METHODS IN RAPID THERMAL ANNEALING

A. T. Fiory

Bell Laboratories, Lucent Technologies, Inc.
700 Mountain Avenue, Murray Hill NJ 07974

This presentation highlights some of the techniques in common practice for rapid thermal annealing (RTA) of silicon. Progress and the level of maturity attained in RTA technology owes much to its widespread use in critical processing steps for the manufacture of silicon integrated circuits. Procedural considerations are reviewed from the viewpoints of process research and development, tool development, and RTA operations in mass production.

INTRODUCTION

From an historical perspective it can instructive to begin by reviewing a case where rapid thermal annealing (RTA) techniques were developed and introduced for Si processing. A rapid anneal for manipulating the placement of dopant atoms in crystal Si was an early critical application. A method called "gold spiking," where Au atoms are diffused and dispersed in the Si to control device storage time, was being used in the manufacture of n-p-n transistors on n/n^+ -type epitaxial Si wafers (1.25 inch dia.) for integrated circuit modules at the former Western Electric Company. It was necessary to use a spike anneal with a rapid quench to achieve high yield in production. Initially, manufacturing used an automated furnace (set at 1250 ^{o}C) with a quartz paddle (a wafer holder) that was rapidly inserted and then withdrawn after reaching the process peak temperature (1160 ^{o}C). Temperature was controlled by using a thermocouple on the holder and a two-step cycle: the first for preheating a bare holder and the second for heating (14 s) and quenching (3 s) the Si wafer [1]. However, improvement in the technique was being compelled by drift and nonuniformity in the quench step in the anneal, mainly from deformation of the quartz holder.

In 1966 W. K. Mammel, at the former Western Electric Engineering Research Center, and colleagues developed an alternative method using incandescent lamps [2]. The prototype spike RTA used a Research Inc. heater with six 2000-watt lamps to achieve better reproducibility and control over the heating cycle time (which varied ± 0.5 s with the furnace method) and faster heating than could be achieved in the furnace, owing to a higher lamp filament temperature (2200 ^{o}C). Many of the benefits and control challenges of incandescent lamp RTA were recognized when this technique was implemented. The heater was modified with reflectors to improve uniformity of illumination. The Si holder used N_2 gas levitation and rotation to improve thermal isolation and uniformity in heating. Radiative heat transfer was modeled on the spectral properties of the Si in the infrared, namely absorptance, transmittance, and reflectance. The relationship between wafer temperature and lamp power was measured by a thermocouple. A practical approach was used to establish the optimum process for gold spiking by surveying results from a series of 5-s treatments at various lamp power settings. Process control relied on a basically open-loop method with preheat and annealing cycles of fixed lamp power and time.

Proceedings of RTP 2000, Eighth International Conference on Advanced Thermal Processing of Semiconductors (Gaithersburg, Md., Sept. 20 - 22, 2000), pp. 15-25.

An issue that remained challenging within the field for the next few decades was the question of how to correctly measure actual temperatures of wafers in RTA chambers. Temperature sensors have been based mainly on thermocouple junctions, infrared detectors, or a temperature dependent physical property of Si like thermal expansion, index of refraction or optical absorption in the infrared, and Lamb wave propagation. Infrared pyrometry has been the preferred method because the nearly logarithmic dependence of temperature on detector signal yields better precision than most alternatives. However, pyrometry requires determining the emissivity of the wafer and separating thermal emissions from the wafer and the heater in an environment where they are mixed by internal reflections. Pyrometry with multipoint reflectivity measurements is one of the techniques being used for heated bodies in a furnace [3]. Some of the currently preferred methods for wafer temperature measurement and control are reviewed below in the context of the RTA systems for which they are being used.

Implementations of RTA have remained essentially the same in that the heat is incident normally on one or both of the broad faces of a wafer to achieve uniform temperature transients on the order of 10 to 10^3 $^{o}Cs^{-1}$. Heating capability is therefore much faster than in batch methods with stacked wafers (e.g., fast ramp vertical furnace), where wafer pitch and other geometrical constraints interfere with lateral heat flow. The RTA method may use either lamps or furnaces, depending on the required temperature range and cycle time. Furnace techniques are well suited for low temperatures, such as in the treatment of dielectric and conducting films at temperatures below about 500 ^{o}C, and for accuracy at high temperatures where extremely sharp spiking may not be required. Optical methods, mainly with incandescent or arc lamps, have general applicability and are particularly suited to short high-temperature cycles.

A number of authors have prepared excellent reviews of RTA methods and the reader is referred to some of the literature [4-9]. This paper touches upon methods for heating, temperature measurement, and process control. Commercial implementations now use principles of holistic design with custom heaters, temperature sensors, and control mechanisms and algorithms [10]. Chambers are designed to control the flow uniformity and thermal buoyancy of purge gases in addition to the handling and heating of wafers. A related topic, which has intersected the history of RTA, is laser annealing [11]. It is receiving renewed interest, e.g., for dopant activation, by either melt quenching a surface layer or as a sub-melting pretreatment prior to a standard RTA.

