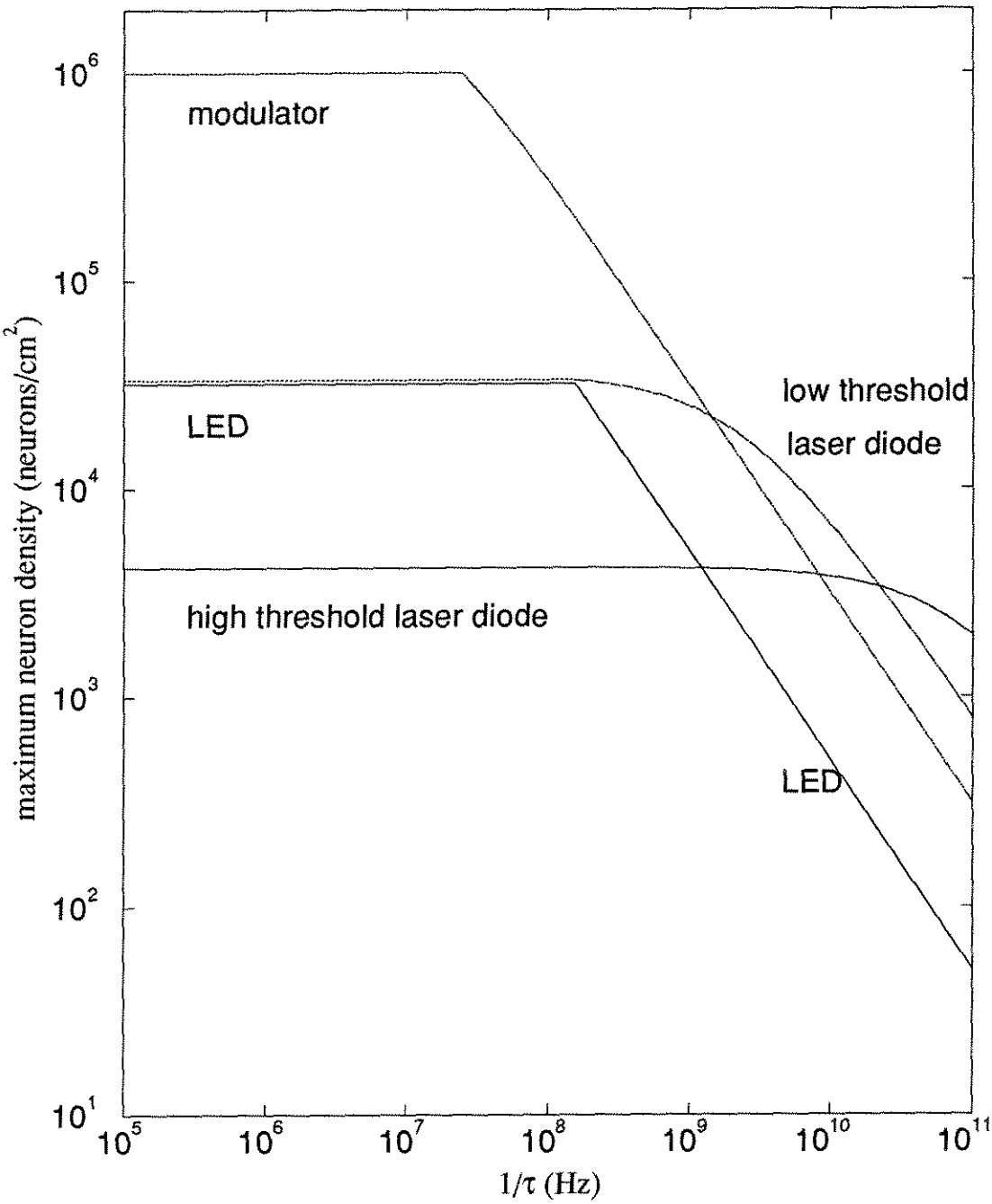


Figure 1.3: The comparison between LEDs, laser diodes and optical modulators. The limit at low frequency comes from the MESFET leakage current. The estimation on very high frequency is not accurate because the cutoff frequency of optical output devices is not taken into consideration.

effect of the finite bandwidth of the LED or laser diode.

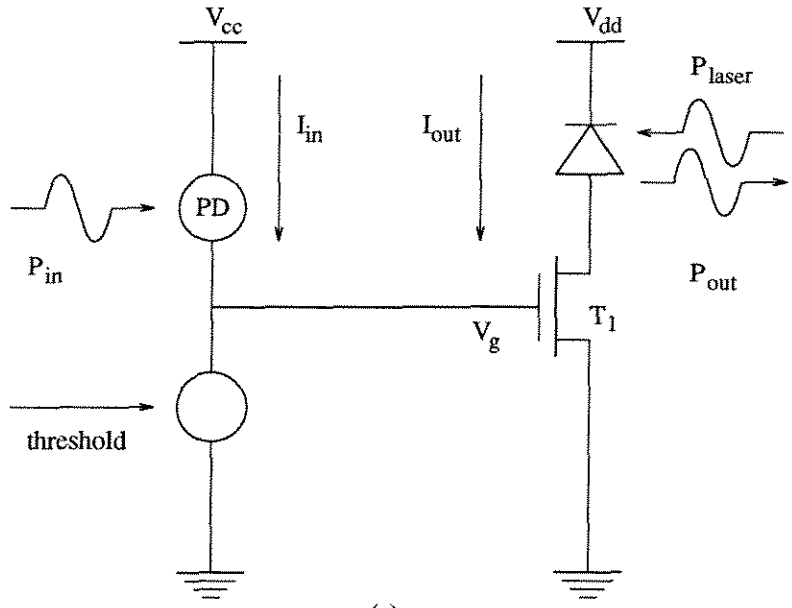
Besides on-chip light sources, the other option for optical outputs is optical modulators [39, 40]. Because the light source is off-chip in this case, we can easily increase the optoelectronic gain by increasing the power level of the laser source, which does not affect the power consumption of the neuron chip. Therefore, we can build arrays with higher density due to the lower power consumption of modulator-based neurons. By using a single laser as the off-chip light source, we can also achieve both temporal and spatial coherence, which is required for the effective implementation of the holographic interconnections in adaptive neural nets [46]. However, the alignment of the optical system is more complicated for modulators due to the addition of off-chip light sources and is also limited by the ability to generate large incident beam arrays. The other disadvantage of modulators comes from its relatively low contrast ratio and the required high driving voltages. For example, a GaAs/AlGaAs multiple quantum well (MQW) modulator, which we will discuss in more detail in Chapter 5, typically has a contrast ratio of 4 : 1 at a bias voltage of 10V [47]. By depositing a low-reflectivity mirror on the modulators, we can get an asymmetric Fabry-Perot cavity modulator [48], which has much higher contrast ratio and requires less driving voltage. However, the fabrication process and wavelength alignment is much more critical in order to achieve the optimum performance of an asymmetric Fabry-Perot cavity modulator.

There are two types of driving circuits for the MQW modulators as shown in Figure 1.4. Figure 1.4(a) is similar to Figure 1.2 except that the diode at the output part of the circuit is now reverse-biased and V_{dd} is higher. For the output branch, when T_1 is turned on, the output current is limited by the photocurrent generated by the reverse-biased modulator, which acts electrically as a photodiode. Therefore, the voltage drop on the modulator is very close to V_{dd} and we have:

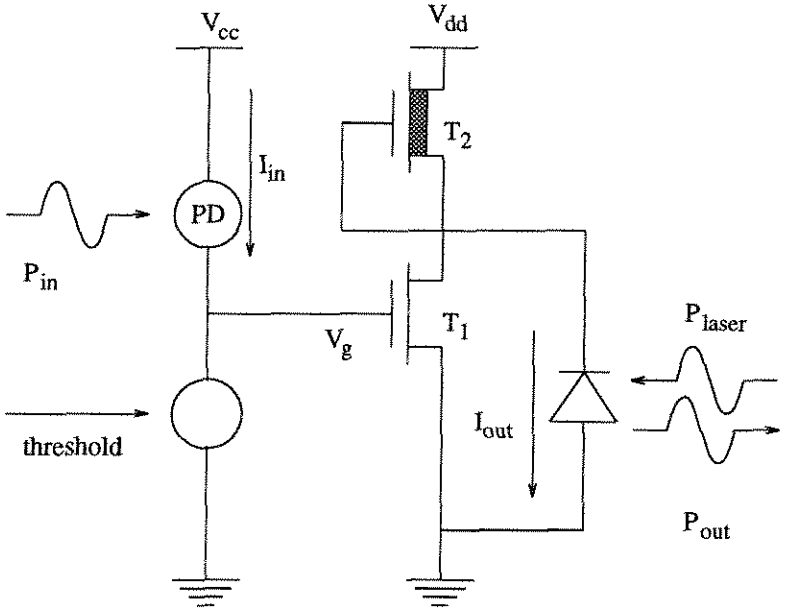
$$I_{out} = \eta_{PD}(V = V_{dd})P_{laser} \tag{1.25}$$

$$P_{out} = (1 - \eta_{PD}(V = V_{dd}))P_{laser} \times \eta_{MOD} \tag{1.26}$$

where P_{laser} is the incident power on the modulator, η_{PD} is the detector responsivity of



(a)



(b)

Figure 1.4: Two threshold neuron circuits using MQW modulators as optical outputs. T_1 is an enhancement-mode MESFET and T_2 is a depletion-mode MESFET.

the modulator/photodiode, which is a function of the bias voltage, and the additional insertion loss is included in the factor of η_{MOD}. When T_1 is turned off, the output current is limited by the drain-source current of T_1 at low gate-source bias. The voltage drop on the modulator in this case is close to 0. Depending on the operating wavelength, given a modulator, $\eta_{PD}(V = V_{dd})$ can be either greater or less than $\eta_{PD}(V = 0)$.

For simplicity, let's assume that $\eta_{PD}(V = V_{dd}) < \eta_{PD}(V = 0)$, which results in a high output (low absorption) when $P_{in} > P_T$. For the reverse case, we can simply switch PD and the threshold unit and still get the same response. The power consumption of such neurons is simply the required power to charge up C_g and the capacitance of a reverse-biased modulator, C_{mod}, plus the power consumption due to I_{out}:

$$P_{neuron} = \frac{C_g V_{cc}^2}{2\tau} + \frac{C_{mod} V_{dd}^2}{2\tau} + V_{dd} \cdot I_{out} \ . \tag{1.27}$$

In order to determine I_{out}, we use the same optical feedback loop as represented by Eq. 1.8. Substituting Eq. 1.8 into Eq. 1.26, we have:

$$P_{laser} = \frac{P_{out}}{\eta_{MOD}(1 - \eta_{PD}(V = V_{dd})} = \frac{G V_{cc} C_g}{\eta_{MOD}(1 - \eta_{PD}(V = V_{dd}))\eta_{DEC}\tau} \tag{1.28}$$

$$\begin{aligned} P_{neuron} &= \frac{C_g V_{cc}^2}{2\tau} + \frac{C_{mod} V_{dd}^2}{2\tau} + V_{dd} \cdot \eta_{PD}(V = V_{dd}) \cdot P_{laser} \\ &= \frac{C_g V_{cc}^2}{2\tau} + \frac{C_{mod} V_{dd}^2}{2\tau} + \frac{\eta_{PD}(V = V_{dd})}{\eta_{MOD}(1 - \eta_{PD}(V = V_{dd}))} \cdot \frac{G V_{cc} V_{dd} C_g}{\eta_{DEC}\tau} \ . \end{aligned} \tag{1.29}$$

The maximum neuron density that can be achieved is:

$$N = \frac{2P_{max}\tau}{C_g V_{cc}^2 + C_{mod} V_{dd}^2 + \frac{\eta_{PD}(V = V_{dd})}{\eta_{MOD}(1 - \eta_{PD}(V = V_{dd}))} \cdot \frac{2G V_{cc} V_{dd} C_g}{\eta_{DEC}}} \ . \tag{1.30}$$

Using the following parameters: $P_{max} = 5\text{W/cm}^2, \tau = 10\text{nsec}$, $C_g = 12\text{fF}$, $V_{cc} = 0.5\text{V}$, $G = 100, \eta_{DEC} = 200\text{A/W}$, $C_{mod} = 3.0\text{fF}$, which represents a $5 \times 5\mu\text{m}^2$ modulator with $1\mu\text{m}$ depletion region, $V_{dd} = 10\text{V}$, $\eta_{PD} = 0.2\text{A/W}$, $\eta_{MOD} = 0.9$, we have $N = 3.1 \times 10^5 \text{neurons/cm}^2$.

Again at low speed operation, the maximum neuron density has a leakage-current limit as discussed in the LED case. However, since modulators can increase P_{out} by simply using brighter light source without affecting the required energy to switch C_{mod}, this effect is not that significant. At low speeds, Eqs. 1.28 to 1.30 become:

$$\begin{aligned} P_{laser,min} &= \frac{P_{out}}{\eta_{MOD}(1-\eta_{PD}(V=V_{dd})} \\ &= \frac{GI_{gs}(V_g=V_{cc})}{\eta_{MOD}(1-\eta_{PD}(V=V_{dd}))\eta_{DEC}} \end{aligned} \tag{1.31}$$

$$\begin{aligned} P_{neuron,min} &= \frac{C_g V_{cc}^2}{2\tau} + \frac{C_{mod}V_{dd}^2}{2\tau} \\ &\quad + \frac{\eta_{PD}(V=V_{dd})}{\eta_{MOD}(1-\eta_{PD}(V=V_{dd}))} \cdot \frac{GV_{dd}I_{ds}(V_g=V_{cc})}{\eta_{DEC}} \end{aligned} \tag{1.32}$$

$$N_{max} = \frac{2P_{max}\tau}{C_g V_{cc}^2 + C_{mod}V_{dd}^2 + \frac{\eta_{PD}(V=V_{dd})}{\eta_{MOD}(1-\eta_{PD}(V=V_{dd}))} \cdot \frac{2\tau G V_{dd} I_{gs}(V_g=V_{cc})}{\eta_{DEC}}} \tag{1.33}$$

When $\tau > \tau_0$, Eq. 1.33 must be used instead of Eq. 1.30. However, remember there is a geometric limit due to the lithography minimum feature. From the discussion in last section, we have:

$$N_{max} \leq 10^6 \text{neurons/cm}^2. \tag{1.34}$$

For the neuron circuit shown in Figure 1.4(b), the analysis is basically the same, except that the term $V_{dd} \cdot I_{out}$ may be different now. When T_1 is off, the current going through transistor T_2 must be limited by I_{out} as given by Eq. 1.25 in order to get high voltage drop on the modulator. So, the saturation current of T_2 must be higher than I_{out}. When T_1 is on, the current will be limited by the saturation current of T_2 in order to get zero voltage drop across the modulator. As a result, such neuron consumes more power than the one shown in Figure 1.4(a). However, Eq. 1.30 and 1.33 are still approximately correct because most of the power consumption comes from the switching energy of C_{mod}.

The result of Eqs. 1.30, 1.33 and 1.34 is also plotted in Figure 1.3. We can see that for all frequency ranges, MQW modulators perform better than LED's in

terms of maximum neuron density. However, because of the easy setup of the LED-based system and the easy fabrication of LED devices, LED's are still important candidates for optical outputs when we design the optical neuron circuits. In the long term, modulators will be our first choice, but in the near term, LED's are a useful alternative.

1.6 Other Consideration for Neuron Arrays

In the previous sections we discussed how to maximize the neuron density based on the consideration of power consumption. We found that modulators and LED's are two candidates as optical outputs. We also found that a high gain photodetector is necessary in order to achieve high density neuron arrays, especially for LED-based neurons.

Besides power consumption, there are other important factors that affect the maximum neuron density. The first issue is uniformity. In the previous discussion, we have assumed that all the neurons are identical and noise free. That's of course not the case in practice. No matter how well-controlled the fabrication process is, there are always variations (or non-uniformity) across a chip. Noise will also be present during the operation of the neuron arrays. When the non-uniformity increases, in order to maintain proper operation of the neural net, the minimum optical input power given by Eq. 1.3 will have to be increased as well because the value of I_{switch} now has a certain distribution among different neurons on an array. Therefore, more output power is needed from the optical feedback consideration, which decreases the neuron density by increasing the power consumption level of all neurons.

By modeling the noise (and non-uniformity) as a Gaussian random variable with zero mean and standard deviation σ_n, the maximum neuron density over an area is given by [49]:

$$N/A = \left\{ \frac{\eta_{DEC}\eta_{LED}P_{max}}{GV_{dd}\sigma_n\sqrt{A}} [pq(1-pq)]^{1/2} \tan(\pi P_e) \right\}^{2/3} \tag{1.35}$$

where p is the probability of neurons on the previous layer in the "on" state, q is the probability of w_{ij} being 1, assuming w_{ij} is either 1 or 0, and P_e is the maximum allowable error rate by the neuron net, which is specified by the learning algorithm. We can see that it is very important to minimize the non-uniformity variance σ_n. However, the control of semiconductor IC fabrication processes is very time consuming and requires a large amount of capital investment which only industrial fabs can support. Such considerations lead to the utilization of industrial foundries discussed in the next chapter of this thesis.

The other important concern is the compatibility of the structure of each components for the OEIC implementation. This is the compromise we have to make due to monolithic integration. If different components, for example, the optical receivers and the electronic IC elements, can not share the same material, then a vertical stack of different structure is required. This will increase the total etching depth in fabrication, which is one of the main sources of non-uniformity. The photolithography requirement is also more complicated because of the non-planar structure. In a world, the growth of the thickness of the total material stack will cause the quality control of the OEIC fabrication process to be more difficult, therefore increases the non-uniformity and reduces the maximum neuron density. In result, we must make our best effort to use compatible structures for individual components.

In summary, we discussed the maximum neuron density that can be achieved using a simple threshold circuit based on the power consumption argument in this chapter. The calculation was conducted for three different output sources: LED, laser diode, and MQW modulator. Based on the result, in order to increase the neuron density, modulators should be chosen as the output source in the ideal case. LED is also a good choice if its external efficiency can be raised at low current levels. In the future, low threshold-current laser diodes will also become a good choice with the development of integration processes and improvement over threshold non-uniformity. For photodetectors, it is concluded that high gain detectors are necessary for high density neuron arrays, especially for those based on LED. Another important conclusion is that non-uniformity in fabricating the neuron array will strongly affect

its maximum density, therefore such non-uniformity should be minimized.

1.7 Thesis Overview

This thesis is arranged in three parts. The first part, including Chapter 1 and Chapter 2, addresses basic questions of OEIC neuron arrays. Chapter 1 discusses the maximum neuron density that can be achieved using simple threshold neuron circuits. We also look at the requirement for individual components of a neuron array due to the goal of maximizing the computation power of neural nets. Chapter 2 describes the three approaches we have investigated in fabricating OEIC neuron circuits, including in-house fabrication and two industrial foundries: the AT&T FET-SEED process and MBE regrowth on MOSIS GaAs MESFET circuits. Detailed description of the fabrication steps and the performance of electronic elements, especially MESFET's, are presented.

Next, the optimization of individual optoelectronic components is discussed in Chapter 3, 4, and 5. Chapter 3 concentrates on the optimization of GaAs/AlGaAs double-heterostructure light-emitting diodes. Surface leakage current and the difficulty of collecting all the optical outputs are found to be the main problems responsible for the low external efficiency of LED's, especially at low driving current. Several fabrication techniques are investigated, including double Zn-diffusion, ion implantation, and micro cavity quantum well LED's. Chapter 4 describes the performance and optimization of a special kind of photodetector: the optical FET detector (OPFET). Its responsivity is in the order of 1000A/W at low optical input powers and it can be fabricated using the same steps used for a MESFET. These two properties make it an excellent choice as photodetectors for high density neuron arrays. The mechanisms for different OPFET's fabricated through the three different integrating approaches are also discussed. Chapter 5 discusses GaAs/AlGaAs multiple quantum well (MQW) modulators. The so-called quantum-confined Stark effect is explained and a simplified model is given for estimating the changes in optical absorption coefficient when an electric field is applied perpendicular to the MQW. This calculation serves as a guid-

ance in designing the actual MQW structure used in the MBE regrowth on MOSIS GaAs circuits. The result is then presented.

Finally, the result of neuron circuits from the three fabrication approaches is presented in Chapter 6, including threshold circuits and "bump" circuits. Threshold levels as low as 1nW optical input is observed for neurons with double Zn-diffusion double-heterojunction LED and OPFET receivers built using in-house facilities. However, the poor non-uniformity is a major concern. Mixed result is obtained from the AT&T FET-SEED process. The main advantage is that no post-processing is needed for this process and the main problem is its limitation on circuit design because only depletion-mode FET's are available. An 10×10 array of bump circuits is built using the third approach: low temperature MBE regrowth on MOSIS GaAs MESFET circuits, which is developed from a collaboration project with the Fonstad group at MIT. The result is discussed.

Chapter 7 gives the conclusion of this thesis, mainly focusing on the comparison between the three approaches and further improvement.

Chapter 2 Fabrication Methods of Monolithic OEIC's

The realization of an OEIC neuron array includes the fabrication and, if allowable, optimization of each individual components and the integration of all components in a single process. Depending on the devices we want to integrate, different epi-layer structures and different integration schemes may be needed for fabricating neuron arrays. In this chapter, we will describe the three different integration approaches we have used for building neuron circuits and briefly discuss their advantages and disadvantages. The main focus is on the performance of electronic devices, especially MESFET's. Full discussion of the result will be presented in Chapter 6 and 7.

2.1 GaAs MESFET

A GaAs metal-semiconductor field-effect transistor (MESFET) is chosen as the active electronic device in designing the neuron circuits because of the requirement for low power dissipation and high gain. We do not select bipolar junction transistor (BJT) due to its low current gain at low current level. Another disadvantage of BJT is the lack of input-output isolation, which will place stringent requirement on the design of threshold photodetectors. On the other hand, MESFET is a voltage-controlled device and provides good input-output isolation compared to BJT. In addition, it has good performance even at low current levels.

The cross-section of a simplified GaAs MESFET is illustrated in Figure 2.1(a). It is a three-terminal device with two ohmic contacts, the source and the drain, and a Schottky contact, the gate. Electric current flows between the two ohmic contacts and the amount is controlled by the bias voltage applied on the gate.

The modeling of MESFET can be found in Reference [37]. A brief summary is

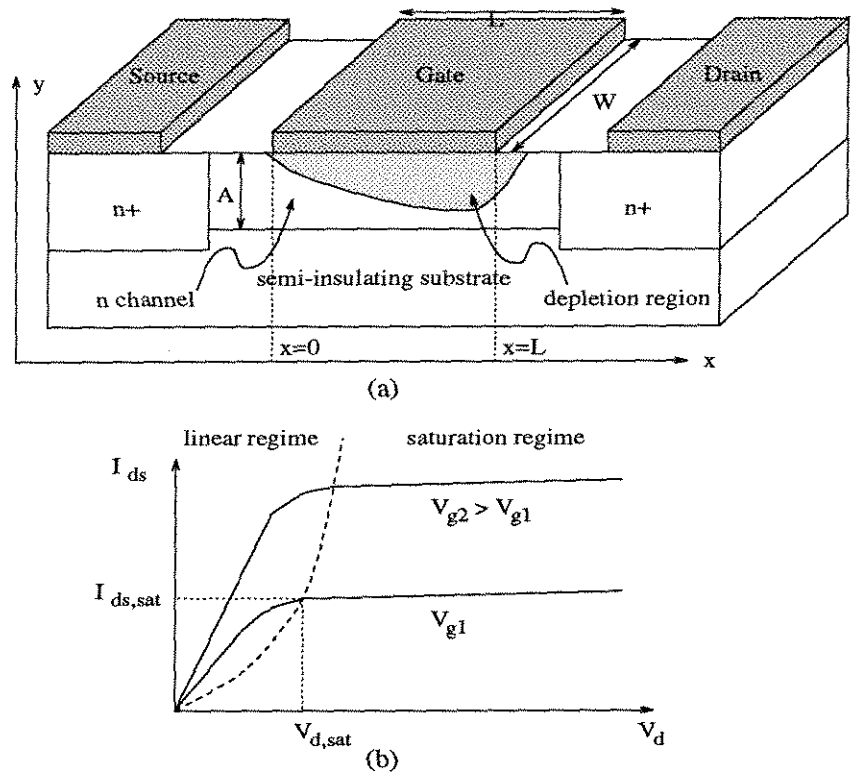


Figure 2.1: Cross-section and I-V curves of a simplified GaAs MESFET. The gate length and width of the MESFET is L and W and the channel thickness is A.

given here to illustrate the basic mechanism. Basically, the drain-source current of a MESFET is determined by the thickness of its conducting channel, which is the difference between the physical thickness of its n-type channel layer and the thickness of its gate depletion region. By applying different gate voltages, the gate depletion region is modulated, resulting in the modulation of drain-source currents.

For a uniformly doped channel, under the gradual channel approximation, i.e., when the channel potential varies slowly with x, the potential across the channel can be solved from the one-dimensional Poisson equation in y-direction. Under the abrupt junction approximation, the depletion region thickness at any point x is given by:

$$A_d(x) = \left\{ \frac{2\epsilon[V(x) + V_{bi} - V_g]}{qN_D} \right\}^{1/2} \tag{2.1}$$

where ϵ is the dielectric permittivity of GaAs, V_{bi} is the built-in voltage of the Schottky gate contact, V_g is the gate potential, N_D is the effective donor concentration in the

channel, and $V(x)$ is the channel potential due to the applied drain-source voltage. In addition, we have the following boundary condictions: $V(0) = 0$ and $V(L) = V_d$.

Further assuming a constant electron mobility, μ_n, when the channel is not totally depleted, we have the following incremental change in channel potential due to electron drifting:

$$\mathrm{d}V = I_{ds}\mathrm{d}R = I_{ds} \cdot \mathrm{d}x/q\mu_n N_D W[A - A_d(x)] \ . \tag{2.2}$$

Substituting Eq. 2.1 into Eq. 2.2 and integrating from $x = 0$ to $x = L$, we can derive the drain-source current as:

$$I_{ds} = g_0 \left\{ V_d - \frac{2[(V_d + V_{bi} - V_g)^{3/2} - (V_{bi} - V_g)^{3/2}]}{3V_{po}^{1/2}} \right\} \tag{2.3}$$

where

$$g_0 = \frac{q\mu_n N_D W A}{L} \tag{2.4}$$

is the channel conductance of the undepleted channel,

$$V_{po} = \frac{qN_D A^2}{2\epsilon} \tag{2.5}$$

is the voltage required to totally deplete the channel and is called pinch-off voltage.

For small V_d, Eq. 2.3 gives $I_{ds} \propto V_d$ and the device is operated in the linear regime as shown in Figure 2.1(b). Further increasing V_d, the drain-source current will saturate because Eq. 2.1 is valid only for those $V(x)$ that satisfies $A_d(x) \leq A$. The saturation occurs when $A_d(L) = A$, or,

$$V_{d,sat} + V_{bi} - V_g = V_{po} \ . \tag{2.6}$$

This is called the pinch-off condition. The saturation current is given by substituting

Eq. 2.6 into Eq. 2.3:

$$I_{ds,sat} = g_0 \left[\frac{V_{po}}{3} + \frac{2(V_{bi} - V_g)^{3/2}}{3V_{po}^{1/2}} - (V_{bi} - V_g) \right] . \tag{2.7}$$

From Eq. 2.7, the gate transconductance in the saturation regime is:

$$\begin{aligned} g_{m,sat} &= \left. \frac{\partial I_{ds}}{\partial V_g} \right|_{V_d=\text{constant}} \\ &= g_0 \left[1 - \left(\frac{V_{bi} - V_g}{V_{po}} \right)^{1/2} \right] . \end{aligned} \tag{2.8}$$

Another important characteristic of a GaAs MESFET is its threshold voltage, given by:

$$V_T = V_{bi} - V_{po} \tag{2.9}$$

which is the gate voltage required to totally deplete the channel at $V_d = 0$V. Depending on the doping concentration and the thickness of the channel, there are two-types of MESFET's: when $V_T < 0$V, the device is normally-on (at $V_g = 0$V) and is called a depletion-mode FET because a negative gate voltage is needed to deplete the channel. When $V_T > 0$V, it is normally-off and is then called an enhancement-mode FET because a positive gate voltage is needed to enhance the channel.

In the above discussion, we made the assumption that the electron has constant mobility. If we consider the field dependence of electron mobility in GaAs, more complicated models are required. An approximate result is given by considering the velocity saturation in GaAs [37]:

$$I_{ds,sat} = \frac{2\epsilon\mu_n v_s W}{A(\mu_n V_{po} + 3v_s L)} \cdot (V_g - V_T)^2 \tag{2.10}$$

where v_s is the saturation velocity of electrons in GaAs, which is slightly higher than 2×10^7cm/sec. This "square law" approximation is accurate for devices with relatively low pinch-off voltages, e.g., $V_{po} \leq 1.5 \sim 2$V. For devices with higher V_{po}, an empirical

model is proposed as [50]:

$$I_{ds,sat} = \frac{\beta(V_g - V_T)^2}{1 + b(V_g - V_T)} \tag{2.11}$$

with experimentally fitted β and b. These models are widely used in circuit simulation tools.

2.2 In-house Fabrication

The most flexible method of fabricating MESFET-based optoelectronic neuron circuits is to develop a process using ones own facilities. With available GaAs wafers, we can finish circuit design, mask making, photolithography and processing, and testing in a continuous way. Therefore, it is very time-effective for developing a new generation of neuron circuits. In addition, the epi-layer of the wafer can be tailored accordingly to achieve the optimum performance. However, such approach suffers from the limited in-house facilities and requires a large amount of capital input. Besides, the quality control of IC circuits is very difficult to maintain in an academic environment.

Our in-house integration process starts with the design of wafer epi-layers. The GaAs/AlGaAs epi-layers are designed to accommodate all the required individual components. The goal here is to optimize the performance of each individual components. However, since all the heterostructures are grown in a single wafer, sometimes compromise has to be made. Once the wafer structure is determined, it is then grown in commercial foundries using metal-organic chemical vapor deposition (MOCVD) or molecular-beam epitaxy (MBE) because we don't have the required equipment.

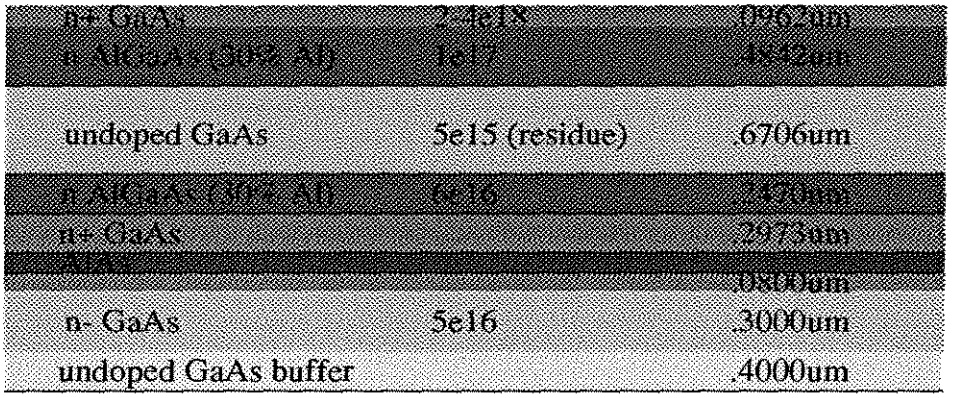
The circuit design is then transferred to layout design using layout tools, such as L-EDIT or MAGIC. It is then converted to postscript files and printed out in films with 3386dpi resolution. Finally, an optical step camera is used to reduce the film by ten times to make the masks. Due to the limitation in printing resolution, the minimum mask feature is limited to about 4μm in this process. By using better mask making techniques, such as e-beam writing, we should be able to reduce the minimum

feature to less than 2μm.

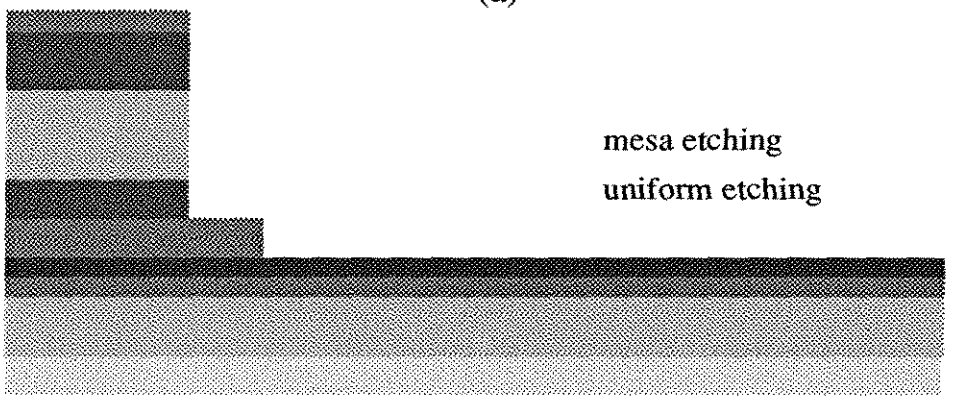
With the desired wafer and necessary photolithography masks, we can begin the fabrication of an integrated circuit. One example is a full run of fabricating LED-based threshold circuits as shown in Figure 2.2. The epi-layers are illustrated in Figure 2.2(a). The LED double-heterostructure is on the top and the MESFET structure is at the bottom with a thin (91Å) AlAs stop-etch layer in between.

First, the wafer is cut into small chips for easy handling and for economic reasons as well. The chip is then cleaned by applying standard cleaning procedure, with 10 minutes soaking in boiled trichloroethylene (TCE), 5 minutes ultrasonic cleaning in acetone, followed by 5 minutes ultrasonic in methanol. A photo-mask is used to define the LED mesa by chemical etching with a phosphoric acid-based etchant, which consists of a mixture of H_3PO_4, H_2O_2, and CH_3COOH in a ratio of 1:1:3. The etching rate is approximately $2 \sim 3\mu$m/minute with agitation at room temperature. The exposed area is etched down to the second n^+ GaAs layer so that ohmic contact can be made for the LED. For the structure shown here, an AlAs stop-etch layer is inserted between the LED structure and the MESFET structure in order to improve etching uniformity. Therefore, the next step is to uniformly etch down to the AlAs layer with the LED area protected by photoresist (Figure 2.2(b)). In some designs, the AlAs layer is not incorporated in the design and this step can be omitted. The selective etch is done with an etchant consisting of citric acid, water, and hydrogen peroxide [51]. First, citric acid is dissolved in DI water with 1:1 weight ratio. The solution is then mixed with 30% H_2O_2 solution in a volume ratio of 10:1. At 18°C, the etching rate is about 0.2μm/minute for GaAs and less than 0.04μm/minute for AlGaAs. At optimum condition, the ratio of the two etching rates can be more than 10:1.

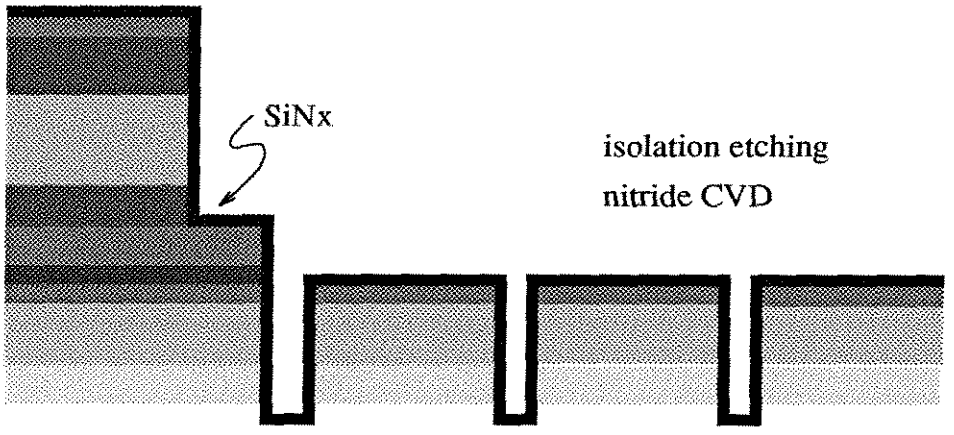
A non-selective etchant is applied to etch the AlAs layer. Because the layer is so thin, the resulting non-uniformity can be reduced. A phosphoric acid-based non-selective etchant is then used to etch down to the semi-insulating GaAs substrate to achieve isolation between different components, followed by a deposition of approximately 1000Å silicon nitride (SiN_x) on the surface by using a thermal chemical vapor


n+ GaAs
2-4e18
.0962um
n AlGaAs (30% Al)
1e17
.4842um
undoped GaAs
5e15 (residue)
.6706um
n AlGaAs (30% Al)
6e16
.2470um
n+ GaAs
.2973um
AlAs
.0800um
n- GaAs
5e16
.3000um
undoped GaAs buffer
.4000um
SI substrate

(a)


mesa etching
uniform etching

(b)


SiNx
isolation etching
nitride CVD

(c)

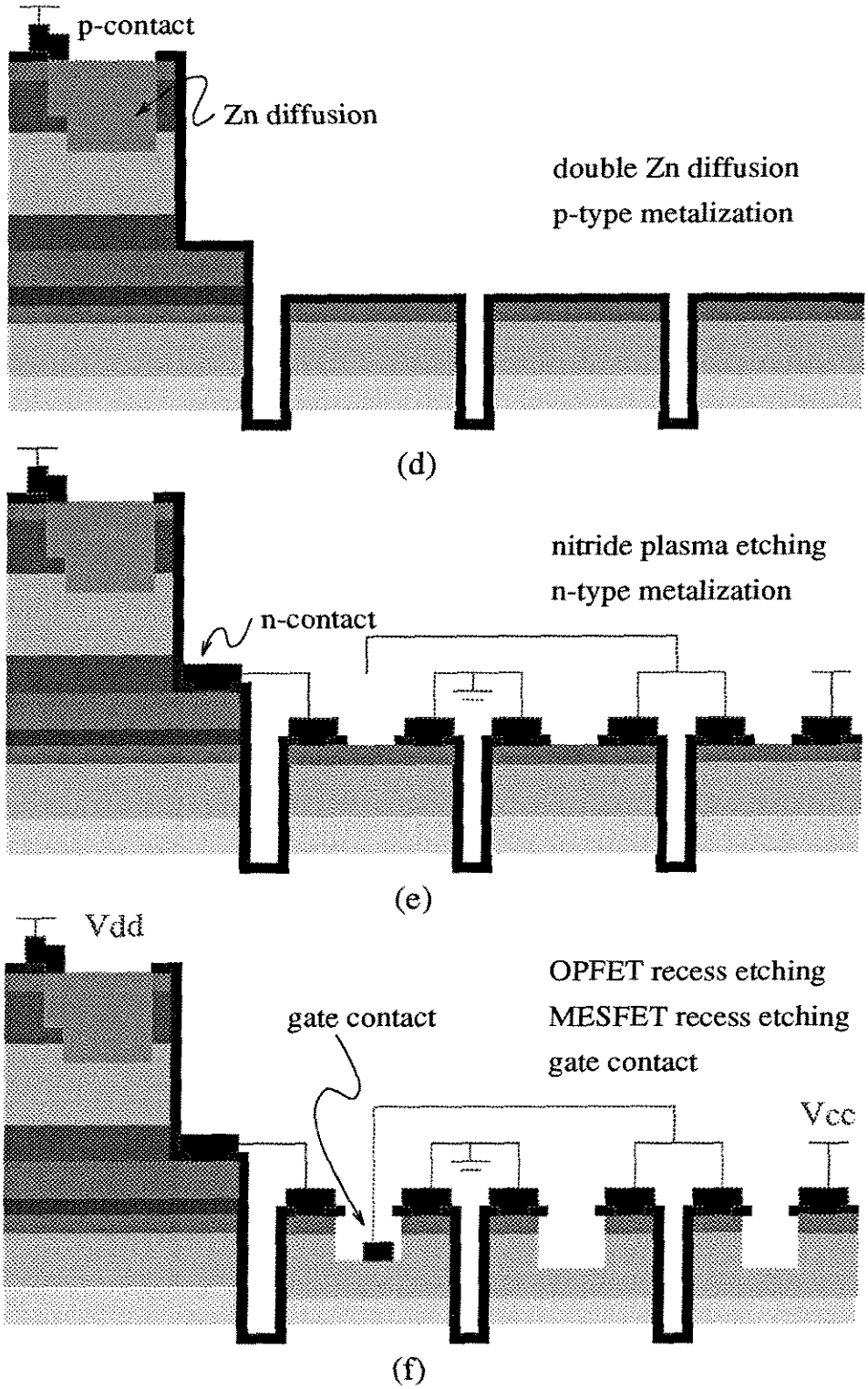


Figure 2.2: Schematic illustration of our in-house fabrication process. Except the growth of GaAs wafer, all steps are done at Caltech.

deposition (CVD) system heated to 610°C. The reacting gases are silane, nitrogen, and ammonia and the thickness is monitored by the color of the deposited nitride (Figure 2.2(c)). The heater is turned off when the deposited nitride shows a color of deep blue to pale blue. Normally the deposition time is between 5 to 10 minutes.

The next step is to use Zn diffusion to convert the top n-layer to p-type in order to provide the current confinement for the LED's, which will be discussed in Chapter 3. The nitride layer is used as the mask for the diffusion while the diffusion windows are opened by CF_4 plasma etching. The chip and Zn_3As_2 source are placed in a quartz ampule, which is then pumped to a vacuum of 8×10^{-8} torr. It is sealed with a torch and inserted into a 640°C furnace for the diffusion process [52]. Depending on the required diffusion depth, the ampule is left inside the furnace for different times, which can be calculated from measured diffusion rates [53], before it is pulled out and quickly quenched. The second diffusion follows with larger nitride windows opened by CF_4 plasma etching. The two diffusion times should be calculated carefully so that the desired doping profile can be obtained. Ohmic contact metal is then deposited using e-beam and/or thermal evaporation. A lift-off technique is used to define the p-contact for the LED. The metal is Cr/Au with a thickness of 150Å/1000Å, respectively for Zn-diffused material. In several particular runs, instead of using Zn diffusion to convert the top n-layer, a p-type GaAs layer is grown by MOCVD and the contact metal is changed accordingly to AuZn/Au (200Å/1000Å) (Figure 2.2(d)). Before the chip is loaded into the evaporator chamber, it is briefly soaked in a dilute HCl solution to etch away the oxidized surface. This step can help to reduce the resistance of deposited ohmic contacts.

The n-contact of the LED and the drain and source contacts of the FET are then deposited using a similar technique. The metal is AuGe/Ni/Au with a thickness of 250Å/100Å/1000Å, respectively after selectively opening Si_3N_4 windows (Figure 2.2(e)). The devices are subsequently annealed at 410°C in nitrogen ambient for 1.5 minutes to achieve ohmic contact.

The next step is the recess etch of gate regions for MESFET's and OPFET's. The etching is performed using the nitride as a mask in order to obtain a self-aligned recess.

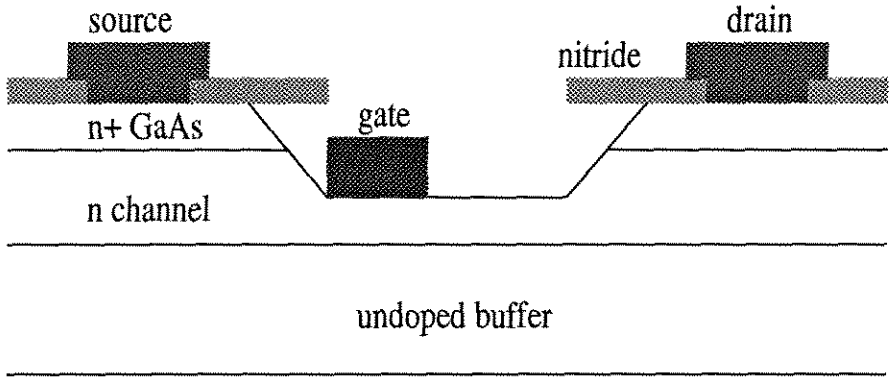


Figure 2.3: Detailed cross-section of a MESFET fabricated using self-aligned process. The undercut is due to wet chemical etching.

The etchant used consists of NH_4OH, H_2O_2, and H_2O in a ratio of 3:1:139 with an etching rate of about 30Å/second. Finally, the Schottky gate metal is deposited using an e-beam evaporator. The metal is Ti/Pt/Au with a thickness of 250Å/100Å/1000Å, respectively, and the fabrication process is finished (Figure 2.2(f)).

The recess etching of MESFET gate region and the Schottky metal deposition are the most critical steps in this process because a small change here will strongly affect the performance of the MESFET fabricated. During the recess etching, we periodically measured the drain-source current of the FET until a preset level is reached. This preset I_{ds} depends on the desired threshold levels. In addition, when the gate metal is deposited, the forming of a Schottky junction will widen the gate depletion region and reduce the drain-source current. This effect has to be taken into consideration beforehand. Otherwise, the channel will be etched too far. When depositing the gate metal, it is asymmetrically placed closer to the edge of the source as shown in the detailed MESFET cross-section in Figure 2.3. There are two reasons: first, it helps to reduce the gate-source parasitic resistance, R_{gs}. In return, the transconductance is increased because of a reduced voltage drop from source to gate for the same I_{ds}. Second, it helps to increase the drain-gate breakdown voltage because the gate is far away from the drain.

In Figure 2.3, the undercut of GaAs layers under SiN_x is due to wet chemical

etching used in our process. Because the undercut profiles at the sidewalls of GaAs mesa are different in different directions, special consideration should be given in designing the metal interconnection lines. In [110] direction, the undercut profile is similar to what is shown in Figure 2.3, so step coverage of metals can be easily achieved from the mesa to the substrate. However, no step coverage can be provided along the $[1\bar{1}0]$ direction because of the undesired undercut profile.

Due to the limited resolution in making masks, the minimum feature can be controlled during in-house fabrication is 4μm. Therefore, the device we can fabricate is far away from state-of-the-art semiconductor devices. Besides, the fabrication lab is under normal atmosphere with limited "clean" area, so the fabricating process is subjected to lots of particle contamination. Therefore, it is very difficult to obtain good quality control. Furthermore, one major problem for fabricating large optoelectronic neuron arrays is the non-uniformity of wet chemical etching. Typically, we observe more than 10% non-uniformity of etching depth across a 1cm$\times$1cm chip. As discussed in Chapter 1, this non-uniformity greatly increases the power consumption of a neuron array and reduces the maximum neuron density that can be achieved. Even worse, the poor uniformity sometimes results in abnormal behavior from the fabricated circuits. Nonetheless, despite all these problems, the in-house fabrication process is still an effective approach in building prototype neuron circuits due to the short time required for a full run of processing.

The performance of a GaAs MESFET built using this process is shown in Figure 2.4 for an enhancement-mode MESFET and Figure 2.5 for a depletion-mode MESFET. From Figure 2.4, the breakdown voltage is found to be $4 \sim 5$V and the saturation voltage is less than 1V. The device has a gate length of 5μm and a gate width of 50μm. The transconductance is measured to be 35mS/mm at $V_{gs} = 0.5$V. For the depletion-mode MESFET shown in Figure 2.5, the transconductance is measured to be 66mS/mm at $V_{gs} = 0$V. However, due to the poor quality control, these numbers vary significantly for different devices in the same run, not to mention devices from different runs. Besides depletion-mode and enhancement-mode MESFET's, the devices can be fabricated in this process include Schottky diode, which is the gate

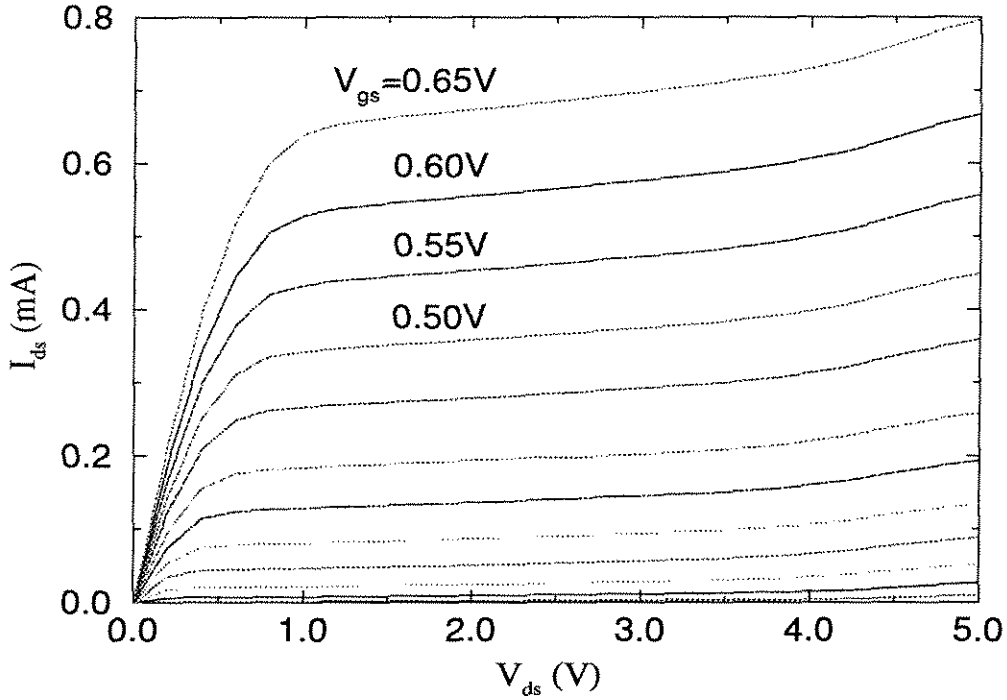


Figure 2.4: I-V curves of an enhancement-mode MESFET built using in-house process. The gate voltage changes from 0.1V to 0.65V at an increment of 50mV. The device has a 5μm gate and its width is 50μm.

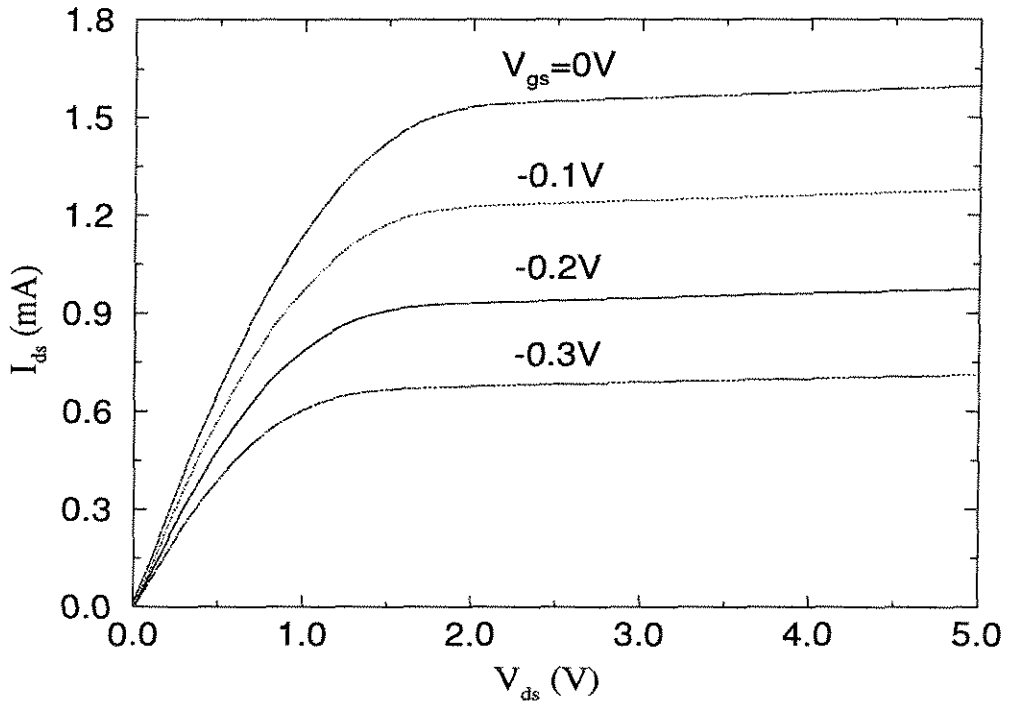


Figure 2.5: I-V curves of a depletion-mode MESFET built using in-house process. The device has similar geometry design as the one shown in Figure 2.4.

contact; MSM photodetector, which consists of two gate contacts on the n-channel; LED (or other output devices), which will be discussed in Chapter 3; and high gain optical FET detector, which will be discussed in Chapter 4.

2.3 FET-SEED Process

As we mentioned in the previous section, the non-uniformity of the OPFET's and MESFET's is a major problem in fabricating practically useful high density OEIC neuron arrays using our in-house process. This non-uniformity comes from the extreme sensitivity of FET's to the depth and doping profile of the FET channel. In our case, the non-uniformity is mainly dominated by the recess etching non-uniformity of the FET gate region. On order to control the uniformity of the FET's, large amount of time and capital would be required if we were to use our in-house facilities. On the other hand, industrial foundries provide the solution to this non-uniformity challenge. Furthermore, ultimately the successful fabrication of large OEIC neuron arrays would require the involvement of semiconductor industry for commercial applications.

In this section, we will describe one industrial approach in fabricating OEIC neuron circuits — the AT&T FET-SEED process. This process integrates GaAs FET's, photodetectors, and multiple quantum well (MQW) modulators in a batch fabrication. Therefore, by careful circuit design, it is possible to build an OEIC neuron array using this process and no further post-processing is required.

2.3.1 MQW Modulators and FET-SEED Workshop

We will discuss the mechanism of GaAs/AlGaAs MQW modulators and the designing of MQW structures in Chapter 5. Basically, electrons and holes inside a quantum well are distributed among discrete energy levels due to the quantum effect. When an electric field is applied perpendicular to the quantum well, the change in energy band will cause changes in those discrete energy levels. In return, the absorption edge shifts toward longer wavelengths. This electro-optic effect is called quantum-confined Stark effect (QCSE). Based on QCSE, MQW modulators can be built. The contrast

ratio is around 4:1 for voltage changes of about 8 ~10V. The MQW modulators are sometimes called self-electro-optic effect device (SEED) due to historical reasons [54].

Since the first observation of the QCSE [55], the MQW modulator has received considerable attention because of its several distinct advantages. As we mentioned in the last chapter, modulators are low power consumption devices. The required switching energy is what is needed to charge up the reverse-biased MQW p-i-n diode. Therefore, high density neuron arrays can be built by using these modulators as optical output devices. In addition, SEED's can be operated up to very high speed and it can be easily integrated with other optical or electronic devices.

The integration of SEED's with field-effect transistors was first demonstrated on individual circuits by AT&T Bell Laboratories [56]. Later it was developed into a batch fabrication process [57, 58] and was applied to the fabrication of optoelectronic smart pixel arrays [59] and other applications [60]. The AT&T FET-SEED process is available to non-AT&T users through the FET-SEED workshop organized by the Consortium for Optical and Optoelectronic Technologies for Computing (CO-OP) [61] and co-sponsored by ARPA and Bell Laboratories. The workshop attendants designed their own circuits during the workshop and submitted the final version to Bell Laboratories one week after that. All the designs were collected and fabricated together using the FET-SEED process at Bell Labs. The finished chips were then sent back to the attendants several months later. The fabrication cost is sponsored by ARPA and AT&T Bell Labs. In this way, relatively low cost customer-designed OEIC's with good uniformity were provided. The main advantage for using this process to fabricating OEIC neuron circuits, compared to other processes, is that no post-processing is required since it provides electronic devices, photodetectors, and optical output devices with a single fabrication process.

2.3.2 Structure and Fabrication

The structure of the FET-SEED's is shown in Figure 2.6. Starting from a semi-insulating GaAs substrate, there is a 500Å undoped GaAs buffer and 16 pairs of

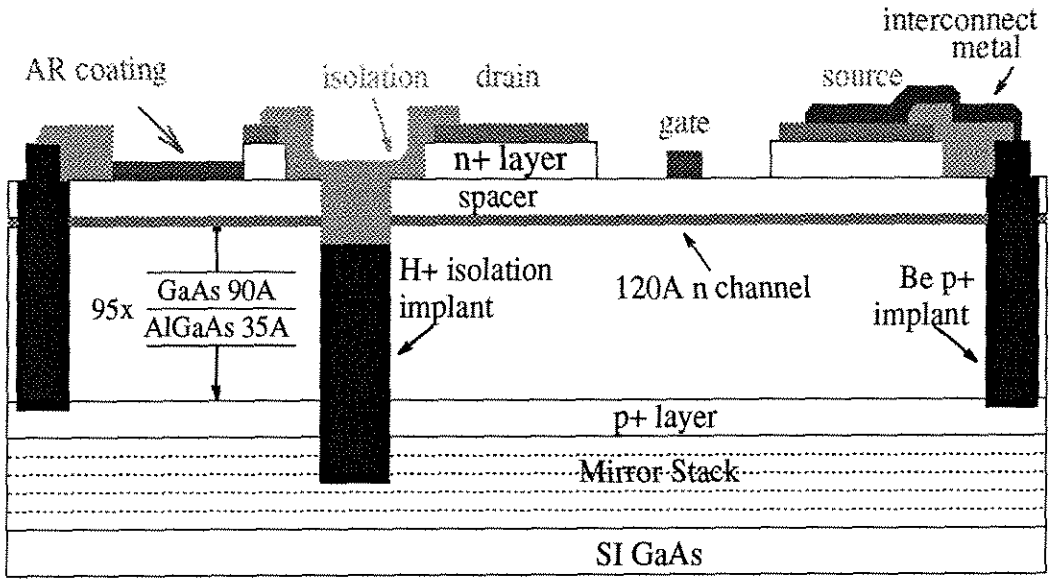


Figure 2.6: AT&T FET-SEED structure. On the left is a MQW modulator with anti-reflection coating; on the right is a GaAs doped-channel FET with the backgate connected to the source.

quarter wavelength distributed Bragg reflector (DBR) stack. Each pair consists of 711Å AlAs and 599Å $Al_{0.11}Ga_{0.89}As$. On the top of the reflector, 5000Å p-doped ($3 \times 10^{18}cm^{-3}$) $Al_{0.11}Ga_{0.89}As$ is grown as the contact layer. Following that, 95 undoped quantum wells are sandwiched between two $Al_{0.11}Ga_{0.89}As$ buffer layers. The quantum well is comprised of 90Å GaAs well and 35Å $Al_{0.3}Ga_{0.7}As$ barrier; and the buffer layers are 500Å on the bottom and 200Å on the top. The doped-channel FET structure is then grown over the MQW: on the top is a 1000Å n-doped ($1 \times 10^{18}cm^{-3}$) GaAs contact layer and on the bottom is the 120Å n channel ($1 \times 10^{18}cm^{-3}$ doped GaAs). An undoped 900Å $Al_{0.11}Ga_{0.89}As$ spacer is grown between the contact layer and the channel. p^+ Be implant or etching is used to achieve electric connection to the p-layer for the SEED. This contact and the top metal contact are then used to apply the required bias voltage to the MQW p-i-n diode. The undoped $Al_{0.11}Ga_{0.89}As$ spacer is used as a stop etch layer for the shallow etching of top GaAs layer to reduce insertion loss for MQW modulators and to define the gate region for the FET. This design is used to increase the uniformity of fabricated circuits. The gate metal on the FET's

is deposited through the self-aligned lift-off technique. The minimum gate length is 1μm and two interconnection metal layers are provided.

The fabrication steps are shown in Figure 2.7 [62]. First, the lift-off technique is employed to deposit fiduciary alignment marks before the first processing step. Next, ion-milling is used to etch slightly into the MQW region to define the mesas for modulators and FET's, followed by a proton ion-implant to achieve device isolation in the buried p^+ layer. Following either p-well etch to expose the p^+ layer or Be ion-implant to make electric connections to the p^+ layer, ohmic contacts to p^+ regions are deposited using lift-off technique and then annealed. What follows are the deposition of the n-metal and its annealing. Deep Au/Ge spikes are formed during the annealing and they can penetrate through the 900Å undoped AlGaAs spacer, resulting in a good conducting path to the n-doped GaAs channel. The first interconnection metal is then deposited directly on the undoped semiconductor material. This metal line does not have good conductivity and should be keep as short as possible during the design. Silicon nitride is then deposited to provide insulating layer between the first and the second interconnection metal. Selective areas are then etched to make optical windows for modulators, to open connection vias, and to expose the gate region for FET's. The most critical step is the creation of the self-aligned FET gates. Reactive ion etch (RIE) selectively etches away the top n^+ GaAs layer and stops at the AlGaAs layer. In this way, good uniformity can be achieved. The gate metal is then deposited in a self-aligned manner similar to what we used in our in-house process described earlier. The last three steps are the deposition of second interconnection metal and the anti-reflection coating, followed by the AR-coating etch to expose metal pads. The interconnection metal and the gate metal are both Pt/Ti/Au.

A photograph of an optoelectronic bump circuit fabricated using the AT&T FET-SEED process is shown in Figure 2.8. The area for the probe pads is $40 \times 40\mu m^2$. The designed is not optimized in order to make the testing easier. Because only depletion-mode FET's are available, the voltages have to be reset after each step, resulting in many bias voltage pads.

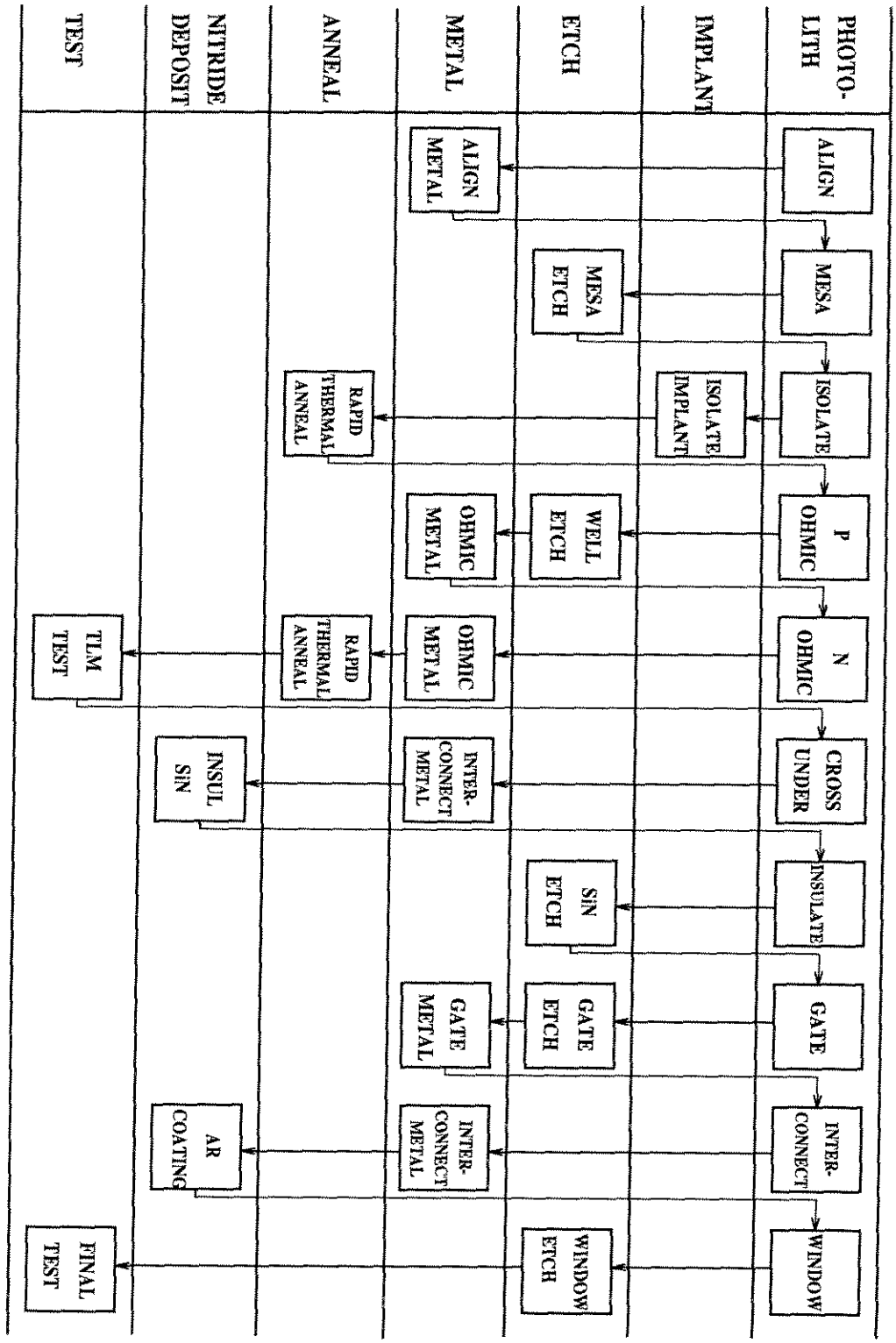


Figure 2.7: Fabrication steps for the AT&T FET-SEED process. Including the first mask for alignment marks, ten masks are need in total.

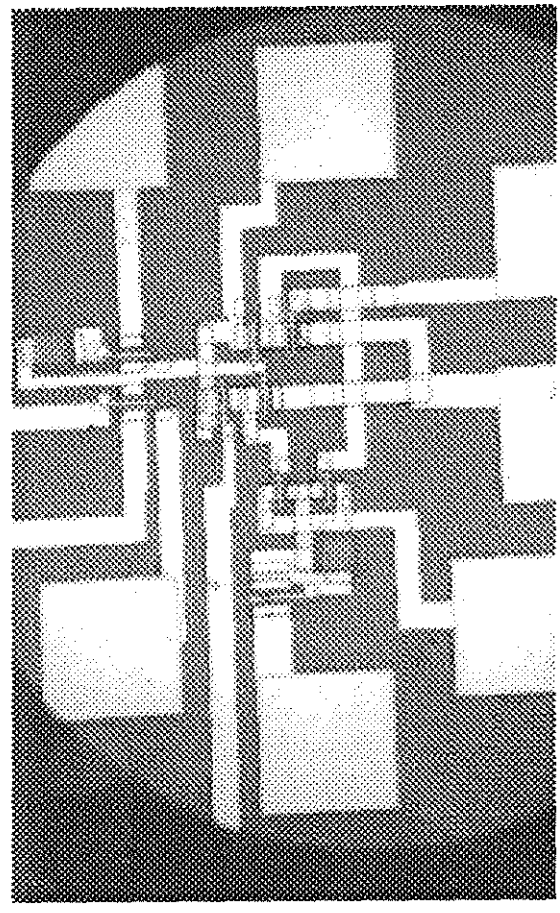


Figure 2.8: Picture of a optoelectronic bump circuit fabricated using the AT&T FET-SEED process. The probe pads are $40\mu m \times 40\mu m$.

