TSV Assignment of Thermal and Wirelength Optimization for 3D-IC Routing

Yi Zhao[12], Cong Hao[13], Takeshi Yoshimura[1]
[1]Graduate School of Information, Production and System, Waseda University, Hibikino 2-7, Wakamatsu, Kitakyushu, 808-0135, Japan
[2]School of Electronic Science and Engineering, Southeast University, #2 Sipailou, Xuanwu, Nanjing, 210018, China
[3]Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China

***Abstract*—The 3D integrated circuit is a complex structure composed of chips in multi-layers fabricated vertically and horizontally. Signal through-silicon-via (TSV) is vertical electric connection enabling to communicate and compact the functional bulk. And thermal TSV is effective to amplify the heat dissipation and reduce temperature by establishing the heat dissipation path. In this work we introduce conditional grid extension method to Integrated Multi-Commodity Min-Cost (IMCMC) problem to relieve capacity limit. Moreover, we propose thermal increase model to simulate thermal distribution and realize temperature reduction. Compared with previous work, several routing revision is adopted to make the result more accurate and reduce wire capacitance. The experimental results show the effectiveness of our algorithm on reducing temperature and optimizing congestion.**

***Index Terms*—3D-IC, TSV assignment, thermal TSV**

I. Introduction

A. The Development of ICs

As is well-known, in 1965, Gordon Moore, one of the founder of Intel pointed in his famous article that the number of transistors an integrated circuit could contain may increase one time every two years around and the energy of each transistor might decrease proportionally to the size of the devices. However, as the integrated circuits develop in the recent years, the fabrication of ICs meets physical limits. [1]

1) The invention of 3D-ICs: In the latest decade, an emerging technology named three-dimension integrated circuits (3D-IC) supported by through-silicon via (TSV) comes up, which greatly extends the space of IC development together with supplementing the drawbacks of 2D-IC. The basic structure of a 3D-IC fabricated in via-middle is shown in the Fig. 1.

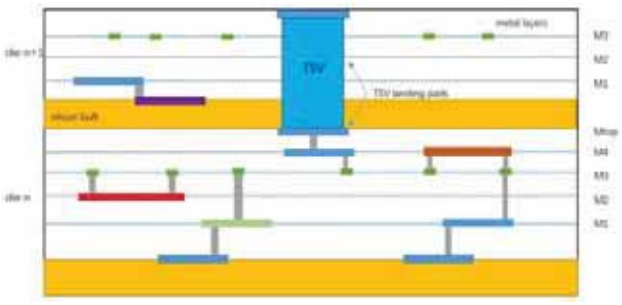


Fig. 1. TSVs, TSV landing pads, and connections to TSV landing pads

The components (Gate Arrays) are located on the different layers of chip dies, communication wires are crisscross in the chip. Chips with different functions pile layer by layer in the vertical directions. And there are bulks whose radius is larger than ordinary vias called TSVs, taking the responsibility of connecting and communicating between dies. Fig. 2 shows an entire structure of complicated 3D-IC.

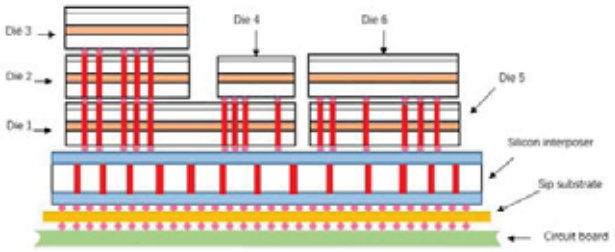


Fig. 2. The entire 3D-IC module

2) The Advantages of 3D-ICs: By vertical electric connection, the chip dies are stacked to form a 3 dimensional IC. Compared with 2D-IC, it has appeared prominent merits and realized high performance. [2]

- The minimized footprint occupies smaller area on the board. It means the chips could be designed smaller, which results in the reduction of fabrication. If the density could be higher, the pressure of physical limit is relieved.
- The chip is divided into parts of different functions so that the fabrication and test could be managed individually.
- The wirelength is reduced so that the power consumption is reduced.
- The congestion is optimized due to optimizing the assignment and partitioning. The performance of chip makes great progress.

3) The Challenges of 3D-ICs: But the complex structure makes trouble to the fabrication, such as micro-channels. The possibility of mistakes inevitably becomes larger and the cost increases. The heat dissipation of the chip is also a vital factor in the IC design. This is because up to millions of components produce great quantity of heat in such tiny room, the temperature may sharply increase to make danger and deteriorate the performance. At this point, cooling method has

978-1-5386-6365-3/18/$31.00

to be made use of and thermal TSV is effective. [3]

