COMPUTER AIDED DESIGN OF MONOLITHIC MICROWAVE
AND MILLIMETER WAVE INTEGRATED CIRCUITS AND SUBSYSTEMS

Interim Technical Report
1 September 1986 - 31 August 1987

AFOSR Grant No. AFOSR-86-0339

Monitored by
Dr. Gerald Witt

Submitted by
Department of Electrical and Computer Engineering
Mail Code C-014
University of California, San Diego
La Jolla, CA 92093

Principal Investigator:
Walter H. Ku

DISTRIBUTION STATEMENT A
Approved for public release. Distribution unlimited.
This interim technical report presents results of research on the computer aided design of monolithic microwave and millimeter wave integrated circuits and subsystems. A specific objective is to extend the state-of-the-art of the computer aided design (CAD) of the monolithic microwave and millimeter wave integrated circuits (MIMIC). In this reporting period, we have derived a new model for the high electron mobility transistor (HEMT). Based on a nonlinear charge control formulation which takes into consideration the variation of the 2DEG distance offset from the heterointerface as a function of bias. Pseudomorphic InGaAs/GaAs HEMT devices have been successfully fabricated at UCSD. For a lum gate length, a maximum transconductance of 320 mS/mm was obtained. In cooperation with TRW, Devices with 0.1 um and 0.25 um gate lengths have been successfully fabricated and tested. New results on the design of ultra-wideband distributed amplifiers using 0.15 um pseudomorphic InGaAs/GaAs HEMT's have also been obtained. In addition, two-dimensional models of the submicron MESFET's, HEMT's and HBT's are currently being developed for the CRAY X-MP/48 supercomputer.
COMPUTER AIDED DESIGN OF MONOLITHIC MICROWAVE AND MILLIMETER WAVE INTEGRATED CIRCUITS AND SUBSYSTEMS

ABSTRACT

This interim technical report presents results of research on the computer aided design of monolithic microwave and millimeter wave integrated circuits and subsystems. A specific objective is to extend the state-of-the-art of the computer aided design (CAD) of the monolithic microwave and millimeter wave integrated circuits (MIMIC). In this reporting period, we have derived a new model for the high electron mobility transistor (HEMT) based on a nonlinear charge control formulation which takes into consideration the variation of the 2DEG distance offset from the heterointerface as a function of bias. Pseudomorphic InGaAs/GaAs HEMT devices have been successfully fabricated at UCSD. For a 1 µm gate length, a maximum transconductance of 320 mS/mm was obtained. In cooperation with TRW, devices with 0.15 µm and 0.25 µm gate lengths have been successfully fabricated and tested. New results on the design of ultra-wideband distributed amplifiers using 0.15 µm pseudomorphic InGaAs/GaAs HEMT's have also been obtained. In addition, two-dimensional models of the submicron MESFET's, HEMT's and HBT's are currently being developed for the CRAY X-MP/48 supercomputer. Preliminary results obtained are also presented in this report.
1. Fabrication of HEMT Devices

1.1 Introduction

High electron mobility transistors (HEMT) have shown great potential for high speed and high frequency applications. Much progress has been made in the AlGaAs/GaAs HEMTs where the larger band gap highly doped AlGaAs supplies electrons to the two dimensional electron gas (2DEG) in the smaller band gap undoped GaAs. Recently, there has been increasing interest in pseudomorphic InGaAs/GaAs HEMTs because of the higher mobility of InGaAs over GaAs and the elimination of the persistent photoconductivity problem associated with the AlGaAs.

Pseudomorphic InGaAs/GaAs HEMT devices with a gate length of 2 microns and transconductance of 120 mS/mm have been fabricated at UCSD. Using the fabrication facilities and process at TRW, a one micron gate length InGaAs/GaAs HEMT with a maximum transconductance of 320 mS/mm was successfully made. These results are comparable to those found in a 1 micron pseudomorphic InGaAs/AlGaAs HEMT. These HEMTs have exhibited a transconductance of 540 mS/mm in a 0.1 micron device.

Material

The material system for these HEMT devices is pseudomorphic In$_{15}$Ga$_{85}$As/GaAs on GaAs substrate, grown by molecular beam epitaxy at UCSD. The advantages of this material over the conventional AlGaAs/GaAs is the higher electron mobility and peak electron velocity in the InGaAs and the avoidance of problems associated with the deep donor - vacancy complexes known as DX centers in the AlGaAs. These cause the threshold voltage to be dependent on temperature and illumination and also lead to a persistent photoconductivity at low temperatures. The lattice mis-match between the GaAs and InGaAs would normally create dislocations at the interface but by making the InGaAs layer thin, the mis-match can be taken up by elastic strain.

Figures 1 and 2 show the material structures of the HEMTs reported here.

---

Device Fabrication

The HEMTs fabricated were split source devices with various gate lengths and widths. A new mask set was made for 2 micron discrete field effect transistors. It was determined that 2 micron geometries was the minimum dimension that could be reliably fabricated with the current HTG aligners at UCSD. The devices would serve as an initial step in the development toward smaller dimension devices and be useful as prototype devices for investigating various material systems. The total gate width was 300 microns and the source to drain spacing was 6 microns, allowing a 2 micron alignment tolerance in the mask set. The overall dimensions are shown in Figure 3a. In addition, a few 1 micron, 2.5 micron and 3 micron gate length devices with source to drain spacings of 5 micron, 7.5 micron and 9 microns, respectively, were included as test structures. The 1 micron gate length device mask set given us by TRW and shown in Figure 3b and 3c, has two split source devices with source to drain spacing of 3 microns and gate widths of 300 and 100 microns, respectively.

The HEMT fabrication process at UCSD consists of three sections, which are mesa isolation, ohmic contact and gate. The mesa isolation is accomplished by a wet etch consisting of phosphoric acid, hydrogen peroxide and water in a 1:1:25 mixture. For the ohmic contacts a AuGe alloy with a thin Ni cap layer is used. Furnace annealing is performed at 450°C. Much effort has been focused on the gate process, which is the most critical portion of the fabrication. The gate lithography involves fine geometries and alignment and the gate recess controls the threshold voltage. Thermally evaporated aluminum is the gate metal. An outline of the UCSD process is attached. The intention was to establish a consistent 2 micron gate process followed by work on a 1 micron gate process at UCSD. Toward this goal, there have been beneficial opportunities to visit and use TRW facilities in Redondo Beach, California.

A two micron photolithography process was developed for the new devices. An HTG aligner with mid - UV (300-320 nm) wavelength source was used. It became apparent that contact between the wafer and the mask is the most critical factor to achieving uniform, consistent and accurate resolution of the mask features. The best results come with use of vacuum contact, a chuck designed for the appropriate sample size, and a flat sample backside. As a result,
the backside of the sample needed to be cleaned and lapped to remove the In residue left after MBE growth. Difficulties arise with small piece samples because good contact is harder to achieve. Among the principle causes are mask bow and non-uniform resist thickness. Mask bow can be reduced with a mask holder and chuck designed for small pieces.

The alignment tools and microscope optics on the HTG are inadequate for alignment within 1 micron. Consistent resolution of 1 micron spaces has been difficult. The Karl Suss mask aligner is designed for use in the 1 micron and submicron range, and is more widely accepted and commonly used. The fabrication facility at TRW has Karl Suss aligners and these were used for the 1 micron InGaAs/GaAs HEMT devices fabricated there.

The gate recess etch is the most critical step in the processing of HEMT devices. It is necessary for the removal of the conduction path in the highly doped top layer and to adjust the threshold voltage. As a result, there was much development for this part of the process. A method was needed to determine the endpoint of the etch. This was accomplished by monitoring the saturation current between the source and drain by making contact to the ohmic pads through the resist. The sample would be etched for a certain time and measured until a pre-determined endpoint saturation current was reached. Two GaAs MESFETs and a InGaAs/GaAs HEMT with five micron gate lengths were fabricated as trial runs for using this technique. The results of these devices are seen in Figures 4, 5, and 6. The gate recess allows the pinchoff in the depletion mode MESFET and the fabrication of enhancement mode devices. The depth of the gate recess may only be a couple of hundred angstroms so a very slow etch is necessary.

