A 3D thermal simulation tool for integrated devices – *Atar*

T. Smy, D. Walkey, S.K. Dew[1]

Department of Electronics
Carleton University
Ottawa, Ont., Canada K1S 5B6
[1] Department of Electrical and Computer Engineering
University of Alberta, Edmonton, AB, Canada T6G 2G7

Abstract

This paper presents a novel 3D thermal simulation tool for semiconductor integrated devices. The simulator is used to automatically generate an accurate 3D physical model of the device to be simulated from layout information. The simulator produces an appropriate mesh of the device based on a rectangular block structure. The mesh is automatically created such that a fine mesh is produced around heat generation regions, but a moderate number of blocks are used for the entire device.

This paper first confirms that the simulator produces an accurate solution to the non-linear differential equation describing the heat flow. Then model generation from three example technologies (silicon trench, GaAs mesa structures, silicon on insulator) is presented. The potential of the simulator to quickly and easily explore the effect of layout and process variations is illustrated, with the simulation of a two transistor GaAs power cell as a large example.

The program incorporates a transient solver based on a transmission line matrix (TLM) implementation using a physical extraction of a resistance and capacitance network. The formulation allows for temperature dependent material parameters and a non-uniform time stepping. An example of a full transient solution of heat flow in a realistic Si trench device is presented.

1 Introduction

The quest for a more accurate understanding of the thermal properties of integrated semiconductor devices is driven both by a need for better models of these devices for circuit level simulation and also reliability concerns. Thermal behaviour of electronic devices has generally been characterized using a thermal resistance (R_{th} defined as the steady state temperature above ambient divided by the power dissipated in the device) and a thermal capacitance (C_{th}). A number of analytic approaches to the steady state determination of R_{th} have been pursued [1, 2, 3]. However, all of these attempts are limited in accuracy by a necessarily simplistic approach to the device geometry. Process factors and layout geometry strongly influence the value of both R_{th} and C_{th}, particularly in substrates with poor thermal conductivities such as GaAs and InP. Determination of these parameters may need to include the effect of a number of factors, including the "backend" (metalization and interconnect structures), complex device geometries such as multiple emitters and thermal coupling between adjacent devices. A significant challenge is the need to evaluate and model a large variety of process technologies varying from silicon on insulator (SOI) technologies to high power III-V technologies.

To illustrate the complexity of the heat flow in an integrated device Fig. 1 shows a 2D cross-section of a generic mesa transistor structure with heat spreader. The heat is shown as being generated in a well defined region at the base/collector junction, a reasonable assumption at moderate current densities. A portion of heat flows into the substrate spreading in three dimensions as it flows towards the bottom of the substrate which is a uniform temperature boundary condition. The remainder of the heat energy flows up the emitter metallization, back down to the substrate surface through the spreader and then to the backside of the substrate. This heat flow can be further complicated by the presence of base and collector metallization and any flow through the insulating layers present in the backend. This figure shows all cooling as being due to conductive paths to the backside of the wafer. Although radiative and convective cooling off the front of the wafer could be present it is negligible at any reasonable device temperatures.

A complete theoretical determination of R_{th} requires the coupled solution of both the semiconductor device equations and the heat equation[4]. This combined electro/thermal problem is a significant challenge. The heat flow from an integrated device is characterized by a number of complications. Specifically, the complications are: heat spreading in 3 dimensions, a wide variety of materials with greatly differing and often temperature dependent thermal conductivities, complex geometries and a variety of boundary conditions. A major feature of the heat problem is the need to simulate a very large region of the device and substrate, and in some cases the package geometry may need to be taken into account. The simultaneous solution of the 3D electro/thermal problem is therefore difficult due to the need for very fine meshing of the device equations at the device junctions and a need for a large simulation region to produce an accurate thermal simulation. However, the thermal properties of the layout and process can be determined without solving the device equations if the power generation region is known *a priori*. Knowing the power generation as a function of position the heat equation can be solved independently to calculate the device temperature profile and R_{th} determined. For a given device the general size, shape and position of the heat generation region is known as a function of the operating point. The thermal resistance can then be correctly calculated ignoring the operating point dependence if it is only minimally sensitive to the expected changes in position and shape of the heat generation region due to the thermal/electrical coupling. Although determination of

subtle effects requires a full electro/thermal approach basic. Basic thermal properties of the device, layout and process can be determined by using an independent thermal simulator with appropriate assumptions[5].

A number of numerical tools exist for specifying and solving 3D thermal problems of this type. They generally have a sophisticated GUI to aid in building the 3D model, specifying material parameters, and setting boundary conditions and heat generation regions. The tools are based on finite element or finite difference solution "engines" and have powerful software tools for producing the discritization of the 3D geometry needed for a solution. A significant drawback in the use of these tools is the time and expertise needed to build a 3D model. The tools have a steep learning curve and even after mastering the tool generation of the 3D model of a single device can take a significant effort. Building the model also requires a detailed knowledge of the IC process technology. Acquiring the knowledge base and obtaining the time required for the building and simulation of a single device geometry is often prohibitive for both device engineers and circuit designers.

The solution to this problem is a simulator that given a technology description and layout information will automatically generate a 3D model of the device, discretize the model, and solve for the temperature distribution in the device (see Fig. 2). In this paper we will present a thermal simulation tool that uses a technology description and layout information to automatically generate a full 3D model, complete with discretization, and then solves for the temperature distribution. A number of technologies will be used as illustrative examples and a comparison between the tool and a commercial simulator will be used to verify the accuracy of the solution.

Although many applications simply require the solution of the steady state problem and then the extraction of an effective thermal resistance. A number of problems require a full solution of the transient response of the device. TLM has shown itself highly appropriate for solving time dependent transient heat flow problems and is particularly attractive for multi-material and nonlinear problems as any temperature or material dependence of the thermal parameters can be locally accommodated[6]. A particular advantage of the TLM method is the ability for time step to be adjusted dynamically during simulation to reduce execution time.

2 *Atar* a 3d thermal simulator

The physics involved in solving for the conductive heat flow requires solving the following partial differential equation[7]:

$$\rho C \frac{\partial T}{\partial t} = \nabla(\kappa(T)\nabla T) + g(x, y, z, t) \tag{1}$$

where $\kappa(T)$ is the temperature dependant thermal conductivity, C the thermal capacity of the material, ρ the material density, and $g(x, y, z)$ describes the heat generation. Relevant solutions to this problem are difficult to obtain due to the complexities outlined above and the non-linear nature of the equation resulting from the temperature dependence of κ. It is important to note that both Si and GaAs display significant drops in κ as the temperature changes from room temperature to 200 °C[8].

To produce a numerical solution an appropriately discretized 3D model must be built, boundary conditions applied, a mathematical model extracted and then solved using a non-linear solution method. Due to the requirement of an automatic generation of the model from layout information it was decided that the

model would be built first from the chip surface down into the substrate and then the backend interconnect structure would be built from the chip surface up. The model is constructed by first defining a surface layer and then projecting this layer either downwards into the the substrate or upwards into the backend.

2.1 3D Model Creation

A significant challenge in the model building is the automatic creation of a model with a high density of elements or nodes in the regions of large heat flow or temperature gradients. This essentially means a large number of small elements must be created near the heat generation regions in the simulation region. The heat generation regions in an integrated device are generally fairly close to the chip surface and well defined by the device layout. For example a vertical silicon BJT would generate heat at the base/collector junction in a thin plate defined essentially by the emitter geometry.

The basic philosophy of the model building is to create a complex mesh geometry using rectangular blocks of varying sizes. The model will be created such that a block and its neighbors form a relatively simple set of topologies where by each block can have a maximum of two blocks on any vertical side and a maximum of four blocks on the top or bottom of a block. From these block topologies a network of thermal resistances and capacitances will be extracted representing the thermal equation to be solved. As will be shown complex meshes can be built using these rules. A second key part to the model building is to build from the chip surface down into the substrate and then up into the backend.

The creation of a model is shown for a simple example in figures 3 and 4. The model is constructed layer by layer starting with the surface layer. The surface layer is created by starting with a very high uniform density of blocks (see Fig 3a). The layout information is then used to allow for the iterative amalgamation of blocks into single blocks from groups of four. This amalgamation is controlled by the position of the mask edges and variables associated with each mask which set the maximum block size allowed inside and outside mask areas. During the iteration groups of four blocks are visited and subject to a number of tests to see if amalgamation is allowed:

1. The blocks are checked to see if all four initial blocks are the same size and of the same material.
2. The final size of the new block to be created is checked against the maximum block size allowed inside any masks covering the area.
3. Neighboring block sizes are checked to see if there will be at most two blocks on any one side of the new amalgamated block.

