

CHAPTER TWO

ETCHING AND POLISHING

2.1 Theory of Etching

Etched silicon wafers are prepared through various chemical and mechanical process as discussed in chapter one. The mechanical damage induced during the previous shaping process (slicing to lapping) is removed by etching. Etching is followed by various operations such as polishing and cleaning before it is ready for device fabrication.

Chemical etching of silicon wafers is accomplished by dipping the wafers in an etchant which is traditionally an acidic mixture of $\text{HNO}_3 + \text{HF}$ and a diluent or a caustic solution of KOH . Acid Etching in $\text{HNO}_3 + \text{HF}$ mixture is reported to proceed with following global reactions [22]:-



The identification of the rate-controlling step in a heterogeneous process like this one (mass-transfer vs. reaction) becomes more critical than the knowledge of the actual chemistry in the design of an etcher, because a reaction-controlled etching requires a different design from a mass-transfer controlled etching to produce uniformly etched silicon wafers.

The magnitude of the transport or diffusion effects is not only a function of the speed of transport given by a transport characteristic time or transport resistance, but also a function of the kinetic resistance or the kinetic characteristic time. The nature of the etched silicon wafer changes with a change in kinetic mechanism as well as transport properties. In acid-based etching the controlling mechanism can impact its signature on the etched surface.

Phenomenological Model: Two-Phase System

The etching process itself typically involves the following steps (Fig 2a) :-

- i) transport of the reactants from the bulk solution to the wafer surface
- ii) effective reaction(s) on the wafer surface, and
- iii) transport of products from the wafer surface to the bulk solution.

Reactants pass through a stagnant liquid film, which offers finite resistance for mass-transfer before reactants reach the surface of the wafer. Products also pass through the mass-transfer film (henceforward, the term film refers to mass-transfer film unless specified otherwise). Hence, the reaction and physical transport of reagents occur as steps in series. The finite rate of chemical kinetics provides a finite kinetic resistance that acts in series with the mass-transfer resistance (Fig 2a and 2b). When the mass –transfer and kinetic (reactive) resistance are comparable in magnitude, both kinetics and mass-transfer affect the rate of etching. However, when the difference in the kinetic and mass-transfer resistance is appreciable, the step with higher resistance controls the rate of acid etching. The phenomenological model shown in Fig 2a and 2b is applicable to any typical solid-liquid systems.

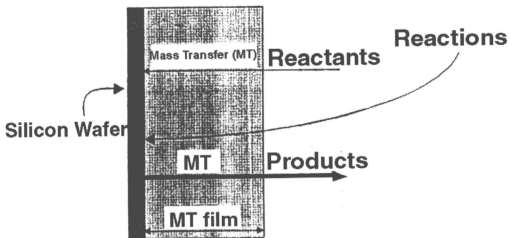


Figure 2 (a) A schematic representation of the mass transfer of resistance in a generic heterogeneous reaction system [24].

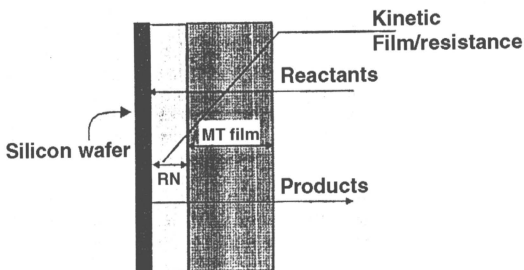


Figure 2 (b) A phenomenological representation of the kinetic and mass-transfer resistance in generic heterogeneous reaction system [24].

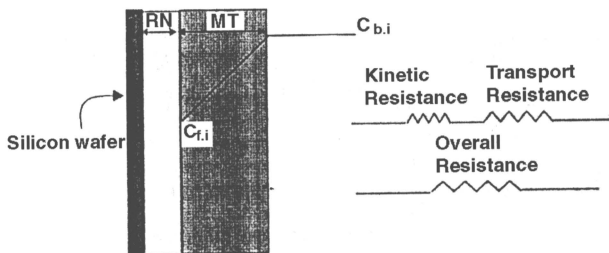


Figure 2(c) A phenomenological representation of the concentration boundary layer for key reactant [24].

