Routing Optimization of Multi-modal Interconnects In 3D ICs

Young-Joon Lee and Sung Kyu Lim
School of Electrical and Computer Engineering
Georgia Institute of Technology
email: {yjlee, limsk}@ece.gatech.edu

Abstract

With the rapid advance of enabling technologies, the era of 3D ICs is near. Yet, there are several practical problems to be solved for 3D ICs. Especially, heat removal and power delivery have become two major reliability concerns in 3D stacked IC technology. Recently, liquid cooling based on micro-fluidic channels is proposed as a viable solution to dramatically reduce the temperature of 3D ICs. In addition, a highly complex power distribution network is required to deliver currents to all parts of the 3D IC stack while suppressing power noise to an acceptable level. These technologies pose major challenges to signal routing. With a given amount of silicon area, signal, power, and thermal interconnects compete against one another for routing space. This paper presents an optimization study on signal, power, and thermal networks in 3D ICs. We demonstrate how to optimize various design parameters by adopting design of experiments and response surface method.

I. Introduction

Historically, advances in the field of packaging and system integration have not progressed at the same rate as ICs. In fact, today's silicon ancillary technologies have truly become a limiter to the performance gains possible from advances in semiconductor manufacturing, especially due to cooling, power delivery, and signaling [1], [2]. Today, it is widely accepted that three-dimensional (3D) system integration is a key enabling technology and has recently gained significant momentum in the semiconductor industry. Three-dimensional integration can be used either to partition a single-core or multi-core chip into multiple dies. The average and maximum distance between the transistors are greatly reduced in 3D ICs, which translates to significant savings on delay, power, and area.

Many efforts have been made to solve heat removal and power delivery concerns in the 3D stacked IC technology. Liquid cooling based on micro-fluidic channels is proposed as a viable solution to dramatically reduce the operating temperature of 3D ICs [3]. In addition, designers use a highly complex hierarchical power distribution network to deliver currents to all parts of the 3D IC while suppressing the power supply noise to an acceptable level [4]. These so called silicon ancillary technologies, however, pose major challenges to routing completion and congestion because these large interconnects need to be routed together with billions of smaller signal interconnects (see Figure 1).

This material is based upon work supported by the National Science Foundation under CAREER Grant No. CCF-0546382 and the Interconnect Focus Center (IFC).

The major contributions of this work are as follows:

- To the best of our knowledge, this is the first work to take a holistic approach on how to provide reliable signal, power, and thermal delivery mechanisms. The goal is to co-optimize the on-chip routing geometries for signal, power, and thermal interconnects in 3D stacked ICs under temperature, power noise, and routability objectives.
- Co-optimization of the above mentioned interconnects is a hard task. Considering vast design space with lots of design freedom and little knowledge on the system poses great challenge. We demonstrate how to perform this task using design of experiments (DOE) and response surface methodology (RSM). The design knobs and the assessing metrics are defined, and the correlations among them are discussed.

The remainder of the paper is organized as follows. In Section 2, routing requirements and ways of computing metrics of the three kinds of networks are discussed. Section 3 discusses the details of DOE and RSM. Experimental results are presented in Section 4, followed by the conclusions in Section 5.

II. Three Types of 3D Interconnects

A. Thermal Distribution Network

3D stacked ICs bring several challenges in thermal management. By stacking layers, the heat dissipation per unit volume and per unit horizontal footprint area are significantly increased. Also, the interior layers of the 3D structure are thermally detached from the heat sink. Heat transfer is further restricted by the low thermal conductivity bonding interfaces and thermal obstacles in multiple IC layers. Unlike conventional air-cooled heat sinks, liquid cooling using micro-fluidic channels (MFCs) offer a larger heat transfer coefficient and chip-scale cooling solution. MFCs can be fabricated on bulk side of Si dies, enabling rejection of heat from every layer.

To analyze thermal performance of MFC cooling for 3D ICs, we run numerical simulations. We used the same thermal simulation method as in [5]. A 3D thermal model in [6] is modified to consider the lateral temperature and fluid flow rate distribution caused by non-uniform power/heat flux distribution. Thermal and fluid flow in MFCs are described by energy and momentum conservation equations. Deionized water is considered as a representative working fluid.

978-1-4244-4476-2/09/$25.00 ©2009 IEEE

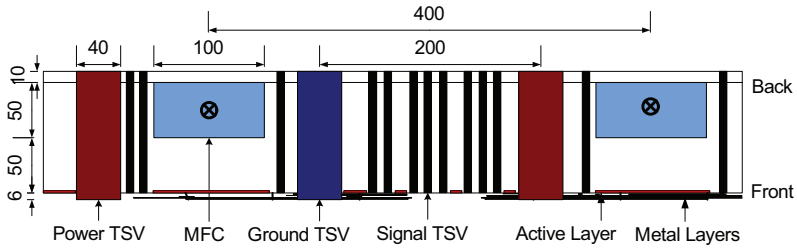

Fig. 2. Side view of a die in a 3D stacked chip. The die is flipped over, and the active layer is facing down. Shapes are drawn to scale. Unit is μm.