The amalgamation of blocks in the surface layer stops when all blocks have reached their maximum size as defined by the masks and a technology file. The net result of the meshing is to produce a 3D quad tree mesh (QTM).

The technology file contains information on the two the types of masks allowed and the impact of these masks on the model. Substrate masks are defined providing information on the material type, maximum block size with in the mask, and allowed physical connections to backend layers. An example of a substrate mask would be a mask defining substrate contacts to an emitter metalization layer or a mask defining a

oxide trench isolation region. The other type of mask is a backend mask. For backend masks information is provided on the material associated with the mask, the thickness of the layer, maximum block sizes in the layer and physical connections to other backend layers. A typical backend mask would define for example an aluminum “metal 1” layer or a “W plug” interconnect.

Using the information from the technology file and the mask geometries the surface layer generated in the model is created such that very a high density of blocks is produced near areas that will define the heat generation region. This is done for example by making the maximum block size inside the emitter contact mask to be very small. This forces a high density of nodes at the emitter with a smooth transition to fewer nodes away from the emitter. A simple example is shown in Fig 3b. As noted above, during the amalgamation of the blocks in the surface layer the block topography is restricted so that on any vertical side of a block the maximum number of blocks is two. This produces a gradual controlled gradient in block size as we move away from the heat generation regions.

Once the surface layer has been formed the substrate is built by duplicating the surface layer. As the substrate layers are created the number of blocks in each layer is reduced by further amalgamating groups of four blocks into a single block. The amalgamation is controlled by both the layout geometry, the block position and also the requirement that above or below a single block there can be no more than four blocks. The result of these requirements is that all blocks have a maximum of two blocks on each side and four on the top or bottom. This allows for a significantly restricted set of block topologies. The thickness of the substrate layers is generally increased non-linearly as we move down into the substrate. This allows for the generation of relatively large blocks at the bottom of the substrate where thermal gradients are small. Fig. 3c and d show two characteristic substrate layers created during the formation of the substrate. This figure clearly shows the coursing of the grid and the increase in thickness from the surface layer shown in Fig. 3d.

The backend is created in a manner similar to the substrate and is shown in Fig. 4. The surface layer is duplicated and layers extruded upwards. These layers are formed of two materials, for example metal and a dielectric (in the figure the dielectric blocks are only shown in outline for clarity). Once the backend layers have been formed, blocks with zero conductivity are removed and an iterative procedure is used to remove extra blocks from the structure by merging blocks of four into single blocks. All the blocks forming the backend have the same limitations on adjacent block topology as the ones in the substrate with a maximum of four blocks on top or below and a maximum of two on any vertical side.

2.2 Extraction of the Thermal Network

After the creation of the 3D model a thermal network is extracted consisting of non-linear resistors and capacitors. Each block is visited and using the topology of the connections with adjacent blocks a network of resistors to the neighboring nodes is created. The value of each resistor is calculated using the common cross-sectional area, the size of the two blocks and the local thermal conductivity. The thermal capacitance of the material is used to calculate the value of a capacitor connected from the block node to ground. As the material properties are, in general, temperature dependant the resistor and the capacitor values are a function of the node temperature. In Fig. 5 the heat flows for a single block are shown, each heat flow would be represented by a thermal resistor. This block has a single block above, four below, one on the front, back and left sides and two on the right side. The heat flow associated with the heat capacity of the block is not

shown.

A number of boundary conditions can also be placed on the model. Regions of the model can be designated to be at a fixed temperature, as would be used for a fixed backside wafer temperature. Constant power generation can be specified in volumes such as the channel of a MOSFET or the base/collector junction of a BJT. The boundary condition on the vertical sides of the simulation region is somewhat problematic. Placing a fixed boundary condition on these surface produces a dramatically incorrect result, unless a very large simulation region is used at the expense of very long simulation run times. A more natural boundary condition is a zero flow condition across these surfaces. However, the use of this boundary condition has implications. A zero flow surface implies a mirroring of all geometry in the simulation region at this surface.

2.3 Steady State and Transient Solvers

Once the thermal model has been extracted and the boundary conditions set the network needs to be solved. The simulation tool allows for either a netlist to be outputed (allowing for a solution to be found using an external circuit simulator such as spice), or an internal solver is used. Currently two steady state internal solvers are available. A direct solver can be used which first sets up a global sparse array and solves the linear set of equations representing the thermal resistor network. If the temperature dependence of the thermal conductivity is to be considered an iterative technique is used using SOR and the repeated solution of the linearized set of equations using the direct solver. A second solver based on SOR of the nodal temperature equations has also been implemented and is particularly useful for large problems. This solver implicitly handles the temperature dependence of the material properties. Once the solution has been found the maximum temperature is obtained and a greyscale representation of the temperature distribution can be displayed.

A transient solver has been written that use the TLM method. This method is particularly suited to the extracted network approach described above. TLM modeling represents a physical model of heat flow as a sequence of voltage pulses travelling through a matrix network of transmission lines. By computing the voltage (temperature) at the nodes connecting the transmission lines, an explicit, unconditionally stable solution to the heat flow equation can be described by a set of equations that are iterated in a straightforward manner.

The fundamental approach to TLM modeling of the QTM is shown for a 2D case in Fig. 6(a). At the center of each block is a temperature node, and a transmission line from this node to all adjacent nodes is created. Each transmission line consists of two impedances and two resistances, one of each associated with each block. This is shown in Fig. 6(b) for two nodes denoted as n_i and n_j. Although the association of R_dC_d with $\rho C/\kappa$ allows for a number of choices in choosing the values of R and Z of each transmission line, a natural choice is to use the physical values of thermal resistance and capacitance associated with nodal connections. A complete description of the use of TLM on a QTM will be presented in reference [9].

The program is written in a combination of C and tcl/tk. The main body of the program and data structures are written in C with the visualization being handled by a set of tcl/tk routines executed using *wish* a tcl/tk interpreter.

3 Simulation results

3.1 Comparison with Patran — simple heat source

In order to confirm that the simulator accurately solved the steady state heat equation the simulation tool was compared with a commercial FEM 3D modeling tool (Patran). A simple case of a 2 μm x 25 μm x 0.15 μm heat source in a rectangular section of GaAs was taken as base line. For this model the heat source was a 100 mW distributed evenly over the generation region and the backside temperature was fixed at 355 K. In Fig. 7 typical models for both simulators are shown with a details of the high density discretization around the heat source also presented. The *Atar* model shows the characteristic gridding pattern produced by the tool and was automatically generated. The Patran model has a sophisticated non-rectangular grid which minimizes the number of elements, but does however require a significant amount of time and expertise to build.

In Fig 8 temperature distributions are shown for both simulators. To allow for an easy direct comparison of the two simulators Fig. 9 shows the temperature plotted along the surface and into the substrate for two identical structures one built using *Atar* and the other Patran. Two cases were compared one for a constant thermal conductivity (GaAs at 355 K) and the other using a non-linear one ($k = k_0(1 + k_{c1}(T - T_O) + k_(T - T_O)^2)$. Two *Atar* simulations were done with differing mesh densities to determine the sensitivity of the temperature distribution to meshing. As can be seen in the figure the both simulators produce essentially identical temperature distributions. The Patran simulation consisted of approximately 10,000 finite elements. The number of elements in the *Atar* models were of the order of 6,000 and 10,000 elements for the course and fine mesh models respectively. The *Atar* simulations were found to match the Patran results marginally better for the finer meshed model, however, further refinement was not found to be useful. The Patran simulation completes in approximately 30 min on a Sun Ultra 1. The course and fine meshed *Atar* simulations complete in 15 min and 45 min respectively.

3.2 Example technologies

The primary use of the *Atar* tool is the quick building and thermal simulation of integrated devices in a wide variety of IC processes. As an example of three substantially different technologies and devices Fig. 10 shows devices simulated in: 1) a high speed GaAs HBT power amp technology. 2) a trench isolated Si BJT technology and 3) a SOI MOSFET technology. For all of the technologies shown only simple backend topographies were defined and the thermal conductivity of the backend oxide was assumed to be zero and is not present in the model.

The GaAs technology is a mesa HBT process incorporating a heat spreading technology whereby the emitter metalization is used to remove heat from the hot emitter stack and sink it into the semi-insulating GaAs substrate. The HBT was assumed to dissipating 100 mW of heat at the base/collector junction in region defined by the emitter geometry.

