

# Taking Thermal Considerations into Account during High-Level Synthesis

JEN-PIN WENG and ALICE C. PARKER\*

*Department of Electrical Engineering—Systems, University of Southern California, University Park, Los Angeles CA 90089-2562*

Submicron feature sizes result in designs in which power density is significantly increased. High-level synthesis strategies are proposed in this paper to relieve potential thermal problems. Operators are placed as close as possible to their data predecessors in order to minimize the interconnection cost while not violating the thermal constraints. Spreading overused functional units away from the thermal problem area often results in performance degradation. Introducing redundant operators is suggested to reduce the module utilization and hence thermal problems among problem modules when system performance is important. Our experimental results show that this technique produces quite satisfactory results for a power-dominated example.

*Keywords:* Scheduling, reliability, floorplanning, thermal analysis

## 1. INTRODUCTION

As integrated circuit feature sizes have decreased, designers have been forced to consider physical effects at higher levels of the design process. An increase in power density results from reduced feature sizes, partly due to higher operating speeds, and partly due to the increased component density on integrated circuits. However, the operating temperature of a chip is limited to a certain range for acceptable system reliability. For silicon devices, this temperature is in the range of 75–85°C [1]. With increasing power density and with limits on the operating temperature, thermal limitations of the chip must be considered during the design process. Consequently, it is imperative to study not only the thermal properties of the device

and package material but also the run-time thermal properties of the chip/die so that, during the design stage, a better thermal layout can be obtained to alleviate potentially high thermal stress. While thermal effects on-chip might be relatively inconsequential for many single MOS chips, the current thermal problems associated with multichip modules indicate that thermal effects cannot be ignored during multichip synthesis.

For the design of electronic circuits for reliability, there are a number of handbooks, specifications and guidelines which help us establish a common basis for comparing, evaluating and predicting related or competitive designs. The Military Handbook MIL-HDBK-217E [2] was developed by the Department of Defense to unify reliability prediction methods for

---

\*Corresponding author. Tel.: (213) 740-4476. Fax: (213) 740-9803. E-mail: parker@eve.usc.edu.

integrated circuits produced by the military. According to this handbook, surface temperature is one of the major factors affecting circuit reliability. Computer-aided design software should predict and analyze thermal properties in circuits while they are under design, highlighting areas which are overused in order to prevent such failures. Engineers can then improve designs so that systems themselves operate at lower temperatures and more reliably.

Localized heat-concentration problems can appear when a high-level synthesis tool or designer uses a greedy approach to the allocation and binding synthesis subtasks which overuses some central resources. The scheduling procedure, the schedule and topology of functional modules and the selection of the package material all affect the design enormously. Since not all functional modules are active all the time, utilizations of functional modules, which are the result of scheduling, can also affect the reliability of the design.

Thermal effects thus can represent a limiting factor in the development of ASIC chips. An accurate model of the thermal behavior of the die structure is necessary in order to make reliable designs. Thermal analysis methods can be categorized into two general approaches, namely, analytical solutions and numerical solutions. The analytical approach involves the search for an exact analytical solution for structures with regular geometries. An analytical solution is not always obtainable for a structure with complex geometries. Therefore, a numerical method is a better approach for a structure with irregular geometries. Numerical techniques such as finite-difference [3] finite-element [4], and boundary element [5] have been widely used to analyze thermal profiles of electronic circuits.

The purpose of this work is to provide a novel method for improving overall thermal characteristics of circuits during the high-level synthesis process. We assume the amount of heat produced by a functional module is proportional to its work load. The basic idea in this work is to alleviate the heat-concentration phenomenon by averaging work loads between modules of the same functionality during scheduling and by balancing power dissipation during floorplanning. Averaging or spreading the power

dissipation of functional modules allows the design to operate with a more balanced thermal profile.

In the following sections, we introduce our approach to the problem, and present thermal models for a chip under thermal consideration. An analytical solution is then derived. A numerical finite-difference method is also presented. An example is used to illustrate thermal impacts on floorplanning and scheduling. Finally, some remarks on the thermal analysis are given.

### 1.1 Problem Approach

Our approach to this problem is to combine thermal analysis with an existing scheduling/floorplanning program called 3D scheduling. 3D scheduling simultaneously constructs a floorplan during scheduling, allocation and binding. 3D scheduling can introduce additional ("redundant") operators to alleviate wiring delays, even if other operators are free during a given clock cycle. We perform thermal analysis on the floorplans produced by 3D scheduling, and illustrate how the thermal profile can be improved by either moving operators away from the high-temperature area (possibly causing performance to degrade) or by creating additional redundant operators to smooth out the thermal profile. Although the thermal analysis programs are coded and tested, we have manually modified the floorplans produced by 3D scheduling, to illustrate our ideas.