Kinetic (reactive) and transport resistance – The two-phase phenomenological model proposed is shown in Fig. 2a and 2b. A finite mass-transfer resistance causes a drop in concentration of a species from the bulk value ($C_{b,i}$) to the solid-liquid interface value ($C_{f,i}$), and this drop is proportional to the mass-transfer resistance. According the classical mass-transfer theory [23], the rate of transport across the mass-transfer film is given by

$$r_{m,i} = k_{m,i}(C_{b,i} - C_{f,i}) \quad (3)$$

$$r_{m,i} = \frac{(C_{b,i} - C_{f,i})}{1/k_{m,i}} = \frac{DF_{m,i}}{Rm,i} \quad (4)$$

where mass-transfer rates ($r_{m,i}$) are defined per cross-sectional area for transport for any given species i . All terms in equations are defined in the List of symbols section. The driving force for the mass-transport is the concentration difference across the mass-transfer film, and the resistance to the transport is given by the inverse of mass-transfer coefficient.

The effective rate of reaction that incorporate adsorption-desorption-diffusion effects in the silicon wafer at the interface is given by

$$(r_{r,i}) = f_r(T, C_{f,i}, C_{f,j} \dots) \quad (5)$$

is an initial approximation, the effect of thermal boundary layer can be neglected; however, analysis of the model does not change if the thermal boundary layer effects are appreciable. Kinetic resistance for systems following simple kinetics is generally defined

as a system constant [23]. Nonetheless, a nonlinear kinetic resistance that varies as a function of interfacial species concentration

$$(R_{r,i}) = \frac{1}{\frac{f_r(T, C_{f,i}, C_{f,j}, \dots)}{C_{f,i}}} \quad (6)$$

Using the above equation, equation 5 can be written as

$$(r_{r,i}) = \frac{DF_{r,i}}{R_{r,i}}$$

where reaction rates for any given species i ($r_{r,i}$) are defined per unit area where reaction takes place. At steady state, at any given location, reaction rates are equal to mass-transfer rates ($r_{r,i} = r_{m,i}$). Thus for any given system, and a given species, the kinetic and mass-transport resistance can be quantified and overall resistance is estimated as

$$= \frac{DF_{o,i}}{R_{o,i}} = \frac{(C_{b,i} - 0)}{R_{o,i}} = \frac{(C_{b,i} - C_{f,i})}{R_{m,i}} = \left(\frac{C_{f,i} - 0}{R_{r,i}} \right) \quad (7)$$

implies $R_{o,i} = R_{m,i} = R_{r,i}$

Characteristic times – The speed of a process is inversely proportional to its characteristic time and also inversely proportion to the resistance of the process. This means that the product of mass-transfer resistance with a characteristic length scale such as mass-transfer film thickness represents a characteristic time for mass-transfer. Similarly, the product of kinetic (reactive) resistance and a reaction length scale such as equivalent reaction film thickness represents a reaction characteristic time.

Thus, as shown in Eq.8, the resistance is a measure of the characteristic time

$$\tau_{m,i} = \delta_{m,i} \times R_{m,i} \quad \text{and} \quad \tau_{r,i} = \delta_{r,i} \times R_{r,i} \quad (8)$$

Application of resistance – the influence of transport and kinetics on a system can be quantified by the ratio of mass-transport resistance to the kinetic resistance. If this ratio is greater than the critical minimum value (that is specific to a system) the system is mass-transport influenced. The system is kinetically controlled when this ratio is below its critical value

$$\frac{R_{m,i}}{R_{r,i}} > \left[\frac{R_{m,i}}{R_{r,i}} \right]_{\min} \quad \text{implies mass transport influenced}$$

and

$$\frac{R_{m,i}}{R_{r,i}} < \left[\frac{R_{m,i}}{R_{r,i}} \right]_{\min} \quad \text{implies kinetically controlled}$$