HEATING METHODS

As indicated in the introduction, techniques for RTA fall into two types based on heating method: furnaces using steady heat sources and mechanical wafer transport; or electrical lamps, either incandescent or arc, with programmed optical output cycles. Lamp systems with shutters have also been designed. Figure 1 shows the temperature vs time profiles obtained for several RTA methods of each type. The challenge for furnace methods is to properly manage the thermal equilibration of a wafer introduced into a furnace environment. The challenge in lamp

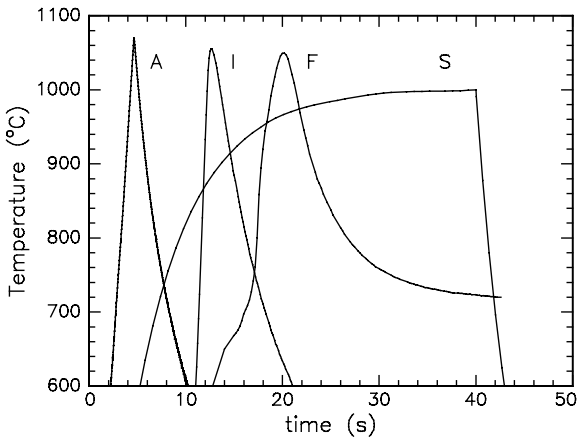

FIG. 1. Temperature sensor signal vs time for four RTA methods. A, arc lamp; I incandescent lamp; F, bell-jar furnace; S, susceptor furnace. Preheat region below 600 ^{o}C varies among methods. Origin of time is start of heating cycle in each case.

based systems is to control the environment in a chamber that is far from thermal equilibrium. Several problems are discussed next which tend to be more difficult for lamp based systems, but generally are to be addressed for accurate heating by RTA.

From basic heat transfer considerations it was recognized that uniform flux from a bank of infrared lamps in a reflective enclosure may not necessarily lead to a constant temperature across the wafer, even in steady state mode [12]. Non-uniformity is caused by variation in the view factors for direct and reflected radiation incident at various points on the wafer, with the wafer edge being a particularly more exposed region. The expedient solution, useful also in furnace methods, is to enlarge the effective wafer diameter by surrounding it, or supporting it, with an annular ring of similarly absorptive material, preferably silicon of the same thickness as the wafer. The remaining variations in heating uniformity are compensated by tuning the power distribution among the lamps. The spatial resolution and residual error is given by the lamp spacing.

A related point is that differences in optical absorptivity across wafers, from the patterns for various regions of the integrated circuits and the areas between them, lead to local temperature variations if left uncompensated [13]. Film layers, doping and texture cause the emissivity of wafer surfaces to depart significantly from that of intrinsic Si [14-16]. Normal occurrences of emissivity variability lead to significant uncertainties in wafer temperature with open-loop methods. One solution is to reduce the effect of a freely radiating environment for the wafer by surrounding the wafer with a thermal shield at nearly the same temperature as the wafer, making a pseudo blackbody cavity for quasi thermal equilibrium [17]. This is implemented by placing a heated wafer-like shield near one or both faces of the wafer, or by using a proximity reflector in which a blackbody illusion is approximated by multiple reflections [18]. The single shield method for lamp heating has to take into account temperature offsets between the shield and the wafer. Differential radiation effects are less pronounced for furnace RTA methods, since the heater and wafer temperatures are closer to each other.

Furnace RTA

A bell jar furnace method for annealing implant monitors was developed at High Temperature Engineering and recently improved with an emissivity compensating temperature measurement scheme [19]. Figure 2 shows a rough schematic. The wafer, on a support platform, is elevated into a heated zone with a vertical temperature gradient. The maximum temperature, at the top of the bell jar, is set to be about 200 ^{o}C above that of the desired process temperature. The comparatively weak sensitivity of wafer temperature to wafer emissivity within the near blackbody environment at high temperature has made this a viable method for open loop operation in some production work.

Elevator motion is sufficiently fast to produce thermal spiking with 0.5 s dwell time within 2 ^{o}C of peak temperature (Fig. 1, curve F). Under closed loop temperature control with

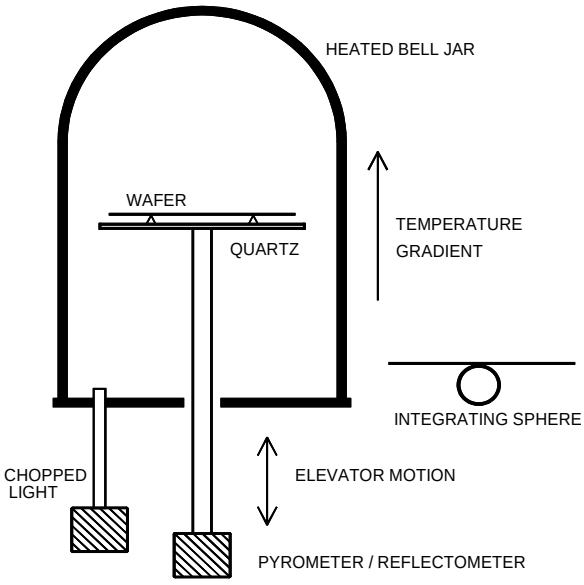

FIG. 2. Furnace bell jar method, after Ref. 19. Emissivity corrected pyrometer uses reflectometers with in-situ chopped-light and ex-situ integrating sphere.

a pyrometer focused on the back side of the wafer, some of the best results for process reproducibility and uniformity have been claimed.

