Self-Consistent Modeling of Heating and MOSFET Performance in Three-Dimensional Integrated Circuits

Akin Akturk, Neil Goldsman and George Metze

Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742, USA

{akturka, neil, metze}@glue.umd.edu

Abstract

We present a new method for finding the temperature profile of vertically stacked three-dimensional (3D) digital integrated circuits (ICs), as shown in Fig. 1. Using our model, we achieve spatial thermal resolution at the desired circuit level, which can be as small as a single MOSFET. To resolve heating of 3D ICs, we solve non-isothermal device equations self-consistently with lumped heat flow equations for the entire 3D IC. Our methodology accounts for operational variations due to technology nodes (hardware: device), chip floor plans (hardware: layout), operating speed (hardware: clock frequency) and running applications (software). To model hardware, we first decide on an appropriate device configuration. We then calculate elements of lumped thermal network using 3D IC layout. To include software, chip floor plan and duty cycle related performance variations, we employ a statistical Monte Carlo (MC) type algorithm. In this work, we investigate performance of vertically stacked 3D ICs, with each layer modeled after Pentium III [1]. Our calculated results show that layers within the stacked 3D ICs, especially the ones in the middle, may greatly suffer from thermal heating.

Keywords: 3D IC Heating, Chip Heating, Non-Isothermal MOSFET Performance, Lumped Thermal Analysis, Heat Flow.
As the industry pushes the down-scaling of devices further to increase the speed of devices, and consequently circuits and chips, a challenge in IC operation has emerged: interconnect and input/output (I/O) delays. This is especially evident where systems require multiple integrated circuits that communicate through printed circuit boards, I/O pads and bond wires. To alleviate the problem, manufacturers are investigating the development of 3-dimensional integrated circuits (3D ICs). 3D designs can obviate the need for many I/O pads, bond wires, package pins and PCB interconnects. Additionally, 3D designs offer substantial real estate gains. However, while chip heating has become a large problem for standard planar integrated circuits [1]-[10], the heat problem becomes exacerbated for 3D IC’s.

High device densities are causing increased temperatures on chip due to elevated power densities. According to traditional device scaling, when device dimensions are scaled downward by a factor S, all other parameters are scaled by the same factor, either downward (physical features, supply voltage…) or upward (frequency and capacitance per area…), in order to maintain a fixed power density per unit area. However, as dimensions become smaller, manufacturers must deviate from this, and especially from voltage scaling, because of intrinsic limitations of silicon bandgap and built-in voltages [1]-[6]. The result is higher power densities because of higher clock frequencies and supply voltages. Also, isolation between supply rails gets smaller in nano-devices, leading to higher leakage levels. In addition, silicon dioxide (SiO₂), which acts like a thermal and electrical insulator between stacked chips in a 3D IC, aggravates heating problem by greatly restricting the flow of heat generated. The main result is increased thermal resistance and power density, leading to higher chip temperatures—temperatures higher than conventional cooling methods can account for. Thus, as feature sizes shrink, the power density is increasing exponentially, demanding a focus on heating and cooling of 3D ICs and planar chips if this barrier is to be overcome [1]-[10].
A good approach to understanding a complex problem is to develop simulators to mimic the problem. For the chip heating problem, the simulator should predict localized and overall chip heating for a given 3D IC architecture. It should also assist in developing alternate IC layouts that could help keep localized temperatures low. A foundation has already been established for estimating chip temperatures [7]-[10]. In this paper, we bring to light the need for a simulator that can connect individual device operations with heating of 3D ICs. Since there can be over a billion devices on a 3D integrated circuit, it is a challenge to calculate the details of device and chip heating simultaneously. Here, we present a method to achieve this connection. First, by self consistently solving coupled quantum and semiconductor equations, we find the electrical characteristics of an n-MOSFET. Next, we take each device on a 3D IC as a cell and model the thermal connections between devices using a lumped circuit type thermal network of thermal resistances, capacitances, and heating sources. From the architectural aspects of the chip layout, we determine the values of the thermal resistances and capacitances in the network. Since the heating source for each device is the driving force in the thermal network, we incorporate the results of the individual MOSFET operations in these thermal elements. We do this using a Monte Carlo type algorithm, which allows us to realize the goal of connecting 3D IC heating to individual device operations. Finally, we suggest chip design solutions for cooling the warmest areas of a chip. We present our device, IC levels, and their collective relation in Fig. 2.

