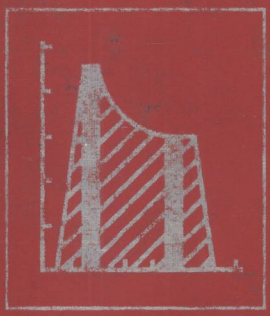
RAPID THERMAL PROCESSING



Science and Technology

Edited by

RICHARD B. FAIR

Rapid Thermal Processing

Science and Technology

Edited by

Richard B. Fair
MCNC
Center for Microelectronic Systems Technologies
Research Triangle Park, North Carolina

ACADEMIC PRESS, INC.
Harcourt Brace Jovanovich, Publishers
Boston San Diego New York
London Sydney Tokyo Toronto

This book is printed on acid-free paper. ♾

ACADEMIC PRESS, INC.
1250 Sixth Avenue, San Diego, CA 92101-4311

United Kingdom Edition published by
ACADEMIC PRESS LIMITED
24-28 Oval Road, London NW1 7DX

Library of Congress Cataloging-in-Publication Data

Fair, Richard B.
Rapid thermal processing : science and technology / Richard B. Fair.
p. cm.
Includes bibliographical references and index.
ISBN 0-12-247690-5
1. Semiconductors--Heat treatment. 2. Semiconductor doping. I. Title
TK7871.85.F299 1993
621.3815'2--dc20 93-6882
CIP

Printed in the United States of America

93 94 95 96 97 BB 9 8 7 6 5 4 3 2 1

Rapid Thermal Processing

Science and Technology

Contributors

Number in parentheses indicate the pages on which the authors' contributions begin.

RICHARD B. FAIR (1, 169). *MCNC, Center for Microelectronic Systems Technologies, P.O. Box 12889, Research Triangle Park, North Carolina 27709 and Department of Electrical Engineering, Duke University, Durham, North Carolina 27706*

J. L. HOYT (13), *Solid State Electronics Laboratory, McCullough Building, Room 226, Stanford, California 94305*

KEVIN S. JONES (123), *Department of Materials Science and Engineering, University of Florida, Gainesville, Florida 32611*

B. LOJEK (311), *Motorola Inc., Advanced Technology Center, 2200 West Broadway Road, Mesa, Arizona 85202*

HISHAM Z. MASSOUD (45), *Semiconductor Research Laboratory, Department of Electrical Engineering, Duke University, Durham, North Carolina 27706*

C. M. OSBURN (227), *MCNC, Center for Microelectronic Systems Technologies, P.O. Box 12889, Research Triangle Park, North Carolina 27511 and Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, North Carolina 27695*

MEHMET C. ÖZTÜRK (79), *North Carolina State University, Department of Electrical and Computer Engineering, Raleigh, North Carolina 27695-7911*

FRED ROOZEBOOM (349), *Philips Research Laboratories, P.O. Box 80,000, NL-5600 JA Eindhoven, The Netherlands*

GEORGE A. ROZGONYI (123), *Department of Materials Science and Engineering, North Carolina State University, Raleigh, North Carolina 27695*

Contents

1 Rapid Thermal Processing—A Justification

Richard B. Fair
MCNC
Center for Microelectronic Systems Technologies
Research Triangle Park
North Carolina
and
Department of Electrical Engineering
Duke University, Durham
North Carolina

I. Manufacturing Issues in the Gigachip Age

The cost of manufacturing submicron, ultralarge scale integration (ULSI) chips is scaling upwards in at least inverse proportion to the downward scaling of device feature sizes. These expenses are driven by a manufacturing budget crisis associated with the technological and complexity limits of integrated circuit (IC) design, costs of research and development to address manufacturing issues, and facility capitalization [1]. Thus, examination of the rate of progress in microelectronics over the past 30 years suggests that the primary challenge in reaching the gigachip age in the year 2001 will be the semiconductor industry's ability to change the cost trend lines; that is, to change the economics of how ICs are developed and manufactured [2].

