CONTACT MECHANICS OF MULTILAYERED ROUGH SURFACES IN TRIBOLOGY

DISsertation

Presented in Partial fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Wei Peng, M. S.

*****

The Ohio State University 2001

Dissertation Committee:

Professor Bharat Bhushan, Advisor

Professor Bernard J. Hamrock

Professor Shoichiro Nakamura

Approved by

Advisor

Department of Mechanical Engineering
ABSTRACT

The deposition of layers is an effective way to improve the tribological performance of rough surfaces. The contact mechanics of layered rough surfaces needs to be studied to optimize layer parameters. Since 1995 a lot of progress has been made in the development of numerical contact models, which analyze the contact behavior of layered rough surfaces with no assumption concerning the roughness distribution as well as the effect of interfacial liquid film on the contact statistics. Based on the formulation of contact problems, these models are classified into three categories: direct formulation, weighted residual formulation, and minimum total potential energy formulation. The numerical methods applied in these models include Finite Difference Method (FDM), Finite Element Method (FEM), and Boundary Element Method (BEM). A 3-D BEM model based on a variational principle is developed for its capability to analyze the layered rough surfaces contact involving a large number of contact points. This model predicts contact pressure profile on the interface and contact statistics, namely fractional contact area, the maximum value of contact pressure, von Mises and principal tensile stresses, and relative meniscus force. The results allow the specification of layer properties to reduce friction, stiction, and wear of layered rough surfaces. Typical examples of layered rough surfaces contact simulated by this model are presented.
The examples contain data for various surface topographies, elastic and elastic-plastic material properties, normal and tangential loading conditions, and dry and wet interfaces. Applications of this model to the magnetic storage devices and MicroElectroMechanical Systems (MEMS) are presented.

Keywords: Contact mechanics, Layer, Rough surface, Friction, Wear
Dedicated to my parents
ACKNOWLEDGMENTS

I wish to thank my advisor, Professor Bharat Bhushan, for intellectual support, encouragement, and enthusiasm which made this thesis possible, and for his patience in correcting both my stylistic and scientific errors. I genuinely enjoyed the time spent with him at OSU and thank him for allowing me to work with him in such an outstanding laboratory. I thank Professor Bernard Hamrock and Professor Shoichiro Nakamura for serving on my candidacy exam and dissertation committee; Professor Si Lee and Professor Anthony Luscher for serving on my candidacy exam committee. They gave valuable advice as members of my candidacy exam committee.

Colleagues and friends in the laboratory offered many discussions that were helpful for this work. I’d also like to thank members of the lab who have offered their assistance. I thank Dr. Xiaodong Li, Dr. Huiwen Liu, Dr. Tiejun Ma, Dr. Zheming Zhao, Dr. M. S. Bobji, Dr. Sumit Chandra, Dr. Sameera Chilamakuri, Dr. William Scott, Sriram Sundararajan, Anton Goldade, Prasad Mokashi, Weihua Sheng, Micah Kattner, and other friends.

I would also like to thank all the reviewer’s of my papers submitted to ASME Journal of Tribology, Wear, Journal of Information Storage Processing Systems, Journal of Engineering Tribology for their valuable comments.
The financial support for this research was provided by the National Science Foundation (Contract No. ECS-9820022) and the membership of the Nanotribology Laboratory for Information Storage and MEMS / NEMS. The content of this information does not necessarily reflect the position or policy of the Government and no official endorsement should be inferred.
VITA

November 18, 1972 …..Born, Hubei, China

1993 …………………..B.S. Mechanical Engineering, Xi’an Jiao Tong University

1996 …………………..M.S. Mechanical Engineering, Shanghai Jiao Tong University

1997 – present …………Graduate Research Associate, The Ohio State University

PUBLICATIONS

Research Publication


FIELDS OF STUDY

Major Field: Mechanical Engineering
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>VITA</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td>NOMENCLATURE</td>
<td>xv</td>
</tr>
<tr>
<td>CHAPTERS:</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 SINGLE ASPERITY CONTACT OF LAYERED SURFACES</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Analytical Methods Based on Weighted Residual Formulation</td>
<td>11</td>
</tr>
<tr>
<td>2.1.1 Least-square error method</td>
<td>13</td>
</tr>
<tr>
<td>2.1.1.1 Quasi-sliding contact of a smooth sphere on a layered half-space</td>
<td>14</td>
</tr>
<tr>
<td>2.2 BEM Numerical Methods Based on Direct formulation</td>
<td>17</td>
</tr>
<tr>
<td>2.2.1 Matrix inversion method</td>
<td>20</td>
</tr>
<tr>
<td>2.2.1.1 Normal contact of a smooth cylinder on a layered half-space</td>
<td>20</td>
</tr>
<tr>
<td>2.3 FEM Numerical Methods</td>
<td>21</td>
</tr>
<tr>
<td>2.3.1 Finite element modeling</td>
<td>23</td>
</tr>
<tr>
<td>2.3.1.1 Sliding contact of a smooth sphere on a layered half-space</td>
<td>25</td>
</tr>
<tr>
<td>3 MULTIPLE ASPERITY CONTACT OF LAYERED ROUGH SURFACES</td>
<td>29</td>
</tr>
<tr>
<td>3.1 BEM Numerical Methods Based on Direct Formulation</td>
<td>33</td>
</tr>
<tr>
<td>3.1.1 Matrix inversion method</td>
<td>34</td>
</tr>
<tr>
<td>3.1.1.1 Quasi-sliding contact of a rough cylinder on a layered half-space</td>
<td>36</td>
</tr>
<tr>
<td>3.1.2 Conjugate gradient method</td>
<td>40</td>
</tr>
<tr>
<td>3.1.2.1 Normal contact of a rough sphere on a layered half-space</td>
<td>43</td>
</tr>
</tbody>
</table>
3.2 BEM Numerical Methods Based on Minimum Total Potential Energy

3.2.1 Influence coefficients ................................................................. 56
3.2.2 Elastic / perfectly-plastic model .................................................. 67
3.2.3 Minimization algorithms ............................................................ 70
3.2.4 Meniscus effect ......................................................................... 73
3.2.5 Program flowchart ................................................................. 75
3.2.6 Quasi-sliding contact of two layered nominally flat rough surfaces ............ 79
  3.2.6.1 Normalization scheme ....................................................... 81
  3.2.6.2 Layer effect ........................................................................ 92
  3.2.6.3 Friction effect ................................................................. 98

4 APPLICATIONS ................................................................................. 101
4.1 Magnetic Storage Devices ........................................................... 101
  4.1.1 Head slider / disk interface (HDI) modeling ................................. 102
  4.1.2 Calculation of critical thickness ................................................... 105
  4.1.3 Results and discussion ............................................................... 106
4.2 MicroElectroMechanical System ..................................................... 109

5 CONCLUSIONS .................................................................................. 111

BIBLIOGRAPHY ..................................................................................... 115

APPENDIXES:
  A. DEFINITION OF THE TERMS \( R_1, R_2, R_3, R_4, R_5, R_6, R_a, R_b, R_c \) AND \( R_d \) ............ 122
  B. INFLUENCE COEFFICIENTS FOR A 3-D LAYERED SOLID IN NORMAL CONTACT ................................................. 123
  C. CALCULATION OF THE EFFECTIVE HARDNESS OF A LAYERED SOLID ............................................................ 125
  D. CALCULATION OF THE COMPOSITE ROUGH SURFACE ................................................................. 127
  E. ROUGH SURFACE GENERATION ................................................................. 130
  F. DRY CONTACT ................................................................................. 148
  G. WET CONTACT ................................................................................. 210
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Variation of fractional contact area, maximum pressure with $\sigma$ and $p_n/E_2$ at various values of $E_1/E_2$. Here $\beta^* = 0.5 \mu m$, $E_2 = 100$ GPa, $H_s/E_1 = 0.05$, $h = 1 \mu m$.</td>
<td>90</td>
</tr>
<tr>
<td>4.1</td>
<td>Typical physical properties and surface topography statistics of magnetic thin-film disks and sliders</td>
<td>105</td>
</tr>
<tr>
<td>4.2</td>
<td>Typical physical properties and surface topography statistics of MEMS materials</td>
<td>110</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Various methods classified into several categories. Among them the analytical weighted residual formulation applies exclusively to single asperity contact, and the numerical direct formulation and minimum total potential energy formulation apply to both single asperity contact and multiple asperity contact.</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Schematic of a smooth cylinder in contact with a layered half-space</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Schematic of a smooth sphere in contact with a layered half-space</td>
<td>9</td>
</tr>
<tr>
<td>2.3</td>
<td>Profile of contact pressures beneath sphere at various $E_1/E_2$ [O’Sullivan and King, 1988, Peng and Bhushan, 2001b]</td>
<td>15</td>
</tr>
<tr>
<td>2.4</td>
<td>Stress components along z-axis at various $E_1/E_2$ and coefficient of friction $\mu$ ($\sigma_{xz}$ is plotted along the line $x/a_0 = 0.5$, $y = 0$) [Peng and Bhushan, 2001b]</td>
<td>16</td>
</tr>
<tr>
<td>2.5</td>
<td>Contours of $\sqrt{J_2/p_0}$ at $y = 0$ at various $E_1/E_2$ and coefficient of friction $\mu$ [Peng and Bhushan, 2001b]</td>
<td>18</td>
</tr>
<tr>
<td>2.6</td>
<td>Variation of the maximum contact pressure $p_{max}$ and half contact width $a$ with layer thickness $h$ at various $\gamma = E_1/E_2$ [Cole and Sayles, 1992]</td>
<td>22</td>
</tr>
<tr>
<td>2.7</td>
<td>3-D finite element mesh of the layered half-space in the first octant indented by a rigid sphere centered along z-axis, the inset is the magnified finer meshed region at the contact interface. [Kral and Komvopoulos, 1997]</td>
<td>24</td>
</tr>
<tr>
<td>2.8</td>
<td>Contours of $\varepsilon'<em>{eq}$ and $\sigma</em>{xx}$ in the quadrant ($y &gt; 0$, $z &gt; 0$) at sliding distance $\Delta x$ equal to (a) 2$a_y$, (b) 4$a_y$, (c) 8$a_y$, and (d) 12$a_y$, with $E_1/E_2 = H_1/H_2 = 2$, $\mu = 0.1$, $W/W_y = 100$. $a_y$ and $W_y$ represent the sliding distance and normal loading corresponding to the initial yield condition of the substrate, respectively. [Kral and Komvopoulos, 1997]</td>
<td>27</td>
</tr>
</tbody>
</table>
3.1 Profiles of (a) the nominally flat rough surface being superposed on the smooth cylinder and (b) corresponding contact pressures. The smooth curve is contact pressure profile of the smooth cylinder in Hertzian contact. [Mao et al., 1996] .................................................................37

3.2 Profiles of contact pressures at various layer thickness $h$ and coefficients of friction $\mu$ with (a) $E_1/E_2 = 3.2$ and (b) $E_1/E_2 = 0.17$. [Mao et al., 1996] ......38

3.3 Profiles of a rough sphere with $R = 0.0137 \mu m$, and corresponding contact pressures and z-direction surface displacements at the contact interface [Nogi and Kato, 1996].............................................................................................44

3.4 Contours of von Mises stresses (in GPa) on x-z plane through the center of the rough sphere in (a) a homogenous half-space and (b) a layered half-space [Nogi and Kato, 1996].................................................................45

3.5 Schematics of (a) a rough surface in contact with a layer rough surface and (b) top view of contact regions.................................................................................................................48

3.6 (a) Definition of strain energy and complementary energy, (b) relationship between elastic strain energy and internal complementary energy for a linear elastic or a linear elastic-perfectly plastic material......................................................50

3.7 Schematics of surface discretization at the contact interface, (a) 3-D view in space domain, (b) top view in space domain, and (c) top view in frequency domain.................................................................................................................52

3.8 (a) Geometrical interference area and real contact area in the normal contact of two identical spheres, and (b) determination of the total prescribed z-direction surface displacement from geometrical interference ........................................56

3.9 Variation of $C^\mu_z(0,0,0)$, the influence coefficient corresponding to z-direction surface displacement at the origin, with layer thickness $h$ at various $E_1/E_2$, for a homogeneous (Love, 1929) and a layered elastic half-space ..................66

3.10 Profiles of von Mises stresses (in pressure unit) on the surface and in the subsurface ($x = 0.3$ unit in length, $y = 0$) at various $E_1/E_2$ and coefficient of friction $\mu$ [Peng and Bhushan, 2002].................................................................................................68

xii
3.11 Schematic of a smooth surface in contact with a composite rough surface in the presence of a liquid film

3.12 Flowchart of a 3-D layered rough surfaces contact model based on a variational principle and FFT

3.13 Profiles of two computer generated rough surfaces. The lower surface ($\sigma = 2$ nm, $\beta^* = 1$ $\mu$m) is magnified twice and cut off at the $20 \times 20$ $\mu$m$^2$ margin to obtain the upper surface ($\sigma = 1$ nm, $\beta^* = 0.5$ $\mu$m)

3.14 Variation of $A_r/A_n$, $p_{max}/E_2$, and $F_m/W$ with $p_n/E_2$ at various $E_1/E_2$ [Peng and Bhushan, 2001a]

3.15 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_t$ plane and shear stresses on the max $\sigma_{xz}$ plane at various $E_1/E_2$, with $\sigma = 1$ nm, $\beta^* = 0.5$ $\mu$m, $p_n/E_2 = 4 \times 10^{-6}$, $h = 1$ $\mu$m. All contours are plotted after taking natural log values of the calculated stresses (in kPa). Negative values of $\sigma_t$ and $\sigma_{xz}$ in the plot represent the compressive stress and shear stress along negative x direction, respectively. [Peng and Bhushan, 2002]

3.16 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_t$ plane and shear stresses on the max $\sigma_{xz}$ plane at various $E_1/E_2$, with $\sigma = 0.1$ nm, $\beta^* = 0.5$ $\mu$m, $p_n/E_2 = 4 \times 10^{-7}$, $h = 1$ $\mu$m [Peng and Bhushan, 2002]

3.17 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_t$ plane and shear stresses on the max $\sigma_{xz}$ plane at various $E_1/E_2$, with $\sigma = 2$ nm, $\beta^* = 1$ $\mu$m, $p_n/E_2 = 4 \times 10^{-6}$, $h = 2$ $\mu$m [Peng and Bhushan, 2002]

3.18 Variation of $A_r/A_n$ with $\frac{p_n}{E_2} \frac{\beta^*}{\sigma}$ and $p_{max}/E_2$ with $\left[\frac{p_n}{E_2} \frac{\sigma}{\beta^*}\right]^{1/3}$ at various $E_1/E_2$. These values are independent of $\beta^*$ and $\mu$. [Peng and Bhushan, 2002]
3.19 Profiles of (a) a composite layered rough surface and (b) corresponding contact pressures at various $h$ and $E_1/E_2$. These values are independent of $\mu$ .........93

3.20 Variation of $A_r/A_n$, $p_{\text{max}}/E_2$, and $F_m/W$ with layer thickness $h$ at various $E_1/E_2$. These values are independent of $\mu$. .................................................................94

3.21 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_r$ plane and shear stresses on the max $\sigma_{xz}$ plane at various layer thickness $h$ with (a) $E_1/E_2 = 2$ and (b) $E_1/E_2 = 0$, $\mu = 0$. [Peng and Bhushan, 2001b] ....95

3.22 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_r$ plane and shear stresses on the max $\sigma_{xz}$ plane at various layer thickness $h$ with (a) $E_1/E_2 = 2$ and (b) $E_1/E_2 = 0.5$, $\mu = 0.5$. [Peng and Bhushan, 2002] ........................................................................................................99

4.1 Cross sectional schematic of a typical thin-film magnetic disk .................102

4.2 Variation of $\sigma$ with scan size for two glass ceramic substrates (GC1 and GC2) and corresponding finished magnetic disks (GC1f and GC2f), Ni-P coated Al-Mg substrate (Ni-P), and glass substrate (Gl) measured by atomic force microscope (AFM), stylus profiler (SP) and non-contact optical profiler (NOP) [Peng and Bhushan, 2000b] .................................................................................................103

4.3 Variation of $A_r/A_n$, $u_c$, $p_{\text{max}}/E_2$, and $F_m/W$ with layer thickness $h$ at various $E_1/E_2$ and (a) $\sigma = 0.1$ nm, (b) $\sigma = 1$ nm, and (c) $\sigma = 10$ nm, in the normal contact of a rigid head-slider on an elastic-perfectly plastic disk in the presence of a liquid film [Peng and Bhushan, 2000b] ........................................107

C.1 Schematics of (a) two identical half circles in contact and (b) a flat surface in contact with a composite half ellipse ..........................................................128
NOMENCLATURE

\( a \)  
Half contact width (2D) or contact radius (3D)

\( a_0 \)  
Hertzian contact radius for homogenous solids

\( A_n, A_r \)  
Nominal and real area of contact

\( C^u, C^\sigma \)  
Influence coefficients corresponding to the displacements and the stresses

\( E_0, E_1, E_2 \)  
Young’s moduli of the indenter, layer and substrate

\( f_z(x, y) \)  
Initial separation of two contact surfaces at point \((x, y)\)

\( F_m \)  
Meniscus force

\( G_1, G_2 \)  
Shear moduli of the layer and substrate

\( h \)  
Layer thickness

\( h_f, h_m \)  
Mean liquid thickness and mean meniscus height

\( H_0, H_1, H_2 \)  
Hardness of the indenter, layer and substrate

\( H_s \)  
Hardness of the softer material in contact

\( H_e \)  
Effective hardness of the layered solid

\( k \)  
Yield stress in pure shear

\( p, p_r, p_{\text{max}} \)  
Contact pressure, average contact pressure, and maximum contact pressure

\( p_0 \)  
Hertzian maximum pressure for homogenous solids

\( p_n \)  
Nominal pressure

\( R \)  
Radius of cylinder (2D) or sphere (3D)
\( u_{z0}, u_{z1}, u_z \)  
z-direction displacements of indenter, layer and the total

\( u_{z0}^*, u_{z1}^*, u_z^* \)  
Prescribed z-direction displacements of indenter, layer and the total

\( U_E, U_E^*, V^* \)  
Energy of elastic strain, internal complementary, and total complementary

\( w(r) \)  
Undeformed profile of the indenter

\( W \)  
Applied normal load

\( x_{\text{max}}, y_{\text{max}}, z_{\text{max}} \)  
Location of maximum von Mises stress

\( x'_{\text{max}}, y'_{\text{max}}, z'_{\text{max}} \)  
Location of maximum principal tensile stress

\( x''_{\text{max}}, y''_{\text{max}}, z''_{\text{max}} \)  
Location of maximum shear stress

\( Y \)  
Yield stress in simple tension

\( \beta_x^*, \beta_y^* \)  
Composite correlation length along x axis and y axis for anisotropic roughness

\( \beta^* \)  
Composite correlation length for isotropic roughness

\( \delta \)  
Relative vertical rigid approach

\( \Delta \)  
Patch width

\( \phi, \psi_1, \psi_2, \psi_3 \)  
Papkovich-Neuber elastic potentials (3D harmonic functions)

\( \mu \)  
Coefficient of friction

\( \nu \)  
Poisson’s ratio, set as 0.3 for all cases

\( \Omega \)  
Prescribed (real) contact domain

\( \sigma \)  
Composite standard deviation of the surface heights

\( \sigma_t \)  
Principal tensile stress

\( \sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{xz}, \sigma_{yz} \)  
Stress components

\( \sqrt{J_2} \)  
von Mises stress
Solid surfaces, irrespective of their method of formation, contain irregularities, or deviations from a prescribed geometric form. The high points on the surfaces are referred to as asperities, peaks, summits, or hills and the low points are referred to as valleys. When two surfaces are placed in contact, surface roughness causes contact to occur at discrete contact spots (junctions), i.e., a multiple asperity contact. Deformation occurs in the region of the contact spots, establishing stresses that oppose the applied load. The sum of the areas of all the contact spots constitutes the real (true) area of contact, and for most materials with applied load, this is only a small fraction of the nominal (apparent) area of contact (that which would occur if the surfaces were perfectly smooth). The real area of contact is a function of the surface topography, material properties and interfacial loading conditions. The proximity of the asperities results in adhesive contacts caused by interatomic interactions. When two surfaces move relative to each other, the adhesion of these asperities and other sources of surface interactions contribute to friction force. Repeated surface interactions and surface and subsurface stresses, developed at the interface, results in the formation of wear particles and eventual failure (Bowden and Tabor, 1950, 1964; Bhushan, 1996a, 1999a, b, 2001). The deposition of thin layers,
ranging in thickness from a couple of nanometers to a few microns, has been proved to be an effective way to improve the tribological performances of contacting solids (Bhushan, 1996a, 2001; Bhushan and Gupta, 1991). For example, without having to change the bulk material, selected layers can reduce the friction, stiction, wear, and failure. Examples are the multilayered construction of magnetic heads and disks in magnetic rigid disk drives, and the use of solid lubricating films such as diamond-like carbon (DLC) films on Si substrate in MEMS devices. Layers applied in these devices are as thin as 1 nm. The tribological performances of these layered solids directly depend on the layer properties, e.g., the layer thickness, surface roughness, coefficient of friction, stiffness- and hardness- ratio of the layer and the substrate. In order to obtain an optimal layer, e.g., a layer possessing a longer life as well as lower friction, it is necessary to investigate friction and wear mechanisms of layered rough surfaces contact through either theoretical or empirical analyses. Among them the contact modeling of layered rough surfaces is widely applied (Bhushan, 1998, 1999a).

Statistical contact models are generally applied to provide main trends for surfaces with a given roughness distribution. These models are highly conceptual and thus a geometrically complicated surface is usually represented by a few statistical parameters, which is readily solved either by analytical or numerical method. The classical statistical model for a combination of elastic and plastic contacts between rough surfaces is that of Greenwood and Williamson (1966). Surfaces are assumed composed of hemispherically tipped asperities of uniform radius of curvature, with their heights following a Gaussian distribution about a mean plane. For a review of this model and its modifications see Bhushan (1998, 1999a). Although the statistical models result in simple
relationships and are able to predict important trends in the effect of surface properties on the real area of contact, their usefulness is very limited because of over-simplifying assumptions about asperity geometry and height distributions, the difficulty in determination of statistical roughness parameters, and the neglecting of interactions between adjacent asperities.

Surfaces usually exhibit multiscale nature. A unique property of rough surfaces is that if a surface is repeated magnified, increasing details of roughness are observed right down to the nanoscale. In addition, the roughness at all magnifications appears quite similar in structure. Such behavior can be characterized by fractal geometry (Majumdar and Bhushan, 1990; Bhushan, 1999b). The main conclusions from these studies are that a fractal characterization of surface roughness is scale-independent and provides information of roughness structure at all the length scales that exhibit the fractal behavior. Based on this observation, Majumdar and Bhushan (1991) and Bhushan and Majumdar (1992) developed a new fractal theory of contact between two rough surfaces. They have reported relationships for cumulative size distribution of the contact spots, portions of the real area of contact in elastic and plastic deformation modes, and the load-area relationships. However, the usefulness is limited since most engineering surfaces can not be described by a simple fractal dimension.

