Thermal and Power Integrity based Power/Ground Networks Optimization*

Ting-Yuan Wang and Jeng-Liang Tsai
Electrical and Computer Engineering
University of Wisconsin-Madison
Madison, WI 53706, USA
{wangt, jltsai}@cae.wisc.edu

Charlie Chung-Ping Chen
Graduate Institute of Electronics Engineering
& Department of Electrical Engineering
National Taiwan University, Taipei 106, Taiwan
cchen@cc.ee.ntu.edu.tw

Abstract

With the increasing power density and heat-dissipation cost of modern VLSI designs, thermal and power integrity has become serious concern. Although the impacts of thermal effects on transistor and interconnect performance are well-studied, the interactions between power-delivery and thermal effects are not clear. As a result, power-delivery design without thermal consideration may cause soft-error, reliability degradation, and even premature chip failures. In this paper, we propose a thermal-aware power-delivery optimization algorithm. By simultaneously considering thermal and power integrity, we are able to achieve high power supply quality and thermal reliability. For a 58×72 *mesh as shown in the experimental results, our algorithm shows that the lifetime of the optimized ground network is 9.5 years. Whereas the lifetime of the ground network generated by a traditional method is only 2 years without thermal concern.*

1. Introduction

The ever-increasing demands for functionality and speed have pushed the VLSI industry towards more aggressive scaling. Since this trend leads to higher power consumption and current density in interconnects, low-power design has become more and more important. However, the relentless push for low-power design has been directed towards the decrease of supply voltage which reduces noise margin. Power delivery noise is becoming a crucial factor for determining the performance and the reliability of VLSI designs. It has been shown that a 10% voltage drop in a $0.18\mu m$ technology increases the propagation delay of the gate up to 8% [14].

Traditional power/ground (P/G) networks design methodologies aim at minimizing the total wire area subject to Electromigration (EM) and IR-drop constraints. Two main physical design approaches are available for improving the quality of power-delivery: wire-sizing [16] [7] [6] [8] [18] [22] [20] [4] and topology-optimization [13] [5] [17]. The major contributions of these papers are based on the improvement of wire area and computational speed.

Recent studies show that in modern VLSI designs, nonuniform temperature distribution on chips has become more and more serious concern [11] [2] [1]. Without thermal management, thermal problems not only lead to timing failures but also degrade chip reliability. Therefore, the constraints of traditional P/G distribution design which limit maximum current density only depending on EM effect without addressing thermal effects are not sufficient and may be too optimistic. The influence of temperature factor on EM effect due to self-heating (SH) also needs to be considered. Therefore, a new constraint which integrates both EM and SH effects is needed.

The objective function of traditional P/G networks optimizations is to minimize the total wire area. Smaller the wire area, larger the power consumption in the network. Larger power consumption results in higher temperature in P/G networks. Therefore, traditional design methods minimizing total wire area may cause the degradation of chip reliability and timing failure in hot spots. To cope up with the thermal and power integrity issues, the objective function based on the trade-off between power consumption and wire area in networks is needed.

In this paper, we propose an algorithm for P/G networks optimization based on thermal and power integrity. First, a new *self-consistent constraint* is defined and used to replace the EM constraint for the thermal reliability concerns. This self-consistent constraint is based on the idea of finding maximum current density which satisfies both EM and SH effects [9] [10]. This approach is to avoid the unexpected reliability failures in hot spots. Second, the objective function is based on minimizing the sum of each wire's weighted sum of average power dissipation and wire area. This approach is to address power and thermal integrity issues. The formulated optimization problem is convex and results in more reliable P/G structure with comparable wire area. This approach solves the reliability problem of wires in hot spots by giving lower current density and wider wire widths.

The remainder of the paper is organized as follows. In Section 2, reliability and thermal effects are discussed; in Section 3, self-consistent approach are described; Section 4 is thermal-aware P/G networks optimization. Section 5 shows the experimental results, which is followed by conclusion in Section 6.

2. Reliability and Thermal Effects

Thermal effects are inseparable aspects of electrical power distribution and signal transmission through the interconnects in VLSI circuits due to SH caused by the flow of current [2]. Therefore, in P/G networks design, the EM constraint must simultaneously address the temperature factor which is determined by the SH and heat spreading.

* This work was partially funded by National Science Foundation under grants CCR-0093309 & CCR-0204468 and National Science Council of Taiwan, R.O.C., under grant NSC 92-2218-E-002-030.

