Surface roughness and its effects in tribology

Mitjan Kalin, Janez Kogovšek

University of Ljubljana, Faculty of Mechanical Engineering
Centre for Tribology and Technical Diagnostics

October 25, 2011
Tribological surfaces

- **Surfaces and surface roughness (engineering perspective):**
  - meaning,
  - measuring surface roughness,
  - characterization of surfaces roughness,
  - real contact area,
  - plasticity index.

- **Friction:**
  - basic concepts,
  - dependence on surface properties.
Surface properties

Engineering surfaces never have an ideal geometrical shape, but instead include different deviations.

With regards to the level of approximation they can be considered:

- smooth and even,
- smooth and wavy,
- rough and even,
- rough and wavy.
Errors of form

- Micro-geometrical deviations – roughness (important for interaction of surfaces)
- Macro-geometrical deviations – waviness

- What is „rough“?

An objective “measure” is required.
Surface roughness measurement methods

1. Stylus instruments
2D
The resolution depends on the stylus tip and the velocity of scanning.
\[ Z = 10 \text{ nm} \]
\[ X = f (v, f) \]
2. Optical interferometer

3D
The resolution depends on optics and light wavelength.

\[ Z < 1 \, \text{nm} \]
\[ Y = f(\lambda; \text{100-200 nm}) \]

3D

The resolution depends on the laser, scanner, feedback loop, software, probe (tip)...

$Z < 0.007 \text{ nm, } XY = 0.1 \text{ nm}$
Analysis of the measured surface roughness parameters

**Basic element:** surface profile.

Traversing length is denoted with $L_T$ and represents the distance that is traversed across the surface by the stylus when characterizing the surface, i.e. measurement length.

Assessment length $L_M$ is the length over which surface data is acquired and assessed.

Sampling length (reference length) is denoted by $L_V$. It is a length of a section inside the assessment length and it is equivalent to wavelength of the filter, $\lambda_C$ (it distinguishes the roughness from the waviness).

Standardized: important to choose the correct reference length and assessment length, so that the macro-geometrical deviations are excluded from the measurement.
Mean line of the profile, $m$

Mean line of the profile is denoted by $m$. It is a line with a shape of geometrical profile (perfect geometric line) and it runs parallel to that profile. The mean line of the profile is determined so that the sum of squared deviations from this line is the smallest.

...or otherwise: *Surface area above and below the mean line of the profile is the same!*

\[ m = \frac{1}{l} \int_{0}^{l} z \, dx \]
Arithmetical mean deviation, $R_a$

The most widely recognized and used parameter for surface roughness characterization.

$R_a$ is arithmetical mean deviation of all the measured values in the assessed profile ($L_M$) from the mean line of that profile.

$$R_a = \frac{1}{l} \int_0^l |z(x)| \, dx$$
Averaging of data:
⇒ $R_a$ does not differentiate between profile peaks and valleys!!

⇒ $R_a$ or any other parameter by itself: not sufficient.

⇒ Additional parameters necessary:
more sensitive & able to distinguish between surfaces with different shapes and/or ratios of peaks and valleys.
Root mean square deviation of the assessed roughness profile, $\sigma = R_q (m=0)$

In statistics: standard deviation, $\sigma$ - not only for surfaces

$R_q$: - more sensitive to different shapes and distributions of valleys and peaks,
- it still does not distinguish one from another,
- also based on averaging of values.

In general, parameter $R_q$ has a slightly higher value than parameter $R_a$ (10-25%).
Amplitude density function $p(z)$

Amplitude density function (= probability distribution of surface heights) is denoted by $p(z)$.

Its value is proportional to probability, that a point of the surface profile exists at a certain height – $z$. (proportion of individual heights of the profile)

A “tool” for further mathematical evaluation of surface roughness.
Skewness, $S_k$ – measure of asymmetry of $p(z)$

Peaks and valleys are distributed symmetrically in relation to mean line:
$\Rightarrow p(z)$ is symmetrical in relation to $m$.

Peaks and valleys are distributed asymmetrically in relation to mean line:
$\Rightarrow p(z)$ is asymmetrical in relation to $m$ – it is shifted “higher” or “lower”.