The main feature of the silicon trench technology is the presence of oxide lined trenches deep into the substrate. These trenches substantially restrict the heat flow in the substrate due to their low thermal conductivity. The trench device was assumed to be generating 10 mW of heat at the base/collector junction.

The SOI technology required the definition of a layered substrate in which the top layer was formed of oxide. A silicon mesa was then placed on the oxide defining the active area and a simple backend defined. The heat dissipated in the device was 1mW and was generated in the channel between the drain and the source.

For both the Si trench and SOI technolgies the backside temperature was set a 300 K.

An assumption in the use of *Atar* is that an independent thermal simulation will provide useful information. As the position of the heat generation region is assumed *a priori* it is necessary to investigate the relative importance of the size and position of the region. As an example the effect of moving the position of the heat generation region in the Si trench device away from the surface and into the substrate was investigated. At low to moderate current densities the bulk of the heat generation will be in the base/collector junction where the electric field is largest. However as the current densities are raised the region of maximum electric field will enlarge and push into the collector due to the Kirk effect. In Fig. 11 the effect on the maximum device temperature of moving a 0.075 μm generation region deeper into the collector region of the device is shown. The temperature drops as the generation region moves away from the device and towards the backside boundary. This is due to a reduction in the thermal resistance between the generation region and the backside boundary condition.

It is important to note that the variation in the position of the generation region in this data is far more than would be expected from either process variations or from that due to changes of the operating point due to temperature variations. Therefore from these results we can conclude that the position and shape of the generation region will have only a small effect on the maximum temperature of the device and for a given power level R_{th} can be determined using an un-coupled thermal analysis.

3.3 Process and Geometrical Variations

The primary advantage of *Atar* is that once a technology has been defined it is very straight forward to generate models with differing geometries or to investigate the effect of small modifications of the technology. A large variety of device geometries and process variations can quickly and efficiently be generated using the simulator in a batch mode. A simple example of this is the optimization of the GaAs HBT heat spreader technology. There is an engineering tradeoff between the size of the heat spreader and the packing density of the devices. The larger the heat spreader the cooler the device will run. However, the increase in the effectiveness of the heat spreader diminishes for very large spreader lengths due to the finite conductivity of the emitter metallization.

This effect is presented in Fig. 12. The first figure shows a plot of the temperature of the surface through the heat spreader. This plot clearly shows the rise in temperature of the GaAs surface under the heat spreader due to the heat flow off the emitter down into the spreader. Also shown in this figure are surface temperatures from an identical Patran simulation, as with the simple simulations above the temperature distributions for the two simulators match well. The second figure displays the maximum temperature of the device under a number of configurations. With no metallization present the device temperature is 110 degrees above the backside temperature of 355 K. If a simple post is present the maximum drops by 10 degrees due to heat spreading through the emitter post. The presence of even a small heat spreader drops the maximum temperature by 15 more degrees. Increasing the length of the heat spreader increases the drop in device

temperature, although with diminishing returns. In addition to studying the effect of the layout on the GaAs HBT thermal response process variations such metal and mesa thicknesses can also be easily studied using *Atar*.

Examples of variations for trench devices can be done very simply using *Atar*. The primary characteristic of the trench device is the effectiveness of the thermally isolating trench structures. The trenches force the heat flow down into the substrate reducing the 3D heat spreading and increasing R_{th}. Moving the trench away from the heat source reduces the effect of the trench in restricting the heat flow and the temperature decreases. A significant rise in temperature is found when the trench is brought closer than 2 μm to the emitter. A rise of 30% is found as the separation is decreased from 2 μm to zero. Conversely, for separations larger than 5 μm little effect is seen. Increasing the trench depth also constricts the heat flow and causes an increase in the device temperature, with an almost linear increase from 308 °C to 316°C as the trench depth is increased from zero to 20 μm.

The device R_{th} is also a function of the emitter geometry. For an initial structure having a 0.5 μm x 10 μm emitter changing the width has only a small effect (20%)for even a increase of 6 times. However, reducing the length has a dramatic effect as the device essentially changes from a line source at long lengths to a square source as the length approaches the width. A 250% increase in R_{th} is found for a reduction in the length by a factor of 6. For an SOI structure the rise in the device temperature as the oxide layer thickness is increased is of interest. As would be expected a nearly linear variation is found with a rise in maximum temperature from 305 °C to 308 °C as the thickness is increased from 0.10 μm to 0.30 μm.

Figure 13 shows the sensitivity of a variety of devices to heat flow through the backend metalization. To determine this sensitivity the emitter metallization was extended and a fixed boundary condition (300 K) placed at it's far end (somewhat similar to having a bond pad at the end of the line). The length of the line to the emitter was then varied from 2 to 60 microns in length and the device temperature determined. For the Si devices little effect can be seen until the fixed boundary condition is very close to the device. A GaAs device on the other hand shows a significant effect due to the high intrinsic R_{th} of the device. Interestingly, the SOI device although having a large R_{th} due to the oxide layer does not show as appreciable an effect. The decrease in the strength of the effect is due to the heat generation region being in the channel of the device. This results in an additional thermal resistance to the heat flow to the source and drain contacts and then through the backend moderating the effect of altering the backend.

3.4 GaAs Power Cell Model

The previous examples presented singled devices with limited metalization as examples of distinctive technologies. However, once a technology has been defined a larger area of the circuit can be simulated. In Fig. 14 a GaAs power cell is shown in the technology described above. Shown in this figure is the original layout, a 3D model as built by the simulator and a temperature distribution. The power cell consists of two power transistors, with 40 and 100 mW of generated power respectively. Each transistor has an emitter heat spreader and all metalization is modeled including a NiCr resistor. The maximum temperature in the cell was at the center of the 100 mW transistor and was 105 K above the backside temperature of 355K.

An aspect of this simulation that is of interest is the thermal coupling between the two transistors. This issue can be investigated easily and effectively using a layout driven simulator, such as *Atar*, and will be

explored in the next section.

3.5 Transient Analysis

To confirm the accuracy of the TLM method a transient solution for the simple case of 3D block of Si with a height of 50 μm and a width and length of 32 μm was done. The bottom face of the block was fixed at 300 K, and a thin heat source region was situated near the top of the block. The heat source of 25 mW is centered at the block, is 8$\times$8 μm in its lateral dimensions, has a thickness of 0.2 μm and is 0.2 μm from the top face. An analytical transient solution of this problem is available (assuming temperature independent material parameters) using Green's functions. A comparison of the time response of the model and the analytical solution is shown in Fig. 15. As can be seen from both the plot of the temperature transient response and the residual error, the QTM TLM solution is very accurate over the entire response time.

Finally, a model of a Si trench device is shown in Fig. 16(a). The figure shows 1/4 of the device which has a central 1 μm x 20 μm emitter with a collector contact on each side. The technology has a PtSi/W plug contact structure and two aluminum metal layers were included in the model with a W plug via. The simulation region is 64 μm square and the wafer thickness was 100 μm. The backside temperature was 300 K and the power generated in the device was 1.75 mW, uniformly generated in a region under the emitter. The ends of the emitter and collector metalization were also fixed at 300 K. The total model consists of 15,000 blocks. A simulation of the transient response of this Si device is shown in Fig. 16(b). The temperature of point at the center of the base/collector junction is presented as a function of time.

4 Conclusion

This paper has presented a new thermal simulator for integrated circuit devices. The simulator automatically produces an accurate 3D model of the device from layout information which is appropriately meshed for solving the non-linear thermal diffusion equation. A 3D model is built up from the surface of the semiconductor surface that represents both the substrate of the device and any metalization that is present. A 3D quad tree mesh is created that produces a fine mesh at the heat generation region of the device, but still uses a moderate number of temperature nodes. From this model a network of thermal resistors and capacitors are extracted that mathematically represents the model which can be solved to obtain the temperature distribution.

The paper first compares results from the simulator with a commercial simulator FEM simulator to confirm that it accurately solves the steady state heat equation. Three example technologies are then used to display the abilities of the simulator to explore the effect of process and layout variations. A complete simulation of GaAs power cell consisting of two power transistors and heat spreaders is presented. Finally, an example of a full transient simulation of a Si trench device using 15,000 blocks was presented, displaying the full capabilities of the simulator.

Acknowledgement

This work was supported by the National Sciences and Engineering Research Council of Canada (NSERC), Manufacturing and Material Ontario (MMO) and the Networks of Centers of Excellence (Micronet).

References

[1] Richard C. Joy and E.S. Schlig, “Thermal Properties of Very Fast Transistors”, IEEE Trans. on ED, vol. 17, no. 8, August, 1970.