Since rough surfaces are normally too complex to be generalized, most recent models take the profiles of rough surfaces as the model input directly, rather than just a few statistics. The rough surfaces are either measured real surfaces or computer generated surfaces. The surface height distributions of these surfaces either follow a single modal distribution, Gaussian or non-Gaussian (Chilamakuri and Bhushan, 1998), or follow
multi-modal distributions (Peng and Bhushan, 2000a). These models start from calculating the stresses and displacements for a minimum set of distinct simple distributed loadings using theory of elasticity. Then the stresses and displacements for a real situation are obtained by proper assembling of solutions for simple loading conditions. The unknown assembling factors are obtained by enforcing the boundary and initial conditions. The number of distinct loadings in the minimum set may become very large with a geometrically complicated surface, which makes analytical solution impossible. This situation leads to approximate numerical analyses, which produce sufficiently precise solutions.

Mathematically the approximate numerical solutions are obtained by essentially solving the governing differential equations numerically. Physically the complicated object is divided (discretized) into small and manageable pieces. Depending on the way the complicated object is discretized, numerical methods are classified into Finite Difference Method (FDM), Finite Element Method (FEM), and Boundary Element Method (BEM) (Fig. 1.1). FDM is limited to regular domain thus is unfeasible for the rough surfaces contact. FEM is more flexible and theoretically is capable to solve rough surfaces contact problems. However, in practice a 3-D layered rough surfaces contact generally requires a large number of fine mesh elements to describe the irregular domain. The situation is especially severe in FEM since the entire volume of the contacting solids is divided into discrete finite elements. It makes commercial FEM packages unfeasible due to the requirement of extremely long computation time. Unlike FEM, in BEM only the contact interface (boundary) requires subdivision, thereby reducing the size of the
Figure 1.1 Various methods classified into several categories. Among them the analytical weighted residual formulation applies exclusively to single asperity contact, and the numerical direct formulation and minimum total potential energy formulation apply to both single asperity contact and multiple asperity contact.

problem and thus dramatically reducing the effort involved in obtaining a solution. BEM is widely used in continuum mechanics, especially in 3D contact problems. The contact interface of the layered rough surfaces is discretized into patches sufficiently small that the shape affect of the asperity is negligible. Therefore no assumption of surface isotropy,
asperity shape, or distribution of asperity heights, slopes and curvatures is required. In order to take the interactions between adjacent asperities into account, the stresses and displacements induced by each asperity in the contact domain is calculated first and assembled together in such a way that the equilibrium is satisfied under the restriction of the boundary, interface, and initial contact conditions.

A lot of progress has been made since 1995 in the development of BEM numerical methods with the advent of computer technology. Various criteria are applied to formulate these BEM models: the weighted residual formulation, the direct formulation, and the minimum total potential energy formulation. Based on the weighted residual formulation, the least-square-error method has been applied to solve layered single asperity contact problems. It generally requires closed form description of contact pressure profile, which can not be obtained for geometrically complicated interface thus not feasible for multiple asperity contact. Based on the direct formulation, matrix inversion method and Conjugate Gradient Method (CGM) have been applied to solve layered multiple asperity contact problems with a moderate number of contact points. In case of large number of contact points, the so-called influence coefficients matrix becomes very large and possibly ill conditioned due to large round off error and these two methods leads to convergence problem. Moreover, the uniqueness of the solution is not guaranteed. In contrast, the minimum energy formulation theoretically guarantees the convergence and the uniqueness of the solution. Based on the minimum energy formulation, the quasi-Newton method has been applied to solve layered multiple asperity contact problems.
In the presence of a thin liquid film with a small contact angle (wetting characteristics), such as a lubricant or an absorbed water layer at the contact interface, curved menisci form around contacting and non-contacting asperities. The attractive meniscus force arises from the negative Laplace pressure inside the curved meniscus as a result of surface tension. This intrinsic attractive force may result in high static friction (stiction), kinetic friction and wear (Bhushan, 1999a). This problem is particularly important in the computer storage industry and MEMS (Bhushan, 1996a, 1999b). Several statistical models have been developed to predict meniscus forces developed at a wet interface. A numerical model based on real rough surface profiles also has been developed to account the effect of surface roughness, thin liquid film, environment and operating conditions on meniscus forces. This numerical model has been adopted to analyze the contact of layered rough surfaces with the presence of a thin liquid film.

It is evident that substantial on-going contributions, particularly since 1995, are being made to the numerical modeling of multilayered rough surfaces contact. This perspective paper presents typical examples of layered surface contact problem, which covers contacts with various surface topographies, elastic and elastic-plastic materials, normal loading and tangential loading, and dry and wet interfaces. Emphases are on the implications and results of these contact models. This paper starts out with a short review of modeling of single asperity layered surface contact, followed by a comprehensive review of the modeling of multiple asperity layered rough surfaces contact in both dry and wet conditions. Examples of applications are shown in the field of the magnetic storage devices and MicroElectroMechanical Systems (MEMS).
CHAPTER 2

SINGLE ASPERITY CONTACT OF LAYERED SURFACES

Single asperity contact is the basic component of multiple asperity contact, therefore is of interest to better understand multiple asperity contact. It is much simpler than multiple asperity contact and usually has a unique solution, which makes it an ideal solution validation case for models developed for multiple asperity contact. A comprehensive review of modeling of both homogenous and layered single asperity contact has been presented by Bhushan (1996b). Here we review the models for layered single asperity contact.

The asperity is usually assumed to be cylindrical (2-D) or spherical (3-D). Figure 2.1 shows a schematic of an elastic cylinder \( E_0 \) of radius \( R \) in contact with a layered elastic half-space. The real contact domain is represented by the half contact width \( a \). Figure 2.2 shows a schematic of an elastic sphere \( E_0 \) of radius \( R \) in contact with a layered elastic half-space. The real contact domain is represented by the contact radius \( a \). Generally in the initial contact condition, \( a \) is prescribed and \( \delta \) is the corresponding relative rigid approach (unknown). Geometrical overlap is not allowed and the adhesive forces are assumed negligible, i.e., no normal tension is allowed and pressures must be
Figure 2.1  Schematic of a smooth cylinder in contact with a layered half-space

Figure 2.2  Schematic of a smooth sphere in contact with a layered half-space
positive or zero everywhere. A single asperity contact problem is then addressed in polar
coordination system as

\[
\begin{align*}
  u_{z0}(r,0) + u_{z1}(r,0) &= u_{z0}^*(r,0) + u_{z1}^*(r,0), \quad |r| \leq a \\
  u_{z0}(r,0) + u_{z1}(r,0) &> u_{z0}^*(r,0) + u_{z1}^*(r,0), \quad |r| > a \\
p(r) &\geq 0, \quad |r| \leq a \\
p(r) &= 0, \quad |r| > a
\end{align*}
\]

(2.1)

where \( p \) is the contact pressure, \( r \) is the distance from the center of the contact zone, \( u_{z0} \)
and \( u_{z1}^* \) are the calculated and prescribed z-direction surface displacements of the
cylinder / sphere and the layered half-space over the real contact domain (identified by
indices 0 and 1), respectively. \( u_{z0}(r,0) \) and \( p(r) \) must also satisfy the governing
formulation of elasticity in terms of surface displacements and contact pressures. Since
the sum of prescribed z-direction surface displacements \( u_{z0}^*(r,0) + u_{z1}^*(r,0) \) in the real
contact domain is a function of \( R \) and \( a \) and thus known, so are \( u_{z0}(r,0) + u_{z1}(r,0) \). The
problem then becomes to find the contact pressure profile \( p(r) \) satisfying the governing
formulation of elasticity with given \( u_{z0}(r,0) \).

Chen and Engel (1972) further expressed \( u_{z0}^*(r,0) + u_{z1}^*(r,0) \) as \( w(r) \), the
undeformed surface profile of the indenter in reference to a cutting plane set at \( \delta \) below
\( u_{z0}^*(r,0) + u_{z1}^*(r,0) \) as \( w(r) + u_{z0}(a,0) + u_{z1}(a,0) \), where \( w(r) \) was the undeformed surface
profile of the indenter in reference to a cutting plane corresponding to \( r = a \).

Nevertheless, due to the geometrical simplicity of the single asperity, the prescribed
surface displacements \( u_{z0}^*(r,0) + u_{z1}^*(r,0) \) in the real contact domain have a one-to-one
explicit relationship with the given initial contact condition, the problem is well-defined and thus a unique solution is guaranteed. Moreover, the single asperity contact problem is generally solved analytically due to its simplicity.

Obtaining an exact solution for the single asperity contact of layered surfaces is usually mathematically complicated and unnecessary. For example, analytical methods based on the direct formulation and minimum total potential energy formulations generally lead to an exact solution, thus are not applied in the single asperity contact. Approximate methods are generally applied to achieve a good compromise between the precision and the computation effort. For example, analytical methods based on the weighted residual formulation are applied since they lead to an approximate solution. All numerical methods, such as FDM, FEM, and BEM automatically lead to an approximate solution. However, FDM only applies to regular domain and not feasible for layered surfaces. BEM and FEM numerical methods based on any of the three formulations can solve the single asperity contact of layered surfaces due to its simplicity. Among them direct formulations is widely applied since it is most intuitive. BEM numerical methods based on direct formulation have been applied to analyze normal and quasi-sliding contact of single asperity. FEM numerical methods have been applied to analyze normal and sliding contact of single asperity. Plumet and Dubourg (1998) also analyzed the sliding and rolling contact of a rigid ellipsoid on a layered elastic half-space. Their results are not shown for brevity.

2.1 Analytical Methods Based on Weighted Residual Formulation.

Analytical methods based on the weighted residual formulation, such as the least-square error method, have been applied to the normal and quasi-sliding contact of single
asperity. The weighted residual formulation first assumes an approximate contact pressure profile satisfying the initial and boundary conditions. Since the assumed contact pressure profile is not exact, its substitution into the Eq. (2.1) leads to some residuals or errors. By minimizing the residual over some selected intervals or at some points, a sufficiently precise contact pressure profile is obtained. Once the contact pressure profile is known, the subsurface stresses are obtained directly using the elasticity theory for subsurface stresses under a given pressure profile.

Chen (1971) and Chen and Engel (1972) obtained the contact pressure profile and subsurface stresses for an elastic layered half-space under indenter of various geometries. The elasticity theory necessary to obtain the subsurface stresses in a layered half-space was initially developed by Burmister (1945) under a prescribed axisymmetric pressure profile and was extended by Chen (1971) to non-axisymmetric pressure profile. The contact pressure profile was obtained by employing the least-square error method (Chen and Engel, 1972). King and O’Sullivan (1987) obtained the contact pressure profile and subsurface stresses in a layered elastic half-space, induced by a 2-D quasi-sliding contact of a rigid cylinder. The elasticity theory necessary to obtain the subsurface stresses in a layered half-space under arbitrary contact pressure profiles was initially derived by Gupta et al. (1973) and the pressure profile under the cylindrical indenter was initially obtained by Chen and Engel (1972). O’Sullivan and King (1988) also obtained the contact pressure profile and subsurface stresses in a layered elastic half-space, induced by a 3-D quasi-sliding contact of a rigid sphere.
2.1.1 Least-square error method.

The least-square error method states that, under the approximated pressure distribution, the square error between the prescribed and calculated surface displacements over the real contact domain should be minimal. First, the unknown pressure distribution under the indenter is approximated in a series of simple basis functions, i.e.,

\[ p(r) = \sum_{j=1}^{N} a_j p_j(r), \]

where \( a_j \) are unknown coefficients. The surface displacements \( u_{zi,j} (r,0) \) due to \( p_j(r) \) acting alone are solved directly either analytically or numerically due to the simplicity of \( p_j(r) \). The total \( u_{zi}(r,0) \) is obtained by assembling \( u_{zi,j}(r,0) \). For materials with linear elastic and small strain behaviors, the assembling is a simple linear superposition. The square of the error over the real contact domain due to the approximation is given as

\[ e^2 = \frac{1}{a^2} \int_0^a \left( [u_{z0}(r,0) + u_{z1}(r,0) - w(r)]^2 r \right) dr \] (2.2)

Substituting \( u_{z}(r,0) \) into Eq. (2.2) and minimizing \( e^2 \) with respect to the unknown coefficients \( a_j \) and \( \delta \) results in a weighted residual formulation, which upon solving gives the coefficients \( a_i \) and the pressure distribution beneath the indenter.

A very important factor affecting the accuracy of the approximation scheme is the choice of the set of basis functions \( p_j(r) \). The Hertz solution for the homogenous contact case provides a good first basis function, while a relatively small number of higher order basis functions (typically less than five) are needed to accurately represent the smooth pressure distributions.
2.1.1.1 Quasi-sliding contact of a smooth sphere on a layered half-space.

A 3-D quasi-sliding contact of a smooth sphere on an elastic layered half-space is shown in Fig. 2.2. The problem is addressed the same as Eq. (2.1), where $a$ is the contact radius and $r = \sqrt{x^2 + y^2}$. The sphere is assumed to be rigid, i.e., $E_0 = \infty$ and $H_0 = \infty$. The layer thickness $h$ is set as 1/10 of the sphere radius $R$. $a_0$ is defined as the contact radius and $p_0$ is the pressure under the center of the sphere when the layer takes the same Young’s modulus as the substrate, i.e., a Hertzian contact. The normal load $W$ is set as the value corresponding to $a_0 = h$, and the corresponding shear traction due to friction is $\mu W$, along positive $x$ direction by default. The results presented here were originally obtained by O’Sullivan and King (1988) with the least-square error method and verified by Peng and Bhushan (2002) using the quasi-Newton method discussed later. It is worth being noticed that the correlation is good although different methodologies were applied. Moreover, O’Sullivan and King (1988) uncoupled the surface traction and solved the normal and tangential contact in sequence, the pressure profile in the contact zone was assumed to be unaltered by friction.

Figure 2.3 shows the contact pressure profile beneath the sphere. As expected, the contact radius decreases while the contact pressure increases by a stiffer layer ($E_1 / E_2 > 1$) and vice versa. The contact variables are readily normalized to their corresponding values at the homogenous case. For example, $x (0 < x < a)$ and $p$ are normalized to $a_0$ and $p_0$, respectively. This normalization scheme is shown in Fig. 2.3 by lower $x$ axis and left $y$ axis. Furthermore, since $a_0 = \left(\frac{3\pi p_n}{4E_2}\right)^{1/3} R$ and $p_0 = \left(\frac{6p_nE_2^2}{\pi^2}\right)^{1/3}$ (Bhushan, 1999b), $x$ and $p$ are further normalized to the prescribed contact variables. The final normalization
Figure 2.3 Profile of contact pressures beneath sphere at various $E_1/E_2$ [O’Sullivan and King, 1988, Peng and Bhushan, 2001b]

scheme is $x/((p_n/E_2)^{1/3})$ and $p/((p_nE_2^2)^{1/3})$, as shown in Fig. 4 by upper x axis and right y axis.

Figure 2.4 shows three subsurface stress components in the layer and substrate as a function of depth. The magnitudes of these stresses components are very high close to the layer surface and decay rapidly into the substrate. A stiffer layer ($E_1/E_2 = 2$)
Figure 2.4  Stress components along z-axis at various $E_1/E_2$ and coefficient of friction $\mu$ ($\sigma_{xz}$ is plotted along the line $x/a_0 = 0.5$, $y = 0$) [Peng and Bhushan, 2001b]
increases the magnitude of these stress components, especially in the area close to the layer surface, and vice versa. The increase of friction moves the maximum shear stress \( \sigma_{xz} \) closer to the layer surface but has negligible influence on other stress components.

Figure 2.5 shows contour plots of \( \sqrt{J_2 / p_0} \) beneath the spherical indenter. The presence of the stiffer layer increases the value of \( \sqrt{J_2} \) in the layer and vice versa. Discontinuities of \( \sqrt{J_2} \) occur at the interface for non-homogeneous cases. The maximum \( \sqrt{J_2} \) occurs beneath the center of the sphere and close to the surface for \( \mu = 0 \). The increase of friction increases the magnitude of the maximum \( \sqrt{J_2} \) and moves it closer to the layer surface. Since friction only changes shear stress \( \sigma_{xz} \) and has negligible influence on other stresses (Fig. 2.4), the change of \( \sqrt{J_2} \) due to friction is mainly from the shear stress \( \sigma_{xz} \).

2.2 BEM Numerical Methods Based on Direct formulation

Analytical methods based on the weighted residual formulation requires the knowledge of basis functions. Without this requirement, numerical methods have been applied to analyze the normal and quasi-sliding contact of single asperity. In these numerical methods, the sum of prescribed z-direction surface displacements

\[ u_{z0}^* (r,0) + u_{z1}^* (r,0) \]

are expressed as \( u(r) \), same as in the analytical methods. However, in the real contact domain, the pressure profile is represented by an array of discrete point loads \( (p_{ri}, i = 1, ..M) \) rather than a few pressure profiles in terms of continuous basis functions. Besides, to determine the unknown coefficient (magnitude) of these point load and \( \delta \), a match check is applied on these discrete points to ensure that the prescribed and
Figure 2.5  Contours of $\sqrt{J_2}/p_0$ at $y = 0$ at various $E_1/E_2$ and coefficient of friction $\mu$ [Peng and Bhushan, 2001b]
calculated surface displacements are the same, i.e., the error
\[ \varepsilon_i = u_{z0}(r_i,0) + u_{z1}(r_i,0) - w(r_i) = 0 \quad (i = 1, \ldots, M), \]
rather than a least-square error check over the whole continuous real contact domain.

As per the physics of formulation, the ideas behind analytical and numerical methods are the same: the stresses and displacements for a minimum set of distinct simple distributed loadings are calculated first using theory of elasticity, the stresses and displacements for a real situation are then obtained by proper assembling of solutions for simple loading conditions. The unknown assembling factors are obtained with an approximate approach to satisfy the boundary and initial conditions.

Among various numerical methods, BEM numerical methods based on the direct formulation, such as the matrix inversion method, is very simple and straightforward, therefore is feasible for single asperity contact. BEM discretizes only the surface of the single asperity, thereby reducing the size of the problem and dramatically reducing the effort involved in obtaining the solution. The direct formulation employs only basic physical principles such as equilibrium, compatibility and stress-strain governing equation, therefore is much intuitive as opposed to other formulations. Since the match check requires \[ u_{z0}(r_i,0) + u_{z1}(r_i,0) = u_{z0}^*(r,0) + u_{z1}^*(r,0) \quad \text{at discrete points } i \quad (i = 1, \ldots, M), \]
the direct formulation is usually expressed in matrix form as
\[ u_{z}^* = C^{u_z} \cdot p, \quad (2.3) \]
where \( p \) is the vector of contact pressures and denoted by \( p^T = (p_1, p_2, \ldots, p_k, \ldots, p_M) \), \( u_z^* \) is the total prescribed z-direction displacements vector and denoted by \( u_z^T = (u_{z1}^*, u_{z2}^*, \ldots, u_{zk}^*, \ldots, u_{zM}^*) \). \( C^{u_z} \) is the so-called influence coefficients matrix.
corresponding to \( u_z^* \), which is developed in Sec. 3.2.1. It contains the surface displacements \( u_{zj}(r,0) \) due to unit point load \( p_j(r) \) acting alone.

### 2.2.1 Matrix inversion method.

In case of single asperity contact, the matrix inversion method is very straightforward: the contact pressure profile is solved directly as \( P = C u_z^* \), where a matrix inversion is involved. The matrix inversion method was first used by Gupta and Walowit (1974) to solve a 2-D cylindrical indentation, as a comparable numerical approach to the analytical basis function approach (King and O’Sullivan, 1987). From the mathematics point of view, comparing to matrix inversion method, fewer unknowns are needed in the analytical basis function approach and no matrix inversion is required. However, in case of a complex pressure profile, e.g., multiple asperity contact, obtaining an expression of the pressure profile in closed form then decomposing it into a few basic functions presents difficulties, thus the analytical basis functions approach is unfeasible. The matrix inversion method still works due to the flexibility of the numerical approach.

#### 2.2.1.1 Normal contact of a smooth cylinder on a layered half-space.

A 2-D normal contact of an elastic smooth cylinder on an elastic layered half-space (cylindrical indentation) is shown in Fig. 2.1. Several simulations are carried out for various \( \gamma = E_1 / E_2 \) and layer thickness \( h \) with constant ratio of normal pressure \( p_n \) to the substrate’s Young’s modulus \( E_2 \). The results are presented in Fig. 2.6, which were obtained by Cole and Sayles (1992) with the matrix inversion method.. \( a_0 \) and \( p_0 \) are defined as the half contact width and the maximum pressure when the layer has the same Young’s modulus as the substrate, i.e., a Hertzian contact. As expected, in case of \( E_1 / E_2 \)
\neq 1$, the half contact width $a$ and the maximum pressure $p_{\text{max}}$ converge to the homogeneous solution $E = E_1$ with the increase of layer thickness, i.e., $a / a_0$ and $p_{\text{max}} / p_0 \rightarrow 0$ as $a_0 / h \rightarrow 0$. The contact variables are readily normalized to their corresponding values at the homogenous case. For example, $p_{\text{max}}$ and $a$ are normalized to $p_0$ and $a_0$, respectively, as shown in Fig. 2.6. Furthermore, since $a_0 = \left(\frac{8p_n}{\pi E_2}\right)^{1/2} R$ and $p_0 = \left(\frac{2p_nE_2}{\pi}\right)^{1/2}$ (Bhushan, 1999b), $a$ and $p_{\text{max}}$ can be further normalized to the prescribed contact variables. Cole and Sayles (1992) normalized the layer thickness $h$ to $a_0$, which is valid only for this specific $p_n / E_2$. A general normalized scheme is to normalize all the direct geometric variables to one among themselves, thus independent of initial and loading conditions. For 2-D cylindrical contact the direct geometric variable is chosen as the cylinder radius $R$. The final normalization scheme is $a / \left(\frac{p_n}{E_2}\right)^{1/2} R$, $p / (p_n E_2)^{1/2}$ and $h / R$.

2.3 FEM Numerical Methods.

Despite the simple geometry, single asperity contact problem may becomes very complex due to contact environments, e.g., complex loading conditions (contact, impact, etc.), complex material behaviors (plasticity, viscoelasticity, creep, etc.), complex structural responses (large deformations, crack propagation, etc.). These problems are usually solved by dividing (meshing) the single asperity into small and manageable elements. FEM methods are widely applied since commercial FEM packages provide advanced mesh generation tools, and well-developed models handling large deformations, which is required in sliding and rolling. FEM theoretically can achieve any...
Figure 2.6  Variation of the maximum contact pressure $p_{\text{max}}$ and half contact width $a$ with layer thickness $h$ at various $\gamma = E_1 / E_2$ [Cole and Sayles, 1992]

precision by reducing the elements’ sizes, i.e., refining the mesh. For the single asperity contact, since only a relatively small number of regular shaped elements are in contact, they can be easily refined without exceed the upper limit of element number. In this case, FEM provides rigorous results. Similar to FDM and BEM numerical models, FEM
methods are classified into three formulations and in single asperity contact, these three formulations generate the same results.

