INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]

Efficient Numerical Thermal Simulation Using a Quad-Tree Mesh

by

Lei Zhang, B.Eng.

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Applied Science

Ottawa-Carleton Institute of Electrical Engineering
Department of Electronics
Carleton University
Ottawa, Ontario, Canada

November 2001

© Lei Zhang, 2001

**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**395 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**395, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-68317-6

Canada

The undersigned recommend to
the Faculty of Graduate Studies and Research
acceptance of this thesis

Efficient Numerical Thermal Simulation Using a Quad-Tree Mesh

Submitted by Lei Zhang, B.Eng.
in partial fulfillment of the requirements for the degree of

Master of APPLIED SCIENCE

Thesis Supervisor

Chair, Department of Electronics

Carleton University
Ottawa, Ontario, Canada

November 2001

Abstract

This research develops an efficient method for the numerical thermal simulation of semiconductor devices for the rapid and accurate prediction of the maximum temperature rise in device junctions. This method is based on quad-tree meshes and consists of approximation approaches regarding the generation of square-block based numerical meshes, optimization of substrate boundary dimensions in the computational domain, and quantization of heat source dimensions and depth from the substrate surface.

A 3D numerical thermal simulator, called Atar, was used throughout the research for creating numerical meshes. Simulation results were compared to those obtained from source simulations that were based on an analytical method.

A case study is included which demonstrates that by using these approximation approaches in Atar, the number of numerical blocks is reduced by a factor of approximately 90, the CPU time is reduced by a factor of almost 120, and the temperature error is within 3% compared to the temperature before quantization in a semi-infinite substrate. In addition, after the quantization and optimization of a given substrate which has a quarter structure of $200\mu\text{m} \times 200\mu\text{m} \times 100\mu\text{m}$, the number of numerical blocks is reduced by a factor of approximately 60, the CPU time is reduced by a factor of almost 40, and the temperature error after quantization is within 5% compared to the temperature before quantization. Therefore, these approximation approaches greatly reduce the complexity and processing time required for simulations yet incur only a small, acceptable level of error, making them very helpful in integrated device thermal modeling.

Acknowledgments

First, I would like to thank my supervisor, Professor David J. Walkey. Thanks for your encouragement, patience, and instructions, all of which have made this research a truly fulfilling experience.

My thanks also give to Professor Tom Smy for his assistance, and continuous improvement of the numerical thermal simulator, Atar, which was used throughout this research.

I would like to thank all the organizations that provide me the financial support for this research.

I would like to thank our system administrator, Jacques Lemieux, for his generosity and maintenances of my huge computer account.

I would also like to thank the staff and faculty in the Electronics Department at Carleton University, my fellow research group members Gerardo Romo and Dritan Celo for their continuous help and sharing with their ideas, as well as Helen Triantafellou and Peter Knazko who helped to edit this thesis.

Last but not the least, I would like to thank my family and friends for encouragement. Thank you all.

Table of Contents

Chapter 1: Introduction	1
1.1 Research Motivations.....	1
1.2 Research Objectives.....	2
1.3 Thesis Organization	3
 Chapter 2: Background	 5
2.1 Thermal Analysis Methods.....	5
2.1.1 Comparison of Numerical Method with Analytical Method.....	5
2.1.2 Numerical Method Approximation.....	6
2.1.3 Grid Refinement in Numerical Method	8
2.2 Boundary Conditions and Simulation Size	9
 Chapter 3: Grid Generation Using Atar.....	 16
3.1 Atar Numerical Thermal Simulator	16
3.2 Model Building Using Atar	19
3.2.1 Mesh Construction Mechanism	20
3.2.2 Rules in Setting Up Parameters	25
3.2.3 Procedures in Setting Up Parameters.....	36
3.2.4 Grid Independence	39
3.3 Multiple Sources and Gridding Mask Overlapping	58
3.4 Conclusion	65
 Chapter 4: Finite Substrate Effects in Thermal Simulation	 66
4.1 Introduction.....	66
4.1.1 Model Description	66

4.1.2 Modeling of Heat Spreading and Constriction	68
4.1.3 Simulation Size and Accuracy	71
4.1.4 Power Dissipation, Temperature Rise (T_{rise}) and Normalization	74
4.2 Simulation Studies Using Atar	78
4.2.1 Temperature Rise (T_{rise}) With Respect to Combination of t_{sub} & L	78
4.2.2 Given Small t_{sub} , Choose L	86
4.2.3 Shrinking of Substrates with a Large Thickness	90
4.2.3.1 Substrate Shrinking for Single Square Sources	91
4.2.3.2 Substrate Shrinking for Single Rectangular Sources	94
4.3 Line Source Simulation	97
4.3.1 Line Source Simulation Results Compared with Single Square Source Simulation Results	97
4.3.2 Line Source Simulation Results Comparing to Single Rectangular Source Simulation Results	100
4.4 Bounding Box Simulation for Multiple Sources	112
4.5 Conclusion	122
Chapter 5: Source Quantization	123
5.1 Purpose	123
5.2 Single Square Source Quantization	132
5.3 Single Rectangular Source Quantization	140
5.4 T_{rise} Sensitivities and Aspect Ratios (l/w)	147
5.4.1 Curve Fitting	147
5.4.2 T_{rise} Sensitivity vs. Aspect Ratios	148
5.5 Evaluation of Quantization Ranges in l and w	151
5.5.1 Problems with Small Source Width and Aspect Ratio (l/w)	152
5.6 Quantization in Source Depth (d) and Thickness (t)	157
5.6.1 Effects of d and t on T_{rise} Sensitivities to l and w	157
5.6.2 Relations Between d and t and Their Effects on T_{rise}	160
5.7 Atar Simulation Using Quantization Rules	166

5.8 Multiple Source Quantization	172
5.9 Conclusion	180
Chapter 6: Case Study	182
Chapter 7: Conclusions and Future Studies	194
7.1 Conclusions.....	194
7.2 Summary of Contributions.....	197
7.3 Future Work.....	198
Reference.....	201

List of Table

Table 2-1 Comparisons of Thermal Simulation Results Based on the Entire Symmetrical Structure and one Quarter	15
Table 3-1 Comparison of Mesh Structures and Temperatures for Example 1	28
Table 3-2 Rules in Defining Gridding Mask Depths	35
Table 3-3 Possible Depth for Each Numerical Region	36
Table 3-4 Comparison of Mesh Structures and Temperatures for Example 2	48
Table 3-5 Comparison of Mesh Structures and Temperatures for the $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ Source and $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ Source	56
Table 3-6 Comparison of Mesh Structures and Temperatures for the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ Source and $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Source	57
Table 3-7 Comparison of Mesh Structures and Temperatures for the Multiple-Source Example 1 ...	59
Table 3-8 Comparison of Mesh Structures and Temperatures for the Multiple-Source Example 2 ...	65
Table 4-1 Comparison of Line-Source Simulation and Square Source Simulation	99
Table 4-2 Comparison of Line-Source Simulation and Rectangular Source Simulation (1)	104
Table 4-3 Comparison of Line-Source Simulation and Rectangular Source Simulation (2)	107
Table 5-1 Comparison of Different Approaches and Temperatures for the Two-Source Example in Figure 5.1	130
Table 5-2 Single Square Source T_{rise} Sensitivity to Source Length l (or source width w)	134
Table 5-3 Comparison of T_{rise} Predicted and T_{rise} Simulated (source depth $d = 0$, source thickness $t = 0.25\mu\text{m}$)	156
Table 5-4 Comparison of T_{rise} Predicted and T_{rise} Simulated (source depth $d = 0.1\mu\text{m}$, source thickness $t = 0.1\mu\text{m}$)	158
Table 5-5 Effect of Source Depth and Thickness on T_{rise} (1)	160
Table 5-6 Effect of Source Depth and Thickness on T_{rise} (2)	165
Table 5-7 Comparison of T_{rise} Before and After Quantization for the $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ Heat Source with $d = 0.1\mu\text{m}$ and $P = 40\text{mW}$	171
Table 5-8 Comparison of T_{rise} Before and After Quantization for the Two-Source Example	180
Table 6-1 Comparison of T_{rise} Before and After Quantization for the	

Four-Source Example in a Semi-Infinite Substrate	192
Table 6-2 Comparison of T_{rise} Before and After Quantization for the Four-Source Example in a Substrate with Given and Wide Lateral Dimensions	193

List of Figures

Figure 2.1 Numerical Network for a Two-Dimensional System.....	7
Figure 2.2 (a) Odd-shaped heat source over a rectangular substrate with adiabatic edges	10
(b) “Method of images” equivalent to (a).....	10
Figure 2.3 An Example of a Symmetrical Structure	12
Figure 2.4 A Quarter of the Symmetrical Structure in Figure 2.3	14
Figure 3.1 (a) Initial high density block structure.....	18
(b) Mesh structure created from the high density block structure after iterative amalgamation controlled by the gridding mask	18
Figure 3.2 Substrate Layer Construction.....	22
Figure 3.3 Surface Layer with Grid Refinement in the Heat Generation Region.....	24
Figure 3.4 Complete Mesh Outlines of the Model Shown in Figure 3.2	24
Figure 3.5 Example 1	
(a) Structure of a semiconductor device thermal model.....	26
(b) A quarter of the structure.....	26
Figure 3.6 Mesh Structures for Example 1	
(a) Well distributed mesh structure based on square blocks	27
(b) Mesh structure with thin but wide blocks (unexpected).....	27
Figure 3.7 Substrate Numerical Regions (top view)	34
Figure 3.8 Example 2	
(a) Structure of a semiconductor device thermal model.....	40
(b) A quarter of the structure.....	40
Figure 3.9 Substrate Numerical Regions (Front View) of Example 2	41
Figure 3.10 Meshes for Example 2 (1).....	42
Figure 3.11 Meshes for Example 2 (2).....	42
Figure 3.12 Meshes for Example 2 (3).....	46
Figure 3.13 Meshes for Example 2 (4).....	46
Figure 3.14 Meshes for Example 2 (5).....	47
Figure 3.15 Meshes for Example 2 (6).....	47

Figure 3.16 Meshes for the $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1).....	52
Figure 3.17 Meshes for the $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2).....	52
Figure 3.18 Meshes for the $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1).....	53
Figure 3.19 Meshes for the $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2).....	53
Figure 3.20 Meshes for the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1)	54
Figure 3.21 Meshes for the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2)	54
Figure 3.22 Meshes for the $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1).....	55
Figure 3.23 Meshes for the $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2).....	55
Figure 3.24 Multiple-Source Example 1	60
Figure 3.25 Meshes for Multiple-Source Example 1 (1).....	61
Figure 3.26 Meshes for Multiple-Source Example 1 (2).....	61
Figure 3.27 Meshes for Multiple-Source Example 1 (3).....	62
Figure 3.28 Multiple-Source Example 2.....	63
Figure 3.29 Meshes for Multiple-Source Example 2 (1).....	64
Figure 3.30 Meshes for Multiple-Source Example 2 (2).....	64
Figure 4.1 Heat Flow Lines and Isotherms in a Semi-Infinite Substrate.....	67
Figure 4.2 Example of Free Heat Spreading	69
Figure 4.3 Example of Restricted Heat Spreading.....	70
Figure 4.4 Thermal Resistance for Limited Heat Spreading.....	70
Figure 4.5 Simulation Model	73
Figure 4.6 The Effect of Power Dissipation P on the Function of T_{rise}/T_{inf} vs. $L = t_{sub}$	77
Figure 4.7 T_{rise} Contour as a Function of t_{sub} and L for a Quarter of Different Sources.....	79
Figure 4.8 3D Heat Flow in Substrate ($L \gg t_{sub}$, t_{sub} is not too small)	81
Figure 4.9 Restricted Heat Flow in Substrate ($L \gg t_{sub}$, t_{sub} is thin).....	81
Figure 4.10 Restricted Heat Flow in Substrate ($t_{sub} \gg L$).....	82
Figure 4.11 An Example of a Substrate With Small Thickness But Wide.....	83
Figure 4.12 An Example of a Semi-Infinite Substrate (Large L and t_{sub}).....	85
Figure 4.13 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	87
Figure 4.14 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	88

Figure 4.15 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	89
Figure 4.16 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	89
Figure 4.17 T_{rise} vs. $L = t_{sub}$ for Single Square Sources.....	91
Figure 4.18 T_{rise} / T_{inf} vs. $L = t_{sub}$ for Different Sizes of Single Square Sources	93
Figure 4.19 T_{rise} / T_{inf} vs. $L = t_{sub}$ for Single Rectangular Sources of Different Respect Ratios (1).....	95
Figure 4.20 T_{rise} / T_{inf} vs. $L = t_{sub}$ for Single Rectangular Sources of Different Aspect Ratios (2).....	96
Figure 4.21 Line Source Simulation	98
Figure 4.22 Line-Source Simulation Solution at 1% Reflection Error for Rectangular Sources with l of $40\mu\text{m}$ and Different Aspect Ratios	102
Figure 4.23 Line-Source Simulation Solution at 2% Reflection Error for Rectangular Sources with l of $40\mu\text{m}$ and Different Aspect Ratios	103
Figure 4.24 Line-Source Simulation Solution at 1% Reflection Error for Rectangular Sources with l of $64\mu\text{m}$ and Different Aspect Ratios	105
Figure 4.25 Line-Source Simulation Solution at 2% Reflection Error for Rectangular Sources with l of $64\mu\text{m}$ and Different Aspect Ratios	106
Figure 4.26 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	108
Figure 4.27 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	109
Figure 4.28 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	109
Figure 4.29 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source	110
Figure 4.30 Surface Plot of Temperature Distributions for a Nine-Source Example as shown in Figure 4.33	113
Figure 4.31 A Three-Source Example.....	115
Figure 4.32 A Seven-Source Example	116

Figure 4.33 A Nine-Source Example	117
Figure 4.34 Multiple Source Simulations	118
Figure 5.1 Top View of a Two-Source Example in a Substrate with Initial Minimum Substrate Blocks of $1\mu\text{m}$ on Each Side	124
Figure 5.2 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.05\mu\text{m}$ on Each Side for the Two-Source Example in Figure 5.1 before Quantization.....	126
Figure 5.3 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the Two-Source Example in Figure 5.1 before Quantization.....	128
Figure 5.4 A Quarter Mesh Structure with Initial Minimum Substrate Block of $1\mu\text{m}$ on Each Side for the Two-Source Example in Figure 5.1 after Quantization	131
Figure 5.5 Single Square Source Quantization.....	139
Figure 5.6 Single Rectangular Source Quantization of Length l	141
Figure 5.7 Single Rectangular Source Quantization of Width w	141
Figure 5.8 Straight Line Fit for the Functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of l	149
Figure 5.9 Straight Line Fit for the Functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of w	149
Figure 5.10 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with Length $l = 20\mu\text{m}$ and Different Aspect Ratios.....	150
Figure 5.11 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with Length $l = 20\mu\text{m}$ and Different Aspect Ratios.....	150
Figure 5.12 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with Different Width and Aspect Ratios	153
Figure 5.13 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with Different Width and Aspect Ratios	153
Figure 5.14 Universal Functions of T_{rise} Sensitivity to l and w vs. Aspect Ratio with $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$	154
Figure 5.15 Effect of Source Depth and Thickness on T_{rise} Sensitivity to Source Length l	159
Figure 5.16 Effect of Source Depth and Thickness on T_{rise} Sensitivity to Source Width w	159
Figure 5.17 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.1\mu\text{m}$	163

Figure 5.18 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.1\mu\text{m}$	163
Figure 5.19 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.5\mu\text{m}$	164
Figure 5.20 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.5\mu\text{m}$	164
Figure 5.21 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ Source Example with Depth $d = 0.1\mu\text{m}$ before Quantization.....	168
Figure 5.22 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $1.2\mu\text{m}$ on Each Side for the $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ Source Example with Depth $d = 0.0\mu\text{m}$ after Quantization	170
Figure 5.23 Multiple Source Temperature Calculation	173
Figure 5.24 A Two-Source Example Before Quantization.....	175
Figure 5.25 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the Two-Source Example before Quantization	176
Figure 5.26 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $1.2\mu\text{m}$ on Each Side for the Two-Source Example after Quantization.....	179
Figure 6.1 (a) Top view of a four-source example.....	184
(b) A quarter structure before quantization for Atar simulation	184
(c) A quarter structure after quantization for Atar simulation.....	184
Figure 6.2 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the Four-Source Example before Quantization	185
Figure 6.3 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $1.6\mu\text{m}$ on Each Side for the Four-Source Example after Quantization	191

List of Abbreviations and Symbols

FDM	Finite Difference Numerical method
$Atar$	a 3D numerical thermal simulator
t_{source}	an analytical thermal simulator
T_{rise}	temperature rise when substrate is constricted
T_{inf}	temperature rise when substrate is semi-infinite
T_{wide}	temperature rise in a substrate with finite thickness and infinite lateral dimensions
l	source length
w	source width
l/w	aspect ratio of a heat source
d	source depth from substrate surface
t	heat source thickness
l_i	gridding mask length
w_i	gridding mask width
d_i	gridding mask depth
L	distance between source edges and substrate lateral boundaries
t_{sub}	substrate thickness
k	thermal conductivity
A	conduction area normal to heat flow
d_{sub}	distance between the cross-sectional area A and the substrate surface
P	power dissipation

s_i	the maximum block size of each numerical region in a substrate
s	initial minimum substrate block size
n_i	number of blocks along the substrate thickness direction in each numerical region
θ	thermal resistance
q	heat generated per unit volume
ρ	density of material
C_θ	thermal capacitance
C_T	specific heat of material
V	volume of material

Chapter 1: Introduction

1.1 Research Motivations

Due to the increasing of signal speeds and integration densities, thermal problems of heat generation from integrated devices become more and more important. Both the performance and reliability of an electronic circuit are strongly influenced by temperature. Exposure to temperatures beyond which a circuit is designed to withstand, may result in failure of the circuit to perform according to specifications, or in failure altogether. Researchers have reported that for every 10°C rise in temperature, the reliability of a circuit is decreased by approximately half [1]. In order to maximize circuit performance and reliability, it is important to predict the rise in temperature of a device due to its own internal power dissipation, in order that electronic circuits can be designed and packed with minimum heat generation by employing adequate thermal management methods.

For electronic devices, in most cases, the data of interest are the temperatures at the power dissipating elements, i.e. device junctions. The thermal characteristics of devices mostly depend on their junction operating temperature which affects device performance and reliability. Today, the need for more accurate junction temperature prediction is increasing[14].

Therefore, the main target of this research is to find the feasible approximate approaches for quickly and yet accurately simulating the device junction temperature rise.

1.2 Research Objectives

In thermal analyses of microelectronic devices, there are mainly two general approaches: analytical and numerical. At present, the Finite Difference Numerical method (FDM) is widely used in solving the partial differential equations of three-dimensional heat transfer in semiconductor devices[18]. The use of FDM is due to its flexibility in solving complex geometries and boundary problems that can not be handled by exact analysis techniques. However, in the numerical method, there is always a trade off between solution accuracy, number of grids, simulation size and calculation time. Unfortunately, there has been little comparative research on numerical gridding, simulation size and solution accuracy in device thermal modeling.

The objectives of this research include:

- 1) Studies of numerical grid generation using Atar - a 3D thermal simulator for integrated devices. Rules and procedures in setting up parameters to create square-block based numerical meshes are discussed. Based on the idea of grid independence, the rules of setting up gridding mask dimensions are also investigated and approximation expressions are created in order to quickly and accurately simulate the junction temperature.

- 2) Effects of a finite substrate on thermal analysis are studied to suggest optimal substrate boundary dimensions in the computational domain for a given heat source structure, in order to minimize simulation size and save simulation time with retention of solution accuracy.

3) Quantization of heat source dimensions and depth are also studied to simplify numerical grid generation in order to quickly and accurately predict junction temperature.

1.3 Thesis Organization

The thesis is organized in the following manner:

Chapter 2 briefly outlines the important background information on thermal analysis methods, and leads into a detailed discussion of numerical method analysis, including numerical method approximation, grid refinement in the numerical method, boundary conditions and simulation size.

Chapter 3 discusses the rules and procedures for creating square-block based numerical meshes using Atar. In addition, the rules for setting up gridding mask dimensions are also investigated, and approximation expressions are presented and verified by showing different models.

In Chapter 4, effects of a finite substrate on thermal analysis are studied and simulation results, which were based on the numerical method using Atar, are compared with the heat spreading concepts. Using the line-source simulation solutions and bounding box ideas, some approximate approaches to quickly and accurately estimate substrate boundary dimensions in the computational domain are established for different heat source and substrate geometries.

In chapter 5, relations between temperature sensitivities and the aspect ratios of heat source dimensions are studied and approximate quantization solutions are provided.

In chapter 6, all the rules for quantizing source dimensions and depth, optimizing substrate boundary dimensions, and generating square-block based numerical meshes are verified via a final example. The efficiency is demonstrated.

Chapter 7 is the conclusion of this research, and some potential topics for future studies.

Chapter 2: Background

2.1 Thermal Analysis Methods

In semiconductor device thermal modeling, the main challenge is to create an accurate model of the device from its physical situations, such that, the maximum temperature rise in the device junction can be quickly and accurately predicted. Therefore, an appropriate thermal analysis method needs to be used in order to precisely represent the device's physical conditions.

2.1.1 Comparison of Numerical Method with Analytical Method

In thermal analyses of microelectronic devices, there are mainly two general approaches: analytical and numerical. The analytical approach often focuses on the derivation of an exact solution for the problems by means of Fourier Series, which provides a theoretically exact result in the form of an infinite series. Methods based on infinite series techniques require the summation of a large number of terms with coefficients that depend on boundary conditions, and lead to lengthy expressions. The relationship between the expressions and the geometry of a device is difficult to establish. In addition, the solution is not always derivable in the case of complex device geometries and boundary conditions.

Numerical methods, on the other hand, are useful for solving problems that cannot be handled by exact analysis techniques. At present, the Finite Difference Method (FDM) is widely used for the solution of partial differential equations of three-dimensional heat transfer in semiconductor devices.

In this work, a 3D thermal simulator, “Atar”, which is based on a finite difference numerical method, and the “tsource” program, which is based on an analytical method, are used alternatively, and their simulation results are compared.

2.1.2 Numerical Method Approximation

When the partial differential heat transfer equation is solved analytically over a given region with specified boundary conditions, the solution satisfies the differential equation at every point in the region. However, when the finite-difference approach is used, the problem is discretized so that only a finite number of nodal points are considered instead of every point over the region. If N nodes are chosen, N algebraic equations are developed by discretizing the differential equations and the boundary conditions for the problem.

Therefore, the first step in the numerical approach is to subdivide the given system into a suitable number of small volumes and assign a reference number, called a node, to the central point in each of these subvolumes. The temperature of a subvolume is often assumed to be that at its center. Figure 2.1 shows a numerical network for a two-dimensional system[2]. These subvolumes, as shown in Figure 2.1(a), are then replaced by a nodal network as illustrated in Figure 2.1(b). In this nodal network, each node is connected to at least one other node by a line. The network of lines that connect the nodes is called a grid or mesh. The approximate numerical approach based on this simplification has great advantages over the exact method because it effectively reduces a multi-dimensional system into a number of elementary one-dimensional systems. Using this approximate numerical approach, the required boundary conditions can be more easily satisfied.

The fundamental heat balance is reverted in finite-difference form, and the temperature is computed at only pre-selected points.

Figure 2.1 Numerical Network for a Two-Dimensional System
(a) Subdivisions of the system (b) Nodal network of the system

Referring to Figure 2.1, select a square subdivision $\delta \times \delta \times 1$ so that for a nodal point such as $\textcircled{1}$, $q_{21} + q_{31} + q_{41} + q_{51} = 0$, where

$$q_{ij} = k_{ij} A \frac{dT}{dx} = k_{ij} (\delta \cdot 1) \frac{\Delta T}{\delta} = k_{ij} (t_i - t_j), \text{ and } k_{ij} \text{ is the internal thermal conductance}$$

between each node[2], which is defined as $k_{ij} = k(A/L)$ (A is the conduction area normal to heat flow, L is the heat flow path length, and k is the thermal conductivity).

A square subdivision has the advantage that each internal conductance is again independent of network size, since $k_{21} = k(\delta \times 1)/\delta = k = k_{31} = k_{41} = k_{51}$. Therefore, $\overline{q_1} = t_2 + t_3 + t_4 + t_5 - 4t_1 = 0$.

Although the numerical network for a two-dimensional system does not represent heat flow in the third dimension, the character of heat flow through each node in a 3D sys-

tem is similar to that in a two-dimensional system. The above equation is written for each set of n nodal points in the numerical network. The result, which represents the fundamental development of the numerical method, is a set of n linear algebraic equations in n unknown temperatures. It needs to be pointed out that the temperature varies through each subvolume with a gradient dT/dx , and the approximation can only be improved if Δx is chosen smaller, such that $\Delta T / \Delta x$ approaches dT/dx .

When Δx is chosen smaller, n becomes large (corresponding to a dense network), making such methods tedious. Therefore, the grid refinement technology needs to be used in numerical methods.

2.1.3 Grid Refinement in Numerical Method

As pointed out in section 2.1.2, in order to ensure solution accuracy using the numerical method approximation, Δx needs to be chosen small. However, when Δx is chosen smaller, n is larger (corresponding to a denser network), meaning that more simulation time is required. Therefore, in the numerical method approach, the ultimate goal is to create a mesh structure that resolves all the relevant details of the model, without significantly impacting computational time or accuracy. Grid refinement is a modeling technique that adds grids in regions that are of interest while using sparse grids in areas of less interest. By creating localized regions of dense grids around relevant elements in the model, the computational time is decreased while accuracy is maintained.

As will be shown in chapter 3, Atar has the advantage of easily generating meshes in such a way that a fine mesh is created around heat generation regions which are of the

most interest, while only a moderate number of grids are used for the entire device. Chapter 3 describes Atar's rules of grid generation and parameter setting. In addition, the grid independence point at which additional grids have little impact on solutions, is also investigated.

2.2 Boundary Conditions and Simulation Size

In device thermal modeling, conduction equations, together with initial and boundary conditions, complete the differential formulation[17][20]. As shown in section 4.1.1, in general, the boundary conditions of a semi-infinite substrate are often modeled with adiabatic top and sides, and an ideal constant backside boundary condition at the bottom. In the steady state, these boundary conditions will play important roles.

Usually, symmetry boundary conditions are used to simplify physical models. These boundary conditions can be applied when the geometry has mirror symmetry. At symmetry boundaries, heat flow across the symmetry plane is considered to be "0"[15][16]. Thus, for single rectangular or square heat sources, because of the symmetry of structures, the entire model could be considered as only one quarter by imposing adiabatic conditions (i.e., zero heat flow) at the central faces of $x = 0$ and $y = 0$. For problems with axisymmetric geometry, the symmetry boundary conditions can be applied at the center line of the geometry.

This method of using symmetry boundary conditions is also called the "image method" and was clearly presented by A. L. Palisoc and C. C. Lee[3]. In their research, the image technique was realized by placing a "mirror" source in a position of symmetry

relative to the adiabatic boundary. This symmetry of position causes heat flow to cancel exactly along the desired adiabatic plane, and results in zero heat flow across the adiabatic boundaries.

As illustrated in Figure 2.2[3], a heat source as shown with shading is mounted in a semi-infinite substrate with adiabatic top and sides, and an ideal 300K boundary condition at the bottom. The physical situation depicted in Figure 2.2(a) is equivalent to that shown in Figure 2.2(b).

Figure 2.2 (a) Odd-shaped heat source over a rectangular substrate with adiabatic edges
(b) "Method of images" equivalent to (a)

In Figure 2.2(b), the image structure is constructed one by one, by implementing the boundary conditions at $x = 0$, $y = 0$ and $x = a$, $y = b$. Image ① is introduced to implement boundary conditions at $x = 0$. Image ② is introduced to implement boundary conditions at $x = a$. Image ③ is introduced to implement boundary conditions at $y = 0$, and image ④ is introduced to implement boundary conditions at $y = b$. Never-

theless, the introduction of images ① and ② needs to further introduce images ⑤ and ⑥ in order to cancel the heat flow across the $y = 0$ boundary. In addition, the introduction of images ⑤ and ⑥ cancels the heat flow across the $x = 0$ and $x = a$ boundaries caused by image ③. Similarly, images ⑦ and ⑧ are introduced to cancel the heat flow across the $y = b$, $x = 0$ and $x = a$ boundaries caused by images ①, ② and ④. This process continues to infinity so that the entire 2D plane is tiled with sources. The solution of temperature over the device is obtained from the superposition of the contribution due to the actual and image sources.

Figure 2.3 shows a two-source example. The sources are mounted in a $192\mu\text{m} \times 160\mu\text{m} \times 75\mu\text{m}$ substrate symmetrically. Indium Phosphide, which has a thermal conductivity of 0.44mW/cmK , is used as the substrate material and will be used for the rest of the examples in this thesis. Each source has a size of $10\mu\text{m} \times 2\mu\text{m} \times 0.25\mu\text{m}$ with depth $d = 0$. The power dissipation of each source is 20mW . Figure 2.3(a) shows the top view of the structure. Simulation was done using the 3D thermal simulator, Atar. Figure 2.3(b) shows the temperature contours on the substrate surface (top view) using the Atar simulator. Figure 2.3(c) shows more detailed information about the temperature contours on the substrate surface by loading the temperature data from the Atar simulator into Matlab and then plotting from Matlab.