## 2. THERMAL MODELS

To calculate the induced thermal stress of a design, the temperature profile must be known. In our analysis, the heat conduction from the top of the die surface to the working fluid (usually, the working fluid is air) is assumed to be negligible, as compared to the heat conducted laterally through the doped substrate to the print circuit board. The thermal conductivities of the substrate and package material are constant in the steady state, and the surface temperature is prima-

rily a function of functional module utilizations and the topology of functional modules.

For a steady-state conductive cooling environment, the single-chip heat transfer environment can be modeled by a network of thermal resistances as shown in Figure 1 [6]. In the thermal resistance model, the temperature and heat flux are analogous to the voltage and current in the analysis of an electronic circuit. The thermal resistances in Figure 1 are completed as follows:

$$R_t = \frac{W_c}{K \cdot A_c} \quad (1)$$

where  $W_c$  the distance between the conduction,  $K$  is the thermal conductivity of the conductive material and  $A_c$  is the cross-sectional area of the conduction. Note that Equation 1 only considers one-dimensional heat flow. Since the amount of heat produced inside a chip and the ambient temperature are known, the temperature on the surface of package can then be derived from the network of thermal resistances.

In order to obtain the analytical solution of the temperature profile on the surface of the silicon substrate, we modified this thermal resistance model to a four-layer thermal model as shown in Figure 2 [7]. The thermal characteristics are retained in this simplified model. For a common silicon device, the first (top) layer is usually silicon. The second layer is usually a bonding material, such as *epoxy* or solder. The third layer represents metallization on the substrate (usually gold). The fourth (bottom) layer, the package, uses material such as alumina. Several assumptions have been made in this four-layer thermal model:

- The surface temperature at any position on the silicon depends on the heat dissipated by the functional modules, the thermal conductivities of the silicon and package material and the temperature of the working fluid (note that the heat dissipated by convention and radiation is not considered in our current approach).
- The heat generated from each functional module on the substrate is assumed to be uniformly distributed. The heat dissipation of a functional module is dependent on the clock frequency and its utilization.
- The average power dissipation of a functional module is used in the thermal analysis, since the thermal time constant is much greater than the period of the clock signal.

Note that the average power dissipation of a functional module is the product of the maximal power dissipated at the working clock frequency and its utilization. For example, the power dissipated by a typical adder circuit when operating at a frequency of 10 MHz and  $V_{DD} = +5$  volts is about 2 mW. If the utilization of this adder on a given design is 80%, the average power dissipation of this adder is 1.6 mW.

### 3. DERIVATION OF THE ANALYTICAL SOLUTION

Due to the small thickness of the bonding and metallization layers, only the first silicon layer and the fourth package layer are considered. Therefore, the four-layer thermal model shown in Figure 2 is further simplified to a two-layer thermal model as shown in

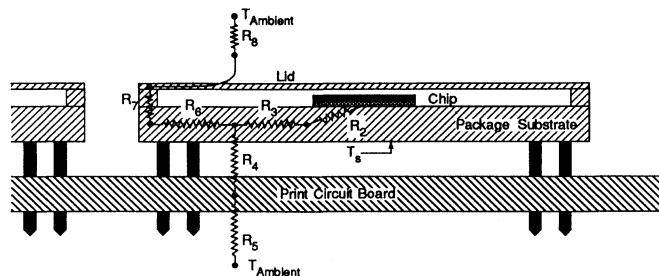


FIGURE 1 A typical single-chip heat transfer model.

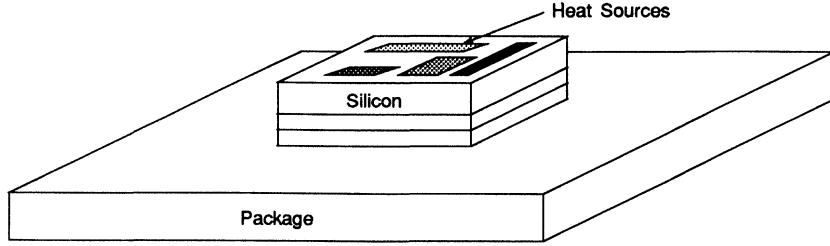


FIGURE 2 A four-layer chip thermal model.

Figure 3 [8]. In this model, the package size is assumed to be the same as the die size; only the thermal properties of the silicon and package layers are considered. To derive the temperature distribution of the die structure in the steady state, the classic heat-flow govern equation must be solved:

$$\nabla^2 T(x,y,z) = 0 \quad (2)$$

where the thermal conductivity of the medium is assumed not to be a function of temperature. The need for a three-dimensional solution is due to the fact that heat transfer in the die is three-dimensional. The following boundary conditions are considered:

- The bottom surface (the package surface) has an arbitrary but known temperature distribution  $T_0(x,y)$ .
- The lateral sides of the layer structure are considered to be adiabatic (adiabatic means no heat transfer between the analyzed structure and the working environment).
- The energy flux dissipated on the top surface of the silicon chip is described by the function  $P(x,y)$ .