The pyrometer for the bell jar method detects the thermal energy flux emitted by the wafer and measures the emissivity of the wafer. Emissivity, *E*, is defined for practical purposes as the ratio of thermal emission received from the wafer relative to that of a blackbody at the same temperature (see Ref. 17 for more scientific rigor). The operating wavelength of the pyrometer, about 0.95 μm, is above (in energy) the Si band gap [14] where optical transmission through the wafer can be neglected. An expression derived from energy conservation and the Kirchhoff law, which reduces to $E = 1 - R$, is used to compute emissivity from a separate measurement of reflectivity, *R*. The reflectivity is determined by first mapping geometrical and background reflection factors as a function of elevator height and then normalizing bidirectional reflection signals to a separate ex-situ precision measurement of *R* at room temperature using hemispherical illumination. Wafer temperature is then computed from the Planck radiation law and a gauge factor for the detector sensitivity.

A different approach to the furnace method is that used by Mattson Technology, which has chambers for two wafers. Wafers are lowered close to a heated susceptor in a thermally baffled enclosure. The equilibration time constant (~10 s at 1000 ^{o}C) determines the heating and cooling profiles, which are monitored by a fiber-optic pyrometer probe (Fig. 1, curve S). The design objective is to take advantage of the thermal equilibrium environment of a conventional furnace. The duration of a typical anneal cycle might be 80 s.

Incandescent lamp RTA

This method uses reflectors to boost the radiative coupling from quartz-halogen-tungsten lamps to the wafer. The sensitivity of heating uniformity to wafer emissivity generally increases with chamber reflectivity and specularity. Two primary approaches have evolved to deal with the uniformity problem, depending on whether the heat is applied to only one face of the wafer, or to both faces (and, in some cases, also to the edges).

Single-sided heating methods were developed for an array of spot lamps connected in annular zones (such as Stanford / Texas Instruments MMST, Applied Materials Centura, AG Starfire) or linear lamps (AG Integra, Applied Materials) with wafer rotation usually incorporated. Multiple temperature sensor inputs and model-based algorithms are used to articulate multiple power outputs to the lamps during the RTA cycle.

Dual sided heating methods can deliver roughly twice the heating power to a wafer in a chamber containing lamps and quartz for ambient isolation and wafer support. A gray and non-specular inner surface for the chamber was adopted by AST Electronik for efficiency in heating, reduced sensitivity to wafer emissivity, and single point temperature sensing. Predetermined power profiles are used in this case to drive the lamp array. The power distribution is determined by optimizing the uniformity of desired processes from wafer map measurements and a chamber model.

Both methods give improved uniformity with wafer rotation to smooth out azimuthal variations in heat transfer. In systems with multiple lamp zones and feedback control, wafer rotation is required to average the readings of the temperature sensors. The controller maps the system to a one-dimensional function of radial distance and has a characteristic response time that scales with the rotation period.

In single-sided heating methods, the temperature sensors are usually located on the opposite side from the lamps. Interference from the lamps is then readily suppressed either by a combination of optical baffling and wafer absorption (Centura, pyrometry near 0.85 μm) or by a cooled selective filter between the lamps and the wafer (AG Starfire, hyrdroxyl doped quartz filter for pyrometry near 2.7 μm). A high reflectivity plate is placed close to the wafer to create a pseudo blackbody cavity by multiple reflections [20]. Figure 3 shows a detail of a

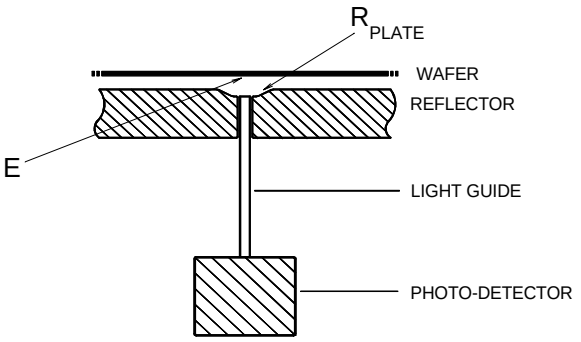

FIG. 3. Detail of a proximity sensor configuration (one of a pair) for sensing wafer temperature with partial compensation for wafer emissivity, E (after Ref. 21).

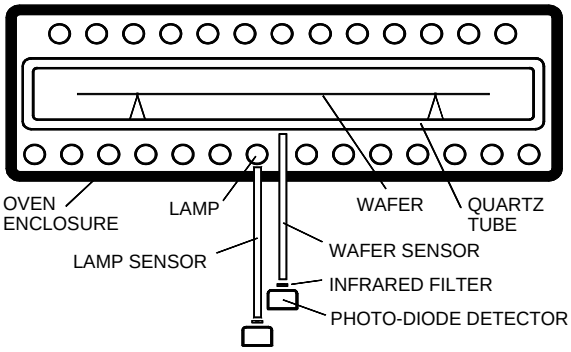

FIG. 4. Configuration of sapphire light guide sensors for a ripple technique in an oven with AC-powered incandescent lamps and quartz isolation tube (after Ref. 22)

proximity sensor using this principle [21]. Modeling the effective reflectivity of the plate as R_{PLATE}, the effective emissivity of the wafer is expressed as $E\,[1 - R_{PLATE}\,(1 - E)\,]^{-1}$. Effective emissivity approaches unity as R_{PLATE} approaches unity. Extrapolation to unity is obtained by using a pair of such sensors with two different and pre-determined $R_{PLATE} < 1$.

