Scientific Research

A Power Grid Optimization Algorithm by Direct Observation of Manufacturing Cost Reduction

Takayuki Hayashi[1], Yoshiyuki Kawakami[1], Masahiro Fukui[2]
[1]Graduate School of Science and Engineering, Ritsumeikan University, Kusatsu, Japan
[2]Department of VLSI System Design, Ritsumeikan University, Kusatsu, Japan
Email: mfukui@se.ritsumei.ac.jp

ABSTRACT

With the recent advances of the VLSI technologies, stabilizing the physical behavior of VLSI chips is becoming a very complicated problem. Power grid optimization is required to minimize the risks of timing error by IR drop, defects by electro migration (EM), and manufacturing cost by the chip size. This problem includes complicated tradeoff relationships. We propose a new approach by observing the direct objectives of manufacturing cost, and timing error risk caused by IR drop and EM. The manufacturing cost is based on yield for LSI chip. The optimization is executed in early phase of the physical design, and the purpose is to find the rough budget of decoupling capacitors that may cause block size increase. Rough budgeting of the power wire width is also determined simultaneously. The experimental result shows that our approach enables selection of a cost sensitive result or a performance sensitive result in early physical design phase.

Keywords: Power Grid Optimization; Manufacturing Cost; Timing Violation; Decoupling Capacitor

1. Introduction

With the advent of super deep submicron technologies, designing stable and dependable physical behavior of LSIs is becoming very difficult and serious problems, due to the IR-drop and the EM. Insertion of decupling capacitances and making wider the power grid wires are most effective for this purpose, but we must pay area penalty which causes cost increase. Conventional approaches [1,2] deal with the chip area or the IR drop as their design constraint or objective function. However, no designer can say adequate goal of the chip area without detailed statistical data about the relations between the chip area and the manufacturing cost. Similarly, no designer can say adequate value of the IR drop constraint without detailed statistical data about the relations between the IR drop and timing error risks. Only experienced manager can indicate those goals and suitable values. Without considering that the manufacturing cost increases exponentially as the chip area increase, it is difficult to develop effective optimization system. Furthermore, there is another aspect of the design optimization. Many conventional power grid optimization algorithms have been proposed [1-5]. But most of them select one metric from IR-drop, EM, wiring congestion, or area. Then it is used for their objective function of the optimization, and other metrics are selected as their constraint

function. We introduced in [6] a new concept of a risk function to deal with those different characteristics of metrics at the natural process of the optimization schedule. Furthermore, we introduced a timing error risk as a direct metrics for optimization instead of IR drop in [7].

In this paper we propose a new efficient and effective power optimization algorithm, appropriate for current large scale chips. It deals directly with the manufacturing cost, which is calculated by the chip area increase caused by inserting the decoupling capacitors. The main design steps of VLSI are composed of system level design, function/logic design, and physical design. The area and timing can be dealt with the physical design phase. Especially, the manufacturing cost information is more effectively optimized in the early physical design phase called floor planning because there is more freedom of shape and size selection of the functional blocks. To reduce the design turnaround time, the power grid optimization is usually divided into two steps, high-level power grid optimization and detailed power grid optimization. The insertion of decoupling capacitors is executed in the high-level optimization to abstracted power grids. After the area of each block is fixed, the detailed power grid optimization makes detailed power grid physical patterns.

Our target is "power grid resource budgeting" in early physical design phase. The power grid resource includes

both of power supply/ground lines and decoupling capacitors. The advantage of our approach is to optimize power/ground supply lines roughly and to insert decoupling capacitors analyzing trade-off between the yield for LSI chip and a manufacturing cost. Since the ground wiring can be treated as well as the power wiring, in the following description, we will explain using only the power supply wiring. As a result, the unnecessary cost-can be eliminated from early design phase, and LSI's design becomes more sophisticated. That is to say, not only the cost of a chip but also IR-drop, EM and wiring congestion are considered simultaneously in this optimization.

The rest of this paper is organized as follows. We discuss the layout model which enables the design exploration in high-level floor-plan in Section 2. In Section 3, we briefly summarize the optimization flow that enables simultaneous optimization of multi-objective optimization. In Section 4, we explain an important concept of risk function which is introduced in the optimization algorithm. The risk function is defined for each objective, the wiring congestion, the EM, the timing error due to IR drop, and the chip cost. Most of these are already proposed in other papers [6,7]. Thus, we spend more space to the chip cost risk. It represents the manufacturing cost characteristic which is associated with the chip area. In Section 5, experimental results are showed and the effectiveness of our proposed algorithm is discussed. The conclusion is stated in Section 6.

2. Layout Model

A power grid model and block layer model are shown in **Figure 1**. The power grid formed the mesh has two-layer structure, the horizontal and the vertical layer. These layers are connected with vias. The power grid optimization is performed by not only changing power wiring width [4] but also insertion of decoupling capacitors. Insertion of decoupling capacitors is effective for reducing IR-drop and inductor noise. Decoupling capacitors are placed in spare area in the block layer [3]. Block layer is mainly covered with standard cells. The ratio of the spare area is preset as the limitation which is able to place decoupling capacitors. The chip area is not increased by insertion of the capacitors if they were placed in the spare area. However the optimization algorithm may require additional decoupling capacitors by increasing the block size. It is necessary to consider resulting in manufacturing cost increase.