As a result, the next issue addressed was to improve the wet chemical etch system. An etch of phosphoric acid, hydrogen peroxide and water in a 1:1:25 ratio had been used. However, the etch rate of 30–40Å/sec was too fast for a shallow gate recess and resulted in a lack of control and uniformity. The citric acid etch system was investigated by Otsubo and found to be an excellent slow and linear etch of GaAs. A mixture of 50% aqueous citric acid with hydrogen peroxide in a 40:1 ratio proved to have an constant etch rate of 10Å/sec. This new etch provided more stability and by increasing the etch time improved the control of the gate recess depth.

Visits to TRW this summer and the opportunity to work in a well equipped facility with an established process was very helpful. Fabrication of InGaAs/GaAs HEMT's at TRW serve as a benchmark for improvements of the process at UCSD and help qualify portions of the process as well as identify problem or critical areas. The 2 micron HEMT devices, #348 A and #462C shown in Figures 7 and 8 had the gate portion of the process done at TRW using the resist and resist process from UCSD. The results indicated that the mesa isolation and the ohmic contact process as well as the gate resist process are adequate to successfully fabricate 2 micron devices at UCSD. The complete fabrication process was then done at UCSD resulting in working devices on #348C and #462D. However, as seen in Figures 9 and 10, the performance of these devices were not as good as those finished at TRW, indicating some improvement is needed, particularly in the gate metallizations.

Sample #462E and #348D were fabricated with the TRW ohmic and gate process. The results from the 2 micron devices on #462E were compared to devices either partially or

---

completely fabricated at UCSD. The performance of these devices on #462E and the 1 micron devices on #348D serve as a target for future devices.

Results and Discussion

The best results for the pseudomorphic InGaAs/GaAs HEMTs are from the 1 micron devices on sample #348D. A short gate device (Figure 11) (gate width of 100 microns) exhibits a transconductance of 320 mS/mm, while the best value for the long gate device Figure 12 (gate width of 300 microns) is 225 mS/mm. The I-V curves shown in Figures 11 and 12 show both positive and negative gate voltages with the peak transconductance occurring with positive gate voltage. They also show good saturation and pinch-off characteristics as well as good current values. These results are comparable to the 1 micron InGaAs/AlGaAs pseudomorphic HEMT by Ketterson et al. which had a transconductance of 270 mS/mm. The potential of the device is shown here and further improvements to the process should increase its performance. High frequency measurements will be made on these devices.

For the 2 micron devices, the highest transconductance, 183 mS/mm for a 300 micron wide gate, was from sample #462E which was fabricated with the TRW process (Figure 13). As a comparison, #462C (Figure 8) and #348A, (Figure 7) which only had gate recess and metallization done at TRW, have transconductance of 135 mS/mm and 175 mS/mm. It was found that the ohmic contact resistance for #462E is half of that for #462C. Devices from the same material but completely fabricated at UCSD, #462D (Figure 10) and #348C (Figure 9), have transconductances of 80 mS/mm and 120 mS/mm. With some further development of the gate recess and metallization the performance of the devices fabricated at UCSD will improve and a 2 micron process will be finalized. This gives us the capability of initial evaluation of different material systems and with a few improvements to the processing equipment, allow us to establish a 1 micron process at UCSD.

---

1. APPENDIX A

UCSD

Mesa Isolation

Lithography

Solvent clean
Prebake (105 C. 20 min)
Hunt HPR 1182 photoresist
Spin (5000 rpm. 20 seconds)
Softbake (105 C. 30 min)
Exposure HTG aligner (*40mJ/cm²*)
Develop (1 min, no agitation)
Hardbake (110 C. 30 min)
Etch (45 sec, H₃PO₄:H₂O₂:H₂O, ratio 1:1:25)
Resist strip in acetone

Ohmic Contact

Lithography

(same as above)

Metal Evaporation

AuGe:Ni (1000Å, thermal evaporation)
Liftoff, in acetone
Alloy

450° C. 30 sec in formine gas using furnace with graphite boat

Gate

Lithography

(Same as above)

Recess etch

Preclean (NH₄OH:H₂O, ratio 1:15)
Etch (citric acid : H₂O₂:H₂O, ratio 20:1:20)
Measure source to drain current

Metal evaporation

Preclean (NH₄OH:H₂O, ratio 1:15)
Al (1000Å, thermal evaporation)
Liftoff in acetone
2. APPENDIX B

Figures

a) Schematic of the UCSD 2 μm gate FET

b), c) Schematic of the two devices on the 1 μm gate FET mask set.

Figure 3
Enhancement mode GaAs MESFET A
Figure 4

Depletion mode GaAs MESFET B
Figure 5

5 um gate enhancement mode InGaAs/GaAs HEMT
Figure 6
UCSD 348A InGaAs/GaAs 2μm HEMT with $g_m = 175 \text{mS/mm}$

Figure 7

UCSD 462C InGaAs/GaAs 2μm HEMT with $g_m = 135 \text{mS/mm}$

Figure 8
UCSD 348C InGaAs/GaAs HEMT with $g_m = 120\text{mS/mm}$
Figure 9

UCSD 462D InGaAs/GaAs HEMT with $g_m = 80\text{mS/mm}$
Figure 10
UCSD 348D InGaAs/GaAs 1 µm HEMT (gate width = 100 µm)
\[ g_m = 320 \text{mS/mm} \]
Figure 11

UCSD 348D InGaAs/GaAs 1 µm HEMT (gate width = 300 µm)
\[ g_m = 225 \text{mS/mm} \]
Figure 12
UCSD 462E InGaAs/GaAs 2μm HEMT with $g_m = 183 \text{mS/mm}$

Figure 13
2. Design of Distributed Amplifiers Using 0.15 μm Pseudomorphic InGaAs/GaAs HEMT's

2.1 Introduction:

Distributed amplifiers using 0.15 μm pseudomorphic InGaAs/GaAs HEMT's for 60 GHz and possibly 94 GHz frequency bands are being designed at UCSD. The potential of distributed amplification for obtaining gain over wide frequency bands has long been recognized.\(^1,2\) Broadband distributed amplifiers using discrete GaAs MESFET's have been demonstrated at microwave frequencies.\(^3,4,5,6,7\) Better performance is expected with the use of HEMT due to its superior high frequency characteristics over MESFETs. The principle behind distributed amplifiers was based on the idea to separate the input capacitance of the active devices by means of artificial transmission lines, while adding their transconductances. As a result, it is possible to achieve extremely wide bandwidths up to very high frequencies with effectively very wide gates. The latter obviously has a beneficial impact on the amplifier's power handling capabilities.

For the distributed amplifier design, we will first develop the D.C. model and small signal equivalent circuit for HEMT. These will be initial steps toward developing a systematic approach to the circuit design of distributed amplifier. In the following sections, the small signal equivalent circuit will be discussed, and be applied to some distributed amplifier design.

3. Small Signal Equivalent Circuit

There are several sets of S-parameters measured on a 0.15μm pseudomorphic HEMT by TRW. One set labeled as HEMT C2 is listed in Table 1. Its frequency performance MAG/MSG, MUG, S\(_{21}\), K vs. frequency are plotted in Fig. 1, and \(f_{\text{max}}\) can be projected to higher than 80 GHz. The current gain \(h_{21}\) vs. frequency is plotted in Fig. 2 and \(f_T\) is projected to 70 GHz.

---

The small signal equivalent circuit of the active device is necessary to predict performance outside the measured frequency band. The circuit derived from the measured data includes many parameters, such as the parasitics associated with the fabrication process. Therefore, it is important to derive an exact equivalent circuit. The parasitic parameter resulting from measurement should be de-embedded out of the small signal equivalent circuit. This de-embedded model is used in the monolithic amplifier circuit design.