[2] Guang-Bo, Ming-Zhu Wang, Xiang Gui, and Hadis Morkoc, “themal Design Studies of High-Power Hetrojunction Bipolar Transistors”, IEEE Trans. on ED, vol. 36, no. 5, May. 1989.

[3] Chang-Woo Kim, Norio Goto, and Kazuhiko Honjo, “Thermal Behavior Depending on Emitter Finger and Substrate Configurations in Power Hetrojunction Bipolar Transistors”, IEEE Trans. on ED, vol. 45, no. 6, June 1998.

[4] A. Kager and J.J. Liou, “Two-dimensional Numerical Analysis of AlGaAs/GAAS Hetrojunction Bipolar Transisitors Includieng the Effects of Graded Layer, SetBack Layer and Self-Heating”, Solid-State Electronics, Vol. 39., no. 2, pp. 193-199, 1996.

[5] C. Bozada et al., “Thermal Management of Microwave Power hetrojunctioni Bipolar Devices”, Solid-State Electronics, Vol. 41, no 10, pp. 1667-1673, 1997.

[6] Xiang Gui, Paul W. Webb, and Guang-bo Gao, “Use of three-dimensional TLM method in the thermal simulation and design of semiconductor devices”, IEEE Trans. Electr. Dev., **39**(6) 1295–1302 (1992).

[7] H. Carslaw, *Conduction of heat in solids*, Oxford, 1959

[8] M. Holland, Phys. Rev. 134, A471, 1964.

[9] T. Smy, D. Walkey, S.K. Dew, “Transient 3D heat flow analysis for integrated circuit devices using the transmission line matrix method on a quad tree mesh”, Submitted to the Int. J. Numer. Eng.

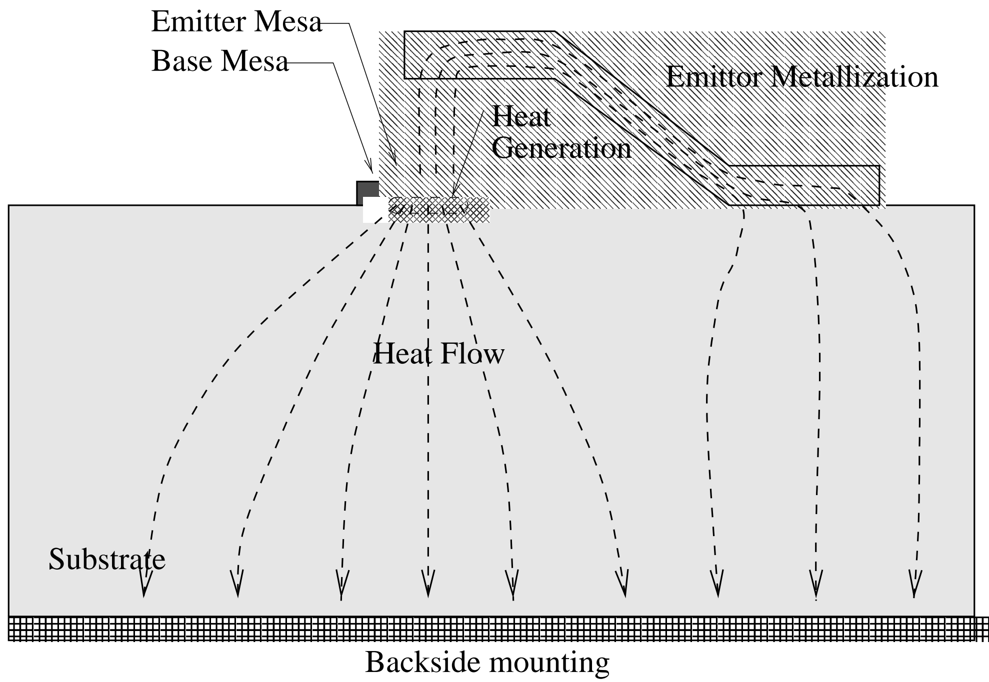


Figure 1: A schematic illustration of the geometry of the heat flow problem for GaAs power amplifier

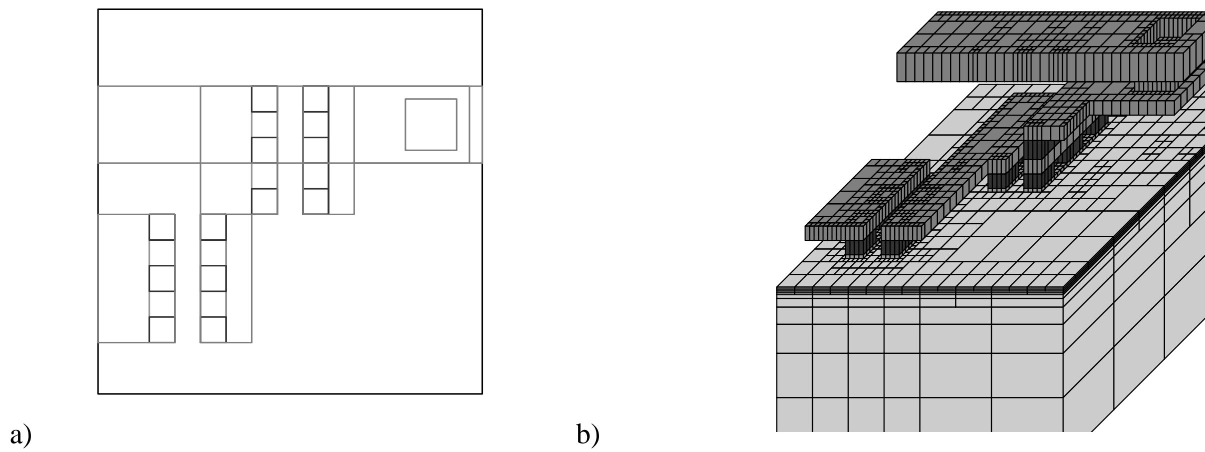
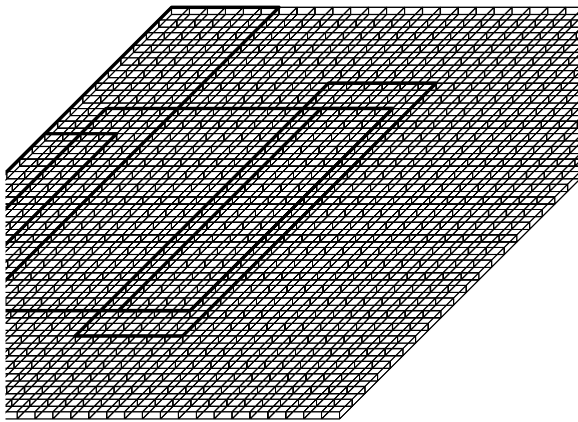
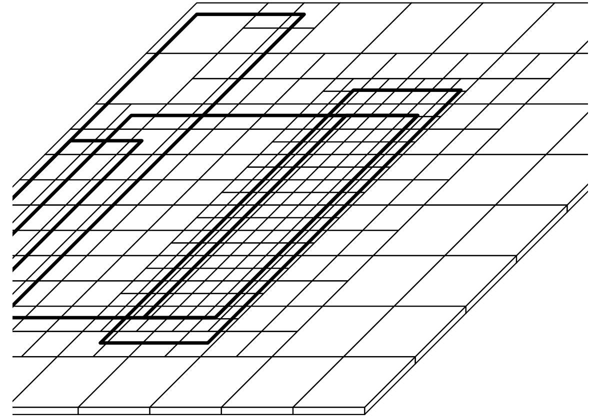


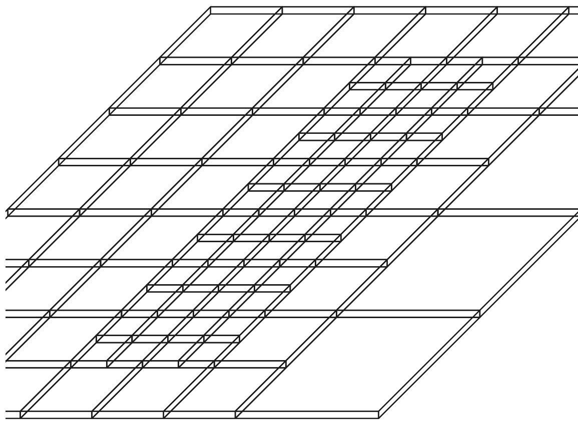
Figure 2: Layout to 3d model



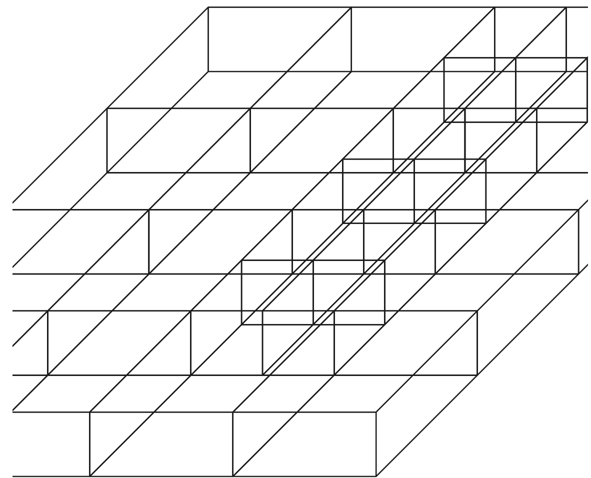
a)



b)