Earlier pyrometer designs for chambers with dual-side heating lamps used a view port in the quartz isolation tube surrounding the wafer and operated at wavelengths where the tube (pure or hydroxyl doped) absorbs the lamp radiation. Curve I in Fig. 1 was obtained by such a pyrometer. The main drawback of this approach is the uncertainty in wafer emissivity, since a fixed emissivity is used to obtain temperature. Also, heat radiated by the quartz scatters into the pyrometer. AG Associates developed a separate sensor for the quartz tube to correct for drift from quartz heating, but errors remain because compensation varies with wafer reflectivity as well as quartz temperature.

A ripple technique was proposed at Accufiber to treat the lamp interference and wafer emissivity problems together [22]. Two sapphire light-guide probes are arranged inside the chamber: a wafer sensor oriented to collect radiation from the wafer and a lamp sensor with opposite orientation to sample incident radiation. Figure 4 shows a ripple technique configuration with sensor probes positioned outside the quartz isolation tube. Rapid AC modulation in the lamp power creates AC components in the signals for the two detectors, the ratio of which is related to wafer reflectance and thence emissivity. The reflectance is also used to subtract out reflected signals from the lamps. The equivalent blackbody wafer signal is then used to derive temperature.

Pyrometers with probes outside the process tube near the oven wall were developed at Bell Labs and have been tested in Lucent manufacturing lines [23]. Thermocouple wafers of varied emissivity are used to profile the sensitivity settings for the two sensors. Versions from Luxtron, AST and Steag have extended the concept of Fig. 4 with additional individual sensors for the lamps [24]. Current ripple pyrometers usually operate at a wavelength near 2.3 μm where quartz absorption and emission can be neglected. Optical transmission through Si (for lightly doped Si wafers at temperatures below 650 °C) is detected by phase shifting the modulation of selected lamps above the wafer.

Arc lamp RTA

Arc lamp heating, which has seen early use in commercial RTP equipment, has been proposed at Vortek as a practical alternative with potentially significant improvement [25]. A simplified schematic is shown in Fig. 5. With back side heating, the device structures are not

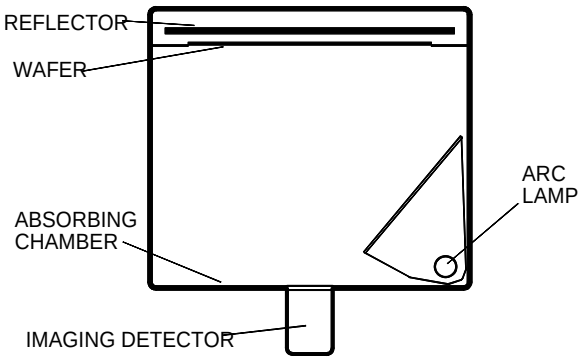

FIG. 5. RTA chamber with water-wall arc lamp and imaging wafer temperature sensor (after Ref. 25).

exposed to direct optical excitation.

Higher deliverable optical flux and faster lamp response time (a few μs for Vortek's water wall electric arc lamps) allow for sharper spiking effects (Fig. 1, curve A). A highly absorptive chamber is used to eliminate unwanted multiple reflections and to give faster radiative quenching on cooldown. Uniform heating radiation is projected onto the wafer and guard/support ring from a primary high power lamp within an appropriately designed reflector. Temperature measurement with emissivity compensated pyrometry is relatively easier to implement. A camera is used to record digital pictures of the whole wafer while the lamp is rapidly switched on and off [26]. The thermal emission, emissivity, and thus temperature patterns on the wafer are derived from the differences. Using signal processing information from the camera, a secondary illuminator projects in real time a negative correction image on the wafer to yield uniform temperature. A proximity reflector is used on the device side of the wafer to manage cooling rates and device pattern effects. Camera techniques are also amenable to measuring wafer expansion and stress, which are indirectly related to wafer temperature.

PROCESS CAPABILITY

Tool selection for RTA processing takes into account robustness of design (modeled and demonstrated), performance reliability data, processing cost (calculated real cost per wafer), and, most importantly, matching tool capability with the process requirements. Taking temperature as the primary process capability parameter, the process capability index Cp is the ratio of the temperature range specified for the RTA process and the 6-sigma temperature variation of the RTA tool. Defined similarly, Cpk indexes the maximum deviation of the process mean from the specification limit. Capable processes require Cp > 1 and Cpk > 1.

The temperature accuracy specified for a number of low temperature processes, such as metal and dielectric anneals, is often sufficiently broad that low cost drives decisions in specialized commercial applications. Silicidation processing requires a better temperature accuracy because Si diffusion reactions are involved (± 10 ^{o}C range is usually adequate). Placing a high premium on repeatability and uniformity tends to favor furnace-based methods. Greater temperature accuracy and uniformity are required for dopant activation and diffusion at high temperatures where device characteristics (junction depths and leakage, channel length, threshold voltage) are more sensitive to process temperature.

