The effect of determining topography parameters on analyzing elastic contact between isotropic rough surfaces

Gorakh Pawar

MS candidate
Dept. of Mechanical Engineering,
University of Utah
June 11, 2012
Introduction (1)

- **Industrialization** and urbanization has increased energy needs significantly in recent years.
- **Economic losses** associated with friction and wear contribute approximately 5% of total GDP of developed countries\(^1,2\).
- Friction and wear contributes to environmental issues such as pollution and global warming.
- Friction and wear are also critical in micro- and nanoscale hi-tech applications.
- **Contact of rough surfaces** needs to be studied in detail to gain fundamental insight in friction mechanisms on the micro and nanoscale.

---

Introduction (2)

• Applications

Design of intricate structures requires in-depth understanding of contact between rough surfaces

Automobile engines
Chip manufacturing

MEMS mechanisms

Hard disk drives

Automobile tires

pvd-coatings.co.uk
intel.com
Abysky.blogspot.com
thenanoage.com
Tristanmac.tripod.com
• Contact between rough surfaces on macro and microscale

- Macroscale contact
- Microscale contact

Nominal contact area

Real contact area defined by contacting asperities

Asperity
• **Contact models** are used to predict contact parameters as a function of the separation between contacting rough surfaces.

- Surface topography parameters
- Multi-asperity contact model
- Contact parameters

- Asperity density
- Mean asperity radius
- Standard deviation of asperity heights

- Number of contacting asperities
- Real area of contact
- Normal load
Introduction (5)

- Perhaps the most widely used multi-asperity elastic contact model is the Greenwood-Williamson (GW) model[3]
- GW model assumes an ideal 3D isotropic surface that can be represented by a single trace
- When characterizing an actual isotropic rough surface experimentally, or when analyzing a numerically generated isotropic rough surface, one finds that the topography parameters depend on the cross-section (trace) from which they are derived

Objective

- The original GW model has been successively improved by relaxing some of its simplifying assumptions, and entirely new theories were also developed.
- Many researchers still apply the original GW model to simulate elastic contact of 3D isotropic rough surfaces.
- **Objective:** Analyze the relationship between the number of contacting asperities, the separation, the real area of contact, and the normal load, using different methods to determine the asperity density $\eta$, mean asperity radius $\rho$, and standard deviation of asperity heights $\sigma_s$ of a rough surface.
Overview

• Numerically generating 3D isotropic rough surfaces
• Different methods to determine the topography parameters of the rough surfaces
• Results
  – Deviations between different contact models
  – Effect of the autocorrelation length of the rough surface
  – Effect of the sampling interval
• Summary
Rough surfaces

• Rough surfaces are fully characterized by three surface topography parameters
  – Asperity density $\eta$
  – Mean asperity radius $\rho$
  – Standard deviation of asperity heights $\sigma_s$

• Isotropic versus anisotropic surfaces

Isotropic surfaces:
Surface topography parameters are identical in all directions

Anisotropic surfaces:
Surface topography parameters depend on the direction in which they are evaluated
Numerically generating isotropic rough surfaces (1)

- **Commonly used** methods to numerically generate 3D rough surfaces
  
  - **Hu and Tonder method**[^4]
    
    - Surface generation based on **finite impulse response** (FIR) filter design
    - Surface generation with specified **autocorrelation function** (ACF)
  
  - **Wu method**[^5]
    
    - Surface generated based on specified ACF and using Fast Fourier Transform


Numerically generating isotropic rough surfaces (2)

- Surface generation based on FIR filter design (Hu and Tonder)

\[ M = N = 64 \]
Numerically generating isotropic rough surfaces (3)

- **Surface generation with specified ACF (Hu and Tonder)**
  ACF is used to describe the *repetitiveness* of the surface profile with itself when it is *spatially delayed* by some distance $\Delta x$ [6]

\[
R_z (\Delta x) = \frac{1}{\sigma^2} \lim_{L \to \infty} \frac{1}{L} \int_0^L z(x) z(x + \Delta x) \, dx
\]

Value of ACF varies between -1 to 1

Numerically generating isotropic rough surfaces (4)