A small signal equivalent circuit proposed by Dr. Ahlgren is shown in Fig. 3(a) to fit the measured data listed in Table 1, and the deembedded model is shown in Fig. 3(b). An amplifier with lumped elements was designed and is shown in Fig. 3(c), and its performance is shown in Fig. 4.

Although the fit between the model and the measured data is very good, a more conventional model is derived based on the physical structure of the device and from the point of view of statistics. In addition, the parasitic resistance $R_g$, $R_e$, and $R_d$ can be obtained from dc measurements as detailed by Fukui. A resulting small signal equivalent circuit is shown in Fig. 5(a), and the de-embedded model is shown in Fig. 5(b).

4. Verification of the Potential of Distributed Amplifier

The performance of distributed amplifier is examined to determine whether its potential warrants the development of a systematic approach for design. The de-embedded model derived by Ahlgren is shown in Fig. 3(b), and its performance is shown in Fig. 6. Transmission line sections, whose characteristic impedances are all 50 ohms, were introduced between gate ends and between drain ends of this de-embedded model to form a distributed amplifier, which is shown in Fig. 7. Its performance within the frequency band 0-30 GHz is shown in Fig. 8. Comparing the curves in Fig. 6 and Fig. 8, it is clear that the gain flatness is improved, and the return losses of input and output are much smaller.

An attempt was made to improve the performance of distributed amplifier by using tapered transmission lines. These transmission line sections have different characteristic impedances from one another with linearly or exponentially tapered relations. The circuit topology is shown in Fig. 9, and its performance is shown in Fig. 10. It is clear that gain is $8.5 \pm 0.5$ dB, and return losses are below -10 dB.

---


Similarly, using the de-embedded model shown in Fig. 5(b), we designed another distributed amplifier with tapered transmission line, which is shown in Fig. 11, and its performance is shown in Fig. 12.

5. Systematic Design approach to be Developed

Knowing that the distributed amplifier has potential to increase the gain-bandwidth product, systematic study should commence to achieve a practical design guideline for distributed amplifiers. These guidelines will be established from examining the first order performance of distributed amplifier designs using the conventional simplified model rather than the de-embedded model. Then the small signal equivalent circuit and de-embedded model should be re-evaluated by comparison with the measured S-parameters and dc measurements. This de-embedded model would be used to optimize the circuit design when running the Touchstone program.

---


6. References:


### Table 1

<table>
<thead>
<tr>
<th>Freq</th>
<th>Gain</th>
<th>Volts</th>
<th>Amp</th>
<th>GS</th>
<th>Gain</th>
<th>Volts</th>
<th>Amp</th>
<th>GS</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.00</td>
<td>0.356</td>
<td>0.275</td>
<td>18.3</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>5.00</td>
<td>0.356</td>
<td>-43.9</td>
<td>-25.9</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>4.00</td>
<td>0.356</td>
<td>-42.4</td>
<td>-18.2</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>3.00</td>
<td>0.356</td>
<td>-40.8</td>
<td>-15.3</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>2.00</td>
<td>0.356</td>
<td>-44.7</td>
<td>-24.6</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>1.50</td>
<td>0.356</td>
<td>-47.4</td>
<td>-26.8</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>1.00</td>
<td>0.356</td>
<td>-49.6</td>
<td>-30.9</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
<tr>
<td>0.50</td>
<td>0.356</td>
<td>-51.4</td>
<td>-33.9</td>
<td>0.048</td>
<td>46.7</td>
<td>0.683</td>
<td>20.4</td>
<td>7.20</td>
</tr>
</tbody>
</table>

**Notes:**
- Gain values are listed in dB.
- Voltages and currents are in appropriate units (e.g., Volts and Amps).
- The table provides a comparison of gain, voltage, and current values across different frequencies.
<table>
<thead>
<tr>
<th>Freq (GHz)</th>
<th>K</th>
<th>SIGN</th>
<th>AVAILABLE</th>
<th>STABLE</th>
<th>UNILATERAL</th>
<th>H21</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.240</td>
<td>+</td>
<td>29.05</td>
<td>33.75</td>
<td>71.22</td>
<td></td>
</tr>
<tr>
<td>2.02</td>
<td>0.239</td>
<td>+</td>
<td>25.99</td>
<td>31.09</td>
<td>35.87</td>
<td></td>
</tr>
<tr>
<td>3.04</td>
<td>0.231</td>
<td>+</td>
<td>24.19</td>
<td>29.60</td>
<td>23.79</td>
<td></td>
</tr>
<tr>
<td>4.06</td>
<td>0.229</td>
<td>+</td>
<td>22.84</td>
<td>26.53</td>
<td>17.57</td>
<td></td>
</tr>
<tr>
<td>5.08</td>
<td>0.228</td>
<td>+</td>
<td>22.05</td>
<td>24.62</td>
<td>14.21</td>
<td></td>
</tr>
<tr>
<td>6.10</td>
<td>0.437</td>
<td>+</td>
<td>21.24</td>
<td>23.34</td>
<td>11.91</td>
<td></td>
</tr>
<tr>
<td>7.12</td>
<td>0.559</td>
<td>+</td>
<td>20.32</td>
<td>22.11</td>
<td>10.00</td>
<td></td>
</tr>
<tr>
<td>8.14</td>
<td>0.695</td>
<td>+</td>
<td>19.83</td>
<td>20.94</td>
<td>8.81</td>
<td></td>
</tr>
<tr>
<td>9.16</td>
<td>0.665</td>
<td>+</td>
<td>19.30</td>
<td>20.15</td>
<td>7.96</td>
<td></td>
</tr>
<tr>
<td>10.18</td>
<td>0.752</td>
<td>+</td>
<td>18.31</td>
<td>19.15</td>
<td>7.13</td>
<td></td>
</tr>
<tr>
<td>11.20</td>
<td>0.831</td>
<td>+</td>
<td>16.32</td>
<td>16.19</td>
<td>6.40</td>
<td></td>
</tr>
<tr>
<td>12.22</td>
<td>0.935</td>
<td>+</td>
<td>13.00</td>
<td>17.49</td>
<td>5.91</td>
<td></td>
</tr>
<tr>
<td>13.24</td>
<td>1.041</td>
<td>+</td>
<td>17.53</td>
<td>16.85</td>
<td>5.42</td>
<td></td>
</tr>
<tr>
<td>14.26</td>
<td>1.158</td>
<td>+</td>
<td>15.31</td>
<td>16.02</td>
<td>5.00</td>
<td></td>
</tr>
<tr>
<td>15.28</td>
<td>1.087</td>
<td>+</td>
<td>15.13</td>
<td>15.39</td>
<td>4.67</td>
<td></td>
</tr>
<tr>
<td>16.30</td>
<td>1.142</td>
<td>+</td>
<td>14.48</td>
<td>14.98</td>
<td>4.45</td>
<td></td>
</tr>
<tr>
<td>17.32</td>
<td>1.224</td>
<td>+</td>
<td>13.56</td>
<td>14.30</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>18.34</td>
<td>1.306</td>
<td>+</td>
<td>12.83</td>
<td>13.64</td>
<td>3.84</td>
<td></td>
</tr>
<tr>
<td>19.36</td>
<td>1.381</td>
<td>+</td>
<td>12.55</td>
<td>13.33</td>
<td>3.73</td>
<td></td>
</tr>
<tr>
<td>20.39</td>
<td>1.414</td>
<td>+</td>
<td>12.06</td>
<td>12.93</td>
<td>3.66</td>
<td></td>
</tr>
<tr>
<td>21.42</td>
<td>1.488</td>
<td>+</td>
<td>11.42</td>
<td>12.31</td>
<td>3.30</td>
<td></td>
</tr>
<tr>
<td>22.45</td>
<td>1.514</td>
<td>+</td>
<td>11.20</td>
<td>12.00</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>23.44</td>
<td>1.473</td>
<td>+</td>
<td>11.14</td>
<td>12.05</td>
<td>3.14</td>
<td></td>
</tr>
<tr>
<td>24.46</td>
<td>1.504</td>
<td>+</td>
<td>10.75</td>
<td>11.41</td>
<td>2.90</td>
<td></td>
</tr>
<tr>
<td>25.46</td>
<td>1.561</td>
<td>+</td>
<td>10.21</td>
<td>11.07</td>
<td>2.73</td>
<td></td>
</tr>
<tr>
<td>26.50</td>
<td>1.564</td>
<td>+</td>
<td>10.13</td>
<td>11.19</td>
<td>2.62</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 (cont'd)
Fig. 5
Z₁ = Z₂ = Z₃ = Z₄ = Z₅ = Z₆ = 50Ω, Z₇ = 30Ω