Komvopoulos (1988, 1989) analyzed a 2-D cylindrical indentation with commercial FEM software ABAQUS in the elastic and elastic-plastic modes, respectively. Diao et al. (1994) analyzed a 2-D quasi-sliding contact of a layered flat half-space under an elliptical contact pressure distribution. Diao et al. verified the accuracy of their meshing scheme by comparing the FEM results for a homogenous case with those of Smith and Liu (1953). They extended the relation between the maximum tensile stress and maximum pressure from homogenous case to layered case based on the FEM results. Kral et al. (1993, 1995a, b) analyzed a 3-D spherical indentation. Stephens et al. (2000) analyzed a 2-D quasi-sliding contact of a layered flat half-space under a cylindrical contact pressure distribution. Their emphases are on the initial yielding behavior of a layered solid with functionally graded substrate, e.g., gradients in yield strength and Young’s modulus. Their results suggest distinct benefits to the durability of layered systems with using a substrate with functionally graded properties. Kral and Komvopoulos (1996a, b, 1997) simulated a 3-D spherical indentation and sliding contact with commercial FEM software ABAQUS.

### 2.3.1 Finite element modeling.

In FEM, the selection of element type and meshing scheme affect the solution directly. An example of finite element modeling for the sliding contact of a smooth sphere on a layered half-space was given by Kral and Komvopoulos (1996a, b, 1997). The meshing scheme of the layered half-space in the first octant is shown in Fig. 2.7. The element type is eight-node linear interpolation reduced integration element, which uses a
Figure 2.7 3-D finite element mesh of the layered half-space in the first octant indented by a rigid sphere centered along z-axis, the inset is the magnified finer meshed region at the contact interface. [Kral and Komvopoulos, 1997]

Gaussian integration scheme to reduce the number of integration point to one per element. The inset of Fig. 2.7 shows the layer and the finer mesh region at the sliding interface. The mesh is refined by using linear constraints for the corner nodes of two
elements lying on the edge of a larger element, and bilinear constraints for corner nodes lying on the face center of a larger element. The dimensions of the smallest cubic elements are roughly equivalent to the contact radius at initial yielding of the substrate material. The plane $y = 0$ is a plane of symmetry, and sliding proceeds in the positive $x$-direction. The plane of symmetry and the bounding plane $y / R = 0.64$ are constrained against displacement in the $y$ direction, the bounding planes $x / R = -0.747$ and $0.747$ are constrained against displacement in the $x$ direction, and the plane $z / R = 0.713$ are constrained against displacement in the $z$ direction. The rigid spherical indenter of radius $R$ is modeled with two-node rigid surface contact elements, the meshing scheme is omitted for brevity. One node of each contact elements corresponds to a common master node through which loads and displacements are applied to the sphere. Indentation is then modeled by incrementally applying normal loads to the master node and hard contact is modeled as applying normal traction once the clearance between a surface node of the deformable layered medium and the rigid sphere surface reaches zero.

During the sliding the domain changes with time (transient) and the finite element mesh must move to satisfy the boundary conditions at the moving interfaces. An updated Lagrangian formulation is usually used for moving domains transient analyses.

### 2.3.1.1 Sliding contact of a smooth sphere on a layered half-space.

A 3-D sliding contact of a smooth sphere on a layered half-space was simulated by Kral and Komvopoulos (1997). The simulation consists of an incremental indentation to the specified normal load, followed by sliding at a contact normal load and coefficient of friction $\mu$, in increments to a specified distance ($2 \sim 3$ times the initial contact radius). Unloading of the sphere after sliding is also simulated by restraining the $x$ and $y$
displacements of the sphere and removing the normal load in the reverse order of the loading steps during the initial indentation. Elastic-perfectly plastic material behavior is adopted for both the layer and the substrate material. Plastic deformation is detected using von Mises yielding criterion. The material model for plastic deformation is based on the usual associated flow rule, and the assumption of negligible plastic volume change is maintained throughout. The plastic flow rule apply only if yielding occurs. Otherwise the usual elastic constitutive equations apply.

The evolutions of the plastic strain and stress component $\sigma_{xx}$ during sliding are shown in Fig. 2.8. The location of the sphere center is noted by the arrow. The top of the Fig. 2.8 shows the development of the equivalent plastic strain with increasing sliding distance. The equivalent plastic strain is defined as

$$\varepsilon_{eq}^p = \left[ \frac{2}{3} \varepsilon_{ij}^p \varepsilon_{ij}^p \right]^{1/2}$$

(2.4)

where $S$ is the strain path. The contours of equivalent plastic strain under indentation ($\Delta x = 0$) are shown in (a). The maximum plastic strain in the layer approaches the outer contact edge, while the maximum strain in the substrate remains at the layer interface near the sphere center. The deformation is still considered within the elastic-plastic indentation regime. As sliding commences for a sliding distance ($\Delta x > 0$) as shown in (b), the maximum equivalent plastic strain in the layer shifts to the front of the deformation region and the peak magnitude is almost doubled. The maximum plastic strain in the substrate occurs under the sphere center. With the progression of sliding as shown in (c), the maximum plastic strain in the layer near the front of the contact region continues to increase, but its location relative to the sphere center remains almost
Figure 2.8 Contours of $\varepsilon_{eq}^p$ and $\sigma_{xx}$ in the quadrant ($y > 0, z > 0$) at sliding distance $\Delta x$ equal to (a) $2a_y$, (b) $4a_y$, (c) $8a_y$, and (d) $12a_y$, with $E_1/E_2 = H_1/H_2 = 2$, $\mu = 0.1$, $W/W_y = 100$. $a_y$ and $W_y$ represent the sliding distance and normal loading corresponding to the initial yield condition of the substrate, respectively. [Kral and Komvopoulos, 1997]

constant. The maximum plastic strain in the substrate shifts slightly behind the sphere center. At the final sliding distance shown in (d), the maximum plastic strain in the layer approaches a constant magnitude, while a steady-state distribution in the substrate
becomes apparent: the plastic zone spreads parallel to the interface and the magnitude of the strain increases toward the interface.

The development of stress component $\sigma_{xx}$ during loading / sliding / unloading are also shown in the bottom of the Fig. 2.8. von Mises stress is calculated accordingly to detect yielding by comparing the maximum von Mises stress with the yield stress. In unloading process, it is re-yielding as the von Mises equivalent stress of the residual stresses exceeds the yield stress. Based on the calculated stresses, conclusion of stress effects are drawn, specially on the tensile normal stresses and on the shear stresses at the layer / substrate interface. The effects of layer properties, normal load and coefficient of friction are demonstrated by varying the contact parameters. These conclusions are quite similar to those drawn from other numerical models introduced later and omitted here for brevity.
CHAPTER 3

MULTIPLE ASPERITY CONTACT OF LAYERED ROUGH SURFACES

When two rough surfaces come in contact, the contact generally occurs at multiple asperities. A comprehensive review of modeling of multiple asperity contact of homogenous rough surfaces has been presented by Bhushan (1998). Here we complete the review with introducing numerical models recently developed for multiple asperity contact of layered rough surfaces contact.

A 3-D rough surfaces contact problem is addressed in Cartesian coordination system as

\[ u_{x_0}(x, y, 0) + u_{z_1}(x, y, 0) = u^*_x(x, y, 0) + u^*_z(x, y, 0), \quad (x, y) \in \Omega \]

\[ u_{x_0}(x, y, 0) + u_{z_1}(x, y, 0) > u^*_x(x, y, 0) + u^*_z(x, y, 0), \quad (x, y) \notin \Omega \]

\[ p(x, y) \geq 0, \quad (x, y) \in \Omega \]

\[ p(x, y) = 0, \quad (x, y) \notin \Omega \]

(3.1)

where \( \Omega \) is the real contact domain, \( p(x, y) \) and \( u^*_z(x, y, 0) \) must also satisfy the governing formulation of elasticity. If the real contact domain \( \Omega \) and corresponding prescribed surface displacement \( u^*_z \) are given in the initial condition as in single asperity contact, the problem is readily solved and the same results are obtained with all formulations. However, in general prescribing \( \Omega \) and \( u^*_z \) is impossible in the multiple asperity contact due to the geometrically complicated interface. Instead, the normal load \( W \) or the relative
rigid approach \( \delta \) are prescribed in the initial contact condition. The real contact domain \( \Omega \) therefore has a many-to-one implicit relationship with the given initial contact condition. The problem becomes ill-defined and the solution uniqueness is not guaranteed unless supplementary criterion is provided.

Generally the multiple asperity contact problem is practically unsolvable with analytical methods due to the geometrically complicated contact interface. Numerical methods are applied to obtain an approximate computational solution. The complicated object is divided (discretized) into small and manageable pieces, i.e., elements. Solutions are obtained for each element and assembled together into a complete solution, allowing for continuity at the inter-elemental boundaries, i.e., collocation at nodal points. Available numerical methods for structural problems are Finite Difference Method, Finite Element Method, and Boundary Element Method. FDM discretizes the entire object into finite elements. An algebraic equation of stresses and displacements is created by enforcing the governing differential formulation of elasticity on each node inside the object. Combining these equations and boundary conditions (including initial contact conditions) together creates a set of simultaneous linear equations which upon solving provides displacements and stresses inside the object. FDM has the advantages of simplicity and parallelization, but is limited for regular domain thus unfeasible for complex geometries or non-isotropic material, e.g., a layered rough surfaces contact.

In contrast to FDM, FEM employs the governing integral formulation of elasticity rather than the governing differential formulation of elasticity in each element inside the object to create an algebraic equation (Zienkiewicz, 1977; Beer and Watson, 1992). Moreover, an approximate continuous interpolation function is presumed in term of nodal
values to represent the distribution of stresses and strains over an element, e.g., as a constant, linear or quadratic variation. FEM has a near-universal adaptability and theoretically is capable to solve rough surfaces contact problems using any of the three formulations mentioned earlier. Komvopoulos and Choi (1992) applied FEM to solve a 2-D (plane-strain) homogenous elastic contact problem with only a few cylindrical asperities in contact. Oden and Martin (1985) and Martin et al. (1990) applied FEM to solve a 2-D homogenous sliding contact problem with two rough surfaces in contact. However, a 3-D layered rough surfaces contact generally requires a large number of fine mesh elements to describe the irregular real contact domain. The situation is especially severe in FEM since the entire volume of the contacting solids is divided into discrete finite elements. It makes commercial FEM packages unfeasible due to the requirement of extremely long computation time. A 3-D FEM model for layered rough surfaces contact has not been reported to date.

BEM employs a different discretization scheme from FDM and FEM (Man, 1994), where only the boundary (surface) rather than the whole object is discretized into finite elements. The governing differential equation set for the continuous region inside the object is transferred to the corresponding boundary integral equation set for the surface of the object. The boundary integral formulation of elasticity is employed on each boundary element to create an algebraic equation. Since BEM avoids discretizing the entire volume into discrete finite elements, it reduces the dimensionality of the problem and thus dramatically reducing the effort involved in obtaining a solution. BEM is widely used in continuum mechanics, especially in 3D rough surfaces contact problems.
BEM generally discretized the contacting rough surfaces into an array of square elements (patches) of equal size. The patch size is chosen to achieve a good compromise between the precision and the computation time. As FEM, an approximate continuous interpolation function is presumed to represent the distribution of pressures and displacements over each patch. The simplest numerical representation of a multiple asperity contact is an array of concentrated point contacts, where the pressure is presumed concentrated at the center point of each patch. The difficulty with this approximation lies in the infinite surface displacement occurring at the center point due to a point load. This difficulty is avoided if a constant (uniform) pressure is presumed acting on each element, i.e., a patch contact. This approach gives rise to a stepwise distribution of pressures over the interface and the surface displacements are finite anywhere (Cole and Sayles, 1992). Furthermore, a triangular pressure can be presumed acting on each element, and by the superposition of overlapped triangular pressures a trapezoidal pressure is obtained afterwards (Gupta and Walowit 1974; Kannel and Dow 1985; Merriman and Kannel 1989). This gives rise to a piecewise-linear distribution of pressures over the interface and the surface displacements are smooth and continuous everywhere (Johnson, 1985). However, the multiple asperity contact is well simulated by patch contacts if the patch size is small enough, which makes the superposition unnecessary.

After the surface discretization, BEM numerical methods are applied to analyze the multiple asperity contact. Since the weighted residual formulation requires closed form description of contact pressure profile, which can not be obtained for geometrically complicated interface thus not feasible for multiple asperity contact. The direct
formulation and the minimal total potential energy formulation have been applied in BEM numerical methods to analyze the normal and quasi-sliding contact of layered rough surfaces.

3.1 BEM Numerical Methods Based on Direct Formulation.

The direct formulation applies basic physical principles such as equilibrium, compatibility and stress-strain governing equation to solve the multiple asperity contact problems. It is much intuitive as opposed to formulations applying complex physical principles, e.g., the minimum total potential energy formulation, and to formulations more mathematically based, e.g., the weighted residual formulation. The direct formulation is usually expressed in matrix form as $u^* = C^{u^*} \cdot p$. Since the prescribed displacements $u^*$ is generally unknown, an initial approximation of $u^*$ is presumed. An iteration process based on $u^* = C^{u^*} \cdot p$ starts to modify the approximate prescribed displacements, until a steady solution satisfying all the initial conditions and boundary conditions is obtained. Since the relative rigid approach $\delta$ is usually prescribed, the initial approximation of $u^*$ is usually taken as the corresponding geometric interference. This approximation is an overestimate since the real contact domain is completely contained within the geometric overlap region (interpenetration area). It guarantees that the approximate solution always remains inside the domain of feasibility. However, since no supplementary criterion is provided, the solution uniqueness is not guaranteed with the direct formulation.
The matrix inversion method and the Conjugate Gradient Method have been applied in the iteration process to obtain the contact pressure profile corresponding to given surface displacements.

3.1.1 Matrix inversion method.

Since both the contact pressure distribution and the real contact domain are unknown, an iteration process is necessary to solve the direct formulation $u_z^* = C^{u z} \cdot p$.

Since the relative rigid approach $\delta$ is usually prescribed, an initial approximation of the real contact domain $\Omega$ and the corresponding prescribed surface displacements $u_z^*$ are decided by the geometric interference. $u_z^* = C^{u z} \cdot p$ is then transformed to $p = C^{u z -1} \cdot u_z^*$, where the influence coefficients matrix is inversed and its matrix product with the approximate surface displacements vector gives the pressure vector. Locations with negative pressure (surface tension) violates the requirement $p \geq 0$ and are excluded from the real contact domain. The prescribed surface displacements are then updated by $u_z^* = C^{u z} \cdot p$ to ensure that no geometrical interference occurs outside the real contact domain. The iteration is repeated in a sequence until convergence occurs with a set of non-negative contact pressures. If in the initial contact condition the normal load $W$ is prescribed instead of the relative rigid approach $\delta$, an outer level of iteration is necessary to adjust $\delta$ until the sum of all pressures balances $W$.

The matrix inversion method was initially used by Gupta and Walowit (1974) to analyze the contact between an elastic cylinder and a layered elastic solid – a single asperity contact problem. Since the surface displacements are prescribed, the pressure distribution is solved without iteration. Kannel and Dow (1985) and Merriman and
Kannel (1989) adopted the 2-D cylindrical single asperity contact model (Gupta and Walowit, 1974) and introduced roughness as the height perturbation relative to the smooth cylinder. Since it was a multiple asperity contact case with contact profiles unknown, an initial approximation to the prescribed surface displacements was chosen as undeformed profile of the indenter in reference to a cutting plane corresponding to \( r = a \) (Fig. 2.1). An iteration process was performed to modify the prescribed surface displacements until a steady solution satisfied the direct formulation and the corresponding boundary condition. Similar to Kannel and Dow (1985) and Merriman and Kannel (1989) approach, Cole and Sayles (1992) performed a series of homogenous and layered, smooth and rough cylindrical contact to verify the matrix inversion method with available theoretical analyses. In the initial contact condition the relative rigid approach \( \delta \) was prescribed instead of the half contact width \( a \). The approximation to the prescribed displacements was assigned as the geometrical interference corresponding to \( \delta \). Their calculations were based on unit pressure instead of unit point load (Gupta and Walowit, 1974; Kannel and Dow, 1985; Merriman and Kannel, 1989), which avoids the assumption on the size of the sampling interval. Mao et al. (1996, 1997) extended Cole and Sayles (1992) approach from normal contact to frictional contact. Tian and Bhushan (1996a) obtained \( p = C u^{-1} \cdot u^* \) by Newton minimization method, and solved it with the Gaussian-Seidel iteration. The inversed influence matrix was triangularly decomposed to guarantee that the matrix remains positive definite and nonsingular, even with finite round-off.
3.1.1.1 Quasi-sliding contact of a rough cylinder on a layered half-space.

A 2-D quasi-sliding contact of an elastic rough cylinder on an elastic layered half space was performed by Mao et al. (1996) with the matrix inversion method. The indenter is defined by superposing a smooth cylinder with a nominally flat rough surface, whose surface profile is shown in Fig. 3.1(a). The corresponding contact pressure profile is shown in Fig. 3.1(b) at $E_1/ E_2 = 1$, i.e., homogenous case. The Hertzian contact pressure corresponding to the smooth cylinder contact is also superimposed in Fig. 3.1 (b). The surface pressure distribution for the real surface profile is found significantly different from the Hertzian smooth contact. Figure 3.2 shows the variation of pressure distribution with coefficient of friction, layer thickness and $E_1/ E_2$. A compliant layer (Fig. 3.2 (a)) enlarges the contact zone and reduces the maximum pressure, while a stiff layer (Fig. 3.2 (b)) reduces the contact zone but increases the maximum pressure. These trends are more marked with an increase of layer thickness. The friction moves the contact zone and the peak pressure in the direction of tangential force. Pressures decrease at the left side of the contact center and increase at the right side due to the indenter’s cylindrical shape. This behavior is more marked with the increase of the coefficient of friction, especially in case of a compliant layer. Mao et al. (1996) also reported that Fig. 3.2 shows that in case of a compliant layer, the increase of friction enlarges the contact zone and reduces the pressures, and in case of a stiff layer it is the opposite. This is not true.
Figure 3.1 Profiles of (a) the nominally flat rough surface being superposed on the smooth cylinder and (b) corresponding contact pressures. The smooth curve is contact pressure profile of the smooth cylinder in Hertzian contact. [Mao et al., 1996]
Figure 3.2  Profiles of contact pressures at various layer thickness $h$ and coefficients of friction $\mu$ with (a) $E_1/E_2 = 3.2$ and (b) $E_1/E_2 = 0.17$. [Mao et al., 1996]
Figure 3.2 (Continued)
3.1.2 Conjugate gradient method.

In case of a large number of contact points, such as in 3-D rough surfaces contact, the influence coefficients matrix is very large and possibly ill conditioned due to the round-off error. The matrix becomes non-positive definite and consequently causes serious non-converging problems with matrix inversion method. Nogi and Kato (1997) and Polonsky and Keer (2000a) applied the Conjugate Gradient Method instead of matrix inversion method to improve the convergence rate (Winther, 1980) and to speedup the computation. The pressure \( p \) is solved directly from equation \( u_z^* = C^{u_z} p \) by CGM and no matrix inversion is involved. Ordinary CGM belongs to a class of iterative methods called minimization methods. CGM solves \( C^{u_z} p - u_z^* = 0 \) by searching for the value of \( p \) that minimizes the quadratic function \( V^* = (1/2) p^T C^{u_z} p - p^T u_z^* \). Since \( C^{u_z} p - u_z^* \) is the gradient of \( V^* \), it equals zero at the minimum \( V^* \). Therefore, once \( \Omega \) and \( u_z^* \) are known, the CGM solution not only leads to function minimization, it also satisfies \( C^{u_z} p - u_z^* = 0 \).

In this case, the direct formulation is mathematically equivalent to the minimum energy formulation. However, in case of unknown \( \Omega \) and \( u_z^* \), the specific versions of CGM designed to solve \( C^{u_z} p - u_z^* = 0 \) does not have a simple connection with function minimization and theoretically does not guarantee a unique solution (Nogi and Kato, 1997; Polonsky and Keer, 2000a). The convergence is not guaranteed either since although the procedure is meant to minimize the energy, the energy itself is never evaluated in the specific versions of CGM.

Nogi and Kato (1997) applied a CGM scheme (Hestenes and Stiefel, 1952) with prescribed rigid approach \( \delta \). The initial approximation to the prescribed surface
displacement $u^*_z$ is assigned as the geometrical interference. The CGM is represented by the following recurrence formulation

$$
\begin{align*}
  p_{i+1} &= p_i + \frac{s_i \cdot t_i}{t_i C^{u*} t_i} t_i \\
  s_{i+1} &= s_i - \frac{s_i \cdot t_i}{t_i C^{u*} t_i} C^{u*} t_i \\
  t_{i+1} &= s_{i+1} - s_{i+1} \cdot t_i s_i
\end{align*}
$$

(3.2)

where $p_0$ is an arbitrary pressure vector and $t_0 = s_0 = u^*_z - C^{u*} p_0$. A truncation of negative pressures in each iteration was applied. The iteration continues until no negative pressure is detected.

The most time-consuming work in CGM is to calculate the matrix product of matrix $C^{u*}$ and vector $t$. It generally requires $O(M^2)$ operations, where $M$ is the total number of contact points among $N^2$ points of the surface. The Fourier transform is applied as a mathematical tool to alter the problem from space domain into frequency domain, where the problem can be more easily solved. By performing Fourier Transform, matrix products in the space domain reduce to simple point-to-point products in the frequency domain. Therefore, in case of a large ratio of $M$ to $N^2$, in order to reduce the computation time, vector $t$ and matrix $C^{u*}$ are expanded to include all the points in the space domain and transferred to $\tilde{t}$ and matrix $\tilde{C}^{u*}$ in frequency domain, i.e.,

$$
\begin{align*}
  \tilde{t}(\xi, \eta, z) &= \Delta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} e^{-i \xi x_i \eta y_j} t(x_i, y_j, z) \\
  \tilde{C}^{u*}(\xi, \eta, z) &= \Delta^2 \sum_{i=1}^{N} \sum_{j=1}^{N} e^{-i \xi x_i \eta y_j} C^{u*}(x_i, y_j, z)
\end{align*}
$$

Matrix product
\[ f(x_i, y_j, z) = \sum_{i,j}^M C_i^u(x_i - x_j, y_j - y_j, z) \cdot t(x_i, y_j, 0) \text{ for } M \text{ contact points then becomes} \]

\[ \tilde{f}(\xi_m, \eta_n, z) = \tilde{C}_i^u(\xi_m, \eta_n, z) \cdot \tilde{t}(\xi_m, \eta_n, 0) \text{ for all } N^2 \text{ points, which needs } O(N^2 \log_2 N^2) \text{ operations instead of } O(M^2). \]

\[ \tilde{f}(\xi_m, \eta_n, z) \text{ is then transferred back to } f(x_i, y_j, z) \text{ in space domain by} \]

\[ f(x, y, z) = \frac{1}{N^2 \Delta^2} \sum_{m=-N/2+1}^{N/2} \sum_{n=-N/2+1}^{N/2} e^{-\sqrt{2}(\xi_m + \eta_n)} \tilde{f}(\xi_m, \eta_n, z). \]

A FFT subroutine performs the Fourier transform and its inverse transform to further improve the computation efficiency.