2.3.3 Device Performance

The FET-SEED process integrates electronic devices, photodetectors, and optical output devices (MQW modulators) in a single run. In this section, we will present the measured result of these individual components, including the MQW photodiode, the MQW modulator, and the FET. It turns out that by proper design, the FET can be also used as a high gain photodetector, called OPFET. The OPFET measurement will be discussed in Chapter 4, together with high gain detectors built in other fabrication approaches because of its importance in enhancing the performance of optical neural networks.

SEED

For the SEED, the quantum confined Stark effect can be easily monitored by measuring the photocurrent from the reverse-biased p-i-n photodiode. For a normal photodiode, electrons (and holes) generated in the depleted i-region due to photon absorption are swept by the high electric field to the n (and p) region, resulting in a photocurrent proportional to the incident optical power. When the reverse bias voltage is higher, the photocurrent will increase slightly because of the widening of the depletion region and the higher reverse current of a diode. However, that's not the case for a MQW photodiode.

When there is no input light, the dark current of the photodiode is very small at reverse bias. We observed a 1.3nA dark current at 10V reverse bias for a $20\mu m \times 20\mu m$ device. When the light is shone on the diode, the responsivity, η, — calculated by dividing the photocurrent by the input power — gives us quantitatively the absorption inside the quantum wells assuming the whole MQW is within the depletion region. We measured the photodiode responsivity at different bias voltages for different wavelengths and the result is plotted in Figure 2.9 for a $30\mu m \times 30\mu m$ SEED. The laser beam, coming from a tunable Ti:sapphire laser, is focused and goes through a beam splitter in our measurement setup. One beam from the beam splitter is used to monitor the incident power and the other as the incident light on the chip. The incident

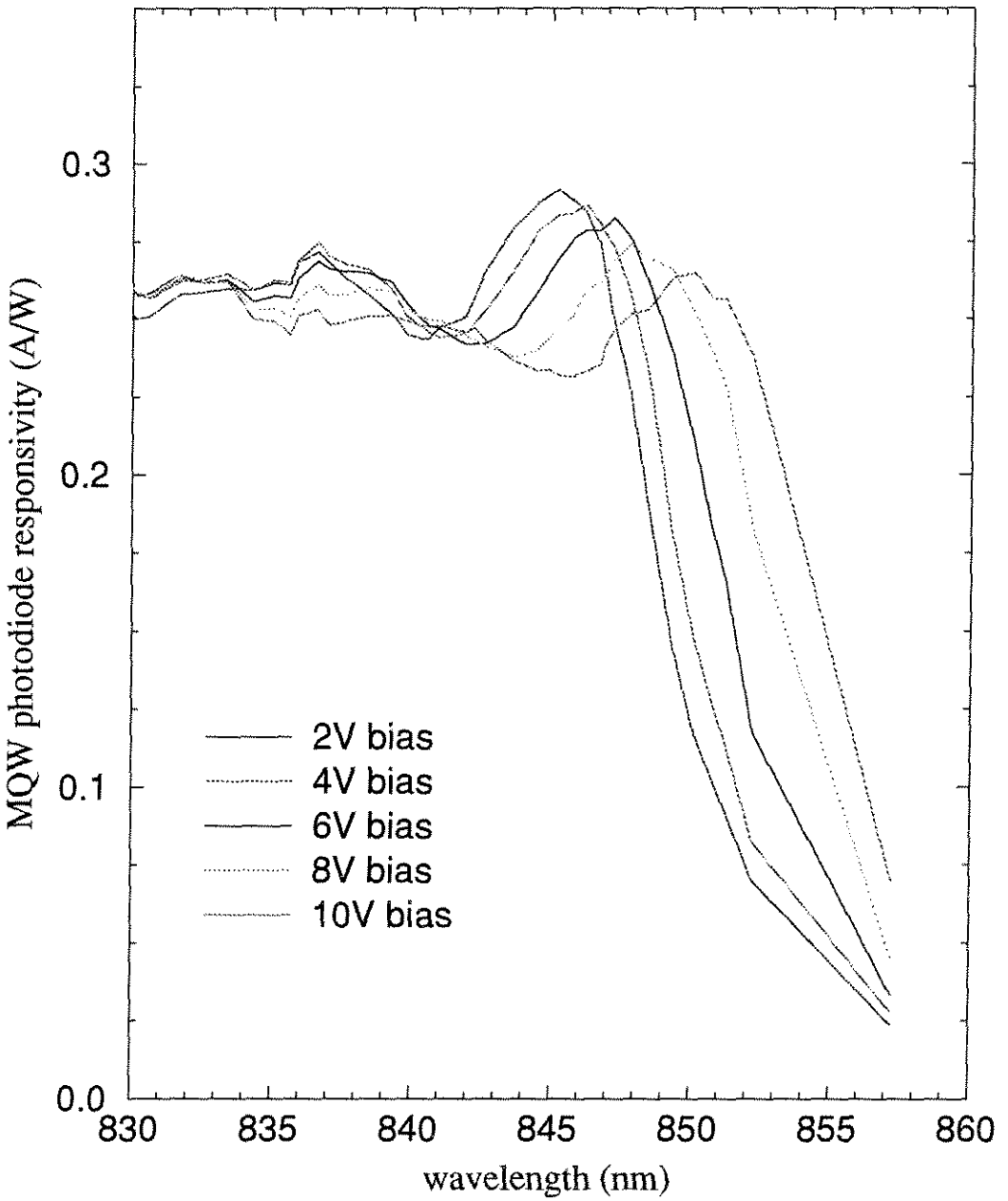


Figure 2.9: Experiment measurement of the quantum confined Stark effect. The device is a 30μm$\times$30μm MQW p-i-n photodiode. The zero bias curve is not shown because the MQW is not totally depleted at this low field.

power is maintained at around 1μW to avoid any saturation effect. Since we did not have a good beam splitter for the 850nm wavelength when the measurement was done, the absolute value of the responsivity, η, may not be accurate. From Figure 2.9, the heavy hole exciton peak, which will be discussed in Chapter 5, can be clearly observed, so can the light hole exciton peak. As the reverse bias increases, the absorption peak shifts toward longer wavelengths, from 845.2nm at 2V bias to 850.1nm at 10V bias. In addition, the resonance width becomes wider and the peak absorption is smaller. All these agree well with the theoretical prediction which we will discuss in more detail in Chapter 5.

The reflection MQW modulators can be operated in either λ_0, which is the wavelength of the zero bias absorption peak, or λ_1, which is the wavelength of the absorption peak at high electric field. From Figure 2.9, we can see that λ_0 is around 845nm and λ_1 is around 850nm for 10V bias. At λ_0, when the modulator has low bias voltage, the absorption coefficient is high as shown in Figure 2.9, therefore, its reflected output is low. When the bias increases, the absorption decreases. In return, the reflected output is higher. This is the mechanism of a MQW modulator. The reversed case occurs at λ_1 where the device is "ON" at low bias and "OFF" at high bias. Even though it looks like the 850nm wavelength would have much larger changes in absorption compared to 845nm, keep in mind that the output is proportional to $1-\eta$. The measured contrast ratios for the λ_0 and λ_1 modes are roughly the same. The measured result of the 30μm$\times$30μm SEED is shown in Figure 2.10. At bias voltage of about 10V, we can get a contrast ratio near 5:1. If our optical system can work with lower contrast ratios, the required bias voltage can be lowered. An 8V bias would give us a contrast ratio near 4:1.

LED

Besides using MQW modulators as optical output devices, the MQW p-i-n diode can be forward-biased and used as a light-emitting diode if on-chip light sources are needed. The turn-on voltage of the MQW LED is around 1.2V. Since these devices are not optimized as LED's, the luminescence efficiency is not good. For a 20μm$\times$20μm

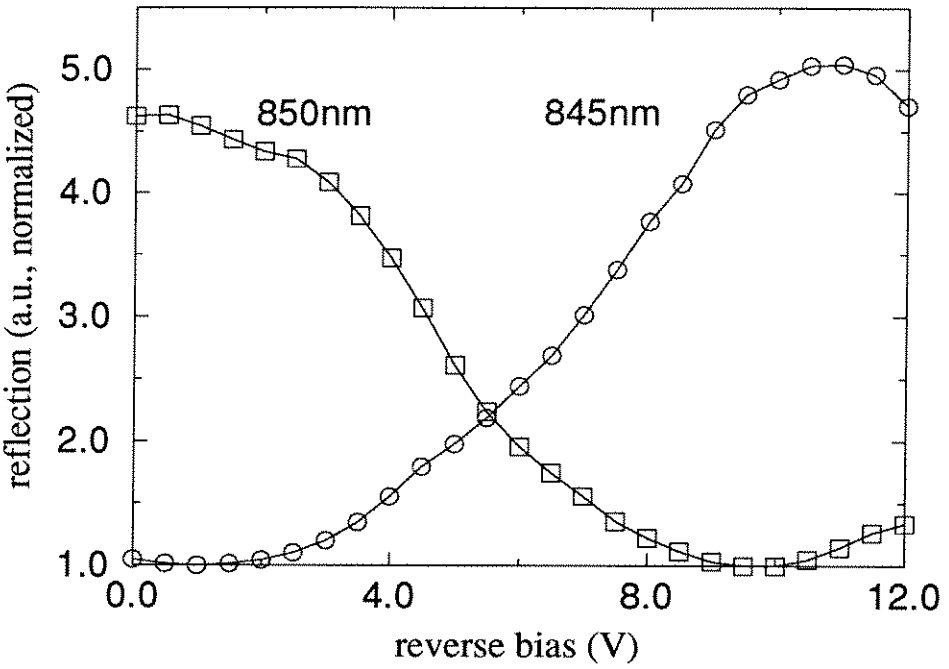


Figure 2.10: λ_0 (845nm) and λ_1 (850nm) operation of the MQW modulator. Contrast ratios around 5:1 can be obtained with reverse biases around 10V. At reduced voltage of 8V, the contrast ratio drops to about 4:1. Incident power is around 200μW.

LED, as shown in Figure 2.11, the differential efficiency is about 0.5×10^{-3}W/A for driving currents more than 2mA. As we will discuss in Chapter 3, at low current levels, the non-radiative recombination along the mesa surface significantly reduces the luminescence and the efficiency drops to 0.22×10^{-3}W/A at 0.5mA and less than 0.1×10^{-3}W/A at 0.1mA.

FET

The FET is a depletion-mode (normally-on) transistor with a threshold voltage of -1.2V. The metal-insulator-semiconductor structure is specially designed so that high breakdown voltages can be achieved. The measured breakdown voltage is more than 9V. By applying voltages on the gate contact, we can change the conductance of the channel. The mechanism is different from what we discussed in Section 2.1. Instead of modulating the channel thickness, the device has an approximately fixed channel thickness of only 120Å. Electrons are confined in this thin layer by the two wide bandgap AlGaAs barrier layers. Therefore, we have a 2-dimensional electron

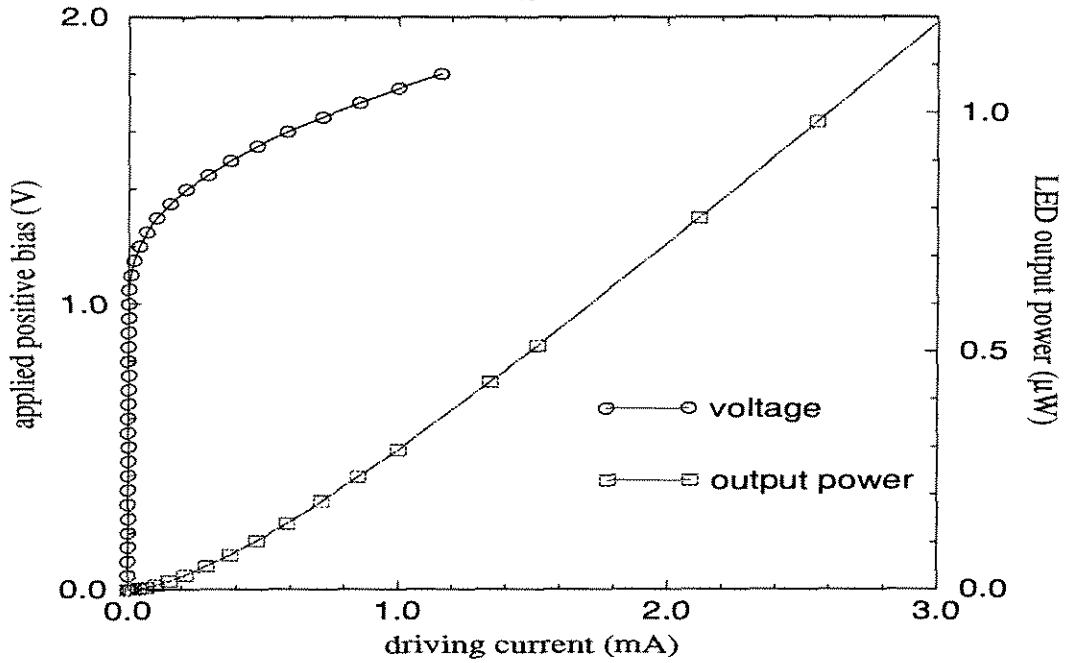


Figure 2.11: The I-V and L-I characteristics of a MQW LED. The size of the LED is $20 \times 20\mu m^2$. Nonradiative recombination is significant for driving currents below 1.0mA.

gas, which provides the conduction channel for the FET. Different gate voltages will change the position of the Fermi-level with respect to the conduction band edge. In return, the concentration of the 2-dimensional electron gas is modulated, resulting in the modulation of drain-source current. The advantage of such a design compared to standard MESFET's is its high breakdown voltage that is required for operating MQW modulators and its easy control of uniformity.

The I-V characteristics of one such FET is shown in Figure 2.12, where the FET has a gate length of $1\mu m$ and a width of $10\mu m$. The transconductance is close to 76mS/mm and the saturation current is about 55mA/mm at $V_{gs} = 0V$. Because of the AlGaAs stop etch layer, the FET's have good uniformity over the whole wafer. We tested 18 devices on 4 chips from the same wafer. The standard deviation of the threshold voltage is 0.065V (5.4%) and the standard deviation of saturation current is 4.3mA/mm (7.8%). For the transconductance, the measured standard deviation is 4.18mS/mm, or 5.5% of the mean value.

Because the MQW structure and the p contact layer are sitting under the FET,

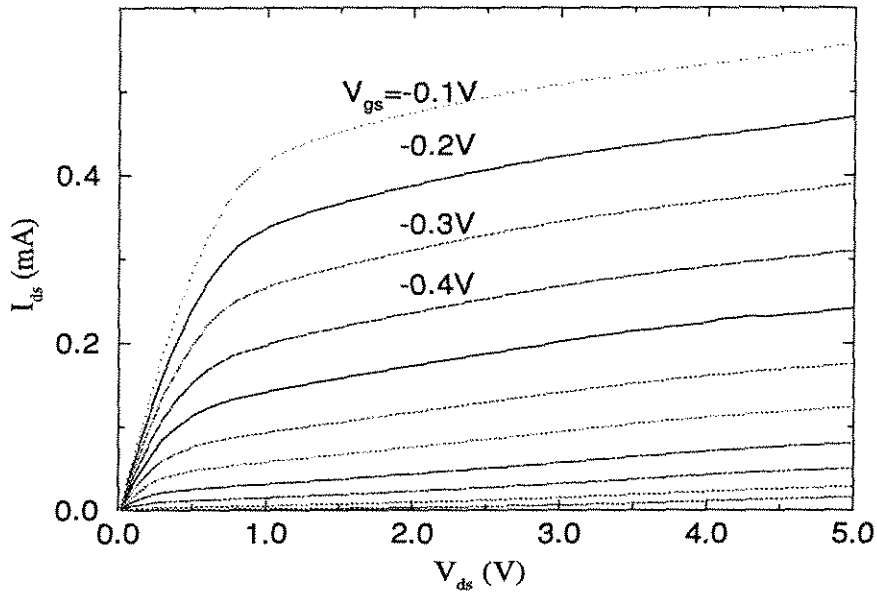


Figure 2.12: I-V characteristics of the doped channel FET with different gate bias voltages. Device has a gate length of 1μm and gate width of 10μm. The threshold voltage is at about -1.2V. The breakdown voltage is at more than 9V.

this transistor is more like a four terminal device with two gates. By changing the bias on the p contact layer, which we will refer to as backgate, we can also change the saturation current of the FET. However, because the thickness of the MQW stack is more than the thickness of the top spacer, a change in the backgate voltage will cause a smaller change in the 2-dimensional electron concentration compared to a similar change in the gate voltage. In return, the backgate transconductance is smaller than the gate transconductance. We measured a value of 50mS/mm at 0V bias when the gate is left unconnected. When the gate bias is fixed, the effect of backgate voltages on the saturation current is reduced but still noticeable. As an electronic device, the backgate of the FET is shorted to its source to avoid any undesirable effect. However, as we will discuss in Chapter 4, the existence of this backgate is very important when the FET is operated as a high gain photodetector.

2.4 MBE Regrowth on MOSIS GaAs Circuits

Even though it was possible to fabricate an OEIC neuron array using the AT&T FET-SEED process as we described in last section, there exists one major disadvantage with the FET-SEED process. Only one type of FET is provided: the depletion-mode FET. Therefore, the design of neuron circuits is more complicated and normally requires more IC elements. It also results in higher power consumption because the FET is normally-on. In return, the neuron density is reduced. In addition, as we can see from Figure 2.8, many bias voltages are needed to reset the output voltages of each stage correctly, resulting in many interconnection metal lines. The end result is a larger area for each neuron and a larger parasitic capacitance. The solution of this problem requires a process that provides better electronic devices, such as both depletion-mode and enhancement-mode MESFET's. In this section, we will describe one of such processes: the low temperature MBE regrowth on MOSIS GaAs MESFET circuits.

Through the MOS Integration Service (MOSIS) [63], one can have custom-designed GaAs MESFET circuits fabricated at reasonable prices, currently \$450/mm^2 for die sizes up to 15mm$\times$15mm. The circuits are fabricated by Vitesse Semiconductor Corporation at Camarillo, California. Currently, there are six runs every year and each fabrication cycle takes about two months. MESFET's with very good uniformity can be obtained from this process. In addition, as we will discuss in Chapter 4, it is also possible to fabricate high gain photodetectors without any modification to this process. However, it does not have any heterostructure which is typically required for GaAs optical output devices. Therefore, some sort of post-processing is needed to grow such devices. MBE regrowth [64] is developed for this purpose and it enables the fabrication of a complete optoelectronic neuron array.

2.4.1 Vitesse GaAs Process

MOSIS collects all circuit designs from its users and submits them to Vitesse Semiconductor Corporation for fabrication. These GaAs MESFET circuits are then fab-

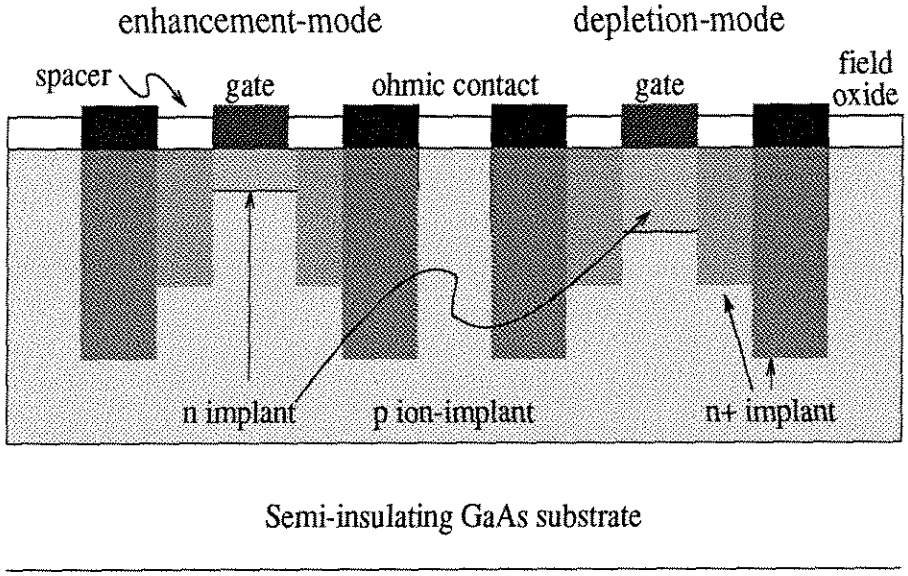


Figure 2.13: The cross-section of MESFET's fabricated using the Vitesse HGaAs3 process. Different ion implantation processes are used to define the two types of MESFET's.

ricated together using the Vitesse high-speed GaAs III (HGaAs3) process on 4-inch-diameter wafers. The HGaAs3 process provides both depletion-mode ($V_T = -0.6$V) and enhancement-mode ($V_T = 0.25$V) MESFET's. The minimum gate feature is 0.8μm and four layers of Aluminum interconnection are provided. The Schottky gate contact can also be used as a Schottky diode if necessary, which mainly serves as a level shifter. Compared to the FET-SEED process, the design of neuron circuits is much simpler now because we can have two types of MESFET's.

The cross-section of MESFET's fabricated using the Vitesse HGaAs3 process is shown in Figure 2.13. It is a process based on ion implantation rather than the chemical etching we used in our own process. Therefore, good uniformity can be achieved by eliminating the non-planar etching. The HGaAs3 process starts with a uniform ion implant to convert the GaAs semi-insulating substrate to slightly p-type. This is because GaAs substrates grown by liquid-encapsulated Czochralski (LEC) normally contain mid-bandgap deep traps to make the material semi-insulating. The density of such deep traps varies significantly from substrate to substrate, which may cause non-uniformity if left untackled. Fortunately, the deep trap density is very

small and a slight p-ion implantation can easily suppress such non-uniformity. The implantation depth is about 1μm in this case.

Next, a thin layer of Si_3N_4 and SiO_2 is deposited on the surface using plasma-enhanced chemical vapor deposition (PECVD). This layer, even though often referred as the field oxide, is different from the field oxide layer applied in a silicon MOS process. Its main function is to provide an insulation layer between the substrate and any interconnection metal. It is then selectively etched away in areas used for transistors. Si ion implantation is then applied to establish the required n-channel. Two ion implant steps are needed here for the two types of MESFET: first, a 0.5μm n-type ion implant for depletion-mode MESFET's, followed by the second n-type ion implant for enhancement-mode MESFET's with a implant depth of about 0.3μm.

Following the n-type ion implantation, self-aligned MESFET process is applied. First, tungsten-based refractory gate metal (WN_x) is deposited by sputtering with a minimum feature of 0.8μm. Then, an n^+ ion-implant is applied to defined the drain and source area by using the gate metal and the field oxide as masks. A thin spacer oxide is then deposited and etched away to result in small spacers on both sides of the gate metal. The purpose of the spacer oxide is to increase the breakdown voltage of the MESFET. Second n^+ ion implant is applied for the drain and source contact regions to reduce the gate-source series resistance. After that, the wafer is deposited with a dielectric layer of Si_3N_4/SiO_2 and then goes through an 800°C thermal annealing to activate the implanted dopant. The Si_3N_4/SiO_2 layer serves as an encapsulating layer to preserve stoichiometry and crystal structure because at high temperature GaAs crystals tend to decompose and emit arsenic.

Next, drain and source ohmic contacts are deposited following the etching of dielectric layers in the specified contact regions. The ohmic metal consists of Ni, Ge, and WN_x. It is then annealed at 550°C to reduce the contact resistance. The fabrication of the MESFET is now finished except for the deposition of Aluminum interconnection layers. Totally, there are four layers of Aluminum interconnections. Between each metal deposition, there is always a layer of SiO_2. The metal is deposited by sputtering and the dielectrics by PECVD. Several spin-on glass layers are applied

to achieve a planar structure during this process. The connection between different metal layers is provided by etching small vias through the dielectric layer. Finally, a passivation layer is deposited on the top of all circuits for protection, with openings for bonding/probe pads (metal 4). The total thickness of the dielectric stack is about 5μm excluding the top overglass layer. The structure is schematically shown in Figure 2.14 [65]. Figure 2.15 is a picture showing one of the circuits fabricated through the MOSIS GaAs run.

The information on the process outlined above was gathered from the Vitesse Foundry Design manual [66] and other sources [65, 67]. The exact process may be slightly different. The important fact is that no gold is used in any metallization. Therefore, the circuit is able to sustain temperatures up to 530°C for several hours [68], which makes it possible for the MBE regrowth discussed later.

2.4.2 Device Performance

MESFET

The I-V curve of an enhancement-mode MESFET fabricated by the Vitesse HGaAs3 process is shown in Figure 2.16. The transistor has a gate length of 1μm and a gate width of 10μm. From Figure 2.16, the saturation voltage is significantly less than 1V and the breakdown voltage is only around 4V when the gate bias voltage is high. Such low breakdown voltage will place some constrain on the design of MQW modulator-based optoelectronic neuron circuits because normally 8 $\sim$ 10V bias is required in order to achieve high contrast ratio (4:1) for MQW modulators. Therefore, we need to use two MESFET's to support such high voltage by certain circuit design which will be addressed in Chapter 6. From Figure 2.16, the gate transconductance is measured to be 116mS/mm at V_{gs} = 0.5V and V_{ds} = 1.0V. The output transconductance, defined as $g_o = \Delta I_{ds}/\Delta V_{ds}$ in the saturation region, is found to be 2.5mS/mm for V_{gs} = 0.5V. This number determines how sharp the threshold response will be.

Compared with the MESFET's we built using our in-house process, the transistors by the Vitesse HGaAs3 process have much better uniformity. The testing result from

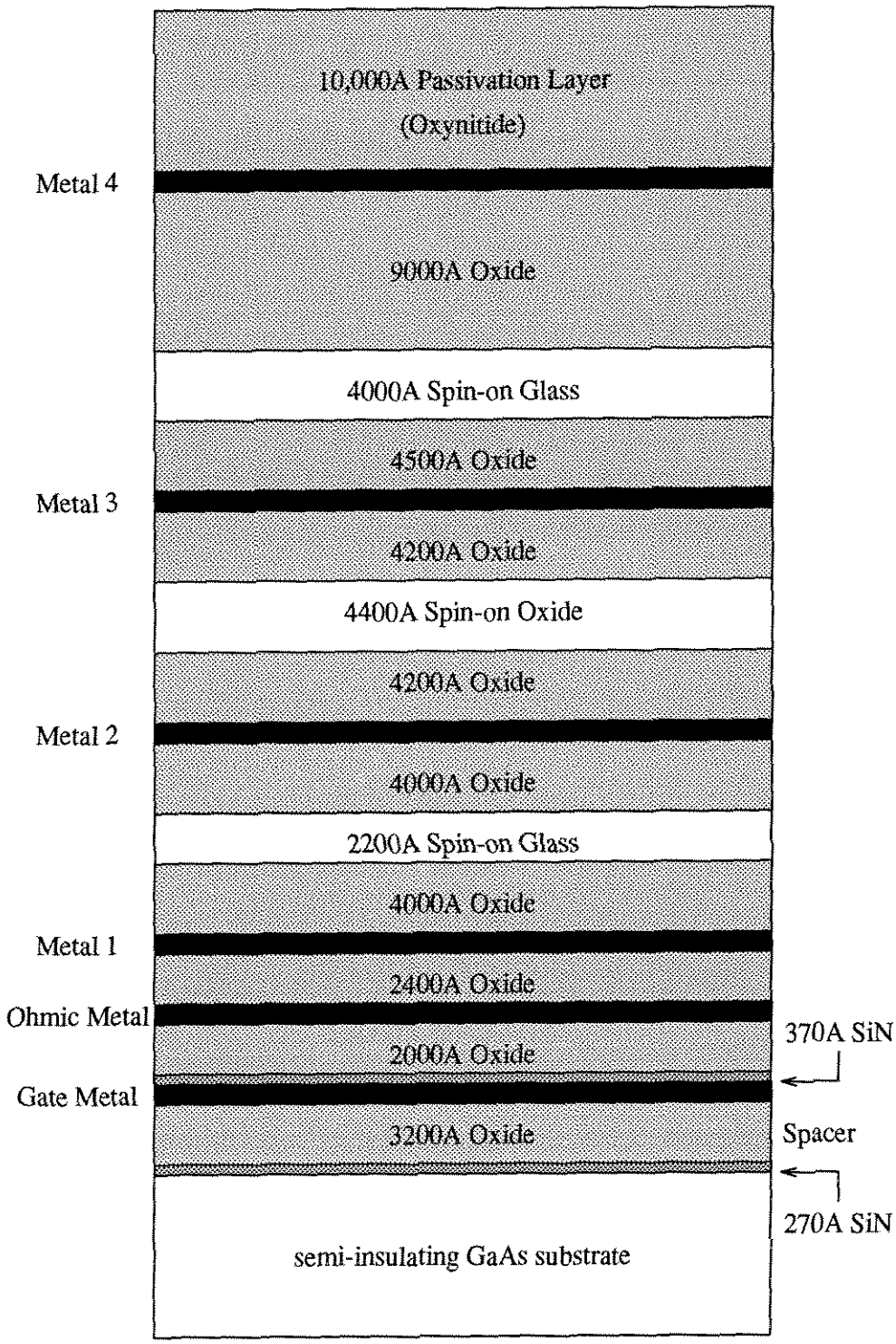


Figure 2.14: Schematic structure of the dielectric stack in the Vitesse HGaAs3 process.

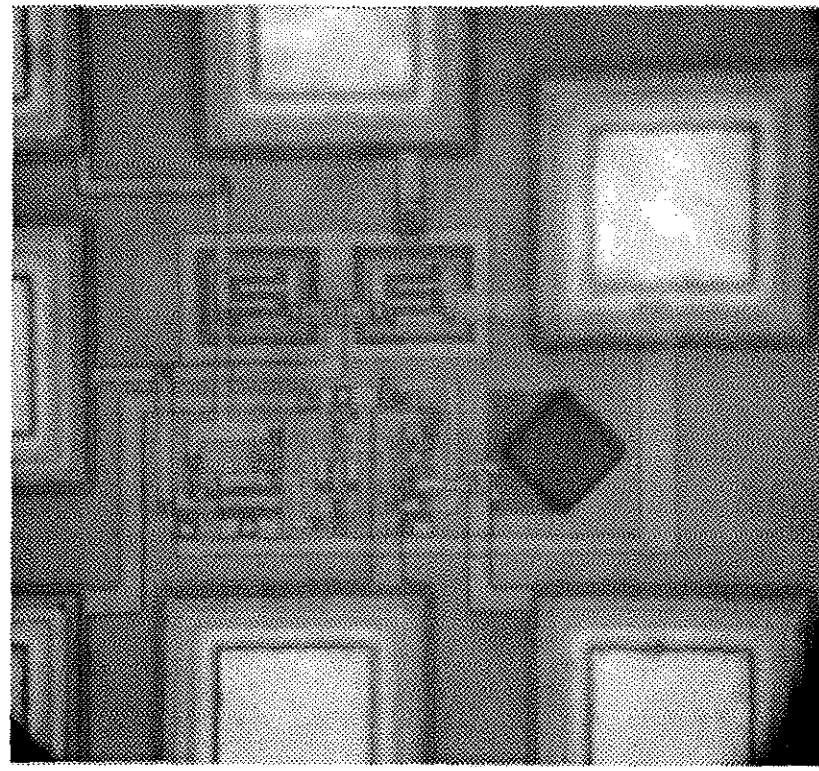


Figure 2.15: Picture of a testing circuit fabricated through the MOSIS GaAs run. The lozenge-shaped opening is the regrowth well and has an area of $25 \times 25\mu m^2$. The testing pads are $80 \times 80\mu m^2$.

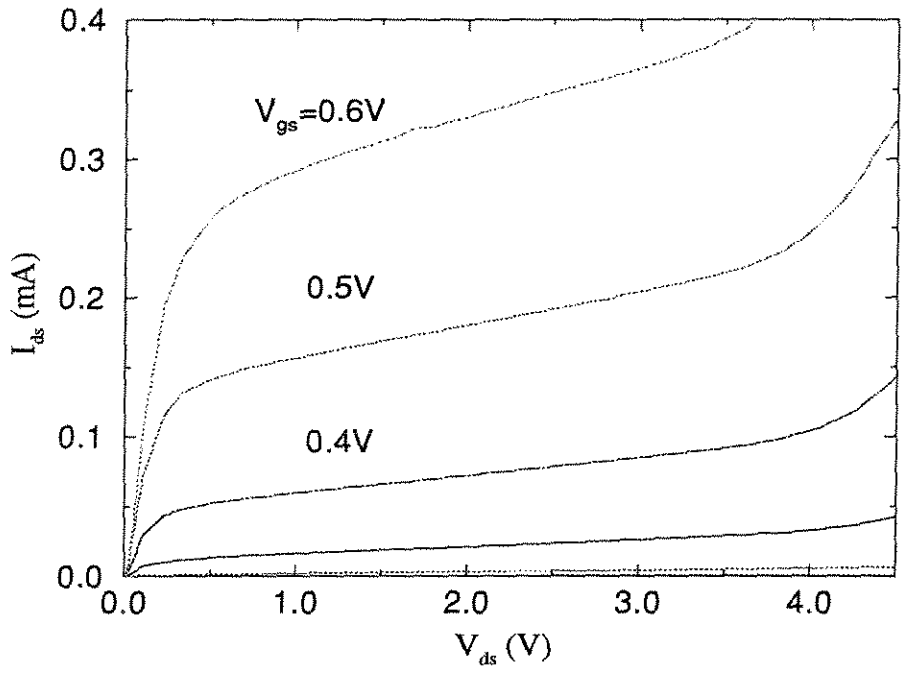


Figure 2.16: I-V characteristics of an enhancement-mode MESFET from MOSIS GaAs circuits. The gate has a length and width of 1μm and 10μm, respectively.

different chips for the same run is presented in Table 2.1 [65]. Each column represents the measurement of eight different FET's and is measured at a gate bias of 0.4V. The threshold voltage is measured following the definition used by Vitesse [66]. The small variation in V_T for the same type of FET's is due to some second-order effects. As we can see, the enhancement-mode FET shows higher non-uniformity compared to DFET's. This is because EFET's have a thinner channel and are therefore more sensitive than DFET's. The DFET's have similar uniformity compared to the transistors from the FET-SEED process.

Schottky Diode

Figure 2.17 shows the I-V curve of a Schottky diode. It is fabricated by connecting the drain and source contacts of an MESFET. Better uniformity can be achieved in this method compared to those that have only one ohmic contact. The main application of a Schottky diode in circuit design is as a level shifter to maintain the correct voltage levels. When the current going through a diode is higher than 0.2mA,

FET type	DFET	DFET	DFET	EFET	EFET
W/L (μm)	1/58	5/58	20/58	1/20	1/40
V_T(V)	-0.733	-0.64	-0.61	0.282	0.304
g_m(mS/mm)	158	49.6	13.8	79	86
g_o(mS/mm)	6.4	0.44	0.27	2.4	0.87
I_{ds}(mA)	5.06	1.03	0.228	0.108	0.271
σ(mA)	0.5 (10%)	0.074 (7%)	0.034 (15%)	0.033 (33%)	0.08 (30%)

Table 2.1: The parameters of the Vitesse depletion-mode FET's and enhancement-mode FET's.

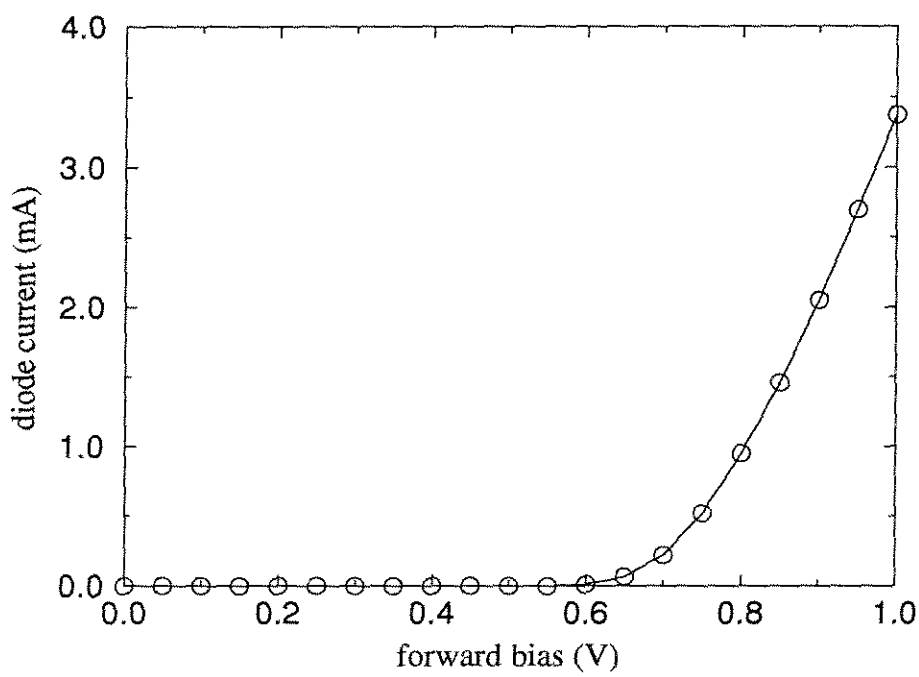


Figure 2.17: I-V curve of a diode level shifter. The Schottky contact is 1μm$\times$10μm.

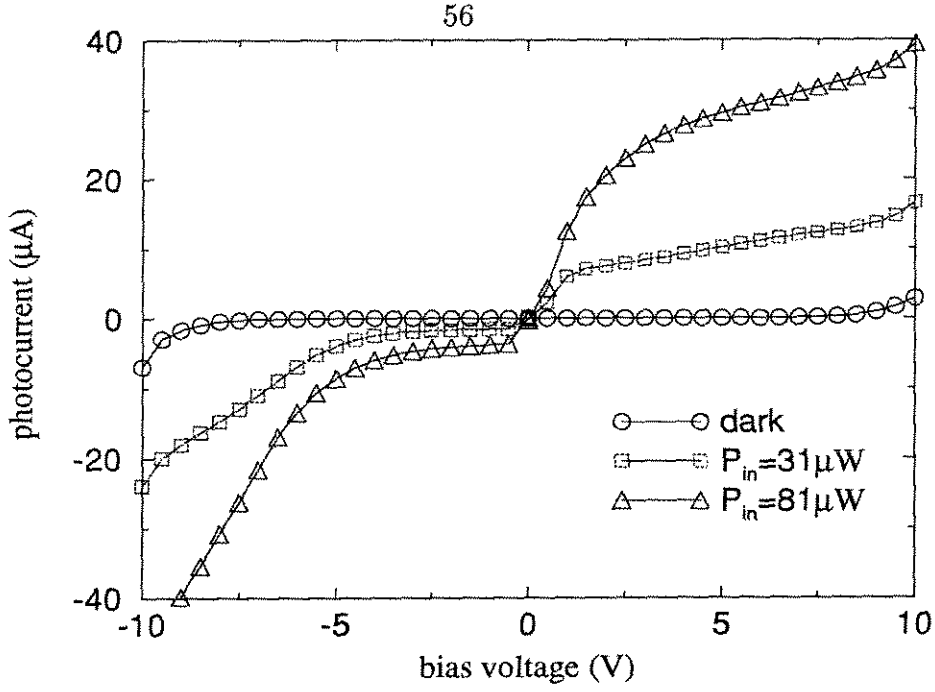


Figure 2.18: I-V curves of a 10μm$\times 15\mu$m MSM photodetector at different optical input.

the voltage drop across the diode will not show significant dependence on the current level, but stay around 0.8V. Therefore, it is a useful device when a certain voltage drop is required in the designed circuit.

MSM Photodetector

Metal-semiconductor-metal photodetectors can be fabricated using the HGaAs3 process by specifying two Schottky contacts on an n-channel. No ohmic contact is needed for this device. The I-V curves of a 10μm$\times 15\mu$m MSM detector is shown in Figure 2.18. The wavelength of the input light is 837nm and three I-V curves are plotted for the input power of 0, 31μW, and 81μW. The curves are slightly asymmetric for positive and negative biases. We attribute this behavior to the position of the input light beam being closer to one Schottky contact. For small bias voltage, the detector efficiency is very poor because the n-channel is highly doped due to the self-aligned process. As a result, the depletion region is very narrow and most photo-generated carriers are not within the high-electric-field depletion region and therefore not col-

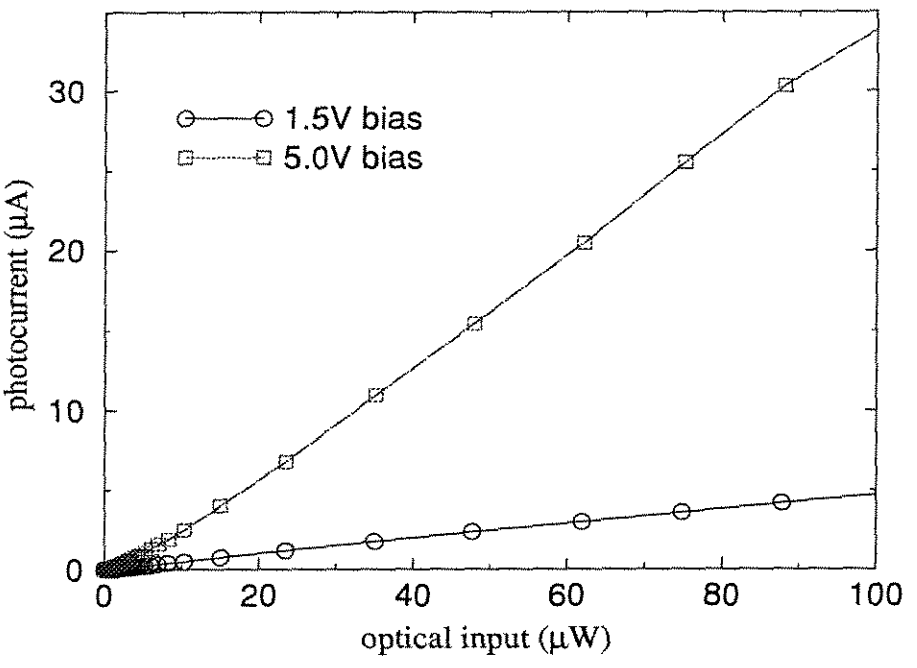


Figure 2.19: The relationship between the photocurrent and the input power of a MSM photodetector at different bias voltages.

lected. As the bias increases, the depletion region is widened and better efficiency is observed. However, the gain of an MSM detector is always less than 1 because there is no amplification mechanism. Since the absorption comes from bulk GaAs material, the spectral response of such devices shows a broad absorption edge and the cut-off wavelength is around 880nm [69].

Figure 2.19 shows the relationship between the optical input power and the photocurrent at two different biases: 1.5V and 5.0V. For 5.0V bias, the curve is slightly super-linear for small input powers. For 1.5V bias, it shows good linearity at a much lower efficiency. At 100μW, the responsivity is 0.34A/W for 5.0V bias, corresponding to an efficiency of 50% . As we discussed in Chapter 1, high density neuron arrays require high efficiency photodetectors. Since no gain can be achieved from MSM detector, neurons based on such devices would require amplification stages or large input power. In either case, the power consumption will increase and the neuron density will decrease. In Chapter 4, we will discuss a new-kind of high efficiency photodetector fabricated in the Vitesse HGaAs3 process.

2.4.3 MBE Regrowth

Even though the HGaAs3 process provides MESFET circuits with good uniformity and high gain photodetectors, optical output devices are not available. Therefore, in order to fabricate OEIC neuron arrays, certain post-processing is needed. The MBE regrowth method is developed for this purpose [64].

The first step for MBE regrowth is the removal of dielectric layers from the regrowth area. This can be down by asking Vitesse to do an RIE etch by specifying a G17 mask. However, it is found that the strong ion bombardment causes damage on the exposed GaAs surface, which will affect any material grown on top of them. So, the RIE etching is modified so as not to etch all the way down to the GaAs substrate. Approximately 2μm of dielectric layer is left. Another side effect of the RIE etching is the formation of some organic residue on the etched well. As a result, the first step after receiving the chip from MOSIS is to etch away this organic residue by long-time (several hours) oxygen plasma etching. The dielectric layer in the well is then etched away using wet chemical etchant, such as buffered-oxide etchant (BOE) or dilute HF solution. In this way, the regrowth well is protected from most damage.

The chip then goes through a final cleaning process before being loaded into the MBE chamber. This was done at MIT in Professor Clifton Fonstad's group. Because all the metals used in the HGaAs3 process are refractory metal, the fully-fabricated circuits can sustain a temperature up to 530°C for several hours without significant degradation [68]. For optoelectronic devices, the growth temperature is normally around 800°C, which is too high for circuits on the chip. So, the chamber temperature is reduced to 520°C. High quality GaAs material can still be obtained under such temperatures by properly adjusting the arsenic over-pressure [70]. The designed heterostructure for the output device, which can be either LED, MQW modulator, or even laser diode, is then grown on the whole chip. On the specified regrowth well, the material is a single crystal because it is grown on the GaAs substrate. On other area, the grown material is poly-crystalline.

The cross-section of the MOSIS chip after MBE regrowth is shown in Figure

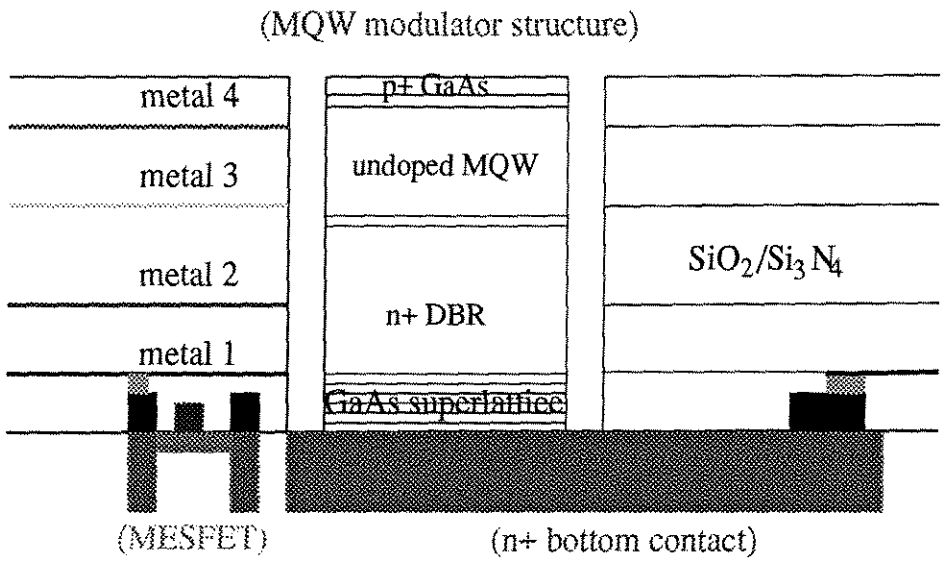


Figure 2.20: The cross-section of MBE regrowth structure of MQW modulator on MOSIS GaAs chips. Only single crystalline material is shown here.

2.20. Only the single crystalline GaAs material grown on the dielectric well is shown here. The poly-crystalline GaAs that is deposited everywhere else is not shown. The regrowth structure starts with a $GaAs/In_{0.2}Ga_{0.8}As$ superlattice buffer layer. The purpose of this layer is to prevent the propagation of defects from the substrate into the quantum well region. On the top of the superlattice is the actual heterostructure required for optical output devices, which is shown to be a MQW modulator in Figure 2.20. The top layer is a GaAs p-contact layer and the n^+ implant of the HGaAs3 process is utilized for the implementation of n-contact. In this way, the regrowth material is sitting on a n^+ region which can be electrically connected to other parts of the circuit. Therefore, only the top contact (p-type) is needed after the regrowth step in the post-processing. The thickness of the regrowth material is designed to be close to the total thickness of the dielectric stack, which is about 5μm. As a result of such planar design, the deposition of top metallization can be done more easily.

After the MBE regrowth is done, the chip is then sent back to Caltech for further processing. First, the poly-crystal material must be etched away. The applied etchant consists of H_3PO_4, H_2O_2, and CH_3COOH in a ratio of 1:1:3. Two etching steps

are applied in order to protect the single crystalline material. First, an area with $5 \sim 10\mu$m larger than the regrowth well in each direction is protected by photoresist and the unprotected material is etched. The next etching defines the real modulator mesa (several microns smaller than the regrowth well) and also serves as an isolation etch to separate the MQW region from any possible short circuit. Silicon nitride is then deposited on the surface using PECVD, followed by selectively etching nitride windows using CF_4 plasma etching. AuZn/Au is deposited as the p-metallization and also as the connection between the top contact to metal 4 pads. It is then annealed at 410°C to achieve ohmic contact. For MQW modulators, an anti-reflection coating is necessary to increase the contrast ratio. Following that, CF_4 plasma etching is applied to expose the contact pads.

In total, six masks are needed for the post-processing for MQW modulators: one mask for the etching of the regrowth well, two masks for the etching of the poly-crystalline material, one mask for plasma etching, one mask for metallization, and one mask for etching the AR coating to expose the contact pads. The processing steps used here are the same as what we described in Section 2.2.

By combining the advanced GaAs MESFET circuits from MOSIS/Vitesse and the MBE regrowth for optical output devices, complex optoelectronic circuits can be fabricated. With the help from the commercial foundry, it is possible to extend such a process from single chip scale to large wafer scale. In return, the cost can be dramatically reduced. However, such a process also has its disadvantages. The MBE regrowth of good quality material is not an easy task. In addition, the reliability of the AuZn/Au connection on aluminum metal 4 pads is still not tested. Even though some problems remain to be solved, the result is promising as we will present in Chapter 6 for the MBE regrowth method.

2.5 Summary

In this chapter, the fabrication methods of OEIC arrays were discussed. First, we briefly address the advantages of MESFET's over BJT's. The mechanism and mod-

eling of GaAs MESFET's were discussed. The three fabrication methods we have investigated were described in detail, including our in-house process, the AT&T FET-SEED process, and the MBE regrowth on MOSIS/Vitesse GaAs circuits. Each process has its own advantages and disadvantages. The in-house process has quick turn-around time and good flexibility. However, the uniformity is extremely poor and the devices we can fabricate are limited by the available facilities. The FET-SEED process can be used to build monolithic optoelectronic arrays with very good uniformity and requires no post-processing. Unfortunately, only depletion-mode FET's are available. Therefore, the circuit design is complicated and usually results in higher power consumption. MOSIS GaAs runs provide complex MESFET circuits with good uniformity. With MBE regrowth, various kinds of optical output devices can be integrated into the optoelectronic array. Therefore, it is a promising process for fabricating OEIC neuron arrays. However, the post-processing is more complicated and may suffer from reliability problems.

Chapter 3 Light-emitting Diodes

As an optical output device, the LED has several advantages: it is very easy to fabricate; it is an on-chip light source, which makes the optical setup simpler compared with modulators; no threshold current is required, which results in lower power consumption compared with laser diodes; the internal quantum efficiency can be close to 100% with careful design. Therefore, it is an important candidate for light sources in fabricating monolithic optoelectronic neuron circuits.

For high density neuron arrays, as we discussed in Chapter 1, the power consumption of each neuron should be minimized in order to increase the neuron density. As a result, for LED's integrated in optoelectronic neuron arrays, the available driving current is usually very limited. For example, assuming that the GaAs wafer has a power dissipation limit of $10W/cm^2$ and we want a neuron density of $10^4 neurons/cm^2$, with a driving voltage of 1.5V for the LED, the maximum driving current allowed is only $667\mu A$. At such low current level, the external efficiency of LED is very low due to surface recombination.

In this chapter, we will discuss the optimization of external efficiencies for LED's at low current levels. First, the theoretical results on the efficiency of double-heterojunction LED's are presented by summarizing published results, followed by the discussion of various factors that affect the LED efficiency. Finally, experimental results are presented and discussed.

3.1 Efficiency of LED

When a semiconductor p-n junction is under forward bias, electrons and holes are injected into the depletion region from its cathode and anode. These n-type and p-type carriers recombine with each other under proper conditions and the energy released from the recombination results in the generation of photons. As a result,

spontaneous emission of radiation is observed and the diode is therefore called a light-emitting diode (LED). Typically, such a phenomenon takes place in direct bandgap semiconductor materials where the conservation of both energy and momentum can be maintained during the recombination process.

In a homojunction structure, the injected electrons and holes will stay in the depletion region for a short time before they are swept away by the electric field. As a result, only part of the injected carriers can recombine with each other and emit photons while a large portion of the injected carriers are wasted. Therefore, the efficiency of such an LED is normally low due to the low injection efficiency. A double heterojunction (DH) structure is used to solve this problem. In this case, the active layer, where the carrier recombination takes place, is sandwiched between two higher bandgap materials. For example, a GaAs active layer between two $Al_xGa_{1-x}As$ barriers, which has a band-gap of [71]:

$$E_g(x) = 1.424 - 1.247x \ \ (\text{eV}). \tag{3.1}$$

Under such a structure, the electrons injected from the n-AlGaAs and the holes injected from the p-AlGaAs are confined within the lower-bandgap GaAs since the potential barrier on the GaAs-AlGaAs interface prevents them from escaping. As a result, they will stay in the active layer for a much longer time and the possibility of radiative recombination is greatly increased.

3.1.1 Efficiency of Double Heterojunction LED

The cross-section of a GaAs/AlGaAs double heterojunction LED is shown in Figure 3.1(a), where we assume the light is extracted from the p-layer. Its energy band diagram is illustrated in Figure 3.1(b), from which we can clearly understand the mechanism of carrier confinement by the two interface barriers. A slightly p-doped material is chosen as the active layer because in GaAs electron injection is more effective than hole injection. In calculating the efficiency of such an LED, several factors have to be considered, including the radiative recombination lifetime, τ_r, the

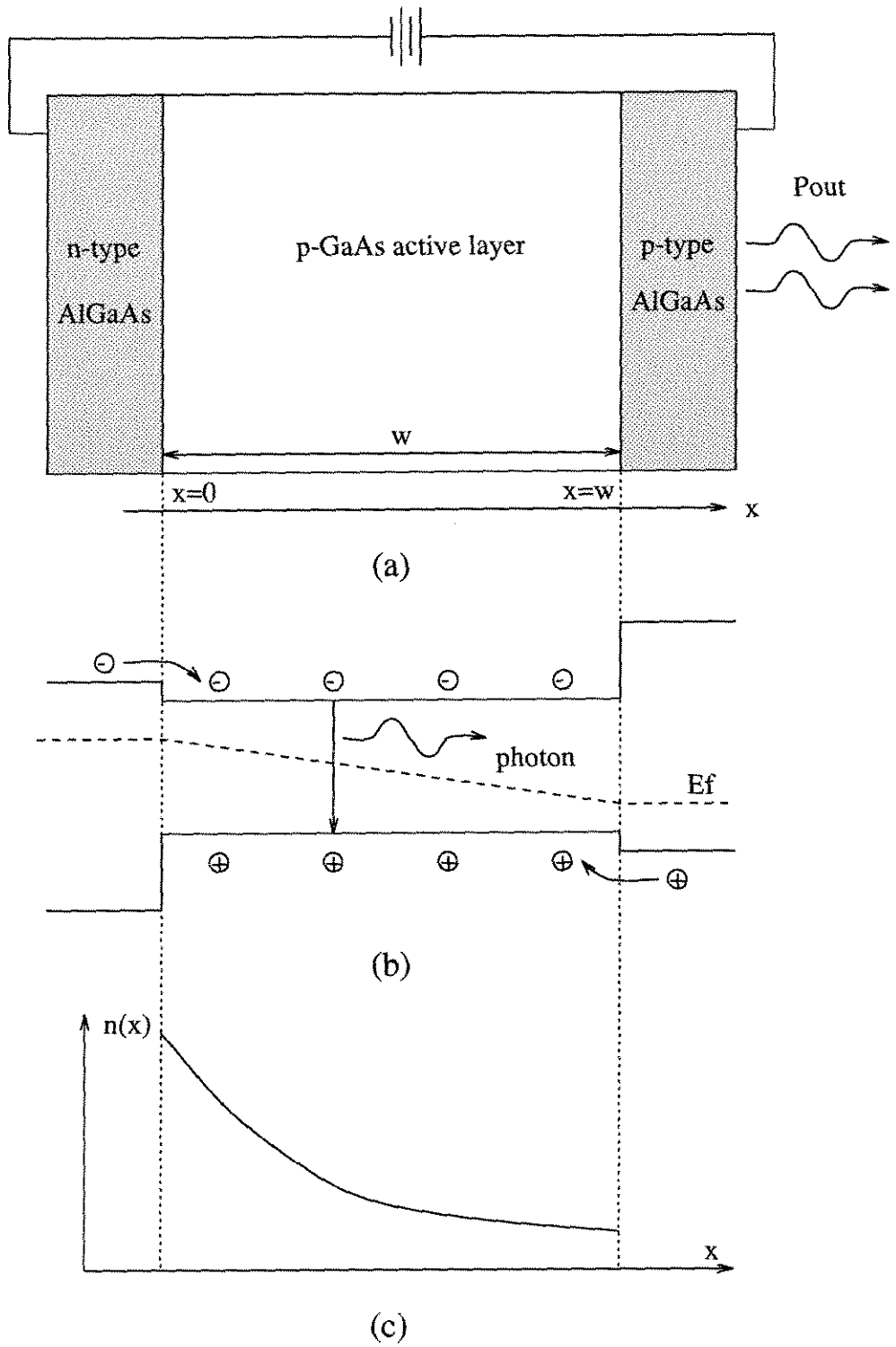


Figure 3.1: Double heterojunction LED. (a) The cross-section of a GaAs/AlGaAs DH LED. (b) Its energy band diagram showing the effect of carrier confinement. (c) The minority carrier concentration inside the active layer.

non-radiative recombination lifetime within the active layer, τ_{nr}, the interface recombination velocity, s, in cm/second, and the self-absorption effect of the generated photons by the active material, represented by the absorption coefficient, α.

Assuming a one-dimensional model which should be applicable if the LED has a large active-area-to-thickness ratio as exists in almost all of the LED's we fabricated and neglecting the nonradiative recombination in the space-charge region, we have the following continuity equation for minority carriers (electrons for Figure 3.1):

$$\frac{\partial n}{\partial t} = G_n - R_n + \frac{1}{q}\frac{\partial J_n}{\partial x} \tag{3.2}$$

where G_n and R_n are the generation and radiative recombination rate for the electron and J_n is the electron conduction current, given by:

$$J_n = q\mu_n nE + qD_n\frac{\partial n}{\partial x} \tag{3.3}$$

where μ_n and D_n are the mobility and diffusion constant of electrons. For the LED structure shown here, we assume that the electron drift current is much smaller than its diffusion current due to the low electron concentration within the active region. Therefore, we have:

$$J_n \simeq qD_n\frac{\partial n}{\partial x}\ . \tag{3.4}$$

At steady state, $\frac{\partial n}{\partial t} = 0$. In addition, we have $G_n = 0$ for generation and $R_n = n/\tau$ for recombination in the LED active layer, where τ is the minority carrier lifetime inside the active layer. Substituting all these conditions into Eq. 3.2, the one-dimensional continuity equation becomes:

$$D_n\frac{d^2n}{dx^2} - \frac{n}{\tau} = 0\ . \tag{3.5}$$

Eq. 3.5 can be rewritten as:

$$\frac{d^2n}{dx^2} - \frac{n}{L_n^2} = 0 \tag{3.6}$$

where $L_n = \sqrt{D_n\tau}$ is the electron diffusion length. The boundary conditions for the

structure shown in Figure 3.1(a) is given by [72]:

$$-\frac{dn}{dx}|_{x=0} = \frac{J}{qD_n} - \frac{s}{D_n}n(0) \tag{3.7}$$

$$\frac{dn}{dx}|_{x=w} = -\frac{s}{D_n}n(w) \tag{3.8}$$

where we have assumed equal surface recombination velocity s at both GaAs/AlGaAs interfaces and J is the injected current density. In addition, we have assumed a perfect carrier confinement so that there is no diffusion current on the interface. From Eq. 3.6-3.8, the electron concentration in the active layer can be solved as:

$$n(x) = \frac{JL_n}{qD_n}\left\{\frac{\cosh(\frac{w-x}{L_n}) + \frac{L_n s}{D_n}\sinh(\frac{w-x}{L_n})}{[(\frac{L_n s}{D_n})^2+1]\sinh\frac{w}{L_n} + \frac{2L_n s}{D_n}\cosh\frac{w}{L_n}}\right\} . \tag{3.9}$$

The result given by Eq. 3.9 is schematically illustrated in Figure 3.1(c). The average electron concentration is:

$$\bar{n} = \frac{1}{w}\int_0^w n(x)dx = \frac{J\tau_{eff}}{qw} \tag{3.10}$$

where τ_{eff} is the effective carrier lifetime and is given by:

$$\tau_{eff} = \tau\left\{\frac{\sinh\frac{w}{L_n} + \frac{L_n s}{D_n}(\cosh\frac{w}{L_n} - 1)}{[(\frac{L_n s}{D_n})^2+1]\sinh\frac{w}{L_n} + \frac{2L_n s}{D_n}\cosh\frac{w}{L_n}}\right\} . \tag{3.11}$$

When the interface recombination velocity is small and the active layer is thin, i.e., $L_n s/D_n \ll 1$ and $w \ll L_n$, Eq. 3.11 can be reduced to a simpler form:

$$\frac{1}{\tau_{eff}} = \frac{1}{\tau} + \frac{2s}{w} . \tag{3.12}$$

We can see that the interface recombination will reduce the effective lifetime of minority carriers, which will result in a lower quantum efficiency of the LED.

On the average, the injected electrons as given by Eq. 3.9 will remain free for the effective lifetime τ_{eff} before they get recombined. The recombination can be

either radiative in the active region or non-radiative on the interface and in the active layer. The generation rate of photon density is simply given by $n(x)/\tau_r$. However, not all of these generated photons can travel through the active layer and reach the surface. Some of them will be absorbed by the active layer since the wavelength of the generated photons is close to the absorption edge of the active layer. As a result, the photon density reaching the surface $x = w$ is $e^{-\alpha(w-x)}n(x)/\tau_r$. In order to calculate the output power of the LED, we need to know its emission spectrum and then integrate the output power over the whole spectrum. To make the matter even more complex, the absorption coefficient α also depends on the photon frequency. Therefore, the calculation will be very complex. For simplicity, we assume that all the output photons have a single frequency ν. Therefore, each photon has an energy of $h\nu$, where h is the Plank's constant. The output power of the LED is the sum of all photons generated per unit time:

$$P_{out} = h\nu \int_0^w \frac{n(x)}{\tau_r} e^{-\alpha(w-x)} dx \cdot A \tag{3.13}$$

where A is the area of the active layer. Substituting Eq. 3.9 into Eq. 3.13, we have the following result:

$$P_{out} = \frac{h\nu}{q} \cdot \eta_i \cdot I = \frac{h\nu}{q} \cdot \frac{\tau}{\tau_r} \cdot \eta_{DH} \cdot I \tag{3.14}$$

where I is the injected current, η_i is the reduced internal quantum efficiency due to interface recombination and self-absorption, and η_{DH} is the reduction factor for double heterojunction structure and is given by:

$$\begin{aligned} \eta_{DH}^{p} &= \frac{1}{2[(S^2+1)\sinh\frac{w}{L_n} + 2S\cosh\frac{w}{L_n}]} \\ &\cdot \left\{ \frac{1+S}{1-\alpha L_n}\left[e^{\frac{w}{L_n}(1-\alpha L_n)} - 1\right] + \frac{1-S}{1+\alpha L_n}\left[1 - e^{-\frac{w}{L_n}(1+\alpha L_n)}\right] \right\} \end{aligned} \tag{3.15}$$

where $S = L_n s/D_n$ is the normalized form for the heterointerface recombination velocity, s.

The above equation is for the case when photons are extracted from the p-side of

the LED. On the other hand, we can also extract photons from the n-side. In that case, Eq. 3.13 is changed to:

$$P_{out} = h\nu \int_0^w \frac{n(x)}{\tau_r} e^{-\alpha x} dx \cdot A \tag{3.16}$$

which results in a reduction factor of:

$$\begin{aligned} \eta_{DH}^{n} &= \frac{1}{2[(S^2+1)\sinh\frac{w}{L_n} + 2S\cosh\frac{w}{L_n}]} \\ &\cdot \left\{ \frac{1-S}{1-\alpha L_n}\left[e^{-\alpha w} - e^{-\frac{w}{L_n}}\right] + \frac{1+S}{1+\alpha L_n}\left[e^{\frac{w}{L_n}} - e^{-\alpha w}\right] \right\} . \end{aligned} \tag{3.17}$$

As we will see later in Figure 3.3, Eq. 3.15 and 3.17 give almost identical results when the active layer thickness is not very large. So, in the following discussion, we will simply assume that light is extracted from the p-side and use Eq. 3.15.

3.1.2 Radiative Recombination Lifetime

Eq. 3.14, 3.15 and 3.17 give the efficiency of a double heterojunction LED at a given injected current. The effect of interface recombination and self-absorption has been considered during the derivation of these two equations. In addition, the radiation recombination process is represented by a constant lifetime, τ_r. Generally, this radiative recombination lifetime depends on the doping level and the injected carrier concentration in the active layer. When the injected current changes, the injected carrier concentration will change according to Eq. 3.9, which results in a change of the radiative recombination lifetime of minority carriers [73]. Such change will affect the ratio of the effective lifetime to the radiative lifetime, τ/τ_r, and result in the change of LED efficiency.