L₁ = 3.34mm  L₄ = 4.72mm  L₇ = 0.43mm
L₂ = 3.96mm  L₅ = 3.87mm  L₆ = 3.25mm

Fig. 7
\[ Z_1 = Z_A \]
\[ Z_2 = Z_A (1 + A) \]
\[ Z_3 = Z_A (1 + 2A) \]
\[ Z_A = 45.02 \]
\[ Z_B = 45.01 \]

\[ Z_4 = Z_B \left(1 + 2B\right) \]
\[ Z_5 = Z_B (1 + B) \]
\[ Z_6 = Z_B \]
\[ A = 0.00007 \]
\[ B = 0.49180 \]

Fig. 9
\[ R = 0.0001 \Omega \]

\[ Z_1, L_1 \]
\[ Z_2, L_2 \]
\[ Z_3, L_3 \]
\[ 0.0001 \Omega \]

\[ 32.6 \Omega, L_7 \]

\[ 71.67 \Omega \]

\[ 1.14 \text{mH} \]

\[ Z_4 = Z_B (1 + 2B) \]
\[ Z_5 = Z_B (1 + B) \]
\[ Z_6 = Z_B \]

\[ Z_A = 50.94 \]
\[ Z_B = 44.01 \]

\[ A = 0.00008 \]
\[ B = 0.75478 \]

Fig. 11
Fig. 13
3. An Analytical HEMT model based on nonlinear charge-control formulation

3.1 Introduction

The high electron mobility transistor (HEMT), a field-effect transistor that takes advantage of the novel properties of the two-dimensional electron gas at the heterointerface has shown promising performance in the of high-speed and high-frequency regimes since its demonstration in 1980. Many models for HEMT's based on a simple linear charge control model similar to that for conventional MOSFET's with only the gate insulator layer. The simulations of the I-V characteristics of these devices have shown fair agreement with experimental data within certain operation regions or in some respects [1]-[4]. Recently, Hughes et al., [5] has pointed out the importance of nonlinear effects on the charge control model and considered the nonlinear variation of \( \Delta d \) (the distance of 2DEG from heterointerface) utilizing a numerical method. These results led to a better fit of the calculated and measured transconductance characteristics. However, the physical mechanism cannot be easily recognized in the derivation of this model. Furthermore, the variation of charge density after current saturation was not considered.

An analytic I-V model of the HEMT, which can simulate the dc with the whole operation range of the dc characteristics through the whole range of operations is derived in this study. This model will be applied to conventional AlGaAs/GaAs HEMT's so that comparisons can be made to the models previously mentioned. In addition, the AlGaAs/GaAs HEMT material system has been much studied so information is available on many of the material parameters needed for the device analysis. This allows the models to focus on representing the basic characteristics and principles behind the operation of HEMT's. The models can later be readily adapted to HEMT's of new material systems, such as the pseudomorphic InGaAs/GaAs HEMT being developed at UCSD, as soon as material parameters become known.

3.2 Nonlinear Charge Control

The first proposed conventional charge control model for HEMT [1] derived from Poisson's equation neglected the Fermi-potential term.

\[
q_n = \frac{\varepsilon_1}{d_1 + d_e} \left[ V_G - V_T - V_F \right] \\
V_T = \phi_B - \frac{\Delta E_c}{q} - \frac{qN_{D1}}{2\varepsilon_1} \frac{d_2}{d_1}
\]

where \( n_s \) is the 2DEG sheet carrier concentration, \( \varepsilon_1 \) and \( d_1 \) are the permittivity and thickness of carrier supplying layer, respectively, \( q \) is the electronic charge, \( V_G \) is the gate-source biased voltage, \( \phi_B \) is the barrier height of the Schottky-gate, \( \Delta E_c \) is the conduction band discontinuity, \( N_{D1} \) is the doping concentration in the top layer, \( d_e \) is the spacer layer thickness, and \( V_F \) is the Fermi-potential with reference to the conduction band edge of the 2DEG channel.
Drummond et al. [2] extended this model to include the variation of Fermi potential of the two-dimensional electron gas by a linear approximation with respect to sheet carrier concentration.

\[ V_F = V_{F0} + a \ n_s \]  

(2)

which yielded

\[ q_n = \frac{\epsilon_1}{d_1+a}\frac{1}{d_1+d_e+\Delta d}\ \left[ V_G - V_{F0} - V_T \right] \]  

(3)

where

\[ \Delta d = \frac{\epsilon_1 a}{q} \]  

(4)

This model was further modified [5] to take into account the nonlinear dependence of the position of 2DEG channel \( \Delta d \) from heterointerface on gate bias. The 2DEG offset is calculated from the weighted average of all energy levels.

\[ \Delta d = \frac{\frac{2}{3} \sum n_{sn} d_n}{n} \]  

n = 0, 1, 2, 3, \cdots

(5)

where

\[ n_s = \sum_{n} n_{sn} \]  

n = 0, 1, 2, 3, \cdots

(6)

Clearly, in the above derivation the computations are very complicated.

According to the triangular potential well approximation the relation between \( n_s \) and the Fermi-level can be established, only if the lower and the excited subband are being considered.

\[ E_F = \frac{kT}{q} \ln \left\{ \frac{1}{2} \left[ e^{\frac{\epsilon_1}{kT}n_{sn}^\frac{1}{3}} + e^{\frac{\epsilon_1}{kT}n_{sn}^\frac{1}{3}} \right] \right\} + \sqrt{\frac{1}{4} \left( e^{\frac{\epsilon_1}{kT}n_{sn}^\frac{1}{3}} - e^{\frac{\epsilon_1}{kT}n_{sn}^\frac{1}{3}} \right)^2 + e^{\frac{\epsilon_1}{kT}n_{sn}^\frac{1}{3}} - e^{\frac{\epsilon_1}{kT}n_{sn}^\frac{1}{3}}} \]  

(7)

As suggested by eq. (7), we can approximate \( E_F \) by

\[ E_F = a(bn_s)^\frac{2}{3} \]  

(8)

with a high accuracy. Substituting eq. (7) into eq. (1), further manipulations yield the 2DEG sheet carrier concentration

\[ n_s = \frac{\epsilon_1}{(d_1+d_e)[1+\frac{P n_s}{q}]^\frac{1}{3}} \left[ V_G - \phi_B + \frac{\Delta E_c}{q} + \frac{qN_D}{2\epsilon_1 d_t^2} \right] \]  

(9)

where

\[ p = \frac{\epsilon_1 a}{d_1+d_e} b^\frac{2}{3} \]  

(10)
Eq. (9) indicates that $\Delta d$ is proportional to $n_1^{-\frac{1}{2}}$. This corresponds to the results calculated from a variational method by Stern and Das Sarma [6].

3.3 Drain Current Characteristics

When a doped layer with wider band-gap is grown on top of an undoped layer with narrower band-gap, a two-dimensional electron gas will form at the heterointerface due to the charge transfer driven by the conduction band discontinuity. In normal operation mode, the depletion of the doped layer is attributed to the charge transfer and Schottky gate depletion. At thermal equilibrium the undepleted channel width of parasitic MESFET, $h$, as shown in Fig. 1, is

$$h = d_1 - \frac{n_s}{N_{D1}} - \frac{2\epsilon_1}{qN_{D1}} (\phi_B - V_n - V_G)$$

(11)

where $V_n$ is the potential difference between the conduction band edge and the Fermi-energy in the neutral substrate.