3. Multi-Objective Optimization Flow

Power grid optimization is a multi-objective optimization problem. We want to optimize many objectives, IR drop, EM, chip area, wiring congestion and so on simultane-

ously. It is, generally, a very difficult problem. One reason is that each objective has different dimension. Second reason is that each of them has a different characteristic curve. We have introduced a new concept of risk functions to deal with those different characteristics at the natural process of the optimization schedule [6,7]. The risk function represents dangerous condition of LSI implementation by using 0% to 100%. Each objective is converted into the same dimension of risk using the risk function. The shape of the risk function should be carefully defined. The combination of IR drop risk, EM, and wiring congestion risk is defined for each grid. Detail definition of the risk functions are stated in [7]. The risk value of the chip area is stated in this paper. The effectiveness is also shown in the section of experimental results.

Figure 2 shows a flow of the power grid optimization. The optimization is scheduled with the gradient method. First, the power grid circuit is constructed with an RC network and initial values of the circuit elements are given (STEP 1). The dynamic current consumption of each functional blockis pre-determined by an RTL power simulation. It is represented as current sources connecting the power grid nodes of the corresponding area. Then, voltage and current of each nodes and edges are calculated by dynamic circuit simulation (STEP 2). Next, a risk value of each grid is calculated (STEP 3). And the worst and four random grids are selected as the candidates of that they may be improved (STPE 4). Then for each candidate grid, the improvement operation *i.e.*, change of power wiring width or insertion of decoupling capacitors, is examined (STEP 5). And a combination of selection of grids which has the highest value of the evaluation function is selected. If the value of the evaluation
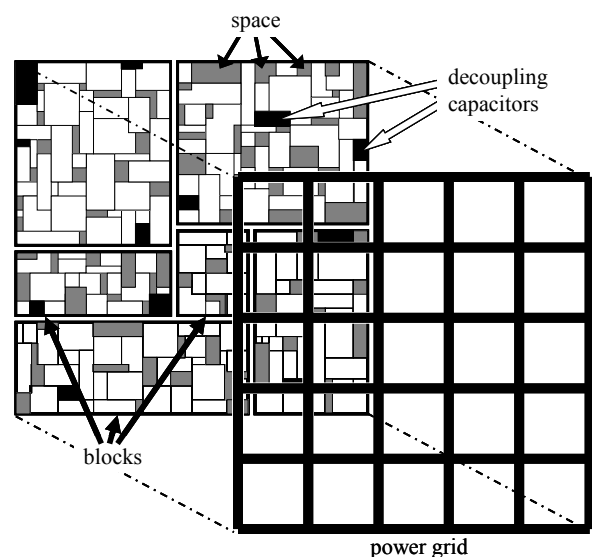

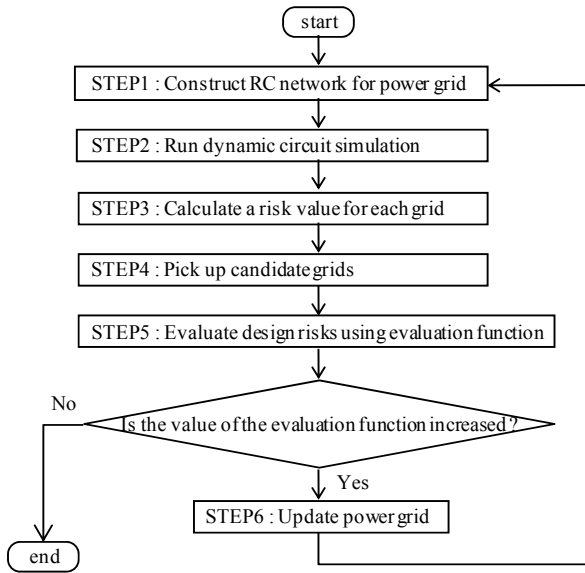Figure 1. Power grid structure.

Figure 2. Power grid optimization flow.

function is increased, the power grid by change of power wiring width and insertion of decoupling capacitors is updated (STEP 6). These operations are repeated as long as the value of the evaluation function is improved.

4. Risk Functions

4.1. Manufacturing Cost Risk Function

This chapter explains the relation between chip area and manufacturing cost. We begin with the definition of yield and critical area, and then we discuss about the manufacturing cost risk. The calculation accuracy of manufacturing cost has been increased by using exact critical area and by revealing the relation between critical area and chip area.

