

Surface texture: two-dimensional

*“Errors, like straws, upon the surface flow;
He who would search for pearls must dive below.”*

(All for Love, Prologue, 25; John Dryden 1631–1700)

Surface texture: an overview

The concept of *surface technology* has previously varied according to whether a scientist, technologist, or engineer has inspected a surface. An individual's viewpoint differs depending upon their specific interest and technical emphasis. Therefore, a surface cannot be thought of in isolation from other features, because each is valid and necessary and needs to be considered to gain an overall impression of its industrial, or scientific usage.

Surface technology encompasses a wide range of disciplines that includes: metrology; metallurgy; materials science; physics; chemistry; tribology and mechanical design. Moreover, it has become apparent that this topic, and associated characteristics that relate to surface technology needs to be recognised as a subject in its own right.

Surface technology provides an important and valuable insight into the practical and theoretical applications of a manufactured surface, most notably for the following reasons:

- the surface is the final link from the original design concept through to its manufacture, with the industrial engineer being the last person to add value to the product prior to shipment;
- industry is continually attempting to improve component power-to-weight ratios, drive down manufacturing costs and find efficient ways of either producing or improving surfaces – this has become an established goal of world-class competing companies today;
- internationally, there is growing demand to recognise the legal implications of a product's performance. Product liability is directly related to the production process technique and, hence, to a surface's quality;
- technical data that in the past was considered valid, has of late been found to be either unreliable or may give insufficient detailed information on surface phenomena. By way of an example, originally it was thought that the surface texture was directly related to a component's fatigue life; however, it has been shown that surface topography must be extended to include sub-surface metallurgical transformations, namely surface integrity;
- the production process selected for the manufacture of a part has been shown to have some influence on the likely in-service reliability of components.

1.1 Introduction

On the macro scale the natural world mimics surfaces found in engineering. Typical of these is that shown in Figure 1, where the desert sand is comprised of sand grains – *roughness*; ripples in the surface – *waviness*; together with the undulating nature of the land – *profile*. From an engineering point of view, surfaces are boundaries between two distinct media, namely, the component and its working environment. When a designer develops a feature for a part – whether on a computer-aided design (CAD) system, or drawing board – neat lines delineate the desired surface condition, which is further specified by its specific geometric tolerance. In reality, this theoretical surface condition cannot exist, as it results from process-induced surface texture modifications. Regardless of the method of manufacture, an engineering surface must have

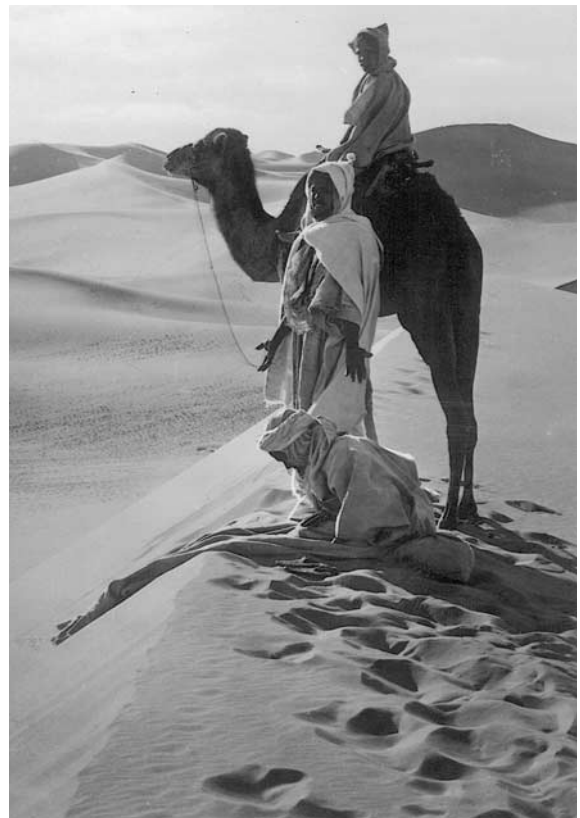


Figure 1. The natural world – the desert – can exhibit a large-scale combination of surface characteristics, for instance:

- agglomeration (i.e. clusters) of sand – *roughness*;
- ripples – *waviness*; and
- the undulating nature of the land – *profile*.

(Source: James Smith, *Prayer in the desert*, circa 1920s.)

some form of texture associated with it, this being a combination of several interrelated factors, such as:

- the influence of the material's microstructure;
- the surface generation method, the tool's cutting action, tool geometry, cutting speed, feedrate and the effect of cutting fluid;
- instability during the manufacturing process, promoted by induced chatter – poor loop-stiffness between the machine-tool-workpiece system, or imbalance in a grinding wheel;
- inherent residual stresses within the part, promoted by stress patterns causing deformations in the component.

Within the limitations resulting from the component's manufacture, a designer must select a functional surface condition that will satisfy the

operational constraints, which might be a requirement for either a "smooth" or "rough" surface. This then begs the question posed many years ago: "How smooth is smooth?" This is not as superficial a statement as it might initially seem, because unless we can quantify a surface accurately, we can only hope that it will function correctly in service. In fact, a surface's texture, as illustrated in Figure 2, is a complex condition resulting from a combination of:

- *roughness* – comprising of irregularities that occur due to the mechanism of the material removal production process; tool geometry, wheel grit, or the EDM spark;
- *waviness* – that component of the surface texture upon which roughness is superimposed, resulting from factors such as machine or part deflections, vibrations and chatter, material strain and extraneous effects;

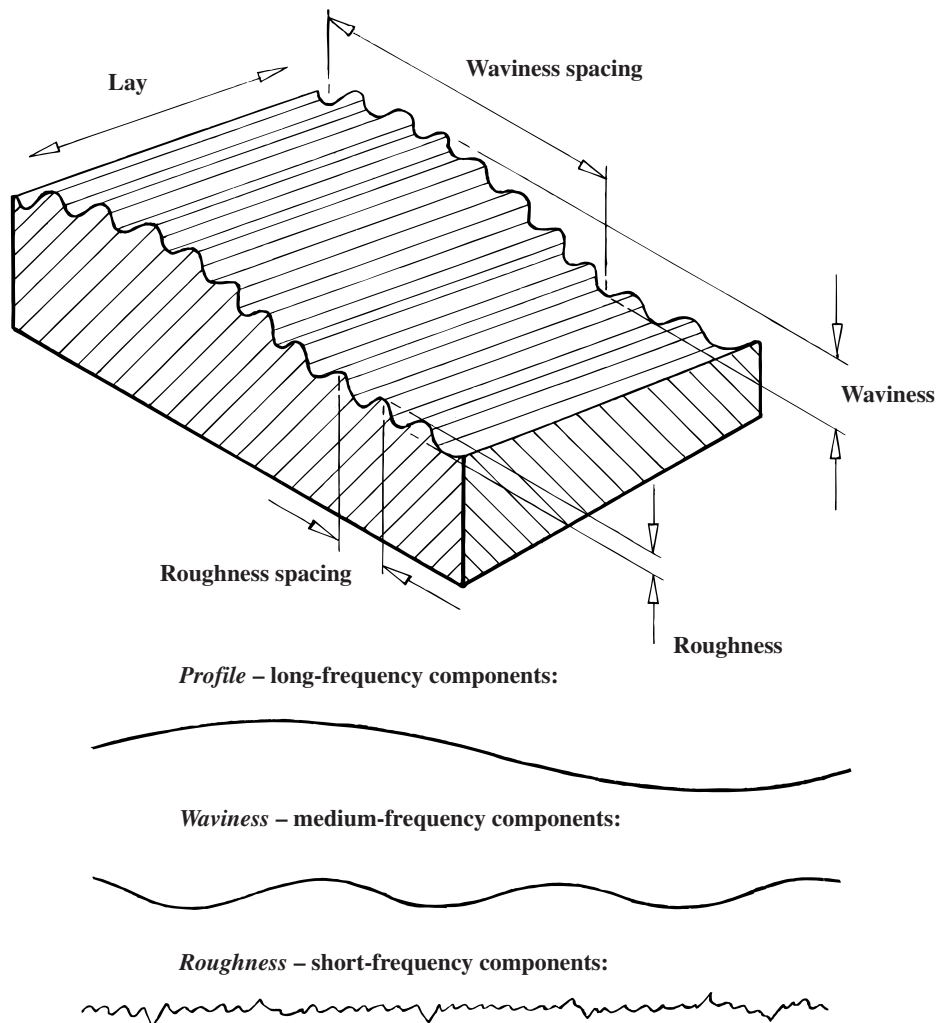
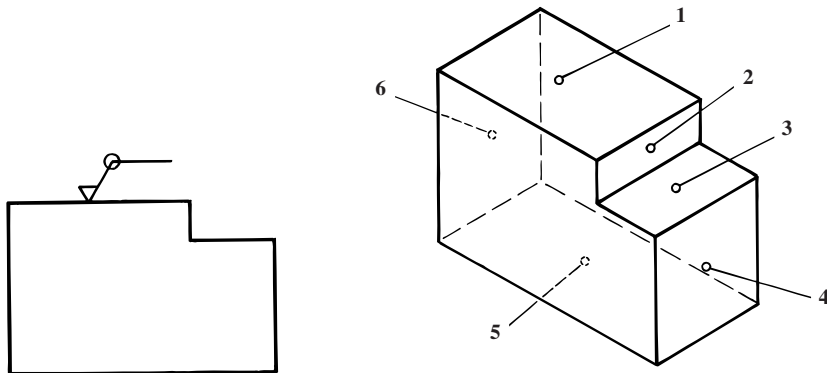
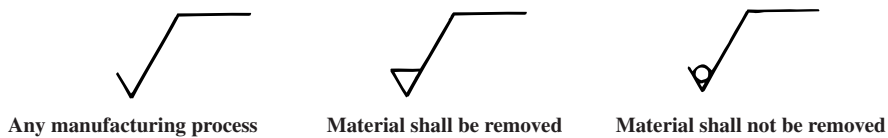
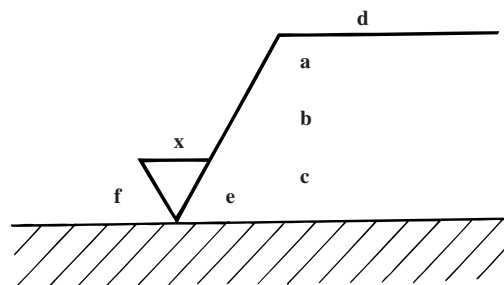


Figure 2. The major components that constitute a typical *surface texture*, also exhibiting some degree of directionality (lay).



*The outline represents the six surfaces shown on the 3D-representation (not the front to rear surfaces).



NB: Surfaces exhibiting a defined directionality, or “Lay” must be measured at right angles to this lay pattern.

KEY

a = 2D parameter 1
 b = 2D parameter 2
 c = 2D parameter 3
 d = process
 e = lay pattern
 f = allowance
 x = not allowed

Figure 3. The composition of the complete graphical symbol for surface texture. [Source: ISO 1302: 2001]

- *profile* – the overall shape of the surface – ignoring roughness and waviness variations – is caused by errors in machine tool slideways.

Such surface distinctions tend to be qualitative – not expressible as a number – yet have considerable practical importance, being an established procedure that is functionally sound (Figure 3). The combination of these surface texture conditions, together with the surface’s associated “lay”, are idealistically shown in Figure 4, where the lay of the surface can be defined as the *direction of the dominant pattern*. Figure 4 also depicts the classification of lay, which tends to be either *anisotropic* – having

directional properties such as feed marks, or *isotropic* – devoid of a predominant lay direction. The lay of any surface is important when attempting to characterise its potential functional performance. If the direction of the trace being produced by a stylus instrument – record of the stylus motion over the assessed surface topography – is not taken into account, then a totally misrepresentative reading will result from an anisotropic surface. This is not the case when measuring an isotropic surface, for example a “multi-directional lay” – as indicated in Figure 4. Under this isotropic condition, perhaps three different measurement paths could be scanned and the worst roughness trace would be utilised to

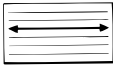
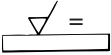
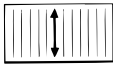
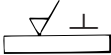

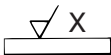
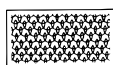


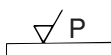




TYPE	LAY	SYMBOL
Parallel		
Perpendicular		
Crossed		
Multi-directional		
Particulate		
Circular		
Radial		

Figure 4. The indication of surface 'lay' as denoted on engineering drawings. [Source: ISO 1302: 2001]

specify the surface. The designer should be aware that the lay condition and other information can be incorporated into the surface texture symbol (ISO 1302: 2001) – which is often misused – see Figure 3 for the positioning of this data and other surface texture descriptors. The surface texture symbol shown in Figure 4, indicated by “surface texture ticks”, highlights the potential process-related lay type and associated directionality (ISO 1302: 2001).

Returning to the theme of an isotropic lay condition, if this surface is assessed at a direction not at right angles to the lay, then a totally unrepresentative surface profile will result, which is graphically illustrated in Figure 5. As indicated in this figure, as the trace angle departs to greater obliquity from condition A then the surface profile becomes flat once the direction has reached condition E. This would give a totally false impression of the true surface topography. If employed in a critical and perhaps highly stressed in-service state, with the user thinking that this incorrectly assessed surface was flat, then, potentially, a premature failure state could arise. Normally, if the production process necessitates, for example, a cross-honing operation,

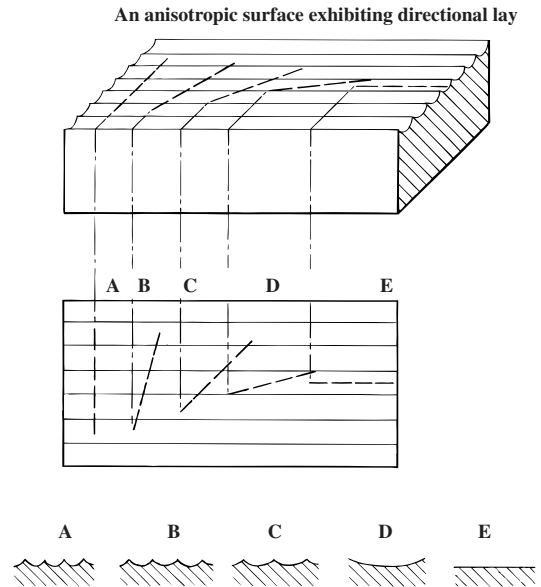


Figure 5. Effect of the relative direction of lay and its associated influence on the measurement of profile shape. (Courtesy of Taylor Hobson.)

the lay condition would be “crossed” – as shown in Figure 4. Under this roughness state, it is usual to measure the surface at 45°, which has the effect of averaging out the influence of the two directions imparted by the cross-honing operation. Where surfaces are devoid of a lay direction – as in the case of “particulate lay” (see Figure 4) from either shot-blasting or the sintering process – then under such circumstances a trace will produce the same surface texture reading, irrespective of the direction of measurement.

The production of certain milled surfaces can sometimes exhibit roughness and waviness conditions at 90° to each other – this is not the same condition as “crossed lay” – because the relative heights and spacings of the two lays differ markedly. Hence, the cusp and chatter marks of the milled topography should be measured in opposing directions. When a turned component has had a “facing-off” operation undertaken, and a “circular lay” condition occurs (see Figure 4), then under these circumstances it is normal to measure the surface texture in a radial direction, otherwise an inappropriate reading will result. Conversely, if a “radial lay” occurs – resulting from “cylindrically grinding” the end face – it is more usual to measure surface roughness at a series of tangent positions with respect to the circumferential direction. When measuring surface texture in the majority of practical situations it can be achieved by direct measurement of the profile, positioning the workpiece in the correct manner to that of the stylus of the surface texture instrument. When the part is

MANUFACTURING PROCESS:	ROUGHNESS (R_a) in μm									
	0.005	0.1	0.2	0.4	0.8	1.6	3.3	6.3	12.5	25
SUPERFINISHING										
LAPPING										
POLISHING										
HONING										
GRINDING										
BORING										
TURNING										
DRILLING										
EXTRUDING										
DRAWING										
MILLING										
SHAPING										
PLANING										

R_a (Nominal) VALUES
in μm :

ROUGHNESS GRADE
NUMBER:

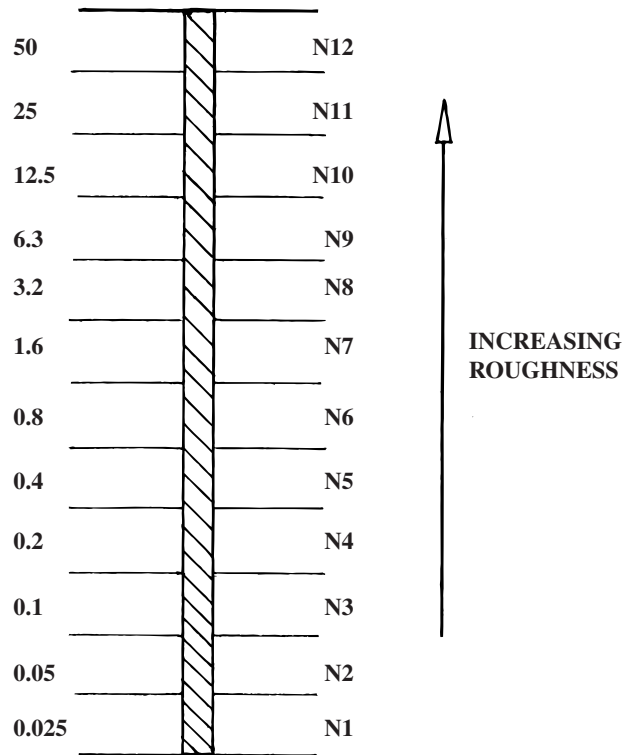


Figure 6. Anticipated process roughness and respective grades. [Source: ISO 1302: 2001.]

either too large or this is impracticable – due to either stylus access to the surface or surface inclination – then placing a small portable instrument on the surface can allow a measurement. Occasions arise when even this is not an option and under these circumstances it may be possible to take a “cast replica” of the surface, then present this to the instrument. Some surface texture measuring instrument manufacturers offer “replica kits” for just this purpose. It should be remembered that this replica of the surface is an inverted image.

1.2 Establishing the R_a numerical value of surface texture from the production process

Numerical data (ISO 1302: 2001) to define the roughness grade for surface texture has been established (see Figure 6) which can be related to the method of production. Caution should be made when using these values for control of the surface condition, because they can misrepresent the actual state of the surface topography, being based solely on a derived numerical value for height. More will be said on this later. Moreover, the “N-number” has been used to establish the arithmetic roughness R_a value, this being just one number to cover a spread of potential R_a values for that production process. Nevertheless, this single numerical value has merit, in that it defines a “global” roughness (R_a) value and accompanying “N-roughness grade”, that can be used by a designer to specify a particularly desired surface condition being related to the manufacturing process. The spread of the roughness for a specific production process has been established from experimental data over the years – covering the maximum expected variance – which can be modified depending upon whether a fine, medium or coarse surface texture is required. Due to the variability in the production process and its stochastic output, such surface texture values do not reflect the likely in-service performance of the part. Neither the surface topography nor its associated integrity has been quantified by assigning to a surface representative numerical parameters. In many instances, “surface engineering” is utilised to enhance specific component in-service conditions, but more will be said on this subject toward the end of Chapter 5.

Previously, it was mentioned that with many in-service applications that the accompanying sur-

face texture is closely allied to functional performance, particularly when one or more surfaces are in motion with respect to an adjacent surface. This suggests that the smoother the surface the better, but this is not necessarily true if the surfaces in question are required to maintain an efficient lubrication film. The apparent roughness of one of these surfaces with respect to the other enables it to retain a “holding film” in its associated roughness valleys. Yet another factor which may limit the designer from selecting a smooth surface is related to its production cost (see Figure 7). If a smooth surface is the requirement then this takes a significantly longer time to produce than an apparently rougher surface, which is exacerbated if this is allied to very close dimensional tolerancing.

1.3 Surface texture roughness comparison blocks and precision reference specimens

In order to gain a basic “feel” for actual replica surfaces in conjunction with their specific numerical value for surface roughness. Then the use of comparison blocks allows an efficient and useful means of interpreting the production process with their expected surface condition. A designer (Figure 8a) can select a surface using a visual and tactile method to reflect the required workpiece condition. Correspondingly, the machine tool operator can utilise similar comparison blocks (see Figure 8b) to establish the same surface condition, without the need to break down the set-up and inspect the part with a suitable surface texture measurement instrument. This is a low-cost solution to on-machine inspection and additionally enables each operator to simply act as an initial source for a decision on the part’s surface texture condition. Such basic visual and tactile inspection might at first glance seem inappropriate for any form of inspection today, but within its limitations it offers a quick solution to surface roughness assessment. Tactile and visual sensing is very sensitive to minute changes in the surface roughness, but is purely subjective in nature and is open to some degree of variation from one person to another. Comparison block surface assessment is purely a subjective form of attribute sampling. Therefore, not only will there be some divergence of opinion between two operators on the actual surface’s numerical R_a value due to their differing sensual assessments, but also this is only a very rudimentary form of inspection.

