

Module Placement for Power Supply Noise and Wire Congestion Avoidance in 3D Packaging

Jacob Minz, Sung Kyu Lim, Jinwoo Choi, and Madhavan Swaminathan
School of Electrical and Computer Engineering
Georgia Institute of Technology
{jrminz, limsk, jwchoi, madhavan.swaminathan}@ece.gatech.edu

Abstract

In this work, we present an automatic module placement algorithm for simultaneous power supply noise and routing congestion minimization for 3D packaging. We employ decoupling capacitance insertion for noise suppression and 3D global routing for congestion avoidance.

I. INTRODUCTION

The true potential of System-On-Package (SOP) technology lies in its capability to integrate both active and passive components into a single high speed/density 3D packaging substrate. 3D packaging offers an order of magnitude saving in area, delay, and power compared to the conventional PCB and MCM technology. We leverage our recent development of 3D packaging CAD algorithms [1], [2], [3], [4] to develop an automatic placement algorithm for 3D packaging to tackle power supply noise and congestion problem that are seriously threatening the performance and reliability of 3D packaging. Existing approaches consider power supply noise and congestion as an afterthought, which may require excessive amount of decoupling capacitance (= decap) to suppress the Simultaneous Switching Noise (SSN) and additional area/layer to alleviate congestion problem. In addition, many iterations are required between full-length SSN/congestion simulation and manual layout repair until we converge to a satisfactory result. Our goal is to overcome this problem with decap/congestion-aware 3D layout tools.

The following are given as the input to our 3D module placement problem:

- a set of blocks $B = \{B_1, B_2, \dots, B_m\}$ that represent the various active and passive components in the given design
- width, height, and maximum switching currents for each block
- a netlist $N = \{N_1, N_2, \dots, N_n\}$ that specifies how the blocks are connected via electrical wires
- the number of placement layers L in the 3D packaging structure
- the number of power/ground signal layers in the 3D packaging structure

The goal of the 3D module placement with decap/congestion minimization is to find the location of each block in the given 3D package layer such that $\alpha \cdot \text{area} + \beta \text{wirelength} + \gamma \cdot \text{congestion} + \delta \cdot \text{decap}$ is minimized.

II. 3D POWER SUPPLY NOISE MODELING

Active devices draw a large volume of instantaneous current during switching, which causes voltage swings at the power sources. The swings are compounded by the presence of several switching entities that cause simultaneous switching noise (SSN). An active device drawing current from a noisy source is likely to cause logic failures due to its decreased drive capability. Hence in order to ensure a high quality design, SSN must be suppressed. Recent works [5], [6], [7], [8] have addressed the issue of decoupling capacitor allocation and power supply noise suppression for 2D circuits and packages. The exact calculation of power-supply noise is, however, too time consuming to be used for placement optimization.

We model the P/G network for a 3-dimensional placement as a 3D grid graph as shown in Figure 1(a). Each placement layer in the multi-layer placement is represented as a mesh. The mesh is connected by edges which represent the via in the P/G network. The edges in the mesh have inductive and resistive impedances. The mesh contains power-supply points and connection points. The connection points consume currents. The current is drawn from all the sources by the consumers and the current drawn along a path is inversely proportional to the impedance of the path in the power supply mesh. If for a particular block I_1, I_2, \dots, I_N are the currents drawn from N power sources in the grid then $I_1 + I_2 + \dots + I_N = I$, where I is the switching current demand of the block. Then, $I_1 Z_1 = I_2 Z_2 = \dots = I_N Z_N$, where Z_i is the impedances of path i . If $Y_j = 1/Z_j$, then

$$I_j = \frac{Y_j}{\sum_{i=1}^N Y_i} I, \quad 1 \leq j \leq N$$

The current distribution for all blocks can be calculated using the above equations. If $\{P_1, P_2, \dots, P_j\}$ are the current paths under consideration for a consumer block B_k , then the current distribution on the paths can be found by the above equations.

