Demonstration of TCAD 2D FEM analysis on DMOS

Edmund Merrow-Smith
College of Engineering
Swansea University
551192

Abstract

In this report, the DMOS is described, the role FEM and TCAD tools play in designing semiconductor devices is outlined, and a demonstration of 2D FEM analysis of a DMOS is given using Silvaco's Atlas and Athena TCAD. In the demonstration, the behaviour of the the simulated device and implications for its design is explained. The simulation compares the output characteristic when lattice temperature is changed and impact ionization is added or removed.

1. The DMOS power transistor

Power electronic devices are (typically silicon) solid state devices designed to control large currents, switch them on and off quickly, block large voltages, in response to electrical signals, without any moving parts. Power devices typically operate on the same principal as common solid state devices used for signal and digital circuitry, but are physically larger or come in parallel arrays of the same device to carry enough current. However, as solid state technology has progressed, new designs have emerged that perform better and make better use of the area of a silicon wafer.

Bipolar junction transistors were the first used for power applications, but suffer from being current controlled and having small gains, so a large control current at the base was needed to control a current not much larger than itself, which was very inefficient. As MOSFET technology improved, it began to replace bipolar junction transistors in power electronics thanks to larger gains at high voltage and being voltage controlled, so only consumed power when switching. As the technology has continued to improve, new layouts to the MOSFET have emerged, such as the LDMOS, the trench MOSFET and the DMOS.

In a common MOSFET, two metal contacts (the source and drain) are electrically isolated by a PNP or NPN junction in the bulk material, and a third contact (the gate) is isolated from the middle of that junction by a thin oxide layer that, when charged, connects the source and drain by forming a minority carrier channel at the oxide interface. The current stays very close to the surface, entering and exiting on the same side of the device, and sees little of the bulk material. This causes three problems in power applications (referring to an n channel device):

- When operating at large voltages, the gate voltage is typically close to the source, creating a large potential across the thin oxide layer separating the gate and drain, which breaks down and conducts a large current which kills the device.
- A thicker oxide layer to solve this undesirably decreases the gain, increases the threshold voltage of the gate, and increases the turn on/off times of the device.
- Moving the drain laterally away from the gate to solve this increases the distance the the current spends confined in the channel, increasing the on resistance, and using a larger surface of the chip.

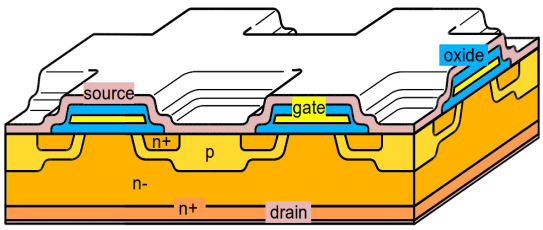


Fig. 1 A DMOS array

The double diffused MOSFET (or DMOS, *example left*) functions largely the same as any other MOSFET, but differs in that the drain is on the underside of the chip, so that after flowing through the channel, the current breaks away from the oxide interface and flows down through the bulk of the material. This disperses the current after the channel throughout the bulk material, reducing the on resistance.

With the drain on the underside, it is further from the gate, preventing oxide breakdown, but the entire structure is also grown atop a lightly doped epitaxial layer designed to increase the maximum drain-source voltage the device can block, and can be made thicker to increase the breakdown voltage without affecting how much area the devices takes up on the wafer.

A DMOS is designed to always be operated with the base at a higher voltage than the source, and the contact between the p region and source forms a diode between the source and drain at the pn junction which prevents the device being used as a transistor in reverse bias and also acts as a freewheel diode when the device is used in voltage converters supplying inductive loads.

2. TCAD & FEM

Being able to understand how and whether a design will meet its specifications, or a solution will adequately solve its problem, is vital to engineering and design. To that end, computer aided design (CAD) is the use of computers and software tools to model, analyse, optimize, and document solutions, and is a very important part of modern engineering. Technology CAD (TCAD) is focused on semiconductor devices, including tool to assist with the design of devices and their fabrication.

The finite element method (FEM) is a tool for solving differential equations, which reality is commonly described by, and are not trivial to solve. It is commonly used as a CAD tool in engineering to model and simulate designs with shapes that are non-trivial to analyse, by breaking the shape down into a connected array of many simple shapes (a mesh). The behaviour of each shape is linked to the behaviour or connected shapes (or elements), which forms matrices

that are solvable by a computer.

Silvaco Inc's Atlas and Athena products are 2D FEM tools that can simulate the fabrication and characterisation of silicon devices. Here, a simple DMOS structure is analysed to demonstrate how being able to predict its behaviour and characteristics can be used in its design.