Precision reference specimens are manufactured to exacting and consistent tolerances, particularly

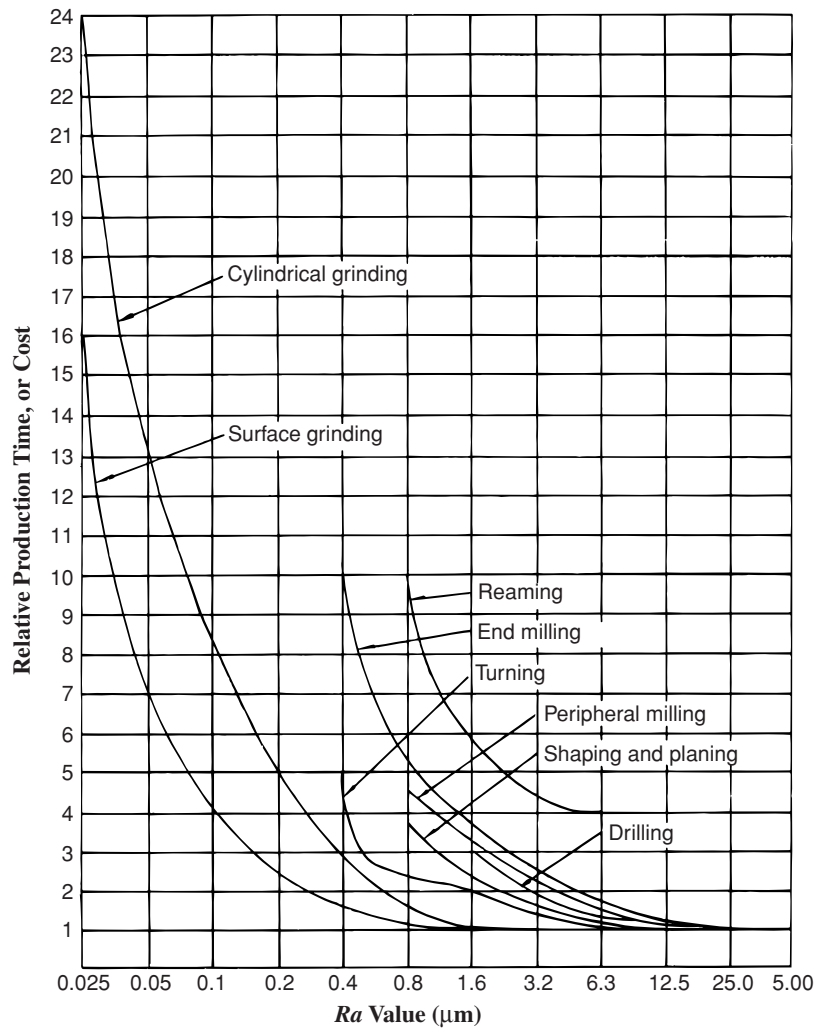


Figure 7. Relative cost, in production time, necessary to produce surface finishes by different production processes. (Courtesy of Taylor Hobson.)

when used to calibrate stylus-based surface texture measuring instruments. A range of standard reference specimens can be purchased for this purpose and will act as a “health check” on both the current stylus condition – whether its point is worn or partially broken – and indicate if the allied instrumentation/electronics has “drifted” since the last calibration check. This enables the inspector to speedily and efficiently remedy the situation and bring the instrument back into calibration. Most of these calibration reference specimens are manufactured from replicas of an original master production surface by nickel electroforming; this enables the block to be a reproduction of the whole surface, with the finest details being approximately $1\ \mu\text{m}$ in size. Each reference specimen offers exacting uniformity of profile shape. In order to minimise wear on these

reference specimens from frequent use a hard boron nitride layer can be specified.

1.4 The basic operating principle of the pick-up, its stylus and skid

Prior to describing some of the more sophisticated instruments by which the surface is measured, it is worth describing the basic function of a typical stylus instrument. In Figure 9, a schematic illustration is depicted of the basic components for a typical surface texture measuring instrument. The

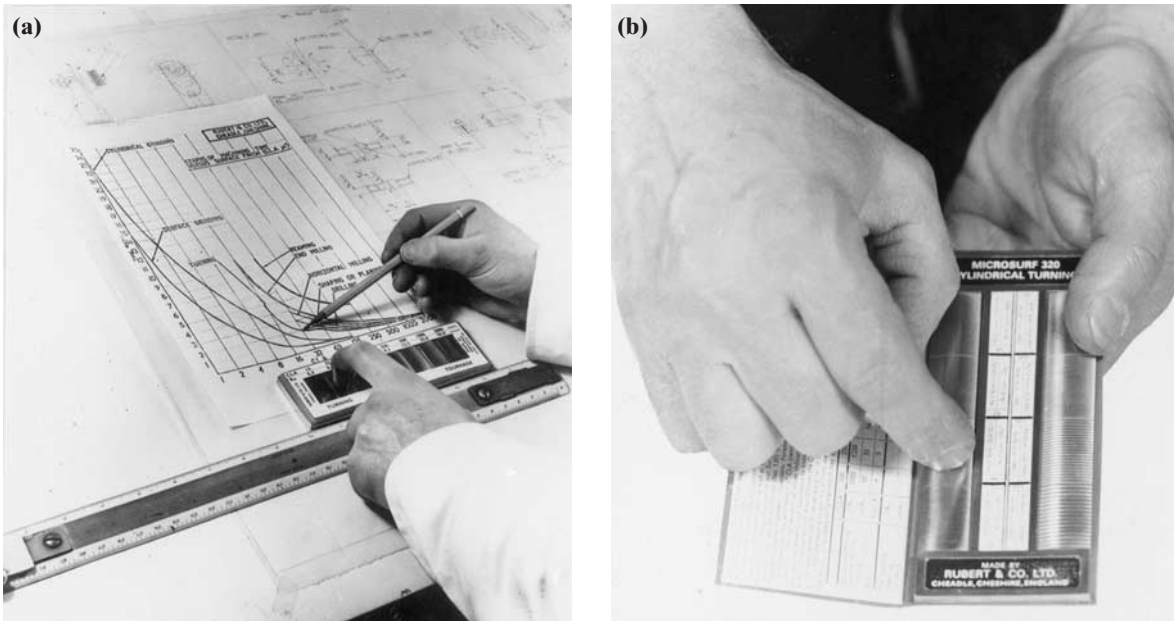


Figure 8. (a) Utilising a comparator gauge to determine the surface finish for a specific manufacturing process. (b) A typical comparator gauge, for both visual and tactile assessment. (Courtesy of Rubert & Co Ltd.)

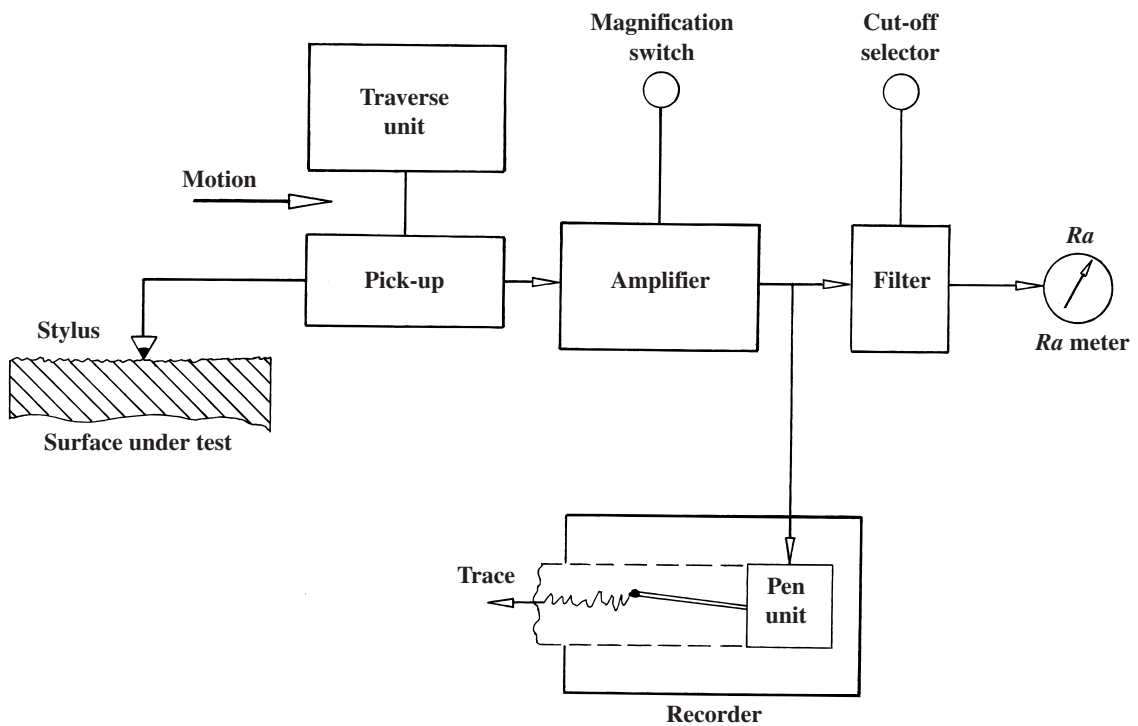


Figure 9. Schematic diagram illustrating the major constituents of a stylus-type of surface texture measuring instrument. (Courtesy of Taylor Hobson.)

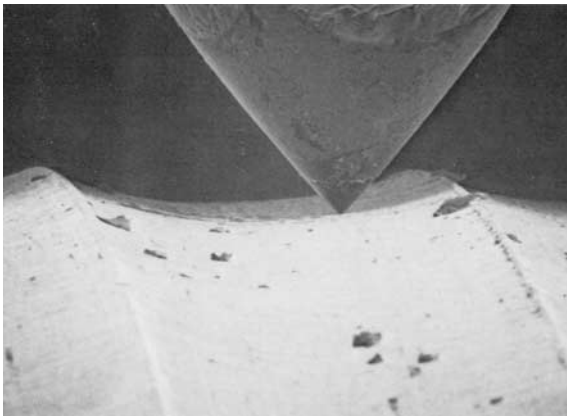


Figure 10. Scanning electron photomicrograph of a stylus (5 μm point radius) on a surface. [Courtesy of Hommelwerke GmbH.]

stylus traverses across the surface and the transducer converts its vertical movements into an electrical signal. This signal is then amplified for subsequent processing, or output to operate a pen recorder. The required parameter value is subsequently derived from the filtered signal, having previously been displayed on screen. In general, the profile is the result of the stylus tracing the *movement* across the surface under test, the contacting of consecutive points of the profile being spaced in *time*. Therefore this relationship between movement and time is closely associated with its “cut-off”. “Cut-off” refers to the limiting wavelength at which components of the profile are passed nominally unchanged by a filter.

The stylus is normally either conical with a spherical tip (see Figure 10), or a four-sided pyramid with a truncated flat tip (Figure 11a). The conical type of styli (Figure 10) have a cone angle of either 60° or 90° , with a tip radius ranging from less than $0.1\ \mu\text{m}$ to $12\ \mu\text{m}$ in size. A pyramidal stylus tends to be approximately $2\ \mu\text{m}$ wide in the traverse direction, but is normally wider transversely to the direction of travel, giving the tip greater strength (see Figure 11a left). The sharp-pointed $0.1\ \mu\text{m}$ wide pyramidal stylus (Figure 11a right), is utilised on the Taylor Hobson “Talystep” and “Nanostep” instruments. However, in use the edges of a pyramidal tip tend to become rounded with wear; conversely, the spherical tip version develops a flat, so the distinction between the two profile geometries becomes less marked with time. If a stylus tip becomes damaged during use, this results in a considerable increase in its width, leading to potentially very serious measuring errors.

When a comparison is made of the scale of an actual surface under test to that of the stylus tip (Figure 10) against a vertical magnification (V_v) of

$\times 5000$ and horizontal magnification (V_h) of $\times 100$, then in this situation the respective size of the stylus is almost indistinguishable from that of a (vertical) straight line or point contact (see Figure 11b). This significant size differential enables the stylus to penetrate into quite narrow valleys, although its finite size affects the accuracy with which the surface profile can be traced. These stylus profile and size limitations influence the following:

- *distortion of peak shape* – as the spherical stylus traverses over a sharp peak, the point of contact shifts across the stylus – from one side to the other (Figure 11c). This results in the stylus having to follow a path which is more rounded than the peak, but due to the stylus being raised to its full height when at the crest the true peak height is measured;
- *penetration into valleys* – due to the tip’s size, it may not be able to completely penetrate to deep and narrow valley bottoms (see Figure 12);
- *re-entrant features cannot be traced* – whenever a stylus encounters a re-entrant feature (Figure 12 detail) the stylus tip loses partial contact with the profile and as a result will remove this feature from the trace. Typical surfaces that exhibit such re-entrant characteristics are powder metallurgy (PM) compacts and other porous parts, together with certain cast irons. This means that misleading surface texture parameter measurements may occur – although recently DIN 4776: 1988 has described a method of obtaining “distortion-free” assessment for PM parts. If transverse sections are taken and positioned within a scanning electron microscope re-entrant features can be visually revealed.

If the slope of a surface texture feature has a steeper angle than that of the half angle of the side of a conical stylus tip, then it will lose contact with the profile, having the same effect as a re-entrant angle on the resulting measurement. The stylus is the only active contact between the surface and the instrument and it has been shown above that its shape and size affect profile trace accuracy. If the force exerted by the stylus is too great, it may scratch or deform the surface, leading to erroneous results. Conversely, the stylus force must be sufficient to ensure that the stylus maintains continuous contact with the surface at the adopted traverse speed. Metrological instrument manufacturers incorporate this into the transducer design, with most surface texture equipment having a tip radius of between 0.1 and $12\ \mu\text{m}$, with a stylus force of $\approx 0.75\ \text{mN}$. This stylus force changes as the tip becomes progressively more blunt as it bears on the surface. An instrument with very fine tip radius (Figure 10a

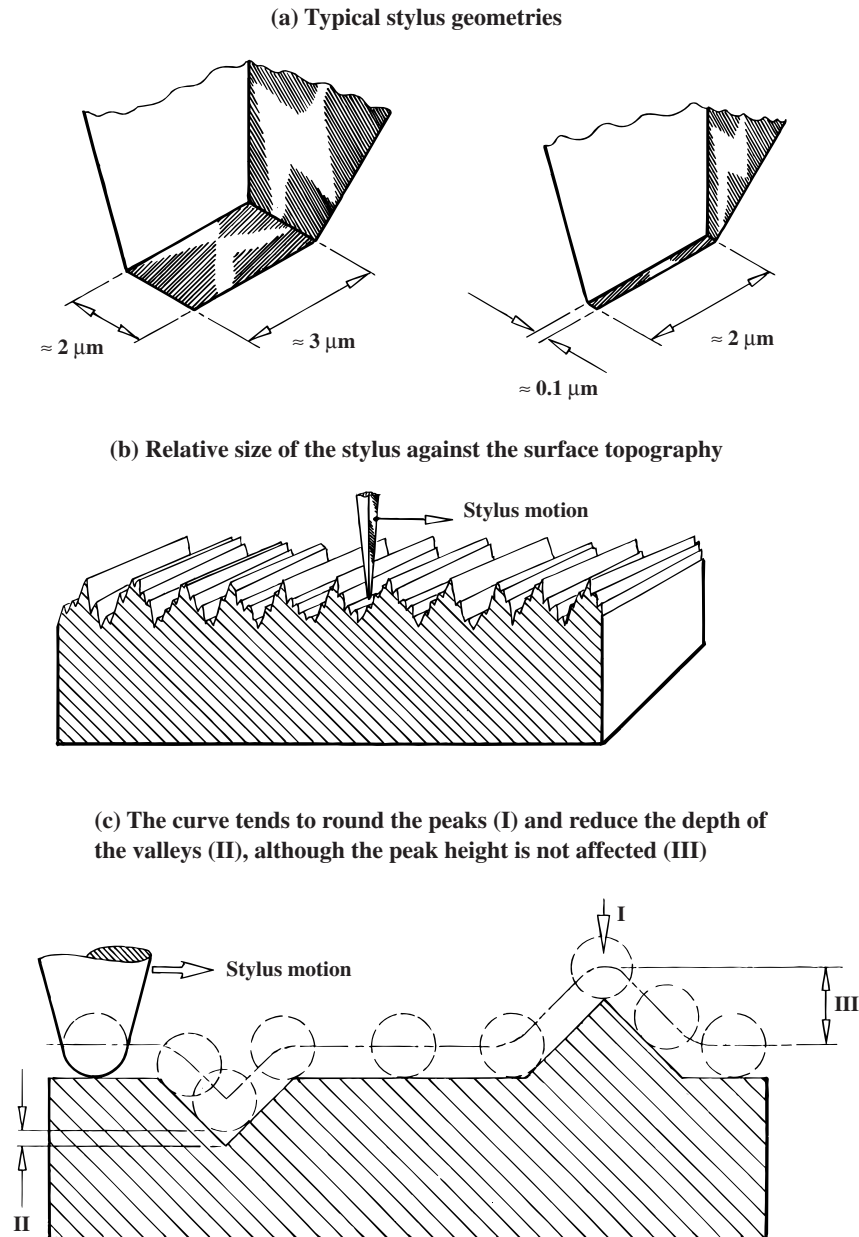
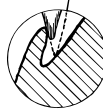
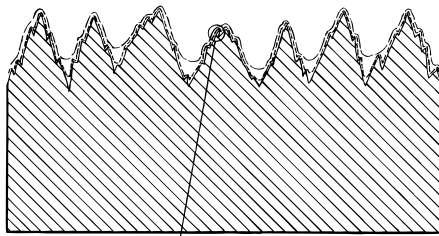


Figure 11. Stylus geometry and its mechanical filtering effect on surface texture. (Courtesy of Taylor Hobson.)

right) must utilise an exceedingly low force, typically $\approx 3 \times 10^{-5}$ N or less. The transducer assembly can incorporate both a stylus and skid, with the skid offering a local datum for this transducer and with that of the surface. Moreover, the skid provides a local datum for the stylus with respect to its vertical and horizontal motion. Therefore, with the skid's large curved radius to that of the relatively small surface under test, it rides along the surface being measured (Figure 13), providing a "local datum". As

long as the skid's radius is greater than the peak spacing, the apparent line of movement of this skid will be virtually a straight line. Under such surface transducer conditions as the vertical skid moves from crest to crest, its relative horizontal skid motion with respect to the test surface can be ignored, as the skid's vertical motion is virtually insignificant. However, once the crests in the surface under test become more widely spaced, this horizontal crest spacing tends to introduce a significant

----- Path traced by a 2.5 μm radius stylus
 - - - - Path traced by a 12.5 μm radius stylus



Re-entrant feature
on a surface

Figure 12. Illustrating how the use of a larger-radius stylus reduces the apparent amplitude of closely spaced irregularities. (Courtesy of Taylor Hobson.)

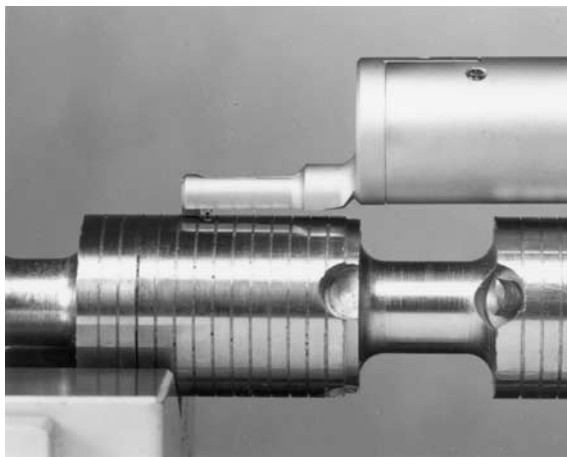


Figure 13. Surface texture of component being inspected using a skid. (Courtesy of Taylor Hobson.)

amount of vertical skid motion. If the skid's vertical motion affects the subsequent surface texture results, it is then preferable to use the instrument in the "skidless mode", as instruments employing a skid should be used for the measurement of roughness parameters (ISO 3274: 1996). To achieve skidless operation and minimise the unwanted affects of vertical skid motion which might otherwise interfere with the results, the skid must be removed, then by utilising the instrument's in-built straight edge as a "sliding datum" the surface assessment can be undertaken, as illustrated in Figure 14. More will be mentioned on the influence of the skid and its associated "phase effect" later.