The *dominant current source* for a block is defined as the voltage source supplying significantly more power to the block than any other neighboring sources. The *dominant path* for a block is a path from the dominant supply to the block causing

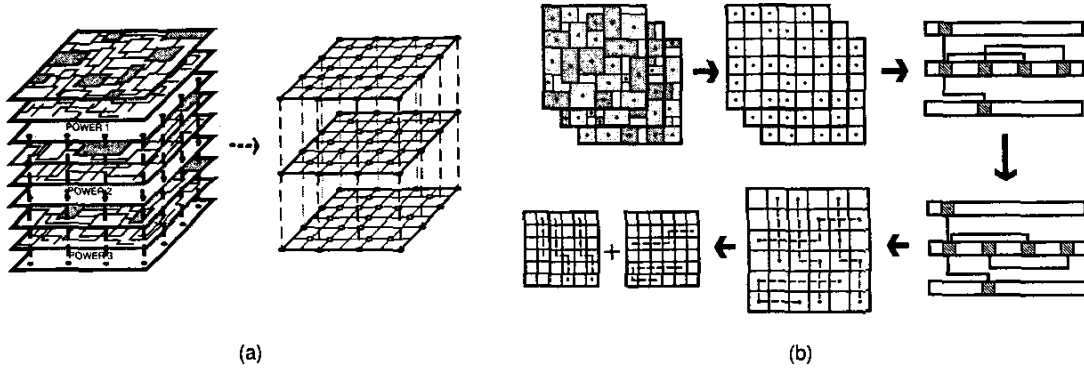


Fig. 1. (a) 3D power supply modeling, (b) 3D congestion estimation

the most drop in voltage. It has been shown experimentally in [5] that the shortest path between the dominant source and the block offers highly accurate SSN estimation within reasonable runtime. In our 3D SSN analysis engine, we compute dominant paths for all blocks. This information is then dynamically updated whenever a new placement solution is evaluated in terms of SSN. Let P_k be the dominant current path for block k . Then $T^k = \{P_j : P_j \cap P_k \neq \emptyset\}$ denotes the set of dominating paths overlapping with P_k (T^k includes P_k itself). Let P_{jk} be the overlapping segments between path P_j and P_k . After the current paths and their values have been determined for all blocks, the SSN for B_k is given by

$$V_{noise}^k = \sum_{P \in T^k} (i_j \cdot R_{P_{jk}} + L_{P_{jk}} \frac{di_j}{dt})$$

where i_j is the current in the path P_{jk} , which is the sum of all currents through this path to various consumer. The weight of i_j and its rate of change are the resistive and inductive components of the path. Let Q^k denote the maximum charge drawn from the power supply by block B_k . If $\theta = \max(1, V_{noise}^k / V_{noise}^{lim})$, where V_{noise}^{lim} is the noise tolerance, the decap allocated to block B_k is given by

$$D^k = \frac{(1 - 1/\theta)Q^k}{V_{noise}^{lim}}, 1 \leq k \leq M$$

where M denotes the total number of blocks to be placed. Finally, the decap cost is given by $D = \sum_{k=1}^M C^k$.

III. 3D ROUTING CONGESTION ESTIMATION

Estimating congestion at a reasonable level of accuracy during the placement process is at least as hard as global routing itself. The congestion profile itself is very sensitive to the placement result, so inaccurate congestion measure can mislead placement easily. The process of global routing for 3D packaging is very different from that of the conventional technologies (PCB, MCM, standard cells) due to the multiple placement layers existing in 3D packaging.

We define *routing interval* to be the interval between two adjacent placement layers. Each routing interval typically contains several routing layers for various connection. There exists two kinds of nets in 3D routing: *i-net* and *x-net*. An *i-net* connects blocks in the same placement layer, whereas an *x-net* connects blocks in different layers. In case of an *x-net*, we decompose it into several segments so that each segment is routed in each routing interval. We model the routing layers in each routing interval with 3D mesh, where each node represents a routing region and each edge represents the boundary between two adjacent routing region. The routing density of an edge is defined to be the total number of nets utilizing the boundary represented by the edge. Let $G_i = (V_i, E_i)$ be the grid graph representing the routing resource for routing interval i . The routing density of an edge e in G_i , denoted d_e^i , is the total number of nets/subnets that use e . Then, $d^i = \max_{e \in E_i} \{d_e^i\}$ is the local congestion in routing interval i . The congestion of the 3D placement with L placement layers is given by $C = \sum_{i=0}^{L-1} d^i$. Finally, the lower bound in the routing layers required for all routing intervals in a 3D package is $\sum_{i \in RI} d^i / cap$, where RI and cap respectively denote the set of all routing intervals and capacity of edges in G_i . A more uniform usage of the grid graph results in low congestion. We perform the following steps to accurately measure C as illustrated in Figure 1(b).