B. Related Research

Many previous dissertations [2]- [14] have already been proposed for TSV assignment of 3D-IC. Current research result is mainly divided into two aspects, one is focusing on multi-level flow in order to optimize the wirelength and congestion, and the other is on thermal dissipation to improve the performance. On the one hand, the assignment of wire and TSV is a significant task because of the great number of components and micro-scaled chip area. The work in [5] paid attention to wirelength optimization in the multi-level. [6] designed an innovative method to solve the multi-level problem by calculating the residual path in the flow network. The work in [7] proposed a multi-flow min-cost flow algorithm to assign the wire in a coarsening-uncoarsening method level by level. [8] investigated the impact of TSV in the 3D-IC of different layers in two schemes called TSV co-placement and TSV site. In [9] in order to make signal TSV planning more flexible, TSV is integrated with pin assignment. [10] fully made use of mathematical tools, relaxing the capacity constraints in the grid, the min-cost multi-commodity flow is transferred to Lagrangian sub-problems, which could be solved by Lagrangian Relaxation. The work in [11] reused 2D blocks to construct 3D rectilinear steiner tree (RST) together with using a post-floorplanning TSV insertion method. What's more, FLUTE algorithm in [12] was also referred to estimate the TSV locations.

On the other hand, resistive thermal network is the most commonly used thermal increase model. [3] used the model to estimate the temperature on the regions of board. But in the calculation, matrix equations need solving, [2] proposed compression storage and left-looking LU decomposition method to reduce the complexity of calculation, which greatly save the storage space and time.

And some works like [13] and [14] proposed many constructive ideas on the material and fabrication of 3D-ICs.

C. Major Contribution

Based on the previous works and the purpose to revise the shortcomings, this paper adopts the **IMCMC network** to search for the minimum wirelength taking efficiency and optimization into consideration at the same time and proposes a conditional grid extension algorithm to relieve the capacity constraints. A **thermal increase model** is introduced to estimate the temperature and decide the maximum reduction of heat. And other modifications are made to make the resolution satisfy the realistic circuits more. We summarize the contributions of this work as follows.

1) We propose a **conditional grid extension** method to extend the adjacent region of the effective grids in the uncoarsening step. In this way, the capacity of the region increases so that more wires could locate in the region.
2) We introduce a **thermal increase model** based on the type of the components. Because a type of components (adder, multiplier or divisor) has a certain area and power, given the number of components in a region, we could calculate the relative temperature and heat reduction potential.
3) We modify the assignment method by **HVH structure** so that the crosstalk from layer beneath and above the routing layer is eliminated and congestion is optimized. We could make best utilization of placement and routing resources. What's more, the capacitance between the wires decreases as well.
4) We revise the location of pins by attaching the **offset distance** to the center of grid because the previous works mainly regard the center as the real location, which led to a considerable error.

The rest of the paper is organized as follows. Section II gives the problem formulation. Section III introduces the IMCMC network and the conditional grid extension method we propose. Then we give the detail about the thermal increase model in Section IV. The experimental results are shown in Section V. Finally Section VI gives the consequences.

II. Problem Formulation

A. Problem Description

In a 3D-IC module, the input is given to us:

- A 3D-IC with n chip dies D=$\{d_1, d_2, \ldots, d_n\}$ and $n-1$ TSV layers L=$\{l_1, l_2, \ldots, l_{n-1}\}$, where l_k TSV layer connects d_k and d_{k+1} chip dies.
- Each chip die is divided into many grids and each grid has a certain capacity c_g defining the number of TSVs could be allocated to. And $\forall c_g \geqslant 0$.
- The 3D netlist includes m nets which describes the situation of connections. Each net is denoted as a 2-pin net composed of $\eta_k = s_k, t_k$. s is the source pin which is transferred from multi-pin on the toplayer by minimum spanning tree, and t is the sink pin. [15]
- Every source pin s and sink pin t locates on a certain grid i associated with a position, denoted as $p_i = (x_i, y_i, z_i)$, $p_s = (x_s, y_s, z_s)$, $p_t = (x_t, y_t, z_t)$, where z is the number of chip and $1 \leqslant z_s < z_t \leqslant n$.
- In the thermal increase model, the device property including the number and area of components, thermal resistance and power dissipation of each layer is also given.

The source pin s and sink pin t of one net $\eta_k = (s, t)$ construct a bounding box. If we project them into one layer, we'll get a 2D rectangular minimum region. The layer that net passes has at most one TSV and the vertical distance between layers is constant so that we could estimate the wirelength by **Manhattan Distance** [19],

$$wl(\eta) = d_{s,\tau_1} + \sum_{1 \leqslant i \leqslant z_t - z_s} d_{\tau_i, \tau_{i+1}} + d_{\tau_{z_t - z_s}, t} \qquad (1)$$

where $d_{p,q}$ is the distance between the two adjacent layer p and q that calculated by Manhattan Distance. In this way we get the estimation of wirelength by summing up every wirelength

between layers and the vertical distance is ignored because it is constant.

The problem has a **constraint** that for each grid i, the number of TSVs are contained mustn't exceed the capacity g_i.