The threshold voltage of the parasitic MESFET's conduction where $h = 0$ may be easily obtained, and can be written as

$$V_c = \phi_B - V_n - \frac{qN_{D1}}{2\epsilon_1} d_1 - \frac{n_{so}}{N_{D1}}$$

(12)

If the gate applied voltage $V_G \geq V_c$, then the doped layer is not fully depleted and the conduction of parasitic MESFET should be taken into account.

Experimental measurement has shown the velocity saturation phenomenon existing in the high electric field regime for III-V compound material. For simplicity in mathematics manipulation, the field-dependent velocity model is approximated by the following model for both the parasitic MESFET Channel and the two-dimensional electron gas channel.

$$v(E) = \frac{\mu E}{(1 + \frac{E}{E_c})}$$

(13)

$$= v_{sat}, \quad E \geq E_c$$

(14)

where $\mu$ is the low-field mobility, and $E_c$ is the critical field at the onset of velocity saturation.

3.4 Normal HEMT Operation Mode

1. Linear Region ($V_D < V_{sat}$)

When the applied drain voltage is not high enough to accelerate carriers up to saturation velocity, the drain current can be evaluated.
\[ I_D = \frac{\bar{w} \varepsilon_1 \mu_2 [V_G - V_T - V]}{(d_1 + d_c) \left[ 1 + \frac{P \alpha (V_G - V_T - V)}{q} \right]^{\frac{1}{3}}} \frac{E}{1 + \frac{E}{E_2}} \]  

where \( \bar{w} \) is the gate width, \( \alpha \) is the nonlinear charge control coefficient and \( V \) is the channel potential.

Integrating eq. (15) \( I_D \) can be written in terms of \( y_o \) and \( y_D \):

\[ I_D = \frac{-3\bar{w} \mu_2 \varepsilon_1 \sum_{n=1}^{\infty} \left[ \frac{\left( -\frac{1}{n} \right)^n}{n} \left( \frac{P \alpha}{q} \right)^{-n} (y_D^n - y_o^n) \right] + \left( \frac{P \alpha}{q} \right)^{\frac{1}{3}} \ln \left[ \frac{y_D + \frac{P \alpha}{q}}{y_o + \frac{P \alpha}{q}} \right]}{(d_1 + d_c) \left[ L + \frac{V_D}{E_{c2}} \right]} \]  

where \( y_o \) and \( y_D \) are defined as

\[ y_o = [V_G - V_T]^{\frac{1}{3}} \]

\[ y_D = [V_G - V_T - V_D]^{\frac{1}{3}} \]  

2. Saturation Region (\( V_D \geq V_{sat} \))

In saturation, we make use of eq. (9) and (14). The 2DEG channel current in saturation region leads to the equation

\[ I_D = \frac{\bar{w} \varepsilon_1 \mu_2 [V_G - V_T - V_{sat}] E_{c2}}{(d_1 + d_c) \left[ 1 + \frac{P \alpha (V_G - V_T - V_{sat})}{q} \right]^{\frac{1}{3}}} \frac{E_{c2}}{1 + \frac{E}{E_2}} \]  

To ensure continuous current transition from linear region to saturation region, we make use of the current continuity between these regions and obtain the following.

\[ -6(y_s + \frac{P \alpha}{q}) \sum_{n=1}^{\infty} \left[ \frac{\left( -\frac{1}{n} \right)^n}{n} \left( \frac{P \alpha}{q} \right)^{-n} (y_s^n - y_o^n) \right] + \left( \frac{P \alpha}{q} \right)^{\frac{1}{3}} \ln \left[ \frac{y_s + \frac{P \alpha}{q}}{y_o + \frac{P \alpha}{q}} \right] = y_s^4 \frac{E_{c2}}{L} (1 - K) + V_G - V_T - y_s^{\frac{1}{3}} \]  

Channel length modulation factor \( K \) and \( y_s \) are defined as

\[ K = \frac{L - L_c}{L} \]

\[ y_s = V_G - V_T - V_{sat}^{\frac{1}{3}} \]
We might determine $K$ and $V_{sat}$ by solving a set of two-dimensional Poisson equations by Laplace's method and keep the lowest space harmonic [7]. However, a simpler one-dimensional approximation to Poisson equation is adopted in the velocity saturation region, which assumes

$$\frac{\partial^2 V}{\partial x^2} = \beta_f$$

where the variation of carrier concentration in the direction perpendicular to the current flow is assumed to be negligible.

Solving eq. (23) subjected to boundary conditions, $E(L_x) = E_{ss}$ and $V(L_x) = V_{sat}$, the voltage drop across the saturated channel region can be written in terms of $K$ and $V_{sat}$.

$$V_{sat} = V_D - \frac{\beta L_x^2 K^2}{2} \frac{\bar{w}e_1 \mu_2 [V_{G}-V_T-V_{sat}] E_{c2}}{\left[ \left( d_1 + d_e \right) \left( 1 + \frac{P}{q} \alpha(V_{G}-V_T-V_{sat}) \right) \right]^{\frac{1}{3}} \left[ 1 + \frac{E_{c2}}{E_2} \right]} - E_{c2} L_K$$

Note that Eq. (20) and Eq. (24) are functions of two dependent variables, $y_s$ and $K$. Thus $V_{sat}$ and $K$ can be obtained by solving this set of coupled equations. Even though the equations are of high-order functions of $y_s$, the range in which the roots $y_s$ and $K$ are located is

$$0 \leq K \leq 1$$

$$-\frac{q}{P} \alpha \leq y_s \leq y_o$$

This will restricts the number of reasonable roots to one.

Once the gate voltage becomes larger than $V_c$, a conduction channel forms in the parasitic MESFET, i.e., the carrier supplying layer is not fully depleted, causing to current contribution from 2DEG channel as well as a MESFET channel underneath the Schottky gate.

### 3.5 Parasitic MESFET Conduction Mode $V_G \geq V_c$

Consider the case where the MESFET channel is pinched off at $x = L_M$ by the gate bias. The channel potential at this point reads

$$V_M = V_p - (\phi_B - V_n) + V_G$$

where the pinch-off voltage is given by

$$V_p = \frac{qN_{D1} \left( d_1 - \frac{n_{so}}{N_{D1}} \right)^2}{2 \varepsilon_1}$$

Within the conduction channel, the 2DEG will not be perturbed by gate and drain voltages, where the channel potential is lower than $V_M$, and will vary at thermal equilibrium values, $n_{so}$, which can be estimated self-consistently by
The current flowing through the MESFET channel can be derived as in the conventional MESFET model, which is given by

\[ I_p = \frac{\mu_n n_{\text{sat}}}{L} \left[ \frac{q}{4} \left( \frac{V_g}{V_T} \right)^{1/2} - \frac{V_D}{V_T} \right] \]

where

\[ n_{\text{sat}} = \frac{\varepsilon_1}{q(d_t + d_d) \left[ 1 + \frac{P}{q} n_{\text{sat}}^{1/2} \right]} \]

Since \( V_D < V_M \), the whole 2DEG channel behaves like a linear resistor. The current flows through the 2DEG channel reads

\[ I_1 = \frac{\bar{\omega} e \mu_e n_{\text{sat}}}{L} \left[ \frac{q}{4} \left( \frac{V_g}{V_T} \right)^{1/2} - \frac{V_D}{V_T} \right] \]

The total drain current \( I_D \) is

\[ I_D = I_1 + I_2 \]

2. Pinched-off MESFET Channel (\( V_M < V_D < V_{\text{sat}} \))

The 2DEG current due to equilibrium two-dimensional electron gas can be calculated as in the linear region.

\[ I_1 = \frac{\bar{\omega} e \mu_e n_{\text{sat}}}{L} \left[ \frac{q}{4} \left( \frac{V_g}{V_T} \right)^{1/2} - \frac{V_D}{V_T} \right] \]