4.1.1. Yield

When chip area size increases, the number of chips produced from one slice of wafer decreases. In addition, defective chips due to dust increase. Therefore, we consider a ratio of the number of the good chips which is produced by one slice of wafer. This ratio is called "yield". The yield equation for LSI chip refers to [8,9]. Equation (1) is the yield function and **Figure 3** shows the yield curve. The D_0 is the average value of defect density in a unit area and the A_{cr} in the Equation (1) is size of critical area. We explain in detail by the following chapter.

$$Y = e^{-D_0 A_{cr}} \qquad (1)$$

4.1.2. Critical Area

Critical area [9] is a layout area which has functional defects caused by particle contamination. Refer to Chapter 2 of [9] for a detailed account of the critical area. The

size of critical area is expressed by Equation (2).

$$A_{cr} = \int_{x_{min}}^{x_{max}} A_c(x) f(x) dx \qquad (2)$$

The A_{cr} is average value of the total of critical area, and it is calculated in critical area for all defect sizes. The $A_c(x)$ is calculated in a critical area for defect size x. The $f(x)$ is the defect size distribution function, and it is defined by Equation (3). The x_0 is defined as a minimum spacing in the design rule of LSI chip.

$$f(x) = \begin{cases} \dfrac{x}{x_0^2} & if \ \ 0 < x \le x_0 \\[2mm] \dfrac{x_0^2}{x^3} & if \ x_0 \le x \le x_{max} \end{cases} \qquad (3)$$

Critical area is defined by the sum of short critical area A_{short} and open critical area A_{open}, and is shown in Equation (4).

$$A_c(x) = A_{short}(x) + A_{open}(x) \qquad (4)$$

The $A_c(x)$ of each defect size is calculated. **Figure 4** shows a short and an open critical area. The functional failure by the short happens when defect size x exceeds the wiring spacing. And the functional failure by the open happens when defect size exceeds the wiring width. They are defined as Equations (5) and (6).
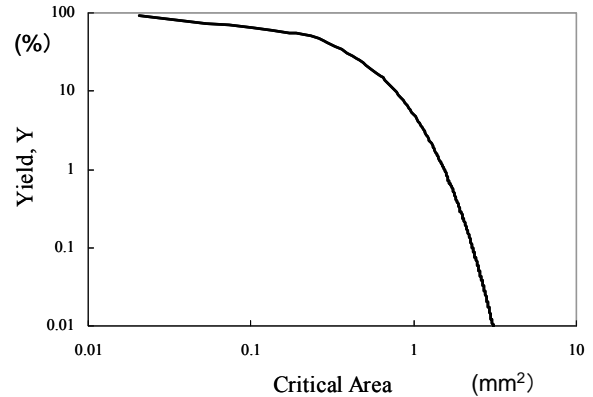

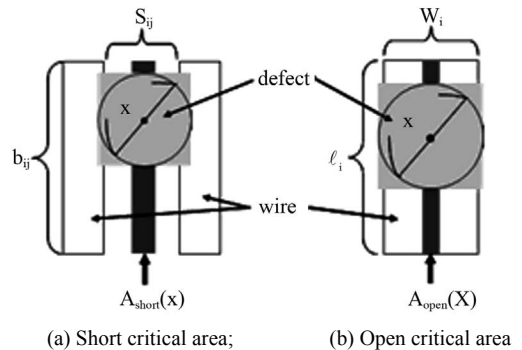**Figure 3. Relation between yield and chip area.**

(a) Short critical area; (b) Open critical area

Figure 4. Short and open critical areas.

CS

$$A_{short} = \begin{cases} 0 & \text{if } x < S_{ij} \\ (x - S_{ij}) \times b_{ij} & \text{if } x \geq S_{ij} \end{cases} \quad (5)$$

$$A_{open} = \begin{cases} 0 & \text{if } x < W_i \\ (x - W_i) \times \ell_i & \text{if } W_i \leq x \leq x_{max} \end{cases} \quad (6)$$

An actual signal wire's width and spacing are not uniform. However, this paper's purpose is the power grid optimization in consideration of manufacturing chip cost. Therefore, we define signal wires are arranged at equal interval, and the number of them is defined depending on the wire congestion in an area. **Figure 5** shows the relation between the chip area and the critical area calculated from signal wires that we defined temporarily. x_{max} of Equation (2) used in this paper defined as a clean level of 0.12 µm.

Generally, if the chip area is only increased without re-design, spacing between the wires will be increased. Thus, critical area does not increase as same ratio as the area increase. It is observed from **Figure 5** that if the chip area is increased three times, the critical area increases about two times. Thus, it is possible to more accurately estimate the chip area, and we can accurately estimate the manufacturing cost described in the next section.

4.1.3. Manufacturing Cost vs Chip Area

The more the chip area increases, the more the yield decreases. The chip cost (=manufacturing cost) function can be represented by the Equation (7).

$$\text{Cost}(A) = \frac{W_{Cost}}{\frac{W_s}{A} \times Y} + a \quad (7)$$

In this Equation, A is a chip area, Y is a yield, W_S is a wafer area, and W_{Cost} is a cost per one wafer. Here, the Y is given by Equation (1). This is influenced by the critical area A_{cr}. Also the A_{cr} is influenced by the chip area A as shown in **Figure 5**. Thus, the Y is a function of A. The term $W_S/A \times Y$ indicates the number of the good quality chips that can be manufactured from one wafer.

The basic cost of the chip is obtained by dividing the W_{Cost} by the term. The parameter a is a cost that does not depends on the chip area. For example, the cost for testing and packaging are included in a. This cost value should be assigned depending on the circuit size and power consumption. The value is not so important for the discussions in this paper, and we set a relatively low cost 25 for a by assuming a low-power and low cost package for our experiment. Consequently the graph of the Equation (7) is shown in **Figure 6**.