- **Surface generation with specified ACF (continued..)**

1. Specify the desired ACF $R_{zz}$
2. Generate a matrix with normally distributed random numbers $R_\eta$
3. Compute FFT $S_{zz}$ and $S_{\eta \eta}$ of ACF $R_{zz}$ and $R_\eta$
4. Calculate $H = (S_{zz}/S_{\eta \eta})^{1/2}$
5. Compute filter coefficients $h_f$ as IFFT of $H$
6. Compute the convolution of $R_\eta$ and $h_f$ to obtain surface heights $z$ with prescribed ACF

- Exponential ACF
- $M = N = 128$
Numerically generating isotropic rough surfaces (5)

- **Generation of numerical rough surfaces with Wu method**
  - Utilizes FFT and user defined ACF or spectral density of rough surface

```
Specify desired ACF  →  Spectral density  →  Surface with normally distributed surface heights $z$
```

OR

```
Spectral density  →  Surface with normally distributed surface heights $z$
```
Numerically generating isotropic rough surfaces (5)

- **Generation of numerical rough surfaces with Wu method**
  - Utilizes FFT and user defined ACF or spectral density of rough surface
  - If spectral density is known, rough surface is simulated as

\[
z_{p,q} = \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} \sqrt{\tilde{S}_{k,l}} \exp \left( i 2\pi \left( \phi_{k,l} + \frac{kp}{M} + \frac{lq}{N} \right) \right)
\]  

(1)

- If ACF is known first spectral density is calculated from known ACF as

\[
\tilde{S}_{k,l} = \frac{1}{MN} \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} \tilde{R}_{r,s} \exp \left( -i 2\pi \left( \frac{kr}{M} + \frac{ls}{N} \right) \right)
\]

(2)

then surface is simulated with Eq. (1)
Numerically generating isotropic rough surfaces (6)

- The method proposed by Wu\textsuperscript{[5]} was used to generate 3D isotropic rough surfaces with an exponential autocorrelation function (autocorrelation length $\beta^*$)
- 512 by 512 data points
- Sampling interval $dx = dy = 1\ \mu m$
- Young’s modulus $E = 210\ \text{GPa}$, hardness $H = 1.96\ \text{GPa}$, and Poisson’s ratio $\nu = 0.3$

Surface 1, $\beta^* = 10\ \mu m$
 Surface 2, $\beta^* = 20\ \mu m$
 Surface 3, $\beta^* = 50\ \mu m$

Str = 0.85
Str = 0.92
Str = 0.96

Equivalent rough surface

- Elastic contact of two rough surfaces with identical autocorrelation function is replaced by that of an **equivalent rough surface and a rigid flat**[7].

- The topography parameters of the equivalent surface are determined using three methods, commonly used in the literature

Determining topography parameters (1)

- Spectral moments approach applied to a single arbitrary cross-section

Spectral moments

\[
m_0 = \text{AVG}\left[\left(z^2\right)\right]
\]

\[
m_2 = \text{AVG}\left[\left(\frac{dz}{dx}\right)^2\right]
\]

\[
m_4 = \text{AVG}\left[\left(\frac{d^2z}{dx^2}\right)^2\right]
\]


Topography parameters

\[
\eta = \left(\frac{m_4}{m_2}\right) / 6\pi \sqrt{3}
\]

\[
\rho = 0.375\left(\frac{\pi}{m_4}\right)^{\frac{1}{2}}
\]

\[
\sigma_s = \left(1 - \frac{0.8968}{\alpha}\right)^{\frac{1}{2}} m_0^{\frac{1}{2}} m_2^{\frac{1}{2}}
\]

with \(\alpha = (m_0 m_4)/m_2^2\)
Determining topography parameters (2)

- Spectral moments approach averaged over a discrete number of cross-sections

Average spectral moments over discrete set of cross-sections

\[ m_0 = \text{AVG} \left[ \left( \frac{dz}{dx} \right)^2 \right] \]

\[ m_2 = \text{AVG} \left[ \left( \frac{d^2 z}{dx^2} \right)^2 \right] \]

\[ m_4 = \text{AVG} \left[ \left( \frac{d^3 z}{dx^3} \right)^2 \right] \]

\[ \eta = \frac{m_4}{m_2} / 6\pi \sqrt{3} \]

\[ \rho = 0.375 \left( \frac{\pi}{m_4} \right)^{1/2} \]

\[ \sigma_s = \left( 1 - \frac{0.8968}{\alpha} \right)^{1/2} \frac{1}{m_0^{1/2}} \]

with \( \alpha = \frac{m_0 m_4}{m_2^2} \)