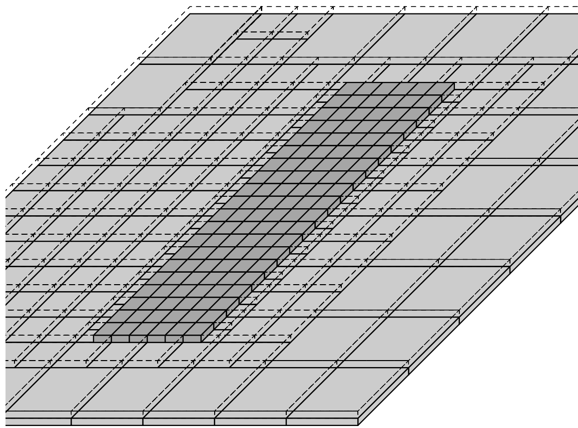


c)

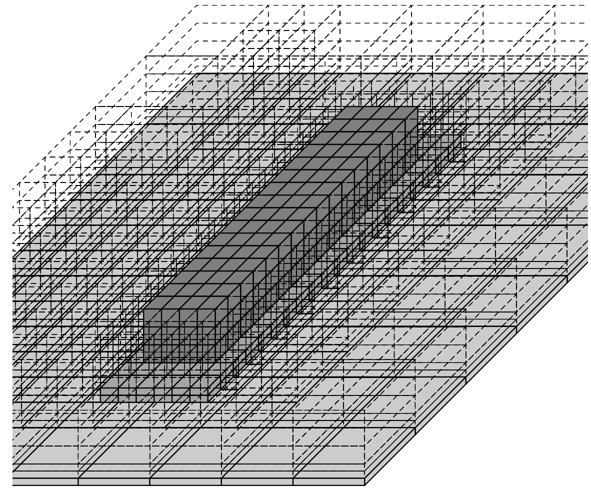


d)

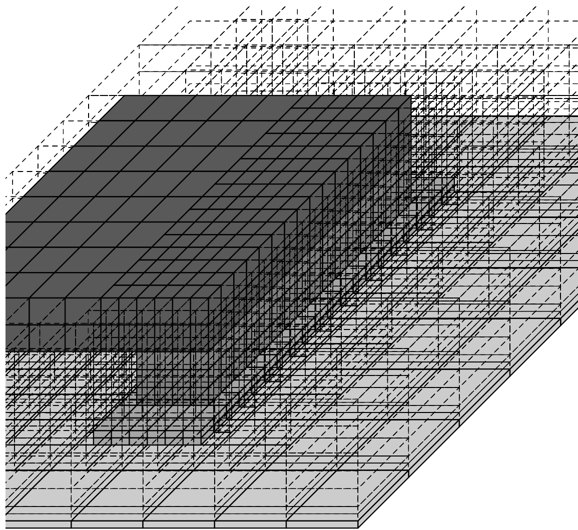
Figure 3: Substrate construction. a) The initial high density block structure showing a simple mask layout. b) The surface structure created from this high density structure after iterative coarsening of the mesh. c) A substrate layer near the surface of the substrate showing more mesh coarsening. d) A substrate layer near the bottom of the substrate showing increased thickness and further coarsening.



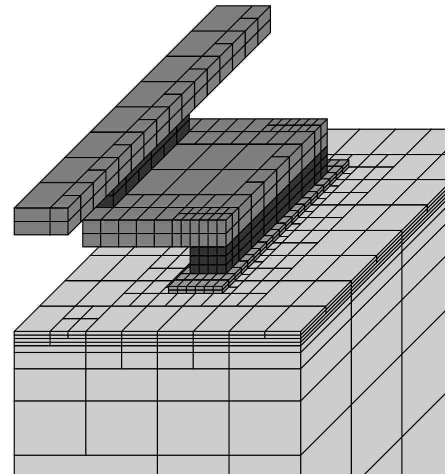
a)



b)



c)



d)

Figure 4: Backend creation and final model. a) Extrusion of the first backend layer a silicide layer. b) Extrusion of a three layer tungsten plug contact structure. c) Creation of a M1 aluminum line. d) The final model showing the substrate and completed backend of a simple emitter structure.

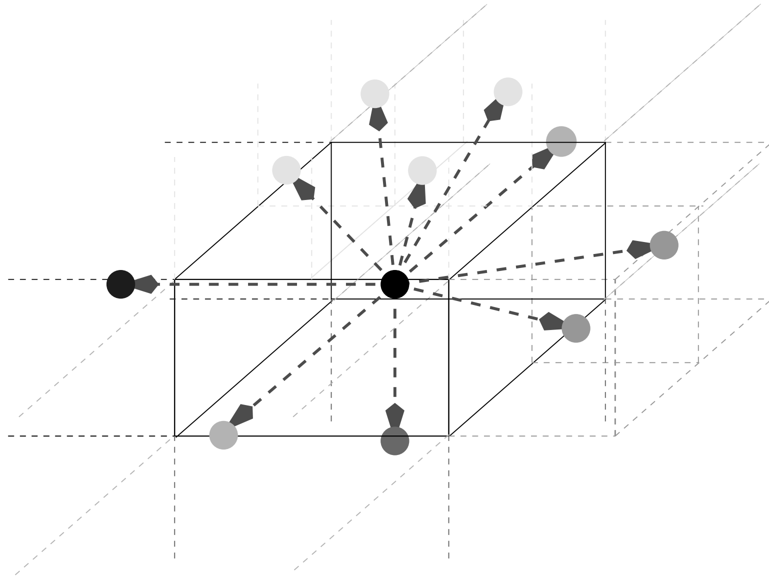


Figure 5: Heat flows between blocks represented by thermal resistors

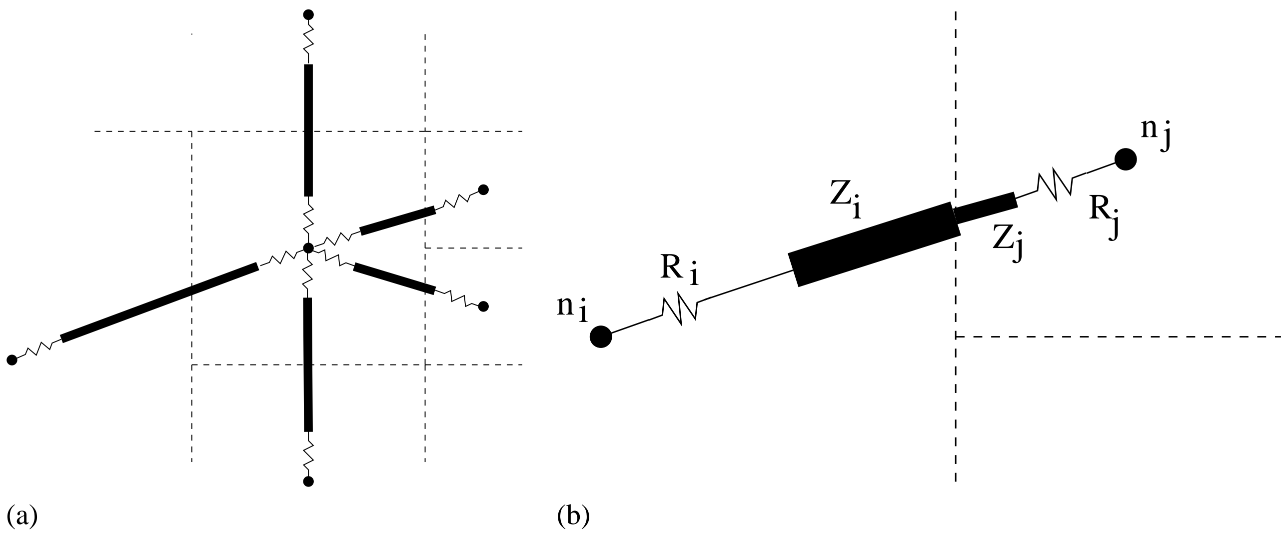
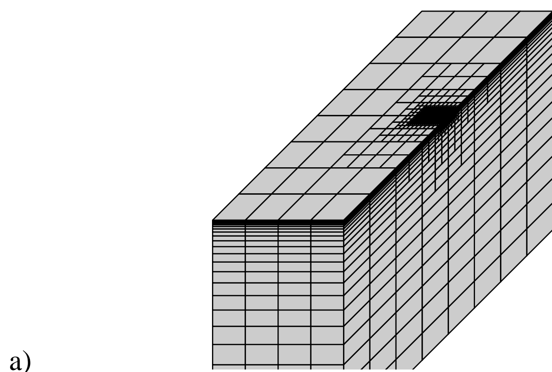
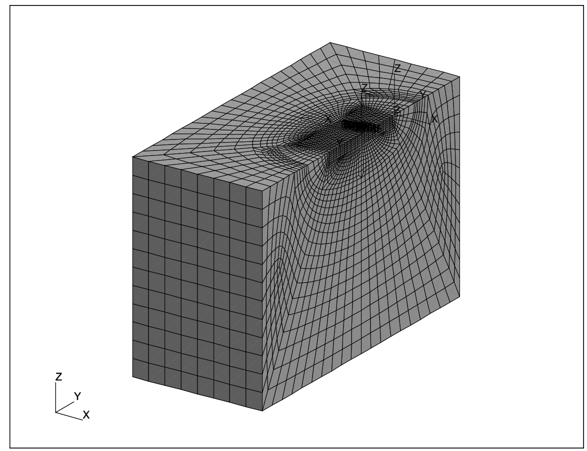


