Analytical Thermal Placement for VLSI Lifetime Improvement and Minimum Performance Variation

Andrew B. Kahng[†]   Sung-Mo Kang[‡]   Wei Li[‡]   Bao Liu[†]

[†]CSE Dept., UC San Diego, CA 92093, Email:{bliu,abk}@cs.ucsd.edu
[‡]CSE Dept., UC Santa Cruz, CA 95064, Email:{kang,ldragon}@soe.ucsc.edu

Abstract

DSM and nanometer VLSI designs are subject to an increasingly significant thermal effect on VLSI circuit lifetime and performance variation, which can be effectively subdued by VLSI placement. We propose analytical placement for accurate and efficient VLSI thermal optimization, and propose minimized maximum on-chip temperature as the thermal optimization objective for improved VLSI lifetime and minimized performance variation. We develop an effective analytical thermal placement technique, as well as an improved analytical placement technique with a new cell spreading function. Our experimental results show that our proposed analytical thermal placement achieves 17.85% and 30.77% maximum on-chip temperature variation reduction as well as 4.61% and 0.45% wirelength reduction respectively for the two industry design test cases compared with thermal-oblivious analytical placement, e.g., APlace.

1 Introduction

Technology scaling has introduced continuously increased VLSI on-chip temperature. Higher component integration induced current density increase, and shrinking interconnect dimensions induced interconnect resistance increased lead to increased VLSI power consumption and heat dissipation. On the other hand, VLSI cooling techniques have largely remained the same in the past decades, the newly introduced low-k material as intra- and inter-layer insulator reduces coupling capacitance but also thermal conductivity, new packaging techniques, e.g., flip chips, MCM, and 3-D chips, all pose new challenges to chip heat dissipation [2].

On-chip temperature increase has significant effects on VLSI performance and reliability. Temperature variation on a chip contributes to performance variation, and must be taken into account by VLSI performance verification. In worst cases, the positive feedback loop between transistor on-resistance, temperature, and power could lead to thermal runaway problem [19]. Furthermore, interconnect life-

time is affected by electromigration effect, i.e., mass shifting along an interconnect over time, which has an exponential dependence on the inverse metal temperature [1, 14]. Consequently, VLSI performance verification and reliability evaluation need to include thermal analysis.

VLSI thermal analysis is based on solving a partial differential heat diffusion equation. Techniques include the finite difference method (FDM), the finite element method (FEM), and the boundary element method (BEM). Zhan and Sapatnekar applied discrete cosine transform (DCT) to achieve orders of magnitude of speed improvement in full-chip thermal analysis [23, 24]. Duality of thermal and electrical simulation has long been observed. Tsai and Kang compute thermal resistance matrices by construction or model order reduction for efficient incremental thermal analysis [20].

Placement is the earliest hence the most effective physical design stage for thermal optimization. Previous thermal placement techniques compute thermal resistance matrices for thermal analysis for each tentative placement solution [3, 7, 20], and apply simulated annealing [7, 20], or min-cut partition [3] for total on-chip temperature [3] or maximum on-chip temperature reduction [5, 7]. Simulated annealing or min-cut partition based thermal placement achieves only limited efficiency and solution quality by targeting a NP-hard problem.

In recent years, analytical placement has been developed as an effective technique, which achieves quality VLSI placement for a wide range of objectives. Analytical placement approximates the original NP-hard placement problem as a continuous optimization problem by applying smooth approximation for the placement objective and the constraints. The approximated problem can then be solved by applying nonlinear optimization techniques. For example, it evenly spreads cell instances in the layout plane by minimizing the deviation of a layout density function which is approximated by a smooth function. Other placement objectives, e.g., wirelength, routing congestion, timing, power, supply voltage drop, etc. are also included in smooth function approximations of cell instance locations, and opti-

mized via a nonlinear optimization solver [12].