The **purpose** of the work is to minimize the sum of wirelength of all of the nets $\eta_i = (s_i, t_i)$ during the TSV assignment.

$$min : \sum_{1 \leqslant i \leqslant m} wl(\eta_i) \tag{2}$$

B. IMCMC Network Description

We formulate the TSV assignment problem as an IMCMC network flow problem [15]. Given $G = (V, E)$, associated with costs and capacities on each edge. The vertices composed of $V = V_s \cup V_t \cup V_g$, and $g_{mn} \subset V_g$ which represents n-th grid on m-th layer. Each vertice g_{mn} has a capacity c_g. The edges $E = \{e|e = (u, v)\}$, where $u \subset V_s, v \subset V_g$, or $u \subset V_g, v \subset V_t$ or $u, v \subset V_g$ in the direction from u to v. Edges have their own costs ω_e defined as the routing costs between the pins and TSVs, which is the same as Manhattan Distance in Eq. (1). When assigning a net in the network, it is equivalent to defining a signal path from s to t and the flow on an edge (u, v) is denoted as $f^{s,t}_{u,v}$. And the solution is to find the min-cost flow of the graph [7]. Fig. 3 sets an example about IMCMC network with two nets $\eta_1 = (s_1, t_1)$, $\eta_2 = (s_2, t_2)$ in yellow and blue respectively. We found that the flow from s to t could determine the assignment solution of nets. $Net1$ passes through g_{11}, g_{22} and g_{34} in 3 layers and correspondingly 3 TSVs are assigned to $Net1$.

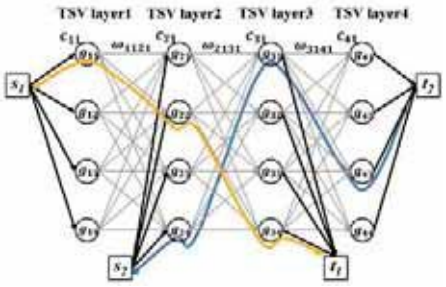


Fig. 3. An instance of TSV assignment of two nets on the IMCMC network

Fig. 4 shows our multilevel assignment flow. The input contains grids and netlists. Generating the IMCMC network, we manage the coarsening and uncoarsening steps to get the initial solution meanwhile optimizing by conditional grid extension. Then the assignment in HVH structure is made. By applying the thermal module and inserting thermal TSVs, we estimate the heat distribution and reduce the temperature. Finally, making revision to pins location to counteract the error.

III. Multilevel Flow Assignment

In the multilevel problem, the most intractable point is to optimize the wire connection in the most efficient method.

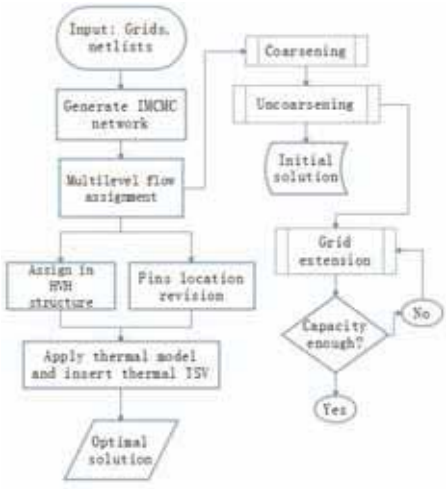
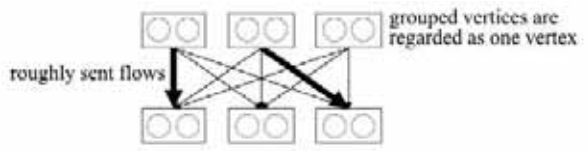
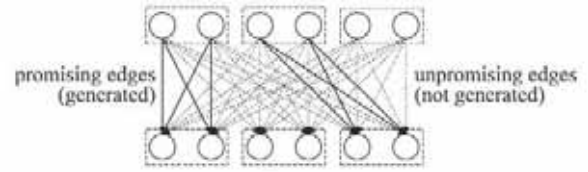


Fig. 4. Algorithm flow of multilevel assignment

Considering the problem we propose **coarsening and uncoarsening algorithm** in the IMCMC network to accomplish the target of reducing the number of edges in calculation. For a chip composed of $k \times k$ grids denoted as level ϵ, we make the adjacent grids into a group denoted as grd^{ϵ}_{ij}. After coarsening for one time, we get $\frac{k}{2} \times \frac{k}{2}$ grids, that is to say, in the coarsening level ϵ the grids become $\frac{k}{2^{\epsilon-1}} \times \frac{k}{2^{\epsilon-1}}$ until 2×2.



(a) Flows are roughly sent on grouped vertices



(b) Only effective edges are generated from rough flows

Fig. 5. Multi-level algorithm

In the flow assignment, we define the grids have positive capacity and could contain TSVs as **effective grids**, effective and ineffective grids are both included in the coarsened graph but only effective grids could generate edges. In the IMCMC network, only a small part of edges are actually used. So we define **promising edges** that have higher possibilities to be used in the solution. Only the promising edges could be generated between effective grids in the uncoarsening step and the unpromising edges are not generated. We send rough flow between the vertices and decide the effective grids in which TSV is inserted in Fig. 5(a). Then promising edges are generated in the next uncoarsening step in Fig. 5(b) but the

unpromising ones aren't. Thus the number of edges reduces. As Fig. 6 shows, promising edges are generated for each effective grid barring the ineffective ones (grey balls).