While the current flows through the MESFET channel it is evaluated with the boundary condition, \( V(L_{\text{sat}}) = V_M \), and can be expressed as

\[ I_1 = \frac{\bar{\omega} e \mu_e n_{\text{sat}}}{L} \left[ \frac{q}{4} \left( \frac{V_g}{V_T} \right)^{1/2} - \frac{V_M}{V_T} \right] \]

\[ I_D = I_1 + I_2 \]
Similar to the formulation in the normal operation mode, the 2DEG current can be explicitly expressed as

\[
I_z = \frac{-3\bar{\omega}_s \varepsilon_s + \frac{V_m[\phi_s - V_e - V_T]}{3\left[1 + \frac{1}{q}\right]} + \sum_{n=1}^{\infty} \frac{(-1)^n}{q} \left( \frac{P\alpha}{q} \right)^n \left( y_0^\alpha - y_M^\alpha \right) + \left( \frac{P\alpha}{q} \right) \ln \left( \frac{y_D^\alpha}{y_M^\alpha} \right)}{(d_j + d_s) \left[ L + \frac{V_D}{E_{s2}} \right]}
\]  

(35)

3. Saturated MESFET Channel \((V_m < V_D)\)

In this operation region, formulation of the current flowing through the MESFET channel is identical to eqn. (34). And the 2DEG current is very similar to that in the pinched-off region and can be derived as

\[
I_z = \frac{-3\bar{\omega}_s \varepsilon_s + \frac{V_m[\phi_s - V_e - V_T]}{3\left[1 + \frac{1}{q}\right]} + \sum_{n=1}^{\infty} \frac{(-1)^n}{q} \left( \frac{P\alpha}{q} \right)^n \left( y_0^\alpha - y_M^\alpha \right) + \left( \frac{P\alpha}{q} \right) \ln \left( \frac{y_D^\alpha}{y_M^\alpha} \right)}{(d_j + d_s) \left[ L + \frac{V_D}{E_{s2}} \right]}
\]  

(36)

Similarly, the two dependent variables, \(y\) and \(K\), can be solved from the following two equations, derived from current continuity requirement, and from a Poisson equation in the saturation region, separately.

\[
-6(y_L + \frac{P\alpha}{q}) + \frac{V_m[\phi_s - V_e - V_T]}{3\left[1 + \frac{1}{q}\right]} + \sum_{n=1}^{\infty} \frac{(-1)^n}{q} \left( \frac{P\alpha}{q} \right)^n \left( y_0^\alpha - y_M^\alpha \right) + \left( \frac{P\alpha}{q} \right) \ln \left( \frac{y_D^\alpha}{y_M^\alpha} \right) = y_L^y \left[ E_{s2} L_c + V_G - V_T - y_L^y \right]
\]

(37)

\[
1 - \frac{B L^2 K^2 \bar{\omega}_s \varepsilon_s E_{s2}}{2(d_j + d_s) \left[ L + \frac{E_{s2}}{E_t} \right]} = y_L^y + \left( \frac{P\alpha}{q} \right)^y \left( V_D - V_G + V_T - E_{s2} LK \right) y_L^y + \left( \frac{P\alpha}{q} \right) \left( V_D - V_G + V_T - E_{s2} LK \right) = 0
\]

(38)

3.6 Discussions

The performance of the model is evaluated with the data from from Drummond et al.[2], we have the model to evaluate its performance. Fig. 2 shows the drain I-V characteristics for the 1 \(\mu m\) device. The used parameters are listed in Table 1. This model gives a better agreement with the experimental I-V curves than does the two-piece model. In this simulation run, the parasitic source and drain resistance are not included so parasitic resistances are not taken into account. However, this kind of gain reduction can be covered by adding two parasitic resistances \(R_s\) and \(R_d\) to the equivalent circuit model externally.
In the deeper gate reverse bias region, the modeled results have a better fit to the experimental data than to existing models due to the more accurate nonlinear charge control model operating in the saturation region. There the carrier concentration in 2DEG channel shows high sensitivity to the variation of gate potential, because the scarce carriers are repelled further from the heterointerface by the gate field. This effect leads to the reduction in transconductance.

The next step to improving the accuracy of this model is further modification of the parasitic MESFET portion. The existing parasitic MESFET model can not describe its external behavior adequately from the view point of gate transconductance. The remedy being planned will include some appropriate terms in the formulation to take care of real space transfer effects. Those current components will be weighted to guarantee the smooth transition from 2DEG current dominated region to the transconductance compression region around the onset of parasitic MESFET conduction.

3.7 Conclusions

We have derived I-V characteristics model for the HEMT's based on a nonlinear charge control formulation which takes into consideration the variation of the 2DEG distance offset from the heterointerface as a function of bias. The model has been kept in a simple and analytical form, making it a possible module to be implemented in the circuit simulation program without adding too heavy of a computation load. Also, the modeled results are in fair agreement with experimental data, which demonstrates that the model could be applied to the optimization of device performance.
Table 1

Parameters Used for the HEMT

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>1 um</td>
</tr>
<tr>
<td>W</td>
<td>145 um</td>
</tr>
<tr>
<td>d</td>
<td>450 A</td>
</tr>
<tr>
<td>dₙ</td>
<td>100 A</td>
</tr>
<tr>
<td>u₁</td>
<td>4050 cm²/V·s</td>
</tr>
<tr>
<td>v₁</td>
<td>1.0 × 10⁷ cm/s</td>
</tr>
<tr>
<td>E₁</td>
<td>3.0 KV/cm</td>
</tr>
<tr>
<td>ε</td>
<td>12.2 ε₀</td>
</tr>
<tr>
<td>ΔE₀</td>
<td>0.32 V</td>
</tr>
<tr>
<td>a</td>
<td>97.6 mV</td>
</tr>
<tr>
<td>b</td>
<td>1.0 × 10¹² cm²</td>
</tr>
<tr>
<td>α</td>
<td>1.0 × 10⁻¹</td>
</tr>
<tr>
<td>β</td>
<td>3.2 × 10⁶</td>
</tr>
<tr>
<td>V₀</td>
<td>1.15 V</td>
</tr>
<tr>
<td>N₀</td>
<td>6.5 × 10¹⁷ cm⁻³</td>
</tr>
<tr>
<td>Vᵣₒ</td>
<td>0.0 V</td>
</tr>
<tr>
<td>T</td>
<td>300.0° K</td>
</tr>
</tbody>
</table>
Fig. 1

Schematic diagram illustrating current in two-dimensional electron gas channel and parasitic MESFET channel.
Comparison of this model with Drummond's two-piece velocity model [2]
References


4. **SDA: A Highly Efficient Two-Dimensional Finite-Element Simulator for Semiconductor Device Analysis**

4.1 **Introduction**

The further reduction in geometrical dimensions of today's semiconductor devices and the advent of novel devices employing new materials, has highlighted the need for simulation, especially numerical simulation. This has become a powerful tool for providing insight into the functioning of devices by examining the distribution of various physical quantities in the interior of a device as well as external device behavior. In addition, the predictive capabilities of detailed numerical device simulations can be used to enhance device characterization techniques and to form a basis for the technological optimization of process, device and circuit designs.

A new approach to nonlinear finite element numerical analysis for semiconductor devices, developed by the VLSI Research Laboratory at UCSD, has been implemented in a general purpose two-dimensional finite element program for Semiconductor Device Analysis (SDA). In addition to the conventional finite-element algorithm constructed to carry out all types of linear and nonlinear finite-element analysis procedures, the following new features are incorporated to enhance computing capability, accuracy, and graphical interaction. In SDA, mesh generation for an arbitrary semiconductor device and the choice of an optimal solving algorithm can be done automatically by SDA or interactively through a dialogue with the user. By using high order error estimation and predictive techniques, a strategy for fully adaptive spatial grid generation and optimal choice of the time steps developed to find a balance between the reduction of discretization errors and the additional amount of computational efforts needed. The high computation speed of SDA is achieved by using simultaneously a modified Scharfetter-Gummel method to solve the two-dimensional Poisson's equation and the two-dimensional continuity equation on an optimal self-adaptive triangular mesh. New physical models have been derived for the analysis of and submicron devices and new III-V compound devices, such as the pseudomorphic InGaAs HEMT. These will be useful for investigating the device physics as well as suggesting new approaches for improvement.