2.1. Electromigration

EM is the transport of mass in metals under the stress of high current density. This metallization failure is the main reliability concern of high performance IC designs. The lifetime of metal interconnects is modeled by Black's equation [3]:

$$MTF = \frac{A}{j^2_{EM,eff}} exp(\frac{E_a}{k_B T_m}) \quad (1)$$

where MTF is the mean-time-failure, A is a constant which depends on the geometry, $j_{EM,eff}$ is the *effective ac value of current density*, E_a is the activation energy, k_B is the Boltzman's constant, and T_m is the interconnect temperature. Detailed studies have shown that the effective ac value of current density for unipolar stressing is the average current density $j_{EM,eff} = j_{avg}$ [10] [19]. For the case of arbitrary bipolar signals like P/G networks system, the effective ac value of current density is given by the Average Current Recovery (ACR) model [10] [19]:

$$j_{EM,eff} = j_{ACR} = \frac{1}{T}\left[\int_0^T |\, j_+(t)\,|\, dt - \gamma \int_0^T |\, j_-(t)\,|\, dt\right], \quad (2)$$

where j_{ACR} is the current density of the ACR model, γ is the recovery parameter (< 1) of the ACR model. It heuristically accounts for the degree of healing of EM void damage that occurs when the current density changes sign.

If the goal of design is to achieve 10 years of operation lifetime under the current density $j_{EM,ref}$ at temperature T_{ref}, the lifetime restriction in Eqn. (1) must satisfy

$$\frac{exp(\frac{E_a}{k_B T_m})}{j^2_{EM,eff}} \geq \frac{exp(\frac{E_a}{k_B T_{ref}})}{j^2_{EM,ref}}. \quad (3)$$

Then EM constraint can be defined as follows:

- **Electromigration constraint**

$$\begin{aligned} |\, j_{EM,eff}\,| &\leq |\, j_{EM,ref}|\, exp\left(\frac{E_a}{2k_B T_m} - \frac{E_a}{2k_B T_{ref}}\right) \\ &= j_{EM,max}(T_m), \end{aligned} \quad (4)$$

  where $j_{EM,max}(T_m)$ is the maximum current density of a wire with temperature T_m, satisfying the required lifetime under the current density $j_{EM,ref}$ at temperature T_{ref}.

Current density $j_{EM,max}$ is a function of the wire temperature. However, traditional design methods suppose that $j_{EM,max}$ is independent of wire temperature. The resulted design is not robust to the nonuniform temperature distribution.

2.2. Self-Heating

Current flowing through metal wires dissipates heat and increases the wire temperature. This phenomenom is referred to as SH or Joule heating and has become more serious issue due to the introduction of the low-k materials and the increase of 3-D thermal coupling [11].

Figure 1 shows a metal wire with one end connected to the substrate through a contact, in which the width, length, thickness, and thermal conductivity are w_m, l_m, t_m, and $\mathcal{K}_m$. The thickness and the thermal conductivity of the underlying insulator are t_{ins} and $\mathcal{K}_{ins}$. The wire resistivity is temperature dependent and is described as follows:

$$\rho_m(T_m) = \rho_o[1 + \beta(T_m - T_o)] \quad (5)$$

where ρ_o is the wire resistivity at reference temperature T_o and β is the temperature coefficient of resistivity.

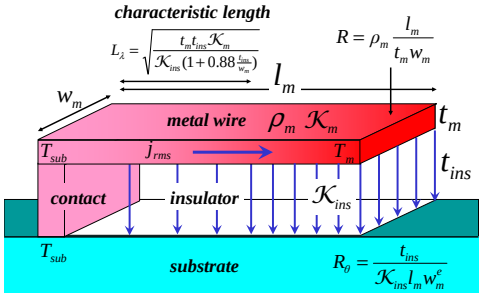

Figure 1. A wire connected to the substrate.

Within a short distance to the contact, the temperature distribution is spatially dependent and temperature will increase from contact end temperature T_{sub} to far end temperature T_m. This distance is defined as *characteristic length* λ [15]. A *thermally-long* wire is a wire whose length is longer than λ. For local wires, most of the wire lengths may not be thermally-long. The temperature increase in these layers is not obvious. However, the temperature increase in global P/G networks cannot be ignored.

The temperature increase in wire compared to the substrate due to SH can be expressed as [15]:

$$\Delta T_{SH} = T_m - T_{sub} = \frac{j^2_{rms} t_{ins} t_m w_m \rho_m(T_m)}{\mathcal{K}_{ins} w^e_m}, \quad (6)$$

where T_{sub} is the substrate temperature and w^e_m is the effective thermal width which can be approximated by $[1 + 0.88(t_{ins}/w_m)]\, w_m$. When $(w_m/t_{ins}) > 0.4$, w^e_m is within 3% accurate. The effective thermal width w^e_m is always greater than w_m, and approaches w_m when $w_m \gg t_{ins}$.

Since EM is the main reliability concern and it is a function of wire temperature, EM must be addressed together with thermal effect which is caused by SH effect.

3. Self-consistent Approach

There is a maximum solution of current density called self-consistent solution satisfying both EM and SH effects [9] [10]. In this section, the proposed self-consistent constraint supplies an integrity approach of EM and thermal reliability.

