

Algorithm of Thermal Optimization of Placement of Basic Elements of VLSI

Victor M. Kureychik
ICTIS, CADS Department
Southern Federal University
Taganrog, Russia
kur@tgn.sfedu.ru

Andrey A. Kulakov
ICTIS, CADS Department
Southern Federal University
Taganrog, Russia
skive@tgn.net.ru

Abstract—A new algorithm for optimizing the thermal layout of basic VLSI elements, taking into account the features of the length of conductors and mutual thermal effects for 3D ICs, is described. It provides a differentiated temperature distribution in different layers of the chip. The algorithm includes parameters such as the length of the conductors, the interlayer transition holes, the power density and, consequently, the temperatures of the corresponding layers, and attempts to ensure a uniform temperature distribution in each layer of the chip. A fundamental difference from existing algorithms is the presence of two localization loops (inside the layer) and global (between layers), as well as using the metric for thermal homogeneity in topology. The novelty of the study lies in the parallel acquisition of a set of alternative solutions and the choice of a quasi-optic one from it. The principal difference lies in the work of the hybrid algorithm with the choice of volatile parameters. The experimental study was carried out for several randomly generated variants of the problem, and a number of well-known comparative tests of MCNC. Based on the results of the experiments, the algorithm showed an improvement in the value of the objective function by 5-10%.

Keywords— *thermal optimization, allocation algorithm, VLSI.*

I. INTRODUCTION

3D placement is an important step in the physical design of IP. It provides the placement of multiple functional cells in multi-stack layers and forms the path to routing. Many 2D placement algorithms extend to 3D placement problems, but heating problems are critical, so they need to be addressed in the course of the physical design process. There are three stages in the allocation problem: global placement, legalization, and detailed placement. Given the initial solution, the global allocation specifies the solution to the cell area in each predetermined area no more than the capacity of this region. These areas are processed in a top-down manner from the aggregated to the detailed level using partitioning methods and multilevel placement methods and processed in a flat format at a detailed level using flat placement methods. After a global placement, legalization is carried out to determine the specifics of the location of all elements without overlays, and detailed placement performs local improvements to obtain the final solution. The experimental investigation was carried out for several randomly generated variants of the problem, and a number of known comparative tests. Based on the results of the experiments, the algorithm showed an improvement in the value of the objective function by 5-10%.

II. BRIEF REVIEW AND ANALYSIS OF THE METHODS OF VLSI DEPLOYMENT

You can structure groups of placement methods into 3D-ICs as follows.

1. Partition-based methods [1] insert segments of the integrated circuit on the layers of the device at separate corresponding stages in the traditional partitioning process. The cost of the partition is measured by the weighted sum of the estimate of the length of the conductors and the number of connections, where the weighted criteria take into account heating, time delays, and traceability.

2. Methods of flat placement in the main for power (or quadratic) placement and their variations, including direct force methods, cell permutation methods, and homogeneous quadratic modeling methods. Direct force methods [2] use a vector, which is called the repulsive force vector. This repulsive force vector is equivalent to the strength of the electric field, in which the charge distribution is the same as the distribution of the cell area. The methods of permutation of cells [3] are similar to direct force methods, in the sense that they also add a vector on the right side of the linear system.

3. The multilevel technique [4] forms a physical hierarchy according to the original logical description of the VLSI, and solves the sequence of allocation problems from aggregated to detailed level.

4. In addition to these methods, in [5] a 3D placement approach is proposed based on the results of the existing 2D layout and builds the 3D placement with the appropriate transformation.

The issues of thermal placement were investigated for both 2D and 3D designs. In [6], a matrix representation of the thermal layout for 2D design was used and an attempt was made to achieve an even power distribution.

Yan Haixia et al. [7] proposed an algorithm for thermal 3D placement using homogeneous quadratic modeling, which combines the thermal task with the placement process to reduce the temperature of the hot spot and obtain a thermally balanced 3D placement. This algorithm includes thermal global placement, thermal layer and, finally, detailed placement. This work offers two quick methods for estimating the temperature of each grate to update the coefficient instead of accurate thermal analysis.