Accurate evaluation of the process capability index, Cp or Cpk, rests on a good estimate of the required temperature control tolerance. Among available sources for process temperature specifications are the device and process models for a given technology. As a simplification, one could use models to focus on the critical parameters. For a MOS transistor, this may be a length, L, like a junction depth or channel length, that is affected by dopant diffusion. For example, let us suppose that the device model specifies that for a given level of device yield the variation, ΔL, must lie within a ± 5% range. Assuming square root dependence of L on a diffusivity, with thermal activation energy E_A, the corresponding temperature control accuracy is estimated as $2\,E_A^{-1}\,k_B\,T^2\,(\Delta L/L)$. For a process at 1000 ^{o}C governed by a typical E_A = 4 eV , the RTA temperature must then be controlled to within a range of ± 3.5 ^{o}C or better for Cp ≥ 1.

Information on tool performance is the other consideration for process capability. While data from device wafers are the most definitive, preliminary information is gathered from special wafers designed to be particularly sensitive to temperature variations in RTA processing. Two common types of tests are presently considered. To evaluate the independence of temperature sensing systems to wafer emissivity, wafers are prepared with various film coatings that vary the emissivity over a broad range (e.g, back surface films for back surface pyrometry). Measurements of process results on the device side of the wafers, discussed in more detail in the next section, are used to deduce temperature performance and 6-sigma variation.

Tests to evaluate temperature uniformity include monitors where the emissivity of the wafer surface(s) exposed to the heating radiation increases strongly with wavelength between 1 and 3 μm. Such wafers show reduced heating efficiency relative to uncoated Si. An example spectral emissivity is shown in Fig. 6. Here, the total wafer emissivity (where half the spectral weight is between 2.4 and 4.8 μm) exceeds total absorptivity over an incandescent lamp spectrum (where half the spectral weight is between 1 and 2 μm). Imaging techniques were used to show the related effect of temperature variations on wafers caused by emissivity patterns [26].

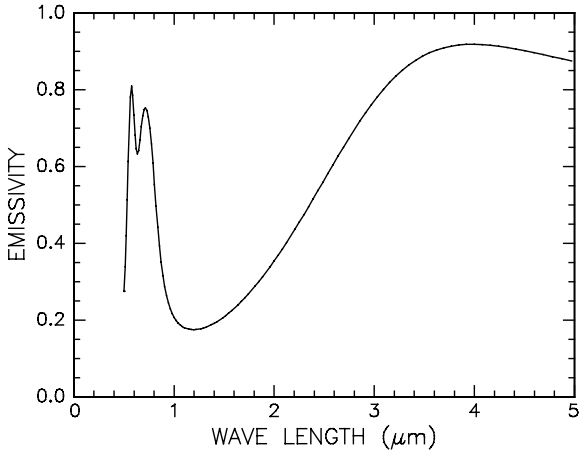

FIG. 6. Spectral emissivity at 1000 °C of Si wafer coated with 70 nm poly-Si over 250 nm SiO_2.

Evaluation with monitor wafers plays a particularly important role in research, development, and generally in the course of introducing new processes. These are stages where device data are statistically less sensitive to RTA performance in the presence of other processing variables.

PROCESS APPLICATION

The tasks for setting up RTA processes in production include producing recipes for the tool sets to execute the prescribed temperature vs time cycles and formulating procedures for monitoring production to assure that device data remain statistically within specifications (i.e., Cpk > 1).

A basic consideration in production is the maximum number of wafers an individual RTA tool may process between routine preventive maintenance procedures. For some low temperature anneals, the interval may be largely a mechanical capability of the RTA tool. For high temperature processes, out-diffusion of dopants from the wafers creates deposits on adjacent quartz windows, wafer support parts, and reflectors. Changes in the optical properties of these components cause drifts in uniformity and errors in temperature measurement. The suitable interval for cleaning such parts derives from experience with the process and tool characteristics and is adjusted according to the results of process monitor wafers. Equipment is periodically monitored for particles added (non-heating test cycles, since thermophoresis removes particles) and for contamination using TXRF and carrier lifetime methods.

Also, irrespective of facility and location, process results must be equivalent and match the given RTA specification. Required are procedures for determining the accuracy of temperature measurement and control within each RTA tool and the error in reproducibility from one RTA tool to another. These errors reduce Cpk. Even with sufficiently accurate temperature measurement, adjustments in nominal tool temperature are nevertheless occasionally used to correct for errors of uncertain origin (e.g., com-

pensation for systematic variation in other processing steps like dopant implants, film depositions, and oxidations). The next sections review several topics associated with temperature measurement and control.

Temperature profiling

The temperature sensors used in RTA are profiled against either a portable secondary standard or a reference involving basic physical principles. Sensors use pyrometry, as discussed above, or thermocouple probes, which are sheathed or coated with SiC for high temperature use. Junction temperatures in thermocouple probes lag and depart from wafer temperatures because of thermal contact resistance and interference from the heaters. Shield plates are used to reduce such errors.

Wired thermocouple wafers with embedded junctions ("TC wafer", e.g., of the SensArray type) are a common method for profiling pyrometer or thermocouple probe signals against wafer temperatures. Figure 7 shows schematically a calibration jig for a thermocouple probe sensor with a shield plate and optional cover wafer. Profile signals are usually recorded during steady temperature operation. Sensor performance is evaluated by running a ramped RTA cycle. A method like in Fig. 7 is also used to profile fixed-emissivity pyrometers in systems that use a shield plate.