II- Device Performance and Full-Chip Heating Model:

We self-consistently solve device performance and full-3D IC heating equations. We first obtain device performance at different temperatures by solving the semiconductor equations along with the Schrödinger equation. Second, we achieve heating figures of vertically stacked 3D ICs by solving a lumped thermal network in conjunction with device performance results and averaged operational statistics.
A. Device Performance

We developed a device simulator that is capable of solving the coupled quantum and semiconductor equations. Pertinent details of the device simulator are given in the Appendix.

Using our simulator, we first investigate the temperature profile within a single MOSFET. Our analyses indicate that temperature variation within a bulk MOSFET channel is small, unless it is a Silicon-On-Insulator (SOI) device. The lattice temperature inside a bulk MOSFET differs only a few percent from the value at the boundary.

We next investigate the effects of parameters including electron and hole mobility, electron and hole saturation velocity, built-in potentials, intrinsic carrier concentration, bandgap of silicon, and the thermal diffusion constant. Our analyses indicate that non-isothermal MOSFET performance (within chip temperature operating limits) is mostly affected through carrier mobility, saturation velocity and built-in boundary potentials. As temperature increases, mobility and saturation velocity decrease, resulting in lower current values. However, change in built-in boundary potentials results in effective threshold voltage lowering as temperature rises, which increases current. For temperatures higher than the upper operating limit of most of today’s devices, such as 100°K above the ambient, MOSFET performance degrades due to chip currents increasing exponentially. This occurs when there is an abundance of carriers due to thermal excitations (intrinsic carrier concentration increases exponentially with temperature).

B. 3D IC Heating Model

We developed a lumped thermal network model based on the differential heat flow equation (A6) to obtain the temperature profile of vertically stacked 3D ICs. In our model, we account for the 3D IC’s layout and floor plan, and the chip transistors’
performance details including heat generated, duty cycle and averaged operational statistics.

Large differences in the scales of an entire 3D IC and a single transistor necessitate use of a lumped thermal network model [7, 17]. To obtain a lumped thermal network model, we first apply the transformation [17] given in (A14) to (A6) to move the space dependent thermal diffusion constant outside the gradient term, and to replace it with a fixed thermal diffusion constant, which is evaluated at room temperature, $\kappa(T_0)$. The result is a modified differential heat flow equation as shown below:

$$\frac{\bar{C}}{\bar{C}} \frac{\partial \bar{T}}{\partial t} = \kappa(T_0) \nabla^2 \bar{T} + H$$  \hspace{1cm} (1)

We then integrate (1) around a single device, which is bounded by a rectangular prism with volume $V$ and surface $S$, and has a single temperature value at its center associated with it. We also note that $-\kappa \nabla \bar{T}$ is the heat flux. Moreover, evaluation of the integrals yields:

$$\bar{C}V \frac{\Delta \bar{T}}{\Delta t} + \sum_{f=1}^{6} \kappa_f \Delta \bar{T}_f S_f = \int \bar{H} dV$$  \hspace{1cm} (2)

This gives a lumped relation for the temperature of a device in relation to the six neighbors in the direction of the six faces, $S_f$, of the enclosing rectangular prism, with separation and temperature difference between different nodes denoted by $l_f$ and $\Delta \bar{T}_f$, respectively.

Equation (2) is analogous to a KCL type nodal equation with capacitive, resistive and source components. Taking $\bar{T}$ analogous to voltage, we can write equivalent thermal resistance, capacitance and current source as shown below:

$$C^h = \bar{C}V$$  \hspace{1cm} (3)

$$R^h_f = \frac{\Delta l_f}{\kappa_f S_f}$$  \hspace{1cm} (4)

$$I = \int_{V} \bar{H} dV$$  \hspace{1cm} (5)
Next, we determine the values of thermal resistances and capacitances for each of the many (several hundred million) 3D IC transistors from the layout and geometrical considerations. Moreover, we use statistics to obtain a value for the thermal current source of each of the 3D IC transistors using a Monte Carlo (MC) type methodology.