Figure 2.3 An Example of a Symmetrical Structure

Both Figures 2.3(b) and 2.3(c) demonstrate that, in this example, because of their symmetrical structure, the temperature contours are symmetrical to the central lines $x = 0$, $y = 0$. This indicates that, for a symmetrical structure, heat flow across its central lines is “0”. Thus, thermal simulation could be simplified by setting the adiabatic conditions to the faces $x = 0$ and $y = 0$, and simulating only one quarter of the entire structure, as shown in Figure 2.4(a).

Figure 2.4(a) shows an equivalent structure of the two-source example in Figure 2.3(a). Because of its symmetrical structure, this two-source example could be considered as a $5\mu\text{m} \times 2\mu\text{m} \times 0.25\mu\text{m}$ heat source with a power dissipation of $P = 10\text{mW}$ (as shown with black shading) mounted into a $96\mu\text{m} \times 80\mu\text{m} \times 75\mu\text{m}$ substrate, a quarter of that in Figure 2.3(a). The source areas enclosed by grey solid lines can be considered as image sources of the $5\mu\text{m} \times 2\mu\text{m} \times 0.25\mu\text{m}$ original source. In Figure 2.4(a), the symmetrical positions of the original source and image source cause zero heat flow across the adiabatic planes $x = 0$ and $y = 0$.

Thus, instead of being simulated with two $10\mu\text{m} \times 2\mu\text{m} \times 0.25\mu\text{m}$ symmetrical sources and adiabatic boundaries as represented by the grey dash lines in Figure 2.4(a), by setting the adiabatic conditions to the faces $x = 0$ and $y = 0$, the entire structure can be simulated at a quarter of the size with only one $5\mu\text{m} \times 2\mu\text{m} \times 0.25\mu\text{m}$ source, as shown in Figure 2.4(a) with shading..

Figure 2.4(b) shows the temperature contours for a quarter of the structure as shown by the shading in Figure 2.4(a). Comparing the contour plot in Figure 2.4(b) with

that in Figure 2.3(c), it could be seen that the temperature contour in figure 2.4(b) is exactly a quarter of that in Figure 2.3(c). The maximum temperature rise found in both cases is the same at 46.31 K. However, using the quarter simulation, CPU time is reduced by 80%. Table 2-1 lists the comparison of thermal simulation results based on both the structure in Figure 2.3(a) and Figure 2.4(a). In each case, the mesh structure was kept the same, and the simulation was done using Atar.

Figure 2.4 A Quarter of the Symmetrical Structure in Figure 2.3

Table 2-1 Comparisons of Thermal Simulation Results Based on the Entire Symmetrical Structure and one Quarter

simulation size	entire structure as in Figure 2.3(a)	a quarter as in Figure 2.4(a)
maximum temperature rise (K)	46.31	46.31
number of numerical nodes (grids)	6816	1704
simulation CPU time (s)	0.5	0.1

Therefore, according to the above discussion, for a single rectangular or square heat source, only one quarter needs to be considered by imposing adiabatic conditions at the faces $x = 0$ and $y = 0$. Meanwhile, for problems with axisymmetric geometry, the symmetry boundary conditions can be applied at the center line of the geometry and consider only half of the entire structure.

Using the image method, a large reduction in CPU time is achieved, especially for small sources over large substrates. Therefore, the image method is used throughout this research. In chapter 4, the effect of a finite substrate and boundary dimensions on device thermal analyses is discussed in detail.

Chapter 3: Grid Generation Using Atar

As discussed in chapter 2, the accuracy of a numerical solution mainly depends on the gridding structure. In this research, new rules were developed for the Atar numerical thermal simulator to control the block agglomeration transitions, in order to avoid non-physical meshing errors. In addition, approximation equations for determining the dimensions of each numerical region were also explored. In this chapter, all these rules and approximation equations are discussed in detail.

3.1 Atar Numerical Thermal Simulator[4]

In order to quickly and accurately predict the device junction temperature using the numerical method, an appropriate numerical mesh structure needs to be created so as to precisely model the physical situation of the device without a significant impact on computational time. The Atar thermal simulator produces a numerical solution for the conductive heat flow in electronic devices. It can be used to easily generate an accurate 3D physical model of the device to be simulated from layout information. The simulator produces an appropriate mesh of the device based on a square block structure. The mesh is automatically created in such a way that numerous fine grids are produced around the heat generation regions which are close to the substrate surface, but only a moderate number of blocks is used for the entire device.

With Atar, the substrate is modeled by different numerical regions using various gridding masks. These gridding masks are used only to generate the numerical meshes for

the substrate, in order to create smaller square blocks around the heat generation region which is close to the substrate surface, while bigger square blocks are generated in the area that is far away from the heat source.

When a numerical mesh is generated using Atar, tiny uniform square blocks are first created throughout the substrate as shown in Figure 3.1(a). These tiny uniform square blocks are called initial minimum substrate blocks. Their sizes are defined in the technology file and the input file, as will be discussed in the following sections.

Once the initial minimum substrate block size is determined, gridding masks are used to divide the substrate into different numerical regions in the computational domain. In each of these regions, numerical grids are created by the iterative amalgamation of blocks from groups of four into single blocks. This amalgamation starts from the initial minimum substrate blocks, and stops when all blocks have reached their maximum size as defined by the gridding masks. In addition, this amalgamation occurs in all the x, y and z directions, where each block can have a maximum of two smaller blocks along each edge, four at the top and four at the bottom. Thus, each block is made of eight smaller blocks. Figure 3.1(b) shows the outlines of the mesh structure after amalgamation using a gridding mask in a substrate.

To illustrate, in Figure 3.1(a), the size of the tiny uniform square block was set to be $2\mu\text{m}$ on each side, while in Figure 3.1(b), the block size of the gridding mask was set to be $4\mu\text{m}$ on each side, and the maximum block size outside the gridding mask in the substrate was set to be $8\mu\text{m}$ on each side. Therefore, the gridding mask in Figure 3.1(b) divides the substrate into two different numerical regions: a region consisting of smaller

grids as defined by the gridding mask, and a region of bigger grids as defined by the maximum substrate block size.

Figure 3.1 (a) Initial high density block structure (b) Mesh structure created from the high density block structure after iterative amalgamation controlled by the gridding mask

In the region defined by the gridding mask, as shown in Figure 3.1(b), numerical grids are created by the iterative amalgamation of blocks from groups of four into single blocks. This amalgamation starts from the initial minimum substrate blocks ($2\mu\text{m}$), and stops when all blocks have reached their maximum size ($4\mu\text{m}$) as defined by the gridding masks. Thus, each block within that region is made of eight smaller blocks as shown with shading in Figures 3.1(b) and 3.1(a), respectively. In the region of the substrate outside the gridding mask, blocks are created by further amalgamating the blocks of $4\mu\text{m}$ on each side into blocks of $8\mu\text{m}$ on each side.

In device thermal modeling, according to the heat source geometry and substrate dimensions, several gridding masks can be used to divide the substrate into several different numerical regions in the computational domain, such that fine grids are created around the heat generation region and bigger grids are created in the regions that are far away from the heat source. The dimensions of each numerical region are defined by the lateral dimensions and depth of each of the corresponding gridding masks. The maximum block size within each numerical region is defined by the block size of the corresponding gridding mask. In the region that has more than one gridding masks superposed with another, the maximum block size within that region is chosen to be the smallest one as defined by each gridding mask. Rules for setting up gridding mask dimensions and block sizes are discussed in the following sections.

3.2 Model Building Using Atar

The basic philosophy behind model building in Atar is based on the creation of a complex mesh geometry using square blocks of varying sizes. The model is created in such a way that a block and its neighbors form a relatively simple set of topologies. Each block can have a maximum of two blocks along each edge (i.e. a bigger block is built up from 8 equal-sized smaller blocks). From these block topologies, a network of thermal resistances and capacitances can be extracted which represent the thermal equation to be solved. As will be shown, complex meshes can be built using these rules.

3.2.1 Mesh Construction Mechanism

In Atar, the model is constructed layer by layer, starting from the substrate surface down into the substrate. As moving from the substrate surface down into the substrate, layers are created by the amalgamation of the initial high density uniform layers (blocks), as shown in Figure 3.2(a). The thickness of each layer depends on the amalgamation of the initial layers, which is controlled by the gridding mask information. Figure 3.2 illustrates a simple example of the creation of the model. The surface layer is created using a very high uniform density of blocks [see Figure 3.2(a)]. The gridding mask information is then used to allow for the iterative amalgamation of blocks from groups of four into single blocks. This amalgamation is controlled by the position of the gridding mask edges and the maximum block size allowed inside and outside each gridding mask. During the iteration, groups of four blocks are visited and subject to a number of tests to see whether amalgamation is allowed:

- (1) The blocks are checked to see whether all four initial blocks are of the same size and material.
- (2) The final size of the new block to be created is checked against the maximum block size allowed inside any gridding masks covering that area.
- (3) Neighboring block sizes are checked to see whether there will be at most two blocks along any one edge of the new amalgamated block.

During the amalgamation of the blocks, the block topography is restricted such that the maximum number of blocks along any edge of a block is two. This topology restriction produces a gradually controlled gradient in block size while moving away from

the heat generation regions. The amalgamation of blocks stops when all blocks have reached their maximum size as defined by the gridding masks.

Using the gridding mask information, the surface layer in the model is created in such a way that a very high density of blocks is produced near the areas that will define the heat generation region [see Figure 3.2(b)]. As moving away from the heat generation region, the size of blocks increases by further amalgamating groups of four blocks into a single block until reaching its maximum size as defined by the gridding masks. The amalgamation is controlled by the gridding mask geometry, the block position, and the restriction that there can be no more than four blocks above or below a single block. The result of these requirements is that all blocks have a maximum of two blocks along each edge and four at the top or bottom, producing a gradually controlled gradient in block size while moving away from the heat generation region. A significantly restricted set of block topologies is thus allowed, along with the generation of relatively large blocks at the bottom of the substrate where thermal gradients are small.

As the substrate layers are created, the number of blocks in each layer is reduced and the maximum block size inside each layer is increased by further amalgamating groups of four blocks into a single block. Meanwhile, the thickness of substrate layers is generally increased while moving down into the substrate by the amalgamation from the high density uniform layer as shown in Figure 3.2(a). Figures 3.2(c) and 3.2(d) show two characteristic substrate layers created during the formation of the substrate. Figure 3.2 clearly shows the generation of the mesh structure and the increase in the thickness of substrate layers from the surface layer.

Figure 3.2 Substrate Layer Construction. (a) Initial high density uniform square block structure. (b) Surface layer created from this high density structure after iterative amalgamation. (c) Substrate layer near the surface of the substrate showing more grid amalgamation. (d) Substrate layer near the bottom of the substrate showing increased thickness and further amalgamation

The heat generation region in an integrated device is generally fairly close to the substrate surface, and has large heat flow and temperature gradients. Therefore, in order to accurately represent the physical situation within the heat generation region, the use of even smaller numerical grids is required. In order to accurately represent the heat flow in

the heat generation region, Atar models the heat generation region by further subdividing the initial minimum substrate block (i.e. the initial high density uniform blocks, as shown in Figures 3.1(a) and 3.2(a)) into tiny uniform square blocks, instead of decreasing the initial minimum substrate block size. This subdividing will not affect the block amalgamation in the substrate. Thus, the solution accuracy can be ensured without having a great impact on the simulation time. This subdividing can be performed using the refine statement in the input file, which will be discussed in detail in the following sections. Figure 3.3 shows the same surface layer as in Figure 3.2(b) after the refinement in the heat generation region. Figure 3.4 shows the complete mesh outlines of the example as shown in Figure 3.2.

In Figure 3.4, a square-block based numerical mesh is illustrated where a number of small square blocks are created around the heat generation region, while larger square blocks are generated in the regions that are far away from the heat source. The blocks inside the heat generation region are created by further subdividing the initial minimum substrate blocks in such a way that tiny uniform square blocks are generated inside the heat generation region in order to accurately represent the heat source in three dimensions. Rules regarding setting the parameters for gridding masks and for grid refinement in the heat generation region, will be discussed in the following sections.

Figure 3.3 Surface Layer with Grid Refinement in the Heat Generation Region

Figure 3.4 Complete Mesh Outlines of the Model Shown in Figure 3.2

3.2.2 Rules in Setting Up Parameters

As discussed in sections 2.1.2 and 2.1.3, the accuracy of a numerical solution depends mainly on the gridding structure. In Atar, the block amalgamation is controlled by the position of gridding mask edges, the maximum block size allowed inside and outside each gridding masks, and the requirement that above or below a single block there can be no more than four blocks. Therefore, a well distributed square-block based mesh structure is critically dependent on the substrate gridding. The substrate gridding defines the initial minimum block size in the substrate, the dimensions and maximum block size of each gridding mask, and the maximum block size allowed in the substrate. It is useful to investigate the rules for setting up these parameters and creating proper mesh structures to precisely represent the physical model.

Figure 3.5(a) shows an example of a semiconductor device thermal model. A $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ heat source is mounted into the center of a substrate with depth $d = 0$. Substrate length and width are $400\mu\text{m}$ respectively, and substrate thickness is $50\mu\text{m}$. Power dissipation is $P = 40\text{mW}$. Base temperature is assumed to be at room temperature, 300K . Using the image method discussed in section 2.3, thermal simulation was done for a quarter as shown in Figure 3.5(b) with $\frac{1}{4}P = 10\text{mW}$. Figure 3.6 shows different mesh structures created by Atar for example 1. The solutions for the maximum temperature rise in the device junction are different. Table 3-1 lists some of the variable values used in each case, and the different temperature results are compared with that obtained from “tsource” simulation, which is based on the analytical method.

Figure 3.5 Example 1 (a) Structure of a semiconductor device thermal model
(b) A quarter of the structure

In both Figures 3.6(a) and 3.6(b), the maximum substrate block size was set to be $10\mu\text{m}$ on each side. In Figure 3.6(a), the initial minimum substrate block size was $1\mu\text{m}$ on each side, and three different numerical regions were defined in the substrate around the thermal source as power region, device region, and circuit region using the corresponding gridding masks. Dimensions and maximum block size within each of these regions are listed in Table 3-1. The minimum block size in the heat generation region was $0.25\mu\text{m}$ by further subdividing the initial minimum substrate block by 2^2 . However, in Figure 3.6(b), the initial minimum substrate block size was set to be $0.5\mu\text{m} \times 2\mu\text{m} \times 1.4\mu\text{m}$, and two different numerical regions were defined in the substrate around the thermal source as power region and device region. Dimensions and maximum block size within each of these regions are also listed in Table 3-1. In Figure 3.6(b), the minimum block size in the heat generation region was the same as the initial minimum substrate block, with no further refinement.

Figure 3.6 Mesh Structures for Example 1 (a) Well distributed mesh structure based on square blocks (b) Mesh structure with thin but wide blocks (unexpected)

Table 3-1 Comparison of Mesh Structures and Temperatures for Example 1

Atar			tsource
parameters	Figure 3.6(a)	Figure 3.6(b)	
initial minimum substrate block size (μm)	$1 \times 1 \times 1$	$0.5 \times 2 \times 1.4$	
maximum substrate block size (μm)	$10 \times 10 \times 10$	$10 \times 10 \times 10$	
subdivision in heat generation region (μm)	$0.25 \times 0.25 \times 0.25$	$0.5 \times 2 \times 1.4$	
circuit region	gridding mask x-y dimensions (μm)	20×20	-
	gridding mask depth (μm)	16	-
	maximum block size (μm)	4	-
device region	gridding mask x-y dimensions (μm)	10×10	12×12
	gridding mask depth (μm)	8	24
	maximum block size (μm)	2	8
power region	gridding mask x-y dimensions (μm)	8×8	8×8
	gridding mask depth (μm)	2	2
	maximum block size (μm)	0.5	0.5
number of blocks		5438	5038
temperature rise (K)		47.78	41.57
			47.86

Theoretically, a high density of grids (/nodes) means more temperature gradients and a large heat flow in three dimensions, resulting in a lower temperature rise in the junction area. However, according to the simulation results listed in Table 3-1, the temperature is decreased in case (b) which has a less dense mesh structure of 5038 numerical blocks, compared with case (a) of 5438 substrate blocks. Comparing the temperature in case (a) and case (b) with that simulated by the “tsource” program, the error in the temperature rise is increased by 13% in (b) compared to (a).

This abnormal phenomenon is caused by meshing errors of the thin but wide blocks near the heat generation region in substrate thickness direction, as shown in Figure 3.6(b). These thin but wide blocks are the result of improper parameter settings for the gridding mask (e.g. dimensions and block size), and the minimum and maximum block size in the substrate. For example, as listed in Table 3-1 case (b), the initial minimum substrate block size was set to be $0.5\mu\text{m} \times 2\mu\text{m} \times 1.4\mu\text{m}$. Nevertheless, according to the mesh construction mechanism discussed in section 3.2.1, the square-block based mesh structure is generated by the amalgamation of the initial high density square blocks. Therefore, those initial rectangular blocks in case (b) cause difficulties for the generation of square blocks by amalgamation of the initial rectangular blocks. Specifically, the thickness of the initial rectangular block in figure 3.6(b) was set to be $1.4\mu\text{m}$, which is neither an even multiple of the lateral dimensions of $0.5\mu\text{m}$ and $2\mu\text{m}$, nor an even fraction of $0.5\mu\text{m}$ and $2\mu\text{m}$. Therefore, it is impossible to generate square blocks along the substrate thickness direction by amalgamation of those initial rectangular blocks.

In addition, there is no refinement in the heat generation region in case (b) (i.e., no subdivisions in the heat generation region). The block size in the heat generation region is the same as that of the initial minimum substrate block, which is $0.5\mu\text{m} \times 2\mu\text{m} \times 1.4\mu\text{m}$, while the thickness of the heat source is $0.25\mu\text{m}$. Thus, the $1.4\mu\text{m}$ thick numerical grid cannot accurately represent the physical situation of a heat source of $0.25\mu\text{m}$ thick.

Moreover, in case (b), 10 uniform layers with a thickness of $0.1\mu\text{m}$ for each were defined near the substrate surface in the technology file. According to the mesh construction mechanism discussed in section 3.2.1, substrate layers are generated by amalgamation

of the initial uniform blocks (i.e., initial minimum substrate blocks), which means that the thickness of substrate layers should be greater than or equal to the thickness of the initial minimum substrate block. However, in case (b), the minimum substrate block thickness is greater than the thickness of each of the first 10 uniform layers. This conflict between the thickness of layers and the initial minimum substrate blocks causes the generation of the thin but wide grids near the substrate surface along the thickness direction.

These thin but wide blocks cause a larger heat spreading in lateral directions in the computational domain. The reason for this, as described in section 2.1.2, is that the temperature of a finite volume is assumed to be that at its center. The results is a larger cross-sectional area normal to the heat flow and a lower junction temperature rise in the computational domain (more discussions in chapter 4), which explains why the temperature in case (b) is lower than that in case (a). However, as shown in chapter 4, it is often assumed that heat spreads in lateral directions with a maximum constant angle of $\alpha \approx 45^\circ$ [1]. Thus, a square-block based mesh needs to be created to represent the 3D heat flow in semiconductor devices. More detailed discussions about heat spreading and temperatures are shown in chapter 4.

Therefore, in order to accurately calculate the junction temperature rise, these thin but wide blocks need to be eliminated when generating mesh structures. This may be done by properly setting up the parameters mentioned above in Atar. Following are the rules that should be obeyed when setting up these parameters.

Minimum Substrate Block Size vs. “Substrate Gridding”, “Substrate Thickness” and “Number of Substrate Layers”

In the technology file, there are three parameters related to the initial minimum substrate block size, and these need to be properly set up. These parameters are “substrate thickness”, “number of substrate layers” (initial uniform layers) and “substrate gridding”. Usually, substrate thickness is given (or is chosen, as will be discussed in chapter 4). Thus, the substrate thickness divided by the number of substrate layers (initial layers) gives the initial minimum substrate block size in the thickness direction. Since the mesh structure is based on a square block topology, this block size should equal the “substrate gridding” parameter that gives the initial minimum substrate block size in x-y directions.

For example, if the substrate thickness is 50 μ m and the “substrate gridding” (initial minimum substrate block size) is set to be 1 μ m on each side, the number of initial substrate layers needs to be set to 50 in order to create square blocks on each side.

“Substrate Gridding” vs. Substrate Lateral Dimensions and “Rows” & “Columns”

In the input file, the “ x_{min} ” and “ x_{max} ” parameters are used to define the substrate length dimension, while “ y_{min} ” and “ y_{max} ” parameters are used to define the substrate width dimension. Along with the “rows” and “cols” parameters in the input file, all these parameters give the initial minimum substrate block size in the x-y direction which should be the same as the “substrate gridding” (initial minimum substrate block size) in the x-y direction that is defined in the technology file.

For instance, using example 1 as shown in Figure 3.5(b), one could set $x_{min}=0$, $x_{max}=200$, $y_{min}=0$, $y_{max}=200$. If the “substrate gridding” (initial minimum substrate block) is set up to be “1 1” in the technology file, which gives the initial minimum sub-

strate block size in x-y direction as $1\mu\text{m}$ each, then “cols” and “rows” should be set to 200, respectively. Their relations could be expressed as follows:

$$\frac{x_{max} - x_{min}}{cols} = \text{initial minimum substrate block size in x direction} \quad (3.1)$$

$$\frac{y_{max} - y_{min}}{rows} = \text{initial minimum substrate block size in y direction} \quad (3.2)$$

Refine Statement in the Input File

The refine statement in the input file is used to define the dimensions of the heat generation region, and subdivide the numerical grids within it from the initial high density uniform square blocks (i.e., the initial minimum substrate blocks). The heat generation region is the one that is most closed to the heat source. The “depth” parameter in the refine statement is one of the most important parameters. It determines how many times the initial block is subdivided. The “depth” parameter should be set up according to the values of source thickness and the initial minimum substrate block size. Let the “depth” in the input file be n , which means that it subdivides the initial minimum substrate block by 2^n . The relations among “depth” n , source thickness and the initial minimum substrate block size could be expressed as follows:

$$m \cdot \frac{\text{initial minimum substrate block size in thickness direction}}{2^n} = \text{source thickness} \quad (3.3)$$

where, m is a count number and $m \geq 1$.

For example, if the source thickness is $0.25\mu\text{m}$, and the initial minimum substrate block is $1\mu\text{m}$ on each side, then n needs to be set to 2, which subdivides $1\mu\text{m}$ by 2^2 and gives $0.25\mu\text{m}$. This could ensure that the source thickness dimension can be represented by a subdivision in the heat generation region.

Maximum Block Size in Each Region and Substrate

In the substrate, different numerical regions are defined by using various gridding masks. Figure 3.7 shows the top view of a substrate. Usually, three different numerical regions are defined around the heat generation region. These are the power region, device region, and circuit region. In some cases, however, a four-region model is also commonly used, where four different numerical regions are defined: power region, device region, cell region, and circuit region. The number of numerical regions depends on the source geometry and substrate dimensions. In the case of a small heat source with a large substrate, more different numerical regions are chosen so that the maximum substrate block size can be increased. As it will be shown in chapter 4, at the points far away from the heat generation region, the temperature gradients are small. Thus, by the generation of relatively large blocks in a region that is far away from the heat source, calculation time can be saved while accuracy is maintained.

Referring to Figure 3.7, let " s_1 ", " s_2 ", " s_3 ", " s_4 " and " s_5 " be the maximum block size of the power region, device region, cell region, circuit region and substrate, respectively. Because of the requirement set for the amalgamation of blocks, on any side of a block, the maximum number of blocks is two. Moreover, because of the need that the

transitions of different block sizes should not be too sharp, the relationship between “ s_i ”.

“ s_2 ”, “ s_3 ”, “ s_4 ” and “ s_5 ” could be expressed as:

$$s_5 = 2 \cdot s_4 = 4 \cdot s_3 = 8 \cdot s_2 = 16 \cdot s_1 \quad (3.4)$$

Figure 3.7 Substrate Numerical Regions (top view)

Values of the maximum block size in each numerical region are controlled by the block size of the corresponding gridding masks and are set up in the technology file. Using more numerical regions can allow for a larger maximum substrate block size. When the maximum block size in a numerical region is smaller than the initial minimum substrate block size, the maximum block size in that region is allowed to be as big as the initial minimum substrate block size. In example 1, although the maximum block size in the power region is set to be $0.5\mu\text{m}$, it is found from the mesh structure that the actual maximum block in that region is $1\mu\text{m}$, as shown in Figure 3.6(a).

An exception is the block size in the heat generation region. It is merely controlled by the refine statement in the input file, which subdivides the initial high density uniform

blocks (i.e. the initial minimum substrate blocks) into even smaller blocks. Thus, there is no block amalgamation in creating the tiny blocks inside the heat generation region.

Gridding Mask Dimensions

The x-y dimensions of each gridding mask are defined in the geometry file, while, its dimension along the substrate thickness direction is defined by the “*depth*” parameter in the technology file. It is found that Atar can automatically generate square blocks in x-y directions by adjusting these gridding mask edges, but it can do nothing to adjust the gridding mask edges in the depth direction. Therefore, in order to successfully create square blocks in all directions, some rules need to be obeyed in defining gridding mask depth. These are listed in Table 3-2.

Table 3-2 Rules in Defining Gridding Mask Depths

regions	power	device	cell	circuit	substrate
max_block_size	s	$2s$	$4s$	$8s$	$16s$
depth of each region	$n_1 s$	$n_2 \cdot 2s$	$n_3 \cdot 4s$	$n_4 \cdot 8s$	

In Table 3-2, n_1, n_2, n_3, n_4 are number of blocks along the “depth” direction in each of these regions. The values of n_1, n_2, n_3, n_4 must be even numbers in order to allow an appropriate space for the next region to create completed blocks. The reason is that the maximum block size of the next region is always double the size of the current region. Moreover, the relationship between n_1, n_2, n_3 , and n_4 is $n_1 s < n_2 \cdot 2s < n_3 \cdot 4s < n_4 \cdot 8s$

For example, assume that the maximum block size in the power region is $1\mu\text{m}$. The maximum block sizes for the device region, cell region and circuit region could then be $2\mu\text{m}$, $4\mu\text{m}$, and $8\mu\text{m}$, respectively. Possible depths for each region are listed in Table 3-3. It should be noticed that the depth of the circuit region should be bigger than that of the cell region; and the depth of the cell region should be bigger than that of the device region, etc.

Table 3-3 Possible Depth for Each Numerical Region
(assume minimum substrate block size is $1\mu\text{m}$ on each side)

power	device	cell	circuit
2	4	8	16
4	8	16	32
6	12	24	48
8	16	32	64
10	20	40	80
12	24	48	96
14	28	56	...
16	32	64	
18	36	72	
20	40	80	
...	...	...	

3.2.3 Procedures in Setting Up Parameters

The above discussion shows the relationship between each parameter and some of the rules that need to be obeyed. However, the procedures in setting these parameters are also important, and worth discussing here. Following is a summary of the main steps.

Step 1. Determine the Initial Minimum Substrate Block Size

The first step in setting up these parameters is to decide the initial minimum substrate block size, on which all other parameters are based. Because in most cases both the heat source thickness and its depth from the substrate surface are very small, the substrate gridding should ensure that by subdividing the initial minimum substrate block by 2^n , the subdivision size can satisfy the requirements of both the source thickness and depth. In other words, the source thickness should be able to be represented by the product of the subdivision size and a count number m_1 , and the source depth should be able to be represented by the product of the subdivision size and a count number m_2 , where $m_1 \geq 1$ and $m_2 \geq 1$.

As an example, if in Example 1, shown in Figure 3.5, the source thickness is $0.2\mu\text{m}$ with depth $d = 0.1\mu\text{m}$, then instead of being set to $1\mu\text{m}$, the initial minimum substrate block size may need to be set to $0.8\mu\text{m}$. Thus, by subdividing the initial minimum substrate block by 2^3 , the subdivision becomes $0.1\mu\text{m}$ on each side. After subdividing, the source depth can be represented by one subdivision in the depth direction, and source thickness can be represented by two subdivisions in the thickness direction. Thus, the source geometry can be accurately represented by the numerical meshes.

Step 2. Set Number of Substrate Layers

Once the initial minimum substrate block size is determined, the number of initial uniform substrate layers could be calculated according to substrate thickness and the initial minimum substrate block size in the substrate thickness direction:

$$\frac{\text{substrate thickness}}{\text{initial minimum substrate block size}} = \text{number of initial uniform layers} \quad (3.5)$$

Step 3. Set Rows and Columns in Input File

As discussed in section 3.2.2, “rows” and “cols” in the input file depend on substrate dimensions and gridding. They can be set up using Equations (3.1) and (3.2).

Step 4. Set Maximum Block Size in Each Numerical Region and Substrate

Let “ s_1 ”, “ s_2 ”, “ s_3 ”, “ s_4 ” and “ s_5 ” be the maximum block size of the power region, device region, cell region, circuit region, and substrate, respectively. According to Equation (3.4) in section 3.2.2, that is, $s_5 = 2 \cdot s_4 = 4 \cdot s_3 = 8 \cdot s_2 = 16 \cdot s_1$, and the fact that the maximum block size of the power region, “ s_1 ”, is usually set to be the same as the initial minimum substrate block size, the maximum block size for each regions can be determined.

Step 5. Set Gridding Mask Dimensions

The last step is to determine the gridding mask dimensions, especially depth. In fact, this is one of the most important steps to ensure the generation of square block mesh structures as discussed in section 3.2.2. Since Atar can automatically generate square blocks in the x-y direction by adjusting mask edges, it is straightforward to set up x-y dimensions for each of these gridding masks. Moreover, based on the heat spreading concept, which is that, in a semiconductor device, heat is often assumed to spread in lateral directions with a maximum constant angle of $\alpha \approx 45^\circ$ [1], the lateral dimensions of each gridding mask are usually chosen in such a way that the distances between each mask in

lateral directions equal to distances between the corresponding gridding masks in the depth direction. However, in order to successfully create square blocks in the depth direction, some rules need to be obeyed in defining gridding mask depths as listed in Table 3-2.