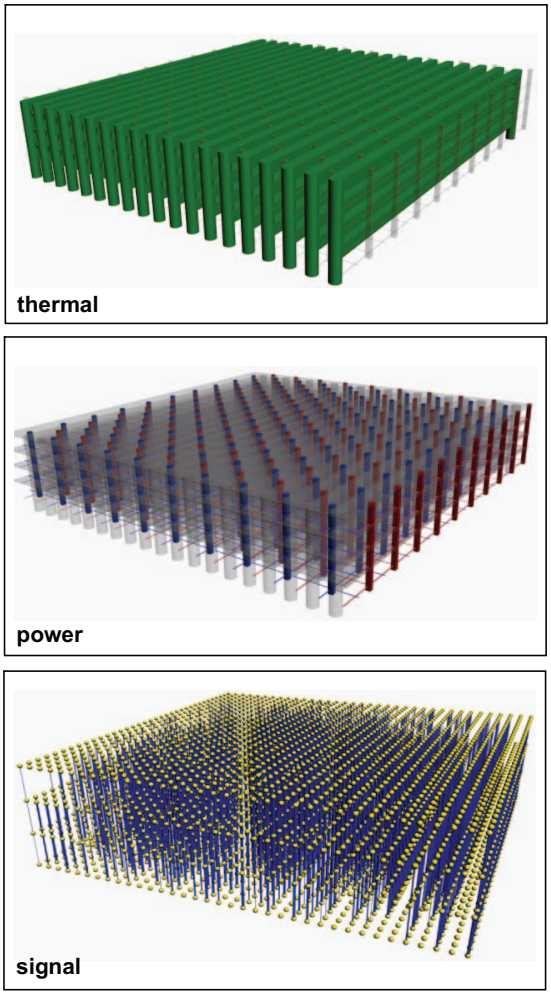

Fig. 1. 3 major types of interconnects in 3D IC. They compete each other for routing space.

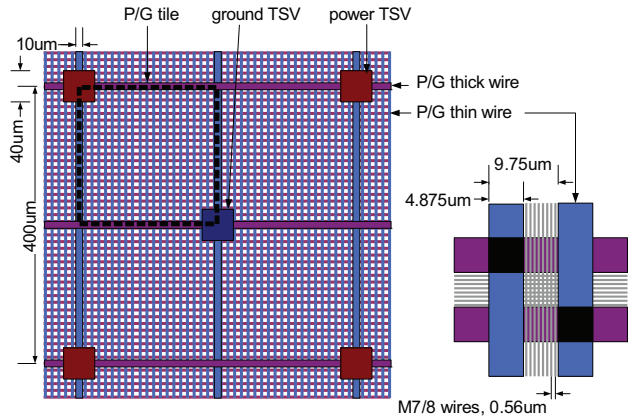

Fig. 3. Top view of our P/G network. Objects are drawn to scale.

The on-chip thermal network is composed of fluidic TSVs, manifolds and MFCs (see Fig. 1). We assume that all the fluidic TSVs and manifolds are located outside the core region in which all gates and metal wires reside. Thus, only MFCs are considered for routing analysis.

TSV connections are made of metal and should not touch any MFCs. The MFCs are wide (around $100\mu m$) and decrease the routing capacity of signal TSVs quite considerably (see Fig. 2). Furthermore, Si dies with MFCs are thicker than the ones without MFCs, and given a TSV aspect ratio, that leads to a larger diameter and lower density of signal TSVs. Thus, care should be taken in designing the width and the depth of MFCs along with the size of signal and power TSVs.

B. Power Distribution Network

In a 3D stacked IC, power should be delivered to all active devices in every layer. The global power distribution network on each die uses grids made of orthogonal interconnects on the top wiring levels. Power is fed from the package through power I/O bumps distributed over the bottom-most die and travels to the upper dies using power/ground (P/G) TSVs and solders [4].

Figure 3 shows the top view of our P/G network. We assume that P/G TSVs are placed regularly in a dual mesh structure. The pitch between two power TSVs is predefined ($400\mu m$) and the same pitch is used for all layers. The diameter of P/G TSVs is $40\mu m$. P/G wires are globally distributed on metal layer 7 and 8. Thick wires have $10\mu m$ width and connect P/G TSVs. And between two thick wires, 20 thin wires are placed. The remaining space is used for signal wires. In our 3D chip, P/G TSVs pierce through all dies for efficient power delivery (see Fig. 2). Thus, no gate cell can be placed at P/G TSV locations. Considering the size of P/G TSVs, this area is not negligible. At routing tiles with pre-placed P/G TSVs, the signal routing capacity is decreased by a large amount.

For a 3D chip with a footprint size of $1cm^2$, we may have thousands of P/G I/Os for each die and millions of

wire segments on the P/G grids in each die. Our power noise analyzer is based on modified nodal analysis (MNA) [7]. We use domain decomposition (DD) [8] to increase the maximum circuit size the analyzer can handle. The DD technique basically decomposes the circuit into several parts and uses a mathematical technique to reduce the time needed for matrix inversion.

C. Signal Delivery Network

With power and thermal interconnects in the system, we perform global routing of signal nets. The reason for global routing, instead of detailed routing, is to obtain quick but accurate pictures of routing congestion so that we can use them in our co-optimization of all 3 types of interconnects. For the signal wires, we used the metal interconnect dimensions similar to Intel's $45nm$ technology node [9]. The TSV integration scheme that we adopted was via-first, and the TSV aspect ratios we assumed were 15:1 or 30:1. Via-first TSVs interfere only with device layer and not with metal layers, so they are less intrusive than via-last TSVs. Although current copper-based TSV technologies require aspect ratio around 5:1 for reliable fill [10], we expect TSV technologies to advance.