Polonsky and Keer (2000a) applied an alternative CGM scheme (Pshenichny and Danilin, 1975) with prescribed normal load \( W \) instead of rigid approach \( \delta \). A load-balance equation is enforced in the iteration and an outer level of iteration is avoided. In this CGM scheme, an initial non-negative pressure distribution is chosen to balance the normal load. The corresponding surface displacements are obtained accordingly and the rigid surface approach \( \delta \) is approximated as the average relative rigid approach of all the points with positive pressure. The re-check of contact is performed by enforcing a penalty function. A truncation of negative pressures in each iteration is applied. A uniform shift of \( p_i \) up and down is performed to satisfy \( \int_\Omega \tilde{p} \, d\Omega = W \) so that the shape of \( p_i \) remains unchanged. The steepest descent direction is applied in each iteration.

A comparison among the CGM schemes and matrix inversion methods such as Gaussian elimination method and Gauss-Seidel method, in solving the \( C_i^u \cdot p - u^*_i = 0 \) was presented by Polonsky and Keer (1999, 2000b). In both matrix inversion method and

42
CGM, the initial guess of the prescribed displacements results in an unstable direct formulation and cannot guarantee the uniqueness of the solution.

3.1.2.1 Normal contact of a rough sphere on a layered half-space.

A 3-D normal contact of a rigid rough sphere on an elastic layered half-space (spherical indentation) was performed by Nogi and Kato (1997) with a CGM-based purely elastic model. The indenter is defined by superposing a smooth sphere with a nominally flat rough surface, whose surface profile is shown in the top of Fig. 3.3. The layer and substrate’s properties are set as $G_1 = 100$ GPa, $G_2 = 79.2$ GPa, $k_1 = 4.11$ GPa, and $k_2 = 0.929$ GPa, i.e., a stiffer and harder layer. The layer thickness $h = 6 \, \mu$m, and applied normal load $W = 14.7$ N. The calculated contact pressures and the surface displacements are shown in the middle and the bottom of Fig. 3.3, respectively. The contact pressures and surface displacement differ considerably from those calculated by the smooth body theory, although the surface deformation is Hertzian on macroscopic scale. The maximum pressure is 6.37 times the Hertzian (smooth and homogenous case) maximum value.

Figure 3.4 shows contour plots of $\sqrt{J_2}$ in the half-space beneath the spherical indenter. The surface pressures associated with asperity contacts produce highly stresses zones where maximum $\sqrt{J_2}$ is very close to the surface. As the depth $z$ increases, the influence of surface roughness reduces and the subsurface stress states approach those of smooth contact. According to von Mises criterion, plastic yielding of materials initiates when the maximum value of $\sqrt{J_2}$ is equal to the shear strength of the material. Hence, it is effective to reduce plastic yield in the contacting body to replace the material in the
Figure 3.3  Profiles of a rough sphere with $R = 0.0137 \, \mu m$, and corresponding contact pressures and z-direction surface displacements at the contact interface [Nogi and Kato, 1996]
Figure 3.4 Contours of von Mises stresses (in GPa) on x-z plane through the center of the rough sphere in (a) a homogenous half-space and (b) a layered half-space [Nogi and Kato, 1996]
highly stressed zone near the surface with materials having a shear strength significantly higher than the substrate.

A comparison of non-layered case (Fig. 3.4 (a)) with stiffer-layered cases (Fig. 3.4 (b)) shows that the influence of the layer on the subsurface stress states is not remarkable because the shear modulus of the layer is not extremely larger than the substrate \((G_1/G_2 = 1.263)\). However, the plastic yielding is considerably reduced because \(k_1/k_2 = 4.423\), thus the von Mises stresses in the layer can rise to a maximum approximately 4.4 times larger than the substrate.

### 3.2 BEM Numerical Methods Based on Minimum Total Potential Energy Formulation.

The matrix inversion method and even CGM may not converge with the presence of a large and ill-conditioned influence coefficients matrix. Moreover, the initial approximation of the prescribed surface displacements creates an uncertainty in the direct formulation, the uniqueness of the solution is not guaranteed. As a supplementary criterion to the direct formulation, the minimum total potential energy formulation allows the use of a quadratic programming method to solve an optimization problem, which theoretically guarantees the convergence and the uniqueness of the solution.

The minimum total potential energy formulation have been applied to rough surfaces contact problems for two reasons: to establish conditions that will determine the shape and size of the contact area and the contact stresses uniquely and to enable well-developed techniques of optimization such as a quadratic programming method to be used in numerical solutions.
Two minimum total potential energy formulations, namely minimum total elastic energy principle and minimum total complementary potential energy principle, apply to elastic and elastic-perfectly plastic contact problems (Richards, 1977). The latter one is more convenient to work in terms of unknown contact pressure, but only valid for small displacements. Since most cases considered in this paper are dealing with small displacements, the minimum total complementary potential energy principle is applied (Kalker and van Randen, 1972; Kalker, 1977). It states that, of all possible equilibrium stress fields for a solid subjected to prescribed loadings and boundary (surface) displacements the true one, corresponding to a compatible strain field, is identified by a minimum value for the total complementary energy. Mathematically the minimum total complementary potential energy principle is to minimize a variational inequality by performing quadratic programming; it also referred as to variational principle. The variational principle was first proposed to solve homogenous rough surfaces contact problems by Tian and Bhushan (1996a). Since the minimum total complementary potential energy principle is still valid for non-homogenous and anisotropic solids, Peng and Bhushan (2001a, 2001b, 2002) applied the variational principle to solve the layered rough surfaces contact problems with a large number of contact points (Fig. 3.5).

Assuming the tangential surface displacements are negligible, the total complementary potential energy $V^*$ of two elastic / perfectly plastic solids is given by (Richards, 1977)

$$V^* = U_E^* - \int_{\Omega} p(u_{z0}^* + u_{z1}^*) d\Omega = U_E^* - \int_{\Omega} pu_{z}^* d\Omega \quad (3.3)$$
Figure 3.5 Schematics of (a) a rough surface in contact with a layer rough surface and (b) top view of contact regions
where $U_{E}^{*}$ is the internal complementary energy of the two stressed solids, $\Omega$ is the real contact domain on which the $z$-direction surface displacements are prescribed as $u_{zi}^{*}$ ($i = 0,1$), $u_{z}^{*}$ is the sum of the prescribed $z$-direction displacements of the two contacting solids and $p$ represents the contact pressures in $\Omega$.

The definitions of strain energy and complementary energy are shown in Fig. 3.6 (a). For linear elastic materials, the internal complementary energy $U_{E}^{*}$ is numerically equal to the elastic strain energy $U_{E}$ (Fig. 3.6 (b)). Since the pressures are zero outside $\Omega$, $U_{E}$ is given by

$$U_{E} = \frac{1}{2} \int_{\Omega} \{pu_{z0} + u_{z1}\} d\Omega = \frac{1}{2} \int_{\Omega} pu_{z} d\Omega \tag{3.4}$$

where $u_{zi}$ ($i = 0,1$) are the calculated $z$-direction surface displacements of the two contacting solids (identified by indices 0 and 1), and $u_{z}$ is the sum of the calculated $z$-direction surface displacements inside the contact domain.

Replacing $U_{E}^{*}$ with $U_{E}$ in Eq. (3.3), the total complementary potential energy is given by

$$V^{*} = \frac{1}{2} \int_{\Omega} pu_{z} d\Omega - \int_{\Omega} pu_{z}^{*} d\Omega \tag{3.5}$$
Figure 3.6  (a) Definition of strain energy and complementary energy, (b) relationship between elastic strain energy and internal complementary energy for a linear elastic or a linear elastic-perfectly plastic material.
In order to apply the BEM numerical method, the contact interface needs to be discretized into boundary elements, usually square in shape, i.e., patches. The patch size is chosen to achieve a good compromise between the precision and the computation time. The pressures on each patch can be assumed either as constant, i.e., patch contact, or piecewise-linear, which is obtained by the superposition of overlapped triangular pressures. The latter produces surface displacements which are smooth and continuous everywhere. However, since the rough surfaces contact is well simulated by patch contacts provided the patch size is small enough, the superposition is unnecessary. The continuous contact is then replaced by a discrete set of patch contacts (Fig. 3.7 (a)) and the boundary conditions are checked at each patch center - the “matching point”.

Equation (3.5) is then discretized to

\[ V^* = \frac{1}{2} \sum_{k=1}^{M} p_k u_{zk} - \sum_{k=1}^{M} p_k u_{zk}^* \]  \hspace{1cm} (3.6)

where \( k \) is the index of the match points inside \( \Omega \), \( u_{zk} \) and \( u_{zk}^* \) are the calculated and prescribed z-direction surface displacement at point \( k \), and \( M \) is the total number of match points inside \( \Omega \) (from a total of \( N^2 \) match points on the contact interface).

To relate displacement \( u_{zk} \) with pressure \( p_l \), \( C_{lk}^u \), the influence coefficient corresponding to z-direction surface displacement, is introduced and developed in the following section. \( C_{lk}^u \) represents the z-direction surface displacement \( u_z \) at a general point \( k \) \((k = 1,2,...,M)\) induced by a uniform pressure centered at a general point \( l \) \((l = 1,2,...,M)\). The total z-direction surface displacement of the two contact surfaces at point \( k \) is then given by
Figure 3.7  Schematics of surface discretization at the contact interface, (a) 3-D view in space domain, (b) top view in space domain, and (c) top view in frequency domain
Substituting $u_{zk}$ in Eq. (3.6), $V^*$ is given by

$$V^* = \frac{1}{2} \sum_{k=1}^{M} p_k \left( \sum_{l=1}^{M} C_{ik}^{u_z} p_l \right) - \sum_{k=1}^{M} p_k u_{zk}^*.$$  

(3.8)

Before proceeding, it is convenient to write Eq. (3.8) in quadratic form as

$$V^*(p) = \frac{1}{2} p^T C^{u_z^*} p - p^T u_z^*,$$  

(3.9)

where $p^T$ is the transpose vector of contact pressures and denoted by

$$p^T = (p_1, p_2, \cdots, p_k, \cdots, p_M), \quad u_z^*$$  

is the total prescribed $z$-direction displacements vector and denoted by $u_z^{*T} = (u_{z1}^*, u_{z2}^*, \cdots, u_{zk}^*, \cdots, u_{zM}^*)$. $C^{u_z}$ is the influence coefficients matrix corresponding to the $z$-direction surface displacements and denoted by

$$C^{u_z} = \begin{bmatrix}
C_{11}^{u_z} & \cdots & C_{1k}^{u_z} & \cdots & C_{1M}^{u_z} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
C_{M1}^{u_z} & \cdots & C_{MK}^{u_z} & \cdots & C_{MM}^{u_z}
\end{bmatrix}$$

In case of $u_z^*$ are known, finding the minimum value of Eq. (3.9) in terms of $p$ gives

$$p = C^{u_z^{-1}} \cdot u_z^*, \quad \text{which is the same as the one derived in the direct formulation.}$$

However, $u_z^*$ are generally unknown in multiple asperity contact and so is $\Omega$.

Kalker (1977) extended the principle of minimal total complementary energy in a manner that the exact $u_{zk}^*$ needs no longer be known beforehand. For example, the actual contact does not need to be established in normal contact and the slip needs not occur in frictional
contact. Since the real contact domain $\Omega$ is unknown, a larger domain embedding the
real contact domain, referred as to the potential contact domain, is selected to replace $\Omega$.
It consists of the real contact domain and part of the domain with prescribed traction ($p =
0$). Kalker (1977) showed that the uniqueness and minimality of the solution is
maintained while replacing $\Omega$ with the potential contact domain.

The equivalent variational description of the principle of minimum total
complementary potential energy then becomes to find $p$ which satisfies

$$\min V^* = \frac{1}{2} p^T C^{zz} p - p^T u^*,$$  \hspace{1cm} (3.10)

The restriction of Eq. (3.1) is modified accordingly and expressed in discretized form as

$$u_{\alpha i} \geq u_{\alpha j}^*, \quad i = 1, \ldots, M$$
$$u_{\alpha i} > u_{\alpha j}^*, \quad i \neq 1, \ldots, M$$
$$p_i \geq 0, \quad i = 1, \ldots, M$$
$$p_i = 0, \quad i \neq 1, \ldots, M$$ \hspace{1cm} (3.11)

Eq. (3.10) under the restriction of Eq. (3.11) is a bounded indefinite quadratic
problem. Well-developed techniques of optimization such as quadratic programming can
be applied to obtain a unique solution.

This potential contact domain can be determined by the geometrical interference
criteria, assuming the contact surfaces are aligned with each other and their reference
planes are parallel. Since the relative rigid approach $\delta$ is usually prescribed, $\Omega$ is
composed of regions with $\delta > f_\alpha(x_k, y_k)$ and $u_{\alpha k}^*$ over $\Omega$ is given by:

$$u_{\alpha k}^* = \delta - f_\alpha(x_k, y_k), \quad (x_k, y_k) \in \Omega$$ \hspace{1cm} (3.12)

where $f_\alpha(x_k, y_k)$ is the initial separation of the two contact surfaces at $k$ and

$f_\alpha(x_k, y_k) = |z_0(x_k, y_k) - z_1(x_k, y_k)|$. Assuming there is no contact between the two surfaces
at the initial position, i.e., \( z_0(x_k, y_k) \geq z_f(x_k, y_k) \), the total prescribed displacement of the two contacting surfaces over the potential contact area is

\[
\overline{u}_{zk}^* = \delta - [z_0(x_k, y_k) - z_1(x_k, y_k)], \quad (x_k, y_k) \in \Omega
\]

(3.13)

An example is shown in Fig. 3.8 as two identical elastic spheres in contact.

If it is the normal load \( W \) is known instead of the relative rigid approach \( \delta \), an outer level of iteration is needed to adjust \( \delta \) until the sum of all positive pressures balance the normal load \( W \). An alternative approach (Stanley and Kato, 1997; Ai and Sawamiphakdi, 1999) is to assign the potential contact area \( \Omega \) as the nominal contact area \( A_n \), and \( u_z^* \) as the gap between the two undeformed surfaces. Since \( u_z^* < 0 \) for all the points in \( \Omega \), the boundary condition \( u_z(x, y, 0) \geq u_z^*(x, y, 0), (x, y) \in \Omega \) is automatically satisfied and thus replaced by an initial contact condition \( \overline{\mathbf{p}} dA = W \) accordingly. Since in general the real area of contact \( A_r \) is just a very small part of the nominal contact area \( A_n \), this approach overestimates the potential contact area, which induces huge overhead and is not efficient.

It is worth being noticed that the effect of tangential surface traction (shear traction) on the complementary energy is essential neglected since the tangential surface displacement is relatively small for most contact situations. For a more robust approach, the total complementary potential energy is modified with tangential energy included, i.e.,

\[
V^* = \frac{1}{2} \int_{\Omega} p u_z d\Omega + \frac{1}{2} \int_{\Omega} \mu p u_z d\Omega - \int_{\Omega} p u_z^* d\Omega - \int_{\Omega} \mu p u_z^* d\Omega
\]

(3.14)

Since the complementary energy principle is extremely flexible as regards the boundary conditions that prevail in a contact, an alternative approach is integrating the shear
Figure 3.8 (a) Geometrical interference area and real contact area in the normal contact of two identical spheres, and (b) determination of the total prescribed $z$-direction surface displacement from geometrical interference

3.2.1 Influence coefficients.

Assuming all the contact pressures $p_l (l = 1..M)$ are known, the stresses and displacements at any point $k (x_k, y_k, z_k)$ are computed by
where the influence coefficients $C^\sigma_{lk}$ represent the stresses and $C^u_{lk}$ represent the displacements at point $k$ $(x_k, y_k, z_k)$ induced by a uniform pressure enforced at point $l$ $(x_l, y_l, 0)$ with the corresponding shear traction due to friction. $C_{lk}$ represents the combination of $C^u_{lk}$ and $C^\sigma_{lk}$. For an isotropic material, the value of the influence coefficient only depends on the geometrical distance between these two points of interest at each axis, i.e.,

$$C_{lk} = C(x - x_l, y - y_l, z - z_l)$$

$$= C(x, y, z)$$

$$= C(0 - x_l, 0 - y_l, 0 - z_l)$$

Therefore, the influence coefficients $C_{lk}$ are calculated as the stresses and displacements at point $(x, y, z)$ induced by a uniform pressure enforced at point $(0, 0, 0)$, with the corresponding shear traction due to friction.

For a homogeneous elastic solid, $C^u_{lk}(x,y,0)$, the influence coefficient corresponding to $z$-direction surface displacement $u_z$, equals

$$\frac{4(1 - \nu^2)}{E} \int_{-\frac{\Delta}{2}}^{x+\frac{\Delta}{2}} \int_{-\frac{\Delta}{2}}^{y+\frac{\Delta}{2}} \left( \frac{1}{\sqrt{x^2 + y^2}} \right) dx dy$$

and approximated as (Love, 1929)
where $E$ is Young's modulus, $\nu$ is Poisson's ratio, and $\Delta$ is the patch width. Tian and Bhushan (1996a) applied this approximation in their numerical model to solve the homogenous rough surfaces contact.

For a 2-D layered elastic solid, the influence coefficients were derived by Gupta and Walowit (1974), Merriman and Kannel (1989), Cole and Sayles (1992) and Mao et al. (1997). The results are omitted for brevity. The influence coefficients for a 3-D layered elastic solid were derived by Nogi and Kato (1997) and Peng and Bhushan (2001a) for the normal contact, and Peng and Bhushan (2002) for the quasi-sliding contact. They followed the same procedure originally developed to analytically calculate the surface displacements under a prescribed surface loading $p(x, y, 0)$ and $q(x, y, 0)$ (O’Sullivan and King, 1988). To apply this procedure to obtain the influence coefficients,
simply set the prescribed surface loading as a uniform pressure centered at the origin, i.e.,
\[ p(0, 0, 0) = 1, \]
with corresponding shear traction due to friction, i.e.,
\[ q(0, 0, 0) = \mu. \]

The boundary conditions on the layer surface \((z_1 = 0)\) are then given by:
\[
\begin{align*}
\sigma_{xz}^{(1)}(x, y, 0) &= -p(x, y, 0) = 1 & x = 0, y = 0 \\
\sigma_{xz}^{(1)}(x, y, 0) &= \mu \cdot p(x, y, 0) & x = 0, y \neq 0 \\
\sigma_{zz}^{(1)}(x, y, 0) &= 0 & x \neq 0, y \neq 0
\end{align*}
\]

where \(\mu\) is the coefficient of friction and the superscript \((1)\) refer to the layer. The minus sign represents compressive stress by convention. The default shear traction is set along the positive \(x\) direction. The interface between the layer and the substrate \((z_1 = h\) and \(z_2 = 0)\) is required to have continuous stresses and displacements, assuming that the layer is fully bonded to the substrate, i.e.,
\[
\begin{align*}
\sigma_{xz}^{(1)}(x, y, h) &= \sigma_{xz}^{(2)}(x, y, 0), & u_x^{(1)}(x, y, h) &= u_x^{(2)}(x, y, 0), \\
\sigma_{yz}^{(1)}(x, y, h) &= \sigma_{yz}^{(2)}(x, y, 0), & u_y^{(1)}(x, y, h) &= u_y^{(2)}(x, y, 0), \\
\sigma_{zz}^{(1)}(x, y, h) &= \sigma_{zz}^{(2)}(x, y, 0), & u_z^{(1)}(x, y, h) &= u_z^{(2)}(x, y, 0).
\end{align*}
\]

where \(h\) is the layer thickness and the superscript \((2)\) refer to the substrate. In the substrate, the stresses and displacements fall to zero at a large distance from the load, i.e.,
\[
\sigma^{(2)}(x, y, \infty) = 0, \quad u^{(2)}(x, y, \infty) = 0
\]

The equilibrium, compatibility and Hooke’s law are automatically satisfied provided that the stresses and the displacements are expressed in terms of stress function or harmonic function. The stress function was introduced first in the solution of 2-D problems by Airy (1862) and also called the Airy stress. The harmonic functions for 3-D problems are Papkovich-Neuber elastic potentials \(\varphi\) and \(\psi = (\psi_1, \psi_2, \psi_3)\) for zero body
force (O’Sullivan and King, 1988). It is known that the number of independent 3-D harmonic functions can be reduced to three by arbitrarily choosing one of \( \psi = (\psi_1, \psi_2, \psi_3) \) to zero (Malvern, 1969). Here let \( \psi_2 = 0 \). In terms of these 3-D harmonic functions, the stresses and displacements are expressed as

\[
2G\varphi_i = \varphi_{,i} + x\psi_{1,i} + z\psi_{3,i} - (3 - 4\nu)\psi_i, \\
\sigma_{ij} = \varphi_{,ij} - 2\nu(\psi_{1,i} + \psi_{2,j})\delta_{ij} - (1 - 2\nu)(\psi_{1,ij} + \psi_{2,j}) + x\psi_{1,ij} + z\psi_{3,ij},
\]

where \( G \) is the shear modulus, the indices \( i \) and \( j \) range over 1, 2 and 3 corresponding to \( x, y \) and \( z \), respectively; and \( \delta_{ij} \) is the Kronecker delta. \( \delta_{ij} \) equals 1 if \( i \) and \( j \) are the same, and 0 otherwise.

The harmonic functions are obtained by enforcing the boundary conditions and interface conditions. The Fourier transform is applied here to speed up the computation. The Fourier transform decomposes the harmonic function into sinusoids of different frequency which sum to the original harmonic function. The transformed Papkovich-Neuber potentials in the layer are given by

\[
\bar{\varphi}^{(1)} = A^{(1)} e^{-\alpha z_1} + \overline{A}^{(1)} e^{\alpha z_1}, \\
\bar{\psi}_1^{(1)} = B^{(1)} e^{-\alpha z_1} + \overline{B}^{(1)} e^{\alpha z_1}, \\
\bar{\psi}_3^{(1)} = C^{(1)} e^{-\alpha z_1} + \overline{C}^{(1)} e^{\alpha z_1},
\]

where \( \alpha = \frac{\xi^2 + \eta^2}{\sqrt{\xi^2 + \eta^2}} \) and the Fourier transform variables \( \xi \) and \( \eta \) correspond to \( x \) and \( y \), respectively. In the substrate these Papkovich-Neuber potentials are given by

\[
\bar{\varphi}^{(2)} = A^{(2)} e^{-\alpha z_2} + \overline{A}^{(2)} e^{\alpha z_2}, \\
\bar{\psi}_1^{(2)} = B^{(2)} e^{-\alpha z_2} + \overline{B}^{(2)} e^{\alpha z_2}, \\
\bar{\psi}_3^{(2)} = C^{(2)} e^{-\alpha z_2} + \overline{C}^{(2)} e^{\alpha z_2}.
\]
where \( A^{(2)} = \overline{B}^{(2)} = \overline{C}^{(2)} = 0 \) since the stresses are zero at an infinite depth as shown in Eq. (3.20).