Surfaces with the same $R_a$ and $R_q$ can have a different $S_k$.

$$S_k = \frac{1}{R_q^3} \frac{1}{l} \int_0^l z^3(x) dx$$

For Gaussian distribution:
$S_k = 0$.

$Z = \text{positive}$
$\ (Sk = \text{negative})$

$S_k < 0$

$S_k > 0$
Kurtosis $K$ – measure of flatness of $p(z)$

Kurtosis $K$ – tells us how high or how flat the $p(z)$ is.

\[ K = \frac{1}{R^4_q} \int_0^l z^4(x) dx \]

For Gaussian distribution: Kurtosis = 3.
Flat peaks and valleys: $K < 3$.
Sharp peaks and valleys: $K > 3$. 
Which of the two surfaces has a greater load-bearing capacity?

cumulative distribution function

Portion of the surface that will carry the load at certain height.
Measurement scale and slope of the peaks

Measurement scale of a profile, e.g. acquired with a stylus profilometer:

M 50:1

Inclination of peaks and valleys:

5° – 20°
Real contact area

What proportion of the nominal contact area is actually included in the real contact area?

$$A_r = \sum_{i=1}^{n} A_i$$
1% $A_{\text{nom}}$

\[ p_r = \frac{F}{A_r} \]

⇒ the real contact pressure is significantly larger than the nominal contact pressure

⇒ conditions on the contact peaks are completely different from our „anticipations“.

A real = 1-10\% (20, 30\%) of the nominal contact area – in the most typical engineering cases.

Diameter of a single peak contact: $d = 1-50 \mu m$. 
Difficult to determine – for usage in “models”. 
(Sometimes the value is measured/determined but is not valid in general...)

Calculations – assumptions!
- Asperities are randomly distributed,
- are of different sizes (height, width),
- asperities in interactions with each other get changed in very short time intervals (position),
- asperities are constantly modified by wear,
- wear particles influence the real contact area.

Measurement methods:
- static (replica...); are not valid for dynamic contacts,
- contact resistance,
- pressure sensor (materials with a known phase transformation),
- contacts cannot be seen (closed); visible contacts (sapphire) are not valid in general.
Plasticity index

Determination of plastic deformation of asperities:

\[ \psi = \frac{E^*}{H} \sqrt{\frac{\sigma}{r}} \]

\( \psi \) ... plasticity index.

\( E^* \) ... Young's modulus,

\( H \) ... hardness of the softer material,

\( r \) ... radius of asperities,

\( \sigma = SD = RMS = R_q \).

\( \psi < 0.6 \) – most of the asperities are deformed *elastically*; plastic deformation occurs only under high contact pressures,

\( \psi > 1.0 \) – most of the asperities will be *plastically* deformed already under low contact pressures,

\( 0.6 < \psi < 1 \) – intermediate area.

Deformation mode of the contact is inseparably related to the surface roughness and material properties (hardness)!
Relation between plasticity index and surface roughness

High $E^*/H$ ratio (e.g. for steel) – high plasticity index:
If we want to be in the elastic area, we have to lower either surface roughness – or contact pressure (at significantly higher surface roughness it is not sufficient!). Only finely polished surfaces will remain in the elastic area (all other surfaces will be in the plastic area).

Low $E^*/H$ ratio (e.g. ceramics):
In this case the majority of the contacts will be elastic even at slightly higher surface roughness.
Surface roughness & friction?
What is friction?

Friction force or friction: a resisting force encountered by one body moving over another = resistance to movement in tangential direction of the contact.

Work to overcome friction = energy „loss“:
- transferred to the surroundings as heat,
- friction forces in a tribological contact: desired minimal;
- e. g. modern internal combustion engine: ~15-20 % of the energy is lost (useless heating of components).

Exceptions (friction of key importance):
tires on the road, brakes, clutches…

In long-term: the influence of friction even more detrimental:
- wear of components: higher friction – greater wear (but not always!),
- increased contact temperature;
- effects on material properties,
- direct mechanical influences (replacement of worn-out components, costs of delays, repairing, maintenance and replacement of machinery).
First “rules” – laws of friction

(1) **1st Amonton’s law**: friction force $F$ between a pair of surfaces is directly proportional to the normal load $W$ (Figure a, b).