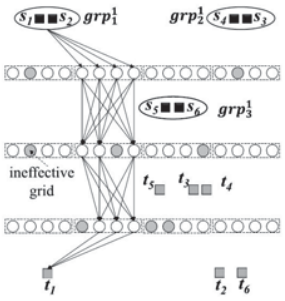


Fig. 6. Edges generating in coarsening graph

A. Conditional Grid Extension

In this solution, the number of effective grids during uncoarsening is usually limited, we cannot find a proper grid to insert TSVs in the adjacent regions, which results in attempts to other grids for more times. In this way, the solution quality would be largely degraded including wirelength could be overestimated and excessive time could be used. Therefore, we propose a **conditional grid extension** to extend the inspective region to get enough effective grids. Denoting the **extending radius** as r, the edges are generated for the effective grids in the region of radius r.

Related research was already done in [7], residual flow will be assigned to other effective in the extended region if the capacity in the initial region isn't sufficient. But the grid extension it proposed sacrifices too much resource (storage and time) extending effective grids. That results from the method is conducted in an iterative way and in the high-level coarsened graph the recursion runs more than needed. The value of r is set to be conditional means the number of grids each level is different, the extending extent also differs. Therefore, we set the different number to r according to the level number in a sequential order. We get the r array by a great number of data statistics to make it satisfy the circuits of different grid size as much as possible. Fig. 7 gives an example, the grids with black point are effective grids. At first, only an initial effective grid exists, so we extend the detective region by Alg. 1 radius by radius. The extension time is 3, 2, 2, 2, 1, 1 as the coarsening level increases. More and more effective grids are included and the capacity is sufficient.

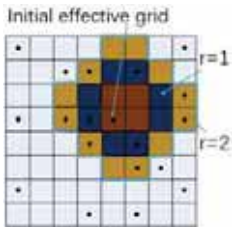


Fig. 7. Conditional grid extension

Algorithm 1 Conditional Grid Extension

Input: $grid(x_0, y_0)$
Output: the set of extended grids $V = (x, y)$

```
d_max_b[] = {0, 3, 2, 2, 2, 1, 1, 0, 0}
d_max ← d_max_b[level]
d ⇐ 0
for d ⩽ d_max do
    i ⇐ 0
    for i ⩽ d do
        V ⇐ V + (x_0 − i, y_0 + d + 1)
        V ⇐ V + (x_0 + i + 1, y_0 + d − i + 1)
        V ⇐ V + (x_0 − i, y_0 − d + i)
        V ⇐ V + (x_0 + i + 1, y_0 − d + i)
        i ⇐ i + 1
    end for
    d ⇐ d + 1
end for
return V
```

B. HVH . . . VH Channel Routing

In the discussion and simulation of net-routing problem, mathematical model is often used to set the routing rules and constraints. As what we discussed last chapter, grid model is frequently used in the assignment. But most of previous works didn't give detail solution of routing rules so that obvious congestion exists on the layers. Providing **H** as horizontal , **V** as vertical, we make use of HVH routing model of **reserved model** in which segments are restricted to particular layers. In the multi-layer IC, the rule is, if the first layer is reserved for vertical components, the next layer is reserved for horizontal components, next is vertical, and so on. Fig. 8 shows a clear perspective of circuits using HVH structure. What's more, HVH channel is used not only in different die layers of 3D chip, but in different metal layers of one chip. And HVH routing model has many merits,

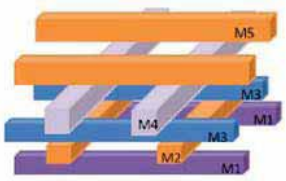


Fig. 8. Metal wire orientation

- The wire width is taken into consideration, defining the measurement of grids, the number of containable wire is confirmed, which makes the solution more realistic.
- It decreases the complexity of circuits, including using less number of vias, because the directions of routing in each layer are determined rather than randomly being assigned.
- The capacitance exists between the wires generated from the electric connection. The capacitance between chip dies is insulated by substrate but still exists between metal

layers in one chip. Obviously, in the HVH structure the capacitance only exists at the position that vertical and horizontal wire cross on the adjacent layers shown in Fig. 9, which is much smaller than randomly routing.

- It makes smaller congestion. Direct connection in Fig. 10 using **Euclidean Distance** [20] neglects the fact that TSV is inserted in the vertical direction. Minimum connection produces giant congestion among different nets due to random routing. HVH channel leads to less unnecessary routes and blocking of nets, which could be easier to get an optimal solution.