The stresses and displacements are expressed as the functions of \( A^{(1)}, \overline{A}^{(1)}, B^{(1)}, \overline{B}^{(1)}, C^{(1)}, \overline{C}^{(1)}, A^{(2)}, B^{(2)} \) and \( C^{(2)} \) accordingly. Enforcing the boundary conditions and interface conditions leads to a linear system of nine equations with nine unknowns. First by solving the following three equations in the frequency domain,

\[
\begin{align*}
\sigma_{xz}^{(1)}(x, y, 0) &= \mu p(x, y, 0) \quad x = 0, y = 0 \\
\sigma_{xz}^{(1)}(x, y, h) &= \sigma_{xz}^{(2)}(x, y, 0), \\
u_x^{(1)}(x, y, h) &= u_x^{(2)}(x, y, 0).
\end{align*}
\]

the three unknowns \( B^{(1)}, \overline{B}^{(1)} \) and \( B^{(2)} \) are obtained as

\[
B^{(1)} = \left[ \frac{(G - 1)e^{-\alpha h}}{(1 + G) + (1 - G)e^{-2\alpha h}} \right] \frac{\mu \overline{p}(\xi, \eta, 0)}{2\alpha(1 - \nu_1)},
\]

\[
B^{(1)} = \overline{B}^{(1)} + \frac{\mu \overline{p}(\xi, \eta, 0)}{2\alpha(1 - \nu_1)},
\]

\[
B^{(2)} = 2\left( \frac{1 - \nu_1}{1 - \nu_2} \right) \left[ \frac{e^{-\alpha h}}{(1 + G) + (1 - G)e^{-2\alpha h}} \right] \frac{\mu \overline{p}(\xi, \eta, 0)}{2\alpha(1 - \nu_1)}.
\]

where \( G_1 = \frac{E_1}{2(1 + \nu_1)}, \ G_2 = \frac{E_2}{2(1 + \nu_2)}, \ G = \frac{G_1}{G_2}, \) and \( p(\xi, \eta, 0) \) is the Fourier transform of \( p(x, y, 0) \) and equals 1 everywhere. The solution of the six equations left in the frequency domain is
\[ \alpha A^{(1)} - \alpha \overline{A}^{(1)} + (1 - 2\nu_1)C^{(1)} + (1 - 2\nu_1)\overline{C}^{(1)} = \alpha R_1 \]
\[ \alpha A^{(1)} + \alpha \overline{A}^{(1)} + 2(1 - \nu_1)C^{(1)} - 2(1 - \nu_1)\overline{C}^{(1)} = \alpha R_2 \]
\[ \alpha A^{(1)} - \alpha e^{2\alpha h}\overline{A}^{(1)} + [(1 - 2\nu_1) + \alpha h]C^{(1)} + [(1 - 2\nu_1) - \alpha h]e^{2\alpha h}\overline{C}^{(1)} - \alpha e^{\alpha h}A^{(2)} - (1 - 2\nu_2)e^{\alpha h}C^{(2)} = \alpha R_3 \]
\[ e^{2\alpha h}\overline{C}^{(1)} - \alpha e^{\alpha h}A^{(2)} - 2(1 - \nu_2)e^{\alpha h}C^{(2)} = \alpha R_4 \]
\[ \alpha A^{(1)} + \alpha e^{2\alpha h}\overline{A}^{(1)} + [2(1 - \nu_1) + \alpha h]C^{(1)} - [2(1 - \nu_1) - \alpha h]e^{2\alpha h}\overline{C}^{(1)} - G\alpha e^{\alpha h}A^{(2)} = \alpha R_5 \]
\[ \alpha A^{(1)} - \alpha e^{2\alpha h}\overline{A}^{(1)} + [(3 - 4\nu_1) + \alpha h]C^{(1)} + [(3 - 4\nu_1) - \alpha h]e^{2\alpha h}\overline{C}^{(1)} - G\alpha e^{\alpha h}A^{(2)} - G(3 - 4\nu_2)e^{\alpha h}C^{(2)} = \alpha R_6 \]

where \( R_1, R_2, R_3, R_4, R_5 \) and \( R_6 \) are given in Appendix 1.

Defining the terms \( \lambda, k, \) and \( S_0 \) by

\[ \lambda = 1 - \frac{4(1 - \nu_1)}{1 + G(3 - 4\nu_2)}, \]
\[ k = \frac{G - 1}{G + (3 - 4\nu_1)}, \]
\[ S_0 = [G + (3 - 4\nu_1)][1 - ke^{-2\alpha h}], \]

the six unknowns \( A^{(1)}, \overline{A}^{(1)}, C^{(1)}, \overline{C}^{(1)}, A^{(2)}, \) and \( C^{(2)} \) are obtained from Eq. (3.25) and given by

\[ C^{(1)} = (1 - \lambda)kS_0R_1 / \{4(1 - \nu_1)(G - 1)[1 - (\lambda + k + 4k\alpha^2h^2)e^{-2\alpha h} + \lambda k e^{-\alpha h}]\} \]
\[ \overline{C}^{(1)} = [2(G - 1)\alpha h e^{-\alpha h}C^{(1)} + e^{-2\alpha h}R_4] / S_0 \]
\[ A^{(1)} = [-(3 - 4\nu_1)C^{(1)} + \overline{C}^{(1)} + \alpha(R_1 + R_3)] / (2\alpha) \]
\[ \overline{A}^{(1)} = [(1 - \lambda)\alpha e^{-\alpha h}C^{(1)} + [(3 - 4\nu_1)(1 - e^{-\alpha h}) - 2\alpha h]\overline{C}^{(1)} + e^{-\alpha h}R_4] / [2\alpha(1 - e^{-2\alpha h})] \]
\[ C^{(2)} = \{4(1 - \nu_1)(1 - \lambda)e^{-\alpha h}C^{(1)} + (1 - \lambda)\alpha e^{-\alpha h}[R_3 - R_4 + R_5 - R_6]\} / \{4(1 - \nu_1)\} \]
\[ A^{(2)} = [2\alpha h e^{-\alpha h}[S_0 - (G - 1)(1 - e^{-2\alpha h})]C^{(1)} - [(3 - 4\nu_2)S_0]C^{(2)} + \alpha e^{-\alpha h}S_0[R_1 + R_2 - R_3 + R_4] - [e^{-\alpha h}(1 - e^{-2\alpha h})]R_4] / (2\alpha S_0) \]

where \( R_{ab}, R_{bc}, R_c \) and \( R_d \) are given in Appendix 1.
Substituting the expressions of $A^{(1)}$, $\overline{A}^{(1)}$, $B^{(1)}$, $\overline{B}^{(1)}$, $C^{(1)}$, $\overline{C}^{(1)}$, $A^{(2)}$, $B^{(2)}$ and $C^{(2)}$ into the Fourier Transformed version of Eq. (3.21), the displacement and stresses are obtained in the frequency domain. The derived coefficients of Papkovich-Neuber potentials - $A^{(1)}$, $\overline{A}^{(1)}$, $B^{(1)}$, $\overline{B}^{(1)}$, $C^{(1)}$, $\overline{C}^{(1)}$, $A^{(2)}$, $B^{(2)}$ and $C^{(2)}$ - are identical to those presented by O’Sullivan and King (1988) and Peng and Bhushan (2002), except that O’Sullivan and King have an erroneous expression of $C^{(1)}$. For normal contact ($\mu = 0$), the solution is simplified and explicit expressions of the influence coefficients $\overline{C}(\xi, \eta, z)$ are given in Appendix 2. These influence coefficients are identical to those presented by Nogi and Kato (1997) and Peng and Bhushan (2001a), except that Nogi and Kato have an erroneous $\xi$ instead of $\eta$ in their $\overline{C}^{\sigma_{yz}^{(M)}}(\xi, \eta, z)$ expression and an erroneous $\eta$ instead of $\xi$ in their $\overline{C}^{\sigma_{zx}^{(M)}}(\xi, \eta, z)$ expression.

Since these displacements and stresses are induced by a uniform pressure centered at origin with the corresponding shear traction due to friction, they represent influence coefficients in the frequency domain, expressed as $\overline{C}(\xi, \eta, z)$. The influence coefficients in space domain $C(x, y, z)$ are obtained by the inverse Fourier Transform of $\overline{C}(\xi, \eta, z)$.

The Fourier Transform and its inverse transform can be performed analytically or numerically. The latter will speeded up the computation tremendously by take discrete data as input (Fig. 3.7 (c)), where $\overline{C}(\xi, \eta, z) = \Delta^2 \sum_{i=0}^{N} \sum_{j=0}^{N} e^{i \sqrt{\pi}(\xi x_i + \eta y_j)} C(x_i, y_j, z)$ and

$$C(x, y, z) = \frac{1}{N^2 \Delta^2} \sum_{m=-N/2+1}^{N/2} \sum_{n=-N/2+1}^{N/2} e^{-i \sqrt{\pi}(\xi x + \eta y)} \overline{C}(\xi_m, \eta_n, z).$$
to further speed up the computation, in which original $N$ point sample is broken up into two ($N/2$) sequences following the divide and conquer approach.

While performing the numerical inverse Fourier Transform, the influence coefficient $\overline{C}(\xi, \eta, z)$ is assumed to be constant and represented by the value of the center point of each small patch in the frequency domain, except in the neighborhood of the origin where $\overline{C}(\xi, \eta, z)$ changes rapidly. These patches near to the origin are further discretized into $8 \times 8$ small patches. A 64-point Gaussian quadrature integration over the patches near the origin computes the average influence coefficient since $\overline{C}(\xi, \eta, z)$ is singular but integrable in these patches (Nogi and Kato, 1997; Peng and Bhushan, 2001a).

The use of the Fourier Transform implies a periodic extension of the contact domain, i.e., a periodic surface. For non-periodic surfaces, applying the Fourier Transform implicitly introduces an error caused by undesired periodic extension of the contact domain (Ju and Farris, 1996). The error is referred to as the periodicity error and particularly pronounced in the calculated interior stress fields. The periodicity error can be reduced by extending the nominal contact area sufficiently far beyond the real contact region and zero-padding the contact pressure array at the added nodes. However, the associated increase in the number of grid nodes significantly degrades the computation efficiency of FFT. An alternative yet fast approach to improve the accuracy was presented by Polonsky and Keer (2000a), where the periodicity error induced by FFT was compensated with a special correction term computed with a variation of the multi-level multi-summation (MLMS) technique (Brandt and Lubrecht, 1990; Polonsky and Keer, 1999). Another approach is developed by Ai and Sawamiphakdi (1999), where the total
pressure is decomposed into two parts: a smooth part and a fluctuation part. The load of the smooth part is set to the prescribed value so the net load of the fluctuation part is always zero, which greatly reduces the pressure fluctuations effect on displacement and interior stresses of the periodical extension of the contact pressure induced by Fourier Transform. In case of two nominally flat rough surfaces in contact, no periodic error occurs since flat surfaces with random roughness are considered periodic.

The sample influence coefficients are calculated with the preceding method. Since essentially a single patch contact problem is simulated to obtain the influence coefficients, the periodic error exists. However, the real contact region is one patch and relatively small to the normal contact region composed of $256 \times 256$ patches, the periodic error is neglected. In case of $E_1 / E_2 = 1$ it shows a good agreement with the homogeneous solid solution (Eq. (3.17)). In case of $E_1 / E_2 \neq 1$, as expected, $u_z$ converges to the homogeneous solution at $E = E_2$ with $h / \Delta \to 0$ and $E = E_1$ with $h / \Delta \to \infty$.

Dimensionless contact results are obtained attributed to the observation that linear relationship reserves throughout the derivation of influence coefficients. The contact variables can be readily normalized to their corresponding values at the homogenous case. For example, the influence coefficient corresponding to $z$-direction surface displacement at the origin $u_z$ is normalized to $u_{z0}$, as shown in Fig. 3.9 by left y-axis. Furthermore, since the derived contact variable $u_{z0} = 352 \Delta \frac{P_n}{E_2} \frac{1-v^2}{\pi}$, $u_z$ is further normalized as $u_z / (\Delta \frac{P_n}{E_2})$, as shown in Fig. 3.9 by right y-axis. The layer thickness $h$ can also be normalized to $u_{z0}$, as shown in Fig. 3.9 by lower x-axis. However, since $u_{z0}$ is a
Figure 3.9 Variation of $C_{zu}^{0}(0,0,0)$, the influence coefficient corresponding to z-direction surface displacement at the origin, with layer thickness $h$ at various $E_1/E_2$, for a homogeneous (Love, 1929) and a layered elastic half-space.

The normalization of $h$ with $u_{z0}$ only applies to the case of $u_{z0} = \Delta$ or $p_n/E_2 = 1$.

derived variable, the normalization only applies to one special case $p_n/E = 1$. A general scheme is to normalize all direct geometric variables to one among themselves, thus independent of other variables such as loading and elastic. For example, the layer thickness $h$ is normalized to the patch width $\Delta$, as shown in Fig. 3.9 by upper x-axis.
Contours of influence coefficients corresponding to $\sqrt{J_2}$ for case $h/\Delta = 12.8$ are plotted in Fig. 3.10. Comparing with Yu and Bhushan’s homogeneous solution (1996), the layered solution shows good agreement at $E_1/E_2 = 1$ and exhibits a consistent trend at $E_1/E_2 \neq 1$ as expected. The contours are symmetric with respect to the z-axis at $\mu = 0$. The maximum value is at the origin. As $\mu$ increase, the contours become asymmetrical and the magnitudes increase, especially in the region beneath the contact location and close to surface. Changing of layer property barely changes the distribution of influence coefficients corresponding to $\sqrt{J_2}$, except that a discontinuity occurs at the interface when the layer and substrate have different material properties.

### 3.2.2 Elastic / perfectly -plastic model.

Plastic deformation surrounding the asperity is of considerable interest in the study of interfacial phenomena, such as friction and wear. It is necessary to extend the contact model from pure elastic to elastic / plastic to produce more realistic results. The von Mises yield criterion is generally used to determine onset of plastic yield. However, the von Mises stresses are the model output and unknown beforehand. An alternative approach applies the theory of indentation hardness and assumes that a contact point deforms plastically once the local contact pressure exceeds the hardness of the softer solid in contact, referred to as $H_s$ (Tian and Bhushan, 1996a; Peng and Bhushan, 2001a).

This assumption generally agrees with the von Mises yield criterion, provided the contact asperities are very sharp, located far from and nearly independent of each other. This provision is satisfied for the contact of nominally flat rough surface under moderate load. For the contact of sphere asperities of large radius, this assumption is not valid.
Figure 3.10 Profiles of von Mises stresses (in pressure unit) on the surface and in the subsurface ($x = 0.3$ unit in length, $y = 0$) at various $E_1/E_2$ and coefficient of friction $\mu$ [Peng and Bhushan, 2002]
To further simplify the elastic / perfectly-plastic model, the region of the plastic deformation is assumed to be confined within a very small region and does not significantly alter the geometry of the elastically deformed region. Outside the plastic region, the relationship between contact pressure and elastic deformation still holds. With plastic dissipation energy included, the total complementary potential energy is modified as

\[ V^* = U^*_E + U_p - \frac{1}{4\pi} \int_\Omega p u^*_z d\Omega = U^*_E + \int_\Omega p_m \Delta u^*_p d\Omega - \frac{1}{4\pi} \int_\Omega p u^*_z d\Omega \]  \hspace{1cm} (3.27)

where \( U_p \) is plastic energy dissipated in plastic deformation, \( p_m \) is the contact pressure which reaches the hardness of the softer solid, i.e., \( p_m = H_s \), and \( \Delta u^*_p \) is the incremental surface displacement in the plastic deformation regime. It is known that for elastic-perfectly plastic contact, the internal complementary energy \( U^*_E \) is also numerically equal to the elastic strain energy \( U^*_E \) (Fig. 3.6 (b)). Therefore,

\[ V^* = U^*_E + \int_\Omega p_m \Delta u^*_p d\Omega - \frac{1}{4\pi} \int_\Omega p u^*_z d\Omega. \]  \hspace{1cm} \Delta u^*_p \text{ is determined as the surface displacement right after the onset of plastic deformation, which is detected by an inner level of iteration. The same minimization algorithm is applied to } V^*.

When the plastic deformation occurs in a very small fraction of the total contact area, i.e., the contact is still considered an elastic contact, the plastic dissipation energy during deformation is negligible. Then the approach simply impose an additional upper bound to the original pure elastic problem

\[ p_k \leq H_s, \ k = 1, \ldots, M \]  \hspace{1cm} (3.28)

It suggests that the maximum contact pressure can not exceed the hardness of the softer solid in contact. If the softer solid in contact is layered, its effective hardness, referred to
as $H_e$, can be decided as shown in Appendix 3. A contact considered whether elastic or plastic is decided by comparing the plastic yielding area $A_p$ with the values of 2 percent of the total real contact area $A_r$ (Greenwood and Williamson, 1966).

### 3.2.3 Minimization algorithms.

Now the problem becomes finding the minimum value of Eq. (3.10) - a standard quadratic function of the contact pressure - under the restriction of Eqs. (3.11) and (3.28). It can be solved by a bounded indefinite quadratic programming method, which combines coordinate searches and subspace minimization steps. In order to accelerate the computation, gradient information of the quadratic function is usually taken into account rather than relying solely on function evaluation. Methodologies based on the quadratic programming method were first developed to solve the 3-D homogeneous multiple asperity contact (Tian and Bhushan, 1996a; Stanley and Kato, 1997; Ai and Sawamiphakdi, 1999), where the influence coefficients corresponding to homogenous cases are substituted into Eq. (3.10). Tian and Bhushan (1996a) obtained the minimum by using the Newton method to search for a zero of the gradient, i.e.,

$$\nabla V_p^* = C^{\alpha \omega} p - u_c^* = 0.$$  

The solution is obtained as $p = C^{\alpha \omega \cdot l} \cdot u_c^*$, which is mathematically equivalent to the matrix inversion method. Even though the quadratic function itself is not evaluated in the minimization process, no non-converge problem was reported due to the relative simplicity of the homogenous case. Stanley and Kato (1997) applied a simplified steepest descent method to incorporate gradient information, and to speedup the minimization procedure. Getting started from a random initial value of pressure $p_0$, which satisfies

$$\int p \, dA = W,$$

iteration of $p_{i+1} = p_i - \nabla V_p^*$, repeated until no negative pressure occurs. The
gradient is calculated as $\nabla V_p^* = C_{uz} \cdot p - u_z^*$. A uniform shift of $p$, up and down is performed to satisfy $\int p dA = W$ so that the shape of $p$, does not change. A truncation of negative pressures is performed in each iteration to satisfy the requirement of non-negative pressures. The steepest decent direction may not, in general, leads to the minimum and therefore does not guarantee converge. This is because the gradient has not been appropriately scaled to assure the energy decreases for each iteration (Press et al., 1999). Moreover, since the quadratic function has never been evaluated, the initial guess of the prescribed displacements results in a unstable formulation and the uniqueness of the solution is not guaranteed. Ai and Sawamiphakdi (1999) applied the Fletcher-Reeves version of CGM to search for the minimum value, and introduced a penalty function to remove the constraints of $\int p dA = W$ and $p \geq 0$. The gradient is calculated as $\nabla V_p^* = C_{uz} \cdot p - u_z^*$. An minimization of the quadratic function at each step is performed along the calculated gradient subjected to boundary condition. The iteration is continued until the error is less than the convergence tolerance. Since the quadratic function is evaluated in the minimization step, the uniqueness of the solution is guaranteed provided CGM converge. However, the robustness of CGM needs to be enhanced by employing the restart procedure (Powell, 1977).

Peng and Bhushan (2001a) applied the Broyden-Fletcher-Goldfarb-Shanno (BFGS) version of quasi-Newton method to solve the 3-D layered multiple asperity contact. The goal of quasi-Newton method is same as that of CGM: to accumulate information from successive line minimizations so that $M$ such line minimizations lead to the exact minimum of a quadratic form in $M$ dimension. With a trivial disadvantage of
requiring a larger storage than CGM, quasi-Newton is widely developed to a greater level of sophistication on issues such as the minimization with round-off error, therefore is feasible for 3-D rough surfaces contact. The basic idea of quasi-Newton method is to build up, iteratively, a good approximation to the inverse influence coefficient matrix $C_{uZ}^{-1}$. It starts out with constructing a positive definite, symmetric matrix $H_0$ to approximate $C_{uZ}^{-1}$ and builds up a sequence of the approximated matrices $H_i$ in such a way that the matrix remains positive definite and symmetric (Press et al., 1999). Getting started from an arbitrary initial point far from the minimum, this method guarantees that one always moves in a downhill direction and thus generate a unique and true solution.

Close to the minimum, the updating $H_i$ will approach $C_{uZ}^{-1}$ with the guarantee of quadratic convergence.

Given the relative rigid approach $\delta$, $u_z^*$ is obtained as the geometrical interference. $H_0$ is initialized as a unit matrix, $p_0$ as a non-negative random vector satisfying $C_{uZ}p_0 \geq u_z^*$, and $d_0$ as $-\nabla V_{p_0}^* = u_z^* - C_{uZ} \cdot p_0$. The quasi-Newton method is iterated with the following recurrence formulation until the error is less than the convergence tolerance,

\[
\begin{align*}
p_{i+1} &= \min_{V_{p_{i+1}}} \nabla V_{p_{i+1}}^* \text{ along direction } d_i, \quad V_{p_{i+1}}^* = \frac{1}{2} p_{i+1}^T C_{uZ} p_{i+1} - p_{i+1}^T u_z^* \\
u &= \frac{(p_{i+1} - p_i) \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)}{(p_{i+1} - p_i) \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)} - \frac{H_i \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)}{(\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)} \cdot H_i \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*) \\
H_{i+1} &= H_i + \frac{(p_{i+1} - p_i) \otimes (p_{i+1} - p_i)}{(p_{i+1} - p_i) \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)} \cdot [H_i \cdot (\nabla f_{p_{i+1}} - \nabla f_{p_i})] \otimes [H_i \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)] \\
&\quad + [(\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*) \cdot H_i \cdot (\nabla V_{p_{i+1}}^* - \nabla V_{p_i}^*)]u \otimes u \\
d_{i+1} &= -H_{i+1} \nabla V_{p_{i+1}}^* \\
&\quad (3.29)
\end{align*}
\]
The requirements $p_{i+1} \geq 0$ and $u_{z} \geq u_{z}^*$ are integrated in the minimization step along
direction $d_i$. The gradient is calculated as $\nabla V_{p_i}^* = C^{u_x} \cdot p_i - u_{z}^*$. 