3.1. Self-Consistent Solution

For a unipolar signal with duty cycle r, the average current density is related to peak current density by $j_{avg} = r j_{peak}$, and the rms current density can be expressed as $j_{rms} = \sqrt{r} j_{peak}$. The relationship between j_{avg} and j_{rms} can be expressed as $r = j^2_{avg}/j^2_{rms}$. For the case of a bipolar signal, the duty cycle is replaced by the effective duty cycle r_{eff} defined as

$$r_{eff} = \frac{j^2_{EM,eff}}{j^2_{rms}} = \frac{j^2_{ACR}}{j^2_{rms}}. \quad (7)$$

The maximum current density satisfying EM reliability and SH effects can be computed by applying $j^2_{EM,eff}$ in Eqn. (3) with equality and j^2_{rms} in Eqn. (6) into Eqn. (7). The self-consistent equation can be expressed as

$$r_{eff} = j^2_{EM,ref} \frac{exp(\frac{E_a}{k_B T_m})}{exp(\frac{E_a}{k_B T_{ref}})} \frac{t_{ins} t_m w_m \rho_m(T_m)}{(T_m - T_{sub}) \mathcal{K}_{ins} w^e_m}. \quad (8)$$

The self-consistent temperature T_m of the wire can be obtained for a given r_{eff}, which is a wave-shape parameter. Once the

the self-consistent temperature T_m is obtained, the corresponding self-consistent current density j_{peak} can also be derived either from Eqn. (6), or Eqn. (3) with equality.

3.2. Self-consistent Constraint

A graphical solution of SH and EM with respect to j_{peak} and T_m is shown in Figure 2 for illustrative purposes. Substrate temperature is assumed to be at $100\ ^oC$. In order to satisfy the required MTF, e.g., 10 years, the current density and wire temperature must satisfy Eqn. (3) under given $j_{EM,ref}$ at temperature T_{ref}, including the line A-B-E for 10 years of lifetime and the area below the line A-B-E for lifetime more than 10 years. For SH effect, the current flowing through the wires and causing temperature increase follows Eqn. (6), which is the curve D-B-C. Note that the MTF on curve D-B-C decreases from point D which has more than 10 years to B which has exactly of 10 years to further less than 10 years at point C. The simultaneous solution of Eqn. (3) and Eqn. (6) is the curve D-B and has maximum values of j_{peak} and T_m at B.

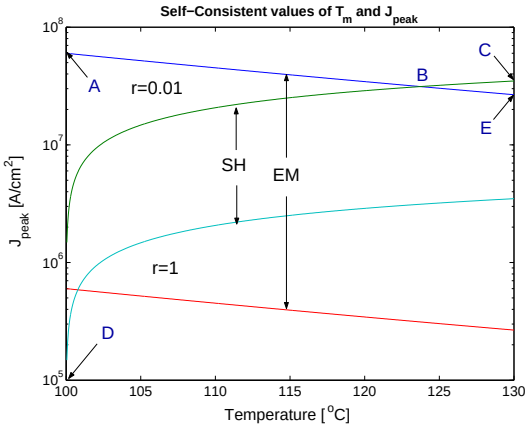

Figure 2. A Graphical solution of EM and SH.

For the traditional design, current density is restricted by EM constraint, Eqn. (4), to satisfy 10 years of MTF under given $j_{EM,ref}$ at temperature T_{ref} without considering wire temperature. The value of $j_{EM,max}$ in Eqn. (4) could be at any point in line A-B-E. If such value is bigger than the self-consistent solution, the MTF of wire is lower than expected due to higher temperature this j_{peak} has generated. If the value of $j_{EM,max}$ is smaller than the self-consistent solution, the resulted wire area may be overestimated because the current density of wires can be bigger. Therefore, we propose a novel approach which replaces the EM constraint with a self-consistent constraint on the P/G networks optimization.

- **Self-consistent constraint**
  Current density of each wire has an upper bound

$$| j | \quad \leq \quad j_{sc}, \tag{9}$$

  where j_{sc} is the current density of one self-consistent solution. For the given duty cycle, all points lie on D-B are solutions, and point B is the maximum solution, Eqn. (8).

4. Thermal-Aware P/G Networks Design

In this section the proposed thermal-aware P/G networks optimization algorithm is presented.

4.1. Problem Formulation

Suppose that a P/G network $G = \{N, B\}$ consists of m non-ground nodes $N = \{1, ..., m\}$ and n branches $B = \{1, ..., n\}$. The currents drawn from the circuit blocks to G are modeled as time dependent current sources. The first and second moments of the current variables, as well as the correlation between the currents, are known. Each wire is modeled as a resistor. Therefore, the P/G network forms a very large scale linear resistive network. The current and voltage drop in branch k are denoted as i_k and v_k. The nodal voltage at node k is V_k, and the external current source drawn from the block at node k is denoted I_k^e. The width, length, and thickness of the wire branch k are w_k, l_k and t_k, respectively. The conductance of the branch k is $g_k = w_k/\rho_s l_k$, where $\rho_s = \rho_m(T_m)/t_k$ is the sheet resistivity and depends on the wire temperature. The current density in branch k is $j_k = i_k/w_k t_k$.