From the theory of bimolecular recombination, the spontaneous emission rate inside the active layer is given by:

$$R_{sp} = B_{eff} np \tag{3.18}$$

where B_{eff} is the effective recombination probability constant, and n and p are the concentration of electrons and holes, respectively. Under external injection, we have $n = n_0 + \Delta n$ and $p = p_0 + \Delta p$, where n_0 and p_0 are the values at thermal equilibrium, determined by the doping level. Furthermore, from charge neutrality, $\Delta n = \Delta p$. As a result, the net rate of radiative recombination is:

$$R_{net} = B_{eff}[(n_0 + \Delta n)(p_0 + \Delta p) - n_0 p_0] \equiv \frac{\Delta n}{\tau_r} . \tag{3.19}$$

Therefore, the radiative recombination lifetime is given by:

$$\tau_r = \frac{\Delta n}{R_{net}} = \frac{1}{B_{eff}(n_0 + p_0 + \Delta n)} . \tag{3.20}$$

Under the situation illustrated in Figure 3.1(a), the active layer is p-doped. Assuming the doping concentration is N_A, we have $p_0 = N_A$ and $n_0 = n_i^2/N_A \ll N_A$, where n_i is the intrinsic carrier concentration. In GaAs, the value of n_i is $1.79 \times 10^6 \text{cm}^{-3}$ [36]. In return, Eq. 3.20 can be reduced as:

$$\tau_r \simeq \frac{1}{B_{eff}(N_A + \Delta n)} . \tag{3.21}$$

Substituting Eq. 3.10 into Eq. 3.21, we have:

$$\tau_r \simeq \frac{1}{B_{eff}(N_A + J\tau_{eff}/qw)} . \tag{3.22}$$

In most LED structures, the interface recombination is made smaller than the radiative recombination in order to increase its efficiency. Therefore, for simplicity, we can use τ_r instead of τ_{eff} in Eq. 3.22 and solve the equation. The result is given by:

$$\tau_r = \frac{\sqrt{N_A^2 + \frac{4J}{qwB_{eff}}} - N_A}{2J/qw} . \tag{3.23}$$

As an estimation, assuming $B_{eff} = 9.0 \times 10^{-11} \text{cm}^3/\text{sec}$ [73], $w = 0.5\mu\text{m}$, $N_A = 10^{16}\text{cm}^{-3}$, and $A = 15\mu\text{m} \times 15\mu\text{m}$, we have $\sqrt{4J/qwB_{eff}} = \sqrt{4I/AqwB_{eff}} = 4.97 \times$

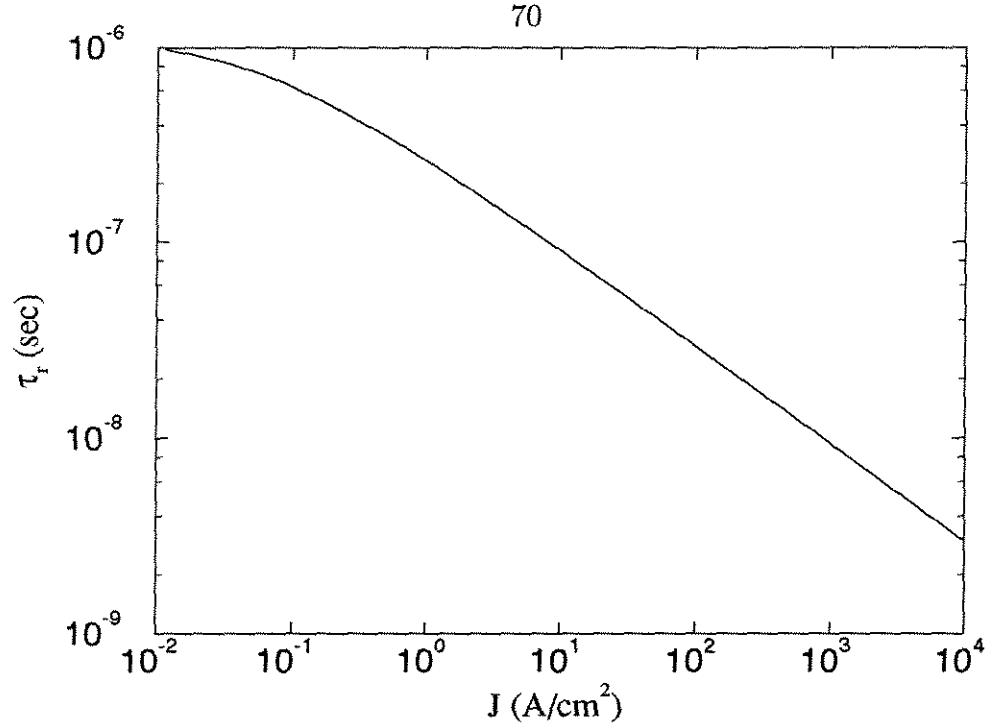


Figure 3.2: Simulation result of radiative recombination lifetime as a function of injected current density.

10^{17}cm^{-3} at $I = 100\mu\text{A}$, which means that even for small injection currents, the injected electron concentration Δn can be much larger than N_A. If so, Eq. 3.22 can be reduced further to:

$$\tau_r \simeq \sqrt{\frac{qw}{B_{eff}J}} \qquad \text{for } N_A \ll \Delta n \ . \tag{3.24}$$

Using the same parameters as above, we can estimate the value of τ_r to be 4.47×10^{-8}sec at $I = 100\mu$A and 1.41×10^{-8}sec at $I = 1.0$mA. Figure 3.2 plots the result of the radiative recombination lifetime given by Eq. 3.23 as a function of injected current density using the same set of parameters. Obviously, by increasing the driving current of a LED, we can decrease the radiative recombination lifetime, which means the radiative recombination process is now more competitive compared to the interface recombination. As a result, the quantum efficiency of the LED will be raised.

3.1.3 Other Factors Affecting LED Efficiency

Besides τ_r, the quantum efficiency of a LED also depends on other factors. In the following, we will discuss these factors.

Nonradiative Recombination

In the above calculation, the effect of nonradiative recombination inside the space-charge region is neglected. In a practical system, however, there always exists such nonradiative recombination processes. If we take the nonradiative lifetime, τ_{nr}, into consideration, the effective lifetime in Eq. 3.11 should be rewritten as follows:

$$\frac{1}{\tau} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} . \tag{3.25}$$

Similarly, for small interface recombination velocity and thin active layer, Eq. 3.12 becomes:

$$\frac{1}{\tau_{eff}} = \frac{1}{\tau_r} + \frac{1}{\tau_{nr}} + \frac{2s}{w} . \tag{3.26}$$

Obviously, the existence of nonradiative recombination in the active layer will reduce the effective lifetime and therefore reduce the quantum efficiency of a LED, which is proportional to τ/τ_r as given by Eq. 3.14.

There are several possible nonradiative recombination processes in the LED structure. If traps are induced during the material preparation, either electrons or holes may be trapped and then recombine with the opposite type of carriers. Because the transition probability depends on the potential difference between the energy level of the traps and the semiconductor energy band edge, such traps are very effective nonradiative recombination centers if their energy level is near the middle of the band-gap. Fortunately, with the modern material growth techniques, the concentration of these so-called deep level traps is very small and we can neglect this type of nonradiative recombination.

For LED's, the most significant nonradiative recombination process is probably the Auger recombination, where the energy released from the carrier recombination

is transformed into the exciting of the energy level of other carriers, rather than the generation of photons in a radiative process. Many different Auger processes are possible [74]. For a three-particle Auger process, if the transition involves donor or acceptor states, the requirement of momentum conservation can be eased due to the uncertainty in the momentum of donor or acceptor states. Therefore, such a process is the most likely one in a LED structure. It is found that the lifetime of the Auger process at low injection level decreases as the doping level increases [72]. Therefore, in order to minimize the effect of Auger recombination, the doping level of the active layer should be reduced to less than 10^{17}cm^{-3}. Under such conditions, the Auger recombination process is normally slower than the interface recombination and can be neglected.

In summary, if we prepare the LED material carefully and the active layer is not highly doped, the nonradiative recombination process is typically very slow and can be neglected compared to interface recombination.

Interface Recombination

As we mentioned earlier, the interface recombination will reduce the effective lifetime and therefore reduce the quantum efficiency of an LED. Interface recombination centers come from the defects during the growth of the LED heterostructure. When one material is grown over another, the difference between their lattice constants, a, will result in strain and defects on the heterointerface, which act as nonradiative recombination centers [75]. The interface recombination velocity is approximately proportional to the mismatch in lattice constant and can be expressed as:

$$s \approx 2 \times 10^7 \cdot \frac{\Delta a}{a} \quad \text{cm/sec.} \tag{3.27}$$

For GaAs/AlGaAs system, the mismatch in lattice constant is very small and the interface recombination velocity can be as low as 10^3cm/sec [72]. On the other hand, for GaAs surface, the dangling bonds of Ga and As atoms result in lots of recombination centers; and s can be as high as 10^6cm/sec. Such high recombination velocity results

in a serious surface nonradiative recombination problem, which will be addressed in Section 3.3.

As an estimation, the nonradiative lifetime due to interface recombination is roughly $w/2s$. Assuming $s = 10^3$cm/sec and $w = 0.5\mu$m, we have $\tau_{nr} = w/2s \approx 2.5 \times 10^{-8}$sec. Comparing with what we calculated earlier, such nonradiative recombination is at least comparable to the radiative recombination for low LED driving currents. Therefore, the efficiency of a LED is significantly limited by the interface recombination process.

Thickness of the Active Layer

In order to reduce the effect of interface recombination, we want to increase the thickness of the active layer so that the nonradiative recombination lifetime $\tau_{nr} = w/2s$ is larger than the radiative recombination lifetime. However, if w increases, there are two undesirable effects that tend to reduce the efficiency of a LED. First, a larger w means the extracted light has to travel through a longer distance before escaping from the surface. So, the self-absorption effect is more significant and a smaller portion of the generated photons can be extracted. In return, even though the quantum efficiency of the LED is raised due to a thicker active layer, the external efficiency may not increase. Second, for the same injected current density, larger w means the injected carriers can be distributed over a larger region, which means the injected carrier concentration is lower. In return, the radiative recombination becomes slower as given in Eq. 3.23. As a result, the radiative recombination is less competitive compared to interface recombination.

Since there are two opposite consequences for any change in the thickness of the active layer, we would expect that there exists an optimum thickness for a given set of parameters. In Figure 3.3, we plot the result of Eq. 3.15 and Eq. 3.17 by using the result of Eq. 3.23 to calculate the minority carrier diffusion length. The following parameters are assumed: $N_A = 10^{16}$cm^3, $s = 1000$cm/sec, $\alpha = 9000$cm^{-1}, $D_n = 80$cm^2/sec [72], $B_{eff} = 9.0 \times 10^{-11}$cm^3/sec, $A = 15\mu$m$\times 15\mu$m. The results are plotted at two different injection currents: $I = 100\mu$A and $I = 1$mA, for light

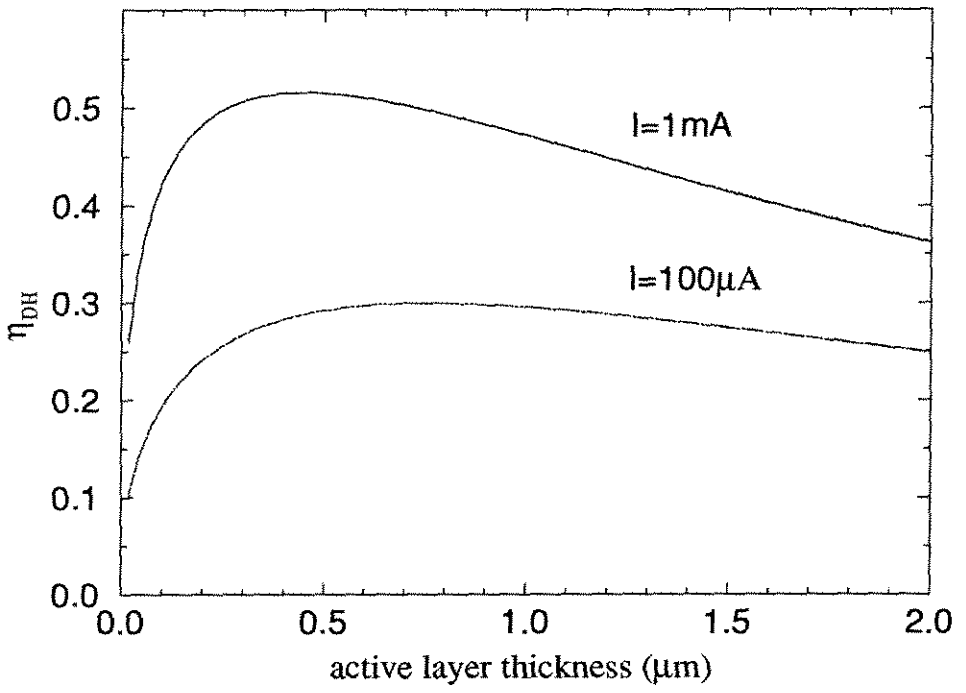


Figure 3.3: Simulation result of LED efficiency as a function of the thickness of active layer. Each curve has actually two almost overlapping curves, corresponding to light being extracted from the p-side and the n-side, respectively.

extraction from either the p-side or the n-side. We can see that no big difference exists between extracting light from the n-side and from the p-side. The optimum active layer thickness is found to be around $0.4 \sim 0.7\mu$m, depending on the current level. Since we are interested in LED's operating at low driving currents in order to reduce their power consumption, the active layer is designed to have a thickness of about 0.6μm in the experiments described later.

Because of the dependence of radiative recombination lifetime on the driving current, the LED efficiency is not a constant for different current levels. The relationship between the internal quantum efficiency and the driving currents is plotted in Figure 3.4 using the result given by Eq. 3.14. The parameters are the same as those used in Figure 3.3 except we have assumed $w = 0.5\mu$m and $\tau_{nr} = 500$nsec. The simulation result is plotted as a solid line. As a comparison, data derived from experimental measurement are also presented here. These data are calculated from the measured external efficiency of quantum well LED and single Zn diffusion LED, which will be given by Figure 3.14 in Section 3.4.2. To estimate the internal efficiency, we assume

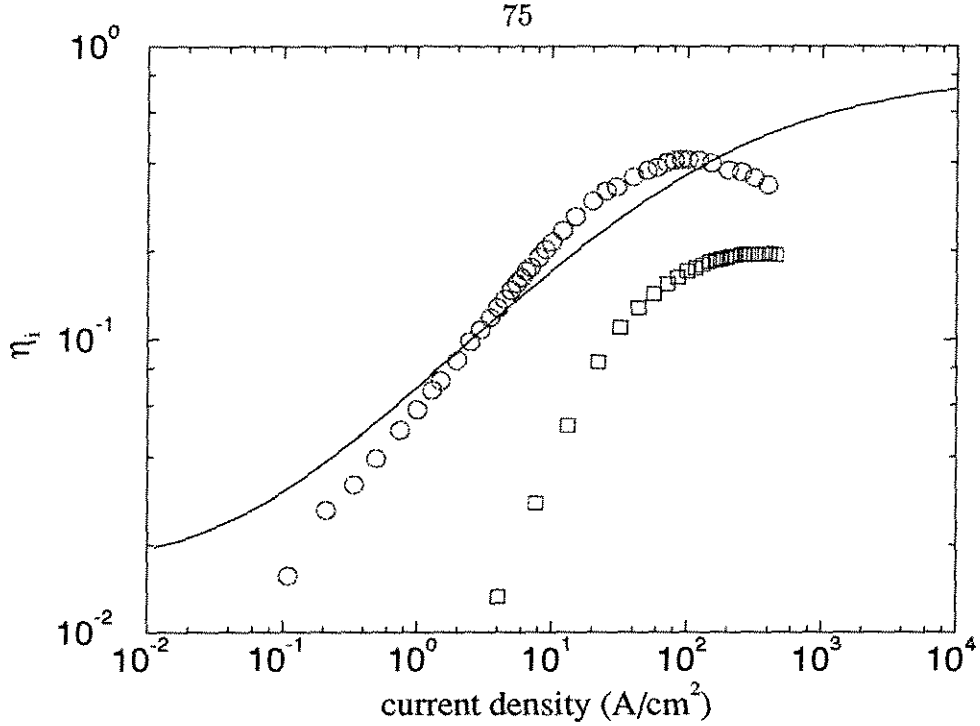


Figure 3.4: Simulation result (solid line) of LED efficiency as a function of the driving current. Light is extracted from the p-side. In comparison is the result from experimental measurement: circles are for quantum well LED and squares for Zinc diffusion LED.

that only 2% of light is extracted due to the critical angle, which will be discussed in Section 3.2.2. In addition, for the Zn diffusion LED, a 30% Fresnel loss is included because no anti-reflection coating was deposited during the fabrication.

3.2 Light Collecting Efficiency

In the last section, we discussed the efficiency of a double heterojunction LED by assuming that all photons generated can be extracted except those being absorbed by the active layer when traveling toward the surface. Unfortunately, this is not the case in a practical system. Not all photons will reach the surface and not all photons reaching the surface can be collected. In this section, we will discuss those factors limiting the collection of generated photons, including the isotropic property of spontaneous emission, the mismatch of refractive index between GaAs and air, and the blocking effect of the contact metal. Due to the limitation in collecting the

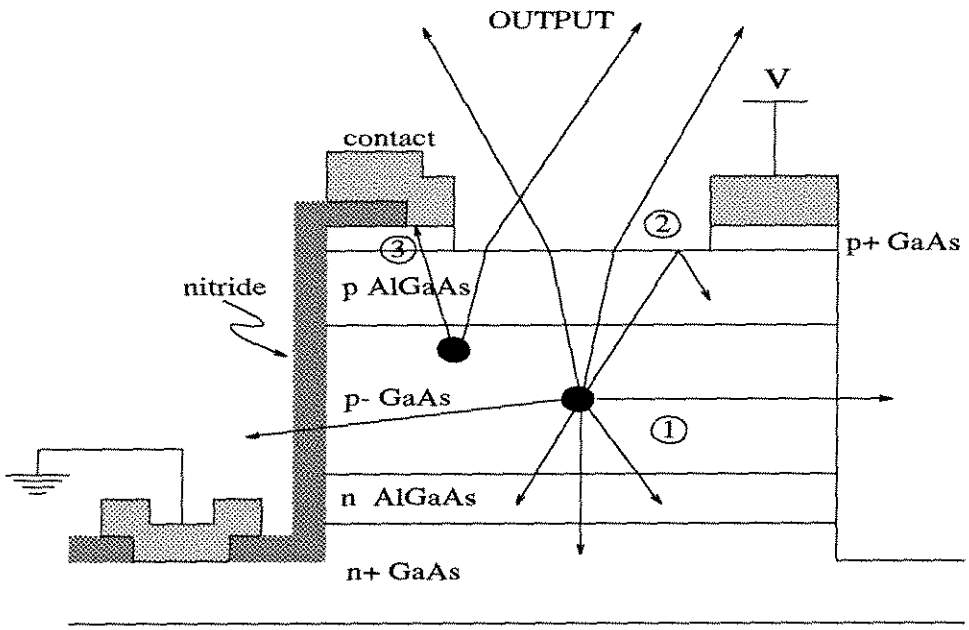


Figure 3.5: Cross-section of a surface-emitting mesa-type LED used in our process. The factors limiting its photon collecting efficiency are illustrated: (1) spontaneous emission, (2) refractive index mismatch, and (3) metal blocking.

generated photons, the output power of a LED, given by Eq. 3.14, should be rewritten as:

$$P_{out} = \frac{h\nu}{q} \cdot \frac{\tau}{\tau_r} \cdot \eta_{DH} \cdot \eta_{ex} \cdot I \tag{3.28}$$

where η_{ex} is the photon collecting efficiency.

Since the LED is to be used in fabricating a two-dimensional optoelectronic neuron array, its output must be perpendicular to the surface. The vertical mesa-type LED, shown in Figure 3.5, is used for this purpose. Figure 3.5 is not plotted in the same scale in horizontal and vertical dimension. Typically, the total thickness is around 1μm and the horizontal size is around $10 \sim 30\mu$m. As discussed in the last chapter, the mesa is defined by chemical wet etching. Ohmic contacts are then deposited as the top and bottom contact. Such a structure can be easily fabricated using our facility. However, there are two main disadvantages: the large surface current, which will be discussed later, and the lack of current confinement, which results in the blocking of emitted light by the top contact. A Burrus-type LED [72] can reduce

these two undesirable effects. However, the fabrication of Burrus-type LED requires lithography and processing on both side of the wafer, which is beyond the capability of our facilities. Therefore, we will focus on the optimization of the mesa-type LED.

3.2.1 Isotropic Emission of LED

Because the light emission in a LED is a spontaneous process, the generated photons are emitted in all direction. Therefore, at most 50% of the light can reach the top surface, and the other 50% will travel down toward the substrate and be absorbed there. In addition, due to the finite size of a LED mesa, some of the light will escape from the edge of the mesa. To make the situation even worse, because the refractive index of the AlGaAs confinement layer is smaller than that of GaAs, the resulting waveguide effect tends to guide more light toward the edge. In actual testing, we can observe very bright luminescence from the perimeter of the LED mesa. For a surface-emitting LED, such effect is undesirable. In order to collect the light emitted from the mesa edge, small 45°C mirrors can be etched using anisotropic etching such as RIE [69]. As a price for increasing the external efficiency, the fabrication process becomes more complex.

For those light rays emitted toward the substrate, if we can insert a mirror between the LED structure and the absorbing substrate, it can be reflected back toward the surface. By doing so, we can double the external efficiency of the LED if there is no self-absorption in the active layer. Such a mirror can be provided by the growth of a distributed Bragg reflector (DBR), which consists of N pairs of alternative layers of low-refractive-index material and high-refractive-index material. For example, we can use $Al_xGa_{1-x}As$ with two different Aluminum percentage, which has a refractive index of [71]:

$$n(x) = 3.590 - 0.710x + 0.091x^2 \ . \tag{3.29}$$

From fundamental optics, when light travels from one material to another material with different refractive index, some portion of the light will be reflected back. The

reflectivity for normal incident beam is given by the Fresnel formula:

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2}\right)^2 \tag{3.30}$$

where n_1 and n_2 are the refractive indexes of the two materials. In a DBR, the transmitted light through the first interface will run into the second interface after traveling through some distance. Some of it will be reflected back and interfere with the light reflected on the first interface. Totally, we will have $2N$ of such reflection interfaces. If we design the thickness of each layer carefully so that all these reflected lights from different interfaces have the same phase, they will add up constructively and we have a good mirror. To satisfy the phase requirement, each layer must has a thickness equal to one quarter of the wavelength in that material. In return, light passing through one pair of the material twice would gain a phase of 2π. For N-pair DBR with such design, the reflectivity is given by [76]:

$$R \simeq \left(\frac{n_1^{2N} - n_2^{2N}}{n_1^{2N} + n_2^{2N}}\right)^2 . \tag{3.31}$$

Obviously, the more pairs the reflector has, the higher its reflectivity will be. In addition, since the thickness of each layer is designed for one particular wavelength, the reflector has a limited range within which it has high reflectivity. By increasing the number of DBR pairs, the high-reflection range also increases. However, the more pairs we need, the longer the growth takes and the more difficult the fabrication becomes. Typically, a DBR with 16 pairs of $Al_{0.11}Ga_{0.89}As$ and AlAs can give a reflectivity of about 99% and the high reflection range is almost 100Å, which is good enough for a LED. We do not want to use GaAs because it will absorb the generated light. In Chapter 5, the result of one such DBR is plotted in Figure 5.12.

3.2.2 Mismatch of Refractive Index

The most important factor limiting the light collecting efficiency is the mismatch of refractive index between GaAs and the air. As we will see next, this problem

significantly reduces the external efficiency of a LED. There are two consequences due to the index mismatch: surface back reflection and the critical angle.

Fresnel Loss (Surface Back Reflection)

As we mentioned in the discussion of DBR design, the interface of two materials with different refractive indexes will reflect some portion of the light. For a GaAs/AlGaAs LED as shown in Figure 3.5, we have two such interfaces: the GaAs/AlGaAs interface and the AlGaAs/air one. From Eq. 3.30, the corresponding reflectivity is 0.086% and 29.6% for normal incidence. Here, we have assumed a 30% Aluminum concentration ($x = 0.3$). For the incident light that is not normal to the surface, the reflectivity is even higher.

To solve the back reflection problem, an anti-reflection (AR) coating can be deposited on the AlGaAs surface. By carefully choosing the refractive index and the thickness of the coating layer, the resulting reflectivity can be reduced even down to zero. For normal incidence, the reflectivity of a coated AlGaAs/air interface can be written as [76]:

$$R = \frac{n_{AR}^2(n_{AlGaAs}-1)^2\cos^2(\frac{2\pi n_{AR}d}{\lambda}) + (n_{AlGaAs}-n_{AR}^2)\sin^2(\frac{2\pi n_{AR}d}{\lambda})}{n_{AR}^2(n_{AlGaAs}+1)^2\cos^2(\frac{2\pi n_{AR}d}{\lambda}) + (n_{AlGaAs}+n_{AR}^2)\sin^2(\frac{2\pi n_{AR}d}{\lambda})} . \qquad (3.32)$$

By setting $R = 0$, we have $n_{AR} = \sqrt{n_{AlGaAs}}$ and $d = (2k+1)\lambda/4\sqrt{n_{AlGaAs}}$, where $k = 0, 1, 2, \cdots$. This is the optimum index and thickness of the AR coating. For $x = 0.3$, we have $n_{AlGaAs} = 3.39$, so the optimum condition is $n_{AR} = 1.84$ and $d = (2k+1) \times 1169\mathring{A}$ for $\lambda = 860$nm. Such refractive index is close to those of some dielectric materials, especially Al_2O_3, so the AR coating can be provided by depositing a thin layer of dielectric material.

Critical Angle

The anti-reflection coating can effectively reduce the back-reflection to zero for the normal incidence light . However, for the light incident with an angle, θ, from the surface normal, the reflection still exists. The larger θ is, the higher the reflectivity

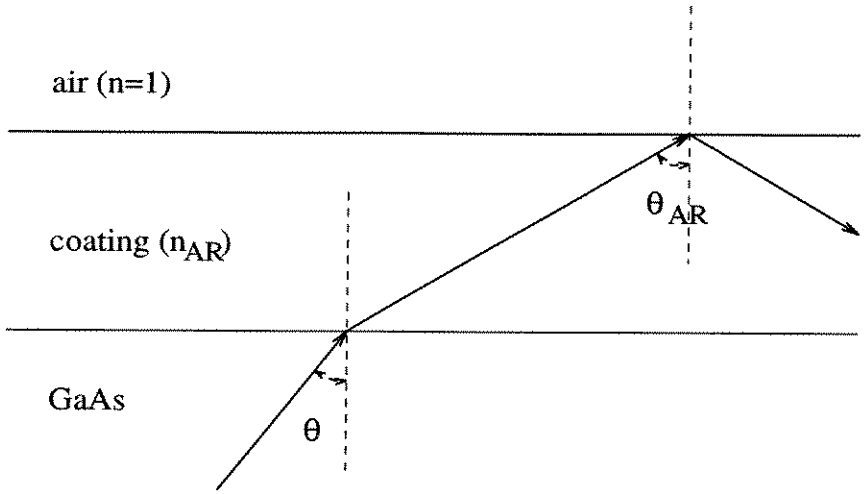


Figure 3.6: The effect of additional coating on the critical angle.

will be. When θ is larger than a certain angle, θ_c, the transmission is reduced to zero and all light will be reflected back. This angle is called the critical angle and is given by:

$$\theta_c = \sin^{-1}(n_{air}/n_{GaAs}) = 16.2^\circ \ . \tag{3.33}$$

No matter how many coating layers there are between the GaAs active layer and the air, any photon that reaches the GaAs interface with an angle larger than θ_c will be totally reflected and unable to escape into the air. This can be explained using Figure 3.6.

In Figure 3.6, given any $\theta > \theta_c$, from fundamental optics we have:

$$\sin\theta_{AR} = \frac{n_{GaAs}}{n_{AR}}\sin\theta > \frac{n_{GaAs}}{n_{AR}} \cdot \frac{1}{n_{GaAs}} = \frac{1}{n_{AR}} = \sin\theta_{c_{AR}} \ . \tag{3.34}$$

Therefore, θ_{AR} is larger than the critical angle of the coating material. As a result, light will be reflected back even with the existence of AR coating.

The effect of the critical angle is a significant reduction on the LED external efficiency because among all the photons emitted by the LED, only those with $\theta < \theta_c$ can escape into air. The percentage can be calculated from the spherical surface area of all $\theta < \theta_c$:

$$\eta_{ex1} = \frac{\int_0^{\theta_c} 2\pi\sin\theta d\theta}{\int_0^{\pi} 2\pi\sin\theta d\theta} = \frac{2\pi(1-\cos\theta)}{4\pi} = 1.985\%. \tag{3.35}$$

This means that even with 100% internal quantum efficiency and 100% transmission of anti-reflection coating for all photons incident with $\theta < \theta_c$, the external efficiency of the mesa-type LED is always less than 2%.

Obviously, among all the limitation discussed so far, this is the most severe one in reducing the LED efficiency. To solve this problem requires some treatment on the emitting surface. One solution is to create an integrated lens on the surface. This can be done by several methods, such as photoelectrochemistry etching [77] where the etching rate is controlled by the light intensity on the surface. By producing spatial intensity variation on the LED surface, lenses with smooth surface can be obtained. The other solution is to create small nanotexture on the emitting surface [78], thereby increasing the possibility for photons to escape from the high refractive index material (GaAs or AlGaAs).

3.2.3 Light Blocking by Metal Contact

In a mesa-type LED, in addition to the two limiting factors discussed above, the LED output power may be reduced further by the top ohmic contact. Because the top contact consists of opaque metal, photons generated under the contact region will be blocked by such contact metal. The result is lower LED output power.

As an example, let's look at a circular-shaped LED with ring contact as shown in Figure 3.7. The diameter of the LED mesa is $2r$ and the contact metal has a width of d. Assuming a uniform distribution of injected current over the whole area, the reduction factor due to metal blocking will be:

$$\eta_{ex2} = \left(\frac{r-d}{r}\right)^2 = \left(1-\frac{d}{r}\right)^2 . \tag{3.36}$$

For $r = 15\mu$m, $d = 5\mu$m, only 44% of the LED output can be collected through the open window.

Obviously, we can reduce the contact width to increase the external efficiency. However, by doing so, we are taking the risk of increasing the ohmic contact series resistance, which has two undesirable effects. As the series resistance increases, to

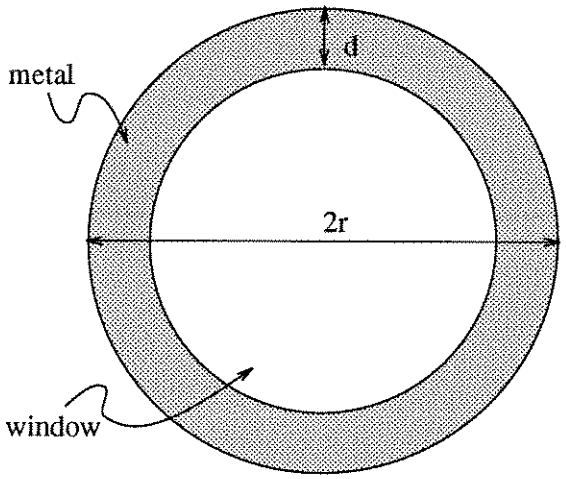


Figure 3.7: The top view of a circular LED with ring contact.

inject the same amount of current, higher voltage is required, which raises the power consumption. Furthermore, as we observed in actual testing, the increase of contact resistance tends to affect more the radiative recombination current than the nonradiative recombination current along the surface, which will be discussed in next section. In return, the LED efficiency at low current level will suffer a lot.

The other method to solve the metal blocking problem is to increase the mesa size while maintaining the same contact width. However, when deriving Eq. 3.36, we have assumed that the current is distributed uniformly over the whole area. In the actual device, this is not true. Due to the finite conductance of the top layer, the injected current density tends to crowd near the contact metal and decrease as it moves toward the center of the mesa. So, if r is too large, we will not get any significant injected current density at the center. Even though the light generated there is not blocked, the efficiency is still low because most light is generated under the metal contact and is therefore blocked as shown in Figure 3.8(a).

The problem of metal blocking comes from the fact that no current confinement is provided, which is different from the case of a Burrus-type LED [72]. Therefore, in order to reduce the effect of metal blocking, some sort of current confinement is needed. Figure 3.8(b) and (c) shows two methods to do that.

Figure 3.8(b) is a double Zn diffusion LED briefly mentioned in the last chapter. The starting structure is a n-i-n double heterostructure. By doing Zn diffusion, the

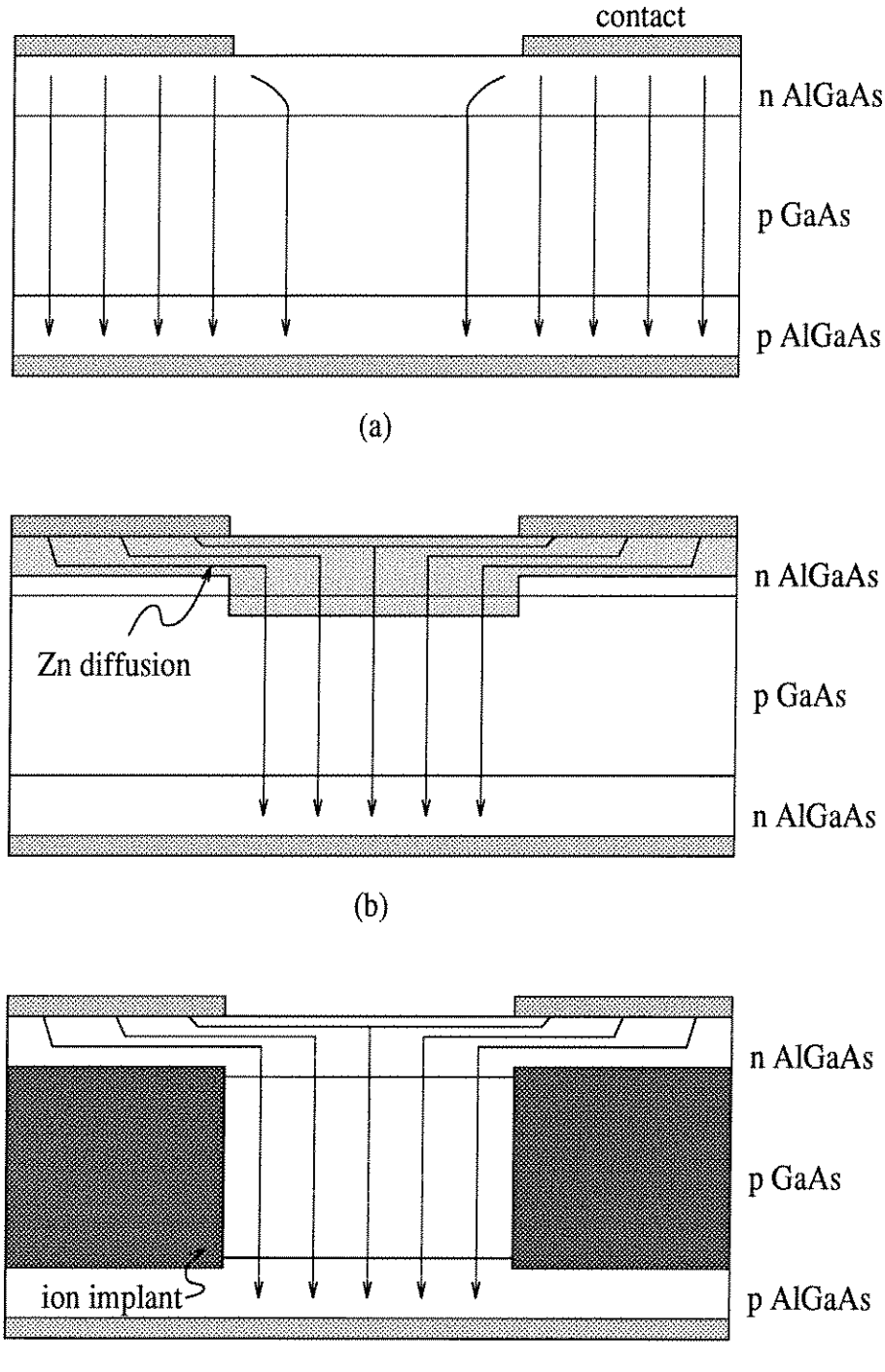


Figure 3.8: Current confinement for LED. (a) No confinement is provided, so most light is blocked. (b) Confinement by double Zn diffusion. (c) Confinement by ion implantation.

top layer is converted into p-type and we can obtain the p-i-n diode. The first diffusion defines the contact region and the second diffusion is applied slightly down to the p^--doped GaAs active layer in a smaller area. Outside this area, the diffusion front is carefully designed so that a p-n-p-n blocking thyristor is achieved. In the middle, the p-i-n diode will force all injected current to go through this area. By depositing the contact metal on the first diffusion region but outside the second diffusion area, the metal blocking effect can be reduced. In actual processing, the order of the two diffusion is reversed because of the diffusion mask we used, but the mechanism remains the same. The important step here is to control the first diffusion front so that it does not touch the p-GaAs active layer. Otherwise, no thyristor can be achieved to block the current and we are back to the situation of Figure 3.8(a).

Figure 3.8(c) provides similar current confinement by buried ion implantation. By choosing the correct energy and dose of the implanted ions, we can selectively damage some part of the GaAs active layer and change it into non-conducting material while maintaining the conductivity of the top layer. As a result, the injected current is only allowed to go through the undamaged region and light will be generated from there. Since the top layer is still conducting, an ohmic contact can be deposited on top of the damaged region and the metal blocking effect is also reduced. The testing result of LED's fabricated using the two methods shown in Figure 3.8(b) and (c) will be presented later in this chapter.

3.3 Effect of Surface Recombination

From Figure 3.4, the quantum efficiency, η_i, of a $15\mu m \times 15\mu m$ LED at 1mA driving current, which equals to a current density of $44 A/cm^2$, is about 51.3%. In the LED's we fabricated, there is typically no DBR mirror, so the external efficiency of such LED's is:

$$\eta = \frac{P_{out}}{I} = \frac{h\nu}{q} \cdot \eta_i \cdot \eta_{ex1} \cdot \eta_{ex2} \tag{3.37}$$

where η_{ex1} and η_{ex2} are the efficiency reduction factors due to refractive index mismatch and metal blocking, respectively. As discussed in the previous section, η_{ex1} comes from the effects of critical angle and the surface back reflection. If the LED does not have an anti-reflection coating, then the Fresnel loss depends on the difference of refractive indexes and we have:

$$\eta_{ex1} \sim 1.985\% \cdot \left[1 - \left(\frac{n_{GaAs} - 1}{n_{GaAs} + 1}\right)^2\right] \sim 1.36\%. \tag{3.38}$$

In addition, we assume $\eta_{ex2} \sim 50\%$ and $h\nu/q = 1.42$, which corresponds to a wavelength of 875nm. The external efficiency of the LED should be:

$$\eta = 1.42 \times 0.513 \times 0.0136 \times 0.5 = 4.95 \times 10^{-3} \quad \text{(W/A)}. \tag{3.39}$$

However, in some of the LED's we fabricated, the efficiency at 1mA driving current can be as low as $10^{-4} \sim 10^{-5}$W/A. This dramatic reduction in the efficiency can not be attributed to the possible errors in our previous calculation. Therefore, there must be some other mechanism that is responsible for this further reduction. This additional effect comes from the surface recombination.

In the discussion of the interface recombination, we pointed out that the mismatch between lattice constants results in defects on the interface, which acts as nonradiative recombination centers. For the mesa-type LED, dramatic change in lattice structure occurs along the GaAs/air surface on the sidewall of the mesa. Similar to the case of heterointerface, such discontinuity will also cause nonradiative recombination even though the recombination centers are not the same kind of defects as those on the interface. Since the discontinuity here is much more significant than the GaAs/AlGaAs interface, the resulting surface recombination velocity is also much higher. Depending on the fabrication process, its value is typically on the order of $10^5 \sim 10^6$cm/sec.

The exact mechanism of GaAs surface recombination is not yet very clear. In bulk material, we know that each gallium atom has three bonds connected to three arsenic atoms and the reverse case happens for arsenic atoms, which thereby forms

a stable zincblende lattice structure. However, on the surface, there will be free Ga or As bonds dangling without being connected to any other atoms. As a result, they will capture all kinds of atoms or ions from the environment they are exposed to. It is believed that some of these dangling bonds are responsible for the high surface recombination velocity in GaAs. The same situation takes place on a silicon surface. However, since silicon has a natural oxide material, silicon dioxide, we can therefore passivate the silicon surface by forming a stable oxide layer on its surface. This is the reason why silicon MOS devices can be successfully fabricated. Unfortunately, GaAs does not have any natural oxide material. As a result, the high surface recombination velocity is a major problem for the mesa-type LED.

For a LED under a certain forward bias voltage, the junction current can be expressed as follows from experiment measurement:

$$I = I_1 \exp[qV/kT] + I_2 \exp[qV/2kT] \ . \tag{3.40}$$

The first term can come from diffusion current, which is very small in a double-heterostructure diode since the effective carrier confinement by AlGaAs barriers make the diffusion process across the interface negligible. The other source of the kT current is the radiative recombination, which has a recombination rate of $B_{eff}np$ as given by Eq. 3.18. From basic semiconductor physics [36], the pn product of a p-n junction is given by:

$$pn = n_i^2 \exp[qV/kT] \ . \tag{3.41}$$

Therefore, the radiative recombination current density is:

$$j_r = qwR = j_{r0} e^{qV/kT} \tag{3.42}$$

where $j_{r0} = qwB_{eff}n_i^2$.

The second term in Eq. 3.40 generally comes from the nonradiative recombination process. Among them, the surface recombination is found to be the main contributor for light-emitting diodes based on III-V material because of their high surface recom-

bination velocity [79, 80]. The recombination rate on the surface can be written as [81] :

$$R_s = s_0(np)^{1/2} \ . \tag{3.43}$$

From the theory given in Ref [81], the surface current density is:

$$j_s = j_{s0}e^{qV/2kT} = qs_0L_sn_ie^{qV/2kT} \tag{3.44}$$

where L_s is the surface diffusion length for the minority carriers, which is found to be about 0.07μm [81].

Assuming the LED is circular-shaped with a radius of r, the forward-biased current is then the sum of the radiative recombination current and the surface current:

$$\begin{aligned} I &= j_r \cdot \pi r^2 + j_s \cdot 2\pi r \\ &= \pi r^2 \cdot qwB_{eff}n_i^2e^{qV/kT} + 2\pi r \cdot qs_oL_sn_ie^{qV/2kT} \ . \end{aligned} \tag{3.45}$$

Due to the addition of surface recombination current, the output power of an LED can be rewritten as:

$$\begin{aligned} P_{out} &= \frac{h\nu}{q}\eta_i\eta_{ex1}\eta_{ex2} \cdot I_r \\ &= \frac{h\nu}{q}\eta_i\eta_{ex1}\eta_{ex2}\frac{I_r}{I_r + I_s} \cdot I_{total} \ . \end{aligned} \tag{3.46}$$

The reduction factor introduced by surface recombination is:

$$\begin{aligned} \eta_s &= \frac{I_r}{I_r + I_s} \\ &= \frac{\pi r^2 \cdot j_{r0}\exp[qV/kT]}{\pi r^2 \cdot j_{r0}\exp[qV/kT] + 2\pi r \cdot j_{s0}\exp[qV/2kT]} \\ &= \frac{1}{1 + \frac{2j_{s0}}{rj_{r0}} \cdot \exp[-qV/2kT]} \ . \end{aligned} \tag{3.47}$$

As an estimation of magnitude, assuming the following parameters: $r = 15\mu$m, $s_0 =$

10^6cm/sec, $L_s = 0.07\mu$m, $n_i = 1.79 \times 10^6\text{cm}^{-3}$, $B_{eff} = 9 \times 10^{-11}\text{cm}^3/\text{sec}$, $w = 0.6\mu$m, we have $j_{s0} = qs_0L_sn_i = 2.00 \times 10^{-12}\text{A/cm}$ and $j_{r0} = qwB_{eff}n_i^2 = 2.77 \times 10^{-21}\text{A/cm}^2$. At a bias of 1.25V, the corresponding current is $I_s = 0.583$mA from Eq. 3.44 and $I_r = 18.7\mu$A from Eq. 3.42. The resulting reduction factor is $\frac{I_r}{I_r+I_s} = 3.1\%$, which is very substantial as confirmed by our actual testing.

The effect of surface recombination is more significant at low driving current levels due to the $2kT$ term in the exponential. For example, if the bias is increased to 1.50V, then $I_s = 73.0$mA and $I_r = 294$mA. The reduction factor is increased to 80% , which is much better than the 3.1% at $V = 1.25$V. For the optoelectronic neuron arrays, this is bad news because we are more interested in operating the device on low current levels in order to increase the neuron density.

In order to solve the problem of surface recombination, the ultimate solution is to reduce s_0. If we can reduce it by an order of magnitude to 10^5cm/sec, the reduction factor will be 24.3% at 1.25V bias ($I_{total} = 77\mu$A) and 97.6% at 1.50V bias ($I_{total} = 301$mA). This is not an easy task. Normally, the so-called passivation layer of Si_3Ni_4 on GaAs does not affect the recombination process much but only provides mechanical protection. One method to reduce s_0 is to find out a better process since the surface recombination velocity of GaAs depends on the environment in which it is etched. We found that when etching the mesa, a dry etch, such as RIE, can produce a surface with better (lower) s_0 compared with the wet etching we normally used in our process, but the reduction is limited.

Another method to reduce s_0 is to treat the newly etched GaAs surface with sulphide [82, 83]. The treatment involves the deposition of Na_2S or $(NH_4)_2S$, which results in the forming of stable Ga-S bonds. Samples with sulphide treatment showed more than one order of magnitude increase in photoluminescence intensity and recombination velocity as low as 1.2×10^3cm/sec was observed [82]. However, the stability of such treatment is still a major problem. The GaAs surface after sulphide treatment usually shows a slower surface recombination velocity for a short time period, ranging from several minutes to several hours, then gradually deteriorates and returns back to the situation before the sulphide treatment. Further progress is needed before

applying this technique in the fabrication of LED's on neuron arrays.

Besides reducing s_0, η_s may be changed by adjusting the size of the LED, r. From Eq. 3.47, it seems like increasing r will result in the increase of LED efficiency. However, for a fixed driving current, larger r means the bias voltage is smaller, therefore η_s is lower. It turns out that these two effects cancel with each other and the reduction factor due to surface current is independent of mesa sizes when the total current is fixed. For the same driving current I_{total}, we can solve the required forward-bias voltage from Eq. 3.45. The result is:

$$V = \frac{2kT}{q} \cdot \ln[\frac{\sqrt{\pi^2 j_{s0}^2 + \pi I_{total} j_{r0}} - \pi j_{s0}}{\pi r j_{r0}}] \,. \tag{3.48}$$

Substitute into Eq. 3.47, we have the following reduction factor due to surface recombination:

$$\eta_s = \frac{\sqrt{\pi^2 j_{s0}^2 + \pi I_{total} j_{r0}} - \pi j_{s0}}{\sqrt{\pi^2 j_{s0}^2 + \pi I_{total} j_{r0}} + \pi j_{s0}} \,. \tag{3.49}$$

When the driving current is small and the surface recombination current is dominating, i.e., when $I_{total} j_{r0} \ll \pi j_{s0}^2$, Eq. 3.49 can be reduced to a simpler form:

$$\eta_s = \frac{1}{4} \frac{I j_{r0}}{\pi j_{s0}^2} \tag{3.50}$$

which means that we should observe the LED efficiency increases linearly as the driving current increases when the surface recombination current is dominating.

From the previous discussion, it looks like besides using different etching process, there is not much we can do in reducing the surface recombination current based on Eq. 3.40 – Eq. 3.47. However, these equations are all based on an one-dimensional model which does not consider the variation of injected carrier concentration within the plane parallel to the emitting surface. If the diffusion process in that plane is taken into consideration, we know that the injected carrier concentration will decay as it moves away from the electrode toward the edge. The diffusion length is given by $L_n = \sqrt{D_n \tau_n}$. For $D_n = 100\text{cm}^2/\text{sec}$ and $\tau_n = 1 \times 10^{-8}\text{sec}$, the diffusion length

is approximately 10μm. Therefore, if the mesa edge is several tens of microns away from the electrodes, only a very small portion of the injected carriers will reach the surface and recombine nonradiatively there. All other injected carriers will be forced to stay within the active layer and result in mostly radiative recombination. By doing so, we can substantially reduce the surface recombination current and increase the LED efficiency. The disadvantage of this technique is the increased lateral dimension of LED and the possible metal blocking effect if no proper current confinement is implemented.

There is one more important factor: the series resistance of ohmic contacts, which is briefly mentioned when discussing the effect of metal blocking. If we consider the finite conductance of the ohmic contact and the contact layers, Eq. 3.45 and 3.47 should be rewritten as:

$$I = \pi r^2 j_{r0} \exp[q(V - IR)/kT] + 2\pi r j_{s0} \exp[q(V - IR)/2kT] \quad (3.51)$$

$$\eta_s = \frac{1}{1 + \frac{2j_{s0}}{rj_{r0}} \cdot \exp[-q(V - IR)/2kT]} \quad (3.52)$$

where R represents the series resistance of the LED. Therefore, the higher R is, the less competitive radiative recombination will be and the lower the LED efficiency will be. Therefore, it is very important to reduce the series resistance in order to reduce the effect of surface recombination. We can do so by increasing the width of contact metal. A possible trade-off exists here because the metal blocking effect will be more significant when the contact width is increased if no proper current confinement is provided. Such trade-off can be easily observed in our experimental measurement, which will be presented in the next section.

3.4 Experiment Result

In the previous sections, we discussed those factors affecting the LED efficiency. Surface recombination current is found to cause substantial reduction in the output power and must be controlled. Metal blocking effect is also important and can be minimized

if current confinement is implemented. We will present the result of experimental testing in this section.

3.4.1 Fabrication

Several kinds of LED fabrication techniques are investigated, including double heterostructure LED with no current confinement, ion-implanted LED, Zn diffusion LED, and quantum well LED. The structures are shown in Figure 3.9. LED's with different mesa sizes and different shapes of metal contact are built for each type of LED. It is found that the optimum condition depends on the level of driving currents and will be discussed later.

The LEDs shown in Figure 3.9(a) – (c) are all double heterostructure LED's, which gives the best carrier injection efficiency as discussed earlier. The fabrication procedure of (a) and (b) are quite similar except (b) goes through an additional step of ion implantation. They also share the same epi-layers as the one shown in Figure 3.10. The active layer is nominally undoped, but there is always some residue doping, typically in the order of 10^{15}cm^{-3}. The mesa was etched using phosphoric acid etchant consisting of H_3PO_4, H_2O_2, and CH_3COOH in a ratio of 1:1:3. After etching down to expose the p^+ GaAs layer, a second etching was performed to achieve device isolation. Thermal CVD at 610°C was then applied to deposit a thin layer of silicon nitride (SiN_x). The deposition thickness was about 1000Å, controlled by monitoring the color of the deposited nitride. For the ion-implanted LED, an oxygen ion-implantation was performed to achieve the buried structure as shown in Figure 3.9(b). The ion energy is 360keV, which has a projected range of 0.548μm, and a projected straggle of 0.167μm [84], and the applied dose is $0.5 \times 10^{12}\text{ions/cm}^2$. The implanted ion concentration has a Gaussian distribution and can be expressed as [36]:

$$n(x) = \frac{\phi}{\sqrt{2\pi}\Delta R_p} \exp\left[-\left(\frac{x - R_p}{\sqrt{2}\Delta R_p}\right)^2\right] \tag{3.53}$$

where ϕ is the ion dose, R_p is the projected range, and ΔR_p is the projected straggle.

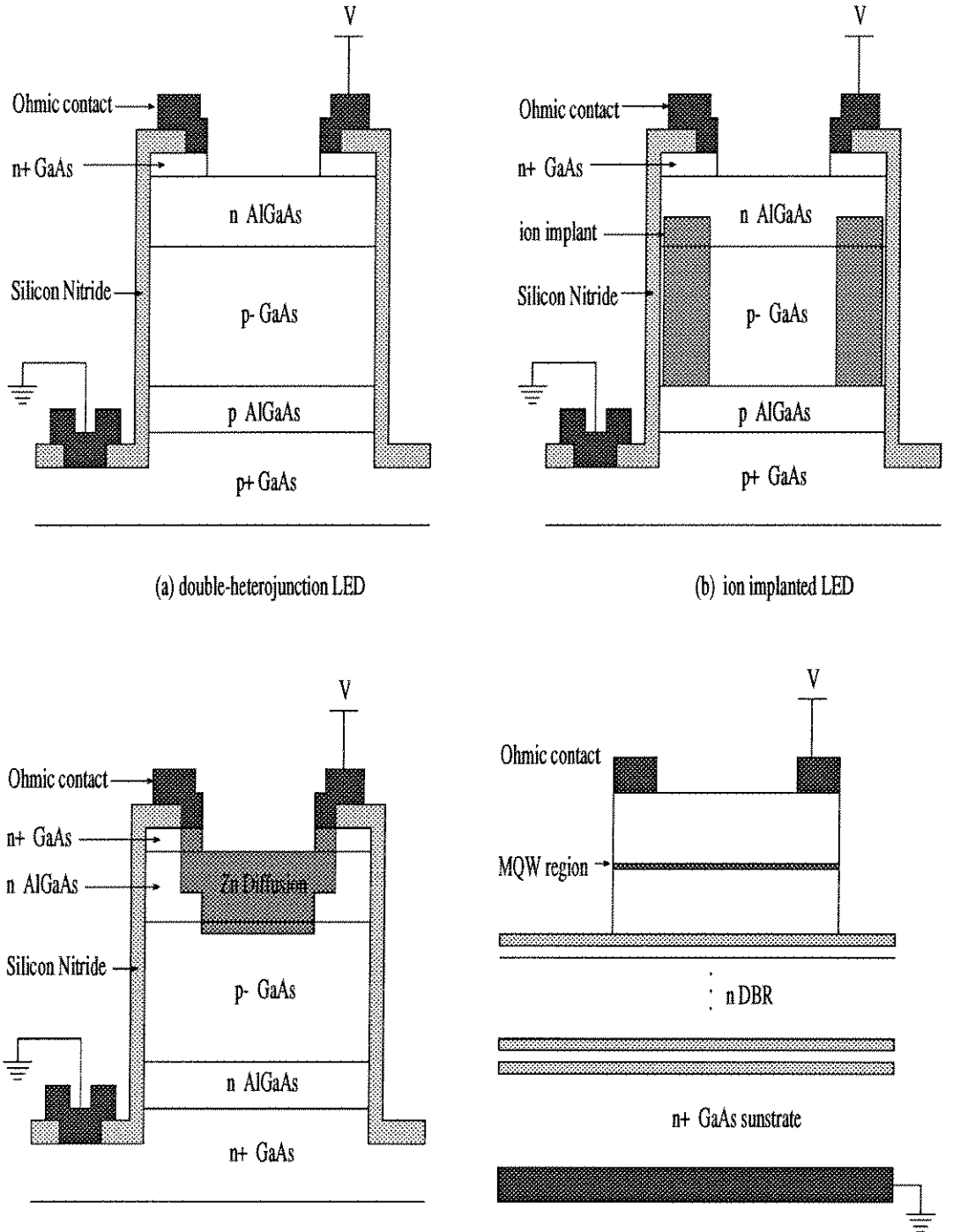


Figure 3.9: Different types of LED structure.

GaAs	n=2-3e18	0.1000 μm
AlGaAs (30% Al)	n=1e18	0.2550 μm
GaAs	undoped	0.6000 μm
AlGaAs (30% Al)	p=1e18	0.2570 μm
GaAs	p=5e18	0.5000 μm

Figure 3.10: Epi-layer for the LED shown in Figure 3.9(a) and (b). Doping concentrations are in unit of cm^{-3}.

Each oxygen ion can cause approximately 50 to 100 damages. The mask used for the implantation was 4μm thick positive photoresist. Nitride windows were then opened using CF_4 plasma etching and ohmic metals were deposited using lift-off technique. The n-metal was AuGe/Ni/Au and the p-metal was AuZn/Au. A thermal annealing at 410°C was the final step in order to achieve ohmic contact. When we looked at the LED output under microscope, the effect of current confinement due to ion implantation is evident: LED's without current confinement had bright light emission from their mesa perimeter while no edge emission could be observed on properly implanted LED's.

Figure 3.11 shows the epi-layer for the device shown in Figure 3.9(c). Starting with a n-p-n structure, the top layer was converted to p-type by Zinc diffusion. The first three steps were the same as what we described for the ion-implanted LED: mesa etch, isolation etch, and nitride deposition. CF_4 plasma etching was applied to open windows for the first Zn diffusion. The sample is then loaded into a high purity quartz ampoule together with Zn_3As_2 source. The ampoule had been soaked in dilute HF solution for two hours and dried up in 200°C oven before being used. It was then

GaAs	n=2-4e18	0.0962 μm
AlGaAs (30% Al)	n=1e17	0.4842 μm
GaAs	p=5e15	0.6706 μm
AlGaAs (30% Al)	n=6e16	0.2470 μm
GaAs	n=2-4e18	0.2973 μm

Figure 3.11: Epi-layer for the double Zn diffusion LED shown in Figure 3.9(c). Doping concentrations are in unit of cm^{-3}.

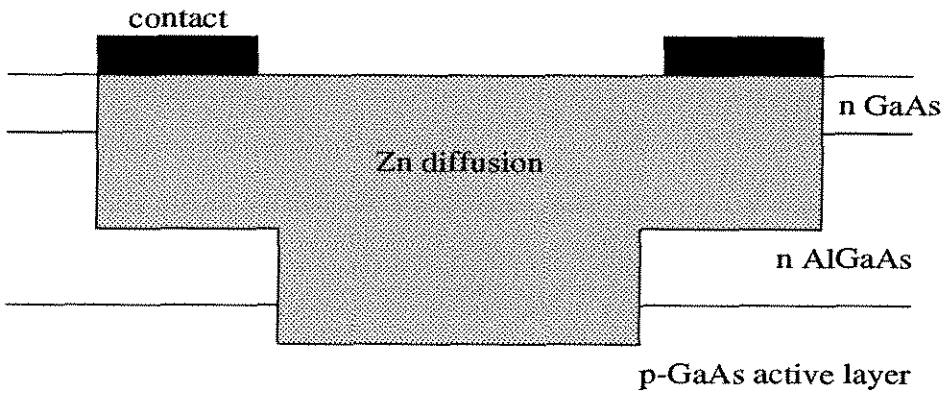
pumped to a vacuum of almost 10^{-8}torr . The ampoule was sealed using a torch in order to maintain the high vacuum during the diffusion process. The sample was put into an oven heated to 640°C. By controlling the time within which the sample remained in the oven, we could control the diffusion depth. It is given by the following formula [52]:

$$d = Ae^{-E_d/kT}\sqrt{t} = X_j\sqrt{t} \tag{3.54}$$

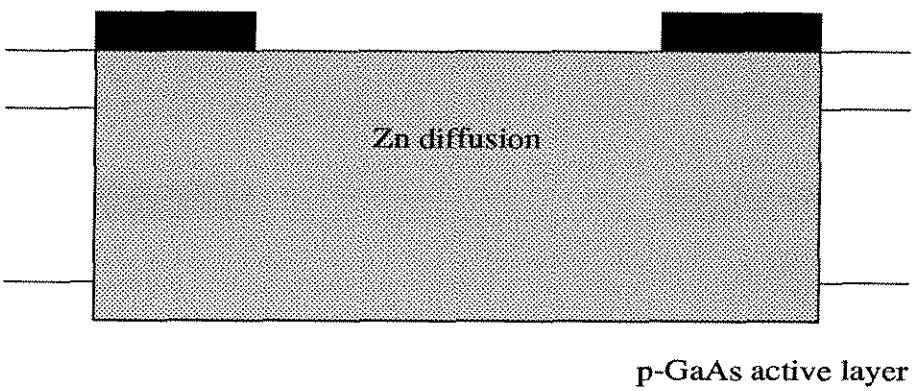
where A is a constant, E_d is the activation energy, and X_j is called the diffusion rate. From data given in Ref [53], the diffusion rate can be calculated as: $X_j = 1.0758\mu\text{m/hr}^{1/2}$ for GaAs and $X_j = 1.768\mu\text{m/hr}^{1/2}$ for $Al_{0.3}Ga_{0.7}As$. The nitride layer was used as the diffusion mask, which means diffusion only occurred in the opened window region. Once the preset time was reached, the sample was pulled out of the oven and quickly quenched. A larger nitride window was etched and second diffusion process was performed. During the second diffusion, the diffusion front from the first diffusion process would also diffuse further into the material and this effect must be considered when determining the two diffusion times so that

the first diffusion front reached the p^- active layer while the second diffusion front remained in the n AlGaAs layer. The doping profile is illustrated in Figure 3.12(a). In the first diffusion region, we have a p-i-n diode which the electric current could go through under proper forward bias. In the second diffusion region, a p-n-p-n thyristor structure was formed which prevented any current from flowing through that region. Therefore, current confinement was implemented and the effect of light blocking by metal contacts could be reduced. For comparison, LED's with overlapping diffusion regions was also fabricated and the resulting doping profile is shown in Figure 3.12(b). Since two diffusion steps are needed, it takes quite a long time to finish the diffusion process. Especially the high vacuum requires a couple of days of pumping using the ion pump we have. In order to save the processing time, a one-step diffusion process was performed in some fabrication runs: instead of the first diffusion, the window region was etched using a slow etchant consisting of NH_4OH, H_2O_2, and H_2O in a ratio of 3:1:139. The etching depth was calculated beforehand. Following the etching, the second diffusion was applied to achieve a doping profile as the one shown in Figure 3.12(c). The current confinement effect was found to be similar to those using a two-step diffusion. However, we found out that the series resistance was higher for the one-step diffusion due to the narrow conducting channel. The result was a reduced efficiency due to the competition between surface current and radiative current, which was discussed in Section 3.3. After the diffusion was done, the sample was deposited with n-metal (AuGe/Ni/Au) and p-metal (Cr/Au) following the CF_4 plasma etching of nitride windows. It was then annealed at 410°C for 1.5 minutes. Effective current confinement was observed under microscope by examining the light emission from LED mesa edge.

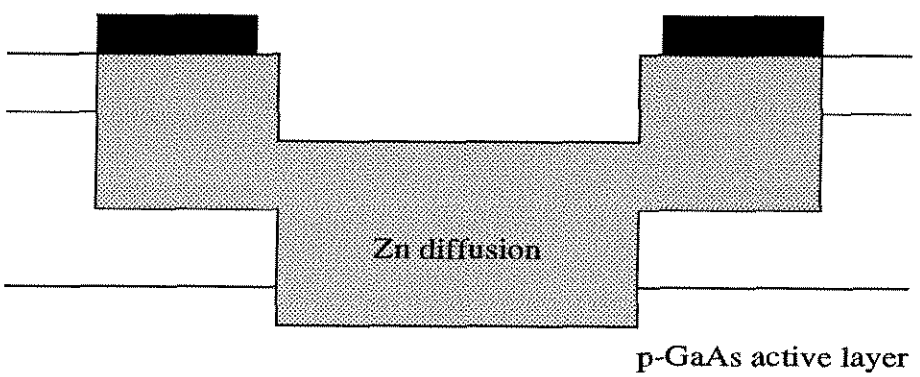
The quantum well micro-cavity LED shown in Figure 3.9(d) was fabricated at Bellcore by Dr. Scherer [85]. Dry etching was applied for etching the LED mesa, which should result in a better surface with lower surface recombination velocity as compared to the chemical wet etching. The structure and fabrication process were similar to those used for VCSEL's [86] except that no top mirror was deposited. Two types of quantum well devices were tested: a three-quantum-well InGaP and a five-



(a) two-step double Zn diffusion



(b) two-step single Zn diffusion



(c) one-step double Zn diffusion

Figure 3.12: Different doping profiles from the Zn diffusion process.

superlattice-pair of AlGaAs. The emission spectra were in the visible range instead of the normal infrared regime for bulk GaAs. The emission peak is around 670nm for the InGaP devices and 720nm for the AlGaAs devices. Totally two runs of MQW LED devices were tested and promising results were obtained from the second run.

There are several advantages of the micro-cavity LED's over conventional LED's as those we fabricated in our lab. First, by placing the active region inside and optical cavity, the spontaneous emission will interact with the optical mode density within that cavity. In return, the spontaneous emission can be either enhanced or inhibited [87]. If the active layer is placed in the optical node position, the spontaneous emission will have a longer lifetime; if it is placed in the antinode position, the lifetime will be shortened. In the devices we tested, the multiple quantum wells were placed into the antinode position, therefore, we should have an enhanced spontaneous emission which will increase the LED efficiency. Second, the bottom DBR will reflect the light that propagates toward the substrate back to the GaAs/air interface. As a result, the efficiency can be increased by a factor of two. In addition, the linewidth of the emission spectrum can be improved.

3.4.2 Measurement

I-V Characteristics

In Section 3.3, Eq. 3.45 gives the electric current of a LED under a forward bias voltage. In a practical situation, it can be expressed as:

$$I = I_0 \exp(qV/nkT) \tag{3.55}$$

where I_0 is called the reverse saturation current and n is called the ideal factor and typically ranges between 1 and 2. When $n = 1$, the current is dominated by radiative recombination current or diffusion current which is negligible in the case of a double heterostructure LED. When $n = 2$, the nonradiative recombination current, either through surface or deep level recombination centers, dominates.

Fabrication Method	ideal factor	I_0 (A)
homojunction	4.286	7.70×10^{-9}
double heterojunction	2.011	6.63×10^{-13}
ion-implant	2.195	2.64×10^{-12}
single diffusion	1.591	1.74×10^{-16}
double diffusion (DD1)	3.374	3.90×10^{-10}
double diffusion (DD2)	3.720	3.01×10^{-8}
quantum well (1st run)	1.838	5.69×10^{-18}
quantum well (2nd run)	2.383	8.19×10^{-19}

Table 3.1: Measured ideal factors and I_0 for different LED's.