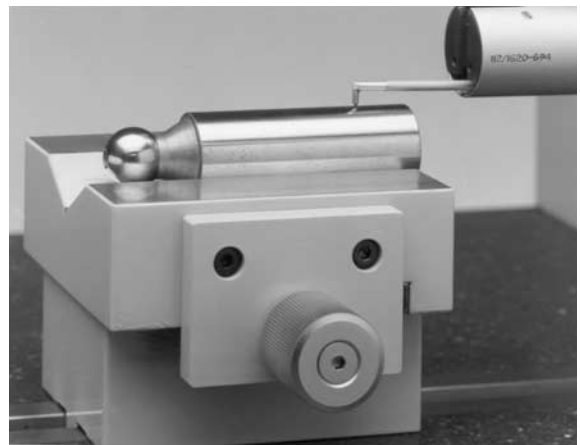


Figure 14. A high-quality component's surface texture being assessed without a skid which tends to be the "approved method", of late. (Courtesy of Taylor Hobson.)

1.5 Filters and cut-off

It was previously suggested that surface geometry is comprised of roughness, waviness and profile (see Figure 2). These interrelated factors also tend to have different relationships to the performance of the component. It is normally the case that they are separated out during analysis. To obtain this separation on surface texture measuring instruments, selectable filters for roughness and waviness are applied. In this section, only a superficial treatment of filtering and its effects will be mentioned, as this will be dealt with in greater depth later in the chapter.

In general, selection of the desired filtering can be established by the following "rules of thumb":

- *roughness filtering* would be applied if control is required of workpiece performance (for example: resistance to stress failure; wear resistance; friction; reflectivity and lubrication properties);
- *waviness filtering* might be selected if control is necessary for workpiece or machine tool performance (for example, vibration or noise generation).

Once the type of filter has been chosen then the filter cut-offs should be selected (see ISO 11562: 1996 for more detailed information) with a range, notably: ". . . mm; 0.08 mm; 0.25 mm; 0.8 mm; 2.5 mm; 8.0 mm; . . ." (ISO 3274: 1996). However, this can only be undertaken when the specific features needing to be measured are known. Generally, the following criteria can be applied:

- a 0.08 mm cut-off length should be selected for high-quality optical components;
- a 0.8 mm cut-off length should be chosen for general engineering components;
- longer cut-offs are necessary when very rough components are to be measured, or if surfaces for cosmetic appearance are important.

NB A practical guideline is that the cut-off chosen should be of the order of 10 times the length of the feature spacings under test. If there is any doubt about the selection of cut-off, initially measure the surface in an unfiltered manner and then make a qualified decision from this test.

So far, the term cut-off has been liberally used, but what is it? As an example of cut-off and assuming a perfect filter characteristic, if a *high-pass filter* has a cut-off of 0.8 mm, this will only allow wavelengths below 0.8 mm to be assessed, with wavelengths above this value being removed. Conversely, when a *low-pass filter* of 0.8 mm cut-off has been selected, this will only allow wavelengths above 0.8 mm to be assessed. As previously mentioned (ISO 3274: 1996), the internationally recognised cut-off lengths for surface roughness measurement are given in Table

Table 1. Typical cut-offs for variously manufactured surfaces

Typical finishing process	Meter cut-off (mm)				
	0.25	0.8	2.5	8.0	25.0
Milling		x	x	x	
Boring		x	x	x	
Turning		x	x		
Grinding	x	x	x		
Planing			x	x	x
Reaming		x	x		
Broaching		x	x		
Diamond boring	x	x			
Diamond turning	x	x			
Honing	x	x			
Lapping	x	x			
Superfinishing	x	x			
Buffing	x	x			
Polishing	x	x			
Shaping		x	x	x	
Electro-discharge machining	x	x			
Burnishing		x	x		
Drawing		x	x		
Extruding		x	x		
Moulding		x	x		
Electro-polishing		x	x		

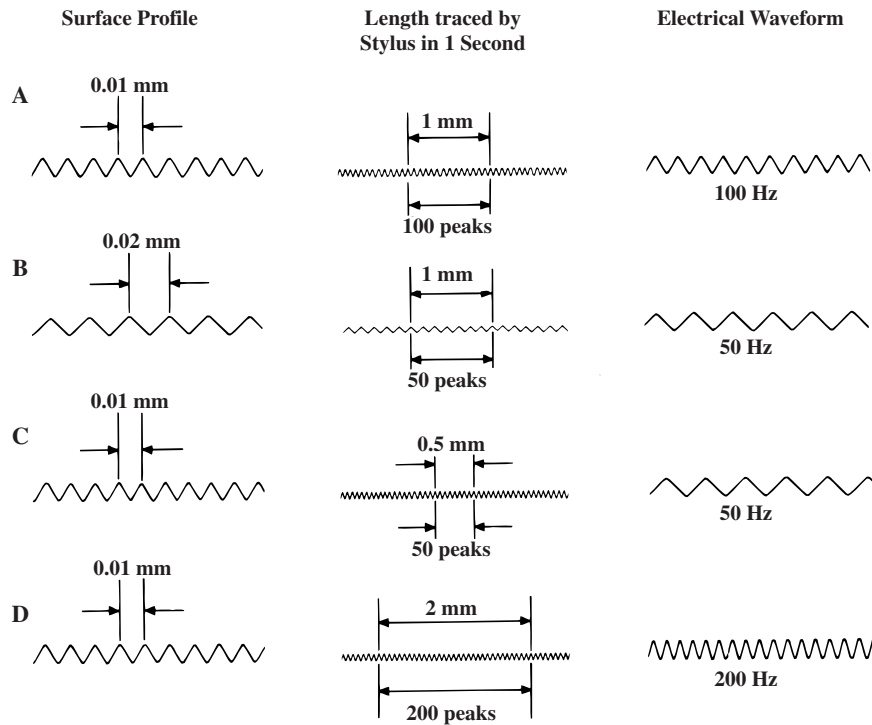
NB The cut-off selected must be one that will give a valid assessment of the surface characteristic of interest. For example, although a cut-off of 0.8 mm can validly be used for almost all of these surfaces, it may not necessarily be suitable for assessing a particular feature of the texture. This entails examining the surface and considering the purpose of the measurement before selecting the cut-off. (Courtesy of Taylor Hobson.)

1. One basic feature of stylus-based instruments is that the profile is traced by *movement* across the test surface, with the consecutive points of the profile in contact being spaced in *time*. Hence, this relationship between movement and time is closely associated with cut-off.

The frequencies present are dependent upon two factors: traverse speed and spacing irregularities. In Figure 15 (A and B), the frequencies of 100 Hz and 50 Hz represent irregularity spacings of 0.01 mm and 0.02 mm respectively. At the same traverse speed, frequencies of 4.0 Hz and 1.25 Hz would be generated by spacings of 0.25 mm and 0.8 mm. If a high-pass electrical filter is used, this will suppress any frequency below 4.0 Hz, enabling only those irregularities having spacings of 0.25 mm or less to be represented in the filtered profile. Therefore, this relationship is ideal to obtain a sampling length of 0.25 mm, hence the term “cut-off” length – since the response now cuts off at irregularity spacings of 0.25 mm, denoted by the international symbol, (λ_c) for cut-off length. By introducing different filters, the most suitable cut-off for the surface can be selected, where Table 1 acts as a guide to the selection of suitable cut-offs for various machining processes.

Until 1990, standard roughness and waviness filters were analogue or 2RC (two resistors and two capacitors) types. Since 1996, the standard (ISO 11562: 1996) has specified the digital filter for surface texture measurements. For example, if a 0.8 mm cut-off value is utilised, then there is a 50% transmission of the profile equal to this 0.8 mm. This type of filtering is analogous to a sieve, only allowing values below the cut-off to be passed through. The main problem with utilising an analogue filter is that it assumes that the surface is basically sinusoidal in nature, but this rarely is the case. This mismatch with the “real” surface under test causes the so-called “Gibb’s phenomenon”. The Gibb’s phenomenon results in an overshooting of the filtered profile when analysing certain extreme forms of profile. The “Gibb’s phenomenon” can be explained by the following example.

Filtering is used to determine the mean line of the roughness profile – being identical to the waviness profile. An analogue filter determines a profile’s mean line by a “moving average” technique. Hence, at any point in time the filter averages the profile previously traversed by the stylus. This has been equated to a person walking backwards! Such a filter cannot anticipate what lies ahead of it in the remainder of the profile, such as abrupt changes in the topography. Under these conditions, it reacts after a significant change has occurred; this means that the true profile is no longer represented by the mean line, and as a result the filtered profile will be distorted vertically



Traverse speed: A = 1 mm s⁻¹; B = 1 mm s⁻¹; C = 0.5 mm s⁻¹; D = 2 mm s⁻¹;

Figure 15. Signal frequency depends on irregularity spacing (compare A and B) and traverse speed (compare A, C and D). (Courtesy of Taylor Hobson.)

and horizontally shifted. Such an action by the analogue filter may lead to serious errors for either plateau honed types of surface or PM (porous) types of topographies.

If a digital filter is employed, such as that shown on the graph in Figure 16(a), then the vertical distortions of the profile caused by abrupt changes in the height of the topography are minimised. The principle operation of the digital filter can be explained as below.

Digital profiles are no longer represented by a smooth curve, but instead, are described by a series of ordinates (numbers) relating to profile heights at regular intervals. In order to estimate the profile's mean line the filter's action is to calculate the average height at a given point, as the arithmetic average of the ordinate points in its vicinity. Hence, if the profile is divided up into "windows" the mean profile height for each window can be estimated. A line joins these window averages together, which represents the profile mean line.

Such a digital filter allows up to 100% transmission of the profile, representing wavelengths shorter than the actual cut-off length, together with zero transmission of wavelengths greater than this cut-

off. Figure 16(bi) shows how the unfiltered profile equates to an electrically filtered one – Figure 16(bii) – for the same value of cut-off, but related to 50% of its depth. The main difference between the (PC) digital filter and the analogue filter is that the former type evaluates the profile mean around a point – enabling it to anticipate changes in the profile height – whereas the latter (analogue) filter cannot. It is possible to emulate the analogue filter's characteristics by employing a "weighting function" for each ordinate height when a determination of the arithmetic average or mean point is required. However, unless there is a large variation in the surface topography, or special processing conditions apply – such as the "Gibb's phenomenon" – then the analogue-to-digital differences remain slight.

1.6 Measuring lengths

In order to determine a workpiece's surface texture, three characteristic lengths are associated with the profile (ISO 4287: 1997). These are:

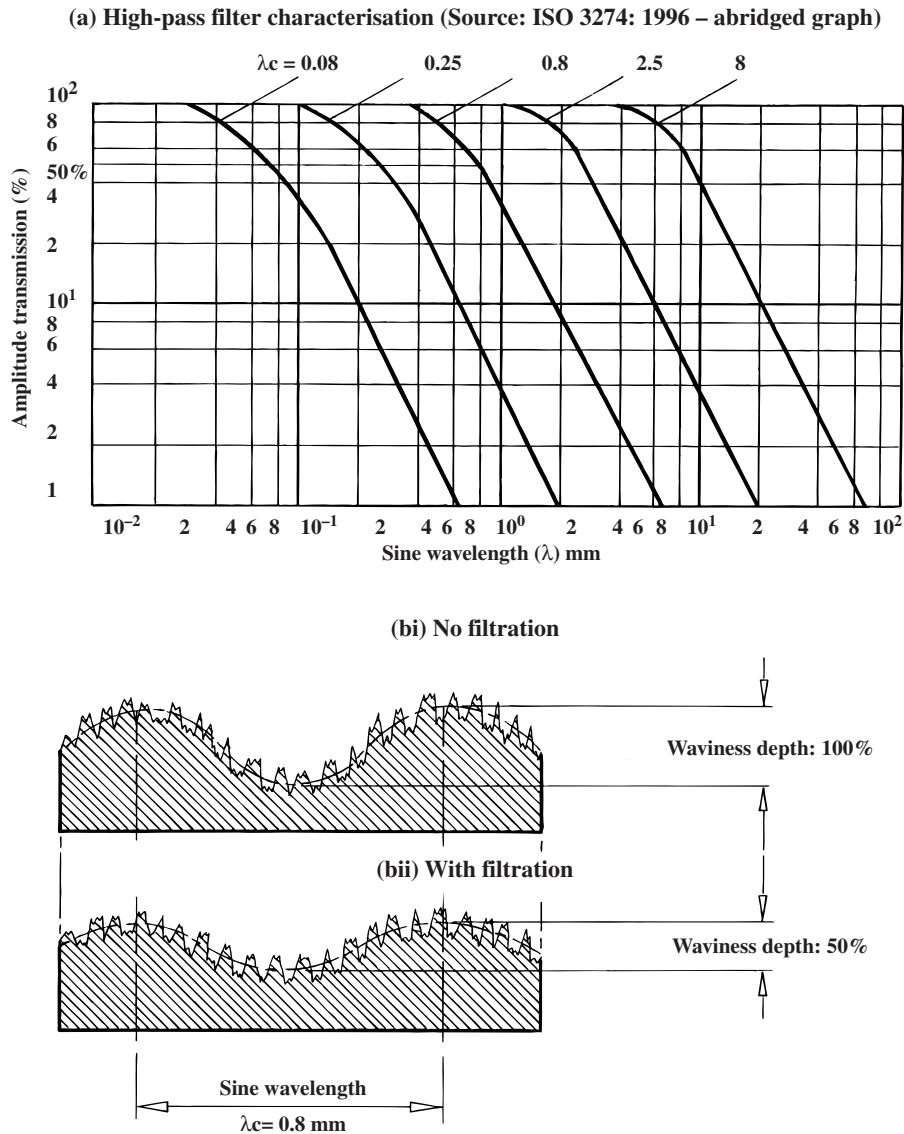


Figure 16. Effect of an electrical filter on the surface texture. (Courtesy of Taylor Hobson.)

- sampling length;
- evaluation (or assessment) length;
- traverse length.

Sampling length (l_p , l_r and l_w)

The surface texture profile illustrated in Figure 17(a) has both roughness and waviness present, with both patterns being generally repetitive in nature. Therefore the roughness pattern A over the distance L might be typical for the whole length of the surface, only differing in minor detail from that indicated at B over a similar distance.

NB If a shorter distance T was selected, the roughness patterns at C and D would not be identical. This would give a misleading picture of their respective roughness heights.

What was true for the reduced distances also applies to waviness, neglecting the roughness (Figure 17b). This is repetitive over distance U, although if the examination is confined to a shorter distance V then the waviness at E and F appear to be markedly different in character. Such a variation demonstrates why it is necessary to select the length of the profile over which a parameter is to be determined, termed its “sampling length”, to suit the surface under test. Yet another feature concerning

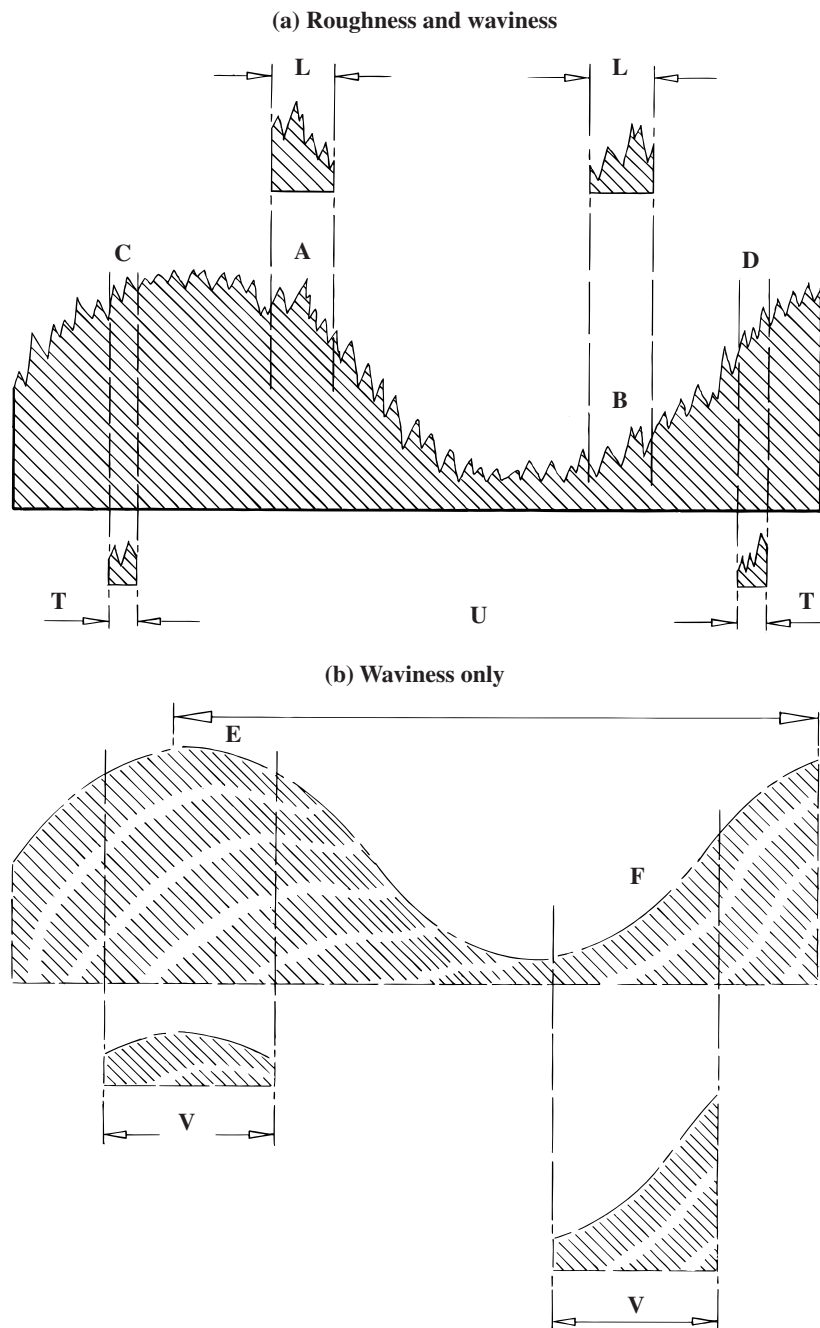


Figure 17. Effect of different sampling lengths. (Courtesy of Taylor Hobson.)

selection of appropriate lengths can be demonstrated in Figure 17(a). Figure 17(a) indicates that while the sampling length L is sufficient to reveal the whole of the roughness pattern, the waviness has very little influence over this length. Even if the sampling length was increased, it would simply include more roughness detail but would have little, if any, effect on a roughness parameter value. In this

case, the waviness would play a greater role and for this reason it is undesirable to increase the sampling length much beyond that required to obtain a representative assessment of roughness.

The sampling length can be defined as *the length in the direction of the x-axis used for identifying the irregularities that characterise the profile under evaluation*. By specifying a sampling length, this implies

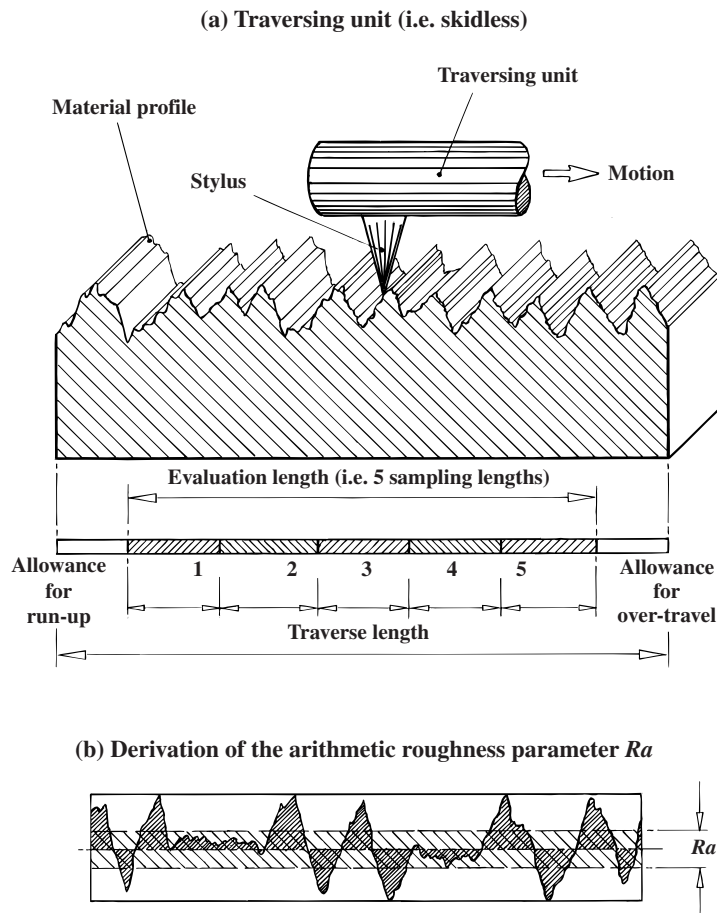


Figure 18. Determining surface texture analysis roughness parameter R_a , together with its arithmetic derivation.

that structure in the profile occurring over longer lengths will not be applicable in this particular evaluation. The sampling length for l_r – roughness – is equal (numerically) to the wavelength of the profile λ_c , whereas the sampling length for l_w – waviness – equates to that of the profile filter λ_f . Virtually all the parameters necessitate evaluation over the sampling length; however, reliability is improved by taking an average of them from several sampling lengths as depicted by the evaluation length in Figure 18(a).