- The energy flux is continuous at the interfaces between layers.

Equation 2 can be solved by separation of variables or other equivalent techniques. The solution that calculates the temperature on the top silicon surface is thus derived as [8]

$$\begin{aligned} T(x,y,0) = & T_0(x,y,h_1 + h_2) + \left( \frac{h_1}{K_1 L_1 L_2} \right. \\ & \left. + \frac{h_2}{K_2 L_1 L_2} \right) \cdot \int_0^{L_1} \int_0^{L_2} P(x,y) dx dy \\ & + \frac{4}{K_1 L_1 L_2} \cdot \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \\ & \frac{\int_0^{L_1} \int_0^{L_2} P(x,y) \cos\left(\frac{m\pi x}{L_1}\right) \cos\left(\frac{n\pi y}{L_2}\right) dx dy}{A_{mn} \cdot (\delta_m + 1) \cdot (\delta_n + 1)} \end{aligned}$$

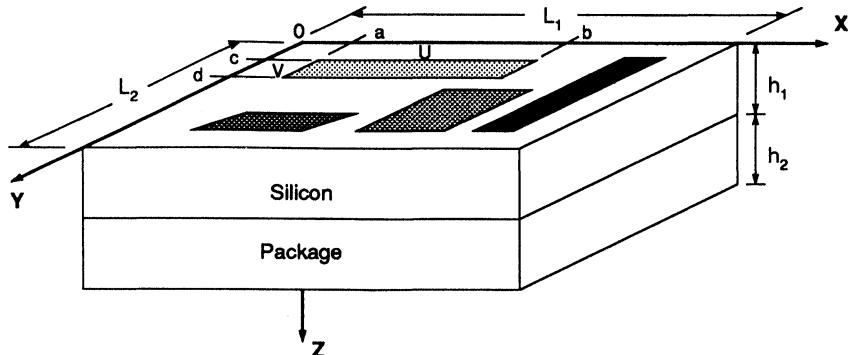


FIGURE 3 A simplified layered chip thermal model finite difference approximation and the corresponding nodal network.

$$\frac{\sinh A_{mn}h_1 + \frac{K_1}{K_2} \cdot \cosh A_{mn}h_1 \cdot \tanh A_{mn}h_2}{\frac{K_1}{K_2} \cdot \sinh A_{mn}h_1 \cdot \tanh A_{mn}h_2 + \cosh A_{mn}h_1} \cdot \cos\left(\frac{m\pi x}{L_1}\right) \cdot \cos\left(\frac{n\pi y}{L_2}\right) \quad (3)$$

where

$$A_{mn} = \sqrt{\frac{m^2\pi^2}{L_1^2} + \frac{n^2\pi^2}{L_2^2}}$$

$\delta_m$  and  $\delta_n$  are the Kronecker delta and  $K_1, K_2$  are the thermal conductivities of the first layer and the second layer, respectively. Note that the exact solution of the four-layer structure can be derived in a similar manner [7].

In the case of ASIC design, the power distribution  $P(x,y)$  can be modeled by a group of uniformly distributed functional modules. This fact suggests that we can represent the function  $P(x,y)$  as a set of box functions (a box function,  $B(x,y)$ , is a function with zero-value everywhere, except a constant value inside a rectangle area), with the amplitude  $Q_{fi}$  for each functional module. Due to the linearity of the Laplace equation (Equation 3), the solution for the structure with a single heat source is first derived. For the structure with multiple sources, the solutions are then obtained by the superposition of corresponding single-source solutions. Two integrals in Equation 3 are thus derived as

$$\int_0^{L_1} \int_0^{L_2} P(x,y) dx dy = \sum_{i \in M} Q_{fi} \quad (4)$$

$$\int_0^{L_1} \int_0^{L_2} P(x,y) \cos\left(\frac{m\pi x}{L_1}\right) \cos\left(\frac{n\pi y}{L_2}\right) dx dy = \sum_{i \in M} Q_{fi} \cdot \left\{ \text{sinc}\left(\frac{m\pi U}{2L_1}\right) \cdot \text{sinc}\left(\frac{n\pi V}{2L_2}\right) \cdot \cos\left(\frac{m\pi(a+b)}{2L_1}\right) \cdot \cos\left(\frac{n\pi(c+d)}{2L_2}\right) \right\} \quad (5)$$

where

$$\text{sinc}(g) = \frac{\sin(g)}{g}$$

$Q_{fi}$  is the heat dissipated by module  $i$  and  $M$  is the allocated module set.