ISBN 0-12-247690-5

The scale of integration of dynamic random-access memory (DRAM) chips has continued to increase by four times every three years! And there is evidence that nationalistic efforts in Japan, Europe, and the United States are attempting to accelerate the integration-versus-time trend curves in the face of shrinking profit margins in order to achieve world dominance. However, there is great risk in these investments. Indeed, the system applications that would utilize higher density DRAMs are not keeping pace with chip availability. It is clear that the semiconductor manufacturers are trying to drive the end-user market! Thus, accelerated leapfrog programs to produce gigachips may have an inadequate market for timely DRAM sales [3]. It is expected that 85% of gross annual sales of 1G DRAMs will be required to pay for research and development and manufacturing costs, assuming normal market growth [4]. This estimate is based on 1G DRAM research and development investment growing to 10–15 times that of the 1M DRAM, 1.2 to 1.3 times more processing steps per generation, 0.9 times fewer chips per wafer per generation in spite of larger diameter wafers, and 40–50 times larger investment in production equipment!

Several approaches have been suggested for reducing the costs associated with developing and manufacturing gigachips, including internationalizing the technology through global partnerships. Requirements on contamination, process control (manufacturing parameter budgets), and cost of manufacturing floor space are driving a paradigm shift to a microprocessing methodology. Thus, single-wafer processing environments with highly controlled, ultraclean ambients clustered together in specialty process modules are being considered. In single chamber machines it is necessary to extract a silicon wafer out of a carrier and present it to the process chamber. Wafer transport among modules can be done best in a modest vacuum (10^{-4} to 10^{-5} torr). *In situ* vacuum processing equipment accounts for 40% of the total equipment today. Rapid thermal processing (RTP) using lamp heating will move thermal processes into cluster tools. Dry cleaning will also move vacuum processes to cluster tools. It is even projected that *in situ* lithography is possible. Thus it is possible that 80% of gigachip equipment could be *in situ* vacuum-based clusters controlled by a factory information system [2].

Rapid thermal processing is a key technology in the cluster tool, single-wafer manufacturing approach. With RTP a single wafer is heated quickly at atmospheric or low pressure under isothermal conditions. The processing chamber is made of either quartz, silicon carbide, stainless steel, or aluminum with quartz windows. The wafer holder is often made of quartz and contacts the wafer at a minimum number of places. A temperature measurement system is placed in a control loop to set wafer temperature. The RTP system is interfaced with a gas handling system and a computer that controls system operation. The small thermal mass inherent in this

Table I Technology comparisons.

Furnace	RTP
Batch	Single wafer
Hot wall	Cold wall
Long time	Short time
Small dT/dt	Large dT/dt
High cycle time	Low cycle time
Temperature measurement	Temperature measurement
—environment	—wafer
Issues	Issues
—small thermal budget	—uniformity
—particles	—repeatability
—atmosphere	—throughput
—vertical furnaces	—stress
	—measurements
	—automation

processing system along with stringent ambient and particle control allow for reduced processing times and improved control in the formation of *pn* junctions, thin oxides, nitrides and silicides, thin deposited layers, and flowed glass structures. In essence RTP provides a controlled environment for thermally activated processes that is increasingly difficult for existing batch furnace systems to achieve. In addition, RTP is fully consistent with the advanced microprocessing-for-manufacturing paradigm.

A comparison between batch furnace and RTP technologies is shown in Table I. In order to achieve short processing times, one trades off a new set of challenges including temperature and process uniformity, temperature measurement and control, wafer stress, and throughput.

A road map showing the introduction of RTP into manufacturing DRAMs is depicted in Table II. Each of the key processing areas that will be impacted by RTP is listed along with estimated timing for the transition from batch to single-wafer technology. Details regarding these processing areas are provided in the subsequent chapters of this book. A discussion of the manufacturing requirements that are driving the expanded use of RTP follows.

II. The Parameter Budget Crisis

The challenge to the semiconductor industry to maintain future viability is to develop equipment and processes that will mass produce ULSI chips with tight tolerance, high reliability, and low costs. This challenge translates to technological problems associated with patterning, doping, interconnections

Table II Rapid thermal processing technology road map.