The network behavior is described by a set of nodal equations $\mathcal{G}_n\mathcal{V}_n = \mathcal{I}_n$, and the branch elements are expressed as $\mathcal{G}_b\mathcal{V}_b = \mathcal{I}_b$, where $\mathcal{V}_b = [v_1, \ldots, v_n]^T$ is the vector of branch voltage, $\mathcal{V}_n = [V_1, \ldots, V_m]^T$ is the vector of nodal voltage, $\mathcal{I}_b = [i_1, \ldots, i_n]^T$ is the vector of branch current, $\mathcal{I}_n = [I_1^e, \ldots, I_m^e]^T$ is the vector of nodal current sources drawn from the block, $\mathcal{G}_b = diag(g_1, \ldots, g_n)$ is the branch conductance matrix, and $\mathcal{G}_n$ represents the nodal conductance matrix. The relation between $\mathcal{G}_n$ and $\mathcal{G}_b$ is

$$\mathcal{G}_n = \mathcal{A}\mathcal{G}_b\mathcal{A}^T = \sum_k^n \frac{w_k}{\rho_s l_k} a_k a_k^T, \tag{10}$$

where $\mathcal{A}$ is the incidence matrix which implies the KCL and KVL, and a_k is the k^{th} column of $\mathcal{A}$. The variables in the design are the widths w_k of the network wires.

For the traditional P/G networks optimization, the objective function is the total wire area $A(w) = \sum_{k=1}^n l_k w_k$ subject to the following constraints.

- **IR-drop constraint**
  IR-drop is the voltage fluctuation due to the resistance of the on-chip P/G networks. Therefore, e.g., the voltage fluctuation from ground pads to the leaf nodes must be restricted

$$V_{k \in N_{leaf}} \leq V_{max}, \tag{11}$$

  where N_{leaf} is a set of leaf nodes and V_{max} is the maximum voltage of ground bounce.

- **Minimum-width constraint**
  According to the semiconductor processing, wire width has a lower bound

$$w_{k \in B} \geq w_{min}, \tag{12}$$

  where w_{min} is the minimum wire width.

- **Electromigration constraint** (Eqn. (4))

An idea based on minimizing a weighted sum of power dissipation in P/G networks and total wire area, $P(w) + \mu A(w)$, was proposed [4], where μ is a positive constant controlling the relative importance of both term. Here we give the physical meaning of μ as *ohmic power density*. The physical meaning of optimizing $P(w) + \mu A(w)$ implies that smaller the wire area, larger the power dissipation in networks. Therefore, the traditional design sacrifices power dissipation to save wire area. However, higher power dissipated in networks causes higher temperature in networks.

4.2. Thermal-Aware P/G Networks Optimization

Our method is based on minimizing the sum of each wire's weighted sum of average power dissipation and wire area, $\sum_{k=1}^{n}[P_k(w)+\mu_k A_k(w)]$, where $P_k(w)$ and $A_k(w)$ are the average power dissipation and wire area of wire k, respectively. Each wire can be assigned a different μ_k depending on the maximum current density restriction. This problem is subject to the following constraints: IR-drop constraint, Eqn. (11); minimum-width constraint, Eqn. (12); and self-consistent constraint, Eqn. (9). Note that the EM constraint is replaced by the self-consistent constraint. Specifically, we consider the following thermal-aware based optimization problem.

Problem TOP *(Thermal-aware Optimization Problem)*

$$\begin{aligned} minimize: \quad & \sum_{k=1}^{n}[P_k(w)+\mu_k A_k(w)] \\ subject\ to: \quad & w_k \geq 0 \end{aligned} \tag{13}$$

Total power consumption in the objective function can be expressed as $\sum_{k=1}^{n} P_k(w) = E[\sum_{k=1}^{n} i_k v_k] = E[\sum_{k=1}^{m} I_k^e V_k] = E[\mathcal{I}_n^T \mathcal{G}_n^{-1}(w)\mathcal{I}_n]$. The expected value of the power consumption is expressed in terms of widths, w, and current source drawn from blocks, $\mathcal{I}_n$, as [4]

$$E[\mathcal{I}_n^T \mathcal{G}_n^{-1}(w)\mathcal{I}_n] = Tr\Gamma\mathcal{G}_n^{-1}(w), \tag{14}$$

where $\Gamma = E[\mathcal{I}_n\mathcal{I}_n^T]$ is the second moment of the $\mathcal{I}_n$, and Tr is the trace of the matrix. Since $Tr\Gamma\mathcal{G}_n^{-1}(w)$ is a differentiable convex function of w, this problem is a convex optimization problem so it can efficiently be solved globally.

