ISSN 0002-306X. Изв. НАН РА и ГИУА. Сер. ТН. 2014. Т. LXVII, № 2.

UDC 621.382.13 RADIOELECTRONICS

A.A. GEVORGYAN, A.M. BAGHDASARYAN, R.E. FATYAN

THERMAL AWARE DYNAMIC FREQUENCY SCALING FOR 3D IC

Thermal dissipation in 3D integrated circuits has become one of the most vital problems for this technology. The hardware and software approaches could be used to minimize the thermal dissipation issue. A dynamic frequency scaling based algorithm that allows to minimize the thermal energy concentration in a chip and keep the task execution deadline is introduced. Experimental results and comparison of algorithms are presented as well and show that the method allows to minimize the heat generation by 8%.

Keywords: 3D IC, DVFS, task scheduler.

Introduction. The number of semiconductor devices in ICs grows according to Moors law [1]. This leads to several problems in the design of ICs. The growing complexity of ICs requires more semiconductor devices and more die area to implement complicated architectures. The sizes of transistors decrease and the nano-scaling effects become more important [2]. The problem becomes more vital as the classical scaling mechanisms are not able to adapt to new technologies. Key parameters such as gate oxide thickness are no longer possible to scale down [2]. This means that the device off-state current has begun to creep up very fast. All these problems limit the possibility to increase the clocking speed of digital ICs. The value of the clock frequency for digital circuits can be increased but issues arise like heat generation during over clocking, error quantity, etc. This leads to the platforms with multiple cores that can be clocked with lower frequency but can perform parallel calculus, and in this way increase the total productivity of the system. The goal is to continue to increase the performance of the system by keeping a relatively low clock frequency (several *GHz*) and adding more cores to perform parallel operations. One of the problems that come up for such kind of platforms is the code parallelism. The software developers should design the programs so that the operations could be done in parallel, otherwise the code will be executed on one core and there would be no benefit from multicore architecture. There are several other problems, that are specific for most types of architectures like crosstalk noise, that become more important as the sizes of semiconductor devices become smaller and the density and total length of the wires grow [3]. The next problem is the signal propagation delays and signal sharpness [3]. The wire length plays a significant role in the signal timing and shape parameters and it's become critically important to minimize the signal prorogation delays and preserve the form of the source signal. Then, for more complex

architectures, a larger die area is needed, so the size should be scaled as well. But the sizes of the die cannot be increased without any side effects. The first problem that arises during this procedure is that the number of defects grows in exponential order as the sizes of the chip grow. The defects can lead to disoperation of the semiconductor devices and the output can become unpredictable.

All the problems described above have a major influence on the chip sustainability and productivity and it becomes obvious that new solutions need to be developed to overcome the described issues.

One of the promising technologies is the three-dimensional integrated circuit (3D IC) which allows to improve the technology performance and minimize the side effects described above [4].

3D IC technology. Advantages and disadvantages. The 3D IC consists of several chips that are stacked together in the same package in a vertical direction. The typical 3D IC structure is shown in Fig. 1.

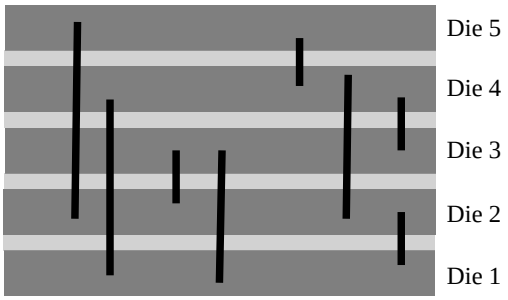

Fig. 1. General structure of 3D IC

This technology allows to have several layers of active logic connected to each other with vertical signaling vias. The benefits of this technology are: the shorter length of global interconnections, the interconnection bandwidth can be much larger compared to the traditional horizontal technology, latencies and propagation delays become much less and the signal edge sharpness is preserved. It also allows to integrate a large amount of low-latency cache memory into the chip structure, and by this to increase the performance of the chip [4]. 3D technology proposes a realistic way to maintain the progress defined by Moore's law.

With all the advantages described above, there are several disadvantages of such kind of architectures. As there are several active layers integrated in the same chip, and the higher integration assumes that more complex functionality is packed into one chip, the probability of a broken chip will be higher. Due to this, the yield percentage is less compared to traditional structures.

One of the most important problems that comes up in 3D IC is the thermal dissipation and heat generation in such structures.