(9)

Thus the system is either controlled by the dominant resistance or influenced by both when neither of the resistance is negligible. One goal of etching is to reduce the surface roughness. Roughness is a measure of nonuniformity of the surface represented as a field of peaks and valleys and is defined as

$$\phi = \frac{\int_0^L \lambda dx}{\int_0^L dx} \quad (10)$$

where ϕ is the length averaged roughness and L is the total length for roughness measurement. Roughness can decrease only if the rate of removal (etching) at peaks ($r_{p,i}$) is greater than the rate of removal at valleys ($r_{v,i}$) such that

$$\frac{dh}{dt} = -\xi_{s-1} (r_{p,i} - r_{v,i}) = -(rl_p - rl_v) \quad (11)$$

where ξ_{s-1} is the conversion factor that converts rates of consumption of species i per unit area, (r_i) to removal rates based on the decrease in thickness of silicon (rl) [24].

2.1 Etching Methods

Two kinds of etching methods that will be used for this experiment are the Concave Etcher (CV-E) and Convex Etcher (CX-E). The wafers processed at these etchers are shaped Concave and Convex respectively. The similarities between these etchers are that both of them are acid etch using HNO_3 and HF. The actual reaction mechanism is quite complicated and involves many elementary reactions.

The silicon surface undergoes oxidation by HNO_3 to form SiO_2 as shown in Eq. 1 and 2 above. The HF reacts on the SiO_2 and is literally ‘stripped off’. Depending on the acid concentration, temperature and process time, the etching rate can be controlled to meet the process specification. The wafers are loaded in Teflon racks and are placed in the process box which host an assembly of rotors that rotate the wafers about their centers. To maintain the concentration of the acids, both the acids are spiked (term used in silicon manufacturing to add acids periodically by means of a pump) on the concave and convex etcher.

2.1.1 Concave Etching

The schematic set-up of the concave etcher is shown in Figure 2.1 a. The mixed acids ($\text{HNO}_3 + \text{HF}$) and clean dry air (CDA) are pump into the static mixer by a high pressure pump. The purpose of using CDA is to generate air bubbles into the tank. The acid distribution pipe is designed to have equally spaced holes for the acid and CDA to be

released to the tank. The CDA, when optionally introduced, remains well mixed in the liquid and is partly in the form of bubbles. The air bubbles are broken down into misty bubbles by the porex sheets. In the concave etcher, bubbling can be used to achieve a more homogeneous turbulence to improve the performance. The bubbles generated by sparging or the introduction of an inert gas into the etch-bath are referred to as the extrinsic bubbles. The extrinsic bubbles increase the mixing intensity and decrease the bubbles masking and effective mass-transfer resistances. Figure 2.1 (b) and 2.1 (c) shows the wafer mapping of the concave etcher.

2.1.2 Convex Etching

The schematic set-up of the convex etcher is shown in Figure 2.1 d. In this etcher, there is no introduction of CDA and the acid laminar flow is from bottom of wafers towards the top. Unlike the concave etcher, the acid circulation is done by compressed air through the sparger tube and the acid level is maintain. The air bubbles from the sparger tube will dissipate on the surface whereby an exhaust hoods will suck the fumes away.

Therefore the convex etcher can be considered as a two phase model interaction between the silicon wafers (solid) and the acids (liquid). The flow is not homogenous unlike the concave etcher and there exist a boundary layer between the surface as seen in Figure 2a. Wafers are etched more to the edges compared to the center due to the higher angular velocities at the edges. Thus, the wafer are convex in shaped as shown in Figure 2.1 (e) and 2.1 (f).