III. FORMULATION OF THE PROBLEM

The initial data for the developed algorithm is a set of basic elements (modules), each of which is associated with a real non-negative number equal to the value of the power density of the element. Each layer of the chip is assigned a set of modules placed on it. The modules assigned to the chip layer must have the maximum number of interconnections. For each such layer, the modules and their power density values are represented as a matrix M_l of nonnegative real numbers representing the power density of the modules, where $l = 1$ to N , N is the number of (active) layers of the chip. $t \times t$ of the sub matrix M_l corresponds to the chip region, and $S_t(M_l)$ denotes the set of all such disjoint sub matrices M_l . The sub matrix that has the largest amount represents the most "hot" area on the chip. For each layer of the chip, the proposed method tries to find the placement of modules in such a way that the temperature of the hottest area is as low as possible. The parameter t of the models denotes the heat transfer rate. In general, the dissipated power of a cell or a logic gate depends on several factors, such as the circuit structure, valve functionality, conductor loads, and input data and so on.

The task of synthesizing the matrix (SM) is accepted as the basic concept of this stage of our work. Models of SM thermal placement offer two ways to solve it. The method of calculating the temperature based on the power estimate for the placement of standard cells was proposed in [8].

The initial data for the statement of the problem can be formulated as follows. For a given set of numbers, the goal of the problem is to construct a matrix of these numbers in such a way that there are no sub-matrices of a certain size that have a large sum. Sub matrix refers to the set of consecutive rows and columns. For a given integer t and a matrix M , let $S_t(M_l)$ be the set of all $t \times t$ sub matrices of M .

Let $\sigma(M)$ be the sum of all records in M . Denote the maximal

value $\mu_t(M) = \max_{S \in S_t(M)} \sigma(S)$. Then the problem can formally

be defined as follows: The integers t, m, n and the list mn of nonnegative real numbers x_0, x_1, \dots, x_{mn} are given. It is necessary to generalize the $m \times n$ matrix from these real numbers so that $\mu_t(M)$ is minimized.

Under these conditions, the definition of the objective function for our problem can be represented as follows. Let $\zeta(M_l)$ denote the thermal boundary of the input matrix M_l . The problem can be formulated as follows: let the module power density values for a given module topology be represented by an $m \times n$ matrix M_l of real numbers. Let S_t be the set $t \times t$ of the sub matrix M_l . Let $\sigma(M_l)$ be the sum of all the entries in M_l . Let

$$\mu_t(M) = \max_{S \in S_t(M)} \sigma(S). \text{ It is necessary to organize the elements}$$

M_l in such a way that the value of $\mu_t(M_l)$ will be minimal.

Hence, by definition, $\zeta(M_l) = \min \mu_t(M)$. Detection of the location in the matrix for the critical threshold is equivalent to finding the location of the modules in such a way that the power density and, consequently, the temperature of the "hot" region itself is the minimum of all possible other alternative locations.

For a number of layers of the chip $l = 1$ to N , let $\beta = \max N_l = l \zeta(M_l)$, and the goal of the task is to minimize the cost of β . For M_l , let $l = 1$ correspond to the lowest layer of the device and $l = N$ corresponds to the top layer of the chip.

In this development, an attempt is made to organize the values of $\zeta(M_l)$ to no increase for $l = 1$ to N . Let D denote a measure of degree in which the values of $\zeta(M_l)$ deviate from this order of non-increasing and is defined as the sum of positive values of $\mu_t(M_l) - \mu_t(M_{l-1})$ for all the layers l . Therefore, the lower the value of D , the better the location of the values of $\zeta(M_l)$, $l = 1$ to N . Let k_l denote the total number of relationships between all pairs of modules in the layer l of the microcircuit. Then, the total number of interconnects for all layers of the chip is.

Then the problem can be formulated as follows: Considering the set of K modules (gates or cells) with their interconnections and N layers of the chip, find the location of these modules in the layers of the chip, such that the objective function (as indicated below) is minimal. The target function of the allocation algorithm f is given by:

$$f = \alpha \cdot D + (1 - \alpha) \cdot \sqrt{\sum_{l=1}^N k_l \cdot \beta} \quad (1)$$

where $0 \leq \alpha \leq 1$

IV. ALGORITHM FOR THE THERMAL ALLOCATION OF BASIC ELEMENTS OF VLSI

The inputs in the proposed algorithm consist of N (the number of layers of the chip), the set K of rectangular modules with their individual power density values, R rectangular positions for placement in each layer, the predetermined topology and the crystal region allowed for interconnections, and the order of the sub matrix t .