Evaluation length (l_n)

A simplistic diagram of the relationship of the stylus to the material profile/topography for a pre-selected traverse length is indicated in Figure 18(a). The evaluation length can be defined as *the total length in the x-axis used for the assessment of the profile under evaluation* (see Figure 18a). As shown, this length may include several sampling lengths – typically five – being the normal practice in evaluating roughness

and waviness profiles. The evaluation length measurement is the sum of the individual sampling lengths. If a different number is used for the roughness parameter assessment, ISO 4287: 1997 advocates that this number be included in the symbol, for example Ra_6 . In the case of waviness, no default value is recommended. From a practical viewpoint, the selection for the correct filter can normally be ensured when at least 2.5 times the peak spacing occurs, with two peaks and valleys within each one. Typically, this would mean that an evaluation length of 0.8 mm would be selected, but there are occasions when either a larger or smaller evaluation length might be preferable for the surface under test. The metrologist's experience and judgement will come into play here.

Traverse length

The traverse length can be defined as *the total length of the surface traversed by the stylus in making a*

measurement. It will normally be greater than the evaluation length, due to the necessity of allowing run-up and over-travel at each end of the evaluation length to ensure that any mechanical and electrical transients, together with filter edge effects, are excluded from the measurement (see Figure 18a). On shorter surfaces, it may be necessary to confine the measurements to one sampling length, under these circumstances the sampling and evaluation lengths are identical, but the over-travel necessary to contribute to the traverse length must be retained.

1.7 Filtering effects (λ_s , λ_c and λ_f)

In Figure 19 a flow chart is illustrated showing the manner in which surface assessment occurs, with the primary profile broken down into filtered elements to obtain the waviness and roughness profiles of the workpiece surface. Algorithms enable suitable characteristic functions and parameters to be obtained, scaled to an appropriate size. As has been mentioned previously in Section 1.5, filters are vital to any form of surface texture analysis. Today, most forms of filtering are normally electrical or computational, although they can occasionally be mechanical for analysis of the range of structure in

the total profile, which can be adjudged to be of most significance in a particular situation. Conversely, the effects of filtering can be considered as a means of removing irrelevant information, such as instrument noise and imperfections. Therefore, filters can select or reject surface topographical structure depending upon its scale in the x-axis, in terms of spatial frequencies, or wavelengths. The terms *low-pass* and *high-pass filters* refer to whether they either reject or preserve surface data. For example, a *low-pass filter* will reject short wavelengths while retaining longer ones, whereas a *high-pass filter* preserves shorter-wavelength features while rejecting longer ones. The term *bandpass filter* refers to the combination of a low-/high-pass filter selected to restrict the range of wavelengths, with both high- and low-pass regions being rejected. Filter attenuation (rejection) should be somewhat gradual, otherwise significantly differing results occur from almost identical part surfaces, the exception being when a significant surface feature causes a slight wavelength shift.

The term *cut-off* refers to the 50% transmission (and rejection) wavelength filter, this being specific to the topic of surface texture. In the case of the large majority of surface texture assessments, it is suggested that when the width of a specific surface feature is significant, but its size may only be 1% of the overall width, this will make it less important. Under these circumstances that affect feature transmission/rejection, it is suggested that *bandpass*

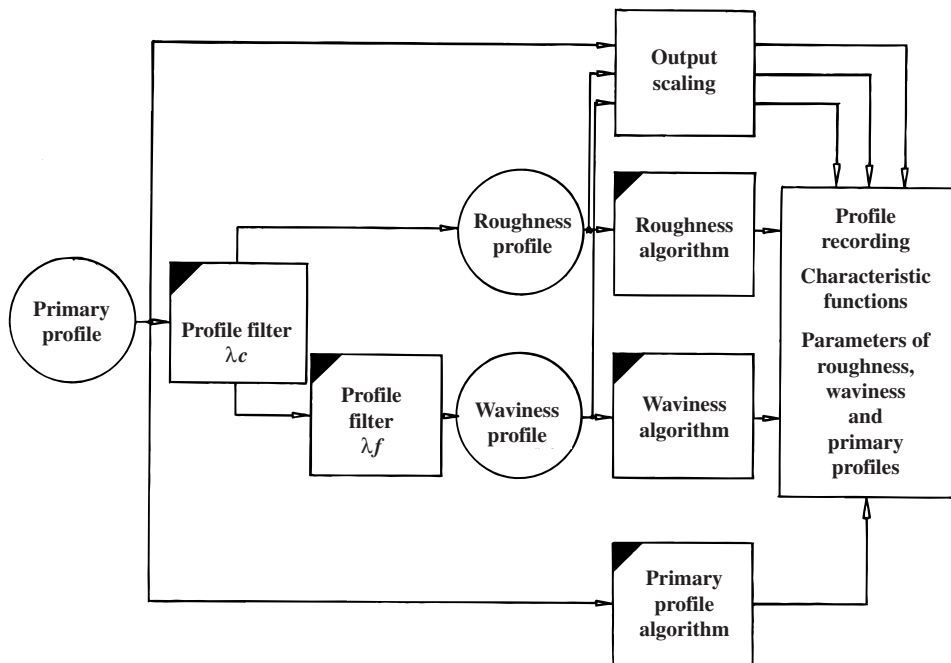


Figure 19. A flowchart for surface assessment. [Source: ISO 4287: 1997.]

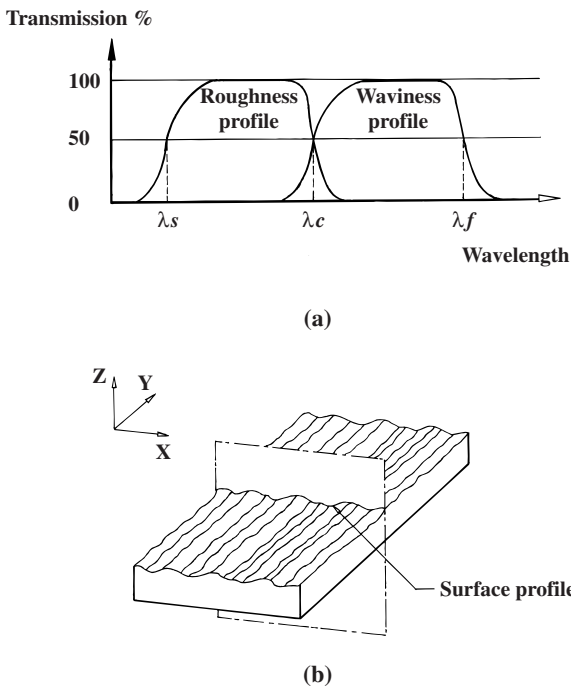


Figure 20. (a) Transmission characteristic of roughness and waviness profiles. (b) Surface profile. [Source: ISO 4287: 1997]

filtering is vital to any potential analysis of surface texture.

Profile filter

The “profile filter” can be defined (ISO 4287: 1997) as a *filter that separates profiles into long- and short-wave components* (see Figure 20a). In particular, three filters can be employed for surface measurement (Figure 20a):

- *profile filter* (λ_s) – this filter defines where the intersection occurs between the roughness and the presence of shorter-wavelength components in a surface;
- *profile filter* (λ_c) – this filter defines the intersection position between the components of roughness and waviness;
- *profile filter* (λ_f) – filters of this type define the intersection point between waviness and the presence of any longer wavelengths in a surface.

Primary profile

The *primary profile* forms the basis for primary profile parameter evaluations, being defined as *the total profile after application of a short-wavelength*

(low-pass) *filter, with a cut-off* λ_s . The finite size of a stylus will eliminate any very short wavelengths, which practically is a form of mechanical filtering, being in the main utilised as a default for the λ_s filter. As the size of styli will vary and the instrument can introduce vibration and/or supplementary noise into the profile signal (having equivalent wavelengths to that of the stylus dimensions) under these circumstances, it is recommended to ignore any λ_s filtration along the total profile.

Roughness profile

The *roughness profile* can be defined as *the profile derived from the primary profile by suppressing the long-wave component using a long-wavelength (high-pass) filter, with a cut-off* λ_c . This roughness profile provides the foundation for the evaluation of the roughness profile parameters and will automatically include λ_f profile filter information, because it is derived from the primary profile.

Waviness profile

The derivation of this waviness profile is the result of the application of the bandpass filter to select the surface structure at somewhat longer wavelengths to that of roughness. The *filter* λ_f will suppress any longer-wave components, namely the *profile* component, while the *filter* λ_c suppresses the shorter-wave components, such as any of the *roughness* components. Hence, the waviness profile forms the basis for the evaluation of the waviness profile parameters.

Roughness profile mean line

The roughness profile mean line is a reference line for parameter calculation, which corresponds to the suppression of the long-wave profile component by the profile filter λ_c .

Waviness profile mean line

The waviness profile mean line is a reference line for parameter calculation, which corresponds to the suppression of the long-wave profile component by the profile filter λ_f .

Primary profile mean line

The primary profile mean line is a reference line for parameter calculation, being a line determined by

fitting a least-squares line of nominal form through the primary profile.

Surface profile

In ISO 4287: 1997, the coordinate system is defined by making use of the rectangular coordinate of a right-handed Cartesian set for the surface profile (Figure 20b). Here, the x -axis provides the *direction* of the trace; the y -axis nominally lies on the *real surface*, with the z -axis facing in an *outward direction* from the material to the surrounding medium. Thus, the *real surface* can be defined as *the surface limiting the body and separating it from the surrounding medium*. The *lay* previously alluded to in Section 1.1 is normally functionally significant for the workpiece, so it is important to specify it on engineering drawings in terms of the type of *lay* and its direction (ISO 1302: 2001 – Appendix B).

1.8 Geometrical parameters

The following discussion is concerned with the calculation of parameters from the measurement data. These parameters are all derived *after* the *form* has been removed. It should be stated that any parameters selected should be appropriate for a given application and that not all of them would be necessary in all circumstances. On an engineering drawing, a designer may have specified a parameter to define the part's functional behaviour, which means that this parameter is outside the control of the user of the instrument. When surface texture parameters have been specified, the user needs to have a clear and unambiguous understanding of the manner in which they are calculated.

In understanding and evaluating surfaces, the philosophy behind the reasons why *peaks* and *valleys* are significant is important, although just how and what represents a peak/valley is not always clear (see further discussion on this topic in the Appendix). When a surface topography shows the presence of a peak or valley that is not significant, the decision concerning its importance for in-service functional characteristics is not always a simple matter to establish. In the surface profile shown in Figure 21(a), for example, are both the peaks here of importance, or even significant? In order to minimise any confusion for the user, the latest standards have introduced the important concept that the *profile element* consists of a *peak and valley event*. Moreover, associated with this

profile element is a discrimination that prevents any minute, unreliable measurement features from affecting the detection of these elements. Several of these surface texture-related elements and associated features are described in ISO 4287: 1997, consisting of:

- *profile element* – the section of a profile that crosses the mean line to the point at which it next crosses this mean line in the same direction – for example, from below to above the mean line;
- *profile peak* – the section of the profile element that is above the mean line, namely, the profile by which it crosses the mean line in the positive direction until it once again crosses this mean line, but now in the negative direction;
- *profile valley* – as described for the “peak” above, but with the direction reversed;
- *discrimination level* – within a profile it may be likely that a minute fluctuation occurs which takes the profile trace across the mean line, then back again almost immediately. In such a condition the departure of the trace from the mean line is not considered to be a real profile peak or valley. In order to prevent automatic systems from counting small trace departures, allowing them to only consider those larger surface features, they specify particular heights and widths to be counted. Hence, default levels are established in the absence of other specifications. These default levels are set such that their profile peak heights or valley depths must exceed 10% of the R_z , W_z or P_z parameter value. Furthermore, the width of either a profile peak or valley must exceed 1% of the sampling length, with both criteria being simultaneously met;
- *ordinate value* $Z(x)$ – this represents the height of the assessed profile at any position x , its height being regarded as negative when the ordinate lies below the x -axis, and otherwise positive;
- *profile element width* X_s – this is concerned with the x -axis segment intersecting with the profile element (see Figure 21a);
- *profile peak height* Z_p – this represents the distance between the mean line on the x -axis and the highest point of the largest profile peak (see Figure 21a);
- *profile valley depth* Z_v – this represents is the distance between the mean line on the x -axis and the lowest point of the deepest valley (see Figure 21a);
- *profile element height* Z_t – this equates to the sum of the peak height and valley depth of the profile element, namely, the sum of Z_p and Z_v (see Figure 21a);
- *local slope* dZ/dX – this represents the slope of the assessed profile at position x_i (see Figure 21b).

The local slope's numerical value critically depends upon the ordinate spacing and thus influences $P\Delta q$, $R\Delta q$ and $W\Delta q$;

- *material length of profile at level "c"* $Ml(c)$ – this represents the section lengths obtained when intersecting with the profile element by a parallel line to the x -axis at a given level "c" (see Figure 21c).

1.9 Surface profile parameters

The following examples of defined parameters can be calculated from any profile. The designation of letters follows the logic that the parameter symbol's first capital letter denotes the type of profile under evaluation, for example the Ra is calculated from the roughness profile, while Wa has its origin from the waviness profile, with the Pa occurring from the primary profile. In the subsequent sections, the definitions and schematic representations of many of these surface profile parameters are described along with examples of their usage.

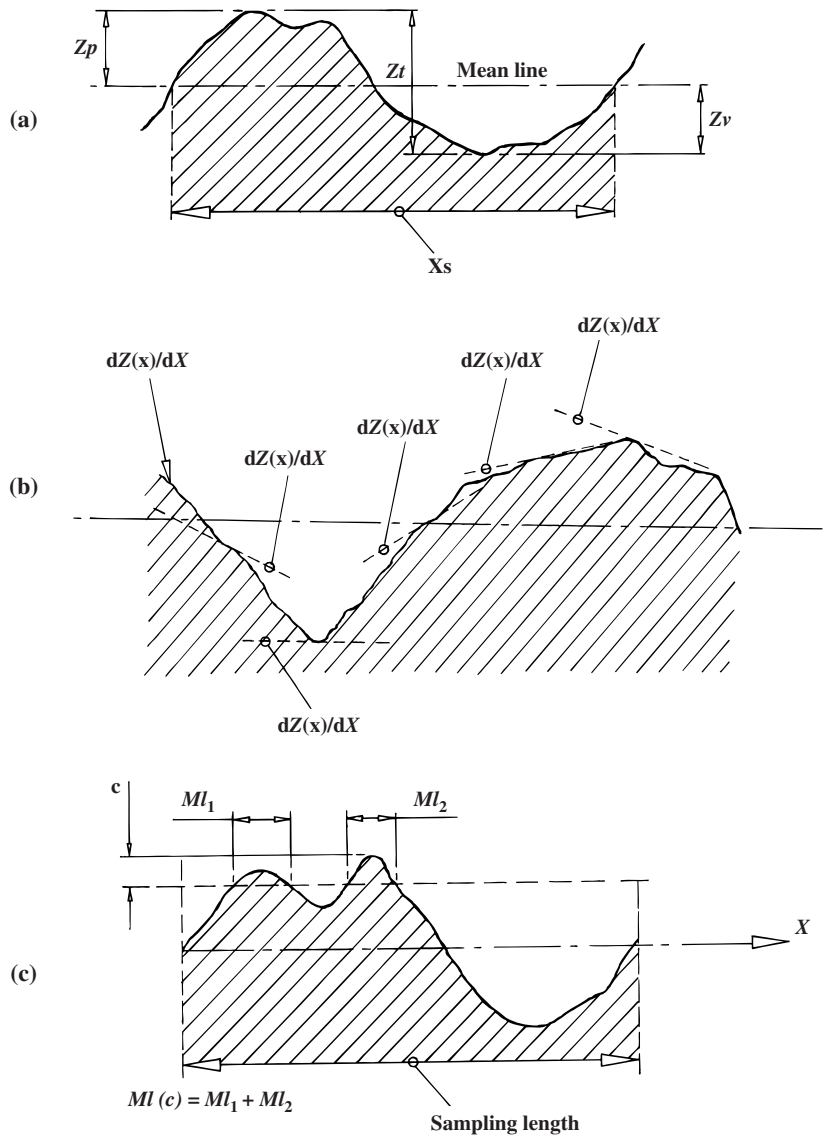


Figure 21. (a) Profile length, (b) local slope and (c) material length of a surface trace. [Source: ISO 4287: 1997]

1.9.1 Amplitude parameters (peak-to-valley)

Maximum profile peak height R_p , W_p and P_p

This parameter is represented by the largest peak height (Z_p) within the sampling length, with its height being measured from the mean line to the highest point (see Figure 22a). The R_p parameter (equating to roughness) is generally less favoured, with preference given to parameters based on the total peak-to-valley height. Often R_p and its associated parameters W_p and P_p are referred to as extreme-value parameters, being somewhat unrepresentative of the overall surface, because their numerical value can vary between respective samples. In order to minimise variation in R_p , it is feasible to average readings over consecutive sampling lengths, although in most cases the value obtained is numerically too large to offer practical assistance. The R_p and its W_p and P_p derivatives are

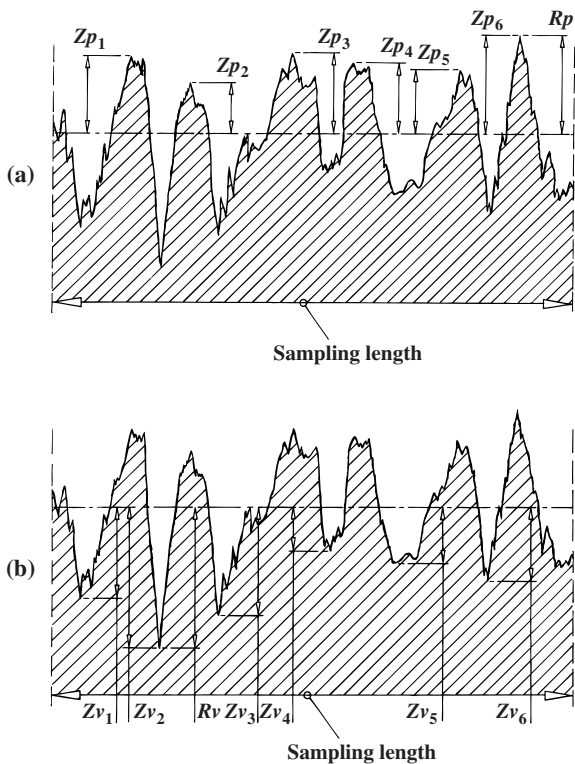


Figure 22. (a) Maximum profile peak height and (b) maximum profile valley depth (examples of roughness profiles). [Source: ISO 4287: 1997]

not without some merit, as they can establish unusual surface features such as either a burr on the surface or sharp spike, and will indicate the presence of scratches or cracks – possibly indicative of low-grade material or poor production processing.

Maximum profile valley depth R_v , W_v and P_v

This parameter is represented by the largest valley depth (Z_v) within the sampling length, with its absolute value R_p (equating to roughness) being obtained from the lowest point on the profile from the mean line (see Figure 22b). As in the case above for peaks, the maximum profile valley depths are extreme-value parameters. In a similar manner to R_p , R_v can establish whether there is a tendency to workpiece cracking, spikes and burrs, detecting such features on the surface.

Maximum height of profile R_z , W_z and P_z

The maximum height parameter (for example, R_z – for roughness) is the sum of the height of the largest profile peak height (Z_p), together with the largest profile valley (Z_v) within the sampling length (see Figure 23a). In isolation, the R_z does not provide too much useful surface information, and therefore it is often divided into R_p and R_v , previously described. In fact, in ISO 4287: 1984, the R_z symbol indicated the *ten point height irregularities*, with many of the previous surface texture measuring instruments measuring this R_z parameter – see the Appendix for more information regarding this previously utilised parameter.