3. Simulation results

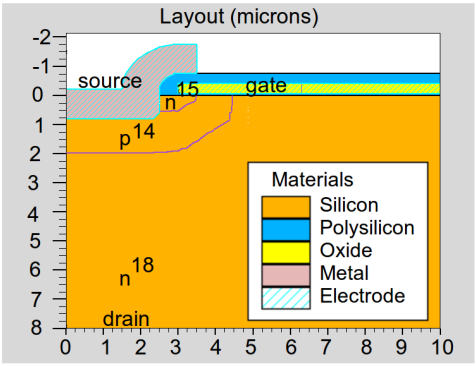


Fig. 2 DMOS structure

The Silvaco provide an example simulation, “vertical DMOS turn-on characteristics”, which provides the transfer characteristic and structure layout of what would be one element of an array of alternating mirrored structures to form a larger single device capable of carrying high current.

The input file was modified to output lattice temperature in the structure file, plot the output characteristic at 4 gate voltages and 4 temperatures, and repeat to simulate and output impact ionization. From the transfer characteristic, it can be seen that the gate threshold voltage is approximately 3V (and is known to vary with temperature), so gate voltages of 3V, 6V, 9V and 12V were chosen. The temperature range was chosen to loosely reflect temperatures seen in extreme aerospace and vehicle conditions. This was achieved by setting the temperature of the metal contacts, which then acted as heat sinks.

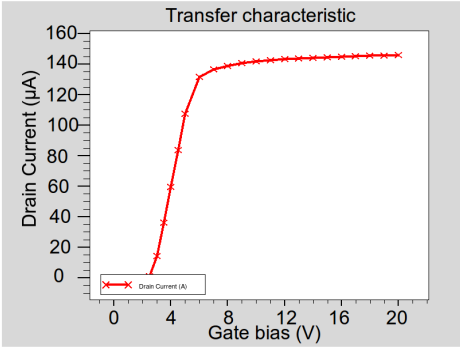


Fig. 3 Transfer characteristic

Fig. 4 Output Characteristic

without impact ionisation **with impact ionisation**

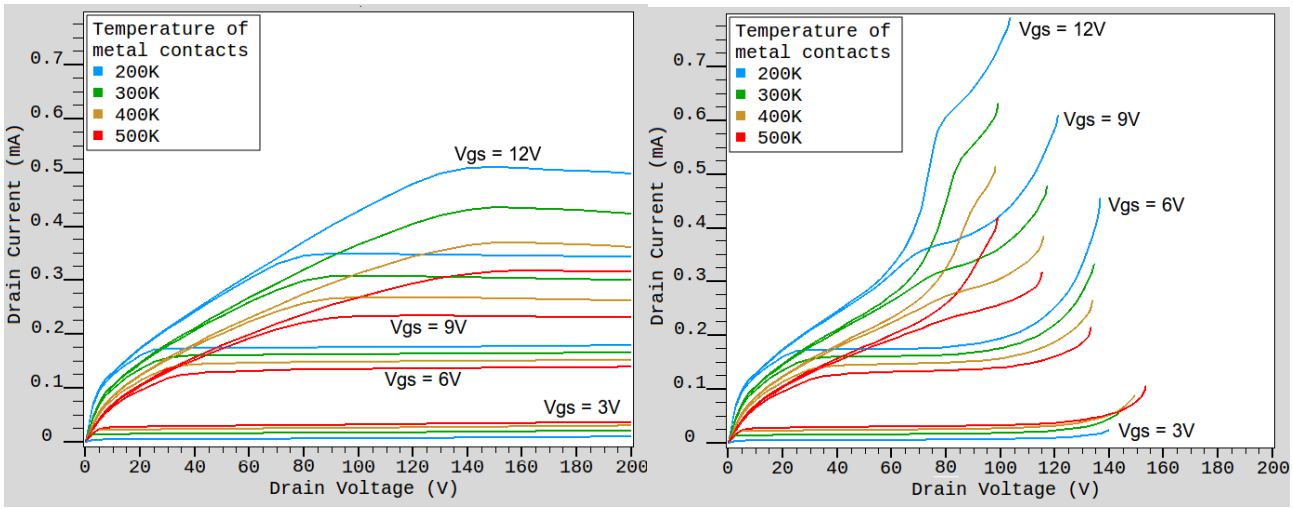


Fig. 5 Final heat power output (mW)