	1985		1990	1995		2000	
Design rule (μm)	2	1.3	0.8	0.5	0.3	0.2	0.15
DRAM equivalent	256k	1M	4M	16M	64M	256M	1G
Wafer size (in.)	5	5	6–8	8	8–10		12
t_{ox} (nm)	35	25	20	15	12	10	8
x_j (μm)	0.3	0.25	0.2	0.15	0.10	0.08	0.06
Repeatability (°C)	±15	±5		±2		±1	
Uniformity (°C)	±7–10	±3–5		±2–3		<±2	
Accuracy (°C)	±20+	±5–10		±3–5		<±3	
Thermal budget x_j	Furnace			RTA			
t_{ox}	Furnace			RTO			
Thin dielectrics	LPCVD			RTCVD-ONO			
Silicides	Furnace	RTA					
Epitaxy	LPCVD, MBE, APCVD, etc.			RTCVD			
Polysilicon	LPCVD			RTCVD-SiGe			
Metals	LPCVD, sputtered, evaporated					RTCVD	
Cleaning	Wet chemistry			RTC			
Processing	Batch		Single wafer	Cluster tools			
Atmospheric budget	0.1/cm^2		0.01/cm^2	0.005/cm^2		0.001/cm^2	
	>0.5 μm		>0.3 μm	>0.2 μm		>0.1 μm	

defect densities, mechanical and structural aspects of handling large-diameter silicon wafers, contamination, and thermal requirements. All of these problems can be discussed in terms of process parameter targets and control tolerances or budgets. The allowed processing parameter budgets are set by device performance and manufacturing requirements, and it is through quantifying these budgets that the requirements for advanced manufacturing are set [1].

A. Thermal Budget

The process thermal budget refers to the allowed time at temperature that can be tolerated to control dopant impurity diffusion and oxide growth [5]. In addition there is a manufacturing thermal budget that deals with temperature control and uniformity across a wafer. Critical concerns for ULSI manufacturing include wafer temperature control during ion implantation, implantation damage annealing, sheet resistance variation of doped layers, oxide thickness control, and absolute and repeatable temperature measurements.

The requirements on equipment to meet the thermal budget needs of junction formation and oxidation include the following:

- a mechanical budget that satisfies the structural and mechanical aspects of processing large-diameter silicon wafers;
- a temperature uniformity budget across each wafer from run to run;
- an atmospheric budget for gases used in the furnace;
- a particle contamination budget;
- a time budget that deals with processing times as small as a few seconds and throughputs measured in hundreds of wafers per hour; and
- absolute, repeatable temperature measurements of wafers.

Improper processing conditions and wafer handling can lead to the nucleation of structural defects in silicon such as slip dislocations or wafer warpage, nonuniform oxide thickness, and irregular silicide contact interfaces. These effects produce concomitant problems with device junction leakage, lithography excursion, dielectric and contact nonuniformities, nonuniform junction sheet resistance, etc. Many such problems have been solved in large, multiwafer furnace annealing systems with large thermal masses. However, these systems have difficulty meeting all the time, particulate, and atmospheric budgets of ULSI technology. An alternative is to go to single-wafer systems using RTP.

Rapid thermal processing uses transient radiation sources such as arc lamps and graphite heaters to produce short-time, high-temperature, isothermal wafer processing. RTP can also be accomplished with continuous heat sources where the wafer is moved rapidly in and out of the vicinity of the heat source.

B. Ambient Control Budget

The smallest fabricated dimension in a MOSFET is the gate oxide thickness, which is grown by thermal oxidation. Ultrathin oxide growth requires careful process control and oxidation furnace optimization. In large-diameter batch furnace tubes, control of the furnace ambient is difficult because of backstreaming of air from the large open ends of the tubes. If control of the partial pressures of the oxidant gases (dry or wet oxygen) were the only variable in achieving 70 ± 3.5 Å oxides, then these partial pressures would have to be maintained at ±6% of nominal [1]. Trace amounts of water must also be minimized to the ppb range [6].

Silicon surface control prior to the growth of thin oxide layers is also important because of the fact that a native oxide grows on bare silicon at room temperature. Thus, 70-Å film growth can be controlled if the Si surface is HF cleaned, leaving the surface Si bonds terminated with H [7].

Desorption of the H at 300°C in a highly pure Ar gas ambient followed by the formation of one monolayer of oxide passivates the surface for subsequent gate oxide growth [8].