The problem TOP has the property that the optimal solution of each wire k is either zero width or has the rms current density $j_{k,rms} = \sqrt{\mu_k/\rho_s t_k^2}$. The necessary and sufficient conditions for the objective function of problem TOP to have optimal solution, subject to the constraint $w_k \geq 0$, are $\frac{\partial}{\partial w_k}(\sum_{k=1}^{n}[P_k(w)+\mu_k A_k(w)]) = 0$ for each wire k with $w_k > 0$, and $\frac{\partial}{\partial w_k}(\sum_{k=1}^{n}[P_k(w)+\mu_k A_k(w)]) \geq 0$ for each wire k with $w_k = 0$. After similar derivation as in [4], we have the following results

$$v_{k,rms} = l_k\sqrt{\rho_s\mu_k} \quad \text{or} \quad j_{k,rms} = \frac{1}{t_k}\sqrt{\frac{\mu_k}{\rho_s}} \tag{15}$$

for each wire with $w_k > 0$.

This property gives a hint for P/G design to satisfy self-consistent constraint. For example, to have current density of wire k with $j_{k,rms}$, satisfying the self-consistent constraint in Eqn. (9), it can be designed by assigning ohmic power density

$$\mu_k = \rho_s t_k^2 j_{k,rms}^2. \tag{16}$$

Then the optimal solution for each non-zero wire has current density $j_{k,rms}$. Note that the current density of each wire need not necessarily to be same, this property is different from that in [4] and more practical. With the design of each wire having the same current density, the resulted wire widths may be too big or too small, and some of the nodes may have big voltage drop.

Next we will use the property of scaling to satisfy the IR-drop and minimum-width constraints. Suppose a set of wire widths, w, solves the problem TOP for a set of ohmic power density, $\mu = \{\mu_1, \ldots, \mu_n\}$. Then λw solves the problem TOP for μ/λ^2, where $\lambda > 0$ is the scaling facor. This scaling method only changes the current density of each wire from j_k to j_k/λ. The current flowing through each wire remains the same, and KCL is still satisfied. With the scaling rule, the violation of minimum-width or IR-drop constraint can be fixed. For example, the maximum allowed voltage drop is V_{max}, Eqn. (11), but the solution has the worst voltage drop $V_{sol} > V_{max}$. Then $\lambda = V_{sol}/V_{max}$ will be used to reduce the voltage of each node. For minimum-width violation, the minimum allowed wire width is w_{min}, but the minimum wire width of the solution is w_{sol}. The scaling factor $\lambda = w_{min}/w_{sol}$ is used to increase the wire widths. Since the current density of each wire decreases after scaling, the self-consistent constraint will still satisfy.

However, the scaling method has potential drawbacks due to the large maximum current density of each wire calculated from self-consistent solution. First, the resulted wire widths can be very small, and the scaling factor of minimum-width violation is very large. Then the wire area after scaling becomes very large, and this is not practical. Second, the wires with large current density have large voltage drop because of $v_k = \rho_m j_k l_k$. This is the source of worst case IR-drop violation. A simple technique can avoid these problems: giving an upper bound of the self-consistent current density, j_{ub}, is helpful.

4.3. Design Algorithm

Algorithm 1: *Thermal-Aware P/G Networks Design*

Given
  Effective duty cycle r_{eff};
  Current sources $\mathcal{I}_n$;
  Initial $w^0 = \mathbf{1}$;
Simulate substrate temperature profile T_{sub}^0;
Iteration $q \leftarrow 1$;
repeat
  Calculate self-consistent current density
    $j_{sc}^q \leftarrow (w^{q-1}, T_{sub}^{q-1})$;
  Solve *Thermal-Aware P/G Networks Optimization*
  {
    Assign current density for each wire
      $j_k^q \leftarrow min(j_{sc,k}^q, j_{ub})$;
    Assign ohmic power density
      $\mu_k \leftarrow \rho_s t_k^2 (j_{k,rms}^q)^2$;
    Optimize problem **TOP**
      $min\ \ E[\mathcal{I}_n^T \mathcal{G}_n^{-1}(w)\mathcal{I}_n] + \sum_{k=1}^{n}\mu_k l_k w_k$
      $s.t.\ \ w_k \geq 0$
  }
  $w^{q-1} \leftarrow w^q$;
  Simulate substrate temperature profile T_{sub}^q;
  $T_{sub}^{q-1} \leftarrow T_{sub}^q$;
  $q \leftarrow q+1$;
until $\| j_k^q - j_k^{q-1} \|_\infty \leq \delta$
Wire widths $w = w^q$;

if (IR-drop constraint is violated)
  $\lambda = V_{sol}/V_{max}$;
  $w \leftarrow \lambda w$;
if (Minimum-width constraint is violated)
  $\lambda = w_{min}/w_{sol}$;
  $w \leftarrow \lambda w$;
Final wire widths $w^* = w$