Define the following parameters of the algorithm: the set of N layers of the chip; set K of rectangular cells (modules); a set of cell power density values P ; matrix of commutation of cells I ; a set of R rectangular positions for placing cells at each level of the chip; the maximum permissible area A of the circuit layers; partition parameter t .

The output of the proposed algorithm is the placement of $II = \pi_l, l = 1, 2, \dots, N$ of all rectangular modules on the N layers of the microcircuit in such a way that the objective function is minimized.

The generalized block diagram of the proposed algorithm is shown in Fig. 1. In general, the algorithm is based on the annealing simulation method, but the fundamental difference from existing implementations is the presence of phases of local and global solution improvement.

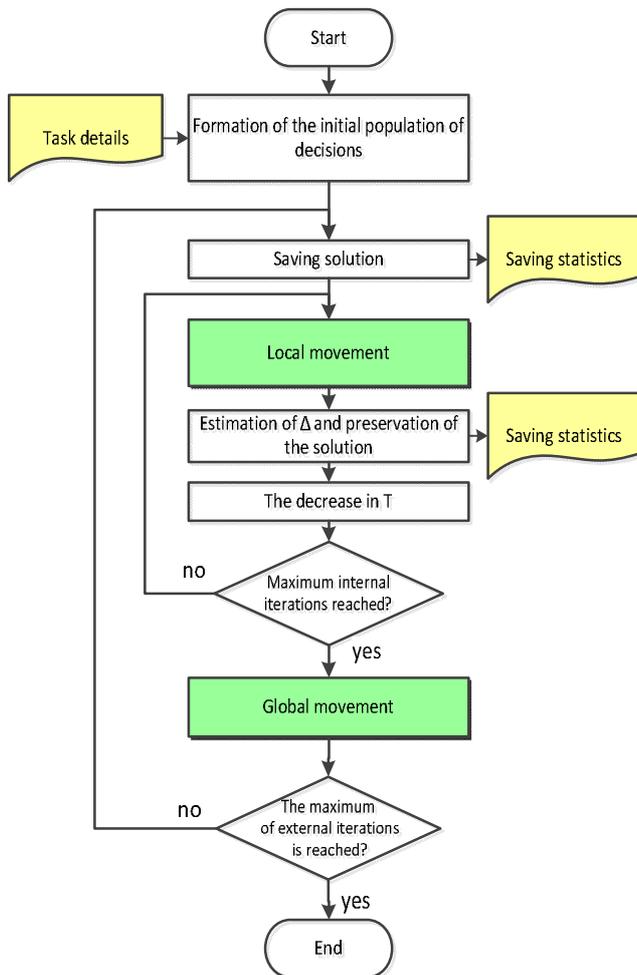


Fig. 1. Generalized structural scheme of the algorithm.

Initially, the set of K modules is divided into N clusters, where each cluster contains the maximum interconnection, where the degree of interconnection of a pair of modules is determined by the number of interconnections between them. Each cluster includes many modules with strong interconnection, and is formed using a technique similar to the

Prima algorithm. Different clusters are formed on the basis of the no increasing degree of interconnection of their component modules, and are assigned in the same order, in different layers, from the bottom to the top of the stack.

Let's consider in detail the main procedures of the algorithm.

In the proposed algorithm, at the stage of the primary module placement, each of the cluster of modules, the microchip is in the order of no decreasing the values of the power density of the cells. It can be assumed that R is an integer multiple of t^2 . Assignment of the modules in the individual matrix elements is performed starting with the upper left corner element, and filled with a "snake". In this case, the cells with the highest and lowest values are placed in the middle of the thermal window (taking into account the elements inside the $t \times t$ sub matrices).

After the formation of the initial population, the algorithm starts the annealing cycle, which includes the operators for the local movement of the modules within the layer, and the global movement of the modules between the various layers of the chip.