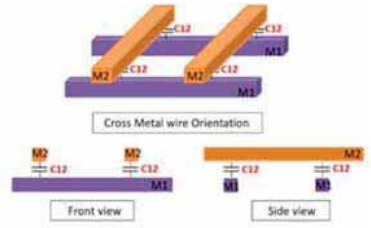


Fig. 9. Cross Wire Orientation

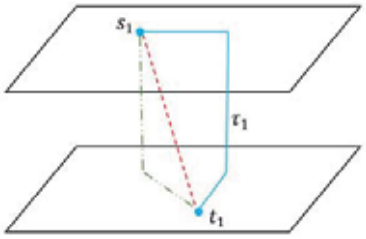


Fig. 10. Connection between layers. Red line is direct connection, green line is minimum connection considered vertical direction of TSV, blue line is HVH channel.

The wirelength in HVH model could be calculated by Manhattan Distance in the bounding box compose of 2 pins as Eq. 3 shows. Because the distance between two layers is constant, it mainly considers the vertical and horizontal distance between two pins.

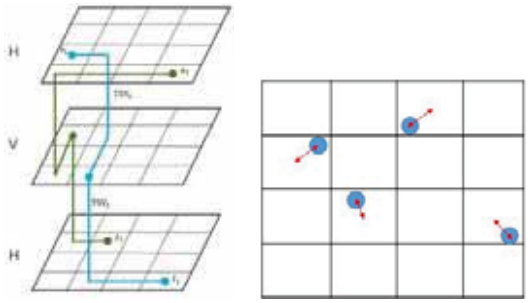
$$\begin{aligned} wl &= \sum_{l=1}^{l_n}(|x_u - x_v| + |y_u - y_v| + h) \\ &= |x_s - x_t| + |y_s - y_t| + l_n * h \end{aligned} \tag{3}$$

Where l is the number of layer, l_n is the number of layer the net passes totally, h is the height between adjacent layers and u, v are edges in the net.

Fig. 11(a) gives an example of routing detail. $Net1$ connects from s_1 to TSV_1 in the horizontal direction, then connects with TSV_2 in the vertical direction and connects with t_1 horizontally. $Net2$ is routed in the same way. Two nets avoid unnecessary overlaps.

C. Modified Locations of Pins

In the coarsening-uncoarsening process, we assign pins and TSVs as well as calculating the wirelength based on the assumption that pins are located at the center of the grid. But it ignores the great number of pins and grid area. Regarding the grid center as the real position will make large error in the wirelength calculation. So we revise the location coordinate by adding the offset distance to pins as Fig. 11(b) shows. We use Eq. 4 to revise the wirelength by adding offset distance.



(a) HVH structure (b) The real location of pins

Fig. 11. Improvement in assignment

$$d_{off} = |node_1\_x - node_2\_x| + |node_1\_y - node_2\_y| \tag{4}$$

IV. Thermal Increase Model and Thermal TSV

A. Overview of the thermal increase model

In the previous 3D resistive thermal-aware research like [3], the model regards the chip as a resistive network. The temperature of a grid is related to the adjacent grid, room temperature, which imitates the electric phenomenon. The heat flowing along the circuits is similar to current going through the resistance so that voltage and current is used to represent temperature and heat respectively.

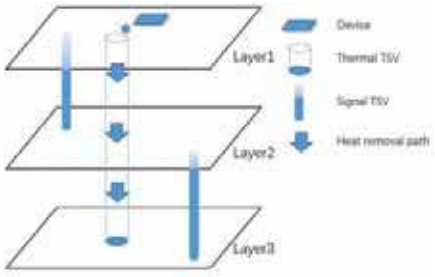


Fig. 12. Heat dissipation path

The model mentioned above is proved to be effective but our model properly takes the effect of layer assignment and thermal TSV into consideration according to the heat dissipation path shown in Fig. 12. And we introduce a concept called thermal resistance, denoted as the ratio of temperature difference between two edges of an object and power of thermal source when heat is transmitting along the object. We could calculate by,

$$R = \frac{T_2 - T_1}{P} \tag{5}$$

B. The effect of layer assignment

We make use of the compact resistive thermal network model [16] to calculate the temperature increase generated from device consuming power. The number of components on the chip is extremely large and each component may produce great quantity of heat so that the number of components in the region is very significant. We specially consider the number of each type of components together with the area. We denote R_j as the thermal resistance of layer j, R_b as the thermal resistance of substrate, d_l as average power dissipation of layer l and L_{max} as the total number of layers. So the amount of temperature increase could be calculated as:

$$T_{rise} = \sum_{l=1}^{L_{max}} d_l(\sum_{j=1}^{l} R_j + R_b) \tag{6}$$

C. The effect of thermal TSV

Here a structure called thermal TSV is introduced. Considering the heat dissipation path, the thermal TSV which passes across the silicon layer in the vertical direction is effective in dissipating the heat generated by components. And the number of thermal TSVs decides the effect of temperature reduction. Provided there are $ttsv_{num}$ thermal TSVs on the grid of 3D-IC and each thermal TSV could reduce temperature by δ. So the temperature reduction in this dissipation path on the grid should be $\delta * ttsv_{num}$ totally.

