(8-2)

CAD OF PROCESSES FOR SILICON HIGH SPEED DEVICES

B. Lojek
MOTOROLA, Inc.
Advanced Technology Center
2200 W. Broadway Rd., Mesa, AZ 85202

ABSTRACT
The complex fabrication processing of advanced silicon devices enhances the importance of process models. Processes such as RTA, RIE, low temperature epitaxy and extensive use of LPCVD polysilicon changes the properties of the silicon diffusion system by process induced defects and contamination by deep level impurities. In these cases phenomenological models often fail. In this review some of these issues are discussed mostly from device engineer (user of process simulation tools) point of view. In the comparison of three fabrication processes the source of main deficiencies in modeling theory is shown. also an improved diffusion model is briefly discussed.

INTRODUCTION

In the ideal situation process modeling should support accurate, predictive simulation of advanced technologies. If this were true, then the majority of the development of new technologies would be accomplished via simulators and the time required for such development could be dramatically abbreviated.

It is widely believed that process simulation of advanced processes will rescue the semiconductor industry from the prohibitive costs of developing and optimizing such processes experimentally. Our experience shows us that reality is different. It is the goal of this work clarify some issues connected with this topic.

Because most commercially available process simulation tools are prepared by the software houses, the using of these tools is so simple that almost everybody has immediate confidence in designing of fabrication processes using the computer. In the beginning of these activities the engineers often discover that the result are not even close to the experimental data. They conclude that CAD process modeling is either useless in the worst case or they try to discover the principle of the used models and localize the source of error in the best case. The conclusions from both groups, lead to a final statement that our knowledge of physics and chemistry of majority processes does not allow us to built process simulation tools which can be used for complex simulation of submicron VLSI fabrication processes.

One can find many reasons for this situation. One is that the modeling theory can not precede (except in a few applications) the progress in equipment used in the production of VLSI technologies. Additional reasons are related to the work of designers of TCAD tools and they are:

- lack of systematic approach in used theory
- poor knowledge of state of the art processing
- lack of discipline with regard to the experimental data

The combination of these and other factors raises many questions. The main questions addressed here are: is process modeling useless ? Where are the main benefits if the answer to the first question is no ? Even if there are serious limitations how we can maximize utilization for the design of submicron processes ?

Our experience show us that process simulation is a very useful complement of experimental work during designing of the new fabrication processes and a major benefit which can be expected from **sophisticated** (and we must emphasize this point) **process modeling is increased probability of first pass success on a new process.**

In the following section, we shall first discuss the details of device shrinking and consequent changes in processing, followed by a presentation of the concentration profile engineering where we will manifest the major deficiencies in TCAD. Finally, we shall consider an example application modeling the submicron structure.

DEVICE SHRINKING AND CONSEQUENT CHANGES IN PROCESSING

To illustrate the type of requirements imposed by technological processing on modeling capabilities of TCAD tools we compare three processes which were developed in the same facility during the last decade. All processes are designed for ECL type of circuits.

From the Table I. one can make the following conclusions

- The most dramatic increases of processing steps is in reactive ion etching and LPCVD polysilicon depositions. Both processes are characterized by high level of contamination.
- Very significant increase can be seen also for Rapid Thermal Annealing steps.
- All devices in new technologies are built on epitaxial silicon with a very low oxygen concentration.

An inevitable and yet commonality of all the process steps that are increasing in frequency most rapidly in advanced technologies is the production of damage of the crystal lattice with production of native defects and contamination by deep impurities.

The native defects (vacancy, self interstitial, interstitial) are believed to play a dominant role in diffusion of impurities and in determining the electrical properties of semiconductor materials. Diffusion of impurities following ion implantation (with damage of the structure after implantation) is often discussed in recent literature and several models have been proposed to explain these phenomena [1]. Intuitively we can expect some anomalies in the redistribution of impurities. However, Misra and Hasell [2] show that even in RIE processing where the wafer temperature goes up to 300 °C (as a function of RF power density) this temperature is sufficient to induce the diffusion of native point defects and therefore change the diffusion behavior. But the annealing behavior of RIE induced defects are relatively unknown. The situation is similar for RTA processes. These circumstances result in a very poor basis for the creating of models for TCAD.