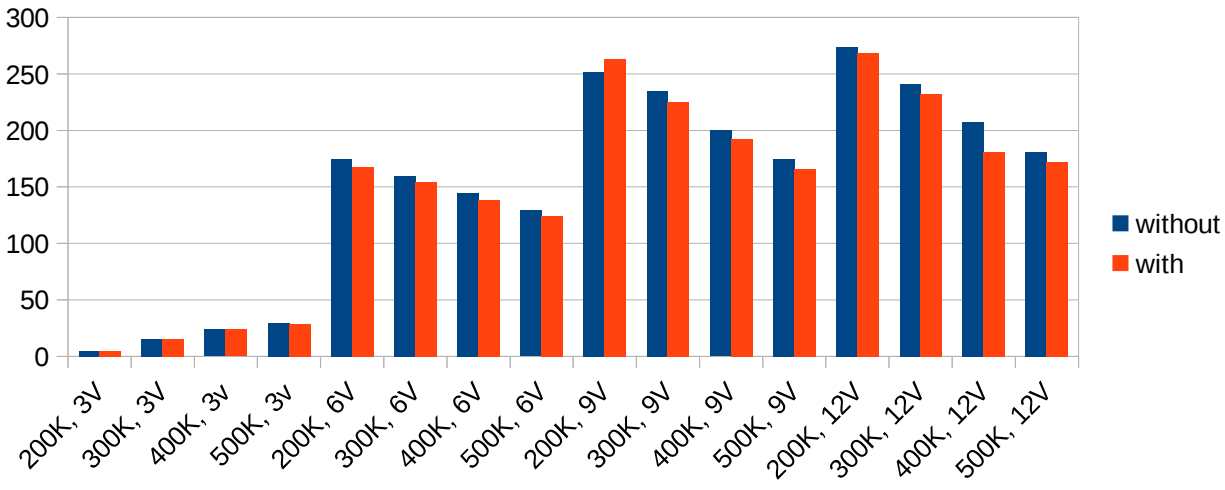
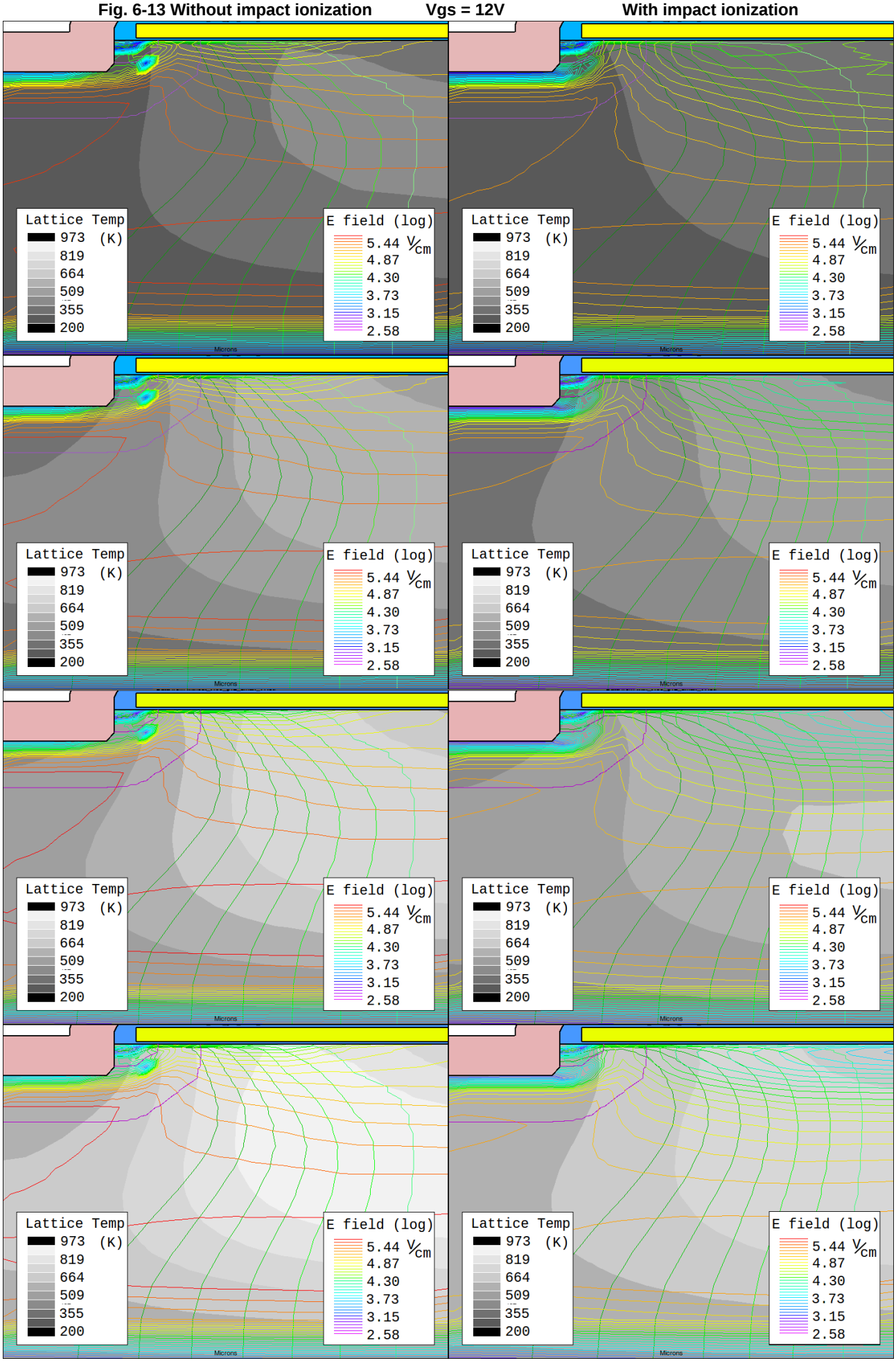