In this paper, we present an analytical thermal placement method. We believe that accurate thermal placement must be analytical, while previous thermal placement methods did not target the exact object. We propose minimized maximum on-chip temperature as the thermal placement objective, and integrate it in an analytical placement framework based on a smooth approximation function. For efficient incremental thermal analysis in placement, we separate thermal resistance map and heat sources, and compute only an initial thermal resistance map which does not need to update during thermal placement. We also propose a new cell spreading function for improved analytical placement. We validate our proposed method for two industry designs with respect to voltage degradation, placed wirelength and runtime criteria. Our experimental results show that our proposed analytical thermal placement achieves 17.85% and 30.77% reduction of maximum on-chip temperature variation as well as 4.61% and 0.45% wirelength reduction respectively for the two test cases compared with thermal-oblivious analytical placement, e.g., APlace.

The rest of this paper is organized as follows. We introduce thermal analysis, thermal effects on performance and lifetime of a VLSI design, and analytical placement in Section 2. We present our problem formulation, thermal model, complexity analysis, and our proposed analytical thermal placement in Section 3. Our experimental results are presented in Section 4, before we conclude in Section 5.

2 Background

2.1 Thermal Analysis

We consider steady state thermal analysis, which is adequate in most cases because the thermal RC time constant (e.g., the product of thermal resistance and thermal capacitance) is in milliseconds and much large than a VLSI clock cycle time [14]. The steady state temperature distribution on a chip is governed by the Poisson's equation

$$\nabla^2 T(\vec{r}) = -\frac{g(\vec{r})}{k(\vec{r})} \qquad (1)$$

where $\vec{r} = (x, y, z)$, $T(\vec{r})$ is the temperature distribution in the chip, $g(\vec{r})$ is the volume power density (W/m^3), and $k(\vec{r})$ is the thermal conductivity ($W/(m{\cdot}K)$) of the layer where point $\vec{r}$ is located. Assuming adiabatic top and side surfaces, and a convective bottom surface of the chip with an effective heat transfer coefficient h ($W/(m^2{\cdot}K)$), the boundary conditions are given by (Fig. 1)

$$\frac{\partial T(\vec{r})}{\partial x}\Big|_{x=0,a} = 0$$
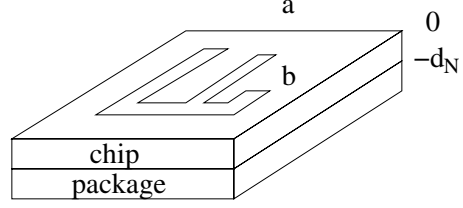$$\frac{\partial T(\vec{r})}{\partial y}\Big|_{y=0,b} = 0$$

Figure 1. A simplified thermal model for a chip of dimensions a, b, **and** d_N **on a package board.**

$$\frac{\partial T(\vec{r})}{\partial z}\Big|_{z=0} = 0$$
$$k_N \frac{\partial T(\vec{r})}{\partial z}\Big|_{z=-d_N} = h(T(\vec{r})|_{z=-d_N} - T_r) \qquad (2)$$

where T_r is the ambient temperature, k_N is the thermal conductivity of the bottom layer of the chip [23, 24].

2.2 Thermal Effect on Performance

Chip temperature increase significantly affects two parameters: carrier mobility and transistor threshold voltage. Higher temperature leads to superlinearly decreased carrier mobility, and almost linearly decreased transistor threshold voltage, e.g., as follows.

$$\mu(T) = \mu(T_r)(\frac{T}{T_r})^{-k_s}$$
$$V_T(T) = V_T(V_r) - k_4(T - T_r) \qquad (3)$$

where T is temperature in Kelvins, T_r is the ambient temperature, k_s is a constant between 1.0 and 2.0, k_4 is between $0.5mV/K$ and $4mV/K$ [21].

Transistor output current I_{DS} could increase with temperature increase due to decreased threshold voltage, or decrease with temperature increase due to decreased mobility [21]. For latest technologies, e.g., $0.35\mu m$, saturated transistor output current I_{DSAT} decreases with temperature increase, leading to performance degradation [13]. Increased interconnect resistance due to decreased carrier mobility also contributes to VLSI performance degradation at high temperature.