Figure 2 shows the side view of a die with all 3 types of wires shown. The diameter of signal TSVs (=$6.67\mu m$) is set to the minimum size to accommodate as many connections as possible. In contrast, the diameter of P/G TSVs is $40\mu m$. Figure 4 shows the routing tile objects. We fix the width of our routing tile to $40\mu m$. Note that some tiles are fully occupied by MFCs and thus are not routable in z-direction. Since a P/G TSV fully occupy one routing tile, the tiles that contain P/G TSVs have no x/y/z-direction routing capacity left. Also note that TSVs are much larger than global interconnects, making them significant routing obstacles in 3D ICs with MFCs.

For each routing tile, there are x-, y-, and z-direction routing capacity values. x- and y-direction capacity represents available routing space in metal layers, while z-direction capacity is for signal TSVs. Basically, x- and y-direction capacity values of a metal layer are calculated from dividing the routing tile size by the pitch of the metal layer. We assume that when the placed cells occupy $x\%$ area of a placement tile, they use $0.8 \times x\%$ of the routing capacity on metal 1-2 of the corresponding routing tile, and the rest $100 - 0.8 \times x\%$ is used for signal routing. Metal 3-6 are dedicated to signal routing. On metal 7 and 8, only part of the space is available due to the P/G nets. The capacity values of metal layers are summed up for each routing tile. If the tile is pre-occupied with P/G TSVs, we decrease the capacity by the area ratio. For z-direction capacity, we calculate the remaining surface area of each routing tile after thermal and P/G routing. Starting from the routing tile area, we subtract the placed cell area and the P/G TSV area. Then, we divide the resulting area by the area of a minimum-pitch signal TSV to get the z-direction capacity of the routing tile.

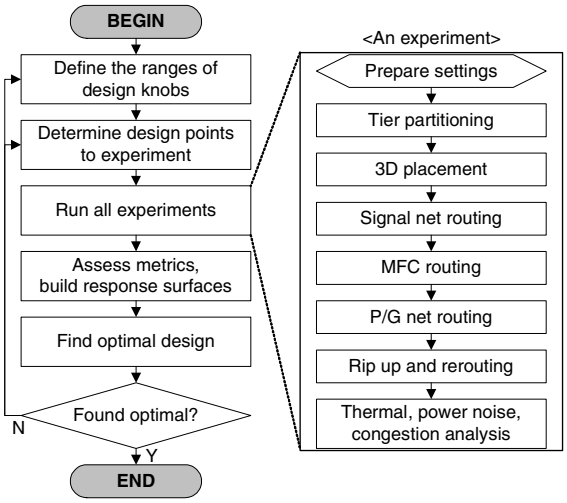

Fig. 5. Overall design flow.

III. Design of Experiments

Since its invention [11], DOE has been used for many science and engineering applications. DOE has been proven to be effective and efficient when an optimization is desired for complex systems with multiple input factors. With the input factors and the region of interest, we assume an underlying metamodel to describe the system. Based on the strategy for covering the region of interest, DOE method suggests the design points to be experimented. After running the experiments and gathering the responses, we draw the response surfaces to understand and optimize the system. This is called response surface methodology [12]. Major benefits of modern DOE techniques are as follows: (1) The number of experimental runs is dramatically decreased compared to that of conventional full factorial design. (2) The knowledge about the target system is not needed. (3) The effect of input factors on responses can be identified.

DOE has also been used in VLSI and CAD community. In [13] the DOE framework for CAD was discussed. A robust interconnect model based on DOE was presented in [14]. And in [15] DOE was used to identify the performance critical buses in a microarchitecture.

A. Overall Design Flow

Overall design flow is summarized in Fig. 5. We start with defining the design knobs (=input factors) and the metrics (=responses). Our single experiment is equivalent to performing gate-level global placement and routing, where we first perform gate-level partitioning and placement, followed by routing thermal, P/G, and signal nets. We then evaluate the metrics of interest and complete the current experiment. Once all the experiments are performed, we construct response surfaces and use them to obtain optimal design solutions. Table I shows our design knobs, and Table II shows our assessing metrics. We use the 3D router presented in [16].

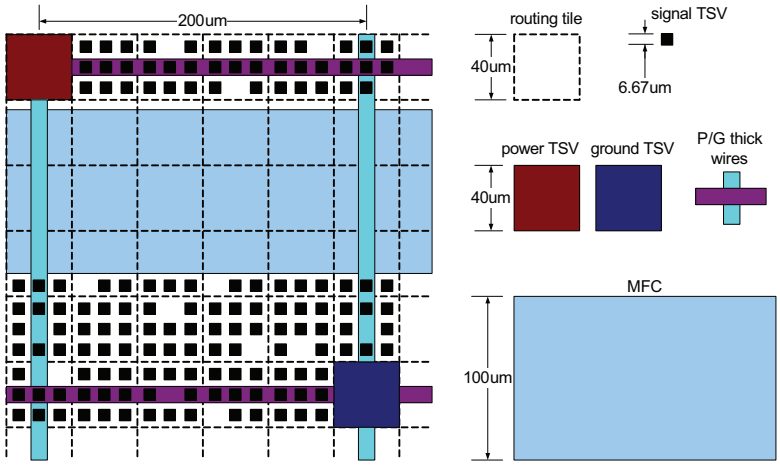

Fig. 4. Top view of routing tiles. Objects are drawn to scale. The sizes of objects may vary depending on the settings of design knobs.