This program can simulate GaAs MESFET and AlGaAs/GaAs heterojunction bipolar transistors as well as conventional silicon MOSFET and bipolar transistors. These devices can have arbitrary geometry and impurity distribution with a wide temperature range.

The program also considers incomplete ionization and variable band structure effects and can employ either Boltzmann or Fermi statistics. The graphics capability makes the graphic interpretation of simulation results very clear and attractive.
4.2 Basic Equations

The two-dimensional numerical model, SDA-1A, developed at UCSD for semiconductor device analysis first examines Poisson's equation

$$\nabla \cdot \nabla \psi = n - p + N,$$

where $n$, $p$ are electron, and hole concentration, respectively $\psi$ is potential and $N_i$ is net impurity concentration, which is defined as

$$N_i = N^* - N_A$$

$N^*$ and $N_A$ the ionized donor and acceptor concentration, respectively, are given by,

$$N^* = \frac{N_D}{1 + g \exp[-\frac{E_D - E_m}{k_BT}]}$$

$$N_A = \frac{N_A}{1 + g \exp[-\frac{E_A - E_m}{k_BT}]}$$

where $E_D$ is the donor energy level and $E_A$ is the acceptor energy level. The electron and hole concentration can be represented as

$$n = n_o e^{-(\phi_e - \psi - \phi_n) / k_BT}$$

$$p = n_o e^{-(\phi_h - \psi - \phi_p) / k_BT}$$

where $\phi_e$ and $\phi_h$ are the electron and hole quasi-Fermi potentials respectively, $V_e$ and $V_p$ are electron and hole band parameters respectively, $n_o$ is intrinsic carrier density for reference material and thermal potential is given by

$$V_t = \frac{k_BT}{q}$$

The equations for the electron and hole band parameters are as follows

$$V_e = -\frac{\chi_e - \chi_n}{q} + \frac{k_BT}{q} \ln \frac{N_i}{n_o} + \frac{k_BT}{q} \ln \frac{F_{e}^{1}(\eta_e)}{e^{\eta_e}}$$

$$V_p = -\frac{\chi_e - \chi_n}{q} - \frac{E_p - E_m}{q} + \frac{k_BT}{q} \ln \frac{N_i}{n_o} + \frac{k_BT}{q} \ln \frac{F_{p}^{1}(\eta_p)}{e^{\eta_p}}$$

The potential difference between the Fermi levels and the conduction or valence band are defined as

$$\eta_e = \frac{E_e - E_m}{k_BT} = \frac{q(\psi - \phi_e + (\chi_e - \chi_n) + \frac{1}{2} k_BT \ln \frac{N_i}{n_o} e^{-\psi/k_BT})}{k_BT}$$

$$\eta_p = \frac{E_p - E_m}{k_BT} = \frac{q(\psi - \phi_h + (\chi_p - \chi_n) - \frac{1}{2} k_BT \ln \frac{N_i}{n_o} e^{-\psi/k_BT})}{k_BT}$$
where \( x \) is electron affinity, \( E_g \) is bandgap, \( N_r, N_v \) are effective density of states for the conductor band and valence band respectively. The subscript \( r \) represents the reference material, and \( F \) is Fermi integral.

The second major equation for device analysis is the continuity equation.

\[
\nabla \cdot \vec{J} = \frac{\partial n}{\partial t} + R
\]

\[
\nabla \cdot \vec{J}_p = - \frac{\partial p}{\partial t} - R
\]

where \( R \) is the recombination rate which includes SRH bulk and surface recombination, Auger recombination and impact ionization. The recombination rate in turn is described by

\[
R = \frac{(n_p-n_n^2)}{\tau_{SNt} n_t + \tau_{SNp} n_p} - \frac{(n_p-n_n^2)}{\tau_{SNp} n_t + \tau_{SNn} n_n} - \frac{1}{q} \left( \alpha_n |J_n| - \alpha_p |J_p| \right)
\]

and \( \vec{J}, \vec{J}_p \) are the electron and hole current densities respectively are defined as

\[
\vec{J}_n = -q \mu_n \nabla \phi_n
\]

\[
\vec{J}_p = -q \mu_p \nabla \phi_p
\]

where \( \mu_n, \mu_p \) are electron and hole mobilities. For GaAs, the following empirical electron mobility model, which is a function of impurity concentration, temperature and field, is used

\[
\mu(N_r, T) = \frac{\mu_N \left( N_t - 10^6 N_{0T} \right)}{N_t + 10^6 N_{0T}} + \frac{\mu_{max} - \mu_{max}}{\left[ 1 - N_t / N_{0T} \right]^{1/3}} \frac{T}{300}
\]

\[
\mu(N_r, T, E) = \frac{V_{sat} \left( \frac{|E|}{E_s} \right)^4}{1 + \left( \frac{|E|}{E_s} \right)^4}
\]

\[
V_{sat} = V_{sat}(300K) - \beta T
\]

where \( E \) is field and \( V_{sat} \) is saturation speed.

It should be emphasized that the model described above has taken into account incomplete ionization of impurities ( \( N_0 = N_x, N_{0V} = N_x \) in (2), (3), (4)), and the different conditions which require either Fermi or Boltzmann statistics. In equations (7) and (8), with Fermi statistics \( F_{1st}^{\eta} = e^\eta, F_{1st}^{\bar{\eta}} = e^{\bar{\eta}} \), and for Boltzmann statistics \( F_{1st}^{\eta} = e^\eta, F_{1st}^{\bar{\eta}} = e^{\bar{\eta}} \). Finally, for heterojunctions we find that \( V_n = 0, V_p = 0 \) in (5), (6).
4.3 Discretization Method Using Scharfetter-Gummel Scheme

The equations (1), (11), and (12) as described above can be expressed in a general form like

\[ \nabla (v \nabla u) = f \]  

where \( v = v \cdot \Phi_s \cdot \Phi_n \).

\[ f = (n \cdot p + N_v \cdot \frac{\partial n}{\partial t} + R_n - \frac{\partial m}{\partial t} \), \]

\[ m = \| \varepsilon^+ - \varepsilon^- \).

After manipulation using a finite discretization scheme, the discretized equation set becomes

\[ F_i = \int_a \nabla \varphi \cdot \nabla u \, d\Omega + \int f \varphi , \, d\Omega = 0 \]  

This is a set of non-linear equations that can be solved by Newton method. The equation set in Newton iteration form is

\[ \sum_{i} (K_{ii} + G_{ii}) u_i = - \sum_{i} K_{ij} u_j + f_i \]  

where

\[ K_{ii} = \int_a \nabla \varphi_i \cdot \nabla u \, d\Omega \]  

\[ G_{ii} = \int_a \frac{\partial f_i}{\partial u} \varphi_i \, d\Omega \]  

\[ f_i = \int_a f \varphi_i \, d\Omega \]  

A seven point Gaussian quadrature method is used in eqs. (22)-(24), which reads as

\[ \int F d\Omega = \sum_{i=1} w_i F_i \]  

where \( w_i \) is weighting factor at the \( i \) point and \( F \) is the value of function \( F \) at the \( i \) point. The finite element mesh is comprised of triangular elements and so the 1st, 2nd and 3rd points are the vertices of the triangle, the 4th, 5th and 6th points are the mid-points of the edges and the 7th point is the center of the triangle.