- 1) Net Segmentation: Nets traversing multiple routing intervals are segmented into subnets.
- 2) Pin Generation: Pins are generated for the nets, and the locations where the net enters or exits the routing interval are determined.
- 3) Net Distribution: each net is assigned to a unique routing interval. For example, nets having all its pin in the same placement layer can either be routed in routing interval right above or below.

- 4) Detailed Pin Distribution: The pins are assigned a legal location in the routing interval. Special care is taken to distribute the pins uniformly in the routing interval.
- 5) Topology Generation: A topology for each subnet is generated in the local grid graph.

IV. DECAP AND CONGESTION AWARE 3D PLACEMENT

Simulated Annealing [9] is a very popular approach for module placement due to its high quality solutions and flexibility in handling various constraints. We extend the existing 2D Sequence Pair scheme [10] to represent our 3D module placement solutions. Simulated annealing procedure starts with an initial multi-layer placement. Based on the current configuration, the area, decap, and congestion are calculated for the initial solution. We then make random perturbation (= move) to the initial solution to generate a new 3D placement solution and measure its cost. If the new cost is lower than the old one, the solution is accepted—otherwise the new solution is accepted based on some probability that is dependent on temperature of the annealing schedule. The temperature is decreased exponentially, and the annealing process terminates when the freezing temperature is reached.

Our congestion estimation involves several time-consuming steps unlike decap estimation, which makes it impractical to perform it for every new candidate 3D placement solution. In order to tackle this problem, we perform congestion estimation at every m moves and use these results to interpolate congestion for every candidate solution. Let the values of area, wirelength, decap and congestion at the n^{th} congestion analysis call be $a1, w1, d1, c1$ and at $n + 1^{\text{th}}$ is $a2, w2, d2, c2$. Then, congestion at move b ($n + m < b < n + 2m$) is interpolated as follows:

$$c3 = c2 + \Delta \left[\alpha \left(\frac{a2 - a3}{a2 - a1} \right) + \beta \left(\frac{w2 - w3}{w2 - w1} \right) + \gamma \left(\frac{d2 - d3}{d2 - d1} \right) \right]$$

where $\Delta c = |c2 - c1|$, and $a3, w3$, and $d3$ respectively denote the area, wirelength, and decap results at move b . We note that $c3$ is only an approximate value for congestion at move b . The estimated value of congestion is reasonably close to actual value only if the next move causes a small perturbation of the current one and does not result in value swings of the individual cost parameters. Since there is no sure and easy way to ensure this, the estimated value maybe quite erroneous in practice. Our solution is to allow the value of estimated congestion to vary only within a certain range ($c2 - L < c3 < c2 + L$). This prevents a pathological configuration from becoming the best solution.

V. EXPERIMENTAL RESULTS

We implemented the proposed algorithms and analysis tools using C++/STL. Our program was evaluated using the GSRC benchmarks. We designed several experiments to test the efficiency of our algorithms. In our Simulated Annealing schedule, twenty temperature levels are defined and hundred moves are made in each temperature level. We chose our baseline as the layout optimized for wirelength. We use the following metrics for comparison: area utilization, wirelength, decap amount and congestion. The weights used for the area, wirelength, decap and congestion are 1, 1, 2, and 1 for the multi-objective placement. The results obtained are highly sensitive to the weights of the metrics in the cost function. In our experiment, we use the same parameters across all the benchmarks. However we noticed that the solution quality can be improved by fine tuning the parameters per benchmark. We summarize our observation from Table I as follows:

- Our decap-aware placement gives reduced decap amount (average of 20% and maximum of 25%) for all circuits at the cost of a small decrease in area utilization (19%).
- Our congestion-aware placement alone obtained 3% improvement in congestion. The maximum improvement of congestion is 15%. Surprisingly, our decap-aware algorithm obtains 8% average improvement in congestion. We believe that this is an indication of high correlation between decap and congestion.
- We were able to achieve improvements in both congestion and decap over the baseline with only a slight increase in area (23% on the average) and wirelength (15% on the average) with only a small increase in runtime.