Thermal dissipation problem in 3D IC. Layers in 3D IC have a stacked structure, and by this thermal energy is accumulated in the layers that are located far from the package. This energy must be effectively removed from the body of the chip, otherwise this can lead to heat concentration and finally the chip can become unusable. Several methods are proposed to solve this problem. In general, the methods can be classified into two main groups:

- IC design modifications (hardware approaches)
- Task scheduling methods (software approaches)

Each of these methods has its own submethods. There are three main methods that are used in case of hardware approach:

- Thermal aware floor planning
- Thermal aware placement
- Thermal via insertion

Each of these methods has its advantages and disadvantages. Thermal aware floor planning assumes that blocks in each layer are placed based on their thermal activity [5]. The goal is to avoid having thermal active blocks vertically in the same direction. Thermal aware placement performs the same operation for semiconductor devices [6]. During this procedure the vertical thermal pattern is also taken into account. The third method assumes insertion of special thermal vias [7]. During this operation thermal vias are placed in heat producing zones and drive heat energy from inner layers to upper layers and finally to the package. In general, there are possibly two types of thermal vias. In the first case, the vias connect only two nearby layers. The second approach is to use vias to connect several layers to one another.

The experimental results show that the most effective method is to place thermal vias [8]. But the disadvantage of this method is that it requires additional area for thermal vias to be placed.

Another issue is the behavior of the chip during the run time. As the chips are stacked together, the overall temperature is very dependent of the temperature of the stack in which the core is located. During the run time, the tasks have different thermal activities, and by this, the thermal profile of the system can vary during the task scheduling. To avoid overheating during the run time the thermal aware scheduling methods can be used, but the currently existing algorithms for thermal aware task scheduling have low performance and cannot be used in most cases.

Dynamic voltage and frequency scaling. One of the methods that is used to control heat generation in a chip is dynamic voltage and frequency scaling (DVFS) for the processor [9]. This is a widely used technique that is intended to reduce power consumption and heat generation in a wide range of devices like embedded systems, laptops, mobile phones and portable devices.

DVFS technique allows to reduce power consumption in CMOS ICs by reducing the supply voltage level and the operational frequency. To represent the power consumption of the processor, the following equation can be used (1):

$$P = CfV^2 + P_{Static}, \quad (1)$$

where C is the capacitance of transistor gates, f is the operational frequency, V is the supply voltage. The voltage value is determined by the frequency value, and is the minimal value at which the circuit is still operational. Therefore, the voltage value can be decreased by decreasing the value of operational frequency. This can lead to significant power minimization as the values for both variables V and f are minimized.

For most CMOS based processors, the frequency depends on the supply voltage and can be given by the following equation [10, 11]:

$$f = a\frac{(V_{dd} - V_t)^2}{V_{dd}}, \quad (2)$$

where V_{dd} is the supply voltage, f is the clock frequency, V_t is the threshold voltage, a is a constant. This equation can be rewritten:

$$f = kV_{dd}, \quad (3)$$

where k is a constant. Thus, frequency has a linear relation with the supply voltage.

By substituting the V value from (3) in (1):

$$P = \frac{C}{k^2} f^3, \quad (4)$$

which is equivalent to

$$P = qf^3, \quad (5)$$

where q is a constant.

As it can be seen, the power consumption is directly proportional to the cube of core frequency [12]. So it becomes obvious that the frequency scaling for 3D IC can have major effect to overcome the thermal dissipation problem.

There are several factors that influence the DVFS efficiency. Several of these factors are described below:

- memory performance,
- cash organization and hitting coefficient,
- sizes of the transistors,
- idle/sleep modes of the system,
- number of core and their complexity,
- clock domain structures.

All these aspects have different impacts on the efficiency of the DVFS technique. The DVFS was first proposed in 1994 when the transistor sizes were approximately

0.8 μm and the supply voltages 5 V. Also, the ratio of dynamic power to static power consumption was high. Since that time, the structure of computational systems has changed significantly. The size of transistors has decreased to 14 *nm* technology and by this the threshold voltage has decreased as well. This means that the efficiency of the DVFS method is reduced as the difference of idle and sleep modes is not so significant as before and the power saving will not be so efficient. The structure and performance of memory has changed as well. The bandwidth of the memory bus has increased and the performance of the memory read/write operations has also increased. This means that the number of the waiting cycles of the processor can be less than before and the efficiency of the frequency scaling is going down. The next point is the organization of the cash memory. In modern systems there are several layers of cash memory and the hit percentage is relatively high compared to the previous architectures. This will lead to the acceleration of the memory response, and the number of the idle cycles of the processor will be low. In addition to all this, the current architectures of computational systems assume more than one processor in the system and multitasking. Modern architectures have a performance-monitoring unit (PMU). Based on this memory requests per cycle and instructions per cycle models have been developed [13].