TABLE I
DESIGN KNOBS

Signal TSV AR	We change the aspect ratio (AR) of signal TSVs. z-direction congestion can be alleviated with higher signal TSV AR. However, higher signal TSV AR may incur more expensive manufacturing processes.
Signal TSV penalty	To control the number of signal TSVs in routing stage, we set the z-direction routing edge distance to be multiples of x/y-direction edges. When this is 0, a z-direction edge distance is the same as an x/y-direction one. And when this is 1, a z-direction edge distance has 1 additional distance of x/y-direction edge.
MFC width	The width of MFCs is varied. With the same pressure drop between inlet and outlet, wider MFC means higher mass flow rate and better cooling capability. We assume all MFCs have the same width.
P/G thin wire ratio	This is the the ratio between P/G thin wires and signal wires on metal layer 7/8. For example, if this value is 0.4, P/G thin wires use 40% (=20% each) of the routing tile, and the rest (=60%) is used for signal routing.

TABLE II
ASSESSING METRICS

Total wire-length	All the wirelengths of nets are summed up. This represents the quality of signal routing.
Total # of signal TSVs	The actual signal TSV usage may vary with the availability of routing capacity. If it helps reduce wirelength, the router may use more TSVs.
Routing congestion	Since it is found that z-direction congestion is more severe than x/y-direction ones, we focus on z-direction congestion. The average z-direction utilization for each layer is calculated, and the maximum value among all layers is used to represent routing congestion. 1.0 (=100%) means signal routing failure due to insufficient z-direction capacity between layers.
Max. Si temp.	The performance of the transistors degrades with higher temperature. It is desired that this is less than $85°C$.
Max. power noise	The power noise should not exceed the noise margin of the transistors. Usually it is assumed to be 10% of V_{dd}.

B. Designing the Experiments

We define the range of the design knobs as shown in Table III. We used the *Model-Based Calibration Toolbox* in MATLAB to design the experiments. Stratified Latin hypercube is selected from space filling design styles. We chose the range of MFC width with the consideration of MFC pitch, P/G TSV diameter and pitch to rule out any physical overlap among objects and conform to spacing requirements.

TABLE III
THE RANGES OF INPUT FACTORS.

Input factor	Value
Signal TSV aspect ratio	15:1, 30:1 (discrete)
Signal TSV penalty	0, 1, ..., 6 (discrete)
MFC width (μm)	50-150 (continuous)
P/G thin wire ratio	0.2-0.8 (continuous)

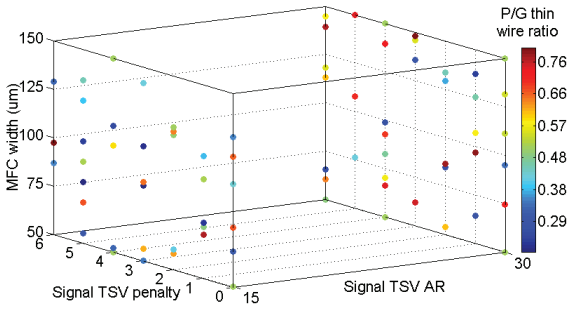

Fig. 6. Selected design points to experiment.

Total 70 design points were generated for experiments. Then, we ran the routing stage and found that 1 design point failed signal routing. Since the failed design point has incorrect response values such as total wirelength and total number of signal TSVs, it degrades the accuracy of response surfaces. Thus we pruned the design point, and the remaining 69 design points were used to construct response surfaces. In addition, 5 design points that are different from the previous ones were generated for model validation purposes. Figure 6 shows the selected design points to experiment.

C. Design Optimization

With multiple assessing metrics and design constraints, there can be multiple optimization scenarios. In this work,

we solve the following problem: minimize the combined cost ($= Cost$) under 100% routability constraint. We want to minimize total wirelength, total number of signal TSVs, maximum Si temperature, and maximum power noise. Each of the metrics under consideration is normalized to $[0, 1]$ and forms a partial cost. Then, we combine them into a single desirability function [17]. The following $Cost$ is used to evaluate the solution:

$$Cost = \sqrt[4]{Cost^*_{wl} \cdot Cost^*_{tsv} \cdot Cost^*_{st} \cdot Cost^*_{pn}}$$

where $Cost^*_{wl}$, $Cost^*_{tsv}$, $Cost^*_{st}$, and $Cost^*_{pn}$ denote normalized total wirelength, total number of signal TSVs, maximum Si temperature, and maximum power noise costs, respectively.

Some of the input factors have discrete levels, thus we made optimization cases for all possible combinations of them. Since signal TSV aspect ratio has 2 levels and signal TSV penalty has 7, total 14 cases were considered. Optimization with respect to $Cost$ was performed for all the cases independently. Sorting the designs with the $Cost$ values in ascending order, we get the best optimal design candidates.