From the discussion above, the temperature increase of the region is:

$$T_{rise} = \sum_{l=1}^{L_{max}} d_l(\sum_{j=1}^{l} R_j + R_b) - \delta * ttsv\_num \tag{7}$$

Fig. 13 is an example about how to make use of our thermal increase model to estimate the temperature and amplify heat dissipation. We assume that the area of an adder is $10\mu m^2$, the area of a multiplier is $20\mu m^2$, the area of a divisor is $30\mu m^2$ and the area of a thermal TSV is $20\mu m^2$; the power dissipation of an adder is $10\mu W$, the power dissipation of a multiplier is $30\mu W$ and the power dissipation of a divisor is $40\mu W$; the region area constraint is $100\mu m^2$; the thermal resistance of each layer is $0.1°\mathrm{C}/\mu W$ (i.e., $R_1 = R_2 = R_3 = 0.1°\mathrm{C}/\mu W$), and the thermal resistance of the substrate is $0.2°\mathrm{C}/\mu W$ (i.e., $R_b = 0.2°\mathrm{C}/\mu W$); the amount of temperature reduction per dissipation path is $5°\mathrm{C}$.

The temperature increase in this example is calculated using Eq.3:

$$[(40+40+30)*(0.1+0.1+0.2)] + [(10+10+30)* (0.1+0.2)] + [10*0.2] = 61°\mathrm{C}.$$

The maximum total area of components is $80\mu m^2$ in layer 3 ($30\mu m^2 + 30\mu m^2 + 20\mu m^2$). $20\mu m^2$ area is available and exactly only 1 thermal TSV could be allocated so that $5°\mathrm{C}$ could be reduced.

Eventually, through adding the thermal TSVs to the region, the final temperature increase is $61°\mathrm{C} - 5°\mathrm{C} = 56°\mathrm{C}$.

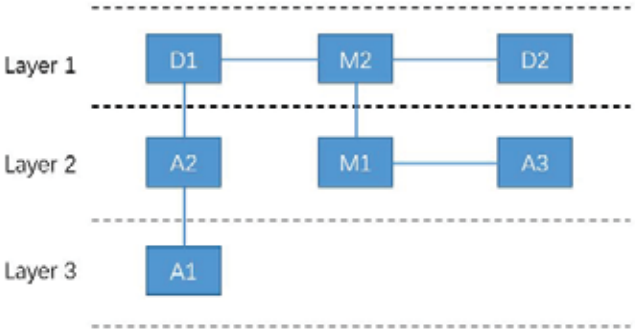


Fig. 13. An instance of application of our thermal increase model

V. Experiment Result

A. Experimental Environment

We test our algorithm using C programming language on the 2.5GHz Intel (R) Core (M) workstation with 8-GB memory. The experiments are carried on two MCNC [17] and four GSRC [18] benchmarks, including ami33, ami49, n50, n100, n200, n300, YMC and the floor-plans are generated randomly. And in all of test-benches we have tried different grid sizes to do the experiments. The data of each testbench is given in Tab. I. And the meaning in the table is as follows:

TABLE I
Floor-plan data

Test Bench	Grid Size	layer	#net	#g_exist	#max_cap
ami33	35×35	3	85	750	5
	48×48	3	85	1418	5
	64×64	4	85	3360	5
ami49	26×26	3	207	416	5
	38×38	3	207	900	5
	52×52	3	207	1657	5
	60×60	4	207	2970	5
n50	33×33	3	382	676	5
	41×41	4	382	1383	5
	60×60	5	382	3722	5
n100	38×38	4	676	1184	5
	49×49	4	676	1975	5
	56×56	4	676	2591	5
n200	28×28	3	848	475	5
	42×42	4	848	1439	5
n300	41×41	4	1246	1355	5
	46×46	4	1246	1739	5
	56×56	5	1246	3224	5
YMC	32×32	4	1500	828	4

- Grid size: Each layer is divided into $g_1 \times g_2$ grids.
- layer: The number of chip dies of 3D-IC.
- #net: The total number of nets in the netlists.
- #g_exist: The number of grids with capacity, which could contain TSVs.
- #max_cap: The maximum capacity of grids.

B. Comparison and Evaluation

We compare our improved algorithm with the previous work [7] in the same model in Table. II. The overlap between the different nets is ignored by most of the previous works,

TABLE II
COMPARISON BETWEEN [7] AND OUR WORK WITH/WITHOUT HVH STRUCTURE AND LOCATION REVISION

Test Bench	Grid Size	#Nets	Previous Works	HVH		Location Revision		HVH & Location Revision	
			Wirelength	Wirelength	Accuracy[1]	Wirelength	Accuracy	Wirelength	Accuracy
ami33	35×35	85	2136	2547	19.2%	2288	7.1%	2699	26.3%
ami49	26×26	207	4356	5189	19.1%	4738	8.8%	5571	27.9%
	38×38	207	5636	6699	18.9%	6010	6.6%	7073	25.5%
n50	33×33	382	9045	10704	18.3%	9711	7.4%	11370	25.7%
	41×41	382	10640	12817	20.5%	11304	6.2%	13481	26.7%
n200	28×28	848	13702	15514	13.2%	15242	11.2%	17054	24.4%
n300	41×41	1246	34123	40390	18.4%	36310	6.4%	42577	24.8%
YMC	32×32	1500	16612	19619	18.1%	19261	15.9%	22268	34%
AVE					18.2%		8.0%		26.2%

[1] $accuracy = \frac{previouswirelength - improvedwirelength}{previouswirelength}$.

