2009

Modeling and control of material removal and defectivity in chemical mechanical planarization

Pavan Kumar Karra
Iowa State University

Modeling and control of material removal and defectivity in chemical mechanical planarization

by

Pavan K Karra

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mechanical Engineering

Program of Study Committee:
Abhijit Chandra, Co-Major Professor
Atul Kelkar, Co-Major Professor
Ashraf Bastawros
Elgin Johnston
Baskar Ganapathysubramanian

Iowa State University

Ames, Iowa

2009

TABLE OF CONTENTS

LIST OF FIGURES iv

LIST OF TABLES vi

ABSTRACT vii

CHAPTER 1. INTRODUCTION 1
1.1 Introduction 1
1.2 CMP Expansion in the Industry 2
1.3 CMP Process 3
1.4 Interactions between components 3
1.5 Material removal in CMP 7
1.6 Thesis Outline 12

CHAPTER 2. ROLE OF PAD AND ITS CHARACTERISTICS ON MATERIAL REMOVAL IN CMP 19
2.1 CMP and Grinding 19
2.2 Asperity Contact Model 19
2.3 Model of Pad Asperity Evolution 22
2.4 Pad Evolution Simulation 25

CHAPTER 3. INFLUENCE OF SLURRY PARTICLE AGGLOMERATION IN MATERIAL REMOVAL 31
3.1 Brownian Motion and Agglomeration 31
3.2 Types of Agglomeration 33
3.3 Modeling of RLA 34
3.4 Effect of Isoelectric point on agglomeration 36
3.5 Agglomeration Kernel 38
3.6 DLA as a Special Case 41
3.7 Simulation of Agglomeration 42
3.8 Simulation Results 43
3.8.1 RLA Simulation Results 44
3.8.2 DLA Simulation Results 46
3.9 Conclusions 47

CHAPTER 4. Scratch Generation in CMP 60
4.1 Material Removal Process in CMP 60
4.2 Indentation Depth 62
4.3 Material Removal 64
4.4 Validation of the model 65
4.5 Fractal Dimensions 66
4.6 Parametric study 67
4.6.1 Effect of Pad and Slurry Evolution 68
4.6.2 Effect of pH 69
4.6.3 Effect of Fractal Geometry 69
4.7 Conclusions 70

CHAPTER 5. CONCLUSIONS AND FUTURE WORK 91
5.1 Conclusions 91
5.2 Future Work 92

BIBLIOGRAPHY 94

ACKNOWLEDGEMENTS 100

LIST OF FIGURES

Figure 1.1. Schematic of orbital CMP Machine 14
Figure 1.2. Interaction of various components of CMP 14
Figure 1.3. Silicon oxide CMP profile evolution monitoring using SEM 15
Figure 1.4. SEM image of cellular pad 16
Figure 1.5. Pad asperity profile 16
Figure 1.6. The speculated domains for details of local pad-wafer-particle contact 17
Figure 1.7. Cell structure of CMP pad 17
Figure 1.8. Bending modes of pad 18
Figure 2.1. Conceptual illustration of Greenwood-Williamson contact model 27
Figure 2.2. Pad asperity height evolution 27
Figure 2.3. Evolution of pad-wafer distance 28
Figure 2.4. Evolution of area fraction 29
Figure 3.1. Schematic of particle agglomeration 49
Figure 3.2. Zeta potential measured as a function of pH and counter-ion concentration 50
Figure 3.3. Particle diameter distribution evolution for pH = 2 51
Figure 3.4. Particle diameter distribution evolution for pH = 4 52
Figure 3.5. Particle diameter distribution evolution for pH = 5 53
Figure 3.6. Particle diameter distribution evolution for pH = 7 54
Figure 3.7. Particle diameter distribution evolution for pH = 12 55
Figure 3.8. Evolution of mean diameter of particles for various pH values 56
Figure 3.9. Particle diameter distribution evolution at isoelectric point 57
Figure 3.10. Particle diameter distribution evolution at isoelectric point, behavior at lower diameter 58
Figure 4.1. Schematic of material removal in CMP. Rake angle here is negative 72
Figure 4.2. Hydrolysis of silica 73
Figure 4.3. Schematic of particle indentation, intersection of wafer and particle denotes cross section area A, angle ϕ determines the mode of contact 74
Figure 4.4. Particle diameter distribution evolution for experimental conditions 75
Figure 4.5. Particle diameter distribution evolution for experimental conditions, tail portion of figure 4.4 76
Figure 4.6. Variation of number of particles with time for experimental conditions 77
Figure 4.7. Variation of particle mean diameter with time for experimental conditions 78
Figure 4.8. Simulated scratch depth distribution for experimental conditions with linear pad. Scratch depth is not linear with load 79
Figure 4.9. Model validated against experimental results at initial time 80
Figure 4.10. Variation of scratch depth with time for a linear pad response and evolving slurry 81
Figure 4.11. Variation of MRR in time with respect to pressure, effect of pad an slurry evolution 82
Figure 4.12. Geometry of fractals, extent of agglomeration determines fractal dimension [55] 83
Figure 4.13. Scratch probability for linear pad for parametric study 84

Figure 4.14. Scratch Probability with evolving pad, scratch depth increases with conditioned pad and scratch becomes uniform as pad evolves 85
Figure 4.15. Scratch probability for pH 3 at 60 sec, much higher compared to figure 4.14 86
Figure 4.16. Scratch probability for pH 7 at 60 sec, all the curves over lap because pad is not evolving and there is no agglomeration when slurry pH is away from isoelectric point of particles 87
Figure 4.17. Fractals increase scratch propensity, if mechanical behavior of fractals is assumed to be similar to spherical particles 88

LIST OF TABLES

Table 2.1. Physical parameters used for pad evolution simulation 30
Table 3.1. Variables used in simulations 59
Table 4.1. Experimental conditions for model validation 89
Table 4.2. Zeta potential of (90%) ceria 90

ABSTRACT

Chemical Mechanical Planarization (CMP) is a necessary step in semi conductor manufacturing. Since its introduction it has been able to provide better local and global planarization. CMP has found applications in emerging technologies such as shallow trench isolation, damascene technologies. As device size shrinks CMP has become increasingly prominent.

CMP process has been analyzed at different length scales such as particle, feature and wafer scales. Models have been developed for each scale. The models initially have been deterministic, accounting for material removal by a single isolated particle. These models predicted the quality control parameters such as material removal rate and planarization at global scales. Recently, probabilistic models have been developed to describe material removal rate from particle scale. However there do not exist models which capture the interaction between parameters at different length scales.

The focus of this thesis is to develop a multi-scale model that considers the interaction between parameters at different length scales. The interaction between macro-scale property pH of the slurry on the micro-scale phenomenon such as particle agglomeration has been studied. Interaction between pad asperity distribution and particle diameter distribution has also been studied.

In the existing probabilistic models the effect of asperities on the material removal has been studied assuming that the pad asperities are supported on a rigid base. In this thesis the pad cellular structure under the asperities is considered to predict scratch performance and material removal rate.

The effect of pad structure and slurry pH on scratch propensity and wafer scale material removal rate is studied. The model prediction of material removal rate has been validated against experimental data. The predicted linear dependence of material removal rate on pressure has been verified. The underlying cellular structure of the pad has been found to have little impact on the dependence of material removal rate on pressure.

Defectivity, as defined by scratch propensity has been studied in this thesis. Scratch performance has been found to be effected mainly by the proximity of slurry pH to the isoelectric point of the slurry particles. The scratch performance has worsened as the pH of the slurry becomes closer to the isoelectric point of particles.

A parametric study has been undertaken to study the effect of pad and slurry evolution on scratch propensity and material removal rate. Based on the study, suggestions have been made to improve scratch performance and material removal rate. Aggressive pad conditioning has been found to improve scratch propensity while it also has been found to maintain material removal rate, which decreases if the pad is allowed to evolve. On the other hand, pad evolution has been found to alleviate scratch problem.

Based on the insights gained in the parametric study, few suggestions about operating conditions have been made. It has been suggested to keep the slurry pH away from the isoelectric point of particles to alleviate scratch propensity. It is suggested that the pad be conditioned when the wafer surface is away from the target profile and the pad be allowed to evolve when the wafer surface is closer to target profile. This strategy maintains the material removal rate during the beginning of the process while avoiding the killer defects towards the end of the process.

CHAPTER 1. INTRODUCTION

1.1 Introduction

Since the process was introduced in the early 1990s to achieve better local and global planarization, Chemical Mechanical Planarization (CMP) has been an essential step in semi-conductor manufacturing, particularly for technology nodes below 0.250 microns [1]. The process has been adapted from the glass polishing industry. With the increase of number of layers and with the shrinking of the feature size in semi-conductor devices, the prominence of CMP has increased immensely over the last decade.

The semi-conductor industry has been keeping pace with the prediction of Moore's law [2] by improvising the technology for each process step in semi-conductor manufacturing. Moore's law states that the number of transistors, which roughly relates to computational power, that can be placed inexpensively on an integrated circuit has increased exponentially, doubling approximately every two years. This also resulted in miniaturization. Moore says [3] that progress in lithography played a crucial role in exploiting the higher densities of transistors. CMP, as explained later, plays a crucial role in the success of lithography and hence is important in the further miniaturization of the semi-conductor devices. It is the miniaturization that has led to pervasive use of semi-conductor devices ranging from micro-wave ovens to defense equipment to portable electronics. This has resulted in the development of consumer based economy.

Another reason for miniaturization is to reduce the circuit delay. The time delay of the signal is one bottleneck in the effort to improve signal speed and thus the speed of operation of the semi conductor device. The delay is proportional to RC where R is the

resistance and C is capacitance. When the semi-conductors devices involve multiple layers there is capacitance effect due to conductive metal circuit lines separated by dielectric layers. There are multiple approaches to solve this. As capacitance decreases with the distance between circuit lines, it is necessary to reduce the size of the circuit to reduce the time delays in the circuit. Another way to reduce the capacitance is to use low-K materials as inter layer dielectric materials and this also requires some tight quality in terms of planarization [4]. Since the time delay is proportional to the resistance R and since Copper has 40% less resistivity compared to Aluminum the shift from aluminum to copper also reduces the delay [5]. All approaches have been aggressively pursued by the industry. The miniaturization is not without drawbacks as the reduction of line width reduces the size but it also reduces the conductor area of cross-section which means that the resistance R increases and the delay increases. Since miniaturization has other benefits the net deteriorating effect on time delay can be offset by choosing proper materials for conductor and dielectric.

1.2 CMP Expansion in the Industry

CMP began as a process to help planarize inter-level dielectrics. This involves copper CMP, inter layer dielectric (ILD, usually silicon oxide) CMP and shallow trench isolation (STI). Multiple layers of circuits have been realized due to this ability of CMP to planarize different materials. Thus CMP enabled sophisticated multi-level metallization integration approach to prosper [4].

The planarization process, when it was initially applied to semi-conductor industry had its origins in glass polishing. In early 1990s the CMP started as an unsophisticated adaptation of glass polishing. As the device size shrank with each year the

precision of photo-lithography became a bottleneck in the final quality of the devices. To keep the wafer surface in focus for wide lengths, the planarization requirements, both at local level and global level became stringent. Thus the necessity to study the process of planarization in detail emerged. Researchers started looking closely at the various components of the CMP process and how each component contributed to the final quality.

1.3 CMP Process

CMP involves a wafer rotating on a polyurethane pad as shown in figure 1.1. Pressure is applied on the wafer and slurry is used in the process. The slurry contains abrasive particles which perform the process of abrasion and aid in mechanical aspect of material removal. CMP involves different length scales with particles of nanometer size to the wafer with a size of few hundred millimeters. Figure 1.2 shows how the different parts at different length scales combine to produce the target surface. In the CMP process the particles are suspended in chemically active slurry which flows continuously onto the pad. The pad is conditioned frequently to keep the asperities distribution closer to a new pad which improves material removal rate. It is the interaction between wafer pad and the abrasive particles along with the chemically active slurry that determines the quality of the CMP. Thus the mechanical properties of the components and the chemical nature of slurry are crucial design parameters in the quality control of the CMP process. The pad and the wafer are placed eccentrically and rotate in opposite directions as shown in figure 1.1. Figure 1.3 shows the evolution of silicon oxide (silica) profile during CMP.

1.4 Interactions between components

CMP components are, as shown in figure 1.1, pad, wafer, slurry, and the CMP polisher. CMP pads are usually made of polyurethane material. Pads are of two types, cellular and fibrous. Cellular pad are the commonly employed pads in CMP and in this thesis only cellular pads are discussed. Cellular pads contain a cellular structure of base and asperities on the base as shown in figure 1.4. The asperities of the pad are shown in figure 1.5.

Pad and wafer are in contact during CMP and the slurry particles are sandwiched in between. The pressure applied on the pad causes the particles to indent into the wafer and the relative motion ensures that particles plough through the wafer surface to remove material. There are three possible mechanisms to explain the wafer-pad-particle interaction ([6]). Figure 1.6 demonstrates the different scenarios. In figure 1.6(a) there is no contact between the wafer asperities and the pad. Instead the asperities apply pressure on the particles interlaced and the particles transfer the load further down on to the wafer thus resulting in indentation. This mode of contact is relevant for cases when the slurry particle concentration is high or when the pad is stiff or when the load is small or a combination of these factors. A stiffer pad does not deform much and hence it is in this mode of deformation.

Figure 1.6 (b) demonstrates the second type of contact in which pad is partly in contact with the wafer along with the particle wafer contact. This happens when the pad is moderately stiff or when the particle concentration is moderate or when the load is moderate, or a combination of these factors. During this mode of deformation the pad wears out during the CMP process.

Figure 1.6 (c) demonstrates the extreme form of contact in which each particle is completely engulfed by a pad asperity. This occurs when the load is higher or when the pad is very soft along with low particle concentration. The material removal process is not affected between the second and third modes because the horizontal force in the third mode cancels the horizontal component from the opposite direction. The pad wear is maximum in this case because of the maximum contact with the wafer.

Another important factor that affects the pad-wafer interaction is the pad cellular structure. The pad has a cell structure from the studies of Gouda [7] as shown in figure 1.7. There are two modes of loading the each cell structure of the pad. There is local indentation and the cell bending mode as depicted in figure 1.8. For small loads and a load span less than the length of the cell, local indentation is active. When the load increases the cell starts to bend. Gouda [7] developed a series spring mechanical model to understand the nature of load distribution on a CMP pad. Even though there has been such detailed study about the structure and mechanics of the pad, few models considered pad cell structure to quantify scratch propensity. The current thesis makes an attempt to include the cell structure in the prediction of scratch propensity. The pad bending modes are important in understanding the mechanical interaction between the pad particle and wafer.

Another interaction that influences the key quality parameters of CMP is the slurry particle-wafer interaction. This is arguably the most crucial interaction in CMP. CMP involves abrasion of the wafer surface by slurry particles which are softer than the wafer surface. However, due to the presence of chemically active (usually acidic or basic) liquid in the slurry the wafer surface becomes softer and thus the material removal becomes possible. The modified softened layer is of the order of nano-meters and each time an abrasion takes

places there is a new layer formed due to the chemical reaction of slurry chemical on the now exposed surface. Thus the material removal takes place incrementally until the target thickness is achieved. There is chemical reaction between wafer surface and slurry, and mechanical abrasion in the material removal process, hence the name Chemical Mechanical Planarization.

In the previous paragraph the mechanical aspect of the material removal is discussed. The chemical aspect arises not only from the softening of the wafer surface but also from dissolution. The models developed in this thesis predominantly deal with the mechanical aspects and the chemical dissolution is not taken up here. However it is important to acknowledge that the material removal takes place through dissolution as well. Yu and Suo [8] proposed a mathematical model which describes the effect of dissolution of the wafer surface roughness. They found that the chemical dissolution is a function of stress applied on the surface. They derived the effects of compressive and tensile stresses on dissolution. Che et al [9] conducted an experimental study on copper CMP which established that there is a wavelength selectivity caused by chemical dissolution. They found that after the dissolution longer wavelengths dominate while shorter wavelengths decay. The critical wavelength for their experimental conditions has been found to be 0.7 micron. It has to be noted that this critical wavelength is a function of the chemical agent used in the dissolution process. In light of these modeling and experimental studies it can be said that despite the trends observed, the dissolution hampers planarity in an arbitrary fashion and can not be controlled by chemistry alone. Hence the mechanical aspects play more important role in reinstating the planarity than the chemical dissolution.

As discussed here, pad asperities are under loading constantly through out the process of CMP which means that the asperities tend to become more uniform in their height as they wear out. This has implications on the material removal and a model quantifying the effect of such wear has been developed by Wang et al [10] based on an earlier analysis conducted by Borucki et al [11] and Greenwood Williamson contact model [12].

All the above mentioned models are based on the assumption that the polyurethane pad used in CMP behaves as linear elastic material. The effects of non linear pad are studied by Gouda [7] assuming a time-dependent viscoelastic pad. In this study the effect of visco-elastic pad on the pad-particle-wafer interaction and on the pad-wafer interaction are explored.

Another time-dependent viscoelastic model has been developed by Guo and Chandra [13] to explain the effects of visco-elastic pad on the die scale evolution, especially in deteriorating the dishing and erosion that happens during CMP. Pad bending is implicated in these problems.

1.5 Material removal in CMP

As CMP has its historical roots in glass polishing and Preston was among the first to study glass polishing phenomenologically it is not surprising that many earlier models developed used Preston's equation [14] in quantifying the wafer scale average material removal rate. The mean removal rate is

$$\overline{MRR} = KPV \tag{1.1}$$

where, P is the mean interface pressure and V is the relative velocity and K is Preston constant. All the other physical variables are combined into the proportionality constant K.

Numerous other models have been developed by using the Preston's equation in some form in the models.

Two kinds of theories have been developed to explain the wafer-pad-slurry interaction aiding in material removal in CMP. One theory hypothesizes the existence of a hydrodynamic layer between the wafer and the pad, called the lubrication theory. Numerous models have been developed to explain the fluid mechanics and mass transport of this layer. The models typically try to predict the thickness of the layer and the hydrodynamic pressure at the interface. Sunderarajan et al [15] developed a 2D model wafer scale model for lubrication and mass transport for the hydrodynamic layer. As mentioned in the review paper on the lubrication theory by Elon et al [16] many of the models started with the Reynolds equation involving the hydrodynamic pressure and the layer thickness. For CMP Couette flow boundary conditions are more relevant. Higgs et al provided a model to predict the pressure based on polar form of Reynolds equation. [17]. The hydrodynamic models assume the existence of an uninterrupted layer of fluid between the pad and wafer. In CMP abrasive slurry particles are of nanometer size and hence penetrate this layer to develop mechanical contact with the wafer. So the assumption of an uninterrupted layer does seem to be a limitation in these models. The hydrodynamic model does not explain dishing, a common problem in CMP.

In these models wafer and pad have been assumed to have different shapes ranging from flat wafer and pad to flat wafer and sinusoidal pad [18] parabolic wafer and flat pad [19].

In the lubrication theory mechanical abrasion plays a diminished role. There is another competing theory which predominantly accounts for the removal via mechanistic

processes. The actual removal is due to a combination of these theories but it is closer to the mechanistic model as the mechanical abrasion accounts for most of the material removed. In this thesis mechanistic and probabilistic model is discussed which is closer to the actual process.

Lubrication theories described earlier account at the best for die scale details. To build a multi-scale model that explains the material removal mechanisms from the smallest scale-particle scale, these models are not sufficient. In the current thesis the model start from the abrasion of the wafer by an individual particle. The model then integrates the particle scale phenomenon to account for the material removal and scratch propensity at the wafer scale. Thus the model helps gain insights into the physical parameters affecting the material removal and scratch propensity from the particle scale. This modeled can hence be useful in designing the strategies to control the material removal at different stages of CMP and also the final surface quality at the end of CMP in terms of the scratch propensity.

The mechanistic models of material removal in CMP can be categorized into two different eras. The models involving deterministic models in which the material removal rate due to a single particle has been derived with no account for the probability distributions of either particle diameter or pad asperities. With the use of Greenwood and Williamson contact model in deriving the material removal rate a whole range of probabilistic models have been proposed.

Both the models predominantly have been based on contact mechanics to capture the pad-wafer-slurry interaction and its role in material removal. So many of the models, just like Preston's model, have pressure (P) and velocity (V) as the only physical variables. There have been variations in the dependence of material removal rate on of the exponent(s) of

pressure and /or velocity. The exponents vary from 0.5 to 9/8. Zhang and Busnania[20] proposed a model that is based on normal and shear stress at the interface and came up with $MRR = K\sqrt{PV}$, which has much lower dependence on pressure and velocity compared to Preston model. Che et al [21] proposed a scratch intersection model of material removal for copper CMP in which material is removed only when two scratches made by abrasive particles intersect. The resulting model derived had a pressure dependence exponent of 1.3 in the expression for material removal rate. This has been proposed after an earlier study into micro-machining by Che et al [22] supported a characteristic length, which quantifies the length of the chip removed due to secondary scratch after a primary scratch has been made. Using FEM the length has been found to be approximately 7.6 times the height of the chip for micro-machining of while the experimental results yielded 3.7. This characteristic length however is a function of material properties.

Fu et al proposed another model assuming that the indentation of abrasive slurry particles is plastic [23]. Material removal rates were derived for both spherical and pyramid particles. The rate varies linearly with velocity. In this model the material removal is derived for both extreme cases of pad-particle-wafer contact depicted in figure 1.6. For extended pad-wafer contact the pressure dependence has been derived to be of exponent 1.5 where as for the no pad-wafer contact regime the pressure dependence has been derived to be of exponent 9/8. Both of these are for spherical slurry particles. Expressions for sharp pyramidal particles have also been derived in the same paper. For extended contact the dependence is of exponent ¾ where as for the no pad-wafer contact regime the dependence in pressure is linear.

The above described models characterize material removal at particle scale. The goal of CMP is to achieve die scale and wafer scale planarity. This means the models that quantify the die scale and wafer scale planarity must be based on particle scale phenomenon. Models that account for die scale and wafer scale material removal have been phenomenological in nature and did not involve particle scale aspects.

Alleviating this deficiency in the models, a series of probabilistic models based on statistics of pad asperities and slurry particles have been proposed using Greenwood Williamson [12] model for contact of two rough surfaces. The model is applied to CMP assuming a flat wafer and a probabilistic nature of pad asperities. The wafer roughness is of the order of nanometers, which is orders of magnitude lower than the pad asperity roughness, hence the assumption is reasonable.

Borucki [11] is among the first to use Greenwood Williamson model to CMP pad evolution which has been used to account for the decrease in material removal rate as the pad wears out. The model assumes a Hertzian elastic contact between the pad and the wafer and calculates the wear rate of pad asperities based on the evolution of pad asperity distribution. Since pad asperity distribution is used to calculate the material removal rate, the evolution of pad asperity distribution provides the evolution of the material removal rate with pad wear.