Since the solution has the form of an infinite series, a criterion is needed to truncate the summation at a given required precision. Through a closer inspection of the infinite double Fourier cosine series, a rule of thumb,  $m = 6L_1/U$ ,  $n = 6L_2/V$ , is used which generally allows the temperature to converge within 1 percent of the final value [7].

#### 4. DERIVATION OF THE NUMERICAL SOLUTION

The basic principle of the finite difference approach is derived from differential equations via Taylor's expansion. Instead of using finite difference equations directly, these equations can be interpreted in a physical way to permit a more convenient application. Consider the die structure shown in Figure 4. The die is divided into a number of cubes. The cross-sectional area of each cube equals the area of a unit cell. We

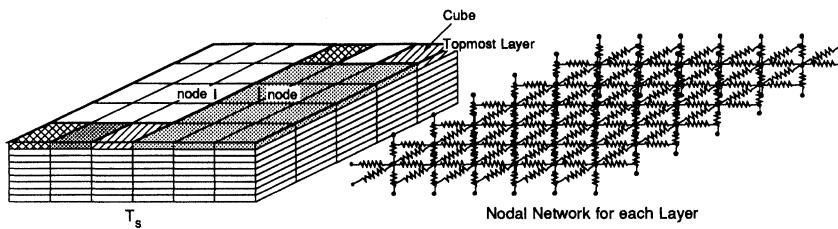


FIGURE 4 Nodal Network for the Simplified Model

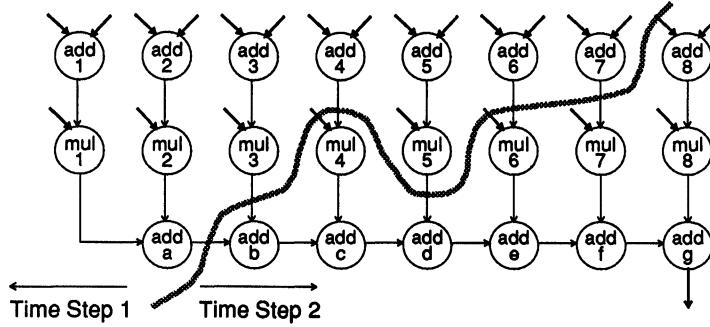


FIGURE 5 A 2-time-step non-pipelined FIR filter design.

use a constructive approach to floorplanning in the preliminary 3D scheduling research. We define a unit cell as a primitive block so a larger functional module (such as a multiplier) is divided into several cells. In this case, the heat dissipated in the module is assumed to equally distribute among all topmost cubes which cover the module. The heat dissipation of a cube is assumed to be a point heat source on the geometric center of the cube. A nodal network is thus obtained, representing the structure under a steady-state condition. Each node  $i$ , which is the geometric center of cube  $i$ , must satisfy the equations

$$\sum_{j \in B_i} \frac{t_j - t_i}{R_{ij}} + q_i = 0 \quad (6)$$

where  $B_i$  is the set of all neighboring nodes adjacent to node  $i$  and  $R_{ij}$  is the thermal resistance between node  $i$  and node  $j$  which equals to  $\delta_{ij} / k_s A_{ij}$ .  $\delta_{ij}$  denotes the conduction distance between node  $i$  and node  $j$  and  $A_{ij}$  is the cross-sectional area for heat conduction normal to  $\delta_{ij}$ .  $q_i$  is the heat produced in the volume lump at  $i$  (i.e. cube  $i$ ) which is the average heat dissipated in the module divided by the number of topmost cubes covering the module. Note that the heat produced by a module is assumed only coming from the topmost layer of cubes, since the power is mainly consumed by semiconductor circuits implanted on the surface of the silicon substrate, during signal switching.

The formulation so far is restricted to interior points of the structure in which the heat conduction is taking place. In order to solve the equation set repre-

sented by Equation 6, imposed boundary conditions must also be satisfied. The boundary condition here is the equilibrium temperature on the bottom of the structure surface which is the surface of the package [6]. The temperature is assumed to be uniformly distributed on the surface of the package due to the high heat conductivity of the package material. By solving the  $n$ -equation set simultaneously, the temperature profile can be obtained. The profile of the thermal stress can then be calculated.

## 5. THERMAL EXPERIMENTS AND RESULTS

The thermal model presented in this chapter was used to experiment with the ADAM high-level synthesis tools developed at USC. Figure 5 shows an example of a data flow graph which is used as an input for high-level synthesis. Table I lists the module library set, which was derived from the Cascade Design Automation ChipCrafter silicon compiler. The data path synthesis program used to experiment with the thermal model is based on the preliminary 3D scheduling research, which incorporates interconnection delays during scheduling using floorplanning. A two-time-step schedule for a non-pipelined FIR filter design is shown with the cross-hatched line in Figure 5.