After we obtain thermal capacitances, resistances, and non-isothermal device performance figures, and decide on an appropriate MC methodology, we determine the temperature of each transistor on the 3D IC, represented by \((i,j,k)\), by solving KCL-type equations like the following:

\[
C_{th}^{i,j,k} \left( \frac{T_{i,j,k}^{l} - T_{i,j,k}^{l-1}}{\Delta t} \right) + R_{i,j,k}^{th} \frac{T_{i,j,k}^{l} - T_{i-1,j,k}^{l}}{R_{i,j,k}^{th}} + R_{i,j,k}^{th} \frac{T_{i,j,k}^{l} - T_{i,j,k+1}^{l}}{R_{i,j,k}^{th}} + R_{i,j,k}^{th} \frac{T_{i,j,k}^{l} - T_{i,j,k-1}^{l}}{R_{i,j,k}^{th}} = I_{i,j,k}^{l} \left( T_{i,j,k}^{l-1} \right)
\]

(6)

Here, \(\pm \frac{1}{2}\) in the subscript gives the resistance between nodes in the given direction. Furthermore, \((i,j)\) represents a device within a layer \(k\). The superscript \(l\) shows the iteration number for our numerical solver.

### III- Coupled Device Performance and 3D IC Heating Model:

#### A- Coupled Algorithm:

To obtain the temperature map of 3D ICs, we solve self-consistently lumped thermal network equations for the entire vertically stacked 3D IC in conjunction with device performance details. These details include non-isothermal device performance figures including current-voltage characteristics, and operational statistics such as duty cycle and functionality. Therefore, we achieve convergence at the device level and the 3D IC level as described below in our coupled algorithm:

**1- Obtain device performance as a function of temperature**

For a given vertically stacked 3D IC, we first find the technology node used for fabrication. We determine the average dimensions of a typical transistor on the chip. (We use a MOSFET as our unit cell, but fundamental logic gates such as an
inverter can also be used instead.) We then input our representative device in our device simulator. We also decide on typical bias conditions and average on-power during switching for that particular digital IC to adjust total Joule heating for one clock cycle. To obtain device performance including current-voltage characteristics and heat generated at different temperatures, we solve quantum device equations, and prepare a look-up table.

2- *Fit device performance results to a polynomial*

We obtain a heat generated, $H$, versus temperature, $T$, curve. Since our KCL-type equations for the lumped thermal network are derived after we applied the transformation (A14) to the differential heat flow equation as described in section II-B, we also produce a heat generated, $H$ versus transformed temperature, $\bar{T}$, curve. We then fit the $H$ vs. $\bar{T}$ curve to a second-order polynomial and obtain an analytical expression for their relationship.

3- *Set spatial resolution for the 3D IC*

We next focus on the geometry of the 3D IC. We first set the spatial resolution in accordance with the average size of the 3D IC’s transistors. We then determine the thermal link between devices by defining the thermal resistances, $R^\text{th}$, and thermal capacitances, $C^\text{th}$, in conjunction with the 3D IC’s layout and device architecture. Thus, we obtain values for all the lumped thermal elements except the current sources shown in Fig. 2. The strengths of the current sources are related to the heat generated by each transistor on the 3D IC. Therefore, we find their actual values along with the temperature of each device at the end of our mixed-mode simulation.

4- *Determine effects of 3D IC’s floor plan, and software application on performance*

To embed effects of 3D IC’s floor plan on performance, we group transistors in each layer into a few functional blocks such as cache, floating point unit, execution unit, clock, etc., as shown in Fig 1b. Next, to embed the effects of the typical software applications on IC performance, we determine consumed
percentage power for each functional block in that layer. Later, to obtain the activity level of a transistor within a functional block relative to one within another functional block, we normalize these percentage powers by the corresponding areas of each block. We then renormalize these percentage power per areas by the maximum for that particular layer.