In Table 3-2, n_1 , n_2 , n_3 , n_4 are the number of blocks along the “depth” direction in each of these numerical regions. According to the source geometry and substrate dimensions, the way to effectively set up the number of numerical regions and the number of blocks along the “depth” direction in each numerical region, involves the concept of grid independence, and is studied in the following sections.

3.2.4 Grid Independence

In modeling, there is often a trade-off between solution accuracy, the number of computational grids and simulation time. As the dimensions of a numerical grid decreases, the number of grids increases along with the simulation time. However, there is a point where a further decrease in the numerical grid dimensions has little effect on the solution due to the small temperature gradient between each finite volume when the subdivision is small. At this point, any solution is said to be grid-independent.

Again, similarly to example 1, Figure 3.8(a) shows an example where a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ heat source is mounted into the center of the substrate with depth $d = 0$. Substrate length and width are $400\mu\text{m}$, respectively, and substrate thickness is $100\mu\text{m}$. Power dissipation $P = 40\text{mW}$. Base temperature is assumed to be room temperature, 300K . Using the image method, thermal simulation was done for a quarter, as

shown in Figure 3.8(b), with $\frac{1}{4}P = 10\text{mW}$. According to the discussions in sections 3.2.2 and 3.2.3, “substrate gridding” (initial minimum substrate block size) was set to be $1\mu\text{m}$ on each side by properly setting up the “rows” and “cols” in the input file and the number of substrate layers in the technology file, according to the rules discussed above. Thus, by setting $depth = 2$ in the input file, the initial minimum substrate block could be subdivided into $0.25\mu\text{m}$, which would satisfy the source thickness requirement. The maximum block size in each numerical region could be $1\mu\text{m}$, $2\mu\text{m}$, $4\mu\text{m}$, $8\mu\text{m}$, ...

Figure 3.8 Example 2 (a) Structure of a semiconductor device thermal model
(b) A quarter of the structure

In this section, the goal is to find out the rules in determining the number of numerical regions in a substrate and dimensions of each of these regions, such that the difference between relevant values in any two meshes is negligible. Based on the quarter structure, Figure 3.9 shows the front view of the substrate numerical regions for example 2. Let l_1 , l_2 , l_3 , l_4 , be the x-y dimensions, and d_1 , d_2 , d_3 , d_4 , be the depth of each of these regions. Simulations were first done to determine the number of numerical regions.

Figure 3.9 Substrate Numerical Regions (Front View) of Example 2

Figure 3.10 shows the mesh structure of example 2 with three different gridding masks: power, device, and cell. Figure 3.11 shows the mesh structure of example 2 with four different gridding masks: power, device, cell, and circuit. The maximum block size of the power region, device region, and cell region, were kept to be $1\mu\text{m}$, $2\mu\text{m}$, $4\mu\text{m}$, respectively in both cases. Inside the power region, a heat generation region is created around the heat source by further subdividing the initial minimum substrate blocks ($1\mu\text{m}$). In both cases, the heat generation region was set to be $6\mu\text{m} \times 6\mu\text{m} \times 2\mu\text{m}$ with a minimum block size of $0.25\mu\text{m}$ using the refine statement in the input file. The x-y dimensions and depth of power region, device region, and cell region, were also set to be the same in both cases.

Figure 3.10 Meshes for Example 2 (1)

Figure 3.11 Meshes for Example 2 (2)

However, in Figure 3.11, because of the additional circuit region which has a maximum block size of $8\mu\text{m}$, the maximum substrate block size could increase from $8\mu\text{m}$, as in Figure 10, to $16\mu\text{m}$. Table 3-4 lists the comparisons of parameters and simulation results for both cases.

From the simulation results, it could be seen that, by properly adding the circuit region, the number of substrate blocks in Figure 3.11 is reduced by 36% and the CPU time is reduced by 17%, while the junction temperature rise is found to be the same as in Figure 3.10. This demonstrates that due to small temperature gradients at the points that are far away from the heat generation region, using large blocks at the region that is far away from the heat source can save calculation time while still remaining accurate.

The comparison of Figures 3.10 and 3.11 indicates that a 4-region model could be used if the dimensions and maximum block size of each numerical region are set up properly according to the rules discussed in sections 3.2.2 and 3.2.3.

Using the same example, to further investigate the effect of gridding mask x-y dimensions and depth on temperature rise, more simulations were done as shown in Figures 3.12 and 3.13. In Figure 3.12, the distances between each gridding mask in the depth direction are greater than those in the x-y directions. While in Figure 3.13, the distances between each gridding mask in the depth direction are less than those in the x-y directions. Four gridding masks were used in both cases. Comparisons and simulation results are also listed in Table 3-4. The simulation results indicate that in such cases as shown in Figure 3.13, there are more small blocks in the substrate lateral dimensions and heat spreading increases in the computational domain, causing a smaller rise in the junction temperature.

Conversely, in the case of Figure 3.12, because there are more small blocks in the substrate thickness direction than in the lateral direction, heat is forced to flow straight down to the substrate base, causing a greater temperature rise in the junction area. Compared with the simulation results obtained from Figures 3.10 and 3.11, because the difference in the temperature rise in each case is within 0.55%, the error caused by the ratio of gridding mask lateral dimensions and depth may be ignored.

To find out the acceptable minimum gridding mask lateral dimensions and depth for Example 2, Figures 3.14 and 3.15 include additional simulations. Comparisons and simulation results are also listed in Table 3-4. In both cases, four gridding masks were used, and the maximum block size of each region was set to be $1\mu\text{m}$, $2\mu\text{m}$, $4\mu\text{m}$ and $8\mu\text{m}$. The maximum substrate block size was $16\mu\text{m}$ in both cases. In Figure 3.14, the depth of each gridding mask was set to be the minimum possible sizes as listed in table 3-3, which were $2\mu\text{m}$, $4\mu\text{m}$, $8\mu\text{m}$ and $16\mu\text{m}$. This makes $d_2 - d_1 = s_2$, $d_3 - d_2 = s_3$, $d_4 - d_3 = s_4$. In addition, for the quarter structure, the lateral dimensions of each gridding mask were set in such a way that $l_2 - l_1 = d_2 - d_1$, $l_3 - l_2 = d_3 - d_2$, $l_4 - l_3 = d_4 - d_3$. The gridding mask lateral dimensions and depth chosen for each numerical region ensure the creation of square blocks in all dimensions, especially in the substrate thickness dimension. However, according to the simulation result listed in Table 3-4, there is an error of approximately 0.9% of the maximum junction temperature rise compared with those in Figures 3.10 and 3.11.

Nevertheless, in Figure 3.15, the depth of each gridding mask was set to be $2\mu\text{m}$, $8\mu\text{m}$, $16\mu\text{m}$ and $32\mu\text{m}$. This makes $d_2 - d_1 = 3s_2$, $d_3 - d_2 = 2s_3$, $d_4 - d_3 = 2s_4$. In

addition, the lateral dimensions of each gridding mask were set as $l_2 - l_1 = d_2 - d_1$, $l_3 - l_2 = d_3 - d_2$, $l_4 - l_3 = d_4 - d_3$. The gridding mask lateral dimensions and depth chosen for each numerical region in Figure 3.15 could ensure the creation of square blocks in all dimensions. According to the simulation result listed in Table 3-4, the error of maximum junction temperature rise in Figure 3.15, compared with those in Figures 3.10 and 3.11, is approximately 0.1%. In both Figures 3.14 and 3.15, the number of substrate blocks and CPU time are approximately the same.

Therefore, from the simulation results in Table 3-4 regarding the structure in Figures 3.14 and 3.15, it can be seen that the mesh structure in Figure 3.14 causes a bigger temperature error. Such an error is most likely to occur because the transitions of each numerical region as shown in Figure 3.14 are too sharp. In order to maintain solution accuracy, the use of more small grids near the heat generation region is required. Hence, based on the quarter structure, for single square sources, it is important to refer to the following relations:

$$l_i - l_{i-1} = d_i - d_{i-1} \geq 2 \cdot s_i \quad (3.6)$$

where, d_i is the current numerical region gridding mask depth,

d_{i-1} is the previous numerical region gridding mask depth,

l_i is the current numerical region gridding mask x-y dimension,

l_{i-1} is the previous numerical region gridding mask x-y dimension,

s_i is the maximum block size in the current region.

Figure 3.12 Meshes for Example 2 (3)

Figure 3.13 Meshes for Example 2 (4)

Figure 3.14 Meshes for Example 2 (5)

Figure 3.15 Meshes for Example 2 (6)

Table 3-4 Comparison of Mesh Structures and Temperatures for Example 2

parameters	Figure 3.10	Figure 3.11	Figure 3.12	Figure 3.13	Figure 3.14	Figure 3.15
initial minimum substrate block size (μm)	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$
maximum substrate block size (μm)	$8 \times 8 \times 8$	$16 \times 16 \times 16$	$16 \times 16 \times 16$	$16 \times 16 \times 16$	$16 \times 16 \times 16$	$16 \times 16 \times 16$
gridding mask x-y dimensions (μm)	-	88×88	40×40	88×88	24×24	40×40
gridding mask depth (μm)	-	80	80	32	16	32
maximum block size (μm)	-	8	8	8	8	8
gridding mask x-y dimensions (μm)	48×48	48×48	24×24	64×64	16×16	24×24
gridding mask depth (μm)	40	40	64	16	8	16
maximum block size (μm)	4	4	4	4	4	4
gridding mask x-y dimensions (μm)	28×28	28×28	16×16	48×48	12×12	16×16
gridding mask depth (μm)	20	20	56	8	4	8
maximum block size (μm)	2	2	2	2	2	2
gridding mask x-y dimensions (μm)	16×16	16×16	10×10	32×32	10×10	10×10
gridding mask depth (μm)	8	8	20	2	2	2
maximum block size (μm)	1	1	1	1	1	1
number of blocks	15537	9933	8443	7168	2254	2611
temperature rise (K)	48.75	48.75	48.76	48.48	49.18	48.69
CPU time (s)	0.6	0.5	0.5	0.5	0.4	0.4

Based on the results from Example 2 in Figure 3.15, an approximation is made in setting up the dimensions of each numerical region for single square or rectangular sources. Let “ l ” be the source length, and “ w ” be the source width. In addition, for the quarter structure, let “ l_i ” be the gridding mask length, “ w_i ” the gridding mask width, “ d_i ” the gridding mask depth, and “ s ” the initial minimum substrate block size. Assume that the heat source is close to the substrate surface, and its thickness is small ($< s$), then the possible values for each region in a quarter structure could be the following:

$$\text{heat generation region: } \begin{cases} l_g \geq \frac{1}{2}l + s \\ w_g \geq \frac{1}{2}w + s \\ d_g \geq 2s \end{cases} \quad (3.7)$$

$$\text{power region: } \begin{cases} l_1 = l_g + 4s \\ w_1 = w_g + 4s \\ d_1 = d_g \end{cases} \quad (3.8)$$

$$\text{device region: } \begin{cases} l_2 = l_1 + d_2 - d_1 \\ w_2 = w_1 + d_2 - d_1 \\ d_2 = 8s \end{cases} \quad (3.9)$$

$$\text{cell region: } \begin{cases} l_3 = l_2 + d_3 - d_2 \\ w_3 = w_2 + d_3 - d_2 \\ d_3 = 16s \end{cases} \quad (3.10)$$

$$\text{circuit region: } \begin{cases} l_4 = l_3 + d_4 - d_3 \\ w_4 = w_3 + d_4 - d_3 \\ d_4 = 32s \end{cases} \quad (3.11)$$

To verify the applicability of the above expressions, more simulations were done for different sizes of single square and rectangular sources. With the same substrate dimensions and power dissipation P as in Example 2, the source sizes were changed to $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$, $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$, $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ and $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$, respectively. Figures 3.16 and 3.17 are mesh structures for a quarter of the $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ source example. Figures 3.18 and 3.19 are mesh structures for a quarter of the $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ source example. Figures 3.20 and 3.21 are mesh structures for a quarter of the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ source example. Figures 3.22 and 3.23 are mesh structures for a quarter of the $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ source example. For each example, mesh (1) was created using the above approximation rules, while mesh (2) was created using more small blocks. Simulation results are listed in Tables 3-5 and 3-6.

Based on the simulation results, it could be seen that using the above approximation rules, for the $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ source example, the error in temperature rise is 0.8% compared with that obtained from (2), which has more small blocks. The CPU time is reduced by 33%. For the $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ source example, the error in temperature rise is 1.3% with a reduction of CPU time of 90% using the above approximation rules. For the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ source example, the error in temperature rise is

1.05% with a reduction of CPU time of 87.5%. For the $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ source example, the error in temperature rise is 0.4% with a reduction of CPU time of 90%.

Therefore, for all these examples, using the approximate rules, the errors in temperature rise are within 1.3%, while the CPU time is reduced greatly. This indicates that the approximation rules presented above can be used for single square or rectangular sources in order to save simulation time while still maintaining accuracy.

Moreover, according to the mesh structures in Figures 3.16, 3.18, as well as Figures 3.20 and 3.22, it could be seen that Atar could automatically generate square blocks in x-y directions by adjusting mask edges. Thus, it is straightforward in setting up the gridding mask x-y dimensions to make square blocks.

Figure 3.18 Meshes for the $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1)

Figure 3.19 Meshes for the $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2)

Figure 3.20 Meshes for the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1)

Figure 3.21 Meshes for the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2)

Figure 3.22 Meshes for the $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Source Example (1)

Figure 3.23 Meshes for the $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Source Example (2)

Table 3-5 Comparison of Mesh Structures and Temperatures for the $5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ Source and $50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ Source

parameters		$5\mu\text{m} \times 5\mu\text{m} \times 0.25\mu\text{m}$ heat source		$50\mu\text{m} \times 50\mu\text{m} \times 0.25\mu\text{m}$ heat source	
		Figure 3.16	Figure 3.17	Figure 3.18	Figure 3.19
initial minimum substrate block size (μm)		$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$	$1 \times 1 \times 1$
maximum substrate block size (μm)		16	16	16	16
heat region	x-y dimensions (μm)	3.5×3.5	6×6	26×26	30×30
	depth (μm)	2	4	2	4
	minimum block size (μm)	0.25	0.25	0.25	0.25
power region	gridding mask x-y dimensions (μm)	7.5×7.5	12×12	32×32	40×40
	gridding mask depth (μm)	2	16	2	8
	maximum block size (μm)	1	1	1	1
device region	gridding mask x-y dimensions (μm)	13.5×13.5	24×24	38×38	64×64
	gridding mask depth (μm)	8	32	8	32
	maximum block size (μm)	2	2	2	2
cell region	gridding mask x-y dimensions (μm)	21.5×21.5	56×56	46×46	96×96
	gridding mask depth (μm)	16	64	16	64
	maximum block size (μm)	4	4	4	4
circuit region	gridding mask x-y dimensions (μm)	37.5×37.5	80×80	62×62	112×112
	gridding mask depth (μm)	32	80	32	80
	maximum block size (μm)	8	8	8	8
number of blocks		1987	15749	18120	222345
temperature rise (K)		97.21	96.44	9.07	8.95
CPU time (s)		0.4	0.6	1.2	11.9

Table 3-6 Comparison of Mesh Structures and Temperatures for the $10\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ Source and $50\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Source

parameters		10 $\mu\text{m} \times 1\mu\text{m} \times 0.25\mu\text{m}$ heat source		50 $\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ heat source	
		Figure 3.20	Figure 3.21	Figure 3.22	Figure 3.23
initial minimum substrate block size (μm)		1 × 1 × 1	1 × 1 × 1	1 × 1 × 1	1 × 1 × 1
maximum substrate block size (μm)		16	16	16	16
heat region	x-y dimensions (μm)	6 × 2	12 × 6	26 × 11	30 × 15
	depth (μm)	2	16	2	4
	minimum block size (μm)	0.25	0.25	0.25	0.25
power region	gridding mask x-y dimensions (μm)	10 × 6	20 × 8	30 × 15	34 × 18
	gridding mask depth (μm)	2	24	2	8
	maximum block size (μm)	1	1	1	1
device region	gridding mask x-y dimensions (μm)	16 × 12	36 × 24	36 × 21	60 × 44
	gridding mask depth (μm)	8	40	8	32
	maximum block size (μm)	2	2	2	2
cell region	gridding mask x-y dimensions (μm)	24 × 20	64 × 48	44 × 29	96 × 80
	gridding mask depth (μm)	16	64	16	64
	maximum block size (μm)	4	4	4	4
circuit region	gridding mask x-y dimensions (μm)	40 × 36	80 × 64	60 × 45	112 × 96
	gridding mask depth (μm)	32	80	32	80
	maximum block size (μm)	8	8	8	8
number of blocks		2061	74714	8860	115414
temperature rise (K)		104.90	103.80	13.96	13.90
CPU time (s)		0.3	2.4	0.4	4.0

3.3 Multiple Sources and Gridding Mask Overlapping

One difficulty with the use of the square-block based numerical mesh is the partial overlapping of some of the nodes (blocks) at or near the interface of different numerical regions. In fact, from the above discussions, it can be seen that, in Atar, meshes are constructed by overlapping each numerical region in the substrate using different gridding masks. In addition, this overlapping commonly occurs in the case of multiple sources. The overlapping increases thermal resistance and capacitance, and can affect the solution accuracy. However, if the overlapping is relatively small compared to the entire region, the error involved is negligible.

From the simulation results of the examples shown previously in section 3.2.4, it is found that for single sources, changing the dimensions of each numerical region, i.e. changing the overlapping area of each numerical region, has little effect on the solution, provided that those rules are obeyed.

Following are two multiple-source examples. In example 1, three $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ sources are located symmetrically in the centre of a $400\mu\text{m} \times 400\mu\text{m} \times 100\mu\text{m}$ substrate. For each source, depth $d = 0$ and power dissipation $P = 40\text{mW}$. Figure 3.24(a) shows the top view of the structure. Because of its symmetry, simulation was done for a quarter of the entire structure as shown in Figure 3.24(b). Figures 3.25, 3.26 and 3.27 illustrate different mesh structures. For each of these mesh structures, the minimum and maximum block sizes in the substrate were $1\mu\text{m}$ and $16\mu\text{m}$ respectively on each side. In Figure 3.25, four numerical regions were created around each corresponding source according to the approximation rules discussed above, and

there was no inter-overlapping between the regions around each source. In Figures 3.26 and 3.27, the depth of each of the numerical regions was kept the same as those in Figure 3.25, but the lateral dimensions of each numerical region were increased. In Figure 3.26, there was partial inter-overlapping between the regions around each source. In Figure 3.27, the corresponding regions around each source were totally merged and had a denser mesh structure.

Simulation results are listed in Table 3-7. It could be seen that among these different mesh structures, the error in temperature rise is within 0.2%. However, using the meshes in Figure 3.25, the number of numerical blocks is reduced by 36%, and the CPU time by 20%.

**Table 3-7 Comparison of Mesh Structures and Temperatures
for the Multiple-Source Example 1**

simulation results	Figure 3.25	Figure 3.26	Figure 3.27
number of blocks	5269	6697	7159
temperature rise (K)	50.52	50.50	50.44
CPU time (s)	0.4	0.5	0.5

Figure 3.24 Multiple-Source Example I
 (a) Top view of the entire structure (b) Top view of a quarter

Figure 3.25 Meshes for Multiple-Source Example 1 (1)

Figure 3.26 Meshes for Multiple-Source Example 1 (2)

Figure 3.27 Meshes for Multiple-Source Example 1 (3)

Similar to Figure 3.24, Figure 3.28(a) shows the top view of another multiple-source example with the same substrate dimensions. Again, for each source, depth $d = 0$. Simulation was done for a quarter of the entire structure as shown in Figure 3.28(b). Figures 3.29 and 3.30 are different mesh structures. In both cases, the minimum and maximum substrate block sizes were $1\mu\text{m}$ and $16\mu\text{m}$, respectively, on each side. In Figure 3.29, meshes were created according to the approximation rules, and there was partial inter-overlapping. In Figure 3.30, the regions around each source were totally merged and had a denser mesh structure. Simulation results are listed in Table 3-8.

Figure 3.28 Multiple-Source Example 2
 (a) Top view of the entire structure (b) Top view of a quarter

Figure 3.29 Meshes for Multiple-Source Example 2 (1)

Figure 3.30 Meshes for Multiple-Source Example 2 (2)

Table 3-8 Comparison of Mesh Structures and Temperatures for the Multiple-Source Example 2

simulation results	Figure 3.29	Figure 3.30
number of blocks	11715	19961
temperature rise (K)	14.05	14.00
CPU time (s)	0.6	0.8

It could be seen that among these different mesh structures, the error in temperature rise is within 0.4%. However, using the meshes in Figure 3.29, the number of numerical blocks is reduced by 41%, and the CPU time is reduced by 25%.

According to the simulation results of the above two multiple-source examples, it can be seen that for multiple-source structures, the approximation rules can also be used to create numerical meshes around each source. Using the approximation rules can save simulation time while maintaining solution accuracy.

3.4 Conclusion

In this chapter, rules and procedures in setting up parameters to create square-block based numerical meshes were discussed. Based on the idea of grid independence, the rules of setting up gridding mask dimensions were also investigated, and approximation expressions were created and verified by running different models. Simulation results indicate that the error caused by the ratio of gridding mask lateral dimensions and depth, and the overlapping between each numerical region, can be ignored provided that all the rules are obeyed.

Chapter 4: Finite Substrate Effects in Thermal Simulation

4.1 Introduction

The main goal of this chapter is to suggest optimal substrate boundary dimensions in the computational domain for a given heat source structure in order to achieve a minimized simulation size and time while maintaining accuracy. Effects of finite substrate on thermal analysis are studied by using the numerical thermal simulator, Atar. Simulation results are compared to the heat spreading concepts. Some approximate approaches in estimating substrate boundary dimensions are also investigated by introducing the line-source simulation solutions and bounding box ideas.

4.1.1 Model Description

Heat transfer from a semiconductor device junction is often modeled as a small heat source mounted into a semi-infinite substrate with constant and isotropic thermal conductivity, as presented by Kraus, Chu and Bar-Cohen[5]. Furthermore, it is assumed that the heat source is close to the substrate surface, and the thermal energy generated by the semiconductor is released uniformly over the junction area. As shown in Figure 4.1, heat flows from the junction through the substrate, spreads in some complex three-dimensional manner, and leaves the substrate through some interfaces or boundaries located at a distance which is large relative to the dimensions of the junction area[6].

The boundary conditions at the sides and top of the substrate where there is no heat source are presumed to be adiabatic, that is, no heat flow exists perpendicular to the boundaries. Meanwhile, the bottom of the substrate is presumed to be isothermal to some constant backside temperature which is usually found to be between 40~85 °C. Nevertheless, since this backside temperature has little effect on the solution of the device junction temperature rise, a 300K backside temperature is considered throughout this research for convenience. In addition, for low power integrated semiconductor devices, since conduction is often presumed to be the dominant heat transfer mechanism[7][19], convection and radiation are neglected. These boundary conditions are the closest to those commonly encountered in real packages where the semiconductor chip is surrounded by thermally isolating media (plastic or air) except for the bottom surface, which is bonded to a good thermal conductor.

Figure 4.1 Heat Flow Lines and Isotherms in a Semi-Infinite Substrate

4.1.2 Modeling of Heat Spreading and Constriction

It has been often assumed that heat spreads with a constant angle from the power dissipating element (heat source). This angle can take any value from zero to 90° depending on the substrate boundaries. However, in isotropic structures where the heat spreads out equally well in all directions, a 45° spreading angle is often used to estimate the junction temperature rise. This spreading angle approximation is most commonly used and holds if no thermal boundaries come into play. However, as demonstrated by R.David, when the normalized thickness t_{sub}/l (l is the length of the heat source, and t_{sub} is the substrate thickness) is less than 2, the 26.6° single angle model is more accurate. However, this model becomes very inaccurate at larger thicknesses[7]. Another approach was also presented by R.David in the same paper suggesting a spreading angle of $\alpha = 32.5^\circ$. This is the best rough fit for the whole range of normalized thicknesses t_{sub}/l from 0.1 to 10. However, the fit is poorer than that of the 26.6° approach when normalized thickness t_{sub}/l is less than 2; it is also poorer than that of the 45° approach when t_{sub}/l is bigger than 2. Based on the physical situation that $\frac{t_{sub}}{l} < 2$ is not useful for device simulation, the 45° constant spreading angle approximation is often used in device thermal modeling.

However, the constant spreading angle model produces a mathematical case where, when the substrate thickness t_{sub} approaches infinity, the cross-sectional area, "A",

of the substrate normal to the heat flow path reaches infinity as well. Thus, thermal resistance:

$\theta = \frac{t_{sub}}{kA}$ approaches 0, where k is the thermal conductivity.

This, however, is physically not correct. In fact, as referred to in Figure 4.1, when substrate thickness moves to infinity, heat flows straight down to the substrate base. Therefore, for a substrate with a large thickness, the constant spreading angle approach is not suitable and new approaches need to be applied.

Moreover, all of the above constant angle models are based on the assumption that the boundaries are far away from the source. If thermal boundaries come into play, different situations need to be discussed.

Define “ L ” as the distance between source edges and substrate lateral boundaries. When substrate lateral boundaries are far away from the heat source and the substrate thickness is not large (i.e., the substrate thickness t_{sub} is smaller than the substrate lateral dimension L) as shown in Figure 4.2, heat spreads freely in lateral directions with a maximum constant angle of $\alpha \approx 45^\circ$.

Figure 4.2 Example of Free Heat Spreading

When the substrate thickness is large but narrow as shown in Figure 4.3, then the analysis needs to be broken down into two components in series.

Figure 4.3 Example of Restricted Heat Spreading

Figure 4.4 Thermal Resistance for Limited Heat Spreading (a) A trapezoidal portion of the heat path with spreading (b) A rectangular portion without heat spreading

The first component would have the trapezoidal portion of the heat path with spreading as shown in Figure 4.4(a). The second, the rectangular section, depicted in Figure 4.4(b), would have no heat spreading. Heat flows straight towards the substrate base with a constant cross-sectional area which is normal to the heat flow lines and forms a one dimensional heat flow, called “1D effect”. The total thermal resistance θ_{total} is the series combination of the thermal resistances of the trapezoidal and rectangular sections[1]:

$$\theta_{\text{total}} = \theta_1 + \theta_2 \quad (4.1)$$

where θ_1 =thermal resistance of the trapezoidal section

θ_2 =thermal resistance of the rectangular section

To solve this problem, a closed-form solution of the junction to substrate thermal resistance in semiconductor chips was created by Francesc N.Masana[8]. His model is based on an extension of the constant angle heat spreading, taking into account the source and substrate boundary dimensions, and changing the spreading angle accordingly. However, like any other analytical method, this model is limited to some simple and regular structures. In many practical situations, multiple heat sources will not be located symmetrically with respect to the substrate. For unsymmetrically located multiple sources, the solution in that model will become more complicated, and its accuracy needs to be verified. Therefore, a better way of accurately evaluating the thermal characteristics between junction and substrate base is to use the finite difference numerical method.

In this chapter, the simulation results are obtained by using the Atar thermal simulator which is based on the numerical method.

4.1.3 Simulation Size and Accuracy

In thermal modeling, there is often a trade-off between solution accuracy and simulation size, when using the finite difference numerical method.

Physically, the spreading of heat flow lines gives rise to a spreading resistance[9]. The maximum temperature rise within the junction area will occur at or near the centroid of the junction area, depending on whether it is symmetric or nonsymmetric.

As a general consideration, for small devices in large substrates, it seems evident that to accurately represent the heat flow and predict the maximum temperature rise in the junction area, the substrate boundaries in the computational domain must be sufficiently far from the heat source for border effects to be neglected. Otherwise, heat spreading will be constricted by the fictitious boundary surfaces, and the reflection of heat from the fictitious adiabatic boundaries will lead to a simulation based temperature increase which introduces a component of heat rise that is not representative of the physical situation[12].

However, increasing boundary distances means increasing the substrate dimensions and the number of grids which, in turn, as discussed in Chapters 2 and 3, will increase computational time enormously. On the other hand, similarly to the “grid independence” discussed in chapter 3, there is also a point where a further increase in the substrate dimensions has little effect on the solution of temperature rise in the junction area. This can be seen from the heat flow chart presented in Figure 4.1. At large substrate lateral dimensions, the isotherms are more straight and not affected when the substrate lateral dimensions are further increased. In addition, with a large substrate thickness, the isotherms become less dense and more straight in the substrate thickness direction, and changes in substrate thickness have little effect on temperature. Therefore, it can be seen that a further increase in substrate dimensions has few effects on temperature. Furthermore, from the heat spreading point of view as discussed above, for any given heat sources and substrate dimensions, the spreading angle varies accordingly. Therefore, for a given source size and substrate thickness, as long as the substrate dimensions chosen for simulation satisfy the condition that heat spreads and flows freely, the reflection error can be ignored.

Figure 4.5(a) shows a simulation model. Dimension “ l ” is the source length and “ w ” is the source width ($l \geq w$). Dimension “ L ” is the distance between the source edges and the substrate lateral boundaries. Dimension “ t_{sub} ” is the substrate thickness. As discussed in chapter 2, because of the symmetrical structure, simulation can be done for a quarter of the entire structure as shown in Figure 4.5(b).