The I-V curves of different LED's are plotted in Figure 3.13 in logarithmic scales. A typical LED shows two steps in its I-V curve. The first one saturates at very low current level, below 1μA. We believe it comes from the nonradiative deep level recombination that occurs in the space-charge region. The low concentration of such deep level centers provided by modern growth technology results in its early saturation. Because different wafer structure were grown by different process, the values of this saturated current changed quite a lot, ranging from 10pA to 10nA. After that, we have a quickly rising current. Carefully examining the rising slope reveals an ideal factor close to 2.0 for most devices, which means it is dominated by surface current. Table 3.1 lists the measured result of I_0 and n for different LED's. Due to the series resistance, the measured ideal factor should be larger than the actual value. As the current increases further, high injection effect and series resistance cause the curve to be more and more horizontal.

In some cases, defects or contamination were introduced during the fabrication process. The result is an abnormally large nonradiative current. Since the surface current is so large, the deep level recombination current will not be observed, as the one shown in Figure 3.13 (device DD2). As a result, a very poor output efficiency is measured for this LED.

For comparison, the result from a homojunction LED is also presented in Figure 3.13. This device was fabricated from Zn diffusion on a 1.5μm thick n-type GaAs

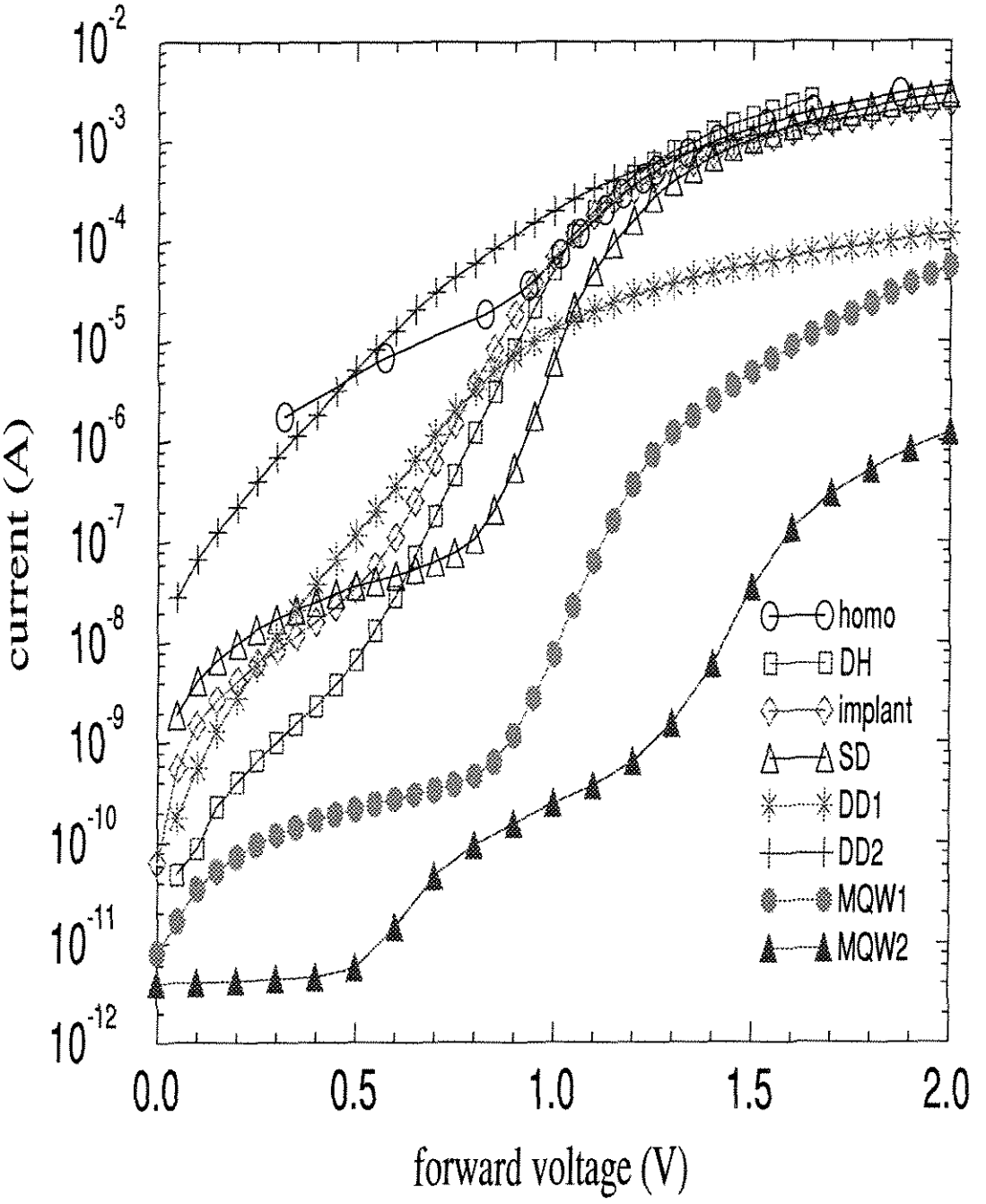


Figure 3.13: I-V characteristics of different LED's: homojunction LED (homo), double heterojunction LED (DH), ion-implanted LED (implant), Zn diffusion LED, including single diffusion LED (SD) and two double diffusion LED's (DD), and two quantum well LED's (MQW).

layer. The diffusion depth is 0.8μm and the same sealed ampoule technique described earlier was applied.

L-I Curves

We measured the output power of a LED using a UDT detector, which is calibrated for both 850nm and 630nm wavelength. The detector head has an area of $\frac{1}{4}$inch$\times\frac{1}{4}$inch. It was placed about 7 millimeters above the LED. The output power of different LED's are plotted as a function of driving current in Figure 3.14 in logarithmic scales. For almost all devices, the relationship between the output power and the driving current can be expressed as:

$$P = P_0 I^m \ . \tag{3.56}$$

Furthermore, we normally observed two different m for each devices. At small driving current, the output power increases roughly quadratically ($m_1 = 2$), which means the efficiency increases linearly as a function of driving current. This is in agreement with Eq. 3.50, confirming that surface recombination current is much larger than the radiative recombination current. As the current level increases, the radiative term gradually takes over and the L-I curve becomes more and more linear ($m_2 \to 1$). However, the transition here strongly depends on the fabrication process and the size of the LED. Other factors affecting the L-I curves at high current levels include the thermal effect and the series resistance problem.

In Table 3.2, we list the experimentally fitted values of m_1, m_2, and P_0 for the different devices we tested. If we compare it with the values given by Table 3.1, we can see that normally m_1 is close to the ideal factor n. In addition, a large I_0 typically results in a small P_0. This is understandable. Because the radiative current $I_{rad} \propto e^{qV/kT}$ and the total LED current $I_{total} \propto e^{qV/nkT}$, the output power $P \propto I_{rad} \propto e^{qV/kT} \propto I^n$. In addition, since surface current is dominating, a larger I_0 means the nonradiative recombination velocity is higher if the devices have the same sizes. In return, the output power should be weaker.

The best efficiencies were obtained from the second run of quantum well LED's.

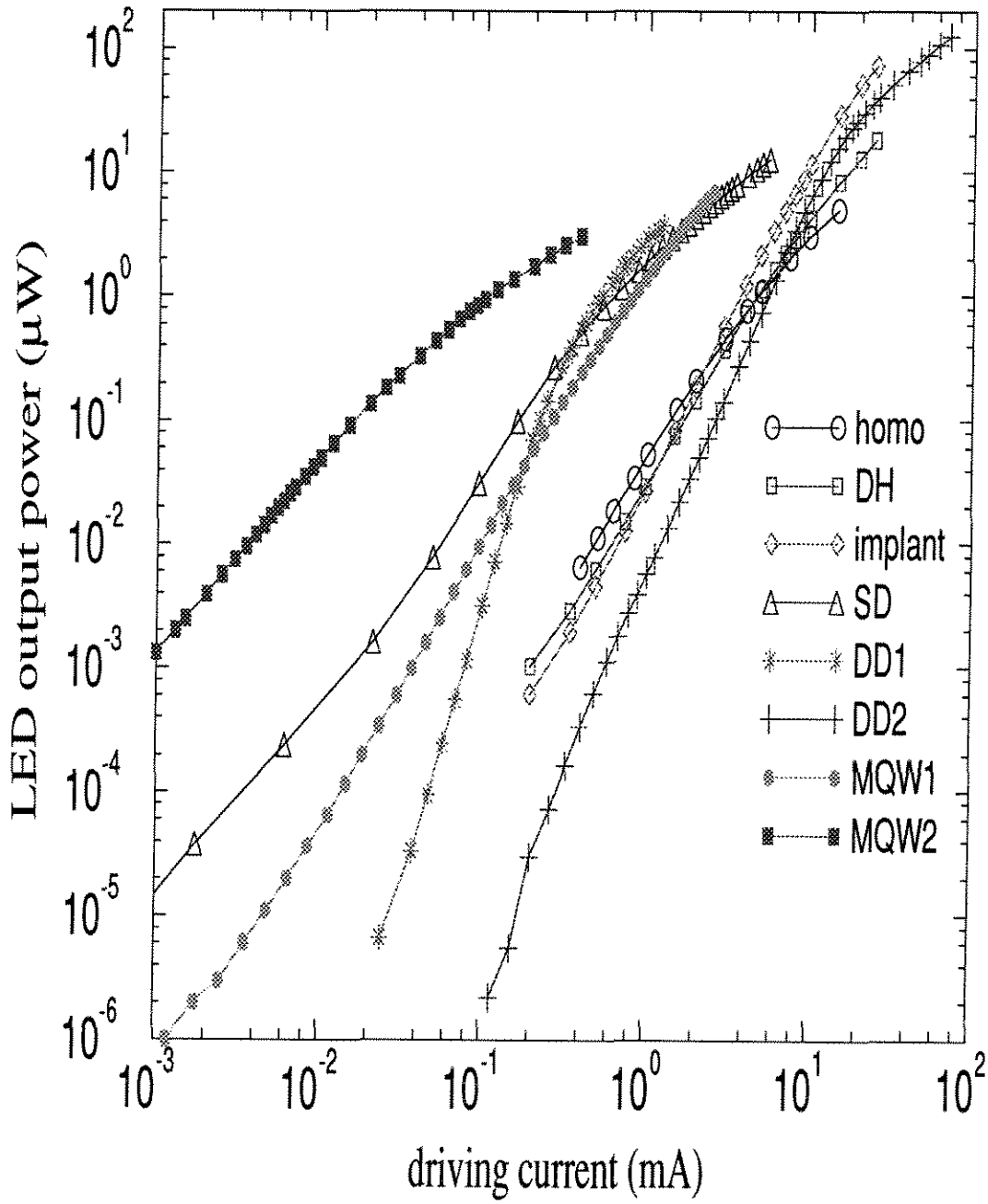


Figure 3.14: L-I curves of different LED's. Detector head is about 7mm over the device and has an area of $\frac{1}{4} \times \frac{1}{4}\text{in}^2$.

	low current level		high current level	
Fabrication Method	m_1	P_0 at 0.1mA (nW)	m_2	P_0 at 1mA (μW)
homojunction	2.184	0.327	1.321	0.145
double heterojunction	2.273	0.176	1.594	5.73×10^{-3}
ion-implant	2.727	0.0578	1.944	6.66×10^{-3}
single diffusion	2.118	35.86	1.049	1.96
double diffusion (DD1)	4.611	3.70	1.266	2.57
double diffusion (DD2)	3.105	4.67×10^{-3}	1.126	0.977
quantum well (1st run)	2.360	11.0	1.515	1.87
quantum well (2nd run)	1.548	1943	0.8254	6.28

Table 3.2: Measured L-I relationship for different LED's.

This device has good efficiency, especially at low current level, which is exactly what we need. However, for some unknown reason, the maximum allowable current is around several hundred micro-amps. Beyond that, the device shows signs of damage and the light output decreases dramatically. Besides quantum well LED, LED's fabricated by Zn diffusion, which can operate up to tens of milliamps, also shows good result at slightly higher current levels. Table 3.3 lists the best efficiency we achieved at different driving current levels. The results do not necessarily come from a single LED. As we will see later, LED's with different sizes have different optimum driving currents. In addition, the efficiency from diffusion LED's could be increased by about 43% if anti-reflection coating were implemented.

To compare with the theoretical discussion presented in Section 3.1, the result from device MQW2 and SD, after the effect of critical angle and Fresnel loss had been compensated, were plotted earlier in Figure 3.4 together with the simulation result. We can see that they agree well qualitatively. However, quantitative difference exists. At low current levels, the diffusion device performed much worse than theoretical prediction because the surface recombination current was not taken into consideration in the simulation. Therefore, the resulting efficiency is much lower than the simulation result. At high current density, the surface current is less important and we observed better agreement between experiment and theoretical calculation. However, the series resistance problem and thermal effect tends to decrease the LED efficiency now.

current (μA)	η for QW LED	η for diffusion LED
10	4.93×10^{-3}A/W	4.82×10^{-5}A/W
20	5.82×10^{-3}A/W	7.32×10^{-5}A/W
30	7.62×10^{-3}A/W	1.08×10^{-4}A/W
50	8.71×10^{-3}A/W	1.70×10^{-4}A/W
70	9.26×10^{-3}A/W	2.41×10^{-4}A/W
100	9.44×10^{-3}A/W	3.64×10^{-4}A/W
200	8.71×10^{-3}A/W	7.44×10^{-4}A/W
300	8.18×10^{-3}A/W	1.20×10^{-3}A/W
500	8.26×10^{-4}A/W	1.76×10^{-3}A/W
700	1.19×10^{-3}A/W	2.39×10^{-3}A/W
1000	1.67×10^{-3}A/W	2.77×10^{-3}A/W
2000	2.58×10^{-3}A/W	4.04×10^{-3}A/W

Table 3.3: Best efficiency at different current levels from all LED's.

3.4.3 Discussion

Quantum well LED

The quantum well LED's prepared by dry etching normally show lower surface recombination currents. In addition, by using the anisotropic etching, the mesa size can be dramatically reduced. Devices with diameters as small as 6μm were fabricated and tested. The reduced mesa size helps to increase the injected current density, thereby reducing the radiative emission lifetime. As a result, even at low driving current levels, the radiative emission is very strong and the result is a high efficiency. However, due to the narrow width of contact metal, the high series resistance is a major problem. A larger forward bias is needed compared to other LED devices, which will reduce the energy efficiency even though the current efficiency is high. In addition, as we mentioned earlier, only a small current is allowed to pass through the LED's fabricated in the second run, which is the one with high efficiency. For the first run, the current efficiency is only comparable to diffusion LED's.

current (mA)	$\Delta = -546$Å	$\Delta = 65$Å	$\Delta = 330$Å	$\Delta = 783$Å	$\Delta = 1405$Å
0.1	100%	78.5%	81.6%	4.3%	63.2%
0.3	100%	91.1%	98.4%	19.3%	77%
0.5	100%	98.8%	93.7%	19.2%	82.4%
1.0	100%	97.6%	102.4%	29.5%	95.2%
2.0	100%	86.6%	101.7%	26%	86%

Table 3.4: The effect of the distance, Δ, between the calculated diffusion front and the AlGaAs/GaAs heterointerface on the LED output power. The result is average over 10 devices. The LED output is presented as the percentage over devices with $\Delta = -546$Å.

Diffusion LED

Double Zn diffusion provides efficient current confinement. In return, the metal blocking effect is reduced. So, at high current levels, when the surface recombination current is less significant, a double diffusion LED has better efficiency than a single diffusion device. However, it seems like the additional doping profile of double diffusion introduces additional nonradiative recombination current and increases series resistance of the conducting channel. As the result, when the driving current is small and surface recombination dominates, the benefit from lower blocking effect is offset by the higher nonradiative currents and the efficiency is actually lower than that of a single diffusion device.

In experiment, we also found that the distance, Δ, between the diffusion front and the AlGaAs/GaAs-active-layer heterointerface strongly affects the output power for the one-step double Zn diffusion devices as shown in Figure 3.12(c). The conclusion here should be able to extend to devices fabricated using the two-step diffusion process even though no experiment was carried out to test this effect due to the long time needed for the fabrication. It is found that when the calculated diffusion front is slightly above the heterointerface ($\Delta = -500 \sim -700$Å), the output power is better than the case when the diffusion front is within the GaAs active layer (positive Δ). For different positive Δ, the improvement is different and we did not observe any simple relationship between Δ and the output power as shown in Table 3.4. Since all

these devices were fabricated in the same run and other parameters remain the same, the only difference comes from the position of the diffusion front and the resulting series resistance. As Δ increases, the first diffusion region is etched further, so the conducting channel has a thinner layer and may result in higher series resistance. However, the output does not decrease monotonically when Δ increases. Therefore, we believe this is not the main reason for the observed behavior.

The only reason we can think of is that damages were introduced by the Zn diffusion process, which resulted in another interface recombination problem. If Δ is slightly negative, this diffusion-resulted interface is above the actual heterointerface, so the electron concentration at this position is very low. Therefore, this additional interface recombination does not cause severe nonradiative recombination current. In the meantime, the thin n-AlGaAs layer with only 546Åthickness can be easily depleted and we don't have any current conducting problem. On the other hand, if the diffusion front is within the active layer, we will have two layers of interface recombination centers, so the efficiency goes down due to the extra nonradiative recombination. When the diffusion interface is deep into the active region, most of the injected electrons that do not involve in radiative recombination will recombine on this diffusion interface. In return, the electron concentration at the original heterointerface becomes small and it acts like we have only one interface again. As a result, the output power is raised a little bit. This can also explain why the reverse saturation current, I_0, is higher for the double diffusion devices than the double heterojunction devices as given in Table 3.1.

Ion-implanted LED

The results from the double heterojunction LED and the homojunction LED are comparable, which is a surprise to us, especially when the reverse saturation current of DH devices is much smaller. In both cases, the efficiency is very poor. The reason comes from the dominating surface recombination current and the poor current confinement. With the help from ion implantation, the metal blocking effect is reduced and the efficiency at high current levels can be increased by a factor of five. However,

at low current levels, the additional damage caused by ion implant results in even more nonradiative recombination current and the efficiency is lower than DH devices.

Trade-off

The L-I curves of different LED's typically show some cross-over within a certain current range. This means devices that have better efficiency at low current levels normally do not perform as well at high current levels and vice versa. This is the result of the trade-off between reducing metal blocking effect and increasing radiative recombination current. Devices with good ohmic contact, which will increase the radiative current, usually have wider metal contact, which results in a more severe metal blocking problem. At low current levels, the surface current is dominating, so the reduction in the series resistance can help a lot in increasing the radiative recombination current. The increase from this factor is much larger than the metal blocking effect. On the other hand, when the radiative recombination current becomes comparable or even higher than the surface current, the increase from reducing series resistance is not that significant, so the difference in metal blocking effect will show up in the output power. The above trade-off can be clearly seen when we compare the result of LED's with different mesa sizes but fabricated using the same process, which will be discussed next.

The effect of LED geometry design

In Section 3.3, we showed that in contrast to the case of VCSEL's, a larger LED mesa did not improve the percentage of radiative current for a fixed current level. Therefore, a smaller mesa size is preferred since such devices has a higher current density for a fixed current level, which results in a faster radiative recombination lifetime and a better efficiency. However, if the mesa size is too small, we will have a problem either with poor ohmic contact or severe metal blocking effects. As a result, an optimum design exists for a fixed current level.

Figure 3.15 shows the output power of three different quantum well LED's. These devices were all circular-shaped LED's with ring contacts as the one illustrated in

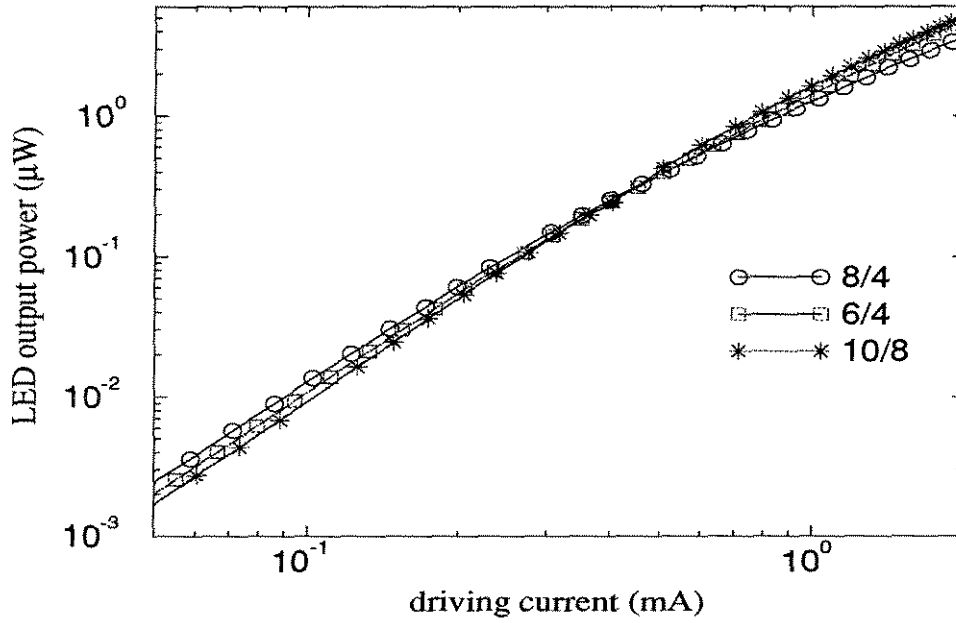


Figure 3.15: The effect of device size on the output of quantum well LED. The top view of these devices was shown in Figure 3.7. The first number given in the legend is the diameter of the LED mesa, in units of micron. The second one is the diameter of the optical window, also in units of micron.

Figure 3.7. For low driving currents, i.e., $I < 0.2$mA, device 8/4 (corresponding to $2r = 8\mu$m and $2(r-d) = 4\mu$m) gives the best efficiency. On the other hand, as current increases beyond 0.5mA, device 10/8 (corresponding to $r = 5\mu$m and d=1μm) has the best performance. This is the result of the trade-off between increasing radiative recombination current and reducing metal blocking effect.

As we mentioned earlier, at low current levels, anything that can improve the radiative recombination lifetime will show up in the output since $I_{total} \approx I_{nr}$. So, a smaller device is better since it has faster radiative recombination process. That's why device 6/4 ($r = 3\mu$m and $d = 1\mu$m) is better than device 10/8 when they have the same ohmic contact design. In addition, if we can reduce the series resistance, the radiative process becomes more competitive compared to surface current. The result is device 8/4, which has twice wider metal contact, has a better performance than device 6/4. That's the case of low current levels where nonradiative current dominates.

As the current increases, the above situation is no longer valid. Up to a certain point, the radiative recombination current becomes the major component. Under such a situation, the metal blocking effect becomes important. Therefore, device 10/8, which has a metal blocking reduction factor of 64% , has the best performance, followed by device 6/4, which has a metal blocking reduction factor of 44% . Device 8/4, which has a metal blocking reduction factor of merely 25% , gives the poorest performance.

3.5 Summary

In this chapter, the mechanism of light-emitting diodes was reviewed. The efficiency of a double heterojunction LED can be calculated by taking the following factors into consideration: the heterointerface recombination, self-absorption, radiative and nonradiative recombination in the active layer. The radiative recombination lifetime is not a constant but depends on the injected current density. The result is an increased efficiency for high current levels. An optimum thickness of the active layer exists due to the trade-off between the interface recombination and self-absorption. This is the case for an ideal LED. However, in practical situations, the external efficiency of a LED is limited by other factors as well, including the light collecting efficiency and the surface recombination current. The mismatch between the refractive indexes of GaAs and the air results in a significant reduction of external efficiency. Even with the help of an anti-reflection coating, the maximum efficiency is only 2% if no surface lithography, such as micro-lenses or nanotexture, and no bottom DBR mirror is implemented. In addition, for a surface emitting mesa-type LED, the light blocking effect by metal contact due to poor current confinement and the nonradiative recombination along the mesa edge due to the large surface recombination velocity of GaAs presents two major problems for increasing the LED efficiency. Several different fabrication techniques were investigated in order to increase the LED efficiency at low current levels. The results were presented in this chapter.

It was found that some of the quantum well LED's had good efficiency, but they

required large forward bias. Zn diffusion LED also gave good performance, but the surface recombination current was dominating at low current levels. Typical efficiency is below 1% due to the high surface recombination velocity. With anti-reflection coating and better current confinement, we can increase the efficiency a little bit. It was also found that there existed a trade-off between increasing the radiative recombination current and reducing the light blocking effect by metal contact. At low current level when the surface current dominates, the metal blocking effect is less important. By decreasing the mesa size and increasing the width of ohmic contact metal to reduce the series resistance of the ohmic contact, we can increase the external efficiency of a LED. On the other hand, at high current levels, the reverse case occurs where the metal blocking effect is more important and efficient current confinement is needed in order to get better efficiency.

Chapter 4 Optical FET Detector

In Chapter 1, when discussing the maximum achievable neuron density on an optoelectronic neuron array, we pointed out that it was essential to use high gain photodetectors in the OEIC neuron circuit. For detectors with high gain, less input light is needed to generate the same switching current, which, in return, means less output power is required for neurons from the previous layer. Therefore, lower power consumption can be achieved and more neurons can be integrated into the same area. In this chapter, we will discuss the issue of photodetectors, especially the high gain optical FET detector (OPFET), which shares the same epi-structure as a normal MESFET and has detector gain as high as 10^4A/W.

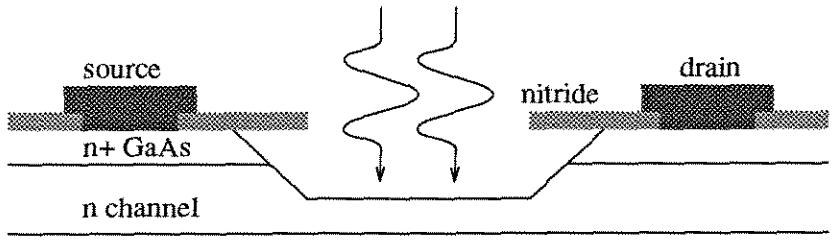
There are several other choices for the photodetector besides the OPFET, including bipolar phototransistors, p-i-n photodiodes, and metal-semiconductor-metal (MSM) detectors. However, these are not good candidates for high density OEIC neuron arrays either due to the incompatibility of epitaxial structures or the low detector efficiencies. The bipolar phototransistor, which has good current gain at high optical input levels, suffers from the fact that it requires a double heterojunction n-p-n structure. Even though the Zn diffusion LED starts with similar n-p-n epitaxial layers, the requirements on the layer thickness and doping levels are quite different for these two devices. Either the efficiency of the LED or the detector gain of the phototransistor has to be compromised if we want to integrate bipolar phototransistors into an optoelectronic neuron circuit and share the same epitaxial structure with the LED. If one set of the n-p-n structure is grown on top of the other, the increased vertical thickness will place stringent requirement on the fabrication process and may reduce the circuit uniformity. Furthermore, at low optical input levels, the detector gain of a bipolar transistor will decrease. Therefore, it is not a good choice for high density arrays.

Unlike bipolar phototransistors, both the photodiode and the MSM detector do

not have the epi-layer compatibility problem. The photodiode can share the same epitaxial layers with either a LED or a MQW modulator since both have p-i-n structures. The MSM detector can be made by simply depositing two Schottky contacts on undoped GaAs, which can be found on the buffer GaAs layer grown on top of the substrate. The relative disadvantage of these two detectors is their low efficiency. For p-i-n photodiodes, the reverse bias voltage totally depletes the intrinsic layer and results in high electric field there. When photons are absorbed by GaAs, electrons and holes are generated. If these carriers are generated within the high electric field region, they will be swept toward the electrodes before recombining with each other. This is how photocurrent is generated. For MSM detectors, we have a similar situation except that the high electric field region is depleted by two Schottky junctions and the applied voltage. Because photocurrents are generated from sweeping the photo-created carriers, the detector gain can not exceed unity gain. The result is higher optical switching power and lower neuron density. Therefore, they are not good choices for high density arrays either.

In contrast to these devices, the OPFET meets the two requirements of epi-layer compatibility and high detector gain. The main disadvantage is its relatively low speed, typically ranging from 100μsec to 10μsec as we will see later. However, for optical neural network applications, the operation speed of the whole system is usually limited by the power consumption on the highly parallel OEIC rather than the intrinsic speed of the individual GaAs components. Therefore, the low speed of optical FET detectors is not a major concern. This makes the OPFET an ideal choice as the high gain photodetector in an OEIC neuron array.

The optical response of an GaAs MESFET detector was first reported in the late 70's [88, 89]. Two different explanations were proposed at that time: the photoconductive effect [89] and the gate voltage modulation mechanism [90]. We find that depending on the epi-layer structure and the fabrication process, different FET detectors have different performance and different mechanisms. This chapter is therefore divided into three sections, corresponding to the three fabrication approaches we have investigated, including the in-house fabrication, the AT&T FET-SEED process, and



(a) cross-section

GaAs	n+ = 2e18	1000A
GaAs	n- = 5e16	3000A
GaAs buffer	undoped	4000A
substrate		

(b) epitaxial layers

Figure 4.1: The cross-section and epitaxial structure of an OPFET built using the in-house fabrication process. Doping concentrations are in units of cm^{-3}.

the MBE regrowth on MOSIS GaAs chips.

4.1 OPFET from In-house Fabrication Process

4.1.1 Structure and Fabrication

The cross-section of an OPFET is shown in Figure 4.1(a), which is basically a GaAs MESFET without the gate Schottky contact. We can use the same MESFET epitaxial layers and the same fabrication procedure except that a different recess etching is needed in order to control the OPFET conducting channel thickness and the detector sensitivity.

The fabrication procedure was described in Section 2.2 and is outlined here. First, starting with the epitaxial structure shown in Figure 4.1(b), the chip was cleaned and

mesas were defined by etching into the substrate using phosphoric-acid-based etchant. Following the thermal CVD deposition of silicon nitride and the nitride opening by CF_4 plasma etching, drain and source contacts were defined by lift-off technique. The contact metal was AuGe/Ni/Au deposited using e-beam and thermal evaporation. It was then annealed at 410°C to reduce the contact resistance. The last step was the gate recess etch to control the channel thickness. The etching depth was measured using a mechanic surface profiler and also monitored by measuring the drain-source current of the FET. The whole fabrication process was finished once the preset drain-source saturation current was reached.

4.1.2 I-V Characteristics and Detector Responsivity

When there was no light illumination, the OPFET operated just like a MESFET. Even though no gate contact was deposited, there existed a depletion layer due to the GaAs surface states. In the previous chapter, we discussed the problem of surface recombination centers, which was a major concern for LED's operating at low driving current levels. Assuming a surface state density of 10^{12}cm^{-2} and a channel doping level of $5 \times 10^{17}\text{cm}^{-3}$, the thickness of the surface depletion layer is about $10^{12}/(5 \times 10^{17})$ cm $= 0.2\mu$m. When a positive voltage was applied on the drain contact, electrons can travel through the undepleted channel and we have a fixed-channel MESFET. When the recessed gate region was illuminated with light, we observed a sharp increase in the FET drain-source current. The I-V curves of one such OPFET, which is very similar to that of an ordinary MESFET, is plotted in Figure 4.2. The dark current was 64nA at $V_{gs} = 2.0$V. For 5.33nW input power, the drain-source current is about 12.5μA, giving us a responsivity of more than 2300A/W.

The responsivity of the OPFET, defined as dividing the photocurrent, which is the difference between the total drain-source current and the dark current, by the optical input power, is plotted in Figure 4.3 as the function of input powers for OPFET devices with different dark currents. Data were taken at a fixed drain-source voltage of 2.0V. The input beam was a focused laser beam from a GaAs/AlGaAs laser diode

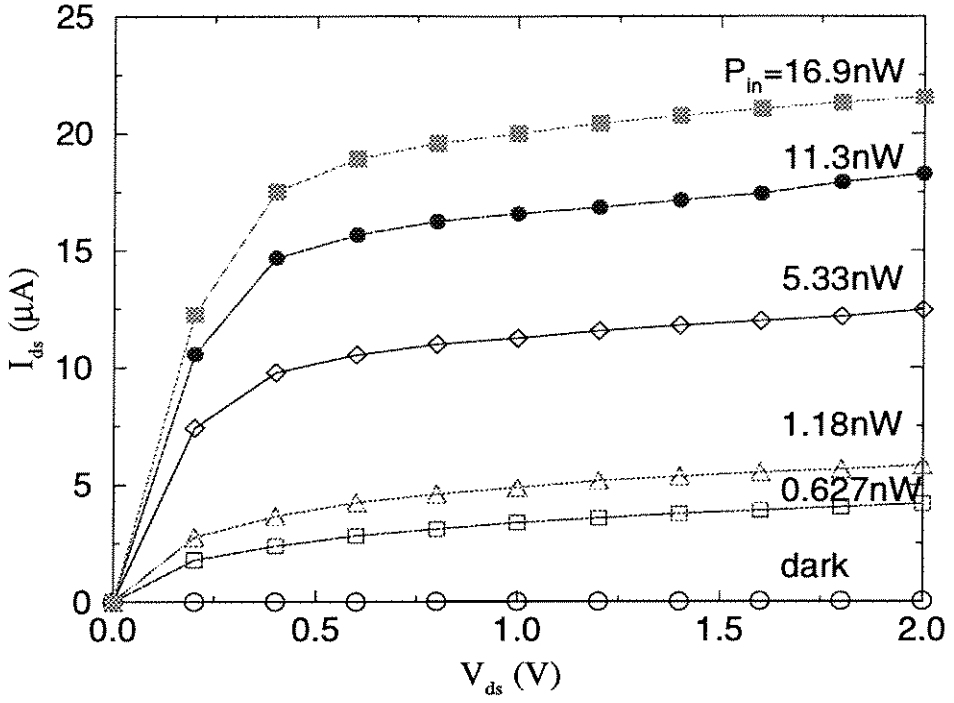


Figure 4.2: The I-V characteristics of an OPFET, fabricated using the in-house process, at different optical input power. The gate region, from drain to source, has a length of 6μm and a width of 20μm. The dark current is 64nA at $V_{ds} = 2.0$V.

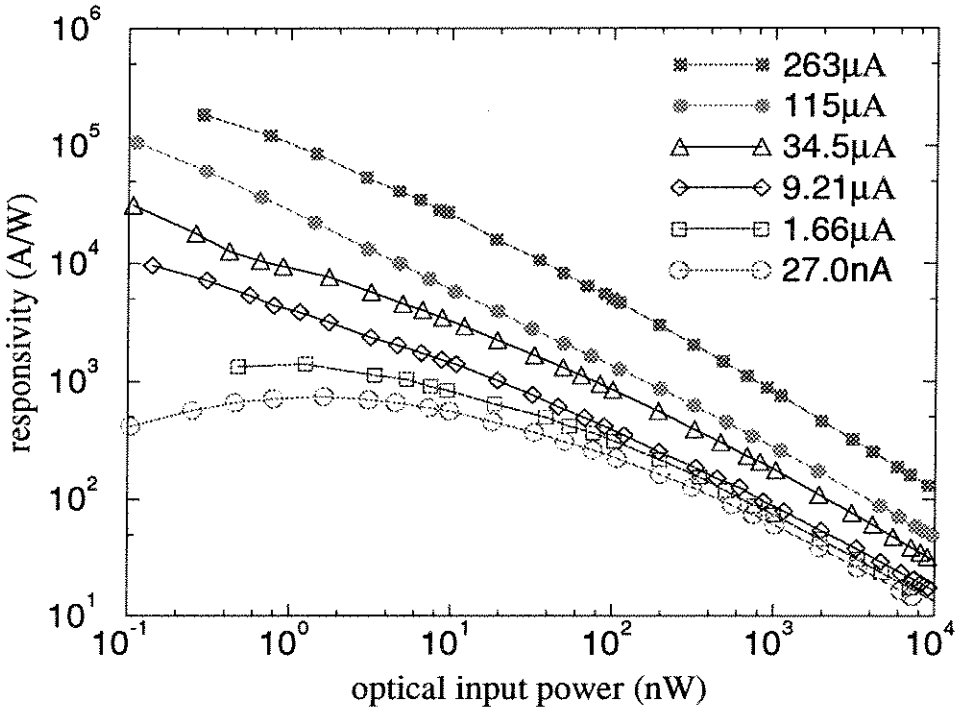


Figure 4.3: The responsivity of OPFET's as a function of the input power at different dark currents (shown in the legend), corresponding to different channel thickness.

with a wavelength of about 837nm. The focused spot had an estimated diameter of $30 \sim 40\mu$m. Different responsivity curves were plotted, corresponding to different recess etching depths.

From Figure 4.3, we find that the detector efficiency is not a constant, but depends on the input power. As the input power increases, the gain goes down. At an optical input power of 10nW, the responsivity is in the order of $10^3 \sim 10^5$A/W, depending on the dark current. For devices with large dark current or large input power, we have roughly $\eta = I_{ph}/P_{in} \propto P_{in}^{-2/3}$, or $I_{ph} \propto P_{in}^{1/3}$. When the channel layer is etched further, the detector gain will decrease. However, the amount of gain change is much less than the amount of dark current change. For example, when the dark current changes from 9.21μA to 27nA, which is reduced by a factor of more than 300, the responsivity is reduced by a factor of only 2.6 for 10nW input power. The reduction factor increases to 27 for 0.1nW input power, which is still much less than the reduction factor for dark currents. We also find that the gain decrease due to dark current reduction is much larger at low input powers than at high input powers. As the result, devices with low dark currents show a maximum gain at a certain input power. For input powers stronger or weaker than this optimum level, the detector gain starts to decrease.

4.1.3 Mechanism

There are two possible mechanisms for the observed detector gain in an OPFET. The first one is the photoconductive effect [89]. When photons are absorbed in the conducting channel, electrons and holes are created. Due to the high electric field, free electrons can be quickly collected by the drain contact. The transit time is given by:

$$\tau_t = \frac{L}{v_n} = \frac{L}{\mu_n \mathcal{E}} \ . \tag{4.1}$$

where L is the gate length, v_n is the electron velocity, μ_n is the electron mobility and $\mathcal{E}$ is the electric field. On the other hand, the holes can remain inside the channel region for a much longer time, due to their slower velocity or being trapped by hole traps. In order to maintain charge neutrality, many electrons have to be injected

from the source contact and the drain-source current can increase a lot. Assuming the average lifetime of holes in the conducting channel is τ_r, the photoconductive gain is then given by:

$$\eta = \frac{\tau_r}{\tau_t} = \frac{\tau_r \mu_n \mathcal{E}}{L} . \tag{4.2}$$

For example, if the electron velocity is 10^7cm/sec, its transit time is only 60psec for a 6μm gate. Suppose the holes are trapped by traps with an average lifetime of 100nsec, the photoconductive gain can be as high as 1.7×10^3.

The other mechanism responsible for the high OPFET responsivity is the gate potential modulation due to incident light. We mentioned before that a surface depletion region was formed due to the GaAs surface states. When electrons and holes are created due to photon absorption, some of the holes will be collected near the surface depletion region. The result is the increase of the surface potential. In return, the surface depletion region shrinks and a larger amount of current is allowed to flow through the conducting channel. High detector gain can be achieved through the amplification mechanism of a MESFET.

In order to find out which mechanism is responsible for the observed results, we need to carefully examine the experimental result presented in Figure 4.3. All the responsivity curves show similar behavior at high input powers. However, at low input levels, we can see a big difference between devices with large dark currents and those with low dark currents. If the photoconductive effect is the predominant mechanism at low incident powers, we should not observe such a result. From Eq. 4.2, when the photoconductive effect is due to the difference between the velocities of electrons and holes, we should observe a constant photodetector gain. When holes are trapped by traps inside the conducting channel, such volume effect will result in decreased detector gain as the input power is raised due to the change of hole lifetime. Depending on the trap distribution, the photocurrent is found to be proportional to P_{in}^{α}, where α ranges between 0.5 to 1 [91]. If such a volume effect is the main mechanism, a linear dependence of the detector gain on the channel thickness should be observed and all devices should show similar behavior, which is in contrast to the

measurement result.

On the other hand, the depletion region modulation effect can explain the behavior at low input powers. When there still remains a large conducting channel, i.e., when the dark current is high, holes generated in the channel will move toward the surface and be trapped by the surface state there. These accumulated holes will increase the surface potential, which acts like applying an positive gate voltage to the MESFET. The detector gain comes from the shrinkage of the surface depletion region due to the change in surface potential. As the input power increases, the increase in surface potential becomes smaller and smaller. Therefore, the detector gain will decrease. When the channel thickness decreases due to the recess etch, the amount of light absorption goes down and the surface potential change is not as strong. As a result, the detector gain at low input levels depends roughly linear to the channel thickness. This is the case when there is a conducting channel.

When we further etch the OPFET channel, finally the conducting layer is totally depleted by the surface states and the MESFET is operating in the subthreshold regime, in which the drain-source current is due to the diffusion of electrons rather than drift. When a small amount of light is absorbed, even though the surface potential increases quite a lot, the gain is still low because not many electrons can flow through the subthreshold channel. It is not until the creation of an above-threshold conducting channel by a large enough input power that we will observe high detector gains. This explains why there exists maximum detector gain at an optimum input power for devices with low dark currents. For input powers lower than this optimum amount, the MESFET channel is not created. For input powers higher than this optimum amount, we already have a large enough channel and the increase in surface potential becomes slower. In both situations, the detector gain is reduced.

In conclusion, the depletion region modulation effect is responsible for the observed behavior at low input levels. As the input power keeps increasing, eventually we will not get a large change in the surface potential. So, as the input power becomes high enough, we believe that the photoconductive effect will take over as the main

mechanism for the high gain.

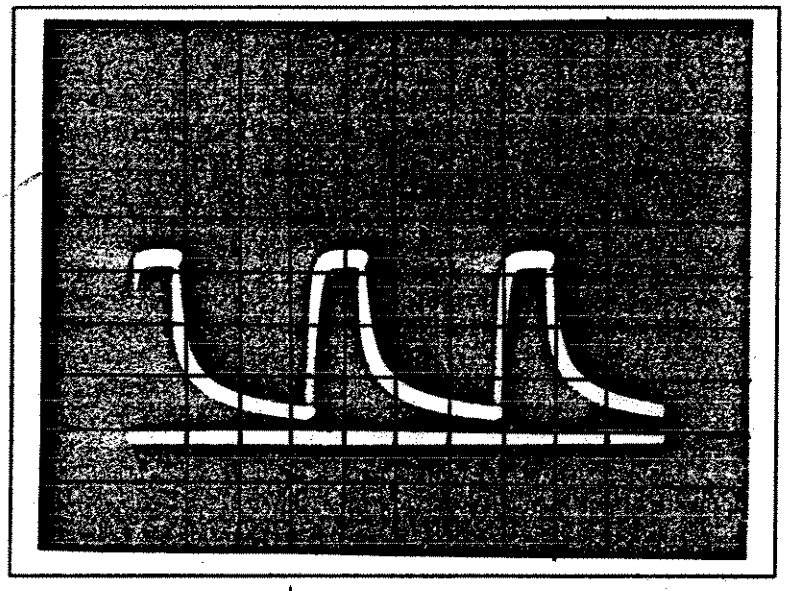
4.1.4 Speed Measurement of OPFET

The above mechanism of the OPFET is further supported by the measured time response of such OPFET, shown in Figure 4.4(a) [65]. It was obtained by applying an optical pulse to the OPFET gate region. We can see that even though the rise time is in the order of 10μsec, the fall time is much longer, in the order of several hundred microseconds, which suggests the involvement of very slow traps. Such a long falling tail can not be attributed to the volume traps in the conducting channel, which usually have faster response time. The only possible explanation is the slow surface states.

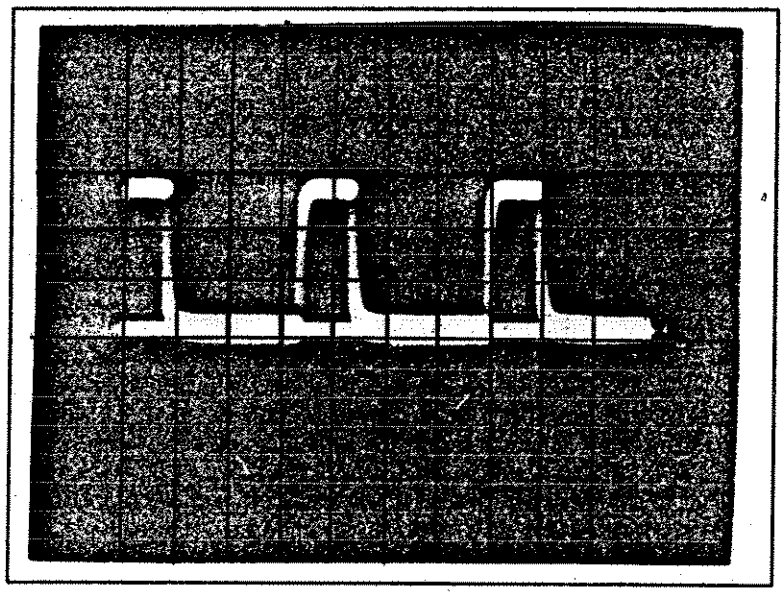
In order to increase the speed of such an OPFET, one can fill all the slow traps by applying a background illumination, as shown in Figure 4.4(b) [65]. We can see that the rise and fall response is more symmetric now and the speed of the OPFET can be increased a lot. The price we have to pay is the decrease detector gain and the increased "dark" current.

4.2 High Gain Detector from FET-SEED Process

The structure and fabrication steps of the AT&T FET-SEED process were described earlier in Section 2.3 together with the device performance. For the FET-SEED process, the only photodetector that had been reported before is the MQW photodiode, which has a responsivity of 0.3 to 0.6A/W, depending on the bias voltage and the operation wavelength. In order to build high density neuron arrays using this industrial process, some sort of high gain photodetector is needed. For the in-house process, we have used the GaAs MESFET as a photodetector to get responsivities higher than 1000A/W, in which the absorption comes from the FET conducting channel consisting of bulk GaAs material. However, for the FET-SEED process, the FET has a GaAs channel of only 120Å thick and the gate region is always covered with Schottky contact metal due to the requirement of self-alignment process, so it is unlikely to give



(a) Without background illumination 0.1msec/div 0.1V/div



(b) With background illumination 0.1msec/div 0.05V/div

Figure 4.4: Time response of an OPFET with and without background illumination.

any significant absorption for the incoming light. We need to find other mechanisms to achieve high gain detectors.

4.2.1 Gate Voltage Changed by Photodiode

One idea that came into our mind was to modulate the FET gate-source voltage by using the photo-generated voltage on an open-circuit p-i-n photodiode. When the light is absorbed by a photodiode, a photocurrent proportional to the input power, P_{in}, is generated if the circuit is shorted. When the diode is open ended, this photocurrent will then charge up the diode voltage, with the p-side being positive. This is called the photovoltaic effect. At equilibrium, we have:

$$\eta P_{in} = I_0 \left(e^{qV_{ph}/nkT} - 1 \right) \tag{4.3}$$

where η is the photodiode responsivity as defined in previous section, I_0 is the reverse saturation current and n is the ideality factor, between 1 and 2, of the diode. Rearranging Eq. 4.3, we have:

$$V_{ph} = \frac{nkT}{q} \ln \left(1 + \frac{\eta P_{in}}{I_0} \right) . \tag{4.4}$$

Figure 4.5 shows the measured result of the photo-generated voltage for a $10 \times 10 \mu m^2$ MQW diode. In the following experiment, the optical input beam was always from a GaAs/AlGaAs laser diode, unless specified otherwise. When this laser diode was lasing, the output wavelength is near 837nm. For small P_{in}, V_{ph} increases very fast. Eventually, the relationship becomes logarithmic, which can be confirmed if we plotted Figure 4.5 in a log-linear scale. At zero P_{in}, the measured voltage is not zero as expected from Eq. 4.4 because the charging/discharging current is so small that a very long time is needed to settle the measurement setup.

As we can see from Figure 4.5, an optical input as small as 50nW is able to generate a voltage in excess of 0.7V. If this voltage is applied to modulate the gate-source bias of a FET, the expected change in the FET drain-source saturation current would be

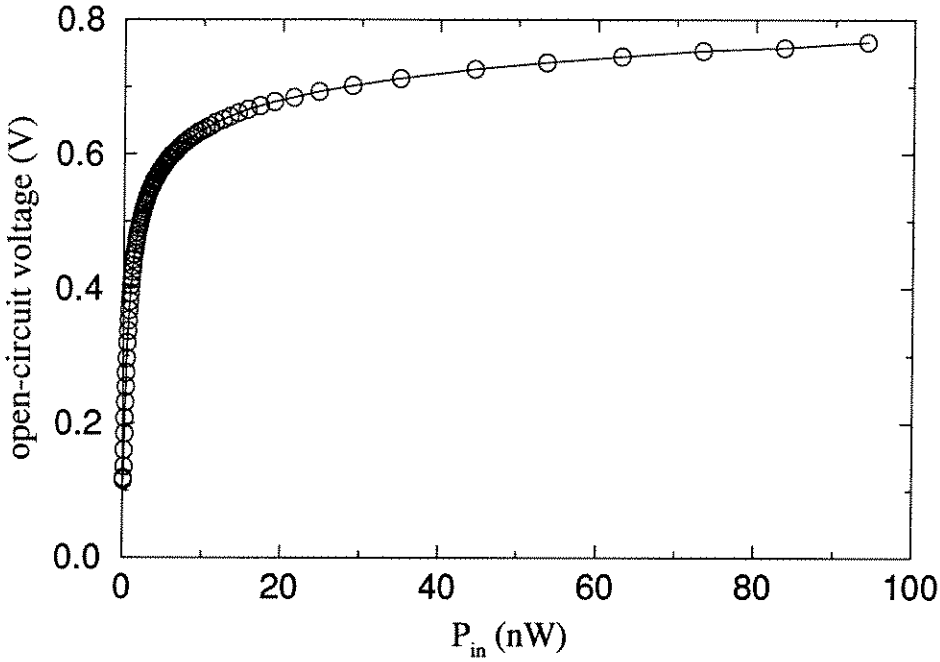


Figure 4.5: Photo-generated voltage of an open-ended MQW photodiode with the p-side being positive. The device is $10 \times 10\mu\text{m}^2$. For $P_{in} > 10\text{nW}$, it is found that the open-circuit voltage increases logarithmically.

approximately:

$$\Delta I_{ph} = \Delta V_{ph} \times g_m \ . \tag{4.5}$$

For a 10μm wide FET, $g_m = 750\mu\text{S}$, we have $\Delta I_{ph} \sim 0.5\text{mA}$, which corresponds to a differential responsivity of $0.5\text{mA}/50\text{nW} = 10^4\text{A/W}$. In experiment, we designed two such FET detectors in tow configurations: one with the source connected to the p-side of a MQW diode and the gate to the n-side, called "v-" configuration; the other one is its reverse case, called "v+" configuration. In Figure 4.6, we plot the FET drain-source saturation current, at $V_{ds} = 2.0\text{V}$, as a function of the optical input power on the diode. As we can see, very significant changes in I_{ds} is achieved with small input powers. The differential gain is up to 10^4 for $P_{in} < 10\text{nW}$. As the input power becomes larger, the change in I_{ds} is not as strong because the photo-generated gate voltage increases only logarithmically. Since for large neuron arrays, we want to push down the operation powers, the high responsivities at low input power is good news.

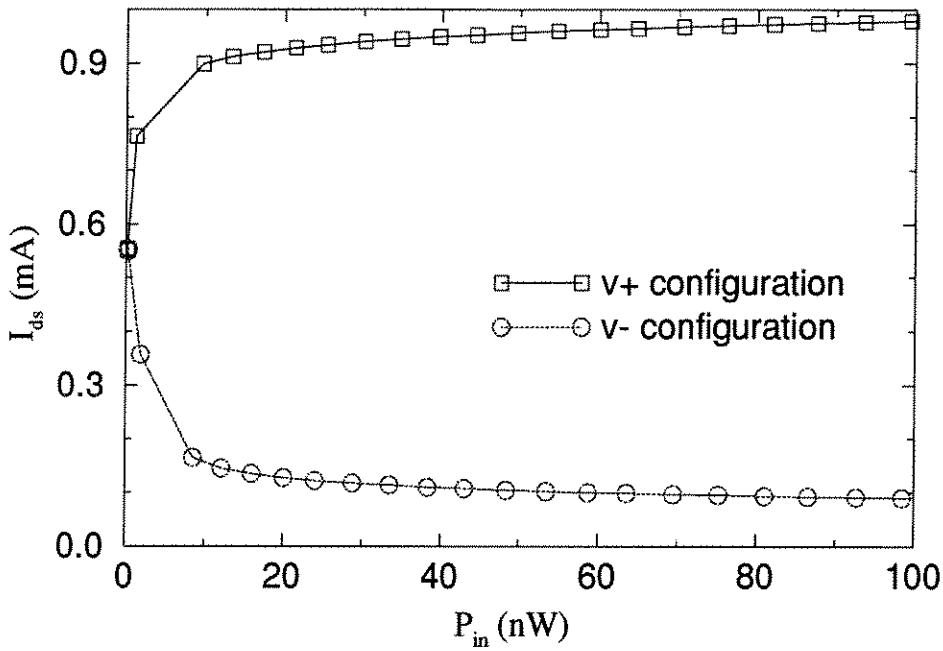


Figure 4.6: The response of 10μm wide FET by using the photo-generated voltage of $10 \times 10\mu m^2$ MQW photodiodes to modulate the gate-source voltage. Circles are for the "v-" configuration where the n-contact of the diode is connected to the gate and p-contact to the source. The reversed "v+" configuration is represented by squares.

Unfortunately, there is one major problem for the v+ and v- configuration. For optical neural network applications, the most commonly used neuron response is the threshold function. In this case the response from the FET detector will be compared with a threshold unit to determine the output as shown in Figure 1.2. For v+ configuration, this is a big problem because the relative change in I_{ds} due to optical input is small even though the absolute value of such change is very significant. As the result, threshold units with v+ configuration FET detectors gives a very "soft" threshold response, which reduces the optoelectronic gain of the neuron circuit. The v- configuration is better for this purpose, but the lowest current level is still too high for a photodetector and results in higher power consumption. In addition, both configurations require spaces for one FET, one photodiode, and the connection metal lines. This is a waste of the precious area for large neuron arrays.

4.2.2 Optical FET Detectors

There are solutions for the above two problems. For the first one, if we can somehow apply a bias so that the FET is operating in sub-threshold regime, then the current level at zero optical input will be very small. Then the v+ configuration can be applied to modulate the gate-source voltage and a high gain photodetector with low dark current can be achieved. This would require a four-terminal FET, which is exactly what we get from the FET-SEED process where the p-contact layer on the bottom can be electrically addressed as a backgate to the FET as we mentioned in Section 2.3.3. By using the MQW structure under the FET channel, we can also save the required area for optoelectronic neuron circuits. In the following, this configuration is referred as optical FET detectors, or OPFET's, since it has similar properties as the OPFET detectors described in Section 4.1.

First Configuration

The OPFET we tested was a 1μm FET with a gate width of 10μm. The spacing between the gate metal and the ohmic contacts were 2μm. This $4 \times 10\mu m^2$ area was the only opening for the incoming laser beams. Limited by the optical system, our focused laser beam had an estimated diameter of $30 \sim 40\mu$m. As the result, only less than 10% of the input power can travel through the uncovered opening and be absorbed by the MQW region. It means that potentially the gain of the OPFET can be almost one order of magnitude higher than what we have observed if this uncovered opening is increased by designing interdigitated gates, which we has successfully implemented for the MOSIS OPFET's described in the next section.

As we mentioned before, a low dark current is desirable for photodetectors used in threshold circuits. Therefore, a negative bias was applied to the gate contact to control the dark current. Instead of connecting the backgate to the source, we left it floating. When light was shined onto the gate region of an OPFET, we observed a dramatic increase in its drain-source saturation current. The I-V characteristic of such OPFET is plotted in Figure 4.7 with V_{gs} fixed at -1.5V, corresponding to $I_{dark} = 0.20\mu$A at

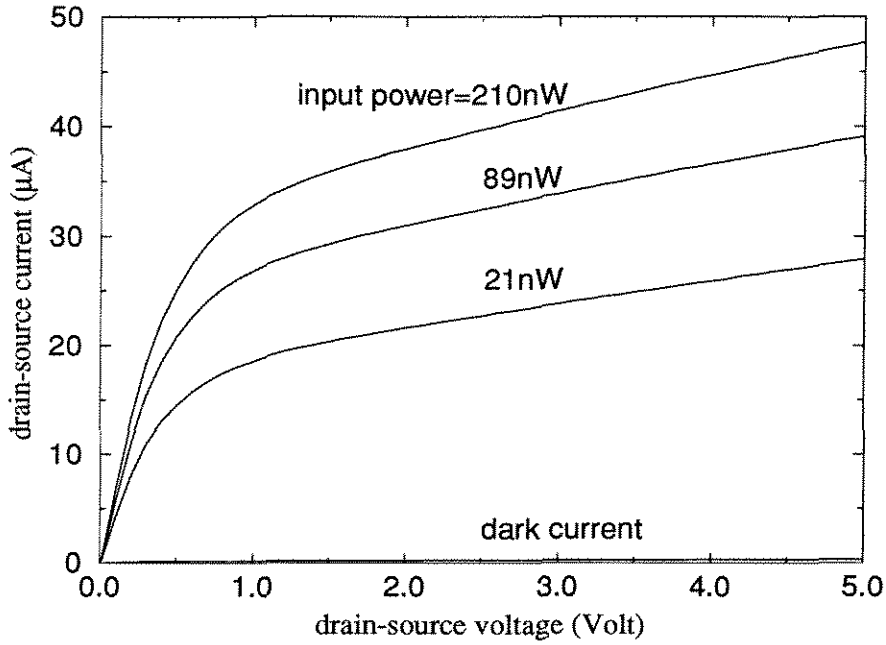


Figure 4.7: I-V curves of an OPFET. Gate voltage is fixed at -1.5V, while the backgate is floating. The dark current is 0.2μA at $V_{ds} = 2.5V$.

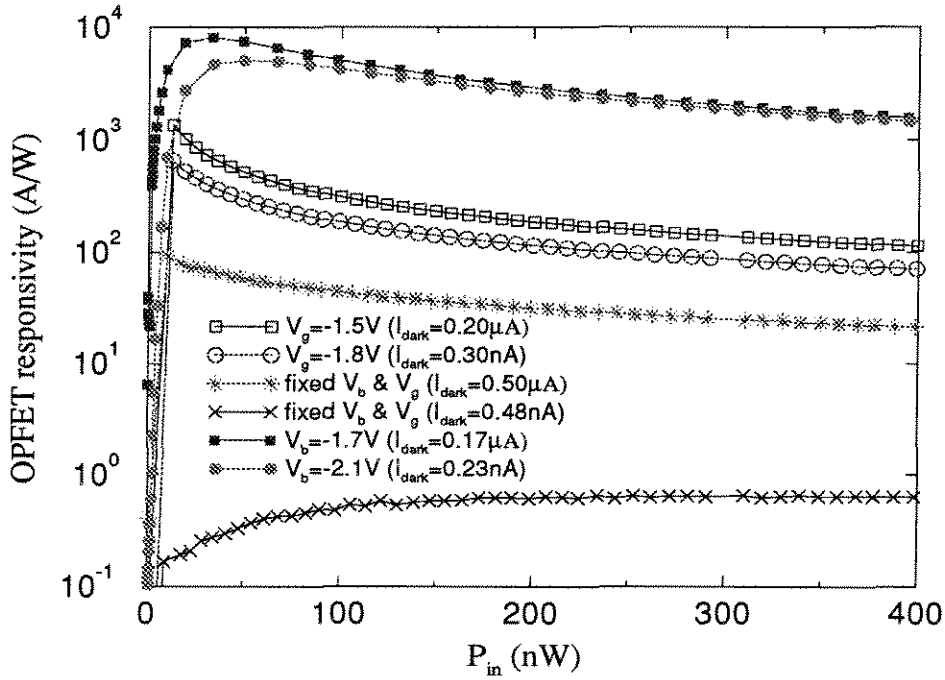


Figure 4.8: The responsivity of an OPFET as a function of optical input power. The result from three configurations are plotted here: open squares and open circles are for fixed gate voltages with floating backgate; filled symbols are for fixed backgate voltage with floating gate; * and $\times$ are for the cases when both gate and backgate voltages were fixed.

$V_{ds} = 2.5V$. It is almost identical to the I-V curves at different electric gate-source bias as shown in Figure 2.12. Therefore, we can think of the OPFET as a FET with optical gate. For 21nW optical input, we measured a photocurrent, calculated from $I_{ds} - I_{dark}$, of more than 21μA at $V_{ds} = 2.5V$, which gives us a responsivity over 10^3A/W. When the gate bias is changed, the measured responsivities will change correspondingly. The relationship between the responsivities at different gate-source bias and the optical input power is plotted in Figure 4.8 with open circles and open squares.

When the optical input power is low, the OPFET has high responsivity, typically in the order of 10^3A/W for $P_{in} < 50$nW. An interesting phenomenon is that its gain does not depend much on its dark current. With the dark current reduced from 0.2μA to 0.3nA, the gain is only reduced by 40%. This suggests that the high gain for the OPFET is not from the 2-dimensional electron gas inside the 120Å channel. Instead, it is the photovoltaic effect from the MQW structure that modulates the backgate voltage, and is then amplified by the FET to give a current output.

To verify this mechanism, we measured the spectral response of the OPFET using a tunable Ti:sapphire laser and compared it with the spectral response of a MQW photodiode. The two spectral responses show identical exciton peaks and absorption edges, as shown in Figure 4.9, which confirms that the MQW structure is where incident photons are absorbed. Actually, we can measure the photo-induced voltage on the backgate contact. It is plotted in Figure 4.10 with filled symbols and is similar to what we got in Figure 4.5. Based on the measured backgate transconductance, we can calculate the expected photocurrent from the photo-induced backgate voltage. The result agrees with our experimental measurement.

Second Configuration

Since the gate and the backgate have some symmetry, we can also operate the OPFET with the gate floating while the dark current is controlled by the backgate bias. The result is also plotted in Figure 4.8 with filled symbols. An even higher gain is observed, but the behavior is different. We observe a peak in the responsivity curve, which is

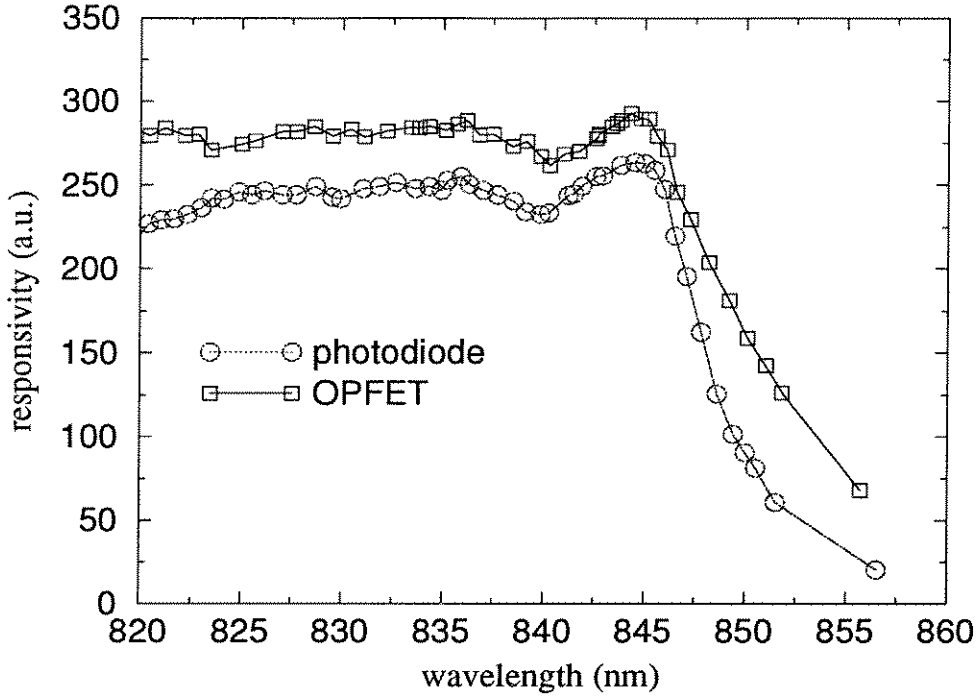


Figure 4.9: The spectral response of an OPFET and a MQW photodiode built using the FET-SEED process. For comparison, their responsivities are plotted in different scales.

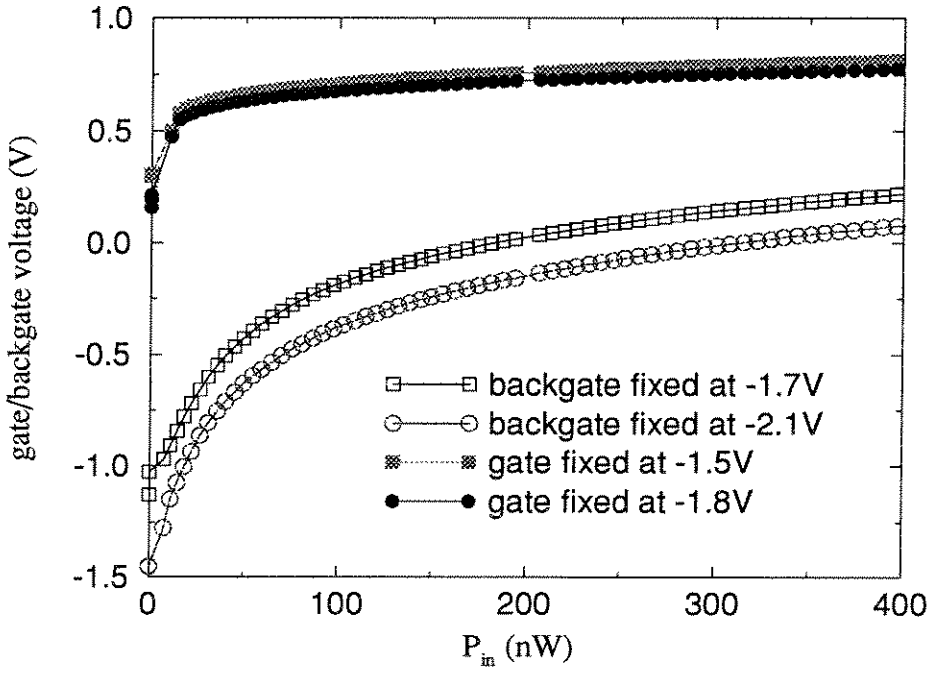


Figure 4.10: Measured photo-induced gate/backgate voltage for the OPFET. The backgate voltage when V_{gs} is fixed is similar to Figure 4.5.

attributed to the peak of FET transconductance when plotted as a function of gate voltage [59].