It is still possible round-off errors may cause the matrix $H_i$ to become nearly
singular or non-positive definite. This could be serious since the supposed search
directions might then not lead downhill, and because nearly singular $H_i$ tends to give
subsequent $H_{i+1}$ that is also nearly singular. An advanced implementation of quasi-
Newton method is to build up an approximation $H_i$ to $C^{u_x}$ instead of $C^{u_x^{-1}}$. A triangular
decomposition of $H_i$ is stored instead of $H_i$ itself and the updating formulation can be
arranged to guarantee that the matrix remains positive definite and nonsingular, even with
finite round-off.

The main feature of the quasi-Newton method is that the unique pressure
distribution corresponding to a given rigid approach $\delta$ is obtained directly from a single
minimization process. Moreover, since the quadratic termination property guarantees that
the solution converges in $M$ steps, the unique pressure distribution will be found in at
most $M$ iterations, which makes this model feasible for 3-D rough surfaces contact
problems with large number of contact points.

3.2.4 Meniscus effect.

The presence of the thin-film liquid at the contact interface, i.e., wet contact,
causes the formation of menisci due to surface energy effects, e.g., the meniscus rings
around contacting asperities and meniscus bridges connecting the non-contacting
asperities touching the liquid. The source of liquid can be either a thin lubricant film
present on the surface or condensed water. An attractive meniscus force arises from
negative Laplace pressure inside the curved meniscus as a result of surface tension (Derjaguin et al., 1987; Adamson, 1990; Israelachvili, 1992). The meniscus force pulls the contacting bodies even further and results in a higher contact normal load. Several statistical models have been developed to predict meniscus forces developed at a wet interface (Li and Talke 1990; Tian and Matsudaira, 1993; Gao et al., 1995; Bhushan et al., 1998). These analyses reveals insights into stiction problem.

Tian and Bhushan (1996b) developed a numerical model to study the micro-meniscus effect of an ultra-thin liquid film on the static friction. Peng and Bhushan (2001a) adopted this model to analyze the wet contact of layered rough surfaces. As shown in Fig. 3.11, for a given relative humidity, a liquid film forms over the deformed rough surfaces with the mean liquid thickness $h_f$ and mean meniscus height $h_m$. The areas where asperities of the composite rough surface touch the liquid are referred to as wetted area $A_w$. Cross cut areas $A_c$ are obtained at the given mean meniscus height $h_m$. $A_c^*$ are those area islands in $A_c$ that overlap the wetted area $A_w$. The total projected meniscus area $A_m$ is then determined by subtracting the real area of contact $A_r$ from $A_c^*$. Applying the extended first principle of the micro-meniscus theory, the meniscus force is given by:

$$F_m = \gamma_l (\cos \theta_1 + \cos \theta_2) A_m$$

(3.30)

where $\gamma_l$ is the surface tension of the liquid, $\theta_1$ and $\theta_2$ are the contact angles of the liquid.
Figure 3.11  Schematic of a smooth surface in contact with a composite rough surface in the presence of a liquid film

3.2.5 Program flowchart.

Figure 3.12 shows the flowchart of the 3-D layered rough surfaces contact model. In the preprocessing phase, normal loading $W$, coefficient of friction $\mu$, and layer thickness $h$ are initialized. The 3-D surface profile and physical properties of each contact surface are read in. Influence coefficient matrices corresponding to each contact surface are constructed. FFT subroutine is used here to speed up the computation. Since the contact between a rigid flat indenter and an elastic rough-surface is simple and
Figure 3.12 Flowchart of a 3-D layered rough surfaces contact model based on a variational principle and FFT

INPUT surface profiles $z_0(x, y)$ and $z_1(x, y)$, friction coefficient $\mu$, layer thickness $h$, normal load $W$, $E_0$, $E_1$, $E_2$, $H_0$, $H_1$, $H_2$, $\nu_0$, $\nu_1$, $\nu_2$

Compute influence coefficient matrix $C^*(x, y, z)$ with Fast Fourier Transformation

Issue a relative rigid approach $\delta$

Identify contacts

No

Yes

Determine prescribed displacements $u^*_c$

Compute pressure distribution $p_k$ with quasi-Newton method subjected to $H_s \geq p_k \geq 0$

No

$W = \sum p_k$?

Yes

Compute deformed surface
Determine the project meniscus area
Computer meniscus force $F_m$

Replace $W$ with $W + F_m$

No

$F_m/W$ is small?

Yes

OUTPUT $A_r/A_n$, $p_{max}/E_2$, $F_m/W$

$\sigma_{zz}$, $\sqrt{J_2}$, $\sigma_t$
straightforward, it is convenient to transform the contact of two elastic rough surfaces to a rigid flat surface in contact with a composite rough surface. The microgeometry of the composite rough surface is the sum of the two original surfaces (for the proof see Appendix 4). For two homogenous elastic surfaces in contact, Young's modulus of composite rough surface $E^*$ is given by

$$\frac{1}{E^*} = \frac{1 - \nu_0^2}{E_0} + \frac{1 - \nu_1^2}{E_1},$$

(3.31)

where the subscripts 0 and 1 denote the two contacting surfaces, respectively. For layered surfaces in contact, instead of calculating $E^*$, the influence coefficients of the composite rough surface are calculated as the sum of the influence coefficients of two contacting surfaces, i.e.,

$$C^*(x, y, z) = C_0(x, y, z) + C_1(x, y, z).$$

(3.32)

After preprocessing phase, the program proceeds by applying the upper body a series of vertical rigid approach until contact is detected. The prescribed z-direction surface displacements are then determined by geometrical interference. The pressure distribution is solved iteratively with the quasi-Newton method subroutine. If the sum of the calculated pressures does not agree with the prescribed normal load, a further vertical rigid approach is applied to the upper body until agreement is obtained. The meniscus force corresponding to the deformed surface is calculated and considered as a variance of the normal load. The varied normal load is feed back into the positive loop until the steady state is achieved.

Upon convergence, the contact pressure distribution is obtained and the postprocessing phase starts. The maximum pressure $p_{\text{max}}$ is pointed out and compared
with $H_s$ to determine the onset of plastic yielding. The real area of contact $A_r$ is the sum of all the patches with non-zero contact pressure. Fractional contact area, defined as the ratio of real to nominal contact area, represents the relative contact area independent of the nominal contact area $A_n$. The relative meniscus force, defined as the ratio of meniscus force to normal load, represents the relative meniscus effect.

All stress components are computed as the matrix product of the corresponding stress influence matrices and the pressure vector. The shear stress distributions are of special interest since the interfacial shear stress is one cause of the delamination in layered solids (shear rupture). $\sigma_{xz}$ component is selected to represent the shear stress since the default friction is along the positive x direction.

The von Mises equivalent stresses $\sqrt{J_2}$ are determined as:

$$ \sqrt{J_2} = \left( \frac{1}{6} \left[ (\sigma_{xx} - \sigma_{yy})^2 + (\sigma_{yy} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{xx})^2 \right] + \sigma_{xy}^2 + \sigma_{yz}^2 + \sigma_{xz}^2 \right)^{\frac{1}{2}} \tag{3.33} $$

They are used to decide the plastic yielding zone, which is the source of residual stresses or micro cracking, and an important factor in predicting the tendency of the contact solids to generate debris. For ductile materials, according to the von Mises criterion, plastic yielding initiates as the maximum $\sqrt{J_2}$ exceeds the yield stress in pure shear, referred as to $k$. $k$ equals $1/\sqrt{3}$ times the yield stress in simple tension $Y$; for metal $Y \approx H / 3$ thus $k \approx H / (3 \sqrt{3})$, where $H$ is the hardness.

The principal tensile stresses are determined as:

$$ \sigma_t = \frac{\sigma_{xx} + \sigma_{yy}}{2} + \frac{1}{2} \sqrt{(\sigma_{xx} - \sigma_{yy})^2 + 4\sigma_{xy}^2} \tag{3.34} $$
Brittle materials subject to high principal tensile stresses are liable to surface cracking or delamination at the interface.

High-performance computing system featured with parallel computing (SGI Origin 2000) is used to speedup both the FFT and the quasi-Newton computation. For an average number of contact points (~200), the CPU time to solve the contact problem and calculate the stress fields is less than 10 minutes.

3.2.6 Quasi-sliding contact of two layered nominally flat rough surfaces

A 3-D quasi-sliding contact analysis of two layered elastic perfectly-plastic nominally flat rough surface in both dry and wet conditions was performed by Peng and Bhushan (2001a, b, 2002) with the quasi-Newton method. A schematic of a typical rough surfaces contact is shown in Fig. 3.5. The rough surface is either a measured real surface or a computer generated surface. Whitehouse and Archard (1970) regarded the profile of a random surface as a random signal represented by a height distribution \( z(x, y) \) and an autocorrelation function

\[
R(\Delta x, \Delta y) = \lim_{L \to \infty} \frac{1}{L^2} \int_{-L}^{L} \int_{-L}^{L} z(x, y) \cdot z(x + \Delta x, y + \Delta y) dx dy.
\]

Engineering surfaces usually have a Gaussian distribution of \( z(x, y) \) (skewness \( Sk = 0 \) and kurtosis \( K = 3 \)) and an exponential autocorrelation function, i.e.,

\[
R(\Delta x, \Delta y) = \sigma^2 \exp\left[-2.3 \left( \frac{\Delta x}{\beta_x^*} \right)^2 + \left( \frac{\Delta y}{\beta_y^*} \right)^2 \right],
\]  

(3.35)

where \( \sigma \) is the standard deviation of \( z(x, y) \) and \( \beta_x^*, \beta_y^* \) are the correlation lengths of surface profiles in the x and y directions. Whitehouse and Archard (1970) further indicated that all features of such engineering surfaces are represented by \( \sigma \) and \( \beta_x^*, \beta_y^* \).

\( \beta_x^* = \beta_y^* \) represents an isotropic surface roughness and \( \beta_x^* \neq \beta_y^* \) represents an anisotropic
one. $\beta^*$ is viewed as a measure of randomness, i.e., how quickly the random event decays. The degree of randomness of a rough surface increases with $\beta^*$. Here the composite rough surface after the preprocessing is assumed following an isotropic Gaussian distribution. A two-dimensional digital filter technique is applied to generate this composite rough surface in the computer. This technique was developed by Hu and Tonder (1992) to generate Gaussian random surfaces with expected $\sigma$ and $\beta^*$ and extended by Chilamakuri and Bhushan (1998) to generate non-Gaussian surfaces. Two sample surfaces are shown in Fig. 3.13.

The nominal contact area $A_n$ and the patch size are chosen to match the scan window size and sampling interval of various contact and non-contact profilers, e.g., Atomic Force microscopy (AFM). Here $A_n$ is taken as $20 \times 20 \, \mu m^2$ and discretized into $256 \times 256$ patches ($0.078 \times 0.078 \, \mu m^2$ each). The indenter is assumed to be rigid with $E_0 = \infty$ and $H_0 = \infty$ after preprocessing. $E_2$ is taken as 100 GPa, while $E_1$ varies with the ratio $E_1 / E_2$. The effective hardness $H_e$ of the equivalent layered rough surface is taken as $0.05 \, E_1$. Since the indenter is rigid, $H_e$ is the hardness of the softer solid in contact $H_s$.

The normal load $W$ is represented by the nominal pressure $p_n = W / A_n$. Here $p_n$ is taken as $4 \times 10^{-6} \, E_2$. For the wet contact analyses, a water film is assumed to exist with $\gamma_f = 7.275 \times 10^2 \, N \, m^{-1}$ and $\theta_f = \theta_2 = 60^\circ$. The film thickness $h_f$ and meniscus height $h_m$ are taken as the same value of $\sigma$.

The contact parameters such as surface roughness, material properties and applied normal load are chosen as typical values of magnetic storage devices. However, the generic trends can be used in variety of applications.
Figure 3.13 Profiles of two computer generated rough surfaces. The lower surface \((\sigma = 2 \text{ nm}, \beta^* = 1 \text{ } \mu\text{m})\) is magnified twice and cut off at the \(20 \times 20 \text{ } \mu\text{m}^2\) margin to obtain the upper surface \((\sigma = 1 \text{ nm}, \beta^* = 0.5 \text{ } \mu\text{m})\)

### 3.2.6.1 Normalization scheme.

To apply the results on various combinations of surface roughness, material properties, and normal load, normalization is the essential step. Similar to the single asperity contact case, the contact variables are readily normalized to their corresponding values at the homogenous case. For example, \(A_r\) is normalized to \(A_{r0}\), \(p_{\text{max}}\) to \(p_{\text{max0}}\), etc.

Similar to single asperity contact case discussed in Secs. 2.1.1.1 and 2.2.1.1, the layer
thickness $h$ is normalized to $\sqrt{A_{r_0}}$ but only applies to the specific case. An example will be presented later in Fig. 3.20.

A more general normalization scheme is laid out as follows. An increase of $p_n/E_2$ will bring more asperities into contact to support the increased normal load, i.e., an increase of $A_r/A_n$. A linear relationship between $A_r/A_n$ and $p_n/E_2$ is shown in Fig. 3.14. This linear relationship can be traced back to the linear stress-strain material and is commonly observed in engineering practice: friction and wear increases with the normal load. On the other hand, it is also known that a rougher surface (higher $\sigma$) brings fewer asperities into contact and reduces $A_r/A_n$, an opposite to the increase of $p_n/E_2$. The pressure profile and stress distributions at $\sigma = 1$ nm and $p_n/E_2 = 4 \times 10^6$ (Fig. 3.15) are compared with those at $\sigma = 0.1$ nm and $p_n/E_2 = 4 \times 10^7$ (Fig. 3.16). By comparing stress distributions in Fig. 3.15 with corresponding ones in Fig. 3.16, it is found that the changes of the normal load ($p_n/E_2$) and the roughness ($\sigma$) by the same factor do not affect the pressure profiles and stress distributions, except only the stresses’ magnitudes are changed by the same factor. Therefore, $A_r/A_n$ can be normalized as $\frac{A_r}{A_n} / \left( \frac{p_n}{E_2} \frac{1}{\sigma} \right)$.

Since the meniscus force $F_m$ is mainly decided by the deformed surface profile, it is considered proportional to $A_r/A_n$ accordingly for the sake of simplification. The meniscus force $F_m$ is then normalized as $F_m / \left( \frac{p_n}{E_2} \frac{1}{\sigma} \right)$. The average pressure $p_r$, defined as $p \frac{A_r}{A_n}$, is proportional to $E_2\sigma$ and is practically independent of $p_n$ and $A_n$. This empirical normalization scheme provides a good match with Onions and Archard (1973)’s analytical one.
Figure 3.14 Variation of $A_r/A_n$, $p_{max}/E_2$, and $F_m/W$ with $p_n/E_2$ at various $E_1/E_2$ [Peng and Bhushan, 2001a]
\[ \sigma = 1 \text{ nm}, \, \beta^* = 0.5 \mu \text{m}, \, E_2 = 100 \text{ GPa}, \, H_1 / E_1 = 0.05, \, p_n / E_2 = 4 \times 10^{-6}, \, h = 10^3 \text{ nm}, \, \mu = 0.5 \]

Pressure distribution at \( z = 0 \) \( \sqrt{J_2} \) at \( z = 0 \)

\[ \sqrt{J_2} \text{ at } y = y_{\text{max}} \]

\[ E_1 / E_2 = 0.5 \]

\[ \sigma_t \text{ at } y = y'_{\text{max}} \]

\[ \sigma_{xz} \text{ at } y = y''_{\text{max}} \]

\[ \text{Max } 12.78 (y_{\text{max}} = 4.6) \]

\[ \text{Max } 12.78 (y_{\text{max}} = 4.6) \]

\[ \text{Max } 12.81 (y'_{\text{max}} = 4.6) \]

\[ \text{Max } 12.69 (y''_{\text{max}} = 4.6) \]

\[ \text{Max } 13.34 (y_{\text{max}} = 4.6) \]

\[ \text{Max } 13.34 (y_{\text{max}} = 4.6) \]

\[ \text{Max } 13.31 (y'_{\text{max}} = 4.6) \]

\[ \text{Max } 13.26 (y''_{\text{max}} = 4.6) \]

\[ \text{Max } 13.82 (y_{\text{max}} = 18.1) \]

\[ \text{Max } 13.85 (y'_{\text{max}} = 10.9) \]

\[ \text{Max } 13.74 (y''_{\text{max}} = 18.1) \]

Figure 3.15 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max \( \sqrt{J_2} \) plane, principal tensile stresses on the max \( \sigma_t \) plane and shear stresses on the max \( \sigma_{xz} \) plane at various \( E_1 / E_2 \), with \( \sigma = 1 \text{ nm}, \, \beta^* = 0.5 \mu \text{m}, \, p_n / E_2 = 4 \times 10^{-6}, \, h = 1 \mu \text{m} \). All contours are plotted after taking natural log values of the calculated stresses (in kPa). Negative values of \( \sigma_t \) and \( \sigma_{xz} \) in the plot represent the compressive stress and shear stress along negative x direction, respectively. [Peng and Bhushan, 2002]
\( \sigma = 0.1 \text{ nm}, \beta^* = 0.5 \mu\text{m}, E_2 = 100 \text{ GPa}, H_f/E_1 = 0.05, p_n/E_2 = 4 \times 10^{-7}, h = 10^3 \text{ nm}, \mu = 0.5 \)

Pressure distribution at \( z = 0 \)
\( \sqrt{J_2} \) at \( z = 0 \)
\( J_2 \) at \( y = y_{\text{max}} \)
\( \sigma_{r} \) at \( y = y'_{\text{max}} \)
\( \sigma_{xx} \) at \( y = y''_{\text{max}} \)

\( E_1/E_2 = 1 \)
\( E_1/E_2 = 2 \)

Figure 3.16 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max \( \sqrt{J_2} \) plane, principal tensile stresses on the max \( \sigma_r \) plane and shear stresses on the max \( \sigma_{xz} \) plane at various \( E_1/E_2 \), with \( \sigma = 0.1 \text{ nm}, \beta^* = 0.5 \mu\text{m}, p_n/E_2 = 4 \times 10^{-7}, h = 1 \mu\text{m} \) [Peng and Bhushan, 2002]
The stresses $\sigma_{ij}$, where the indices $i$ and $j$ range over $x$, $y$ and $z$ respectively, increase linearly with $p_n$ at constant $A_r/A_n$, e.g., in case of single patch contact. However, the increase of $\sigma_{ij}$ is much moderated in case of multiple asperity contact, since $A_r/A_n$ increases with $p_n$ at the same time and consequently reduces $\sigma_{ij}$. A normalization scheme is then given as $\sigma_{ij} \propto \frac{p_n}{(A_r/A_n)^m} \propto p_n^{1-m} E^m \sigma^m$, where $A_r/A_n$ is reduced to a power of $m (m < 1)$ considering the different weights of $A_r/A_n$ and $p_n$. Based on trial and error, $m$ is found to be about $2/3$. The stress components $\sigma_{ij}$ are then approximately normalized as $\sigma_{ij} / (p_n^{1/3} E_2^{2/3} \sigma^{2/3})$. $p_{\text{max}} / E_2$ is normalized to $\left(\frac{p_n}{E_2} \sigma^2\right)^{1/3}$ accordingly.

All direct geometric variables should be normalized to one among themselves, thus independent of initial and loading conditions. For example, in case of single asperity contact, all direct geometric variables are normalized to a lateral length-variable, i.e., patch width in single patch contact and sphere radius in spherical indentation. In case of multiple asperity contact, $\beta^*$ is chosen as the relevant lateral variable corresponding to patch width and sphere radius. All length-variables are then normalized to $\beta^*$: layer thickness $h$ is normalized as $h/\beta^*$, standard deviation of the surface heights $\sigma$ as $\sigma/\beta^*$, and all the Cartesian coordinates $(x,y,z)$ as $(x,y,z)/\beta^*$. Verification results are shown in Figs. 3.25 and 3.27. The pressure profile and stress distributions at $\sigma = 1$ nm, $\beta^* = 0.5$ $\mu$m and $p_n/E_2 = 4 \times 10^{-6}$, $h = 1$ $\mu$m (Fig. 3.15) are compared with those at $\sigma = 2$ nm, $\beta^* = 1$ $\mu$m, $p_n/E_2 = 4 \times 10^{-6}$, $h = 2$ $\mu$m (Fig. 3.17). By comparing stress distributions in Fig. 3.15
$\sigma = 2 \text{ nm, } \beta^* = 1 \mu\text{m, } E_2 = 100 \text{ GPa, } H_i / E_1 = 0.05, \ p_n / E_2 = 4 \times 10^{-6}, \ h = 2 \mu\text{m, } \mu = 0.5$

Pressure distribution at $z = 0$  

Figure 3.17 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_t$ plane and shear stresses on the max $\sigma_{xz}$ plane at various $E_1 / E_2$, with $\sigma = 2 \text{ nm, } \beta^* = 1 \mu\text{m, } p_n / E_2 = 4 \times 10^{-6}, \ h = 2 \mu\text{m}$ [Peng and Bhushan, 2002]
with corresponding ones in Fig. 3.17, it is found that the changes of the layer thickness $h$ and the roughness ($\sigma$) with $\beta^*$ by the same factor do not affect the stresses’ magnitudes. The shape of each stress distribution remains the same except it is scaled up proportionally. Therefore, $\beta^*$ is a proper normalization factor for all direct geometric variables.

The final normalization scheme is

$$\frac{A_r}{A_n} \left( \frac{P_n}{E_2} \frac{\beta^*}{\sigma} \right), \frac{P_{\text{max}}}{E_2} \left( \frac{P_n}{E_2} \sigma^2 \right)^{\frac{1}{3}} \text{and } F_{\text{nn}} \left( \frac{P_n}{E_2} \right).$$

A further verification is performed by showing the linear relationship between dimensional data and their corresponding normalization factors in Fig. 3.18. All the data in Fig. 3.18 are the numerical contact analyses’ results listed in Table 3.1. As expected, $A_r / A_n$ is proportional to $\frac{P_n}{E_2} \frac{\beta^*}{\sigma}$, while $\frac{P_{\text{max}}}{E_2}$ is proportional to $\left( \frac{P_n}{E_2} \sigma^2 \right)^{\frac{1}{3}}$ until it reaches the hardness of the softer solid in contact $H_s$. After that, plastic deformation occurs and $\frac{P_{\text{max}}}{E_2}$ remains constant.