Figure 4.5 Simulation Model (a) The entire structure of the simulation model
(b) A quarter of the model

As an example, a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square heat source with depth $d = 0$ was simulated by a quarter using Atar. The power dissipation for a quarter structure was $\frac{1}{4}P=10\text{mW}$. It was found that when simulating using a large substrate of $L = t_{sub} = 400\mu\text{m}$, the junction temperature rise was 49.52 K, the number of substrate block was 17250, and the CPU time was 10.1s. When simulating using $L = t_{sub} = 96\mu\text{m}$, the temperature rise was 49.00 K, the number of substrate block was 1812, and the CPU time was 0.1s. In both cases, the substrate base temperature was set to be 300 K, and the maximum substrate block size was set to be $16\mu\text{m}$ on each side. The

dimensions and maximum block sizes for each of the gridding masks in the substrate were built according to the rules discussed in chapter 3 and were kept the same in both cases.

It can be seen that, by reducing substrate dimensions, the temperature rise changes by approximately 1%, whereas the number of substrate blocks falls by 91.3% and the simulation CPU time falls by a factor of 100. These results indicate that by choosing a proper substrate size, it is possible to optimize simulation size and time while still maintaining accuracy.

Hence, the goal here is to study the finite substrate effect in thermal simulation and suggest optimal substrate boundary dimensions in the computational domain (i.e. for thermal simulation) for a given heat source, in order to minimize simulation size and save simulation time with retention of accuracy when using the numerical simulator, Atar.

4.1.4 Power Dissipation, Temperature Rise (T_{rise}) and Normalization

Thermal characteristics can be simplified by considering the thermal properties analogous to electrical properties. The temperature drop from the junction to the base of the substrate is analogous to the voltage drop V . The rate of heat flow is analogous to current flow. Thermal resistance θ is analogous to electrical resistance R . And the thermal conductivity k is analogous to electrical conductivity σ .

When current flows through a semiconductor device, power is dissipated which results in a rise in the device temperature. The temperature drop between a transistor junction and the base of a silicon chip substrate depends on the power to be dissipated as well as the geometry of the device[10][11].

In this research, only steady state is considered. Generally, in the case of steady state with uniform conductivity k , for a single source, conductive heat transfer could be defined as:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + q/k = 0 \quad (4.2)$$

where, k is thermal conductivity, and q is heat generated per unit volume. However, this equation is generally reduced to the Fourier equations[1], which express the heat transfer in each direction independently. Thus, a set of one-dimensional equations can be used to represent a three-dimensional heat flow as expressed below:

$$\begin{aligned} q_x &= -k\Delta y\Delta z \frac{\partial T}{\partial x} \\ q_y &= -k\Delta x\Delta z \frac{\partial T}{\partial y} \\ q_z &= -k\Delta x\Delta y \frac{\partial T}{\partial z} \end{aligned} \quad (4.3)$$

Hence, the temperature drop between a transistor junction and the base of a silicon chip depends on the power to be dissipated (or the heat generated) as well as the geometry of the device. In order to optimize the simulation size, one needs to simplify the problem down to the device geometry only. This simplification can be accomplished by introducing the normalized junction temperature rise.

Since k is assumed to be constant and not a function of temperature, the relationship between q and T_{rise} is linear. The normalization may be done by normalizing temperature rise with the ratio of T_{rise}/T_{inf} , where, T_{inf} is the temperature rise when the

substrate is semi-infinite, and T_{rise} is the temperature rise when the substrate is constricted, as in the computational domain. Thus, the effect of q on T_{rise}/T_{inf} can be ignored. Therefore, the ratio of T_{rise}/T_{inf} is simply related to the geometry of the heat source and the dimensions of the substrate.

In order to demonstrate that the ratio of T_{rise}/T_{inf} is a function of substrate dimensions and source geometry but is unaffected by power dissipation, P , more simulations were done using the model defined in Figure 4.5, with different sizes of single square sources. As an example, single square sources with sizes of $16\mu\text{m} \times 16\mu\text{m} \times 0.25\mu\text{m}$, $48\mu\text{m} \times 48\mu\text{m} \times 0.25\mu\text{m}$, $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$, and $112\mu\text{m} \times 112\mu\text{m} \times 0.25\mu\text{m}$ were simulated using Atar. The functions of T_{rise}/T_{inf} vs. $L = t_{sub}$ for each of these sources are plotted in Figure 4.6. In each case, dimensions and block sizes of the gridding masks in the substrate were built according to the rules discussed in Chapter 3 and kept the same while varying the finite substrate lateral dimension (L) and thickness (t_{sub}). Using T_{rise} at $L = t_{sub} = 400\mu\text{m}$ as a solution of source in semi-infinite substrate, i.e. T_{inf} , and changing the substrate lateral dimension L and thickness t_{sub} simultaneously, the effect of a different power dissipation on the function of T_{rise}/T_{inf} vs. $L = t_{sub}$ was compared for each source. The plot in Figure 4.6 shows that the power dissipation P (or q , the heat generated per unit volume) is not a factor in evaluating the function of T_{rise}/T_{inf} vs. $L = t_{sub}$. For example, for a $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ single square source, when being simulated at a quarter of the entire structure with a power dissipation

of $\frac{1}{4}P = 250\text{mW}$ (i.e., the heat generated per unit $q \approx 0.63\text{mW}/\mu\text{m}^3$), the function of T_{rise}/T_{inf} vs. $L = t_{sub}$, as shown by the solid line, is identical to that with power dissipation of $\frac{1}{4}P = 10\text{mW}$ (i.e., the heat generated per unit $q \approx 0.025\text{mW}/\mu\text{m}^3$) as shown in the plus sign.

Figure 4.6 The Effect of Power Dissipation P on the Function of T_{rise}/T_{inf} vs. $L = t_{sub}$

After normalizing the junction temperature rise with T_{rise}/T_{inf} , the problem is narrowed to investigating only the effects of source geometry and substrate dimensions on temperature rise. For a given heat source, when simulation substrate dimensions are large enough (close to infinity), T_{rise}/T_{inf} becomes close to "1". When substrate dimensions

are reduced in the computational domain, the constriction of heat flow will lead to simulation based temperature errors. By defining reflection error $= 1 - T_{rise}/T_{inf}$, this normalization of temperature rise could be used to evaluate the heat reflection error caused by the fictitious adiabatic boundaries in the computational domain. Therefore, this normalization is used throughout all of the simulation result analyses presented in this chapter.

4.2 Simulation Studies Using Atar

In order to determine an optimal substrate size and minimize simulation time, a large number of simulations were done using Atar to investigate the effect of simulation size on predicted temperature rise.

4.2.1 Temperature Rise (T_{rise}) With Respect to Combination of t_{sub} & L

Using the model as defined in Figure 4.5, simulations were first done for different sizes of single heat sources with a constant power density of $q = 1.6\text{mW}/\mu\text{m}^3$ and depth $d = 0$.

For each source, the maximum substrate block size was set to be $16\mu\text{m}$ on each side. In order to characterize reflection effects, the simulations were done in finite substrates with varying substrate lateral dimension (L) and thickness (t_{sub}), while the dimensions and block sizes of each gridding mask were kept the same. Figure 4.7 illustrates the contour plots of T_{rise} vs. the combination of t_{sub} and L in each case.

Figure 4.7 T_{rx} Contour as a Function of l_{sub} and L for a Quarter of Different Sources

The plots in Figure 4.7 indicate that when substrate thickness (t_{sub}) increases at constant L , temperature rise (T_{rise}) increases as well, whereas, when lateral dimension (L) increases at constant t_{sub} , T_{rise} decreases. This can be explained by the conductive heat transfer equation. From the above Fourier Equations (4.3), the temperature difference (ΔT) between the junction and the base of the substrate could be written in the form of

$$\Delta T = \frac{t_{sub}}{kA}q \quad (4.4)$$

where, A is the cross-sectional area normal to the heat flow path in substrate, and changes at different substrate thickness due to the 3D heat spreading in the substrate: t_{sub} is the substrate thickness, k is thermal conductivity, and q is the heat generated per unit volume (i.e. average power dissipation).

Thus, when increasing substrate thickness at constant L , temperature rise (T_{rise}) increases as well; Whereas, when increasing substrate lateral dimension L at constant t_{sub} , heat spreads more freely in the lateral directions, which means an increase in the cross-sectional area A , and, in turn, a decrease in T_{rise} .

However, as discussed in section 4.1.2, heat spreads with constant angles if the substrate thickness is not large and lateral boundaries are far away from the heat source as shown in Figure 4.8. Increasing further the substrate lateral dimension L , will not change the cross-sectional area A , and T_{rise} will also remain the same. This can be seen from the T_{rise} contour plots in Figure 4.7. When $L \gg t_{sub}$, T_{rise} is almost insensitive to L but is sensitive to t_{sub} if t_{sub} is small. The reason is that when the substrate is very thin and wide,

due to the close proximity of the heat source to the bottom of the substrate, no spreading can occur, as shown in Figure 4.9. Therefore, T_{rise} is directly proportional to t_{sub} .

Figure 4.8 3D Heat Flow in Substrate ($L \gg t_{sub}$, t_{sub} is not too small)

Figure 4.9 Restricted Heat Flow in Substrate ($L \gg t_{sub}$, t_{sub} is thin)

For a substrate with large thickness but narrow, heat spreading is restricted by the lateral boundaries, as shown in Figure 4.10. In this case, both changes in substrate lateral dimensions and thickness will affect T_{rise} dramatically. The plots in Figure 4.7 also show that when $t_{sub} \gg L$, T_{rise} is sensitive to t_{sub} and L . However, for a large thickness but narrow substrate, the maximum spreading angle is usually assumed to be 45° . By further increasing L , T_{rise} becomes less sensitive to L and its sensitivity to t_{sub} will also be reduced when increasing the substrate lateral dimensions and thickness. This can be explained by the 3-D effect of the heat flow[13]. The 3-D effect means that heat sources play the role of hot spots when the size of a heat source is much smaller than that of the substrate. As the

substrate dimensions become larger, the 3-D effect becomes dominant and the heat concentrates in the vicinity of the source area. In this case, the junction temperature decreases very gradually with the thinning of the substrate. On the other hand, as the substrate lateral dimension becomes smaller, the 3-D effect weakens and the heat flows over the entire substrate, causing the junction temperature to decrease significantly when substrate becomes thinner.

Figure 4.10 Restricted Heat Flow in Substrate ($t_{sub} \gg L$)

From a practical point of view, substrate thickness (t_{sub}) is usually provided and there are often two different cases: a substrate with a small thickness but with enough width, as shown in Figures 4.2 and 4.8, and a substrate with a large thickness and lateral dimensions such that the substrate may be considered to be semi-infinite.

In the first case, heat spreads freely in the lateral dimensions all the way towards the base of the substrate. Therefore, for a given heat source and substrate thickness (t_{sub}), in order to minimize simulation size, the main point is to keep the substrate thickness (t_{sub}), and reduce the lateral boundary dimensions (L) in the computational domain. In

this way the reflection error caused by the lateral boundaries is small enough compared to T_{rise} when the substrate is infinite in the lateral dimensions at a given thickness. For example, using the model as defined in Figure 4.5, assume that a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square heat source of power dissipation $\frac{1}{4}P = 10\text{mW}$ is located at the center of a substrate with thickness $t_{sub} = 100\mu\text{m}$ and depth $d = 0$. Figure 4.11(a) shows the front view for a quarter of the entire structure. A, B, C and D are points where L equals to $100\mu\text{m}$, $200\mu\text{m}$, $300\mu\text{m}$ and $400\mu\text{m}$, respectively. According to the T_{rise} contour plots as shown in Figure 4.7, for the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ heat source example, when $t_{sub} = 100\mu\text{m}$, the reflection error of T_{rise} is within 2% when L is reduced from $400\mu\text{m}$ to $100\mu\text{m}$. Figure 4.11(b) illustrates the situation by annotating the A, B, C and D points into the T_{rise} contour plot.

Figure 4.11 An Example of a Substrate With Small Thickness But Wide
 (a) Front view for a quarter of the entire structure (b) T_{rise} contour plot of the example

For the second case, when substrate dimensions are large (semi-infinite), the 3-D effect becomes dominant and the junction temperature decreases very gradually with the decrease of both the substrate thickness and the lateral dimensions. Therefore, in the case of the semi-infinite substrate, efforts are made to find the possibility of “shrinking” the substrate, i.e. to reduce both substrate thickness and lateral dimensions in the computational domain, in order to save simulation time while still maintaining accuracy. Again, use the model as defined in Figure 4.5. Assume that a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square heat source of power dissipation $\frac{1}{4}P = 10\text{mW}$ is located at the center of a substrate with thickness $L = t_{sub} = 400\mu\text{m}$ and depth $d = 0$. Figure 4.12(a) shows the front view for a quarter of the entire structure. Points A, B, C and D are where $L = t_{sub} = 100\mu\text{m}$, $L = t_{sub} = 200\mu\text{m}$, $L = t_{sub} = 300\mu\text{m}$, and $L = t_{sub} = 400\mu\text{m}$, respectively. Figure 4.12(b) illustrates the T_{rise} contour plot by annotating points A, B, C and D. It can be seen in this example that by reducing the substrate dimensions from $L = t_{sub} = 400\mu\text{m}$ to $L = t_{sub} = 100\mu\text{m}$, the reflection error of T_{rise} is within 2%.

Figure 4.12 An Example of a Semi-Infinite Substrate (Large L and t_{sub})
 (a) Front view for a quarter of the entire structure (b) T_{rise} contour plot of the example

Following are discussions for both of these two cases.

4.2.2 Given Small t_{sub} , Choose L

As discussed in the previous sections, it is clear that for a given substrate thickness which is not large, and for the lateral boundaries far away from the heat source, heat spreads laterally with different constant spreading angles from the heat source all the way towards the substrate base, depending on source geometries and substrate dimensions. To accurately evaluate T_{rise} in the junction area, the L chosen in the computational domain must satisfy for the heat to spread freely to ensure that the reflection error from the fictitious lateral boundaries is small enough.

Again, consider the above examples shown in Figure 4.7. In each case, for each t_{sub} , use T_{rise} at $L = 400\mu\text{m}$ as the solution for the heat source in substrates with finite thickness and infinite lateral dimensions, called “ T_{wide} ”. Ideally, for a given substrate thickness t_{sub} , when L moves to infinity, T_{rise} moves to a constant and $\frac{T_{rise}}{T_{wide}}$ becomes close to 1. Defining the reflection error function as $f = \left(\frac{T_{rise}}{T_{wide}} - 1 \right)$, Figure 4.13 shows percentage reflection errors of 1%, 2% and 5% (which are 0.01, 0.02 and 0.05 in the plot) on the T_{rise} contour plots for a quarter of the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ heat source example, where power dissipation is $\frac{1}{4}P=10\text{mW}$, annotating with $L = \frac{1}{2}t_{sub}$, $L = \frac{2}{3}t_{sub}$, $L = t_{sub}$, $L = \frac{3}{2}t_{sub}$, and $L = 2t_{sub}$ lines. The plot in Figure 4.13 shows that in the case of a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square source, for a given substrate thickness t_{sub} , choosing $L = t_{sub}$ could make the reflection error smaller than 1%.

Figure 4.13 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=10\text{mW}$)

To verify that choosing $L = t_{sub}$ can ensure acceptably small reflection errors for a given substrate thickness t_{sub} , additional simulations were run for different sizes of single square and rectangular sources. Figures 4.14, 4.15 and 4.16 show the simulation results for a $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ square source, a $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ rectangular source and a $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ rectangular source, respectively. The same analyses are performed as for the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square source, and reflection errors are plotted also. For each of these examples, unit power is set to be the same as $q = 1.6\text{mW}/\mu\text{m}^3$. Atar automatically simulates for the maximum temperature (T_{max}) at the center of the heat source based on the substrate base temperature, which is set

to room temperature, 300 K. Thus, the temperature rise (T_{rise}) could be calculated as

$$T_{rise} = T_{max} - 300.$$

These plots further demonstrate that, for a given small t_{sub} , choosing $L = t_{sub}$ could make the reflection error within 1%~2%, which is acceptable. This coincides with the assumption that when substrate lateral boundaries are far away from the heat source, and the substrate thickness is not large (i.e., the substrate thickness t_{sub} is smaller than the substrate lateral dimension L), heat will spread freely in lateral directions with a maximum constant angle of $\alpha \approx 45^\circ$.

Figure 4.14 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=640\text{mW}$)

Figure 4.15 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=60\text{mW}$)

Figure 4.16 T_{rise} Contour with Reflection Errors Compared to T_{wide} for a Quarter of the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=307\text{mW}$)

4.2.3 Shrinking of Substrates with a Large Thickness

Given the substrate thickness t_{sub} , choosing $L = t_{sub}$ may be satisfactory for small reflection errors in most cases. However, in the case of a given large substrate thickness t_{sub} , choosing $L = t_{sub}$ means a larger substrate size and more simulation time. On the other hand, from the above discussions, when substrate is thick and wide, T_{rise} becomes less sensitive to substrate boundary dimensions, where a further increase of the lateral boundary dimensions has little impact on T_{rise} . Moreover, if substrate is thick and wide, T_{rise} becomes less sensitive to substrate thickness when the simulation size is further increased in the thickness direction. Therefore, it is helpful to investigate the possibility of substrate shrinking to ensure small reflection errors while making substrate size as small as possible.

For a substrate with large thickness and lateral dimensions, substrate shrinking means reducing both substrate thickness and lateral dimensions in the computational domain. Based on the idea that for a given heat source, the maximum spreading angle is usually assumed to be 45° , to ensure that the reflection error caused by the lateral boundaries are sufficiently small, let $L = t_{sub}$ in order to satisfy the requirements for the maximum spreading angle. The main aim here is to find the substrate thickness (t_{sub}), i.e. the 1D dimension as discussed above, in order to suggest an optimal substrate size for simulating, and to minimize the simulation time while maintaining accuracy.

4.2.3.1 Substrate Shrinking for Single Square Sources

Choosing $L = t_{sub}$, simulations were first done for single square sources with different sizes. For each source, the dimensions and block size of each gridding mask were maintained and the substrate thickness (t_{sub}) and the lateral dimension (L) were varied simultaneously in order to characterize the reflection effect. Again, the depth of sources was set to $d = 0$. For each source, the power dissipation was the same, $\frac{1}{4}P=10\text{mW}$, and source thickness was $0.25\mu\text{m}$. Simulations were done for a quarter of the structures. Figure 4.17 shows the plots of T_{rise} vs. $L = t_{sub}$ for different sizes of single square sources from $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ to $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$.

Figure 4.17 T_{rise} vs. $L = t_{sub}$ for Single Square Sources

The plot in Figure 4.17 shows that for single square sources from $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ to $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$, T_{rise} is almost constant as

$L = t_{sub}$ moves to $400\mu\text{m}$. Therefore, use T_{rise} at $L = t_{sub} = 400\mu\text{m}$, or now “ T_{inf} ”, as a solution to the source in a semi-infinite substrate. Ideally, when $L = t_{sub}$ increases, T_{rise} becomes a constant and T_{rise}/T_{inf} becomes close to 1. Figure 4.18 shows the functions of T_{rise}/T_{inf} vs. $L = t_{sub}$ for different single square sources from sizes $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ to $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$. Defining “reflection error” to be $(1 - T_{rise}/T_{inf})$, points with percentage reflection errors of 1% and 2% (corresponding to 0.99 and 0.98 in the plot) are also annotated.

The plot in Figure 4.18 shows that for small square sources, the simulation substrate size may be reduced while maintaining accuracy. For example, for a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ source, choosing $L = t_{sub} \geq 61\mu\text{m}$ can make the reflection error $\leq 2\%$. Choosing $L = t_{sub} \geq 99\mu\text{m}$ can make the reflection error $\leq 1\%$. Hence, if there is a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ heat source with a given substrate thickness of $200\mu\text{m}$, instead of simulating at $L = t_{sub} = 200\mu\text{m}$ to maintain accuracy as discussed in section 4.2.2, the substrate may be able to be simulated at $L = t_{sub} = 61\mu\text{m}$ with a reflection error of 2%, which is acceptable compared with that when substrate is semi-infinite.

However, for large sizes of single square sources, in order to remain accuracy, the shrinking of the substrate size is limited. For example, for the $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ square source, in order to make the reflection error $\leq 2\%$, one must choose $L = t_{sub} \geq 253\mu\text{m}$. Thus, in this case, for any given t_{sub} that is smaller than $253\mu\text{m}$, one must choose $L = t_{sub}$ to maintain accuracy, as discussed in section 4.2.2.

Figure 4.18 T_{rise} / T_{inf} vs. $L = t_{sub}$ for Different Sizes of Single Square Sources

This result coincides with the 3D effect as discussed in section 4.2.1, that when a heat source is small, the 3D effect becomes dominant, the heat concentrates in the vicinity of the source area, and the junction temperature decreases very gradually with the thinning of the substrate. Therefore, in this case, the substrate size can be greatly shrunk. Nevertheless, as the source area becomes larger, the 3D effect weakens and the heat spreads over the entire chip. In this case, the junction temperature decreases significantly as the substrate becomes thinner. Any attempt to further shrink the substrate size may cause significant reflection errors.

4.2.3.2 Substrate Shrinking for Single Rectangular Sources

The same simulations as were done for single square sources were also done for single rectangular thermal sources with different aspect ratios (l/w) and sizes. Figures 4.19 and 4.20 show the plots of T_{rise}/T_{inf} vs. $L = t_{sub}$ for sources with different aspect ratios ($l/w=1,2,3,4$). In Figure 4.19, source length (l) was kept as $40\mu\text{m}$, and source width (w) was changed according to the aspect ratios. In Figure 4.20, source length (l) was kept as $64\mu\text{m}$, and source width (w) was changed. From these plots, the relations between source size, aspect ratios, and their effects on reflection error functions are not obvious. In order to determine the computational substrate dimensions according to source geometry, so as to retain solution accuracy, a predictive capability is required. This requirement is pursued by attempting to determine substrate dimensions based on substrate thickness (t_{sub}) and distances from source edges to substrate lateral boundaries (i.e. “ L ” as defined in section 4.1.3 Figure 4.5), where $L = t_{sub}$. For this purpose, line-source simulations are introduced.

Figure 4.19 T_{rise} / T_{inf} vs. $L = t_{sub}$ for Single Rectangular Sources with Different Aspect Ratios (1)

Figure 4.20 T_{rise} / T_{inf} vs. $L = t_{sub}$ for Single Rectangular Sources of Different Aspect Ratios (2)

4.3 Line Source Simulation

This section introduces the concept of line-source simulations to further investigate the relations between source dimensions, aspect ratios, and their effects on substrate shrinking. Line-source simulations are based on the idea that reflection error functions are related to source dimensions and aspect ratios. Therefore, simulations are done by taking each edge of a thermal source as a single line source whose width and thickness are small (e.g., set line source width the same as source thickness, $0.25\mu\text{m}$, respectively), characterizing reflections at each edge using these line sources, applying the results to the edges of the rectangular source, and comparing the simulation results with those when simulating at single square source or rectangular source as shown in Figures 4.18, 4.19 and 4.20. Figure 4.21 shows the plots of the reflection error (defined as $1 - T_{rise}/T_{inf}$) vs. $L = t_{sub}$ for various line sources with lengths from $10\mu\text{m}$ to $80\mu\text{m}$.

4.3.1 Line Source Simulation Results Compared with Single Square Source Simulation Results

Comparing the plots in Figure 4.21 with those in Figure 4.18, it is found that for a single square source with a certain substrate size of $L = t_{sub}$, its reflection error is approximately the sum of the reflection errors caused by each source edge, considering each edge of the source as a line source. That is, for a $l \times l$ square source with certain $L = t_{sub}$, its reflection error is close to four times that of a line source of length l at the same $L = t_{sub}$.

For example, for a $40\mu\text{m} \times 40\mu\text{m} \times 0.25\mu\text{m}$ single square source, from the simulation results plotted in Figure 4.18, when $L = t_{sub} = 238\mu\text{m}$, $T_{rise}/T_{inf} = 0.99$ (i.e. reflection error = $0.01 = 1\%$). When $L = t_{sub} = 175\mu\text{m}$, $T_{rise}/T_{inf} = 0.98$ (i.e. reflection error = $0.02 = 2\%$). Whereas, from the line-source simulation results plotted in Figure 4.21, when $L = t_{sub} = 237\mu\text{m}$, reflection error = $0.0025 = 0.25\%$. When $L = t_{sub} = 175\mu\text{m}$, reflection error = $0.005 = 0.5\%$. Table 4-1 lists the comparison results of line-source simulation and single square source simulation.

Table 4-1 Comparison of Line-Source Simulation and Square Source Simulation

source length (l)	10	16	20	24	30	32	40	48	64	80
$L(\text{or } t_{sub})$ at error = 0.0025×4 from line-source simulation	126	161	179	194	213	219	237	254	277	295
$L(\text{or } t_{sub})$ at error = 1% from square source simulation	99	139	162	182	206	213	238	258	287	308
$L(\text{or } t_{sub})$ at error = 0.005×4 from line-source simulation	79	105	120	133	150	156	175	192	218	239
$L(\text{or } t_{sub})$ at error = 2% from square source simulation	60	89	107	123	144	150	175	195	228	253

These comparison results demonstrate that the L (and t_{sub}) found from line-source simulations are close to those found from single square source simulations. This comparison provides a feasible approximate approach to estimating substrate thickness (t_{sub}) and lateral dimension (L) for thermal simulation of single square sources at certain reflection errors (where $L = t_{sub}$). That is, to find the value of $L = t_{sub}$ for a $l \times l$ square source

at a certain reflection error, it is possible to use the simulation results for the line source of length l , and find $L = t_{sub}$ at one-fourth of the reflection error of that from a single square source. A possible explanation for this approximation is that because of the symmetrical structure described above, for a single square source, the heat flow across each side is equal. Thus, the reflection error of a single square source can be calculated as approximately four times that of a line source with the same length of the square source. The advantage of this result is the narrower simulation size of a source and the ability to investigate the relations among source dimensions, aspect ratios and reflection errors.

4.3.2 Line Source Simulation Results Comparing to Single Rectangular Source Simulation Results

Similar analyses and comparisons are performed for both line source and single rectangular source simulations. Again, the idea is that for a single square source or rectangular source with certain $L = t_{sub}$, its reflection error is approximately the sum of the reflection errors caused by each source edge, considering each edge of the source as a line source. Therefore, for a $l \times w$ rectangular source with certain $L = t_{sub}$, its reflection error can be estimated as two times the error from the line source of length l plus two times the error from the line source of length w at the same $L = t_{sub}$. Figures 4.22 and 4.23 show the line-source simulation solutions for the same source structures as shown in Figure 4.19. In Figure 4.22, points of $L = t_{sub}$ that make the reflection errors of the rectangular sources equal to 1% are annotated, and in Figure 4.23, points of $L = t_{sub}$ that make the reflection errors of the rectangular sources equal to 2% are annotated for single

rectangular sources with the same length of $40\mu\text{m}$ and different aspect ratios, as shown in Figure 4.19.

Comparing the plots in Figure 4.19 with those in Figures 4.22 and 4.23, it can be seen that for a $40\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ rectangular source, when being simulated as a single rectangular source, at $L = t_{sub} = 203\mu\text{m}$, reflection error = 1% as shown in Figure 4.19; while from the line-source simulation results plotted in figure 4.22, when $L = t_{sub} = 213\mu\text{m}$, reflection error = $(0.0033 + 0.0017) \times 2 = 0.01 = 1\%$. In addition, from Figure 4.19, when $L = t_{sub} = 142\mu\text{m}$, reflection error = 2%; while from Figure 4.23, when $L = t_{sub} = 151\mu\text{m}$, reflection error = $(0.0066 + 0.0033) \times 2 = 0.02 = 2\%$. Hence, the solutions for $L = t_{sub}$ from line-source simulations are also close to those from single rectangular simulations. More detailed comparisons are listed in Table 4-2.

Furthermore, from the plots in Figures 4.22 and 4.23, it is found that for a rectangular source, the ratio of reflection errors caused by each edge is approximately the same as its aspect ratio (l/w). Taking the same example of a $40\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ rectangular source, its aspect ratio is 2. When $L = t_{sub} = 213\mu\text{m}$, the reflection error caused by its $40\mu\text{m}$ line source (edge) is 0.0033, about 2 times that caused by its $20\mu\text{m}$ line source (edge), which is 0.0017. However, when $L = t_{sub} = 151\mu\text{m}$, the reflection error caused by its $40\mu\text{m}$ line source is 0.0066, about 2 times that caused by its $20\mu\text{m}$ line source, which is 0.0033. Detailed comparisons for each case are also listed in Table 4-2.

Figure 4.22 Line-Source Simulation Solution at 1% Reflection Error for Rectangular Sources with l of $40\mu m$ and Different Aspect Ratios

† points of $L = t_{sub}$ that make the reflection error of the rectangular sources equals to $2 \times 0.5\% = 1\%$

Figure 4.23 Line-Source Simulation Solution at 2% Reflection Error for Rectangular Sources with l of 40 μm and Different Aspect Ratios
 † points of $L = t_{sub}$ that make the reflection error of the rectangular sources equals to $2 \times 10^{-2} = 2\%$

Table 4-2 Comparison of Line-Source Simulation and Rectangular Source Simulation (1)

source length (l)		40	40	40	40
source width (w)		10	20	30	40
aspect ratio (l / w)		4	2	1.3	1
reflection error=1%	$L(\text{or } t_{sub})$ from rectangular source simulation	176	203	223	238
	$L(\text{or } t_{sub})$ from line-source simulation	197	213	226	237
	error from line source of l	0.0039	0.0033	0.0028	0.0025
	error from line source of w	0.0011	0.0017	0.0022	0.0025
	ratio of reflection errors from l and w	3.6	1.9	1.3	1
reflection error=2%	$L(\text{or } t_{sub})$ from rectangular source simulation	118	142	160	175
	$L(\text{or } t_{sub})$ from line-source simulation	137	151	164	175
	error from line source of l	0.0078	0.0066	0.0057	0.0050
	error from line source of w	0.0022	0.0034	0.0043	0.0050
	ratio of reflection errors from l and w	3.5	1.9	1.3	1

Similar analyses and comparisons are done for rectangular sources with a length of $64\mu\text{m}$ and different aspect ratios, as shown in Figure 4.20. Simulation results as plotted in Figure 4.20 are also compared with those from line-source simulation solutions plotted in Figures 4.24 and 4.25. Comparison results are listed in Table 4-3 which, again, verifies the above assumption.