Summit identification method

Determining the summits of the surface as local maxima using an 8-nearest neighbor summit identification scheme

Equivalent rough

Summit identification method

Summit locations

\[ \sigma_s \] : Determined from the \( n \) asperities identified on the surface

\[ \eta \] : Determined as \( n/A_n \)

\[ \rho \] : Average radius of all asperity summits on the surface

The summit curvature is determined for each asperity \( i \) in two orthogonal directions

\[ \kappa_{x,i} = \frac{d^2 z}{dx^2} \quad \text{and} \quad \kappa_{y,i} = \frac{d^2 z}{dy^2} \]

of that summit is computed as

\[ \rho_i = -\left[ \frac{\left( \kappa_{x,i} + \kappa_{y,i} \right)}{2} \right]^{1/2} \] [9]

Determining topography parameters (4)

- Summit identification method (continued..)

4-nearest neighbor square SID scheme

8-nearest neighbor SID scheme

4-nearest neighbor diagonal SID scheme
Determining topography parameters (5)

- Summit identification method (continued..)

    False summit identification

    4-nearest neighbor versus 8-nearest neighbor SID

    - Possible reasons for false summits
      - Presence of saddle points
      - Presence of ridges

Determining topography parameters (6)

- When determined from an arbitrary cross-section, $\rho$, $\eta$, $\sigma_s$ vary significantly (30 – 75%)
- The values of $\sigma_s$ and $\rho$ obtained by using the summit identification method (SID) fall within the range of the corresponding values determined from an arbitrary cross-section
- The difference between $\sigma_s$ and $\rho$ obtained from the SID method and the average spectral moment method is less than 15%, for each of the surfaces
- $\eta$ obtained with the SID method is significantly smaller than that obtained with any of the other approaches
Determining contact parameters (1)

The topography parameters are used to determine the contact parameters between the equivalent rough surface and a rigid flat.

\[ \sigma_s \quad \eta \quad \rho \]

- Lower extreme
- Upper extreme
- Average
- SID

\[ n_p = \eta A_n \int_{d^*}^{\infty} \Phi(z^*) dz^* \]

\[ A^* = \pi \beta \int_{d^*}^{\infty} (z^* - d^*) \Phi(z^*) dz^* \]

\[ P^* = \frac{4}{3} \rho \sigma_s \eta \int_{d^*}^{\infty} (z^* - d^*)^{\frac{3}{2}} \Phi(z^*) dz^* \]

\[ \Phi(z^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \]

- Summit identification model (SID)

\[ A^* = \frac{\pi \rho \sigma_s}{A_n} \sum_{i=1}^{n_p} (z_i^* - d^*) \]

\[ P^* = \frac{4}{3} \rho^2 \sigma_s^2 \frac{1}{A_n} \sum_{i=1}^{n_p} (z_i^* - d^*) \]

Determining contact parameters (2)

- The results in this work are limited to the range $0 \leq h^* \leq 3$
- The SID model is assumed to provide the most reliable results compared to the other models:
  - The SID surface topography parameters are derived from the 3D surface and the actual asperities, not from statistical analysis based on limited data, i.e., a discrete number of cross-sections
  - The number of contacting asperities is determined more accurately, even at large separations
- The deviation between the different models is calculated by comparing the results for each model relative to the results obtained with the SID model
- All following results for surface 2 with $\beta^* = 20 \ \mu m$
Contact parameters (1)

- **Number of contacting asperities** $n_p$ **versus nondimensional separation based on surface heights** $h^*$

![Graph a) showing number of contacting asperities vs. nondimensional separation](image)

- The models diverge with decreasing $h^*$ as more asperities come into contact, increasing the effect of the different $\eta$ in each model.
- The deviation varies between $-26\%$ and $187\%$ for the extreme cases where $n_p$ is derived from a single cross-section.
- When using a single cross-section, the actual error will be contained within the envelope constrained by the lower and upper extreme errors.
Contact parameters (2)

- Nondimensional real area of contact $A^*$ versus nondimensional separation based on surface heights $h^*$

Real area of contact $A^*$ is directly related to $n_p$ and the contact area per asperity.