2.3 Thermal Effect on VLSI Lifetime

Chip temperature increase also significantly degrades VLSI reliability. For example, interconnect lifetime due to electromigration (or similarly other failure mechanisms such as gate oxide/dielectric breakdown) is modeled as a function of temperature as follows [15].

$$\int_0^{T_f} j(t)(\frac{exp(\frac{-Q}{kT(t)})}{kT(t)})dt = D \qquad (4)$$

where T_f is the interconnect mean time to failure (MTF), $j(t)$ and $T(t)$ are time-dependent current density and temperature, respectively, Q is the activation energy (e.g., $1.0eV$ for copper interconnect), kT is the thermal energy, and D is a constant determined by the structure of the interconnect.

If we regard VLSI lifetime as a resource that is consumed by the system over time, then higher temperature implies a large consumption rate for the VLSI lifetime, and this consumption rate increases superlinearly with increased on-chip temperature [15]. This can be seen by rewriting (4) in terms of average current density J and steady state temperature T as follows.

$$T_f^{-1} = \frac{Je^{-\frac{Q}{kT}}}{DkT} \qquad (5)$$

2.4 Analytical Placement

Recent development in analytical placement has lead to a highly effective placement technique of high solution quality and strong extensibility [16, 11, 22]. Analytical placement approximates placement objective and constraints in smooth and differentiable functions, thus transforms the NP-hard combinational problem into a continuous global optimization problem. It applies local optimization techniques (e.g. conjugate gradient) for local optimum, and approximates global optimum by recursive relaxation and near-convexity of smooth approximation functions. For example, in APlace [11], a component c of rectangle (x_1, y_1, x_2, y_2) has contribution to layout density at location (x, y) of

$$d(c, x, y) = \begin{cases} 1 & (x_1 < x < x_2, y_1 < y < y_2) \\ 0 & otherwise \end{cases} \qquad (6)$$

and is approximated by

$$d(c, x, y) = b(x - \frac{x_2 - x_1}{2}, x_2 - x_1) \cdot b(y - \frac{y_2 - y_1}{2}, y_2 - y_1) \qquad (7)$$

where $b(l, r)$ is a bell-shaped function (Fig. 2)

$$b(l, r) = \begin{cases} 1 - 2l^2/r^2 & (0 < l < r/2) \\ 2(l - r)^2/r^2 & (r/2 < l < r) \end{cases} \qquad (8)$$

The half-perimeter wirelength of a net n with k terminals $(x_1, y_1), ...(x_k, y_k)$ in a Manhattan plane is given by

$$l(n) = Max_i(x_i) - Min_i(x_i) + Max_i(y_i) - Max_i(y_i) \qquad (9)$$

and is approximated by

$$l(n) = \alpha(\log(\sum_i e^{x_i/\alpha}) + \log(\sum_i e^{-x_i/\alpha}) + \\ \log(\sum_i e^{y_i/\alpha}) + \log(\sum_i e^{-y_i/\alpha})) \qquad (10)$$
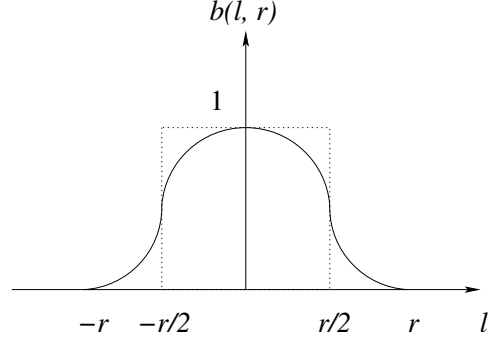

Figure 2. A bell-shaped function $b(l, r)$ **(in solid curves) which approximates the pulse function (in dotted lines).**

With these smooth and differentiable approximation functions for the placement objectives, analytical placement applies a nonlinear optimization solver, e.g., based on conjugate gradient computation in APlace [11] for the following objective

$$minimize \qquad \sum_{n \in N} l(n) + \frac{1}{\beta} \sum_{(x,y) \in G} (D(x, y) - \bar{D})^2 \qquad (11)$$

where

$$\begin{aligned} D(x, y) &= \sum_{c \in C} d(c, x, y) \\ \bar{D} &= \sum_{(x,y) \in G} D(x, y)/(m \cdot n) \qquad (12) \end{aligned}$$

G is a $m \times n$ array of grid crosspoints (e.g., of global cells in placement). β is a weighting factor for cell spreading or layout density deviation minimization, which continuously decreases during placement, such that the placement starts with minimum wirelength and congested components to solutions with more evenly distributed components.