Another very important conclusion can be deduced from Table I. The next generation of technological processes will extensively use single wafer processing equipment where there is no possibility for monitor wafers. Therefore we cannot (unless we add an additional test structure) verify the results and reproducibility of processing. This point has very important consequences for the acquisition of experimental data which we need for the design of models appropriate for TCAD. Generally the data can be extracted from the special objects where processing conditions could be slightly or significantly different than the process conditions on the structure which is under investigation.

CONCENTRATION PROFILE ENGINEERING

The typical approach of developing new fabrication processes usually assumes that we know target concentration profiles for the devices. The target profile of devices is determined by a variety of approaches: intuitively, based on our previous experiences or as a result of relatively predictable device simulations. Profile engineering is a process of designing of individual technological operations in such a way that we can create the desired structure and at the same time meet the requirements defined by the target profile.

The process steps used in VLSI fabrication technologies can be divided into three modules. Every module is typically formed of the following substructures:

- isolation structure and bulk of active devices
- active devices
- interconnections and contact

Each module has a unique position in the process flow and is characterized by range of processing temperatures. Fig. 1 shows a simplified diagram of processing temperatures and times for three processes which we mentioned earlier. For example, process M5 is 100 minutes in a range of temperatures higher than 1000°C, 855 minutes in a range of temperatures between 900°C and 1000°C and 120 minutes at temperatures between 800°C and 900°C. Surprisingly and contrary to common myth, the thermal budget of advanced processes is not lower than "conventional" (not submicron) processes. The time of processing with temperatures higher than 1000°C is decreasing, but the total time of thermal processing is longer. If we introduce the performance index as product of the processing time and the processing temperature we can find:

process	performance index [°C min]
M1	550E3
M3	890E3
M5	960E3

From the viewpoint of diffusion and redistribution of impurities, this does not mean a trend in the wrong direction because the main part of the higher thermal budget is for the more complicated isolation structure. If we compare the same performance index for steps after deep collector formation:

	M1	M3	M5
TECHNOLOGY	BJT	BJT	BiCMOS
YEAR	1982	1986	1990
PRIMARY APPLICATION	ECL	ECL	ECL
ISOLATION	LOCOS	LOCOS	TRENCH
LITOGRAPHY (AS DROWN) [μm]	2	1 75	1 2
BASE THICKNESS [A]	2500	1200	900
# MASK STEPS APPLIED	14	16	20
# IMPLANT STEPS	7	7	15
# REACTIVE ION ETCH STEPS	0	12	25
# CRITICAL WET ETCH STEPS	0	0	0
# FURNACE DIFFUSION STEPS	1	0	0
# THERMAL OXIDATION STEPS	3	5	6
# HIGH PRESSURE OXIDATION STEPS	1	1	3
# RAPID THERMAL ANNEAL STEPS	0	1	5
# FURNACE ANNEAL STEPS	3	6	7
# LPCVD OXIDE DEPOSITION STEPS	0	3	7
# LPCVD NITRIDE DEPOSITION STEPS	2	4	4
# LPCVD POLY DEPOSITION STEPS	0	4	8
# EPITAXIAL GROWTH STEPS	1	1	3
# ETCH BACK PLANARIZATION	0	1	2
# SILICIDE DEPOSITION STEPS	0	1	1
# METAL 1	1	1	1
TOTAL NUMBER STEPS	33	63	107

Table I. Comparison of processes M1,M2 M3

NOTES
1 process M1 uses implanted emitter
2 process M3 and M5 use a polysilicon emitter
3 All processes are selfalignment processes
4 In M1 deep collector follow the base/emitter formation
5 Number of implantation steps in M5 include the threshold and punchthrough implants

process	performance index [°C min]
M1	74E3
M3	110E3
M5	135E3

In all cases, after formation of the isolation and wells we have a concentration profile which is exposed to thermal treatment shorter than 120 minutes. This treatment has a small impact on the bulk concentration profile. This is illustrated in Fig. 2 and Fig. 3 where the changes after the projected thermal budget for typical processing of an active device (i.e. gate, source/drain, base, emitter) at 950°C are shown for the NPN and N–channel transistor respectively. From both plots it is clear that the changes in redistribution of impurities are very small.