As can be seen from the discussion above there are several methods for DVFS implementation. However this method can be a good heat minimization technique for 3D IC. In this case there are several problems that need to be solved to ensure a good performance of this method and power saving.

Due to the specific structure of 3D IC the issues arising are as follows:

- more than one processor,
- strong thermal correlations in vertical direction between the heat sources,
- several clock domains.

In the following section, the algorithm allowing to perform DVFS for 3D IC based on the task scheduler aimed at decreasing the heat generation and concentration in the chip will be considered.

Methodology. In this section, the technique for performing DVFS for 3D IC is considered. Linux kernel governor is chosen as a basic governor structure. The governor should perform DVFS based on the thermal activity of the tasks and perform the task rescheduling to achieve an acceptable thermal profile for the processors' stack in 3D IC. The governor consists of two parts: the decision making logic, and the frequency scaling logic. The general structure is shown in Fig 2.

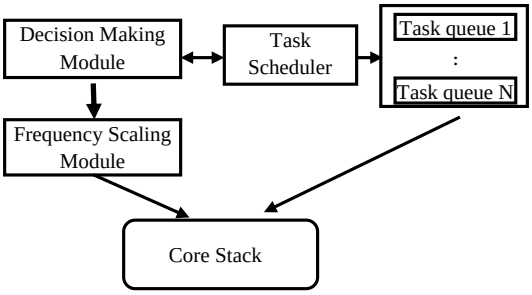

Fig. 2. DVFS governor structure of 3D IC

The overall system consists of the task scheduler and the DVFS modules. The scheduler is responsible for performing task assignment to the cores in one stack and rearrangment of tasks that are in active and waiting states based on feedback from the DVFS module.

The DVFS module is an always running module that performs monitoring of cores in a stack, and, based on thermal profile, performs the frequency scaling to keep the desired temperature value.

There are two possible methods of frequency scaling optimization:

- static slack optimization,
- dynamic slack optimization.

The proposed method uses static slack optimization. This technique uses the idle CPU time for performing a frequency scaling. When the CPU time requested by the application is less than 100%, there are time frames when the CPU is idle and this time can be used to perform frequency scaling and power saving. Fig. 3 illustrates this concept,

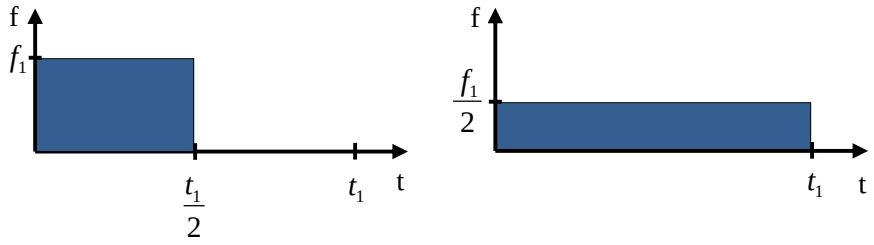

Fig. 3. Frequency scaling by CPU idle time

where t_1 is the deadline time for the task, f is the frequency. As can be seen, the idle time can be used to decrease the frequency value and, by that, decrease the power consumption. For the example in Fig. 3, the power saving calculation can be carried out based on (5) and will be 25%.

The algorithm performs selecting and rescheduling tasks based on the pending task list Table 1.

Table 1

Task execution and deadline times

Task	Execution time	Deadline
T_1	et_1	dt_1
T_2	et_2	dt_2
...	...	...
T_n	et_n	dt_n

When the temperature of a core increases and reaches the allowed threshold value, the DVFS module sends notification to DMM(Decision Making Module). The DMM performs the calculation of the target frequency to decrease the temperature of the core. Then it notifies the task scheduler to perform the rescheduling of the tasks so that the task with maximal allowed execution time for the newly calculated frequency value should be chosen.