Table II shows power supply noise simulation results for three 3D placement schemes—no-decap aware, decap-aware, and decap-aware+decap-placement. The P/G plane structure size is $246mm \times 254mm$, and the top P/G plane pair was modeled using cavity resonator model [11] and simulated in HSPICE. The placement layer that uses this P/G plan includes 14 active devices. The DC 5V sources are located at four edges in the plane pair and fourteen current sources exist in the plane pair. As can be seen from the Table II, the SSN for the decap-aware algorithm is lower compared to non-decap-aware algorithm. The SSN of the noisiest block blk4 is 1.58V, which is reduced to 1.42V by our decap-aware scheme. With the insertion of decap, the noise is suppressed to 0.22V. In addition, the total amount of decap required for non-decap-aware algorithm is 26.7, which is reduced to 19.9 with our decap-aware scheme. The largest amount of decap is used for blk5 (0.50nF), because of an increase in its SSN after optimization. The numbers show that the SSN was efficiently suppressed and the amount of decap reduced by using our algorithms.

TABLE I
AREA/WIRE-DRIVEN VS DECAP-DRIVEN VS CONGESTION-DRIVEN

| ckts | | | area/wire-driven | | | | decap-driven | | | | congestion-driven | | | |
|-------|------|-----|------------------|--------|-------|------|--------------|--------|-------|------|-------------------|--------|--------|------|
| name | size | lyr | util | wire | decap | cong | util | wire | decap | cong | util | wire | decap | cong |
| n50 | 50 | 4 | 0.81 | 47599 | 26.7 | 23 | 0.57 | 58503 | 19.9 | 25 | 0.44 | 62514 | 49.1 | 23 |
| n50b | 50 | 4 | 0.73 | 45711 | 27.1 | 27 | 0.62 | 52566 | 18.5 | 24 | 0.32 | 61222 | 51.45 | 23 |
| n50c | 50 | 4 | 0.71 | 52804 | 30.0 | 26 | 0.60 | 51638 | 16.4 | 24 | 0.49 | 54306 | 37.0 | 24 |
| n100 | 100 | 4 | 0.78 | 84469 | 90.6 | 40 | 0.59 | 112131 | 75.5 | 38 | 0.58 | 105136 | 106.60 | 44 |
| n100b | 100 | 4 | 0.71 | 69554 | 98.6 | 34 | 0.60 | 88277 | 78.3 | 33 | 0.49 | 89359 | 102.0 | 38 |
| n100c | 100 | 4 | 0.74 | 82728 | 100.6 | 41 | 0.66 | 105769 | 71.7 | 32 | 0.41 | 117786 | 116.9 | 36 |
| n200 | 200 | 4 | 0.81 | 171096 | 226.3 | 87 | 0.64 | 211171 | 209.6 | 67 | 0.26 | 246638 | 264.0 | 82 |
| n200b | 200 | 4 | 0.81 | 181526 | 233.2 | 83 | 0.67 | 221619 | 214.5 | 71 | 0.43 | 249897 | 257.6 | 82 |
| n200c | 200 | 4 | 0.80 | 168831 | 237.4 | 62 | 0.69 | 195118 | 214.4 | 67 | 0.80 | 168831 | 237.4 | 62 |
| n300 | 300 | 4 | 0.84 | 286218 | 393.8 | 100 | 0.63 | 178150 | 382.7 | 90 | 0.61 | 331029 | 402.9 | 94 |
| RATIO | | | 1.00 | 1.00 | 1.00 | 1.00 | 0.81 | 1.12 | 0.80 | 0.92 | 0.62 | 1.26 | 1.25 | 0.97 |
| TIME | | | 32 | | | | 40 | | | | 70 | | | |