TABLE I. Library set created by ChipCrafter silicon compiler.

1.2 $\mu\text{m}$ Technology	8 bit adder		8 bit multiplier	
	Delay	Area	Delay	Area
	13 ns	75 mil $^2$	32 ns	903 mil $^2$

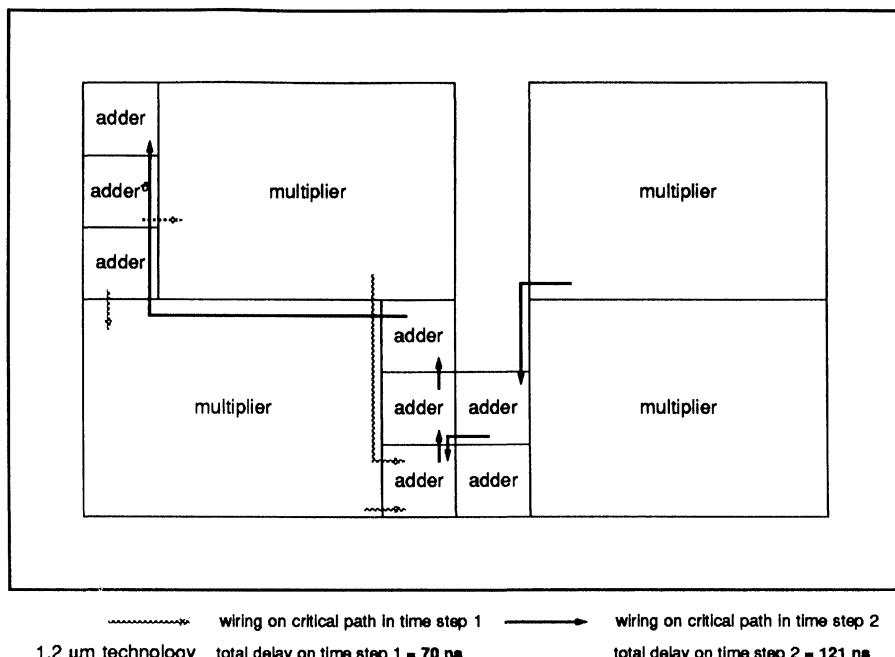


FIGURE 6 The floorplan for the FIR filter design with minimum operators.

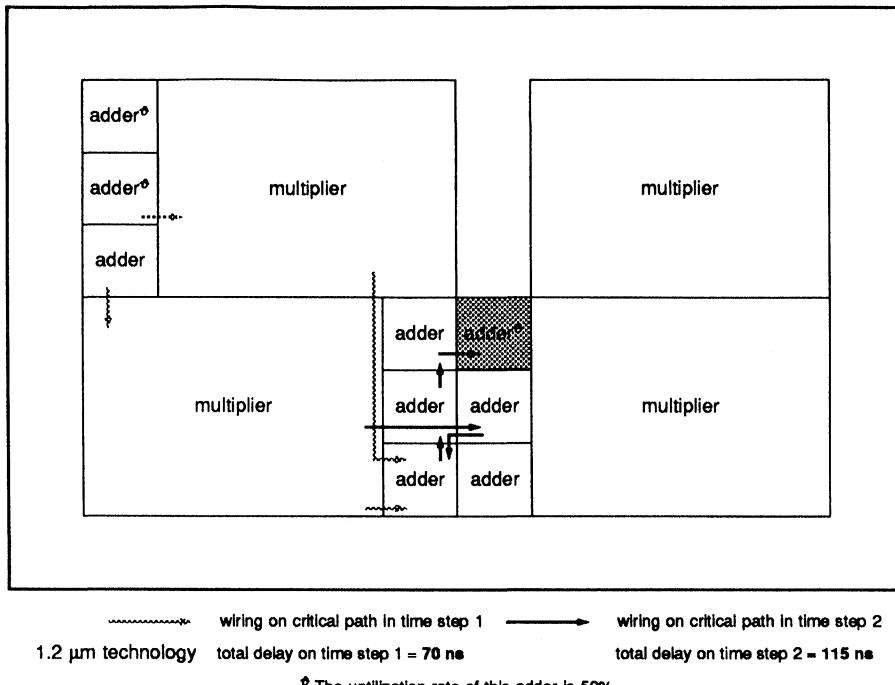


FIGURE 7 The floorplan for the FIR filter design with a redundant adder.

The 3D scheduling algorithm minimizes interconnection delays along critical paths to improve the design performance. The software tries to introduce *redundant operators* (redundant operators are operators not required for the feasible design with minimum operators) to alleviate interconnection delays, when the synthesized design cannot achieve user specified timing constraints [9]. For example, the floorplan of the two-time-step non-pipelined FIR filter design in Figure 5 is shown in Figure 6. In this example, the feasible design with minimum operators allocated contains 8 adders and 4 multipliers. The software found that interconnection delays along critical paths are reduced by introducing one more adder, which is shown shaded in Figure 7, in this two-time-step non-pipelined FIR filter design.