A linear relationship between the contact statistics and the stiffness-ratio of the layer to the substrate ($E_1 / E_2$) is observed in Fig. 3.18 by comparing the contact statistics at various $E_1 / E_2$. For example, comparing to the value in case of $E_1 / E_2 = 1$ (homogenous case), $A_r / A_n$ is approximately doubled in case of $E_1 / E_2 = 0.5$ and reduced by half in case of $E_1 / E_2 = 2$. $p_n / E_2$ has exact opposite relationship with $E_1 / E_2$ provided the hardness ratio of the layer to the substrate ($H_1 / H_2$) is the same as $E_1 / E_2$. Therefore, the contact statistics at arbitrary stiffness- and hardness- ratio of the layer to the substrate can be obtained by adjusting the homogenous solution accordingly.
Figure 3.18 Variation of \( \frac{A_r}{A_n} \) with \( \frac{p_n}{E_2} \) and \( \frac{p_{\text{max}}}{E_2} \) with \( \left[ \frac{p_n}{E_2} \left( \frac{\sigma}{\beta^*} \right)^2 \right] \times 10^3 \) at various \( E_1 / E_2 \). These values are independent of \( \beta^* \) and \( \mu \). [Peng and Bhushan, 2002]
<table>
<thead>
<tr>
<th>$E_1/E_2$</th>
<th>$P_n/E_2$</th>
<th>$\sigma$ (nm)</th>
<th>Fractional contact area (%)</th>
<th>Maximum pressure /$E_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>4.00E-07</td>
<td>0.1</td>
<td>0.2261</td>
<td>0.0007</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-07</td>
<td>1</td>
<td>0.0246</td>
<td>0.0039</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-07</td>
<td>10</td>
<td>0.0046</td>
<td>0.0227</td>
</tr>
<tr>
<td>0.5</td>
<td>2.00E-06</td>
<td>0.5</td>
<td>0.2322</td>
<td>0.0033</td>
</tr>
<tr>
<td>0.5</td>
<td>2.00E-06</td>
<td>2</td>
<td>0.0615</td>
<td>0.0095</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-06</td>
<td>0.5</td>
<td>0.4813</td>
<td>0.0036</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-06</td>
<td>1</td>
<td>0.2322</td>
<td>0.0066</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-06</td>
<td>2</td>
<td>0.1230</td>
<td>0.0115</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-06</td>
<td>10</td>
<td>0.0246</td>
<td>0.0250</td>
</tr>
<tr>
<td>0.5</td>
<td>1.00E-05</td>
<td>10</td>
<td>0.0708</td>
<td>0.0250</td>
</tr>
<tr>
<td>0.5</td>
<td>4.00E-05</td>
<td>10</td>
<td>0.2753</td>
<td>0.0250</td>
</tr>
<tr>
<td>1</td>
<td>4.00E-07</td>
<td>0.1</td>
<td>0.1215</td>
<td>0.0012</td>
</tr>
<tr>
<td>1</td>
<td>4.00E-07</td>
<td>1</td>
<td>0.0200</td>
<td>0.0065</td>
</tr>
<tr>
<td>1</td>
<td>4.00E-07</td>
<td>10</td>
<td>0.0015</td>
<td>0.0260</td>
</tr>
<tr>
<td>1</td>
<td>1.00E-06</td>
<td>10</td>
<td>0.0077</td>
<td>0.0490</td>
</tr>
<tr>
<td>1</td>
<td>2.00E-06</td>
<td>0.5</td>
<td>0.1215</td>
<td>0.0058</td>
</tr>
<tr>
<td>1</td>
<td>2.00E-06</td>
<td>2</td>
<td>0.0292</td>
<td>0.0163</td>
</tr>
</tbody>
</table>

Table 3.1 Variation of fractional contact area, maximum pressure with $\sigma$ and $P_n/E_2$ at various values of $E_1/E_2$. Here $\beta^* = 0.5 \mu m$, $E_2 = 100$ GPa, $H_s/E_1 = 0.05$, $h = 1 \mu m$. (CONTINUED)
TABLE 3.1: CONTINUED

<table>
<thead>
<tr>
<th>$E_1$ / $E_2$</th>
<th>$p_\alpha$ / $E_2$</th>
<th>$\sigma$ (nm)</th>
<th>Fractional contact area (%)</th>
<th>Maximum pressure /$E_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.00E-06</td>
<td>0.5</td>
<td>0.2338</td>
<td>0.0066</td>
</tr>
<tr>
<td>1</td>
<td>4.00E-06</td>
<td>1</td>
<td>0.1215</td>
<td>0.0115</td>
</tr>
<tr>
<td>1</td>
<td>4.00E-06</td>
<td>2</td>
<td>0.0584</td>
<td>0.0188</td>
</tr>
<tr>
<td>1</td>
<td>4.00E-06</td>
<td>10</td>
<td>0.0200</td>
<td>0.0500</td>
</tr>
<tr>
<td>1</td>
<td>1.00E-05</td>
<td>10</td>
<td>0.0308</td>
<td>0.0500</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-07</td>
<td>0.1</td>
<td>0.0631</td>
<td>0.0019</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-07</td>
<td>1</td>
<td>0.0108</td>
<td>0.0111</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-07</td>
<td>10</td>
<td>0.0015</td>
<td>0.0260</td>
</tr>
<tr>
<td>2</td>
<td>1.00E-06</td>
<td>10</td>
<td>0.0015</td>
<td>0.0650</td>
</tr>
<tr>
<td>2</td>
<td>2.00E-06</td>
<td>0.5</td>
<td>0.0631</td>
<td>0.0093</td>
</tr>
<tr>
<td>2</td>
<td>2.00E-06</td>
<td>2</td>
<td>0.0215</td>
<td>0.0272</td>
</tr>
<tr>
<td>2</td>
<td>2.00E-06</td>
<td>10</td>
<td>0.0062</td>
<td>0.0977</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-06</td>
<td>0.5</td>
<td>0.1277</td>
<td>0.0116</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-06</td>
<td>1</td>
<td>0.0631</td>
<td>0.0185</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-06</td>
<td>2</td>
<td>0.0308</td>
<td>0.0323</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-06</td>
<td>10</td>
<td>0.0108</td>
<td>0.1000</td>
</tr>
<tr>
<td>2</td>
<td>4.00E-05</td>
<td>10</td>
<td>0.0708</td>
<td>0.1000</td>
</tr>
</tbody>
</table>
Since the changes of fractional contact area, $p_{\text{max}}/E_2$, and relative meniscus force due to friction are negligible provided the coefficient of friction is not very large, the normalization formulation is valid for both normal and quasi-sliding contact.

3.2.6.2 Layer effect.

The layer effects on the pressure profile and contact statistics are shown in Figs. 3.19 and 3.20. In case of $E_1/E_2 \neq 1$, as expected, the contact converges to the homogeneous solution at $E = E_1$ with $h/\beta \to 0$, and to the homogenous solution at $E = E_2$ with $h/\beta \to \infty$ accordingly. For a stiffer layer ($E_1/E_2 > 1$), the fractional contact area decreases compared to the homogeneous case. From Tabor’s classical theory of adhesion, the friction due to adhesion is proportional to the real area of contact (Bhushan, 1999a). Therefore, a stiffer layer is beneficial in reducing the adhesional friction. The probability of wear-particle formation on a stiffer layer also decreases with smaller real area of contact (Bhushan, 1999a). The decrease of relative meniscus force on a stiffer layer is beneficial in stiction prevention. In contrast, a more compliant layer ($E_1/E_2 < 1$) increases the real area of contact, which consequently results in higher adhesional friction, stiction, and wear-particle formation. A linear relationship between the contact statistics and the stiffness-ratio of the layer to the substrate ($E_1/E_2$) is also observed in Fig. 3.20.

The layer effects on the distribution of surface and subsurface stresses are shown in Fig. 3.21. Dimensional data are shown here since unlike the contact statistics, the contours are valid only for this specific case. $\sqrt{J_2}$, $\sigma_t$, and $\sigma_{xz}$ are plotted with taking natural log values of the calculated stresses in kPa, since these stresses are of a large
Figure 3.19 Profiles of (a) a composite layered rough surface and (b) corresponding contact pressures at various $h$ and $E_1/E_2$. These values are independent of $\mu$. 

93
\[ \frac{H_s}{E_1} = 0.05, \frac{h_f}{\sigma} = 1, \frac{h_m}{\sigma} = 1, A_n = 16 \beta^2 \]

- \( E_1 / E_2 = 0.5 \)
- \( E_1 / E_2 = 1 \)
- \( E_1 / E_2 = 2 \)

The normalization of \( h \) with \( \sqrt{A_n} \) only applies to the case of \( A_n / A_0 = 0.12 \% \) or \( (p_n \beta^*) / (E_2 \sigma) = 2 \times 10^{-3} \)

Figure 3.20  Variation of \( A_r/A_n \), \( p_{\text{max}}/E_2 \), and \( F_m/W \) with layer thickness \( h \) at various \( E_1/E_2 \). These values are independent of \( \mu \).
Figure 3.21 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_t$ plane and shear stresses on the max $\sigma_{xz}$ plane at various layer thickness $h$ with (a) $E_1/E_2 = 2$ and (b) $E_1/E_2 = 0$, $\mu = 0$. [Peng and Bhushan, 2001b] (CONTINUED)
Figure 3.21: CONTINUED

\[ \sigma = 1 \text{ nm}, \beta = 0.5 \mu\text{m}, E_2 = 100 \text{ GPa}, E_1 / E_2 = 0.5, \ H_1 / E_1 = 0.05, \ p_n / E_2 = 4 \times 10^{-6} \]

Pressure distribution at \( z = 0 \):
\[ \sqrt{J_2} \text{ at } z = 0 \]
\[ \sqrt{J_2} \text{ at } y = y_{\text{max}} \]
\[ \sigma_t \text{ at } y = y'_{\text{max}} \]
\[ \sigma_{zz} \text{ at } y = y''_{\text{max}} \]

- \( h = 0 \text{ nm} \)
- \( h = 10 \text{ nm} \)
- \( h = 10^2 \text{ nm} \)
- \( h = 10^3 \text{ nm} \)
- \( h = \infty \)
range. There is no symmetrical plane to show a typical stress distribution, each individual plane is of unique stress distribution. Normally a “y = constant” plane is chosen to make contour plots for sub-surface stresses. Since location of maximum stress is of interest in failure prevention, the “y = constant” plane is set as the plane including the maximum stress.

The von Mises stress distributions on the surface are shown in the plots labeled $\sqrt{J_2}$ at $z = 0$, with the contact pressure profiles shown aside as a reference. Since each contact pressure produces a high stress zone around it, concentrated contours of $\sqrt{J_2}$ are observed and centered around each contact pressure. The sub-surface von Mises stress distributions are shown in the plots labeled $\sqrt{J_2}$ at $y = y_{\text{max}}$. A comparison between Figs. 3.21 (a) and 3.21 (b) shows that in a stiffer layer the maximum $\sqrt{J_2}$ is much higher than in a more compliant layer. If wear mechanisms caused by plastic deformation are the main concern, the layer hardness $H$ should be high enough to compensate the increase of maximum $\sqrt{J_2}$. For example, it is known that an increase in effective hardness and Young’s modulus of the disk structure improves its wear resistance.

The sub-surface principal tensile stress distributions are shown in the plots labeled $\sigma_t$ at $y = y'_{\text{max}}$. High $\sigma_t$ are located close to the surface and the maximum $\sigma_t$ always occurs on the surface. Tensile radial stresses make brittle materials susceptible to ring cracks and is the cause of surface brittle failure. Since materials in contact applications are often surface hardened, the surface could be brittle. A comparison between Figs. 3.21 (a) and 3.21 (b) shows that a more compliant layer reduces both the value of the
maximum $\sigma_t$ and the region of high $\sigma_n$, which consequently decreases the potential of crack failure.

The shear stress distributions are shown in the plots labeled $\sigma_{xz}$ at $y = y^{''}\text{max}$. The maximum $\sigma_{xz}$ always occurs on the surface. High $\sigma_{xz}$ are located near the surface and decays rapidly into the depth. For a thin layer, high $\sigma_{xz}$ may extend to the interface. In terms of adhesional failure during sliding contact, e.g., delamination, maintaining a low interfacial shear stress is important. A comparison between Figs. 3.31 (a) and 31 (b) shows that a more compliant layer will reduce the maximum value of $\sigma_{xz}$ and is beneficial in reducing the potential of delamination failure.

### 3.2.6.3 Friction effect.

The contact results shown in Fig. 3.21 are obtained at $\mu = 0$, i.e., a normal contact. Figure 3.22 shows the corresponding contact results at $\mu = 0.5$, i.e., a quasi-sliding contact. The pressure profiles are found unaltered by friction, so are the fractional contact area, $p_{\text{max}}/E_2$, and relative meniscus force. However, the friction effect on the subsurface stresses is marked: the maximum $\sigma_{xz}$ and $\sqrt{J_2}$ move closer to the layer surface with the increase of the shear force. The magnitudes of some stress components, such as shear stress $\sigma_{xz}$, increases with the shear force. The magnitudes of $\sigma_t$ and $\sqrt{J_2}$ increase in turn. As a result, the increase of friction increases the potential of wear mechanisms such as crack failure, delamination and residual stresses.
$\sigma = 1 \text{ nm, } \beta^* = 0.5 \mu m, E_2 = 100 \text{ GPa, } E_1/E_2 = 2, \ H_1/E_1 = 0.05, p_n/E_2 = 4 \times 10^{8}, \mu = 0.5$

Figure 3.22 Profiles of contact pressures, contours of von Mises stresses on the surface, von Mises stresses on the max $\sqrt{J_2}$ plane, principal tensile stresses on the max $\sigma_t$ plane and shear stresses on the max $\sigma_{xz}$ plane at various layer thickness $h$ with (a) $E_1/E_2 = 2$ and (b) $E_1/E_2 = 0.5$. $\mu = 0.5$. [Peng and Bhushan, 2002]

(CONTINUED)
Figure 3.22: CONTINUED

\[ \sigma = 1 \text{ nm}, \beta^* = 0.5 \text{ \(\mu\)m}, E_2 = 100 \text{ GPa}, E_1 / E_2 = 0.5, \ H_1 / E_1 = 0.05, \ p_n / E_2 = 4 \times 10^{-6}, \ \mu = 0.5 \]

Pressure distribution at \( z = 0 \) \[ \sqrt{J_2} \] at \( z = 0 \) \[ \sqrt{J_2} \] at \( y = y_{\text{max}} \) \[ \sigma_{zz} \] at \( y = y'_{\text{max}} \) \[ \sigma_{zz} \] at \( y = y''_{\text{max}} \]

- \( h = 0 \text{ nm} \)
- \( h = 10 \text{ nm} \)
- \( h = 10^2 \text{ nm} \)
- \( h = 10^3 \text{ nm} \)
- \( h = \infty \)
CHAPTER 4

APPLICATIONS

4.1 Magnetic Storage Devices.

Magnetic disk drives are used for data recording applications. The read-write operation is performed under steady conditions where a load-carrying film is formed to separate the slider from the disk, and only isolated asperity contacts may occur between them. Physical separation in today’s drives is about 15 nm. During the start-stop operations of the disk drive, physical contact occurs and friction is generated during the sliding motion. Thin magnetic films used in the construction of thin-film disks are metallic such as Co-Cr alloys, which are soft and have poor wear and corrosion resistance. Protective diamond-like carbon (DLC) coatings with a thin lubricant overlay are used to provide low friction, low wear and corrosion resistance (Bhushan, 1996a, 1999a, 1999c). A cross sectional schematic of a typical thin-film metal disk is shown in Fig. 4.1. The protective films must be as thin as possible (3 ~ 5 nm) in order to keep the head-disk spacing small, which is required for high-density recording. On the other hand, since the disk surface is kept rough to reduce friction and stiction, the protective films must have thickness homogeneity as high as possible to provide good protection. Therefore, there should be a critical thickness which is optimum for the overcoat system.
4.1.1 Head slider / disk interface (HDI) modeling.

The surface topographies of various disks were measured with contact and non-contact profilers by Poon and Bhushan (1995) and Bhushan (1999b). As shown in Fig. 4.2, measurements of $\sigma$ are made on magnetic disks and substrates by atomic force microscope (AFM), stylus profiler (SP) and non-contact optical profiler (NOP) with different scan sizes and sampling intervals (Poon and Bhushan, 1995). The scan sizes used in AFM measurements are 4, 8, 16, 32, 64 and up to maximum 100 $\mu$m and each measurement contained $256 \times 256$ data points. The scan size and sampling interval used in SP measurements are 80 $\mu$m and 0.02 $\mu$m, 400 $\mu$m and 0.2 $\mu$m, and 800 $\mu$m and 0.4 $\mu$m, respectively. The scan size and sampling interval used in NOP measurements are $256 \times 256 \mu m^2$ and 1 $\mu$m. For conventional magnetic disks, it is found that $\sigma$ of each sample disk surface initially increases with scan size and approaches a constant value.
when the scan size reaches to about 16 µm, which suggests a long-wavelength limit \( L = 16 \mu m \) of disk surfaces. Consequently the scan window size (profile lengths in the x and y directions) of this model is taken as \( 20 \times 20 \mu m^2 \) to cover the long wavelength. A suitable sampling interval is related to the waviness structure of the surface (Poon and Bhushan, 1995; Bhushan, 1999b) from the physics point of view, and also a compromise between the precision and the computation time from the mathematics point of view. Here \( 256 \times 256 \) sample data are taken in the \( 20 \times 20 \mu m^2 \) scan window.

Table 2 shows the physical properties and surface topography statistics of a conventional magnetic thin-film (metal) disk and an \( \text{Al}_2\text{O}_3\)-TiC slider. Roughness \( \sigma \) of
today’s disk and sliders are on the order of 1 nm. Since the major concern is the DLC overcoat effect on the disk tribological performance, underlying sublayers are simplified as a uniform substrate taking the properties of the magnetic film. A layered rough surface is then generated in the computer to resemble the composite rough surface of the disk and the slider. The roughness of the layered rough surface is first taken as an isotropic Gaussian rough surface ($\sigma = 1$ nm, $\beta_x^* = \beta_y^* = 0.5$ µm, skewness $Sk = 0$ and kurtosis $K = 3$), while the real surface roughness for a given head-disk interface is taken into account afterwards by applying the normalization scheme. The slider’s Young’s modulus $E_0$ is taken as 450 GPa and hardness $H_0$ as 23 GPa. The magnetic film’s Young’s modulus $E_2$ is taken as 200 GPa and hardness $H_2$ as 9 GPa. The DLC overcoat’s Young’s modulus $E_i$ varies with the ratio of $E_i / E_2$, which is taken as 0.75 and 1.75. Two additional ratios of $E_i / E_2$ (0.1, 10) are shown as extreme cases. The DLC overcoat’s hardness $H_i$ is taken as 0.1 $E_i$. All Poisson’s ratios are taken as 0.25. A 30% pico-slider with a nominal area of 0.47 mm$^2$ (contact area between the rails and the disk) and a normal load of 2.5g (24.5 mN) is considered, and corresponding nominal pressure $p_n$ is about 50 kPa assuming the normal load is uniformly distributed over the entire nominal contact area. The hardness of a layered solid is known to be independent of the substrate for indentation depth less than about 0.3 of the film thickness, after that the hardness varies with the presence of the substrate (Bhushan, 1996a). Since in all the cases considered here the vertical rigid approaches $\delta$ (will be shown in Fig. 4.3) are less than 0.3 of the corresponding DLC overcoat’s thickness, so are the indentation depths. The effective hardness of the
<table>
<thead>
<tr>
<th>Disk materials / slider</th>
<th>$E$ (GPa)</th>
<th>$\nu$</th>
<th>$H$ (GPa)</th>
<th>$\sigma$ (nm)</th>
<th>$\beta^*$ (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLC overcoat</td>
<td>140$^1$, 280$^1$</td>
<td>0.25$^1$</td>
<td>15, 24$^1$</td>
<td>2.1, 4, 6, 7.3$^2$</td>
<td>36, 3, 4$^2$</td>
</tr>
<tr>
<td>Co-Cr alloy magnetic coating</td>
<td>190$^2$</td>
<td>9$^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ni-P on Al-Mg alloy</td>
<td>130$^2$</td>
<td>0.36$^3$</td>
<td>7$^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Mg alloy</td>
<td>70$^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glass ceramics</td>
<td>150$^4$</td>
<td>0.25$^5$</td>
<td>13$^4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al$_2$O$_3$-TiC slider</td>
<td>450$^2$</td>
<td>0.23$^3$</td>
<td>23$^2$</td>
<td>1.5$^5$</td>
<td>20$^2$</td>
</tr>
</tbody>
</table>

$^1$ Bhushan (1999c); $^2$ Bhushan (1996a); $^3$ Chilamakuri and Bhushan (1998); $^4$ Bhushan et al. (1996); $^5$ Poon and Bhushan (1996).

Table 4.1 Typical physical properties and surface topography statistics of magnetic thin-film disks and sliders

magnetic disk $H_e$ is then taken as the DLC overcoat’s hardness $H_f$. A liquid film of perfluoropolyether lubricant with surface tension $\gamma_f = 0.025$ N/m and contact angle $\theta = 10^\circ$ is considered. The film thickness $h_f$ and meniscus height $h_m$ are taken as $2\sigma$.

**4.1.2 Calculation of critical thickness.**

In general with an increase of the layer thickness $h$, the layer effect increases and the substrate effect decreases. However, the increase of layer effect with $h$ is normally non-linear. There exists a transition zone of layer thickness decided by two threshold values of layer thickness. Inside the transition zone, the contact behavior is decided by both the layer and the substrate and thus very sensitive to the change of $h$; outside the
transition zone the contact is considered solely decided either by the layer or the substrate and thus insensitive to the change of $h$. These two threshold values are defined as ceiling / floor critical thickness, above / under that the layer / substrate effect becomes dominant and substrate / layer effect is negligible. The floor critical thickness represents a necessary thickness for layer to take effect and the ceiling critical thickness represents a sufficient thickness to fully achieve the effect of the layer. Therefore, a change of the layer thickness before / after it exceeds floor / ceiling value barely changes the contact performance. It is necessary to control the layer thickness in the transition zone in order to bring the layer into full play and ensure an optimum performance.