Continuing with the model Wang [24] has assumed the pad asperities to have Pearson type distributions and has calculated the effects of evolution of asperities on material removal rate. Stability of the differential equation governing the pad asperity evolution has been analyzed. Material removal rate decay with time has been obtained and compared with experimental results. Similar to Borucki [11], Wang has also found that the material removal

rate drops with the pad wear. A regime of partial wafer-particle-pad contact is assumed as shown in figure 1.6(b).

Greenwood Williamson model has been used to build models that quantify the material removal rate by calculating the distribution of the force on each particle and the distribution of indentation depth by a particle under an asperity. A series [25, 26, 10] of models have been developed to quantify the material removal rate at wafer scale from the calculation of material removal by individual particles. Probability distributions haven been used to account for the diameter distribution of slurry particles.

Among the above models Wang [13] has worked on the effects of variation of slurry particle distribution with material removal rate. The model predicts that the material removal rate increases as the mean particle size increases and also increases with particle standard deviation for same contact area fraction.

With these models probability made inroads into CMP modeling before which deterministic models had been used. The deterministic models could provide the mechanistic material removal rates for a single particle. With the advent of probabilistic models, mean material removal has been calculated using distributions of particles and pad asperities.

1.6 Thesis Outline

The thesis consists of five chapters. Chapter two deals with the pad model that quantifies pad evolution and pad effects. Chapter three involves detailed modeling of slurry agglomeration and its role in generating scratches larger than expected. The variation of agglomeration rate as a function of slurry pH is discussed. Chapter four combines the effects of pad and slurry and other process conditions on material removal in the CMP. The final

chapter discusses the conclusion of this work and possible improvements to the CMP process from the perspective of this thesis.

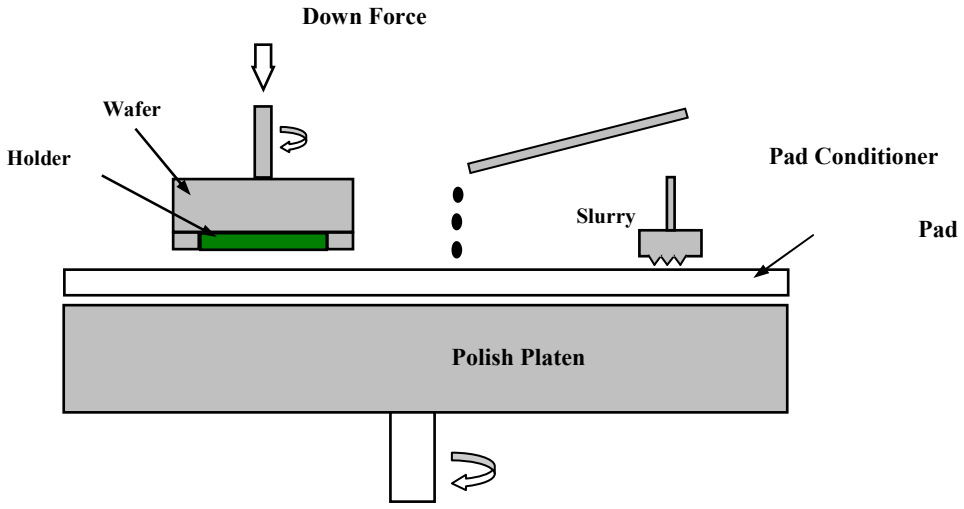

Figure 1.1. Schematic of orbital CMP Machine (adapted from [6])

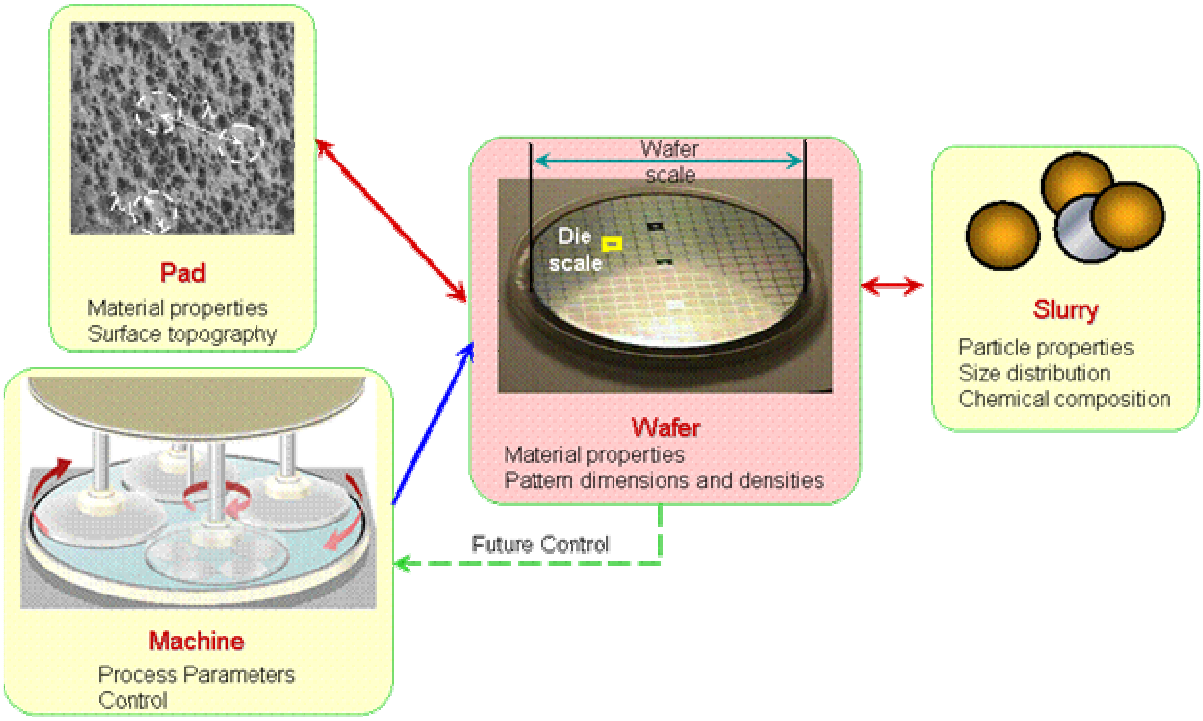

Figure 1.2. Interaction of various components of CMP

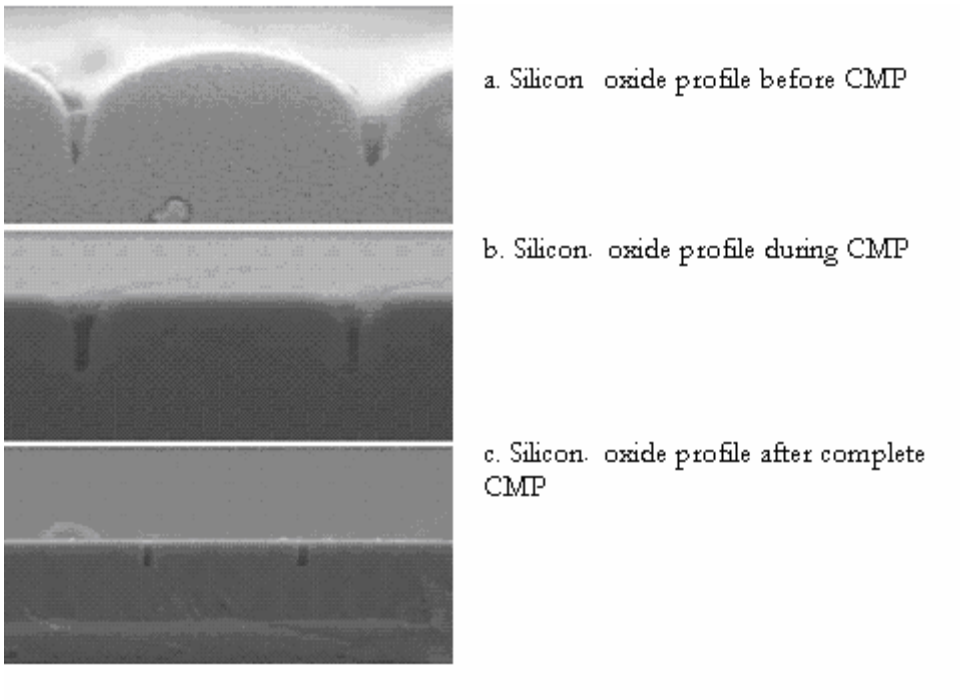

Figure 1.3. Silicon oxide CMP profile evolution monitoring using SEM (adapted from [27], [28])

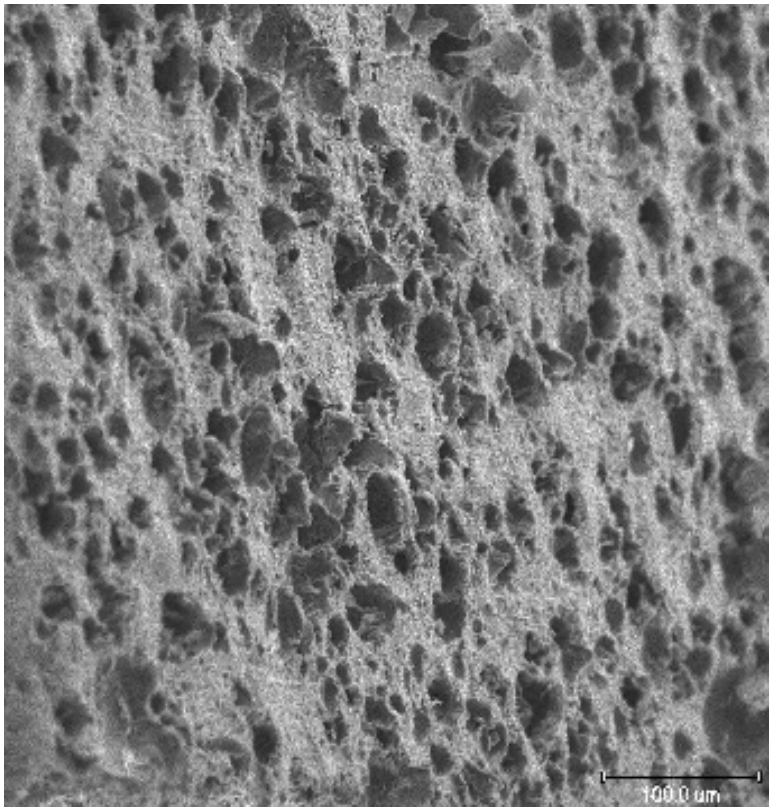

Figure 1.4. SEM image of cellular pad (adapted from [29])

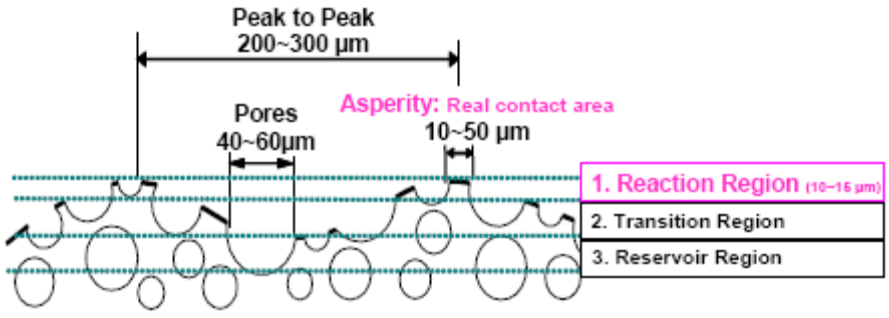

Figure 1.5. Pad asperity profile (adapted from [30])

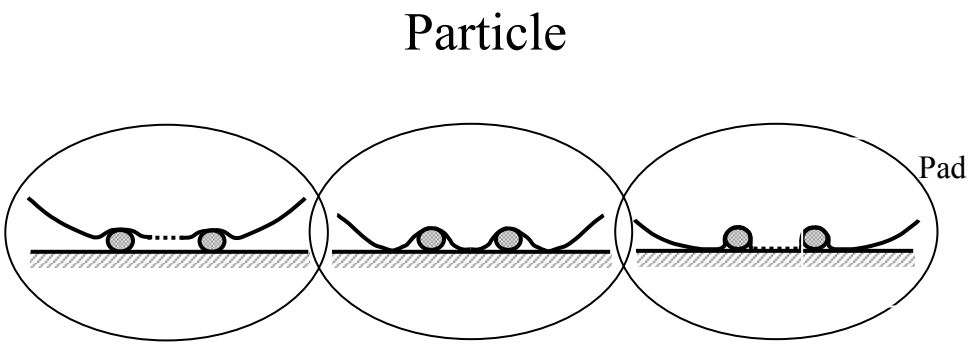

Figure 1.6. The speculated domains for details of local pad-wafer-particle contact (adapted from [29])

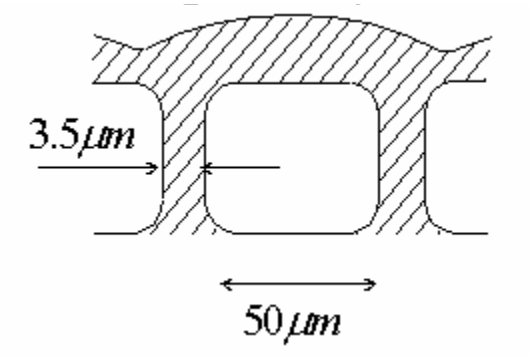

Figure 1.7. Cell structure of CMP pad (adapted from [7])

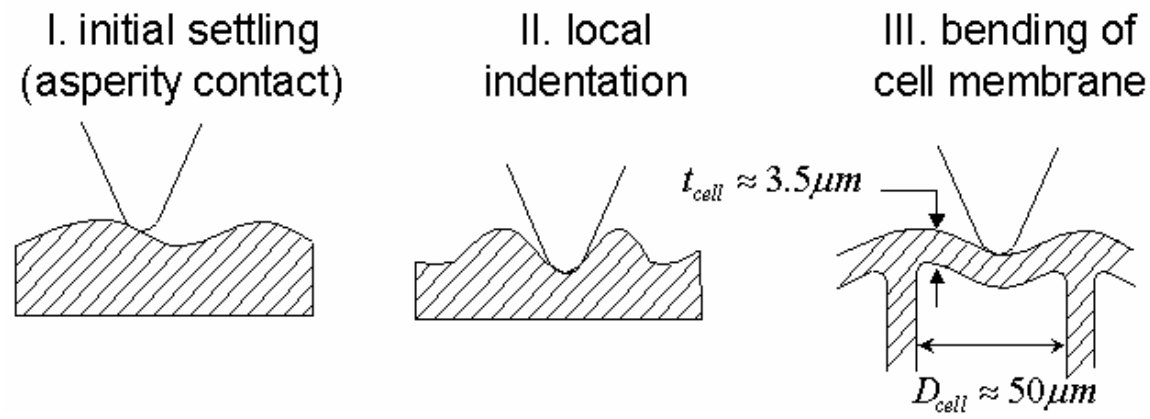

Figure 1.8. Bending modes of pad (adapted from [7])

CHAPTER 2. ROLE OF PAD AND ITS CHARACTERISTICS ON MATERIAL REMOVAL IN CMP

2.1 CMP and Grinding

CMP is a free abrasive grinding process. In grinding process the abrasives are fixed. Grinding can be modeled as a special case of CMP with a pad much harder than the polyurethane pads used in CMP. The abrasives in grinding process are much rougher as well. In CMP the chemical slurry together with pad acts as free abrasives. In contrast grinding has bonding material which binds all the abrasive grits together.

Another difference between CMP and grinding is that in grinding the bonding material does not touch the work piece. However in CMP, the pad material is much softer and deforms along the contour lines of the patterned wafer. This accelerates the pad wear.

2.2 Asperity Contact Model

The pad-wafer contact is assumed to be Hertzian elastic in this thesis. The elastic model is used by Greenwood and Williamson [12] to develop a model that describes the contact between two nominally flat but rough surfaces. In the present study, the Greenwood Williamson model has been simplified to account for the contact between a rigid, flat and infinitely smooth plane and another surface with surface roughness equivalent to the combined effect of the collective roughness of the mating surfaces. This second surface (or body) is also endowed with an effective or equivalent elastic modulus representing the combined elasticity of both contacting bodies. In this thesis the wafer is assumed to be flat and since the pad is the only rough surface the model used here can be describe as a limiting

case for the original Greenwood and Williamson contact model for two nominally flat rough surfaces.

Figure 2.1 shows the Greenwood and Williamson model applied to the current problem. The pad asperities are shown to be in contact with the wafer. The variable z in the figure is the asperity height and the variable s is the mean separation between the wafer and the pad. In this model pad asperities have a probabilistic distribution $\phi(z)$ with the asperity height Z being the random variable. It can be easily deduced that the distribution of asperity heights determines the mean separation distance and that only the asperities taller than s actually contact the wafer surface. The separation distance s is a function of the applied nominal pressure applied P.

The pad asperities are assumed to be standing over a flat pad layer of around 1.8 mm. The pad asperity density is assumed to be η_s and for an area of A_s the number of asperities $N = \eta_s A_s$. Each of the N asperities $\{Z_i\}_{i=1}^{N}$ is assumed to have independent and identical distribution (IID). Using load balance we have [31]

$$L_i = \begin{cases} \dfrac{4E}{3\sqrt{\kappa_s}}(Z_i - s)^{\frac{3}{2}}, & Z_i > s \\ 0 & Z_i \le s \end{cases} \tag{2.1}$$

with the area of contact using Hertzian linear elastic contact solution is given by

$$A_i = \begin{cases} \pi \dfrac{(Z_i - s)}{\kappa_s}, & Z_i > s \\ 0, & Z_i \le s \end{cases} \tag{2.2}$$

where s is the mean separation distance between the pad and the wafer, κ_s the asperity tip curvature and $\frac{1}{E} = \frac{1-\upsilon_p^{\,2}}{E_p} + \frac{1-\upsilon_w^{\,2}}{E_w}$ with E being the equivalent modulus of the contact between wafer and pad. E_p, E_w are the Young's modulii of pad and wafer while ν_p and ν_w are Poisson's ratio of pad and wafer respectively. Because the wafer modulus is approximately three orders of magnitude higher than that of pad, the expression approximately results in $E = \frac{E_p}{1-\upsilon_p^{\,2}}$. The local contact pressure for an individual asperity then becomes

$$P_{cont} = \frac{L_i}{A_i} = \frac{4E\sqrt{\kappa_s}}{3\pi}\sqrt{Z_i - s} \tag{2.3}$$

Since Z_i is a random variable any expression involving the height of the asperity becomes a random variable. Thus contact pressure and contact area are both random variables. Since the load, as evident from equation (2.1), is carried only by the "active" asperities (defined by height $z > s$), the real contact area is only a fraction of the expected value. Thus the load carried by an arbitrarily selected asperity is

$$E[L] = \int_{-\infty}^{s} 0\phi(z)dz \;+\; \int_{s}^{\infty} \frac{4E}{3\sqrt{\kappa_s}}(Z_i - s)^{\frac{3}{2}}\phi(z)dz \;= \frac{4E}{3\sqrt{\kappa_s}}\int_{s}^{\infty}(Z_i - s)^{\frac{3}{2}}\phi(z)dz \tag{2.4}$$

Hence the total expected load carried by all the active asperities is $N\frac{4E^*}{3\kappa_s^{1/2}}\int_s^\infty (z-s)^{3/2}\phi(z)dz$, since N is the total number of asperities.

Similarly, as a result of equation (2.1) the actual (real) contact area is given by [10]

$$A_c = E[A_{actual}] = \sum_{i=1}^{N} E[A_i] = N\int_s^{\infty} \frac{\pi}{\kappa_s}(z-s)\phi(z)dz = N\frac{\pi}{\kappa_s}\int_s^{\infty}(z-s)\phi(z)dz \tag{2.5}$$

Let $\overline{p}$ be the nominal contact pressure applied on the wafer. The total load carried by the asperities is a random variable and so is contact pressure on each asperity. The expected value of the nominal pressure which is the ratio of the total load to nominal contact area, is given by

$$\overline{p} = E[p] = E[L]/A_0 = \frac{4\eta_s E}{3\sqrt{\kappa_s}}\int_s^{\infty}(Z_i - s)^{\frac{3}{2}}\phi(z)dz \tag{2.6}$$

2.3 Model of Pad Asperity Evolution

The model explained so far is based on Greenwood Williamson model. The model only captures the effect of asperities. As shown in figures 1.4 and 1.7 the pad has a base of a cellular structure which carries the load down to the base. But the pad structure beneath the asperities has been known [7] to play a role in its response to loading. This response is sensitive to the amount of load applied because of non-linearity in the pad response. The non-linear effect begins to show their effect after a certain loading at which the slope of the load-displacement curve suddenly increases. This effect of non-linear response of the pad is explained in chapter 4.

In this current work an effort has been made to capture the effects on the entire structure of the pad including the cellular base and the asperities. In [7] an effort has been made to capture the response of the pad by studying the response of effect of individual components on the loading of the pad and then combining the responses to predict the total

response of the pad. As the loading of pad progresses, it can be categorized into two regimes. The first involves the indentation of the asperities and the second involves bending of the cells that define the structure of the pad below the asperities (figure 1.8). As loading begins the asperities come into contact. The load then is transferred to the structure beneath the asperities. Thus the cells of the pad start to absorb the load and deform. This analysis will be used in chapter 4 in scratch prediction. The asperities have a radius of curvature of about 50 micron.

Thus the contact between the wafer and a single asperity is modeled as contact between two spheres, one with radius of $R_1 = 50$ μ and the other with extremely large radius of curvature R_2. The equivalent radius of curvature is given by [31] $\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \approx \frac{1}{R_1}$, since R_2 is large compared to R_1. Let the asperity be compressed by δ (=z-s) by the load applied on the asperity. Then, from Hertzian normal contact mechanics of elastic solids [31] the load supported by the asperity is given by

$$P = \frac{4\delta^{1.5} R^{0.5} E}{3} \tag{2.7}$$

The contact stiffness for the assumed spherical tip is then given by

$$K_{sph} = \frac{P}{\delta} = \frac{4\delta^{0.5} R^{0.5} E}{3} \tag{2.8}$$

The load is then transferred to the base structure of the pad which contains the cellular structure of the pad. The stiffness of the cell is given by [7] as

$$K_{cell} = \frac{4\pi\, E_{pad}\, t_{cell}^3}{3\left(1-\nu_{pad}^2\right) r_{cell}^2} \tag{2.9}$$

As shown in figure 1.7 r_{cell} is the radius of the cell and t_{cell} is the thickness of the cell. The other variables in the equation have already been discussed. The pad used in a typical CMP process has a thickness of 2mm. The asperities only account for about 100-200 μm of that. The remaining of the structure is made of cells. Each cell has a height of approximately 20 μm which implies that there can be up to 100 layers of the cell structures in the base of a pad.