Figure 6: (a) TLM transmission line links in a quad tree and (b) details of a single link in the quad tree mesh.

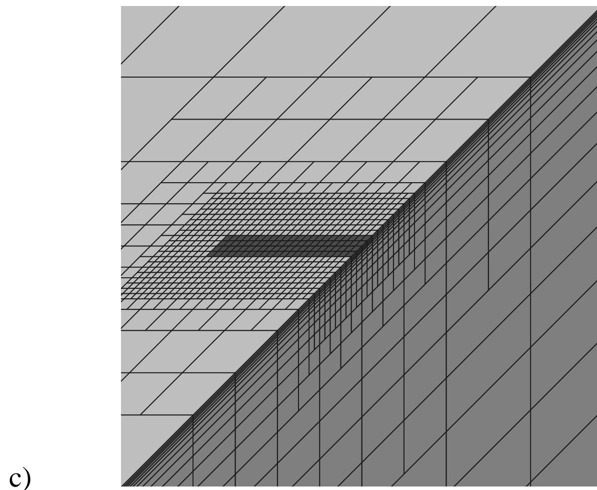
a)



b)



c)



d)

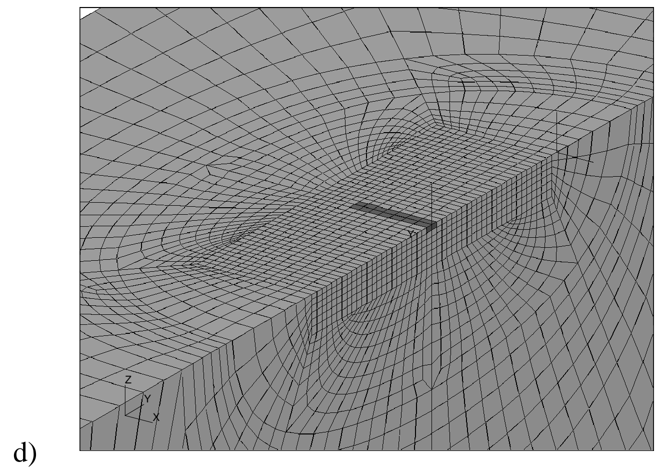


Figure 7: Typical meshs for a simple heat source in a substrate region. a) *Atar* and b) Patran. Details of the heat source region c) *Atar* and d) Patran

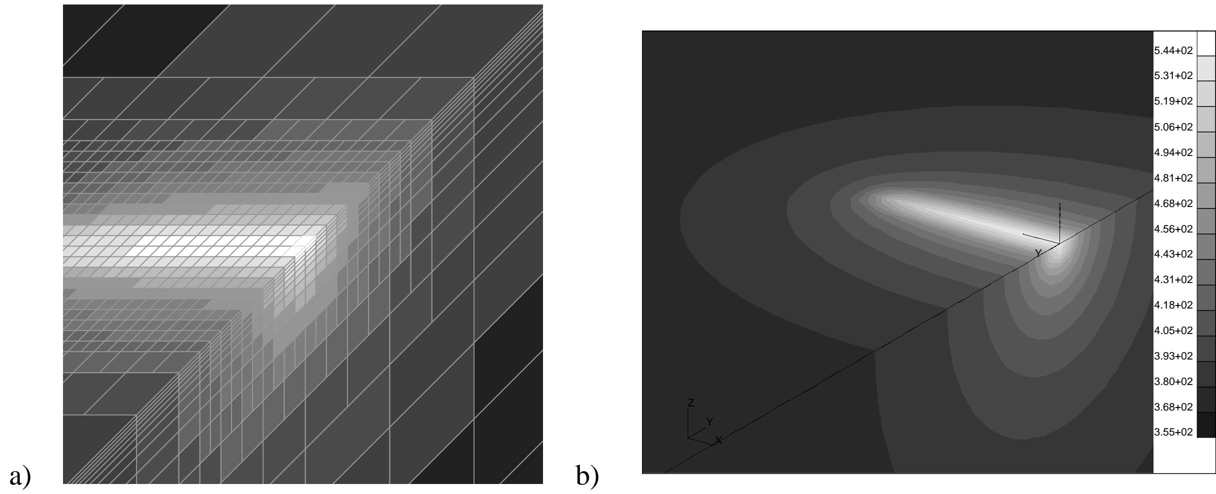


Figure 8: Temperature distribution for a rectangular heat source in a GaAs substrate with a fixed 355 K backside temperature. (a) *Atar* (d) PATRAN.

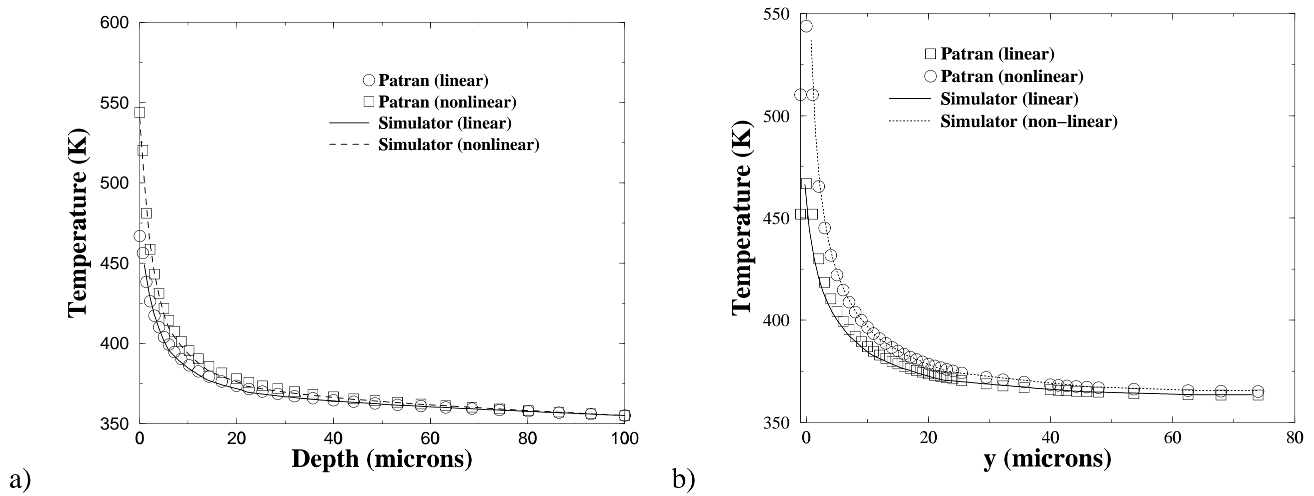


Figure 9: Temperature profiles. a) From the emitter down into the substrate. b) Along the surface.