3 Analytical Thermal Placement

3.1 Problem Formulation

Existing thermal placement methods pursue several different objectives, including (1) minimized sum of on-chip component temperatures [3], (2) minimized maximum on-chip temperature [5, 7, 20], or (3) minimized on-chip temperature variation [20].

We propose minimized maximum on-chip temperature as the thermal placement objective. Minimized maximum on-chip temperature implies maximized VLSI lifetime (with current density J and structure constant D weighting factors based on (5)), and minimized temperature-induced performance variation in a VLSI design.

Our thermal placement optimization objective is the static, i.e., average on-chip temperature over the lifetime of a VLSI system, which requires the *average* power consumption for each component on a chip. A more sophisticated performance optimization thermal placement method would take *worst case* power consumptions for each component for traditional min/max timing requirements, or *statistical* power consumption for each component for timing yield requirements in a statistical timing analyzer.

Our thermal placement problem formulation is as follows.

Problem 1 (Analytical Thermal Placement) *Given*

1. *chip dimensions $0 < x < a$, $0 < y < b$, $0 < z < d$,*

2. *thermal parameters:*

   (a) *thermal conductivity k on the top of the chip,*
   (b) *thermal conductivity k_N on the bottom of the chip,*
   (c) *effective heat transfer coefficient h on the bottom of the chip, and*
   (d) *ambient temperature T_r,*

3. *components C,*

4. *power consumption $p(c)$ for each component $c \in C$,*

5. *netlist N,*

find a placement which minimizes

$$\sum_{n \in N} l(n) + Max_{c \in C}T(c) \tag{13}$$

where $l(n)$ is the wirelength of net n, $T(c)$ is the temperature of component c.

3.2 Thermal Model

A heat transfer system is analogous to an electrical circuit, i.e., the heat flow (W) passing through a thermal resistor (K/W) is equivalent to the electrical current (Ampere) through an electrical resistance (Ohm), and the temperature difference (K) corresponds to voltage difference (Volt), as follows (Fig. 3).

$$T = T_r + R^t P \tag{14}$$

where T is the temperatures of n heat sources, T_r is the ambient temperature, R^t is the $n \times n$ thermal resistance matrix, and P is the power consumptions of the n heat sources [18]. Comparing with (1), P and R^t are the volume integral of $g(\vec{r})$ and $k(\vec{r})^{-1}$, respectively.

Several approaches are available to construct such a thermal resistance network. We can compute volume integral of the inverse of thermal conductivity $k(\vec{r})^{-1}$ at each location on the chip, and reduce the size of the matrix by model order reduction techniques. Using a commercial field solver
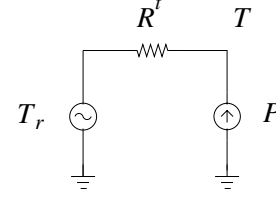

Figure 3. An equivalent DC circuit for a steady state heat transfer system. The node voltage (temperature T) is given by the DC source voltage (ambient temperature T_r), the resistance (thermal resistance R^t), and the source current (power dissipation P of the heat source).

for thermal analysis, we can apply a unit power source at the location of one heat source and find temperature increases at the locations of all heat sources, which gives a column of the thermal resistance matrix R^t. Or, we can reduce thermal resistance matrix given by a field solver by applying model order reduction techniques [20].

Given a thermal resistance matrix, temperature variation due to placement is simply computed by applying (14) for different heat source P permutations.

3.3 Complexity

We simplify the thermal placement problem as to assign the components to an array of pre-defined slots, for a given thermal resistance map R^t, and given power consumptions P for the components. Among the existing thermal placement objectives, minimized sum of on-chip component temperature is achievable by a greedy algorithm [3]. Achieving minimized maximum on-chip temperature is NP-hard.