D. Enhancing the Model Accuracy

As a starting point, we used hybrid radial basis function (RBF) as the metamodel class. Linear part was set with a polynomial of order 3 for input factors and order 3 for interaction factors. The RBF kernel that we tried first was *multiquadric*. After running all the design points and gathering the resulting metrics, MATLAB builds the metamodel equations and draws the response surfaces of the metrics. The accuracy of these equations and response surfaces are important because we use them in the optimization process.

We observed that on the corners of the region of interest, the error is relatively higher. To increase model accuracy, we added 8 corner design points that are carefully chosen. It is also found that increasing model order may help to some extent. We checked if increasing the polynomial order of main factors helps improve the model accuracy. Although it depends on which metric we look at, order of 7 for main factors was good enough. The order of interaction factors was kept at 3 because increasing it did not help.

Furthermore, several kinds of RBF kernel functions were tested. We found that for total signal wirelength, total number of signal TSVs, and maximum Si temperature, *multiquadric* is the best in terms of the statistics values such as the coefficient of multiple determination (R^2) and the root mean squared error of prediction error sum of squares (PRESS RMSE). For routing congestion and maximum power noise, Gaussian and thin-plate kernels were the best, respectively. The RBF kernel function equations are as follows:

$$\Phi_{multiquadric}(r) = \sqrt{r^2 + \beta^2}, \beta > 0$$
$$\Phi_{Gaussian}(r) = exp(-\beta r^2), \beta > 0$$
$$\Phi_{thin-plate}(r) = r^2 ln(r)$$

Too many parameters for a model may incur overfitting problem which means the response curve follows not only the underlying truth but also unwanted noise with it. Since the value of PRESS RMSE is not much bigger than RMSE, the response surfaces are not overfitted.

We also ran experiments with 5 validation design points to see how the models predict unseen design points. The resulting validation RMSE was also considered in choosing the best RBF kernel function. The models of responses with chosen RBF kernels were used to make the combined cost ($Cost$). Table IV shows the summary of models for responses.

TABLE V
VARIOUS TECHNOLOGY AND SETTING PARAMETERS USED IN OUR EXPERIMENTS.

Item	Value
Chip size (μm)	4000×4000
Number of dies	4
Bonding type	face-to-back
Si base layer thickness (μm)	50
Bonding layer thickness (μm)	10
Routing grid size (μm)	40
MFC depth (μm)	50
MFC pitch (μm)	400
P/G TSV diameter (μm)	40
P/G TSV pitch (μm)	400
Pressure drop (kPa)	100
Gate oxide thickness (nm)	1
Inductance of package pins (nH)	0.3
Resistance of package pins ($m\Omega$)	3

IV. Experiments Results

A. Experimental Settings and Preliminary Results

We implemented our design package in C++/STL and MATLAB. The simulations were executed on a 64-bit Linux server with two quadcore Intel Xeon 2.5GHz CPUs and 16GB main memory. The technology and setting parameters are shown in Table V. We synthesized a RISC processor circuit named OpenRISC (*or1200*) from [18] using Synopsys Design Compiler and a 45nm technology library from North Carolina State University [19]. The synthesized circuit has about 330K gates and nets. With baseline settings, the x- and z-direction routing edge utilization is shown in Fig. 7. Since layer 1 is the most crowded, we present the layer 1 results. z-direction congestion is more severe compared to x-direction one.

Each hotspot is rectangular with randomly chosen dimensions ranging $40 - 200\mu m$ and a power density of $400 - 800W/cm^2$. The power map was generated after routing stage, and fed into our thermal analyzer. The same power map is also used for power noise analysis. The pressure drop between inlet and outlet of MFCs was assumed to be the same for all MFCs. Inlet working fluid temperature was set to $20°C$. The thermal analyzer was written in MATLAB, and the runtime was about 3 minutes. The resulting maximum Si temperature is shown in Fig. 8. Due to MFCs, the entire chip is cooled down as desired.

For power noise analysis, P/G grid was formed by superimposing RLC mesh structures of a half P/G TSV pitch on each layer. The power consumption at each grid location was modeled as a current source. In order to determine the decap

TABLE IV
SUMMARY OF MODELS. R^2_{adj} IS THE ADJUSTED R^2 STATISTIC.

Response	Model type	# parameters	PRESS RMSE	RMSE	R^2	R^2_{adj}	Validation RMSE
Total signal wirelength	RBF, Poly7, multiquadric	24	294600	225515	0.988	0.985	441849
Total # signal TSVs	RBF, Poly7, multiquadric	25	1191.01	1001.33	0.994	0.995	2192.06
Routing congestion	RBF, Poly7, Gaussian	22	0.00749	0.00609	1.000	1.000	0.01201
Max. Si temp.	RBF, Poly7, multiquadric	17	0.11425	0.10220	1.000	1.000	0.09139
Max. power noise	RBF, Poly7, thin-plate	26	0.26274	0.21853	1.000	1.000	0.23605

TABLE VI
COMPARISON OF DESIGN SETTINGS AND RESULTS. DOE-PREDICTED SHOWS THE OPTIMAL DESIGN SETTING WITH RESPECT TO $Cost$. PERCENTILES ARE THE RELATIVE VALUE COMPARED TO BASELINE.