5- Statistically extend effects of operational device variations to the entire 3D IC

To extend the effects of operational device variations to the entire 3D IC, we employ a statistical Monte Carlo-type methodology. We first generate a random number for each transistor as a function of the calculated normalized percentage power per area corresponding to that device. We then assign this calculated random number to the corresponding 3D IC’s transistor as an indicator of the likelihood of the full power that the particular device is consuming on average. This procedure is applied to each transistor in the 3D chip. In essence, we statistically determine the relative power consumed by each transistor in the IC.

6- Compilation of data

At this point, we know the following:

- Device performance details including heat generated \( H \) versus transformed temperature \( \tilde{T} \) curve as well as an analytical expression for a second order polynomial fit,
- 3D IC geometry and layout dependent thermal resistances and capacitances between 3D IC’s transistors, and devices and ambient,
- Statistically determined normalized powers for each transistor that are obtained using the given 3D IC floor plan and the typical running application on that 3D IC.

7- Mixed-mode solution

We now can solve the KCL-type lumped thermal network equations given in (6). From layout, we know the coefficients of the temperature \( \tilde{T} \) on the left hand side of (6). We also know the steady-state heat generated as a function of temperature.
Moreover, we know the percentage of the heat generated consumed by each transistor. We have as many equations as the number of transistors on the 3D IC. Each equation is non-linear due to the second-power dependency of heat generated on temperature. To solve, we first assign the heat generated at room temperature to all nodes (devices). We then use a preconditioned bi-conjugate gradient solver to obtain nodal temperatures. We next update the heat generated of each transistor in conjunction with its calculated temperature value. To get a self-consistent solution for the full-3D IC temperature, we iterate temperature and heat generated until convergence. The solution gives the temperature map of the 3D IC as well as the heat generated of each device.

For easy reference, we summarize our algorithm in Fig. 3.

**B- Application and Results**

After establishing our methodology, we test it on hypothetical digital 3D ICs that have layers modeled after a Pentium III, as shown in Fig. 1. We take 0.13µm as the technology node for that chip, and model a device after [18]. We then obtain device performance and heat generated as a function of temperature. We next determine the thermal network associated with this 3D IC, representing a single transistor by a thermal node. We last obtain nodal temperatures (temperatures of each transistor on the 3D IC).

To obtain device performance as a function of temperature, we simulate a 0.13µm N-MOSFET with drain-to-source and gate-to-source biases of 1.5V, at different temperatures, by solving the device equations given in the Appendix. We then fit the device performance results to a polynomial function. We also weight the calculated steady state powers by the percentage of the on-power during switching.

We next set spatial resolution for our 3D IC by taking a single transistor as our unit cell. Consequently, we have roughly forty million devices in each layer of about one square centimeter. To simplify the problem, we take the 3D IC’s transistors to be laid out
uniformly in each layer. Using the 3D IC’s layout and package details, our calculations yield thermal resistances of 25 K/W between nodes in the same layer, and $5 \times 10^5$ K/W between nodes in the vertical direction between layers, respectively.

We next work on the solution of this thermal network, which consists of forty million nodes (corresponding to all transistors) in each layer. To make the problem tractable, we reduce the associated number of equations while increasing the bandwidth of the connectivity matrix that defines the connection between each node (A regular node in a 3D rectangular mesh is connected to six other nodes.). To achieve this, we replace sub-blocks in each layer by their Thevenin equivalent circuits, reducing the size of the system of equations. We first enclose a sub-block of twenty two by twenty two nodes in each layer, a half resistance away from the outer nodes. We then short its borders on each side yielding six new nodes, We next obtain the six-port Thevenin equivalent circuit [19] seen from these nodes, with equivalent thermal resistances, a capacitance and a heat source attached to each. As shown in Fig. 4b (a), the resulting graph for 3D (2D) has tetrahedral shape (diamond) unit cells, where each node has explicit connection to eight (six) other nodes as opposed to six (four) other nodes in the rectangular grid. This reduces the number of simultaneous equations that we solve from approximately 200million to a more tractable 3million.