The matrix of R elements is divided into R/t^2 sub matrices, each matrix of order $t \times t$. For each of these sub matrices, the module with the maximum total length associated with other modules of connections (conductors). In layer R , the modules are sorted in order of no decreasing their temperatures, and the sorted set is divided into R/t^2 subsets.

Global movement between different layers includes the following steps in the order indicated: selecting a pair of layers of the device; select a pair of modules in these layers, one from each layer; exchange selected modules between the selected pair of layers.

One of the main goals of global migration is to reduce the number of interlayer connections. First, pairs of layers are sought in which replacement is allowed. They can be obtained from a list of pairs where there is an increase in thermal deflection. Next, the work continues with the pairs in the following order. If layer 0 is the bottom layer and layer n is the top layer, then layer 0 is stored as fixed. The other layers are checked sequentially, starting from layer 1 to n , to find the first pair in which the deviation increment occurred. If the pair is not found, then layer 1 is selected. The procedure will continue until the last possibility of reaching a pair of layers, i.e., layer $n-1$ and layer n . For each pair of deviation layers, the following steps are performed.

- Check if there is a deviation $D > 0$ or not. If yes, do the following.

- Let there be a deviation D between the layers $n-1$ and n .

- Find the elements of the sub matrix that contribute to the maximum amount for an individual layer.

- For the layer $n-1, t^2$ there are elements for the sub matrix of order t . Now for each element of layer $n-1$ find the total number of layers n (let it be $C1$) interconnected with all modules, and also the total number of connections in the same layer (let it be $C2$). Now, we follow the assumption that only

that module is assigned for permutation between layers, which has the least number of links in the current layer, and the largest one in the other layer. Therefore, it can be ensured that the permutation will reduce the number of interlayer connections, as well as the length of the conductor. Hence, let's take the ratio $(C1/C2)$ as a weight factor and perform the following steps:

- 1) Calculate $C1$ and $C2$ for each of the modules t^2 of the $n-1$ layer. It can be done within one pass; therefore, the complexity will not increase.
- 2) Calculate for each module: $\beta (C1/C2) + (1-\beta) T_{i,j}$, where $\beta = 0,5$. The module that promotes the maximum value will be the candidate module of layer $n-1$, which will be replaced.
- 3) Identify the candidate module of layer n .
- 4) Swap two modules between two layers

We will determine the time complexity of constructing the initial placement of modules in 3D VLSI. Let N be the number of layers, and K the total number of modules of the microcircuit. We take K exactly N , forming a cluster for each layer of the chip is carried out similarly to the Prima algorithm, starting with the initial module. The set of interconnection lengths is calculated during the time $O(K^2)$, sorting by the non-increase of the order of the length of the interconnections is performed in time $O(K^2 \log K)$. After sorting the set, it turns out that the initial module for any cluster can be obtained in time $O(1)$. In addition, for the first cluster, the remaining $(K/N - 1)$ modules can be obtained for an approximate time as soon as the initial module is obtained, $\Sigma (K - i + 1) = O((K/N) 2)$. For each subsequent cluster, K/N is subtracted from the full set of all modules. Thus, the total complexity of constructing the initial placement of modules in the layers $O((K/N)^3)$.

V. EXPERIMENTAL EVALUATION OF THE ALGORITHM FOR THE ALLOCATION OF BASIC ELEMENTS OF VLSI

The proposed algorithm is implemented in the "C" language on the Dell Precision T1700 workstation at 3 GHz. The proposed algorithm is performed for various cases of comparative tests of the Meschach Library MCNC [9, 10]. For all experiments, the initial temperature for simulating the annealing is set to 10,000, and the temperature is reduced in steps of 10 in each of the global iterations. Both internal and external iterations are set to a value of 5. The power density of the modules is generated in the range (0, 100) through a single random number generator. The value of α is taken to be 0.5, and the number of layers of the microcircuit is adopted. 3. The applicability limits of the proposed method are multilayer VLSI with a number of cells up to 104, for a greater number of elements the method successfully finds acceptable solutions, however, the execution time of the algorithm increases dramatically.