Mean height of profile elements R_c , W_c and P_c

Parameters of this type evaluate surface profile element heights (Z_t) within the sampling length (see Figure 23b) and to obtain this parameter it requires both height and spacing discrimination, previously mentioned. When these values are not specified then the default height discrimination used shall be 10% of R_z , W_z and P_z , respectively, with the default spacing discrimination being stated as 1% of the sampling length. Both of these height and spacing requirements must be met. Normally in practice it is extremely rare to utilise this parameter because of its difficulty in interpretation and, as a result, it would rarely be used on an engineering drawing.

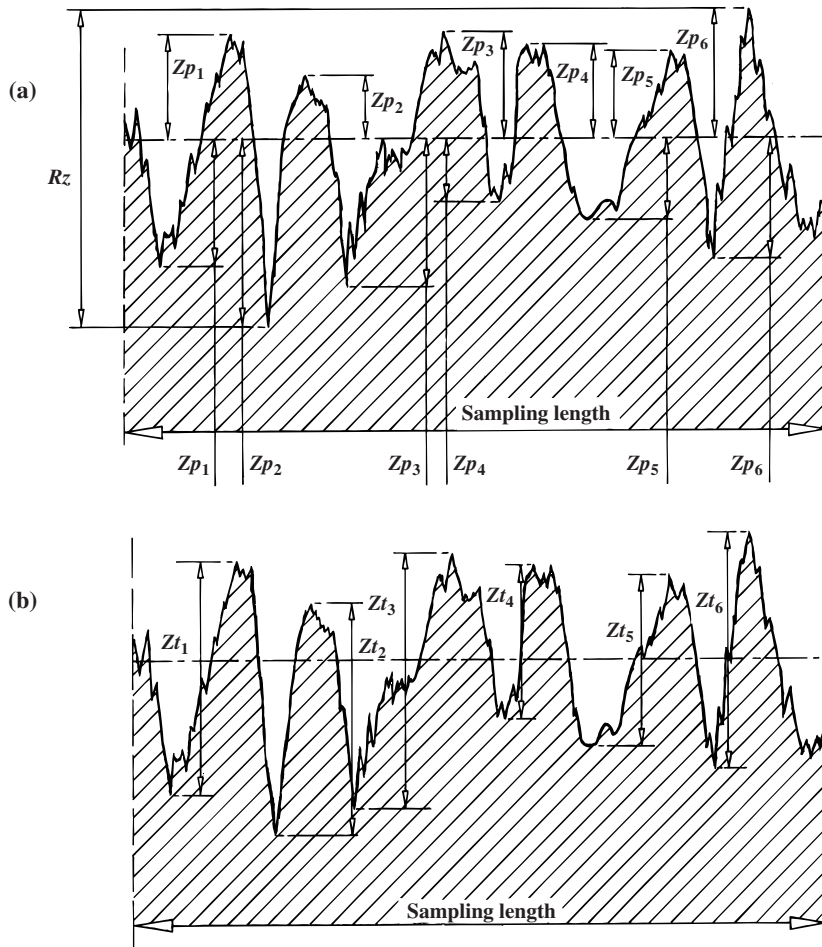


Figure 23. (a) Maximum height of profile and (b) height of profile elements (examples of roughness profiles). [Source: ISO 4287: 1997]

Total height of profile Rt , Wt and Pt

This parameter relates to the sum of the height of the largest peak height (Zp) and the largest profile valley depth (Zv) within the evaluation length. It is defined over the evaluation length rather than the sampling length and in this manner has no averaging effect. This lack of averaging means that any surface contamination (dirt) or scratches present will have a direct effect on the numerical value of, say, Rt .

1.9.2 Amplitude parameters (average of ordinates)

Arithmetical mean deviation of the assessed profile Ra , Wa and Pa

The arithmetic mean parameter is the absolute ordinate value $Z(x)$ within the sampling length. The numerical value of Ra is able to vary somewhat without unduly influencing the performance of a surface. Often on engineering drawings a designer will specify a tolerance band, or alternatively a maximum value for Ra that is acceptable. The mathematical expression of this parameter is given below for these absolute ordinate values as:

$$Z_a = \frac{1}{N} \sum_{i=1}^N |Z_i|$$

where N = number of measured points in a sampling length (see Figure 24b).

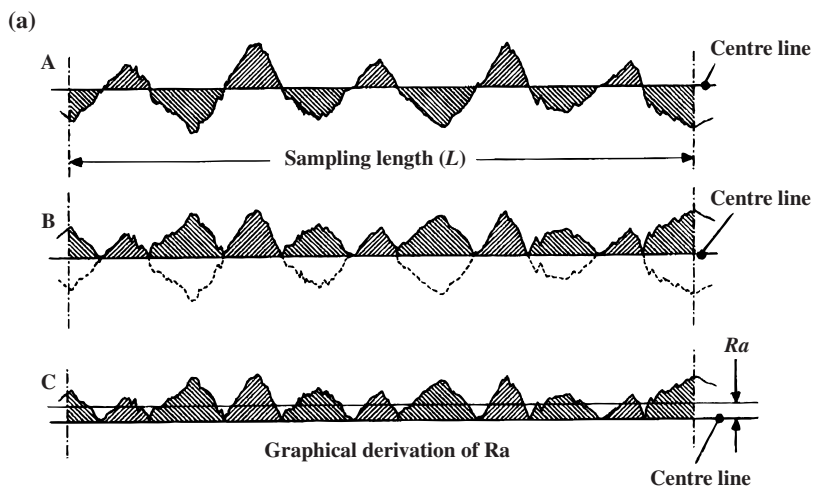
In Figure 24(a) is illustrated the graphical derivation of this parameter within the sampling length, with the shaded areas of the graph below the centre line in A being repositioned above this centre line in B. Here, the R_a value is the mean line of the resulting profile indicated in C.

The absolute average roughness over one sampling length is given by the R_a value. This means that the influence of a discrete but non-typical peak or valley on the numerical value of R_a is not really significant. Good metrological practice is to ensure that a number of assessments of R_a occur over consecutive sampling lengths, enabling the acceptance of an average of these values. This sampling strategy ensures that the numerical value of R_a is typical for the inspected surface. As has been previ-

ously mentioned on several occasions, if a defined “unidirectional lay” is present on the surface (Figure 4), then any measurements should be undertaken perpendicularly to this “lay” if misinterpretation of the surface is to be avoided (Figure 5).

The numerical value of R_a does not impart any information regarding the surface shape, nor of any irregularities present. In fact, it is quite possible to have identical R_a values for surfaces that are markedly different in their profiles (Figure A in the Appendix) and this will affect their potential in-service performance, with engineering drawings often quoting the type of production process (Figure 3c). Historically, in Great Britain the most widely used of all the surface texture parameters has been R_a , but this should not discourage industrial and research organisations from using others, as they may offer more in terms of information regarding a surface’s functional performance.

The mathematical derivation of both the commonly utilised parameters R_a and R_q are given below (see Figure 24b):



Derivation of the arithmetical mean deviation

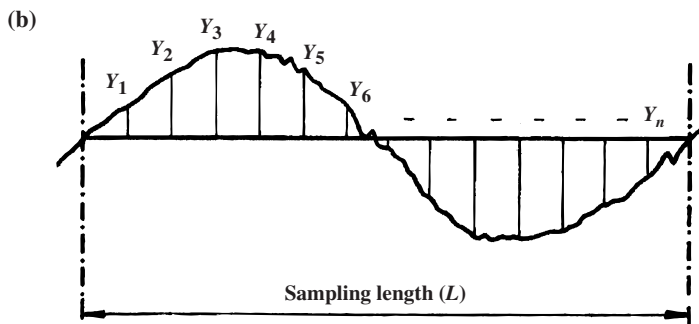


Figure 24. Determination of R_a and R_q . (Courtesy of Taylor Hobson.)

$$Ra = |Y_1| + |Y_2| + |Y_3| \dots + |Y_n|/n$$

$$Rq = \sqrt{(Y_1^2 + Y_2^2 + Y_3^2 \dots + Y_n^2)/n}$$

Root mean square deviation from the assessed profile Rq , Wq and Pq

The root mean square of the ordinate values $Z(x)$ within the sampling length is established by the departures from mean line of the profile and can be mathematically expressed in the following way:

$$Zq = \sqrt{\frac{1}{N} \sum_{i=1}^N Z_i^2}$$

If a comparison is made between the arithmetic average and the root mean square parameters, the latter has the effect of providing additional weighting to the numerically greater values of surface height. One of the reasons for the harmony between both the Ra and Rq parameters is mainly from a historical perspective. The Ra parameter is much easier to determine graphically from a profile recording and for this reason, it was primarily adopted prior to automatic surface measurement instruments being developed. Once the roughness parameters utilising automatic surface texture instrumentation were made available, the parameter Rq had the advantage of being able to neglect the *phase effect* from the electrical filters, while the Ra parameter using arithmetic average will be affected by these phase effects and cannot be disregarded. To compound the problem, Ra has been virtually universally adopted for machining specifications to the detriment of Rq . The Rq parameter is still employed in many optical applications for the assessment of lenses, mirrors and the like for surface optical quality.

Skewness of the assessed profile Rsk , Wsk and Psk

This parameter represents the quotient of the mean cube value of the ordinate values $Z(x)$ and the cube of Pq , Rq or Wq , respectively, within the sampling length (Figure 25). The skewness is derived from the amplitude distribution curve, representing the symmetry about the mean line. This parameter cannot distinguish if profile spikes are evenly distributed above or below the mean line, being considerably influenced by any isolated peaks or valleys present within the sampling length. Skewness can be expressed in mathematical form with Rsk , equating to:

$$Rsk = \frac{1}{Rq^3} \frac{1}{N} \sum_{i=1}^N Z_i^3$$

The skewness parameter of an amplitude distribution curve as depicted in Figure 25(a) indicates a certain amount of bias that might be either in an upward or downward direction. The amplitude distribution curve shape is very informative as to the overall construction of the surface topography. If this curve is symmetrical in nature, then it indicates symmetry of the surface profile; conversely, an unsymmetrical surface profile will be indicative of a skewed amplitude distribution curve (Figure 25b and c). As can be seen in Figure 25(b) and (c), the direction of skew is dependent upon whether the bulk material appears above the mean line – termed *negative skewness* (Figure 25c) – or below the mean line – *positive skewness* (Figure 25b). Utilising this skewness parameter distinguishes between profiles having similar if not identical Ra values.

Further, one can deal with this amplitude or height data statistically, in the same manner as one might physically measure anthropometric data such as a person's average stature, within a specific population range. As with the statistical dispersion of population stature, engineering surfaces can exhibit a broad range of profile heights.

For example, a boring operation with a relatively long length-to-diameter ratio may cause deflection (elastic deformation of the boring bar) and occasion the cutting insert to deflect, producing large peak-to-valley undulations along the bore (waviness). Superimposed onto these longer wavelengths are small-amplitude cyclical peaks (periodic oscillations) indicating vibrations resulting from the cutting process. The resultant surface profile for the bored hole would depict the interactions from the boring bar deformations and any harmonic oscillations. This boring bar motion reflected in the profile trace would exhibit low average profile height, but with a large range of height values.

Surface texture data can be statistically manipulated, beginning with the profile trace's height, or amplitude distribution curve – this being a graphical representation of the distribution of height ordinates over the total depth of the profile. The characteristics of amplitude distribution curves can be defined mathematically by several terms called “moments”, but more will be mentioned concerning this aspect in the Appendix.

A numerical value can be given to Rsk and in the case of Figure 25(b) the “positive skewness” may eventually obtain an adequate bearing surface, although it is unlikely to have oil-retaining abilities. This type of surface can typically be exploited for adhesive-bonding applications. The surface charac-

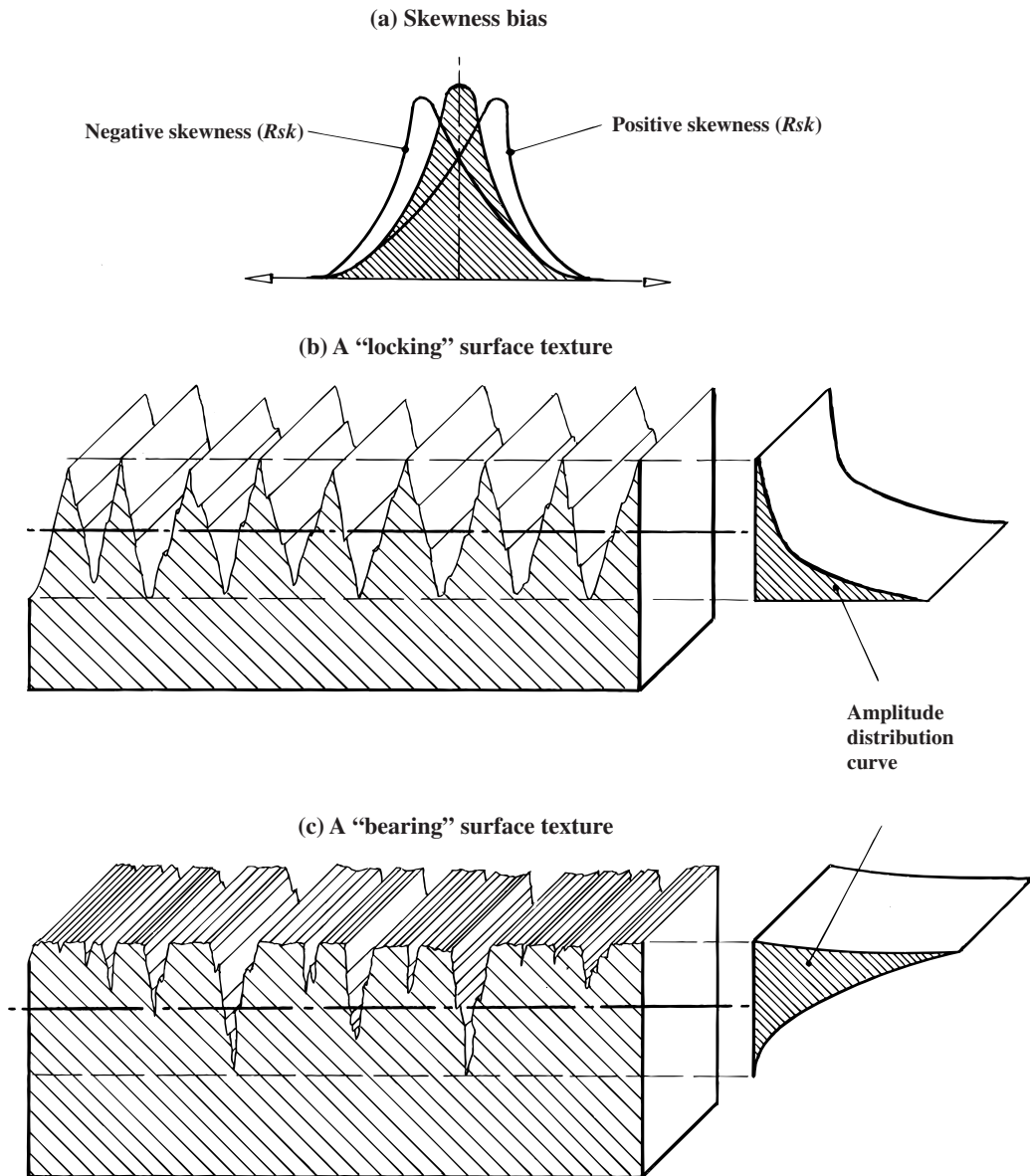


Figure 25. How surface texture topography influences the amplitude distribution curve.

terised by Figure 25(c) might occur in the cases of porous, sintered or cast-iron surface topography – having comparatively large numerical values of negative skewness. The surface is sensitive to extreme ordinate values within the profile under test; this is due to Rsk being a function of the cube of the ordinate height. As a result of this peak sensitivity, it is a hindrance when attempting to inspect plateau-type surfaces, although the Rsk parameter indicates a reasonable correlation with a component's potential load carrying ability, or its porosity.

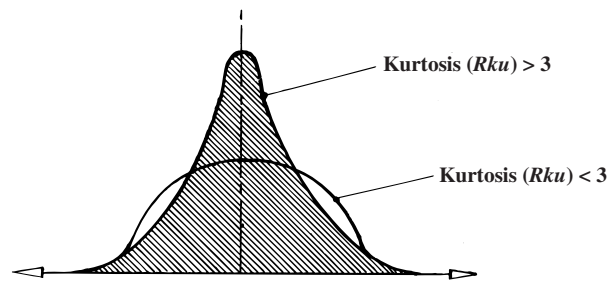
The shape, or “spikeness” of the amplitude distri-

bution curve can also relay useful information about the dispersion, or “randomness” of the surface profile, which can be quantified by means of a parameter known as kurtosis (Rku).

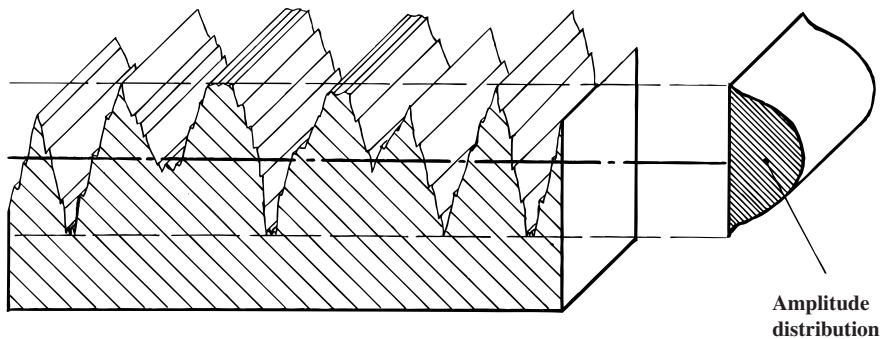
Kurtosis of the assessed profile Rku , Wku and Pku

The kurtosis parameter, typified by Rku , is the mean quadratic value of the ordinate values $Z(x)$ and the

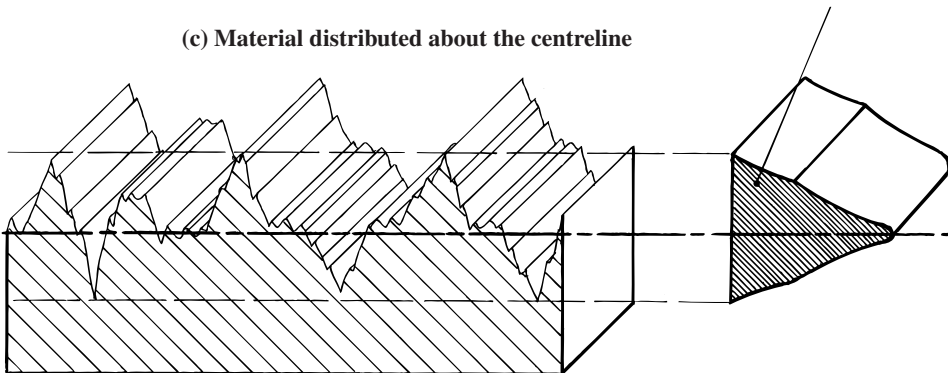
(a) Kurtosis is influenced by the distribution shape



(b) Material distributed evenly across the whole of the surface topography



(c) Material distributed about the centreline

**Figure 26.** Variation in surface topography influences the shape and height of the amplitude distribution curve.

fourth power of Pq , Rq or Wq , respectively, within the sampling length. However, unlike the skewness parameters previously described, namely Rsk and its derivatives, Rku can detect if the profile peaks are distributed in an even manner across the sampling length trace, as well as providing information on the profile's sharpness (see Figure 26).

Kurtosis provides a means of measuring the *sharpness* of the profile, with a “spiky” surface exhibiting a high numerical value of Rku (Figure 26c); alternatively, a “bumpy” surface topography will have a low Rku value (Figure 26b). As a consequence of this ability to distinguish variations in the actual

surface topography, Rku is a useful parameter in the prediction of in-service component performance with respect to lubricant retention and subsequent wear behaviour.

It should be mentioned that kurtosis cannot convey differences between either peaks or valleys in the assessed profile.

The kurtosis parameter indicative of Rku , can be mathematically expressed in the following manner:

$$Rku = \frac{1}{Rq^4} \frac{1}{N} \sum_{i=1}^N Z_i^4$$

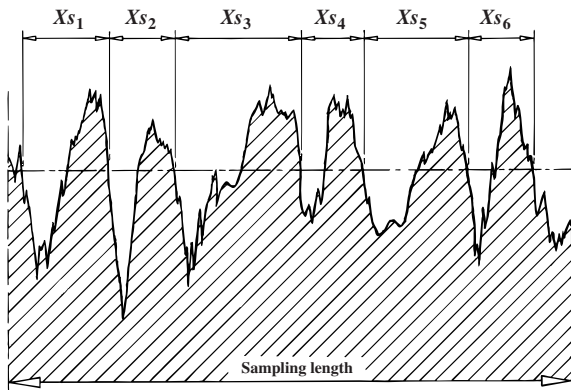


Figure 27. Width of profile elements. [Source: ISO 4287: 1997]

Later, the parameters Rsk and Rku will be applied to specific manufacturing processes to produce *manufacturing process envelopes* describing the functional aspects of a particular generated surface condition.