Calculation of the values of \( n, p, e^+, e^- \) at the 4th, 5th, 6th, and 7th points is very important and critical because it has a significant effect on the accuracy of discretization and the convergence speed. First, we apply the Scharfetter-Gummel (S-G) formula into Poisson's equation as well as into the continuity equations using finite element methods, so that the computational effort required for convergence is reduced and the accuracy of solutions is improved significantly.
The S-G formulas assumed in SDA-1A are as follows:

\[ n(x) = (1 - g(x, \Delta \psi_x)) n_x + g x, \Delta \psi_x, n_x \]  
\[ p(x) = (1 - g x, -\Delta \psi_x) p_x + g x, -\Delta \psi_x, p_x \]  
\[ e^\psi \Delta \psi_x = B \Delta \psi_x e^\psi \]  
\[ e^{-\psi} \Delta \psi_x = B(-\Delta \psi_x) e^{-\psi} \]

where

\[ g(x, y) = 1 - \exp \frac{y - x}{b_x} / 1 - \exp y \]  
\[ B(x) = x / [\exp(x) - 1] \]  
\[ \Delta \psi_x = \psi_x - \psi_x \]  
\[ b_x = x - \chi \]  

To illustrate the effectiveness of this new scheme described above, which will be called method II, an abrupt n-P junction is simulated, and results are shown as Table 1. Method I is the classical finite element method, using low order polynomials to approximate n, p, e+ and e-.

Table 1. Number of iterations required for convergence (NUM) versus impurity density (Nd) with method II (the number of mesh points Np = 24, Na = 1 × 10^15 cm^-3, zero bias)

<table>
<thead>
<tr>
<th>Nd (cm^-3)</th>
<th>5 × 10^15</th>
<th>5 × 10^16</th>
<th>5 × 10^17</th>
<th>5 × 10^18</th>
<th>5 × 10^19</th>
<th>1 × 10^20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Np</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>NUM</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>
Table 2. NUM and the fewest number of mesh points \(N_{\text{mesh}}\) required for convergence versus \(N_d\) with method I (classical method) (\(Na = 1 \times 10^{15}\) cm\(^{-3}\), zero bias)

<table>
<thead>
<tr>
<th>(N_d) (cm(^{-3}))</th>
<th>(5 \times 10^{15})</th>
<th>(5 \times 10^{16})</th>
<th>(5 \times 10^{17})</th>
<th>(5 \times 10^{18})</th>
<th>(5 \times 10^{19})</th>
<th>(1 \times 10^{20})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N_{\text{mesh}})</td>
<td>24</td>
<td>37</td>
<td>149</td>
<td>194</td>
<td>352</td>
<td>446</td>
</tr>
<tr>
<td>NUM</td>
<td>11</td>
<td>36</td>
<td>17</td>
<td>70</td>
<td>134</td>
<td>260</td>
</tr>
</tbody>
</table>

Table 3. A comparison of discretization errors of Poisson's equation with method I and with method II, as a function of minimum mesh size \(\Delta X_{\text{min}}(\mu)\) (\(Na = 1 \times 10^{15}\) cm\(^{-3}\), \(N_d = 5 \times 10^{15}\) cm\(^{-3}\), zero bias)

<table>
<thead>
<tr>
<th>(\Delta X_{\text{min}}(\mu))</th>
<th>0.15</th>
<th>0.075</th>
<th>0.0375</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERR</td>
<td>46.07</td>
<td>21.30</td>
<td>9.30</td>
</tr>
<tr>
<td>(%)</td>
<td>13.68</td>
<td>8.23</td>
<td>1.58</td>
</tr>
</tbody>
</table>

Table 4. A comparison of discretization errors of continuity equations with method I and that with method II, as a function of \(\Delta X_{\text{min}}\) (\(Na = 1 \times 10^{15}\) cm\(^{-3}\), \(N_d = 5 \times 10^{15}\) cm\(^{-3}\), \(V_{\text{app}} = 0.6V\))

<table>
<thead>
<tr>
<th>(\Delta X_{\text{min}}(\mu))</th>
<th>0.15</th>
<th>0.075</th>
<th>0.0375</th>
<th>0.15</th>
<th>0.075</th>
<th>0.0375</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERR</td>
<td>34.19</td>
<td>19.86</td>
<td>10.58</td>
<td>24.19</td>
<td>12.21</td>
<td></td>
</tr>
<tr>
<td>(%)</td>
<td>20.20</td>
<td>5.53</td>
<td>2.01</td>
<td>12.01</td>
<td>5.76</td>
<td></td>
</tr>
</tbody>
</table>

From Tables 1 and 2, it is shown that method II is sufficiently accurate so that the convergence of the problem is almost independent of the impurity density and the grid size of mesh. This is in contrast to method I where the convergence is very sensitive to the impurity density and fineness and layout of the mesh. This is due to large integration errors, especially in highly doped regions. When \(N_d = 10^{20}\) cm\(^{-3}\) and \(N_d = 5 \times 10^{15}\), the new method only needs 7 iterations on a rather rough mesh (the number of points in the mesh \(N_p = 24\), while the classical method needs 246 iterations on a rather fine mesh. \(N_p = 446\). This shows the convergence rate increases by a factor of 37. The saved time is greater because the time per iteration on rough mesh is far less than on fine mesh.
From Tables 3 and 4, it is shown that with method II, we can obtain more accurate solutions than with method I on the same mesh particularly for Poisson's equation. In addition, the results indicate that applying method II to the continuity equations will be able to take the same advantage of the S-G scheme as those well known in the finite difference counterpart. When using the mesh whose minimum mesh size $\Delta x_{\min} = 0.0375\mu$ the accuracy of solution of $\nu, n, p$ for method II is increased by a factor of 5.8, 5.26, 3.20, respectively, over method I. The smaller $\Delta x_{\min}$ becomes the higher the improvement of accuracy.

### 4.4 Non-Linear Iteration Algorithm

It is well known that decoupled methods have the advantage of higher speed and less storage requirements compared with coupled methods in the low and medium current regime. However, its speed is much lower than that of coupled method in the high current regime. This is due to a considerable increase in the number of relaxation iterations between equations needed for convergence when there is strong coupling among the equations. Therefore, we proposed the following schemes to overcoming the drawbacks of decoupled methods.

1. **Adaptive control of inner iteration number in solving a single equation.**
   
   The inner iteration is Newton iteration and the out iteration is S-G iteration. Because the error function per outer iteration usually decreases only by a factor of 0.3 - 0.5, it is not necessary to have the inner iteration reach convergence during the initial phase of outer iteration before the start of next outer iteration. Therefore, the following strategy is adopted:

   If $\|F_e\| > C$, the inner iteration proceeds until its error function decreases by an order of two. Otherwise the iteration keeps going until the convergence is reached.

   The computation shows that the desired computation quantity is decreased considerably by the method.

2. **Variable order prediction technique and adaptive control of the step for bias voltage.**
   
   A variable order and step method, which is similar to the auto-integration algorithm, but in terms of voltage step instead of time step, is presented. It is shown that the number of outer relaxation iterations is decreased significantly by the method.

3. **The method to speed up convergence by using SOR and extrapolation in outer iteration is assumed.**
4.5 The Self-Adaptive Mesh Refinement Technique

The initial mesh for finite element discretization is generated by the simulator according to device structures provided by the users, and the self-adaptive mesh refinement procedure will be invoked to optimize the mesh whenever the speed-up of convergence rate is possible. The self-adaptive mesh requirement technique employed in SDA has eliminated the tedious mesh specification and has enhanced the program's efficiency substantially.

4.6 Some examples of Device Analysis

For purposes of illustration, the SDA program has been used to simulate GaAs MESFET device. The device dimensions and refined grid are shown in Fig. 1. Contour maps of potential distribution and three dimensional electron density distribution at several biases under thermal equilibrium condition are plotted in Fig. 2 and Fig. 3, respectively. Fig. 4 illustrates the external I-V characteristics obtained in this sample run, which demonstrated that this SDA simulator can effectively model semiconductor devices of arbitrary structures within a wide range of bias conditions, and can satisfy any user requirement.

4.7 Further Work

Revisions to the SDA-1A program are underway to include new function and models, such as hot electron models which are pertinent to devices with small size, and quantum well and bulk electron effects which are needed for HEMT simulation.