For ambient budget control, single-wafer RTP processing chambers offer a microenvironment approach that can satisfy the stringent requirements for ultrathin oxide growth using rapid thermal oxidation (RTO).

C. Mechanical Budget

The mechanical budget for ULSI manufacturing impacts on the patterning budget and the budgets for wafer defects and dopant profile control. Included in this budget are mechanical systems for alignment, wafer film stresses, wafer handling, and wafer flatness.

The patterning budget, overlay registration accuracy, Δ_t, is an important concern for ULSI. There are numerous contributions, Δ_i, to Δ_t that, if mutually independent, can be summed together in quadrature [9]:

$$\delta_t = \left(\sum_i \Delta_i^2\right)^{1/2} \tag{1.1}$$

Mechanical contributions to Δ_t include the following:

- Alignment system errors—due to limitations of the systems in lithographic printers for registering alignment marks or the masks to the wafer alignment marks.
- Wafer processing errors—due to changes in wafer feature dimensions from mechanical stresses of deposited films, high temperature processing, and etching tolerances.
- Mask and wafer mounting errors—due to deformations of the mask as mounted in the exposure tool or local changes in wafer flatness during chucking of properly selected, low-warpage wafers [10].

D. Contamination Budget

Scaling transistors to smaller dimensions has a profound effect on the manufacturing yield and reliability of integrated circuits. Processing complexity (i.e., the numbers of lithography levels) increases as devices become smaller. This added complexity is a result of the need for additional levels of metal to interconnect the increased number of subcircuits on a chip. Each added metal layer requires two or more film layers and two masks. Processing complexity is also increased by the need to overcome those material or circuit parameters that do not scale with decreasing device dimension. Such parameters include the metal–semiconductor work

function, silicon conductivity, and circuit operating voltages. A doubling of mask levels and films is expected as the technology is scaled down from 2 to 0.25 μm. In addition a similar doubling is expected in the number of process steps for manufacturing a chip [11].

These trends make devices more susceptible to contamination introduced by particulate and chemical impurities. Beside the increased amount or processing and, thus, exposure to impurities, smaller devices are susceptible to smaller defects and smaller amounts of chemical impurities that may cause chip loss [11]. For example, smaller devices have larger perimeter-to-area ratios, and defects along pattern edges are more likely to cause problems. Thinner oxides are vulnerable to smaller particles. Smaller devices biased with voltages that are not scaled produce higher internal electric fields that aggravate hot electron effects and oxide breakdown.

Particles can cause yield loss through the presence of random defects in the patterning of film levels. Chip yield is expressed in terms of the defect density through various statistical models such a Poisson distribution [12]:

$$\text{Yield} = e^{-A\rho} \tag{1.2}$$

where A is the chip area and ρ is the density of defects per unit area. On the basis of the device design parameter trends and forecasts of the allowable killer surface particle sizes and densities to achieve a *total* allowed defect density of 0.25/cm^2, it has been shown that 0.001 particles/cm^2 per step are required. The objective of 0.25 defects/cm^2 provides a yield of 78% for a 1-cm^2 chip for a Poisson distribution. Both the defect density and the killer defect size decrease as device dimensions decrease. These results are based on the rule of thumb that killer defects are at least 1/3 the size of a lithographic feature or 1/2 of a film thickness. With a gate oxide thickness of 70 Å, a 35-Å particle could be fatal! As a result, scaling device dimensions means that improved means must be found for controlling particle sizes and numbers in the processing environment.

The particle budget crisis is illustrated in Fig. 1. Measured particle densities are plotted versus particle size for airborne particles in a state-of-the-art semiconductor clean room [13], in bulk gases [14–16], in different semiconductor chemicals [17], and the minimum reported size distribution in deionized water. All the distributions in Fig. 1 show increasing particle densities with decreasing particle size. And this is the environment in which smaller devices will be made. Device scaling by a factor of two takes place in a processing environment in which the number of potentially fatal particulates in the air increases by four to eight times! The impact on yield may be devastating. Under these conditions a process that yields 25% and is limited by particle contamination would yield nothing after scaling down dimensions by a factor of two [11].