1.9.3 Spacing parameters

Mean width of profile elements Rsm , Wsm and Psm

The mean width parameter of the profile element widths Xs within the sampling length relates to the average value of the length of the mean line section that contains adjacent profile peaks and valleys (Figure 27). The parameter needs both height and spacing discrimination, and if not specified the default height bias utilised shall be 10% of Pz , Rz or Wz , respectively. Moreover, the default spacing discrimination shall be 1% of the sampling length, with both of these conditions requiring to be met.

The spacing parameters are particularly useful in determining the feedmarks from a specific machining operation, as they relate very closely to that of the actual feed per revolution of either the cutter or workpiece, depending on which production process was selected.

NB More is mentioned on spacing parameters and their affects on the surface topography in the Appendix.

1.9.4 Hybrid parameters

Root mean square slope of the assessed profile $R\Delta q$, $W\Delta q$ and $P\Delta q$

The root mean square slope parameter refers to the value of the ordinate slope dZ/dX within the sampling length, depending upon both amplitude and spacing information, hence the term *hybrid parameter*. The slope of the profile is the angle it makes with a line that is parallel to the mean line, with the mean of the slopes at all points in the profile being termed the average slope within the sampling length. By way of illustration of its use, it might be necessary to determine the developed or actual profile length, namely the length occupied if all the peaks and valleys were laid out along a single straight line. Hence, the steeper the slope the longer will be the actual surface length. In practical industrial situations, the parameter might be utilised in either a plating or painting operation, where the surface length for *keying* of the coating is a critical feature. Moreover, the average slope can be related to certain mechanical properties such as hardness, elasticity or more generally to the *crushability* of a surface. Further, if the root mean square value is small, this is an indication that the surface would have a good optical reflection property.

NB In the Appendix are listed some of the previous and current techniques for the assessment for these *hybrid parameters*.

1.9.5 Curves and related parameters

Curves and their related parameters are defined over the evaluation length rather than the sampling length.

Material ratio of the profile $Rmr(c)$, $Wmr(c)$ and $Pmr(c)$

The material ratio of a profile refers to the ratio of the bearing length to the evaluation length and is represented as a percentage. This bearing length can be found by the sum of the section lengths obtained by cutting the profile line – termed *slice level* – that is drawn parallel to the mean line at a specified level. The ratio is assumed to be 0% if the slice level is at the highest peak; conversely, at the deepest valley this would represent 100%. Parameters $Rmr(c)$, $Wmr(c)$ and $Pmr(c)$ will determine the percentage of each bearing length ratio of a single slice level, or

say, 19 slice levels that are drawn at equal intervals within the R_t , W_t or P_t , respectively.

NB More is mentioned on this topic in the Appendix.

Material ratio of the curve of the profile (Abbot–Firestone or bearing ratio curve)

The material ratio curve represents the profile as a function of level. Specifically, by plotting the bearing ratio at a range of depths in the profile, the manner by which the bearing ratio changes with depth provides a method of characterising differing shapes present on the profile (Figure 28a). The bearing area fraction can be defined as *the sum of the lengths of individual plateaux at a particular height, normalised by the total assessment length* – this parameter is designated by Rmr (Figure 28a). The values of Rmr are occasionally specified on engineering drawings, although this may lead to large uncertainties (see Chapter 6, Section 6.4) if the bearing area curve is

referred to by the highest and lowest profile points.

In the majority of circumstances mating surfaces require specific tribological functions; these are the direct result of particular machining operational sequences. Normally, the initial production operation will establish the general shape of the surface – by roughing-out – providing a somewhat coarse finish, with subsequent operations necessary to improve the finish, resulting in the desired design properties. This machining strategy provides the operational sequence that will invariably remove surface peaks from the original process, but often leaves any deep valleys intact. This roughing and finishing machining technique leads to a surface texture type that is often termed a *stratified surface*. In such cases the height distributions are negatively skewed; therefore this would make it otherwise difficult for an average parameter such as R_a to represent the surface effectively for either its specification or in the case of quality control matters.

NB This topic is discussed in further detail in the Appendix.

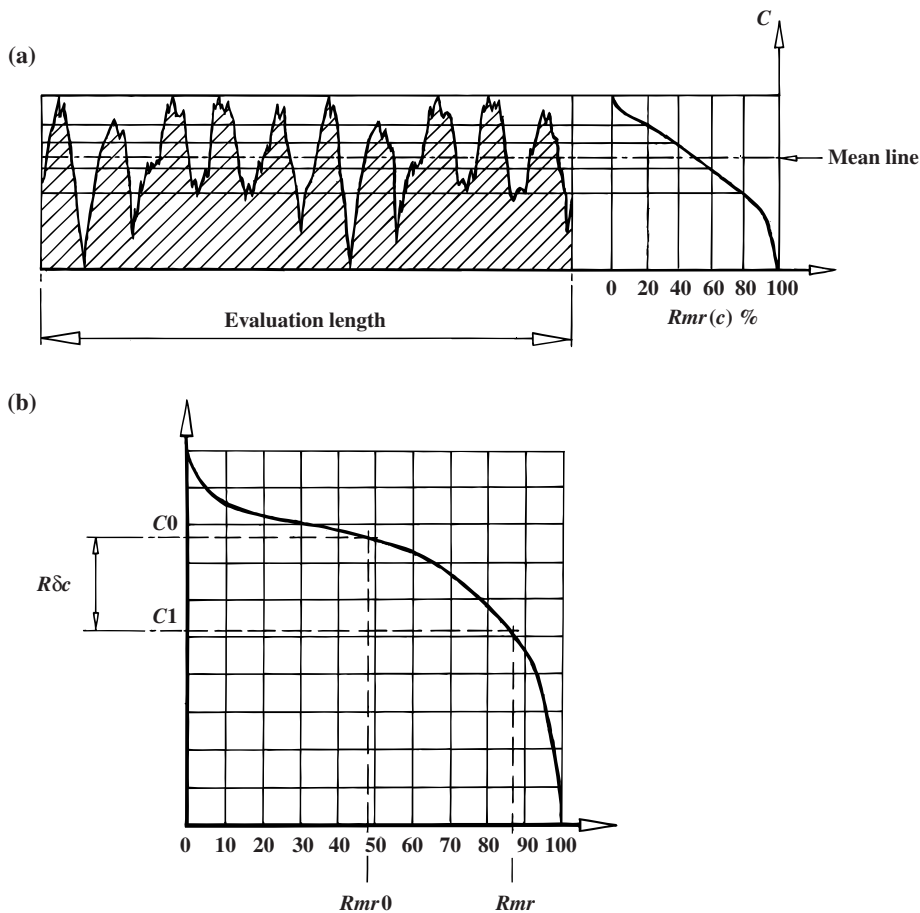


Figure 28. (a) Material ratio curve. (b) Profile section level separation. [Source: ISO 4287: 1997]

Profile section height difference $R\delta c$, $W\delta c$ and $P\delta c$

This parameter can be represented as the vertical distance between two section levels of a given material ratio curve (see Figure 28b).

Relative material ratio Rmr , Wmr and Pmr

The relative material ratio can be established at a profile section level $R\delta c$, being related to a reference $C0$ (see Figure 28b), where:

$$C1 = C0 - R\delta c \text{ or } (W\delta c \text{ or } P\delta c)$$

$$C0 = C(Rmr0, Wmr0, Pmr0)$$

Rmr refers to the bearing ratio at a predetermined height (see Figure 28b). One method of specifying the height is to move over to a certain percentage – reference percentage – on the bearing ratio curve, then move down to a certain depth (slice depth),

with the bearing ratio at the corresponding position here being the Rmr (Figure 28b). The reason for the reference percentage is to eliminate any spurious peaks from the assessment, as they would tend to be worn away during the initial *burn-in/running-in* period. The slice depth will then correspond to a satisfactory surface roughness, or to an acceptable level of surface wear.

Profile height amplitude curve

This represents the sample probability density function of the ordinate $Z(x)$ within the evaluation length. The amplitude distribution curve, as it is usually known, is a probability function giving the probability that a profile of the surface has a particular height at a certain position. Like many probability distributions, the curve normally follows the contours of a Gaussian – *bell-shaped* – distribution (Figure 29a). The amplitude distribution curve informs the user of how much of the profile is situated at a particular height, from the aspect of a histogram-like sense.

This curve illustrates the relative total lengths

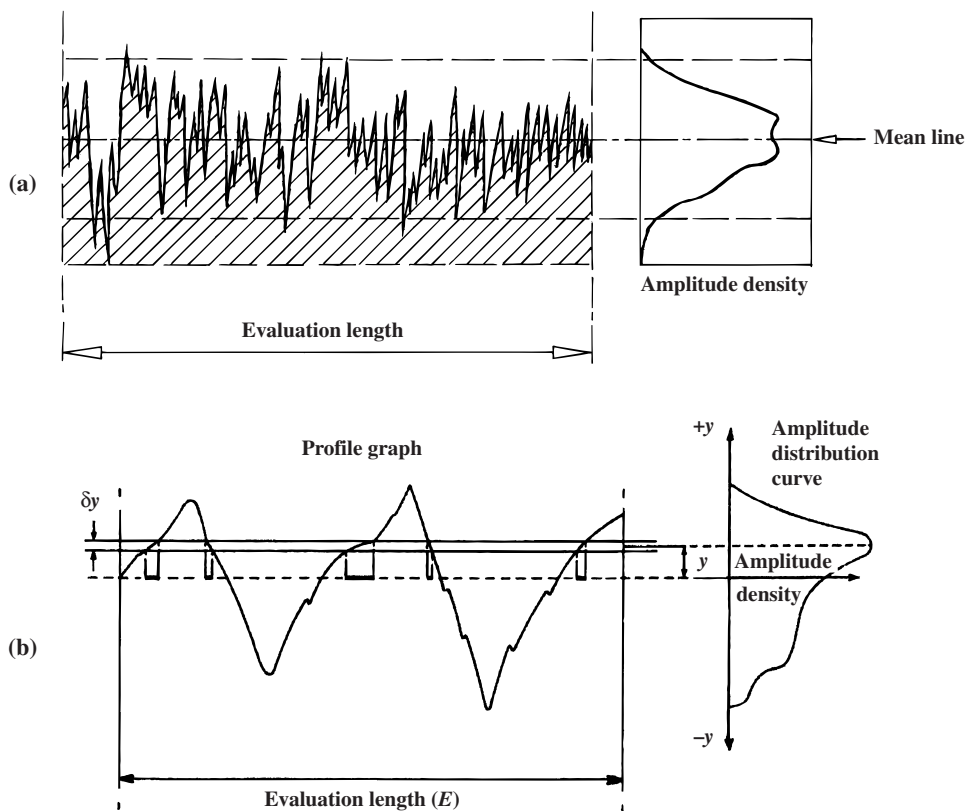


Figure 29. Profile height and amplitude distribution curve. [Source: ISO 4287: 1997]

over which the profile graph will attain any selected range of heights that are above or below the mean line (see Figure 29b). The profile's horizontal lengths are included within a narrow zone width δy at a height z and are represented in Figure 29(b) by "a, b, c, d and e". Therefore, by expressing the sum of these lengths as a percentage of the evaluation length, this will give a measure of the relative amount of the profile at height z .

This graph is known as the amplitude distribution at height z , so by plotting density against height the amplitude over the whole profile can be determined – producing the amplitude density distribution curve.

1.9.6 Overview of parameters

In Table 2 are shown the more recent and previously utilised parameters, together with information concerning whether the parameter has to be calculated over the sampling or evaluation length.

1.10 Auto-correlation function

Every day of one's life, the technique of correlation is applied to compare sights, sounds, tastes, objects – the list is almost limitless – this being a process of comparison. The auto-correlation function (ACF) is essentially a process of determining the relationship of any point on the profile to all other points. Some

manufactured surfaces clearly indicate visible repetitive marks on either the material itself or indirectly via the profile graph, while on other surfaces it becomes difficult to distinguish any repetitive irregularities from random occurrences. This visual difficulty is particularly true when the amplitudes of the repetitive irregularities are less than the random occurrences. Moreover, the presence of repetitive surface features – however small – can indicate factors such as tool wear, machine vibration or deficiencies in the machine, with identification of them being an important factor. Any random pattern occurring on the surface can reveal, for example, whether a built-up edge condition has transpired on the cutting tool resulting in a degree of tearing of the machined surface – this tends to be of a random nature and is not predictable.

The extent of this randomness of the surface can be monitored and assessed by isolating random from repetitive textural patterns, this being achieved by auto-correlation. When a profile is perfectly periodic in nature – typified by a sine wave – the relationship of a particular group of points repeats itself at a distance equal to the wavelength. Conversely, if the profile under inspection is comprised entirely from random irregularities, the precise relationship between any specific points will not occur at any position along the trace length, hence any repetitive feature or group of features can be identified. Computers equipped with fast digital processors have significantly reduced the tedious task of determining a surface profile's auto-correlation.

The technique exploited by auto-correlation is to compare different parts of the surface profile; in this

Table 2. Past and present parameters for surface texture

Parameters 1997 edition	1984 edition	1997 edition	Determined within:	
			Evaluation length	Sampling length
Maximum profile height	<i>R_p</i>	<i>R_p</i>		X
Maximum profile valley depth	<i>R_m</i>	<i>R_v</i>		X
Maximum height of profile	<i>R_y</i>	<i>R_z</i>		X
Mean height of profile	<i>R_c</i>	<i>R_c</i>		X
Total height of profile	–	<i>R_t</i>	X	
Arithmetical mean deviation of the assessed profile	<i>R_a</i>	<i>R_a</i>		X
Root mean square deviation of the assessed profile	<i>R_q</i>	<i>R_q</i>		X
Skewness of the assessed profile	<i>S_k</i>	<i>R_{sk}</i>		X
Kurtosis of the assessed profile	–	<i>R_{ku}</i>		X
Mean width of profile elements	<i>S_m</i>	<i>R_{Sm}</i>		X
Root mean square slope of the assessed profile	Δq	<i>RΔq</i>		X
Material ratio of the profile		<i>R_{Mr(c)}</i>	X	
Profile section height difference	–	<i>Rδc</i>	X	
Relative material ratio	<i>t_p</i>	<i>R_{mr}</i>	X	
Ten-point height (deleted as an ISO parameter)	<i>R_z</i>	–		

Source: ISO 4287: 1997 (E/F).

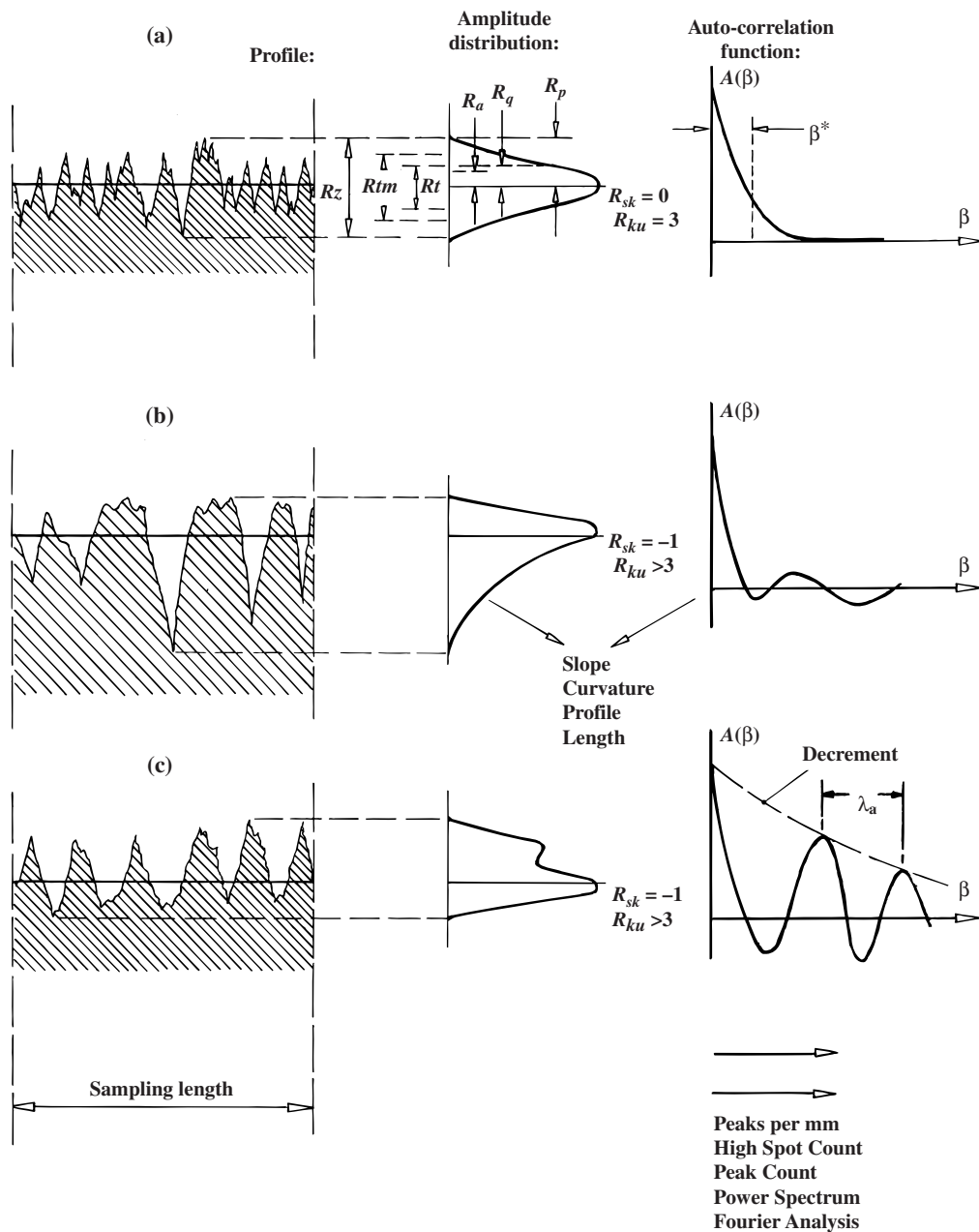


Figure 30. Comparison of surface parameters for typical surfaces showing those currently in use. [After Whitehouse, 1978.]

manner profile repetitions or similarities can be discovered. When the profile of the surface displays either an isotropic or random shape then no portions will be similar, exhibiting low auto-correlation function. On the other hand, if an anisotropic or periodic surface topography occurs – for example as in single-point turning operations – then repetition of the surface profile occurs and auto-correlation will be high.

Figure 30 shows an example of the ACF relating to specific surface profile conditions. As mentioned above, the ACF contains information relating to the spacings of peaks and valleys in the profile. An alternative to the ACF is the equivalent term known as the power spectrum, or power spectral density, in which spacings are replaced by frequencies – one being a Fourier transform of the other. For the ACF examples shown in Figure 30, the scale of the hori-

zontal axis can be the mean distance between peaks or the average distance between mean line crossings.

The ACF is derived from the normalised auto-covariance function (ACVF) – this being as follows:

$$ACF(\Delta x) = \frac{ACVF(\Delta x)}{(Rq)^2}$$

where

$$ACVF(\Delta x) = \lim_{lm \rightarrow \infty} \int_0^{lm} y(x) y(x + \Delta x) dx$$

where lm = assessment length, $y(x)$ = profile height at position x and Rq = rms parameter.

The ACF does not contain information about the amplitude of the profile, with its values ranging from plus one – giving perfect correlation – to negative one – this having a correlation of the inverted but shifted profile. Due to the fact that the ACF is independent of the amplitude of the profile it is more popular than the ACVF. In the case of an ACF having an isotropic (random) profile, it quickly decays to zero (Figure 30a), with the profile trace being of an unreplicative nature. Conversely, the tendency for a truly anisotropic (periodic) generated surface is that it does not decay to zero, with Figure 30(c) illustrating some periodicity in its profile trace – indicating that decay here is significantly less than that for Figure 30(a). Despite the fact that both Figure 30(a) and (c) have similar Rz values, their respective ACF are distinctly different and can be used as a means to discriminate between surfaces and indicate whether they will fulfil the desired in-service applications, which might not otherwise be apparent.