Fig. 6-13 Without impact ionization Vgs = 12V With impact ionization



4. Analysis

Immediately apparent is that the output characteristic plots stop short of the simulation's 200V maximum drain voltage, as the device has broken down and the model no longer converges.

The similarity in lattice temperature between each pairs of runs with and without modelling impact ionisation, the total output power of their final steps is roughly equal. This explains why the lattice temperature plots look similar for their structure files. However, this is a coincidence, and would not be the case had a different final drain voltage been chosen.

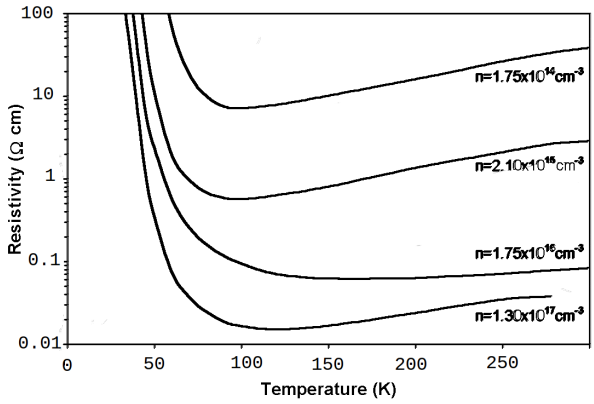


Fig. 14 resistivity of silicon vs temperature

Without impact ionization, in the active region of the output characteristic, for gate voltages of 9V and 12V, the graph slopes down, with higher drain voltages allowing less current through, for all temperatures. This indicates that the device is suffering from self heating, where the power flowing through the device is heating the device enough to significantly alter the material's resistivity.

Left is the resistivity-temperature plot of silicon for various doping concentrations. The resistivity rises with temperature due to the increase in phonons in the lattice, and thus the mean phonon scattering frequency, which reduces the carrier mobility, and thus the conductivity of the material.

It must be noted that the current drops slower than the voltage rises, which using P = IV tells us that the power output increases with voltage, consistent with the explanation that it is the increased power output that heats the device further to a point of equilibrium for each step in voltage. It is also telling that the 500K plots sloped far less than the 200K plots, indicating that the effect is temperature dependant, which is consistent with this being a heat-related effect.

Table 1 - Highest temperatures reached with Vgs = 12V

Original Temperature	Without	Difference	With	Difference
500 K	896 K	396 K	741 K	241 K
400 K	741 K	341 K	644 K	244 K
300 K	587 K	287 K	509 K	209 K
200 K	432 K	232 K	355 K	155 K

This table shows that with impact ionization, the device has likely heated up to the point of destroying itself. However, without impact ionization, the device still has a significant temperature difference inside the bulk of the material, which will cause stress from heat expansion. Such a device would need a way to better dissipate its heat evenly within itself, and more evenly distribute area that generate heat if it were to be used in such a manner that it regularly turned on and off, or the continuous cycling application of thermal stress would wear the device quickly.

It is apparent by comparing the output characteristics that impact ionization is having an effect on the device characteristics, and by looking at the impact generation rate overlaid with the electric field it becomes apparent that the bulk of the device is in avalanche breakdown. The critical field of silicon is 10^5 V/cm, which at Vds = 50V, the device is clearly reaching, and the impact ionization rate confirms that this strong field is inducing impact ionization.

Looking at the green flowlines, it is apparent that the current is sticking to the oxide surface when Vgs=12V, as is expected of a DMOS, yet when Vgs=3V, the current quickly separates from the oxide surface. In both cases much of the bulk material is in avalanche breakdown, so it can be concluded then that this device is designed to deliver most of its current through the bulk material with the avalanche effect. This is punchthrough current and while parasitic in some cases, as punchthrough current happens deeper in the bulk material so is poorly controlled by the gate voltage, in this case it is possibly deliberately encouraged by the device design to overcome its self heating effects and conduct better.

Fig. 15 impact ionization, with electric field contours and current flowlines

T = 200K, Vgs = 3V Vds = 50V T = 200K, Vgs = 12V