We assume the size of the floorplan is 1.74 mm by 1.16 mm. The energy dissipated by an adder and multiplier is 100 mW and 10 mWat 100% utilization, respectively. The top layer of the die is silicon its thickness is 0.38 mm and thermal conductivity is 46  $W/m^2C$ . The bottom layer of the die is alumina and its thickness and thermal conductivity are 0.785 mm and 20  $W/m^2C$ , respectively. The thermograms of the design with minimum operators and the design with a redundant adder are shown in Figures 8a and 8b, respectively. The temperature shown in thermograms is the temperature difference between the top surface of the silicon substrate and the bottom surface of the package. The thermal profiles were calculated by both analytical and numerical methods. The results of these two solutions match, but the analytical solution provides a more flexible and efficient way to compute simulation results. The complexity of the analytical solution is dependent on the number of heat sources and the number of terms in the infinite Fourier series to be summed. For example, in the design with minimum operators, it took 2.2 seconds of CPU time on a Solbourne 5e/900 machine (or 2.25 seconds of CPU time on a SUN 4/460 machine) on the average to compute one sample temperature when there are twelve heat sources on the top surface of the silicon substrate.

In order to demonstrate the problem of heat concentration, adders in our experiments are considered

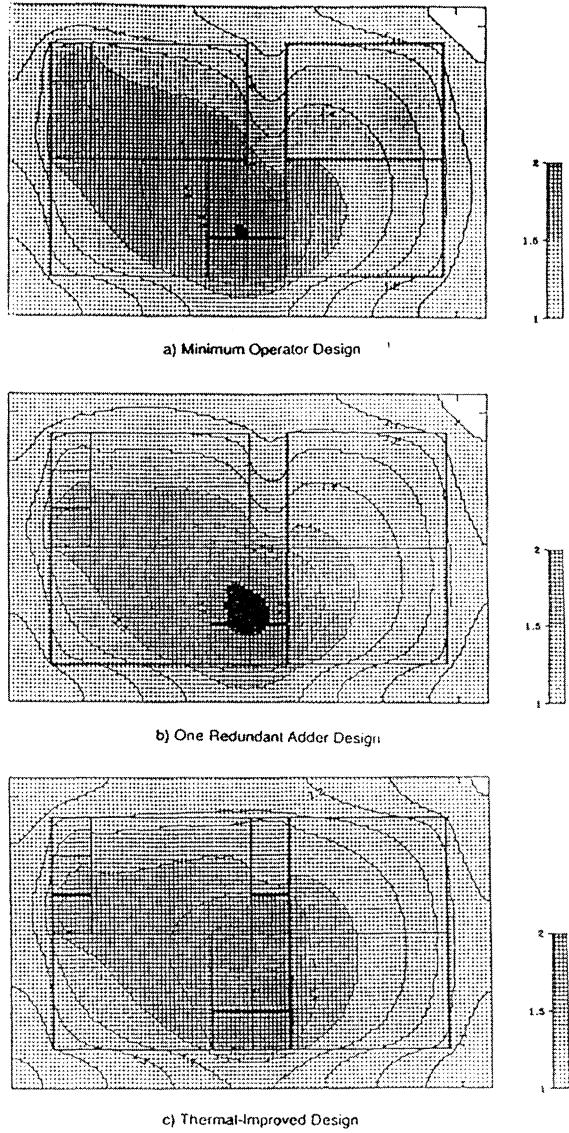


FIGURE 8 The thermograms of the FIR filter design.

to be “hot-problem devices” which produce three times the amount of heat that multipliers produce when both adders and multipliers are fully utilized. The thermogram of the design with minimum operators in Figure 8a presents a smaller “hot-spot” as compared to the thermogram of the design with a redundant adder in Figure 8b. The reason was obvious after these two floorplans were compared and analyzed. The design with a redundant adder design