Theorem 1 *To find a placement which minimizes the maximum on-chip temperature is NP-hard.*

Proof. We prove that to find a permutation of P in (14) which minimizes $Max_i T_i$ is NP-hard, which can be reduced to the NP-hard two-way balanced partition problem. Consider a thermal resistance matrix

$$R^t = \begin{bmatrix} . & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & . \\ \vdots & & & \vdots & \vdots & & & \vdots \\ 1 & \cdots & \cdots & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & \cdots & \cdots & 1 \\ \vdots & & & \vdots & \vdots & & & \vdots \\ . & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & . \end{bmatrix} \tag{15}$$

Cell Density Variation
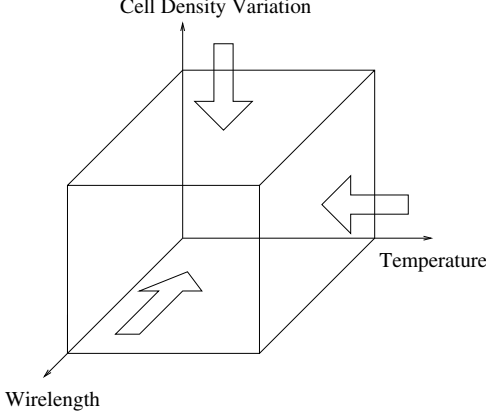
Temperature

Wirelength

Figure 4. Thermal placement solution space of cell density variation, wirelength, and maximum on-chip temperature variation.

where the non-zero entries are only on the first $n/2$ columns of the p-th row, and the last $n/2$ columns of the q-th row. For this R^t, to minimize the maximum temperature T_p and T_q for components p and q is to find a balanced partition of P. $\square$

3.4 Approximation

Analytical placement applies continuous optimization techniques based on smooth approximation of the layout density and the wirelengths on a chip. Similarly, we approximate the max function for on-chip temperature in a smooth function, e.g., a log-sum-exp function as follows.

$$\text{Max}_{c \in C} T(c) = \alpha \cdot \log(\sum_{c \in C} e^{T(c)/\alpha}) \qquad (16)$$

The log-sum-exp function is continuously differentiable and converges to the maximum on-chip temperature degradation as α converges to 0. Compare with the average on-chip temperature variation objective, the smoothing parameter α in (16) is also a weighting factor in minimizing on-chip temperature variation at different locations. I.e., the effect of cell movement on the largest on-chip temperature variations are emphasized.

3.5 Analytical Thermal Placement

We integrate thermal optimization in an analytical placement framework by including (16) as an extra thermal optimization term in the analytical placement objective (Fig. 4).

$$minimize \qquad \sum_{n \in N} l(n) + \gamma \cdot \alpha \cdot \log(\sum_{c \in C} e^{T(c)/\alpha})$$
$$+ \frac{1}{\beta} \sum_{(x,y) \in G} (D(x,y) - \bar{D})^2 \qquad (17)$$

where γ is the weight for the thermal objective, and is given by

$$\gamma = \delta \frac{\sum_{(x,y) \in G} \frac{\partial \sum_{n \in N} l(n)}{\partial x} + \frac{\partial \sum_{n \in N} l(n)}{\partial y}}{\sum_{(x,y) \in G} \frac{\partial \text{Max}_{c \in C} T(c)}{\partial x} + \frac{\partial \text{Max}_{c \in C} T(c)}{\partial y}} \qquad (18)$$

such that the cell density, the wirelength, and the temperature gradients are comparable. δ is a user-adjustable constant which differs for different test cases and tradeoffs between wirelength minimization and thermal optimization. δ is also updated during analytical placement, e.g., it decreases in each placement iteration, such that the initial placements are temperature optimized, while the final placement optimization stages focus on cell spreading and wirelength minimization.

We apply bi-linear interpolation to translate a discrete thermal resistance matrix to a continuous thermal resistance map, and apply two levels of interpolation for a thermal resistance between two locations, one for each location respectively.

3.6 Congestion Penalty Improvement

We also improve the performance of the existing analytical placement technique by a new cell spreading function. We observe that the existing analytical placement spreads cells evenly across the layout plane, by minimizing the cell density deviation. However, it is unnecessary to evenly spread the cells across the whole layout plane. In a sparse design, allowing cell instances to locate closer to each other gives shorter wirelength, as long as the cell instances do not overlap with each other, and the nets are routable (without large detours).

A layout congestion penalty function provides more accurate placement control, e.g., to guarantee the legality of the placement result, and to take into account the congestion induced wirelength increase.