The stiffness of the cell mentioned above is per layer. Taking multiple layers of cells in account makes the effective stiffness of the pad much lower, since the layers are in series. If n is the number of layers of cells present in a pad the cellular structure has an effective stiffness of $K_{total\_cell} = \frac{K_{cell}}{n}$. The asperities and the cellular structure are in series and hence the effective stiffness K of the pad is given by

$$\frac{1}{K} = \frac{1}{K_{total\_cell}} + \frac{1}{K_{sph}} \tag{2.10}$$

Thus the pad is softer than its individual components. The deflection of the pad is the sum of deflection of asperities and the deflection of the base. The effective stiffness will be used in chapter 4 for scratch propensity calculations. The separation distance s can be established by solving the following load balance integral equation iteratively until equilibrium is reached.

$$\overline{p}A = \int_{s}^{\infty} K\phi(z)(z-s)dz \tag{2.11}$$

In the above integral equation, $\overline{p}$ is the nominal pressure; A is the nominal area; and ϕ is the distribution of pad asperities and z is the random variable denoting asperity height.

The equation has to be solved by iteration. The load carried by individual asperity is then calculated and thus the pressure carried out by each individual asperity is computed. Archard's law [32] is then used to find the pad asperity wear rate for individual asperity treating each asperity height as a random variable. This forms the following governing equation for the pad asperity evolution.

$$\frac{dZ}{dt} = C_w p(Z) \tag{2.12}$$

From the above equation it can be seen that the wear rate of an asperity is a function of the asperity height. This is because the taller asperities undergo larger deformation and hence experience higher pressure. This implies that the polishing process makes the asperities more uniform in height. To confirm this, the probability density function of pad asperity height needs to be evolved. From equations (2.7), (2.12) and [11] the evolution of probability density function for $Z > s$ is given by the partial differential equation

$$\frac{d}{dt}\varphi(Z,t) = \frac{4C_a E\sqrt{\kappa_s}}{3\pi}\frac{d}{dz}\left\{\sqrt{z-s(t)}\phi(Z,t)\right\} \tag{2.13}$$

2.4 Pad Evolution Simulation

Simulation has been performed to predict the pad asperity evolution under the conditions given in table 2.1. All the conditions are kept constant from [10,11]. The partial differential equation (2.13) is solved with a time step of 0.5 sec.

Simulation has been carried out for the parameters listed in table 2.1. The time step used is 0.5 sec and the resolution of the domain, which is pad asperity height, is 0.01 micron. The simulation has been run for 3000 seconds which is of the order of time a pad is used in CMP (or typical conditioning interval). The evolution of pad asperities is shown in figure

2.2. For a new pad only the asperities with heights greater than a threshold value are in contact with the wafer. These asperities carry the entire load. As the pad asperities wear out more asperities, the ones with smaller initial height also come into contact and undergo the wear process. With time more asperities come into contact. The wearing process makes the asperities uniform in height. In other words, the pad asperity height evolution process may also be viewed as a homogenization process [33]. Since the number of asperities in contact is cumulative in nature along with the fact that the wearing process makes the asperities uniform, it is expected that the asperities towards the right end of the distribution have higher frequency in occurrence as time progresses. This is the reason why the probability density function develops a spike which becomes increasingly sharp with time.

The wearing of the pad means that with time taller asperities are worn out and hence the distance between pad and wafer decreases as evident from figure 2.3. For a new pad the taller asperities carry the entire load and hence the area fraction in equation (2.6) is small. As more asperities come into contact, the area fraction increases as shown in figure 2.4.

The evolution of pad asperity height, pad-wafer distance and the contact area fraction will be used to predict scratch propensity in chapter 4.

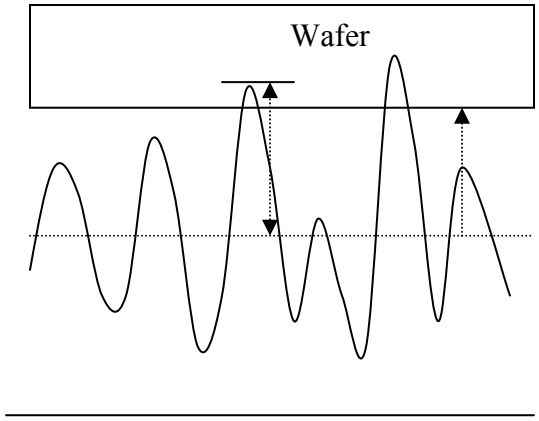

Figure 2.1. Conceptual illustration of Greenwood-Williamson contact model [12]

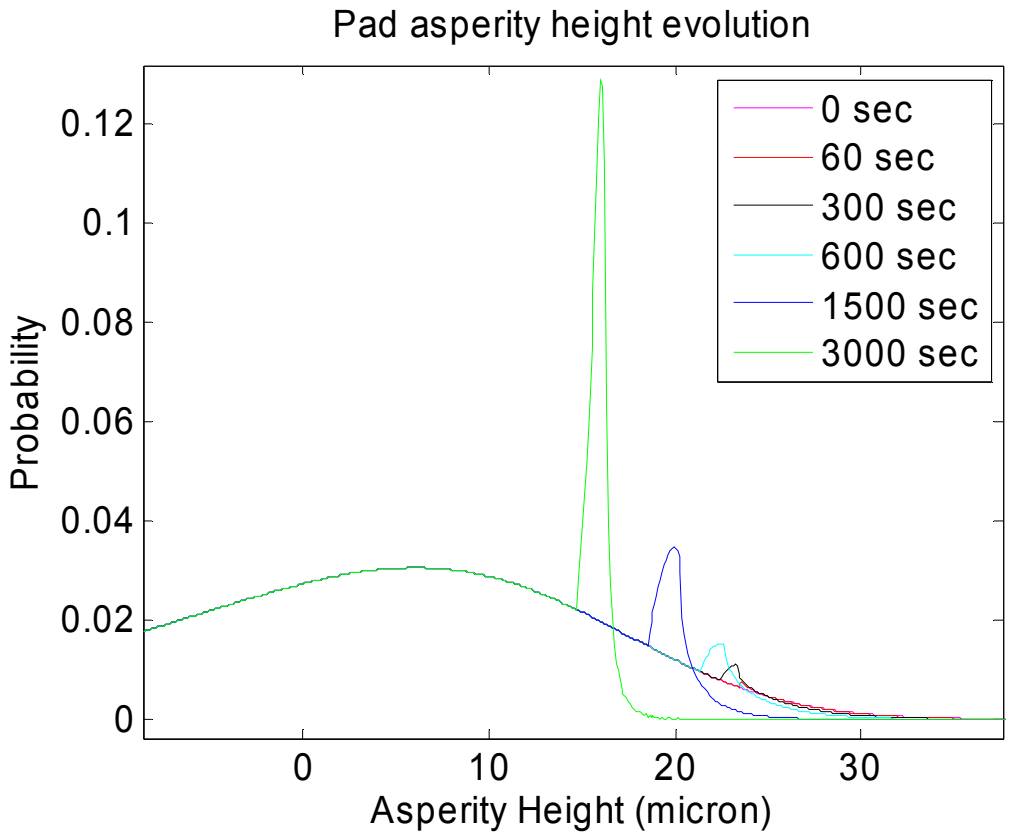

Figure 2.2. Pad asperity height evolution

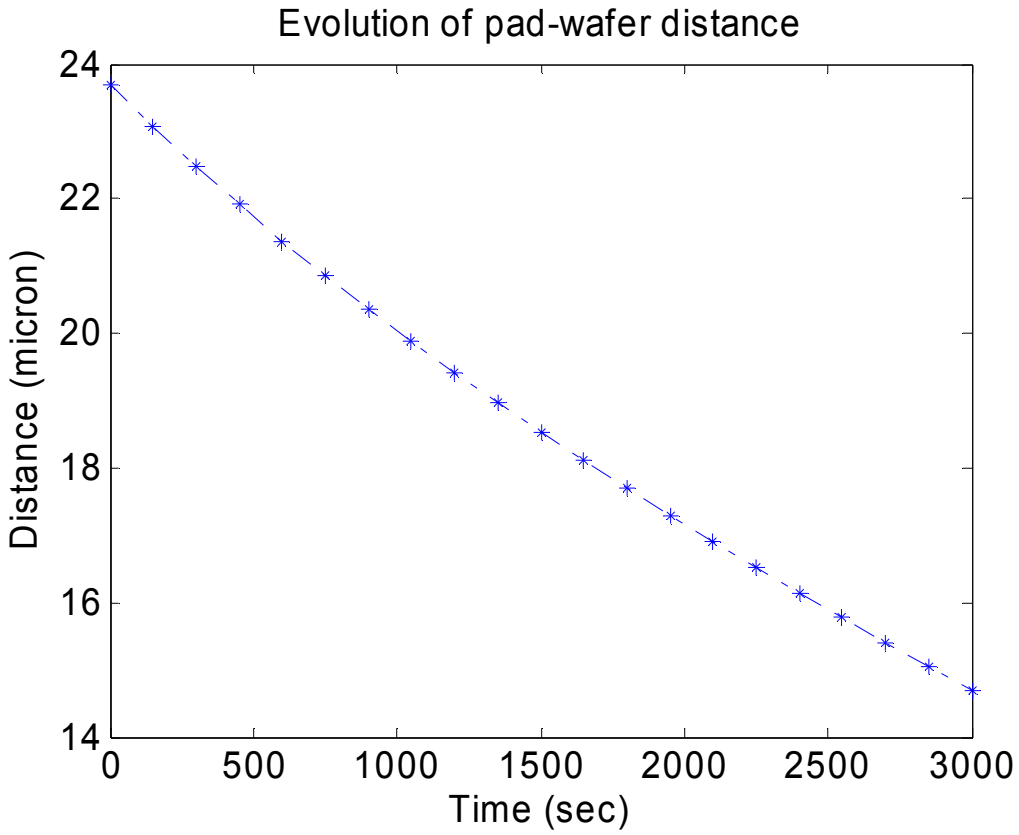

Figure 2.3. Evolution of pad-wafer distance

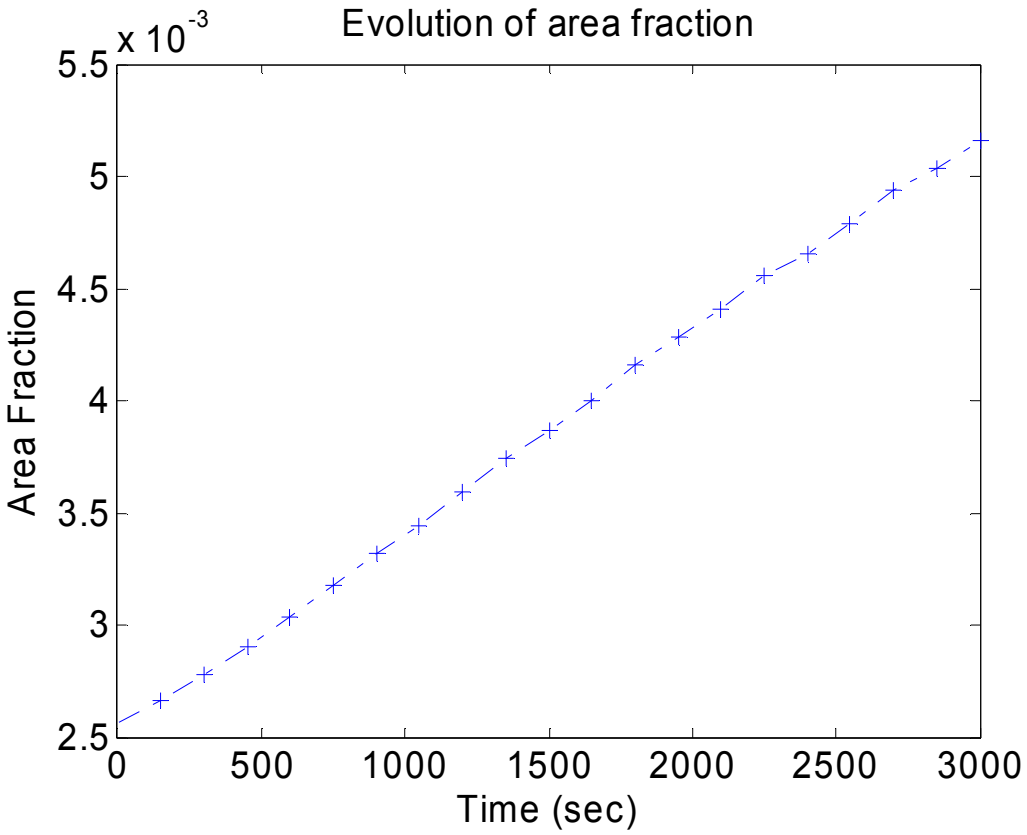

Figure 2.4. Evolution of area fraction

Table 2.1. Physical parameters used for pad evolution simulation [11]

Variable	Value	Units
Asperity mean	0	micron
Standard deviation	15.625	micron
Skewness	-1.25	
Kurtosis	6.875	
Pad Modulus (E_p)	100E6	Pa
Poisson's ratio	0.4	
Pad asperity density	2E8	m^{-2}
Pad asperity wear rate	1.5E-8	$Pa^{-1}min^{-1}$
Asperity tip radius	50	micron
Pressure	50E3	Pa

CHAPTER 3. INFLUENCE OF SLURRY PARTICLE AGGLOMERATION IN MATERIAL REMOVAL

Slurry particle agglomeration exerts a crucial influence on the material removal mechanism in CMP. Since the geometry of the particles participating in the polishing process inherently induces a negative rake angle [34], the material removal mechanism may be described as an inclined indenter, subject to both normal and shear loading. This implies that the indentation characteristics play a crucial role in the material removal process. Thus, the indentation depth becomes a function of the particle size, which determines the magnitude of the negative rake angle as well as the magnitudes of normal and shear forces carried by a particle. It is interesting to note that the force per particle is determined not only by the particle carrying the load, but also by its neighboring particles, though this effect is not included in this thesis. Thus, the neighborhood or the particle cluster in which a particle resides also influences the material removal characteristics. This influence of particle size and its distribution, which is altered in time due to agglomeration, is expressed in a material removal process via the indentation depth under the particle(s) of interest. The fact that the agglomerates exhibit fractal behavior [35, 36] further complicates this expression.

3.1 Brownian Motion and Agglomeration

Slurry particles tend to agglomerate before (on the shelf) and during a CMP process (depending on the conditions in which the slurry has been stored. Figure 3.1 shows a schematic of the agglomeration process. Since particles of smaller sizes agglomerate to form larger particles, the phenomenon in a closed system is mass-conserving. The debris generated

due to material removal also plays a role in the process but since the debris usually constitute a different material, only the agglomeration of slurry particles is discussed in this thesis. Work is currently in progress on understanding the agglomeration process in composite slurries constituting of multiple materials.

The predominant aspect of agglomeration is Brownian motion of nano-particles which are suspended in a liquid. Brownian motion gains the name as it was first studied by Robert Brown in 19th century. The problem was mathematically solved by Louis Bachelier [37] in 1900 but the phenomenon remained an enigma till Einstein's [38] and Somluchowski's [39] research brought the solution to the attention of physicists, which also indirectly confirmed the existence of atoms and molecules.

Fluid molecules are not as rigidly bound as molecules in a solid. As temperature of a fluid rises, the vibration of the molecules in a fluid increases. The dependence of velocity on temperature increases from liquid state to gaseous state. The motion of molecules in a fluid is random due to quantum mechanical effects. Hence the solid particle that is suspended in a fluid is constantly bombarded with the fluid molecules from all direction. Due to the random nature of the bombardment, not only the direction of the resultant force of the collision(s) is random, but the equilibrium position of a solid particle is also not sustainable. Thus the particles suspended in a fluid are under the influence of forces in random directions with random (but finite) magnitude. The magnitude of motion however has an upper limit determined by the temperature of the fluid. The average magnitude, as will be seen later in this chapter, is determined by the Boltzman constant. It relates the energy at the microscopic scale to the temperature at the macroscopic scale. The Brownian motion is responsible for particles coming into each others proximity, and also makes them collide with each other. In

general, the particles which come into contact with each other have an affinity to stick to each other and form agglomerates. This tendency is of course mitigated by several factors or properties of the particle, e.g., charge, net momentum change in a collision, forces holding the agglomerate together, etc. Based on the "strength" of the affinity, an agglomeration process is, in general, classified into two regimes: Diffusion Limited Agglomeration (DLA) and Reaction Limited Agglomeration (RLA). Only the mitigating effects of charge are discussed in this thesis.

3.2 Types of Agglomeration

The particle agglomeration process, based on the nature of the agglomeration is divided into two regimes, Diffusion Limited Agglomeration (DLA) and Reaction Limited Agglomeration (RLA). In DLA, the agglomeration process is limited by the ability of the particles to diffuse through the fluid medium. This regime takes place when the particles suspended in the fluid carry no net charge. Under such circumstances, it is reasonable to assume that whenever two or more particles come together, they stick to each other immediately with a probability of 1. In this chapter it will be explained how the lack of net charge results in the probability of 1. DLA is the fastest among known forms of agglomeration.

RLA occurs when the particles suspended in the fluid carry net charge on their surface. Due to the charge present on the particles they tend to electrostatically repel each other and hence it takes much more collisions for a particle to stick to another particle. Particles of same chemical composition, when suspended in a fluid, carry same charge. Since particles of same chemical composition will always carry the same charge, the electrostatic

force between them cannot be attractive. Thus, there exist only two regimes of agglomeration in slurry containing only a single species of particles.

In this thesis an attempt has been made to unify DLA and RLA regimes of agglomeration. Thus, under assumptions of a uni-species slurry, DLA will be presented as a special case of RLA, the case when charge on particle surface is zero. In this development, it is assumed that a large particle cannot break up into smaller pieces or fragments. Such an assumption is reasonable when the energy contained in the Brownian motion is below a critical threshold that is needed to cause such debonding of agglomerates. This is due to the fact that in a typical CMP process temperatures are well within the threshold.

3.3 Modeling of RLA

Modeling of RLA paves the way to account for a general form of particle agglomeration, thus making it easy to explain the scratches generated in CMP that are larger than expected. Since the scratches generated are directly proportional to the particle diameter the ideal way to express the agglomeration of particles is in terms of their diameter distribution as a function of time. The distribution as a function of time is dependent on the environment that surrounds the particles. The environment that affects the agglomeration process the most is the chemistry of the fluid in which the nano-particles are suspended. Other parameters that affect the agglomeration include temperature and pressure. Only the effect of temperature is discussed in this thesis.

The agglomeration of particles is described in terms of volume by Smoluchowski's equations [39] as described below.

$$\frac{d}{dt}N(M) = \frac{1}{2}\sum_{K=1}^{M-1} a(M,K)N(M-K)N(K) - \sum_{K=1}^{\infty} a(M,K)N(M)N(K) \qquad (3.1)$$

It should be noted that the equation describes a physical process in a statistical sense. This means that the particle volume discussed here is a distribution. In the equation, the left hand side describes the rate of change of frequency of particles of volume *M*. The frequency discussed here is in the sense of probability distribution. The first term on the right hand side describes the gain in the frequency of the particles of volume *M*. This is contributed by the collision of particles of volumes *M-K* and *K*, whose sum is M. The index of summation goes from *K=1* till *K = M-1*, which means that each pair whose volumes adds up to *M* is counted twice. Hence there is a factor of half in front of the term. The second term on the right hand side describes the loss in the frequency of particles of volume *M*, which agglomerate with other particles ranging from the smallest to the largest. The net change in the frequency of particles of volume *M* is thus the difference between the frequency of particles that are gaining volume to become particles of volume *M,* and the frequency of particles of volume *M* that agglomerate with other particles, to form even bigger particles. Only low energy collisions are considered here, and situations involving particle fragmentation are neglected.

The term that is common among both the terms on the right hand side is the agglomeration kernel *a*. The agglomeration kernel determines the speed of agglomeration. This means that the kernel determines the speed or rate at which particles of volume *M* aggregate with other particles and also the rate at which particles of volume lower than *M* aggregate to become particles of volume *M*. From the earlier discussion, it becomes apparent that the kernel is related to the extent of Brownian motion of nano-particles in the fluid. In

the generalized case of RLA, the kernel is dependent on the surface area of the two particles about to agglomerate. The reason for this will be explained in the following sections.

3.4 Effect of Isoelectric point on agglomeration

Each material has an isoelectric point. Isoelectric point is defined as the pH of the solution at which the particles suspended in the fluid carry no charge on their surface. At this pH the particle surfaces carry no charge which implies that there is no hindrance for particle agglomeration. The Brownian motion is generally sufficient to allow particles to come in contact with each other subject to the concentration of the particles and viscosity of the slurry. At the isoelectric point, the particles agglomerate or stick to each other upon their first contact.

Amphoteric substance is one which can react both as an acid and as a base depending on the medium they are surrounded with. Most metalloids such as B, Si, As, Ge have amphoteric oxides. Cerium also has oxides that can behave both as an acid and a base. Apart from these oxides, ampholytes, which have both acidic and basic functional groups, are also amphoteric. Examples of these include polymers, amino acids, buffers and many other organic molecules. In CMP, typical slurry particles are made of oxides of silicon and/or cerium. Recently polymer particles have been found useful in alleviating the scratch generation problem in CMP. These polymer particles can also behave as both acidic and basic which implies that there is an optimal range of pH at which a specific polymer slurry particle may be used.

For polymer particles that have x acidic functional groups and y basic functional groups, the isoelectric point can be estimated theoretically using the formula $pI = (x\, pK_1 + y$

$pK_2)/(x+y)$, where pK_1 is acid-ionization constant and pK_2 is base-ionization constant. For the amphoteric oxide there is no simple formula and hence the isoelectric point has to be established from experiments.

When slurry particles are suspended in a fluid whose pH is away from the particles' isoelectric point, the particles carry a net surface charge which is proportional to how far the pH is from the isoelectric point. Since the particles carry a net charge on the surface, there is electro-static repulsion between the particles. This force restricts particles coming into contact with each other. Hence it takes more collisions for two particles to agglomerate at this pH compared to that at the isoelectric point. Since the collisions and the process of agglomeration is statistical in nature, the probability of agglomeration decreases. However, if the kinetic energy of the approaching particle(s) is high enough, they can cross the critical threshold distance between them after which van der Waal's forces become predominant and bind the two particles together. The kinetic energy required to penetrate the repulsive force is provided by Brownian motion.

There are three factors that affect the agglomeration of particles. The repulsive force exists because the particles carry similar charge on their surface. The van der Waal forces exist due to intermolecular attractions between different atoms carrying opposite charges. These forces are much smaller in range since the forces occur at atomic or molecular scale which is much smaller than particle scale, which is of the order of nanometers. Van der Waals forces are responsible for the particles to stick to each other when they come in contact. The third factor that affects agglomeration is the Brownian motion, which catapults the particles to approach each other closer than the repulsive forces allow them. This offsets the hindrance provided by the electro-static repulsive force.

3.5 Agglomeration Kernel

The kernel for RLA is given by [27]

$$a(M,K) = \frac{8K_B T N_0}{3\eta} \frac{(M^{1/3} + K^{1/3})^2}{M^{2/3} + K^{2/3}} e^{-\frac{U(r)}{k_B T}} \tag{3.2}$$

The kernel is a product of three different terms each describing the effect of one or more factors discussed in the previous section. The inverse of kernel determines the average time between two collisions. This means that the higher the kernel the smaller the time between collisions and the faster the agglomeration process.