The mechanism here is slightly different. Since the backgate is now electrically connected to a voltage source, photo-generated holes will be swept away while electrons are moved to the n-channel. These electrons have longer transition time compared to holes, so they will accumulate in the n-channel and result in negative charges. In order to maintain charge neutrality, we believe that positive charges are generated near the gate region through deep levels or surface states. As the result, the gate voltage increases and increased photocurrent is observed. In the experiment, we did observe such voltage increase on the gate contact and it is also plotted in Figure 4.10 with open symbols. Again, we can calculate the expected photocurrent using this gate voltage and g_m measurement. It agrees well with our experimental results.

In the two high gain detector configurations, it is important to have one gate unconnected so that its voltage can be charged up when photons are absorbed. If both the gate and the backgate voltages are fixed, this photovoltaic mechanism is no longer valid and the detector gain is at least one order of magnitude lower as we present in Figure 4.8. Due to photoconductive effect, the FET detector still shows some gain. However, its gain shows obvious dependence on the 2-dimensional gas concentration, represented by the dark current.

Time Response

The time response of the photocurrent was monitored through a current-sense amplifier. The whole measuring system had a parasitic response time faster than 1μsec when we tested it by applying an square-wave electric signal to the gate of the FET. We then illuminated the gate region with a square-wave optical signal. The signal changes from zero to variable high levels. Figure 4.11 shows the resulting time-depending response of a backgate floating OPFET with V_{gs} fixed at -1.8V at P_{in} = 136nW (the high level). We observe an asymmetric rise and fall time response. From Figure 4.11, we measure the 10% to 90% rise time and fall time to be 2.19μsec and 8.80μsec, respectively. The exact reason for this behavior is not yet clear to

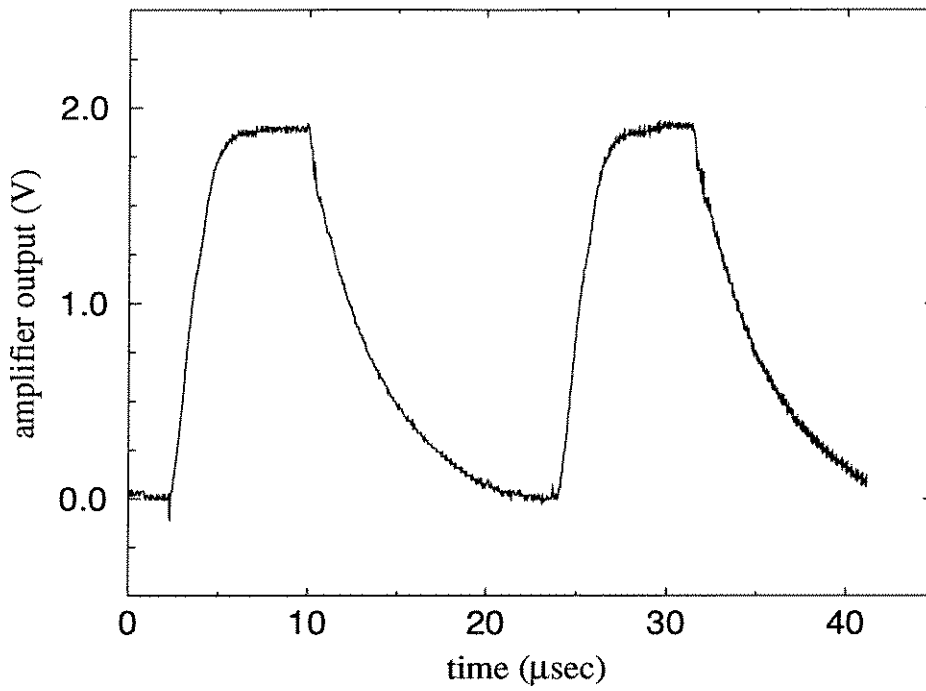


Figure 4.11: Time response of an OPFET measured by a current-sense amplifier with 100kΩ resistor. The optical input power is 136.3nW. Following parameters are used: backgate floating, $V_{ds} = 2.5$V, $V_{gs} = -1.8$V, which gives a dark current of 0.30nA.

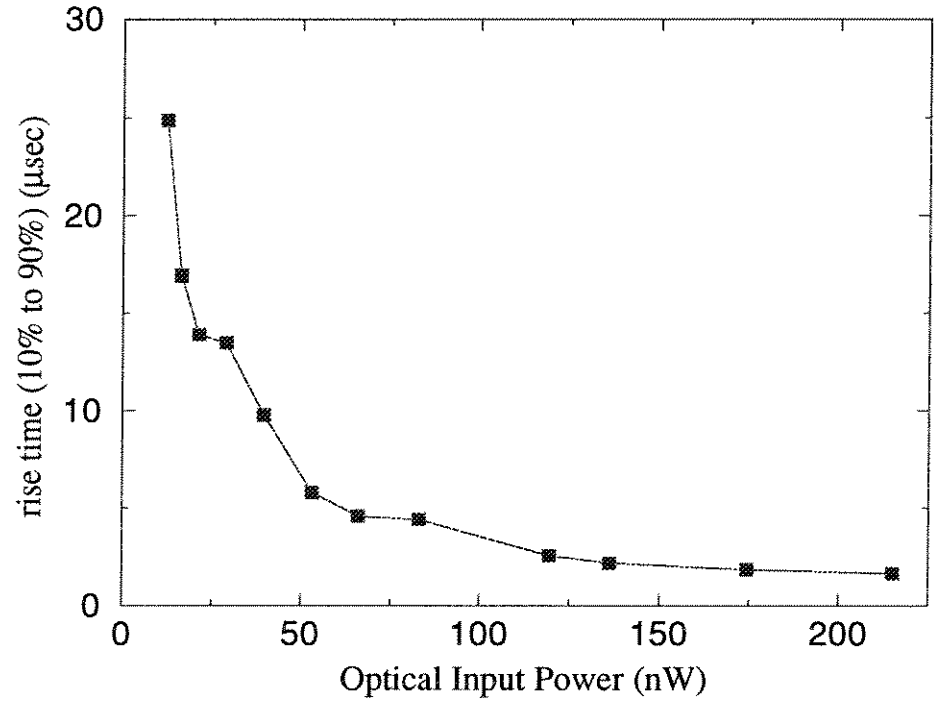


Figure 4.12: The rise time of an OPFET, under the same condition as Figure 4.11.

us. Our guess is that the time response is basically the charging and discharging of the MQW backgate voltage. When P_{in} increases, the photo-induced backgate voltage increases logarithmically as expressed in Eq. 4.4. Since the charging current is proportional to P_{in}, the charging time should decrease. On the other hand, the falling response is due to the discharging of the diode, so it is determined by the MQW diode itself and is therefore longer than the rise time. The exact explanation is difficult to obtain quantitatively because some sort of photoconductive effect is also involved when the 2-dimensional electron gas has high enough concentration, which contributes to the detector gain when both gate and backgate voltages are fixed as explained before.

The rise time of the backgate-floating OPFET at different input levels is plotted in Figure 4.12 and is in the order of 10μsec. As we expected, when P_{in} increases, the device becomes faster. As an estimation, the rise time is approximately:

$$\tau_{rise} \simeq \frac{C\Delta V}{I} \tag{4.6}$$

where C is the capacitance of the i-MQW region, ΔV is the voltage swing on the backgate and is plotted in Figure 4.10, and I is the photo-generated charging current. For $P_{in} = 215$nW, if the focused spot has a diameter of 30μm, only 12.2nW power is absorbed by the MQW region through the $4 \times 10\mu m^2$ opening on the surface, which gives us a charging current of 4.88nA assuming the efficiency of the MQW photodiode is 0.4A/W. The total area under the FET is $16 \times 10\mu m^2$ for our design, giving us a capacitance of 15.6fF for the 1.19μm MQW stack. From Figure 4.10, the voltage swing is 0.56V for $P_{in} = 215$nW. Based on these numbers, we can estimate the rise time to be 1.79μsec, which is very close to our measured result of 1.65μsec. Therefore, we believe the main factor here is the charging of the MQW diode.

For OPFET's with fixed backgate voltage and floating gate, the response time is longer due to the suspected involvement of deep levels or surface states. Also observed is a much longer falling tail when P_{in} increases. Using the same measurement setup, we measured the rise time to be in the order of 100μsec to 1msec. If we apply

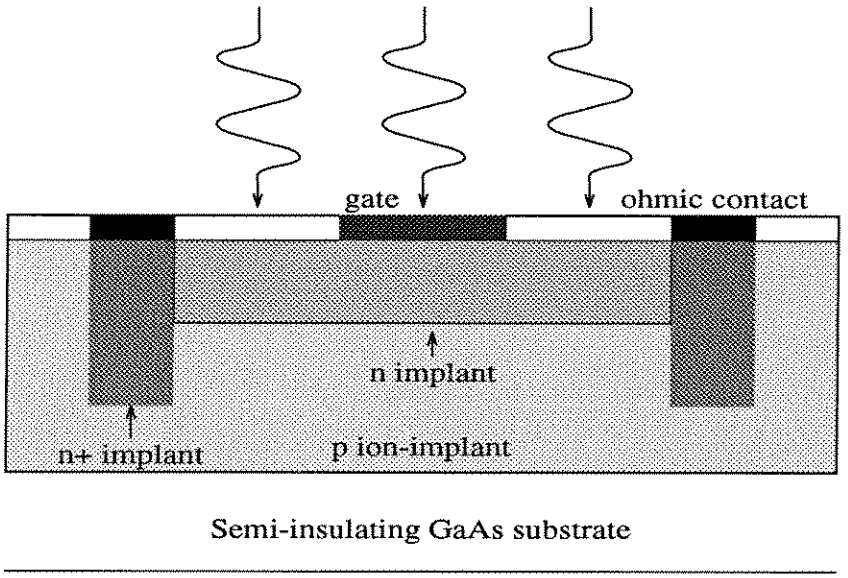


Figure 4.13: Cross-section of the MOSIS enhancement-mode FET used as a high gain photodetector.

a background illumination on the OPFET, the long falling tail is truncated just as what we observed for the OPFET's built in our lab, shown in Figure 4.4. The possible explanation is the saturation of deep level states or surface states.

4.3 OPFET from HGaAs3 Process

4.3.1 Structure and Fabrication

The fabrication procedure of the Vitesse HGaAs3 process was outlined in Section 2.4. This process, available through MOSIS, offers both enhancement-mode and depletion-mode GaAs MESFET's with 0.8μm minimum feature size and four levels of Aluminum interconnect metallization. MSM detectors can be fabricated by defining two Schottky contacts on active area and its performance was presented in Section 2.4 as well. Even though the MSM detector can operate up to very high speed, the major concern is its low efficiency when integrated in high density OEIC neuron arrays.

Since GaAs OPFET's had been successfully fabricated using our own facilities, it was natural for us to investigate the possibility of using MESFET's from the HGaAs3

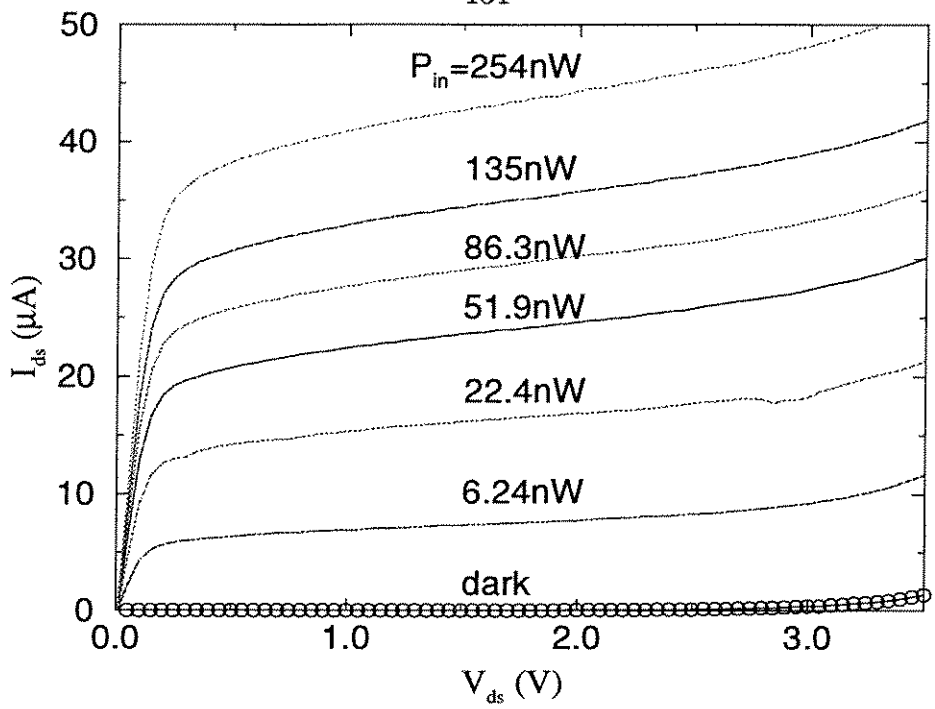


Figure 4.14: I-V curves of a MOSIS OPFET with different optical input powers. Its gate was left floating in order to achieve high detector gain and the dark current was 28.8nA at 1.0V drain-source bias.

process as high gain photodetectors. The cross-section of such device is shown in Figure 4.13. In order to reduce the dark current, only enhancement-mode FET (EFET) was tested. As we described in Section 2.4.1, p-type background ion implant was performed with implantation depth of about 1μm in order to compensate the non-uniformity of GaAs substrate doping levels. The MESFET channel is also defined by ion implant with a depth of 0.3μm. Due to the requirement of self-alignment process, unlike the OPFET's built in our lab, the gate metal must always be deposited as a mask for the drain and source n^+ implantation. Therefore, as a photodetector, the gate-ohmic-contact opening serves as the only light detecting window. Later in this section, we will discuss the geometry design issues, including the optimum distance of gate-ohmic-contact spacing.

4.3.2 Measurement Result

The I-V characteristics of a MOSIS EFET with different electric gate voltages was plotted in Figure 2.16. The EFET has a threshold voltage of 0.30V and the dark

current is typically in the order of nanoamps for 10μm gate width devices at $V_{gs} = 0V$. Figure 4.14 shows the I-V characteristics of an OPFET under different illumination powers from a laser diode ($\lambda = 837$nm) when its gate was left floating. We used a similar testing setup as the one used in testing the AT&T FET-SEED chips. The EFET had a gate width of 10μm and a gate length of 2μm with a spacing of 2μm between the gate and the ohmic contacts. The two $2 \times 10\mu m^2$ opening was the only light detecting window while the focused laser beam has a diameter of about 30 to 40μm. Therefore, the photocurrent can be increased by more than one order of magnitude if the focused input beam matches the opening window as we will show later. The dark current of this FET was 28.8nA at V_{ds}=1.0V with floating gate. Notice how the response due to optical signals is similar to the response due to applying electrical gate bias, plotted in Figure 2.16. Thus, one can think of the OPFET as having an optical gate. The conductance of the channel is modulated by the optical input power, which seems like having the effect of changing the MESFET gate-source voltage. With 22.4nW optical input, the OPFET drain-source current is 15.4μA at V_{ds}=1.0V, which corresponds to a responsivity of almost 700A/W.

The detector responsivity, defined in Section 4.1, is plotted in Figure 4.15 as a function of the optical input power in logarithmic scales. Two curves are plotted for the cases when the gate was unconnected (filled triangle symbols) and when the gate was shorted to the source (filled square symbols). From the plotted curves, we find that the detector gain does not change much for low input powers. For the 2μm gate device, the gain is around several hundred for floating gate configuration. When the gate-source voltage was clamped at 0V, the responsivity was greatly reduced to around 1, which suggests that the gate voltage change is responsible for the high gain behavior. Once the input power is higher than a certain amount, P_{change}, the responsivity shows obvious reduction. The relationship can be fitted as:

$$\eta \propto P_{in}^{-\alpha} \tag{4.7}$$

with α ranges from 0.6 to 0.75, according to our measured results. In experiment, we

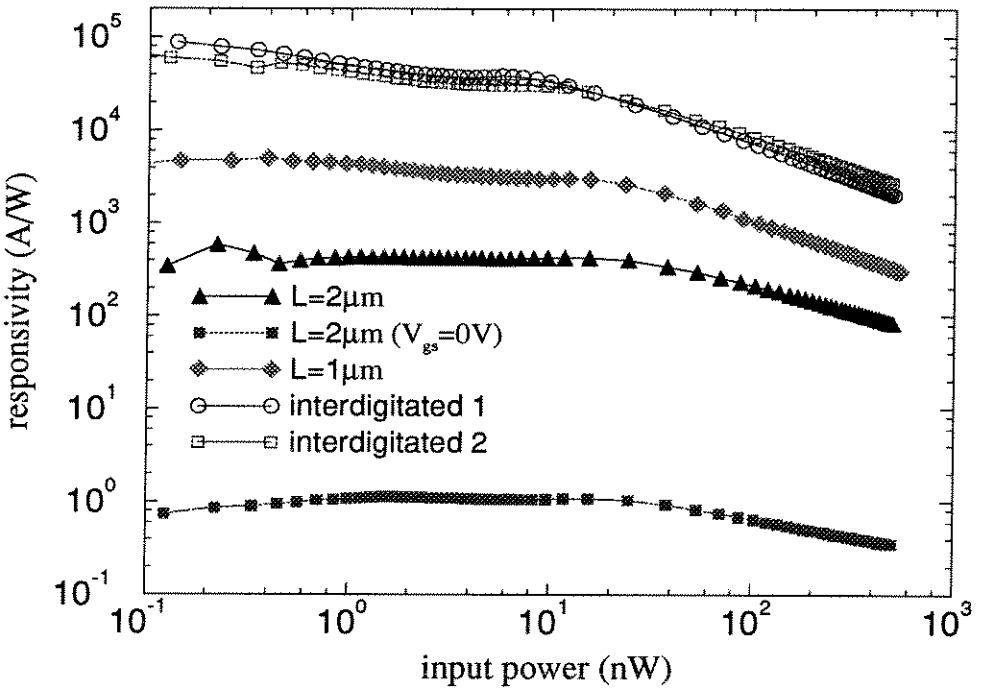


Figure 4.15: OPFET responsivity for a 2μm$\times 10\mu$m device with 2μm gate-ohmic-contact spacing when the gate was floating (filled triangles) or the gate was tied to the source (filled squares). Also plotted is the result of gain improvement by better geometry design discussed later: filled diamonds for a 1μm$\times 10\mu$m device with 8μm gate-ohmic-contact spacing, open circles and open squares for the interdigitated gate design shown in Figure 4.21.

also measured the gate-source voltage induced due to input light. We found out that the value of this P_{change} was closely related to the threshold voltage of the MESFET. Once the induced V_{gs} was larger than the threshold voltage, Eq. 4.7 took effect.

Using the same current-sense amplifier setup as in testing the FET-SEED chips, the time response of a gate-floating OPFET was measured by illuminating the gate region with a square-wave optical signal. This signal changes from zero input to variable high levels. In contrast to the case of OPFET's from the FET-SEED process, we observed a roughly symmetric rise time and fall time. Figure 4.16 shows the measured 10% to 90% rise time as a function of the signal level (high). At 45nW input, a rise time of 28.4μsec was measured, which was higher than the rise time of OPFET's from the FET-SEED process. The OPFET bandwidth shows a roughly linear dependence on the optical input power.

When integrating the MOSIS GaAs circuits with MQW modulators, it is important that the detector works at the appropriate wavelength, typically ranging from 840nm to 850nm. Since the photon absorption of the MOSIS OPFET comes from bulk GaAs material, this is not a problem as confirmed by the measured spectral response, plotted in Figure 4.17. It was measured using the laser output from a tunable Ti:sapphire laser. We notice that the detector gain does drop a little for wavelengths longer than 830nm, however, the reduction is only about 10% for wavelengths up to 860nm. Due to the limitation of output wavelength range of the Ti:sapphire laser, we did not have measured results for longer wavelength. However, we believe that the OPFET response should agree with the absorption edge of bulk GaAs material.

4.3.3 Discussion on OPFET Mechanism

The response of ion-implanted GaAs MESFET's to optical illumination has been studied extensively [92, 93, 94]. The main focus of these papers was on the modulation of threshold voltages by optical inputs and it was believed that the sensitivity of the MESFET's is due to deep-level traps in the substrate. Because the doping profile of commercial GaAs MESFET's is unknown, it is impossible to do computer simulation

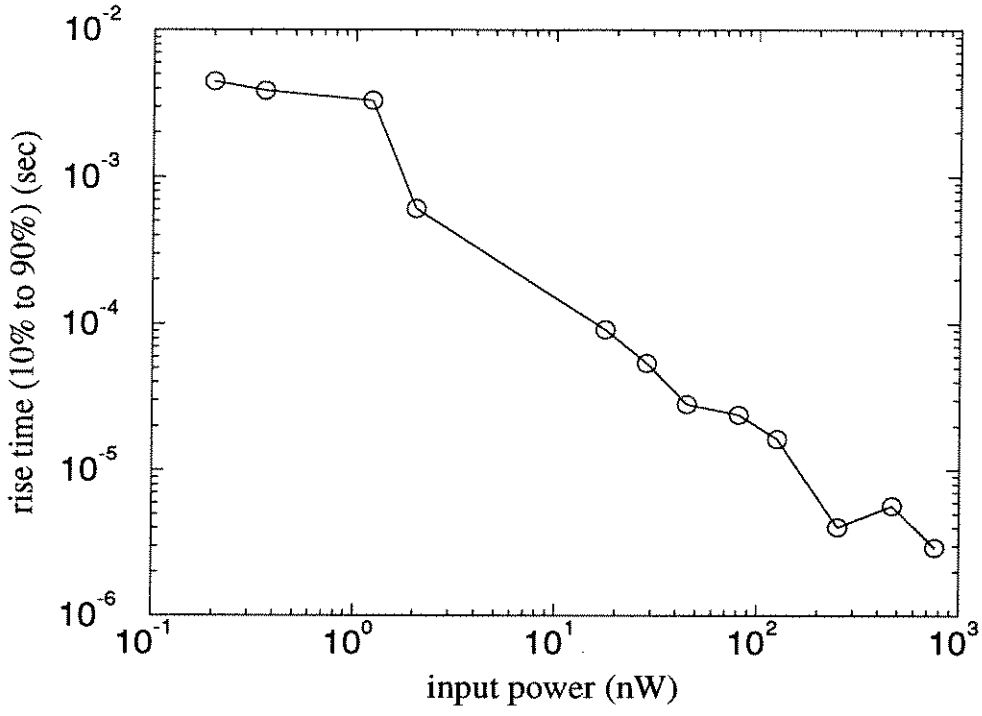


Figure 4.16: OPFET response time with gate floating. Square-wave optical signal was used for the measurement.

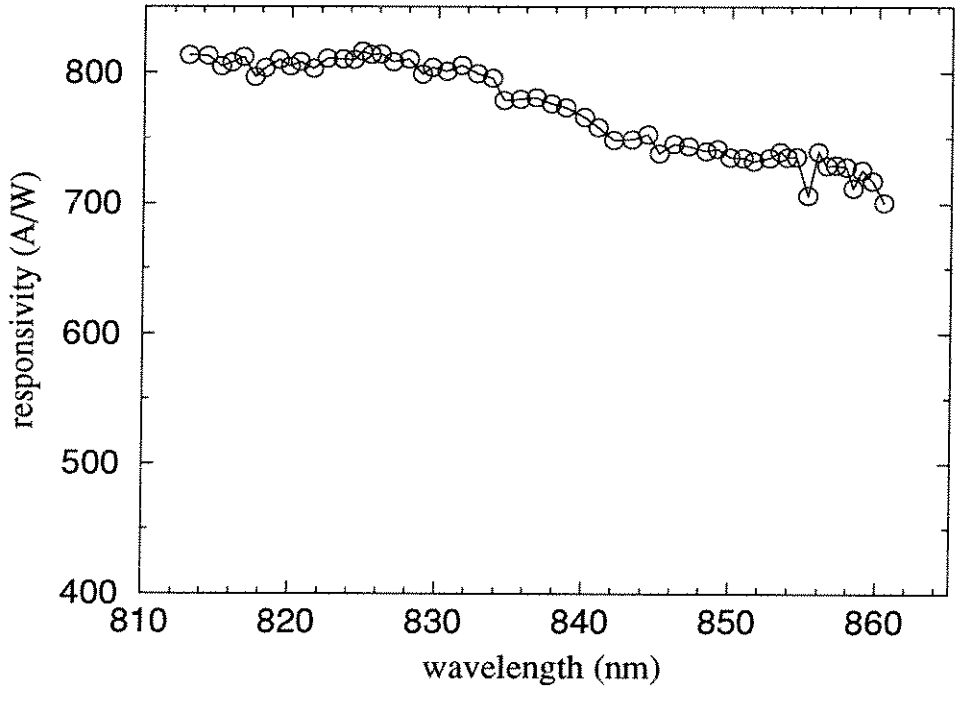


Figure 4.17: Spectral response of MOSIS OPFET.

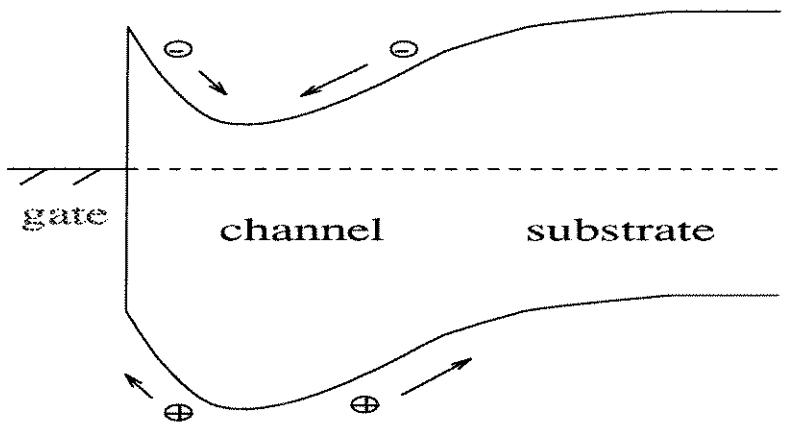


Figure 4.18: Band diagram of an HGaAs3 OPFET, showing the collection of photo-generated carriers.

for the tested devices. However, based on our observation, since the detector gain drops dramatically as the gate voltage is clamped at 0V, the gate Schottky contact must play a very important role here besides the deep-level traps on the backgate side.

Figure 4.18 shows the band diagram of a MOSIS MESFET with p-type background doping. Since the OPFET operates at the sub-threshold regime, the model given in Section 2.1 is no longer correct. In this case, the MESFET conductance is controlled by the electron concentration in the channel region, which is controlled by the gate voltage. In addition, similar to the FET-SEED structure, we have a backgate formed by the channel-p-substrate junction, which can also modulate the transistor conductance. As shown in Figure 4.18, when electrons and holes are created due to the absorption of photons, electrons will move toward the channel region while holes will move either toward the gate contact or the backgate junction. Since both junctions are floating, holes will accumulate there, resulting in the increase of gate voltage and backgate potential. The result is a big change in the channel conductance and high detector efficiency as we observed. When the gate is shorted to the source contact, only the backgate potential is increased by the photo-generated holes. Since the channel electron concentration is mostly controlled by the gate voltage and the backgate bias is only effective when there are significant amount of electrons in the

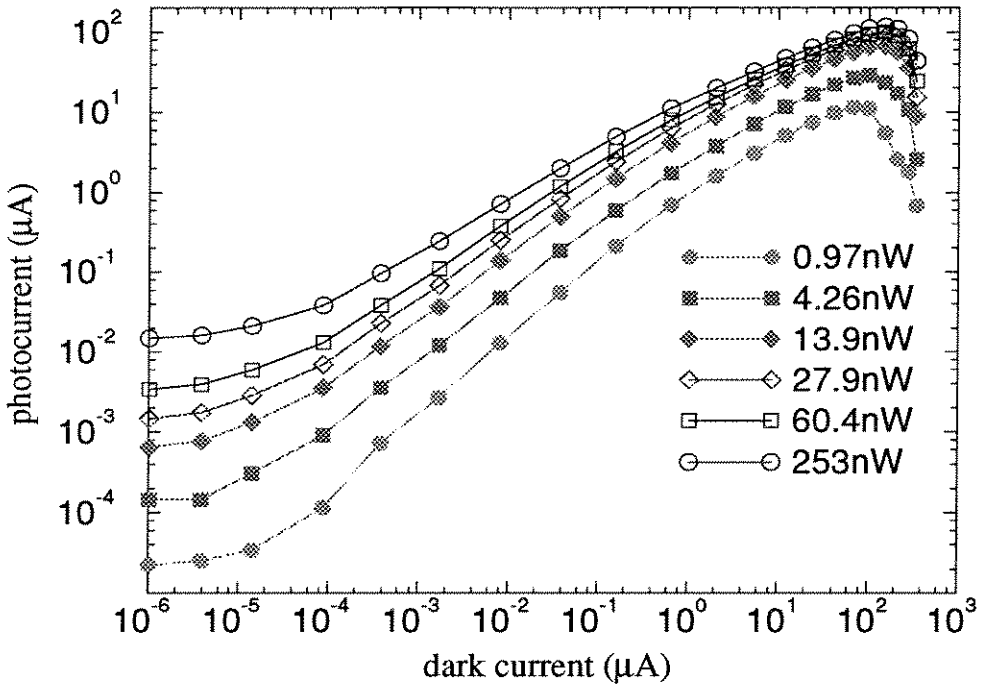


Figure 4.19: The relationship between photocurrent and the MESFET dark current at different optical input powers.

channel, we will not get much increase in the drain-source current. This explains the dramatic reduction of OPFET gain when the gate-source voltage is fixed at 0V.

From the above mechanism, if we fix the gate voltage at higher bias, since there are more electrons in the channel now, we would expect higher detector gain due to the modulation of the backgate junction. Plotted in Figure 4.19 are the experimental results. It was done by measuring I_{ds} at different optical inputs and changing the gate-source voltage from -0.20V to 0.80V by an increment of 50mV. The photocurrent, which is the difference between I_{ds} and the dark current, is then plotted as a function of the dark current, corresponding to different gate-source voltages. For sub-threshold region, i.e., when $V_{gs} \leq V_T = 0.30$V or $I_{\text{dark}} \leq 2\mu$A, we observe an almost linear relationship between the photocurrent and the dark current except at some negative V_{gs} when the photocurrent from the direct collection of photo-generated electrons is too large to be neglected. Once the MESFET channel has been created, the modulation effect of backgate potential becomes less important. At this moment, the electron concentration is roughly determined by the doping level. The effect of gate or backgate bias is to change the thickness of the channel. Up to a certain point when the

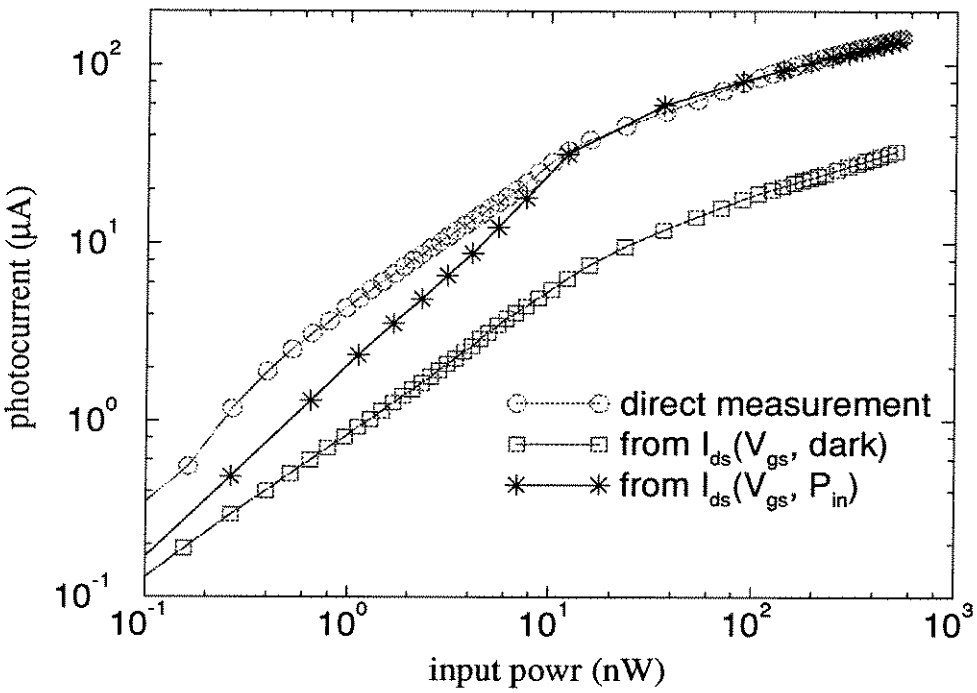


Figure 4.20: The comparison between actual measurement and the deduced results.

channel is extended to its physical limit by implant depth, the increase in backgate potential will have less effect and we begin to observe decrease of photocurrent when increasing V_{gs}.

To test the accuracy of our model, we measured the gate-source voltage induced by the input light when the gate was left floating. The result was similar to the case of FET-SEED OPFET's, shown in Figure 4.10. Since we had also measured the drain-source current at different V_{gs} when no light input was applied, from the photo-induced gate-source voltage we could calculate how much photocurrent was expected. The result is plotted in Figure 4.20 in comparison with the direct measurement result. Both curves show almost identical behavior but the deduced result is about 4.4 times less than the actual measurement. This is because we did not account for the effect of backgate potential increasing. So, when doing the calculation, instead of using $I_{ds}(V_{gs})$ at zero light input, we used the result of drain-source current at different fixed V_{gs} under light illumination and it is plotted in Figure 4.20 as well. For P_{in} >10nW, it agrees almost identically with the actual testing. However, for low input powers, inconsistency still exists. The reason is not yet clear to us. We suspect that it is due

gate length (μm)	1	2	4	8
g_m (10μS)	127.44	47.403	13.495	8.721
η(A/W)	266.8	95.33	25.01	16.34
η/g_m (V/μW)	0.2094	0.1998	0.1853	0.1873

Table 4.1: The relationship between OPFET responsivity and the device transconductance. All devices are 10μm wide and the gate-ohmic spacings are 2μm. The transconductance was measured at $V_{gs} = 0.45$V.

to the error when measuring the photo-induced gate voltage. At low input powers, due to the small opening window, only a small portion of incoming light is absorbed. Therefore, the charging current for the gate voltage is very small and longer time is needed to complete the measurement and we expect a larger experimental error as the result.

4.3.4 Discussion on Gate Geometry Design

Gate Length

Since the high gain of OPFET's comes from the amplification of MESFET gate voltages, we can change the OPFET gain by adjusting the width/length ratio of a MESFET. Table 4.1 shows the measured OPFET responsivity for devices with different gate lengths. All the devices have gate width of 10μm and the gate-ohmic-contact spacing is fixed at 2μm. The transconductance was measured at $V_{gs} = 0.45V$ and the input power was fixed at around 420nW. We can clearly see the linear relationship between the OPFET responsivity and its transconductance. If the FET gate length is reduced, its transconductance will increase as discussed in Section 2.1. In return, the OPFET will have a higher responsivity.

Gate-Ohmic Spacing

Besides the gate length, we can change the gate-ohmic-contact spacing to increase the detector gain. since the focused spot is much larger than the FET open window, if we increase the gate-ohmic spacing, more incoming light can be absorbed by the

gate-ohmic-contact spacing (μm)	1	2	4	8	16
gain at P_{in} = 1nW (A/W)	3130	3550	4390	4810	4120
gain at P_{in} = 320nW (A/W)	531	562	686	654	543

Table 4.2: The relationship between OPFET responsivity and the gate-ohmic spacing. All devices are rectangular-shaped MESFET with 1μm gate length and 10μm gate width.

GaAs channel, which will increase the photo-induced gate voltage and result in higher responsivity. However, there are two undesirable effects if the gate-ohmic spacing becomes too large. First, since the gate junction has a limited depletion region, in the order of 1μm, only those holes and electrons that can diffuse into this depletion region will contribute to the charging of the gate potential. Therefore, once the gate-ohmic spacing increases beyond a certain distance, the gate can not collect more photo-generated carriers and the detector gain will no longer increase. Second, since the gate-ohmic spacing is increased, the gate-source series resistance will be raised, which reduces the MESFET transconductance. For a FET with g_m transconductance, if there is gate-source resistance R_{gs}, the actual transconductance will be:

$$\begin{aligned} g'_m &= \frac{I_{ds}}{V_{gs}} = \frac{I_{ds}}{V'_{gs} + I_{ds}R_{gs}} \\ &= \frac{g_m}{1 + g_m R_{gs}} . \end{aligned} \tag{4.8}$$

In return, the detector responsivity will decrease rather than increase. Table 4.2 shows the result of our measurement for OPFET devices with 1μm gate length and 10μm gate width. The gate-ohmic spacing is 1μm, 2μm, 4μm, 8μm, and 16μm, respectively. We can see that the devices with 4μm or 8μm spacing are the optimum design. Since the effect of gate-source series resistance is more significant for larger currents from Eq. 4.8, 8μm spacing device performs better at low input powers but worse at high input levels compared to the 4μm spacing OPFET. The result of the 8μm spacing OPFET was also presented in Figure 4.15 (filled diamond symbols). The improvement in detector gain is obvious compared to the 2μm gate device with 2μm

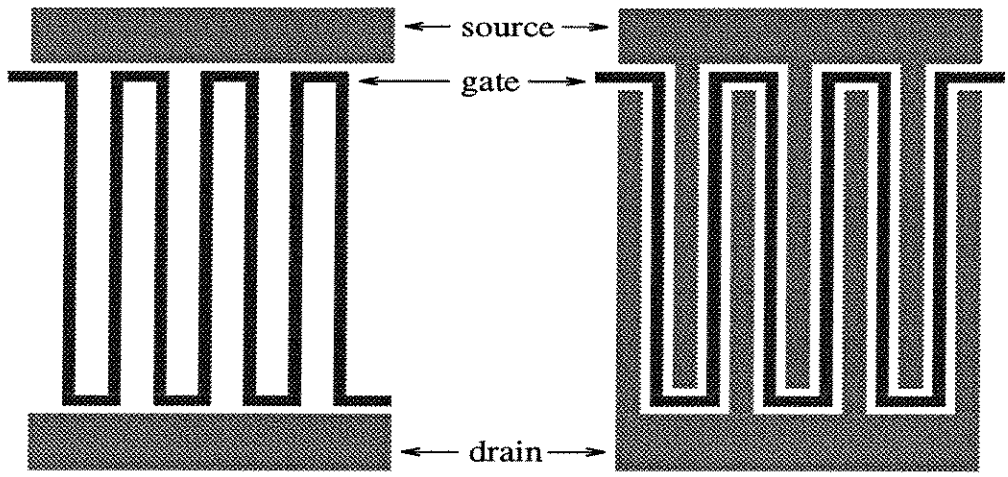


Figure 4.21: Interdigitated gate design to increase the detection area.

spacing: about 10 times larger gain at low input powers and about 4 times higher at high input powers.

Detector Area

In the above discussion, we try to increase the light detecting area by increasing the gate-ohmic-contact spacing. The undesirable effect is the increase of gate-source series resistance, resulting in the decrease of detector gain at high input powers due to the decreased MESFET transconductance. We can avoid this problem by using the interdigitated gate design as shown in Figure 4.21 to match the area of the focused input light spot. Two designs can be implemented, shown in Figure 4.21(a) and 4.21(b). The advantage of the first design is its larger detecting area because no ohmic contact metal is deposited in the center region. The disadvantage is the higher series resistance as compared to the second configuration. The measurement results for both designs were presented earlier in Figure 4.15. The area of the center region is $30\mu\text{m}\times30\mu\text{m}$ for both devices, which is close to the focused laser spot. Compared to the $1\mu\text{m}\times10\mu\text{m}$ device with two $8\mu\text{m}$ gate-ohmic spacings, the responsivities increased by about one order of magnitude. As for the spacing between different gate fingers (for the first design) or the gate-ohmic spacing (for the second design), since the light

first interdigitated gate design

gate finger spacing (μm)	1.5	2	3	4
gain at P_{in} = 1nW (A/W)	3.05×10^4	3.13×10^4	4.99×10^4	4.57×10^4
gain at P_{in} = 320nW (A/W)	2610	2700	2940	2780

second interdigitated gate design

gate-ohmic-contact spacing (μm)	1	2	4
gain at P_{in} = 1nW (A/W)	3.63×10^4	4.26×10^4	2.39×10^4
gain at P_{in} = 320nW (A/W)	4350	3670	2580

Table 4.3: The relationship between OPFET responsivity and the light-detecting spacing for interdigitated gate design. All devices are 30μm×30μm.

detecting area does not increase much by increasing the spacing, the optimum design is different from the rectangular OPFET. Table 4.3 lists the result of devices with different spacings. We find that devices with $1 \sim 1.5\mu$m spacing for each gate finger give the best performance since larger spacing increases gate-source resistance. The undesirable effect of the interdigitated gate design is the increased dark current since the effective gate width is increased to more than 10 times longer. Typically, the dark current is in the order of sub-microamp regime.

4.4 Summary

In this chapter, we discussed the result of fabricating high gain photodetectors. GaAs FET's were successfully applied as detectors with very high responsivity, in the order of more than 10^3A/W. By adjusting the light detecting area, detector gains as high as $10^4 \sim 10^5$ can be achieved. Even though different mechanism is responsible for the different devices fabricated using the three fabrication processes, the basic idea is to change the gate and/or backgate voltage by collecting the photo-generated carriers. Since this process involves the charging and discharging of the gate and/or backgate capacitor, the OPFET has a relatively low speed. When surface states or deep levels are involved, the result is even worse. However, for optical neural network applica-

tions, this is not a major problem. Therefore, the OPFET serves as an ideal candidate for the photodetectors required in high density optoelectronics neuron arrays.

Chapter 5 MQW Modulators

Since the first observation [55] of large optical absorption changes in GaAs/AlGaAs multiple quantum well (MQW) due to an applied electric field, MQW modulators have received considerable attention. Based on the quantum-confined Stark effect (QCSE) [95], various kinds of optoelectronic devices have been proposed and implemented in the past decade, including transmission and reflection modulators [55, 47], self-electrooptic effect devices (SEED's) [54], Fabry-Perot asymmetric cavity modulators [96], etc. Furthermore, there has been great interest in its integration with other electronic devices, including the FET-SEED's [56], which is the SEED on GaAs field-effect transistors described in Chapter 2, and the SEED on silicon CMOS [22]. Similar approaches have also been investigated in InP based, longer wavelength systems [97, 98].

MQW modulators receive so much interest because they offer several advantages: first, a MQW modulator can be very small and has low switching energy. For example, a $5 \times 5\mu m^2$ modulator with $1\mu m$ thickness has a capacitance of about 3.0fF. With a typical driving voltage of 8V, the switching energy is only $\frac{1}{2}CV^2 = 96fJ$. In the case of OEIC neuron arrays, as we discussed in Chapter 1, modulator-based neurons typically consume less power and enable a higher neuron density in a 2-dimensional array compared with those based on light-emitting diodes. Second, the speed of MQW modulators is essentially limited by the time taken to apply the electric field. Operation speed more than 20GHz has been demonstrated by Wakita et al. [99]. Finally, these modulators can be easily integrated with other optical or electronic devices, which enables the implementation of a complex optoelectronic system.

In this chapter, we will discuss the basic mechanism of a GaAs/AlGaAs MQW modulator, with emphasis on the quantum-confined Stark effect. This is basically the summary of some published materials. A simplified model on calculating the change of absorption coefficient near the absorption edge will be presented, serving

as guidance for the SEED structure design. The result of MQW modulators, grown by low temperature MBE at MIT by the Fonstad group and processed and tested in our lab at Caltech, will be given at the end of this chapter.

5.1 Quantum-Confined Stark Effect

The mechanism of a GaAs MQW modulator is the so-called quantum-confined Stark effect (QCSE). Electrons and holes inside a quantum well have discrete energy levels. When a perpendicular electric field is applied, the resulting modulation of the energy bands causes the shift of these discrete levels. As a result, the absorption edge of the quantum well system shifts toward longer wavelengths. Using this electro-optic effect, we can achieve light intensity modulation.

In this section, we will derive a simplified model for estimating the performance of a GaAs MQW modulator. First, the energy levels inside a quantum well without any electric field will be derived. Then we will figure out the amount of Stark shift based on a perturbation approximation. Once the absorption edge is known, an empirical model is applied to predict the absorption coefficient, α, and the resulting change, $\Delta\alpha$, when electric fields are applied. With that, we can predict the modulator contrast ratio and optimize our design.

5.1.1 Quantum Well at Zero Electric Field

With the advance in semiconductor material technology, it is now possible to grow one semiconductor material sandwiched between a different type of material. Normally, these two materials have different bandgap and what we get is a double heterojunction structure as the one we used for the LED discussed in Chapter 3. In that case, an active GaAs layer is sandwiched between two layers of $Al_xGa_{1-x}As$, which has a direct bandgap of [71] :

$$E_g(x) = 1.424 + 1.247x \text{ (eV)} \tag{5.1}$$

for $0 < x < 0.35$. Using techniques such as molecular beam epitaxy (MBE) or metal-organic chemical vapor deposition (MOCVD), we are able to grow very thin layers of semiconductor materials, in the order of nanometers. When that occurs, the GaAs/$Al_xGa_{1-x}As$ double-heterostructure are called a quantum well structure because a thin narrow-bandgap GaAs sees two $Al_xGa_{1-x}As$ barriers and the quantum effect is very significant for such small dimensions. An extension of the single quantum well structure is the multiple quantum well (MQW) which is a multiple layered structure consisting of thin alternative layers of two different semiconductor materials. Figure 5.1 shows schematically a typically GaAs/$Al_xGa_{1-x}As$ MQW and its band structure. In this case, GaAs layers are referred to as "wells" because of the narrower bandgap and AlGaAs layers as "barriers."

For electrons and holes in a single quantum well, their energy levels are discrete in the direction perpendicular to the quantum well layers, usually called the z-direction, due to the quantum effect. These discrete eigen states can be calculated by solving the one-dimensional Schrödinger equation:

$$H_{0z}\Phi_n(z) = E_n\Phi_n(z) \tag{5.2}$$

where $\Phi_n(z)$ is the eigen function, E_n is the eigen energy and H_{0z} is the zero-field quantum well Hamiltonian in the z-dimension, given by:

$$\begin{aligned} H_{0z} &= \frac{\hbar^2 k_z^2}{2m^*} + V(z) \\ &= -\frac{\hbar^2}{2m^*}\frac{d^2}{dz^2} + V(z) \end{aligned} \tag{5.3}$$

where m^* is the effective mass of electrons or holes and $V(z)$ is their potential function. For simplicity, we will only consider the case of electrons in the following discussion unless specified otherwise.

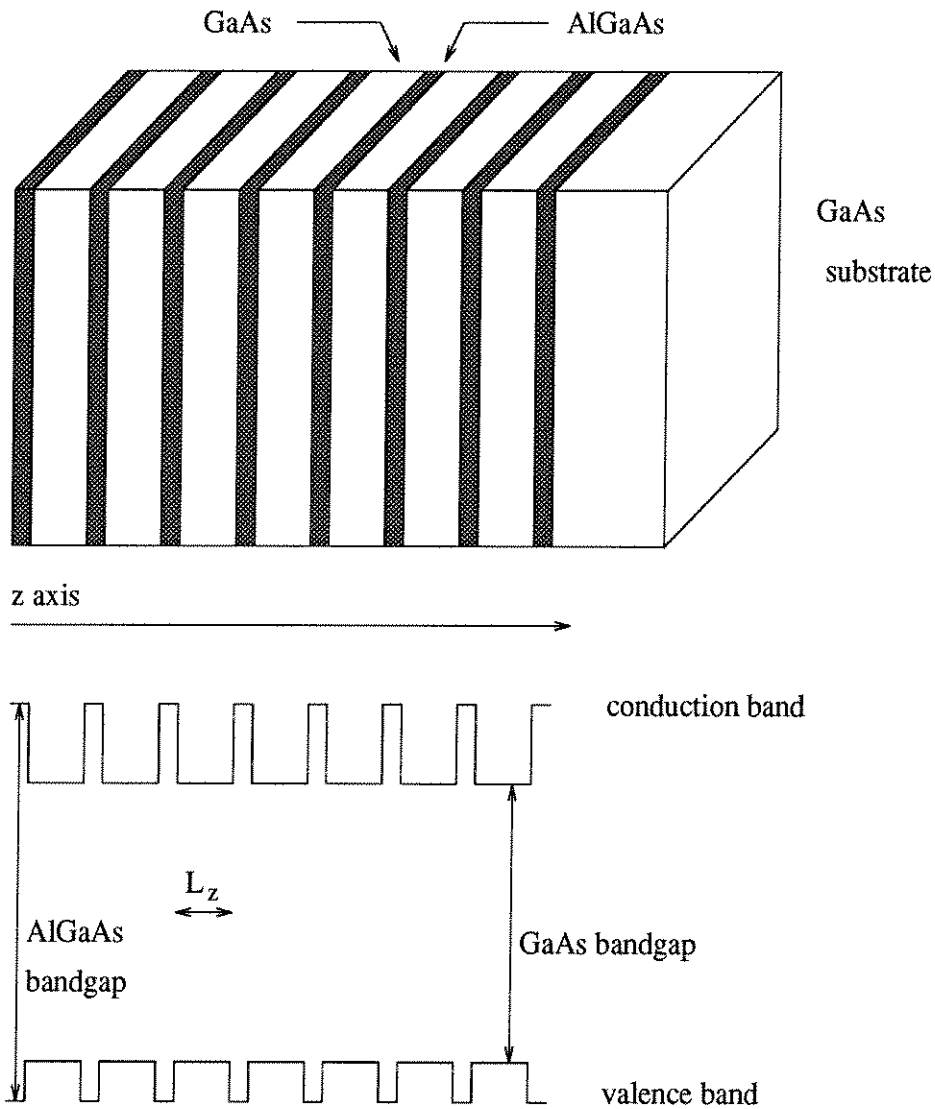


Figure 5.1: Schematic illustration of a GaAs/$Al_xGa_{1-x}As$ quantum well structure. Both the actual layers and the band structure are shown. L_z is the width of quantum wells.

In the case of an infinite quantum well, i.e.,

$$V(z) = \begin{cases} \infty & z < 0 \text{ or } z > L_z \\ 0 & 0 < z < L_z \end{cases} \tag{5.4}$$

where L_z is the width of the quantum well, the eigen equation 5.2 has exact and analytic solutions:

$$\begin{cases} E_n^{\infty} &= \frac{\hbar^2}{2m^*}\left(\frac{n\pi}{L_z}\right)^2 \\ \Phi_n^{\infty}(z) &= A\sin\left(\frac{n\pi z}{L_z}\right) \end{cases} \quad n = 1, 2, 3, \cdots \tag{5.5}$$

In practice, the barrier is limited by the conduction band discontinuity, ΔE_c, for electrons — or the valence band discontinuity, ΔE_v, for holes — and we have a finite quantum well with:

$$V(z) = \begin{cases} \Delta E_c & \mid z \mid < L_z/2 \\ 0 & \mid z \mid > L_z/2 \end{cases} . \tag{5.6}$$

Here we have shifted the z-direction origin to the center of the quantum well in order to simplify the following derivation process. The eigen states can be obtained by solving Eq. 5.2 with the following boundary conditions: Φ and $\frac{1}{m^*}\frac{d\Phi}{dz}$ continuous at $\mid z \mid = L_z/2$ [100]. As a result, the discrete eigen values are given by:

$$\begin{cases} \sqrt{E_n}\tan\left[\frac{\pi}{2}\left(\frac{E_n}{E_1^{\infty}}\right)^{1/2}\right] = \left[\frac{m_w^*}{m_b^*}(\Delta E_c - E_n)\right]^{1/2} & \text{even parity} \\ -\sqrt{E_n}\cot\left[\frac{\pi}{2}\left(\frac{E_n}{E_1^{\infty}}\right)^{1/2}\right] = \left[\frac{m_w^*}{m_b^*}(\Delta E_c - E_n)\right]^{1/2} & \text{odd parity} \end{cases} \tag{5.7}$$

where E_1^{∞} is the ground level energy for infinite wells as given by Eq. 5.5, m_w^* and m_b^* are the effective masses of electrons in the well and the barrier, respectively. The eigen functions are sinusoidal in the well and exponentially decaying in the barriers. For even parity solutions, the wave functions are:

$$\Phi_n(z) = \begin{cases} A\cos(k_w z) & \mid z \mid < L_z/2 \\ A\cos(k_w L_z/2)\exp[-k_b(\mid z \mid - L_z/2)] & \mid z \mid > L_z/2 \end{cases} \tag{5.8}$$

where A is a normalization factor, k_w and k_b are the electron wave vectors and are given by:

$$\begin{cases} k_w = \sqrt{2m_w^* E_n/\hbar^2} \\ k_b = \sqrt{2m_b^*(\Delta E_c - E_n)/\hbar^2} \end{cases} . \tag{5.9}$$

The odd parity wave functions are the same except that we need to substitute sine function for the cosine function in Eq. 5.8. Similar results can be obtained for holes in the valence band by substituting the valence band discontinuity, ΔE_v, for ΔE_c and hole effective masses for m_w^* and m_b^*.

Now, lets look at the optical absorption behavior. When the incident power is not high enough, the absorption is in the linear region and the absorption coefficient is approximately proportional to the density of states for a given photon energy near the absorption edge. For a bulk semiconductor material, the density of states is [101]:

$$\rho_{3D}(E) = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2} E^{1/2} . \tag{5.10}$$

Therefore, the absorption coefficient can be expressed as:

$$\alpha(\omega) = K(\hbar\omega - E_g)^{1/2} . \tag{5.11}$$

However, in the case of a quantum well structure, the electrons and holes are confined by the barriers in the z-dimension. Consequently, the carrier gases are reduced to a 2-dimensional system and we have:

$$\rho_{QW}(E) = \sum_n \frac{m^*}{\pi\hbar^2} H(E - E_n) \tag{5.12}$$

where $H(x)$ is the Heaviside function which is equal to 1 when $x > 0$ and zero when $x < 0$. In addition, besides energy conservation, the other selection rule for bulk material is the momentum conservation, which results in $\Delta\vec{k} = 0$ since the momentum of photons are much smaller than that of electrons or holes. For quantum well structures, the momentum conservation still gives $\Delta k_x = \Delta k_y = 0$ in the 2-dimensional

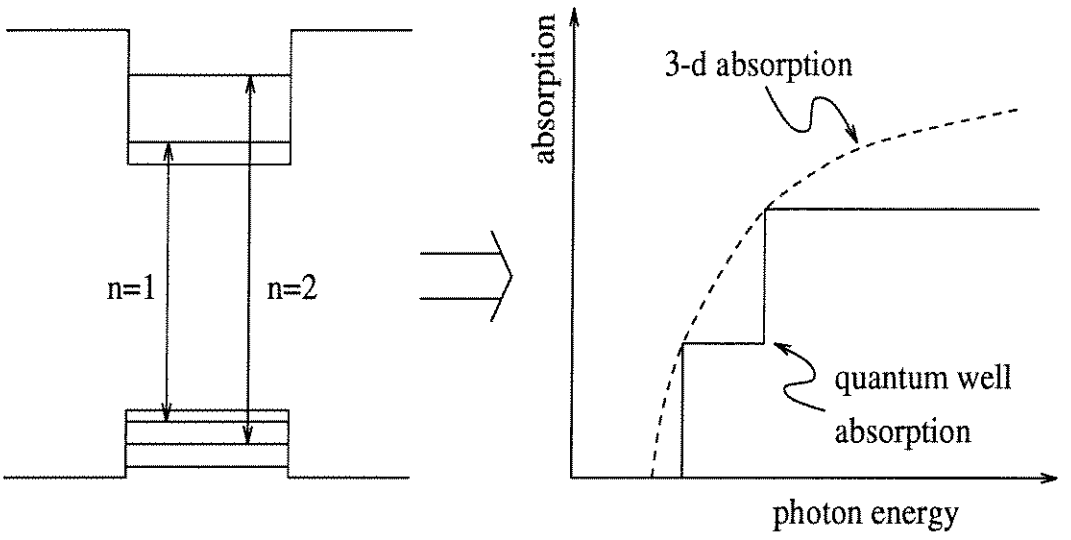


Figure 5.2: Quantization in z-dimension results in a staircase function for the absorption coefficient. In comparison is the result for bulk material.

space parallel to the quantum well. In z-dimension, however, the quantization results in a selection rule: $\Delta n = 0$. Therefore, the absorption edge will happen at $E_g + E_1^e + E_1^h$ and the absorption coefficient becomes a staircase function as shown in Figure 5.2 where the photo energy of each step can be calculated from the result of Eq. 5.7.

To get a feeling for the magnitude, let's assume the following parameters for electrons in a GaAs/$Al_xGa_{1-x}As$ quantum well: $x = 0.30, L_z = 90\mathring{A}, m_w^* = 0.0665m_e, m_b^* = (0.0665 + 0.0835x)m_e = 0.09155m_e$ [102], $\Delta E_c = 213.2$meV, which represents a bandgap difference of 374.1meV with a 57:43 split ratio between ΔE_c and ΔE_v [102]. In early references, a split ratio of 85:15 between ΔE_c and ΔE_v was assumed. However, if we follow the same procedure as described in this section, the resulting Stark shift is higher and does not agree with experimental results as well as the result from assuming a 57:43 split ratio. Therefore, the parameters given in Ref. [102] are assumed. The result for electrons is $E_1^\infty = 69.81$meV and there are two discrete levels for a finite well: $E_1^e = 34.09$meV, $E_2^e = 132.0$meV. Due to the spin-orbit coupling, there are two types of holes in the valence band: heavy holes and light

State	E_n(meV)	$\Phi_n(z)$ for $\mid z \mid < \frac{L_z}{2}$	$\Phi_n(z)$ for $\mid z \mid < \frac{L_z}{2}$
electron 1	34.09	$0.393\cos(0.244z)$	$0.179\exp[-0.656(\mid z \mid -\frac{L_z}{2})]$
electron 2	132.0	$0.377\sin(0.480z)$	$\pm 0.314\exp[-0.442(\mid z \mid -\frac{L_z}{2})]$
heavy hole 1	9.208	$0.427\cos(0.287z)$	$0.128\exp[-1.36(\mid z \mid -\frac{L_z}{2})]$
heavy hole 2	36.69	$0.425\sin(0.572z)$	$\pm 0.228\exp[-1.23(\mid z \mid -\frac{L_z}{2})]$
heavy hole 3	81.64	$0.420\cos(0.854z)$	$-0.321\exp[-0.984(\mid z \mid -\frac{L_z}{2})]$
heavy hole 4	140.0	$0.391\sin(1.12z)$	$\mp 0.371\exp[-0.505(\mid z \mid -\frac{L_z}{2})]$

Table 5.1: Eigen states for electrons and heavy holes in a finite quantum well at zero electric field without considering the Coulomb attraction with each other. z is in the unit of nanometer.

holes, with $m^*_{hh} > m^*_{lh}$. From Eq. 5.5, we know that light holes have higher energy levels. As a result, the absorption edge is determined by heavy holes. For heavy holes: $m^*_w = 0.34m_e, m^*_b = (0.34 + 0.42x)m_e = 0.466m_e, \Delta E_v = 160.9$meV. We have $E^\infty_1 = 13.65$meV and there are four discrete energy levels inside the well: $E^h_1 = 9.208$meV, $E^h_2 = 36.69$meV, $E^h_3 = 81.64$meV, $E^h_4 = 140.0$meV. In return, instead of an absorption edge of 1.424eV — corresponding to a wavelength of 871nm — in a bulk GaAs, the absorption edge of a quantum well is shifted to $E_g + E^h_1 + E^e_1 = 1.467$eV, which corresponds to a wavelength of 845.6nm. Table 5.1 lists the calculated eigen states from Eq. 5.7 and 5.8.

In the above discussion, we calculated the eigenstates for a finite quantum well for electrons and holes, respectively. However, there exists interaction between the electron and the hole due to the Coulomb force. In a bulk material, the total Hamiltonian is:

$$H = H^e + H^h = \frac{\hbar^2 k^2}{2\mu} - \frac{e^2}{4\pi\varepsilon_0\varepsilon r} \tag{5.13}$$

where μ is the reduced mass, given by:

$$\frac{1}{\mu} = \frac{1}{m^*_e} + \frac{1}{m^*_h} . \tag{5.14}$$

We can recognize that this is a hydrogen atom problem and the result is well known

as [103]:

$$E_b^n = -\frac{\mu e^4}{2\hbar^2\varepsilon^2 n^2}\cdot\frac{1}{(4\pi\varepsilon_0)^2} \quad \text{(n=1, 2, 3, } \cdots\text{)} \tag{5.15}$$

The quantization of this interaction is called an exciton. For bulk GaAs, this exciton binding energy E_b (ground level) can be calculated to be 4.2meV. In a quantum well, because electrons and holes are confined in the z-dimension, the Coulomb attraction is enforced due to the reduction in the size of an exciton. The Hamiltonian can be expressed as:

$$H_1 = H_{0z}^e + H_{0z}^h - \frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} - \frac{e^2}{4\pi\varepsilon_0\varepsilon\sqrt{r^2+(z_e-z_h)^2}} \tag{5.16}$$

where H_{0z} is given by Eq. 5.3 and r is the relative position of electron and hole in the x-y plane. We neglect the center-of-mass kinetic energy of electron-hole pairs in the x-y plane because the associated momentum is too large to be observed in a photon absorption process. In the limit of $L_z \to 0$, we can solve the eigen equation of H_1 by separating variables because $z_e - z_h \to 0$, which results in a two dimensional hydrogen system in the x-y plane. The binding energy is:

$$E_b = 4\times\frac{\mu e^4}{2\hbar^2\varepsilon^2}\cdot\frac{1}{(4\pi\varepsilon_0)^2} \sim 17 \text{ meV}. \tag{5.17}$$

When L_z is not zero, the binding energy can be obtained by approximate approaches such as variational calculation [95], which gives 8.6meV for the binding energy of the heavy-hole-electron exciton.

The existence of exciton has two effects on the optical absorption phenomenon: first, the absorption edge is shifted toward lower energy, i.e., longer wavelength, due to the binding energy. Second, exciton peaks will be superimposed upon the staircase absorption coefficient shown is Figure 5.2 with both heavy and light hole exciton peaks. As we will see later, the existence of such exciton peaks increases the contrast ratio of a MQW modulator.

5.1.2 Quantum Well with Applied Electric Fields

So far, we have only considered the case of no applied electric field. When an electric field is applied to the quantum well structure, depending on the direction of the field, different phenomena will be observed.

When the applied field is parallel to the quantum well plane, i.e., in the x-y plane, obviously, the quantization in the z-dimension has no effect and we will get similar behavior as those observed in the bulk material: the field-ionization of the exciton. The exciton resonance peaks can be easily destroyed by an electric field as small as 1.6×10^4V/cm as observed by D. A. B. Miller et al. [102]. The main effect of parallel field is therefore the broadening of the absorption spectrum.

When the applied field is perpendicular to the quantum well plane, the Hamiltonian can be expressed as:

$$H_2 = H_{0z}^e + e\mathcal{E}z_e + H_{0z}^h - e\mathcal{E}z_h - \frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial r^2} - \frac{e^2}{4\pi\varepsilon_0\varepsilon\sqrt{r^2 + (z_e - z_h)^2}} \tag{5.18}$$

where $\mathcal{E}$ is the electric field strength. The eigen states of H_2 are difficult to solve due to the coupling term from Coulomb potential energy. For simplicity, we can neglect this term first and treat it as a perturbation later. Then, we have a separable-variable problem and in the z-dimension the Hamiltonian is:

$$H_{2z} = H_{0z}^e + e\mathcal{E}z_e + H_{0z}^h - e\mathcal{E}z_h \ . \tag{5.19}$$

For an infinite well, the eigen states of H_{2z} can be solved exactly and the resulting wave functions are Airy functions [102]. The eigen energy levels can be determined by:

$$0 = \mathrm{Ai}(Z_+)\mathrm{Bi}(Z_-) - \mathrm{Ai}(Z_-)\mathrm{Bi}(Z_+) \tag{5.20}$$

where $Z_\pm = (-\pi E_1^\infty / e\mathcal{E}L_z)^{2/3}\left(\frac{E_n}{E_1^\infty} \pm \frac{e\mathcal{E}L_z}{2E_1^\infty}\right)$, Ai(Z) and Bi(Z) are the Airy functions. Using the series expansion of the Airy functions, Eq. 5.20 can be solved numerically.

For finite wells, only approximate solution exists. One way to solve the eigen

equation:

$$(H_{0z} \pm e\mathcal{E}z) \mid \Phi_n >= E_n \mid \Phi_n > \tag{5.21}$$

is by assuming that the solution is a linear expansion of the unperturbed eigen functions as:

$$\mid \Phi_n >= \sum_i a_{ni} \mid \Phi_i^0 > \tag{5.22}$$

where $\mid \Phi_n^0 >$ is obtained from solving the following equation:

$$H_{0z} \mid \Phi_n^0 >= E_n^0 \mid \Phi_n^0 > \ . \tag{5.23}$$

E_n^0 and $\mid \Phi_n^0 >$ are given by Eq. 5.7 and 5.8. Substitute Eq. 5.22 into Eq. 5.21 and multiply both side by $< \Phi_j^0 \mid (j = 1, 2, ...)$ and using the orthonormal condition $< \Phi_i^0 \mid \Phi_j^0 >= \delta_{ij}$, we have the following matrix equation:

$$\begin{pmatrix} E_1^0 - E_n & H_{12} & H_{13} & \cdots \\ H_{12} & E_2^0 - E_n & H_{23} & \cdots \\ H_{13} & H_{23} & E_3^0 - E_n & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} a_{n1} \\ a_{n2} \\ a_{n3} \\ \vdots \end{pmatrix} = 0 \tag{5.24}$$

where $H_{ij} =< \Phi_i^0 \mid \pm e\mathcal{E}z \mid \Phi_j^0 >$. From symmetry, we have $H_{ij} = 0$ for even $i + j$ and $H_{ij} = H_{ji}$. Eq. 5.24 can be solved for a given electric field. Table 5.2 lists the result for the ground state at different fields by assuming the same parameters as we used in previous section.