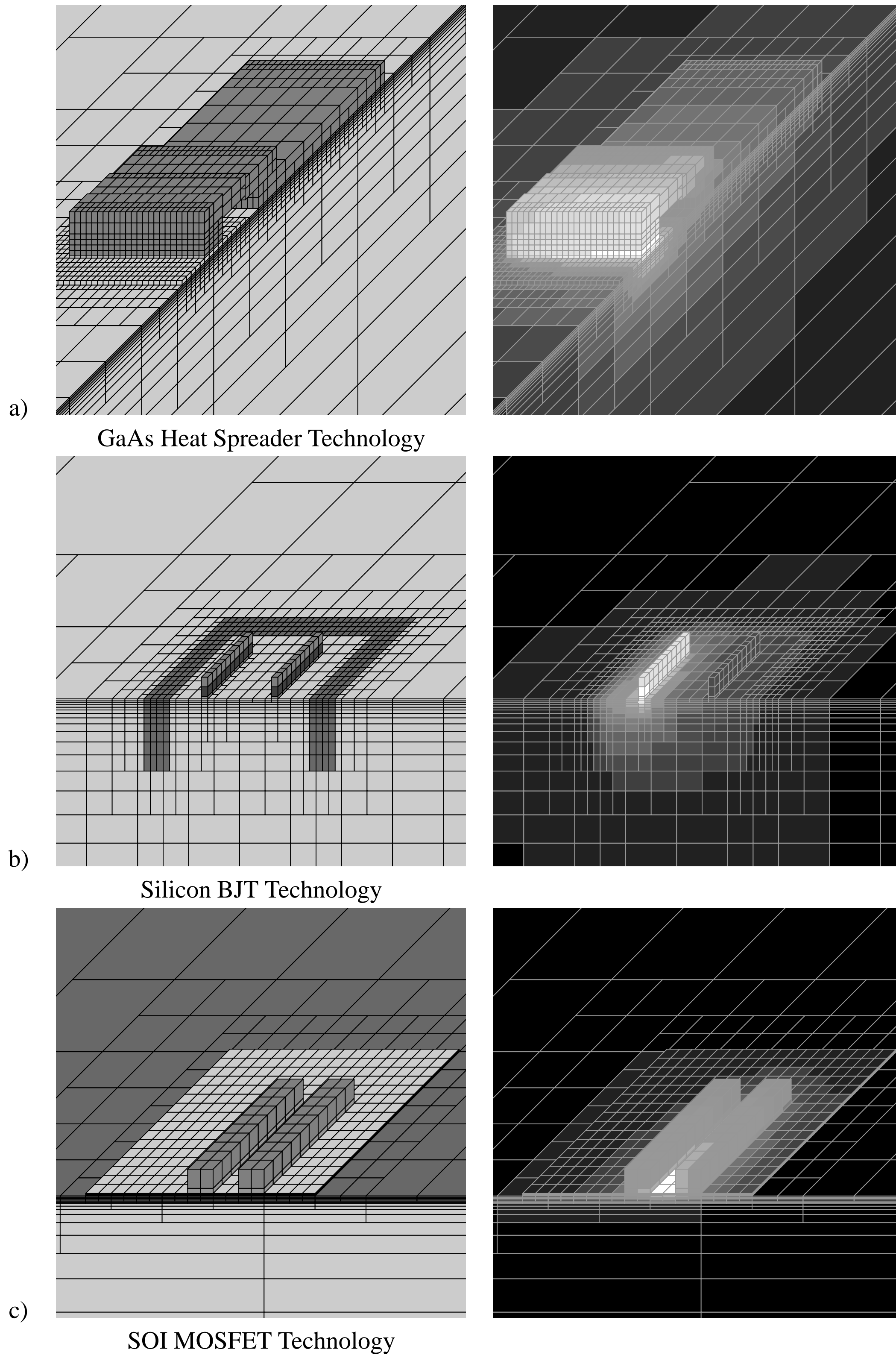


Figure 10: Example 3D *Atar* models and temperature distributions. (a) GaAS heatspreder. (b) Si trench. (c) SOI

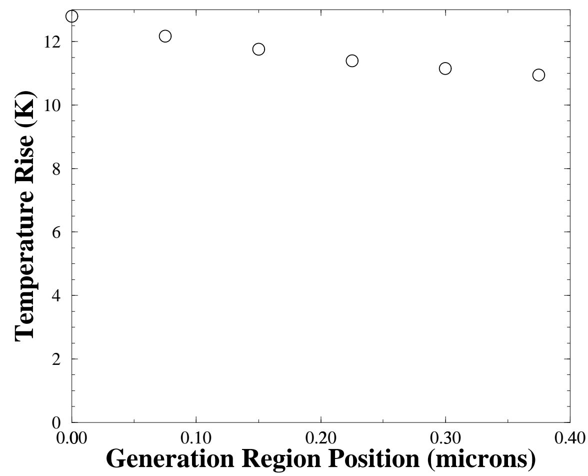
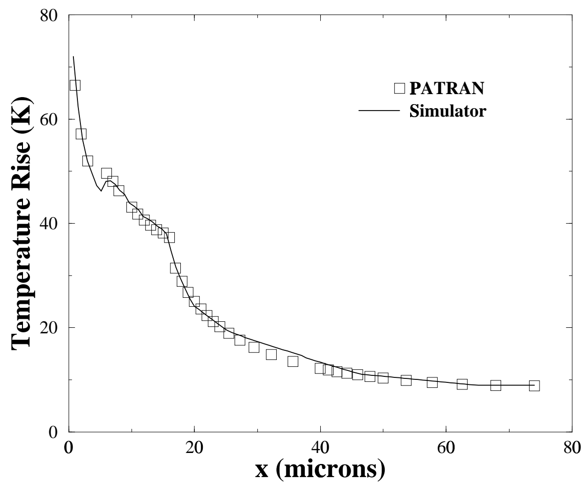


Figure 11: Si trench device operating temperature as a function of the position of the heat generation region.



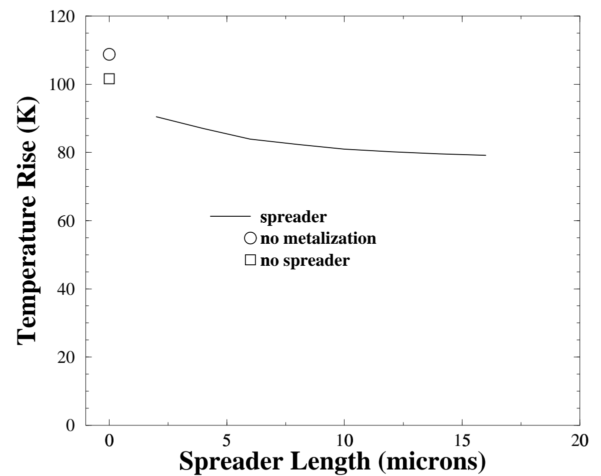


Figure 12: GaAs HBT device. a) Surface temperature. b) Device temperature as a function of spreader length.

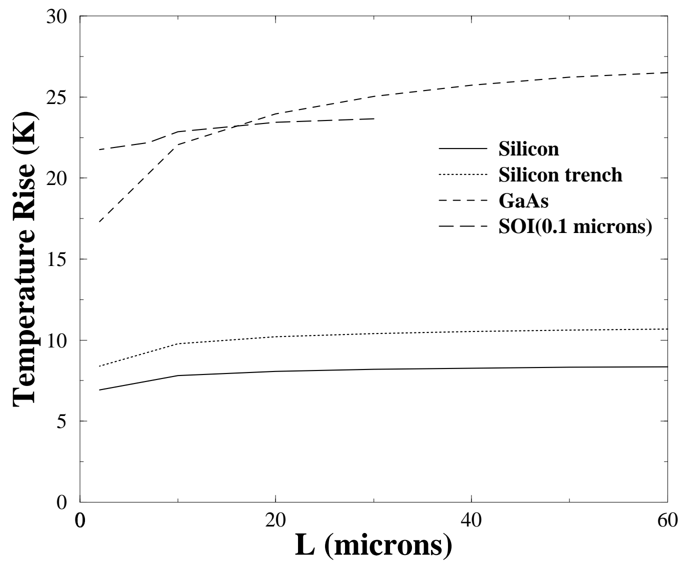
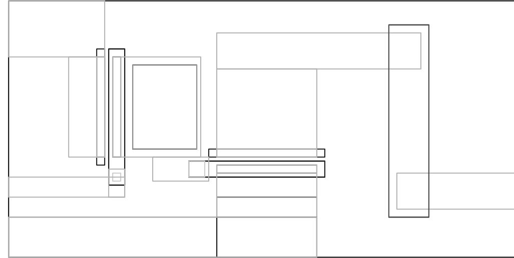
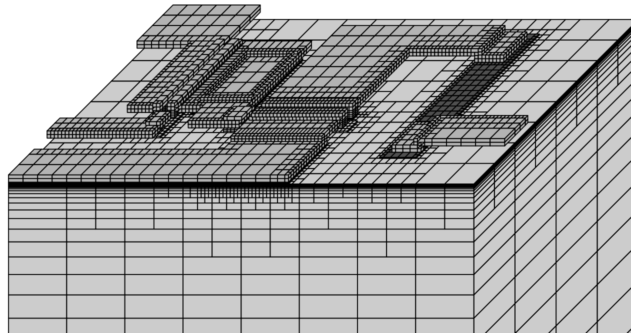