The first term is the description of speed or rate of agglomeration. The agglomeration speed is directly proportional to the temperature of the particles. When the temperature is high the particles have higher kinetic energy and hence the particles are more likely to come into contact with each other as Boltzmann constant relates the energy at particle level with temperature observed at the bulk level. The speed of agglomeration is also directly proportional to the concentration of the particles present in the slurry. The more particles present in a unit volume the more there is chance for the particles to come into contact and hence the possibility of agglomeration. The first term also describes how the agglomeration rate is inversely proportional to the viscosity of the fluid (slurry). Since viscosity makes it difficult for the particles to move around, it reduces the rate of agglomeration.

The second term describes the effect of surface area on the speed of agglomeration. The numerator represents the surface area of the resultant particle whereas the denominator represents the sum of surface areas of the particles before agglomeration. Surface area plays a role in agglomeration because it is this quantity that determines the magnitude of electro-

static repulsion between particles. The higher the denominator, the higher the repulsion between the interacting particles, and hence the agglomeration rate is slower.

The third term contains the effect of electro-static repulsion as a function of the slurry pH. The third term is the exponent $e^{-\frac{U(r)}{k_B T}}$. In this term *U(r)* is the potential energy between two approaching particles whose agglomeration is to be determined. It is clear from the expression that as the potential energy (which is positive for repulsion) increases, the agglomeration rate decreases exponentially. The denominator of the exponent is $k_B T$, which represents the kinetic energy of the particles. As the kinetic energy of the particles increases it can withstand higher repulsive forces and still approach each other to an extent that van der Waal forces become predominant and bind the particles together.

The potential energy of the pair of particles in question has to be determined to calculate the rate of agglomeration for that pair. Since the slurry involves a range of particles with different sizes, the pair potential energy has to be calculated for every possible interaction. The potential energy is given by [40]

$$U = \frac{1}{2}\varepsilon r \psi_0^2 \ln\left[1 + e^{-\kappa l}\right] \tag{3.3}$$

$$\psi_0 = \zeta\left(1 + \frac{z}{r_s}\right)e^{\kappa z} \tag{3.4}$$

In the equations above

ε is the dielectric permittivity,

r is the mean radius of the particle,

ψ_0 is the particle surface potential,

κ is the Debye-Huckel parameter,

ζ is the zeta potential of the slurry

z is the distance between surface and slipping plane

r_s is Stokes radius

l is the separation between particles.

The maximum repulsion experienced by the particles is when they are closest at which point the term κl tends to zero. The distance between the surface and the slipping plane z is typically 1.5 nm [41]. Stokes radius is slightly smaller than the effective radius of the particles used in CMP. Stokes radius defines the maximum radius of influence of the particle. Since the distance between slipping plane and the surface is much smaller than the radius of particles used in CMP, the ratio $\frac{z}{r_s}$ is negligible. The fraction is typically relevant when calculating the surface potential of molecules whose sizes are of or below the order of nanometer. After these relevant assumptions, the potential energy for an approaching pair reduces to

$$U = \frac{1}{2}\varepsilon r \zeta^2 e^{2\kappa z} \ln 2 \tag{3.5}$$

The variable κ, which represents Debye length, needs to be estimated. This is the length in which mobile charge carrier screens out the external electrical field. This variable is given by the expression [42]

$$\kappa = \sqrt{\frac{4\pi e^2}{\varepsilon k_B T} \sum_i n_i z_i} \tag{3.6}$$

In the expression for Debye Huckel parameter,

e is the charge on an electron,

n_i is the number of ions that have valance z_i.

3.6 DLA as a Special Case

The agglomeration equation for RLA has been established in previous section. In RLA the particles carry a net charge on the surface which impedes the agglomeration process. As the pH of the slurry moves towards the isoelectric point of the particles the charge on the surface of the particles decreases and when the pH of the slurry is at its isoelectric point the particles have no charge on the surface and are free to agglomerate without any hindrance. This is the DLA regime.

When there is no charge on the surface, the pair potential energy U of approaching particles is zero. Thus in equation (3.2) the term that accounts for electro-static repulsion becomes unity. The second term which takes into account the surface area of the particles also becomes irrelevant. The term goes to unity as well since agglomeration no longer has surface area dependence. The resulting agglomeration kernel is independent of the volume of the particles and is given by

$$a = \frac{8K_B TN}{3\eta} \tag{3.7}$$

Here N is number of particles per unit volume at any time. Thus the particle agglomeration is limited only by the ability of the particles to diffuse, which is entirely dependent on the temperature, slurry particle concentration and the viscosity of the slurry liquid. Hence the name diffusion limited agglomeration (DLA) is apt.

3.7 Simulation of Agglomeration

The agglomeration equations require that the particles be characterized according to their volumes. This is because volumes are linearly additive and this makes the differential equation linear. The typical measurements of nano-particles are done using dynamic light scattering (DLS) [35, 43]. In this technique particles are characterized by their diameter in probability sense. The particle mean diameter and standard deviation are the available measurements from the technique.

Since the particles need to be characterized by their volumes, in agglomeration equation it is necessary to convert the diameter distribution yielded by the DLS. Given a diameter distribution of the particles $f_D(d)$, it needs to be converted to volume distribution $f_V(v)$ via the transformation $V = \frac{\pi}{6} D^3$. The volume distribution then becomes [44]

$$f_V(v) = f_D\left(\left[\frac{6v}{\pi}\right]^{\frac{1}{3}}\right)\frac{d}{dv}\left(\frac{6v}{\pi}\right)^{\frac{1}{3}} = \frac{1}{3} f_D\left(\left[\frac{6v}{\pi}\right]^{\frac{1}{3}}\right)\left(\frac{6}{\pi v^2}\right)^{\frac{1}{3}} \tag{3.8}$$

Once the simulations are run the time evolution of Smoluchowski equations yield the results in terms of volume of particles. This needs to be converted to the diameter distribution so that the result can be compared with initial particle diameter distribution.

The conversion is also necessary because the indentation depth calculations performed in later chapters use particle diameter distribution. The conversion is similar to equation (3.8) and is given by

$$f_D(d) = f_V\left(\frac{\pi d^3}{6}\right)\frac{d}{dd}\left(\frac{\pi d^3}{6}\right) = \frac{\pi d^2}{2} f_V\left(\frac{\pi d^3}{6}\right) \tag{3.9}$$

For the simulations, another input parameter, the pH of the slurry liquid is required for evolving the particle volume distribution. This parameter is required since pH directly influences zeta potential of the slurry. Zeta potential is the electric potential in the interfacial double layer at the location of slipping plane versus a point in the bulk fluid away from the interface. For typical silica based slurries used in CMP which are away from isoelectric point, zeta potential is typically under 60mV per unit concentration of ions present in the slurry [45, 46, 47] as shown in the figure 3.2.

Simulations have been carried out for silica particles used in CMP for pH values of 2, 4, 7 and 12. Isoelectric point of silica is 3 from figure 3.2. The input parameters used for the simulations are listed in table 3.1. The parameters are listed in various units that are conventionally used in physics. Though the units of input parameters may be different, sufficient care has been taken in most to ensure that proper conversion is made when using the parameters in the simulation. In the remaining cases the simulations need non-dimensional numbers and hence the units of two or more different parameters with same dimensions cancel, in which case the units of these input parameters are not relevant.

3.8 Simulation Results

The results of simulations will be presented and discussed in two subsections. One subsection discusses the generalized RLA case where as the other subsection discusses the special case of DLA. These subsections will discuss the evolution of particles diameter distribution as a function of time and as a function of pH. The dependence on pH will be discussed from the reference point of isoelectric point. Each subsection also discusses the evolution of mean diameter as a function of time.

3.8.1 RLA Simulation Results

Simulations have been carried out with different pH values as shown in table 3.1. Among them, all the cases with pH values other than 3 qualify to be deemed RLA since in all those cases the particles carry a significant net charge on their surfaces. Figure 3.3 through 3.7 describe the particle diameter evolution for different pH values.

From the figures 3.3-3.7, the effect of pH on slurry agglomeration is evident. In these cases the agglomeration takes place in reaction limited agglomeration regime and hence is slow. The speed of agglomeration decreases as the pH of the slurry moves away from the particles' isoelectric point.

Figures 3.3 and 3.4 are so similar to each other that there is no difference in the evolution of particle diameter distribution. This is because the isoelectric point is 3 and the pH values for the figures are 2 and 4 respectively. Since the difference in pH values from the isoelectric point is same, the magnitude of the charge carried by particles in each case is same. The only difference is that in one case the particles' surfaces carry negative charge while in the other case the particles' surfaces carry positive charge. From figure 3.2 it is clear that the zeta potential of the solution in the two cases is same in magnitude but opposite in

sign. From equation (3.4) this implies that the surface potential ψ_0 in both cases is same in magnitude but opposite in sign. But equation (3.3) for pair potential energy contains the term ψ_0^2, which makes the sign of the potential irrelevant. This is consistent with the fact that energy is a scalar. These two cases demonstrate that for pH values closer to isoelectric point (low zeta potential) the particle surfaces do not carry sufficient charge to makes the colloids stable. This result is consistent with [48], which concludes that for this value of zeta potential the colloidal stability is described as "incipient instability". The simulations are run only for 50 minutes of real time and within this time the agglomeration is significant. In CMP the slurry's pH is adjusted much before the process, which means the slurry has more time to agglomerate. For these periods the slurry behavior is clearly unstable.

Figure 3.5 shows the diameter distribution evolution for pH value 5. It is clear that for the pH difference of 3 from isoelectric point, the slurry has sufficient zeta potential to prevent aggressive agglomeration. The slurry in this case can be considered moderately stable compared to the pH values of 2 and 4.

Figures 3.6 and 3.7 represent the particle diameter evolution for the pH values of 7 and 12 respectively. It can be easily noticed that in both cases the particles have not agglomerated at all. This is because the particle surfaces carry very high charge which is represented in terms of high zeta potential. The particle behavior is stable for the timeframe of simulations. The slurry in this case is deemed stable.

Figure 3.8 represents the evolution of mean diameter of the particles for different pH values. From the figure it is clear that for pH values close to isoelectric point the evolution of mean diameter is significant. For the pH values far away from isoelectric point mean

diameter is almost constant. The increase in mean diameter happened due to reduction in the number of clusters when agglomeration takes place and the volume of the particles is conserved during agglomeration.

3.8.2 DLA Simulation Results

In DLA the agglomeration is much faster and hence most of the agglomeration happens during the first few seconds. The third term of agglomeration kernel in equation (3.4) is unity, which implies that the agglomeration is limited by diffusion alone. Thus there are more collisions per unit time, which implies faster agglomeration. There are few problems that arise due to the speed of agglomeration. Any computer simulation is limited by the necessity to define the upper bound of the entity to be simulated, volume in this case. From equation (3.4) it is clear that, as this upper bound of volume increases the simulation time increases as a square function of increase of the upper bound. This upper bound when applied to distributions is termed bin size. In DLA situation the agglomeration is so fast that the bin size is exhausted within seconds. Hence the study of diameter distribution evolution or mean diameter evolution is not possible beyond the first few seconds.

It is evident from figure 3.9 that by 15 seconds the bin size is already exhausted. Hence the analysis that follows is going to be limited to first 12 seconds. Figure 3.10 shows the slurry behavior at lower diameter. This is a zoomed in version of figure 3.9.

To alleviate this problem an analytical expression has been developed [49, 50] in the past. This is done exclusively in empirical fashion with no phenomenological support. Hence the terms of the expression on either side of the equality are not dimensionally consistent.

$$N(M) = \frac{N_0}{\bar{M}}\left(1 - \frac{1}{\bar{M}}\right)^{M-1} \tag{3.10}$$

Here $\bar{M}$ is mean volume and N_0 is initial particle concentration. The mean volume $\bar{M}$ at time *t* is given by

$$\bar{M} = tC + 1 \tag{3.11}$$

The error in the analytical expression mentioned above is that the expression assumes constant agglomeration kernel. It can be observed from equation (3.7) that agglomeration kernel itself is a function of number of particles per unit volume. As agglomeration proceeds, the number of particles decreases and hence the kernel decreases. This effect is significant for DLA since the agglomeration is very fast, which means that the rate of decrease of kernel is also very fast. This is an effect that can be captured by analytical expression since there is no solution for Smoluchowski equations with time dependent kernel. This is an obstacle for progress in solving the differential equations analytically. The simulations have the hurdle in terms of limited bin size. Thus a precise solution to DLA regime remains elusive. Approximations for limited time and limited bin size however are available. The solutions for longer time are not available. The DLA case is unstable.

3.9 Conclusions

For the first time a generalized RLA model is develop to predict the agglomeration behavior of homogenous nano particles suspended in a liquid. The model is able to predict the behavior for wide range of pH. It has been shown that the particle diameter evolution is entirely dependent on the proximity of slurry pH to the isoelectric point of slurry particles.

These solutions for generalized RLA case will be used in the following chapters for the prediction of scratch intensity and scratch frequency.

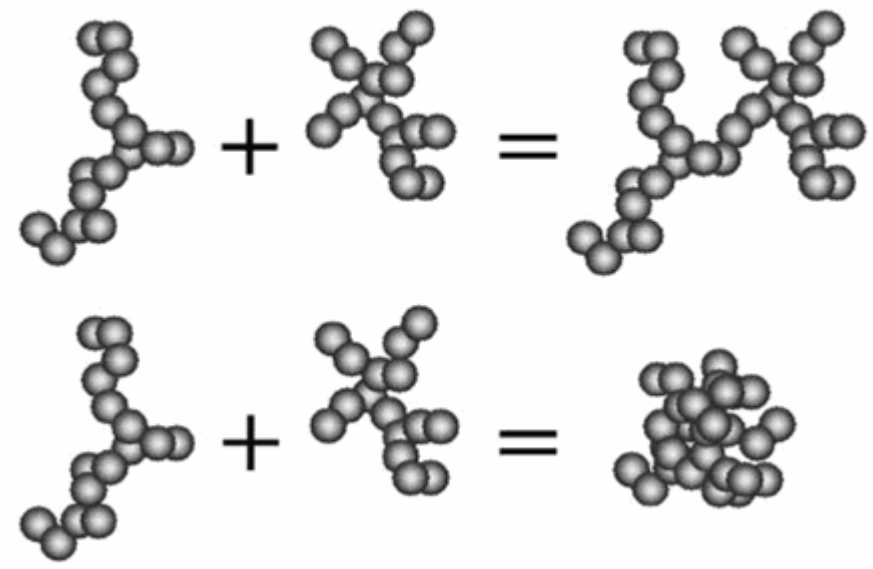

Figure 3.1. Schematic of particle agglomeration (adapted from [51])

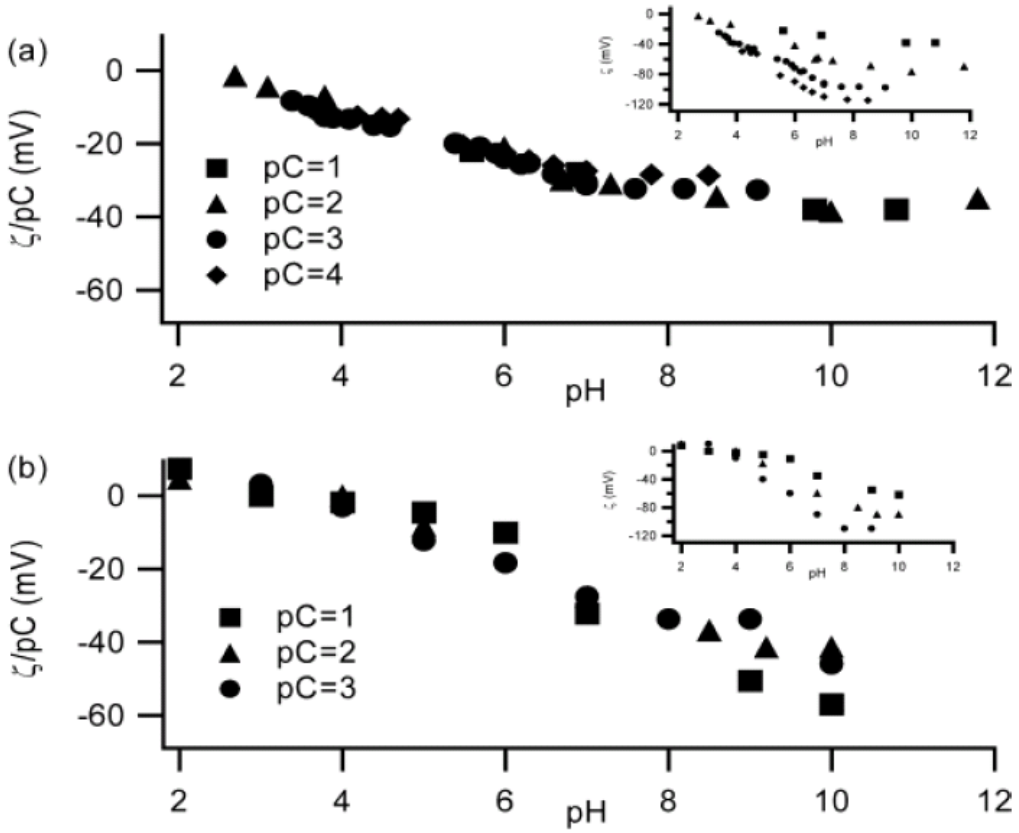

Figure 3.2. Zeta potential measured as a function of pH and counterion concentration. Top: data from Ref. [44]; bottom: data from Ref. [47]. In both cases, z is plotted normalized by pC as a function of pH. For both investigations, the z *vs.* pH relationship can be collapsed to a single curve over decades of counterion concentration change. Insets: plots of absolute z show variation if pC normalization is not employed. (Adapted from [45])