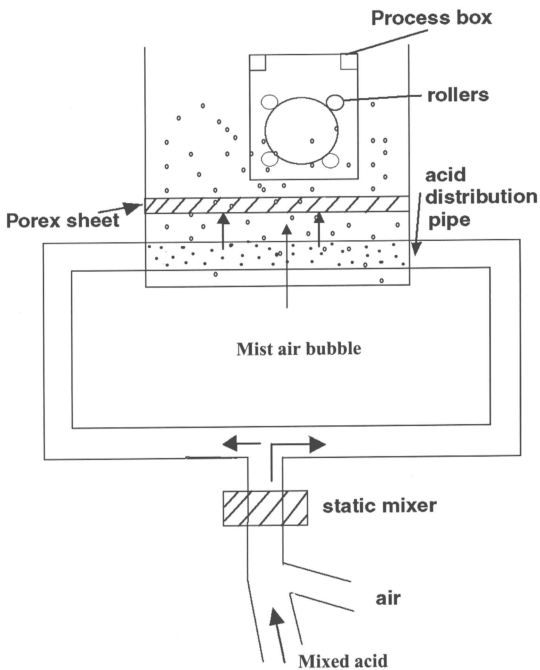


Figure 2.1 (a) Schematic set-up of the Concave Etcher

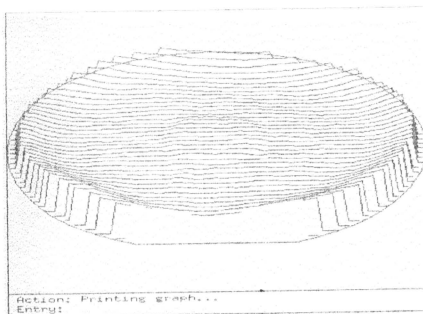


Figure 2.1 (b) Wafer mapping of 125mm processed at Concave Etcher

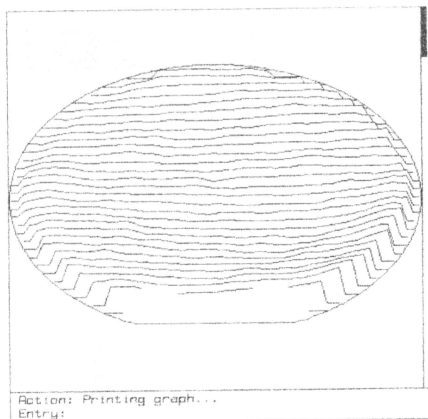


Figure 2.1 (c) Wafer mapping of 100mm processed at Concave Etcher

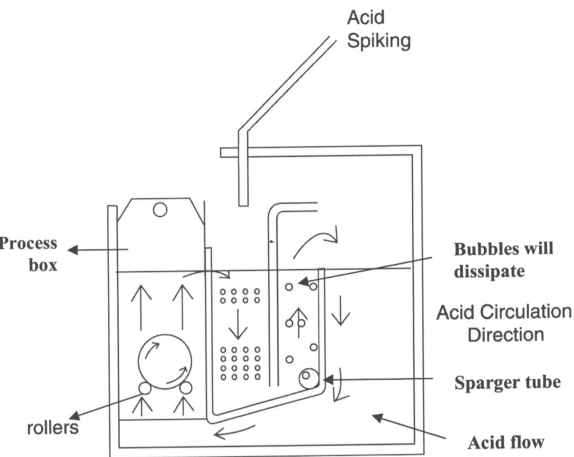


Figure 2.1 (d) The schematic set-up of the Convex Etcher

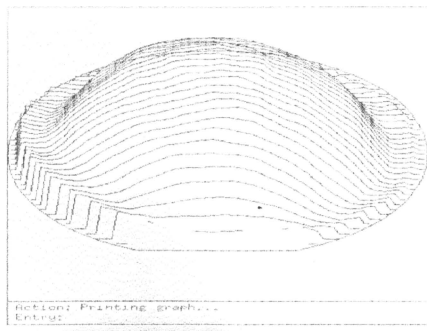


Figure 2.1 (e) Wafer mapping of 125mm wafer processed at Convex Etcher

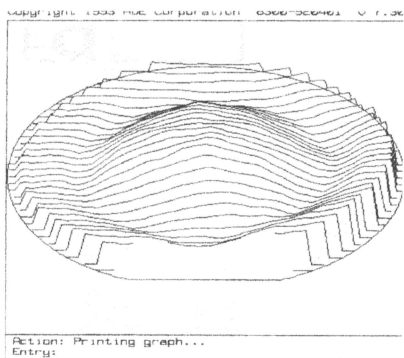


Figure 2.1 (f) Wafer mapping of 100mm wafer processed at Convex Etcher

2.2 Polishing

The basic principle of processing the stain-free mirrorlike surface of crystal materials by polishing is to achieve smoothness and flatness in the atomic unit without disturbing the configurations of the crystal lattice on the surface. Consequently, processing has to be done on the atomic level so that only several atomic layers on the processed surface are removed. Mechanochemical polishing is a typical example of chemical and mechanical composite processing that activates a solid-phase reaction (in the case of the dry type) or solid-liquid phase reaction (in the case of wet type) initiated by a mechanical action based on abrasion to promote processing [25].

Polishing force effect on chemical action

The rate of chemical action (V_c) in mechano-chemical polishing can be expressed by the following equation:

$$V_c = V_o \exp \left(-\frac{E_o - \Delta E}{RT} \right) \quad (11)$$

Where V_o is a constant, R is gas constant and E_o is the activation energy. E_o in Eq (11) is apparently reduced by ΔE through mechanical actions. This means that chemical interaction with the work surface is facilitated. As E_o can be obtained from Arrhenius plots in etching, $\Delta E (=E_o - E_1)$ can be quantitatively determined by obtaining apparent activation energy E_1 . E_1 can be obtained from Arrhenius plots of the mechanochemical polishing rate by controlling work surface temperature to various values. On the other

hand, ΔE must be related to frictional calorific value Q , based on the friction between the work and abrasive or polisher. Q is given by Eq.(12).

$$Q = \frac{F \cdot v}{J} \quad (12)$$

where v is relative polishing velocity, F is tangential polishing force and J is mechanical equivalent of heat. Because Q can be quantitatively determined by measuring F , a functional relation between ΔE and Q can be revealed experimentally [26]. Figure 2.3 shows the mechano-chemical polishing system.