Accuracy with TC wafers is limited by variations in the Seebeck coefficients of the wires and heat flow near the junction attached to the wafer, the latter being minimized by bonding both the junction and some length of the wires near the junction. Although they may not provide the best absolute accuracy available with Pt and Pt-Rh materials, chromel-alumel (type K) wires are often preferred because of high tensile strength (allowing thin wires, e.g., 75 μm dia.) lower thermal conductivity (because of alloying), and less reactivity with Si at high temperature (reduced metal diffusion in bonding cement). Temperature offsets under 1 ^{o}C between wafer and junction temperature can be obtained by covering the TC wafer with a second cover wafer, as shown in Fig. 7, and making a furnace-like local environment. Improved designs for wafers with encapsulated or deposited thin-film thermocouples are emerging as candidates for portable temperature standards that are traceable, e.g., to NIST standards [27].

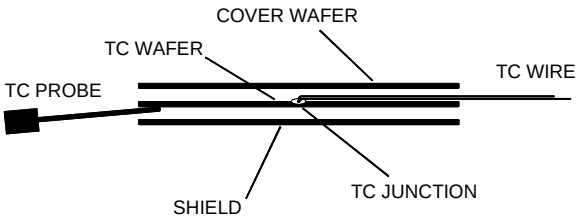

FIG. 7. Apparatus for sensitivity profile of a thermocouple probe. TC wafer has embedded thermocouple junction. Cover wafer, TC wafer and shield plate are heated together (oven, heater and quartz components not shown).

Sensors using ripple techniques require profiling of the signals from wafer and lamp detectors for wafer emissivity and temperature. A group of TC wafers with a broad range of emissivities at the operating wavelength has been used for this purpose. Profiling uses fixtures like Fig. 7, but without the shield. The need for a cover wafer depends on the type of TC wafers being used.

Pyrometers use photo-diode detectors and metal-oxide film interference filters for stable response to IR radiation. Si photo-diodes, notable for low dark current and high stability, are ideal for operation at 0.8 to 0.95 μm. Extended wavelength InGaAs (up to 2.3 μm) or InAs detectors (up to 3.3 μm) are used in mid-IR.

A number of pyrometer systems use sapphire fiber optic probes (as in Figs. 3 or 4). Stable and reproducible sensitivity requires fixed optical transfer coupling factors between adjacent components that include sapphire rods, flexible silica fiber in some designs, lenses, filters, and photodetectors. The primary relationship between detector signal and temperature is obtained from ex-situ measurements with the receptive end of a light pipe in a blackbody furnace. Since the pattern of wafer radiation that is

incident on the probe tip in the pyrometer configuration differs from that in a blackbody furnace, particularly for the backward direction, a separate measurement of a "sensor factor" in the installed configuration is required. An LED illuminator to mimic the thermal output of the wafer has been used by Peuse et al. to reduce the error of probes like in Fig. 3 to less than 2 ^{o}C relative to the blackbody furnace baseline. [21].

In-line Monitoring

Process results from RTA in production are monitored by in-line and final electrical test data from devices on product wafers. The sheet resistance of highly doped crystal and poly-Si van der Pauw patterns and lines, for example, are suitable as indirect process temperature sensors for some of the high-temperature RTA processes. Solid phase diffusion and dissolution of dopants (e.g., B or P) in crystal and poly-Si can yield sheet resistances that decrease monotonically with anneal temperature and time, and are sufficiently sensitive to RTA. Here, the technique is to use a measurement specification in place of a temperature specification.

Process control for individual RTA machines involves charting the results from periodic processing of monitor wafers, which are designed to be somewhat more sensitive to process temperature than the device wafers. Over the temperature range used for monitoring, the

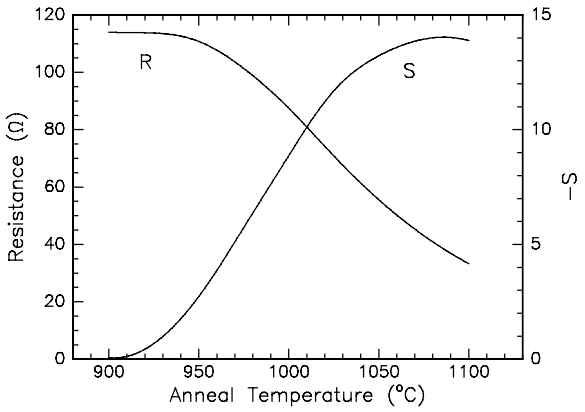

FIG. 8. Variation of resistance and sensitivity parameter of an As implant for RTA temperature monitoring. 25 keV, 10^{16} cm^{-2} implant; 20 s RTA (after Ref. 28).

sensitivity to anneal temperature of a resistance, film thickness or other measurement, M(T), can be represented by a dimensionless sensitivity parameter, $S = T\Delta M\,(M\Delta T)^{-1}$. This relates small temperature deviations, ΔT, to measurement deviations, ΔM. To be suitable for temperature monitoring, one generally selects an annealing reaction with high $|\,S\,|$.