Figure 5.3 shows the resulting eigen functions for electrons and holes from Eq. 5.24 at different electric fields. We can clearly see the separation between electrons and holes due to the applied electric fields. However, unlike in a bulk material where the Coulomb potential energy between an electron and a hole is easily diminished to zero and the exciton is destroyed when the electron and the hole are pulled away by applied electric fields, in a quantum well, the separation between the heavy hole and the electron is limited by the potential barriers. Therefore, there still exists a large overlap between their wave functions. In return, the binding energy of the

ground level of electrons

$\mathcal{E}$ (kV/cm)	E_1^e (meV)	a_{11}	a_{12}
0	34.09	1	0
10	34.03	0.9997	-0.0243
20	33.86	0.9988	-0.0485
30	33.57	0.9974	-0.0724
40	33.17	0.9954	-0.0960
50	32.66	0.9929	-0.1191
60	32.05	0.9899	-0.1416
70	31.33	0.9866	-0.1635
80	30.51	0.9828	-0.1846
90	29.60	0.9788	-0.2050
100	28.60	0.9745	-0.2246

ground level of heavy holes

$\mathcal{E}$ (kV/cm)	E_1^h (meV)	a_{11}	a_{12}	a_{13}	a_{14}
0	9.208	1	0	0	0
10	9.065	0.9974	0.0718	-0.0022	-0.0012
20	8.641	0.9899	0.1413	-0.0086	-0.0027
30	7.949	0.9782	0.2067	-0.0187	-0.0045
40	7.008	0.9632	0.2667	-0.0319	-0.0068
50	5.840	0.9460	0.3206	-0.0474	-0.0096
60	4.469	0.9273	0.3684	-0.0645	-0.0131
70	2.916	0.9080	0.4103	-0.0827	-0.0171
80	1.203	0.8886	0.4468	-0.1014	-0.0216
90	-0.6520	0.8695	0.4784	-0.1203	-0.0266
100	-2.634	0.8509	0.5057	-0.1389	-0.0319

Table 5.2: Lowest energy level for electrons and heavy holes in a finite quantum well at different electric field strength. Calculated from perturbation approximation. $L_z = 9.0$nm, $x = 0.30$.

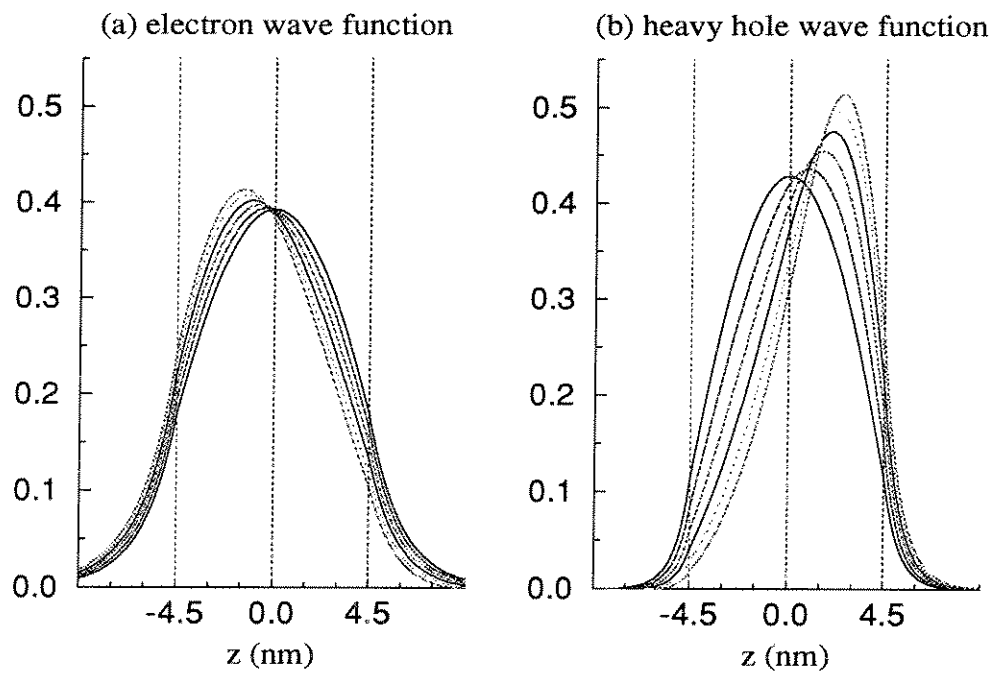


Figure 5.3: Calculated wave functions at different electric field strength for a finite well: (a) electrons (b) heavy holes. The six curves correspond to six different $\mathcal{E}$: 0kV/cm, 20kV/cm, 40kV/cm, 60kV/cm, 80kV/cm, 100kV/cm. The dotted lines show the center and boundaries of the quantum well. At zero field, the wave functions are symmetric. Under applied fields, electrons move to the left and holes to the right. $L_z = 9.0$nm, $x = 0.30$.

electron-heavy-hole exciton, even though reduced, remains strong. In order to find out the binding energy E_b, we need to solve the eigen equation of H_2 as given by Eq. 5.18. In Ref. [102], Miller et al. used a variational method by choosing an 1S-like orbital for the x-y plane motion of the exciton. It was found that even with an applied field of 100kV/cm, the calculated exciton binding energy was still in excess of 6meV, giving a reduction of only 2.6meV compared to the 8.6meV binding energy at zero field. Compared with the shifts in electron and hole energy levels, this is about an order of magnitude less. This is due to the saturation in electron-hole Coulomb interaction as they are pulled to the opposite side of the quantum well. Therefore, when determining the optical properties of a quantum well, we can approximately regard E_b as a constant instead of solving the actual Coulomb interaction term.

5.1.3 Optical Property of GaAs Quantum Wells

The wave function of the quantum well system can be written as:

$$\mid \Phi_{lqs} >= \mid \Phi_l^e > \mid \Phi_q^h > \mid \Psi_s^{2D} > \tag{5.25}$$

where $\mid \Phi^e >$ and $\mid \Phi^h >$ are given by the solutions of Eq. 5.24, $\mid \Psi^{2D} >$ is the x-y plane wave function and can be approximated by the eigen function of a hydrogen system, l, q,and s are quantum numbers. From time-dependent perturbation theory, the absorption coefficient is given by [104]:

$$\alpha(\hbar\omega) = B \sum_i \sum_q \sum_s |< \Phi_l^e \mid \Phi_q^h >|^2 \mid \Psi_s(0) \mid^2 \delta(E_l^e + E_q^h - \mid E_s \mid + E_g - \hbar\omega) \ . \tag{5.26}$$

Near the long wavelength side of the absorption edge, the transition between two ground levels is the dominant factor, so,

$$\alpha(\hbar\omega) \propto |< \Phi_1^e \mid \Phi_1^h >|^2 \delta(E_g + E_1^e + E_1^h - E_b - \hbar\omega) \ . \tag{5.27}$$

Two conclusions can be drawn from this equation. First, the absorption edge

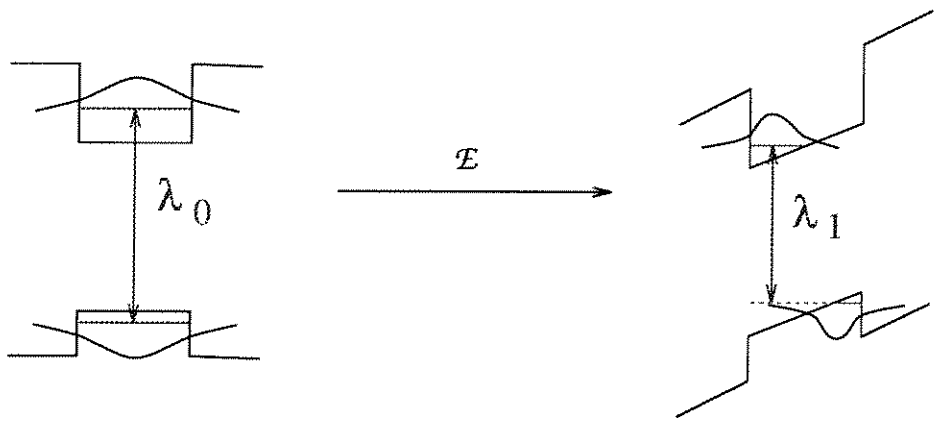


Figure 5.4: Intuitive illustration of quantum confined Stark effect. An applied electric field perpendicular to a quantum well causes a red shift in the wavelength of the absorption peak.

occurs at $E_g + E_1^e + E_1^h - E_b$. Due to the lowering of E_1^e and E_1^h with an applied electric field, the absorption edge will shift to the longer wavelength side when an electric field is applied. In principal, such electro-absorption effect is similar to the Stark effect in a hydrogen atom except the addition of a confining potential, therefore called the quantum confined Stark effect (QCSE), which is the basis of MQW modulators. Second, the height of the exciton absorption peak is mainly determined by the overlap integral between the wave functions of electron and heavy hole ground levels. As the applied field becomes stronger and stronger, electrons and holes will move more and more toward the opposite sides of the quantum well, resulting in a smaller and smaller overlap integral. Therefore, the exciton peak will be lower. A schematic illustration is given by Figure 5.4. Table 5.3 lists the calculated Stark shift and overlap integral at different applied fields.

Knowing the wave function overlap integral, we can calculate the absorption coefficient from Eq. 5.27, which is valid at low temperature. At room temperature, electrons and holes are distributed near the calculated energy levels with a certain distribution. Therefore, the δ-function in Eq. 5.27 becomes a line-shape function with a certain width. From experiment data, it was found that the long wavelength

$\mathcal{E}$ (kV/cm)	ΔE (meV)	λ (nm)	$\mid< \Phi_1^e \mid \Phi_1^h >\mid^2$
0	0	845.6	0.9753
10	0.2008	845.7	0.9670
20	0.7981	846.0	0.9428
30	1.778	846.6	0.9053
40	3.119	847.4	0.8578
50	4.797	848.3	0.8040
60	6.784	849.5	0.7469
70	9.056	850.8	0.6886
80	11.59	852.3	0.6322
90	14.35	853.9	0.5777
100	17.33	855.7	0.5262

Table 5.3: Calculated result of quantum confined Stark effect. Column 2 shows the shift in absorption peak. Column 3 is the actual absorption edge without the correction of exciton binding energy. Column 4 lists the overlap integral of electron and hole wave functions, which determines the maximum absorption coefficient.

side of this absorption edge fits well with a Lorentzian line-shape. Based on this observation, an empirical model was proposed for the estimation of the absorption coefficient [105]:

$$\alpha(\mathcal{E}, \hbar\omega, L_z) = \alpha_{hh} \left\{ 1 + \left(\frac{E_g + E_1^e + E_1^h - E_b - \hbar\omega}{\Gamma_{hh}} \right)^2 \right\}^{-1} . \qquad (5.28)$$

The remaining question is the estimation of the maximum absorption coefficient, α_{hh} and the half-width, Γ_{hh}. Compared with Eq. 5.27, we know that the peak absorption coefficient is linearly related to the wave function overlap integral. Therefore, we have:

$$\alpha_{hh}(\mathcal{E}, L_z) = C \mid< \Phi_1^e \mid \Phi_1^h >\mid^2 + \alpha_0 \qquad \text{for } L_z > 5\text{nm} \qquad (5.29)$$

where both C and α_0 are fitted constants. Ref. [105] found that $C \propto 1/L_z$ and they gave $C = 16,000\text{cm}^{-1}$ for $L_z = 10.5\text{nm}$ and $\alpha_0 = 5500\text{cm}^{-1}$. Γ_{hh} in Eq. 5.28 is the half-width at half maximum, which includes homogeneous broadening of the exciton energy due to phonon process and inhomogeneous broadening of the absorption edge due to variations in L_z — from growth non-uniformity — and $\mathcal{E}$ —

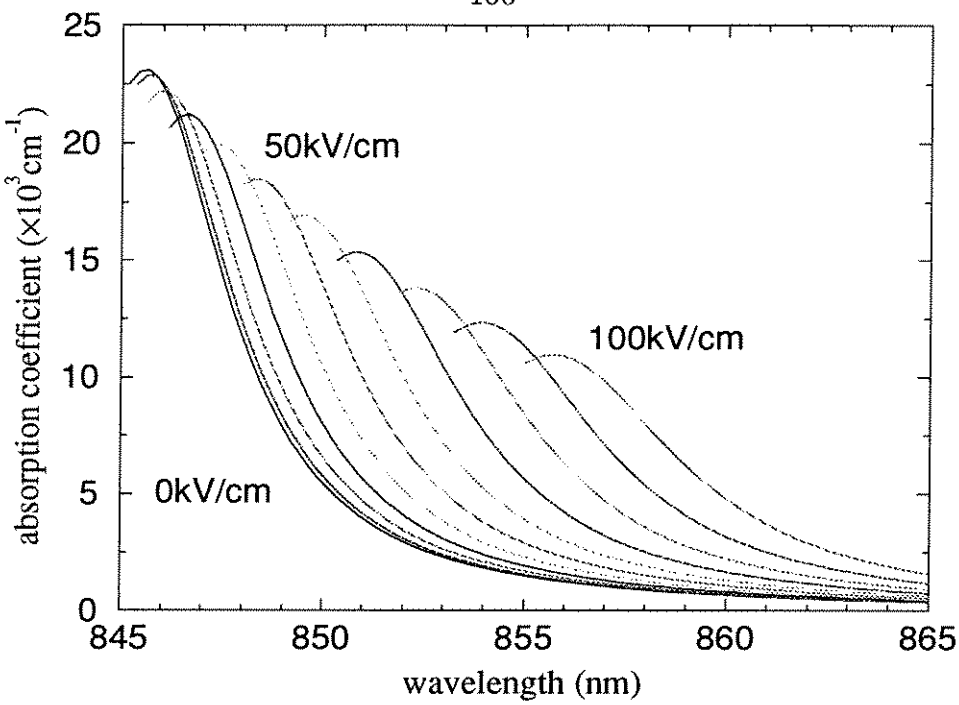


Figure 5.5: Absorption coefficient at different applied electric fields based on theoretical calculation result of the quantum confined Stark effect.

from residue doping. As we mentioned earlier, electrons and holes are separated with a perpendicular electric field. In return, the Coulomb attraction is reduced and we get a smaller exciton binding energy. Therefore, for the phonon broadening at normal temperature, the line width Γ_{hh} will increase as the applied field becomes stronger. Here, we used the value given by Ref. [105]:

$$\Gamma_{hh} = 7.347 - 0.511 L_z + 0.0182 L_z^2 + 0.054 \mathcal{E} + 0.0161 \mathcal{E}^2 \ . \tag{5.30}$$

Following this model and the result listed in Table 5.3, we can calculate the value of α for the longer wavelength side of an absorption peak. It is plotted in Figure 5.5. From this plot, we can theoretically calculated the change of α at a given electric field and a given wavelength. Based on that, the contrast ratio of a MQW modulator can be estimated as we will discuss later. Notice that in Figure 5.5 we did not use the parameters of C and α_0 given by Ref. [105] because the resulting $\Delta\alpha$ from these parameters is larger than what we observed. Therefore, results from D. A. B. Miller et al. [106] was fitted into Eq. 5.29 and C and α_0 were determined to be 27090cm^{-1}

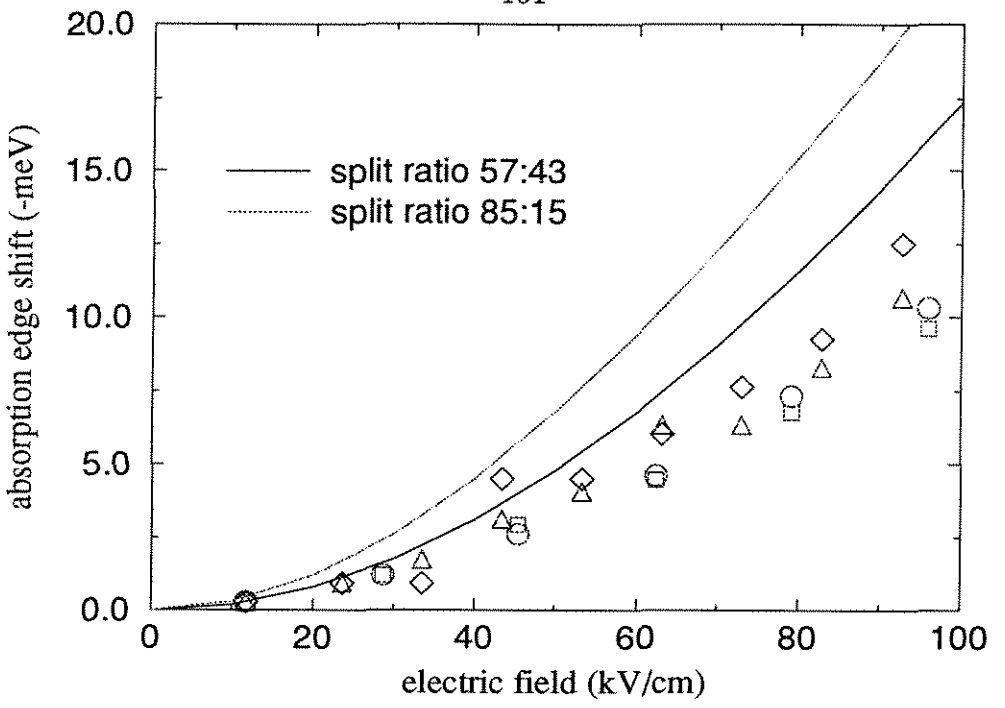


Figure 5.6: The Stark shift due to applied electric field for GaAs/$Al_{0.3}$ $Ga_{0.7}$As quantum well. Both theoretical result and experimental measurement are shown. Circles and squares are for modulators from the FET-SEED process. Triangles and diamonds are for the Caltech-MIT SEED grown at 500°C.

and -3300cm^{-1}, respectively.

To compare with the experimental results, we look at the measurement from the AT&T FET-SEED modulators, shown in Figure 2.10. The contrast ratio we get is 4.6 for $\lambda = 850$nm at bias voltage of $9.5 \sim 10$V, which corresponds to an electric field of about $(10+1.4)\text{V}/1.18\mu\text{m} \approx 10^5\text{V/cm}$, where the 1.4V comes from the built in voltage of the p-i-n diode and the total thickness of the MQW stack is 1.18μm. Because the exact location of the exciton peak depends on the calibration of the quantum well width, we can only compare the amount of exciton shift. This is plotted in Figure 5.6. The theoretical calculation results are presented by two lines, representing the result using 57:43 split ratio or 85:15 split ratio for ΔE_c and ΔE_v. We can see that the result from the 57:43 split ratio agrees well with the experimental measurement, especially at low electric fields. For high fields, remember that we did not consider the change in exciton binding energy, which is about 2.6meV at 10^5V/cm according to Ref. [102]. If we include this effect in our calculation, the result would agree very well with the

experimental measurement. If we look at Figure 2.5, for an electric field of 100kV/cm, we have a calculated $\Delta\alpha$ of 9670cm^{-1} compared to zero field. Therefore, the expected contrast ratio of the 95 pairs of 90Å-GaAs/35Å-AlGaAs quantum well device from the FET-SEED process would be $\exp(2\Delta\alpha L) = \exp(2\times 9670\times 95\times 90\times 10^{-8}) = 5.22$, which is quite close to the measured result of 4.6.

In the foregoing discussion, our main focus has been on the single quantum well. In the case of MQW structures, there are two possibilities. When the barrier width is small enough, electrons (or holes) will be able to tunnel through the barrier and couple with other electrons (or holes). Under this situation, we have to consider the whole MQW structure as a system and the result is remarkably different from those associated with single quantum well [107]. However, if the AlGaAs barrier is wide enough, the coupling effect between electrons (and holes) in nearby quantum wells becomes negligible and we can simply use the result from Eq. 5.24 to get an accurate estimation on the magnitude of the quantum confined Stark effect.

In summary, compared to bulk material, the Stark effect in a quantum well structure is much more significant. Due to field ionization, the Stark shift in a bulk material is typically less that 10% of the exciton binding energy (4.2meV). In an MQW, the potential barriers prevent the ionization of electron-hole excitons and we can observe a Stark shift well beyond the exciton binding energy (8.6meV for ~10nm wells) without destroying the exciton resonance peak. Furthermore, in a 2-dimensional system, the exciton binding energy is much stronger, therefore we have a higher and narrower absorption peak, which is very important for the contrast ratio of the MQW modulators we will discuss next. Even at strong fields, there still remains a strong Coulomb attraction and a strong exciton resonance peak.

5.2 MQW Modulators

We have shown that an electric field perpendicular to the quantum well results in the so-called quantum confined Stark effect, i.e., a red shift of the absorption edge. As a result of this shift, when the exciton peak moves through a certain wavelength

near the absorption edge, the absorption coefficient incurs dramatic changes as given by Eq. 5.28 and plotted in Figure 5.5, therefore we can obtain intensity modulation. This is the basic mechanism of MQW modulators.

5.2.1 Basic Configuration

The operation of such modulators requires a mechanism to apply a high electric field, up to 10^5kV/cm, to the quantum well material. One way to achieve that is by using a reverse-biased p-i-n diode as shown in Figure 5.7 where the quantum well region is nominally undoped. The advantage of such approach is the low electric current of reverse-biased diodes. When no voltage is applied to the device, there is a weak electric field inside the diode caused by the build-in potential due to the different doping levels in the p and n region. Semiconductor physics [37] gives:

$$V_{bi} = \frac{kT}{q} \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{5.31}$$

where N_A and N_D are the doping levels in the p and n region, respectively, by assuming fully ionization of dopant, and n_i is the intrinsic doping level of the semiconductor material. For GaAs, $n_i = 1.79 \times 10^6 \text{cm}^{-3}$, we have $V_{bi} \sim 1.40$V at room temperature for a doping level of $N_A = N_D = 10^{18}\text{cm}^{-3}$.

If the i-quantum well region is truly undoped, the MQW will be totally depleted by this build-in voltage for any reasonable length of d. When a reverse bias (positive V) is applied, the electric field in the MQW will be uniform and is given by:

$$\mathcal{E} = \frac{V_{bi} + V}{d} . \tag{5.32}$$

However, for actual materials prepared by either MBE or MOCVD, there are always residue doping in the i-region, typically around $10^{14} \sim 10^{15}$ cm^{-3}. Therefore, the maximum length of a totally depleted region is:

$$d_{max} = \sqrt{\frac{2\varepsilon\varepsilon_0(V_{bi} + V)}{qN_R}} \tag{5.33}$$

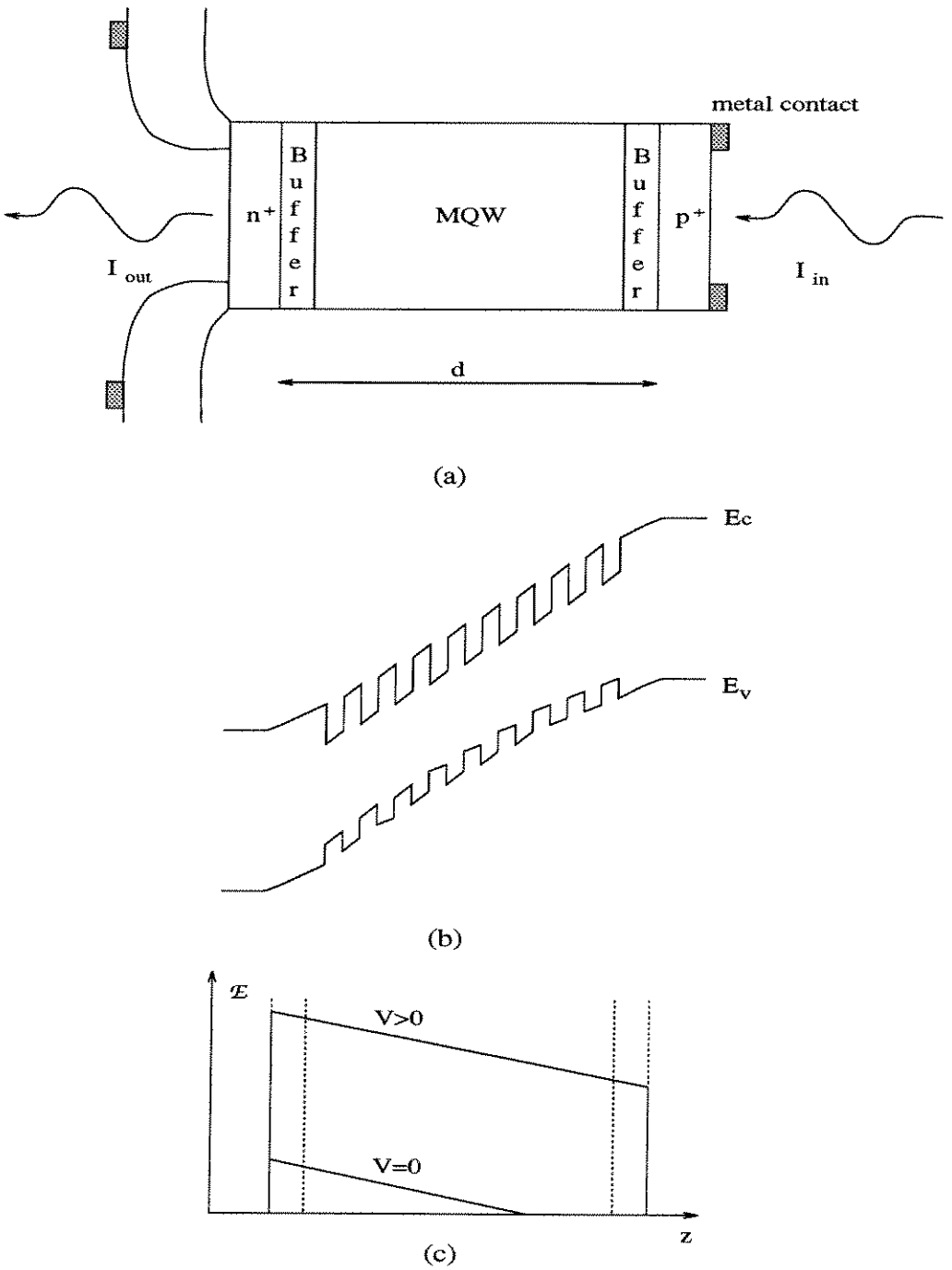


Figure 5.7: Schematic view of a transmission MQW modulator. (a) The actual p-i-n structure with an undoped quantum well region. The width of each quantum well is L_z, and the barriers have widths of L_b. The total thickness of the MQW stack is d, among which L is the total length of quantum wells. (b) Energy band diagram. (c) The distribution of electric field due to residue doping.

for a residue doping of N_R. The other effect of the residue doping is the non-uniform distribution of electric field inside the i-region, resulting in inhomogeneous broadening of the exciton resonance peak discussed in last section and a reduced absorption coefficient change.

5.2.2 Contrast Ratio

When we apply a reverse bias to the p-i-n diode, the eigen states of electrons and holes inside the quantum well are affected and result in the shift of absorption peak and the change in absorption coefficient, $\Delta\alpha = \mid \alpha(0) - \alpha(V) \mid$, as given by Eq. 5.28. Therefore, for a given wavelength, the light output changes from $I_{in}e^{-\alpha(0)L}$ to $I_{in}e^{-\alpha(V)L}$, where L is the total interaction length of the MQW region, and intensity modulation is achieved. An important characteristic of modulators, the contrast ratio, is therefore given by:

$$R = \frac{I_{ON}}{I_{OFF}} = \exp(\Delta\alpha L) \ . \tag{5.34}$$

In order to increase the contrast ratio, we can increase either L or $\Delta\alpha$. Unfortunately, these two requirements are not compatible, which results in the trade-off on L_z and L.

- The trade-off on the width of single quantum well, L_z: Since a smaller L_z will increase the maximum absorption coefficient, α_{hh}, it will thereby increase the achievable $\Delta\alpha$. However, smaller L_z also means electrons and holes are more confined in the well, which reduces the amount of Stark shift for a given $\mathcal{E}$. We have to increase the applied voltage to get the same $\Delta\alpha$. Such effect is undesirable because a higher voltage will increase the switching energy $\frac{1}{2}CV^2$ and makes the integration of modulators with other IC elements more difficult. Furthermore, since electrons and holes are closer, their overlap integral will be higher. In return, $\Delta\alpha$ goes down. So, an optimum L_z exists.

- The trade-off on L: Increasing the total interaction length between the light and the MQW material gives a better contrast ratio. However, the total thickness

of the MQW stack d is also larger if we increase the number of quantum wells. The result is a weaker electric field for a given voltage, which will reduce $\Delta\alpha$. In addition, when d is large, we will run into a problem in depleting all the MQW region and the inhomogeneous broadening effect due to field variation is more significant. All these will reduce the contrast ratio.

The dilemma on L and V can be solved to a certain degree. First, by putting a DBR stack, discussed in Section 3.2.1 on one side of the modulator, the modulated light will pass through the MQW twice and we can double the interaction length while keeping the same voltage requirement. This is the reflection modulator and the contrast ratio is increased to:

$$R = \exp(2\Delta\alpha L) \ . \tag{5.35}$$

From the reflection configuration, it is natural to think that if the modulated light makes multiple traveling through the MQW, the contrast ratio will be further increased or we can reduce the operation voltage and get a smaller switching energy. This can be done by depositing a partially reflecting mirror on the other side of the modulator and is called asymmetric Fabry-Perot cavity modulators. However, this requires very critical control on the total cavity length and the operating wavelength because we need a constructive interference between the multiple reflected light beams. Typically we can do an after-growth shallow etching to adjust the cavity length before depositing the top mirror. The successfully fabrication of Fabry-Perot modulators requires very uniform material growth and may be time consuming due to the requirement of fine-tuning the wavelength. Therefore, we will concentrate on reflection modulators in our research.

A special configuration of MQW modulators is the self-electrooptic effect device (SEED), in which the reverse-biased p-i-n diode is used as both a MQW modulator and a photodiode. For a normal photodiode, electrons (and holes) generated in the depleted i-region due to photon absorption is swept by the high electric field to the n (and p) region, resulting in a photocurrent proportional to incident optical

power. When the reverse bias voltage is higher, the photocurrent will increase slightly because of the widening of depletion region and the higher reverse current of a diode. However, in the case of a SEED, it is possible that the absorption coefficient for a specific wavelength will be lower when the reverse bias voltage is increased due to the Stark shift. As the result, the photocurrent decreases when the bias voltage is higher, resulting in a negative differential resistance. Therefore, bistable states can be obtained and can be used as an all-optical switch. This is called SEED [54]. However, there are disadvantages for this configuration: the all-optical switch has usually limited functionality and it requires high optical switching energy because of the lack of amplification devices. By integrating MQW modulators with active electronic devices, such as FET's, we can reduce the optical power requirement by using the amplifying power of FET's and enhance the functionality of the system. As a result, the name SEED has been extended to the MQW p-i-n diode structure rather than limited for its special original configuration discussed above.

For a MQW modulator, there are two operation modes, called λ_0 and λ_1 operation. The λ_0 mode is operated at a wavelength corresponding to the zero-bias exciton peak. As we increase $\mathcal{E}$, the absorption peak shifts to longer wavelength and $\alpha(\lambda_0)$ decreases. Therefore, we have a high absorption (OFF state) at zero bias and low absorption (ON state) at high bias. λ_1 operation is the reverse case where the modulator is operated at a wavelength longer than zero-bias exciton peak, so it is on the ON state at zero bias. As $\mathcal{E}$ increases, the exciton peak moves toward λ_1 and we have higher absorption. The OFF state is obtained when we have such a bias voltage that the exciton peak is exactly at λ_1. From experiment measurement, both modes have roughly the same contrast ratio. However, the λ_1 mode is more desirable because the insertion loss is smaller due to the longer wavelength.

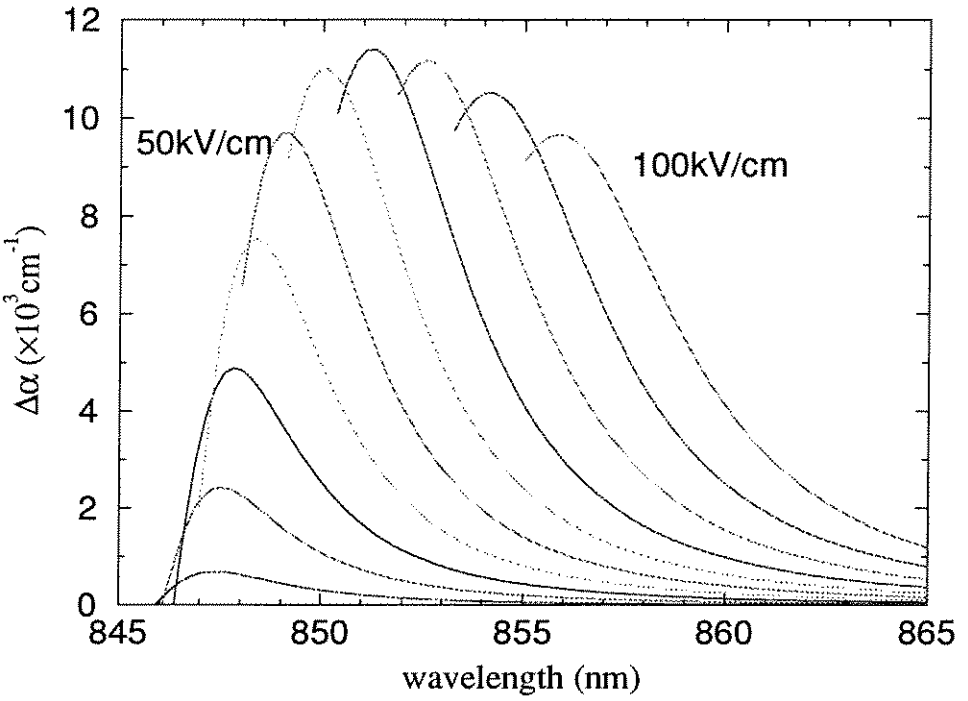


Figure 5.8: Calculated absorption coefficient change for different electric fields.

5.3 Experimental Results

5.3.1 Structure Design

The result for the MQW modulators from the AT&T FET-SEED process was given in Section 2.3.3. For 95 pairs of 90-Å-GaAs/35-Å-$Al_{0.3}Ga_{0.7}As$ quantum well, the best contrast ratio obtained was near 5:1 for a bias voltage of $10 \sim 11$V. When integrating the modulators with MOSIS GaAs circuits, we need to reduce the voltage requirement because the breakdown voltage of the GaAs MESFET is only 4V. By carefully designing the circuit, we are able to obtain a voltage swing of 7V by distributing this voltage on two MESFET's connected in series. The question is how to achieve the best contrast ratio by changing the number of quantum wells.

There are two choices here: we can either use a large number of quantum wells and reduce the working electric field or we can reduce the number of quantum wells to maintain the electric field. In order to figure out which one will give the best contrast ratio, we need to know the change of absorption coefficient for a given electric field. This can be done theoretically as discussed in Section 5.1.3 and presented in Eq. 5.28

and Figure 5.5. From that, we plot in Figure 5.8 the result of $\Delta\alpha$ as a function of wavelength at different electric fields. It is found that when $\mathcal{E} = 70\text{kV/cm}$, maximum $\Delta\alpha$ can be achieved. Therefore, when designing the modulator structure, we should try to choose such a MQW stack thickness so that we can apply electric fields up to 70kV/cm for a voltage of 7V or less.

Figure 5.9 shows the final design of the MQW modulator integrated with the MOSIS GaAs circuits. There are 75 quantum wells in this structure, with a thickness of 9375Å. The total thickness of the undoped region is 0.991μm, including the two AlGaAs spacer. The top layer of 1000ÅGaAs is the contact layer, which will be etched away after depositing ohmic contacts in order to reduce the insertion loss of the modulator. The conducting layer is therefore provided by the 120Åp^+ GaAs. Under the MQW structure, a 24-pair DBR mirror is grown on top of the thick buffer layer. The total thickness of the regrowth stack is about 5.8μm. For different MOSIS runs, sometimes the foundry will change their process and the dielectric thickness will change accordingly. We can always control the total regrowth thickness by changing the buffer layer so that it matches the dielectric stack on the MOSIS GaAs chips.

5.3.2 Fabrication and Measurement

The MQW structure was grown by our collaborator Hao Wang (Fonstad's group at MIT) [108] using molecular beam epitaxy (MBE) at 500°C. The total growth time was almost 10 hours because only one Al cell was working when growing this structure.

The fabrication of the modulators was simple. Because n^+ substrates were used, only the top p contact was needed. First, the wafer was cleaned using standard procedure. A shallow etching was performed to achieve isolation between different modulator devices. This was done using a mixture of NH_4OH, H_2O_2, and H_2O in a ratio of 2.9:1:139. The wafer was etched slightly into the intrinsic MQW region. Next, silicon nitride was deposited on the surface using thermal CVD at 610°C. Finally, following the nitride opening using plasma etching, p-type metal (AuZn/Au) was deposited using e-beam evaporator and annealed at 410°C. The testing was done

GaAs	p+=3e18	1000A
Al(0.11)GaAs	undoped	700A
GaAs	p+=3e18	120A
Al(0.3)GaAs	undoped	35A
	⋮	
undoped MQW 75 X	GaAs	90A
	Al(0.3)GaAs	35A
	⋮	
Al(0.3)GaAs	undoped	500A
	⋮	
n+ DBR 24 X	Al(0.11)GaAs	599A
	AlAs	711A
	⋮	
GaAs Buffer	n+=1e18	15000A
n+ GaAs substrate		

Figure 5.9: MQW structure for the MBE regrowth on MOSIS GaAs circuits.

on a control wafer, so we did not have to worry about the performance of the MOSIS chip, which will be discussed in the next chapter.

The measurement set-up was similar to what we used in testing the FET-SEED chips. The laser output from a Ti:sapphire tunable laser was focused onto the modulator surface. We measured both the intensity of the reflection light and the photocurrent collected by the modulator/photodiode at different reverse bias voltages. Several different samples were tested. The result for sample 8090, which was a 70-quantum-well device with 22-pair DBR, is presented in Figure 5.10 and Figure 5.11. From Figure 5.10, we can clearly see the Stark shift of exciton peak when we increase the applied voltage. The resulting shift was represented by triangle symbols in Figure 5.6 when comparing with theoretical calculation results. The behavior is just as what we expected except for the 0V bias curve, which is because the MQW region is not totally depleted. However, Figure 5.11 is quite different from our expectation. Because no anti-reflection coating was deposited on this test device, we would expect a higher contrast ratio for λ_1 operation and for longer wavelengths the reflectivity should increase monotonously. However, the observed reflectivity function was suppressed by a common function for both short and long wavelengths as we can see from Figure 5.11. Later we found out that it was due to the misalignment of the DBR mirror, which has a center reflection wavelength at 800nm.

Due to the absence of AR coating and the mistuning of DBR mirror, poor contrast ratio was obtained for both sample 8039 and 8090 as listed in Table 5.4. Therefore, several MBE runs were spent on improving the reflectivity of DBR mirrors at MIT. Especially, we wanted to get maximum reflectivity from 830nm to 860nm. Finally, we were able to control the DBR center reflection region and the result was shown in Figure 5.12. Reflectivity as high as 99% was achieved for a 20 pair AlGaAs/AlAs DBR. The high reflection region was centered at around 855nm and extended for about 80nm.

However, even with the correctly-tuned DBR mirror, the contrast ratio measured from sample 9123 is still very poor. This is due to the effect of Fabry-Perot cavity for devices without anti-reflection coating. What we measured is therefore the com-

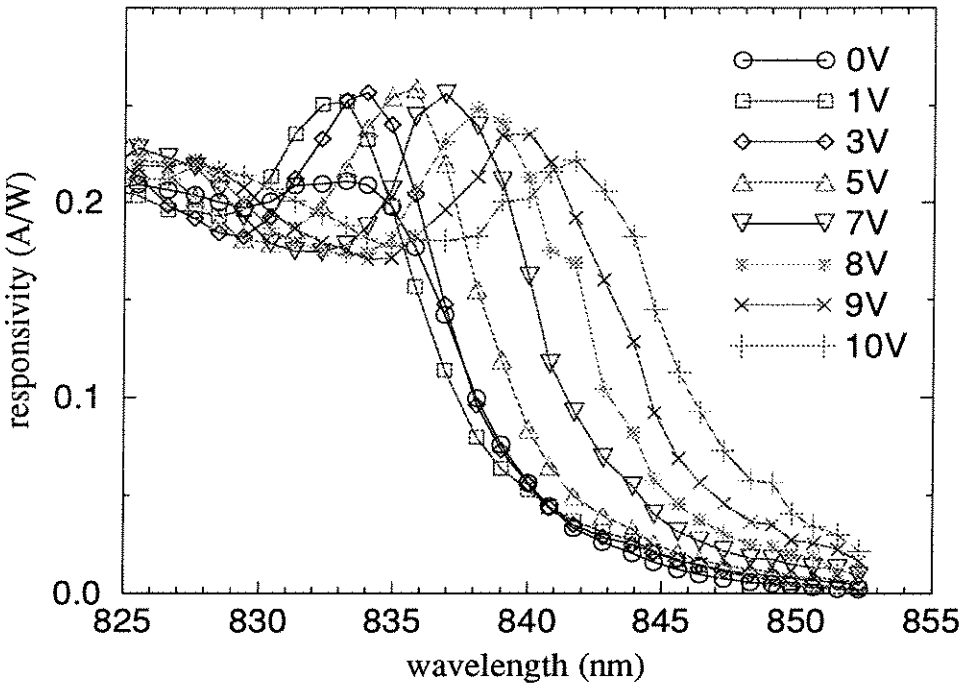


Figure 5.10: The measured responsivity of a SEED from sample 8090, showing the Stark shift when electric fields were applied.

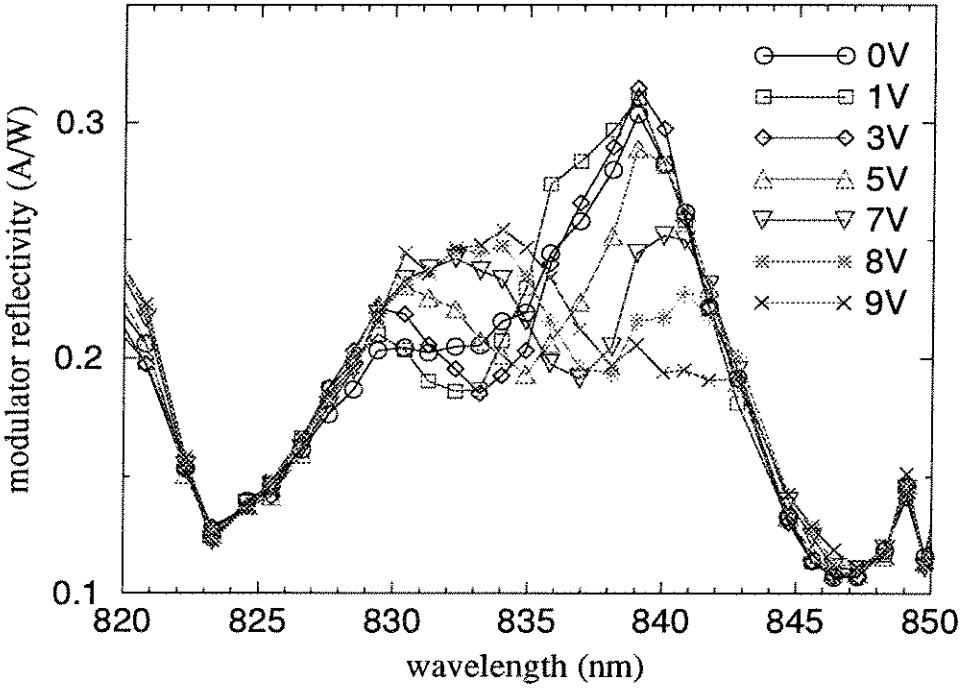


Figure 5.11: Measured reflectivity of a MQW modulator from sample 8090. The DBR mirror was not tuned to the correct wavelength. Combining with the effect of a weak Fabry-Perot cavity, it results in the low contrast ratio.

sample number	8039	8090	2D-MQW	9123
number of MQW	70	70	55	75
number of DBR pairs	16	22	0	24
λ for exciton peak at 1V	838.1nm	832.1nm	839.5nm	839.9nm
best contrast ratio at λ_0	1.20	1.37	N/A	2.47
best contrast ratio at λ_1	1.75	1.55	N/A	3.27
AR coating	No	No	N/A	Yes

Table 5.4: Measured result for the MQW samples grown by low temperature MBE. The contrast ratio of sample 8039 and 8090 were poor because no AR coating was applied and (likely) the DBR mirror was mistuned.

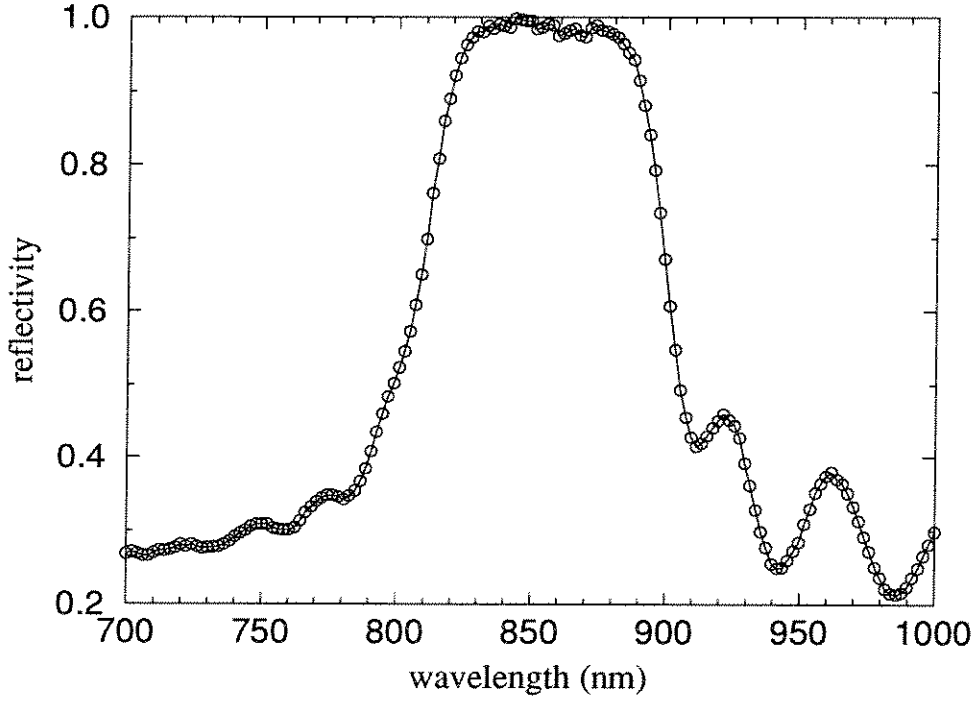


Figure 5.12: The reflectivity of a 20 pair AlGaAs/AlAs distributed Bragg reflector.

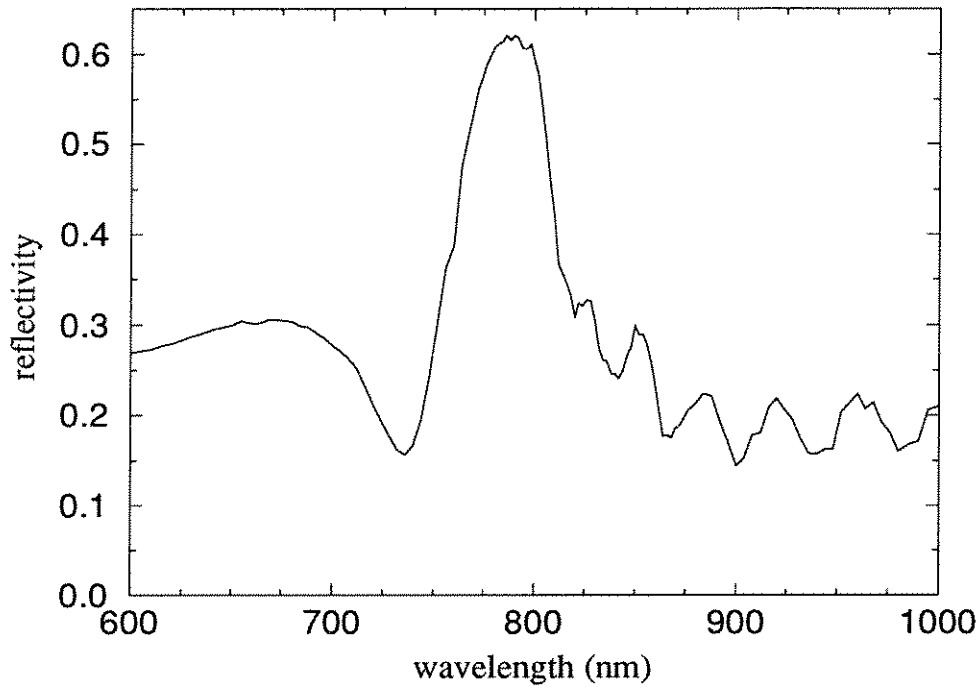


Figure 5.13: The reflectivity of a GaAs/AlGaAs modulator on top of a DBR mirror, showing the effect of Fabry-Perot cavity. No electric bias was applied.

bination of the modulator reflectivity and the FP cavity reflection. The resulting reflectivity is presented in Figure 5.13. Typically, the wavelengths we are interested in are located in one of the side lobes of such reflector, resulting in high insertion loss and low contrast ratios.

With the deposition of a simple AR coating consisting of one quarter-wavelength dielectric layer, sample 9123 shows significant improvement over other samples. The measured reflectivities at different reverse bias voltages are plotted in Figure 5.14. Contrast ratio of 3.27:1 was obtained for 10V bias at 849.1nm. For 8V bias, the measured contrast ratios were 2.43 for 845.3nm and 2.33 for 839.9nm. The normalized light modulation from this modulator at these wavelengths are presented in Figure 5.15.

5.4 Summary

In this chapter, we discussed the mechanism of GaAs/AlGaAs quantum well modulators. The quantum-confined Stark effect was explained from theoretical calculation

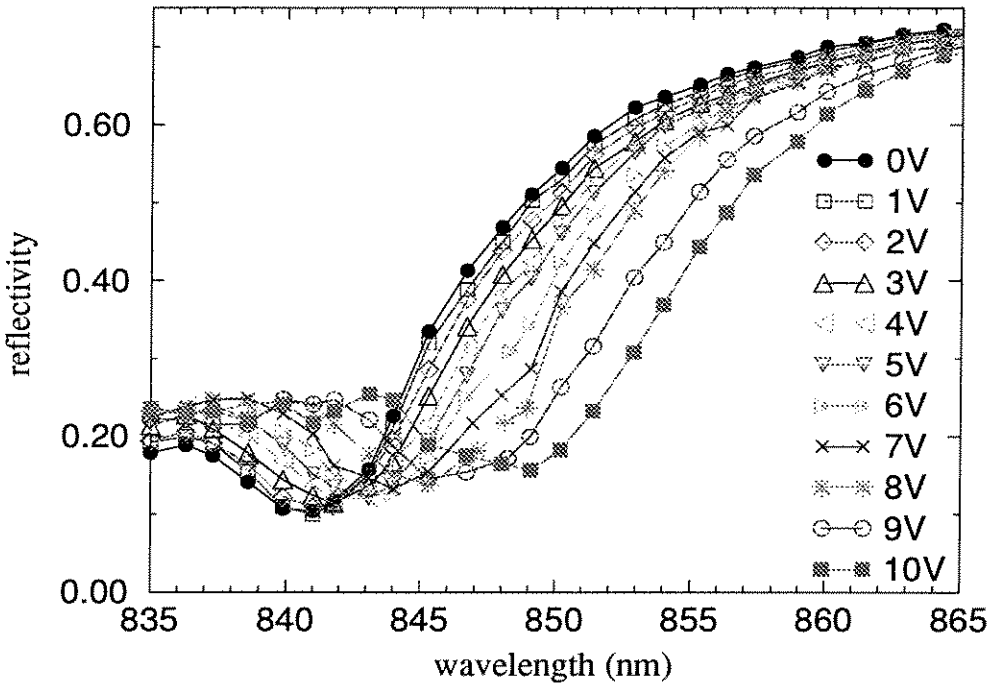


Figure 5.14: Measured reflectivity of a MQW modulator from sample 9123 with proper AR coating and correctly-tuned DBR mirror.

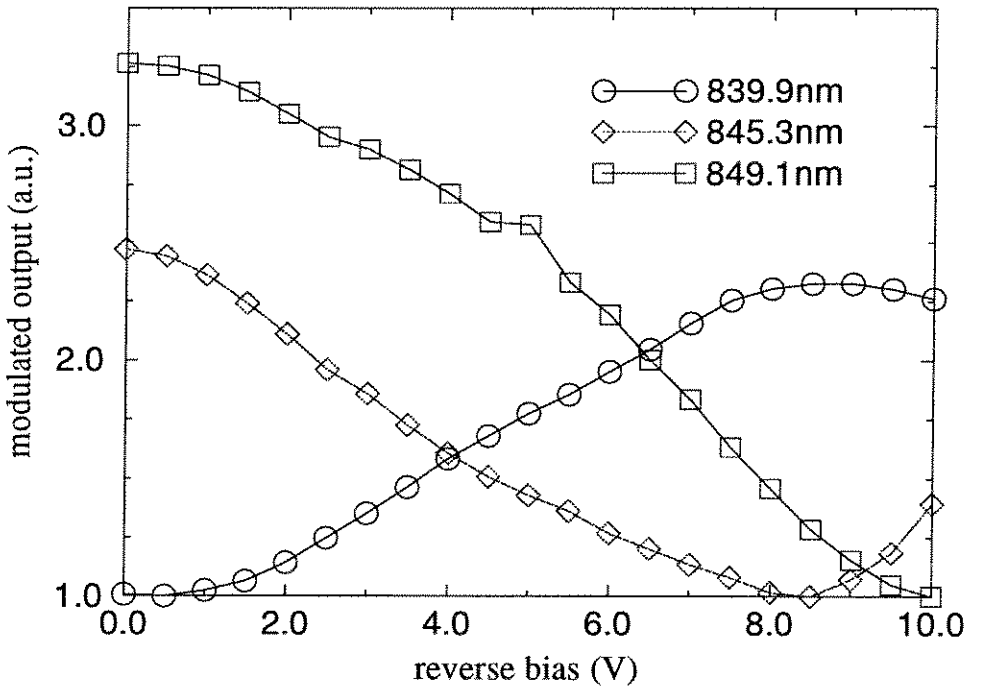


Figure 5.15: The performance of a MQW modulator from sample 9123 at different wavelengths.

and an empirical model was presented for simulation purpose. The result from experiment measurement was compared with the simulation result. They agreed well to first order. Based on the QCSE, MQW modulators can be fabricated with different configurations, including transmission mode, reflection mode, and Fabry-Perot micro cavity modulators.

Based on the theoretical model, we designed the MQW structure for the modulators integrated with MOSIS GaAs chips, which could provide a bias voltage of up to 8V. Two major problems for the low contrast ratio were the mismatch of DBR mirror center wavelength and the requirement of AR coating. With improvement in these two areas, better contrast ratios were obtained: 3.27:1 for 10V bias and 2.4:1 for 8V bias.

Chapter 6 Performance of Neuron Circuits

From Chapter 3 to Chapter 5, we discussed the performance and optimization of individual components for optoelectronic neuron circuits, including light-emitting diodes, high gain photodetectors, and GaAs/AlGaAs multiple quantum well modulators. In this chapter, we will present the result of integrating these devices together into an OEIC circuit for the application of fabricating high density neuron arrays. Since three different fabrication approaches have been investigated, this chapter will be divided into three sections, corresponding to the results from in-house process, AT&T FET-SEED process, and MBE regrowth on MOSIS GaAs circuits, respectively. The detailed description on the fabrication procedure for these processes were given earlier in Chapter 2.

6.1 Result from In-house Process

Because of the poor uniformity of the MESFET's and OPFET's built using our own facilities, which was introduced by the chemical wet etching, it was difficult to test complex circuits on this in-house process. Therefore, we only fabricated simple threshold circuits as the one shown in Figure 1.2. The LED was chosen as the output device because it could be designed and fabricated easily. The driving current of the output LED was provided by a GaAs enhancement-mode MESFET, which had a gate voltage controlled by the output node voltage of the thresholding branch. The desired response of this neuron circuit is:

$$P_{out} = G \cdot \sigma(\sum_i w_{ij} P_{in}^i) \ , \tag{6.1}$$

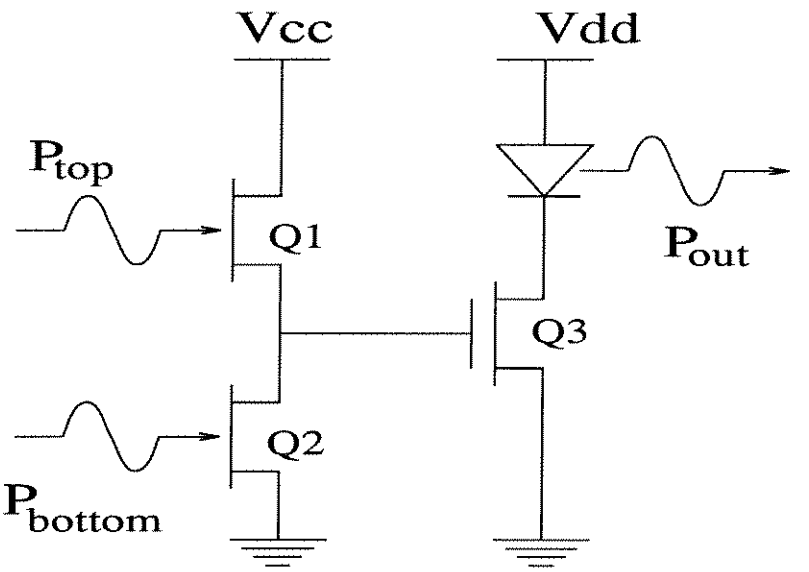
where P_{in}^i is the input for this neuron, which comes from the output of previous neuron layer in an actual optical neural network, G is the optoelectronic gain provided by the

circuit, and w_{ij} is the interconnection weight of the network. $\sigma(x)$ is a thresholding function, which has an output of either 1 when x is larger than the threshold level, or 0 when x is less than the threshold level. This is called a "hard" thresholding function. However, in practical system, the result is usually a "soft" thresholding function where $\sigma(x)$ has a smooth transition from 0 to 1. Its advantage is that such function has finite derivative at any point, which is required for certain neural network learning algorithm.

An OPFET was chosen as the high gain photodetector to detecting the incoming P_{in}, which would be compared with the threshold value. In optical neural network applications, some algorithms require both positive and negative interconnection weights. Because we can only detect the intensity of the input light, two photodetectors are needed in this case, one for the positive input and the other one for the negative input. Figure 6.1 shows an example of such circuits, which was built using our own facilities.

In this circuit, the output node voltage of the thresholding branch is controlled by the comparison result of the two optical input signals. When P_{top} is higher than P_{bottom}, the node voltage will be close to V_{cc} and high output from the LED can be obtained. For the reverse case, the node voltage will be close to ground and the circuit will be turned off. Here P_{top} represents the positive input and P_{bottom} represents the negative input. This is the case of a normal threshold circuit where the bottom OPFET is the thresholding unit. On the other hand, we can operate it as an optical inverter by using the top OPFET as the thresholding unit. In this case, the output is high when P_{bottom} is low and the circuit will be turned off when we increase P_{bottom} beyond P_{top}.

The response of such an optical inverter has been plotted in Figure 6.1 as well. The two OPFET's have a gate opening area of $7 \times 40\mu\text{m}^2$. The driving FET has a gate length of 5μm and a gate width of 100μm. The LED was fabricated using the double Zn diffusion technique described in Chapter 3 and the center diffusion area is $25 \times 25\mu\text{m}^2$. The two bias voltages are: $V_{cc} = 0.7\text{V}$, $V_{dd} = 2.5\text{V}$. When the circuit was on, the LED driving current was slightly more than 1.0mA. When it was off,



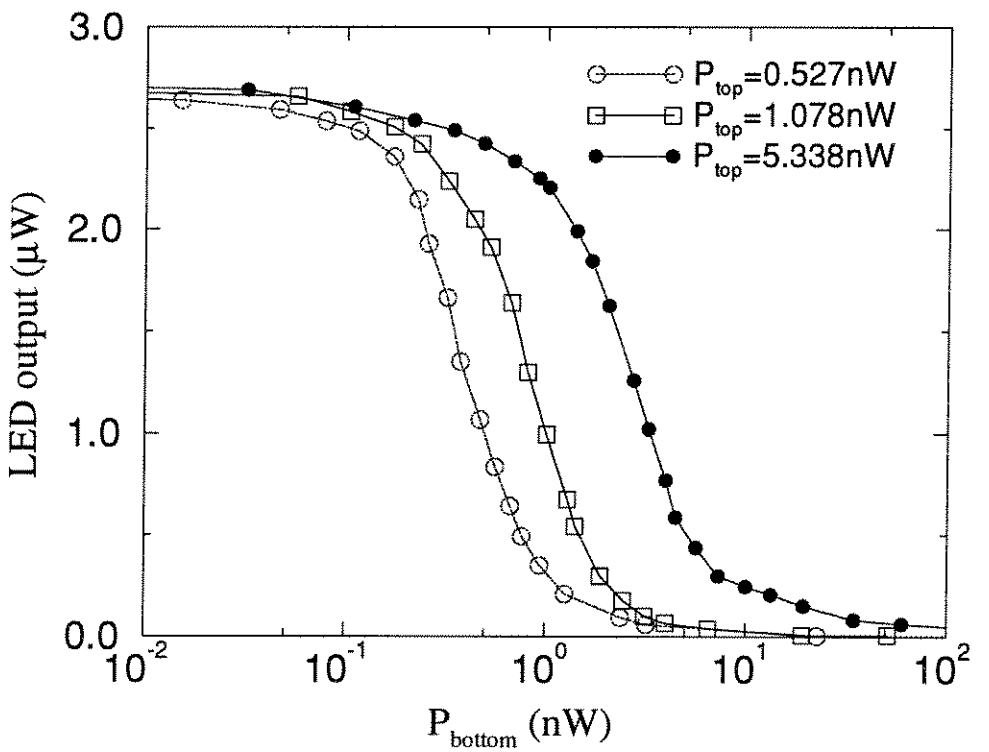


Figure 6.1: A threshold circuit fabricated using the in-house process. The threshold level was set optically by an OPFET. The response shows the result of operating this circuit as an optical inverter.

the current was very small. Therefore, this circuit had a power consumption of more than 2.5mW when it was on. If the power consumption limit of the GaAs wafer is 5W/cm^2, theoretically we can achieve neuron densities as high as $4 \times 10^3 \text{neurons/cm}^2$ assuming that half of the neurons are on and half are off.

From the result shown in Figure 6.1, we found that the threshold level of the optical inverter was controlled by the optical input power on the top OPFET. The threshold level could be as low as 0.527nW. In this case, the thresholding was completed when $P_{\text{bottom}} = 3\text{nW}$. The maximum output was slightly over $2.6\mu\text{W}$, giving us an optoelectronic gain of almost 900. If the threshold level was increased to 5.338nW, the optoelectronic gain dropped to less than 90 if we did not increase the driving current of the output LED.

Even though we achieved good performance from this circuit. The poor uniformity of the fabrication process presented a major problem. For the same testing circuits, we observed threshold levels at $P_{\text{top}} = 0.5\text{nW}$ changing from 0.1nW to as high as 10nW for the same circuit but at different locations on the chip. Therefore, in order to achieve the same response for every neuron in an array, more than 100nW input power is needed for each neuron to turn it on. For the same LED driving current, we can get an optoelectronic gain of only 25. If we want to increase the optoelectronic gain to 100 in order to compensate the high loss in the optical interconnection media, the driving current must be increased to 4mA and the power consumption for each neuron will be more than 10mW when it is turned on. Higher gain requirement means that the power consumption would be even higher. The result is low neuron density.

As the conclusion, for the in-house fabrication process, even though we can achieve good performance from some single circuits, the achievable neuron density for an array is poor due to the non-uniformity introduced in the fabrication process. To solve this problem, a large amount of capital and time is needed, which is more than an academic institute can support. We have to look for industrial foundries to solve this problem.

6.2 Result from AT&T FET-SEED Process

With the successfully demonstration of GaAs MISFET's as high gain photodetectors for the AT&T FET-SEED process, it is possible to reduce the required optical input levels of each neuron. As a result, the achievable neuron density will increase as we discussed in Chapter 1. Several testing circuits were fabricated using the FET-SEED process and the result is presented here.

When designing the FET-SEED circuits during the FET-SEED workshop, we could not simulate the designed circuits due to the lack of a simulation tool. As a result, some of the electric circuits did not perform as we wanted due to the incorrect W/L ratio for different FET's. To correct this expected problem, we broke some of our designs into several parts and put several different design versions for each part. During the testing, contact probes were used to connect these different branches of a circuit together to get the best performance. Such an arrangement introduced additional parasitic capacitance into the integrated circuits. However, this was the best we could do at that time. Later, an extraction and simulation tool for the FET-SEED process was developed by A. Z. Shang et al. at McGill University [111].