2.2.1 Slurry for silicon Polishing

For silicon material, this slurry consists of silica (silicon dioxide) in aqueous suspension. In contrast, the lapping slurry contains alumina (Al_2O_3) or silicon carbide (SiC) in glycerin or other liquid. Both alumina and silicon carbide are considerably harder than silica or silicon in terms of abrasive strength. Silica has about the same hardness as the silicon, so as soon as the silicon surface is exposed after removal of the top layer, the abrasive action stops. The removal rate is proportional to, among other things, the force applied to the wafer surface, the relative speed of the wafer surface with respect to the polishing pad, and of course, the duration of polishing time.

The polishing slurry for silicon is in a high pH ($9 < pH < 12$) medium. The alkaline ($pH > 7$) nature of the medium helps to oxidize silicon to silicon dioxide. But it has been observed

that at pH values 12 and greater, the removal rate of the material from the wafer surface decreases significantly. At this high pH, the polished surface also shows haze (micro-roughness). Silica needed for the slurry is produced by gaseous reaction or from a precipitation reaction in a solution. The silica particle sizes obtained through these processes have a range of sizes from 40Å to 1000 Å in diameters. Silicon is removed at a rate of about 0.6 µm/minute. A typical polishing cycle lasts about 15-30 minutes, removing about 9-18 µm silicon from the surface.

2.2.2 Three Step Polishing

The actual polishing is normally done in three steps, although a two-step process also has been used. The three steps are primary, intermediate, and final.

2.2.3 Primary Polish

In the first step (primary polish), the main purpose is to remove silicon at a high rate. For this step the polishing slurry contains a higher concentration of silica (normally, the solid content in the slurry is 3% to 4%) and the polishing pad is hard with a coarse surface. High slurry flow is maintained with recirculation. The high flow rate washes away chemical by-products and also prevents build-up of excessive heat that can melt the backside wax. The polishing temperature due to friction between the workpiece (ceramic block with wafer) and the polishing pad can exceed 60 degrees Celsius. The silicon removal rate at this step can be as high as one micron per minute and about 15 microns

are removed in the primary step. Etch pits and some mechanical damage is removed at this stage but the silicon surface is still rough and has many small pits.