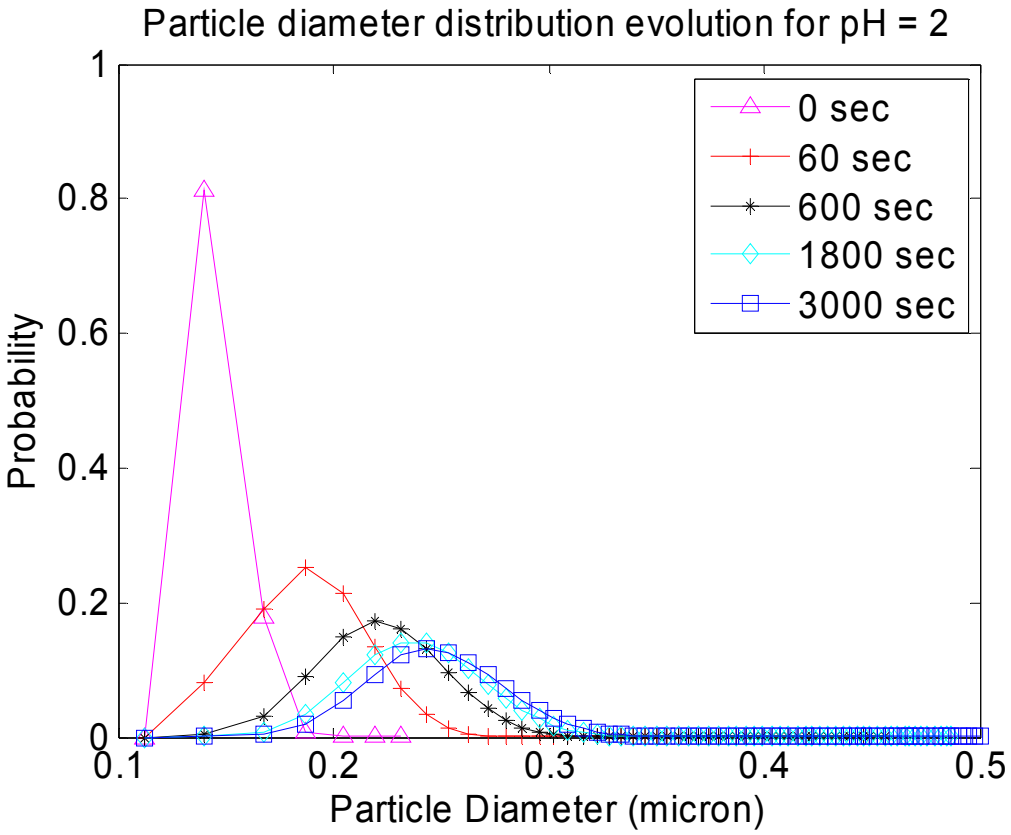

Figure 3.3. Particle diameter distribution evolution for pH = 2

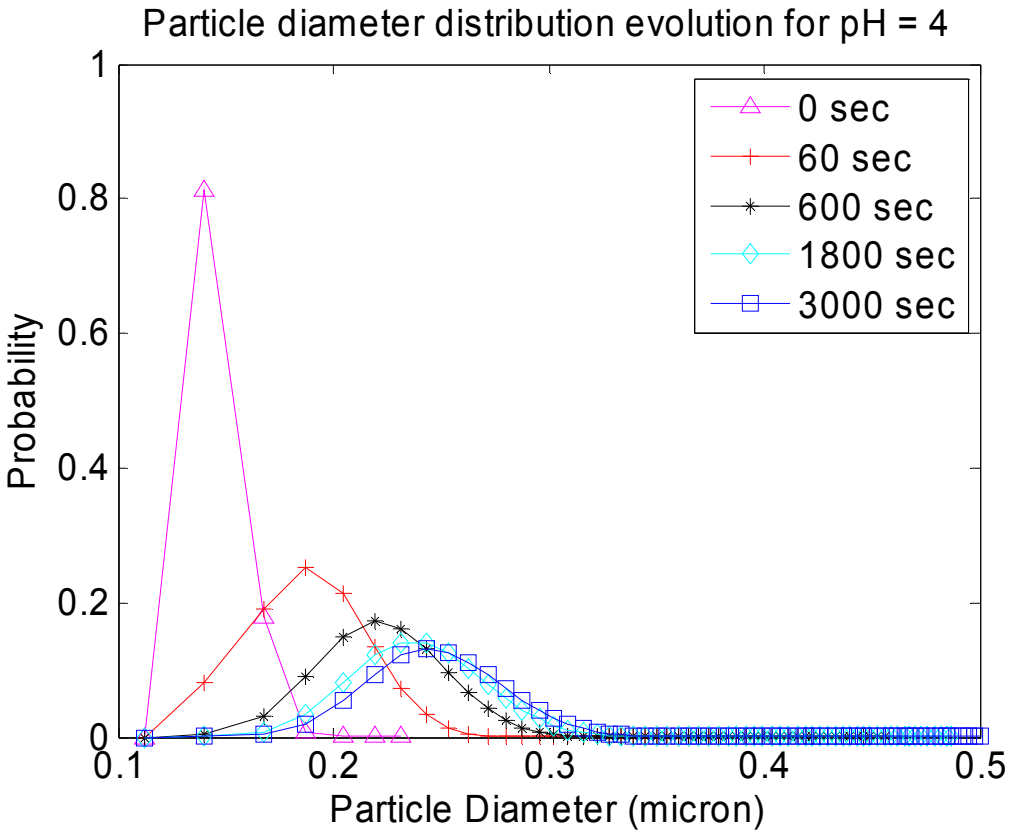

Figure 3.4. Particle diameter distribution evolution for pH = 4

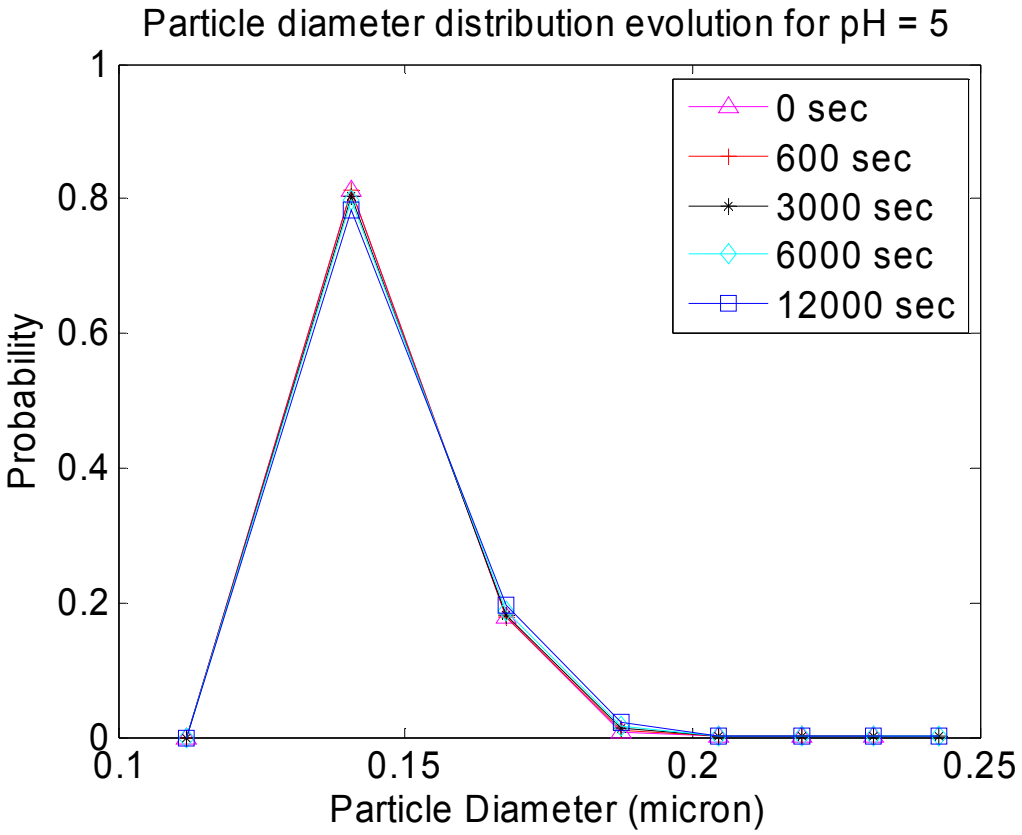

Figure 3.5. Particle diameter distribution evolution for pH = 5

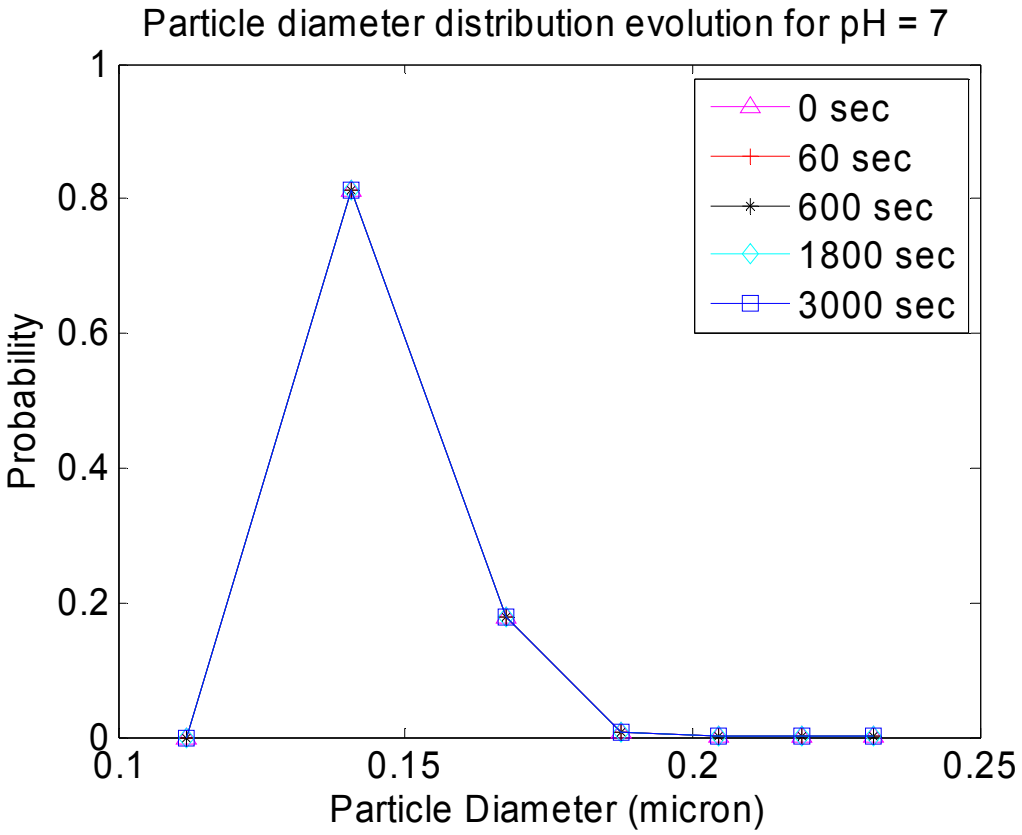

Figure 3.6. Particle diameter distribution evolution for pH = 7

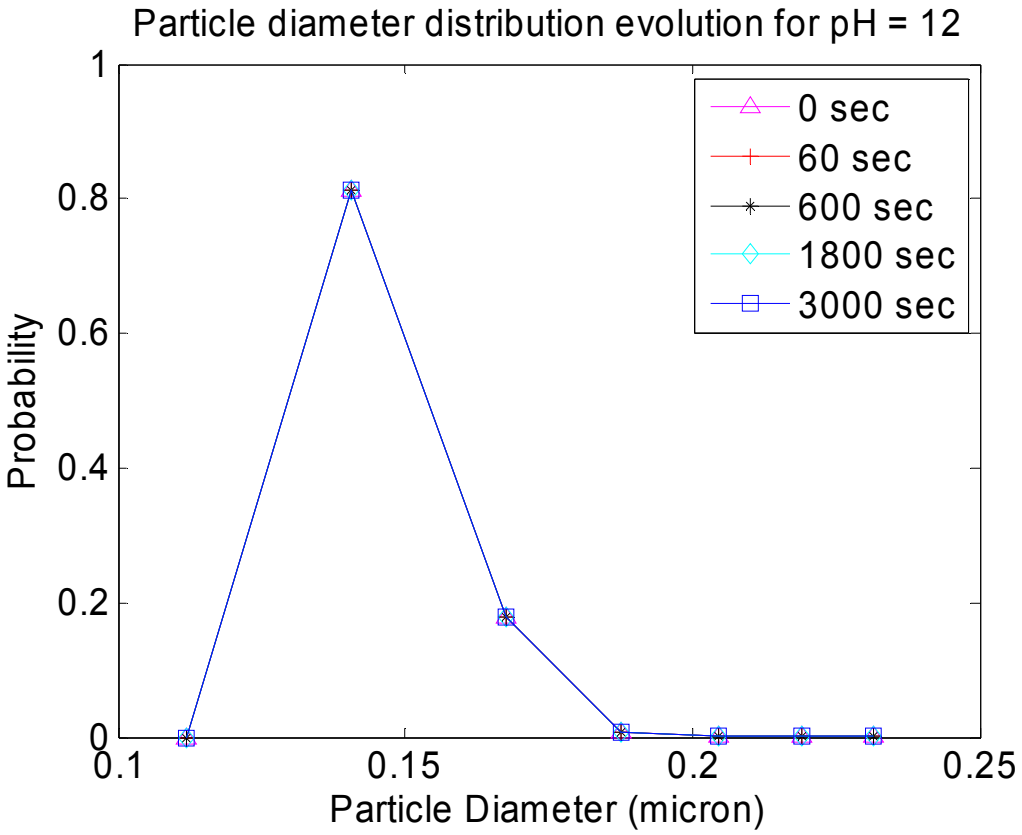

Figure 3.7. Particle diameter distribution evolution for pH = 12

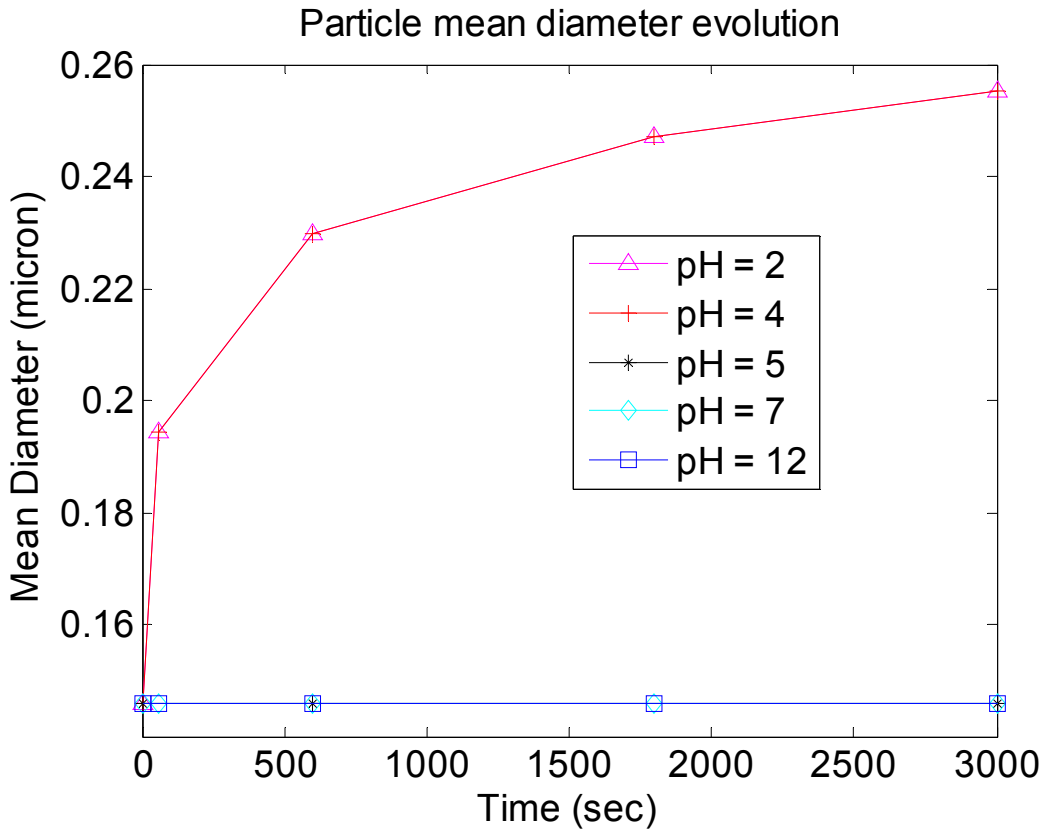

Figure 3.8. Evolution of mean diameter of particles for various pH values

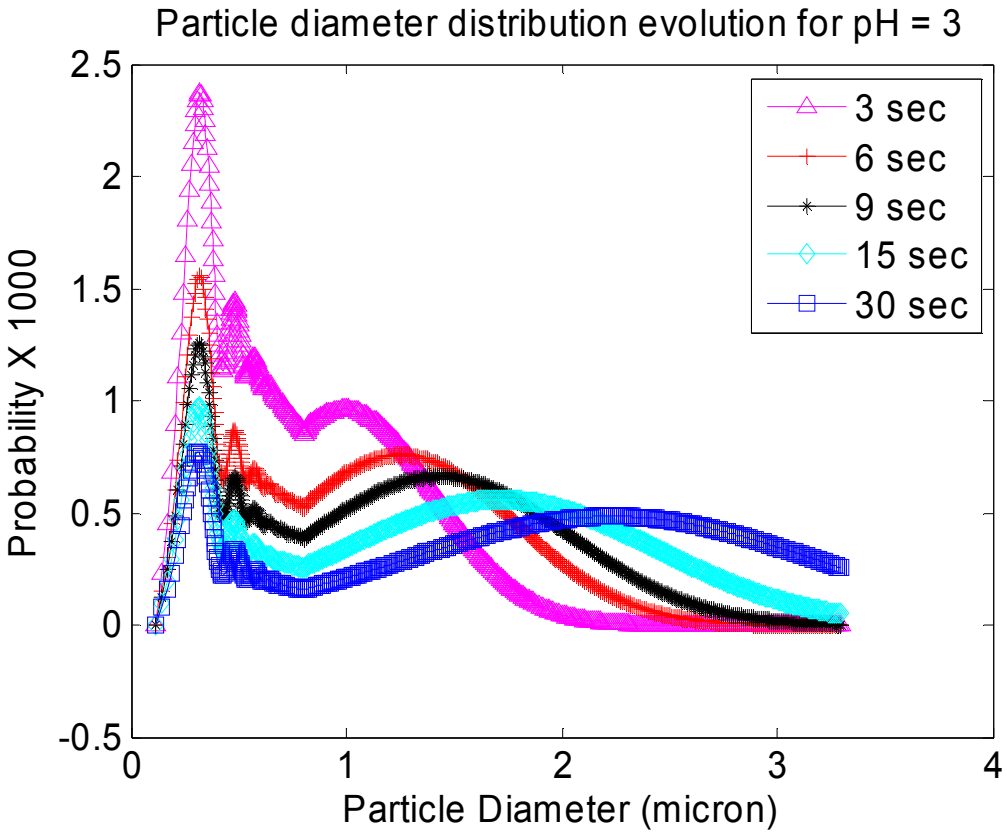

Figure 3.9. Particle diameter distribution evolution at isoelectric point

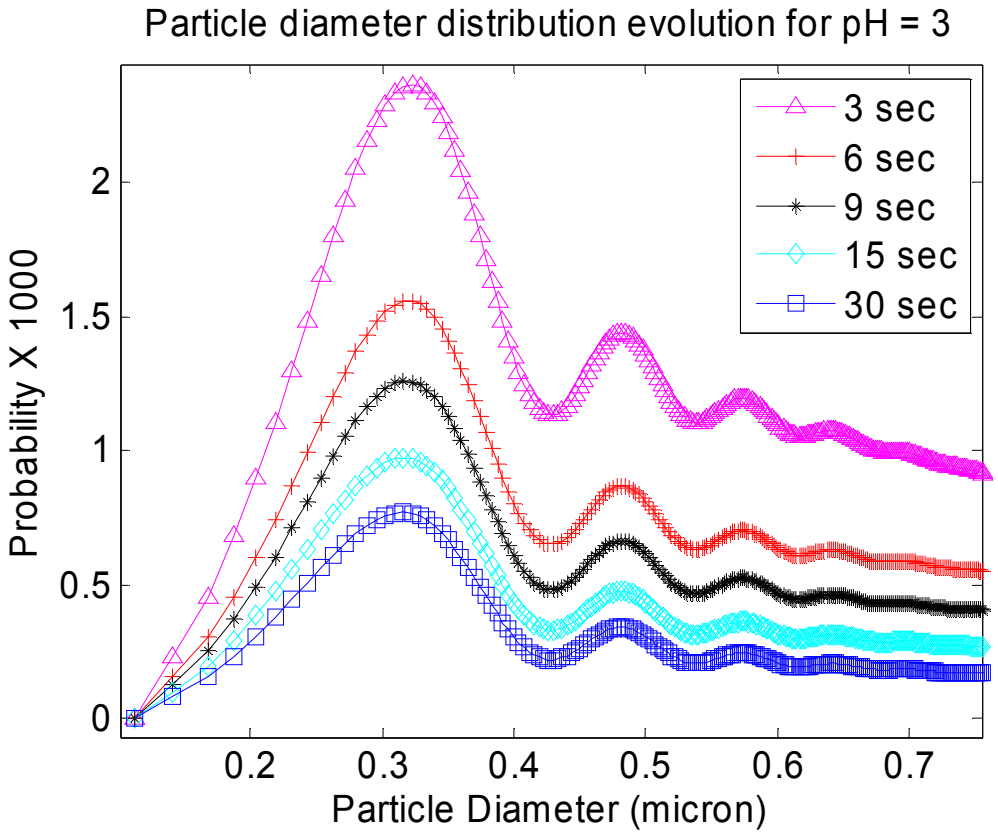

Figure 3.10. Particle diameter distribution evolution at isoelectric point, behavior at lower diameter

Table 3.1. Variables used in simulations

Variable	Value	Units
Particle mean diameter	100	Nm
Particle Standard Deviation	15	Nm
Viscosity of Slurry	8.9E-4	Pa.s
Temperature	300	K
Boltzmann Constant	1.38E-23	J/K
pH	2, 3, 4, 5, 7, 12	-
Charge of electron	1.6E-19	Coloumb
Volume concentration	0.02	-

CHAPTER 4. SCRATCH GENERATION IN CMP

Chapter 2 describes the evolution of pad asperity height with time. Chapter 3 describes the evolution of slurry particle distribution with time. In this chapter the two will be combined to predict the scratch generation propensity in CMP. It should be noted that the models described in the previous chapters as well as the model proposed in this chapter are probabilistic in nature and hence the quantities that are predicted in this thesis are probabilistic in nature.

The probabilistic nature of the quantification of scratch generation is due to the disparity between wafer-scale and particle-scale descriptions of the CMP process. The nature of predicting wafer scale quantities such as the material removal rate from the particle scale parameters and micron scale pad asperity height distribution necessitates the introduction of such models. Such probabilistic descriptions allow us to establish ranges for wafer-scale output variables.

In this chapter the probability of scratch generation in CMP will be predicted. In the next chapter an effort will be made to explain how the schematic of traditional CMP causes the relatively high scratch propensity.

4.1 Material Removal Process in CMP

Scratch generation is an essential ingredient of the CMP process as shown in figure 4.1. Material removal in CMP takes place via abrasion of slurry particles on the wafer surface. Wafer surface in oxide CMP is harder than the particles. However, chemical reaction between slurry liquid and the wafer surface makes the wafer surface softer than the slurry

particle thus making the abrasion and hence material removal, possible. In oxide CMP as well as in glass polishing this occurs as silica surface reacts with the chemically active slurry to form silicon hydroxide layer as shown in figure 4.2. This layer is much softer than silica. In Silicon Nitride wafers the stress applied during polishing along with the presence of water (typically) makes the Silicon Nitride wafer transform the top layer first into silica and then into silicon hydroxide [52].

$$Si_3N_4 + 10H_2O + CeO_2 \rightarrow 3SiO_2 + 4NH_4OH + CeO_2$$

$$SiO_2 + H_2O \rightarrow Si(OH) + O_2$$

The nature of pad-particle-wafer interactions essentially affects the scratch generation process. Different regimes of pad-particle-wafer contact have been proposed by Bastawros and Che [6] as shown in figure 1.6. In the figure the first mode of contact involves the transfer of load from pad to wafer via particle. In this mode pad has no contact with the wafer. Thus the entire load is applied on the particle. This happens for relatively low loads and stiff pad. In the second mode pad is partially in contact with wafer while transferring the load through the particle as well as well as directly. The load transmission directly from the pad to wafer is relatively low compared to the load transmitted through the particle in the second mode. This occurs for moderate loads and pads with relatively low stiffness. In the third mode of contact, the pad completely engulfs the particle and significant amount of load is transmitted through the pad to wafer as well, thus bypassing the particle. In this mode the particle experiences relatively lower loads compared to total load exerted by pad asperities.

In this thesis the first mode of contact is assumed. In light of high pad bending stiffness and pad bending factor [13] at the feature scale, this assumption is reasonable for relatively moderate loads applied here. The particle is surrounded by pad when the pad transfers the load through the particle to the wafer. For any symmetric revolution of angle ϕ, shown in figure 4.3, the force of contact between pad and the particle is given by

$$F(d) = \frac{1}{4}\int_{\theta=\phi}^{\frac{\pi}{2}}\int_{\psi=\phi}^{\frac{\pi}{2}} P_{cont} d^2 \sin\theta \sin\psi \, d\theta d\psi = \frac{P_{cont} d^2 \cos^2\phi}{4} \tag{4.1}$$

When the contact reaches at least half the diameter the angle ϕ becomes zero. This is the case in all the modes shown in figure 1.6. For this case the force on each particles reduced to $\frac{P_{cont} d^2}{4}$. This is the load on each particle of diameter d under an asperity which experiences a contact pressure P_{cont}.

4.2 Indentation Depth

Using the force transmitted by a particle under an asperity indentation depth is calculated. A particle under ith asperity experiences a pressure given by

$$P_{cont}(i) = \frac{L_i}{A_i} = \frac{4E\sqrt{\kappa_s}}{3\pi}\sqrt{Z_i - s(t)} \tag{4.2}$$

And hence jth particle of diameter d under such an asperity carries a load of

$$F(i,j) = \frac{1}{4} d^2(i,j) P_{cont}(i) \tag{4.3}$$

The wafer is assumed to behave like a rigid - perfectly plastic body [25, 6]. For oxide polishing, this assumption is reasonable since the elastic regime of response is almost negligible. The indentation depth w, is inversely proportional to the hardness of the wafer H_w and the expression for indentation depth by j^{th} particle under i^{th} asperity, assuming that the particle diameter is much larger than indentation depth, is given by [25]

$$w(i,j) = \frac{2F(i,j)}{\pi H_w d(i,j)} = \frac{P_{cont}(i)}{2H_w} d(i,j) = \frac{2E\sqrt{\kappa_s}}{3H_w \pi^2} d(i,j)\sqrt{Z_i - s(t)} \tag{4.4}$$

The expression (4.4) calculates the indentation depth for a single particle. The indentation depth on the left hand side is a random variable since it is a product of functions of two other random variables, namely particle diameter and asperity height. These two random variables are assumed to be independent of each other in this thesis. This is reasonable since there is no dependence of diameter on the height of the asperity under which it is present. There is no known physical law that reveals a pattern of particle distribution on the pad. Hence it is assumed that any particle can be present under an asperity of any particular height.

Let f_D be the probability density function of particle diameter and f_Z be the probability density function of asperity height. The cumulative density function of indentation depth F_W is given by

$$F_W(w) = \int_0^{D_{\max}} \int_0^C f_D(d) f_{\bar{Z}}(z) d(d) d\bar{z} \tag{4.5}$$

In the expression $\bar{z} = z - s$, and $C = \left(\frac{3wH_w\pi^2}{2E\sqrt{\kappa_s}} \right)^2$. The probability density function can be obtained by differentiating the cumulative density function with respect to indentation depth.

4.3 Material Removal

Since the particle diameter distribution and particle indentation depth distribution are known the material removed by each particle can be estimated. This can be done using the similar technique (equation 4.5) described in the previous section.