The floor / ceiling critical thickness $h^*$ is chosen as the thickness beyond / under which the increase of the overcoat thickness has no appreciable effect on the contact parameters $Y$, e.g., maximum deformation for single patch contact, contact radius for spherical contact, fractional contact area for rough surfaces contact. A definition is given as following:

$$\left| \frac{Y(h_{ceiling}^*) - Y(\infty)}{Y(\infty) - Y(0)} \right| \leq 5\% \quad \text{and} \quad \left| \frac{Y(h_{floor}^*) - Y(0)}{Y(\infty) - Y(0)} \right| \leq 5\% .$$  \hspace{1cm} (4.1)

4.1.3 Results and discussion.

Performing the normal contact analysis on HDI yields insight into the effect of DLC layer on the contact behavior of magnetic thin-film disks at HDI. Fig. 4.3 shows the variation of contact statistics with layer thickness $h$. The results corresponding to the basis case ($\sigma = 1$ nm) were obtained by Peng and Bhushan (2000b) applying the layered rough surfaces contact model. The other two data sets ($\sigma = 0.1, 10$ nm) are obtained by applying the normalization scheme. Minor adjustments may be necessary concerning the
Figure 4.3  Variation of \( A_r/A_n \), \( u_z \), \( p_{max}/E_2 \), and \( F_m/W \) with layer thickness \( h \) at various \( E_1/E_2 \) and (a) \( \sigma = 0.1 \) nm, (b) \( \sigma = 1 \) nm, and (c) \( \sigma = 10 \) nm, in the normal contact of a rigid head-slider on an elastic-perfectly plastic disk in the presence of a liquid film [Peng and Bhushan, 2000b]
effective hardness and corresponding plastic deformation. For example, in case of $\sigma = 10$ nm, $E_1/E_2 = 0.1$, and $h = 5$ nm, the rigid approach $\delta$ is about 3 nm so that the effective hardness $H_e$ can not be replaced by $H_1$ (Appendix 3). $H_e$ needs to be recalculated and so does $p_{\text{max}}/E_2$. Nevertheless, the main trends in these cases are the same. The contact performance of the magnetic thin-film disk is highly dependent on the following factors: surface roughness of DLC overcoat; DLC overcoat thickness; stiffness- and hardness- ratio of the DLC overcoat to the sublayers; and normal load. DLC overcoat with higher Young's modulus and hardness are beneficial in reducing the adhesive friction and abrasive wear. The improvement in the contact performance due to DLC overcoat becomes distinct with an increase of normal load, especially in reducing the friction by introducing a stiffer DLC overcoat.

In order to bring the DLC overcoat into full play, the overcoat thickness should be larger than the floor critical thickness - 5 nm for the basis case ($\sigma = 1$ nm, $\beta_x^* = \beta_y^* = 0.5 \mu$m). The ceiling critical thickness is about $10^4$ nm for $A_r/A_n$ and $F_m/W$, and $10^2$ nm for $p_{\text{max}}/E_2$. A further normalization to $\beta^*$ changes the critical thickness accordingly. Conventional layer thickness (3 ~ 5 nm) is on the order of floor critical thickness. To make DLC overcoat more profitable, one can increase either the DLC overcoat thickness, or the stiffness- and hardness- ratio of the DLC overcoat to the sublayers, or both. However, the increase of the stiffness ratio results in a larger ceiling critical thickness. Therefore, these two factors should be coordinated with each other to ensure an optimum performance.
4.2 MicroElectroMechanical System.

Another major application field of micro/nano scale contact modeling is MicroElectroMechanical System. Using established semiconductor processing techniques, researchers have fabricated a wide range of miniature components (referred to as MEMS) such as pressure and acceleration sensors, linear actuators, valves, grippers, tweezers, motors, gear trains, turbines, nozzles, and pumps with dimensions of a couple to couple of hundred microns. MEMS devices are used primarily for their miniature size, and also for their high reliability and low-cost manufacturing techniques.

In MEMS devices, various forces associated with the device scale down with the size. Surface forces such as friction, adhesion, meniscus forces, and viscous drag in MEMS become more important than inertial forces. For example, as one group elements of MEMS, structural members are constructed in close proximity to each other (within several microns); they are highly compliant due to their extreme length-to-thickness aspect ratio; they have large surface-to-volume ratios which increases the relative importance of adhesive surface forces. The tribological concerns become critical because friction, stiction, wear and surface contamination affect device performance and in some cases, can even prevent MEMS devices from working. Many coating processes of solid and liquid lubricant and hard films have been conducted to minimize the friction and wear (Bhushan, 1999b, 2001). The interacting effects of coating material, roughness, and environment on adhesion and friction must be understood in order to allow proper operations of MEMS devices over their lifetime.

A sample layered MEMS material is chosen as DLC coated Silicon (100). The physical properties and surface topography statistics are shown in Table 3. $E_0$ and $H_0$ are
### Table 4.2 Typical physical properties and surface topography statistics of MEMS materials

<table>
<thead>
<tr>
<th>Material</th>
<th>$E_1$ (GPa)</th>
<th>$\nu$</th>
<th>$H_1$ (GPa)</th>
<th>$\sigma$ (nm)</th>
<th>$\beta^*$ (µm)</th>
<th>$\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DLC overcoat</td>
<td>140, 280</td>
<td>0.25</td>
<td>15, 24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Si</td>
<td>169</td>
<td>0.28</td>
<td>8</td>
<td>0.09</td>
<td>0.06</td>
<td></td>
</tr>
</tbody>
</table>

$^1$ Bhushan (1999c); $^2$ Bhushan and Venkatesan (1993); $^3$ Bhushan and Kulkarni (1996); $^4$ Bhushan and Gupta (1991); $^5$ Bhushan (1999b).

$\mu$ is the coefficient of micro friction measured using AFM / FFM over a scan size of 10 × 10 µm.

Assumed to be infinite. Silicon (100)'s modulus $E_2$ is taken as 170 GPa and hardness $H_2$ as 8 GPa. The DLC overcoat’s Young’s modulus $E_1$ varies with the ratio of $E_1/E_2$, which is taken as 0.82 and 1.65. The DLC overcoat’s hardness $H_1$ is taken as 0.1 $E_1$. All Poisson’s ratios are taken as 0.25.

The ratios of $E_1/E_2$ of the sample MEMS materials (0.82 and 1.65) are contained in those of HDI (0.75 and 1.75). Since a linear relationship between the contact statistics and $E_1/E_2$ exists, by inference the contact performances of these two MEMS materials should be inside the range of these two HDI cases too. The contact statistics can be obtained by applying the normalization scheme and are omitted here for brevity.
CHAPTER 5

CONCLUSIONS

When two rough surfaces are placed in contact, surface roughness causes contact to occur at multiple discrete contact asperities. Deformation occurs in the region of the contact spots, establishing stresses that oppose the applied load. The mode of surface deformation is either elastic or elastic-plastic. Relative sliding introduces a tangential force (referred to as friction force) at the contact interface. Repeated surface interactions and surface and subsurface stresses, developed at the interface, results in the formation of wear particles and eventual failure. A liquid film at the interface results in attractive meniscus forces that can also increases friction and wear. The deposition of thin layers, ranging in thickness from a couple of nm to a few microns, has been proved to be an effective way to improve the tribological performance.

Contact models of layered rough surfaces are developed to study the contact mechanics and predict optimal layer parameters. Among them statistical contact models and fractal models perform contact analyses on conceptual rough surfaces represented by a few statistical parameters or fractal functions. These models provide a main trend for each type of rough surfaces. However, their usefulness is very limited since they oversimplify the asperity geometry and height distribution, and neglect the interactions between adjacent asperities. This limitation is resolved by taking real rough surface
profiles directly as the model input rather than a few statistical parameters. The input data is generally very large for geometrically complicated rough surfaces, which makes an analytical solution impossible and numerical methods necessary.

Numerical methods physically divide the layered solids with rough surfaces into small and manageable pieces. Depending on the way of dividing these numerical methods are classified into FDM, FEM, BEM (Fig. 1.1). FEM and BEM are more flexible than FDM and theoretically both applicable to multiple asperity contact. However, BEM discretizes only the rough surface into small patches, as opposed to the whole layered solid being discretized by FEM. Therefore, comparing to FEM, BEM reduces the problem dimensionality and consequently the effort involving in obtaining a solution. It is widely used to solve 3-D layered rough surfaces contact problems.

Contact problems can be formulated with different principles and classified into three formulations. The analytical and numerical methods fall into three categories – direct formulation, weighted residual formulation, and minimum total potential energy formulation. For example, the matrix inversion method and CGM belong to the direct formulation, the least-square error method belongs to the weighted residual formulation, the steepest descent method and quasi-Newton method belong to the minimum total potential energy formulation (Fig. 1.1). If the boundary (surface) displacements are prescribed in the initial contact conditions, e.g., in case of single asperity contact, these formulations are mathematically equivalent so that these methods generate the same results. In case of multiple asperity contact, the least-square error method is unfeasible since obtaining an expression of the pressure profile in closed form presents difficulties. BEM numerical methods based on the direct formulation, such as the matrix inversion
method and CGM generate a feasible approximate solution upon converging. However, these methods may not converge with the presence of a large and ill-conditioned influence coefficients matrix. Moreover, the initial approximation of the prescribed displacements creates an uncertainty in the direct formulation and the uniqueness of the solution is not guaranteed. As a supplementary criterion to the direct formulation, the minimum total potential energy formulation allows the use of a quadratic programming method, such as the quasi-Newton method to solve an optimization problem, which theoretically guarantees the convergence and the uniqueness of the solution. It also significantly reduces computing time and thus feasible to solve the 3-D rough surfaces contact problems with large number of contact points.

These models predict contact pressure profile on the interface and contact statistics, namely fractional contact area, the maximum value of contact pressure, von Mises and principal tensile stresses, and relative meniscus force. Normalization schemes are developed with reasonable self-similarity of scaling relations. Contact problems are solved for various surface topographies, elastic and elastic/plastic material properties, normal loading and frictional loading, and dry and wet contact interface. The effects of the contact statistics on friction, stiction and wear problems such as debris generation, brittle failure, and delamination of layered media are addressed. The results allow the specification of optimum layer properties, according to the contact statistics, to reduce friction, stiction, and wear of materials.

In general, a stiffer layer is beneficial in reducing the adhesional friction, stiction and wear-particle formation by reducing the real area of contact. However, the increase of tensile stress $\sigma_t$ on the surface and interface shear stress $\sigma_{xz}$ increases the potential of
wear mechanisms such as crack failure and delamination. In contrast, a more compliant layer increases the real area of contact, which causes higher friction, stiction, and wear-particle formation. However, a more compliant layer reduces $\sigma_t$ and $\sigma_{xz}$, which is beneficial in reducing cracking and delamination failure. Besides the Young’s modulus effect, the increase of layer hardness improves the effective hardness of a layered solid. It reduces the tendency of plastic deformation, which consequently reduces the potential of cracking and residual stresses.

Friction has negligible affect on the contact pressure profile, provided the coefficient of friction is not very large. Introducing the friction force increases the magnitude of $\sigma_{xz}$, $\sigma_t$ and $\sqrt{J_2}$ and moves the maximum $\sigma_{xz}$ and $\sqrt{J_2}$ closer to the surface, which consequently increases the potential of wear mechanisms such as crack failure, delamination and residual stresses.

This model is extensively applied to study the layer effect on the tribological performances of the magnetic storage devices and MicroElectroMechanical Systems (MEMS).
BIBLIOGRAPHY


Ai X. and Sawamiphakdi K. (1999), Solving elastic contact between rough surfaces as an unconstrained strain energy minimization by using CGM and FFT techniques, ASME J. Tribol. 121, 639-647.


APPENDIX A

DEFINITION OF THE TERMS $R_1$, $R_2$, $R_3$, $R_4$, $R_5$, $R_6$, $R_a$, $R_b$, $R_c$ AND $R_d$

The terms $R_1$, $R_2$, $R_3$, $R_4$, $R_5$ and $R_6$ are given by

$$-\alpha^2 R_1 = i\xi (B^{(1)} - \overline{B}^{(1)}) + i\alpha^2 (B_\xi^{(1)} - \overline{B}_\xi^{(1)})$$

$$-\alpha^2 R_2 = 2i\xi (1 - v_1)(B^{(1)} + \overline{B}^{(1)}) + i\alpha^2 (B_\xi^{(1)} + \overline{B}_\xi^{(1)}) + \overline{p}(\xi, \eta)$$

$$-\alpha^2 R_3 = i(\xi + \xi \alpha h) B^{(1)} - i(\xi - \xi \alpha h) e^{2\alpha h} B^{(1)} - i\xi e^{e^{2\alpha h} B^{(1)}} + i\alpha e^{2\alpha h} B^{(1)} - i\xi e^{2\alpha h} B^{(1)}$$

$$-i\alpha^2 e^{a h} B_\xi^{(2)}$$

$$-\alpha^2 R_4 = [2i(1 - v_1)\xi - i\xi \alpha h] B^{(1)} + [2i(1 - v_1)\xi + i\xi \alpha h] e^{2\alpha h} B^{(1)} -$$

$$[2i(1 - v_2)\xi] e^{a h} B^{(2)} + i\alpha^2 B_\xi^{(1)} + i\alpha^2 e^{2\alpha h} B_\xi^{(1)} - i\alpha^2 e^{a h} B_\xi^{(2)}$$

$$-\alpha^2 R_5 = -i\xi \alpha h B^{(1)} + i\xi \alpha h e^{2\alpha h} B^{(1)} + i\alpha^2 B_\xi^{(1)} + i\alpha^2 e^{2\alpha h} B_\xi^{(1)} - i\alpha^2 e^{a h} B_\xi^{(2)}$$

$$-\alpha^2 R_6 = i(\xi - \xi \alpha h) B^{(1)} - i(\xi + \xi \alpha h) e^{2\alpha h} B^{(1)} - i\xi e^{a h} B^{(2)} + i\alpha e^{a h} B^{(1)}$$

$$-i\alpha^2 e^{a h} B_\xi^{(1)} - i\alpha^2 e^{a h} B_\xi^{(2)}$$

The terms $R_a$, $R_b$, $R_c$ and $R_d$ are given by

$$R_a = (G - 1)\alpha (R_1 + R_2) - G\alpha (R_3 + R_4) + \alpha (R_5 + R_6)$$

$$R_b = \alpha (R_2 - R_1) + \alpha e^{-2\alpha h} (R_3 - R_4)$$

$$R_c = \frac{4(1 - v_1)}{1 - \lambda} \left( \frac{2\alpha}{S_0} \right) R_a + R_b$$

$$R_d = \alpha [R_1 - R_2 - R_3 + R_4] + \frac{(1 - \lambda)\alpha}{4(1 - v_1)} (R_3 - R_4 + R_5 - R_6)$$

122
APPENDIX B

INFLUENCE COEFFICIENTS FOR A 3-D LAYERED SOLID
IN NORMAL CONTACT

In case of $\mu = 0$, the nine unknowns were given by Nogi and Kato (1997) and Peng and Bhushan (2001a) as

\[
A^{(1)} = \frac{1}{\left(1 - (2\nu_1)[1 - (1 - 2\alpha h)k e^{-2\alpha h}] + 0.5(k - \lambda - 4k\alpha^2 h^2) e^{-2\alpha h}\right) \Re^{-\alpha z}},
\]
\[
\bar{A}^{(1)} = \frac{1}{\left((1 - 2\nu_1) k (1 + 2\alpha h - \lambda e^{-2\alpha h}) + 0.5(k - 4k\alpha^2 h^2)\right) e^{-2\alpha h} \Re^{-\alpha z}},
\]
\[
C^{(1)} = \frac{1}{\left[1 - (1 - 2\alpha h) k e^{-2\alpha h}\right] \alpha \Re^{-\alpha z}},
\]
\[
\bar{C}^{(1)} = \frac{1}{\left(1 + 2\alpha h - \lambda e^{-2\alpha h}\right) k e^{-2\alpha h} \alpha \Re^{-\alpha z}},
\]
\[
A^{(2)} = \frac{-1}{\left(1 - (2\nu_2)(1 - \lambda)[1 - (1 - 2\alpha h)k e^{-2\alpha h}] + 0.5(\lambda - k)(1 - e^{-2\alpha h}) + \alpha h\right) \left[1 - k - k(1 - \lambda)e^{-2\alpha h}\right] e^{-2\alpha h} \Re^{-\alpha z}},
\]
\[
C^{(2)} = \frac{1}{\left[1 - (1 - 2\alpha h) k e^{-2\alpha h}\right](1 - \lambda)e^{-\alpha h} \alpha \Re^{-\alpha z}},
\]
\[
B^{(1)} = \bar{B}^{(1)} = B^{(2)} = 0.
\]

where \( R = \frac{1}{\left[1 - (\lambda + k + 4k\alpha^2 h^2) e^{-2\alpha h} + \lambda k e^{-4\alpha h}\right] \alpha^2}. \)

The influence coefficients $\bar{C}(\xi, \eta, z)$ corresponding to $z$-direction surface displacement and various stresses were given as
\[
C^{(1)}_{x} (\xi, \eta, 0) = -\left(1 - \frac{v_c}{G_i}\right) \left(1 + 4\alpha h k e^{-2\alpha h} - \lambda k e^{-4\alpha h}\right) \alpha R
\]

\[
C^{(M)}_{xx} (\xi, \eta, z) = -\xi^2 (A^{(M)} + \overline{A}^{(M)}) + 2v_M \alpha (C^{(M)} - \overline{C}^{(M)}) - \xi \eta^2 (C^{(M)} + \overline{C}^{(M)})
\]

\[
C^{(M)}_{yy} (\xi, \eta, z) = -\eta^2 (A^{(M)} + \overline{A}^{(M)}) + 2v_M \alpha (C^{(M)} - \overline{C}^{(M)}) - \xi \eta^2 (C^{(M)} + \overline{C}^{(M)})
\]

\[
C^{(M)}_{zz} (\xi, \eta, z) = \alpha^2 (A^{(M)} + \overline{A}^{(M)}) + 2(1 - v_M) \alpha (C^{(M)} - \overline{C}^{(M)}) + \xi \eta^2 (C^{(M)} + \overline{C}^{(M)})
\]

\[
C^{(M)}_{xy} (\xi, \eta, z) = -\xi \eta (A^{(M)} + \overline{A}^{(M)}) - \xi \eta (C^{(M)} + \overline{C}^{(M)})
\]

\[
C^{(M)}_{yz} (\xi, \eta, z) = -i \left[ (\eta \alpha (A^{(M)} - \overline{A}^{(M)}) + (1 - 2v_M) \eta (C^{(M)} + \overline{C}^{(M)}) + \xi \alpha (C^{(M)} - \overline{C}^{(M)})ight]
\]

\[
C^{(M)}_{xz} (\xi, \eta, z) = -i \left[ (\xi \alpha (A^{(M)} - \overline{A}^{(M)}) + (1 - 2v_M) \xi (C^{(M)} + \overline{C}^{(M)}) + \xi \alpha (C^{(M)} - \overline{C}^{(M)})ight]
\]

where the layer and the substrate were identified by the superscript M
When the indentation depth is less than about 0.2-0.3 of the layer thickness, the hardness of a layered solid is independent of the substrate (Bhushan, 1999b). As the indentation depth increases, the effective hardness of a layered solid becomes a function of the layer thickness, and the elastic-plastic behavior of both the layer and the substrate. Applying the finite element method, Bhattacharya and Nix (1988) developed an empirical formulation to calculate the effective hardness of a layered solid. In case of the layer is softer than the substrate, the effective hardness $H_e$ is given by

$$
\frac{H_e}{H_2} = 1 + \left( \frac{H_1}{H_2} - 1 \right) \exp \left[ - \frac{Y_1 / Y_2}{E_i / E_j} \left( \frac{h_c}{h} \right)^2 \right],
$$

where $E_i$, $Y_i$, and $H_i (i = 0, 1)$ are the Young’s moduli, yield strengths and hardnesses of the layer and substrate, respectively. $h$ is the layer thickness. $h_c$ is the contact indentation depth, which is given by

$$
h_c = h_{\text{max}} - \varepsilon W_{\text{max}} / S_{\text{max}}.
$$

where $\varepsilon = 0.72$ for a conical indenter, $W_{\text{max}}$ is the peak load and $h_{\text{max}}$ is the corresponding displacement, and $S_{\text{max}}$ is the Young’s modulus given by the slope of unloading curve $(dW/dh)$ at the peak load (Oliver and Pharr, 1992).
Similarly in case of the layer is harder than the substrate, the effective hardness $H_e$ is given by

$$\frac{H_e}{H_2} = 1 + \left( \frac{H_1}{H_2} - 1 \right) \exp\left[ -\frac{H_1 / H_2}{(Y_1 / Y_2)(E_1 / E_2)^{1/2}} \left( \frac{h_e}{h} \right) \right].$$
APPENDIX D

CALCULATION OF THE COMPOSITE ROUGH SURFACE

The contact of two rough surfaces can be transformed to a smooth flat surface in contact with a composite rough surface, provided the microgeometry of the composite rough surface is the sum of the two original surfaces. This inference is obtained by a simple geometrical transformation. Given one set of rough surfaces \( z_0(x, y) \) and \( z_1(x, y) \), with reference surface set as \( z = 0 \) by default, a change in the reference plane from \( z = 0 \) to \( z = z_1(x, y) \) leads to an equivalent set of surfaces \( z_0'(x, y) = z_0(x, y) - z_1(x, y) = |z_0(x, y)| + |z_1(x, y)| \) and \( z_1'(x, y) = 0 \). \( z_0'(x, y) \) represents the sum of the microgeometry of two original surfaces, and \( z_1'(x, y) = 0 \) represents a smooth surface.

The inference is mathematically correct. Since in the contact model, regardless of the surfaces profiles themselves, the initial separation of the two contact surfaces, referred as to \( f_z(x, y) = |z_0(x, y) - z_1(x, y)| \), is the program input concerning the surface topography. For various combinations of surface profiles, if \( f_z(x, y) \) are the same, the program inputs are the same. Regarded as program outputs, the contact behaviors of these surface profiles are the same.
The inference is also physically correct. According to Hertz contact theory, the contact is decided by the effective or composite radius of curvature at the contact point, which is given by

\[ r_c = \frac{r_0 r_1}{r_0 + r_1}. \]

Taking a simple 2-D case as an example, the original set of surfaces in contact are two identical half circles with radius \( r \), whose profiles are (Fig. 1(a))

\[
\begin{align*}
F_0 &= r \sin 2 \theta \\
F_0' &= r - r \cos 2 \theta \\
F_1 &= r \sin 2 \theta
\end{align*}
\]

and

\[
F_1' = r \cos 2 \theta - r
\]

![Diagram of two identical half circles in contact](image1)

(a)

![Diagram of a flat surface in contact with a composite half ellipse](image2)

(b)

Figure C.1 Schematics of (a) two identical half circles in contact and (b) a flat surface in contact with a composite half ellipse
where $\pi/4 \geq \theta \geq -\pi/4$. For these two half circles in contact, $r_0 = r_1 = r$, therefore, $r_c = r/2$.

The transformed surfaces in contact are a smooth flat surface and a composite rough surface whose microgeometry is the sum of the two original surfaces. Adding these two half circles, the composite rough surface is

$$
\begin{align*}
\mathbf{x}_0 &= x_0 = x_1 = r \sin 2\theta \\
\mathbf{z}_0 &= z_0 - z_1 = 2r - 2r \cos 2\theta,
\end{align*}
$$

which is a half ellipse (Fig. 1(b)). The radius of curvature of a half ellipse is given by (Stewart, 1995)

$$
r(2\theta - \pi/2) = \frac{r(4\cos^2(2\theta - \pi/2) + \sin^2(2\theta - \pi/2))^{3/2}}{2},
$$

and at the contact point $\theta = 0$, $r_{\theta=0} = \frac{r}{2}$.

For a half ellipse and a flat surface in contact, $r_0' = r/2$ and $r_f' = +\infty$, therefore, $r_c' = r/2$. It has the same radius of curvature at the contact point as the original set of surfaces, thus having the same contact behaviors.