THE MAJOR DEFICIENCIES IN PROCESS MODELING

Before we go further in our discussion we need to clarify questions regarding the dimensionality of process simulation tools. The typical opinion is that submicron devices necessitate two–dimensional process simulation. In our view the major problems where lateral diffusion is important can be solved in one dimension with appropriate interpolation into the second dimension. Where two dimensional solutions are needed is for diffusion in/from anisotropic materials (as is polysilicon with a columnar structure). In Fig 4 and Fig 5 a comparison is shown between boron diffusion from polysilicon in two structures with an oxide/nitride window. The interface area between the polysilicon and silicon is the same in both cases. From calculated data we can see that the error in junction depth determination is higher than 25%. This factor, and the neglecting of anisotropic diffusivity in polysilicon, leads to a poor prediction of emitter junction depth. This is true where arsenic is implanted into nonplanar polysilicon and diffuses into silicon through a tapered oxide spacer window with a very big aspect ratio w_{top}/w_{bottom}.

The difficulties accompanying process modeling are very well know; for a general review of this issues see for example [3]. If we focus only on diffusion modeling then the main deficiencies can be listed as:

–Diffusion principle is not defined.

–Vacancies are considered to promote diffusion by formation of defect–impurity pairs, but the process of formation of pairs is not explicitly included.

–Nonequilibrium phenomena are not included.

–Annealing of point defects is not included.

–The contribution of individual point defect or defect–impurity pairs to diffusion is not known.

–No diffusivity dependence on heating and cooling rates are accounted for

–No dependence of diffusion behavior on the thermal history of the wafer is included.

–An inconsistency in low and high temperature diffusion behavior

Our experience shows us that there is another source of difficulties which is not very often remembered – immense discrepancies between active and chemical concentration (especially if RTA is used in processing) as will be shown in our comparison of activation efficiency by RTA compared to a conventional furnace anneal.

In all commonly available process modeling programs, basic phenomenological diffusion models are used. After many years of application they have been relatively well "calibrated" (fitted) to a large class of equilibrium diffusion problems. Unfortunately, these diffusion models fail in some increasingly important cases such as diffusion at high concentration, diffusion at low temperature (reverse annealing or deactivation), redistribution of impurities from diffusion sources with very steep concentration gradients or diffusion during rapid thermal (RTA) processing. The acquisition of concentration profiles from experiments to "calibrate" the models depends on one or more analytical techniques: SIMS, RBS, incremental Hall and sheet resistance measurement, CV profiling and spreading resistance measurement. All methods, but especially the last two methods suffer from certain limitations. The diffusion length involved in these phenomena is so small that even small measurement error leads to a large uncertainty in the determination of the diffusion coefficient. This is why the analytical data used for "calibration" of these models has been primarily based on the chemical concentration of impurities due to the good accuracy of SIMS. However, the electrically active concentrations are of primary interest.

Two experiments were conducted to investigate arsenic diffusion during RTA. In the first, we have measured chemical and active concentration profile, for samples annealed under similar conditions in the RTA equipment (PEAK ALP 6000) and furnace (TYLAN). The substrates used in this experiment were all 100 mm diameter <100>, CZ, silicon wafers of resistivity 6.0–8.5 Ωcm, boron doped. One group of wafers was thermally oxidized to an oxide thickness of 40nm. LPCVD polysilicon of thickness 180nm was deposited on the second group of wafers after an HF etch. Arsenic with a dose of 1E16 cm^{-2} and an energy of 50 keV was implanted into both group of wafers. Annealing was performed in a nitrogen ambient with the temperature–time profile shown in Fig.6. The intention was to create the same thermal cycle in both the RTA and furnace anneals. SIMS and SRP were used to determine the concentrations profiles (Fig.7 and Fig 8). Significant differences were observed in the active concentration profiles even though the chemical profiles were essentially identical. The rate of temperature increase and the time at temperature (60 sec at 900°) for the isothermal portion of the anneals were the same. Special attention was given to the determination of the cooling rate of all wafers in these experiments. For samples annealed in the RTA system the cooling rate from 900° to 750 °C was 70 °C/sec. For samples annealed in the furnace and cooled in free air, the cooling rate in the same range of temperatures was 40 °C/sec