The thermal-aware P/G networks design algorithm is shown

in Algorithm 1. Suppose that the currents drawn are variables, and the currents causing worst case IR-drop are included. The initial wire widths and effective duty cycle are given. According to these information, the substrate temperature profile can be simulated, e.g., by 3D Thermal-ADI [21]. According to the substrate temperature profile and initial wire widths, maximum current density of each wire that satisfies self-consistent constraints is calculated, j_{sc}. Assigning the current density of each wire according to the calculated self-consistent constraint and the upper bound, j_{ub}. After deciding the current density and ohmic power density of each wire, solve problem TOP and obtain a set of wire solutions. The new substrate temperature profile is simulated again according to the new set of wire widths. Since the assigned current density of each wire is according to the initial wire widths and substrate temperature profile, this procedure is repeated until the difference of current density of two consecutive iteration reaches an error tolerance δ. After finding the solution, check the violations of IR-drop and minimum-width.

4.4. Computational Issues

Problem TOP can be translated into an unconstrained problem and can be solved with the barrier method. We use the logarithmic barrier method by augmenting the objective function with a logarithmic barrier function.

Problem TOP-LBM *(TOP with Logarithmic Barrier Method)*

$$\phi(w) = Tr\Gamma\mathcal{G}_n^{-1}(w) + \sum_{k=1}^{n} \mu_k l_k w_k - \beta \sum_{k=1}^{n} \log w_k \qquad (17)$$

Here $\beta > 0$ is referred to as barrier parameter [12]. This function is defined for each wire $w_k > 0$, and it is smooth and convex on its domain. The minimizer of problem TOP-LBM is suboptimal of problem TOP with an accuracy of at least $n\beta$. Therefore, to solve problem TOP to an accuracy of ϵ, we need to have $\beta = \epsilon/n$. Instead of solving problem TOP, we solve problem TOP-LBM repeatedly for a sequence of decreasing β until the accuracy is achieved.

For the unconstrained minimization problem TOP-LBM, a limited-memory aquasi-Newton lgorithm L-BFGS-B is used for solving this problem [23]. It is useful for solving a large problem in which the Hessian is dense like our problem. L-BFGS-B only needs the gradient, and for (17), it is

$$\frac{\partial\phi(w)}{\partial w_k} = -\frac{1}{\rho_s l_k} a_k^T \mathcal{G}^{-1}(w)\Gamma\mathcal{G}^{-1}(w)a_k + \mu_k l_k - \frac{\beta}{w_k}. \qquad (18)$$

The numerical issues are in the following discussion. For the first iteration in Algorithm 1, we scale the wire widths to make power term $Tr\Gamma\mathcal{G}^{-1}$ and weighted area term $\sum_{k=1}^{n} \mu_k l_k w_k$ equal. After optimization, the scaled wire widths are translated back to non-scaled wire widths. The initial value of β is set as $0.05(Tr\Gamma\mathcal{G}^{-1})/n$ suggested by [4]. The β value is decreased by a factor of 10 until $n\beta$ is smaller than the required accuracy.

5. Experimental Results

The proposed algorithm is implemented with C++ language and executed on a computer with 1.6 GHz Intel Pentium 4 CPU and 640 MB memory. The proposed optimization algorithm is applied on the P/G networks design of an industrial test chip with size $11.3mm \times 14.4mm$ and power consumption $48W$. The P/G grid is 58×72 with wire length $200\mu m$. The large scale network problem can be extended by the multigrid technique, e.g. [20]. The ground nodes are placed at every 4 nodes in the x and y directions. The minimum wire width is limited by $0.6\mu m$. The power supply is $1.5V$ and the maximum IR-drop is $V_{max} = 150mV$. The sheet resistance is $0.027\ \Omega$ at temperature 120^oC and the temperature coefficient is $6.8 \times 10^{-3}K^{-1}$. We assume that wires have 10 years of operation lifetime under the current density $10^6 A/cm^2$ at $105\ ^oC$. The effective duty cycle is 0.1. The initial guess of each wire is $1\mu m$. We test the thermal-aware P/G networks optimization method and compare that with traditional methods without thermal integrity. The simulation results are shown in Table 1.

Thermal-Aware	Yes	No
Grid Size	58×72	58×72
Total wires	8482	8482
Removed wires	2157	2255
Wire width (μm)	$0.1 \sim 33.5$	$0.1 \sim 16.1$
Minimum Lifetime (years)	9.5	2
Lifetime Violations ($> 5\%$)	0	372
Target j_{rms} ($mA/\mu m^2$)	$3.6 \sim 7.5$	7.5
Final j_{rms} ($mA/\mu m^2$)	$3.6 \sim 7.5$	$6.6 \sim 8.9$
Worst v_{rms} mV	97.0	113.1
Area (μm^2)	3.141×10^6	2.980×10^6
Runtime (sec)	1208	943

Table 1. Comparison of the P/G networks design with and without thermal integrity.