From the experimental results with a change in the optimal cost for the same number of modules and with a different number of layers of the chip, it was noted that the cost of the solution decreases with the increase in the number of layers of the microcircuit. This is due to the fact that as the number of

layers increases, for a fixed number of modules, the number of modules in the layer decreases. This reduces the number of interconnections in the layer.

VI. CONCLUSION

A new algorithm for optimizing the thermal layout of basic VLSI elements, taking into account the features of the length of conductors and mutual thermal effects for 3D IC, is described, and provides a differentiated temperature distribution in different layers of the chip. The algorithm includes parameters such as the length of the conductors, the interlayer transition holes, the power density and, consequently, the temperatures of the corresponding layers, and also attempts to ensure a uniform temperature distribution in each layer of the chip. A fundamental difference from existing algorithms is the presence of two localization loops (inside the layer) and global (between layers), as well as using the metric for thermal homogeneity in topology.

The proposed algorithm is an algorithm of "thermal" placement, reducing the lengths of conductors and TSV for 3D ICs having standard cells or valves, and providing a differentiated temperature distribution in different layers of the chip, in which the lowest and upper layers are "hot and cold", respectively.

The novelty of the study lies in the parallel acquisition of a set of alternative solutions and the choice of a quasi-optic from it.

The principal difference between the algorithm and the existing ones is the use of local and global displacement operators, as well as metrics for thermal uniformity in a topology that identifies thermal boundaries and heuristics, based on increasing the connectivity of the cell in the current layer and reducing the connectivity in other layers.

Based on the results of the experiments, the algorithm showed an improvement in the value of the objective function by 5-10%.

The proposed algorithm can be useful for 3D-VLSI CAD developers to improve the quality of design solutions covering several important design parameters of three-dimensional integration schemes - the maximum temperature of the circuit, the length of the conductors and the number of interlayer connections.

ACKNOWLEDGMENT

The work was carried out at the expense of partial financing by the Ministry of Education and Science of the Russian Federation "Organization of scientific research" code 2.5537.2017 / VU

REFERENCES

- [1] B. Goplen, S. Sapatnekar "Placement of 3D ICs with thermal and interlayer via considerations" in Proceedings of the 44th Annual Conference on Design Automation, 2007, pp. 626-631
- [2] I. Kaya, et al. "Wirelength reduction using 3-D physical design" in Proceedings of the 14th International Workshop on Power and Timing Optimization and Simulation, 2004, pp. 453-462.

- [3] R. Hentschke, et al. "3D-vias aware quadratic placement for 3D VLSI circuits." in ISVLSI '07 Proceedings of the IEEE Computer Society Annual Symposium on VLSI, 2007, pp. 67–72.
- [4] J. Cong, G. Luo "A multilevel analytical placement for 3D ICs" in Proceedings of the 2009 Conference on Asia and South Pacific Design Automation. - Yokohama, Japan, 2009, pp. 361–366.
- [5] J. Cong, et al. "A thermal-driven floorplanning algorithm for 3D ICs" in Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design //IEEE Computer Society, Washington, DC., 2004, pp. 306-313.
- [6] C. Chu, D. F. Wong "A matrix synthesis approach to thermal placement" in IEEE Transactions on Computer-Aided Design, Vol. 17, №11, Nov. 2008, pp. 1166-1174.
- [7] Y. Haixia, et al. "Thermal aware placement in 3D ICs using quadratic uniformity modelling approach" in Integration, the VLSI Journal, Vol.42, 2009, pp. 175-180.
- [8] C. Tsai, S. Kang "Cell-level placement for improving substrate thermal distribution" in IEEE Transactions on Computer-Aided Design, Vol. 19 №2, Feb. 2010, pp. 253-266.
- [9] A. A. Kulakov, V. M. Kureichik "The method of placement of VLSI blocks of three-dimensional integration" in Problems of the development of promising micro- and nanoelectronic systems- 2014. Collected works / under general. Ed. Academician of the Russian Academy of Sciences A.L. Stempkovsky, Moscow: IPPM RAS Part I, pp. 167-170, 2014.
- [10] A. A. Kulakov, V.M. Kureichik "A subsystem of planning of super-large integrated circuits of three-dimensional integration taking into account mutual thermal influence of elements" in Software products and systems, No. 1, 2014, pp. 131-136