We then extend our calculated heat generated results to the 3D IC volume using a Monte Carlo (MC) type methodology. We use an MC algorithm to statistically determine each equivalent node’s source strength. Our MC algorithm makes use of the floor plan shown in Fig. 1b with percentage powers and areas given in Table 1. After we set up our thermal network including the source components, we solve the reduced system of equations for nodal temperatures using a bilateral conjugate gradient method.

In Fig. 5, we show steady state device performance figures including current-voltage characteristics and heat generated as a function of temperature. Figure 5a indicates that as temperature increases, current decreases both in the linear and saturation regions. This is in accordance with the downward slope of the heat generated versus temperature curve,
as shown in Fig. 5b. (We note that temperatures calculated have not gone beyond device operating limits where intrinsic carrier concentration approaches that of the doping.)

In Fig. 6a, we show a five layered vertically stacked 3D IC with a Pentium III chip in each layer. Our calculated temperature map for the middle layer of that 3D IC is shown in Fig. 6b. We note the dramatic increase for the peak temperature value, as shown in Fig. 6c, from one layered, 320°K, to five layered, 420°K, configuration. We attribute this to the low thermal diffusion constant of the SiO₂, which traps heat in sandwiched layers. Thus, maximum 3D IC temperature, as well as the peak temperature of the bottom layer, increases as we increase the number of layers in a stacked 3D IC configuration, as shown in Fig. 6c. In addition, we also note that the location of the peak temperature moves from the clock block in one-to-three layered 3D ICs to the issue unit in a five layered 3D IC, as shown in Fig. 6b (in relation to the layout shown in Fig. 1b). We associate this with the increase of equivalent thermal resistance with stacking for each node. Moreover, high temperature variations on a 3D IC are likely to have detrimental effects on device and circuit operations. For example, temperature related phase delays may result in the failure of synchronous circuit operation. In Fig. 6d, we show the oscillation frequency of a thirty one stage ring oscillator as a function of temperature. This shows that if such a circuit is used as a clock generator for each layer, how much the speed of each layer will deviate from the others, even tough they all have the same operating frequency corresponding to the room temperature when the 3D IC is first turned on.

The temperature map of a 3D IC can also be used in conjunction with computer aided design (CAD) tools to relieve problems related to hot spots and high temperature gradients on the chip. To achieve this, chip floor plans can be rearranged to distribute active units over the whole volume. Additionally, thermal contacts can be utilized to pull high temperatures to low at problematic regions. We test the effects of perfect thermal contacts (shorts to ambient) on a layer that has the temperature profile given in Fig. 6b. Utilization of one thermal contact near the peak temperature location pulled the maximum temperature couple of degrees, however, an array of ten by ten thermal contacts pulled the peak temperature about fifty degrees down.
IV- Conclusion:

We present a new method for finding the temperature profile of complex digital 3D ICs. Using the new methodology, we achieve a spatial resolution of a single device. We first obtain device performance figures such as heat generated as a function of temperature. We then calculate values for thermal lumped elements using the 3D IC geometry. After extending our device results to each transistor on the 3D IC using an MC type algorithm, we iteratively solve for nodal temperatures to obtain the thermal map of the 3D IC in conjunction with each transistor’s performance. Details of our algorithm can easily be modified for other planar (2D) or 3D ICs with different designs and operating conditions. Knowing potential hot spots can facilitate new design strategies for 3D ICs that are less susceptible to thermal damages. It can also offer new floor plans and ways to monitor effects of thermal contacts.
We developed a device simulator that is capable of solving the coupled quantum and semiconductor equations. To add quantum corrections to the calculation of the density of electrons in the MOSFET channel, we solve the Schrödinger equation to determine band splitting. Below is a set of our quantum semiconductor equations in the order of the Schrödinger, density, Poisson, electron current continuity, hole current continuity, and lattice heat flow equations.