We understand that chip costs increase at the same time when chip areas increase, and the area of the chip doesn't necessarily increase because it may be arranged
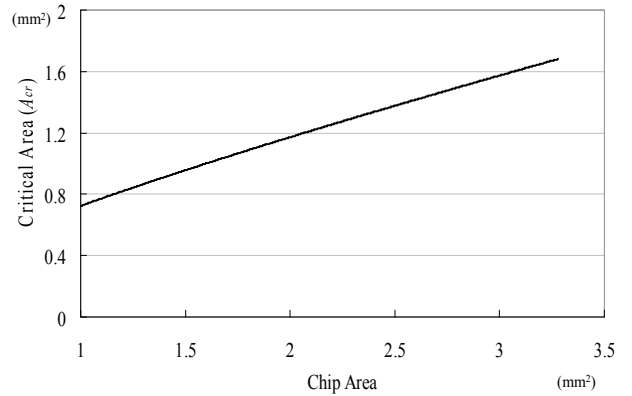

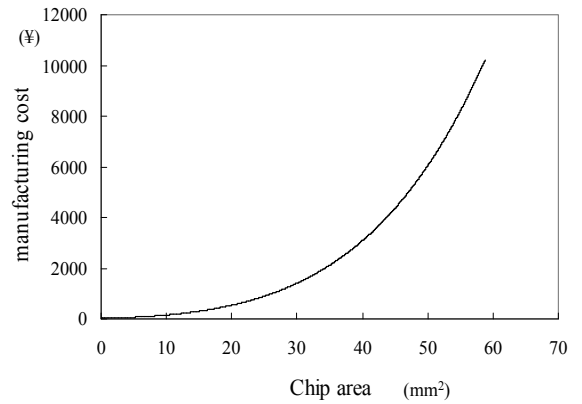**Figure 5. Relation between chip area and critical area.**

Figure 6. Manufacturing cost and chip area.

in the space margin of the block, even if many decoupling capacitors are arranged. However the chip area increases when more decoupling capacitors are arranged than a certain number. Thus, the chip area is able to be represented by Equation (8).

$$A = S + S_{DC} \times (C_D - C_{D_0}) \quad (8)$$

The S_{DC} is the size of a piece of decoupling capacitor, e.g., 0.05 mm × 0.05 mm, C_D is the total number of decoupling capacitors. The C_{D_0} is the number of decoupling capacitors filled into the space area. No area penalty is paid if the placed decoupling capacitors are less than C_{D_0}. The S is initial chip area without decoupling capacitors.

4.1.4. Manufacturing Cost Risk Definition

Chip area increases depending on a total number of decoupling capacitors. Chip cost is proportional to number of decoupling capacitors. Therefore the manufacturing cost risk function becomes a function to change depending on the number of decoupling capacitors arranged to a chip area. The borderline of whether there is a profitable chip or not depends on the quality of chip design. Thus we define the cost risk R_{cost} by the Equation (9), where-

---

MAX$_{cost}$ is the border value.

$$R_{cost} = \frac{Cost(A)}{MAX_{cost}} \times 100 \qquad (9)$$

4.2. Timing Error Risk Function

We have used the IR-drop risk function to eliminate the timing error caused by IR-drop. LSI's functional blocks are constructed by a lot of transistors. To supply enough electrical power to the transistors is essential for desirable operation of LSI because their operational speed degradation is caused by significant IR-drop. The low operational speed causes the timing error of the LSI. Hence, we have used "timing error risk function caused by IR drop" such as in **Figure 7**. This risk function defines the value of IR-drop in the horizontal axis and this risk value is defined as R_{ir}. Details are explained in [7].

4.3. EM Risk Function

EM risk represents the danger of EM [6,7]. In super deep submicron technology, wire width of power grid becomes more complicated and narrower, and power consumption of LSI becomes larger due to increased transistors. The narrow wire tends to cut off by high current density so that larger current flows in the power grid. To design a chip which has high reliability, it is essential to eliminate the danger of EM, and so we have formulated EM risk function.

The EM risk function is defined the current density in a horizontal axis and the EM risk value is defined as R_{em} in the vertical axis. The EM risk is defined with the maximum current density, σ_{max}. The value of R_{em} is 100% when current density is σ_{max}. The value of R_{em} is defined 0% when the current density is less than σ_p. The σ_p is shown in the Equation (10). The risk value between σ_{max} and σ_p is used a linear function proportional to current density.

$$\sigma_P = \sigma_{max} \times \gamma \qquad (10)$$

4.4. Wiring Risk Function

Wiring risk represents the danger of unwired failure on power grid. To optimize the power grid, we need to
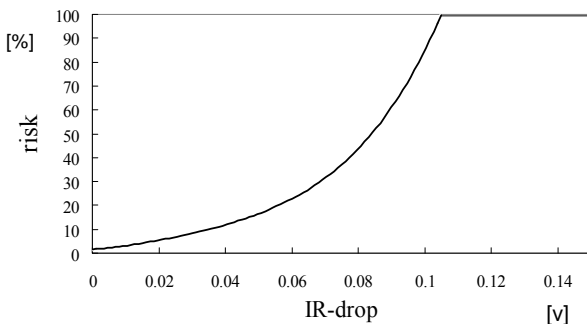

Figure 7. IR-drop Risk Function.

change wire width and insert decoupling capacitors. If the changed wire width is too wide, the LSI chip cannot be designed in desired chip area, and the power grid may become an unfeasible circuit [9]. Thus we have defined wiring risk function to get feasible power grid.