Monitors for use at high temperatures, approximately in the range 900 to 1100 ^{o}C, can use high dose (5×10^{15} to 10^{16} cm^{-2}) and relatively shallow (~100 nm) B or As implants in n-type or p-type lightly-doped Si, respectively. Annealing produces an electrically conducting surface layer, which is sensitive to temperature and isolated by a diffused p-n junction. Specifications on upper and lower limits on the process temperature are used to determine limits on the mean and standard deviation of the sheet resistance of monitor wafers after RTA. The data are obtained by mapping a discrete number of sites across the monitor wafer (e.g., 50 to 100 points, with 3 to 5 mm edge exclusion) with a scanning 4-point probe. Example data from an implant for RTA monitoring is illustrated in Fig. 8.

Another convenient monitoring process is thermal oxidation to grow thin SiO_2 films (2.5 eV activation energy in pure O_2), where the measurable is oxide thickness from scanning ellipsometry. Monitoring with thermal oxidation entails lower expenditure for wafers, which are readily stripped of the oxide and reused. A pre-oxidation clean leaving a hydrophillic surface is preferable for stable wafer storage with minimal particle accumulation (e.g., strip oxide in HF, then peroxide treatment by P-clean or SC2) [29].

A significant deficiency with monitoring temperature by means of implant anneal or thermal oxidation is that the function M(T) is not unique, apart from dependence on RTA time. For implant annealing, the post RTA sheet resistance depends on implant energy, dose, and the thickness of a screen oxide (often used to control the dose and protect against out-diffusion), making a universal M(T) impractical. Oxidation is sensitive to wafer preparation and moisture in the O_2 ambient, in addition to larger

systematic errors in metrology for oxide thickness, compared to sheet resistance.

Monitoring at lower temperature, e.g., for silicidation processes, can use Ti silicidation reactions as a temperature sensor. The sheet resistance of metallic Ti deposited on clean Si varies with temperature as Ti (and oxide) react with N_2 and Si. The sheet resistance is a strongly decreasing function of annealing temperature in the range 600 to 700 ^{o}C. However, this method is also poorly suited as a stable absolute temperature reference, owing to sensitivity to the Ti film thickness, deposition conditions, and Si surface preparation.

Recently, Timans et al proposed using the epitaxial regrowth of amorphous Si on crystal Si as a reproducible physical process for temperature monitoring. Ellipsometry was used to measure changes in layer thickness. The growth velocity activation energy is 2.68 eV. The method is applicable for ~500 to ~670 ^{o}C for spike to 20-s annealing [30].

In addition to the above wafer-based methods, statistical process control techniques have also been applied to data generated by the RTA tool itself. Upper and lower limits are set for operating parameters that are generally correlated with wafer temperature. Control settings are specific to the recipes and product codes run in production. Examples of statistical variables are wafer to wafer excursions recorded for sensor temperatures, emissivities, soaking heat (either lamp power or elevator position), and offsets between redundant temperature readings (e.g., pyrometers vs thermocouple probes).

CONCLUSIONS

The methods for rapid thermal annealing of Si wafers employ either fixed-temperature furnaces with wafer movement during heating or programmable heating lamps of the incandescent or arc varieties. Closed loop temperature control, now the norm for rapid cycling, employs single or multipoint sensors based on infrared pyrometry with wafer emissivity compensation or radiation shielded thermocouple probes. Various approaches are used to achieve satisfactory heating uniformity. Among them are modulated distribution of lamp power, multiple input multiple output controllers, black or gray inner chamber surfaces, and passively heated shield plates and/or reflectors near the wafer. Heating discontinuity at the wafer edge is minimized by a support or guard ring. In order of temperature responsiveness, arc lamp, incandescent lamp, bell-jar furnace, and susceptor furnace methods have been considered. Process control techniques include data from device testing and blanket monitor wafers.

ACKNOWLEDGEMENTS

To the numerous individuals who have contributed to the author's work and education on rapid thermal processing science and technology, it is a pleasure to offer personal acknowledgement and appreciation. This paper includes work supported by Lucent Technologies, Inc.

REFERENCES

[1] R. M. Drexinger and L. R. Durn, "Gold spike machine - military integrated circuit transistor manufacture" (Western Electric, Dec. 25, 1965, unpublished).

[2] W. K. Mammel, "Method for heat treatment of workpieces," US Patent 3,627,590 (1971). R. M. Drexinger, L. R. Durn, and W. K. Mammel, "Infrared spiking process and apparatus – progress report" (Western Electric, Dec., 6, 1966, unpublished).

[3] A. Stein, "Apparatus for remote measurement of temperatures," U.S. Patent 4,708,493 (1987).

[4] A. Kamgar, "Rapid Thermal Processing of Silicon," in *Submicron Integrated Circuits*, edited by R. K. Watts (Wiley Interscience, New York, 1989).

[5] F. Roozeboom and N. Parekh, "Rapid thermal processing systems: a review with emphasis on temperature control," J. Vac. Sci. Technol. **B 8**,1249 (1990).

[6] C. Davis, M. M. Moslehi, A. Bowling, and J. D. Luttmer, "Microelectronics manufacturing science and technology: equipment and sensor technologies," Texas Instruments Technical Journal **9**, 20 (1992).

[7] M. M. Moslehi, C. Davis, and A. Bowling, "Microelectronics manufacturing science and technology: single-wafer thermal processing and wafer cleaning," Texas Instruments Technical Journal **9**, 44 (1992).