Fourier analysis (FFT)

Any surface profile will normally be of some complexity, this being comprised of an array of differing waveforms that are superimposed onto one another. An actual profile's specific composition will be dependent on the shape and size of the waveforms existing within the profile; both their amplitude and frequency, termed "harmonics", can be established by the application of Fourier analysis.

The technique known as a "fast Fourier transform" – FFT for short – enables one to establish a series of sine waves being generated from the surface profile, and when combined together contribute to the original profile. Individual harmonics represent a specific frequency-wavelength, in combination with their associated amplitude.

The surface profile depicted in Figure 31(a) was selected to illustrate a periodic form, which in isola-

tion may mask significant hidden harmonic detail which could play an important role in its later in-service application. Results from the FFT analysis of the profile (Figure 31b) reveal the differing sine wave amplitudes necessary to generate the original profile. In this case, the largest sine wave amplitude equals a value of 0.7 mm, this being termed the "dominant wave".

It should be emphasised that the information obtained from the FFT analysis will be exactly the same as that acquired by the auto-correlation method, but obviously it is displayed in an altered form. Due to the isolation of the harmonics into specific bands for a particular surface profile, FFT can impart valuable understanding as to how the production process might have been influenced by variables in processing for that product. For example, the dominant harmonic in the FFT display for a profile might give a valuable insight into the vibrational frequency for a specific machine tool. Knowledge of this vibrational effect on the resulting surface profile may enable the attenuation of this vibrational signature by judicious adjustment of either cutting data, or specific maintenance to the machine tool to improve the subsequent surface profile's condition.

1.11 Appearance of peaks and valleys

For many engineering and statistical graphs no *intrinsic* relationship exists between the plotted measurement units; for example, when plotting the rise in temperature of a heat treatment oven against elapsed time the ordinates would be represented by both degrees Celsius and hours – two quite dissimilar units. As a consequence of these disparate units, acceptance of the shape of the graph is acknowledged, whatever scales are applied for the ordinates. It is quite different when dealing with a surface profile graph, where both the horizontal and vertical ordinates are represented by the same units (i.e., length). At first introduction, it may be difficult to appreciate that due to the difference of scale the graph will not instantaneously give an indication as to the *shape* of the irregularities. Figure 32 illustrates the effect, with this disparity being one of the most difficult factors in surface texture analysis to comprehend.

It is very important to note that although the *visual* profile's shape is distorted due to the application of different horizontal and vertical magnifications, *measurements scaled from the graph are*

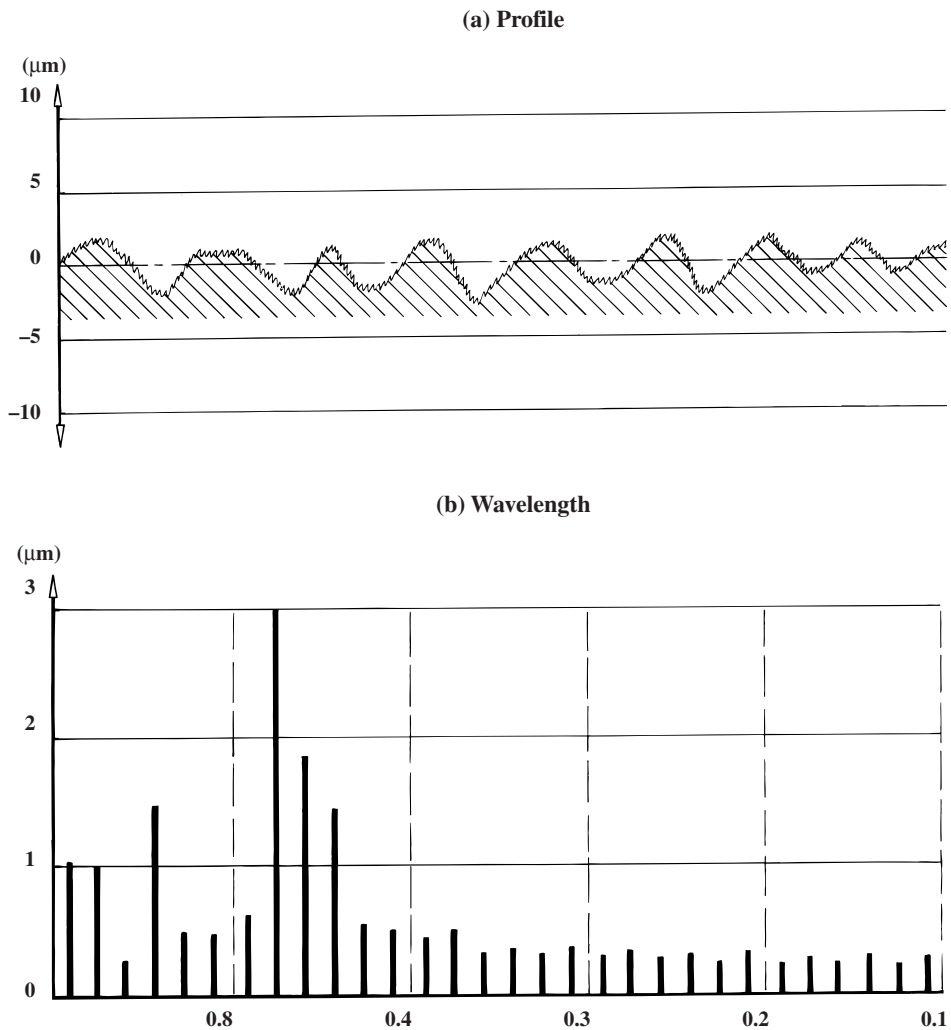


Figure 31. Fourier analysis of an idealised (i.e. periodic) profile.

correct. Such effects can be visually demonstrated with reference to the peak angles of the profile illustrated in Figure 32, indicating how the profile shape varies as horizontal/vertical magnification is altered. Figure 32(III) verifies this fact, as the horizontal magnification of the surface profile is increased (Figure 32II), the length $X-X$ expands to X^1-X^1 and the peaks; A, B, C and D appear flatter. Increasing the horizontal magnification still further – until it equals the vertical magnification (Figure 32I) – expands the sampling length $Y-Y$ to Y^1-Y^1 . Once again, with this still higher magnification, the peaks E and G and valleys F and H now have a much flatter appearance. A point worth making is that the actual difference in respective heights of corresponding peaks and valleys in I, II and III are identical.

As a result of the spiky appearance of both the

peaks and valleys on the profile graph, this can lead to considerable confusion when visually assessing the actual graph. However, the geometric relationship between angles on the graph and those that occur on the physical profile of the component under test are more easily determined. As an example of this interpretation, Figure 33(a), on the profile graph, depicts a symmetrical peak with an included angle 2α . Figure 33(b) exhibits the corresponding peak, on the actual surface, having an angle 2β . The ratio of the angles is determined by

$$\frac{\tan \beta}{\tan \alpha} = \frac{V_v}{V_h}$$

This ratio is strictly true for symmetrical peaks, although it is only slightly different for asperities of

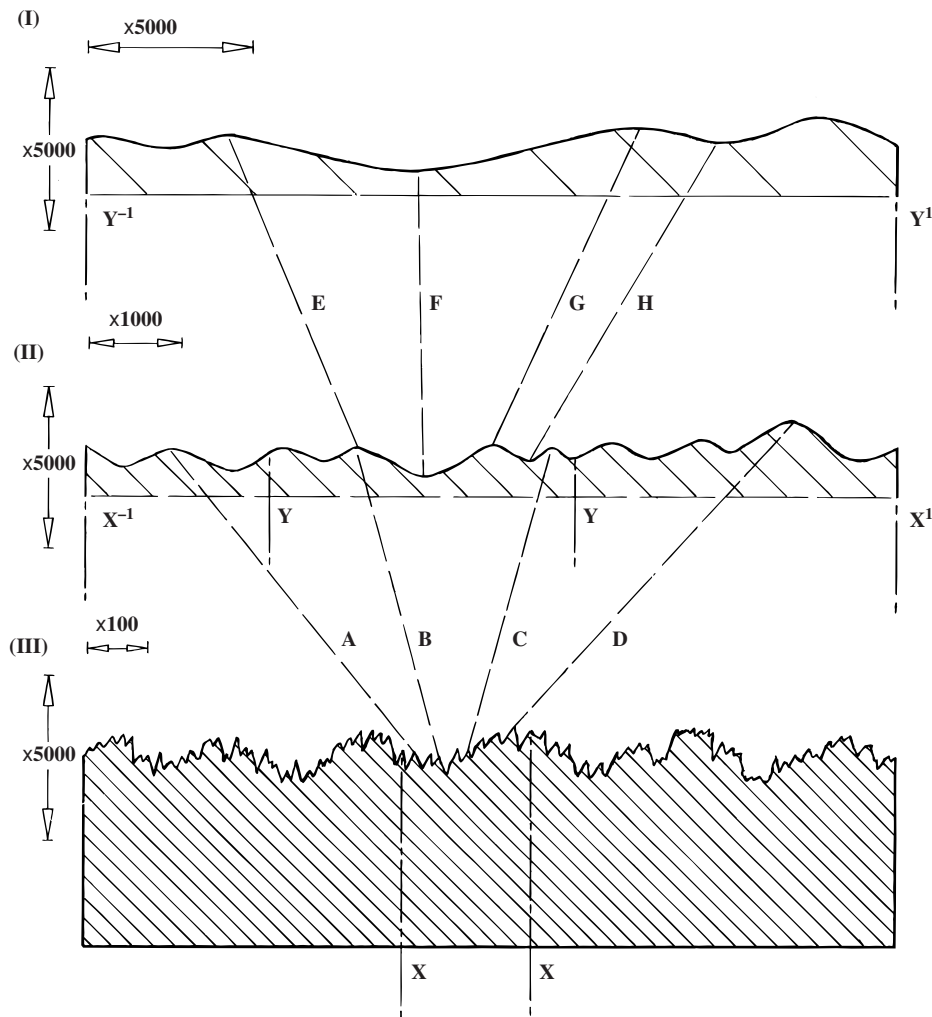


Figure 32. Diagram illustrating how profile shape varies as the horizontal magnification is reduced: (I) surface profile magnified $\times 5000$ equally in all directions; (II) profile with $V_v:V_h$ ratio of 5:1; (III) profile graph recorded with a $V_v:V_h$ ratio of 50:1. (Courtesy of Taylor Hobson.)

an asymmetrical nature. For example, the data obtained from a profile graph in which the vertical and horizontal magnifications and peak geometry might be $V_v = \times 5000$, $V_h = \times 50$ and $2\alpha = 10^\circ$. This will therefore represent

$$\alpha = 5^\circ$$

$$\tan \alpha = 0.0875$$

hence

$$\begin{aligned} \tan \beta &= 0.0875 \times 5000/50 \\ &= 8.75 \end{aligned}$$

giving

$$\beta = 83.5^\circ$$

Thus

$$2\beta = 167^\circ$$

This example of peak geometry (Figure 33a) represents a fairly sharp peak on the profile graph, but only a gradual slope – rising and falling at 6.5° from the horizontal – on the actual component's surface (Figure 33b). Due to the fact that it becomes difficult to accurately measure the peak angle on the graph, and it is rare that an accurate assessment of the surface angles is required, an approximation of $2\alpha : 2\beta$ is normally sufficient.

Table 3 provides approximate conversions for several $V_v:V_h$ ratios.

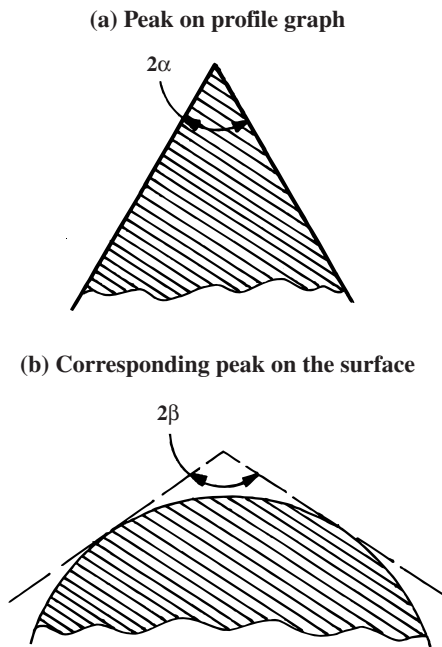


Figure 33. Relationship between peak angles on the graph and the actual surface. (Courtesy of Taylor Hobson.)

It has been previously stated by way of a light-hearted analogy for the representation of two rough contacting surfaces that they resemble “Switzerland on top of Austria”, although this simply shows the influence of the visual aspect of the profile graph. A more preferable analogy might be “The Netherlands on top of Iowa”, as the peaks on the surface are in general of quite blunt and low aspect.

When considering the profile with its associated peaks and valleys, it has been presumed that the peak height on a profile graph is characterised by the summit of individual peaks traced by contact with the stylus for a surface under test. This may be

true for the major ridges for a surface exhibiting marked directional lay (see Figure 5A) being traced at 90° to the direction of the dominant surface pattern lay. Conversely, for a surface displaying random textural influences, such as the “multi-directional” example in Figure 4, it is extremely unlikely that a stylus’s path will trace across the summit of every peak. In these circumstances, some peaks and valleys will be displaced from the path of the stylus by varying distances, with the result that the stylus will also traverse across many flanks. Stylus vertical displacement, although somewhat less than the full peak height or valley encountered, will moreover be correct for the traced feature. Assumptions made by many are that every peak on the graph is the summit of a peak on the actual surface, but where random surfaces occur there will be sufficient actual summits encountered to prevent any significant inaccuracy as a result.

1.12 Stylus-based and non-contact systems

Effect of stylus

It should be appreciated by now that the stylus and skid, when employed, are the only parts which are in active contact with the surface under test. Therefore both the shape and dimensions of this pick-up are critical factors and will strongly influence information retrieved from the surface. The geometry of a typical stylus – its shape and size – was discussed in Section 1.4 and illustrated in Figure 11(a); it influences the accuracy obtainable from the surface. Due to the finite size of the stylus some surface imperfections, such as visible scratches, are

Table 3. Peak angles (2β , Figure 33a) on the surface corresponding to peak angles of 2α on the graph

Ratio: V_v/V_h	Peak angle on graph (2α , Figure 33)							
	1°	5°	10°	15°	20°	30°	40°	50°
10	10°	47°	82°	105°	121°	139°	149°	155°
25	25°	95°	131°	146°	154°	162°	167°	170°
100	82°	154°	167°	171°	173°	176°	177°	177°
250	131°	169°	175°	176°	177°	178°	179°	179°
500	154°	175°	177°	178°	178°	179°	–	–
1000	167°	177°	179°	179°	179°	–	–	–
2500	174°	179°	–	–	–	–	–	–
5000	177°	–	–	–	–	–	–	–

Courtesy of Taylor Hobson.

too fine to be completely penetrated, even with the smallest tip available. This is not too much of a hindrance to assessment as the Ra is an “averaging parameter”, so the omission of the finest textural features will not unduly affect the resulting value. To illustrate this point, many years ago a series of tests were conducted on ground surfaces by Reason et al. (1944), these samples had four increasing levels of roughness. The reduction in measured Ra value when utilising either a $1.25\ \mu\text{m}$ or $10\ \mu\text{m}$ stylus was 4.5 to 3.25, 8 to 7.5, 11 to 10 and 26 to 25 μm , respectively. This small variation occurred despite the fact that the larger stylus was eight times greater than the smaller.

For undistorted surfaces (periodic), there is a predictable relationship between the stylus radius and the resulting error introduced into the Ra reading. Figure 34 charts the approximate values of this error, being applicable to regular profiles with 150° included angle triangular peaks and valleys – typical of surfaces utilised for calibration standards.

For surfaces other than of this regular triangulated geometry, the values tend to be only approximate, although they do indicate the order of any error likely to be occasioned. The horizontal ordinates in Figure 34 are the ratio of the stylus tip radius to the actual Ra value of the surface, with the vertical ordinates being the amount by which the measured Ra value is too low. The chart demonstrates that only when the tip radius is greater than approximately 20 times the Ra value is this error of any significance. By traversing a stylus across a standard incorporating narrow grooves of known widths – 20, 10, 5 and $2.5\ \mu\text{m}$ – this is a technique for estimating stylus wear or damage. Obviously, the blunter the tip the less distance it can descend into the groove, this penetration depth being measured from the profile graph. The effective stylus radius can be established from the calibration curve supplied with the artefact. Stylus cone angle can also be measured by this method; because the stylus wears with use, the change in effective tip size can be monitored.

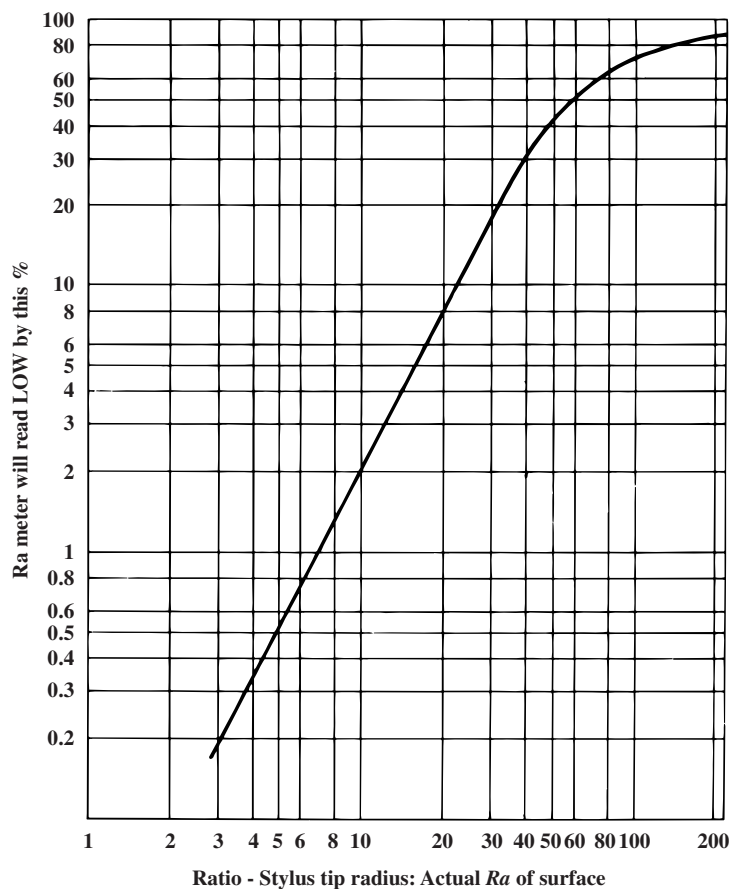


Figure 34. Error due to finite stylus tip radius, for a specific surface. (Courtesy of Taylor Hobson.)

Illustrative example: When measuring a surface of $0.5\ \mu\text{m}$ Ra with a stylus tip of $5\ \mu\text{m}$ (10:1 ratio) the display will be approximately 2% low. If the stylus has double this radius (i.e., $10\ \mu\text{m}$) this error increases to 8%.

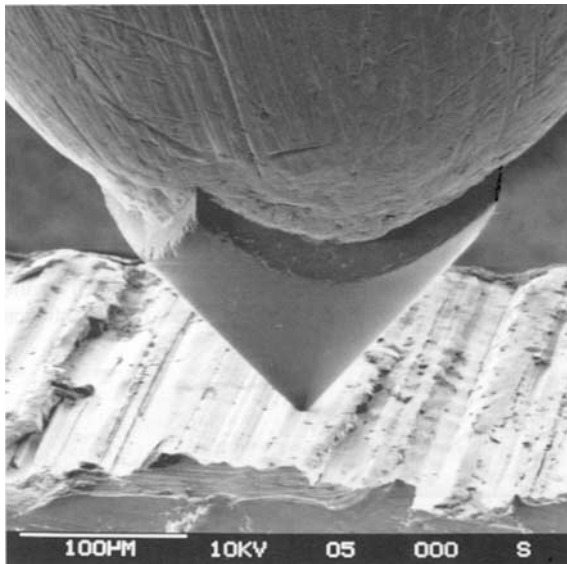


Figure 35. 90° conical stylus tip mechanically tracing across a portion of a honed surface. SEM photomicrograph, X200 magnification. [Courtesy of PTB Braunschweig/Feinpruf Perthen GmbH.]

Stylus force on the surface must not be great enough to deform or scratch the surface as this may promote erroneous results. Conversely, the force imparted by the stylus must be of a sufficient level to ensure that it maintains continuous contact with the surface at the desired traverse speed. This factor is important in the design of the pick-up, with most surface texture instruments having a stylus tip radius of between 1 and 10 μm . However, the greater the radius of the tip, the larger the allowable force that can bear on the surface. If a tip radius is small, say 0.1 μm (Figure 11a), then a very low force must be used, typically around 0.75 mN.