TABLE II

POWER SUPPLY NOISE SIMULATION RESULTS. WE REPORT SSN NOISE FOR EACH BLOCK PLACED IN THE TOP PLACEMENT LAYER. (A) NON-OPTIMIZED CASE WITHOUT ANY DECOUPLING CAPACITOR. FROM TABLE I WE NEED 26.7 DECAP UNITS TO SUPPRESS SSN NOISE. (B) OPTIMIZED CASE WITHOUT ANY DECOUPLING CAPACITOR. FROM TABLE I WE NEED 19.9 DECAP UNITS. (C) OPTIMIZED CASE WITH DECOUPLING CAPACITORS INSERTED.

| blk | no decap | decap | decap | blk | no decap | decap | decap |
|------|-------------|-------|-------------|-------|----------|-------------|-------|
| | aware | aware | used | | aware | aware | used |
| blk1 | 1.21 | 1.01 | 0.16 | blk8 | 1.33 | 1.21 | 0.36 |
| blk2 | 1.31 | 1.15 | 0.17 | blk9 | 1.44 | 1.16 | 0.32 |
| blk3 | 1.33 | 1.18 | 0.20 | blk10 | 1.34 | 1.38 | 0.32 |
| blk4 | 1.58 | 1.42 | 0.22 | blk11 | 1.43 | 1.44 | 0.20 |
| blk5 | 1.26 | 1.41 | 0.50 | blk12 | 1.41 | 1.35 | 0.33 |
| blk6 | 1.54 | 1.27 | 0.31 | blk13 | 1.37 | 1.27 | 0.27 |
| blk7 | 1.47 | 1.33 | 0.43 | blk14 | 1.05 | 1.15 | 0.22 |

VI. CONCLUSION

We presented an automatic module placement algorithm for simultaneous power supply noise and routing congestion minimization for 3D packaging. We performed decoupling capacitance insertion for noise suppression and 3D global routing for congestion avoidance. Our ongoing work focuses on thermal issues in 3D packaging layout.

REFERENCES

- [1] J. Minz and S. K. Lim, "Layer assignment for system on packages," in *Proc. Asia and South Pacific Design Automation Conf.*, 2004.
- [2] J. Minz, M. Pathak, and S. K. Lim, "Net and pin distribution for 3D package global routing," in *Proc. Design, Automation and Test in Europe*, 2004.
- [3] R. Ravichandran, J. Minz, M. Pathak, S. Easwar, and S. K. Lim, "Physical layout automation for system-on-packages," in *IEEE Electronic Components and Technology Conference*, 2004.
- [4] P. H. Shiu, R. Ravichandran, S. Easwar, and S. K. Lim, "Multi-layer floorplanning for reliable system-on-package," in *Proc. IEEE Int. Symp. on Circuits and Systems*, 2004.
- [5] S. Zhao, C.-K. Koh, and K. Roy, "Decoupling capacitance allocation and its application to power supply noise aware floorplanning," *IEEE Trans. on Computer-Aided Design*, pp. 81-92, 2002.
- [6] J. Choi, S. Chun, N. Na, M. Swaminathan, and L. Smith, "A methodology for the placement and optimization of decoupling capacitors for gigahertz systems," in *VLSI Design Symposium*, 2000.
- [7] H. Su, S. Sapatnekar, and S. R. Nassif, "An algorithm for optimal decoupling capacitor sizing and placement for standard cell layouts," in *Proc. Int. Symp. on Physical Design*, 2002, pp. 68-73.
- [8] H. Chen, L. Huang, I. Liu, M. Lai, and D. Wong, "Floorplanning with power supply noise avoidance," in *Proc. Asia and South Pacific Design Automation Conf.*, 2003.
- [9] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, pp. 671-680, 1983.
- [10] H. Murata, K. Fujiyoshi, S. Nakatake, and Y. Kajitani, "Rectangle packing based module placement," in *Proc. IEEE Int. Conf. on Computer-Aided Design*, 1995, pp. 472-479.
- [11] N. Na, J. Choi, M. Swaminathan, J. P. Libous, and D. P. O'Connor, "Modeling and simulation of core switching noise for asics," *IEEE Trans. Advanced Packaging*, pp. 4-11, 2002.