[8] R. B. Fair (ed.), *Rapid Thermal Processing: Science and Technology* (Academic Press, New York, 1993).

[9] F. Roozeboom (ed.), *Advances in Rapid Thermal and Integrated Processing* (Klewer Academic Publishers, Dordrecht, The Netherlands, 1996).

[10] S. Vosen, P. Timans, J. Li, and N. Acharaya, "Optimization of fast-ramp RTP performance," *Proceedings of the Seventh International Conference on Advanced Thermal Processing of Semiconductors*, edited by H. Kitayama, B. Lojek, G. Miner, and A. Tillmann (RTP '99, Colorado Springs, Co., 1999). p. 281.

[11] B. R. Appleton and G. K. Celler (eds.), *Laser and Electron Beam Interactions with Solids* (Elsevier North Holland, New York, 1982).

[12] R. Kakoschke, E. Bußmann and H. Föll, "The appearance of spatially nonuniform temperature distributions during rapid thermal processing," Appl. Phys. A **52**, 52 (1991).

[13] P. Vandenabeele, K. Maex, and R. DeKeersmaecker, "Impact of patterned layers on temperature non-uniformity during rapid thermal processing for VLSI applications," Mat. Res. Soc. Symp. Proc. **146**, 149 (1989).

[14] T. Saito, "Spectral emissivity of silicon," Japn. J. Appl. Phys. **6,** 339 (1967).

[15] P. J. Timans, "The role of thermal radiative properties of semiconductor wafers in rapid thermal processing," Mat. Res. Soc. Symp. Proc. **429**, 3 (1996).

[16] H. Rogne and H. Ahmed, "Emissivity of coated silicon at elevated temperatures," Mat. Res. Soc. Symp. Proc. **525**, 27 (1998).

[17] T. J. Quinn, *Temperature* (Academic Press Ltd., London, 1983).

[18] J. P. Hebb, "Patterns Effects in Rapid Thermal Processing," PhD Thesis, Massachusetts Institute of Technology (1997).

[19] J. Hebb and A. Shajii, "Temperature measurement, uniformity, and control in a furnace-based rapid thermal processing system," in *Advances in Rapid Thermal Processing*, edited by F. Roozeboom, J. C. Gelpy, M. C. Öztürk, and J. Nakos, PV 99-10 (Electrochemical Society, Pennington NJ, 1999), p. 375.

[20] J. Nakos, unpublished; C. Schietinger, unpublished.

[21] B. Peuse, G. Miner, M. Yam, and C. Elia, "Advances in RTP temperature measurement and control," Mat. Res. Soc. Symp. Proc. **525**, 71 (1998).

[22] C. Schietinger, B. Adams, and C. Yarling, "Ripple technique: a novel non-contact wafer emissivity and temperature method for RTP," Mat. Res. Soc. Symp. Proc. **224,** 23 (1991); U.S. Patent 5,154,512 (1992).

[23] B. Nguyenphu and A. T. Fiory, "Wafer temperature measurement in a rapid thermal processor with modulated lamp power," J. Electron. Mater. **28**, 1376 (1999).

[24] M. Hauf, H. Balthasar, Ch. Merkl, S. Muller, and Ch. Striebel, "Emissivity compensated wafer temperature measurement using intensity-modulated lamp light," in *Advances in Rapid Thermal Processing*, edited by F. Roozeboom, J. C. Gelpy, M. C. Öztürk, and J. Nakos, PV 99-10 (Electrochemical Society, Pennington NJ, 1999), p. 383.

[25] D. A. Camm and B. Lojek, "High power arc-lamp RTP system for high temperature annealing applications," in *Proceedings of the Second International Rapid Thermal Processing Conference*, edited by R. B. Fair and B. Lojek (RTP '94, Round Rock Texas, 1994). p. 259. D. A. Camm and M. E. Lefrançois, unpublished.

[26] M. Lefrançois, D. M. Camm, and B. J. Hickson, "Emissivity independent process control in a short wavelength arc lamp RTP chamber," Mat. Res. Soc. Symp. Proc. **429**, 213 (1996).

[27] C. W. Meyer, D. W. Allen, D. P. DeWitt, K. G. Kreider, F. L. Lovas, and B. K. Tsai, "ITS-90 Calibration of radiometers using wire/thin-film thermocouples in the NIST RTP tool: Experimental procedures and results," *Proceedings of the Seventh International Conference on Advanced Thermal Processing of Semiconductors*, edited by H. Kitayama, B. Lojek, G. Miner, and A. Tillmann (RTP '99, Colorado Springs, Co., 1999), p. 136.

[28] L. Lozada, unpublished.

[29] A. T. Fiory, "Thin SiO_2 films grown for brief oxidation times," J. Electron. Mat. **28**, 1358 (1999).

[30] P. J. Timans, N. Acharya, and I. Amarilo, "Low temperature measurements and monitors for rapid thermal processing," in *Rapid Thermal and Other Short-Time Processing Technolgies II*, edited by F. Roozeboom, J. C. Gelpy, M. C. Öztürk, and K. Reid, and D. L. Kwong, PV 2000-9 (Electrochemical Society, Pennington NJ, 2000), p. 375.