Each grid area includes power supply wires and signal wires. If a ratio of the total wiring area to the grid area increases, unconnected wire may occur. S_g is a grid area, S_p is an area that the power supply wires occupy, S_w is an area of the signal wires, and W_c formulated by Equation (11) represents a ratio of wiring area. The risk value is defined as R_{rw}. When the value of W_c is 0.2 or less, the risk value of R_{rw} is 0. When the value of W_c is 0.6 or more, this risk value of R_{rw} is defined as 100%. The risk value of R_{rw} between 0.2 and 0.6 is represented in shape similar to the EM risk. This function has been obtained by many experimental results.

$$W_c = \frac{S_p + S_w}{S_g} \qquad (11)$$

4.5. Evaluation Function

We define new evaluation function which timing error due to IR-drop, EM, wiring congestion and chip cost are able to be all optimized simultaneously. If one of three risks, timing error, EM and wiring risks, becomes 100%, it is clear not to achieve feasible power grid. Hence we have defined a safety which a risk value is subtracted from 100. We have got a safe function by multiplication of three safeties from timing error, EM and wiring congestion. However, a safety value of manufacturing cost is changed by number of decoupling capacitors arranged in the entire chip. So we have not used manufacturing cost safety to evaluate each grid area. The safety function Safe(B) of a grid area B is shown as follows.

$$Safe(B) = \frac{(100 - R_{ir})(100 - R_{em})(100 - R_{rw})}{100^2} \qquad (12)$$

Since R_{ir}, R_{em} and R_{rw} are within 0% and 100%, Safe(B) is within 0% and 100%.

We define the evaluation function by a minimum safety value, the sum of safety value of all grids and safety value of manufacturing cost risk. A value of manufacturing cost risk is decided by the number of decoupling capacitors arranged in the whole chip. And a value of cost risk is not derived in each grid. So the manufacturing cost element is not added in the safety function. Instead, we can add a safe degree of manufacturing cost risk to the evaluation function because the evaluation function represents a safety of entire grid area.

Our evaluation function F(B) is defined as follows.

$$F(B) = M \times \min_B Safe(B) + \sum_B Safe(B) + N \times (100 - R_{cost}) \qquad (13)$$

The evaluation function $F(B)$ consists of three of safeties. It is necessary to raise the minimum value of the safety of each grid in order to lead more safely for the entire power grid. The first term (1st safety) is the lowest safety in the entire power grid. The second term (2nd safety) is the total sum of the safety of each grid. A good solution cannot be obtained only by the 2nd safety. The 2nd safety is a role for preceding the optimization schedule smoothly. The reason why M is multiplied to the minimum Safe(B) is to improve a worst evaluated grid preferentially. Thus the M should be a big value to make the 1st safety bigger enough than the 2nd safety. The third tern (3rd safety) is the safety value of the manufacturing cost risk.

The 1st safety and the 3rd safety decide the quality of optimization. That is, by controlling the M and N, it is possible to change the quality of the solution. When N is small enough compared with M, the safe value of the manufacturing cost reached almost to 0, and stop at the manufacturing cost limit. When N is gradually increased, the constraint of the manufacturing cost increases, and the safety of the electric and physical constraints (=1st safety) is slightly lower. **Figure 8** shows a transition of each safety to M/N. The data used is 1.3 mm × 1.3 mm sized circuit data described in the next section.

As shown in **Figure 8**, N/M = 0.02 - 0.05 is the range which electrical and physical constraints are dominant because the manufacturing cost does not almost change. On the contrary, N/M = 0.12 is the range which the manufacturing cost constraints are dominant. When the adequate values of M and N are selected, tradeoff analysis is performed well, and it is considered to be able to reach better manufacturing cost result still keeping the good electric and physical conditions.

Figures 9-11 show transition of each safety to each M/N. **Figure 10** shows an example when the tradeoff analysis is performed well. When the N is larger enough, only the cost optimization is performed as shown in **Figure 11**. This is the case that we do not expect. We rather expect the case of **Figure 9** or **Figure 10** in general. These can be performed with less computation time of optimization. We may expect the case of **Figure 10** optimally. But it requires try and errors for selection of adequate M and N values to each LSI chip, and takes large computation time.