In the second experiment, the role of temperature ramp rate during RTA was further explored. The experiment was performed on the same type of samples with a 40 nm thick thermal oxide and an arsenic implant of 1E16 [cm^{-2}] at 50keV. The first group was heated up with a "slow" programmed heating rate of 0.333 °C/sec and cooled down with a "fast" programmed rate of 47.5 °C/sec. The relative ramp rates for the second group of wafers was reversed, keeping the time at temperature for the isothermal portion of the anneal (10sec at 950°) and the total thermal budget the same (see Fig 9). SIMS and differential Hall measurement of carrier concentration and mobility were used for acquisition of chemical and electrically active concentration profiles, respectively. As seen from the measured data (Fig 10), the "fast" cooling rate, which was approximately 150 times greater than the "slow" cooling rate, gave an active concentration that was only 2 times higher. We can conclude that the faster cooling rate in the RTA anneal can contribute to higher activation efficiency, but another variable is clearly responsible for the majority of the effect.

The experiments described above were not simulated correctly by using process simulation tools.

Many of the problems listed above can be solved (or at least improved) by correct determination of the Fermi level of the diffusion system and appropriate dependence of diffusivity on concentration of point defects. A relationship which describes the concentration of point defects in the form n/n_i [*] does not follow the temperature dependence of the point defects, the dependence of diffusivity as a function of doping concentration with this approximation does not lead to "saturation" of the concentration of point defects. Also, the relative changes of concentration of individual charged defects with doping level are independent of activation energy. However, it is known from the behavior of degenerate semiconductors that, in the limit, further increases in impurity concentration cannot increase the concentration of point defects.

ADVANCED MODEL FOR IMPURITY DIFFUSION IN SILICON

Several new models for modeling of diffusion processes including RTA have been developed recently [4]. However, these approaches use a physical formalism which has the same deficiencies mentioned above, and are improved only in that a higher number of fitting parameters in the diffusion coefficient allow for higher degree of freedom in calibration.

We extended the previous work on the point defect diffusion mechanism in silicon and propose an improved version of the diffusion model. We solve for the Fermi level and eliminate certain ideas lacking acceptable physical background, for example the fractional interstitial factor f_I. The relative importance of each mechanism **depends on the time evolution of all species present in the system**. This model is described by the following equations:

$$\frac{\partial[A_s]}{\partial t} = k_i^-[A_i] - k_i^+[A_s][\,I\,] + k_v^-[A_i][V] - k_v^+[A_s] - k_{A,V}^-[A_s][V] + k_{A,V}^+[A_sV] \quad (1)$$

$$\frac{\partial[A_i]}{\partial t} = \nabla\left[D_A\nabla[A_i]\right] - k_i^-[A_i][\,I\,] + k_i^+[A_s][\,I\,] - k_v^-[A_i][V] + k_v^+[A_s] \quad (2)$$

$$\frac{\partial[A_sV]}{\partial t} = \nabla\left[D_{A,V}^{ch}([V(\Phi)])\nabla[A_s] + D_{A,V}^{-}[A_s]\nabla[V] + \frac{q}{kT}D_{A,V}^{d}[A_s]\nabla\Phi\right] + k_{A,V}^-[A_s][V] - k_{A,V}^+[A_sV] \quad (3)$$

$$\frac{\partial[V]}{\partial t} = D_V\nabla^2[V] - k_R\left([\,I\,][V] - [\,I^*\,][V^*]\right) - k_{A,V}^-[A_s][V] + k_{A,V}^+[A_sV] \quad (4)$$

$$\frac{\partial[\,I\,]}{\partial t} = D_I\nabla^2[\,I\,] - k_R\left([\,I\,][V] - [\,I^*\,][V^*]\right) \quad (5)$$

and the total (chemical) concentration of impurities is

$$[A_T] = [A_s] + [A_i] + [A_sV] \quad (6)$$

For a list of the used symbols and the method of determination of the Fermi level see [5]. This model is currently undergoing computer implementation. The numerical values of the parameters resulting from the analytical calculations (with theoretical physical support) seem to be realistic and they are verified with systematic experimental data.

We believe that a new generation of diffusion models will be based on a similar set of equations. The major benefit of this approach is that the active concentration is solved during time evolution of the diffusion system and not "after processing" from knowing the chemical concentration.