\[
E_i \psi_i(y) = -\frac{\hbar^2}{2m^*} \frac{d^2 \psi_i(y)}{dy^2} - q\phi(x,y)\psi_i(y) \tag{A1}
\]

\[
n = \frac{m^* kT}{\pi \hbar^2} \sum_i |\psi_i|^2 \ln \left(1 + e^{(E_F - E_i)/kT}\right) \tag{A2}
\]

\[
\nabla^2 \phi = -\frac{q}{\varepsilon} (p - n + D) \tag{A3}
\]

\[
\frac{\partial n}{\partial t} = \nabla \cdot (-n\mu_n \nabla \phi + \mu_n V_{th} \nabla n) + GR_n \tag{A4}
\]

\[
\frac{\partial p}{\partial t} = \nabla \cdot (p\mu_p \nabla \phi + \mu_p V_{th} \nabla p) + GR_p \tag{A5}
\]

\[
C \frac{\partial T}{\partial t} = \nabla \cdot (\kappa \nabla T) + H \tag{A6}
\]

In (A1)-(A6), along with familiar parameters we introduce \(E_i, \psi_i, E_F, \phi, n, p, D, GR_{n-p}, T, H, C\) and \(\kappa\), which are the sub-band energies, wave functions, Fermi level, electrostatic potential, electron concentration, hole concentration, net dopant concentration, electron-hole net generation-recombination rates, lattice temperature, heating, heat capacity and thermal diffusion constant, respectively.

We resolve non-isothermal device behavior by solving (A6), the differential heat flow equation. Here, the heat generated within the MOSFET channel is calculated finding the Joule heating, which is a function of electrostatic potential, and electron and hole
concentrations \((H = -J \nabla \phi)\), where \(J\) is the total current density). We then explicitly let the parameters in (A7)-(A13) change value with the current temperature [11]-[13] in reference to their room temperature, \(T_o\) (300° K) values.

\[
\mu(T) = \mu(T_o) \left( \frac{T}{T_o} \right)^{-2.5}
\]

(A7)

\[
\nu_{sat}(T) = \nu_{sat}(T_o) \left( \frac{1 + e^{-T/2T_o}}{1 + e^{-T/2T_o}} \right)
\]

(A8)

\[
\phi_{built-in}(T) = V_{TH}(T) \ln \left( \frac{n}{n_o(T)} \right)
\]

(A9)

\[
V_{TH}(T) = V_{TH}(T_o) \left( \frac{T}{T_o} \right)
\]

(A10)

\[
n_o(T) = n_o(T_o) \left( \frac{T}{T_o} \right)^{1.5} e^{\left( \frac{-E_g(T)}{2kT} \right) \left( \frac{1}{\frac{T}{T_o} E_g(T)} \right)}
\]

(A11)

\[
E_g(T) = E_g(T_o) \left( 1 - 2.4 \times 10^4 (T - T_o) \right)
\]

(A12)

\[
\kappa(T) = \frac{\kappa(T_o)}{\left( 1 + \frac{D}{2.8 \times 10^{19}} \right) \left( \frac{T}{T_o} \right)^{-4/3}}
\]

(A13)

Here, our list includes electron and hole mobility, \(\mu(T)\), electron and hole saturation velocity, \(\nu_{sat}(T)\), built-in potentials, \(\phi_{built-in}(T)\), thermal voltage, \(V_{TH}(T)\), intrinsic carrier concentration, \(n_o(T)\), bandgap of silicon, \(E_g(T)\), and the thermal diffusion constant, \(\kappa(T)\).

To obtain non-isothermal device performance, we first solve (A3)-(A6) in conjunction with (A7)-(A13). To obtain the quantum corrected values of the state variables \(\phi, n, p, T, \psi_i, \) and \(E_F\) [14]-[16], we then add quantum corrections to the semi-classical solution by
self-consistently solving equations (A1)-(A6) along with temperature relations (A7)-(A13).