Figure 4.24 Line-Source Simulation Solution at 1% Reflection Error for Rectangular Sources with l of $64\mu m$ and Different Aspect Ratios

* points of $L = l_{sub}$ that make the reflection error of the rectangular sources equals to $2 \times 0.5\% = 1\%$

Figure 4.25 Line-Source Simulation Solution at 2% Reflection Error for Rectangular Sources with l of $64\mu m$ and Different Aspect Ratios
+ points of $L = t_{sub}$ that make the reflection error of the rectangular sources equals to 2×10^{-4}

Table 4-3 Comparison of Line-Source Simulation and Rectangular Source Simulation (2)

source length (l)		64	64	64	64
source width (w)		16	32	48	64
aspect ratio (l / w)		4	2	1.3	1
reflection error=1%	$L(\text{or } t_{sub})$ from rectangular source simulation	228	256	274	287
	$L(\text{or } t_{sub})$ from line-source simulation	239	254	267	277
	error from line source of l	0.0040	0.0033	0.0029	0.0025
	error from line source of w	0.0010	0.0017	0.0021	0.0025
	ratio of reflection errors from l and w	4	1.9	1.4	1
reflection error=2%	$L(\text{or } t_{sub})$ from rectangular source simulation	164	192	212	228
	$L(\text{or } t_{sub})$ from line-source simulation	176	192	206	218
	error from line source of l	0.0079	0.0066	0.0057	0.0050
	error from line source of w	0.0021	0.0034	0.0043	0.0050
	ratio of reflection errors from l and w	3.8	1.9	1.3	1

Thus, the line source approximate solution provides a straightforward method of investigating the relations between source dimensions, aspect ratios, and their effects on reflection errors. Based on the above discussion, it is possible to estimate the optimal substrate size ($L = t_{sub}$) for a certain rectangular source, first by calculating the aspect ratio (l / w) from the given source length l and source width w ($l \geq w$). Then, according to the reflection error specification, the acceptable reflection error for the line source of length w

can be defined as $\left(\frac{\text{specified reflection error}}{2 \times (1 + l/w)} \right)$. Finally, from the line-source simulation results plotted in Figure 4.21, the optimal substrate size ($L = t_{sub}$) can be found.

Recalling the examples discussed in section 4.2.2, the T_{rise} at $L = t_{sub} = 400\mu\text{m}$, or " T_{inf} ", is now used as a solution for source in the semi-infinite substrate, and define "reflection error" to be $(1 - T_{rise}/T_{inf})$. Figures 4.26 to 4.29 show percentage reflection errors of $\pm 1\%$, $\pm 2\%$ and $\pm 5\%$ (which are ± 0.01 , ± 0.02 and ± 0.05 in the plots) on the T_{rise} contour plots for a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square source, a $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ square source, a $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ rectangular source and a $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ rectangular source, respectively, annotating with $L = \frac{1}{2}t_{sub}$, $L = \frac{2}{3}t_{sub}$, $L = t_{sub}$, $L = \frac{3}{2}t_{sub}$, $L = 2t_{sub}$ lines.

Figure 4.26 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=10\text{mW}$)

Figure 4.27 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=640\text{mW}$)

Figure 4.28 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=60\text{mW}$)

Figure 4.29 T_{rise} Contour with Reflection Errors Compared to T_{inf} for a Quarter of the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ Thermal Source ($\frac{1}{4}P=307\text{mW}$)

The plot in Figure 4.26 visually shows that for a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square source, when $L = t_{sub} \geq 100\mu\text{m}$, reflection error is $\leq 1\%$; when $L = t_{sub} \geq 60\mu\text{m}$, reflection error $\leq 2\%$. This result coincides with the square source simulations shown in Figure 4.18 and listed in Table 4-1. However, from the line-source simulation solution discussed above, when the reflection error specification is 1%, for a square source ($l/w=1$), the acceptable reflection error for the single line source l is: specified reflection error / $(2 \times (1 + l/w)) = 0.25\%$; when reflection error specification is 2%, the acceptable reflection error for the line source l is 0.5%.

From the plot in Figure 4.21, for a $10\mu\text{m}$ line source, when reflection error is $\leq 0.25\%$, $L = t_{sub} \geq 126\mu\text{m}$; for reflection error $\leq 0.5\%$, $L = t_{sub} \geq 79\mu\text{m}$. Therefore, for a $10\mu\text{m} \times 10\mu\text{m} \times 0.25\mu\text{m}$ single square source, from the line-source simulation solution, when $L = t_{sub} \geq 126\mu\text{m}$, reflection error is $\leq 1\%$; when $L = t_{sub} \geq 79\mu\text{m}$, reflection error is $\leq 2\%$. This result is close to that obtained from the plot in Figure 4.26.

In addition, the plot in Figure 4.27 shows that for a $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ single square source, when $L = t_{sub} \geq 300\mu\text{m}$, reflection error is $\leq 1\%$; when $L = t_{sub} \geq 250\mu\text{m}$, reflection error is $\leq 2\%$. Again, this result is close to the results obtained from the single square source simulation and line-source simulation as shown in Figure 4.18, Figure 4.21 and Table 4-1. The differences are within 4%.

As for the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ rectangular source, its aspect ratio is $l/w = 1.5$. When the reflection error specification is 1%, the acceptable reflection error for the $20\mu\text{m}$ line source is: specified reflection error / $(2 \times (1 + l/w)) = 0.2\%$

Similarly, when reflection error specification is 2%, the acceptable reflection error for the $20\mu\text{m}$ line source is 0.4%. Refer to the plots in Figure 4.21, for a $20\mu\text{m}$ line source, when reflection error is $\leq 0.2\%$, $L = t_{sub} \geq 200\mu\text{m}$; when reflection error is $\leq 0.4\%$, $L = t_{sub} \geq 140\mu\text{m}$. In addition, from the plots in Figure 4.28, when $L = t_{sub} \geq 190\mu\text{m}$, reflection error is $\leq 1\%$; when $L = t_{sub} \geq 130\mu\text{m}$, reflection error is $\leq 2\%$. Hence, the results are approximately the same.

Again, as for the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ rectangular source, its aspect ratio is $l/w = 3.0$. When the reflection error specification is 1%, the acceptable reflection error for the $32\mu\text{m}$ line source is 0.125%; when the reflection error specification is 2%, the acceptable reflection error for the $32\mu\text{m}$ line source is 0.25%. Refer to the plots in Figure 4.21, for a $32\mu\text{m}$ line source, when reflection error is $\leq 0.125\%$, $L = t_{sub} \geq 280\mu\text{m}$; when reflection error is $\leq 0.25\%$, $L = t_{sub} \geq 220\mu\text{m}$. In addition, from the plots in Figure 4.29, when $L = t_{sub} \geq 280\mu\text{m}$, reflection error is $\leq 1\%$; when $L = t_{sub} \geq 220\mu\text{m}$, reflection error is $\leq 2\%$. Therefore, the results are also the same.

All these comparisons further demonstrate that the line-source simulation solution provides a simplified approximate approach to estimating an optimal substrate size for thermal simulation.

4.4 Bounding Box Simulation for Multiple Sources

The above discussions are based on single heat source simulations. However, in practice, there are usually several sources that are tiny in thickness, but thin and long in lateral dimensions near the substrate surface, which represent transistors in realistic situations. Therefore, it is useful to investigate the possibilities of bounding box simulations which consider the entire source structures as a bounding box, and estimate the substrate boundary dimensions according to the bounding box size. By doing this, the line-source simulation solutions discussed above can be applied in estimating an optimized substrate size for thermal simulation, and can save simulation time.

In the case of multiple sources, heat transfer is more complicated. Figure 4.30 shows the area plot of temperature distributions for a half of the nine-source example shown in Figure 4.33. Axis “x” is the substrate length direction, and “y” is the substrate width direction. The maximum temperature rise can occur at the center of any of these sources depending on the power dissipation on each source and the source structure. In this example, the maximum temperature rise is found to be at the center of the third source from the left.

Figure 4.30 Surface Plot of Temperature Distributions for a Nine-Source Example as shown in Figure 4.33

However, because the main purpose here is to determine the substrate boundary dimensions for certain combinations of multiple sources so that the reflection error, which

is defined as $1 - T_{rise}/T_{inf}$, is acceptably small, the effect of P (or q) is not a factor here, as discussed in section 4.1.4. Besides, discussions of the exact temperature distribution among multiple sources exceed the target of this research. Hence, effort has been placed on simplifying the source geometry in order to estimate the substrate boundary dimensions while still maintaining accuracy in evaluating reflection errors.

A feasible approach is to consider the multiple-source structure as a bounding box, and use the line-source simulation solutions discussed above in order to estimate the substrate boundary dimensions for simulation. Similarly to the single-source model, again, the substrate lateral dimension L is defined as the distance between the edge of the bounding box and the edge of the substrate. Figures 4.31(a), 4.32(a) and 4.33(a) show the top view of some multiple-source examples. Figure 4.34 shows the comparison of multiple-source simulation results and bounding box estimation results for those examples. Because of the axisymmetrical structures of these multiple-source examples, simulations are done for half of the whole structures as shown in (b) for each of these examples. Assume the thickness and depth of each source is $0.25\mu\text{m}$ and $0\mu\text{m}$, respectively.

(a)

(b)

Figure 4.31 A Three-Source Example (a) Top view of the whole structure
(b) Top view of a half structure along the symmetrical axis

Figure 4.32 A Seven-Source Example (a) Top view of the whole structure
(b) Top view of a half structure along the symmetrical axis

(a)

(b)

Figure 4.33 A Nine-Source Example (a) Top view of the whole structure
(b) Top view of a half structure along the symmetrical axis

4.3.4 Multiple Source Simulations

Figure 4.34 shows the reflection error function, i.e. $(1 - T_{rise}/T_{inf})$ vs. $L = t_{sub}$ for each one of these examples, using the full solution of multiple-source simulations. Points of 1% and 2% reflection error are annotated in “*”. In addition, an extracted view from Figure 4.21 regarding line-source solutions based on the size of bounding boxes for each one of these examples is enclosed in the up-right corner of Figure 4.34. The estimated substrate dimensions ($L = t_{sub}$) with 1% and 2% reflection errors found from the line-source solutions, which are based on the bounding box sizes for each one of these examples, are also annotated in “o” on those reflection error functions. These functions are based on the full solution of multiple-source simulations. Solutions from the line source approach are compared with those full solutions.

As an example, for the seven-source structure as shown in Figure 4.32, the size of the bounding box is $64\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$. According to the line-source simulation solutions discussed in section 4.3, to keep the reflection error of the bounding box smaller than 1%, the acceptable reflection error for the $32\mu\text{m}$ line source is:

$$\frac{1\%}{2 \times (1 + l/w)} \approx 0.17\% = 0.0017.$$

From the line-source simulation results plotted in Figure 4.21 (or its extracted view on the up-right corner of Figure 4.34), it can be approximately seen that when $L = t_{sub} \geq 250\mu\text{m}$, the reflection error for the line source of length $32\mu\text{m}$ is ≤ 0.0017 . Thus, to make the reflection error of the $64\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ bounding box smaller than 1%, one should choose $L = t_{sub} \geq 250\mu\text{m}$.

Similarly, to keep the reflection error of the bounding box smaller than 2%, the acceptable reflection error for the $32\mu\text{m}$ line source is: $\frac{2\%}{2 \times (1 + l/w)} \approx 0.33\% = 0.0033$.

From Figure 4.21 and its extracted view on the up-right corner of Figure 4.34, it can also be found approximately that when $L = t_{sub} \geq 193\mu\text{m}$, the reflection error for the $32\mu\text{m}$ line source is ≤ 0.0033 . Therefore, to make the reflection error of the $64\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ bounding box smaller than 2%, one should choose $L = t_{sub} \geq 193\mu\text{m}$.

Meanwhile, from the reflection error functions using the full solution of multiple-source simulations, it is found that when $L = t_{sub} \geq 242\mu\text{m}$, the reflection error for the $64\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ seven-source example is $\leq 1\%$, and when $L = t_{sub} \geq 177\mu\text{m}$, the reflection error is $\leq 2\%$ (as shown in Figure 4.34 with “*”).

However, when plotting $L = t_{sub} = 250\mu\text{m}$ and $L = t_{sub} = 193\mu\text{m}$ with 1% and 2% reflection errors (which are obtained from the line-source solution based on the $64\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ bounding box for the seven-source example) onto the reflection error function plots which are based on the full solutions of multiple-source simulations in Figure 4.34, it is found that the reflection error at $L = t_{sub} = 250\mu\text{m}$ is 0.9% and the reflection error at $L = t_{sub} = 193\mu\text{m}$ is 1.7% for the $64\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ seven-source example (as shown in Figure 4.34 with “o”). Comparing the results from the multiple-source full solution and the bounding box solution, the difference of the reflection errors is within 0.3%.

The same analyses are done for the three-source example with a bounding box size of $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$, as shown in Figure 4.31, and the nine-source example with a bounding box size of $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$, as shown in Figure 4.33. The line-source simulation solution based on bounding box size, as shown in the extracted view from Figure 4.21, indicates that to make the reflection error of the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ bounding box smaller than 1%, one should choose $L = t_{sub} \geq 199\mu\text{m}$, and to make its reflection error smaller than 2%, one should choose $L = t_{sub} \geq 138\mu\text{m}$. In addition, to make the reflection error of the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ bounding box smaller than 1%, one should choose $L = t_{sub} \geq 277\mu\text{m}$, and to make its reflection error smaller than 2%, one should choose $L = t_{sub} \geq 219\mu\text{m}$.

However, from the reflection error plots in Figure 4.34 which are based on the full solution of multiple-source simulations, it is found that for the $30\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ three-source example, when $L = t_{sub} \geq 199\mu\text{m}$, reflection error is $\leq 0.7\%$; when $L = t_{sub} \geq 138\mu\text{m}$, reflection error is $\leq 1.4\%$. Also, for the $96\mu\text{m} \times 32\mu\text{m} \times 0.25\mu\text{m}$ nine-source example, when $L = t_{sub} \geq 277\mu\text{m}$, reflection error is $\leq 0.8\%$; when $L = t_{sub} \geq 219\mu\text{m}$, reflection error is $\leq 1.6\%$.

Comparing the results from the multiple-source full solution and the bounding box solution, the differences in reflection error are within 0.6% for all these examples.

Besides, in all these examples, the $L = t_{sub}$ found from the multiple-source full solutions are smaller than those from the bounding box solutions. Hence, if the $L = t_{sub}$ chosen can satisfy the reflection error specifications when using bounding box simulation, then this $L = t_{sub}$ can also ensure that the reflection error is within specifications when doing multiple-source full simulation.

Thus, the bounding box idea provides a feasible approach for quickly estimating the approximate substrate boundary dimensions in the cases of multiple sources.

4.5 Conclusion

In this chapter, line-source simulation solutions were used as an approximate approach for quickly estimating substrate boundary dimensions for different heat source and substrate geometries. Along with the bounding box idea, the line-source solution provides a feasible approximate approach for multiple-source structures. The reflection error differences between the full solution of multiple-source simulation and the line-source solution, which is based on the bounding box size, are within 0.6%.

Chapter 5: Source Quantization

5.1 Purpose

In chapters 3 and 4, rules for generating numerical meshes and choosing computational substrate dimensions were discussed. All of those discussions were based on the main purpose of this research, that is, to develop methods of efficient numerical thermal simulation for semiconductor devices in order to quickly and accurately predict their maximum junction temperature rise.

However, the approximate approach discussed in section 3.2.4, regarding the generation of square-block based numerical meshes, is suitable for single sources or multiple sources with such structures that both the distances between each one of the heat sources and the geometries of the sources themselves can be represented by multiples or subdivisions of the initial minimum substrate block as defined in chapter 3. However, in reality, these situations are not always satisfied. In the case of multiple heat source structures, if the above conditions cannot be satisfied simultaneously, difficulties increase when generating square-block based numerical meshes to accurately represent the physical situation. This is especially true when the distances between each source are fractions of the initial minimum substrate block size.

The following example illustrates this difficulty. Two $10.1\mu\text{m} \times 2.1\mu\text{m} \times 0.25\mu\text{m}$ heat sources are located symmetrically in the centre of a substrate. Small substrate dimensions are defined, as $64\mu\text{m} \times 64\mu\text{m} \times 16\mu\text{m}$, in order to

illustrate the problems using a shorter simulation time. For each source, depth $d = 0$ and power dissipation $P = 20\text{mW}$. Assume the initial minimum substrate block size is $1\mu\text{m}$ on each side, and the distance between these two sources is $5.4\mu\text{m}$. Figure 5.1 shows the top view of the structure.

Figure 5.1 Top View of a Two-Source Example in a Substrate with Initial Minimum Substrate Blocks of $1\mu\text{m}$ on Each Side

According to the discussions in chapter 3, for this example, in order to precisely represent each source in all dimensions using the above numerical mesh, the initial minimum substrate blocks ($1\mu\text{m}$) need to be subdivided into even smaller uniform blocks ($0.05\mu\text{m}$ or its subdivisions) in the heat generation regions around each source. By setting “depth” equal to 3, 4, 5, 6, ... in the “refine” statement of the input file, which subdivides the initial minimum substrate block by 2^3 , 2^4 , 2^5 , 2^6 , ..., the minimum block size in the

heat generation region is further subdivided into $0.125\mu\text{m}$, $0.0625\mu\text{m}$, $0.03125\mu\text{m}$, $0.015625\mu\text{m}$, ... After subdividing, the minimum block size in the heat generation region is getting closer to $0.05\mu\text{m}$ or its subdivisions. However, the problem lies in finding n such that 2^{-n} is acceptably close to the required resolution. In this example, in order to precisely represent the exact accuracy of $0.05\mu\text{m}$ using the subdivisions, n needs to be very large.

Moreover, even though the source dimensions can be represented by further subdividing the minimum substrate blocks in the heat generation regions around each source, the distance between these two sources ($5.4\mu\text{m}$) still cannot be precisely represented by the $1\mu\text{m}$ numerical grid unless the numerical blocks between these sources are also refined.

For this example, in order to solve these problems, two alternative approaches can be applied. The first approach is to redefine the initial minimum substrate block size as $0.05\mu\text{m}$ on each side, such that both the source dimensions and the distance between the sources can be precisely represented by the minimum substrate block. In this approach, since the initial minimum substrate block size already satisfies the condition for precisely representing the source dimensions and the distance between the sources, no refinement is needed in the heat generation region around the heat sources. However, in order to reduce the number of substrate blocks, more numerical regions need to be defined in the substrate. Thus, the maximum substrate block size can be increased in the region that is far from the heat source. Figure 5.2 shows the mesh structure. Simulation results are listed in Table 5-1.

Figure 5.2 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.05\mu\text{m}$ on Each Side for the Two-Source Example in Figure 5.1 before Quantization
(a) Entire view (b) Partial enlarged view

In the first approach, for this example, six different numerical regions were defined. The maximum block sizes in each region were set to be $0.05\mu\text{m}$, $0.1\mu\text{m}$, $0.2\mu\text{m}$, $0.4\mu\text{m}$, $0.8\mu\text{m}$ and $1.6\mu\text{m}$ on each side, respectively. The maximum substrate block size was set to be $3.2\mu\text{m}$ on each side. Because of the symmetrical structure, simulation was done for a quarter of the entire structure.

The second approach in this example is to redefine the initial minimum substrate block size to be multiples of $0.05\mu\text{m}$ (e.g. $0.8\mu\text{m}$) on each side, such that, after subdividing the initial block ($0.8\mu\text{m}$) by 2^4 , small blocks with a minimum block size of $0.05\mu\text{m}$ on each side can be created in the heat generation regions around each of these heat sources. Thus, the source dimensions could be precisely represented by the subdivisions. However, as discussed in chapter 3, only the numerical blocks in the heat generation region can be further subdivided. However, for this example, in order to precisely represent the distance between these sources, the initial blocks ($0.8\mu\text{m}$) between these sources also need to be subdivided. Therefore, in this approach, the heat generation region needs to be generated to enclose both of these two sources. Thus, by subdividing the initial minimum substrate blocks in the heat generation region into $0.05\mu\text{m}$, both the source dimensions and the distance between these sources can be precisely represented.

Using this approach, since the initial minimum substrate block size is being increased, three different numerical regions were defined in the substrate as the power region, device region and circuit region. Inside the power region, a heat generation region was defined, and the blocks in the heat generation region were subdivided. Figure 5.3 shows the mesh structure. Simulation results are also listed in Table 5-1.

Figure 5.3 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the Two-Source Example in Figure 5.1 before Quantization
(a) Entire view (b) Partial enlarged view

In this second approach, the maximum block sizes in the power region, device region and circuit region were set to be $0.8\mu\text{m}$, $1.6\mu\text{m}$, and $3.2\mu\text{m}$ on each side respectively, and the maximum substrate block was set to be $6.4\mu\text{m}$ on each side. Because of the symmetrical structure, simulation was done for a quarter of the entire structure.

Both of the above two approaches are based on decreasing the sizes of numerical grids in order to ensure solution accuracy. However, the smaller the grid size, the larger the number of numerical grids, and the longer the simulation time, especially when the substrate dimensions and the distances between multiple sources are large. According to the main purpose of this research, that is, to quickly and accurately predict the junction temperature rise, new approaches need to be created.

In this chapter, a new approach, called “source quantization”, is introduced and studied. Instead of decreasing numerical block sizes, the quantization approach is mainly based on quantizing source dimensions in order that the source dimensions can be represented by a minimum number of subdivisions from the initial substrate blocks. In addition, approaches of adjusting the distances between multiple sources are also investigated in order that the distances between sources may be represented by a minimum number of multiples of the initial substrate blocks.

Again, use the same example as shown in Figure 5.1. However, instead of redefining the initial minimum substrate blocks, the source dimensions and distance between sources were adjusted to $10\mu\text{m} \times 2\mu\text{m} \times 0.25\mu\text{m}$ and $6\mu\text{m}$, respectively. After this adjustment, both the source lateral dimensions and the distance between sources could be represented by multiples of the initial minimum substrate blocks ($1\mu\text{m}$). In addition, after

subdividing the initial blocks by 2^2 in the heat generation regions around the sources, the source thickness could be represented by a subdivision of the initial blocks.

Using the quantization method, for this example, three different numerical regions were defined around each source in the substrate as power region, device region and circuit region. The maximum block sizes in the power region, device region and circuit region were set to be $1\mu\text{m}$, $2\mu\text{m}$, and $4\mu\text{m}$ on each side, respectively, and the maximum substrate block size was set to be $8\mu\text{m}$ on each side. Because of this symmetrical structure, simulation was done for a quarter of the entire structure. Figure 5.4 shows the mesh structure. Simulation results are also listed in Table 5-1.

Table 5-1 Comparison of Different Approaches and Temperatures for the Two-Source Example in Figure 5.1

parameter	Atar simulation			tsource simulation
	before quantization		after quantization	
	Figure 5.2	Figure 5.3	Figure 5.4	
initial minimum substrate block size (μm)	$0.05 \times 0.05 \times 0.05$	$0.8 \times 0.8 \times 0.8$	$1 \times 1 \times 1$	
maximum substrate block size (μm)	$3.2 \times 3.2 \times 3.2$	$6.4 \times 6.4 \times 6.4$	$8 \times 8 \times 8$	
dimensions of each heat source in a quarter (μm)	$5.05 \times 2.1 \times 0.25$	$5.05 \times 2.1 \times 0.25$	$5 \times 2 \times 0.25$	
size of subdivisions in the heat generation region (μm)	$0.05 \times 0.05 \times 0.05$	$0.05 \times 0.05 \times 0.05$	$0.25 \times 0.25 \times 0.25$	
number of numerical regions in substrate	6	3	3	
number of blocks	56045	240672	1187	
CPU time (s)	8.5	10.8	0.1	
temperature rise (K)	47.14	47.07	48.03	47.31

Figure 5.4 A Quarter Mesh Structure with Initial Minimum Substrate Block of $1\mu\text{m}$ on Each Side for the Two-Source Example in Figure 5.1 after Quantization
(a) Entire view (b) Partial enlarged view

From the simulation results, it can be seen that, by quantizing the source dimensions and the distance between the sources, the number of numerical blocks is reduced by up to 99.5%, and the CPU time is reduced by 99.1%. Meanwhile, the errors in temperature rise obtained from using the quantization approach compared to those obtained by using tsource simulation and the other two approaches, are within 2%.

Therefore, using the quantization method can efficiently decrease the number of numerical grids and save simulation time. This is because, after quantization, the number of subdivisions in Figure 5.4 is greatly reduced compared to that of Figures 5.2 and 5.3. However, the maximum number of subdivisions in the heat generation region depends on how the source dimensions and the distances between multiple sources are quantized. Hence, quantization rules need to be investigated.

5.2 Single Square Source Quantization

In order to investigate quantization rules, simulations are first done for some simple structures such as single square sources in semi-infinite substrates using the tsource program. Recall that the tsource program is based on the exact analytical method. Thus, the simulation errors caused by constricted substrate boundary dimensions and numerical meshing errors can be neglected. The main purpose here was to characterize the junction temperature rise T_{rise} sensitivity to source length l (or source width w , where $l = w$ for square sources), so as to determine the reasonable quantization range for given single square heat source dimensions.

Hence, simulations were done for different sizes of single square sources. For each case, the depth and thickness of the heat source were kept to be $0\mu\text{m}$ and $0.25\mu\text{m}$, respectively, whereas the source length l (or source width w) was changed by keeping the source width w (or source length l). As an example, Table 5-2 lists the simulation results for a $20\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ heat source and a $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ heat source. In both cases, power dissipation $P = 40\text{mW}$.

In Table 5-2, the quantization ratio is defined as:

$$\text{quantization ratio} = \frac{\text{source length } l \text{ after quantization}}{\text{original source length } l} \quad (5.1)$$

T_{rise} fractional change to source length l is defined as:

$$\begin{aligned} T_{rise} \text{ fractional change to } l &= \frac{\Delta T_{rise} / \text{original } T_{rise}}{\Delta l / \text{original } l} \\ &= \frac{\text{quantized } T_{rise} / \text{original } T_{rise} - 1}{\text{quantization ratio} - 1} \end{aligned} \quad (5.2)$$

where, “quantized T_{rise} ” is T_{rise} at quantized source dimensions, and “original T_{rise} ” is T_{rise} at original source dimensions.

The simulation results listed in Table 5-2 indicate that, at an equal quantization ratio, the T_{rise} fractional change to source length l (or width d) almost stays the same for different sizes of single square sources. For example, when the quantization ratio is 0.85, T_{rise} fractional change to l is -0.55 for both the $20\mu\text{m} \times 20\mu\text{m} \times 0.25\mu\text{m}$ heat source and the $80\mu\text{m} \times 80\mu\text{m} \times 0.25\mu\text{m}$ heat source. This means that the T_{rise} fractional change to l (or w) is almost independent of the exact value of the source dimensions, but is dependent

on quantization ratios. At a smaller quantization ratio, the absolute value of T_{rise} fractional change to l is higher. This can be explained by the conductive heat transfer equation presented in chapter 4.

**Table 5-2 Single Square Source
 T_{rise} Fractional Change to l (or d) vs. Quantization Ratios**

lateral source dimensions (μm)	20×17	20×18	20×19	20×20	20×21	20×22	20×23
T_{rise} (K)	27.24	26.51	25.82	25.18	24.57	24.00	23.46
quantization ratio	0.85	0.90	0.95	1.00	1.05	1.10	1.15
quantized T_{rise} / original T_{rise}	1.082	1.053	1.025	1.000	0.976	0.953	0.932
T_{rise} fractional change to l (or w)	-0.55	-0.53	-0.50	-	-0.48	-0.47	-0.45
lateral source dimensions (μm)	80×68	80×72	80×76	80×80	80×84	80×88	80×92
T_{rise} (K)	6.83	6.65	6.47	6.31	6.16	6.01	5.87
quantization ratio	0.85	0.90	0.95	1.00	1.05	1.10	1.15
quantized T_{rise} / original T_{rise}	1.082	1.054	1.025	1.000	0.976	0.953	0.930
T_{rise} fractional change to l (or w)	-0.55	-0.54	-0.50	-	-0.48	-0.47	-0.46

 original source dimensions and T_{rise}

As discussed in chapter 4, from the Fourier equations, the temperature difference ΔT (or T_{rise}) between the junction and the base of the substrate can be written in the form:

$$\Delta T = \frac{t_{sub}}{kA} q \quad (5.3)$$

where, A is the cross-sectional area normal to the heat flow path in substrate, and changes along the substrate thickness direction due to 3D heat spreading; t_{sub} is the substrate thick-

ness, k is thermal conductivity, and q is heat generated per unit volume (i.e. average power dissipation).