CONCLUSION

The merit of an application of process modeling depends on the examiner's viewpoint. Certainly, the contemporary available process modeling tools are sufficient for educational purposes. However, for the design of processes for submicron devices where not only the thermal treatment but also the shape of the structure determine the diffusion behavior, these tool are still insufficient. Thus a major part of the design phase of new technological processes must still be based on experimental work.

In engineering practice, combination of experimental data and simulation is necessary to achieve the acceptable predictability of TCAD tools. For example the module which creates the isolation and bulk of the devices can be characterized almost fully by measurement on "big" test objects. Loading of the concentration profile from these measurement into a grid prepared according to SEM cross–sections we can eliminate errors connected with the epitaxy (autodoping) and high space changes of structure during high temperature processing. These problems are not traditionally predicted satisfactorily by TCAD tools.

Even though we discussed many problems with process modeling, in our opinion the process simulation can contribute to our knowledge of processing phenomena and we can improve our level of understanding of the underlying theory with practical results. Even if we are not able create the better numerical models we can benefit from qualitatively higher level of understanding of our structures. This fact is not negligible.

Our discussion was limited to high speed bipolar processes, therfore it is fair to note that the predictability of the simulation results for MOS processes is slightly better than for bipolar processes.

[*] This quantitative estimation of the dependence of the equilibrium concentration of vacancies on the doping level is built on a thermodynamic approach applying the method of Lagrangian multipliers for minimization of free energy of the system. The thermodynamic approach does not show the source of point defects and this leads to incorrect conceptualization of the defect fluxes from or to the surface interface.

REFERENCES:

[1] R.B.Fair, in "Impurity doping processes in Silicon", ed. F.F.Y.Wang, North Holland, Amsterdam 1981

[2] D.Misra, E.L.Heasell, "Annealing behavior of Reactive Ion Etching induced deep levels J.Electrochem. Soc. Vol.137 (1999) p.1559

[3] Y.Kim, H.Z.Massoud, R.B.Fair "The effect of ion–implantation damage on the dopant diffusion in silicon during shallow junction formation", J.Electrochem. Mat., Vol. 18 (1989), p.143.

[4] M.Heinrich, M.Budil, W.Potzl, " Simulation of Arsenic and Boron diffusion during Rapid Thermal Annealing in silicon", ESSDERC'90, Univ. of Nottingham, 1990

[5] B.Lojek, "Concentration of charged point defects in silicon diffusing system", Phys. Rev. Lett., submitted to publication

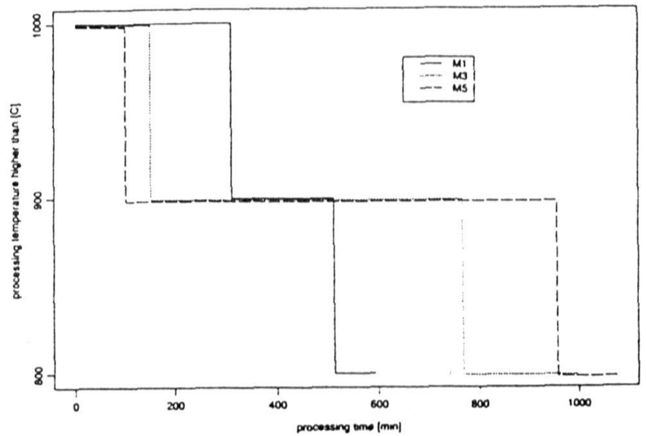

Fig.1 Simplified diagram of processing temperatures and processing times for processes M1,M2,M3.

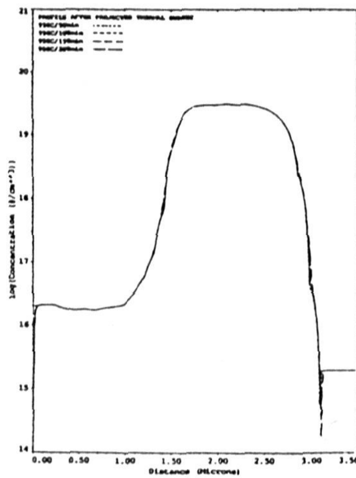

Fig.2. Changes in redistribution of impurities after projected thermal budget for NPN transistor.