The particle diameter d and particle indentation depth w are known. Let a be the semi contact radius of the indented surface and 2θ be the angle subtended by the arc at the center of the particle. Let A be the area of the cross section ploughed by the particle in the direction of velocity V. Then from geometry the following can expressions can be written. Here m is material removal rate.

$$a = \sqrt{dw - w^2} \tag{4.6}$$

$$\theta = \tan^{-1}\left(\frac{a}{\frac{d}{2} - w} \right) \tag{4.7}$$

$$A = \frac{\theta d^2}{4} - a\left(d - 2w\right) \tag{4.8}$$

$$m = N_{act}\left[\frac{\theta d^2}{4} - a\left(d - 2w\right)\right]V = N_{act}\left[\frac{\tan^{-1}\left(\frac{\sqrt{dw - w^2}}{\frac{d}{2} - w}\right)d^2}{4} - \sqrt{dw - w^2}\left(d - 2w\right)\right]V \tag{4.9}$$

MRR m, is a random variable since it depends on other random variables indentation depth (w) and particle diameter (d). The material removal rate m can be calculated from the distributions of diameter and indentation depth. Thus, following a procedure similar to equation (4.5) the probability density function of MRR, can be computed and the expected value of the distribution is the mean material removal rate. Thus the mean material removal over a finite time may be estimated from particle diameter and mean indentation depth. It should be noted here, that the usual Preston equation [14] producing a linear MRR in PV is not utilized here. Instead, a Preston type response at the macro-scale is recovered from the micro-scale based mechanistic model developed in this thesis.

4.4 Validation of the model

Experiments have been conducted [52] on silicon nitride wafers with water based slurry using ceria particles (isoelectric point 6.2). The parameters used in the experiment are listed in table 4.1. Zeta potential of ceria slurry is shown in table 4.2. The pH of slurry is reported to change during the process from 6.5-8. DI water, though neutral on acid-base scale, immediately after exposure to air, reacts with the CO_2 in the air to form carbonic acid.

This makes the slurry slightly acidic before the process begins. The experiment is ideal for the model because it contains no other chemical substance in the slurry other than DI water. In traditional CMP slurries of different pH values are used. Since this involves addition of reactive chemical agents, part of the material removal comes from the dissolution of wafer surface. In the current experiment this effect is minimized because no other chemical agents are present.

Simulations have been run with experimental conditions as input using both linear and non-linear pad responses.

Particle agglomeration effect is shown in figures 4.4-4.7. The pH of the slurry is close to the isoelectric point of the slurry particles, hence the agglomeration is rapid, which is evident from rapid shrinking of the initial peak of diameter distribution in figure 4.5. The effect is shown close up in figure 4.5. The mean particle diameter increases with time and the number of particles decreases accordingly.

From the figure 4.8 it can be seen that scratch increases with load but not linearly. This is because, as load increases more asperities come into contact and the load is shared, hence the number of scratches increases.

Figure 4.9 compare the material removal obtained by using the process described in section 4.3. The comparison is done by fitting one point (150kPa) to the experimental data and then scaling the other data points using the same constant. The material removal due to linear pad changes linearly with time and matches the experimental data well.

4.5 Fractal Dimensions

Agglomerates tend to have geometry different from classical geometry as evident from figure 4.12. First introduced by Benoit Mandelbrot in 1975 [53], this geometry is based on self-similarity in s structure. This means that a fractal has similar structure at different length scale and hence called self-similar. It has been useful in explaining many natural phenomena since its introduction and also has helped design antenna for modern cell phones which is required to absorb a wide range of frequencies.

Fractal geometry is different from Euclidean geometry in that the exponents of radius in defining volume or area need not be a natural numbers. They can be rational numbers. It has been established that [35, 43] agglomerates have fractal dimension ranging from 1.86 to 2.1. This is because there is plenty of empty space in the agglomerates which makes the volume not proportional to r^3 but proportional to r^n where n = 1.86-2.1. In a strict sense as the shape tends to be spherical, meaning no agglomerates, the exponent tends to 3.

Since the agglomerates in CMP exhibit fractal nature in their volume the particle diameter is larger than for a sphere of the same volume. For DLA the fractal dimension has been established [35] at 1.86 while for RLA this is closer to 2.1. This is reasonable since there is aggressive agglomeration in DLA and the resulting structure farther away from spherical geometry whereas in RLA the agglomeration is slow and limited and hence the structure is closer to spherical geometry.

As far as indentation is concerned experiments need to be conducted to study the behavior of agglomerates under stress. Since the agglomerates are conglomerates of particles they have more compliance under stress than a spherical particle and equation (4.4), strictly speaking, is not applicable. This effect is not included in this thesis as there is not enough

research on the behavior of agglomerates under stress. Hence equation (4.4) is applied for indentation depth calculations to determine the worst case scenario due to agglomerates.

4.6 Parametric study

Parametric study has been conducted to ascertain the effect of pad linearity, pH of slurry and pad evolution on scratch propensity. For the parametric study the parameters described in chapters 2 and 3 are used. The oxide film on the wafer is assumed to have a hardness of 4GPa.

Two different sets of parameters have been used for the study. The first set of parameters is the same as the one used in the validation process. This is used for section 4.6.1 which studies evolution of pad and wafer.

Second set of parameters are the ones used in chapter 2 and 3. This is necessary to study the effect of pH on agglomeration and the effect of fractal dimensions on scratch depth as the experimental conditions involves single pH.

4.6.1 Effect of Pad and Slurry Evolution

Particles agglomerate as shown in figure 4.4. Pad evolution is discussed in chapter 2. The scratch depth and scratch frequency change as the slurry and pad evolves. Figure 4.10 shows the evolution of scratch with time. Figure 4.10 shows the effect of evolution of both slurry and pad on material removal rate. The longer the pad is used the more the asperities are worn out. This causes decrease in separation distance and hence more asperities come into contact and the contact pressure on asperities decreases. This causes reduction in material removal rate. Since pad evolves slowly, the effect is slow. As slurry evolves, mean particle diameter increases as different particles add in volumes and the number of particles

decreases linearly. The diameter of particles, which dictates indentation depth, does not increase as fast as the number of particles decreases and hence the material removal rate decreases. This effect is seen in figure 4.11.

Figure 4.14 show the effect of pad evolution alone for the conditioned used in parametric study. For evolving pad the maximum scratch depth decreases with time. Thus the frequency of shallower scratches reduced while the frequency of deeper scratches reduces. As explained earlier this is due to increase in the number of asperities that are contact, thus reducing the mean contact pressure. The non-evolving pad can be assumed to be aggressively conditioned. For all times the scratch propensity, measured in terms of the frequency of maximum scratch, is higher for conditioned pad compared to evolving pad. Let threshold scratch depth be defined as the depth, scratches deeper which cause defects. For any value of threshold and for any time the conditioned pad performs worse. Thus an evolving pad helps to alleviate scratch problem.

4.6.2 Effect of pH

As discussed in chapter 3 the speed of slurry agglomeration depends on the proximity of slurry pH to the isoelectric point of the slurry particles. Scratch probability for three different values of pH is calculated. For pH values of 2, 3, 7 the scratch probability is shown in figures 4.13, 4.15 and 4.16 respectively. It should be noted that for figure 4.16 all the curves overlap as there is no agglomeration and pad evolution is not considered. The pH value of 3 corresponds to fastest agglomeration regime of DLA. Since the agglomeration is very fast the bins in simulation get exhausted or filled at a very rapid rate. Hence scratch is predicted for only one instant of time. For this time instant it can be easily seen that the

maximum scratch depth is about 14 nm compared to maximum scratch depth of 1.75 nm for pH of 2 and 0.75 nm for pH of 7.

4.6.3 Effect of Fractal Geometry

The effect of fractal geometry is shown in figure 4.17. The case with pH of 3 is undertaken. Scratch probability for exponents of 3 and 1.86 is plotted. The first exponent means that the agglomerates are assumed to be spherical while the second exponent means that the agglomerates have extreme fractal geometry. For the fractal case it is assumed that the indentation behavior of the fractal is no different from spherical particle.

From the figure it can be easily seen that the fractal geometry has significant impact on the scratch propensity. The spherical particle produces a scratch of depth 15nm whereas the fractal agglomerate of same volume produces a scratch of 50 nm. If the fractals have similar indentation characteristics the impact on scratch probability is significant. Even for a fractal agglomerates with some compliance the scratch probability is still greater compared to a spherical particle. Thus fractal nature of agglomerates has deteriorating effect on the final surface quality.

The fractal nature of agglomerates yields much higher scratch intensity and material removal rate. However this is not used in the validation process because the effect of stress on particles with fractal geometry is not yet known.

4.8 Conclusions

The model developed in this thesis is validated. Effect of different parameters on defectivity has been studied. Conditioned pad or used (evolved) pad can be used according to the requirements in defectivity. To improve defectivity, the pad can be conditioned during

the initial stage of CMP and the conditioning can be stopped (or made less aggressive) when the wafer surface approaches target profile.

From the studies it is clear that the parameter that has most significant impact on scratch is the proximity of slurry pH to the isoelectric point of slurry particles. In the control of defectivity, slurry pH plays the most important role. The other factor considered is the fractal nature of agglomerates. The speed and extent of agglomeration is entirely dependent on slurry pH and hence the fractal exponent is also entirely dependent on slurry pH. The lower the exponent the worse the scratch if the indentation behavior of these particles under stress is assumed to be similar to spherical particles. Hence pH plays a role in producing deeper scratch not only by increasing the volume of each particle but also by reducing the fractal exponent.

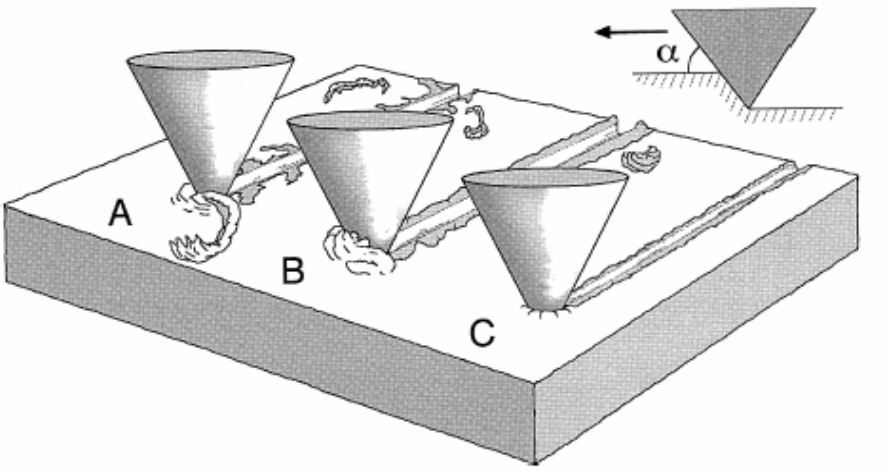

Figure 4.1. Schematic of material removal in CMP. Rake angle here is negative (adapted from [22])

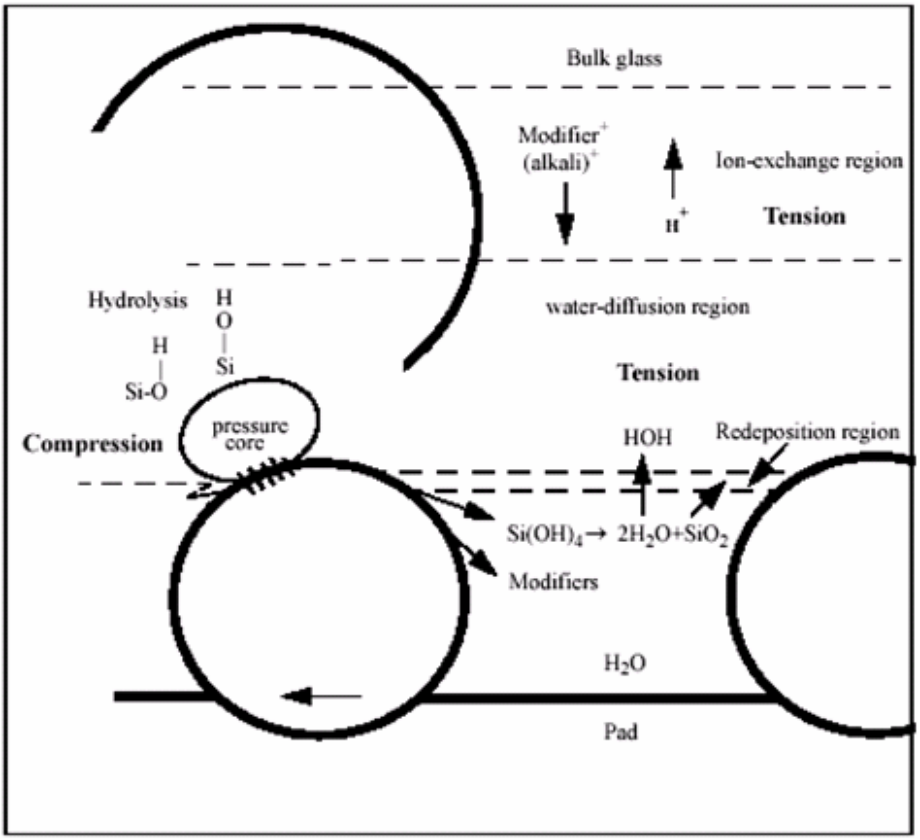

Figure 4.2. Mechanism of hydrolysis of silica [54]

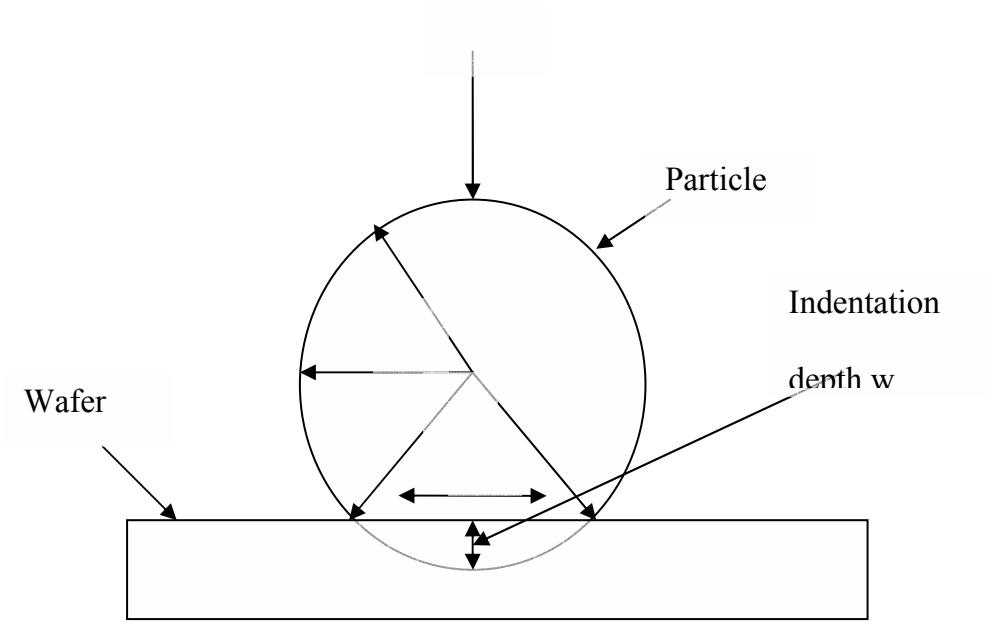

Figure 4.3. Schematic of particle indentation, intersection of wafer and particle denotes cross section area A, angle ϕ determines the mode of contact

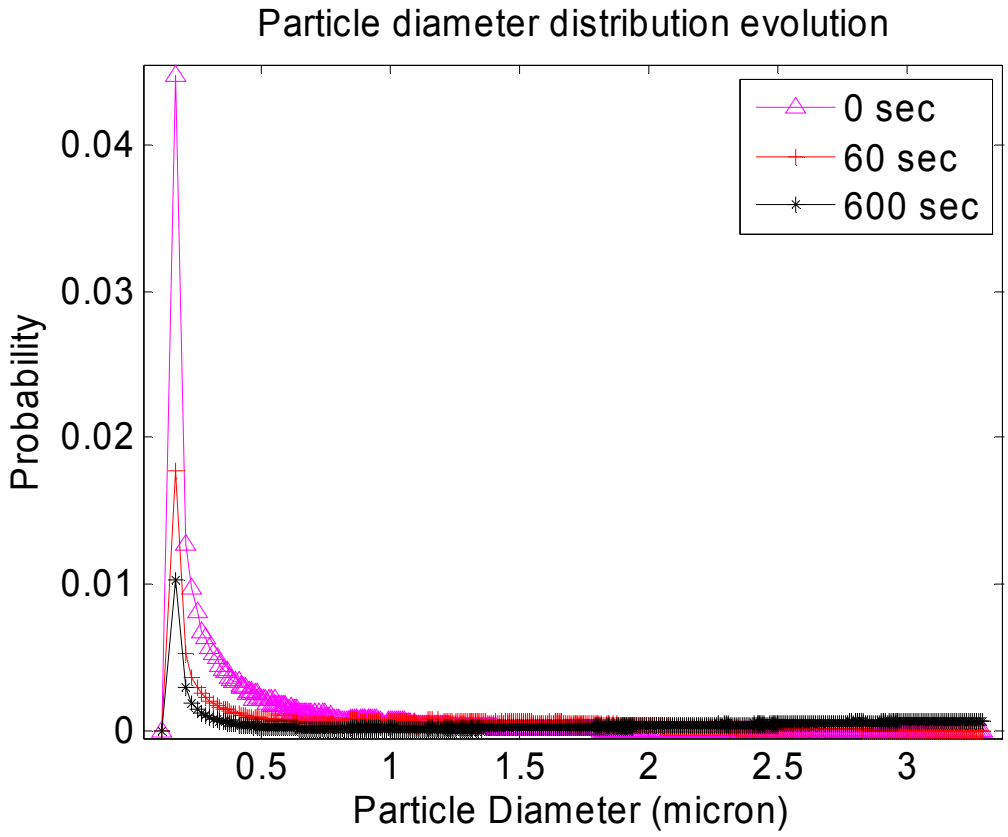

Figure 4.4. Particle diameter distribution evolution for experimental conditions

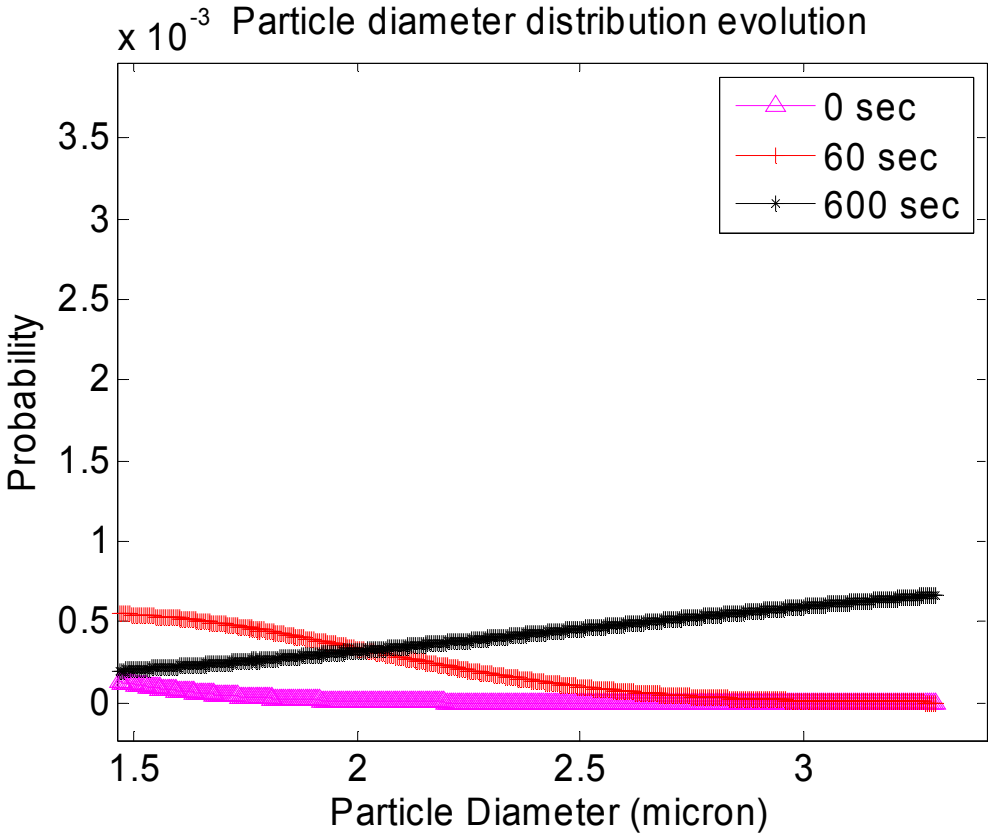

Figure 4.5. Particle diameter distribution evolution for experimental conditions, tail portion of figure 4.4

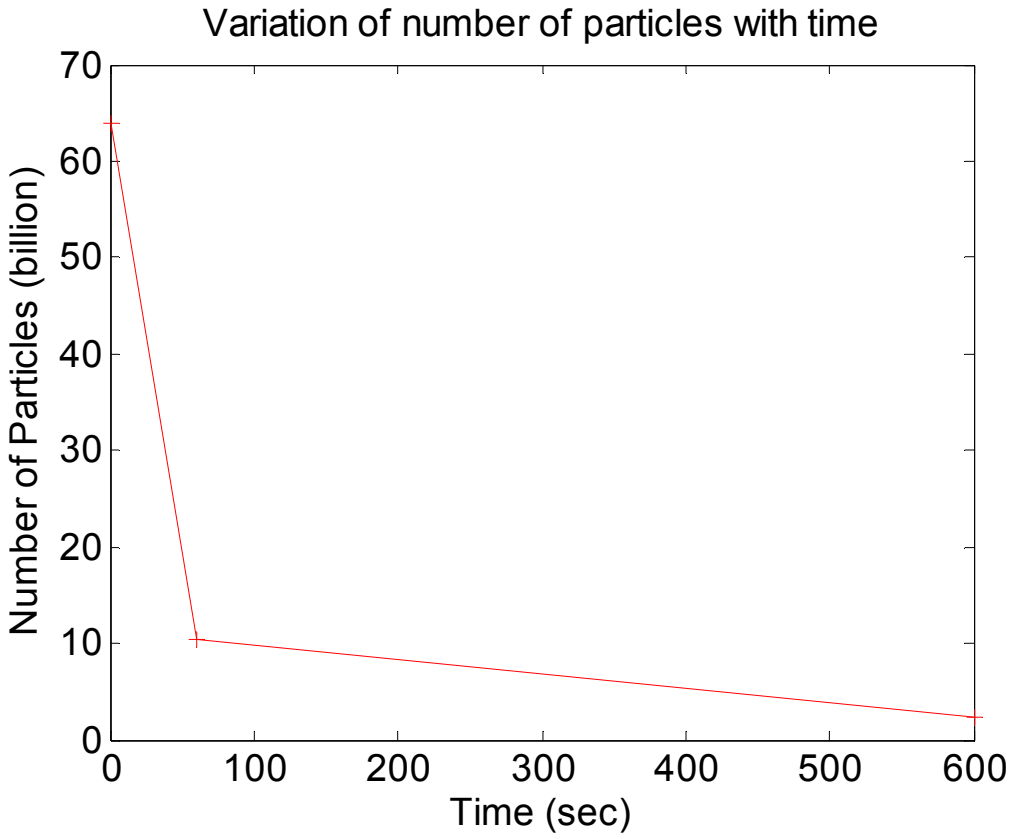

Figure 4.6. Variation of number of particles with time for experimental conditions