Experiments show that the proposed DVFS governor allows to decrease the temperature of the stack and keep the allowed deadline for tasks. The algorithm sorts the task based on their run time and the deadline time. During the frequency scaling procedure the next task that is chosen is the one with the maximum allowed running time The general structure of the algorithm is shown below:

1. Monitoring the temperature of a stack.
2. Indicating the cores with the temperature value near to the threshold level.
3. Calculating the desired frequency based on current power consumption.
4. Notifying the task scheduler to sort the pending task list and choose the task with a maximum admissible running time.
5. Scheduling the selected task for execution.
6. Going to step 1.

In the first step, a set of tasks to perform thermal simulation for the task scheduling procedure is selected. To ensure the best memory performance, all the test are carried out with a virtual file system created in RAM. As a task set SPEC CPU2006 benchmark is selected. It includes two types of benchmarks: integer and floating point.

The workflow of thermal aware task scheduler is based on temperature balancing for each stack. The tasks are chosen based on their thermal activity so that the total heat generation in one stack will not exceed the threshold value for the stack. The DVFS algorithm is performed based on the task priority and thermal activity. When the heat generation exceeds the threshold value, DVFS is performed. As DVFS decreases the core performance, it is necessary to perform the task rescheduling to achieve optimal performance. From this point of view, the task rescheduling is performed in two stages. At the first stage, the task rescheduling is done across the vertical stack, and if no possible option is found, rescheduling is done globally including other stacks one by one. It should be mentioned that this procedure requires

some time and if that time interval is larger than the task performing time, it is not effective and there is no reason to perform rescheduling. To overcome this problem and achieve a maximum performance for the task scheduling, all the tasks are stored in sorted order in queues. When rescheduling needs to be done, a binary search for the task queue can be performed and by this optimal time is achived.

Experimental results. The experimental platform includes PCs with Intel I5 CPU that are connected to the network and are monitored from one controller unit. The selection of I5 cores allows to have many power states that are shown below in Table 2 [14]. These values describe the P-state of the processor.

Table 2

Power states of Intel i5 processor

Power state	Frequency in *MHz*
P_0	2400
P_1	2399
P_2	2266
P_3	2133
P_4	1999
P_5	1866
P_6	1733
P_7	1599
P_8	1466
P_9	1333
P_{10}	1199

The controller software simulates the 3D IC thermal model based on the data from the client machines. The Linux kernel is used as a basic software environment. To simulate the thermal effects, a modified version of HotSpot software is used. It allows to simulate the thermal profile based on the data collected from client systems.
For the task simulation, SPEC CPU2006 is used. It contains a large variety of benchmark tests with different thermal profiles. In HotSpot the i5 processor is used to simulate one layer of 3D IC. The experimental results are shown in Table 3. In average, after performing frequency scaling the temperature is reduced by 8%.

Table 3

Power states of Intel i5 processor

Benchmark type	T_{avg} w/o DVFS	T_{avg} with DVFS
Int arithmetic	33.74	31.04
Floating point arithmetic	49.0	45.08
Dynamic memory allocation/freeing	36.68	33.13
Iterative counting	37.1	34.1
Recursive counting	37.52	34.28

Conclusion. In this article, a method for the dynamic voltage and frequency scaling and the thermal aware task scheduling for 3D IC is considered. An algorithm for performing DVFS and rescheduling tasks to keep the optimal performance of the system is proposed. To maintain the system performance during a low frequency period, task rescheduling is carried out. The rescheduling procedure is done in two phases: at first it is done in a core stack, if no possible solution can be found, rescheduling is performed by including other stacks and, by this, keeping the minimal number of the task switching. Experimental results show that by using the proposed DVFS with the thermal aware task scheduling allows to minimize heat generation by 8%. This method can also be used for multi-core dies.