Recall that formula (17) resembles a penalty function method which translate a constrained optimization problem into an unconstrained optimization problem. In this principle, we define a new penalty function, by replacing even cell spreading by upper bounding cell density for each layout area. We optimize the following objective in congestion driven analytical placement.

$$minimize \qquad \sum_{n} l(n) + \gamma \cdot \alpha \cdot \log(\sum_{c \in C} e^{T(c)/\alpha})$$
$$+ \frac{1}{\beta} \sum_{g} P(g) \qquad (19)$$

where γ is given the same as in (18), P is the congestion penalty function as follows.

$$P(g) = \begin{cases} (D(g) - U)^\epsilon & D(g) > U \\ 0 & D(g) < U \end{cases} \qquad (20)$$

where g is one of the $m \times n$ placement global cells, U is the density upper bound for each global cell to avoid congestion, e.g., $0 \le U \le \bar{D}$, and ϵ is a constant, e.g., $\epsilon = 4$.

We have two parameters U and ϵ in this new congestion penalty function. Layout density upper bound U balances global placement wirelength and routability for minimum routing wirelength. For sparse designs, a loose layout density bound enables the cell instances to locate closer to each other for shorter wirelength. For dense designs, a tight layout density bound improves routability, and reduces the wirelength increase (detours) during legalization.

Exponent ϵ gives the sharpness of congestion penalty increase (such a penalty function is called *sharp penalty function* in constrained non-convex optimization). $(D(g) - U)^{\epsilon-1}$ can be seen as a weighting factor for the congestion $(D(g) - U)$. For $\epsilon = 2$, the congestion itself is a weighting factor. An increased exponent ϵ provides increased efficiency for global placement than the existing analytical placement implementation, e.g., APlace, as we see in the experiments.

4 Experiments

In the following experiments, we apply analytical thermal placement to two industry design test cases, and compare with thermal-oblivious analytical placement in terms of maximum on-chip temperature variation, wirelength, and CPU runtime.

Our proposed analytical thermal placement takes the following inputs: (1) netlist and cell libraries in LEF/DEF files, (2) a thermal resistance matrix for the chip, and (3) power consumption for the cell instances. We achieve optimized maximum on-chip temperature variation with negligible increase of total wirelength and cell density variation.

We improve thermal placement efficiency by separating thermal resistance map and heat sources, such that the thermal resistance map needs to be computed once initially and does not need to be updated during thermal placement. We compute the thermal resistance map via network reduction [20]. We first construct a network of thermal resistors based on the thermal conductivity of the bulk silicon and the thermal coefficients for the top, the bottom, and the side boundaries, then apply the network order model reduction techniques [17]. Table 1 gives the inputs of the thermal analysis which are chip dimensions, grid dimensions, thermal conductivities for the bulk silicon, for the top, the bottom, and the side boundaries.

We apply analytical thermal placement to two industry design test cases. The first test case is a $10W$ $130nm$ design with $13,397$ cell instances in 129 rows of gate arrays with 60% utilization, the second test case is a $10W$ $180nm$ design with 7128 cell instances in 251 rows of gate arrays with 43% cell utilization in the presence of 5 macro blocks

Table 1. Thermal analysis inputs: bulk silicon thermal conductivity k ($W/(m \cdot K)$), boundary thermal conductivity k_N ($W/(m \cdot K)$), the bottom, the top, and the side boundary heat transfer coefficients h_{bottom}, h_{top}, h_{side} ($W/(m^2 \cdot K)$), chip dimensions a, b, and d_N (μm) and the ambient temperature T_r (K).

	k	k_N	h_{bottom}	h_{top}	h_{side}	a	b	d_N	T_r
I	148	148	50	1000	500	947	946	100	300
II	148	148	50	1000	500	2760	2760	100	300

Table 2. Test case characteristics: the number of cells, blocks, rows of the cell instances, technology nodes, layout area utilization, and total power consumption of the two test case designs.