At the end of the primary step, the pressure on the plate is reduced and slurry flow is stopped. Flow of deionized water (DIW) removes all of the slurry. It is necessary to flush out the slurry as quickly as possible, because any slurry remaining after polishing steps will etch the wafer in a non-uniform way. The water flow also helps to cool down the workpiece to prevent slurry residue from drying on the wafer.

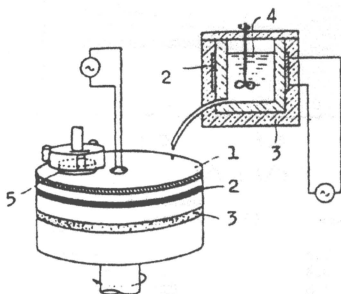
2.2.4 Intermediate Polish

In the intermediate polishing step, the same type of slurry is used. But it is not recirculated, rather a constantly fresh flow is maintained with a reduced slurry flow rate (about one sixth of the flow for the primary stage). One can also use a reduced slurry concentration instead of reduced slurry flowrate or a combination of both. The polishing pad used is softer and made of porometric polyurethane. This film has an open pore structure in the surface that gives softness to it. This step removes the microscratches caused by the recirculating slurry used in the primary step. About 5 additional microns of silicon is removed at this step. This completely removes any residual mechanical damage and produces a very smooth surface. But still the surface is not perfect and contains millions of tiny microscopic pits called “haze”. Haze is caused by various factors and appear as many different types of surface defects such as rough texture, precipitates,

stacking faults, films, and residual damage. Visually they are identified as tiny spots as a result of light dispersion from the microscopically uneven surface.

2.2.4 Final Polish

In the final polishing, the haze is removed from the wafer surface. The polishing pad used is very soft and a different polishing slurry is used. Smaller silica particles and a less aggressive oxidizing chemical is used. The pressure on the wafer is also lower than that used for the previous steps, resulting in a lower polishing temperature. All these factors combine to give a silicon surface that is haze free, very smooth, and highly reflective. Final polish temperature and time are always optimized. Otherwise, a large number of particles on the wafer surface results after cleaning of the wafers. The resulting silicon surface is hydrophobic and very reactive. It is important that the final rinsing with DIW of the clean wafers is kept to a minimum and transferred to a DIW overflow tank to be submerged as soon as possible. This is to prevent the particle-laden air from the polish shop does not have much time to contact the wafer surface. This is performed because of the hydrophobic nature of the surface. Thin water layer acts as a barrier to particle contamination, but on a hydrophobic surface there is no such barrier. A particle landing on such a surface will contact the silicon surface directly and there will be van der Waals type attraction between them. This type of particle is difficult to remove. As a result, the wafers from the polishing area are transported to the pre-dimensional cleaning and inspection area in cassettes submerged in water.



1. Polisher, 2. Heater, 3. Adiabatic material, 4. Polishing slurry, 5. Work piece

Figure 2.3 Mechanochemical Polishing System [26].

Flatness Requirements for Silicon Wafer Polishing

Dram Requirements	1Mbit	64 Mbit
Wafer Diameter (mm)	125-150	200-250
TTV ^a (μm)	5.0	1.5
LTV ^b (μm)/site	1.2 / 15x15mm	0.3 / 20x20 mm
Roughness, R_a (Angstrom)	10	2

a = total thickness variation over entire wafer

b = local thickness variation over the specified site size.

Table 2.4 Flatness requirement for silicon wafers [28].