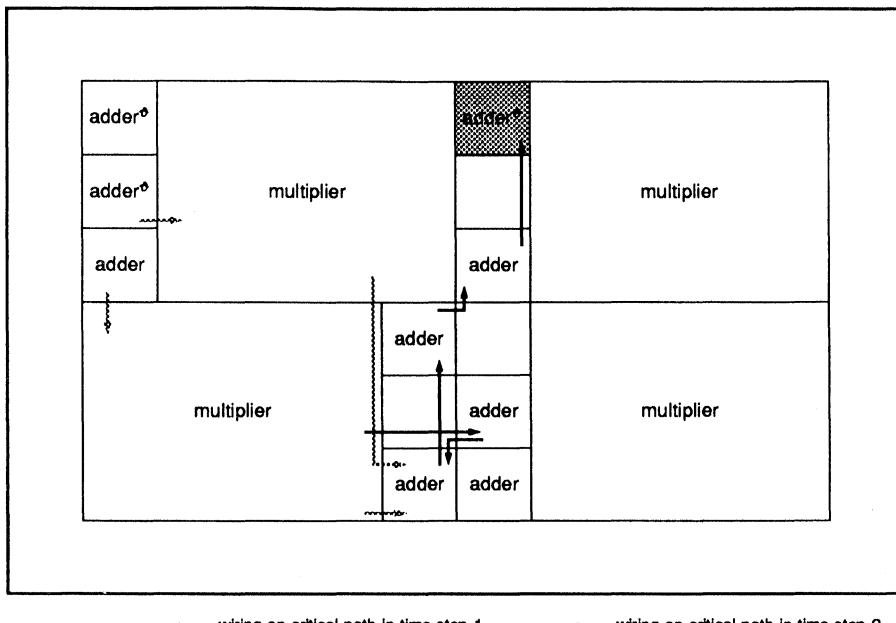


FIGURE 9 The floorplan for the heat-balanced FIR with a redundant adder.

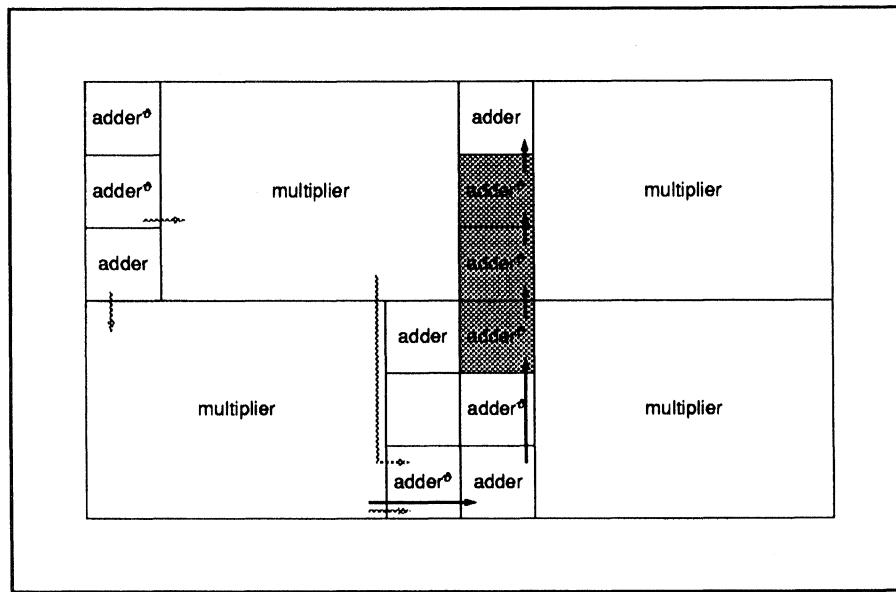


FIGURE 10 The floorplan for the heat-balanced FIR with redundant adders.

does produce better system performance as compared to the design with minimum operators. However, the design with a redundant adder attempts to use modules around the central area intensively to shorten wiring delays, which results a large amount of heat being produced in the central area, causing heat dissipation problems.

To resolve the heat-concentration problem, two solutions are proposed. First, by spreading adders around the problem area over the unused space on the floorplan, the heat-concentration problem can be alleviated. We applied this strategy to the FIR filter example. The resultant floorplan and thermogram are shown in Figures 9 and 8c, respectively. The results are successful. The temperature dropped about 20% (the dropped percentage is the ratio of reduced temperature versus overall temperature difference). This heat-balancing strategy resolved the heat-concentration problems at the cost of performance degradation.

In a design with critical timing constraint, the synthesis program or designer may not allow any degradation of the system performance. In this case, we propose another strategy by allocating additional redundant adders to reduced utilizations of adders around the central area of a die or module. The resulting floorplan is shown in Figure 10. Three redundant adders were allocated to reduce the heat-concentration problem in this example. Utilizations of some adders are reduced in this floorplan. The related thermogram shown in Figure 11 reveals that both the thermal constraint and performance constraint are achieved simultaneously. The thermal improvement achieved is the same as in the previous case; the tem-

perature is reduced about 20% (Indeed, the thermal profile of this floorplan is slightly better than the previous one.). These two design cases whose thermal properties have been improved illustrate that design tradeoffs are possible among area, performance and reliability factors. We believe that tradeoffs can be done much better if thermal effects are considered during the data path synthesis process instead of considering them as two problems separately and/or sequentially.