Assuming that heat spreads freely in the substrate with a constant spreading angle of 45° , then, at d_{sub} , the cross-sectional area A can be calculated as:

$$A = (w + 2d_{sub})(l + 2d_{sub}) \quad (5.4)$$

where, w is source width, l is source length, d_{sub} is the distance between the cross-sectional area and the substrate surface, and d_{sub} is not too large.

Let l be the original source length and l_1 and l_2 the quantized source lengths.

Assume $\frac{l_1}{l} = \alpha$, $\frac{l_2}{l} = \beta$, and $1 < \alpha < \beta$ (i.e. quantization ratio $1 < \frac{l_1}{l} < \frac{l_2}{l}$). Then, the

original T_{rise} before quantization can be calculated as:

$$\Delta T = \frac{t_{sub}}{kA} q = \frac{t_{sub}}{kA} \cdot \frac{P}{wt} \quad (5.5)$$

where, P is power dissipation, t is source thickness, w is source width ($w = l$ for square sources). Assume t and w stay the same while l changes.

Similarly, the T_{rise} at quantized source length l_1 and l_2 can be calculated as:

$$\Delta T_1 = \frac{t_{sub}}{kA_1} q_1 = \frac{t_{sub}}{kA_1} \cdot \frac{P}{wl_1t} \quad (5.6)$$

$$\Delta T_2 = \frac{t_{sub}}{kA_2} q_2 = \frac{t_{sub}}{kA_2} \cdot \frac{P}{wl_2t} \quad (5.7)$$

where, $A_1 = (w_1 + 2d_{sub})(l_1 + 2d_{sub}) = (w + 2d_{sub})(\alpha l + 2d_{sub})$ (5.8)

$A_2 = (w_2 + 2d_{sub})(l_2 + 2d_{sub}) = (w + 2d_{sub})(\beta l + 2d_{sub})$ (5.9)

Thus, according to Equation (5.2), the T_{rise} fractional change to l at a quantization ratio of $\frac{l_1}{l}$ can be calculated as:

$$T_{rise} \text{ fractional change to } l = \frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l} - 1} = \frac{\frac{A}{A_1 \alpha} - 1}{\alpha - 1} < 0$$

In addition, the T_{rise} fractional change to l at a quantization ratio of $\frac{l_2}{l}$ can be calculated as:

$$T_{rise} \text{ sensitivity to } l = \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{l_2}{l} - 1} = \frac{\frac{A}{A_2 \beta} - 1}{\beta - 1} < 0$$

Hence,

$$\begin{aligned} & \frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l} - 1} - \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{l_2}{l} - 1} \\ &= \frac{\frac{A}{A_1 \alpha} - 1}{\alpha - 1} - \frac{\frac{A}{A_2 \beta} - 1}{\beta - 1} \\ &= \frac{(A - A_1 \alpha) A_2 \beta (\beta - 1) - (A - A_2 \beta) A_1 \alpha (\alpha - 1)}{A_1 \alpha (\alpha - 1) A_2 \beta (\beta - 1)} \end{aligned}$$

$$\text{Let } M = (A - A_1\alpha)A_2\beta(\beta - 1) - (A - A_2\beta)A_1\alpha(\alpha - 1)$$

$$N = A_1\alpha(\alpha - 1)A_2\beta(\beta - 1)$$

Since $\alpha > 1$ and $\beta > 1$, therefore, $N = A_1\alpha(\alpha - 1)A_2\beta(\beta - 1) > 0$

$$\begin{aligned} \text{Meanwhile, } M &= (A - A_1\alpha)A_2\beta(\beta - 1) - (A - A_2\beta)A_1\alpha(\alpha - 1) \\ &= AA_2\beta(\beta - 1) - A_1\alpha[A(\alpha - 1) + A_2\beta(\beta - \alpha)] \end{aligned} \quad (5.10)$$

According to the discussion in chapter 4, at the same position along the semi-infinite substrate thickness direction, the cross-sectional area normal to the heat flow path is larger for a bigger heat source than that for a smaller heat source. Therefore, $A < A_1 < A_2$.

Inserting Equations (5.4), (5.8) and (5.9) into Equation (5.10), M can be calculated as:

$$\begin{aligned} M &= AA_2\beta(\beta - 1) - A_1\alpha[A(\alpha - 1) + A_2\beta(\beta - \alpha)] \\ &= (w + 2d_{sub})^2 \{ 4d_{sub}^2(1 - \alpha)(\beta - 1)(\beta - \alpha) \\ &\quad + 2ld_{sub}(\alpha - \beta)(\alpha - 1)(\beta - 1)(1 + \alpha + \beta) \\ &\quad + l^2(\alpha - \beta)(\alpha - 1)(\beta - 1)(\alpha\beta + \alpha + \beta) \} \end{aligned}$$

Since $1 < \alpha < \beta$, therefore, $M < 0$

$$\text{Thus, } \frac{\frac{\Delta T_1}{l_1} - 1}{\frac{l_1}{l} - 1} - \frac{\frac{\Delta T_2}{l_2} - 1}{\frac{l_2}{l} - 1} = \frac{M}{N} < 0$$

Meanwhile, because $\frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l} - 1} < 0$ and $\frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{l_2}{l} - 1} < 0$,

$$\text{hence, } \left| \frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l} - 1} \right| > \left| \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{l_2}{l} - 1} \right|$$

This means that at a smaller quantization ratio, the absolute value of T_{rise} fractional change to l is higher.

Figure 5.5 shows the plots of quantized T_{rise} /original T_{rise} vs. quantization ratio of l for square sources. Different sizes of single square sources are compared.

According to the plots in Figure 5.5, the functions of quantized T_{rise} /original T_{rise} vs. quantization ratio of l are almost the same for different sizes of single square sources, and they are close to linear. At each quantization ratio, the derivative of the curve represents T_{rise} sensitivity to source length l . In this case, it is around -0.5. When the quantization ratio changes from 0.85 to 1.15, the change of T_{rise} sensitivity to l is small. Hence, an approximation can be made that, for single square sources ($l \geq 5\mu\text{m}$), when quantization ratio is between 0.85 and 1.15, the T_{rise} sensitivity to l is about -0.5 despite of the source size. In addition, the T_{rise} sensitivity to l is approximated as:

$$\begin{aligned}
T_{rise} \text{ sensitivity to } l &\approx T_{rise} \text{ fractional change to } l \\
&= \frac{\Delta T_{rise} / \text{original } T_{rise}}{\Delta l / \text{original}} \\
&= \frac{\text{quantized } T_{rise} / \text{original } T_{rise} - 1}{\text{quantization ratio} - 1}
\end{aligned} \tag{5.11}$$

Figure 5.5 Single Square Source Quantization

The above approximation indicates that, for a single square source ($l \geq 5\mu\text{m}$), when its length l (or width w) is changed by $\pm 10\%$, the T_{rise} error caused by the quantization of source length l (or width w) is about $\mp 5\%$. This can be verified by the simulation results listed in Table 5-2. For example, when the length (width) of the $20\mu\text{m} \times 20\mu\text{m}$ square source is reduced by 5% (which is $20\mu\text{m} \times 19\mu\text{m}$), the T_{rise} error is 2.5% (i.e. quantized $T_{rise} / \text{original } T_{rise} = 1.025$). When its length (width) is increased by 5%

(which is $20\mu\text{m} \times 21\mu\text{m}$), the T_{rise} error is -2.4% (i.e. quantized T_{rise} / original $T_{rise} = 0.976$). In addition, for the $80\mu\text{m} \times 80\mu\text{m}$ square source, when its length (width) is reduced by 5% (which is $80\mu\text{m} \times 76\mu\text{m}$), the T_{rise} error is 2.5% (i.e. quantized T_{rise} / original $T_{rise} = 1.025$). When its length (width) is increased by 5% (which is $80\mu\text{m} \times 84\mu\text{m}$), the T_{rise} error is -2.4% (i.e. quantized T_{rise} / original $T_{rise} = 0.976$). Therefore, in order to keep the T_{rise} error within $\pm 5\%$ for single square sources with $l \geq 5\mu\text{m}$, the quantization range of source length l (or width w) should be within $\mp 10\%$.

5.3 Single Rectangular Source Quantization

Similar simulations and analyses were done for single rectangular sources with different sizes and aspect ratios l/w , where, l is source length and w is source width ($l > w$). Again, for each case, the depth and thickness of the heat source were maintained at $0\mu\text{m}$ and $0.25\mu\text{m}$, respectively, whereas the source length l (or source width w) was changed by keeping the source width w (or source length l). Figure 5.6 shows the plots of quantized T_{rise} / original T_{rise} vs. quantization ratio of l for different sources. Meanwhile, Figure 5.7 shows the plots of quantized T_{rise} / original T_{rise} vs. quantization ratio of w for different sources. In addition, the functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of l (and w) for sources with the same aspect ratios but different sizes were also compared.

Figure 5.6 Single Rectangular Source Quantization of Length l

Figure 5.7 Single Rectangular Source Quantization of Width w

Again, these plots further demonstrate that, at a certain aspect ratio l/w (e.g. $l/w = 5$), the functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of l (and w) are almost the same for different sizes of single rectangular sources (e.g. the $20\mu\text{m} \times 4\mu\text{m}$ heat source and the $60\mu\text{m} \times 12\mu\text{m}$ heat source), and they are close to linear. This means that the T_{rise} sensitivity to l and w is almost independent of the exact value of source dimensions, but is dependent on the aspect ratio l/w .

Moreover, according to the plots, the curves are close to linear, which indicates that, at different quantization ratios, the change of T_{rise} sensitivity to l and w is small. Hence, another approximation is made that, for sources with the same aspect ratio l/w , T_{rise} sensitivities to source length l (or source width w) are constant.

Furthermore, these plots in Figures 5.6 and 5.7 also show that when aspect ratio l/w increases, the absolute value of T_{rise} sensitivity to l (i.e. slopes of curves) increases, whereas, the absolute value of T_{rise} sensitivity to w decreases. Meanwhile, for rectangular sources, the absolute value of T_{rise} sensitivity to source length l is higher than that to source width w . This can also be explained by the conductive heat transfer equation.

Assume l_0 is the original rectangular source length, and w_0 is the original rectangular source width, $l_0 > w_0$. Let l_1 be the source length after quantization ($w_1 = w_0$), w_2 be the source width after quantization ($l_2 = l_0$), and $\frac{l_1}{l_0} = \frac{w_2}{w_0} = \alpha > 1$ (i.e. same quantization ratio). According to the discussion in section 5.2,

$$T_{rise} \text{ sensitivity to } l = \frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l_0} - 1} = \frac{\frac{A_0 l_0}{A_1 l_1} - 1}{\frac{l_1}{l_0} - 1} < 0$$

$$T_{rise} \text{ sensitivity to } w = \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{w_2}{w_0} - 1} = \frac{\frac{A_0 w_0}{A_2 w_2} - 1}{\frac{w_2}{w_0} - 1} < 0$$

$$\frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l_0} - 1} - \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{w_2}{w_0} - 1} = \frac{\frac{A_0}{\alpha} \left(\frac{1}{A_1} - \frac{1}{A_2} \right)}{\alpha - 1} = \frac{A_0}{\alpha(\alpha - 1)} \cdot \frac{A_2 - A_1}{A_1 A_2}$$

Since $\alpha > 1$,

$$\begin{aligned} A_2 - A_1 &= (l_0 + 2d_{sub})(\alpha w_0 + 2d_{sub}) - (\alpha l_0 + 2d_{sub})(w_0 + 2d_{sub}) \\ &= 2d_{sub}(1 - \alpha)(l_0 - w_0) < 0 \end{aligned}$$

$$\text{Thus, } \frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l_0} - 1} - \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{w_2}{w_0} - 1} = \frac{\frac{A_0}{\alpha} \left(\frac{1}{A_1} - \frac{1}{A_2} \right)}{\alpha - 1} = \frac{A_0}{\alpha(\alpha - 1)} \cdot \frac{A_2 - A_1}{A_1 A_2} < 0$$

$$\left| \frac{\frac{\Delta T_1}{\Delta T} - 1}{\frac{l_1}{l_0} - 1} \right| > \left| \frac{\frac{\Delta T_2}{\Delta T} - 1}{\frac{w_2}{w_0} - 1} \right|$$

This means that for rectangular sources, the absolute value of T_{rise} sensitivity to source length l is higher than that to source width w .

Furthermore, assume there are two rectangular sources with aspect ratios

$\frac{l_{10}}{w_{10}} = \beta_1$ and $\frac{l_{20}}{w_{20}} = \beta_2$, where, l_{10} and w_{10} are the original length and width of heat

source 1, l_{20} and w_{20} are the original length and width of heat source 2, and $\beta_1 > \beta_2 > 1$.

Let l_{11} be the length of source 1 after quantization ($w_{11} = w_{10}$), l_{21} be the length

of source 2 after quantization ($w_{21} = w_{20}$), and $\frac{l_{11}}{l_{10}} = \frac{l_{21}}{l_{20}} = \alpha_1 > 1$ (i.e. same quantiza-

tion ratio of l). According to the discussion in section 5.2, the T_{rise} sensitivity to source length l can be calculated as follows:

$$\text{for heat source 1: } T_{rise} \text{ sensitivity to } l = \frac{\frac{\Delta T_{11}}{\Delta T_{10}} - 1}{\frac{l_{11}}{l_{10}} - 1} = \frac{\frac{A_{10}l_{10}}{A_{11}l_{11}} - 1}{\frac{l_{11}}{l_{10}} - 1} = \frac{\frac{A_{10}}{A_{11}\alpha_1} - 1}{\alpha_1 - 1} < 0$$

$$\text{for heat source 2: } T_{rise} \text{ sensitivity to } l = \frac{\frac{\Delta T_{21}}{\Delta T_{20}} - 1}{\frac{l_{21}}{l_{20}} - 1} = \frac{\frac{A_{20}l_{20}}{A_{21}l_{21}} - 1}{\frac{l_{21}}{l_{20}} - 1} = \frac{\frac{A_{20}}{A_{21}\alpha_1} - 1}{\alpha_1 - 1} < 0$$

$$\begin{aligned}
& \frac{\frac{\Delta T_{11}}{\Delta T_{10}} - 1}{\frac{l_{11}}{l_{10}} - 1} - \frac{\frac{\Delta T_{21}}{\Delta T_{20}} - 1}{\frac{l_{21}}{l_{20}} - 1} \\
&= \frac{\frac{A_{10}}{A_{11}} - \frac{A_{20}}{A_{21}}}{\alpha_1(\alpha_1 - 1)} \\
&= \frac{\frac{(w_{10} + 2d_{sub})(l_{10} + 2d_{sub})}{(w_{10} + 2d_{sub})(\alpha_1 l_{10} + 2d_{sub})} - \frac{(w_{20} + 2d_{sub})(l_{20} + 2d_{sub})}{(w_{20} + 2d_{sub})(\alpha_1 l_{20} + 2d_{sub})}}{\alpha_1(\alpha_1 - 1)} \\
&= \frac{2(\alpha_1 - 1)d_{sub}(l_{20} - l_{10})}{\alpha_1(\alpha_1 - 1)(\alpha_1 l_{10} + 2d_{sub})(\alpha_1 l_{20} + 2d_{sub})}
\end{aligned}$$

Since $\frac{l_{10}}{w_{10}} > \frac{l_{20}}{w_{20}}$, assuming $w_{10} = w_{20}$, then $l_{10} > l_{20}$.

Therefore,
$$\frac{\frac{\Delta T_{11}}{\Delta T_{10}} - 1}{\frac{l_{11}}{l_{10}} - 1} - \frac{\frac{\Delta T_{21}}{\Delta T_{20}} - 1}{\frac{l_{21}}{l_{20}} - 1} < 0$$

$$\left| \frac{\frac{\Delta T_{11}}{\Delta T_{10}} - 1}{\frac{l_{11}}{l_{10}} - 1} \right| > \left| \frac{\frac{\Delta T_{21}}{\Delta T_{20}} - 1}{\frac{l_{21}}{l_{20}} - 1} \right|$$

This indicates that at a larger aspect ratio l/w , the absolute value of T_{rise} sensitivity to l is higher than that at a smaller aspect ratio.

Similarly, let w_{12} be the width of source 1 after quantization ($l_{12} = l_{10}$), w_{22} be the width of source 2 after quantization ($l_{22} = l_{20}$), and $\frac{w_{12}}{w_{10}} = \frac{w_{22}}{w_{20}} = \alpha_2 > 1$ (i.e.

same quantization ratio of w). Thus, the T_{rise} sensitivity to source width w can be calculated as:

$$\text{for heat source 1: } T_{rise} \text{ sensitivity to } w = \frac{\frac{\Delta T_{12}}{\Delta T_{10}} - 1}{\frac{w_{12}}{w_{10}} - 1} = \frac{\frac{A_{10}w_{10}}{A_{12}w_{12}} - 1}{\frac{w_{12}}{w_{10}} - 1} = \frac{\frac{A_{10}}{A_{12}\alpha_2} - 1}{\alpha_2 - 1} < 0$$

$$\text{for heat source 2: } T_{rise} \text{ sensitivity to } w = \frac{\frac{\Delta T_{22}}{\Delta T_{20}} - 1}{\frac{w_{22}}{w_{20}} - 1} = \frac{\frac{A_{20}w_{20}}{A_{22}w_{22}} - 1}{\frac{w_{22}}{w_{20}} - 1} = \frac{\frac{A_{20}}{A_{22}\alpha_2} - 1}{\alpha_2 - 1} < 0$$

$$\begin{aligned} & \frac{\frac{\Delta T_{12}}{\Delta T_{10}} - 1}{\frac{w_{12}}{w_{10}} - 1} - \frac{\frac{\Delta T_{22}}{\Delta T_{20}} - 1}{\frac{w_{22}}{w_{20}} - 1} \\ &= \frac{\frac{A_{10}}{A_{12}} - \frac{A_{20}}{A_{22}}}{\alpha_2(\alpha_2 - 1)} \\ &= \frac{\frac{(w_{10} + 2d_{sub})(l_{10} + 2d_{sub})}{(\alpha_2 w_{10} + 2d_{sub})(l_{10} + 2d_{sub})} - \frac{(w_{20} + 2d_{sub})(l_{20} + 2d_{sub})}{(\alpha_2 w_{20} + 2d_{sub})(l_{20} + 2d_{sub})}}{\alpha_2(\alpha_2 - 1)} \\ &= \frac{2(\alpha_2 - 1)d_{sub}(w_{20} - w_{10})}{\alpha_2(\alpha_2 - 1)(\alpha_2 w_{10} + 2d_{sub})(\alpha_2 w_{20} + 2d_{sub})} \end{aligned}$$

Since $\frac{l_{10}}{w_{10}} > \frac{l_{20}}{w_{20}}$, assuming $l_{10} = l_{20}$, then $w_{20} > w_{10}$.

$$\text{Therefore, } \frac{\frac{\Delta T_{12}}{\Delta T_{10}} - 1}{\frac{w_{12}}{w_{10}} - 1} - \frac{\frac{\Delta T_{22}}{\Delta T_{20}} - 1}{\frac{w_{22}}{w_{20}} - 1} > 0$$

$$\left| \frac{\frac{\Delta T_{12}}{\Delta T_{10}} - 1}{\frac{w_{12}}{w_{10}} - 1} \right| < \left| \frac{\frac{\Delta T_{22}}{\Delta T_{20}} - 1}{\frac{w_{22}}{w_{20}} - 1} \right|$$

This indicates that at a larger aspect ratio l/w , the absolute value of T_{rise} sensitivity to w is lower than that at a smaller aspect ratio.

According to Figures 5.6 and 5.7, and the above discussions, it can be seen that the T_{rise} sensitivities to l (and w) are dependent on source aspect ratios (l/w), but almost independent of the quantization ratios when quantization ratio are within $\pm 15\%$. As referred to in Figures 5.6 and 5.7, the functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of l (and w) are close to linear (i.e. the slope of the curve at different quantization ratios is almost constant). Therefore, the next step is to find an approximate relationship between the slope (i.e. T_{rise} sensitivity to source length l or source width w) and the aspect ratio.

5.4 T_{rise} Sensitivities and Aspect Ratios (l/w)

In order to find out T_{rise} sensitivity to l (or w) as a function of aspect ratio (l/w), a curve fitting technology is introduced using Matlab.

5.4.1 Curve Fitting

As discussed in section 5.3, in order to find out the constant value of T_{rise} sensitivities to source length (l) and source width (w) at different aspect ratios l/w , straight line

(linear) approximations are made for the simulation data by using Matlab built-in function “polyfit”. Figures 5.8 and 5.9 show the curve fitting results for the functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of l and w , respectively. Moreover, by using Matlab, it is not difficult to determine the slopes of the curves which represent T_{rise} sensitivity to l and w at different aspect ratios. They are also annotated in Figures 5.8 and 5.9. The simulation data in Figures 5.8 and 5.9 are based on the simulation results for sources with length of $20\mu\text{m}$ and various aspect ratios, from 1 to 40.

5.4.2 T_{rise} Sensitivity vs. Aspect Ratios

The relationship between aspect ratios and T_{rise} sensitivities are studied in this section based on the plots in Figures 5.8 and 5.9 and the approximation that, for sources with the same aspect ratio l/w , T_{rise} sensitivity to source length l (or source width w) is constant.

Using the slopes (i.e. T_{rise} sensitivities) found in Figures 5.8 and 5.9, Figures 5.10 and 5.11 show T_{rise} sensitivity to l and w as a function of aspect ratios (l/w). In both cases, since l is defined as source length, w is defined as source width, and $l \geq w$, only aspect ratio $l/w \geq 1$ needs to be considered.

Again, these plots in Figures 5.10 and 5.11 are based on sources with length $l = 20\mu\text{m}$ and aspect ratios from 1 to 40.

Figure 5.8 Straight Line Fit for the Functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of l

Figure 5.9 Straight Line Fit for the Functions of quantized T_{rise} / original T_{rise} vs. quantization ratio of w

Figure 5.10 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with Length $l = 20\mu\text{m}$ and Different Aspect Ratios

Figure 5.11 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with Length $l = 20\mu\text{m}$ and Different Aspect Ratios

5.5 Evaluation of Quantization Ranges in l and w

Given source length l and source width w , in order to find an acceptable quantization range in l and w , one first finds the aspect ratio l/w . Then, one finds the approximate T_{rise} sensitivity to l and w from the plots of T_{rise} sensitivity to l and w vs. aspect ratio (l/w). Finally, the quantization range in l and w could be evaluated using Equation (5.11).

For example, a given source has a size of $20\mu\text{m} \times 4\mu\text{m} \times 0.25\mu\text{m}$ with a depth of $0\mu\text{m}$. Therefore, its aspect ratio l/w can be calculated as 5. According to the plot in Figure 5.10, when the aspect ratio is 5, T_{rise} sensitivity to source length l is around -0.7 , meaning that if the source length is increased by 10% (i.e. quantization ratio of l is 1.1), T_{rise} decreases by 7% (i.e. quantized T_{rise} / original T_{rise} is 0.93). This outcome can be verified in Figure 5.6. In other words, if the change in T_{rise} for this example needs to be within $\pm 5\%$, then the change in l should be within $\mp 7\%$.

Similarly for this same example, according to the plot in Figure 5.11, when the aspect ratio is 5, T_{rise} sensitivity to source length w is approximately -0.28 . Therefore, if the source width is increased by 10% (i.e. quantization ratio of w is 1.1), T_{rise} decreases by 2.8% (i.e. quantized T_{rise} / original T_{rise} is 0.972). This can also be verified in Figure 5.7. In other words, for this example, if the change in T_{rise} needs to be within $\pm 5\%$, then the change in w should be within $\mp 17.9\%$.

5.5.1 Problems with Small Source Width and Aspect Ratio (l/w)

Practically speaking, source width is usually within a range of $0.5\mu\text{m}$ to $2\mu\text{m}$, whereas source length varies with different aspect ratios (l/w). In order to characterize the universal function of T_{rise} sensitivity to source length l (and source width w) vs. aspect ratio, more simulations are done. Using the same approaches as introduced in Figures 5.8, 5.9, 5.10 and 5.11, Figure 5.12 shows the plots of T_{rise} sensitivity to l vs. aspect ratio (l/w) for sources with various widths, from $0.5\mu\text{m}$ to $10\mu\text{m}$, and various aspect ratio, from 1 to 40. Figure 5.13 shows T_{rise} sensitivity to w vs. aspect ratio (l/w). Again, for each source, depth $d = 0$, and the source thickness is $0.25\mu\text{m}$. It is found that at the same aspect ratio, for sources with small dimensions (i.e. small width and aspect ratio), the T_{rise} sensitivity changes greatly with source dimensions. Thus, the functions of T_{rise} sensitivity to l and w vs. aspect ratio as plotted in Figures 5.10 and 5.11 may not be suitable for precisely representing a situation involving small sources.

For example, according to the plots in Figure 5.12, when aspect ratio is 2 and source width is $0.5\mu\text{m}$ (i.e. a $1\mu\text{m} \times 0.5\mu\text{m}$ source), the T_{rise} sensitivity to l is about -0.49. When aspect ratio is 2 and source width is 1 (i.e. a $2\mu\text{m} \times 1\mu\text{m}$ source), the T_{rise} sensitivity to l is about -0.54. When aspect ratio is 2 and source width is 10 (i.e. a $20\mu\text{m} \times 10\mu\text{m}$ source), the T_{rise} sensitivity to l is about -0.59. This indicates that at the same aspect ratio of 2, the error of T_{rise} sensitivity to l is almost 17% for a $1\mu\text{m} \times 0.5\mu\text{m}$ source compared to a $20\mu\text{m} \times 10\mu\text{m}$ source. However, at the aspect ratio of 2, when source width is larger than $2\mu\text{m}$, the functions of T_{rise} sensitivity to l vs. aspect ratio (l/w) are almost the same.

Figure 5.12 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with Different Width and Aspect Ratios

Figure 5.13 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with Different Width and Aspect Ratios

In addition, the plots in Figures 5.12 and 5.13 illustrate that, at larger aspect ratios, the T_{rise} sensitivities to l (or w) change only slightly for sources with smaller or larger widths.

Therefore, based on the practical situation that the source width is usually within range $0.5\mu\text{m}$ to $2\mu\text{m}$, it would be useful to find a universal function of T_{rise} sensitivity to l (or w). It is feasible to choose the function of T_{rise} sensitivity to l (or w) vs. aspect ratio when source width is $1\mu\text{m}$ as the universal function. Thus, by using this universal function, for sources with various width from $0.5\mu\text{m}$ to $2\mu\text{m}$, and aspect ratios from 1 to 40, the errors of T_{rise} sensitivity to l (or w) vs. aspect ratio are within 10%. Figure 5.14 shows the universal function of T_{rise} sensitivity to l and w vs. aspect ratios based on the sources with a width of $1\mu\text{m}$.

Figure 5.14 Universal Functions of T_{rise} Sensitivity to l and w vs. Aspect Ratios with $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$

According to the definition of T_{rise} sensitivity in Equation (5.11), when the source length l or width w are changed by 10%, a 10% error in T_{rise} sensitivity can cause a 1% error in T_{rise} , which is acceptably small. Therefore, the universal functions of T_{rise} sensitivity to l and w vs. aspect ratios found from sources with a width of $1\mu\text{m}$, can be used to predict the change of T_{rise} and determine the quantization ratio of l and w for sources with widths from $0.5\mu\text{m}$ to $2\mu\text{m}$, and aspect ratio from 1 to 40.

As an example, a given heat source has a size of $1\mu\text{m} \times 0.5\mu\text{m} \times 0.25\mu\text{m}$. Its depth is $0\mu\text{m}$. Power dissipation is $P = 40\text{mW}$. Using tsource simulation, when the substrate is semi-infinite, the original $T_{rise} = 518.9087\text{K}$. According to the universal function of T_{rise} sensitivity to l vs. aspect ratio as plotted in Figure 5.14, it is found that for an aspect ratio of 2, T_{rise} sensitivity to l is -0.54. Therefore, when quantizing the source length by $\pm 10\%$ while maintaining the source width and thickness (producing source lengths of $0.9\mu\text{m}$ and $1.1\mu\text{m}$, respectively), T_{rise} can be predicted.

$$\text{According to Equation (5.11), } -0.54 = \frac{\frac{T_{0.9}}{T_{1.0}} - 1}{\frac{0.9}{1.0} - 1}, \text{ where } T_{0.9} \text{ is } T_{rise} \text{ when}$$

$l = 0.9\mu\text{m}$, and $T_{1.0}$ is T_{rise} when $l = 1.0\mu\text{m}$.

$$\text{Therefore, } T_{0.9} = (1 + 0.054) \times 518.9087 = 546.93\text{K}.$$

$$\text{Similarly, } T_{1.1} = (1 - 0.054) \times 518.9087 = 490.89\text{K}.$$

Whereas, the tsource simulation for this example indicates that, $T_{0.9} = 545.33\text{K}$, $T_{1.1} = 495.16\text{K}$. Table 5-3 lists the comparison of the T_{rise} obtained from tsource simulation and predicted using the universal function of T_{rise} sensitivities vs. aspect ratios as plotted in Figure 5.14. The T_{rise} for sources with different width and aspect ratios were simulated and predicted. The errors between the simulation results and predicted results are within 1%.