2.2.6 Maintaining Wafer Flatness

Maintaining wafer flatness during the polishing operation is of primary concern. It has been mentioned that the polishing process is not a flatness improving process [27]. But all the subsequent steps are optimized so that the flatness obtained at the lapping step is not degraded. With decreasing feature sizes, the flatness requirements are becoming more and more challenging. Table 2.4 shows how flatness requirements change as the density of devices increase (with proportional decrease in feature size.)

This type of flatness is required mainly for the lithographic process that defines the features of the final devices. Also from the roughness requirements (10 \AA for 1 Mbit DRAM), it is easy to see why the lapping operation is not suitable to polish the wafer, since any grinding operation will leave residual damage on the surface and beneath it. On the other hand, even in the polishing operation, a total of more than $20\mu\text{m}$ of silicon is removed from the surface. For the flatness required, any thickness of silicon $> 1 \mu\text{m}$ removed from the surface can potentially degrade it from its global initial value of $< 1.5\mu\text{m}$.

2.2.7 Critical Parameters

Flatness during the polishing process is maintained by making sure that all surfaces that hold the wafer or pad are very flat and parallel to one another. The flatness of the following components will directly affect the final outcome.

- 1) Ceramic plate
- 2) Wax layer for wafer mounting
- 3) The vacuum chuck to place the wafer on the ceramic plate
- 4) The bottom plate holding the polishing pad
- 5) The polishing pad

The ceramic plate is extremely flat and because of the way wax is applied to the ceramic surface, the wax layer is of very uniform thickness. The wafer coming out of the lapping operations also are very flat. The polishing machines are designed so as to allow the polishing pad to be perfectly parallel to the ceramic plate and the wafer surface at maximum polishing temperature. Because of these attempts to maintain parallelism and flatness, flatness degradation of the wafer surface does not normally occur in the wax mounted polishing step.

2.2.8 Forced rings

In order to avoid wafer taper during polishing, some force adjustment implements called forced rings pads (hub pads) are used with the ceramic plate. This tapering might occur because of the heat generated due to friction between the polishing pad and the wafers. This heat is not distributed uniformly throughout the ceramic plate and as a result it tends to deform. The hub pads are used inside the assembly to put appropriate force to the ceramic plate to drive it to parallelism. Figure 2.5 shows the schematics of these forced rings.

2.2.9 Parameters for Polishing

The final outcome of polishing process in terms of surface quality depends upon many factors. Some of the important parameters are listed below.

- 1 Polishing temperature
- 2 Plate pressure and rotation rate
- 3 Slurry concentration
- 4 Type of polishing pad

Part of the polishing process takes place as chemical reaction. Chemical processes are highly sensitive to temperature. Too high a polishing temperature may cause too fast etching of the surface and this will make it difficult to maintain wafer flatness over the entire wafer surface. The chemical reaction itself also may generate or absorb heat, which will require additional control. For example, if the reaction generates heat (exothermic

reaction), then the plate may have to be cooled to maintain the proper polishing temperature. Plate pressure and the rotation will create additional heat making it necessary to cool the plate. Slurry concentration and delivery rate will have a direct impact on etch rate. Finally, depending upon the stage of polishing, the right type of polishing pads must be chosen.

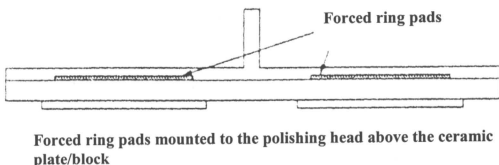


Figure 2.5 Forced ring pads to compensate for ceramic plate deformation [27].

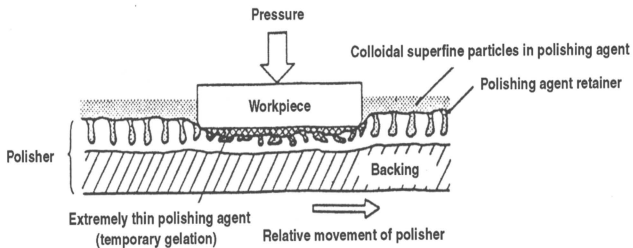


Figure 2.6 Constitutional diagram of the colloidal silica polishing [25].