Design	# Cells	# Blocks	# Rows	Tech (nm)	Utilization	Total Power (W)
I	13397	0	129	130	0.60	10.0
II	7128	5	251	180	0.43	10.0

(Table 2). The thermal objective weight γ is initially assigned based on the δ parameter as is shown in Table 2, and updated during placement optimization by a ratio of 8 for the two test cases. The layout density bound U is $0.5\bar{D}$ half of the average layout density for the test case I and 0 for the test case II, while the layout density penalty exponent $\epsilon = 4$ for both test cases. For each layout density objective weight β and thermal objective weight γ in (17), gradient-based nonlinear optimization is applied until no improvement is available. A new iteration of optimization is initiated by updating the β and γ weighting factors, e.g., decreasing β and γ, such that the placement starts with high congestion, minimum wirelength, and minimum temperature variation, and proceeds to have more degrees of freedom in spreading the cell instances. The resultant maximum on-chip temperature variations, total wirelengths given by Cadence TrialRoute, and CPU runtime in a $i686$ Linux system with a $2.8GHz$ processor and $512MB$ memory are given for the thermal-aware and thermal-oblivious analytical placers in Table 3.

We observe that our analytical thermal placement achieves 17.85% maximum on-chip temperature variation reduction and 4.61% wirelength reduction for the test case I, and 30.77% maximum on-chip temperature variation reduction and 0.45% wirelength increase for the test case II. These significant reductions of maximum on-chip temperature variation contribute to lifetime reliability improvement and performance variation reduction at the cost of minimal wirelength increase.

Table 3. The maximum on-chip temperature variations (K), the total half-perimeter wire-lengths (mm) and the CPU runtime (s) of a thermal-oblivious analytical placer, and an analytical placer based on a maximum on-chip temperature variation gradient.

Test Case	Placer	δ	Max Temp (K)	Max Temp (%)	HPWL (mm)	HPWL (%)	CPU (s)
I	APlace	0.00	12.10	100.00	504.80	100.00	250.15
	ATP	0.00	11.33	97.47	465.35	92.19	718.53
		0.02	11.16	92.23	466.78	92.47	636.13
		0.05	11.01	91.74	467.24	95.56	968.26
		0.10	9.94	82.15	481.54	95.39	800.52
II	APlace	0.00	2.73	100.00	923.55	100.00	211.37
	ATP	0.00	2.99	109.52	905.62	98.06	204.94
		0.02	2.72	99.63	905.34	98.03	496.91
		0.05	2.25	82.72	909.64	98.49	548.67
		0.10	1.89	69.23	919.42	99.55	507.94

5 Conclusion

We propose analytical placement for VLSI thermal optimization in this paper. We minimize maximum on-chip temperature for VLSI lifetime reliability improvement and performance variation reduction. We show that the problem is NP-hard. We compute a thermal resistance map, approximate on-chip temperature in a smooth function, and apply analytical placement for effective thermal optimization. By separating thermal resistance map and heat sources, we only need to compute an initial thermal resistance map which does not need to update. We develop an effective analytical thermal placement technique, with a new cell spreading function. Our experimental results on two industry design test cases show that our proposed analytical thermal placement achieves 17.85% and 30.77% reduction of maximum on-chip temperature variation as well as 4.61% and 0.45% wirelength reduction respectively for the two test cases compared with thermal-oblivious analytical placement, e.g., APlace.

Our ongoing research efforts address thermal effect on VLSI performance variation based on an analytical placement and a thermal-effect-aware timing analyzer, and thermal placement in 3-D structure systems.

References

[1] J. R. Black, "Electromigration - A Brief Survey and Some Recent Results", *IEEE Trans. Electron Devices*, vol. ED-16, pp. 338-347, 1969.

[2] K. Banerjee, M. Pedram and A. H. Ajami, "Analysis and Optimization of Thermal Issues in High-Performance VLSI", in *Proc. International Symposium on Physical Design*, pp. 230-237, 2001.

[3] K.-Y. Chao and D. F. Wong, "Thermal Placement for High-Performance Multichip Modules", in *Proc. International Conference on Computer Design*, pp. 218-223, 1995.

[4] J.-L. Tsai, C. C.-P. Chen, G. Chen, B. Goplen, H. Qian, Y. Zhan, S.-M. Kang, M. D. F. Wong, and S. S. Sapatnekar, "Temperature-Aware Placement for SOCs", in *IEEE Special Issue on On-Chip Thermal Engineering*, Aug., 2006, pp. 1502-1518.