Lowest values of $A^*$ are obtained for the models based on the SID topography parameters.

Models diverge with decreasing $h^*$.
Contact parameters (3)

- Nondimensional separation based on surface heights $h^*$ versus nondimensional normal load $P^*$

  \[ (h_{GW}^* - h_{SID}^*) / h_{SID}^* \]

- GW with SID topography parameters & SID model overlap for $0 < h^* < 2$
- For $h^* > 2$, these two models diverge, as a result of the different treatment of contacting asperities in the SID model as opposed to the GW model
- The SID topography parameters in the GW model yield the smallest % errors
- The error for the case of the GW extreme and average models approaches infinity when the load increases
Contact parameters (4)

- Nondimensional real area of contact $A^*$ versus nondimensional normal load $P^*$

The data are shown for $A^* \leq 0.14$ and $P^* = 7.0 \times 10^{-4}$ since this upper limit coincides with $h^* = 0$ for the SID model.
Contact parameters (5)

- The differences between the contact parameters predicted by the different models result from two main reasons
  - At small separations ($0 < h^* < 2$): Different asperity density $\eta$ used in the GW and SID models plays the most important role
  - At large separations ($h^* > 2$): Different treatment of the number of contacting asperities is the dominant factor
    **SID model**: A discrete number of asperities makes contact and zero normal load and real contact area are achieved above a certain finite separation.
    **GW model**: Overestimates the number of asperities in contact at large separations because $n_p \to 0$ only when $h^* \to \infty$
Effect of autocorrelation length

- **Autocorrelation length:** surface 1: 10 μm, surface 2: 20 μm, surface 3: 50 μm

- Only the results for the GW average model and the SID model are shown for clarity
- For a constant $P^*$, the smoothest surface (surface 3) will result in the largest $A^*$ and the smallest $h^*$
- The limits of the error of $n_p$, $A^*$, and $h^*$ increase with increasing autocorrelation length
Effect of sampling length (1)

- Sampling interval $dx = 1$, $2$, and $3$ $\mu$m, $3$ different surfaces

![Graphs showing the effect of sampling length on different surfaces.](image-url)
Effect of sampling length (2)

- The topography parameters obtained for different sampling intervals varied least for surface 3 (largest autocorrelation length – smoothest surface)
- At $h^* = 3$ the difference between the results for the SID and GW average model is maximum for the largest sampling interval (3 μm)
- At $h^* = 0$ the difference between the results for the SID and GW average model is maximum for the smallest sampling interval (1 μm)
- For increasing autocorrelation length the differences in $h^*$ obtained with the SID and GW average model decrease, in particular at large separations
Summary (1)

- The *topography parameters* obtained from a *single arbitrary cross-section* of an isotropic 3D rough surface using the spectral moments method *vary significantly* depending on which cross-section is selected.

- An approach that results in a *unique solution* is to average the various different spectral moments obtained from a set of many single cross-sections and calculate the *topography parameters* from these *average values*.

- The *asperity density* is the main difference between the topography parameters obtained using the *spectral moments* method and the *SID* method.

- As a result of the *different topography parameters* obtained for each method, the relationship between $n_p$, $A^*$, and $P^*$ varies substantially for the GW and SID models.
Summary (2)

- The **GW w/SID** seems to be the better choice among all the other GW model options that are considered.
- The main cause of the error of the various contact parameters is twofold:
  - $(0 < h^* < 2)$: the **different asperity density** $\eta$ used in the GW and SID models is the most important factor.
  - $(h^* > 2)$, the **different treatment of the number of contacting asperities** is the dominant parameter.
- The **GW model** overestimates the different contact parameters at large **separations** because $n_p \to 0$ only when $h^* \to \infty$.
- With **increasing autocorrelation length**, contact parameters become **less sensitive to the sampling interval**, irrespective of the method to determine the topography parameters.