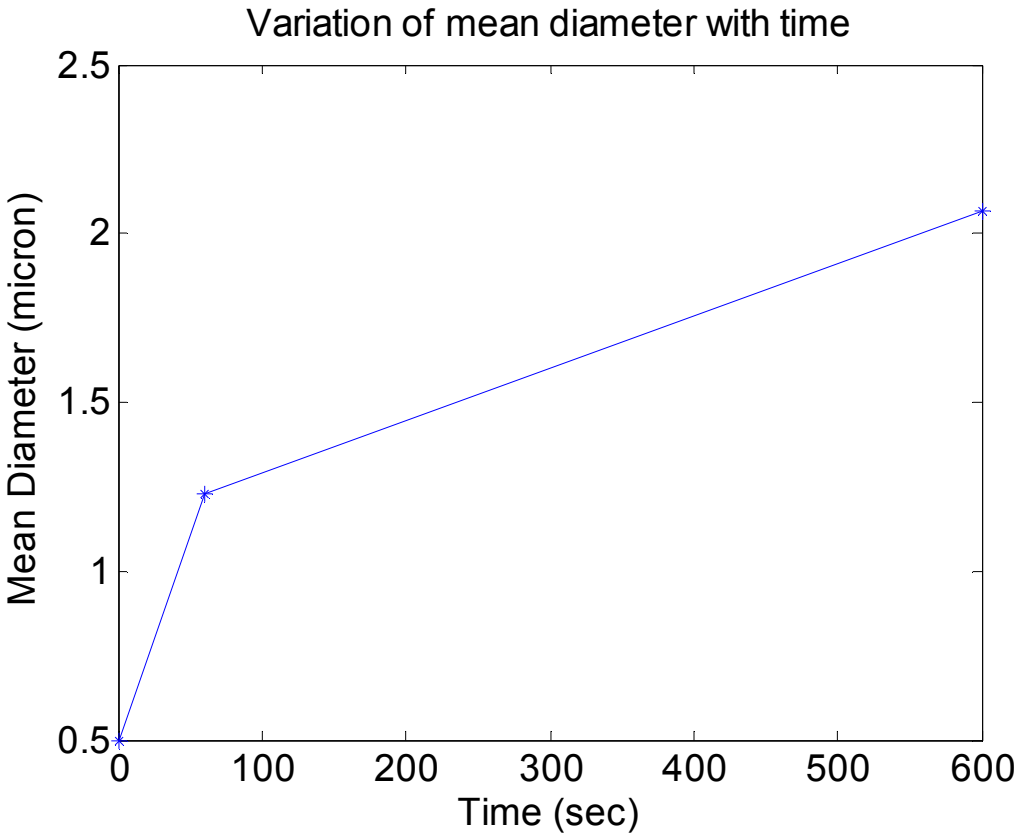

Figure 4.7. Variation of particle mean diameter with time for experimental conditions

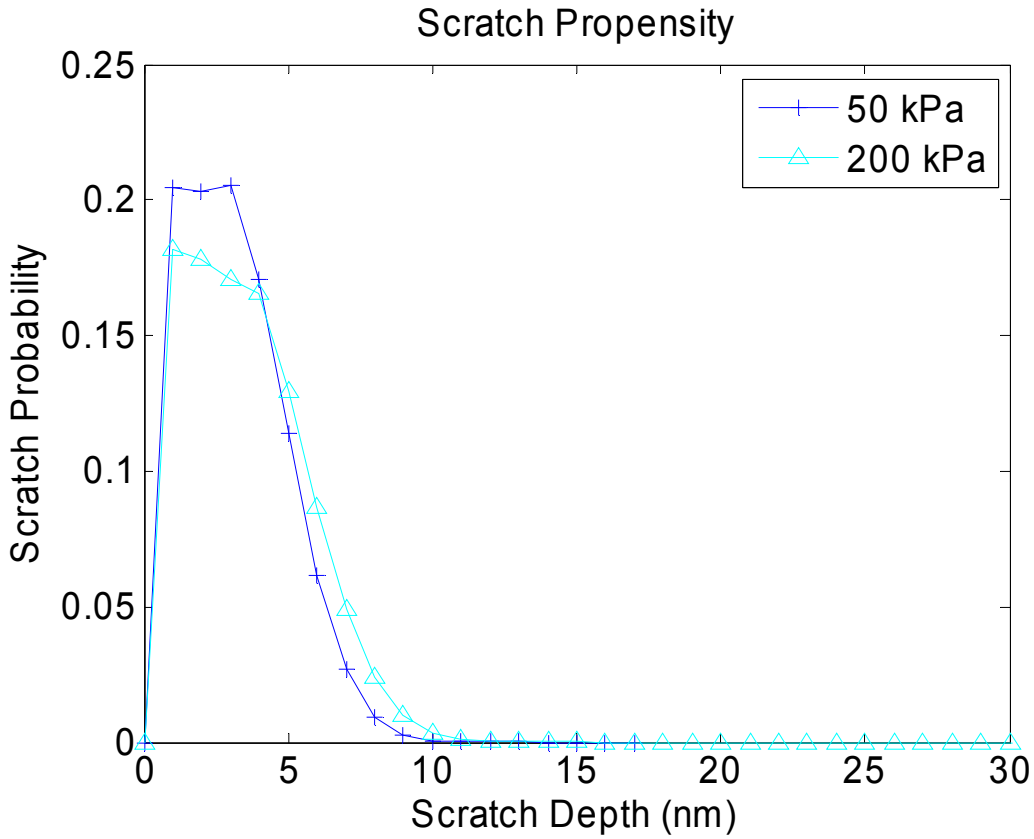

Figure 4.8. Simulated scratch depth distribution for experimental conditions with linear pad. Scratch depth is not linear with load

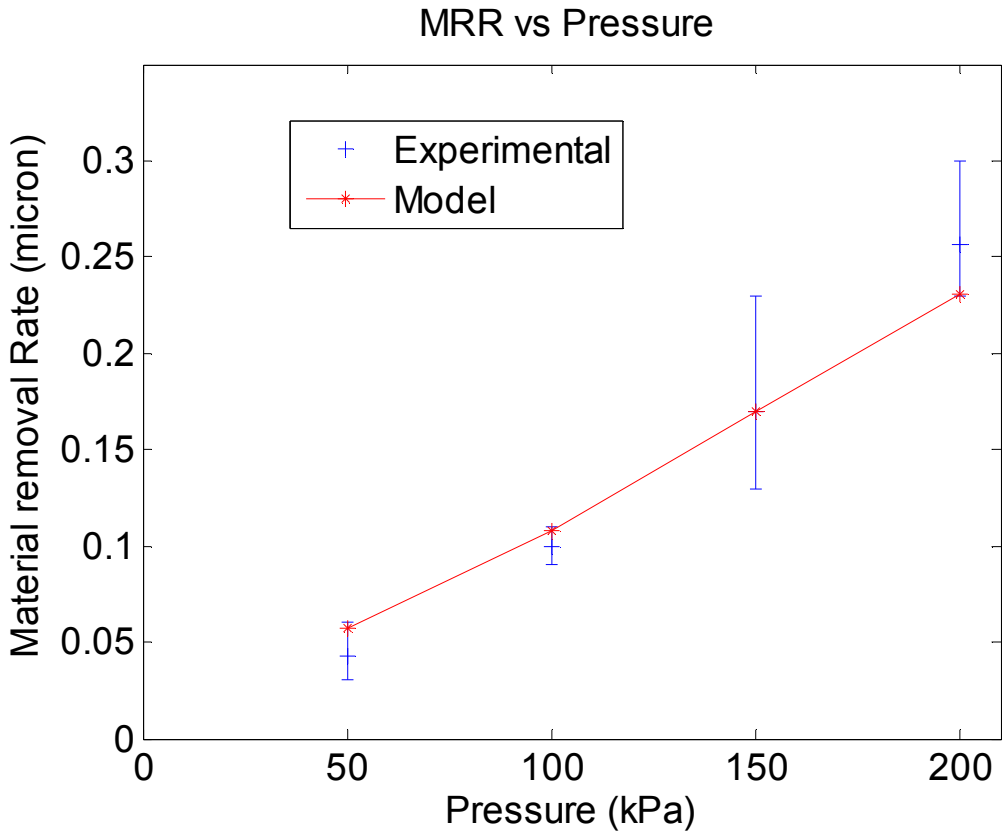

Figure 4.9. Model validated against experimental results at initial time

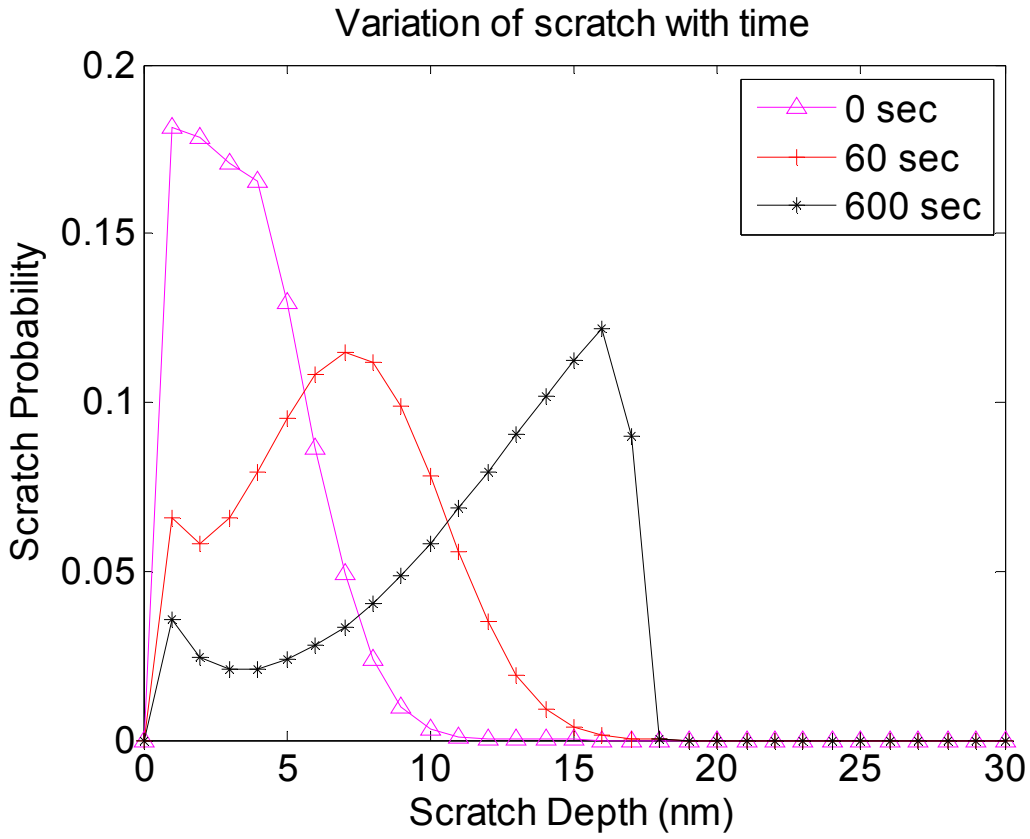

Figure 4.10. Variation of scratch depth with time for evolving slurry

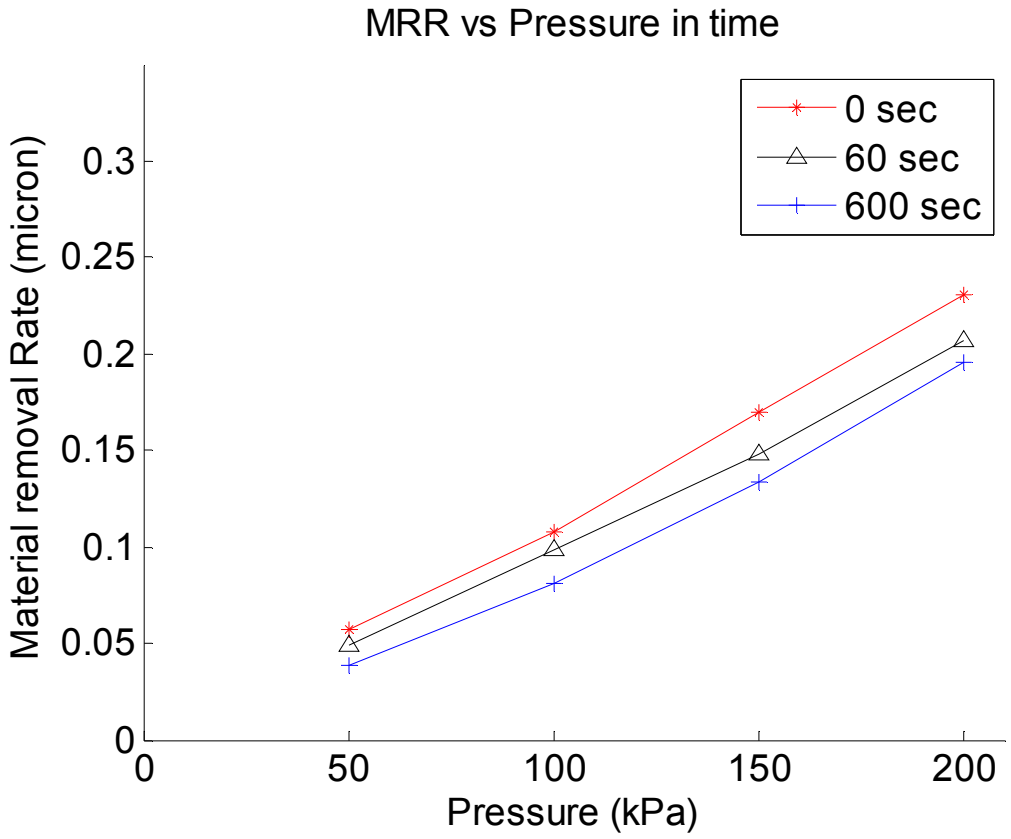

Figure 4.11. Variation of MRR in time with respect to pressure, effect of pad and slurry evolution

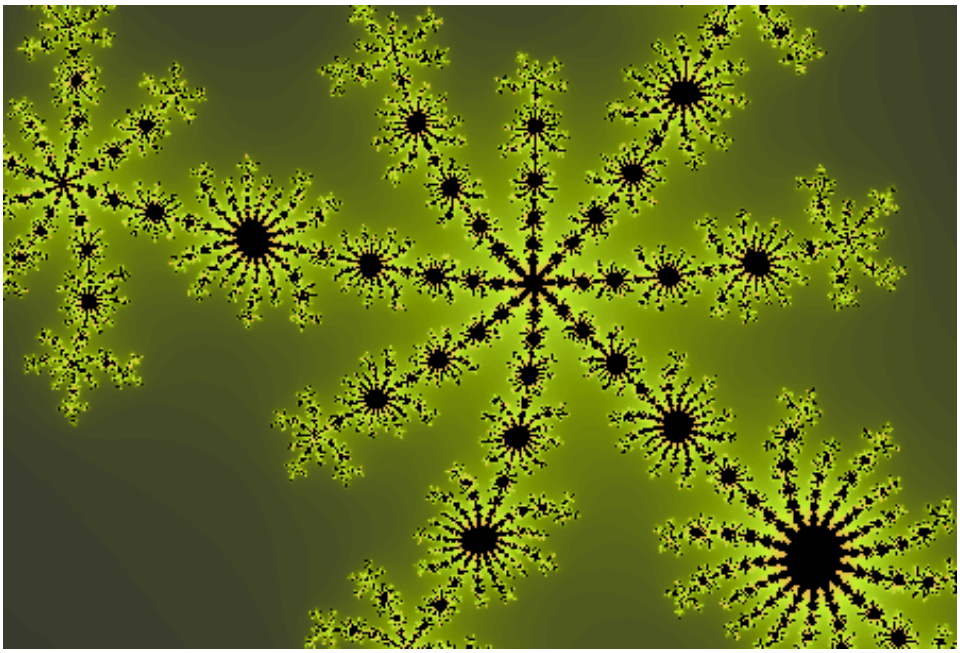

Figure 4.12. Geometry of fractals, extent of agglomeration determines fractal dimension [55]

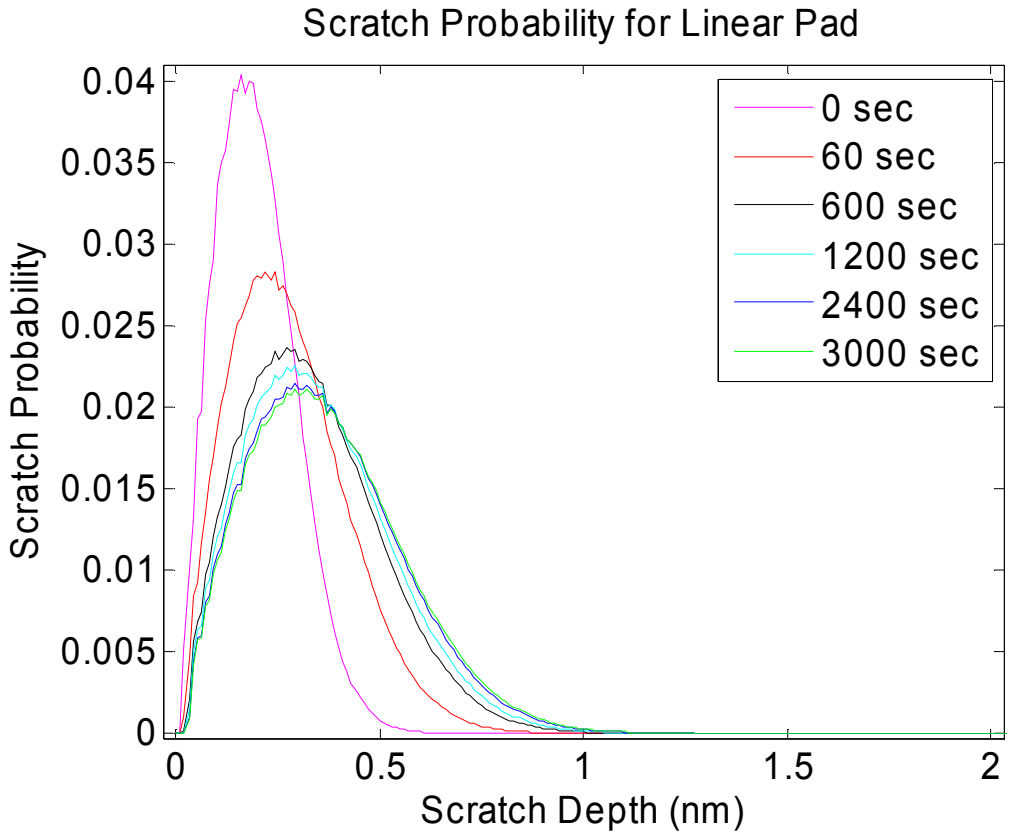

Figure 4.13. Scratch probability for linear pad for parametric study

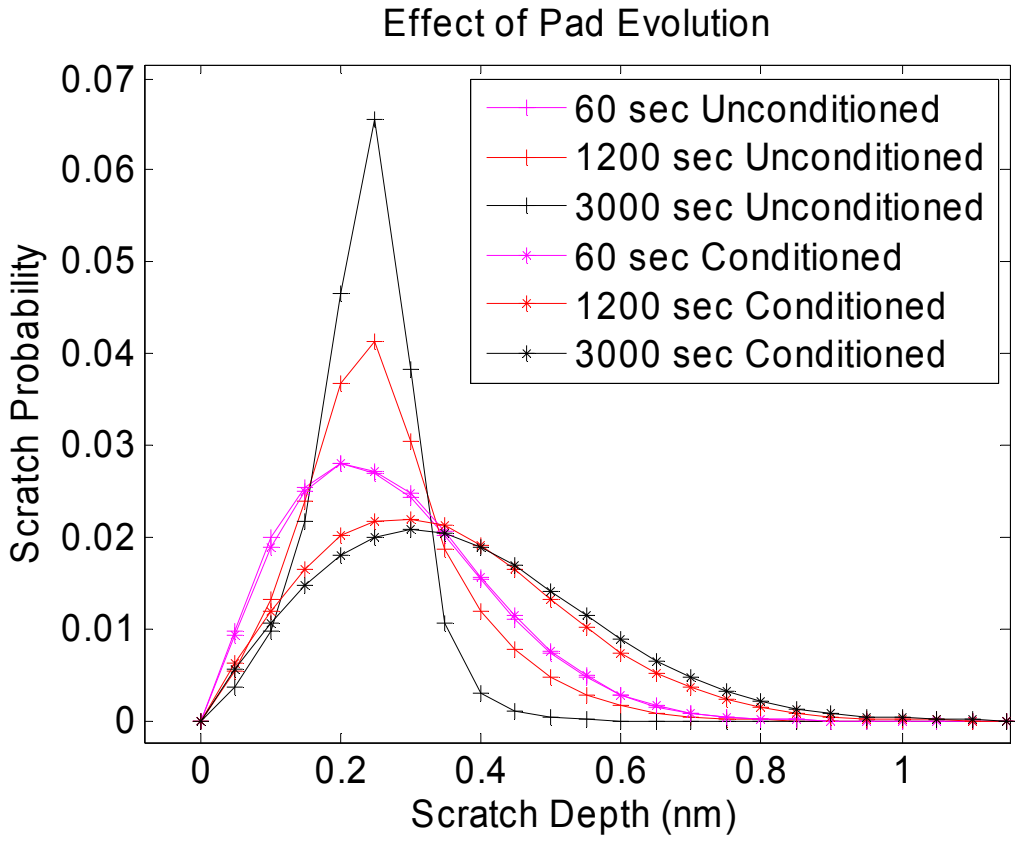

Figure 4.14. Scratch Probability with evolving pad, scratch depth increases with conditioned pad and scratch becomes uniform as pad evolves