## 6. CONCLUSIONS AND FUTURE RESEARCH

The main objective of this work is to floorplan heat-balanced designs by avoiding overuse of the functional modules around the central area. We have demonstrated a localized heat concentration problem which induces a high thermal stress in the problem area and causes reliability problems. It also shows that thermal problems may be successfully alleviated by rearranging functional modules around the problem area or by introducing extraredundant operators. Both analytical and numerical solutions were investigated in our study. However, the analytical solution gives us more flexibility and efficiency during the computation of thermal profiles.

In the future work, a more exact thermal model will be studied which allows more layers to be considered in the computation of thermal profiles. Computation time may be reduced by using an infinite plate model [10], when the dimension ratio of heat sources to die structure is large.

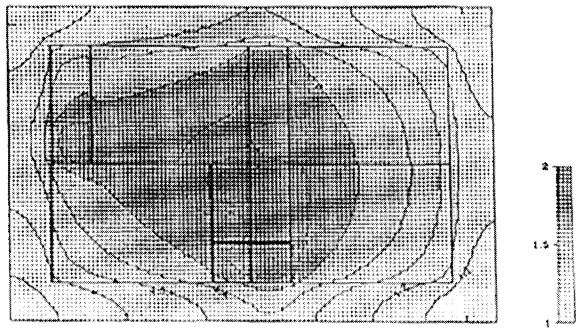


FIGURE 11 The thermogram of the heat-balanced FIR with redundant adders.

### Acknowledgements

This work was supported by the Semiconductor Research Corporation under contract 92-DJ-075.

### References

- [1] A. J. Boldgett Jr. Microelectronic Packaging. *Scientific American*, pages 86–96, July 1983.

- [2] Military Handbook MIL-HDBK-217E. *Reliability Prediction of Electronic Equipment*. Rome Air Development Center, February 1990.
- [3] L. M. Mahalingam, J. A. Andrews, and J. E. Drye. Thermal Analysis on Pin Grid Array Packages for High Density LSI and VLSI Logic Circuits. *IEEE Transactions on Components, Hybrids, Manufacturing Technology*, CHMT-6:246–256, September 1983.
- [4] D. L. Waller, L. R. Fox, and R. J. Mannemann. Analysis of Surface Mount Thermal and Thermal Stress Performance. *IEEE Transactions on Components, Hybrids, Manufacturing Technology*, CHMT-6:257–266, September 1983.
- [5] C. C. Lee, A. L. Palisoc, and J. M. Bayham. Thermal Analysis of Solid-State Devices Using Boundary Element Method. *IEEE Transactions on Electronic Devices*, 35(7):1151–1153, July 1988.
- [6] R. Tummala and E. Rymaszewski. *Microelectronics Packaging Handbook*, chapter 4, pages 167–224. Van Nostrand Reinhold, 1989.
- [7] C. C. Lee, A. L. Palisoc, and Y. J. Min. Thermal Analysis of Integrated Circuit Devices and Packages. *IEEE Transactions on Components, Hybrids, Manufacturing Technology*, 12(4):701–709, December 1989.
- [8] R. Castello and P. Antognetti. Integrated-Circuit Thermal Modeling. *IEEE Journal of Solid-State Circuits*, SC-13(3):363–366, June 1978.
- [9] J. Weng and A. C. Parker. 3D Scheduling: High-Level Synthesis with Floorplanning. In *Proceedings of 28th ACM/IEEE Design Automation Conference*, pages 668–673, July 1991.
- [10] C. C. Lee, Y. J. Min, and A. L. Palisoc. A General Integration Algorithm for the Inverse Fourier Transform of Four-Layer Infinite Plate Structure. *IEEE Transactions on Components, Hybrids, Manufacturing Technology*, 12(4):710–716, December 1989.

where she has been on the faculty since 1980. She currently serves the university as Vice Provost for Graduate Studies and Vice Provost for Research. Formerly, she was an Assistant Professor at Carnegie Mellon University. She obtained her B.S.E.E. and Ph.D. degrees from N.C. State University and her M.S.E.E. from Stanford University. Her research interests are high-level synthesis of digital systems, CAD frameworks, design representation, and engineering ethics. She is on the Board of Directors of the IEEE Computer Society.

Jen-Pin Weng received the B.S. degree in Mechanical Engineering from National Taiwan University, Taiwan, in 1986 and the M.S. and Ph.D. degrees in Computer Engineering from University of Southern California, CA in 1990 and 1993, respectively. Since 1993 he has been with Synopsys, Inc., Mountain View, CA, where he is involved in the development of a system-level mapping/partitioning system. His main research interest is in the area of VLSI CAD considering physical design parameters.

## Authors' Biographies

Dr. Alice C. Parker is a Professor of Electrical Engineering at the University of Southern California,