TABLE III
EVALUATION OF CONDITIONAL GRID EXTENSION ALGORITHM

Test Bench	Grid Size	#Nets	Without Using		Unconditional		Conditional		
			Wirelength	Overflow[2]	Wirelength	Time (ms)	Wirelength	Overflow	Time (ms)
ami33	35×35	85	2469	28	2142 (0.867)	124	1936 (0.784)	1 (0.036)	78 (0.629)
ami49	26×26	207	4306	23	4385 (1.018)	203	4356 (1.012)	8 (0.378)	179 (0.882)
	38×38	207	5522	111	5420 (0.981)	378	5236 (0.948)	3 (0.027)	221 (0.585)
n50	33×33	382	9178	60	8955 (0.976)	755	8745 (0.953)	4 (0.067)	546 (0.723)
	41×41	382	11569	3	10726 (0.927)	1149	10140 (0.876)	1 (0.333)	719 (0.626)
n200	28×28	848	14171	293	13453 (0.949)	1730	13202 (0.932)	176 (0.600)	1284 (0.742)
n300	41×41	1246	36431	63	33778 (0.927)	7874	33123 (0.909)	13 (0.206)	5405 (0.686)
YMC	32×32	1500	17547	863	16652 (0.949)	1944	16312 (0.930)	769 (0.891)	1447 (0.744)
AVE			1.00**X**	1.00**X**	0.949**X**	1.00**X**	0.918**X**	0.673**X**	0.698**X**

[2] Overflow is the excessive flow that cannot be allocated to effective grids in un-coarsening assignment.

resulted from they mainly focus on the locations of pins. The distance between pins is calculated by **Euclidean Metric**. So the wirelength is actually largely underestimated. In our HVH structure, we restrict the routing to a determined direction layer by layer and establish the bounding boxes. Then we calculate the wirelength using the **Manhattan Distance**. In this way, crosstalk in the routings is effectively avoided though wirelength becomes larger than the previous works. It shows that our algorithm with HVH structure is 13.2% to 20.5%, 18.2% on average more accurate than [7]. And in the assignment procedure, pins are located at the center of the grids in most occasions, which a little differs from the real locations. Error accumulates to a non-negligible value due to the great number of pins. So the location revision step could make the results more accurate, 8% on average compared with the previous works. And applying HVH structure and location revision at the same time, we'll get a result 26.2% more accurate than the formal works.

In Table. III the conditional grid extension and unconditional grid extension together with the previous work [6] are evaluated. By using the grid extension method, we could assign more TSVs on effective grids, which increases efficiency and avoids unnecessary residual path calculation. It shows on average 32.7% overflows are eliminated. And wirelength is also reduced because we can easily get shorter path led by capacity limits. As the chart shows, unconditional grid extension reduces the wirelength by 5.1% and conditional grid extension reduces by 8.2% compared with [6]. Although in some case, like ami49 26×26, the result becomes worse. There is a trend that the circuits with lager grid size tend to get shorter wirelength. Moreover, conditional grid extension is better than the unconditional one, because grid extension is conducted in the iterative way which is so time-consuming and conditional method saves the resources in the high-level coarsened graph. Conditional grid extension method could save the time cost by 30.2% on average.

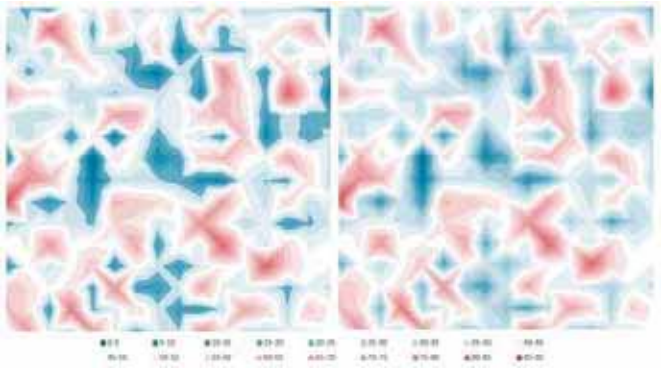


Fig. 14. Evaluate temperature of each grid

As Fig. 14 shows, we detect the heat dissipation on each grid of the circuit YMC 32×32 grids based on the thermal increase model we propose and present the temperature distribution on a 2-D isothermal level graph. The temperature is increasing as the color going to red from blue. According to the heat dissipation path and temperature increase formula, we insert a number of thermal TSV in the proper region. And it is obvious that the color of some regions of the circuits fades proving the heat dissipation path makes sense.