6.2.1 Threshold Circuits

Threshold Set Electrically

As we mentioned before, the most commonly used nonlinear response in neural network applications is the threshold function. For the FET-SEED process, first, we tested a circuit where the threshold could be set electrically (and therefore globally for an array) as shown in Figure 6.2. For simplicity, a LED was chosen as the output device. All FET's shown in Figure 6.2 are depletion-mode FET's with 1μm gate length. The device sizes are: 3μm for Q1, 10μm for Q2, 15μm for Q3, and $20\times 20\mu\text{m}^2$ for the LED. The bias voltages are: $V_+ = 1.0\text{V}$, $V_- = -1.8\text{V}$, $V_{dd} = 4\text{V}$. The circuit is similar to the optical inverter shown in Figure 6.1 except that the threshold is controlled by a depletion-mode FET rather than an optical input signal detected by

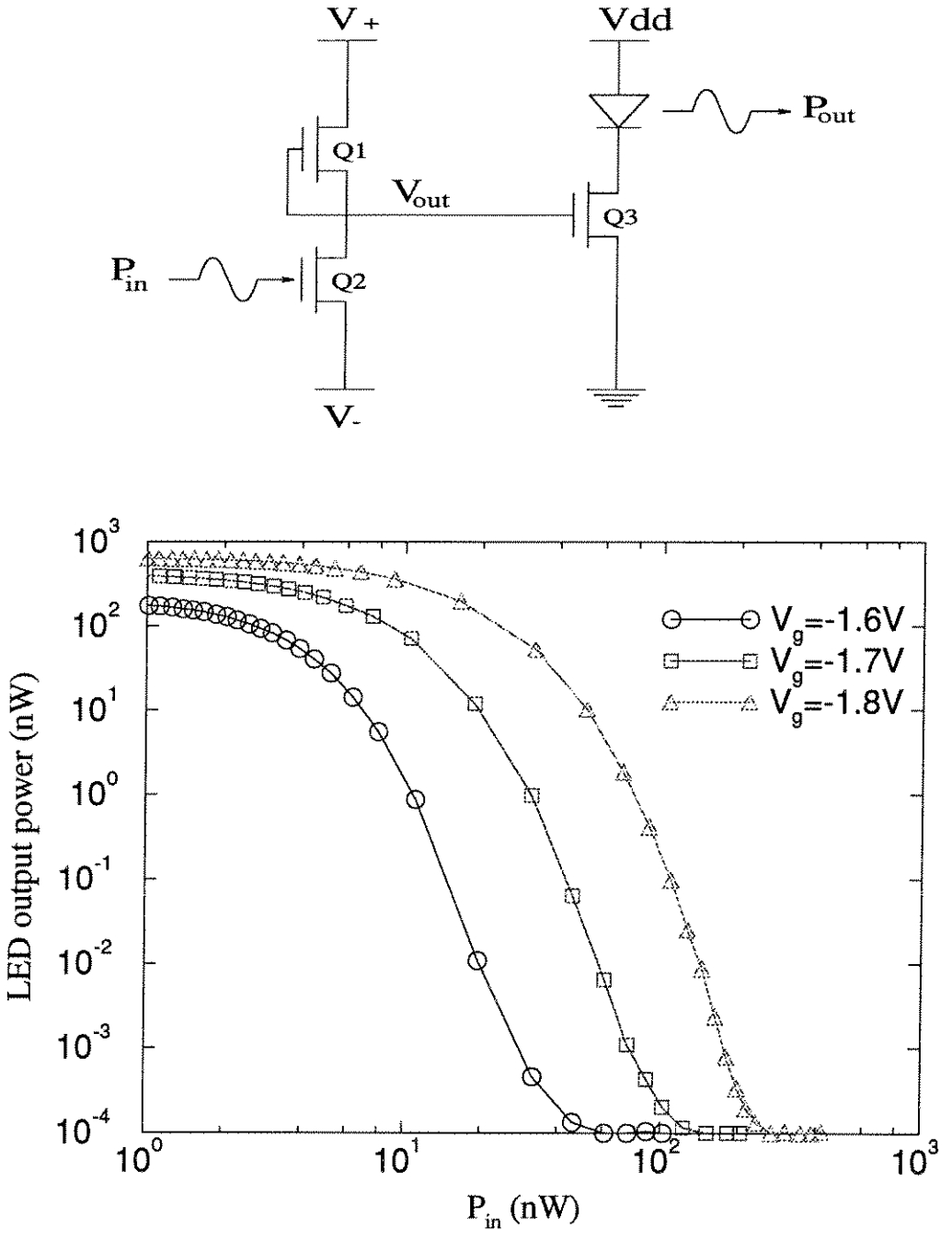


Figure 6.2: The response of an optical inverter from the FET-SEED process. LED was chosen as the output device. The threshold was set electrically by applying different gate bias voltages to the OPFET. On the top is the circuit; on the bottom is the response from the neuron. V_g is the fixed gate bias on the OPFET Q2.

an OPFET. When P_{in} is low, the bottom FET Q2 is turned off by applying negative bias on its gate contact while its backgate contact is left floating in order to achieve high detector gain. Therefore, V_{out} is pushed up to V_+ and Q3 is turned on. When P_{in} is high, the high photocurrent of Q2 will pull V_{out} down to V_- and switch the neuron off.

As we found out from Figure 6.2, the threshold level of the inverter could be set electrically by adjusting the gate voltage of OPFET Q2. The required switching power changed from 10nW to more than 100nW depending on V_g. However, the optoelectronic gain was not high enough: for $V_g = -1.6$V, the switching was completed at $P_{in} \sim 40$nW while the highest output was only 200nW, giving a gain of only 5. This was due to the poor efficiency of the LED (only in the order of 10^{-4}A/W at 1mA current) because the structure and the process were mainly designed for MQW modulators. If we could get similar efficiencies as those fabricated in our lab using double Zn diffusion technique, we could have an output power of 4μW for the same amount of driving current, which would give us an optoelectronic gain of 100.

We also tested the same circuit but operated the OPFET with a floating gate and fixed backgate voltage, which was discussed in Section 4.2.2. The result is presented in Figure 6.3. Since the detector efficiency was improved for OPFET's operating in this configuration, the optoelectronic gain was increased to 170 with the threshold level being reduced to 3nW for $V_{bg} = -2.1$V, even with the same low efficiency LED device. Besides, the output for different threshold levels did not show the big change observed when we operated the OPFET in the earlier configuration. The bad news was that the speed was slower because of the reduced speed of such OPFET.

There was one more problem for the inverter shown in Figure 6.2: the circuit consumed high power. At OFF state, the power consumption was dominated by the saturation current of Q1, which is 13.5μA, resulting in $P_{elec,off} = 2.8\text{V} \times 13.5\mu\text{A} = 37.8\mu\text{W}$. At ON state, the big driving current for the LED was 1.5mA, giving a power consumption of 6mW, which was even higher than the power consumption of the circuit built in our lab. The problem of low optoelectronic gain and high power consumption was mainly due to the low efficiency of the LED, which was not

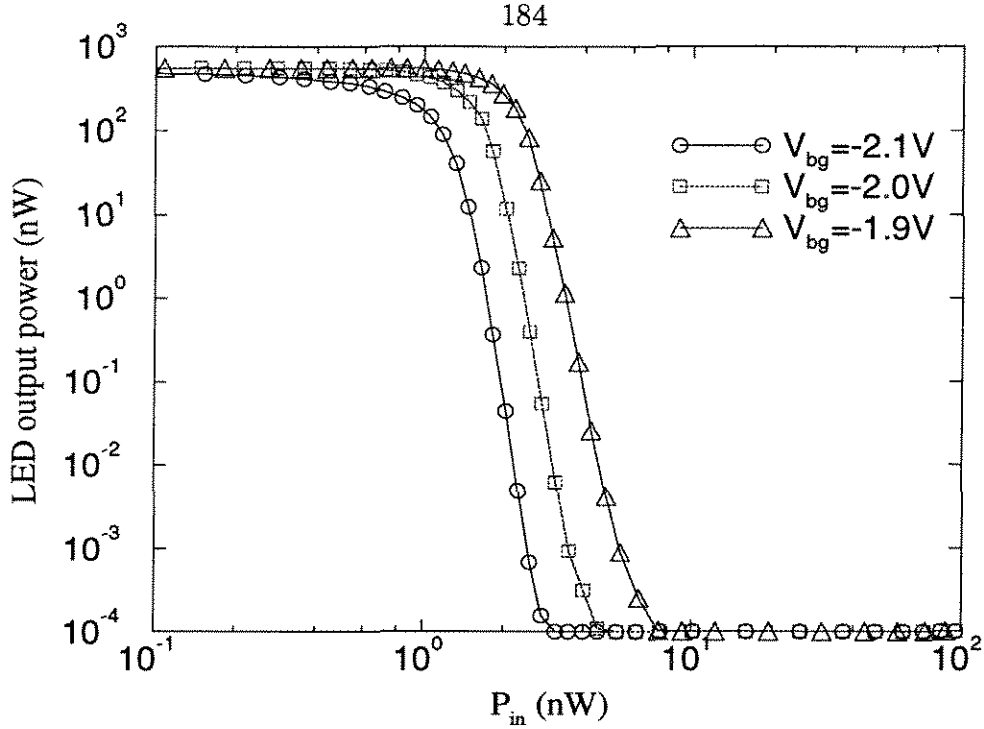


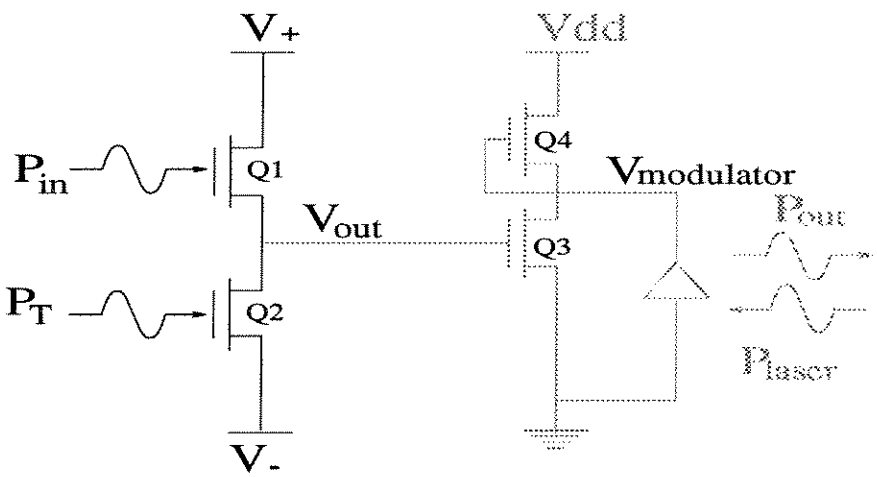
Figure 6.3: The response of an optical inverter. Same circuit and parameter as Figure 6.2 except that the OPFET Q2 was operated with floating gate and fixed backgate voltage.

optimized in this process. By changing the optical output device from LED to MQW modulators, we should be able to get a better result. As a sacrifice, the contrast ratio will drop to 4:1 $\sim$ 5:1.

Threshold Set Optically

As mentioned in Section 6.1, bipolar representation of the interconnection weights are desirable for some neural network algorithm. Therefore, two photodetectors are needed to represent the positive and negative optical inputs. Figure 6.4 is one of such examples. Similar to the circuit shown in Figure 6.1, the threshold branch consists of two OPFET's. However, instead of using a LED output, the optical output is provided by a GaAs/AlGaAs MQW modulator in order to reduce the power consumption of this circuit.

All FET shown in Figure 6.4 are depletion-mode MISFET with 1μm gate length. The device sizes are: 10μm for both Q1 and Q2, 8μm for Q3, 3μm for Q4, and $7\times7\mu m^2$ for the MQW modulator. The total area of the testing circuit is $50\times50\mu m^2$,



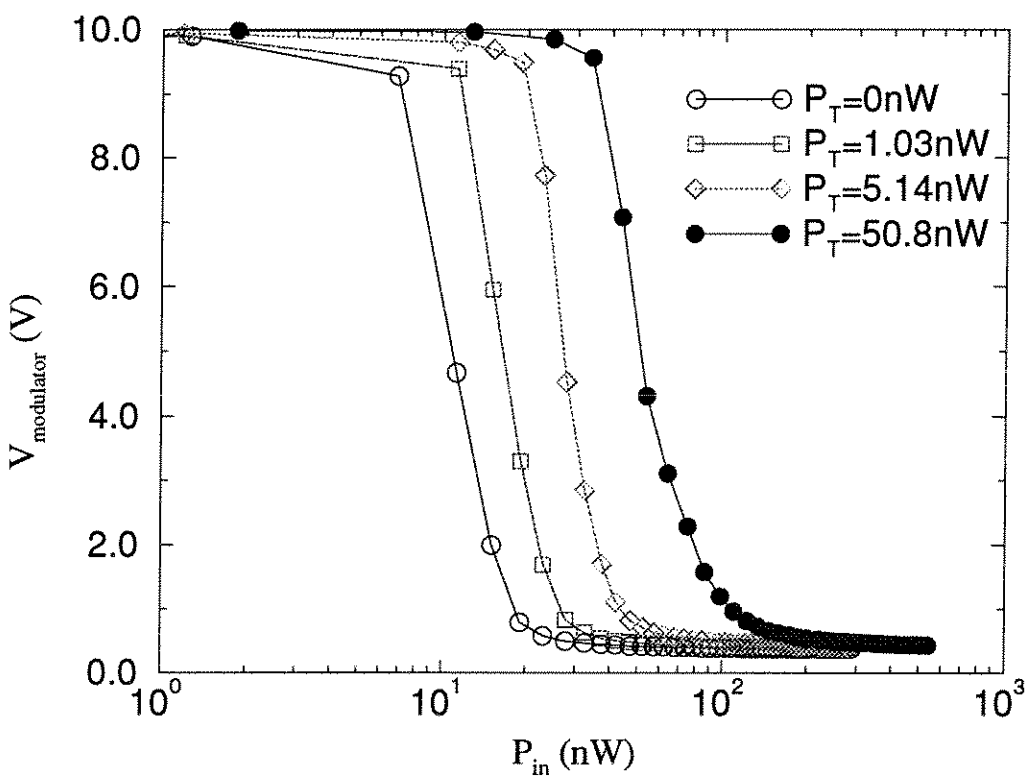


Figure 6.4: The response of a threshold neuron circuit from the FET-SEED process. The threshold was set optically with two OPFET's operating with fixed gate voltages and floating backgates.

excluding the contact pads. More rigorous design should be able to fit all devices into an area of less than $30 \times 30\mu m^2$. The bias voltages are: $V_+ = 0.2V$, $V_- = -2.2V$, $V_{dd} = 10V$. The two OPFET's are operated with fixed gate voltages: $V_{g1} = -2.6V$, $V_{g2} = -3.7V$. The input laser power, P_{laser} is fixed at $10\mu W$. In the experiment, due to the limitation of our measurement setup, we could not measure all four optical powers simultaneously. Therefore, we only measured the driving voltage for the modulator instead of the modulated light from the modulator. Since the reflectivity of this modulator was known from Figure 2.10, we could easily figure out the reflection light from its driving voltage. This was the standard measurement setup we used during the testing of all FET-SEED circuits involving modulators.

The response of this circuit was plotted in Figure 6.4. The testing wavelength was near λ_0 for the modulator in order to increase the detector responsivity. Therefore, high driving voltage means high output from the modulator. In return, since the response in Figure 6.4 gives high $V_{\text{modulator}}$ for low P_{in}, the resulting response is an optical inverter. On the other hand, if input laser on the modulator were near λ_1, high driving voltage would give low output from the modulator and what we get would be a normal threshold circuit. Since the detector responsivity was reduced at λ_1, the required switching power would be larger, so would P_T.

As shown in Figure 6.4, the threshold was set optically by the input power on the bottom OPFET, P_T. Since the response of the OPFET was not linear, the change in threshold levels was not the same as ΔP_T. The actual threshold level changed from 10nW to 100nW, similar to what we got in the electrically-controlled threshold circuit shown in Figure 6.2. As an estimate of the optoelectronic gain, the modulator has reflectivities of 0.23 at 10V bias and 0.05 at zero volt bias for $\lambda = 845nm$ according to the testing data provided by the foundry. At 850nm, it was much higher: 0.65 at zero bias and 0.14 at 10V bias. Therefore, a $10\mu W$ incident power on the modulator will give us $2.3\mu W$ at ON state for λ_0, corresponding to an optoelectronic gain of 100 for the case of $P_T = 0nW$. We can easily double this gain by doubling the incident power, P_{laser}, with only marginal increase on the power consumption of the whole circuit.

For the case of $P_T = 0$nW, the current was about 160nA between the V_+ and the V_- terminals and 4μA into the V_{dd} terminal at ON state. So, $P_{elec,on} = (0.16 \times 2.2 + 4 \times 10)\ \mu$W$= 40.4\mu$W. At the OFF state (high P_{in}), the current for V_+ and V_- did not change much, but it was increased to about 50μA for V_{dd}. So, $P_{elec,off} = 500\mu$W. For other threshold levels, the power consumption of the two OPFET's increased slightly, but the main consumption came from the 10V V_{dd} power supply. For power dissipation limit of 5W/cm^2, we would be able to fabricate neuron arrays with densities up to 1.85×10^4neurons/cm^2 assuming the same 50% ON state probability.

6.2.2 Bump Circuits

For some neuron training algorithms, such as back error propagation training, a non-linear response corresponding to the derivative of the threshold function is needed. Such functions are called "bump" function since the output is high for the center region and low for the two ends. Such function can also be used to compared the similarity between two analog signals. It is only when two signals are similar that we will get high output from the neuron. Besides applications in back error propagation algorithm, such functions also have the basic functionality of the radial-basis functions. Unlike the threshold circuits built from the FET-SEED process, all the bump circuits built on the FET-SEED chip were broken into several parts and then reconnected together when doing the testing due to the lack of simulation tools.

Electrically-controlled Bump Circuits

Since the response of a bump circuit is similar to a NOR gate of digital circuits, we used the NOR gate as the basic output circuit in our design. The position of the bump circuit can be controlled by the same thresholding branch used earlier. What we need is only an electric inverter to invert the output from the threshold branch. Figure 6.5 shows a bump circuit where the position of the bump function can be controlled electrically by changing the gate voltage on the OPFET detector. The thresholding branch consists of Q1 and Q2 and the output branch consists of Q3, Q4, Q5, and the

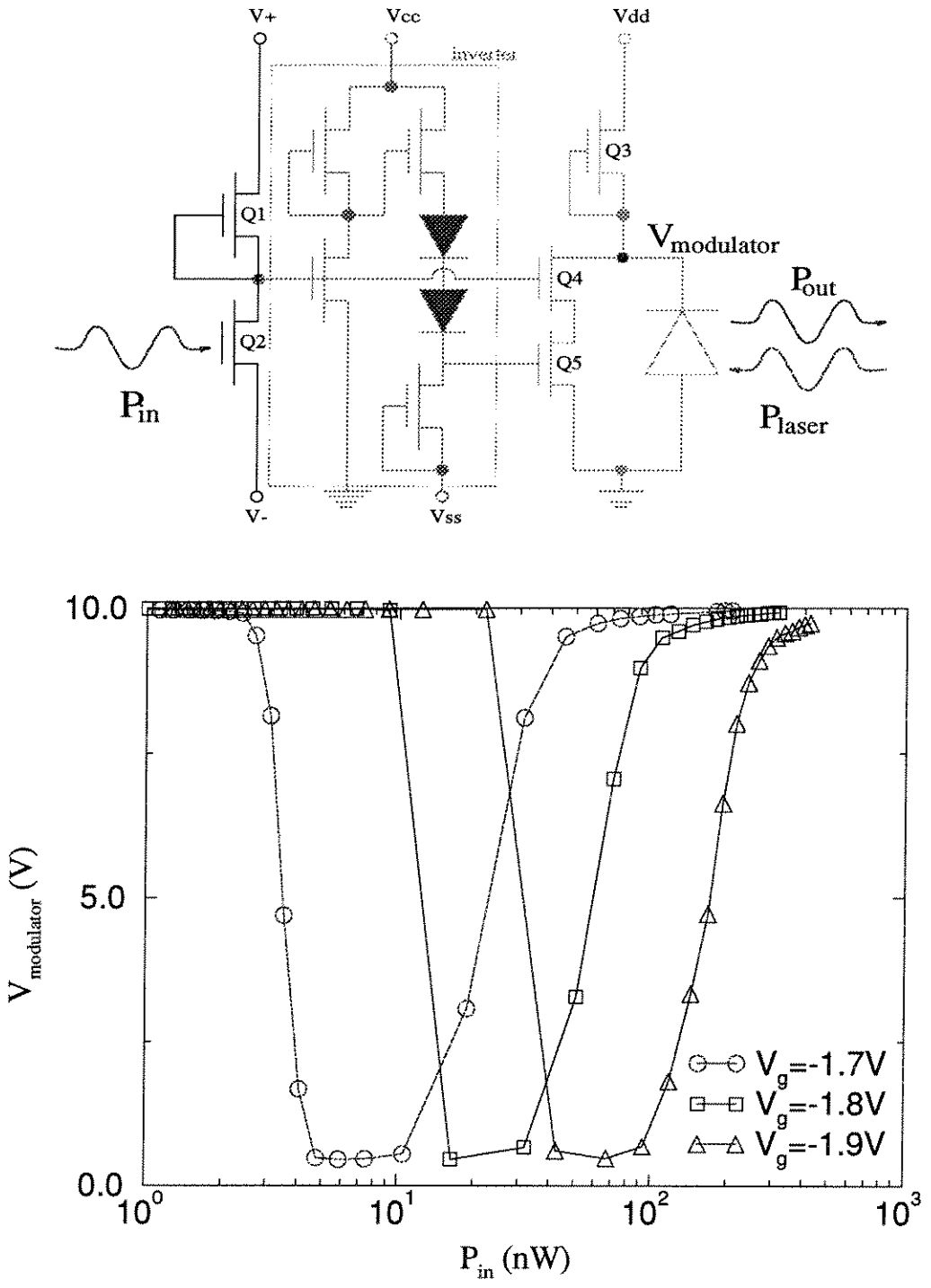


Figure 6.5: The response of a bump circuit. The position was set electrically by applying different gate bias voltages to the OPFET. V_g is the fixed gate bias on the OPFET Q2.

output modulator. In between is the diode-shifted inverter. Two level-shifting diodes are needed in order to get the correct bias result.

As for the device dimension, all FET's are 1μm devices. The gate widths are: 3μm for Q1 and Q3, 10μm for Q2, and 8μm for Q4 and Q5. Actually, we can improve the design by using a double-gate FET instead of using both Q4 and Q5. All FET's in the buffered FET logic (BFL) inverter have 10μm gate width. The output modulator is 7μm$\times$7μm. Because only depletion-mode FET's are available from this process, several bias voltages are needed in order to set all FET's correctly: $V_+ = 0.2$V, $V_- = -2.0$V, $V_{dd} = 10$V, $V_{cc} = 2.5$V, and $V_{ss} = -3.2$V. The response was presented in Figure 6.5 as well. We can clearly see the bump-shaped response and the position-control by the electric gate bias on the OPFET detector. The lowest threshold level was $5 \sim 10$nW. The optoelectronic gain was more than 230 for P_{laser} fixed at 10μW, calculated from the same method described in Section 6.2.1. Similar to the circuit shown in Figure 6.2, the optoelectronic gain could be increased by operating the OPFET with fixed backgate voltage and floating gate contact. The response is plotted in Figure 6.6. Due to the increase of detector responsivity, the position of the bump function was shifted to 1nW and the optoelectronic gain was increased to 2000.

The power consumption of this neuron was much higher due to the addition of the buffered FET logic inverter circuit. For the output branch, it was the same as the output circuit in Figure 6.4, so the power consumption was 500μW for OFF state and 40μW for ON state. For the thresholding branch, it was the same as the thresholding circuit used in Figure 6.2 but with different bias voltage. The current was measured to be 200nA when P_{in} was low and 13μA when P_{in} was high, corresponding to a power consumption of 0.44μW and 28.6μW, respectively. The main power consumption comes from the BFL inverter. The V_{ss} – V_{cc} branch was always on and the current was about 0.55mA, giving a power consumption of 3.135mW. In addition, when the output from the thresholding branch was high, an additional 0.55mA current went through the V_{cc} – ground circuit, giving a power consumption of 1.375mW. If we assumed that all circuits had 50% possibility of staying in either state except for

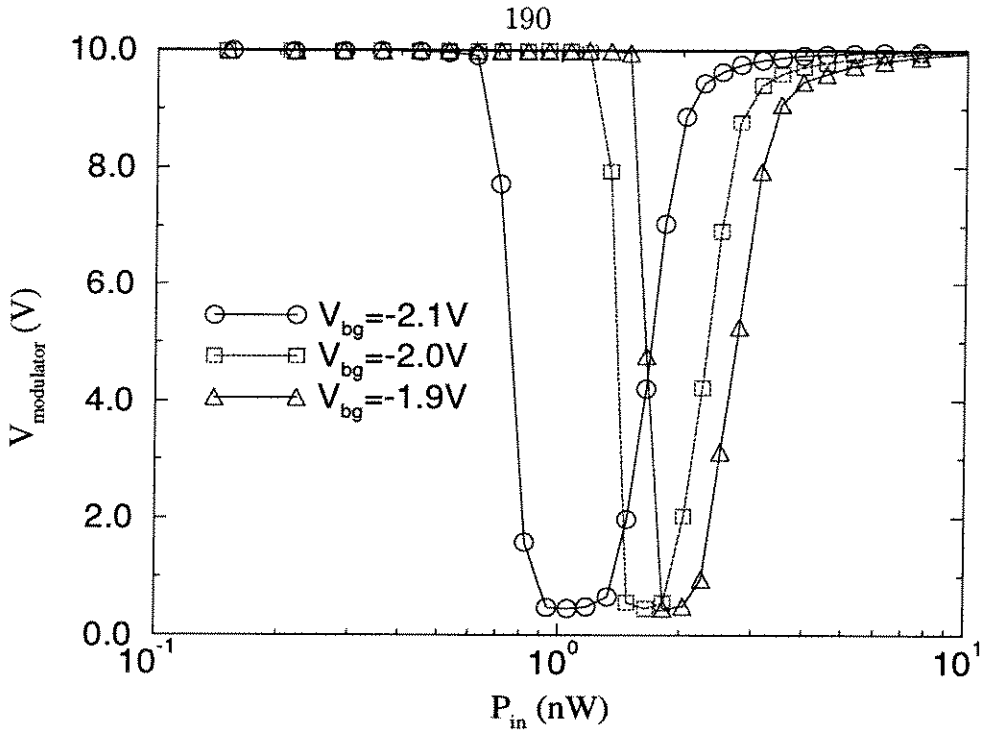


Figure 6.6: The response of a bump circuit from the FET-SEED process. Circuits were similar to the one shown in Figure 6.5 except that the OPFET were operated with fixed backgate voltage and floating gate contact.

the V_{ss} – V_{cc} circuit which was always turned on, the average power consumption of the bump circuit was $0.50 \times (500 + 40 + 0.44 + 28.6 + 1375)\mu$W+3.135mW, which summed up to be 4.1mW. As a result, the maximum neuron density for an array can be 1.2×10^3neurons/cm^2 assuming 5W/cm^2 power dissipation limit for the wafer.

Optically-controlled Bump Circuits

When the position of the bump function need to be set optically, two photodetectors are required as the circuit shown in Figure 6.7. It is similar to the one shown in Figure 6.5 except that the thresholding circuit is changed to two OPFET's. The two OPFET's are operated with floating gates. Their backgate voltages were fixed at $V_{bg1} = -3.2$V and $V_{bg2} = -4.3$V.

From the result plotted in Figure 6.5, we found that the response from this circuit was not as good as what we had expected. First, the width of each bump circuit was too wide, which was due to the inappropriate W/L ratio of the FET's and could be fixed if we had a simulation tool when doing the circuit design. However, a more

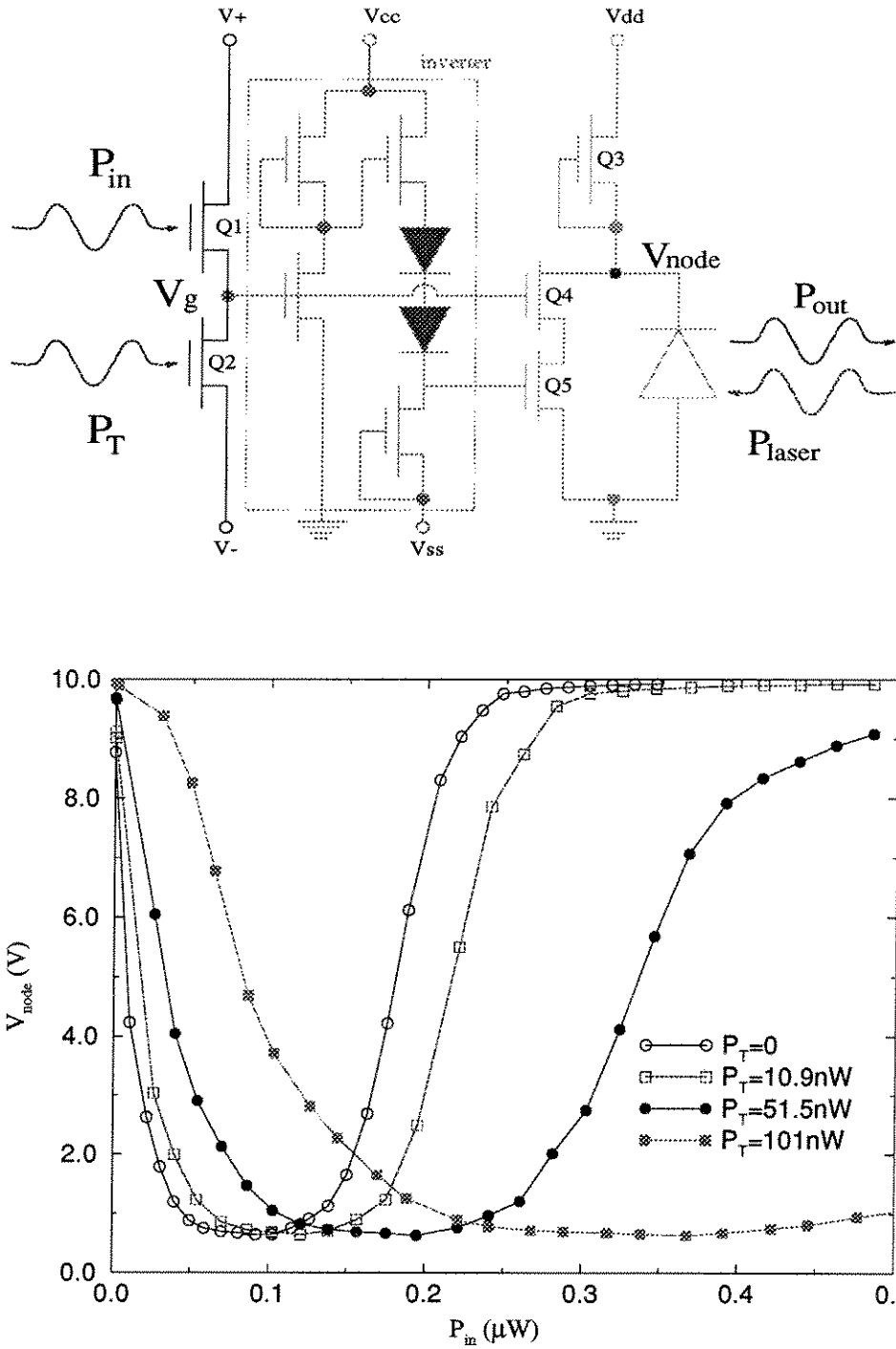


Figure 6.7: The response of a bump neuron circuit from the FET-SEED process. The position of the bump function was set optically. OPFET's had fixed backgate voltages but with gate floating.

severe problem was that the circuit was very difficult to turn off even when P_{in} was smaller than the threshold level. On the other hand, for large threshold levels, it was difficult to turn off for the high input side as well. The problem came from the fact that the FET's were all depletion-mode FET's. As a photodetector, we had to bias its gate voltage to subthreshold level in order to reduce the dark current and increase the optoelectronic gain of the circuit. Originally, when P_{in} was small, we biased the two OPFET's so that V_g was at an appropriate level. As P_{in} increased, it would try to push V_g up toward V_+. However, as V_g went up, since the backgate bias for the top OPFET Q1 was fixed, the backgate-source bias voltage for this detector decreased. The result was reduced photocurrent through Q1 and V_g would not increase as fast as we wanted it to be. As V_g increased more and more, this problem became more and more severe. Therefore, the result was that V_g increased slower and slower. For the case of threshold neuron circuits, this problem could be fixed to a certain degree since we had another thresholding response from the output branch. For the bump circuit, however, it became very difficult to fix this problem. The result was that either we would not be able to turn off the circuit for high input levels or we could not turn it off for low input levels.

If the two OPFET's were operated with gate voltages fixed and the backgate floating, this situation became even worse because the FET current was more sensitive to the gate-source voltage. To solve this problem, we had to place more stringent requirement on the circuit design or add another electrical threshold circuit. Even if we could turn the circuit off at both extreme of the input range, the slow rise of V_g for large inputs would give a low optoelectronic gain as the result. As for the power consumption, it was similar to the case of electrically-control bump circuits since the only difference was the thresholding branch, which consumed little power compared to the BFL inverter circuit and the output circuit.

In conclusion, two different kinds of optoelectronic neuron circuits were successfully fabricated using the FET-SEED process. If LED's were chosen as the output devices, the power consumption of the circuit was normally large due to the poor efficiency of the LED. When modulators were the output devices, we could get low

power consumption neurons. However, three problems exist for the FET-SEED process due to the fact that only depletion-mode FET's are available. First, a lot of bias voltages are required in order to set the circuit working at the appropriate point. For example, for the circuit shown in Figure 6.7, totally seven voltage lines were needed, including the five voltages shown and two backgate bias voltages for the OPFET's. For an OEIC array, the distribution of so many voltages requires a certain area, which will reduce the maximum neuron density. Second, the absence of enhancement-mode FET makes the circuit design more complex. Usually, we end up with more transistors and high power-consuming design. For example, if we have enhancement-mode FET, the BFL inverter can be reduced to just two FET's and the power consumption will be greatly reduced as we will see later when discussing the MOSIS GaAs circuit design. Finally, when two optical inputs are needed, the response from connecting two OPFET's in series is not as good as what we want it to be due to the change in gate-source bias of the top OPFET. To solve these problems, a foundry that provides better electronic circuits is needed.

6.3 MBE Regrowth on MOSIS GaAs Circuits

The MOSIS/Vitesse HGaAs3 process provides both enhancement-mode and depletion-mode MESFET's, so it can accommodate more flexible circuit design as compared to the FET-SEED process. However, the heterostructure for optical output devices must be grown on fully-fabricated chips in order to implement optical I/O for the neuron circuit. Since all the contact and interconnection line on the HGaAs3 process are refractory metals, the fabricated chip can withstand temperatures up to 530°C for several hours. As a result, we can use MBE to grow the output heterostructure at temperatures below 530°C. The fabrication procedure was described in Chapter 2. The regrowth of the LED structure has been reported before [64], so we will concentrate on the regrowth of GaAs/AlGaAs MQW modulators in this thesis.

6.3.1 Circuit Design

The addition of enhancement-mode MESFET's and the availability of simulation tools make the circuit design a much simpler task. We used HSPICE for the design simulation and MAGIC for converting circuit designs into physical layout designs. All device parameters used in the simulation were provided by the foundry. After finishing the layout design, the data file was sent to MOSIS via e-mail and it was then put into fabrication.

The circuit tested for the MOSIS run is a 10×10 array of bump circuits. Threshold circuits can be easily fabricated as well since all we need is a pair of photodetector or one photodetector with a DFET as current source. The output can be driven by an EFET if an LED is the output source. Compared to the FET-SEED circuit shown in both Figure 6.2, there is no major difference except that the driving FET is changed to EFET, so we can connect V_{-} to ground and reduce the required voltage swing. However, the power consumption of the circuit will not change much because a large driving current is still needed for the LED. If MQW modulators are the output devices, as the one shown in Figure 6.4, the MOSIS circuit design is actually more complicated. This is because the breakdown voltage of the MOSIS MESFET is only 4V, which is less than the 8-10V required for modulators. In return, either we have to reduce the contrast ratio from the modulator or a more complex circuit is needed, which will increase the power consumption compared to similar circuits on the FET-SEED chip. Therefore, for simple threshold circuits, it is hard to say which process is better. We need to consider the optoelectronic gain, layout area, voltage requirement, and power consumption.

Electrical Bump Circuit

For complex circuits, such as the bump circuit shown in Figure 6.7, the addition of EFET makes the designing much easier even though we need to worry about the breakdown voltage for circuits with modulators as the output. As we pointed out, the main power consumption of the bump circuit on FET-SEED chips came from

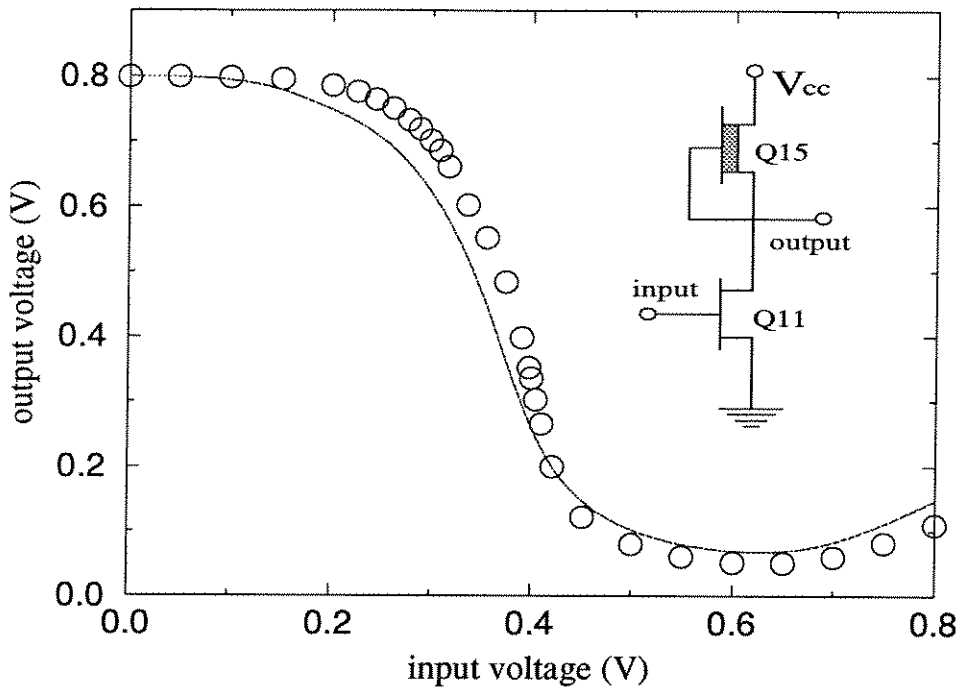


Figure 6.8: The performance of a simple inverter from MOSIS GaAs circuits. Circles corresponds to actual testing result while the line represents HSPICE simulation.

the buffered FET logic inverter. For the MOSIS circuits, the inverter can be simply comprised of a EFET and a DFET and the power consumption can be greatly reduced. Both simulation and actual measurement results are plotted in Figure 6.8 for such an inverter. The circuit is also shown there. The DFET Q15 has gate length of 4.0μm and gate width of 2.0μm; the EFET Q11 has gate length of 1.0μm and gate width of 6.5μm. The bias voltage is $V_{cc} = 0.8$V. We can see that the simulation result is very close to the actual measurement. While the BFL inverter requires four transistors and two level shifting diodes, which consumes almost 4.0mW power, we can do the same thing with just two transistors in the MOSIS GaAs run. In addition, only one low bias voltage is needed, instead of two high bias voltages.

With the simple inverter, we can build the bump circuit by adding a NOR gate design, now consisting of one DFET and two EFET's. The circuit and measurement results are shown in Figure 6.9. The DFET Q14 has gate length of 7.0μm and gate width of 2.0μm; the two EFET Q12 and Q13 have the same gate design, with gate length of 1.0μm and gate width of 18.0μm. We can see that the actual testing has similar performance as the simulation but the width is slightly smaller. When the

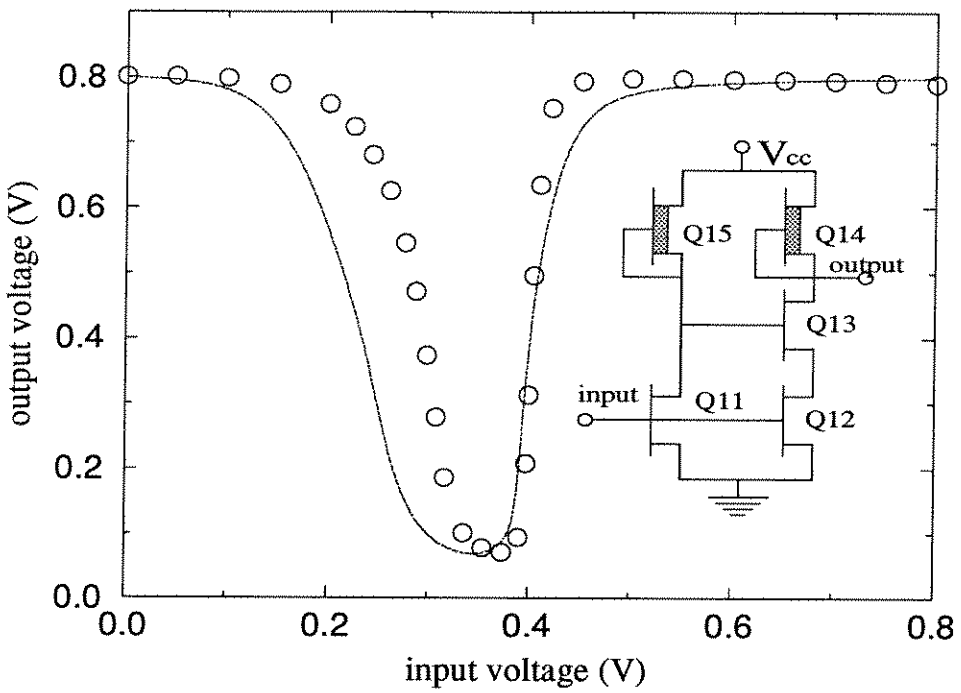


Figure 6.9: The performance of a bump circuit from MOSIS GaAs run. Circles corresponds to actual testing result while the line represents HSPICE simulation.

input voltage is small, this circuit consumes very little power. As the input power increases, the power consumption goes up. For input voltages higher than 0.4V, the current does not change much but stays at about 80μA, corresponding to a power consumption of 32μW, which is much less than the 4mW power consumption of the bump circuit built from FET-SEED process.

Voltage Control Circuit

There remains one major problem for the circuit shown in Figure 6.9. Its output voltage only swings from 0 to 0.8V. By using a separate power line for the output driver, it can be increased to about 4V, which is limited by the breakdown voltage of the MOSIS GaAs MESFET. However, to get good contrast ratios from MQW modulators requires voltages up to 10V. Therefore, a voltage control circuit is needed in order to support such high voltage. This can be done by distributing the high voltage on two FET's connected in series such as the one shown in Figure 6.10. The circuit was proposed by Yakov Royter [109] at MIT with some modification made at

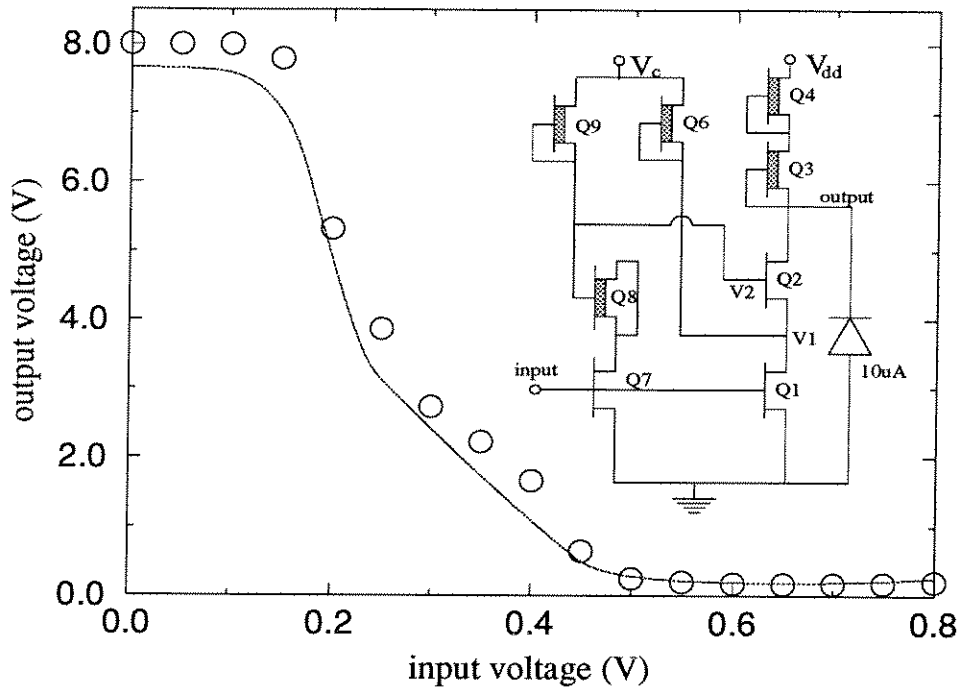


Figure 6.10: The voltage control circuit for supporting high driving voltage on MQW modulators. Circles corresponds to actual testing result while the line represents HSPICE simulation.

Caltech.

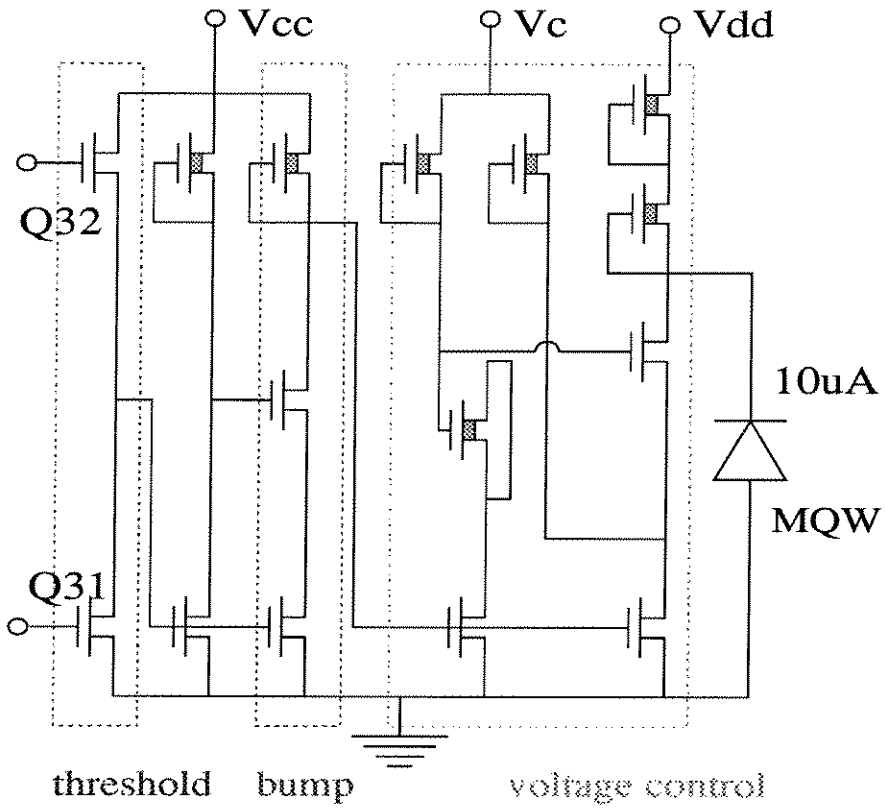
For V_{dd} set at 8V, as we can find from Figure 6.10, a voltage swing up to almost 8V is achieved. This was done by distributing the 8V voltage on either DFET Q3 and Q4 or EFET Q1 and Q2. The effect of Q6 is to make sure that when Q1 is turned off by a low input, the node voltage V_1 will not surpass the 4V limit set by V_c. When Q1 is turned on, Q8, which is just a level-shifting diode fabricated by shortening the drain and source contact of a transistor, and Q9 will push V_2 to be about 0.8V over V_1 so that Q2 is also turned on. The widths and lengths of these MESFET's are: 14×1.0 for Q1, 10×1.0 for Q2, 2.0×2.5 for Q3, 2.5×2.0 for Q4, 2×3.5 for Q6 and Q9, 3.5×1.0 for Q7, and 10×3.5 for Q8, all in units of micron. The HSPICE simulation was done by assuming a constant current of 10μA flowing through the MQW modulators.

Due to the additional trouble of distributing an 8V voltage over MESFET's with 4V break-down voltage, the circuit consumes more power than its counterpart on FET-SEED chips, which is comprised of only two DFET's. At low input level, the

currents through V_c and V_{dd} are negligible, less than 1μA. However, as the input level goes up, I_c and I_{dd} increases gradually and saturates for $V_{in} \geq 0.55$V. At $V_{in} = 0.8$V, we measured $I_c = 0.22$mA and $I_{dd} = 0.16$mA, corresponding to a static power consumption of $4\times0.22+8\times0.16(\text{mW}) = 2.16$mW. As a result, even though the power consumption of an electronic bump circuit on MOSIS chips is two orders less than that of similar circuits on the FET-SEED chip, the voltage control circuit consumes half the power of the whole FET-SEED bump circuit. Considering that the voltage control circuit only consumes power as it is turned on, the average power consumption of the bump circuit and the voltage control circuit is $50\% \times (2.16+0.032) = 1.10$mW, which is 27% of the power consumption of the FET-SEED bump circuit. To further reduce this amount, there are two possible solutions. First, we can adjust the widths and lengths of all FET's of the voltage control circuit. By optimizing the design, we should be able to reduce the required amount of current. Ultimately, if the required driving voltage of MQW modulators can be reduced to 4 or 5V, we will be able to drive the modulator with just two MESFET's, just like the FET-SEED circuit. In return, the static power consumption will be greatly reduced to about 100μW, assuming the same 10μA current on MQW modulators. This can be done by using asymmetric triangular quantum wells [110] instead of square quantum wells, which is currently under investigation at MIT [69].

Optoelectronic Bump Circuit

To control the input level of the bump response, we put a pair of OPFET's in the optoelectronic bump circuit. They were connected in series as a thresholding branch, with the central node voltage as the input to the electrical bump circuit shown in Figure 6.9. The final version of the circuit design is shown in Figure 6.11. To increase the responsivity of the OPFET, OPFET's with serpentine gate design, as shown in Figure 4.21(b), were adopted with 1.0μm gate length. The total gate width of the OPFET detector is 55μm and its gate opening window has an area of $16.5 \times 17\mu\text{m}^2$. To simulate the response of such circuits using HSPICE, we need to develop a model for the OPFET. As discussed in Chapter 4, the OPFET operates like a MESFET



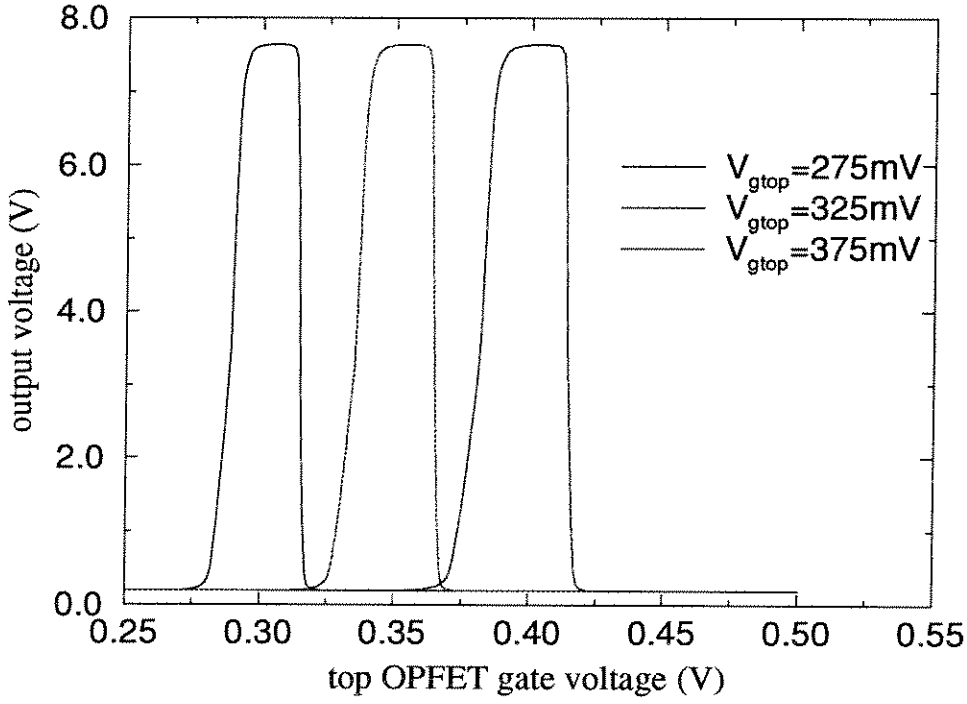


Figure 6.11: The optoelectronic bump circuit fabricated through MOSIS GaAs N4CJ run and its HSPICE simulation result. Gate voltages on the two OPFET's corresponds to different optical input powers.

with gate voltage controlled by the optical input power. Therefore, we can simply adopt the HSPICE model for MESFET's. When interpreting the result, just keep in mind that the gate voltage of an OPFET is actually corresponding to a certain optical input level. The simulation result of the optoelectronic bump circuit is presented in Figure 6.11 where we plot the output voltage as a function of different gate biases on the bottom OPFET. Different gate voltages on the top OPFET was assumed to show the controlling of the location of bump response. As we increase $V_{g,top}$, i.e., increase the optical input power on the top detector, the position of the bump response moves to higher input levels.

Since the HSPICE model does not include the breakdown behavior of MESFET's, the resulting current from simulation was less than the experimental measurement result. At $V_{g,top} = 0.325$V, the simulation gave us $I_{cc} = 0.207$mA, $I_c = 17.67\mu$A, and $I_{dd} = 10.11\mu$A at high $V_{g,bot}$. At the bump response, we had $I_{cc} = 0.573$mA, $I_c = 0.186$mA, and $I_{dd} = 0.129$mA. Therefore, the average power consumption (static) is about:

$$50\% \times [(0.207 + 0.573) \times 0.8 + (0.01767 + 0.186) \times 4 + (0.01011 + 0.129) \times 8] \text{ (mW)}$$

which gives us a total of 1.27mW. Assuming 5W/cm^2 power dissipation limit for the GaAs wafer, the maximum integration density for this optoelectronic bump circuit design is 3.94×10^3neurons/cm^2.

The design submitted to MOSIS is a 10×10 array of the optoelectronic bump circuit. A picture of the fully fabricated chip from MOSIS is shown in Figure 6.12. Each pixel occupies an area of $170 \times 165\mu$m^2, slightly larger than the limit predicted by HSPICE simulation. This design is mainly limited by the circuit power consumption since the area occupied by all transistors is only $80 \times 90\mu$m^2 and could be further reduced. A $25 \times 25\mu$m^2 regrowth area is specified for each pixel based on results from previous experiments [65]. Unfortunately, as we will see later, a change in the foundry process results in a much thicker dielectric stack and requires larger regrowth opening.

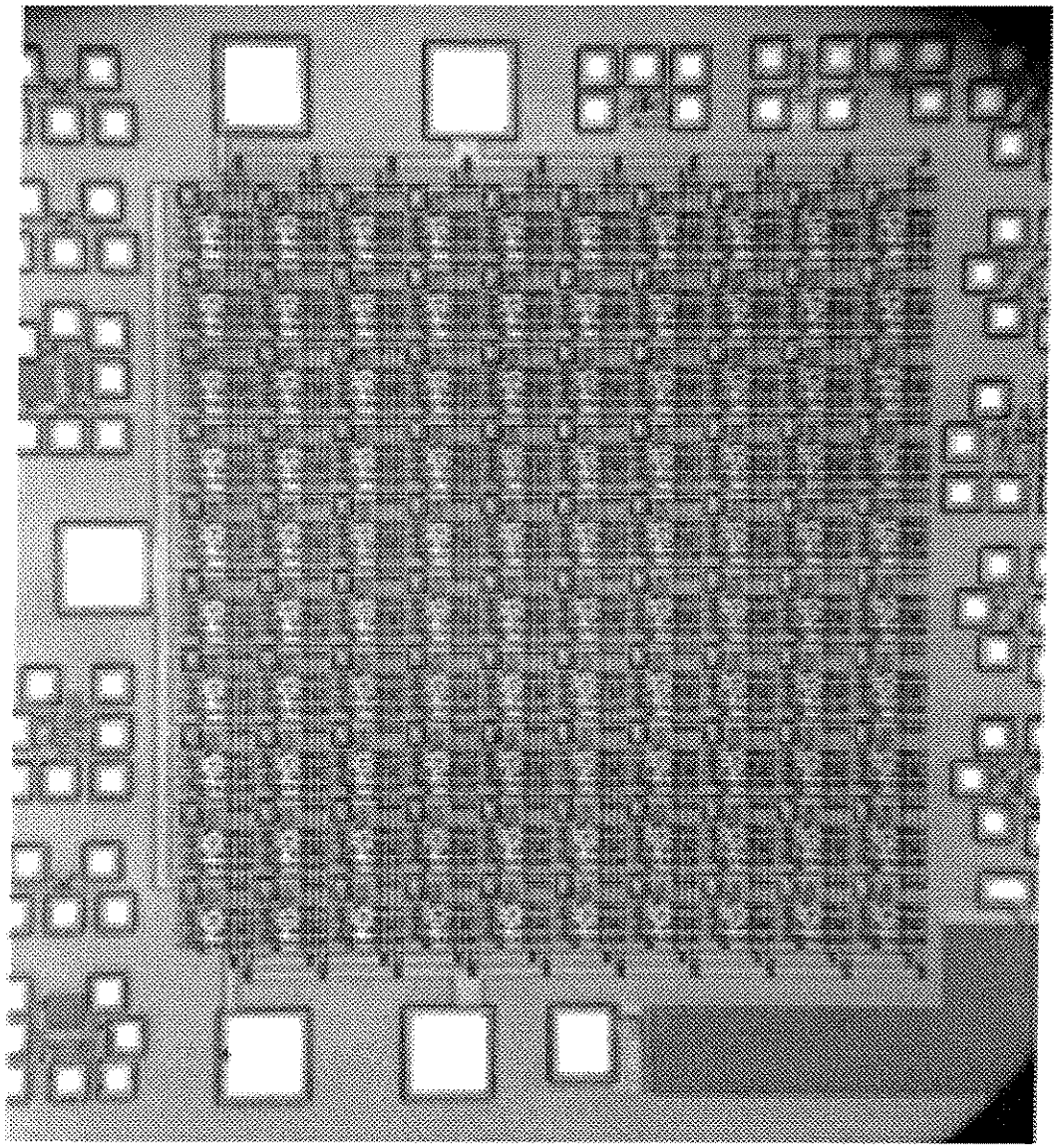


Figure 6.12: Fully fabricated MOSIS chip consisted of a 10×10 array of optoelectronic bump circuits and some testing designs. The large contact pads, designed for wire bonding, are $200\times200\mu m^2$.

6.3.2 Circuit Measurement Results

The measurement results of the inverter, the electrical bump circuit, and the voltage control circuit were presented earlier with simulation results in Figure 6.8, 6.9, and 6.10, respectively. Here we will concentrate on the results involving optical signals, including the thresholding pair of OPFET's and the whole optoelectronic bump circuit.

Optical Threshold Response

The response of a thresholding pair of OPFET's should be straightforward if the two devices were isolated from each other. Under this situation, when the top FET detects more input power, it will generate more photocurrent which then charges up the central node voltage. On the other hand, when the input level on the bottom OPFET is higher, the central node voltage will be pulled down to low level (ground). However, as we pointed out earlier, the MESFET's fabricated by Vitesse are sitting on top of a uniformly p-doped background, which means that all the devices have common backgate potentials in general. When the central node voltage increases, we have a higher source potential for the top OPFET. As a result, the bottom transistor will have a positive backgate-source bias while the top transistor will have a negative backgate-source bias, which gives a tendency of pulling down the central node voltage. In order to finish the switching of the threshold circuit, a much higher input power is required for the top OPFET. This will push power consumptions up and is not good for high density integration.

In order to solve this problem, we can put a positive guard ring (n-type ohmic contract) around each FET detector [65], which generates a reverse biased p-n junction. When the bias on the guard ring, V_{ring}, is high enough, this reverse bias will totally deplete the p-layer and the two FET detectors will be separated from each other. This is what we did on the bump circuit design. The improvement of the guard ring bias on the threshold circuit response is very obvious as we can see from Figure 6.13 where the central node voltage is plotted as a function of either the optical input

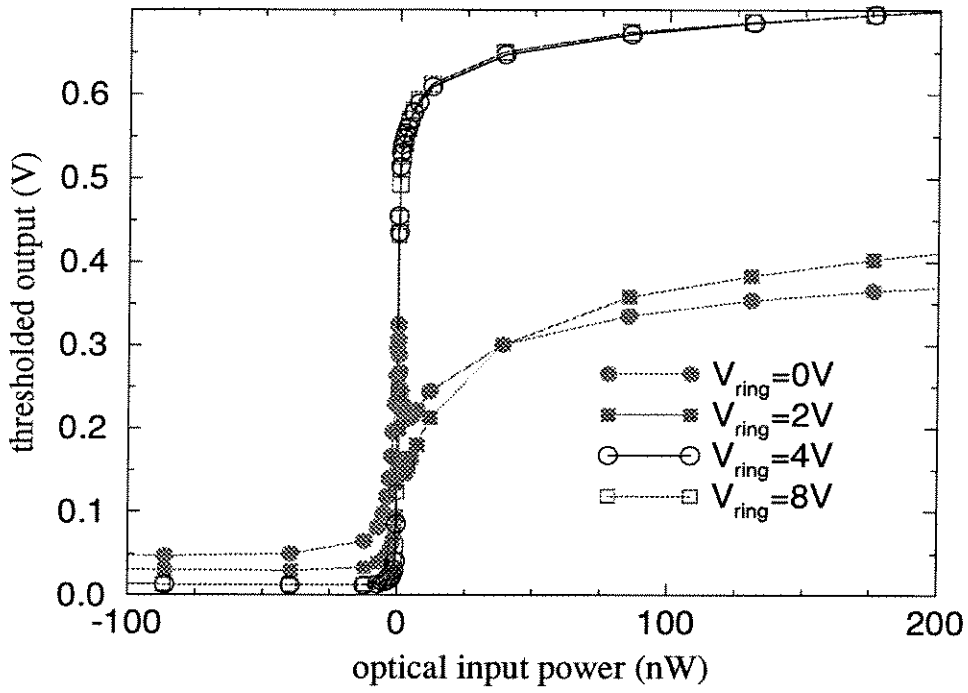


Figure 6.13: The switching characteristics of optical threshold circuit at different guard ring biases.

on the top OPFET while the bottom OPFET is in dark — define as positive input — or the optical input on the bottom OPFET while the top detector is in dark – defined as negative input. Different curves correspond to different guard ring biases. For $V_{ring} \leq 4.0$V, as the optical input increases, the central node voltage increases very fast up to 0.6V for a positive input. Since the thresholded output is connected to the gate contact of a MESFET, when this output voltage is higher than 0.6V, the gate contact starts to turn on. Therefore, the central node voltage increases slowly above 0.6V. On the other hand, when the guard ring bias is not high enough, the backgate of the two OPFET's are connected together and strange behavior is observed: the central node voltage goes down in the beginning and then recovers. The thresholding circuit has not completed the switching even when the optical input is as high as 400nW. For the turning-off behavior, the effect of the guard ring bias is not as significant even though some improvement exists. In Figure 6.14, we plotted the response of the thresholding OPFET's as a function of the guard ring voltage at different optical input levels. We find that a minimum of 4.0V bias is required in order to solve the backgating problem.

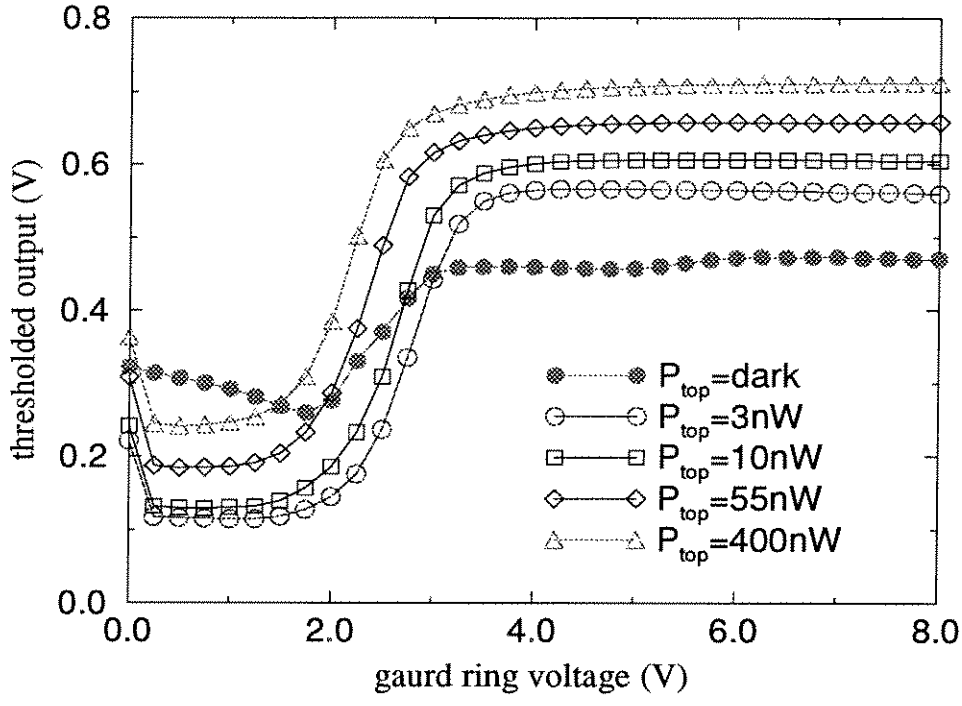


Figure 6.14: The effect of the guard ring bias on the output of the threshold circuit.

In Chapter 4, we pointed out that the operation of high gain OPFET detectors depends strongly on their backgate potential. Therefore, the adoption of a guard ring contact will affect the photocurrent of such devices. Figure 6.15 shows the measured photocurrent of a OPFET designed for the bump circuit. We find that the protection of the guard ring bias increases the OPFET gain. At 1nW input, the photocurrent increases from 2.4μA at V_{ring} = 0V to 10μA at V_{ring} = 4.0V. This is due to the reduction of backgate junction area by guard ring isolation. Therefore, for the same charging current, a higher voltage can be built up to push more current through the OPFET.