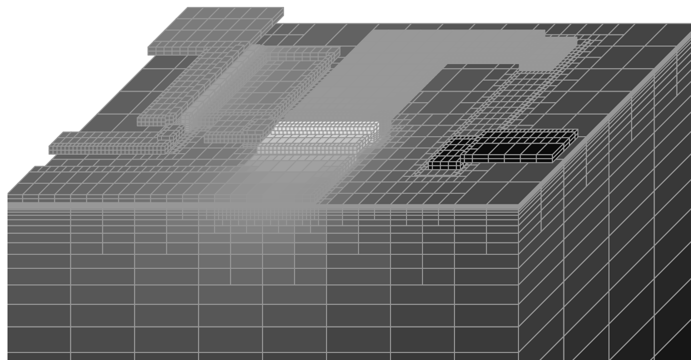
Figure 13: Device temperature sensitivity to back end heat flow.



a)



b)



c)

Figure 14: Example 3D Gaas Power Cell with metallization. a) Layout. b) Model. c) Temperature profile.

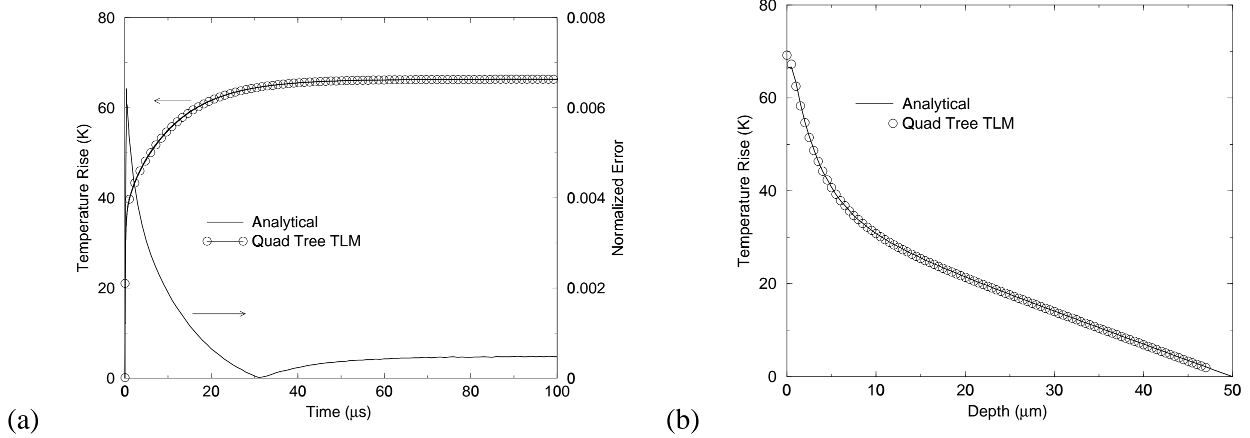


Figure 15: Temperature response within a 3D rectangular solid showing (a) transient response of maximum temperature and (b) temperature profile as a function of depth near the steady state.

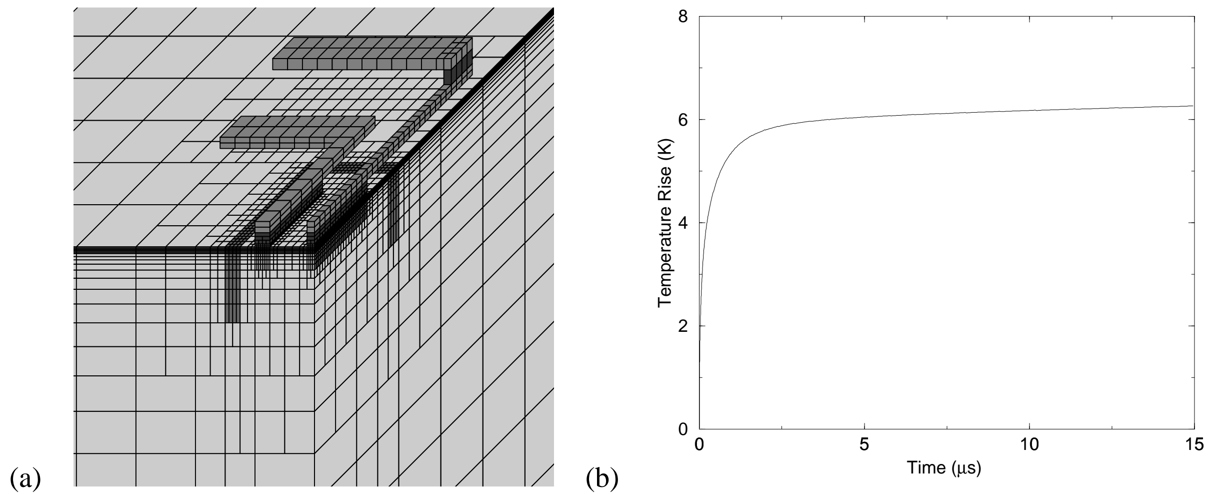


Figure 16: (a) Details of a Si trench device showing structure and metalization and (b) transient thermal response of the device.

Response to the Review

In response to the editors comments and reviewers remarks for manuscripts 921 and 941:

- 921: “A 3D thermal simulation tool for integrated devices – *Atar*”, **T. Smy, D. Walkey**
- 941: “Transient 3D heat flow analysis for integrated devices using TLM on a quad tree mesh”, **T. Smy, D. Walkey, S. Dew**[1]

We wish to first thank the reviewers for their constructive comments. It was suggested that 921 and 941 should be combined and some material removed from 921 (reviewer 921:#1, point 2). We have taken this approach and produced a new manuscript. However, 941 was a fairly long and involved description of a new implementation for TLM on a 3D quad tree mesh. To allow the new manuscript to be of reasonable length we have added to 921 a brief description of the TLM transient solution (referencing a submitted paper on the details) . This has allowed us to add some material to the results section containing the most useful results on transient analysis from 941. We have also, as suggested by reviewer 921:#1, removed a number of figures from 941 and condensed the discussion.

A number of points were specifically brought up by the reviewers which we will now address.

- Electro thermal effects (921:#1-point2, 921:#2-point1, 941:#2-point2)

  The reviewers point out that the electrothermal (ET) interaction will effect device operation and that is very true. We have designed the code with the intention of incorporating ET effects into the thermal calculations. Work is still ongoing and proving to be a rich area of investigation. The results presented here are still relevant at moderate power levels and clearly illustrate the potential of the tool.

- Temperature Interpolation (921:#2-point6, 921:#2-point3)

  The plotted temperatures were interpolated from the 3D temperature distribution given by the node temperatures.

- Applications (921:#1-point1)

  Currently the tool is primarily for internal use. We are in the process of developing interfaces with standard VLSI design tools such as Cadence. If interest is sufficient this work will be accelerated.

- Fig 2. (921:#2-point2) – Fig. 2 is now referenced in the manuscript.
- Fixed T BC’s (921:#2-point3)

  The use of a fixed boundary condition on vertical sides of the simulation region causes an artificial constraint on the temperature profile. This can be overcome by using a very large region of simulation but that is computationally expensive. A better solution is to use zero flow boundary conditions and be aware that this implies mirroring at that boundary.

- Focus on BJT's (921:#2-point5)

  The focus on BJT's was primarily due to the authors industrial experience with high power BJT's where self heating is a concern. As the paper is essentially illustrating the various uses of the tool the use of BJT structures seems appropriate. The tool itself is of course not limited to BJT's and indeed we are currently employing it to look at large power MESFET structures.

- Thermal conductivity (941:#2-point1) – The units for thermal conductivity were mistyped using "cm" not "microns" accounting for the factor of 10,000.

A number of other points were brought up by the reviewers, however, they either dealt with material not incorporated from 941 into the new paper or material removed from 921.

Thanks for you consideration.