VI. Conclusion

In this work we proposed a multilevel algorithm of assignment and thermal optimization for 3D-IC. We construct a thermal increase model to estimate the heat distribution and reduce the temperature on the board by inserting thermal TSVs. We apply conditional grid extension method to extend more effective grids, which results in 8.2% reduction in wirelength and 30.2% saving in time consumption. The crosstalk of nets is greatly optimized in the HVH structure and wirelength accuracy is risen by 18.2%. Moreover the revision of pins locations realizes 8% improvement in wirelength accuracy.

References

[1] Erik P. DeBenedictis, Mustafa Badaroglu, An Chen, Thomas M. Conte and Paolo Gargini, "Sustaining Moore's Law with 3D Chips," Computer, vol. 50, pp. 69–73, August 2017.

[2] Yuxin Qian, Cong Hao and Takeshi Yoshimura1, "3D-IC Signal TSV Assignment for Thermal and Wirelength Optimization,"2017 27th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS).

[3] Jason Cong and Yan Zhang, "Thermal-Driven Multilevel Routing for 3-D ICs," Design Automation Conference, 2005. Proceedings of the ASP-DAC 2005. Asia and South Pacific.

[4] Hua-Hsin Yeh, Chen-Yu Huang and Shih-Hsu Huang, "Temperature Rise Minimization through Simultaneous Layer Assignment and Thermal Through-Silicon-Via Planning," IEEE Catalog Number: CFP1359B-ART ISBN: 978-1-4799-0667-3.

[5] Jason Cong, Min Xie and Yan Zhang, "An Enhanced Multilevel Routing System," MARCODARPA Gigascale Silicon Research Center (GSRC) and the National Science Foundation under the award CCR–0096383.

[6] Cong, Hao, D. Nan, and Y. Takeshi,"An Efficient Algorithm for 3DIC TSV Assignment," 14th IEEE International NEWCAS Conference (NEWCAS), pp.1C-4, June, 2016.

[7] Cong Hao and Takeshi Yoshimura, "An Efficient Multi-Level Algorithm for 3D-IC TSV Assignment," IEICE TRANS.FUNDAMENTALS, VOL.E100–A, NO.3 MARCH 2017.

[8] Dae Hyun Kim, Krit Athikulwongse and Sung Kyu Lim, "Study of Through-Silicon-Via Impact on the 3-D Stacked IC Layout," IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS, VOL.21, NO.5, MAY 2013.

[9] Xu He, SheqinDong and YuchunMa, "Signal through-the-silicon via planning and pin assignment for thermal and wirelength optimization in 3D ICs," INTEGRATION, theVLSIjournal43 (2010)342C352.

[10] Song Chen, Liangwei Ge, Mei-Fang CHIANG and Takeshi Yoshimura, "Lagrangian Relaxation Based Inter-Layer Single Via Assignment for 3-D ICs," IEICE TANS.FUNDAMENTALS, VOL.E92–A,NO.4APRIL 2009.

[11] Dae Hyun Kim, Rasit Onur Topaloglu and Sung Kyu Lim, "Block-level 3D IC Design with Through-Silicon-Via Planning," 2012 IEEE.

[12] Chris Chu and Yiu-Chung Wong, "FLUTE: Fast Lookup Table Based Rectilinear Steiner Minmal Tree Algorithm for VLSI Design," IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, VOL. 27, NO. 1, JANUARY 2008.

[13] Mitsumasa Koyanagi, Takafumi Fukushima and Tetsu Tanaka, "High-Density Through Silicon Vias for 3-D LSIs," Proceedings of the IEEE, Vol.97, NO.1, January 2009.

[14] Muhannad S.Bakir, Calvin King, Deepak Sekaar, Hiren Thacker, Bing Dang, Gang Huang, Azad Naeemi and James D.Meindl, "3D Heterogeneous Integrated Systems: Liquid Cooling, Power Delivery, and Implementation," IEEE 2008 Custom Integrated Circuits Conference (CICC).

[15] X. Liu, G. Yeap, J. Tao, and X. Zeng, "Integrated Algorithm for 3-D IC Through-silicon Via Assignment," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 22, no. 3, pp. 655–666, 2014.

[16] H.H. Yeh, S.H. Huang, and K.H. Li,"3D IC Design Partitioning for Temperature Rise Minimization," Proc. of IEEE International Microsystems, Packaging, Assembly and Circuits Technology Conference, pp. 447–450, 2011.

[17] CBL, The physical design 1991 benchmark (mcnc benchmarks). http://www.cbl.ncsu.edu:16080/benchmarks/PDWorkshop91/bench/.

[18] Gsrc floorplan benchmarks. http://vlsicad.eecs.umich.edu/BK/BlockPacking/.

[19] https://en.wikipedia.org/wiki/Taxicab_geometry.

[20] https://en.wikipedia.org/wiki/Euclidean_distance