Optoelectronic Bump Circuit

The performance of the bump circuit shown in Figure 6.11 is presented in Figure 6.16, with the location of the bump response controlled by an optical signal on the top OPFET detector. The measurement was done with the following parameters: V_{cc} = 0.7V, V_c = 4.0V, V_{dd} = 8.0V, and V_{ring} = 5.0V. It was found that the bump circuit could finish the whole switching process for optical inputs less than 3.0nW

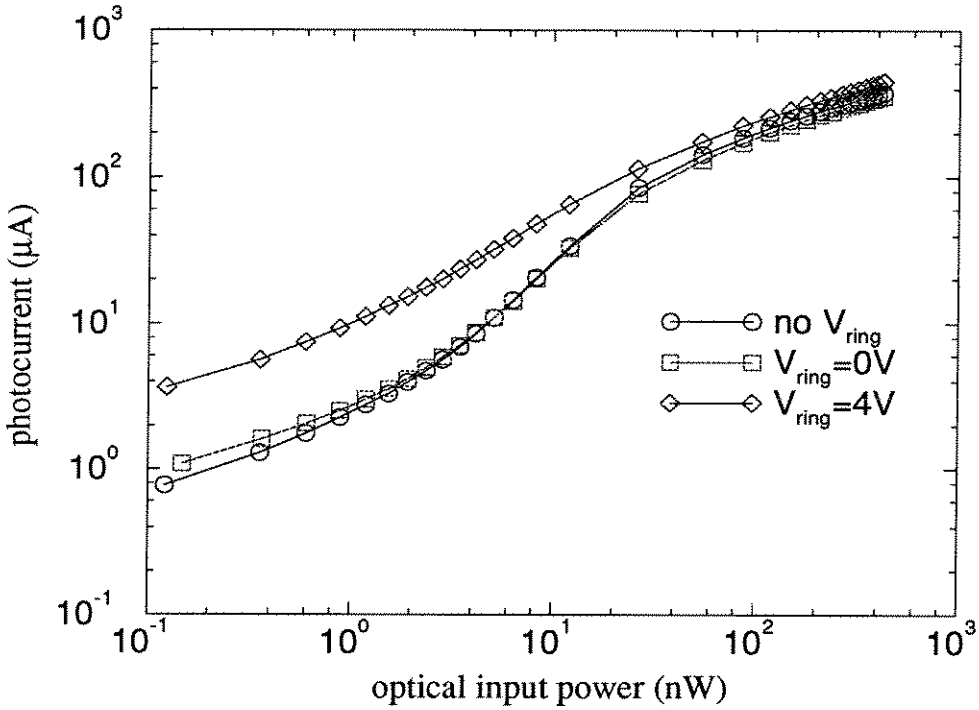


Figure 6.15: The response of OPFET detector at different guard ring biases.

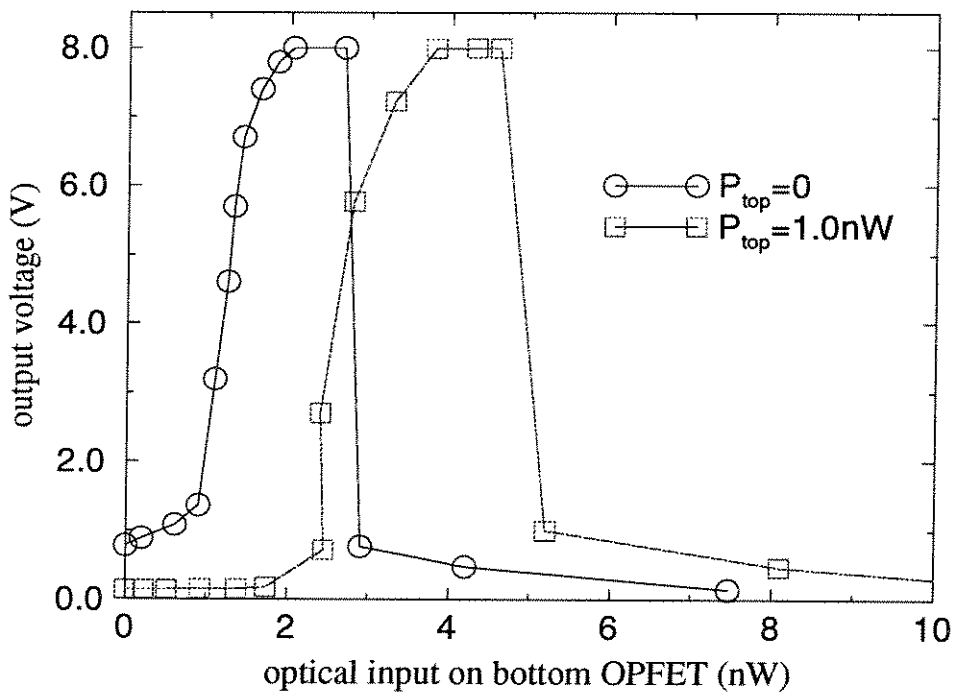


Figure 6.16: The response of the optoelectronic bump circuit with the location of the bump response controlled locally by optical signals detected using the top OPFET.

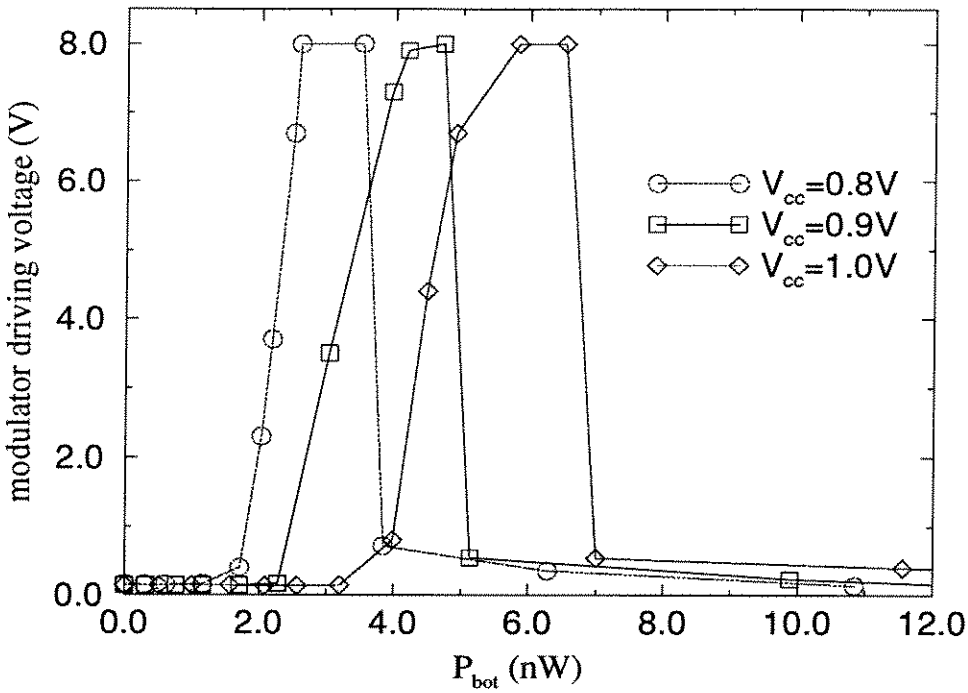


Figure 6.17: The performance of optoelectronic bump circuit. The location of the bump response can be set globally by changing V_{cc} bias.

when the top OPFET was in dark. In addition, we found that the location of the bump response could be shifted by an optical signal as small as 1nW. Due to the nonlinear response of the OPFET detectors, when the bump circuit shifted to the right, its width also increased.

In the above discussion, we always assumed that the location of the bump response was controlled by adjusting the two optical inputs of the thresholding OPFET's while V_{cc}, V_c, and V_{dd} were assumed to be fixed. In practice, the location of the bump response can be set globally, i.e., for the whole array, by adjusting V_{cc} as shown in Figure 6.17, which shows the modulator driving voltage as a function of optical inputs on bottom OPFET — top OPFET is in dark — at different V_{cc}. V_{dd} is set at 8.0V, V_{ring} at 5.0V, and V_c at 4.0V. From Figure 6.9, we knew that the bump circuit will switch as the output from the thresholding OPFET's reached $0.3 \sim 0.4$V. At $V_{cc} = 0.8$V, the thresholded output is around 0.4V for symmetric design when both top and bottom OPFET's are in dark. Therefore, a small increase in the optical input on the bottom OPFET is good enough to pull down the thresholded output voltage into the bump response region. As we increase V_{cc}, the thresholded output is raised

as well when both top and bottom OPFET's are in dark. In return, more optical input is needed to switch the neuron circuit. This is what we observed in Figure 6.17.

6.3.3 Regrowth and Testing

Post Processing

After receiving the fully fabricated GaAs chips (N4CJ run) from MOSIS, we need to clean the dielectric layers in the regrowth wells in order to grow single crystalline material on them. We asked the foundry, Vitesse, to use their powerful RIE to etch a portion of the dielectric stack rather than etching all the way down to the GaAs substrate in order to avoid surface damages which would likely result in defects in the regrowth material. In earlier regrowth experiments with the LED structure, the dielectric stack could be easily etched away by dipping into BOE or HF solution. It was found that the total thickness of such dielectric stack was about 5μm. The adoption of such wet etching resulted in a clean and undamaged GaAs surface.

However, for the N4CJ chips with 10×10 bump circuit arrays, we found out later that there were likely some change on the foundry process. Even with 2μm of dielectric layer removed by the foundry, there were about 4μm left on the specified regrowth well, giving a total thickness of 6μm. In addition to the increase of the total thickness of the dielectric stack, its composition was also changed because we found out that the etching rate of BOE etch was reduced and such wet etching became very anisotropic. The etching rate on the sidewalls were much faster than the etching rate in the vertical direction. As a result, the cleaning process had to be modified.

The cleaning and MBE regrowth were done at MIT. First, a long oxygen plasma etching was applied to clean the residues on the regrowth well introduced by the Vitesse RIE etch. Aluminum masks were then deposited on the fully fabricated chips to protect the dielectric layers except on the regrowth wells. Metal protection was used instead of photoresist because of the long time required to etch all dielectrics. Either RIE or a combination of RIE/BOE etch was then applied. As we mentioned earlier, for the BOE etching, a very wide undercut on the regrowth well sidewalls was

observed, which might damage circuits defined there. Once the regrowth well was cleaned, the chips were loaded into the MBE chamber together with a control GaAs wafer, sample 9123. Following a four minute oxide desorption at 580°C, the modulator structure was grown at 500°C. The total growth time for the 5.8μm stack was almost 10 hours because only on Aluminum cell was working at the time of regrowth while materials with different Al percentage were grown. Such a long exposure time to high temperature was very bad for the fully fabricated chips and resulted in circuit damaging as we will discuss later. If the second Al cell can be fixed in the future, the regrowth time could be able to reduced to three or four hours. Furthermore, the oxide desorption can be done at 400°C with improved technique [108]. We expect these two improvements to be done in the near future.

Figure 6.18 shows the picture of one MOSIS chip with regrowth material on it. The bright reflection comes from single-crystalline material while poly-crystalline GaAs is everywhere else. For the 10×10 array, we can see that single-crystalline modulators are grown on all specified wells. Under microscopes, the regrowth material showed good quality except for some large defects on the large testing area on the lower right-hand corner.

Once the regrowth chips were received from MIT, the poly-crystalline materials were etched away following the procedures described in Chapter 2. An insulating nitride was deposited at 310°C using PECVD. Contact metal was evaporated on areas with nitride opened by CF_4 plasma etching. Finally, an anti-reflection coating was deposited on the modulator surface using Si sputtering at room temperature. The final version of the chip is shown in Figure 6.19. Besides the array, some modulators were fabricated on the lower right-hand corner with contact pads for testing. Figure 6.20 shows the pictures of a pixel in the 10×10 array: before regrowth, after regrowth, and after final processing. The single-crystalline modulator turned out to be much smaller than the $25 \times 25 \mu m^2$ regrowth well due to problems in cleaning the dielectric layers in such a small area.

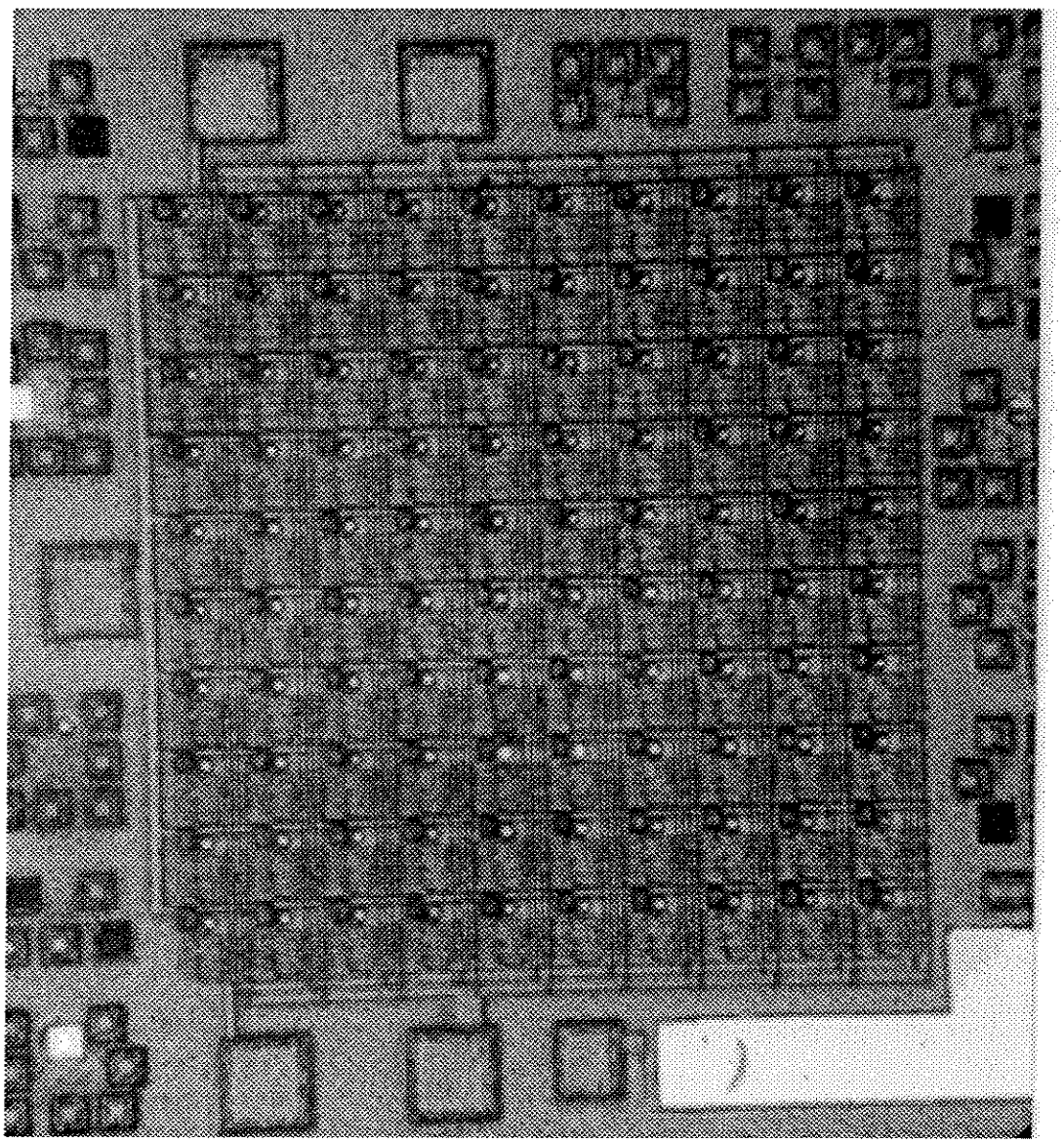


Figure 6.18: MOSIS chip after MBE regrowth of MQW modulator structure. The bright areas are single-crystal GaAs while poly-crystalline material are everywhere else. Some defects are presented in the large single-crystalline GaAs on lower right-hand corner.

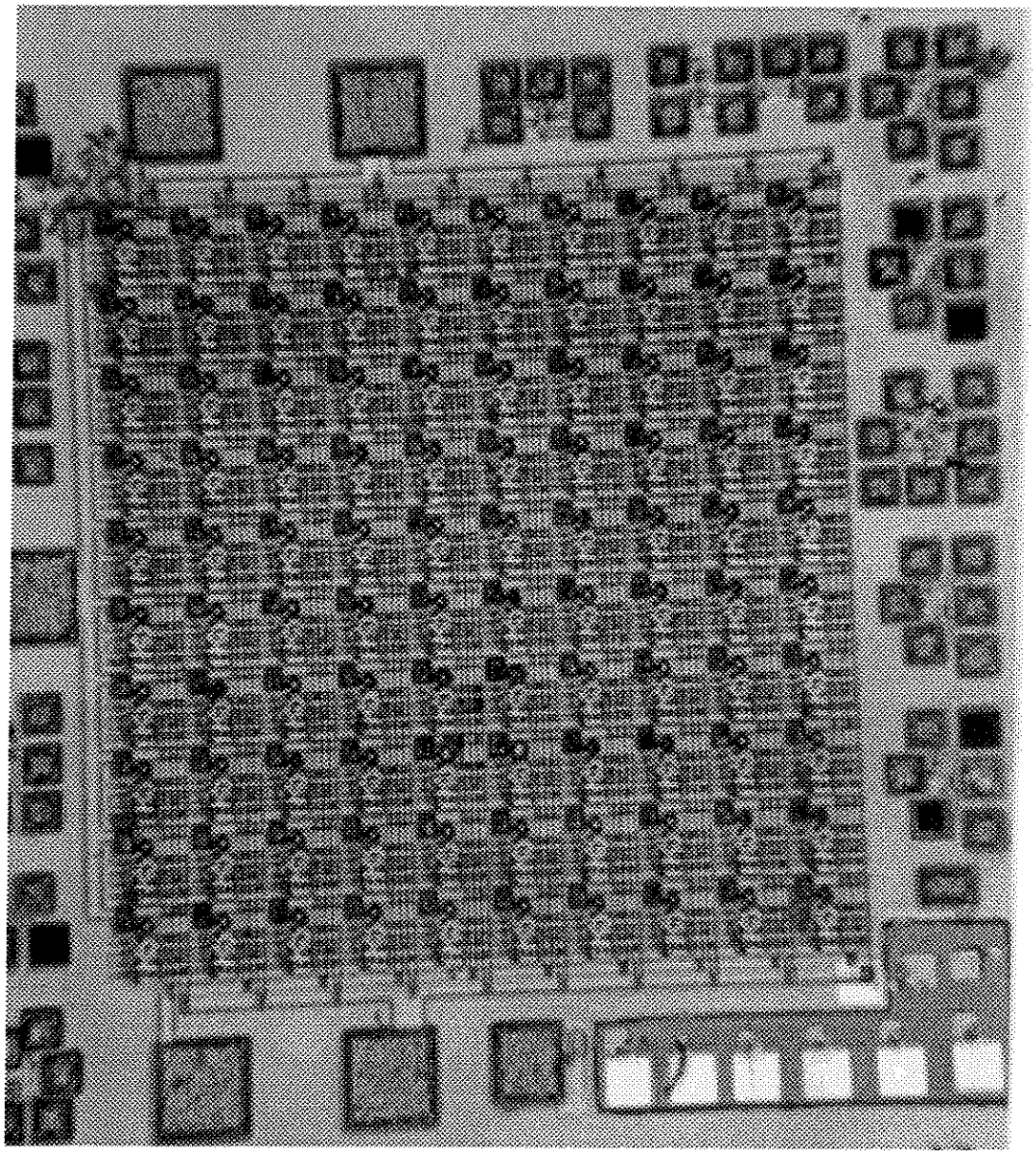


Figure 6.19: Finished MOSIS/regrowth chip. The bright objects on the lower right corner are deposited metal contacts for testing modulators.

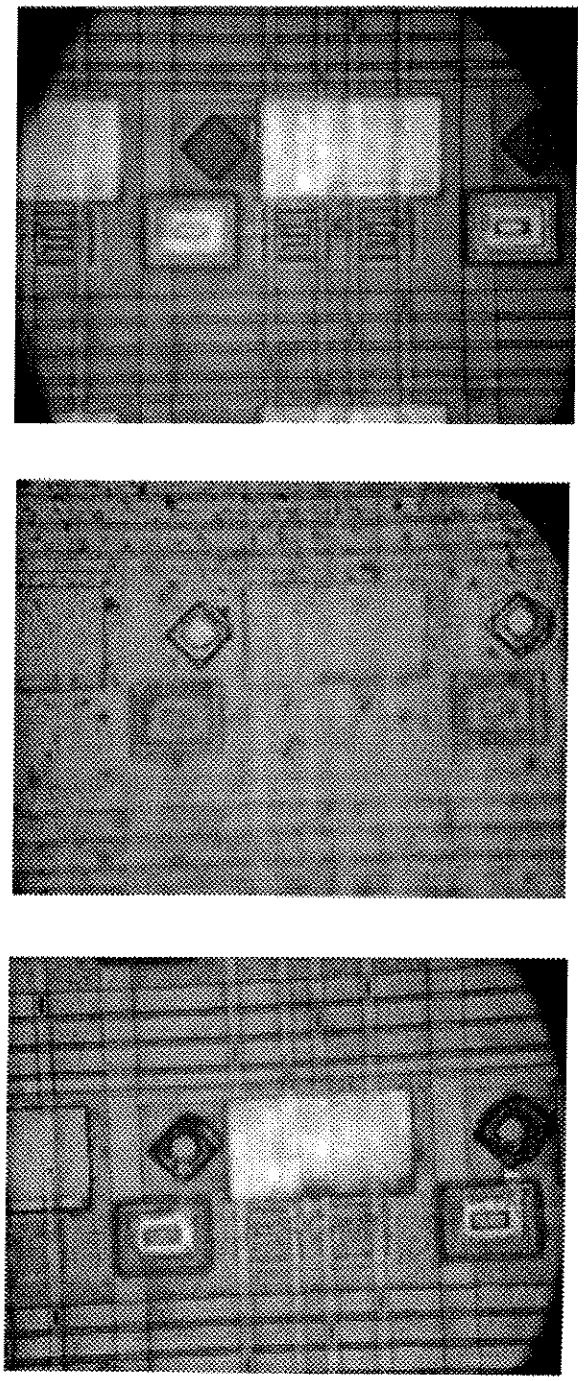


Figure 6.20: A pixel on the 10×10 bump circuit array. From top to bottom: before regrowth, after regrowth, and finished device. All MESFET's are covered by metal in order to avoid any undesirable optical effect.

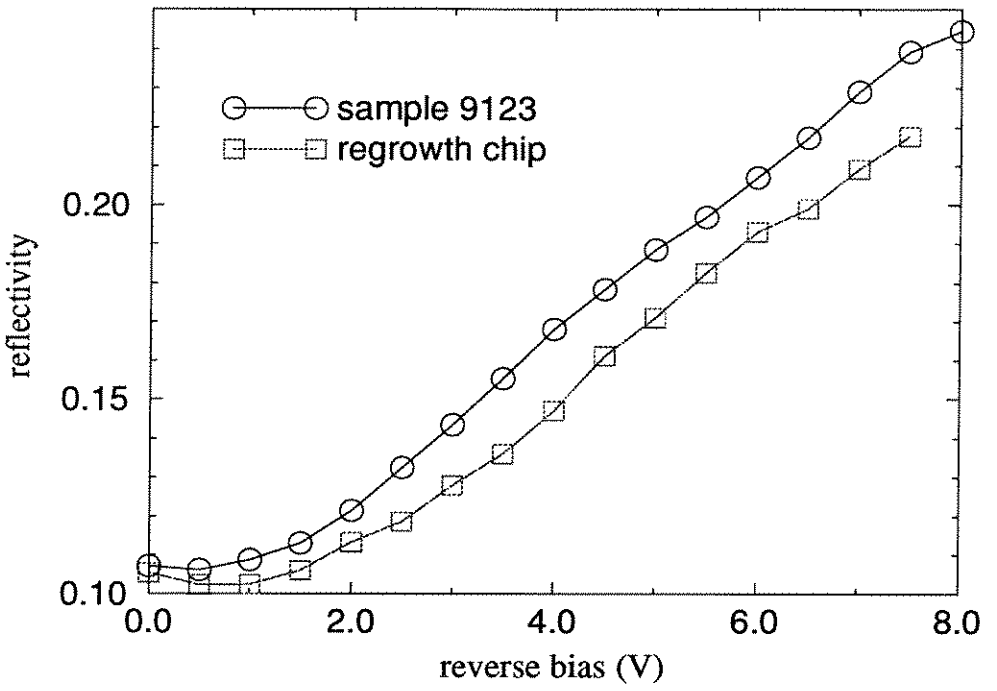


Figure 6.21: The comparison between modulators fabricated from regrowth chips and those from control wafer at $\lambda = 839.9$nm. The data for regrowth chip are multiplied by a factor of 2 for easier comparison.

Testing of Modulators

The modulators fabricated from the regrowth chips showed similar response as those fabricated on the control wafer, sample 9123. Figure 6.21 and 6.22 show the modulation results for $\lambda = 839.9$nm and $\lambda = 845.3$nm, respectively. Almost identical responses were observed even though the reflectivity of the regrowth modulator was considerably lower. This was due to the same reason mentioned earlier when discussing the OPFET responsivity in Chapter 4. The regrowth modulator had small area, $20\times20\mu m^2$, while the incident beam has a diameter of about $40 \times 40\mu m^2$. This resulted in the loss of some light that was not reflected from the modulator opening. Contrast ratio of 2.2:1 was obtained for the regrowth modulator with bias voltages up to 7.5V.

However, we observed a large reverse current for modulators fabricated on the regrowth chips. At 6V bias, we measured a reverse current of more than 0.5mA. On the control wafer, sample 9123, the measured reverse current was less than 1μA even

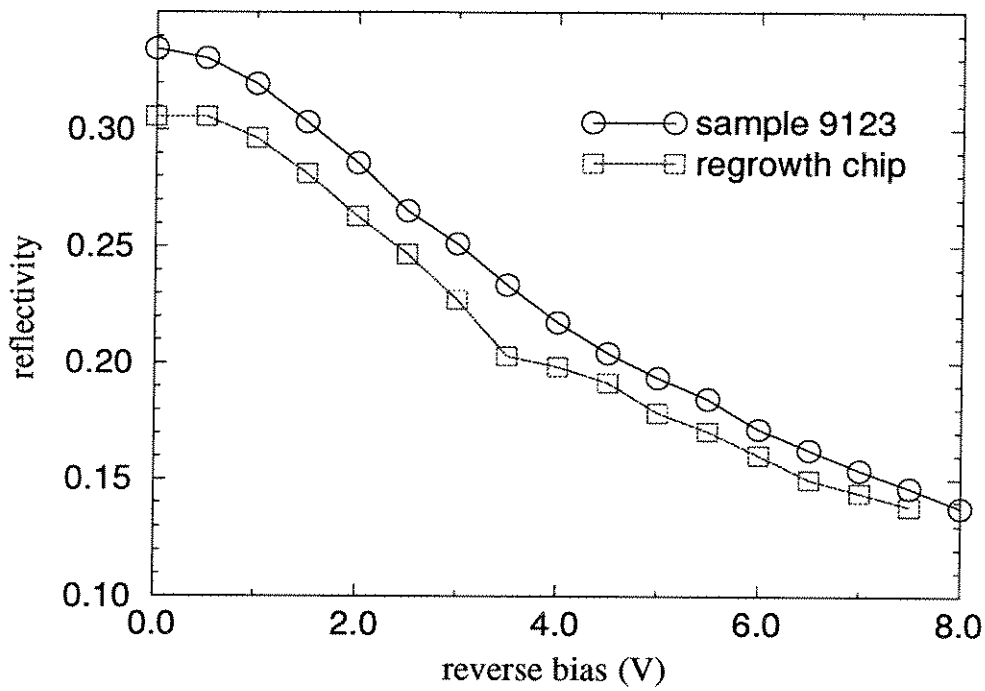


Figure 6.22: The comparison between modulators fabricated from regrowth chips and those from control wafer at $\lambda = 845.3$nm. Regrowth chip data are multiplied by a factor of 2 for easier comparison.

at 10V bias. Such high reverse current may be due to errors during the processing if proper isolation etch was not provided or it may come from surface recombination. With the 4μm resolution limit of our lithography capability, the 25μm opening proved to be more than we can handle and some of the isolation nitride may be overetched. That was another possible reason. The problem of such high reverse current was that the driving circuit shown in Figure 6.11 was not designed to support such amount of current. Therefore, we have to reduce this high reverse current before we are able to operate the 10×10 bump circuit array. In addition, we need to increase the opening of the regrowth well so that more uniform devices can be processed due to the resolution limit of our lithography.

Testing of Circuits on Regrowth Chips

Because of the long regrowth time explained earlier, the circuit showed obvious degradation. On the two regrowth chips, one chip still works. The electrical bump circuit showed a small shift with a much wider bump response while the voltage control

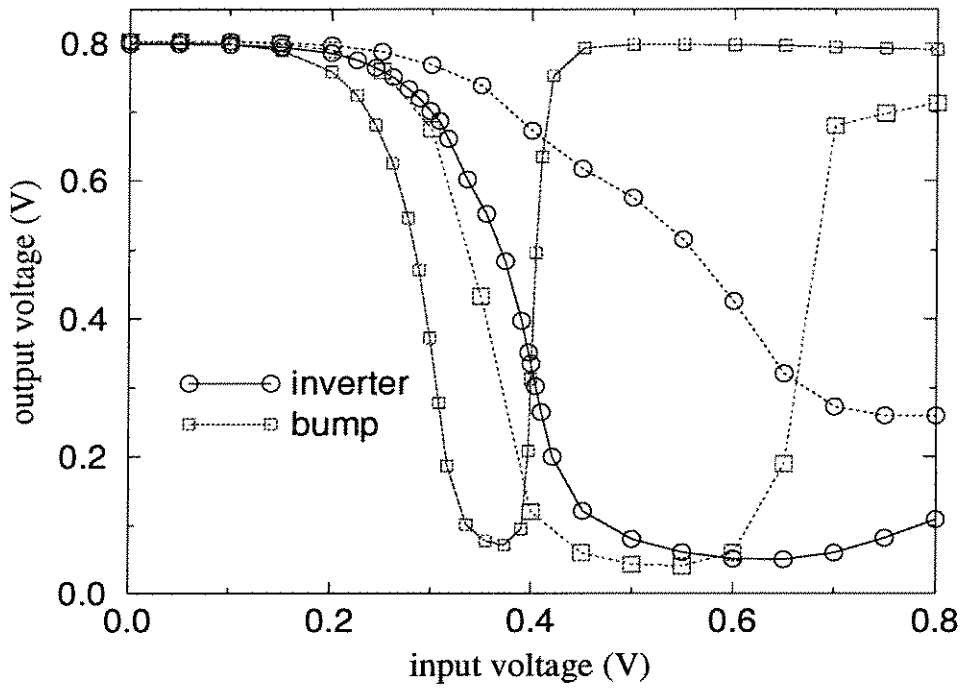


Figure 6.23: Test results of electrical bump circuits on one regrowth chip (in dotted line), in comparison with chips directly received from MOSIS (in solid line).

circuit worked almost the same as shown in Figure 6.23 and 6.24. However, on the other chip, even though the voltage control circuit still worked well, the bump circuit operated abnormally. We found out that it was due to some damage in the inverter circuit so that the inverted signal of the thresholded output was always larger than 0.4V. Therefore, one of the inputs to the bump circuit was always "ON" no matter what the thresholded result was. From earlier regrowth experiments on LED structures, we expect such degradation to be improved greatly when the total regrowth time can be reduced to three to four hours when the second Al cell in the MBE machine is fixed. We can also reduce the degree of degradation by designing more robust circuits. As we can see from Figure 6.23, it is the inverter circuit that causes the problem. If we optimize the W/L ratio of the two transistors in the inverter, such degradation could be less significant.

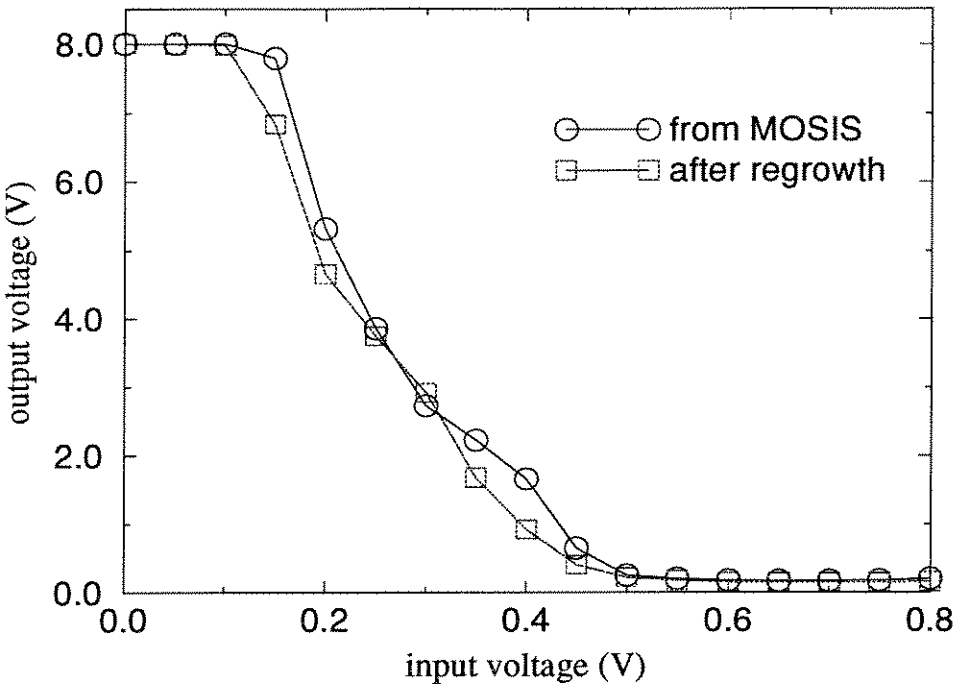


Figure 6.24: Test results of voltage control circuits on one regrowth chip, in comparison with chips directly received from MOSIS.

6.4 Summary

In this chapter, the results of different optoelectronic circuits were presented. They were fabricated by three different approaches: the in-house process, the FET-SEED process, and the MBE regrowth on MOSIS GaAs circuits. We mainly focused on the device design, circuit response, optoelectronic gain, and the power consumption. Simple threshold circuits could be fabricated through our own processing facilities, however, the uniformity was very poor. The uniformity can be improved by using industrial processes. The FET-SEED process can be used to fabricate simple circuit arrays. By using modulators as the optical output devices, the power consumption of each neuron can be reduced. However, for complex circuits, the absence of enhancement-mode FET makes the design more complicated and increases the power consumption. The MOSIS GaAs run provides flexible MESFET circuits with both depletion- and enhancement-mode FET's. The power consumption can be greatly reduced for the same electrical response. However, the low breakdown voltage makes it difficult to design the driving circuits for modulators and will result in the increase

of power consumption. For the bump circuits, we found out that the power consumption of circuits from MOSIS was about four times less than that of circuits from FET-SEED process. MBE regrowth of GaAs/AlGaAs MQW modulators on MOSIS chips with 2.2:1 contrast ratio at 7.5V bias was demonstrated.

Chapter 7 Conclusion

In summary, different versions of optoelectronic threshold circuits and bump circuits were fabricated using three different approaches: the in-house fabrication, the AT&T FET-SEED process, and the MBE regrowth on MOSIS/Vitesse GaAs MESFET circuits. The goal here is to build high density optoelectronic neuron arrays for optical neural network applications, which has the benefit of parallel processing and can greatly increase the computation power of the system.

To increase the neuron density of an OEIC array, the power consumption of each neuron must be minimized because the power dissipation limit of GaAs wafers is the bottleneck. Several factors will affect the neuron power consumption, including the choice of photodetectors, the choice of output devices, the circuit design, and the uniformity of the fabrication process. To reduce the power consumption, a cascade network essentially requires high gain detectors and high efficiency transmitters in order to compensate the high loss introduced by the optical interconnection media.

From our research, it is found that the GaAs field-effect transistors are the ideal choice as a high gain photodetector when the speed consideration is not very important, as in the case of an optical neural network. There are different structures and different configuration for the GaAs optical FET detector, but the basic mechanism is more or less similar. By charging up the gate and/or backgate voltage of a GaAs FET using the photo-generated carriers, detector gains as high as 10^4 can be achieved. As the result of such high gain photodetector, the requirement of the output power from neurons on previous layers in a cascade network is greatly reduced, thereby reducing the power consumption of each neuron.

As for the choice of output transmitters, multiple quantum well modulators are the first choice due to their low switching energy requirement. When the neurons are not operating at high speed, the power consumption from the modulator output is very low even with a large bias voltage. The disadvantage is their low contrast ratio

and complex optical setup. For some fabrication process, it would be desirable if the requirement of bias voltage can be reduced. On the other hand, light-emitting diodes are another choice for light sources if we can increase their efficiency at low current levels, which is mainly restricted by the surface non-radiative recombination and the low photon-collecting efficiency. Several schemes can be implemented to reduce the severity of these problems, such as dry etching for defining the LED mesa, double Zn diffusion for current confinement, and micro-cavity quantum well LED.

With the theoretical discussion, it is important to establish a fabrication scheme to integrate the photodetectors, the transmitters, and the circuit into a single IC. Monolithic integration based on GaAs wafers is our focus because it provides overall advantages over hybrid methods. Three different fabrication approaches have been investigated by our group. Table 7.1 lists their characteristics, including main advantages and disadvantages.

The fabrication process developed using our own facilities usually ends up with poor uniformity, especially when chemical wet etching is applied for the gate recess etching. Even though it is possible to fabricate some simple circuits with good performance, the fabrication of OEIC neuron arrays using this process is undesirable because the operation point of each individual neurons will probably be at different levels. For complex circuits, this process just does not work because the result will be unpredictable and unrepeatable due to the high non-uniformity. Therefore, even though it has the advantage of fast turn-around time and very good flexibility, it is not an applicable method for fabricating high density optoelectronic neuron arrays. For that purpose, industrial foundries that provide uniform electronic circuits are needed.

For the two industrial foundries, each one has its own advantages and disadvantages. The FET-SEED process is established and there is no requirement of any post-processing. All we need is to design the optoelectronic circuit, which has become simpler since the introduction of a simulation tool developed for the FET-SEED process. On the other hand, since this not yet a commercial process, the support provided by the foundry is limited and it takes much longer time for the foundry to

process	in-house	FET-SEED	regrowth
turn-around time	fast	slow	medium + post-processing
high gain detector	OPFET with responsivities up to 10,000A/W 0.01 to 0.1msec response time		
fast detector	photodiode & MSM	p-i-n photodiode	MSM detector
output device	LED on-chip	modulator/LED on-chip	LEDor modulator MBE regrowth
electronic devices	EFET & DFET	DFET only	EFET & DFET
minimum gate length	> 4um	1um	0.8um
interconnection metal layers	one	two	four
uniformity	poor	good	good (unknown for regrowth)
circuit design	simple circuit only	complicated	easy
flexibility	good	poor	good
power consumption	low for individual high for array	high	low
main advantage	flexibility fast turn-around time	no post-processing	flexibility low power consumption
main disadvantage	non-uniformity	no EFET	low breakdown voltage reliability (?)

Table 7.1: The comparison between three different fabrication approaches.

process a fabrication run compared to the MOSIS service. The main problem of the FET-SEED process is the limitation of its electronic circuits. Since only depletion-mode FET's are available, the capability of its circuit is greatly limited. For simple circuits, such as the threshold circuit with modulator output shown in Figure 6.4, it is possible to use the FET-SEED process to achieve OEIC neuron arrays with high density and good uniformity. On the other hand, for complex circuits, the circuit design is usually much more complicated and it typically results in a circuit requiring many bias voltages and consuming a large amount of power, such as the bump circuit we tried to fabricate, shown in Figure 6.6. As a result, the maximum neuron density on an array will be reduced.

In conclusion, if we want to develop the FET-SEED process into an applicable integration scheme for high density optoelectronic neuron arrays, this process must provide more flexible electronic components, at least with both enhancement-mode and depletion-mode FET. With the feedback to the foundry from the FET-SEED workshop attendants, AT&T Bell Laboratories are providing a similar service now, using the SEED device to provide the optical output. It is the SEED on silicon CMOS circuits, a flip-chip bonding hybrid approach. Such service is provided by similar workshop organized by CO-OP [61]. It provides the flexibility of circuit design, power consumption, turn-around time, etc., by using the well-established CMOS technology. The CMOS circuits are fabricated at a commercial foundry and Bell Lab provides arrays of SEED devices with preset sizes and preset spacing. These two chips are then flip-chip bonded to provide both the electronic and optical devices. The question for this improvement is the reliability of the flip-chip bonding as mentioned in Chapter 1. In addition, such techniques must be extended to wafer-scale integration in order to reduce the fabrication cost.

In contrast to the FET-SEED process, the MOSIS GaAs circuits fabricated by Vitesse Semiconductor Corporation can accommodate flexible circuit design by providing both depletion-mode and enhancement-mode GaAs MESFET's. This technology is a well-established commercial fabrication process. Uniform arrays of electronic circuits can be fabricated through the MOSIS service in a reasonable time at a rea-

sonable cost. By using the MBE regrowth, it is possible to provide the optical output devices by growing light-emitting diodes, multiple quantum well modulators, or even laser diodes, such as VCSEL's. Since this is a monolithic integration scheme, we can easily extend it into a wafer-scale process with the help from the foundry, which can greatly reduce the cost of optoelectronic circuits in general, not just for the analog optical neuron arrays.

However, the MBE regrowth method is a newly-developed method and is still in the developing stage. Many questions remain to be answered. For example, it requires a strong liaison with the foundry. For the N4CJ run, when the foundry adjusted their process, we ran into many problems with cleaning the dielectric layers. At this moment, chemical wet etching, which is the only choice for us, is applied to etch the poly-crystalline regrowth material. This is definitely something that can be improved by using dry etching, which has better selectivity and less undesirable undercut on the sidewall. If we change the post-processing from the procedures we use right now to an industrial process, can the circuit withstand the thermal cycle? The list of these questions goes on and on. All those questions must be answered before the MBE regrowth method can be applied in industrial production.

Even though the MBE regrowth on MOSIS GaAs MESFET circuits is still in its beginning stage, our preliminary result is very promising as presented in this thesis and other publications [64, 65, 67]. Similar techniques are being developed on silicon wafers as well at this moment [112]. The author believes that the successful development of such a technique can greatly push forward the fabrication of optoelectronic circuits. Furthermore, by providing optical input/output devices on conventional integrated circuits, we can improve the processing power of such IC's in general.

Bibliography

[1] D. Zhou, F. P. Preparata, and S. M. Kang, "Interconnection delay in very high-speed VLSI," *IEEE Transactions on Circuits and Systems,* Vol. 38, pp. 779, 1991.

[2] K. C. Saraswat and F. Mohammadi, "Effect of scaling of interconnections on the time delay of VLSI circuits," *IEEE Transactions on Electron Devices,* Vol. ED-29, pp. 645, 1982.

[3] G. E. Blonder, R. A. Gottscho, K. L. Tai, "Interconnection processes and materials," *AT&T Technical Journal,* vol. 69, No. 6, pp. 46, 1990.

[4] J. W. Goodman, F. I. Leonberger, S. K. Kung, R. A. Athale, "Optical interconnections for VLSI systems," *Proceedings of the IEEE,* Vol. 72, pp. 850, 1984.

[5] M. R. Feldman, C. C. Guest, T. J. Drabik, and S. C. Esener, "Comparison between electrical and free space optical interconnects for fine-grain processor arrays based on interconnect density capabilities," *Applied Optics,* vol. 28, pp. 3820, 1989.

[6] F. E. Kiamilev, P. Marchand, A. V. Krishnamoorthy, S. C. Esener, and S. H. Lee, "Performance comparison between optoelectronic and VLSI multistage interconnection networks," *Journal of Lightwave Technology,* Vol. 9, pp. 1674, 1991.

[7] R. A. Athale, "Device-architecture interaction in optical computing," in *Spatial Light Modulators and Applications,* OSA Technical Digest Series, vol. 9, pp. 75, Optical Society of America, Washington DC, 1995.

[8] F. B. McCormick, A. L. Lentine, R. L. Morrison, J. M. Sasian, T. J. Cloonan, R. A. Novotny, et al., "155-Mb/s operation of a FET-SEED free-space switching network," *IEEE Photonics Technology Letters,* Vol. 6, pp. 1479, 1994.

[9] J. J. Hopfield, "Neurons with graded response have collective computation properties like those of two-state neurons," *Proceedings of the National Academy of Sciences of the United States of America Biological Sciences,* vol. 81, 3088, 1984.

[10] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the theory of neural computation,* Chapter 1, Redwood City, CA: Addison-Wesley Publishing Company, 1991.

[11] Y. Arima, M. Murasaki, T. Yamada, A. Maeda, and H. Shinohara, "A refreshable analog VLSI neural network chip with 400 neurons and 40k synapses," *IEEE Journal of Solid State Circuits,* Vol. 27, pp. 1854, 1992.

[12] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing,* D. E. Rumelhart and J. L. McClelland eds., vol. 1, chapter 8. MIT, Cambridge, Mass., 1986.

[13] D. Psaltis and N. H. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," *Optics Letters,* vol. 10, pp. 98, 1985.

[14] D. Z. Anderson and D. M. Linninger, "Dynamic optical interconnects: volume holograms as optical two-port operators," *Applied Optics,* vol. 26, pp. 5031, 1987.

[15] J. Moody and C. Darken, "Fast learning in networks of locally tuned processing units," *Neural Computation,* vol. 1, pp. 281, 1989.

[16] F. H. Mok, M. C. Tackitt, and H. M. Stoll, "Storage of 500 high-resolution holograms in a $LiNBO_3$ crystal," *Optics Letters,* Vol. 16, pp. 605, 1991.

[17] E. G. Paek, and D. Psaltis, "Optical associative memory using Fourier transform holograms," *Optical Engineering,* Vol. 26, pp. 428, 1987.

[18] D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," *Applied Optics,* Vol. 27, pp. 1752, 1988.

[19] H. Y. S. Li, Y. Qiao, and D. Psaltis, "Optical network for real-time face recognition," *Applied Optics,* Vol. 32, pp. 5026, 1993.

[20] M. J. Goodwin, A. J. Moseley, M. Q. Kearley, R. C. Goodfellow, and I. Bennion, "Optoelectronic component arrays for optical interconnection of circuits and subsystems," *Journal of Lightwave Technology,* vol. 9, pp. 1639, 1991.

[21] W. R. Imler, K. D. Scholz, and M. Cobarruviaz, "Precision flip-chip solder bump interconnects for optical packaging," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology,* vol. 15, pp. 977, 1992.

[22] K. W. Goossen, J. A. Walker, L. A. D'Asaro, S. P. Hui, B. Tseng, et al., "GaAs MQW modulators integrated with silicon CMOS," *IEEE Photonic Technology Letters,* Vol. 7, pp. 360, 1995.

[23] K. M. Johnson and G. Moddle, "Motivations for using ferroelectric liquid crystal spatial light modulators in neurocomputing," *Applied Optics,* vol. 28, pp. 4888, 1989.

[24] D. J. McKnight, K. M. Johnson, and R. A. Serati, "Electrically addressed 256 by 256 liquid-crystal-on-silicon spatial light modulator," *Optics Letters,* vol. 18, pp. 2159, 1993.

[25] C. Camperiginestet, Y. W. Kim, N. M. Jokerst, M. G. Allen, and M. A. Brooke, "Vertical electrical interconnection of compound semiconductor thin-film devices to underlying silicon circuitry," *IEEE Photonics Technology Letters,* vol. 4, pp. 1003, 1992.

[26] P. Demeester, I. Pollentier, P. Dedobbelaere, C. Brys, and P. Vandaele, "Epitaxial lift-off and its applications," *Semiconductor Science and Technology,* vol. 8, pp. 1124, 1993.

[27] H. H. Kim, R. G. Swartz, Y. Ota, T. K. Woodward, M. D. Feuer, and W. L. Wilson, "Prospects for silicon monolithic opto-electronics with polymer light-emitting diodes," *Journal of Lightwave Technology,* vol. 12, pp. 2114, 1994.

[28] P. Steiner, F. Kozlowski, M. Wielunski, and W. Lang, "Enhanced blue-light emission from an Indium-treated porous silicon device," *Japanese Journal of Applied Physics, Part 1,* vol. 33, pp. 6075, 1994.

[29] W. B. Pearson, *Handbook of Lattice Spacings and Structure of Metals and Alloys,* Pergamon, New York, 1967.

[30] G. Borghs, J. Deboeck, I. Pollentier, P. Demeester, C. Brys, and W. Dobbelaere, "III-V on dissimilar substrates - epitaxy and alternatives," *Institute of Physics Conference Series,* vol. 136, pp. 441, 1994.

[31] M. Razeghi, M. Defour, R. Blondeau, F. Omnes, P. Maurel, et al., "First cw operation of a $Ga_{0.25}In_{0.75}As_{0.5}P_{0.5}$-InP laser on a silicon substrate," *Applied Physics,* vol. 53, pp. 2389, 1989.

[32] K. J. Lodge and D. J. Pedder, "The impact of packaging on the reliability of flip-chip bonded CES," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology,* vol. 13, pp. 847, 1990.

[33] N. Bar-Chaim, S. Margalit, A. Yariv, and I. Ury, "GaAs integrated optoelectronics," *IEEE Transactions on Electron Devices,* Vol. ED-29, pp. 1372, 1982.

[34] D. D. Gupta, "A novel active area bumped flip-chip technology for convergent heat-transfer from Gallium Arsenide power devices," *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part A,* vol. 18, pp. 82, 1995.

[35] Y. S. Liu and D. A. Smith, "The frequency response of an amplitude-modulated GaAs luminescence diode," *Proceedings of the IEEE,* vol. 63, pp. 542, 1975.

[36] S. M. Sze, *Physics of Semiconductor Devices,* John Wiley & Sons, New York, 1981.

[37] M. Shur, "Modeling of GaAs and AlGaAs/GaAs Field Effect Transistors," in *Introduction to Semiconductor Technology: GaAs and Related Compounds* (Edited by C. T. Wang), chapter 2, John Wiley & Sons, New York, 1990.

[38] C. R. Crowell, J. C. Sarace, and S. M. Sze, "Tungsten-semiconductor Schottky-barrier diodes," *Transactions of the Metallurgical Society of AIME,* vol. 233, pp. 478, 1965.

[39] L. K. Cotter, T. J. Drabik, R. J. Dillon, and M. A. Handschy, "Ferroelectric liquid crystal silicon integrated circuit spatial light modulator," *Optics Letters,* vol. 15, pp. 291, 1990.

[40] D. A. Miller, "Quantum wells for optical information processing," *Optical Engineering,* vol. 26, pp. 368, 1987.

[41] I. Schnitzer, E. Yabonovitch, C. Caneau, and T. J. Gmitter, "Ultrahigh spontaneous emission quantum efficiency, 99.7-percent internally and 72-percent externally, from AlGaAs/GaAs/AlGaAs double heterostructures," *Applied Physics Letters,* vol. 62, pp. 131, 1993.

[42] K. Iga, F. Koyama, and S. Kinoshita, "Surface emitting semiconductor lasers," *IEEE Journal of Quantum Electronics,* vol. 24, pp. 1845, 1988.

[43] T. Numai, H. Kosaka, I. Ogura, and K. Kurihara, "Indistinct threshold laser operation in a pnpn vertical to surface transmission electrophotonic device with a vertical cavity, " *IEEE Journal of Quantum Electronics,* vol. 29, pp. 403, 1993.

[44] Y. Hayashi, T. Mukaihara, N. Hatori, N. Ohnoki, et al., "Record low-threshold index-guided InGaAs/GaAlAs vertical-cavity surface-emitting laser with a native oxide confinement structure," *Electronic Letters,* vol. 31, pp. 560, 1995.

[45] information provided by Dr. A. Scherer, mailing address Caltech 200-36, Pasadena, CA 91125.

[46] D. Psaltis, D. Brady, X. G. Gu, and S. Lin, "Holography in artificial neural networks," *Nature,* vol. 343, pp. 325, 1990.

[47] G. D. Boyd, D. A. B. Miller, D. S. Chemla, S. L. McCall, et al., "Multiple quantum well reflection modulator," *Applied Physics Letters,* vol. 50, pp. 1119, 1987.

[48] R. J. Grindle, J. E. Midwinter, and J. S. Roberts, "High contrast, low-voltage, symmetrical-self-electro-optic effect device (S-SEED)," *Electronics Letters,* vol. 27, pp. 2327. 1991.

[49] S. Lin, A. Grot, J. Luo, and D. Psaltis, "GaAs optoelectronic neuron arrays," *Applied Optics,* vol. 32, pp. 1275, 1993.

[50] H. Statz, P. Newman, I. W. Smith, R. A. Pucel, and H. A. Hans, "GaAs-FET device and circuit simulation in SPICE," *IEEE Transactions on Electron Devices,* vol. ED-34, pp. 160, 1987.

[51] C. Juang, K. J. Kuhn, and R. B. Darling, "Selective etching of GaAs and $Al_{.3}Ga_{.7}As$ with citric acid/hydrogen peroxide solution," *Journal of Vacuum Science and Technology B,* vol. 8, pp. 1122, 1990.

[52] Y. Matsumoto, "Study on Zn diffusion in GaAs and $Al_xGa_{1-x}As$ ($x \leq 0.4$) at temperatures from 726°C to 566°C," *Japanese Journal of Applied Physics,* vol. 22, pp. 829, 1983.

[53] C. P. Lee, S. Margalit, and A. Yariv, "Dependence of Zn diffusion on the Al content in $Ga_{1-x}Al_xAs$," *Solid State Electronics,* vol. 21, pp. 905, 1978.

[54] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, et al., "Novel hybrid optically bistable switch: the quantum well self-electro-optic effect device," *Applied Physics Letters,* vol. 45, pp. 13, 1984.

[55] T. W. Wood, C. A. Burrus, D. A. B. Miller, D. S. Chemla, et al., "High-speed optical modulation with GaAs/GaAlAs quantum wells in a *p-i-n* diode structure," *Applied Physics Letters,* vol. 44, pp. 16, 1984.

[56] D. A. B. Miller, M. D. Feuer, T. Y. Chang, S. C. Shunk, et al., "Field-effect transistor self-electrooptic effect device: integrated photodiode, quantum well modulator, and transistor," *IEEE Photonics Technology Letters,* vol. 1, pp. 62, 1989.

[57] T. K. Woodward, L. M. F. Chirovsky, A. L. Lentine, L. A. D'Asaro, et al., "Operation of a fully integrated GaAs-$Al_xGa_{1-x}As$ FET-SEED: a basic optically addressed integrated circuit," *IEEE Photonics Technology Letters,* vol. 4, pp. 614, 1992.

[58] L. A. D'Asaro, L. M. F. Chirovsky, E. J. Laskowski, S. S. Pei, et al., "Batch fabrication and structure of integrated GaAs-$Al_xGa_{1-x}As$ field-effect transistor self-electro-optic devices (FET-SEEDs)," *IEEE Electron Device Letters,* vol. 13, pp. 528, 1992.

[59] L. A. D'Asaro, L. M. F. Chirovsky, E. J. Laskowski, S. S. Pei, et al., "Batch fabrication and operation of GaAs-$Al_xGa_{1-x}As$ field-effect transistor-self-electro-optic device (FET-SEED) smart pixel arrays," *IEEE Journal of Quantum Electronics,* vol. 29, pp. 670, 1993.

[60] D. A. B. Miller, "Noval analog self-electrooptic effect devices," *IEEE Journal of Quantum Electronics,* vol. 29, pp. 678, 1993.

[61] At this moment, Dr. Ravi Athale at ECE Department, George Mason University, Fairfax, VA 22030 is in charge of the operation of CO-OP and related workshops.

[62] L. M. F. Chirovsky, "FET-SEED Fabrication," ARPA CO-OP FET-SEED Workshop Notes, June 21-24, 1993.

[63] Contact mosis@mosis.edu or Sam Reynolds at USC/Information Sciences Institute, 4676 Admiralty Way, Marina Del Rey, CA 90292 for more information.

[64] A. C. Grot, D. Psaltis, K. V. Shenoy, and C. G. Fonstad, Jr., "Integration of LEDs on GaAs MOSIS circuits by MBE regrowth," *IEEE Photonics Technology Letters,* vol. 6, pp. 819, 1994.

[65] A. C. Grot, *Analog GaAs optoelectronic integrated circuits for large scale arrays,* Ph.D. thesis, Caltech, 1994.

[66] Vitesse Semiconductor Corporation, *Foundry Design Manual, Version 6.0,* 1993. This manual is available from MOSIS.

[67] K. V. Shenoy, *Monolithic optoelectronic VLSI circuit Design and fabrication for optical interconnects,* Ph.D. thesis, MIT, 1995.

[68] K. V. Shenoy, C. G. Fonstad, and J. M. Mikkelson, "High temperature stability of refractory-metal GaAs VLSI MESFETs," *IEEE Electron Device Letters,* vol. 15, pp. 106, 1994.

[69] K. V. Shenoy, formerly of Fonstad's group at MIT, current address is Caltech 216-76, Pasadena, CA 91125, kvshenoy@vis.caltech.edu, private communication 1994-1995.

[70] S. Miyazawa and Y. Sekiguchi, "Low-temperature molecular beam epitaxy growth of single quantum well GaAs/AlGaAs lasers," *Japanese Journal of Applied Physics,* vol. 30, pp. L921, 1991.

[71] H. C. Casey, Jr. and M. B. Panish, *Heterostructure Lasers,* Academic, New York, 1978.

[72] T. P. Lee and A. G. Dentai, "Power and modulation bandwidth of GaAs-AlGaAs high-radiance LED's for optical communication systems," *IEEE Journal of Quantum Electrons,* vol. QE-14, pp. 150, 1978.

[73] H. Namizaki, H. Khan, M. Ishii, and A. Ito, "Current dependence of spontaneous carrier lifetime in GaAs-$Ga_{1-x}Al_xAs$ double-heterostructure lasers," *Applied Physics Letters,* vol. 24, pp. 486, 1974.

[74] P. T. Landsberg and D. J. Robbins, "The first 70 semiconductor Auger processes," *Solid-State Electronics,* vol. 21, pp. 1289, 1978.

[75] H. Kressel, M. Ettenberg, J. P. Wittke, and I. Kadany, *Semiconductor Devices for Optical Communications,* Topics in Applied Physics, vol. 39, Spring-Verlad, 1980.

[76] E. Hecht and A. Zajac, "Optics," Addison & Wesley, 1979.

[77] F. W. Ostermayer, Jr., P. A. Kohl, and R. H. Burton, "Photoelectrochemical etching of integral lenses on InGaAsP/Inp light-emitting diodes," *Applied Physics Letters,* vol. 43, pp. 643, 1983.

[78] I. Schnitzer, E. Yablonovitch, C. Caneau, T. J. Gmitter, and A. Scherer, "30% external quantum efficiency from surface textured, thin-film light-emitting diodes," *Applied Physics Letters,* vol. 63, pp. 2174, 1993.

[79] O. Leistiko, Jr. and C. A. Bittmann, "Surface effects of $GaAs_{0.6}P_{0.4}$ light emitting diodes," *Solid-State Electronics,* vol. 16, pp. 1321, 1973.

[80] J. M. Ralston, "Detailed light-current-voltage analysis of GaP electroluminescent diodes," *Journal of Applied Physics,* vol. 44, pp. 2635, 1973.

[81] C. H. Henry, R. A. Logan, and F. R. Merritt, "The effect of surface recombination on current in $Al_xGa_{1-x}As$ heterojunctions," *Journal of Applied Physics,* vol. 49, pp. 3530, 1978.

[82] B. Corbett and W. M. Kelly, "Surface recombination in dry etched AlGaAs/GaAs double heterostructure p-i-n mesa diodes," *Applied Physics Letters,* vol. 62, pp. 87, 1993.

[83] W. S. Hobson, U. Mohideen, S. J. Pearton, R. E. Slusher, and F. Ren, "SiN_x/sulphide passivated GaAs/AlGaAs microdisk lasers," *Electronics Letters,* vol. 29, pp. 2199, 1993.

[84] J. F. Gibbons, W. S. Johnson, S. W. Mylroie, *Projected Range Statistics, Semiconductor and Related Materials,* 2nd Edition. Stroudsburg, PA: Dowden, Hutchinson, & Ross, 1975.

[85] Dr. Axer Scherer, former member of Bellcore, current address is Caltech 200-36, Pasadena, CA 91125, email address etcher@styx.caltech.edu, private communication.

[86] J. L. Jewell, J. P. Harbison, A. Scherer, Y. H. Lee, and L. T. Florez, "Vertical surface emitting lasers: design, growth, fabrication, and characterization," *IEEE Journal of Quantum Electronics,* vol. 27, pp. 1332, 1991.

[87] F. De Martini, G. Innocenty, G. R. Jacobovitz, and P. Mataloni, "Anomalous spontaneous emission time in a microscopic optical cavity," *Physical Review Letters,* vol. 59, pp. 2955, 1987.

[88] C. Baack, G. Elze, and G. Walf, "GaAs M.E.S.F.E.T.: a high-speed optical detector," *Electronics Letters,* vol. 13, pp. 193, 1977.

[89] J. C. Gammel and J. M. Ballantyne, "The OPFET: a new high speed optical detector," *Proceeding of the International Electron Device Meeting,* pp. 120, IEEE, New York, 1978.

[90] T. Sugeta and Y. Mizushima, "High speed photoresponse mechanism of a GaAs-MESFET," *Japanese Journal of Applied Physics,* vol. 19, pp. L27, 1980.

[91] H. Beneking, "On the response behavior of fast photoconductive optical planar and coaxial semiconductor detectors," *IEEE Transactions on Electron Devices,* vol. ED-29, pp. 1431, 1982.

[92] B. B. Pal and S. N. Chattopadhyay, "GaAs OPFET characteristics considering the effect of gate depletion with modulation due to incident radiation," *IEEE Transactions on Electronic Devices,* vol 39, pp. 1021-1027, 1992.

[93] S. Lo and C. Lee, "Numerical analysis of the photoeffects in GaAs MESFET's," *IEEE Trans. Elect. Devices.,* vol 39, pp. 1564-1570, 1992.

[94] P. Chakrabarti, N. L. Shrestha, S. Srivastava, and V. Khemka, "An improved model of ion-implanted GaAs OPFET," *IEEE Trans. Elect. Devices.,* vol 39, pp. 2050-2059, 1992

[95] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, et al., "Band-edge electroabsorption in quantum well structures: the quantum confined Stark effect," *Physical Review Letters,* vol. 53, pp. 2173, 1984.

[96] R. J. Simes, R. H. Yan, R. S. Geels, L. A. Coldreen, et al., "Electrically tunable Fabry-Perot mirror using multiple quantum well index modulation," *Applied Physics Letters,* vol. 53, pp. 637, 1988.

[97] K. Tai, J. Hegarty, and W. T. Tsang, "Observation of optical Stark effect in InGaAs/InP multiple quantum wells," *Applied Physics Letters,* vol. 51, pp. 152, 1987.

[98] A. Tomita, Y. Kohga, A. Suzuki, T. Terakado, and A. Ajisawa, "5:1 ON-OFF contrast InGaAs/InP multiple quantum well Fabry-Perot etalon modulator," *Applied Physics Letters,* vol. 55, pp. 1817, 1989.

[99] K. Wakita, I. Kotaka, and H. Asai, "High-speed InGaAlAs/InAlAs multiple quantum well optical modulators with bandwidths in excess of 20 GHz," *IEEE Photonics Technology Letters,* vol. 4, pp. 29, 1992.

[100] G. Bastard, "Superlattice band structure in the envelop-function approximation," *Physical Review B,* vol. 24, pp. 5693, 1981.

[101] A. Yariv, *Quantum Electronics,* 3rd edition, John Wiley & Sons, New York, 1989.

[102] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, et al., "Electric field dependence of optical absorption near the band gap of quantum-well structures," *Physical Review B,* vol. 32, pp. 1043, 1985.

[103] C. Kittel, *Introduction to Solid State Physics,* 6th edition, John Wiley & Sons, New York, 1986.

[104] D. A. B. Miller, J. S. Weiner, and D. S. Chemla, "Electric-field dependence of linear optical properties in quantum well structures: waveguide electroabsorption and sum rules," *IEEE Journal of Quantum Electronics,* vol. QE-22, pp. 1816, 1986.

[105] G. Lengyel, K. W. Jelley, and R. W. H. Engelmann, "A semi-empirical model for electroabsorption in GaAs/AlGaAs multiple quantum well modulator structure," *IEEE Journal of Quantum Electronics,* vol. 26, pp. 296, 1990.

[106] D. A. B. Miller, "Introduction to quantum well electroabsorptive devices," ARPA CO-OP FET-SEED Workshop Notes, pp. 19, June 21-24, 1993.

[107] H. Q. Le, J. J. Zayhowski, and W. D. Goodhue, "Stark effect in $Al_xGa_{1-x}As$/GaAs coupled quantum wells," *Applied Physics Letters,* vol. 50, pp. 1518, 1987.

[108] Wang Hao, Fonstad's group at MIT, current address is 13-3070 MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, email whao@mit.edu, private communication, 1995.

[109] Yakov Royter, Fonstad's group at MIT, current address is 13-3078 MIT, 77 Massachusetts Avenue, Cambridge, MA 02139, email yakroyt@mit.edu, private communication, 1995.

[110] G. D. Sanders and K. K. Bajaj, "Absorptive electro-optic spatial light modulators: effect of well profile on device performance," *Applied Physics Letters,* vol. 55, pp. 930, 1989.

[111] A. Z. Shang, D. V. Plant, and H. S. Hinton, *Extraction and simulation of FET-SEED devices and circuits,* McGill University, 1994.

[112] M. Tachikawa, T. Yamada, T. Sasaki, H. Mori, and Y. Kadota, "Laser-diode-quality InP/Si grown by hybride vapor phase epitaxy," *Japanese Journal of Applied Physics,* vol. 34, pp. L657, 1995.