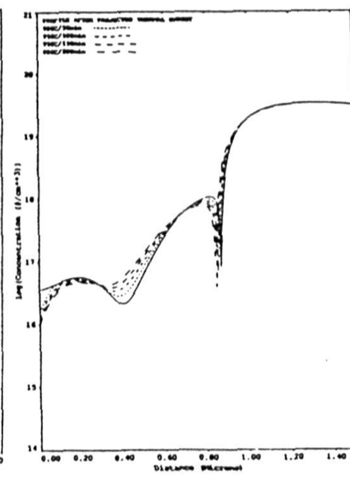

Fig.3. Changes in redistribution of impurities after projected thermal budget for N–channel transistor.

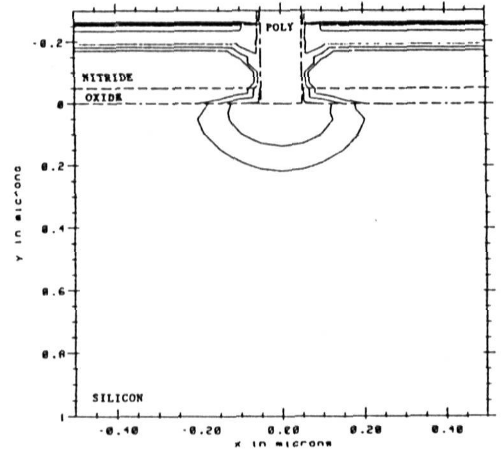

Fig.4. Diffusion from boron doped polysilicon in rectangular window.

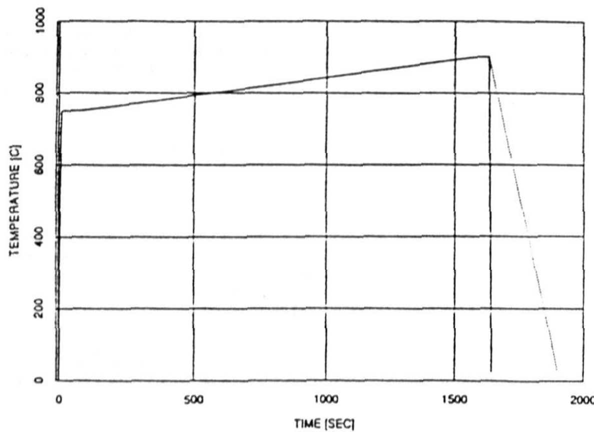

Fig.6. Temperature–time profile for anneal of wafers in RTA (solid line) and furnace (dotted line).

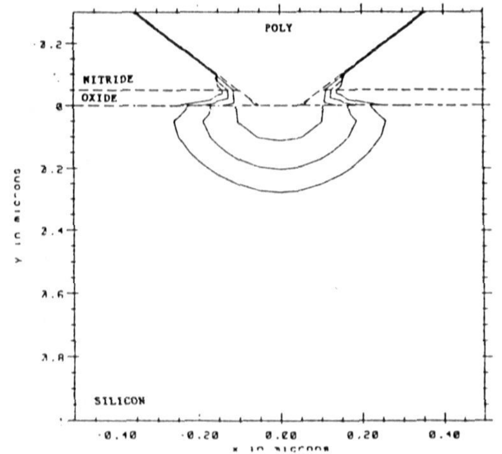

Fig.5. Diffusion from boron doped polysilicon in tapered window.

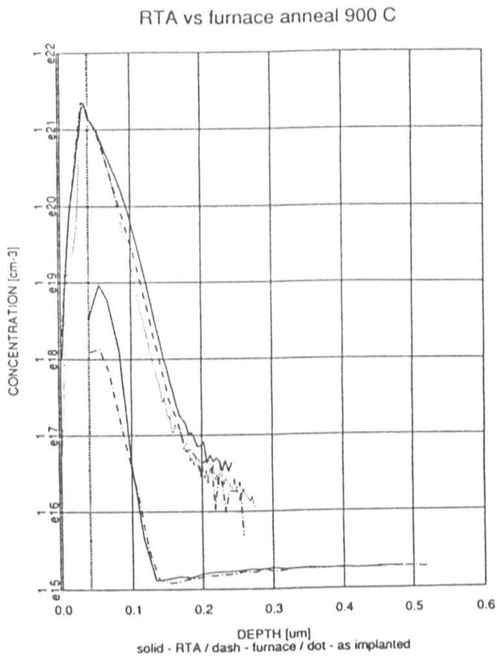

Fig.7. Concentration profiles for arsenic implanted into silicon; as implanted (dotted line), for active (lower 2 curves) and chemical (upper 2 curves) after anneal by RTA (solid line) and furnace (dashed line).