First, the chip reliability by way of the histogram of wire lifetime is shown in Figure 3. Without thermal handling, there are 372 wires violating the 10 years lifetime and the minimum lifetime is only 2 years. In general, the violating wires are in hot spots. With thermal integrity design, the lifetime of all wires is within 5% of expected 10 years, and the minimum lifetime is 9.5 years. This 5% error comes from the experimental setting of the stopping criteria δ in Algorithm 1. In this case, δ is set to be $0.1mA/\mu m^2$. The minimum lifetime can be improved by decreasing δ, but the number of iterations will get increased.

If we allow each wire to reach maximum current density, the lifetime of each wire should be 10 years. Due to the practical concerns as discussed, an upper bound of current density $j_{ub} = 7.5mA/\mu m^2$ is given to avoid high current density wires. This can be observed in Table 1 for the case with thermal integrity that wire rms current density is in the region $3.6 \sim 7.5mA/\mu m^2$. For the case without thermal integrity, the EM constraint gives the target rms current density $7.5mA/\mu m^2$. However, the SH effect makes the final current density in the region $6.6 \sim 8.9mA/\mu m^2$ which is much higher than the case with thermal integrity.

The worst value of rms voltage is $97\ mV$ for the case with thermal integrity, and $113.1\ mV$ for the case without thermal integrity. These higher voltage nodes for the case without thermal integrity are distributed in the hot spots. The average voltage is $55.64mV$ for the thermal integrity design, and $58.07mV$ for the case without thermal integrity. From the discussion of wire reliability, current density, and voltage distribution, the power delivery quality is better with the thermal-aware design. The area of

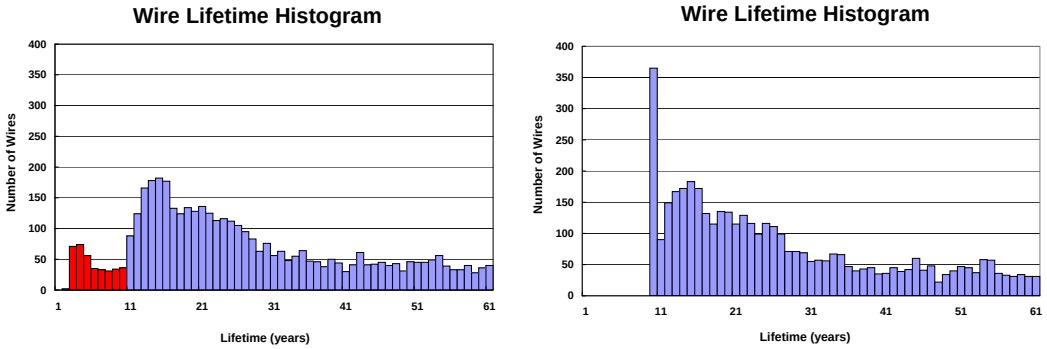

Figure 3. The wire lifetime histogram for the case without (left) and with (right) thermal integrity.

thermal integrity design is only 5.4% larger.

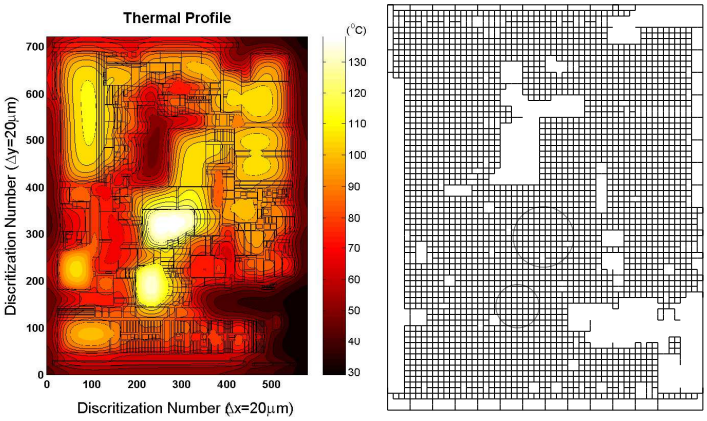

Figure 4. Optimal solution of ground network (Right) and temperature profile (Left).

According to the optimal solution, the simulated substrate temperature profile is shown on the left of Figure 4. The temperature distribution varies from $30\ ^oC$ to $135\ ^oC$. Therefore, it is important to consider thermal issue on P/G design to ensure the reliability. The optimal ground network for the thermal integrity P/G network design is shown on the right of Figure 4. It can be observed that some wires in the circles are dark due to their wide widths. Comparing these two figures, it is obvious that these wires are in the hot spots. The high temperature causes wires to have low current density because of the self-consistent constraint.