5. Experimental Results

We have applied to three different sized circuits to show the effectiveness of our proposal technique. N/M = 0.07 has used in Equation (13). The experiments are performed on P-4 processor with a speed of 3.4 GHz and 4 GB RAM. The program uses C language.
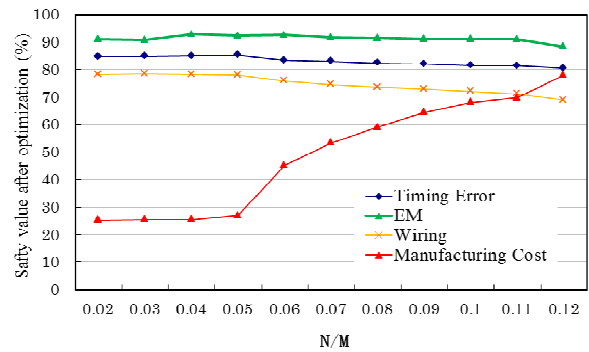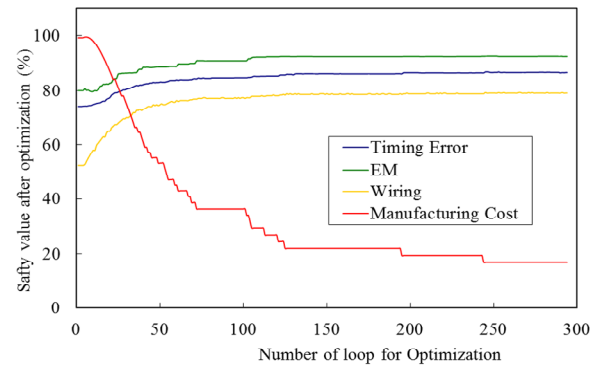

Figure 8. Transition of each safety to *M/N*.

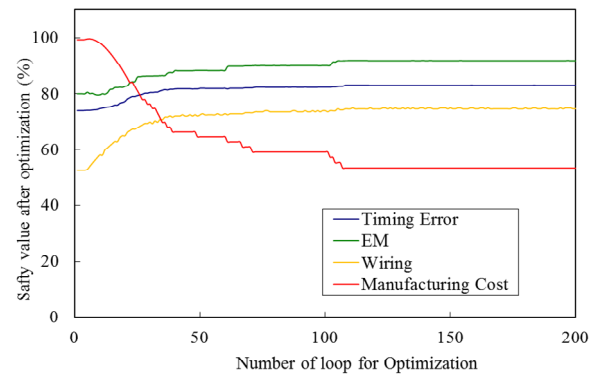**Figure 9. Transition of each safety (*N/M* = 0.05).**

Figure 10. Transition of each safety (*N/M* = 0.07).
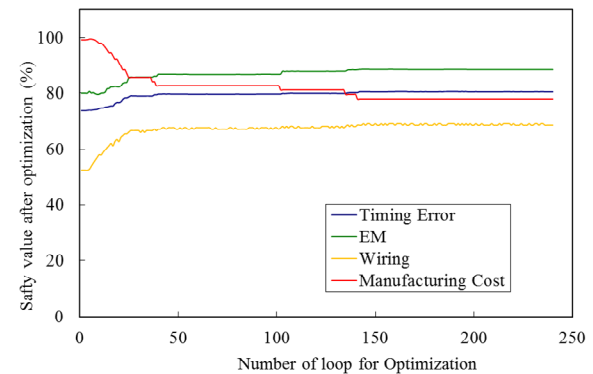

Figure 11. Transition of each safety (*N/M* = 0.12).

The effects of safety improvement are shown in **Table 1**. In the table, "Average safety" is the average safety in the entire chip. IR-drop, EM, wiring and manufacturing cost have been all optimized in this experimentation.

For example, Transition of each safety for CHIP 1 is shown in **Figure 10**. Safety of the manufacturing cost has deteriorated from start to about ten times optimization because the priorities of IR-drop, EM and wiring are higher than the manufacturing cost. However, the safety of manufacturing cost is improved later and finally stabilized on about 100 times in this case. As a result, manufacturing cost is slightly higher, but electrical constraints, IR-drop and EM are satisfied with all. The reason with sufficient electric constraints is an effect of insertion of decoupling capacitors. The distribution of decoupling capacitors placed on chip after the optimization of power grid is shown in **Figure 12**.

Although the processing time of CHIP 1 is 104 sec in **Table 1**, CHIP 3 reaches about 38 times longer than CHIP 1. There is room for improvement for the chip size.

Risk distributions of the whole circuit before and after optimization for CHIP 1 are shown in **Figure 13**. From **Figure 9**, safe degree of EM comes up to the upper limit value soon after the optimization starts. IR-drop and EM risk values are noticeable at grids of both (1,1) and (6,6) in **Figures 13(a)** and **(b)**. Because two power supply resources are connected to the points. They have no voltage drop due to the power supply connections, but EM risk is high because a large current density flows there. The placement of decoupling capacitors has also the same tendency as EM in **Figure 13(b)**. Not so many decoupling capacitors are placed around the power supply resources, and the number of the capacitors increases on the area far away from the power sources.

Table 1. Result of power grid optimization for 3 chips.