[5] C. C. N. Chu and D. F. Wong, "A Matrix Synthesis Approach to Thermal Placement", in *Proc. International Symposium on Physical Design*, pp. 163-167, 1997.

[6] R. Cobbold, "Temperature Effects on MOS Transistors", *Electronic Letters*, vol. 2, pp. 190-192, 1966.

[7] J. Cong, J. Wei, and Y. Zhang, "A Thermal-Driven Floorplanning Algorithm for 3D ICs", in *Proc. International Conference on Computer-Aided Design*, 2004.

[8] H. Eisenmann and F. M. Johannes, "Generic Global Placement and Floorplanning", *Proc. Design Automation Conf.*, pp. 269-274, 1998.

[9] J. Gu and X. Huang, "Efficient Local Search with Search Space Smoothing: A Case Study of the Traveling Salesman Problem (TSP)," in *IEEE Trans. Systems, Man and Cybernetics* 24(5) (1994), pp. 728-735.

[10] S. Im, N. Srivastava, K. Banerjee, and K. E. Goodson, "Scaling Analysis of Multilevel Interconnect Temperatures for High-Performance ICs", *IEEE Trans. on Electron Devices*, 52(12), 2005.

[11] A. B. Kahng and Q. Wang, "Implementation and Extensibility of an Analytic Placer", *Proc. Int. Symp. Physical Design*, 2004, pp. 18-25.

[12] A. B. Kahng, B. Liu and Q. Wang, "Supply Voltage Degradation Aware Placement", in *International Conference on Computer Design*, pp. 437-443, 2005.

[13] A. A. Keshavarz, P. Khare and R. K. Sampson, "Comprehensive Modeling of MOS Transistors in a $0.35\mu m$ Technology for Analog and Digital Applications", in *Proc. International Conference on Modeling and Simulation of Microsystems*, 2002.

[14] W. Liao, L. He, and K. Lepak, "Temperature-Aware Performance and Power Modeling", *Technical Report UCLA Engr. 04-250*, 2004.

[15] Z. Lu, J. Lach, M. Stan and K. Skadron, "Temperature-Aware Modeling and Banking of IC Lifetime Reliability", *University of Virginia Technical Report CS-2005-10*, 2005.

[16] W. Naylor et al., "Non-Linear Optimization System and Method for Wire Length and Delay Optimization for an Automatic Electric Circuit Placer", *US Patent* 6301693, Oct. 2001.

[17] A. Odabasioglu, M. Celik and L. T. Pileggi, "PRIMA: Passive Reduced-Order Interconnect Macromodeling Algorithm," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 17(8), 1998, pp. 645-654.

[18] M. Pedram and S. Nazarian, "Thermal Modeling, Analysis and Management in VLSI Circuits: Principles and Methods", *IEEE Special Issue on On-Chip Thermal Engineering*, Aug. 2006, pp. 1487-1501.

[19] R. Severns, "Safe Operating Area and Thermal Design for MOS Power Transistors", in *Siliconix Application Note AN83-10*, 1983.

[20] C.-H. Tsai and S.-M. Kang, "Cell-Level Placement for Improving Substrate Thermal Distribution", in *IEEE Trans. on Computer-Aided Design*, 19(2), pp. 253-266, 2000.

[21] Y. P. Tsividis, *Operational Modeling of the MOS Transistor*, McGraw-Hill, Inc., 1987.

[22] N. Viswanathan and C. C.-N. Chu, "FastPlace: Efficient Analytical Placement Using Cell Shifting, Iterative Local Refinement and a Hybrid Net Model", *Proc. Int. Symp. Physical Design*, 2004, pp. 26-33.

[23] Y. Zhan and S. S. Sapatnekar, "Fast Computation of the Temperature Distribution in VLSI Chips Using the Discrete Cosine Transform and Table Look-up", in Proc. Asia and South Pacific Design Automation Conference, pp. 87-92, 2005.

[24] Y. Zhan and S. S. Sapatnekar, "A High Efficiency Full-Chip Thermal Simulation Algorithm", in *Proc. International Conference on Computer-Aided Design*, 2005.