	baseline	DOE-predicted	DOE-actual
Signal TSV AR	15	15	15
Signal TSV penalty	2	6	6
MFC width (μm)	100	150	150
P/G thin wire ratio	0.5	0.8	0.8
Total signal wirelength (μm)	271,712,320	276,101,125 (101.62%)	276,345,760 (101.71%)
Total # signal TSVs	119,752	98,753 (82.46%)	98,942 (82.62%)
Routing congestion	0.85	0.848 (99.76%)	0.84 (98.82%)
Max. Si temp. ($^\circ C$)	71.205	57.038 (80.10%)	56.644 (79.55%)
Max. power noise (mV)	82.7404	67.5300 (81.62%)	67.2737 (81.31%)
$Cost$	0.35277	0.17250 (48.90%)	0.17049 (48.33%)

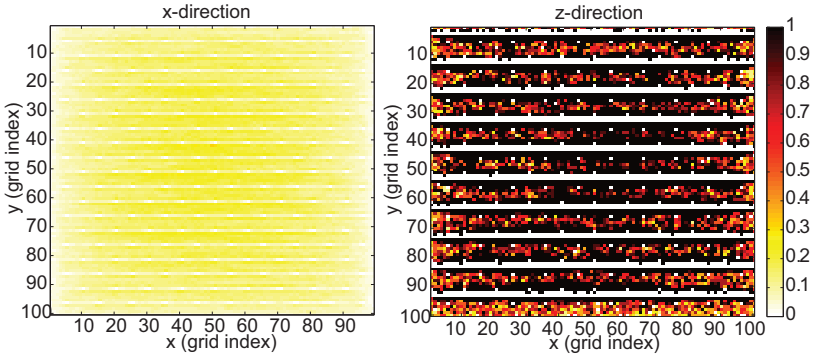

Fig. 7. x- and z-direction routing utilization on layer 1 with baseline settings. White lines in z-direction figure denote MFCs.

area ratio at each power grid point, we calculated the unused Si area per routing tile, and assumed that 80% of the unused area is used for decap. The gate oxide thickness and the parasitic values of package are shown in Table V. We assumed that 1/8 of each layer in the entire stack is turned on at once, and the rise time of current profile was set to $5ns$. After simulation, we gather the peak power noise voltage for each grid point. The runtime of a power noise simulation was about 10 minutes. Figure 9 shows the maximum power noise of the top layer. Power noise is mostly at the switching region.

B. Response Surfaces

Figure 10 shows the comparison between the predicted and the actual total wirelength. It shows that the prediction values match the actual data very well. R^2 was 0.988, which means about 98.8% of design points can be explained by the model. Figure 11 shows the response surfaces of all metrics. For each metric, we show the two input factors per response that are the most influential, except for total wirelength for which three input factors are shown. For total wirelength, the most significant knobs are signal TSV AR, signal TSV penalty and MFC width. And for total number of signal TSVs, signal TSV AR and signal TSV penalty were major influential factors. Routing congestion was mostly influenced by signal TSV AR and MFC width. For maximum Si temperature, the most influential input factor was MFC width, and for power noise, the most influential input factor was P/G thin wire ratio.

C. Correlations Among Knobs and Metrics

When signal TSV AR is 15, z-direction routing congestion is rather severe. Increasing MFC width also increases z-direction routing congestion, resulting in longer total wirelength due to detours. When signal TSV AR is 30, z-direction congestion is less severe, so increasing MFC width does not affect total wirelength much.

A routing grid size in x/y-direction is $40\mu m$ and the one in z-direction is $116\mu m$ which is about 3 times the x/y-direction one. Thus giving signal TSV penalty 2 will make z-direction edge distance almost the same as x/y-direction ones, and we may expect that it will minimize the total wirelength. When signal TSV AR is 15, the total wirelength is minimum at signal TSV penalty of 2, however when it is 30, the total wirelength