Name	Size (mm²)	IR-drop		EM		Wiring		Manufacturing cost	Runtime (sec)
		Average Safety	Improvement (%)	Average Safety	Improvement (%)	Average Safety	Improvement (%)	Increased cost ratio (%)	
CHIP 1	1.3 × 1.3	83.1	9.1	91.7	11.6	74.6	22.1	11.5	104
CHIP 2	5 × 5	79.8	3.6	99.5	0.1	43.2	6.8	3.5	2322
CHIP 3	10 × 10	79.5	5.5	99.9	0.1	39.8	5.7	5.5	38,403
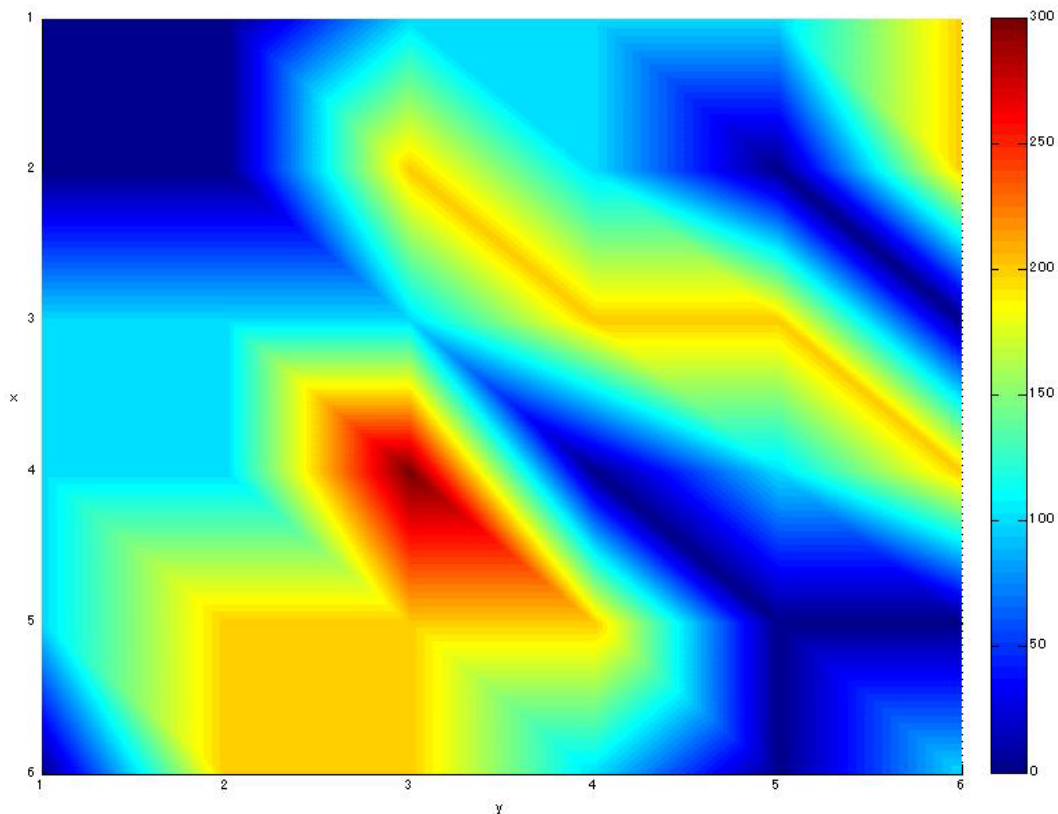
Figure 12. Placement distribution of decoupling capacitors for CHIP 1.