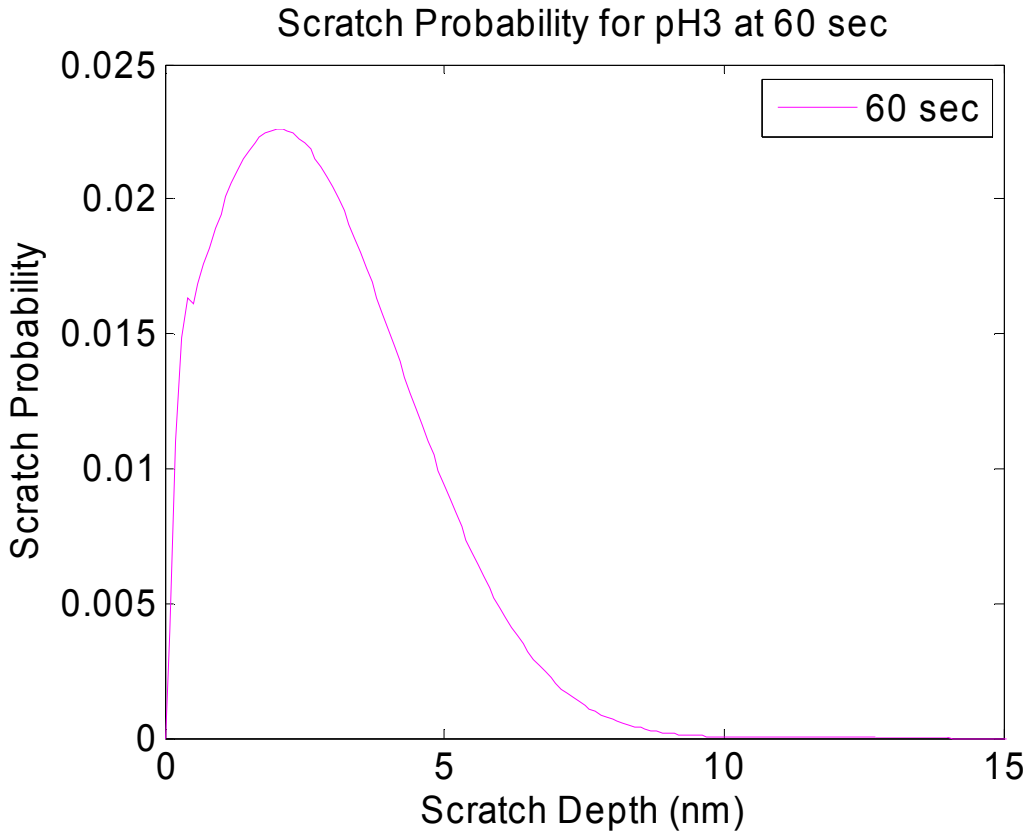

Figure 4.15. Scratch probability for pH 3 at 60 sec, much higher compared to figure 4.14

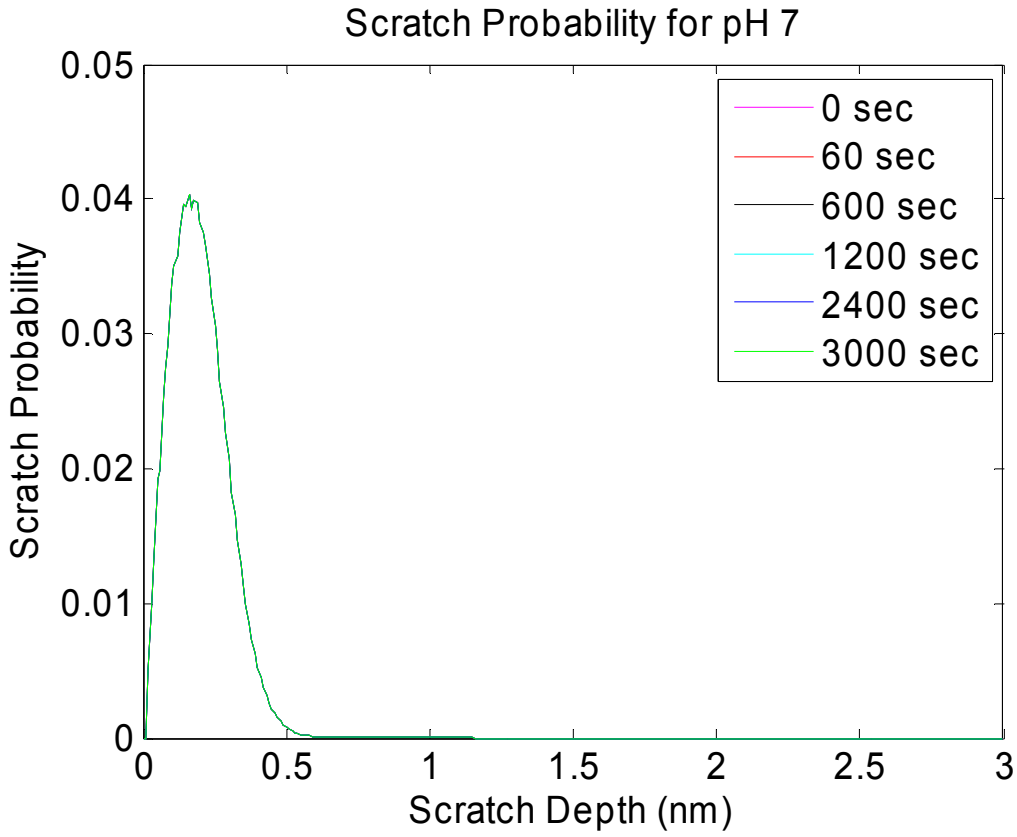

Figure 4.16. Scratch probability for pH 7 at 60 sec, all the curves over lap because pad is not evolving and there is no agglomeration when slurry pH is away from isoelectric point of particles

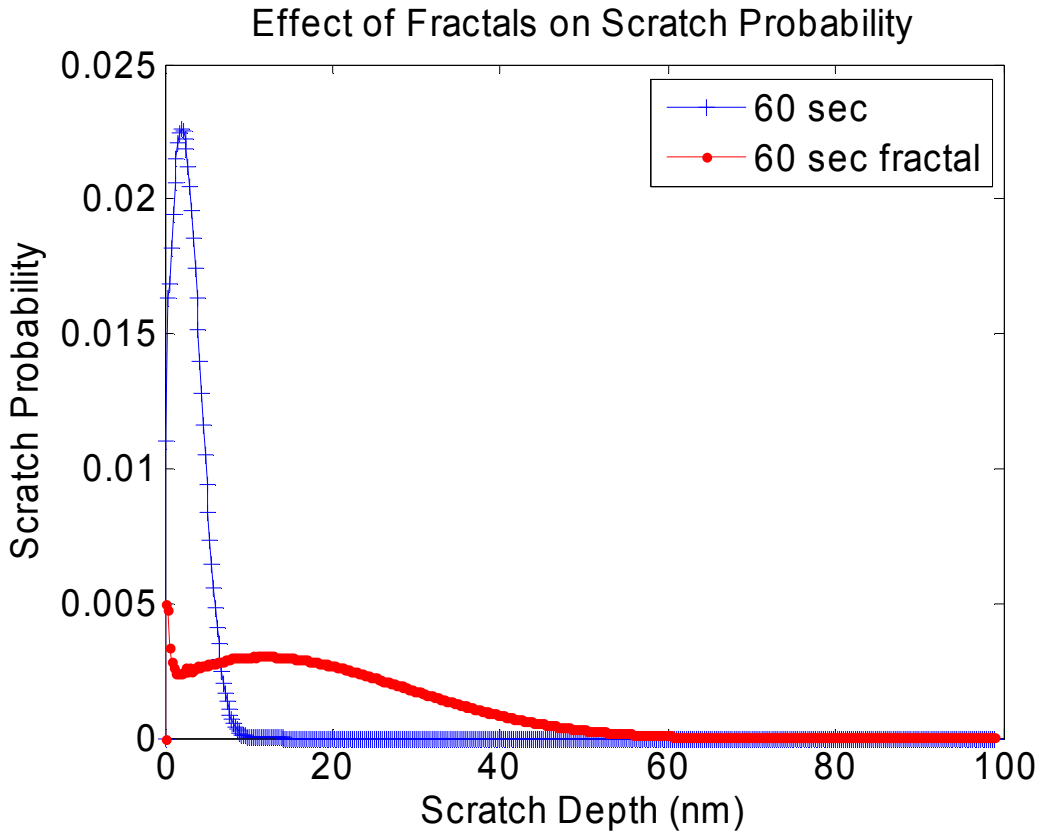

Figure 4.17. Fractals increase scratch propensity, if mechanical behavior of fractals is assumed to be similar to spherical particles

Table 4.1. Experimental conditions for model validation [52]

Property	Value	Units
Pad Modulus	25	MPa
Asperity Curvature	0.02	micron^{-1}
Particle mean Dia	500	nm
Particle Std Deviation Dia	500	nm
Relative Velocity	2.03	m/s
Slurry particle concentration	5%	vol
Slurry flow rate	5.33	l/min
pH of slurry	6.5	
Wafer/Pad Diameter	50/300	mm
Slurry Viscosity(Water)	0.001	Pa.s
Temperature	300	K
Wafer Hardness(hydrated)	15.2	GPa
Nominal Pressure	200	kPa

Table 4.2. Zeta potential of (90%) ceria [56]

pH	Zeta potential
3	15
4	10
5	5
6.2	0
7	-5
8	-10

CHAPTER 5. CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

CMP is a crucial step in semi conductor manufacturing. Its significance has been on the rise with the shrinking device size. To improve the efficacy of the process, mechanistic models are needed to reliably predict product quality and process efficiency. The models will help calibrate the input parameters to achieve optimal output characteristics in terms of product quality and process efficiency. They can also provide a measure for robustness of the process. In CMP, most of the models developed so far have been deterministic. Recently probabilistic models have gained prominence, however the models to predict macroscopic properties from particle scale material removal processes have been lacking.

The macroscopic phenomenological Preston law has been extensively used in CMP to predict material removal rate. Despite its success at macro-scale the law does not explain the underlying mechanisms of material removal.

In the current work a model has been developed that can predict the macroscopic material removal rate in CMP from mechanisms that occur at micro-scale. The current model requires multiple input parameters from all components of CMP to predict material removal rate. The mechanism by which each parameter influences the material removal rate is investigated. Models for individual components are combined to predict wafer-scale material removal rate.

The qualitative effect of pH on CMP process has been established by the industry. One of the main objectives of the thesis has been to quantify the effect of agglomeration as a function of pH. Existing literature on the quantitative influence of pH on zeta potential has

been used. Zeta potential has been used to estimate the repulsion between nano-particle and their tendency to agglomerate to form larger clusters. Thus the time evolution of nano-particles as a function of pH has been established.

The particle evolution is then combined with pad evolution to predict the scratch propensity at feature scale and material removal rate at wafer-scale. This behavior is combined with slurry evolution to predict scratch propensity as a function of slurry evolution.

The model prediction is compared with experimental data the results match well. Parametric study has been conducted to evaluate the effect of different parameters used in the model. Pad evolution is recommended for the alleviation of scratch problem. However, this comes at the expense of material removal rate. The pad structure can be altered to achieve a better defectivity performance. The pad cellular structure can be shrunk so that there are more cells to absorb the load from asperities. This also increases the stiffness of pad cellular structure. An investigation needs to be conducted to determine which effect dominates the other for a given pad.

The quantitative effect of slurry pH on scratch propensity is also studied. As the pH of slurry becomes closer to the isoelectric point of slurry particles, the agglomeration becomes aggressive and the scratch performance worsens. The agglomeration does not help material removal either.

5.2 Future Work

Despite involving many parameters from many components used in CMP, the model still does not account for certain aspects of CMP process. These aspects can be pursued as an extension to the current work.

The first one is the visco-elastic behavior of pad. The pad is made of polyurethane and hence has a visco-elastic component to its behavior. This behavior makes the pad respond not only to the amplitude of loading but also the frequency of loading. The pad becomes stiffer as the frequency of loading increases. Each CMP machine has its own vibration characteristics and the effect of machine vibration on pad response can be included in the model.

The second one is the mechanical behavior of fractals. Agglomerated particles have fractal geometry. It is not known how fractals respond to mechanical stress. This can be studied experimentally, which would further enhance the current model for cases close enough to DLA regime.

Finally, CMP is a multi-physics phenomenon representing interactions of both chemical and mechanical effects. In the present model, the mechanical effects are given primary importance, and chemical effect is incorporated via softening of the control layer at the wafer surface. However, it is noted ([57], [58]) that chemical dissolution also is affected by mechanical stresses and that can cause modulations on the surface profile. Work is currently in progress to incorporate such multi-physics effects.

BIBLIOGRAPHY

[1] Steigerwald, J. M., Murarka, S. P., Gutmann, R. J., (1997), "Chemical mechanical planarization of microelectronic materials," John Wiley & Sons Pub., New York.

[2] Moore, E. G., "Cramming More Components onto Integrated Circuits," Electronics Magazine, Vol. 38, No. 8, April 19, 1965.

[3] Gordon, E. M., Lithography and the Future of Moore's Law, Proc. SPIE Vol. 2437, May 1995.

[4] Oliver, M. R., CMP of Semi Conductor materials, Springer-Verlag, 2004, p 37.

[5] Lide, D. R., CRC Handbook of Chemistry and Physics, CRC Press, Boca Raton, FL, 79th, 1998-1999.

[6] Bastawros, A., Chandra, A., Guo, Y. J., and Yan, B., Pad effects on material-removal rate in chemical-mechanical planarization. Journal of Electronic Materials, Vol. 31, No. 10, pp. 1022–1031, 2002.

[7] Gouda, S., A quantitative analysis of multi-scale response of CMP pad: quasi-static and dynamic characterization of dry and wet pad, Masters Thesis, Iowa State University, 2004.

[8] Yu, H. H., and Suo, Z., Stress-dependent surface reactions and implicationsfor a stress measurement technique. Journal of Applied Physics,Vol 87, No. 3, 1 Feb 2000.

[9] Che, W., Bastawros, A., Chandra, A., Lonardo, P. M., Surface evolution during the chemical mechanical planarization of copper, CIRP Annals – Manufacturing Technology, Vol. 55, No. 1, pp: 605-608, 2006.

[10] Wang, C., A Stochastic Model for the Effects of Pad Surface Topography Evolution on Material Removal Rate Decay in Chemical Mechanical Planarization (CMP), Journal of the

IEEE Transaction on Semiconductor Manufacturing, Vol. 18, No. 4, Pages: 695-708, Nov 2005.

[11] Borucki, L., Mathematical modeling of polish-rate decay in chemical-mechanical polishing, Journal of Engineering Mathematics, Vol. 43, pp.105-114, 2002.

[12] Greenwood, J. W., Contact of nominally flat surfaces. Proceedings Royal Society of London Bulletin, A295, 1966.

[13] Guo, Y., Chandra, A., and Bastawros, A. F., Analytical dishing and step height model for CMP with a viscoelastic pad, IEEE Transactions On Semiconductor Manufacturing, Vol. 16, No. 3, August 2003.

[14] Preston, F. W., The theory and design of plate glass polishing machines, Journal of the Society of Glass Technology, Vol.11, 1927.

[15]. Sundararajan, S., Thakurta, D. G., Schwendeman, D. W., Murarka, S. P., and Gill, W. N., Journal of the Electrochemical Society, Vol. 146, p. 761, 1999.

[16] Elon J. T, and Higgs III, C. F., Hydrodynamics of Slurry Flow in Chemical Mechanical Polishing - A Review, Journal of The Electrochemical Society, Vol. 153, No. 6, pp K15 - K22 , 2006.

[17] Higgs III, C. F., Ng, S. H., Borucki, L., Yoon, I., and Danyluk, S., Journal of the Electrochemical Society, Vol. 152, G193, 2005.

[18] Nishioka, T., Sekine, K., and Tateyama, Y., in IEEE 1999 International Interconnect Technology Conference, San Francisco, CA, 1999.

[19]. Chen, J. M., and Fang, Y. C., IEEE Transactions on Semiconductor Manufacturing, Vol. 15, No. 39, 2002.

[20] Zhang, F., Busnaina, A., and Ahmadi, G., Particle adhesion and removal in chemical mechanical polishing and post-CMP cleaning, Journal of Electrochemical Society, Vol.146, pp.2665-2669, 1999.

[21] Che, W., A Scratch Intersection Model of material removal during CMP, Journal of Manufacturing Science and Engineering, Aug 2005, Vol. 127, p. 545.

[22] Che, W., Mechanistic Understanding of Material Detachment During Micro-Scale Polishing, Journal of Manufacturing Science and Engineering, Vol. 125, p 731, Nov, 2003.

[23] Fu, G., Chandra,. A., A Plasticity-Based Model of Material Removal in Chemical–Mechanical Polishing (CMP), IEEE Transactions on Semiconductor Manufacturing, Vol. 14, No. 4, pp: 406, November 2001.

[24] Wang, C., Sherman, P., Chandra, A., Pad surface roughness and slurry particle size distribution effects on material removal rate in chemical mechanical planarization, CIRP Annals - Manufacturing Technology, Vol. 54, No. 1, Pages 309-312, 2005.

[25] Luo, J., and Dornfeld, D., Material Removal Mechanism in Chemical Mechanical Polishing: Theory and Modeling, IEEE Transaction on Semiconductor Manufacturing, Vol.14, pp.112-133, May 2001.

[26] Bastaninejad, M., and Ahmadi, G., Modeling the Effects of Abrasive Size Distribution, Adhesion, and Surface Plastic Deformation on Chemical-Mechanical Polishing, Journal of The Electrochemical Society, Vol. 152, No. 9, G720-G730, 2005.

[27] Han, Y.Y., Scratch generation probability in chemical mechanical polishing (CMP) process, Masters Thesis, Iowa State University, 2006.

[28] Runze Chang and Costas J. Spanos, Full profile inter-layer dielectric CMP analysis. ISM, 2001.

[29] Bastawros et al, Pad Effects on Material-Removal Rate in Chemical-Mechanical, Planarization, Journal of Electronic Materials Vol. 31, No. 10, 2002.

[30] Sunghoon Lee, Hyoungjae Kim, David Dornfeld, Development of a CMP pad with controlled micro features for improved performance, IEEE International Symposium on Semiconductor Manufacturing, 2005.

[31] K. L. Johnson, Contact Mechanics, Cambridge University Press, 2008.

[32] J.F Archard, Contact and rubbing of flat surfaces. Journal of Applied Physics, Vol. 24, pp.981-985, 1953.

[33] Chandra, A., Dorothy, M., Karra, P., Multi scale modeling, Unpublished work, 2008.

[34] Metal Cutting Principles, Milton C. Shaw, Oxford University Press, USA; 2nd edition July, 2004.

[35] Lin, M. Y., Lindsay, H. M., Weitz, D. A., Ball, R. C., Klein, R., Meakin, P., Universality of fractal aggregates as probed by light scattering, Proceedings of the Royal Society, London, A, Vol. 423**,** pp. 71-87, 1989.

[36] Chandra A, Karra P, Bastawros A F , Biswas R, Sherman P J , Armini S , Lucca DA, Prediction of scratch generation in chemical mechanical planarization, CIRP Annals – Manufacturing Technology, Vol. 57, No. 1, pp: 559-562**,** 2008.

[37] Bachelier, L., "Théorie de la spéculation", Annales Scientifiques de l'École Normale Supérieure, Vol 17, No. 3, pp. 21-86, 1900.

[38] Einstein, A., On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat, Annalen der Physik, May 1905.

[39] Smoluchowski, M., Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen, Annalen der Physik, Vol. 21, pp.756-780, 1906, http://www.physik.uni-augsburg.de/annalen/history/historic-papers/1906_326_756-780.pdf.

[40] Gun'ko, V. M., et al., Aqueous suspension of fumed oxides: Particle size distribution and zeta potential, Advances in Colloid and Interface Science, Vol. 91, pp.1-112, 2001.

[41] Kovacˇevic´ . D., Kallay, N., Antol, I., Pohlmeier, A., Lewandowski, H., Narres, H. D., The use of electrokinetic potential in the interpretation of adsorption phenomena: Adsorption of salicylic acid on hematite, Colloids and Surfaces A: Physicochemical and Engineering Aspects Vol. 140, No. 2, Pages 261-267, 30 September 1998.

[42] Adamson, A. W., Gast, A. P., Physical Chemistry of Surfaces, Wiley-Interscience Inc, Sixth Edition, p172, 1997.

[43] Lin, M. Y., Lindsay, H. M., Weitz, D. A., Ball, R. C., Klein, R., Meakin, P., Universal diffusion-limited colloid aggregation, Journal of Physics: Condensed Matter, Vol. 2, pp. 3093-3113, 1990.

[44] Casella, G., Berger, R. L., Statistical Inference, Duxbury Press; 2nd Edition, June 18, 2001.

[45] Kirby, B. J., and Hasselbrink Jr., E. F., Zeta potential of micro fluidic substrates, Electrophoresis, Vol. 25, pp.187–202, 2004.

[46] Scales, P. J., Grieser, F., Healy, T. W., Electrokinetics of the silica solution interface - a flat-plate streaming potential study, Langmuir, Vol. 8, pp. 965–974, 1992.

[47] Kosmulski, M., Matijevic, E., Zeta-potentials of silica in water alcohol mixtures, Langmuir, Vol. 8, pp. 1060–1064, 1992.

[48] Zeta Potential of Colloids in Water and Waste Water, ASTM Standard D 4187-82, American Society for Testing and Materials, 1985.

[49] Cohen, R. J., and Benedek, G. B., Journal of Chemical Physics, Vol. 86, pp. 3696, 1982.

[50] Meakin, P., Vicsek, T. and Family F, Physical Review B, Vol. 31, pp. 564, 1985.

[51] Blum, J., Dust agglomeration, Advances in Physics, Vol. 55, No. 7–8, pp. 881–947, 2006.

[52] Chandra, A., Zunke, R., Karra, P., Modeling of Material Removal in Polishing of Advanced Ceramics, 12th CIRP Conference on Modeling of Machining Operations, May 2009.

[53] Mandelbrot, B. B., The Fractal Geometry of Nature, New York, W. H. Freeman and Co., 1982.

[54] Brown, N. J., Some speculations on the mechanisms of abrasive grinding and polishing, Precision Engineering, Vol 9, No 3, pp. 129-138, July 1987.

[55] http://www.codeproject.com/KB/recipes/fractalssse.aspx

[56] Suphantharida, P., and Osseo-Asare, K., Cerium Oxide Slurries in CMP. Electrophoretic Mobility and Adsorption Investigations of Ceria/Silicate Interaction, Journal of the Electrochemical Society, Vol. 151, No. 10, pp. G658-G662, 2004.

[57] Kim KS, Hurtado JA, Tan H, Evolution of a surface-roughness spectrum caused by stress in nanometer-scale chemical etching, Physical Review Letters, Vol. 83, No. 19, pp. 3872-3875, Nov 1999.

[58] Yu, H., H., Suo, Z., Stress-dependent surface reactions and implications for a stress measurement technique, Journal of Applied Physics, Vol. 87, No. 3, pp.1211-1218, Feb 2000.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those who helped me with various aspects of research and the writing of this thesis. First and foremost, would like to thank Professor Abhijit Chandra for his guidance, patience and support throughout this research. I do not recollect one instance where I did not feel motivated and encouraged after a meeting with Professor Chandra. His insights have often helped me solve the problems I was facing in the research. Also, I would like to thank Professor Bastawros who helped in research. I would like thank Professor Atul Kelkar for his support during initial stages of my program. I would like to express my gratitude to Professor Elgin Johnston Professor Thomas Rudolphi and Professor Baskar Ganapathysubramanian for their efforts and contributions to this work as committee members. I would like to thank Dr. Elgin Johnston for his inspirational teaching of Real Analysis.

I am grateful to the department of Mechanical Engineering for the financial support throughout the program.