6. Conclusion

We presented a thermal-aware P/G networks optimization algorithm which considers thermal and power-delivery integrity issues. This design method improves power delivery quality and thermal reliability. The major advantage of the proposed algorithm is its ability to handle the thermal issues. The proposed self-consistent constraint gives a maximum current density which simultaneously accounts for both EM and SH effects. In addition, the power-area tradeoff approach can improve the thermal reliability through wire current density controlling. The experimental results show that the thermal-aware design improves thermal reliability and power delivery quality from the comparison of wire lifetime, voltage drop distribution, and wire rms current density.

References

[1] A. H. Ajami, K. Banerjee, A. Mehrotra, and M. Pedram. Analysis of IR-Drop Scaling with Implications for Deep Submicron P/G Network Designs. *ISQED 2003*, pages 35–40.

[2] K. Banerjee and A. Mehrotra. Global (interconnect) Warming. *IEEE Circuits and Devices Magazine*, 17(5):16–32, Sep. 2001.

[3] J. R. Black. Electromigration–a Breief Survey and Some Recent Results. *IEEE Trans. on Electron Devices*, ED-16(4):338–347, Apr. 1969.

[4] S. Boyd, L. Vandenberghe, A. Gamal, and S. Yun. Design of Robust Global Power and Ground Networks. *ISPD 2001*, pages 60–64.

[5] H. Cai. Multi-Pads Single Layer Power Net Routing in VLSI Circuit. *DAC 1988*, pages 183–188.

[6] S. Chowdhury. Optimum Design of Reliable IC Power Networks Having General Graph Topologies. *DAC 1989*, pages 787–790.

[7] S. U. Chowdhury and M. A. Breuer. Minimal Area Design of Power/Ground Nets Having Graph Topologies. *IEEE Trans. on CAS*, CAS-34(12):1441–1450, December 1987.

[8] R. Dutta and M. Marek-Sadowska. Automatic Sizing of Power/Ground (P/G) Networks in VLSI. *DAC 1989*, pages 783–786.

[9] W. R. Hunter. Self-Consistent Solutions for Allowed Interconnect Current Density – Part I: Implications for technology evolution. *IEEE Trans. on Electron Devices*, 44(2):304–309, Feburary 1997.

[10] W. R. Hunter. Self-Consistent Solutions for Allowed Interconnect Current Density – Part II: Application to Design Guidelines. *IEEE Trans. on Electron Devices*, 44(2):310–316, February 1997.

[11] A. S.-V. Kaustav Banerjee, Amit Mehrotra and C. Hu. On Thermal Effects in Deep Sub-Micron VLSI Interconnects. *DAC 1999*, pages 885–891.

[12] J. Nocedal and S. J. Wright. *Numerical Optimization*, chapter 17. Springer, 1999.

[13] J. Oh and M. Pedram. Multi-Pad Power/Ground Network Design for Uniform Distribution of Ground Bounce. *DAC 1998*, pages 287–290.

[14] R. Saleh, S. Z. Hussain, S. Rochel, and D. Overhauser. Clock Skew Verifi cation in the Presence of IR-Drop in the Power Distribution Network. *IEEE Trans. on CAD*, 19(6):635–644, Jun. 2000.

[15] H. A. Schafft. Thermal Analysis of Electromigration Test Structures. ED-34(3):664–672, Mar. 1987.

[16] H. Su, K. Gala, and S. S. Sapatnekar. Fast Analysis and Optimization of Power/Ground Networks. *DAC 2000*, pages 477–482.

[17] Z. A. Syed and A. E. Gamal. Single Layer Routing of Power and Ground Networks in Integrated Circuits. *Journal of Digital Systems*, 6(1):1441–1450, 1982.

[18] X. Tan, C. J. R. Shi, D. Lungeanu, and L. Y. J. Lee. Reliability-Constrained Area Optimization of VLSI Power/Ground Networks via Sequence of Linear Programmings. *DAC 1999*, pages 78–83.

[19] L. M. Ting, J. S. May, W. R. Hunter, and J. W. McPherson. AC Electromigration Characterization and Modeling of Multilayered Interconnects. *Proc. Int. Reliability Physics Symposium*, pages 311–316, 1993.

[20] K. Wang and M. Marek-Sadowska. On-Chip Power Supply Network Optimization Using Multigrid-based Technique. *DAC 2003*, pages 113–118.

[21] T.-Y. Wang and C. C.-P. Chen. 3D Thermal-ADI: A Linear-Time Chip Level Transient Thermal Simulator. *IEEE Trans. on CAD*, 21(12), Dec. 2002.

[22] T.-Y. Wang and C. C.-P. Chen. Optimization of the Power/Ground Network Wire-Sizing and Spacing based on Sequential Network Simplex Algorithm. *ISQED 2002*, pages 157–162.

[23] C. Zhu, R. H. Byrd, and J. Nocedal. L-bfgs-b: Algorithm 778: L-bfgs-b, Fortran Routines for Large Scale Bound Constrained Optimization. *ACM Trans. on Mathematical Software*, 23(4):550–560, 1997.