Table 5-3 Comparison of T_{rise} Predicted and T_{rise} Simulated
(source depth $d = 0$, source thickness $t = 0.25\mu\text{m}$)

source dimensions (μm)		T_{rise} from tsource simulation (K)	T_{rise} prediction (K)	error
quantization of l	$1.0 \times 0.5 \times 0.25$	518.91	—	0
	$0.9 \times 0.5 \times 0.25$	545.34	546.93	0.29%
	$1.1 \times 0.5 \times 0.25$	495.16	490.89	0.86%
	$4.0 \times 2 \times 0.25$	160.70	—	0
	$3.6 \times 2 \times 0.25$	170.46	169.37	0.64%
	$4.4 \times 2 \times 0.25$	152.12	152.02	0.07%
quantization of w	$20 \times 0.5 \times 0.25$	68.70	—	0
	$20 \times 0.45 \times 0.25$	69.44	69.66	0.32%
	$20 \times 0.55 \times 0.25$	67.99	67.74	0.37%
	$80 \times 2 \times 0.25$	18.75	—	0
	$80 \times 1.8 \times 0.25$	19.06	19.01	0.26%
	$80 \times 2.2 \times 0.25$	18.46	18.49	0.16%

The comparison results listed in Table 5-3 indicate that the T_{rise} predicted from the approximate universal functions of T_{rise} sensitivity to l (and w) vs. aspect ratio is pretty close to the simulation result. The temperature error caused by the T_{rise} sensitivity error is small. Therefore, the universal functions of T_{rise} sensitivity to l and w vs. aspect ratios found from sources with width of $1\mu\text{m}$, can be used to represent the T_{rise} sensitivities for sources of width from $0.5\mu\text{m}$ to $2\mu\text{m}$ and aspect ratio from 1 to 40. The universal function can be used to evaluate quantization ratios of l and w , and to predict T_{rise} .

5.6 Quantization in Source Depth (d) and Thickness (t)

All the above discussions were based on fixed source depth and thickness which were set to be $0\mu\text{m}$ and $0.25\mu\text{m}$ respectively, and only l and w have been considered. But the depth and thickness can change between different families of devices. Therefore, depth and thickness need to be taken into account as well. In this section, the effects of source depth and thickness on T_{rise} sensitivities to source length (l) and source width (w) are investigated, and the relations between source depth and source thickness are studied.

5.6.1 Effect of d and t on T_{rise} Sensitivities to l and w

In order to investigate the effects of source depth d and thickness t on T_{rise} sensitivities to l and w , similar simulations were done for single square and rectangular sources with various width from $0.5\mu\text{m}$ to $2\mu\text{m}$, and aspect ratios l/w from 1 to 40, as discussed in section 5.4. The only changes were the source depth and source thickness, which were $0.1\mu\text{m}$, respectively. Figure 5.15 shows the plots of T_{rise} sensitivity to l vs. aspect ratio

for sources with different depths and thicknesses, while Figure 5.16 shows the plots of T_{rise} sensitivity to w for sources with different depths and thicknesses.

The plots in Figures 5.15 and 5.16 indicate that, if the source depth and thickness are small, approximating transistor structures, the changes in source depth and thickness have little effect on the function of T_{rise} sensitivity to l and w vs. aspect ratios. The universal function of T_{rise} sensitivity to l and w found at $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$, can still be applied. Table 5-4 compares T_{rise} obtained from both the tsource simulation and the predicted value.

Table 5-4 Comparison of T_{rise} Predicted and T_{rise} Simulated
(source depth $d = 0.1\mu\text{m}$, source thickness $t = 0.1\mu\text{m}$)

source dimensions (μm)		T_{rise} from tsource simulation (K)	T_{rise} prediction (K)	error
quantization of l	$1.0 \times 0.5 \times 0.25$	500.99	—	0
	$0.9 \times 0.5 \times 0.25$	526.56	528.04	0.28%
	$1.1 \times 0.5 \times 0.25$	478.12	473.94	0.87%
	$4.0 \times 2 \times 0.25$	158.35	—	0
	$3.6 \times 2 \times 0.25$	167.87	166.90	0.12%
	$4.4 \times 2 \times 0.25$	149.98	149.80	0.12%
quantization of w	$20 \times 0.5 \times 0.25$	67.60	—	0
	$20 \times 0.45 \times 0.25$	68.31	68.55	0.35%
	$20 \times 0.55 \times 0.25$	66.92	66.65	0.40%
	$80 \times 2 \times 0.25$	18.63	—	0
	$80 \times 1.8 \times 0.25$	18.93	18.89	0.21%
	$80 \times 2.2 \times 0.25$	18.35	18.37	0.11%

Figure 5.15 Effect of Source Depth and Thickness on T_{rise} Sensitivity to Source Length l

Figure 5.16 Effect of Source Depth and Thickness on T_{rise} Sensitivity to Source Width w

5.6.2 Relations Between d and t and Their Effects on T_{rise}

The simulation results further indicate that when source depth is increased, T_{rise} decreases. When source thickness is reduced, T_{rise} increases. In addition, it is interesting to see that the change in T_{rise} due to the change in source depth is double that resulting from a change in source thickness. As an example, Table 5-5 lists the t_{source} simulation results for a $20\mu\text{m} \times 20\mu\text{m}$ heat source with different depths and thicknesses. Power dissipation is $P = 40\text{mW}$. The substrate was set to be semi-infinite.

Table 5-5 Effect of Source Depth and Thickness on T_{rise} (1)

depth (μm)	<input type="text" value="0"/>	0.1	0	0	0.1
thickness (μm)	<input type="text" value="0.25"/>	0.25	0.15	0.05	0.05
T_{rise} (K)	<input type="text" value="25.1809"/>	24.9582	25.2931	25.4057	25.1812
quantized T_{rise} /original T_{rise}	1	0.9911	1.0045	1.0089	1

original data

In this example, when the depth changes from 0 to $0.1\mu\text{m}$ (which is 40% of the thickness) and the thickness stays the same, T_{rise} decreases by 0.89%. However, when the depth remains at 0 and the thickness changes from $0.25\mu\text{m}$ to $0.15\mu\text{m}$ (a 40% reduction), T_{rise} increases by 0.45%. If the thickness changes from $0.25\mu\text{m}$ to $0.05\mu\text{m}$ (reduces by 80%), T_{rise} increases by 0.89%. The far right column shows that when the depth increases by $0.1\mu\text{m}$ and the thickness reduces by $0.2\mu\text{m}$, T_{rise} is almost the same as the original T_{rise} at depth = 0 and thickness = $0.25\mu\text{m}$.

This result implies that if the source depth is increased by “ d ”, then its thickness needs to be decreased by “ $2d$ ” to maintain the same T_{rise} as original. In other words, for a

source with depth “ d ” and thickness “ t ”, the source may be able to be simulated with depth = 0 and thickness = $2d + t$. This can greatly increase the grid size and simplify the mesh structure in the heat generation region when using the Atar simulator.

It should be pointed out that the above relations may only hold true for sources with small depths and thicknesses. Moreover, the functions of T_{rise} sensitivities to l and w vs. aspect ratios may also change when the source depth and thickness change, especially if the source depth and thickness are large. Therefore, the quantization in d and t must ensure that, after quantization, the error of T_{rise} sensitivities to l and w is within a certain range (e.g. 50% when quantization ratio of l or w is 10%) which will ensure that the total temperature error of T_{rise} is within 5%.

Because of the fact that the functions of T_{rise} sensitivity to l and w vs. aspect ratio (l/w) vary with different source depth d and thickness t when d and t are large, it is difficult to use a single universal function of T_{rise} sensitivity to l (or w) vs. aspect ratio (l/w) to represent all the situations for sources with large depths and thicknesses. However, in practice, source depth and thickness are usually quite small. Therefore, by properly quantizing the depth and thickness, it is feasible to use a universal function of T_{rise} sensitivity to l (or w) vs. aspect ratio (l/w) to represent the situations for sources having small depths and thicknesses.

Hence, instead of finding all the functions of T_{rise} sensitivity to l (or w) vs. aspect ratio (l/w) for sources with various depths and thicknesses, jobs were done to find out the

range of d and t such that the approximate universal function of T_{rise} sensitivities to l and w vs. aspect ratios found at $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$ can still be applied.

Therefore, in order to apply the above approximation for a source with depth “ d ” and thickness “ t ”, the source may be quantized using depth = 0 and thickness = $2d + t$. The next step is to find out when source depth $d = 0$, how much the source thickness should be, such that the approximate universal function of T_{rise} sensitivities to l and w vs. aspect ratios found at $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$ can still be applied to sources with widths from $0.5\mu\text{m}$ to $2\mu\text{m}$, and lengths from $2\mu\text{m}$ to $20\mu\text{m}$, as these represent transistors in most cases.

Using the same approaches as discussed in sections 5.4 and 5.5, but changing the source thickness by keeping the source depth $d = 0$, Figures 5.17 and 5.18 show T_{rise} sensitivities to l and w vs. aspect ratio (l/w) for sources with depth $d = 0$ and thickness $t = 0.1\mu\text{m}$; while Figures 5.19 and 5.20 show T_{rise} sensitivities to l and w vs. aspect ratio (l/w) for sources with $d = 0$ and $t = 0.5\mu\text{m}$. These plots also contain annotations for the universal functions of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) found from sources with $d = 0$, $t = 0.25\mu\text{m}$ and width $w = 1\mu\text{m}$. It can be seen that for all the sources with $d = 0$, $0.5\mu\text{m} \leq w \leq 2\mu\text{m}$, $1 \leq l/w \leq 40$, and $0.1\mu\text{m} \leq t \leq 0.5\mu\text{m}$, the approximate universal function of T_{rise} sensitivities vs. l/w can still be applied. The errors of T_{rise} sensitivities vs. l/w are within 50% which, when the quantization ratio of w or l is 10%, may cause an acceptable 5% T_{rise} temperature error.

Figure 5.17 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.1\mu\text{m}$

Figure 5.18 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.1\mu\text{m}$

Figure 5.19 T_{rise} Sensitivity to Source Length l as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.5\mu m$

Figure 5.20 T_{rise} Sensitivity to Source Width w as a Function of Aspect Ratio (l/w) for Sources with $d = 0$ and $t = 0.5\mu m$

According to the above discussion and the assumption that for a source with depth “ d ” and thickness “ t ” the source may be quantized using depth = 0 and thickness = $2d + t$, the universal function of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) found from sources with $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$ may be applied only if the source depth d and thickness t satisfy the relation of $0.1\mu\text{m} \leq 2d + t \leq 0.5\mu\text{m}$.

To verify the reliability of this assumption, more simulations were done using tsource simulation. For each case, the substrate was set to semi-infinite and power dissipation $P = 40\text{mW}$. Table 5-6 lists the results. It can be seen that the simulation results well demonstrate the above assumption that for sources with any combinations of $0.1\mu\text{m} \leq 2d + t \leq 0.5\mu\text{m}$, these sources can be quantized using depth = 0 and thickness = $2d + t$. The T_{rise} errors are within 5%.

Table 5-6 Effect of Source Depth and Thickness on T_{rise} (2)

$l \times w$ (μm)	2×0.5	20×0.5	2×2	20×2
T_{rise} (K) ($d = 0.1\mu\text{m}$, $t = 0.2\mu\text{m}$)	318.28	64.85	214.96	53.63
T_{rise} (K) ($d = 0.0\mu\text{m}$, $t = 0.4\mu\text{m}$)	318.25	64.85	214.96	53.63
T_{rise} (K) ($d = 0.2\mu\text{m}$, $t = 0.1\mu\text{m}$)	314.38	64.24	208.40	52.84
T_{rise} (K) ($d = 0.0\mu\text{m}$, $t = 0.5\mu\text{m}$)	297.94	62.72	206.52	52.70
T_{rise} (K) ($d = 0.1\mu\text{m}$, $t = 0.1\mu\text{m}$)	345.04	67.60	224.34	54.62
T_{rise} (K) ($d = 0.0\mu\text{m}$, $t = 0.3\mu\text{m}$)	342.09	67.31	224.01	54.60

However, it is found that, when $2d + t > 0.5\mu\text{m}$, the T_{rise} error increases greatly by quantizing the source with depth of 0 and thickness of $2d + t$. In addition, when $2d + t > 0.5\mu\text{m}$, using the universal function of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) found from sources with $d = 0$, $t = 0.25\mu\text{m}$ and width $w = 1\mu\text{m}$ can also cause big errors of T_{rise} sensitivities vs. l/w . Therefore, the approximations discussed above is only applicable for sources with $2d + t < 0.5\mu\text{m}$.

5.7 Atar Simulation Using Quantization Rules

Based on the above discussions regarding the quantization of source length l , width w , depth d , and thickness t for a given heat source with small depth and thickness ($0.1\mu\text{m} \leq 2d + t \leq 0.5\mu\text{m}$), when using the Atar simulator to generate numerical meshes, one can first quantize the source depth and thickness to be 0 and $2d + t$, respectively. Then, the initial minimum substrate block size can be determined such that, by subdividing the initial minimum substrate block as discussed in chapter 3, a resolution of $2d + t$ is obtained. Finally, the source length l and width w are quantized according to the approximate universal function of T_{rise} sensitivities to l and w vs. aspect ratio (l/w), as plotted in Figure 5.14. After quantization, both l and w are then represented by multiples of $2d + t$ while accuracy is still maintained.

An example follows. A $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ heat source is mounted into the center of a semi-infinite substrate. Source depth $d = 0.1\mu\text{m}$. Power dissipation $P = 40\text{mW}$. Base temperature is assumed to be room temperature, 300K. Using the

image method discussed in section 2.3, thermal simulation was done using the Atar simulator for a quarter, with $\frac{1}{4}P = 10\text{mW}$. Without quantization, the initial minimum substrate block size needs to be defined as multiples of $0.05\mu\text{m}$ (e.g. $0.8\mu\text{m}$) on each side such that, after subdividing the initial block ($0.8\mu\text{m}$) by 2^4 , small blocks with a minimum block size of $0.05\mu\text{m}$ on each side may be created in the heat generation region around the heat source. Thus, both the source dimensions ($5\mu\text{m} \times 0.25\mu\text{m} \times 0.1\mu\text{m}$ for a quarter) and the source depth ($0.1\mu\text{m}$) can be precisely represented by the subdivisions. In addition, without optimization, substrate dimensions were chosen as $316.8\mu\text{m} \times 312\mu\text{m} \times 312\mu\text{m}$ (i.e. $L = t_{sub} \approx 312\mu\text{m}$ as defined in chapter 4) in a quarter. The substrate chosen for simulation is large in order to get rid of the reflection temperature errors caused by substrate border effects, as well as to represent the substrate dimensions with the initial minimum substrate blocks. Figure 5.21 shows the mesh structure.

In Figure 5.21, four different numerical regions were defined in the substrate as power region, device region, cell region, and circuit region. The maximum block sizes of each region were defined as $0.8\mu\text{m}$, $1.6\mu\text{m}$, $3.2\mu\text{m}$ and $6.4\mu\text{m}$, respectively. The maximum substrate block size was defined as $12.8\mu\text{m}$. Inside the power region, a heat generation region was defined and the blocks in the heat generation region were subdivided into $0.05\mu\text{m}$. Simulation results are listed in Table 5-7.

Figure 5.21 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ Source Example with Depth $d = 0.1\mu\text{m}$ before Quantization

In this example, both the source depth and thickness are small. Without quantization, in order to precisely represent the source structure, the numerical blocks in the heat generation region around the source need to be further subdivided. This further subdividing results in a large number of blocks and a longer simulation time.

Using the quantization approach described above, one first quantizes the source depth to be 0 and source thickness to be $2d + t = 0.3\mu\text{m}$. Then, the initial minimum substrate block size are defined to be $1.2\mu\text{m}$ such that, after subdividing the initial block ($1.2\mu\text{m}$) by 2^2 , small blocks with a minimum block size of $0.3\mu\text{m}$ on each side can be cre-

ated in the heat generation region around the heat source. Finally, the source length and width are quantized such that both the source length and width may be represented by multiples of the subdivisions ($0.3\mu\text{m}$) in the heat generation region.

For this example, the aspect ratio of heat source is $l/w = 20$. According to the plots in Figure 5.14, at aspect ratio 20, T_{rise} sensitivity to l is approximately -0.78, and T_{rise} sensitivity to w is approximately -0.16. In the quarter structure, the source length is $5\mu\text{m}$, and source width is $0.25\mu\text{m}$. Therefore, in order to represent the source width with the subdivisions ($0.3\mu\text{m}$) in the heat generation region, the source width may be quantized as $0.3\mu\text{m}$. According to the definition of T_{rise} sensitivity in Equation (5.11), the T_{rise} error after quantization of w can be predicted as:

$$\begin{aligned} \frac{\text{quantized } T_{rise} - \text{original } T_{rise}}{\text{original } T_{rise}} &= (\text{quantization ratio of } w - 1) \cdot (T_{rise} \text{ sensitivity to } w) \\ &= -3.2\% \end{aligned}$$

Therefore, in order to compensate the T_{rise} error caused by the quantization of source width w , the source length needs to be reduced such that

$$(\text{quantization ratio of } l - 1) \cdot (T_{rise} \text{ sensitivity to } l) = +3.2\%$$

Since the T_{rise} sensitivity to l at aspect ratio of 20 is approximately -0.78,

$$\frac{\text{quantized } l}{\text{original } l} = \frac{+3.2\%}{(T_{rise} \text{ sensitivity to } l)} + 1 = \frac{3.2\%}{-0.78} + 1 \approx 0.96$$

Thus, for the quarter structure, quantized $l = \text{original } l \times 0.96 = 4.8\mu\text{m}$.

Therefore, in the quarter structure after quantization, the source length becomes $4.8\mu\text{m}$, source width becomes $0.3\mu\text{m}$, source depth becomes 0, and source thickness becomes $0.3\mu\text{m}$. These can all be represented by multiples of the subdivisions ($0.3\mu\text{m}$) in the heat generation region. Figure 5.22 shows the mesh structure after quantization. Again, four different numerical regions were defined in the substrate as power region, device region, cell region and circuit region. The maximum block sizes of each region were defined as $1.2\mu\text{m}$, $2.4\mu\text{m}$, $4.8\mu\text{m}$ and $9.6\mu\text{m}$, respectively. The maximum substrate block size was defined as $19.2\mu\text{m}$. Inside the power region, a heat generation region was defined and the blocks in this region were subdivided into $0.3\mu\text{m}$. Simulation results are also listed in Table 5-7.

Figure 5.22 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $1.2\mu\text{m}$ on Each Side for the $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ Source Example with Depth $d = 0.0\mu\text{m}$ after Quantization

Meanwhile, simulations were also done using tsource simulation. The substrate was assumed to be semi-infinite.

Table 5-7 Comparison of T_{rise} Before and After Quantization for the $10\mu\text{m} \times 0.5\mu\text{m} \times 0.1\mu\text{m}$ Heat Source with $d = 0.1\mu\text{m}$ and $P = 40\text{mW}$

parameters	Atar simulation		tsource simulation	
	before quantization	after quantization	before quantization	after quantization
	Figure 5.21	Figure 5.22		
source dimensions (μm) in a quarter	$5 \times 0.25 \times 0.1$	$4.8 \times 0.3 \times 0.3$		
source depth d (μm)	0.1	0		
initial minimum substrate block size (μm)	$0.8 \times 0.8 \times 0.8$	$1.2 \times 1.2 \times 1.2$		
maximum substrate block size (μm)	$12.8 \times 12.8 \times 12.8$	$19.2 \times 19.2 \times 19.2$		
size of subdivisions in the heat generation region (μm)	$0.05 \times 0.05 \times 0.05$	$0.3 \times 0.3 \times 0.3$		
number of numerical regions in substrate	4	4		
number of blocks	89074	6025		
CPU time (s)	4.7	0.8		
temperature rise (K)	115.77	115.14	115.19	115.57

From the simulation results listed in Table 5-7, it can be seen that, by quantizing the source dimensions and its depth from the substrate surface, the number of numerical blocks is reduced by a factor of 15 and the CPU time is reduced by a factor of 6. Mean-

while, the T_{rise} error after quantization is within 1%, compared to the T_{rise} before quantization, using both Atar and tsource simulations.

5.8 Multiple Source Quantization

So far, only single source quantization has been discussed. For multiple-source structures, the quantization includes not only source dimensions and depth, but also the positions of each source. Difficulties arise in quantizing source positions because they are related to the geometries of each source and how the dimensions and depth of each source are quantized. Detailed studies on the quantization of source positions in a multiple source structure are beyond the scope of this research but could be a topic for future studies.

Regardless, there are still several ways that may be considered for performing multiple source quantization.

Generally speaking, according to Equation (4.2) discussed in Chapter 4, for the steady state case of a single source with uniform conductivity k , conductive heat transfer

may be defined as: $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + q/k = 0$, where, k is thermal conductivity, and q

is heat generated per unit volume. Since k is often assumed to be constant and not a function of temperature, the relationship between q and T_{rise} is linear. Therefore, thermal analysis is assumed to be linear and follows the principle of superposition.

Thus, in multiple-source structures, the temperature at the center of a source can be considered to be the superposition of temperatures generated by each source at that point.

Figure 5.23 shows a two-source example.

In Figure 5.23, two heat sources are mounted into a substrate at positions ① and ②. The temperature (T_1) at the center of source 1 (p_1) can be considered to be the superposition of the temperatures generated only by source 1 (T_{11}) or source 2 (T_{12}) at p_1 , as shown in Figures 5.23(b) and 5.23(c). Similarly, the temperature (T_2) at the center of source 2 (p_2) can be considered to be the superposition of the temperatures generated only by source 1 (T_{21}) or source 2 (T_{22}) at p_2 . Therefore, T_1 and T_2 can be calculated as $T_1 = T_{11} + T_{12}$, $T_2 = T_{21} + T_{22}$.

(a)

ADD TEMP:

$$T_1 = T_{11} + T_{12}$$

$$T_2 = T_{21} + T_{22}$$

(b)

(c)

Figure 5.23 Multiple Source Temperature Calculation (a) A two-source example
(b) Only source 1 (c) Only source 2

Therefore, in order to ensure the accuracy of T_1 and T_2 , the errors of T_{11} , T_{21} , T_{12} and T_{22} caused by the quantization of source 1 and source 2 need to be minimized. This minimization can be done using the compensation approaches discussed in the example in section 5.7.

In addition, based on the heat spreading mechanism discussed in Chapter 4, it is obvious that increasing the distances between each source reduces the temperature on each source. This fact can be used to compensate the temperature error caused by the quantization of dimensions and depth for each source. An example follows.

Figure 5.24 shows a two-source example. Two $10\mu\text{m} \times 1\mu\text{m} \times 0.1\mu\text{m}$ heat sources are located symmetrically in the centre of a substrate. Similar to the example presented in section 5.7, substrate dimensions are adjusted to $633.6\mu\text{m} \times 624\mu\text{m} \times 312\mu\text{m}$. For each source, depth $d = 0.1\mu\text{m}$ and power dissipation $P = 20\text{mW}$. The distance between these two sources is $10\mu\text{m}$.

Again, because of the symmetrical structure, thermal simulation was done for a quarter using Atar simulation. Without quantization, the initial minimum substrate block size needs to be defined as multiples of $0.1\mu\text{m}$ (e.g. $0.8\mu\text{m}$) on each side, such that, after subdividing the initial block ($0.8\mu\text{m}$) by 2^3 , small blocks with a minimum block size of $0.1\mu\text{m}$ on each side can be created in the heat generation region around the heat source. Thus, all the sources with dimensions ($5\mu\text{m} \times 1\mu\text{m} \times 0.1\mu\text{m}$) and depth ($0.1\mu\text{m}$) in a quarter structure can be precisely represented by the subdivisions. However, in order to precisely represent the distance between these two sources ($5\mu\text{m}$ in a quarter structure)

using the initial minimum substrate block ($0.8\mu\text{m}$), the initial blocks between these two sources must also be subdivided. Therefore, the heat generation region should be created in such a way that both of these two sources are enclosed within it. Figure 5.25 shows the mesh structure. Four different numerical regions were defined in the substrate as power region, device region, cell region and circuit region. The maximum block sizes of each region were defined as $0.8\mu\text{m}$, $1.6\mu\text{m}$, $3.2\mu\text{m}$ and $6.4\mu\text{m}$, respectively. The maximum substrate block size was defined as $12.8\mu\text{m}$. Inside the power region, a heat generation region was defined where the blocks were subdivided into $0.1\mu\text{m}$. Simulation results are listed in Table 5-8.

Figure 5.24 A Two-Source Example Before Quantization

Similarly to the single-source example presented in section 5.7, in this two-source example, both the source depth and thickness are small. In order to precisely represent the source structure without quantization, the numerical blocks in the heat generation region around the sources would need to be subdivided further, which would result in a larger number of numerical blocks and longer simulation time.

Figure 5.25 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8\mu\text{m}$ on Each Side for the Two-Source Example before Quantization

The quantization approach described in section 5.7 now follows. First, the source depth are quantized to be 0 and the source thickness to be $2d + t = 0.3\mu\text{m}$. Next, the initial minimum substrate block size are defined to be $1.2\mu\text{m}$ such that, after subdividing the initial block ($1.2\mu\text{m}$) by 2^2 , blocks with a minimum block size of $0.3\mu\text{m}$ on each side are

created in the heat generation region around heat sources. Finally, the source length and width are quantized such that they are both represented by multiples of the subdivisions ($0.3\mu\text{m}$) in the heat generation region.

For this example, the aspect ratio of each heat source is $l/w = 10$. According to the plots in Figure 5.14, at an aspect ratio of 10, T_{rise} sensitivity to l is approximately -0.74 and T_{rise} sensitivity to w is approximately -0.19. In the quarter structure, the source length is $5\mu\text{m}$, and the source width is $1\mu\text{m}$. Therefore, in order to represent the source width with the $0.3\mu\text{m}$ subdivisions in the heat generation region, the source width was quantized as $1.2\mu\text{m}$. According to the definition of T_{rise} sensitivities in Equation (5.11), the T_{rise} error after quantization of w can be predicted as:

$$\begin{aligned} \frac{\text{quantized } T_{rise} - \text{original } T_{rise}}{\text{original } T_{rise}} &= (\text{quantization ratio of } w - 1) \cdot (T_{rise} \text{ sensitivity to } w) \\ &= -3.8\% \end{aligned}$$

Therefore, in order to compensate the T_{rise} error caused by the quantization of source width w , the source length needs to be reduced such that

$$(\text{quantization ratio of } l - 1) \cdot (T_{rise} \text{ sensitivity to } l) = +3.8\%$$

Since the T_{rise} sensitivity to l at aspect ratio of 10 is approximately -0.74,

$$\frac{\text{quantized } l}{\text{original } l} = \frac{+3.8\%}{(T_{rise} \text{ sensitivity to } l)} + 1 = \frac{3.8\%}{-0.74} + 1 \approx 0.95$$

Thus, for the quarter structure, quantized $l = \text{original } l \times 0.95 = 4.75\mu\text{m}$.

However, in order to represent the source length with the $0.3\mu\text{m}$ subdivisions in the heat generation region, source length was quantized as $4.8\mu\text{m}$ for the quarter. Thus, the T_{rise} error caused by the quantization of l may be predicted as:

$$\begin{aligned} \frac{\text{quantized } T_{rise} - \text{original } T_{rise}}{\text{original } T_{rise}} &= (\text{quantization ratio of } l - 1) \cdot (T_{rise} \text{ sensitivity to } l) \\ &= +2.96\% \end{aligned}$$

Therefore, after quantization, the total T_{rise} error caused by the quantization of the source length and source width, is approximately -0.84%. This error can be compensated by reducing the distance between the sources and thereby increasing the temperature. Since the quantization of multiple source positions is difficult to be characterized, in order to represent in this example the distance between these two sources with the initial minimum substrate blocks ($1.2\mu\text{m}$), the distance between the sources was quantized as $9.6\mu\text{m}$ (i.e. $4.8\mu\text{m}$ in the quarter).

Thus, after quantization, in the quarter structure, the source length becomes $4.8\mu\text{m}$, source width becomes $1.2\mu\text{m}$, source depth becomes 0, and source thickness becomes $0.3\mu\text{m}$. All of these dimensions can be represented by multiples of the $0.3\mu\text{m}$ subdivisions in the heat generation region. In addition, after quantization, half the distance between the sources in the quarter structure becomes $4.8\mu\text{m}$. This distance may be represented by multiples of the initial minimum substrate blocks ($1.2\mu\text{m}$). Figure 5.26 shows the mesh structure after quantization. Again, four different numerical regions were defined in the substrate as power region, device region, cell region and circuit region. The maximum block sizes of each region were defined as $1.2\mu\text{m}$, $2.4\mu\text{m}$, $4.8\mu\text{m}$ and $9.6\mu\text{m}$, respectively. The maximum substrate block size was defined as $19.2\mu\text{m}$. Inside the power

region, a heat generation region was defined in which the blocks were subdivided into $0.3\mu\text{m}$. Simulation results are also listed in Table 5-8.

Figure 5.26 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $1.2\mu\text{m}$ on Each Side for the Two-Source Example after Quantization

From the simulation results listed in Table 5-8, it can be seen that, by quantizing the source dimensions, the depth of each source from the substrate surface and the distances between these sources, the number of numerical blocks is reduced by a factor of 8 and the CPU time is reduced by a factor of 6. Meanwhile, the T_{rise} error after quantization is within 2% compared to the T_{rise} before quantization, using both the Atar and tsource simulations. Therefore, this example shows that the quantization approach used for single sources can still be applied to multiple sources as long as the compensation of temperature

errors is taken into account. Again, for this example, the substrate was assumed to be semi-infinite when performing the tsource simulation.