The constant that gives the proportionality between the normal and tangential force, is generally known as coefficient of friction $\mu$.

(2) **2nd Amonton’s law**: friction force $F$ is independent of the nominal (apparent) contact area (Figure a, c).

This arises from the fact that the real contact area, $A_r$, is proportional to the normal load $W$ and independent of the nominal contact area!

(3) **Coulomb’s law** of friction: coefficient of friction is independent of sliding velocity once the movement is established.

Once the contact temperatures must be accounted for, it becomes increasingly less accurate (tribofilms, reactions…).
Understanding friction

\[ \mu = \frac{F_t}{W} = \text{const?} \]

- \( \mu \) – coefficient of friction,
- \( F_t \) – tangential force,
- \( W \) – load.

\[ F_t = A_r \times \tau \]

- \( A_r \) – real contact area,
- \( \tau \) – shear strength = specific resistance to shearing.

\( \tau \) depends on physical (chemical) properties of shearing of surfaces: formation and type of bonds between surfaces and on the manner these bonds are broken.

\( A_r \) depends on mechanical properties of the contact surface: on the deformation – it depends on roughness, hardness, elasticity, toughness, collapse mode of material.

The value of friction coefficient:
- related to the material pair in contact (supposedly constant for each material pair),
- but also to the physical conditions of sliding!
- Universal model for friction (not constant) does not exist!
Mechanisms (causes) of friction

I. Proportion of friction due to adhesion

- Finite number of contact points (real contact area),
- contact stress significantly larger than the nominal value of the contact pressure,
- atoms of one surface may come very close to the atoms of the other surface,
- strong bonds can be formed (electron exchange).

- Shearing of these connections to get tangential motion (tearing/pulling bodies apart),
- pronounced with ductile and soft materials (copper) or noble metals (don’t form stable oxide layers).

Additional explanation: Junction growth.
Junction growth

Single asperity contacts:
- in a **region of plastic deformation** (high contact pressure),
- easily **additionally deformed** (tangential force due to sliding),
- **material flows** – it „fills“ all the valleys and peaks,
- increase in the real contact area,
- larger adhesive connections,
- **greater tangential force required** to break these connections,
- higher coefficient of friction.

Tangential force and real contact area will be increasing as long as the maximum shear strength of material is not reached.
II. Proportion of friction due to abrasion

Surfaces after sliding interaction:
⇒ abrasion scars,
⇒ (direction of the scars indicates sliding direction).

⇒ Result of "scratching" of harder particle;
⇒ energy losses contribute to the total friction loss.

Particles or asperities have a complex shape:
⇒ the geometry of the particles can be simplified,
⇒ two extreme cases: sharp conical particles and spherical particles.
II. Proportion of friction due to abrasion

**Sharp, conical particles:**
- COF only depends on slope (not height),
- surface asperities,
- slope is small (≈ 10 °),
- \( \Rightarrow \) COF is small \(< 0.1\).

**Spherical particles:**
- COF depends on radius and penetration depth,
- contribution to COF can be high \( (> 0.2)\),
- correlated to wear particles, particles from surroundings,
- filtering to eliminate them.
III. Proportion of friction due to deformation of asperities

Coefficient of friction depends on:
- slope of asperities,
- engagement angle of asperities,
- angle of deformation of material.

Coefficient of friction due to deformation of asperities is indicated to be very high (0.4-1.0)!

High COF values similar to static friction!
⇒ it is likely that most deformations occur before the beginning of sliding.
Summary

- Surface topography includes roughness, wavines and form.

- Surface roughness is measured by profilometry, optical interferometry, AFM.

- Some of the main parameters for characterization are: $R_a$, $R_q$, $R_z$, $S_k$, $K$, $t_p$.

- Surface roughness has influence on the real contact area and type of surface deformation (elastic/plastic – plasticity index).

- Surface roughness influences all of the main causes of friction: adhesion, abrasion, deformation of asperities.

Thank you for attention!