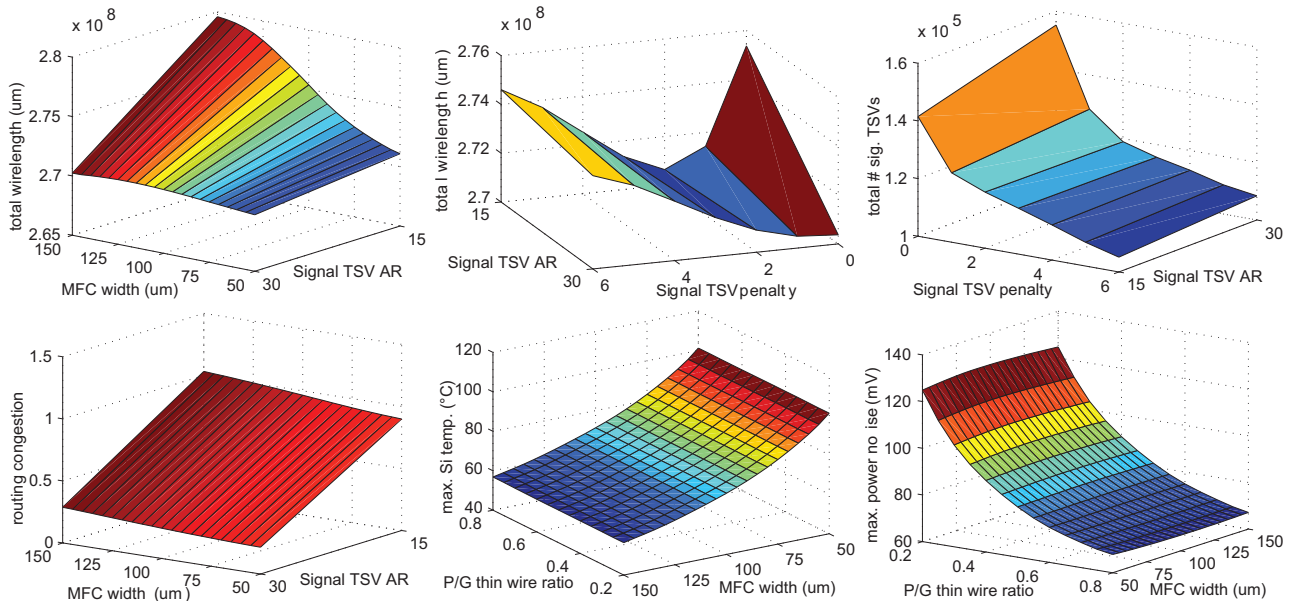

Fig. 11. Response surfaces. Only two significant axes per response are shown, except for the total wirelength for which we show three input factors in combinations.

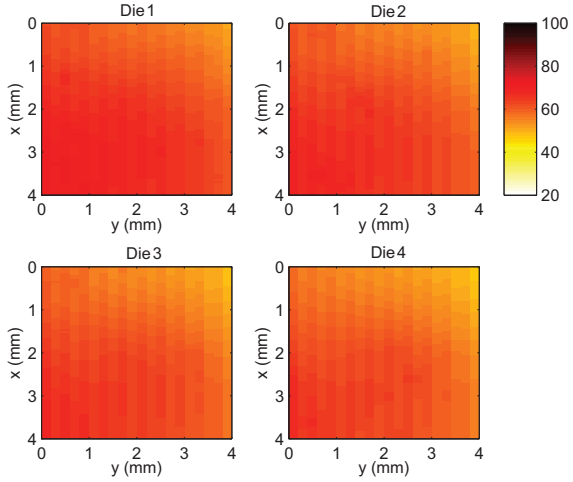

Fig. 8. Maximum Si temperature of layers with baseline settings.

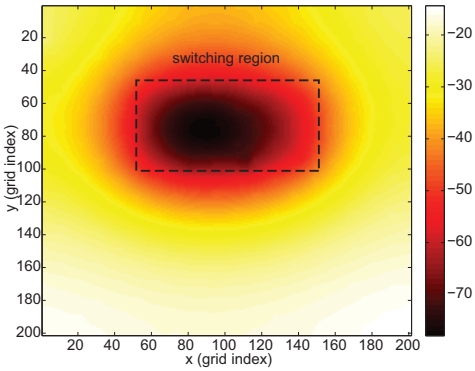

Fig. 9. Maximum power noise map of the top layer with baseline settings. Unit is mV.

is minimum when signal TSV penalty is 0. Total number of signal TSVs is well controlled by signal TSV penalty as shown in top right of Figure 11. When signal TSV AR is 15, signal TSV penalty of 6 does not seem to be a saturation point. Further decrease is possible with higher signal TSV penalty. However, considering total wirelength increase with higher signal TSV penalty, we decided to focus on signal TSV penalty range of 0-6.

If the routing congestion metric is greater than or equal to 1, the signal routing fails. When signal TSV AR is 30, there is no routing congestion, however when signal TSV AR is 15, there is an infeasible region due to this constraint. Yet, this graph is drawn with signal TSV penalty is set to 0. When signal TSV penalty is higher, the router uses less number of signal TSVs, and the routing failure due to congestion goes away.

As shown in the bottom middle of Figure 11, maximum Si temperature is entirely controlled by MFC width. No other input factors were influential. Meanwhile, the graph in the bottom right shows that maximum power noise is mostly affected by P/G thin wire ratio. Although other input factors may affect decap sizes, they did not affect maximum power noise significantly.

D. Optimization Results and Comparison

With the response surface models, the $Cost$ function was formed and minimized. Then, with the optimum design settings we ran the experiments to see the actual result of the optimal design. We compare the following three cases:

- **baseline**: This is based on a typical setting of design knobs.

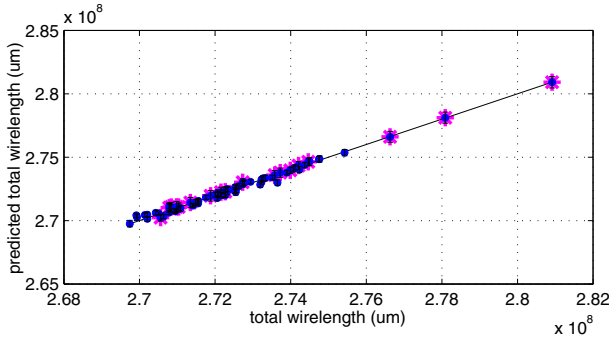

Fig. 10. Comparison of total wirelength between predicted and actual values. The crosshair marks represent RBF centers.