REFERENCES

1. http://www.intel.com/content/www/us/en/history/museum-gordon-moore-law.html
2. Scaling, power, and the future of CMOS / **M. Horowitz, E. Alon, D. Patil et al** // IEEE Intl. Electron Devices Meeting, IEDM Technical Digest.- Washington, DC, Dec. 2005.- USA.- P. 9-15.
3. **Kahng A.B., Muddu S. and Vidhani D.** Noise and delay uncertainty studies for coupled rc interconnect // Proceedings of ASIC/ SOC Conference.- 1999. –P. 3–8.
4. **Yuan** X.**, Jason C., Sachin S.** Three-Dimensional Integrated Circuit Design Series: Integrated Circuits and Systems.- Springer, 2009.
5. Interconnect and thermal-aware floorplanning for 3D microprocessors / **W. L. Hung, G. M. Link, Y. Xie, et al //** Proc. Intl. Symp. Quality of Electronic Design,.- March 2006.- P. 98-104.
6. **Jason C., Guojie L., Jie W. and Yan Z.** Thermal-aware 3D IC Placement Via Transformation //Proc of ASP-DAC.- 2007.- P.780-785.
7. **Sapatnekar S. and Goplen B.** Placement of 3D ICs with thermal and inter-layer via considerations // Design Automation Conference.- June 2007.- P. 626-631.
8. **Brent Goplen, Sachin S. Sapatnekar.** Placement of Thermal Vias in 3-D ICs Using Various Thermal Objectives // IEEE Transactions on computer-aided design of integrated circuits and system. - Apr.2006.- Vol. 25, no. 4.
9. **Graybill R. and Melhem R.** Power Aware Computing.- Kluwer Academic/Plenum Publishers, 2002.
10. **Martin T.L.** Balancing batteries, power, and performance: system issues in CPU speed-setting for mobile computing: PhD thesis, Carnegie Mellon University.- Pittsburgh, PA, USA, 1999.
11. **Weissel A., Bellosa F.** Process cruise control. Event-driven clock scaling for dynamic power management //CASES.- Grenoble, France, Oct 8–11 2002.
12. **Snowdon D. C.** OS-Level Power Managemen: PhD thesis, School Comp. Sci. & Engin., University NSW.- Sydney, 2010.
13. Advanced configuration and power interface specification / **HP, Intel, et al.**- 2011.
14. http://www.intel.com

Synopsys Armenia CJSC. The material is received 10.04.2014.

Ա.Ա. ԳԵՎՈՐԳՅԱՆ, Ա.Մ. ԲԱՂԴԱՍԱՐՅԱՆ, Ռ.Է. ՖԱՏՅԱՆ

ԵՌԱՉԱՓ ԻՆՏԵԳՐԱԼ ՍԽԵՄԱՆԵՐՈՒՄ ՋԵՐՄՈՒԹՅԱՆ ՆՎԱԶԵՑՄԱՆՆ ՈՒՂՂՎԱԾ ՀԱՃԱԽՈՒԹՅԱՆ ԴԻՆԱՄԻԿ ՄԱՇՏԱԲԱՎՈՐՈՒՄ

Եռաչափ ինտեգրալ սխեմաներում ջերմային ցրումը դարձել է այս տեխնոլոգիայի հիմնական խնդիրներից մեկը: Վերջինիս ազդեցությունը նվազեցնելու համար օգտագործվում են սարքային և ծրագրային մեթոդներ: Ներկայացված մեթոդը հիմնված է հաճախականության դինամիկ մասշտաբավորման վրա և թույլ է տալիս նվազեցնել ջերմային էներգիայի կուտակումը միկրոսխեմայի կառուցվածքում՝ ապահովելով խնդրի կատարման վերջնաժամկետը: Ներկայացված են փորձնական արդյունքները և ալգորիթմների համեմատությունը: Ցույց է տրված, որ կիրառված մեթոդը թույլ է տալիս ապահովել ջերմային էներգիայի կուտակման նվազեցում 8%-ով:

Առանցքային բառեր. եռաչափ ինտեգրալ սխեմա, լարման և հաճախության դինամիկ մասշտաբավորում, խնդիրների պլանավորիչ:

А.А. ГЕВОРКЯН, А.М. БАГДАСАРЯН, Р.Э. ФАТЯН

ДИНАМИЧЕСКОЕ МАСШТАБИРОВАНИЕ ЧАСТОТЫ ДЛЯ МИНИМИЗАЦИИ ТЕМПЕРАТУРЫ В ТРЕХМЕРНЫХ ИНТЕГРАЛЬНЫХ СХЕМАХ

Концентрация термальной энергии в трехмерных интегральных микросхемах является одной из важнейших проблем этой технологии. Для минимизации проблемы концентрации термальной энергии используются аппаратные и программные методы. Представлен метод, основанный на масштабировании частоты синхросигнала и позволяющий минимизировать концентрацию термальной энергии в структуре микросхемы и в то же время не превысить лимит времени выполнения задания. Проведен анализ экспериментальных результатов и дано сравнение алгоритмов. Показано, что данный метод позволяет сократить генерацию термальной энергии на 8%.

Ключевые слова: трехмерная интегральная схема, динамическое масштабирование напряжения и частоты, планировщик задач.