Table 5-8 Comparison of T_{rise} Before and After Quantization for the Two-Source Example

parameters	Atar simulation		tsource simulation	
	before quantization	after quantization	before quantization	after quantization
	Figure 5.25	Figure 5.26		
source dimensions (μm) in a quarter	$5 \times 1 \times 0.1$	$4.8 \times 1.2 \times 0.3$		
source depth d (μm)	0.1	0		
half the distance between sources in a quarter (μm)	5	4.8		
initial minimum substrate block size (μm)	$0.8 \times 0.8 \times 0.8$	$1.2 \times 1.2 \times 1.2$		
maximum substrate block size (μm)	$12.8 \times 12.8 \times 12.8$	$19.2 \times 19.2 \times 19.2$		
size of subdivisions in the heat generation region (μm)	$0.1 \times 0.1 \times 0.1$	$0.3 \times 0.3 \times 0.3$		
number of numerical regions in substrate	4	4		
number of blocks	57371	6795		
CPU time (s)	4.9	0.8		
temperature rise (K)	58.34	57.98	59.00	58.74

5.9 Conclusion

In this chapter, a quantization approach in integrated device thermal modeling is introduced and investigated. Instead of decreasing numerical block sizes, the quantization

approach is based on adjusting the heat source dimensions, depth from the substrate surface, and distances between multiple sources, in order to represent source structures with a small number of numerical blocks to save simulation time, yet maintain accuracy.

The quantization of source depth and thickness greatly increases the numerical grid size in the heat generation region during an Astar simulation. In addition, the universal function of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) found from sources with $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$ is demonstrated to be suitable for sources with $0.5\mu\text{m} \leq w \leq 2\mu\text{m}$, $1 \leq l/w \leq 40$, $0.1\mu\text{m} \leq 2d + t \leq 0.5\mu\text{m}$, which approximate transistor structures in most cases. By using the universal function of T_{rise} sensitivities to l and w , the quantization ratio of l and w can be determined, and T_{rise} can be predicted.

Moreover, the T_{rise} temperature error caused by the quantization of the source width can be compensated by the quantization of source length. In this way the total amount of T_{rise} error caused by the quantization of source width and length can be minimized. An example was presented which showed that this compensation idea also works for multiple-source structures.

In multiple-source structures, the total amount of T_{rise} error is minimized by minimizing the T_{rise} error from each source after quantization of l and w , using the compensation idea. Though the quantization of source positions in a multiple-source structure was beyond the scope of this research, one method presented for compensating the residual T_{rise} error in these structures was to adjust the distances between each source.

Chapter 6: Case Study

So far, quantization in source dimensions and depth, optimizations of substrate boundary dimensions, and rules for generating square-block based numerical meshes have been discussed. All of these discussions are based on the main purpose of this research, which is to quickly and accurately predict the maximum temperature rise in device junctions. In this chapter, all the rules are verified via a final example, and the efficiency is demonstrated.

Figure 6.1(a) shows a four-source structure for this final example. The sources are symmetrically located within a $24\mu\text{m} \times 16\mu\text{m}$ bounding box in the center of a substrate. For each source, depth $d = 0.1\mu\text{m}$, thickness $t = 0.2\mu\text{m}$, and power dissipation $P = 20\text{mW}$. Two situations are considered. First, the substrate is assumed to be semi-infinite. Second, the substrate is assumed to be with a given thickness which is not large, but wide in lateral dimensions, which are $200\mu\text{m} \times 200\mu\text{m} \times 100\mu\text{m}$ per quarter.

Because of the symmetrical structure, thermal simulation was done for a quarter using the Atar simulator, as shown in Figure 6.1(b). Without quantization, the initial minimum substrate block size needs to be defined as multiples of $0.1\mu\text{m}$ (e.g. $0.8\mu\text{m}$) on each side, such that, after subdividing the initial block ($0.8\mu\text{m}$) by 2^3 , small blocks with a minimum size of $0.1\mu\text{m}$ on each side can be created in the heat generation region around the heat sources. Thus, all the source dimensions ($5\mu\text{m} \times 0.5\mu\text{m} \times 0.2\mu\text{m}$, and $2\mu\text{m} \times 1\mu\text{m} \times 0.2\mu\text{m}$) and depth ($0.1\mu\text{m}$) in a quarter structure can be precisely repre-

sented by the $0.1\mu\text{m}$ subdivisions. However, in order to precisely represent the distances between these sources ($7.5\mu\text{m}$ and $10\mu\text{m}$ in a quarter structure) using the initial minimum substrate blocks ($0.8\mu\text{m}$), the initial blocks between these sources also need to be subdivided. Therefore, the heat generation region should be created to enclose all of these sources.

In addition, as a general consideration, for small devices in large substrates, in order to accurately represent the heat flow and predict the maximum temperature rise in the junction area, the substrate boundaries in the computational domain must be sufficiently far from the heat source in order that border effects can be neglected. Otherwise, heat spreading will be constricted by the fictitious boundary surfaces, and the reflection of heat from the fictitious adiabatic boundaries will lead to a simulation-based temperature increase which introduces components of heat rise that are not representative of the physical situation. Therefore, in the first case of this example where the substrate is assumed to be semi-infinite, without using the approximation rules as discussed in Chapter 4 to optimize the substrate boundary dimensions, substrate thickness t_{sub} and lateral dimension L were chosen $L = t_{sub} = 400\mu\text{m}$, as shown in Figure 6.1(b). In the second case, the substrate dimensions were assumed to be $200\mu\text{m} \times 200\mu\text{m} \times 100\mu\text{m}$ in a quarter. Figure 6.2 shows the quarter mesh structure.

Figure 6.1 (a) Top view of a four-source example (b) A quarter structure before quantization for Atar simulation (c) A quarter structure after quantization for Atar simulation

Figure 6.2 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of $0.8 \mu\text{m}$ on Each Side for the Four-Source Example before Quantization

According to the approximation rules discussed in Chapter 3, four different numerical regions were defined in the substrate as power region, device region, cell region and circuit region. The maximum block sizes of each region were defined as $0.8 \mu\text{m}$, $1.6 \mu\text{m}$, $3.2 \mu\text{m}$ and $6.4 \mu\text{m}$, respectively. The maximum substrate block size was defined as $12.8 \mu\text{m}$. Inside the power region, a heat generation region was defined, in which the numerical blocks were subdivided into $0.1 \mu\text{m}$. Simulation results are listed in Tables 6-1 and 6-2.

In this example, both the source depth and thickness are small. Without quantization of the source dimensions and depth, in order to precisely represent the source structure, the numerical blocks in the heat generation region around the sources need to be

further subdivided. This further subdividing results in a large number of numerical blocks and more simulation time, as does the neglecting to optimize the substrate boundary dimensions.

The quantization approach described in section 5.7 now follows. First, for each source, the depth is quantized to be 0, and thickness to be $2d + t = 0.4\mu\text{m}$. Then, the initial minimum substrate block size is defined to be $1.6\mu\text{m}$ such that, after subdividing the initial block ($1.6\mu\text{m}$) by 2^2 , blocks with a minimum block size of $0.4\mu\text{m}$ on each side can be created in the heat generation region around the heat sources. Finally, the source length and width are quantized such that they may be represented by multiples of the $0.4\mu\text{m}$ subdivisions in the heat generation region.

For this example, the aspect ratio of the $2\mu\text{m} \times 2\mu\text{m}$ heat source (heat source 2) is $l/w = 1$. According to the plots in Figure 5.14, at an aspect ratio of 1, T_{rise} sensitivity to l (and w) is approximately -0.4. In the quarter structure, the length of source (2) is $2\mu\text{m}$, and the width of source (2) is $1\mu\text{m}$. Therefore, in order to represent the width of source (2) with the subdivisions ($0.4\mu\text{m}$) in the heat generation region, source width was quantized as $1.2\mu\text{m}$. According to the definition of T_{rise} sensitivities in Equation (5.11), the T_{rise} error after quantization of w can be predicted as:

$$\begin{aligned} \frac{\text{quantized } T_{rise} - \text{original } T_{rise}}{\text{original } T_{rise}} &= (\text{quantization ratio of } w - 1) \cdot (T_{rise} \text{ sensitivity to } w) \\ &= -8\% \end{aligned}$$

Therefore, in order to compensate the T_{rise} error caused by the quantization of source width w , the source length needs to be reduced such that

$$(\text{quantization ratio of } l - 1) \cdot (T_{rise} \text{ sensitivity to } l) = +8\%$$

Since the T_{rise} sensitivity to l at an aspect ratio of 1 is approximately -0.4,

$$\frac{\text{quantized } l}{\text{original } l} = \frac{+8\%}{(T_{rise} \text{ sensitivity to } l)} + 1 = \frac{8\%}{-0.4} + 1 = 0.8$$

Thus, for the $2\mu\text{m} \times 2\mu\text{m}$ heat source in the quarter structure, quantized $l = \text{original } l \times 0.8 = 1.6\mu\text{m}$.

Similarly, the aspect ratio of the $10\mu\text{m} \times 0.5\mu\text{m}$ heat source (heat source 1) is $l/w = 20$. According to the plots in Figure 5.14, at an aspect ratio of 20, T_{rise} sensitivity to l is approximately -0.78, and T_{rise} sensitivity to w is approximately -0.16. In the quarter structure, the length of source (1) is $5\mu\text{m}$, and the width of source (1) is $0.5\mu\text{m}$. Therefore, in order to represent the width of source (1) with the subdivisions ($0.4\mu\text{m}$) in the heat generation region, source width was quantized as $0.4\mu\text{m}$. According to the definition of T_{rise} sensitivities in Equation (5.11), the T_{rise} error after quantization of w can be predicted as:

$$\begin{aligned} \frac{\text{quantized } T_{rise} - \text{original } T_{rise}}{\text{original } T_{rise}} &= (\text{quantization ratio of } w - 1) \cdot (T_{rise} \text{ sensitivity to } w) \\ &= +3.2\% \end{aligned}$$

Therefore, in order to compensate the T_{rise} error caused by the quantization of source width w , the source length needs to be increased such that

$$(\text{quantization ratio of } l - 1) \cdot (T_{rise} \text{ sensitivity to } l) = -3.2\%$$

Since the T_{rise} sensitivity to l at an aspect ratio of 20 is approximately -0.78,

$$\frac{\text{quantized } l}{\text{original } l} = \frac{-3.2\%}{(T_{rise} \text{ sensitivity to } l)} + 1 = \frac{-3.2\%}{-0.78} + 1 = 1.04$$

Thus, for the $10\mu\text{m} \times 0.5\mu\text{m}$ heat source in the quarter structure, quantized $l = \text{original } l \times 1.04 = 5.2\mu\text{m}$.

Therefore, after quantization, both the depth and dimensions of each source can be precisely represented by the $0.4\mu\text{m}$ subdivisions in the heat generation region. The total amount of T_{rise} error caused by the quantization of source length l and width w for each source is compensated to “0”.

However, in order to represent the distances between each source using the initial minimum substrate blocks ($1.6\mu\text{m}$), the source positions need to be adjusted. Based on the compensation idea, for this example, the distance between the two $10\mu\text{m} \times 0.5\mu\text{m}$ sources was increased to be $16\mu\text{m}$, while the distance between the two $2\mu\text{m} \times 2\mu\text{m}$ sources was reduced to be $19.2\mu\text{m}$.

Furthermore, according to the discussion in chapter 4, when substrate dimensions are large (semi-infinite), it is possible to shrink the substrate, i.e. to reduce both substrate thickness and lateral dimensions in the computational domain in order to save simulation time while still maintaining accuracy. In addition, in order to maintain accuracy, when substrate thickness is given and is not large, but is wide in lateral dimensions, the substrate

lateral dimensions can be reduced in the computational domain while the substrate thickness is maintained in such a way that $L = t_{sub}$.

Therefore, the multiple-source structure is considered as a bounding box and the line-source simulation solutions discussed in Chapter 4 are used to estimate substrate boundary dimensions for simulation.

For this example, the size of the bounding box is $24\mu\text{m} \times 16\mu\text{m}$. Based on the line-source simulation solutions discussed in section 4.3, in order to keep the reflection error of the bounding box smaller than 2%, the acceptable reflection error for the line source of length $16\mu\text{m}$ is $\frac{2\%}{2 \times (1 + l/w)} = 0.004$.

From the line-source simulation results plotted in Figure 4.21, it can be approximated that when $L = t_{sub} \geq 125\mu\text{m}$, the reflection error for the line source of length $16\mu\text{m}$ is ≤ 0.004 . Thus, to make the reflection error of the $24\mu\text{m} \times 16\mu\text{m}$ bounding box smaller than 2% compared to the T_{rise} in a semi-infinite substrate, the substrate dimensions should be chosen as $L = t_{sub} \geq 125\mu\text{m}$ in the computational domain. However, in order that the substrate dimensions can be represented by multiples of the initial minimum substrate blocks ($1.6\mu\text{m}$), the substrate thickness for this example was chosen as $t_{sub} = 128\mu\text{m}$, and L was adjusted to be $128\mu\text{m}$ and $129.2\mu\text{m}$, as shown in Figure 6.1(c). Figure 6.1(c) shows a quarter structure after quantization, and Figure 6.3 shows its mesh structure after quantization.

However, as discussed in Chapter 4, in order to maintain solution accuracy for the $24\mu\text{m} \times 16\mu\text{m}$ bounding box when substrate thickness is given as $t_{sub} = 100\mu\text{m}$, one must maintain t_{sub} , and reduce L so that $L = t_{sub} = 100\mu\text{m}$. Similarly, in order that the substrate dimensions can be represented by multiples of the initial minimum substrate blocks ($1.6\mu\text{m}$), the substrate thickness was chosen as $t_{sub} = 99.2\mu\text{m}$, and L was adjusted to $110.4\mu\text{m}$ and $108.8\mu\text{m}$, instead of $129.2\mu\text{m}$ and $128\mu\text{m}$, as shown in Figure 6.1(c).

Again, according to the rules discussed in Chapter 3, four different numerical regions were defined in the substrate as power region, device region, cell region and circuit region. The maximum block sizes of each region were defined as $1.6\mu\text{m}$, $3.2\mu\text{m}$, $6.4\mu\text{m}$ and $12.8\mu\text{m}$, respectively. The maximum substrate block size was defined as $25.6\mu\text{m}$. Inside the power region, a heat generation region was defined, in which the numerical blocks were further subdivided into $0.4\mu\text{m}$. The simulation results are also listed in Tables 6-1 and 6-2.

From these simulation results, it can be seen that, by optimizing the substrate boundary dimensions in the computational domain and quantizing the source dimensions, the depth of each source from the substrate surface, and the distances between these sources, the number of numerical blocks is reduced by a factor of 87, CPU time is reduced by a factor of 120, and the T_{rise} error on each source after quantization is within 3% compared to the T_{rise} on each source before quantization in a semi-infinite substrate, using both Atar and tsource simulations. In addition, for a given substrate thickness which is not large, but is large in substrate lateral dimensions, after optimizations and quantization, the number of numerical blocks is reduced by a factor of 60, the CPU time is reduced by a fac-

tor of 39, and the T_{rise} error on each source after quantization is within 5% compared to the T_{rise} on each source before quantization, using both Atar and tsource simulations. Therefore, it is well demonstrated that the approximation approaches discussed in the previous chapters are helpful in integrated device thermal modeling.

Figure 6.3 A Quarter Mesh Structure with Initial Minimum Substrate Blocks of 1.6 μm on Each Side for the Four-Source Example after Quantization

Table 6-1 Comparison of T_{rise} Before and After Quantization for the Four-Source Example in a Semi-Infinite Substrate

parameters	Atar simulation		tsource simulation	
	before quantization	after quantization	before quantization	after quantization
	Figure 6.2	Figure 6.3		
$t_{sub} = L$ (μm)	400	128	∞	∞
source thickness (μm)	0.2	0.4	0.2	0.4
source depth d (μm)	0.1	0	0.1	0
initial minimum substrate block size (μm)	$0.8 \times 0.8 \times 0.8$	$1.6 \times 1.6 \times 1.6$		
maximum substrate block size (μm)	$12.8 \times 12.8 \times 12.8$	$25.6 \times 25.6 \times 25.6$		
size of subdivisions in the heat generation region (μm)	$0.1 \times 0.1 \times 0.1$	$0.4 \times 0.4 \times 0.4$		
number of numerical regions in substrate	4	4		
number of blocks	114649	1322		
CPU time (s)	12.0	0.1		
temperature rise at source (1) (K)	70.66	72.38	70.38	69.78
temperature rise at source (2) (K)	121.92	121.95	121.44	122.20

Table 6-2 Comparison of T_{rise} Before and After Quantization for the Four-Source Example in a Substrate with Given $t_{sub} = 100\mu\text{m}$ and Wide Lateral Dimensions

parameters	Atar simulation		tsource simulation	
	before quantization	after quantization	before quantization	after quantization
	Figure 6.2	Figure 6.3		
substrate dimensions (μm) in a quarter	$200 \times 200 \times 100$	$110.4 \times 108.8 \times 99.2$	$t_{sub} = 100\mu\text{m}$ $L = \infty$	$t_{sub} = 100\mu\text{m}$ $L = \infty$
source thickness (μm)	0.2	0.4	0.2	0.4
source depth d (μm)	0.1	0	0.1	0
initial minimum substrate block size (μm)	$0.8 \times 0.8 \times 0.8$	$1.6 \times 1.6 \times 1.6$		
maximum substrate block size (μm)	$12.8 \times 12.8 \times 12.8$	$25.6 \times 25.6 \times 25.6$		
size of subdivisions in the heat generation region (μm)	$0.1 \times 0.1 \times 0.1$	$0.4 \times 0.4 \times 0.4$		
number of numerical regions in substrate	4	4		
number of blocks	72089	1202		
CPU time (s)	3.9	0.1		
temperature rise at source (1) (K)	68.99	72.01	68.48	67.88
temperature rise at source (2) (K)	120.25	121.59	119.54	120.30

Chapter 7: Conclusions and Future Studies

7.1 Conclusions

In this research, three approximation approaches based on quad-tree meshes have been explored for efficient numerical thermal simulation of semiconductor devices. These approaches cover the generation of square-block based numerical meshes, optimization of substrate boundary dimensions, and quantization of heat sources.

Among these three approximation approaches, the most important is the quantization of heat source dimensions and depth from the substrate surface. Instead of decreasing numerical block sizes, the quantization approach is based on adjusting the heat source dimensions, the depth from the substrate surface, and the distances between multiple sources, in order to represent source structures with a small number of numerical blocks, while saving simulation time and maintaining accuracy. This quantization approach was performed first by quantizing the source depth d and thickness t to “0” and “ $2d + t$ ”, respectively, and then quantizing source length l and width w according to temperature sensitivities to l and w . The approximate universal function of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) found from sources with $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$ was demonstrated to be applicable for sources with $0.5\mu\text{m} \leq w \leq 2\mu\text{m}$, $1 \leq l/w \leq 40$, and $0.1\mu\text{m} \leq 2d + t \leq 0.5\mu\text{m}$, which approximate transistor structures in most cases. Therefore, this universal function of T_{rise} sensitivities to l and w was used to determine the quantization ratio of l and w , and to predict the junction temperature rise T_{rise} .

Using the quantization approach, the T_{rise} temperature error caused by the quantization of source width was compensated by the quantization of the source length. The total amount of temperature error for multiple-source structures was minimized by minimizing the temperature error from each source after quantization of source length and width, using the compensation idea.

The simulation results presented in sections 5.7 and 5.8 illustrated that the quantization of source dimensions and depth from the substrate surface, greatly increases the numerical grid size in the heat generation region and simplifies the mesh structure when using the Atar numerical thermal simulator. The CPU time was reduced dramatically with the reduction of the number of numerical blocks, while solution accuracy was maintained.

The optimization of substrate boundary dimensions is another important approach explored in this research. In this optimization approach, a normalized junction temperature rise, T_{rise}/T_{inf} , was first introduced, where, T_{inf} is the temperature rise when the substrate is semi-infinite, and T_{rise} is the temperature rise when the substrate is constricted, as in the computational domain. The reflection temperature error was then defined as “ $1 - T_{rise}/T_{inf}$ ”. Thus, the evaluation of the heat reflection error caused by the fictitious adiabatic substrate boundaries in the computational domain was narrowed down to investigating only the effects of the source geometry and substrate dimensions on temperature. It was found that the effect of power dissipation can be ignored.

In addition, the effects of a finite substrate on thermal analysis were also studied. Based on the heat spreading mechanism, line-source simulation solutions and bounding

box ideas were introduced in optimizing the substrate boundary dimensions according to source structures.

The line-source simulation is based on the idea that reflection error functions are related to source dimensions and aspect ratios. Therefore, simulations were done by taking each edge of a thermal source as a single line source whose width and thickness are small, characterizing reflections at each edge using these line sources, applying the results to the edges of the rectangular sources, and comparing the simulation results with those when simulating at a single square source or rectangular source. Bounding box simulations consider the entire source structure as a bounding box. Thus, the line-source simulation solutions can still be applied for multiple-source structures in estimating an optimal substrate size for thermal simulation, and can save simulation time.

Simulation results indicate that for substrates with a small thickness but a large width, their lateral dimensions can be reduced by maintaining their thickness. For substrates with both a large thickness and large width, both their thickness and lateral dimensions can be reduced simultaneously.

The 3D numerical thermal simulator, Atar, was also used throughout this research for creating square-block based numerical meshes in such a way that numerous fine grids were produced around heat generation regions which were close to the substrate surface, but only a moderate number of numerical blocks were generated for the entire device. By using Atar, the substrate was modeled by different numerical regions using various gridding masks. In each of these regions, numerical grids were created by the iterative amalgamation of blocks from groups of four into single blocks. This amalgamation

commenced from the initial minimum substrate blocks, and stopped when all blocks reached their maximum sizes as defined by the gridding masks. An exception was the heat generation region. This region was created around the heat source by further subdividing the initial minimum substrate blocks. Rules and procedures for generating square-block based numerical meshes were investigated and, based on the grid independence concept, approximation expressions for setting up gridding mask dimensions and the maximum block size within each numerical region were presented.

The simulation results presented in the case study chapter demonstrate that, by using these approximation approaches, the number of numerical blocks for the example in a semi-infinite substrate is reduced by a factor of approximately 90, the CPU time is reduced by a factor of almost 120, and the temperature error is within 3% compared to the temperature before quantization and optimizations, using both Atar and tsource simulations. In addition, for the example with a given substrate thickness which is not large, but is large in substrate lateral dimensions, after quantization and optimization, the number of numerical blocks is reduced by a factor of approximately 60, the CPU time is reduced by a factor of almost 40, and the temperature error after quantization is within 5% compared to the temperature before quantization. Therefore, these approximation approaches are useful in integrated device thermal modeling.

7.2 Summary of Contributions

This research has two main contributions:

1) The quantization approach is introduced to simplify the numerical mesh generation using Atar.

2) The optimization approach is developed to suggest the optimal substrate boundary dimensions for simulation.

Both of these two approaches result in a great reduction in the number of numerical blocks and CPU time, while the simulation results remain accurate.

7.3 Future Work

Eventually, the Atar 3D numerical thermal simulator would be linked into a CAD environment. The area of heat source should be automatically determined from layout, and the source depth and thickness from technology information. The rules and approximation approaches explored in this research can be used to simplify the numerical mesh generation when using Atar thermal simulator, and reduce the number of numerical blocks and CPU time while retaining accuracy. The simulator should be able to automatically determine the quantized source structure and optimized substrate dimensions for simulation. Different numerical regions should be able to be automatically generated using various gridding masks around the heat source in the substrate, according to the mesh generation rules discussed in this research. Thus, ultimately, the entire thermal simulation process can be more transparent to the user.

However, some issues remain to be explored.

For example, the universal function of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) found from sources with $d = 0$, $t = 0.25\mu\text{m}$ and $w = 1\mu\text{m}$, was demonstrated to be applicable for sources with $0.5\mu\text{m} \leq w \leq 2\mu\text{m}$, $1 \leq l/w \leq 40$, $0.1\mu\text{m} \leq 2d + t \leq 0.5\mu\text{m}$. However, the functions of T_{rise} sensitivities to l and w vs. aspect ratio (l/w) vary with different source depth d and thickness t when d and t are large. For sources with a large depth and thickness, the approximate universal function may not be accurate. Therefore, further investigations into the effects of large source depth and thickness on T_{rise} sensitivities to l and w vs. aspect ratio (l/w) are required, and can be a topic for future researches.

In addition, for multiple-source structures, the quantization includes not only source dimensions and depth, but also the positions of each source. Difficulties arise in quantizing source positions because they are related to the geometries of each source and how the dimensions and depth of each source are quantized. Detailed discussions on the quantization of source positions in a multiple-source structure are beyond the scope of this research but is a topic worthy of future study.

Moreover, it should be pointed out that the approximation approach discussed in chapter 4 regarding the optimization of substrate boundary dimensions is only applicable for steady state thermal analysis. If the transient response is considered, the general differential equation for heat flow under transient conditions with uniform thermal conductivity k is represented as[1]:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} + q/k = \rho C_\theta \frac{\partial T}{\partial t} \quad (7.1)$$

The term $\rho C_{\theta} \frac{\partial T}{\partial t}$ in equation (7.1) describes the time-varying portion of the heat flow. The solution of this differential equation leads to an equation for thermal capacitance[1]:

$$C_{\theta} = C_T \rho V \quad (7.2)$$

where, C_{θ} is thermal capacitance in W.sec/°C

ρ is density of the material in kg/m³

C_T is specific heat of the material in W.sec/kg°C

V is volume of the material in m³

In finite difference numerical thermal analysis, equation (7.2) is applied by calculating an effective thermal capacitance associated with each numerical block. According to equations (7.1) and (7.2), it can be seen that the thermal transient depends on the amount of material presented. If the substrate dimensions are reduced, the total amount of material is changed. Therefore, there is likely some effect on the transient behaviour.

In order to accurately represent the thermal transient conditions, new approaches need to be developed. This can also be a topic for future studies.

Reference

- [1] J. Sergent and A. Krum, *Thermal Management Handbook: for Electronic Assemblies*. New York, NY: McGraw-Hill, 1998.
- [2] P. J. Schneider, *Conduction Heat Transfer*. Reading, MA: Addison-Wesley Publishing Company, 1955.
- [3] A. L. Palisoc and C. C. Lee, "Exact thermal representation of multilayer rectangular structures by infinite plate structures using the method of images," *J. Appl. Phys.*, vol. 64, no. 12, pp. 6851-6857, 1988.
- [4] T. Smy, D. Walkey, and S. K. Dew, "A 3-D thermal simulation tool for integrated device - Atar," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 20, no. 1, pp. 105-115, 2001.
- [5] A. D. Kraus, R. C. Chu, and A. Bar-Cohen, "Thermal management of microelectronics: pass, present, and future," *Computers in Mechanical Engineering*, vol. 1, no. 2, pp. 69-79, 1982.
- [6] M. M. Yovanovich and V.W. Antonetti, "Application of thermal contact resistance theory to electronic packages," in *Advances in thermal Modeling of Electronic Components and systems, Volume 1* (A. Bar-Cohen and A. D. Kraus). Hemisphere Publishing Corporation, 1988, pp. 79-128.
- [7] R. David, "Computerized Thermal Analysis of Hybrid Circuits," *IEEE Transactions on Parts, Hybrids, and Packaging*, vol. PHP-13, no.3, pp. 283-290, 1977.
- [8] F. N. Masana, "A Closed Form Solution of Junction to Substrate Thermal Resistance in Semiconductor Chips", *IEEE Transactions on Components, Packaging, and Manufacturing Technology-Part A*. Vol. 19, No. 4, pp. 539-545, 1996.
- [9] D. P. Kennedy, "Spreading resistance in cylindrical semiconductor devices," *J. Appl. Phys.*, vol.31, no. 8, pp.1490-1497, 1960.
- [10] R. D. Lindsted and R. J. Surty, "Steady-state junction temperatures of semiconductor chips," *IEEE Transactions on Electron Devices*, vol. 19, no. 1, pp. 41-44, 1972.
- [11] R. C. Joy and E. S. Schlig, "Thermal properties of very fast transistors," *IEEE Transactions on Electron Devices*, vol. 17, no. 8, pp. 586-593, 1970.

- [12] C. C. Lee, A. L. Palisoc, and Y. J. Min, "Thermal analysis of integrated circuit devices and packages," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 12, no. 4, pp. 701-709, 1989.
- [13] C. W. Kim, N. Goto, and K. Honjo, "Thermal behavior depending on emitter finger and substrate configurations in power heterojunction bipolar transistors," *IEEE Transactions on Electron Devices*, vol. 45, no. 6, pp. 1190-1195, 1998.
- [14] C. Lasance, H. Vinke, and H. Rosten, "Thermal characterization of electronic devices with boundary condition independent compact models," *IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part A*, vol. 18, no. 4, pp. 723-731, 1995.
- [15] C. C. Lee, Y. J. Min, and A. L. Palisoc, "A general integration algorithm for the inverse fourier transform of four-layer infinite plate structures," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology*, vol. 12, no. 4, pp. 710-716, 1989.
- [16] D. J. Walkey, T. J. Smy, D. Marchesan, H. Tran, and M. Schroter, "A scalable thermal model for trench isolated bipolar devices," *Solid-State Electronics*, vol. 44, pp. 1373-1379, 2000.
- [17] V.S. Arpaci, *Conduction Heat Transfer*. Reading, MA: Addison-Wesley Publishing Company, 1966.
- [18] M. N. Ozisik, *Finite Difference Method in Heat Transfer*. Boca Raton, FL: CRC Press, 1994.
- [19] D. Poulikakos, *Conduction Heat Transfer*. Englewood Cliffs, NJ: Prentice Hall, 1994.
- [20] M. N. Ozisik, *Boundary Value Problems of Heat Conduction*. Scranton, PL: International Textbook Company, 1968.