- **DOE-predicted**: This is the optimal design by model prediction.
- **DOE-actual**: This is the actual results obtained from experimenting the optimal setting in DOE-predicted. Comparison between DOE-predicted and DOE-actual reveals the accuracy of the DOE prediction model.

Table VI shows the knob settings obtained from the three cases as well as the comparison of design results. Compared to the baseline, DOE method found a better solution with about 52% less $Cost$. Since increasing signal TSV penalty led to decreased number of signal TSVs, the optimal signal TSV penalty was the maximum of its range, though it also increased total wirelength. And P/G thin wire ratio also increased to maximum because it decreased power noise and did not exacerbate routing congestion much.

Comparing DOE-predicted and DOE-actual, we see that the DOE prediction was quite accurate on all metrics. The errors between DOE-predicted and DOE-actual per response were around or less than 1%. We can say that the optimized design with DOE and RSM is a good quality solution in terms of $Cost$.

V. Conclusions

In this paper, we presented an optimization study of three types of interconnects in 3D stacked ICs: signal, power, and thermal networks. The effectiveness of optimization based on DOE and RSM has been demonstrated. Carefully tuned response models led to reliable optimization results. These models can be reused if the optimization goal is changed by system designers. Our future work includes the comparison of heat removal efficiency between MFCs and thermal TSVs, and incorporation of optical interconnects in the system.

References

[1] G. G. Shahidi, "Evolution of CMOS technology at 32 nm and beyond," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2007, pp. 413–416.

[2] M. Bakir, B. Dang, and J. Meindl, "Revolutionary nanosilicon ancillary technologies for ultimate-performance gigascale systems," in *Proc. IEEE Custom Integrated Circuits Conf.*, 2007, pp. 421–428.

[3] D. Sekar, C. King, B. Dang, T. Spencer, H. Thacker, P. Joseph, M. S. Bakir, and J. D. Meindl, "A 3D-IC Technology with Integrated Microchannel Cooling," in *Proc. Int. Interconnect Technol. Conf.*, 2008.

[4] G. Huang, D. Sekar, A. Naeemi, K. Shakeri, and J. D. Meindl, "Compact Physical Models for Power Supply Noise and Chip/Package Co-Design of Gigascale Integration," in *IEEE Electronic Components and Technology Conf.*, 2007, pp. 1659–1666.

[5] Y. J. Lee, Y. J. Kim, G. Huang, M. Bakir, Y. Joshi, A. Fedorov, and S. K. Lim, "Co-Design of Signal, Power, and Thermal Distribution Networks for 3D ICs," in *Proc. Design, Automation and Test in Europe*, 2009.

[6] J.-M. Koo, S. Im, L. Jiang, and K. E. Goodson, "Integrated microchannel cooling for three-dimensilonal electronic architecture," *J. Heat Transfer*, vol. 127, pp. 49–58, 2005.

[7] C.-W. Ho, A. E. Ruehli, and P. A. Brennan, "The Modified Nodal Approach to Network Analysis," *IEEE Transactions on Circuits and Systems*, vol. 22, no. 6, pp. 504–509, June 1975.

[8] Q. Zhou, K. Sun, K. Mohanram, and D. C. Sorensen, "Large power grid analysis using domain decomposition," in *Proc. Design, Automation and Test in Europe*, vol. 1, 2006, pp. 1–6.

[9] K. M. et al., "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," in *IEEE International Electron Devices Meeting*, 2007, pp. 247–250.

[10] S. J. K. et al., "Wafer-level 3D integration technology," *IBM Journal of Research and Development*, vol. 52, no. 6, pp. 583–597, 2008.

[11] R. A. Fisher, *The Design of Experiments*. London: Oliver and Boyd, 1935.

[12] R. H. Myers and D. C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley and Sons Inc., 1995.

[13] F. Brglez and R. Drechsler, "Design of Experiments in CAD: Context and New Data Sets for ISCAS'99," in *Proc. IEEE Int. Symp. on Circuits and Systems*, vol. 6, 1999, pp. 424–427.

[14] Q. Zhang, J. J. Liou, J. McMacken, J. Thomson, and P. Layman, "Development of Robust Interconnect Model Based on Design of Experiments and Multiobjective Optimization," *IEEE Transactions on Electron Devices*, vol. 48, no. 9, pp. 1885–1891, Sep. 2001.

[15] V. Nookala, Y. Chen, D. J. Lilja, and S. S. Sapatnekar, "Microarchitecture-Aware Floorplanning Using a Statistical Design of Experiments Approach," in *Proc. ACM Design Automation Conf.*, 2005, pp. 579–584.

[16] M. Pathak and S. K. Lim, "Thermal-aware Steiner Routing for 3D Stacked ICs," in *Proc. IEEE Int. Conf. on Computer-Aided Design*, 2007, pp. 205–211.

[17] G. Derringer and R. Suich, "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, vol. 12, no. 4, pp. 214–219, 1980.

[18] D. Lampret, "opencores.org." [Online]. Available: http://www.opencores.org

[19] N. C. S. University, "NCSU FreePDK." [Online]. Available: http://www.eda.ncsu.edu/wiki/FreePDK