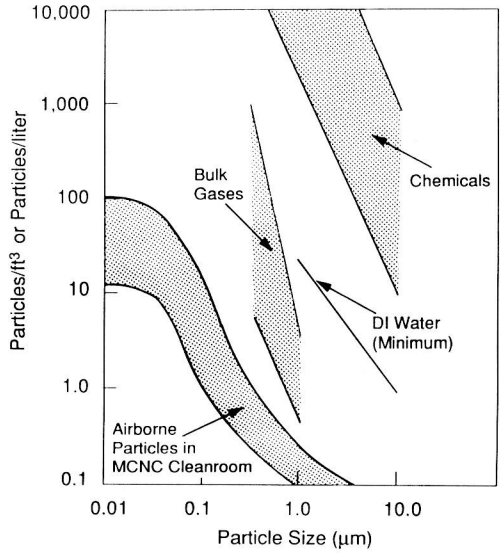


FIGURE 1. Comparison of particle size distributions in bulk gases, chemicals, deionized water and in processed air in the MCNC semiconductor clean room. The particle densities increase rapidly with decreasing size, which is the environment in which traditionally processed silicon devices are being scaled (after Fair, Ref. 1).

In contrast to what is depicted in Fig. 1 the National Advisory Committee on Semiconductors the United States has come out with consensus targets for particle densities in chemicals, in gases, and clean rooms. The target for Microtech 2000 in chemicals is 2000 particles/L of size greater than 0.02 μm, a reduction by a factor of 1000. Bulk gas targets for 0.02 μm particles are 0.02 particles/ft^3, and for airborne particles of size greater than 0.02 μm, the target is 1/ft^3! By extrapolation from current clean room practices and chemical purification methods, these targets will come only at great expense, if at all. In addition, the target for chemicals is much too high, paving the way for dry processing to eliminate wet chemistry completely. For example, it has been demonstrated that the number of particles added in an aqueous native oxide clean-up step is five times greater than in a vapor HF process [2]. Such gains in particle control will drive new strategies in silicon processing such as *in situ* dry cleaning.

Particulate and contamination control possible with *in situ* processing is also expected to be important in chemical vapor deposition (CVD) of thin layers. Rapid thermal CVD (RTCVD) processing is also consistent with the need for multiple, sequential processes such as RTO, rapid thermal nitration, and CVD polysilicon and etching. Other requirements for

RTCVD films that have not been quantified for gigachip manufacturing include film conformity and planarity, film stress, integrity, and the degree to which such films absorb or evolve water.

E. Electrical Budget

The electrical budget refers to the control of electrical device parameters that are determined by device scaling rules. Included in this budget are specifications for contacts, interconnections, electric-field levels, and process-induced electric charge. The electrical budget is shown in Table III.

For ULSI it is essential that low-resistance contacts to semiconductor junctions be made with high yield. These contacts must also be reliable, serving as barriers to unwanted metal reactions with silicon. Barrier layers of titanium/tungsten and titanium nitride have proven to be good choices.

Low-resistance contacts are imperative. If contact dimensions are halved, contact resistance increases by a factor of four. Thus, specific contact resistances must be decreased by factors of 10 or better.

Refractory silicides of transition metals have been used to improve contact resistances. Recent work with Al-TiW-$TiSi_2$ contacts to shallow n^+ junctions has been reported [19]. By performing sputter etching of the $TiSi_2$ surface to remove any oxides prior to TiW deposition in the same vacuum environment, specific contact resistances below 2×10^{-8} ohm·cm^2 can be achieved. Thus, the contact resistance budget is driving the use of *in situ* vacuum processing in an RTP, single-wafer module.

Table III Electrical budgets.

- Contacts
  - —High yield
  - —Low resistance (10^{-8} ohm·cm^2)
  - —Reliable
- Interconnects
  - —High conductivity (higher is better)
  - —Compatible
  - —Multilevel (2–4)
  - —Low electromigration ($J = 5 \times 10^5$ A/cm^2)
  - —Good step coverage
  - —Low stress
  - —Low interlevel capacitance
- V_t control
- Hot carrier injection (limiting E-fields)
- Radiation damage