(a) IR Drop

(b) EM
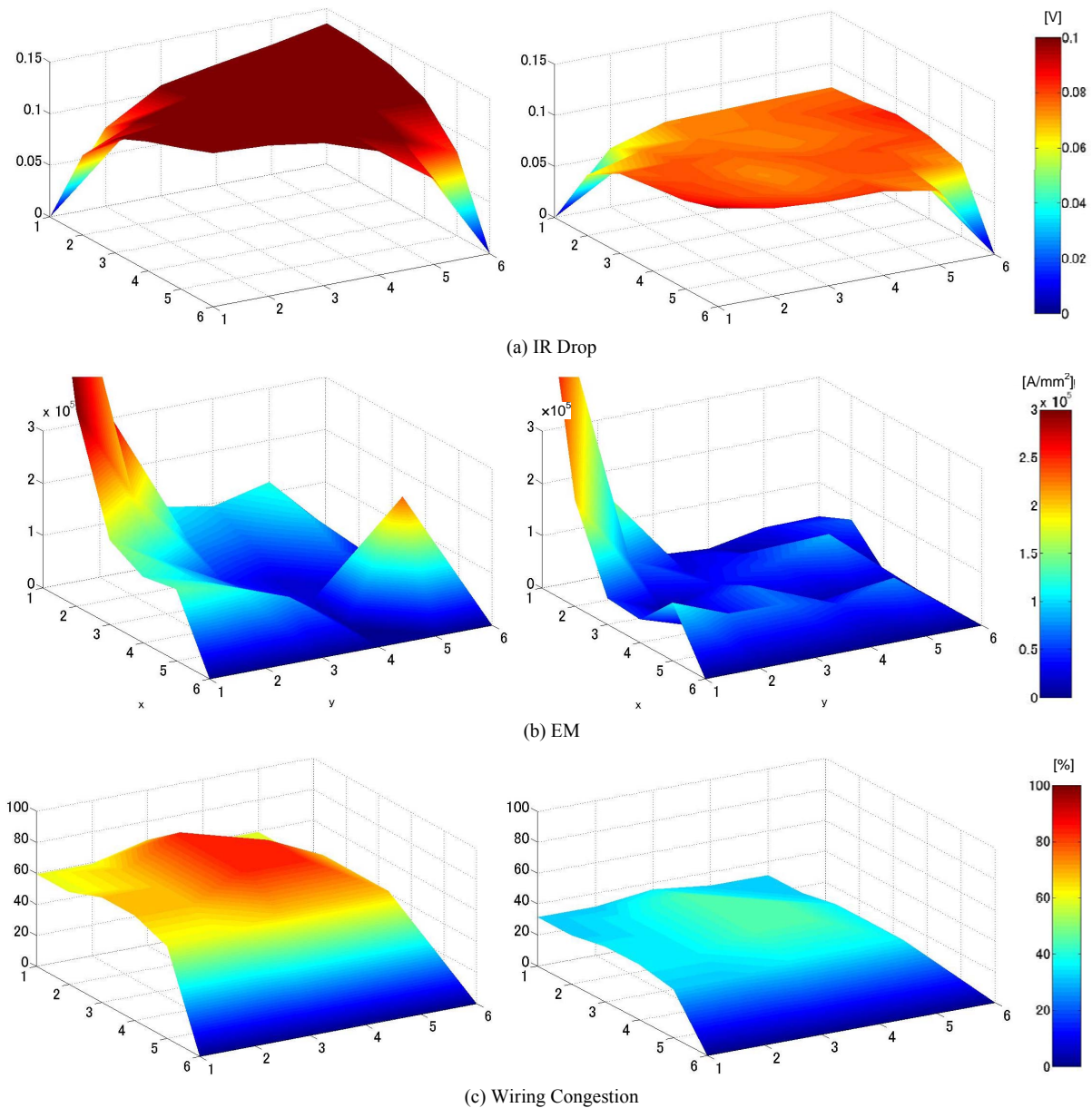

(c) Wiring Congestion

Figure 13. Optimization results for CHIP 1.

6. Conclusion

We have proposed a methodology to optimize a power grid system in the floor plan for the physical design phase. Our approach can consider the risk of physical constraints and timing errors simultaneously, such as IR drop, EM and manufacturing cost. In particular, the manufacturing cost is calculated based on chip yield. In addition, since considering placement of decoupling capacitance at the same time in the chip, power wiring strong against EM noise can be constructed. The experimental results have shown that power wiring optimization can be done while taking the balance of chip reliability and manufacturing cost. However, relatively large chip such as the 10×10 mm^2 will require longer processing time. A key issue for the future is speed up.

REFERENCES

[1] H. Su, K. H. Gala and S. S. Sapatnekar, "Fast Analysis and Optimization of Power/Ground Networks," *International Conference on Computer Aided Design*, San Jose, 5-9 November 2000, pp. 477-480. doi:10.1109/ICCAD.2000.896518

[2] K. Wang and M. M. Sadowska, "On-Chip Power Supply Network Optimization Using Multigrid-Based Technique," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 3, 2005, pp. 407-417. doi:10.1109/TCAD.2004.842802

[3] J. Fu, Z. Luo, X. Hong, Y. Cai, S. X.-D. Tan and Z. Pan, "A Fast Decoupling Capacitor Budgeting Algorithm for Robust On-Chip Power Delivery," *Proceedings of the Asia and South Pacific Design Automation Conference*, Yokohama, 27-30 January 2004, pp. 505-510.

[4] T.-Y. Wang and C. C.-P. Chen, "Optimization of the Power/Ground Network Wire-Sizing and Spacing Based on Sequential Network Simplex Algorithm," *Proceedings of International Symposium on Quality Electronic Design*, San Jose, 21 March 2002, pp. 157-162. doi:10.1109/ISQED.2002.996721

[5] X. Wu, X. Hon, Y. Ca, C. K. Cheng, J. Gu and W. Dai, "Area Minimization of Power Distribution Network Using Efficient Nonlinear Programming Techniques," *International Conference on Computer Aided Design*, San Jose, 4-8 November 2001, pp. 153-157.

[6] H. Ishijima, K. Kusano, T. Harada, Y. Kawakami and M. Fukui, "An Algorithm for Power Grid Optimization Based on Dynamic Current Consumption," *International Association of Science and Technology for Development on Circuits*, *Signals and Systems*, Vol. 531, 2006, pp. 114-119.

[7] Y. Kawakami, M. Terao, M. Fukui and S. Tsukiyama, "A Power Grid Optimization Algorithm by Observing Timing Error Risk by IR Drop," *The IEICE Transactions on Fundamentals of Electronics*, *Communications and Computer Sciences*, Vol. E91-A, No. 12, 2008, pp. 3423-3430.

[8] The Institute of Electronics and Communication Engineers, "LSI Handbook," Ohmsha, 1984.

[9] C.-C. Chiang and J. Kawa, "Design for Manufacturability and Yield for Nano-Scale CMOS," Springer, Berlin, 2007.