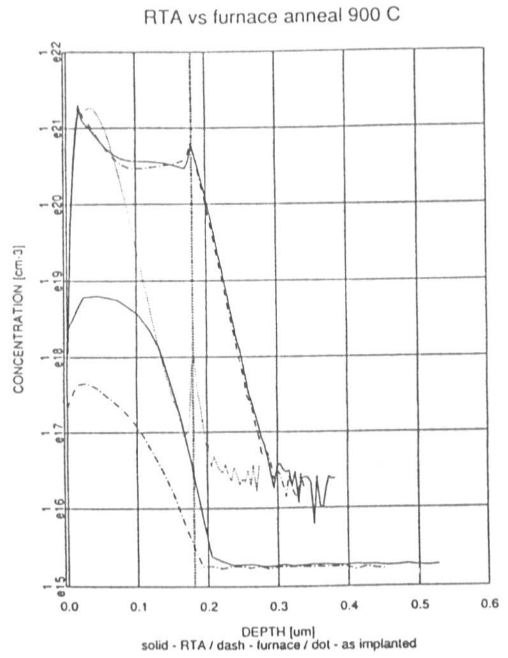

Fig.8. Concentration profiles for arsenic implanted into and diffused from polysilicon; as implanted (dotted line), for active (lower 2 curves) and chemical (upper 2 curves) after anneal by RTA (solid line) and furnace (dashed line).

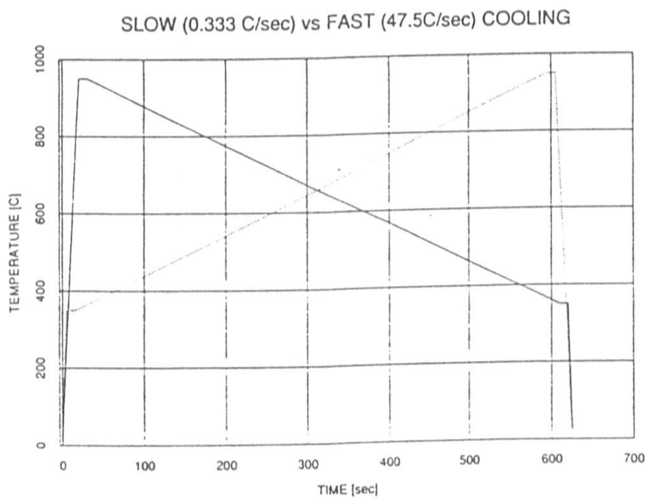

Fig.9. Temperature–time profile for RTA annealing with fast (dotted line) and slow (solid line) cooling rates

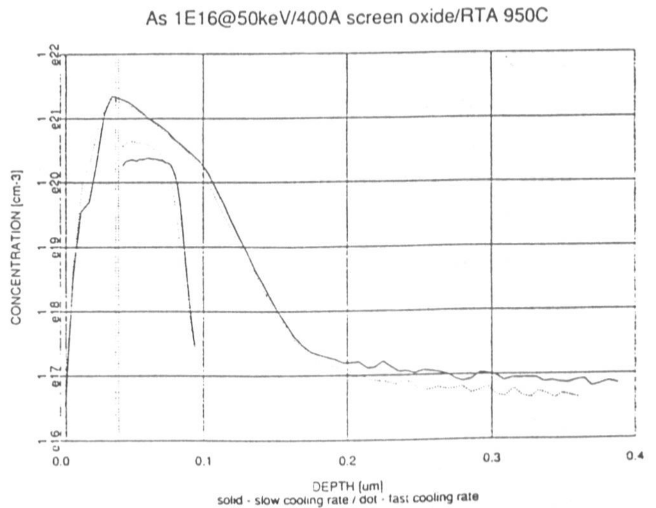

Fig.10 Active (lower 2 curves) and chemical (upper 2 curves) concentration for fast (dotted line) and slow (solid line) cooling rate during RTA anneal