B-Temperature Transformation

We utilize the following transformation in conjunction with (A13) to reduce the complexity of the differential heat flow equation by replacing temperature dependent $\kappa(T)$ with the constant $\kappa(T_o)$ in (A6).

$$\bar{T} = T_o + 3T_o \left(1 - \left(\frac{T_o}{T}\right)^{\frac{1}{2}}\right)$$  \hspace{1cm} (A14)
VI- References:


Table 1: Percentage areas and powers of functional blocks in a Pentium III chip [19, 20]

<table>
<thead>
<tr>
<th>Pentium III Unit</th>
<th>Power (%)</th>
<th>Area (%)</th>
<th>(Power/Area) divided by max(Power/Area)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock (CLK)</td>
<td>5.2*</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Issue Logic (ISL)</td>
<td>14.1</td>
<td>9.5</td>
<td>0.71</td>
</tr>
<tr>
<td>Memory Order Buffer (MOB)</td>
<td>4.7</td>
<td>3.3</td>
<td>0.68</td>
</tr>
<tr>
<td>Register Alias Table (RAT)</td>
<td>4.7</td>
<td>3.3</td>
<td>0.68</td>
</tr>
<tr>
<td>Bus Interface Unit (BIU)</td>
<td>5.9</td>
<td>4.3</td>
<td>0.66</td>
</tr>
<tr>
<td>Execution Unit (EU)</td>
<td>13.0</td>
<td>9.5</td>
<td>0.66</td>
</tr>
<tr>
<td>Fetch</td>
<td>16.9</td>
<td>12.5</td>
<td>0.65</td>
</tr>
<tr>
<td>Decode Unit (DU)</td>
<td>17.2</td>
<td>14.6</td>
<td>0.57</td>
</tr>
<tr>
<td>L1 Data Cache (L1C)</td>
<td>9.8</td>
<td>12.5</td>
<td>0.38</td>
</tr>
<tr>
<td>L2 Data Cache (L2C)</td>
<td>8.5</td>
<td>29.8</td>
<td>0.14</td>
</tr>
</tbody>
</table>

* 60% consumed in the clock network, which is uniform over the chip surface.
Figure 1: a) A vertically stacked 3D IC, where each layer is modeled after a Pentium III [1]. b) Floor plan of each layer in conjunction with Table 1.
Figure 2: a) To analyze 3D IC heating, each MOSFET (M) device is replaced by an RC circuit. b) 3D IC’s transistors interact thermally with each other as a result of thermal coupling.
Input:
A digital planar or 3D IC:
- Technology node
- Layout
- Floor plan
- Statistics for typical running applications

Device Level:
- Solve non-isothermal DD-QM Eqns.
- Obtain joule heating as a function of temperature

3D IC Level:
- Obtain values for $R_C^{th}$ thermal network elements
- Obtain relative generated heat of each transistor
- Solve $R_C^{th}$ thermal network
- Obtain temperature profile as well as generated heat of each transistor

Output:
- Device Characteristics
- 3D IC Temperature Profile

Figure 3: Coupled algorithm flowchart.
Figure 4: a) We apply size reduction methods to a planar chip with one hundred mesh points. We divide it up into four blocks. We then replace original mesh with sixteen nodes corresponding to four-port Thevenin representations of each block. (Bold resistors are for package.) b) In 3D, we have six-port tetrahedral shape Thevenin representations for cubes of grid points.
Figure 5: a) Temperature dependent current-voltage characteristics of a 0.13µm N-MOSFET for $V_{GS}$=1.0, 1.5V. a) Steady-state heat generated ($V_{GS}$= $V_{DS}$=1.5V) as a function of temperature ($T$) and $\bar{T}$ ($T_b$). Conversion from $T$ to $\bar{T}$ is given in (A14).
c) 

*Figure 6:* a) A 3D IC with five layers of stacked Pentium III chips. b) Our calculated temperature map corresponding to the middle layer shown in a). Location of the peak temperature moves from clock (in one layered IC) to issue unit. c) Maximum temperature of the middle (also the maximum of the entire 3D IC) and bottom layers as a function of number of layers. d) Oscillation frequency of a thirty one stage ring oscillator calculated by Cadence [22] decreases as temperature increases. Here, ambient is at room temperature (300°K).