The conical stylus geometry was previously mentioned in Section 1.4 and illustrated in Figure 10, invariably having a 90° tip angle. In action (Figure 35), the conical tip angle of 90° can easily cope with surface features such as that shown, having the anisotropic profile of a honed surface. When surface features become greater than 45° to the horizontal, or the valley is smaller than, say, 5 μm , then the mechanical motion of the tip becomes to some extent distorted, as illustrated in Figure 11(c). Specifically, under these conditions, the tip acts in a similar fashion to a low-pass filter with high-frequency features not being recorded. With a tip angle of 90°, this means that the stylus will foul any peak or valley having an included angle of less than 45° – which may at first seem a considerable limitation, but in reality does not cause too much of a problem.

Chisel-edge styli (see Figure 11a) are a practical variation on the cone-shaped stylus (Figure 35), being particularly successful at entering porous

Table 4(a). Typical (traceable) markings on pick-up

Identification marks:	± 100	5	90	0.8	30/6	1/0
Stylus data:						
Measuring range (μm)	X					
Tip Radius (μm)		X				
Tip angle (°)			X			
Measuring force (mN)				X		
Skid data:						
Skid radius, longitudinal/traverse					X	
Distance skid-to-stylus, longitudinal/traverse						X

Table 4(b). Configurations for some standard pick-ups

Stylus						
Tip angle (°)	60° \pm 5° or		90° \pm 5°			
Tip radius (μm)	2 \pm 1	5 \pm 2	10 \pm 3			
Static measuring force (ie at electric zero, mN)	\leq 0.7	\leq 4	\leq 16			
Variation in static measuring force (mN/m)	\leq 0.035	\leq 0.2	\leq 0.8			
Skid						
Distance from stylus dependent upon pick-up design;						
Radius						
(longitudinal, mm)	0.3	1.3	3	10	30	infinity
(lateral, mm)	depends upon design of pick-up					
Surface roughness (μm)	\leq 0.1					
Skid force (N) for:						
Hard surface	\leq 0.5					
Soft surface	\leq 0.25					
Linearity of system (%)	\pm 1					

materials such as sintered parts. One problem associated with the chisel-type stylus is the difficulty in maintaining its attitude at right angles to the lay, being more sensitive to error from peaks rather than valleys. Stylus misalignment can be minimised by the application of “smart parameters” that mathematically ignore certain features. Normally, it is recommended to utilise a conventional stylus geometry, unless particular problems are encountered that may cause the standard stylus to impair the validity of the resulting data.

Manufacturers of stylus/skid geometries normally apply identifying marks on the pick-up, or at the very least on its accompanying transportation case. Table 4(a) illustrates the pick-up markings, whereas Table 4(b) highlights the standard configurations.

1.12.1 Pick-up

The function of the transducer or pick-up is to convert the minute vertical movements of the stylus, as it progresses along the surface, into proportionate variations of an electrical signal. The sensitivity demanded of the pick-up is such that it should be able to respond to stylus movements of approximately 0.1 nm or better. This output is minute and of necessity must undergo significant amplification to enable stylus trace movements of between 50,000 and 100,000 times greater than the stylus movement ($V_v \times 50,000$ or $\times 100,000$).

In general, pick-ups can be classified into two groups, depending on their operating principle:

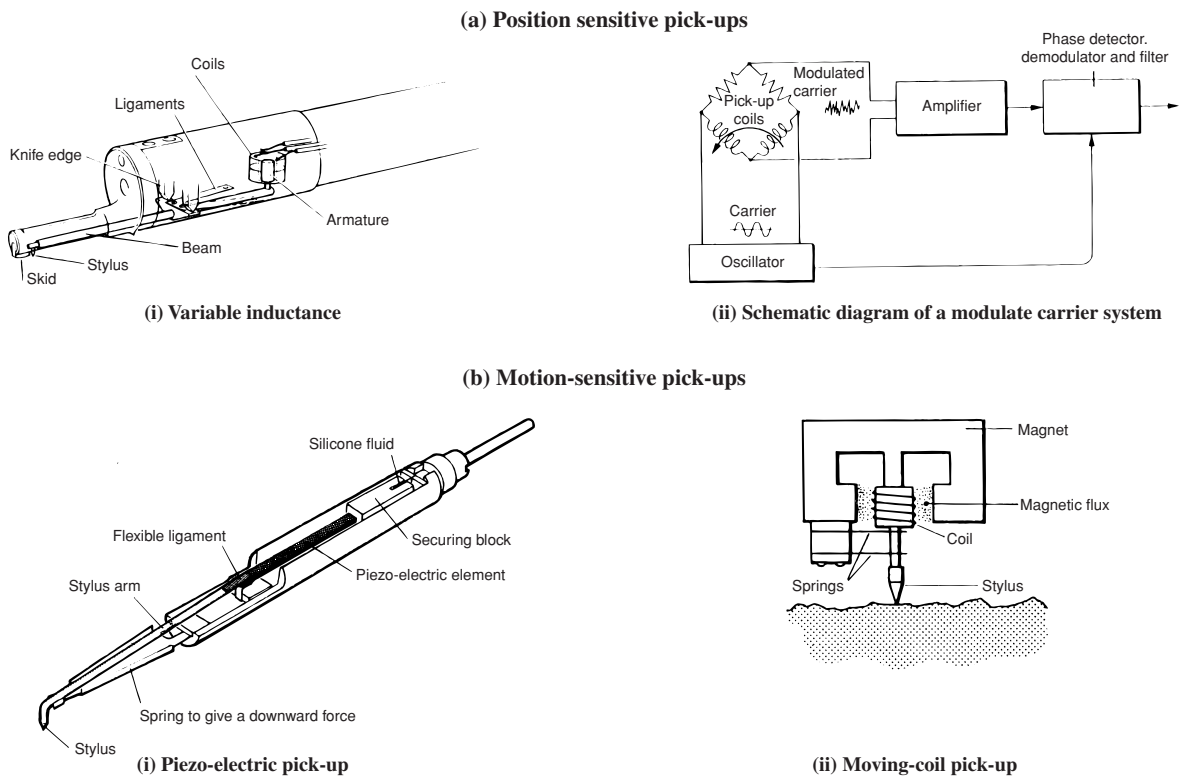
1. Analogue transducers (Figure 36)

- **Position-sensitive pick-ups** (Figure 36a) give a signal proportional to displacement, even when the stylus is stationary. Output is independent of the speed at which the stylus is displaced and is only related to the position of the stylus within a permissible range of vertical movement. A big

advantage of this type of pick-up is that it enables a true recording of waviness and profile to be obtained.

- **Motion-sensitive pick-ups** (Figure 36b) produce an output only when the stylus is in motion, with the output being related to the speed at which the stylus is displaced, dropping to zero when stationary. If displacement is very slow, perhaps due to widely spaced waviness or to change in form, then the output for practical purposes is almost zero. This low output means that variations of waviness and form are excluded from the profile. These pick-ups tend to be utilised if instruments do not have recording facilities, such as some portable equipment.

The *variable-inductance pick-up*, illustrated in Figure 36(ai), has been widely used for many years. The stylus is situated at one end of a beam pivoted at its centre on knife edges, while at the opposite end an armature is carried which moves between two coils, changing the relative inductance. This pick-up's operating principle is as follows: coils are connected in an AC bridge circuit (Figure 36a(ii)) such



NB: The skid is not shown, but is fixed to the housing near the stylus

Figure 36. Basic design elements of pick-ups of the analogue transducer configuration. (Courtesy of Taylor Hobson.)

that when the armature is centrally positioned between the bridge it is balanced, giving no output. Movement of the armature unbalances the bridge, providing an output proportional to its displacement; the relative phase of the signal depends on the direction of movement. This signal is amplified and compared with that of the oscillator, to determine in which direction it has moved from the central (zero) position. It is necessary to utilise an oscillator to produce a constant AC output, because the pick-up – unlike a motion-sensitive pick-up (Figure 36b) – does not generate any output; it merely serves to modify the carrier. Simultaneously, the knife edges exert light pressure from a very weak spring acting on the beam, enabling subsequent stylus contact with the surface under test. The ligaments prevent unwanted motion of the beam in the horizontal plane, with the result that stylus movement is only possible normal to the surface being assessed.

This type of pick-up is also fitted to newer versions of measuring instruments, particularly where the range-to-resolution ratio has increased from around 1000 to 64,000. Improvements in electronics for the latest pick-up designs of this type have obviated the need for reliance on a skid, meaning that precise pivot bearings have replaced the knife-edge pivots.

The *piezoelectric pick-up* (Figure 36b) has been widely used in the past, but now it tends to be used for the less sophisticated hand-held measuring instruments. When the stylus deforms the piezoelectric crystal, it has the property of developing a voltage across electrodes, the advantage being that it is virtually instantaneous. As the stylus follows the contours of the surface, the piezoelectric crystal distorts by bending a flexure that causes crystal compression and it becomes charged. The resulting charge is then amplified and electronically integrated, producing signals proportional to the surface profile. A piezoelectric transducer exerts a proportionally larger stylus force to that of an equivalent inductive transducer, with the former pick-up possibly damaging softer and more delicate surfaces.

The operating mechanism of this piezoelectric transducer can be described in the following manner: the flexible ligament interposed between the stylus arm and the piezoelectric element has enough stiffness to transmit normal vertical motion of the stylus to the crystal, but will flex as the stylus is suddenly subjected to instantaneous shock and in this way protection of the somewhat fragile piezoelectric element is achieved. A light spring force provides downward force on the stylus, keeping it in contact with the surface. A small drop of silicone oil is held by capillary force between the securing block and a thin metal blade fastened to the rear of the housing, giving the pick-up suitable dynamic char-

acteristics. The silicone fluid exhibits low stiffness and at lower frequencies; therefore large but slow stylus displacements allow free movement of the far end of the crystal, hence no bending occurs and zero output is generated. At higher frequencies the fluid stiffness increases, effectively preventing the local end of the piezoelectric crystal from moving; therefore the bridge deforms generating an output.

A piezoelectric device is a position-sensitive device. A voltage on the electrodes persists while the element is deformed (providing no current is drawn from it), due to practical considerations such as finite input impedance of the amplifier and cable losses. These considerations ensure that it can only be utilised in the motion-sensitive mode.

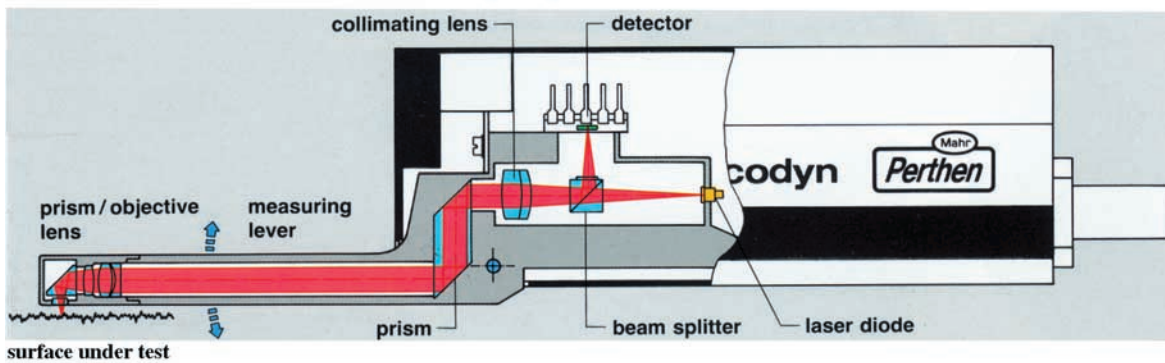
The *moving-coil transducer pick-up* (Figure 36bii) operates on the same principle as a DC generator or an electric dynamo. In operation, as the stylus motion occurs the coil moves inside a permanent magnet, inducing a voltage in the coil. This voltage is proportional to the velocity of the coil. This design allows the transducer element (the coil) to be coupled directly to the stylus without the need for an extended arm or hinged beam. Because of its overall size it is somewhat restrictive in use and is not generally popular today.

In a similar fashion to the piezoelectric transducer, the moving-coil pick-up does not measure displacement, but velocity. Stylus velocity has to be integrated in order that the absolute position of the stylus can be determined. As a result of the coil velocity rather than measurement of stylus displacement, integration errors will increase as the signal frequencies decrease. Hence, this type of pick-up is unsuitable for the measurement of either profile or waviness (low characteristic frequencies). Therefore the piezoelectric transducer is only suitable for the assessment of roughness. Yet another disadvantage of the moving coil transducer is its poor linearity. It should only be used in the form of a comparator by comparing similar component surface roughnesses against a known roughness calibration standard.

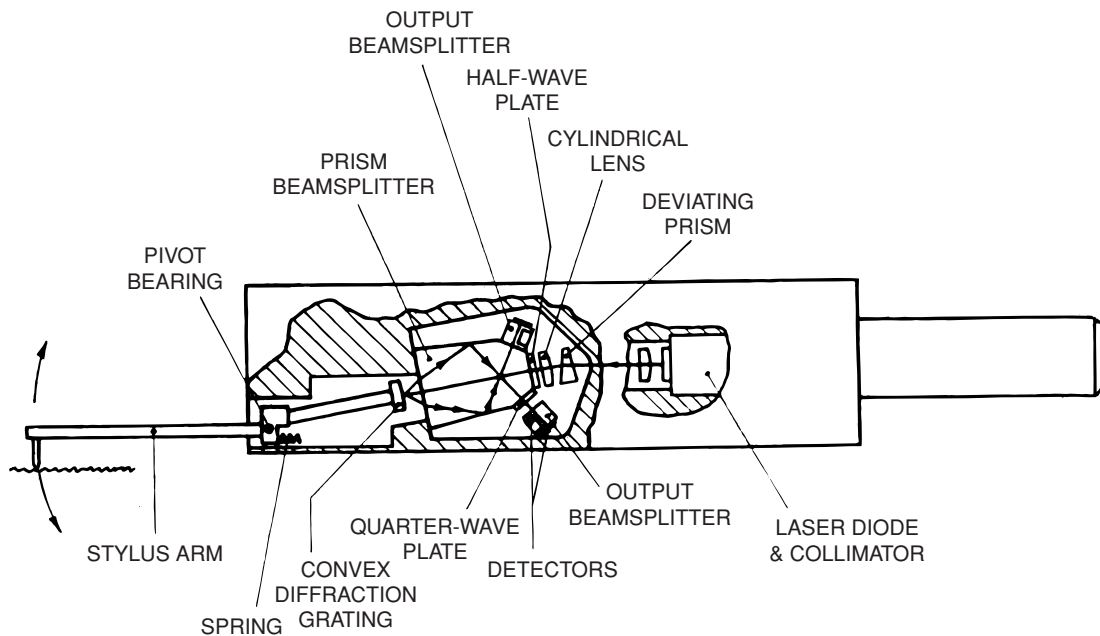
2. Digital transducers (Figure 37)

With stylus motion along the surface under test, pulses occur that correspond to multiples of the transducer resolution which are fed to an up-down electronic counter that displays the gauge displacement, with its range being determined by physical constraints of the gauge. Typical range-to-resolution ratios are of the order of 700,000, hence the advantage is that there is no need for range switching, enabling the maximum resolution to be available over the complete operational pick-up range. The displayed displacement is relative to the stylus position, where the counter is zeroed. This means that every time the electronics are switched off and on

(a) Non-contact laser pick-up. (Courtesy of Mahr Perthen.)



(b) Phase grating interferometric (PGI) pick-up. (Courtesy of Taylor Hobson.)

**Figure 37.** Basic design elements of pick-ups of the digital transducer configuration.

again, the counter must be zeroed at a particular datum position. In a similar manner, if a certain maximum stylus speed has been exceeded the counter may lose count, requiring zeroing again at the datum position.

The following pick-ups are typical of those currently in general use for surface texture instrumentation:

- *Non-contact laser* (Figure 37a): These optical configurations vary depending upon the measuring instrument company, but most systems utilise a miniature Michelson laser interferometer with a laser wavelength of 632.8 nm to provide

the measurement reference. Interference patterns are detected by multiple photodiodes, enabling an output signal to be interpolated, typically giving a basic resolution of 10 nm with an operating range of around 6 mm. A focal point of 2 μm is typical, allowing soft or delicate surfaces to be assessed.

- *Phase grating interferometric* (Figure 37b). This pick-up has been developed to complement the laser interferometric type of pick-up as it offers a greater range, with a corresponding reduction in physical size, by employing a laser diode instead of a gas (HeNe) laser source.

The non-contact laser pick-up (Figure 37a) operates on the following principle: infra-red light is emitted by a laser diode; its path through the optical system to the micro-objective is shown in Figure 37(a), with the beam being focused onto the surface to a focal point of 2 μm diameter. Light is reflected back by the workpiece surface, returning along an identical path, later being deflected onto a detector, whereupon an electrical output is generated that corresponds to the distance between the focal point and the surface feature. A powerful linear motor continuously readjusts the hinged measuring lever during the measuring cycle, enabling the focal point to coincide with the surface feature(s) being measured. The focus follows the surface as its motion translates across that portion of workpiece under test, in a similar fashion to a stylus-based pick-up. The vertical movements of the measuring lever are converted into electrical signals by an inductive transducer. Due to the relatively small size of the point of focus (approximately 2 μm), quite minute profile irregularities can be assessed.

Some caution should be expressed in general about non-contact pick-ups, as the surface features being assessed can have a tendency to act as secondary sources of light. When the radius of curvature of a part feature is smaller than 10 μm , diffraction effects appear which can affect the beam's edge. This is not a serious problem for large surface features but becomes crucial for finer surface finishes, illustrating why optical instruments tend to produce higher values for surface texture than their stylus-based equivalents. Other limitations for non-contact pick-ups are that they cannot easily measure small bores, they traverse widely changing surface features, and they cannot "sweep aside" dirt on the surface, unlike stylus instruments.

The phase grating interferometric pick-up (Figure 37b) has positioned on the end of the arm of the pivoted stylus a curved phase grating, which is the moving element in the interferometer. The pitch (grating wavelength) provides the measurement reference. An interference pattern is detected by four photodiodes, enabling the output signal to be measured, giving a basic resolution of 12.8 nm with a range of 10 mm – ensuring that it is a useful pick-up for form measurement.

1.12.2 Skid or pick-up operation

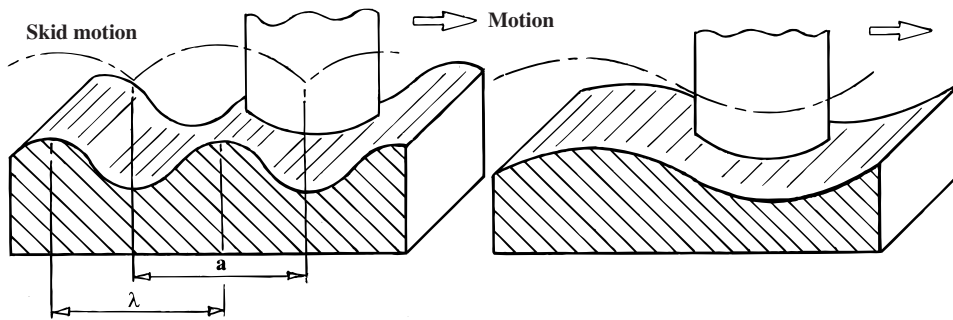
Two types of pick-up operation are normally employed for surface texture assessment:

- with a skid;
- skidless with a reference plane.

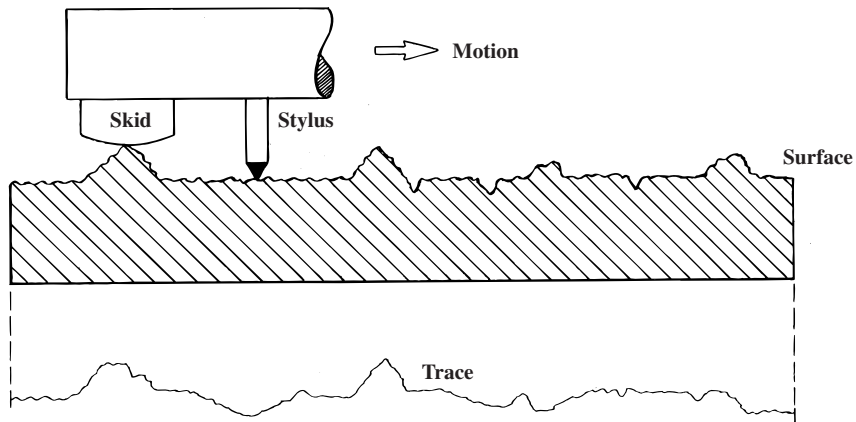
The geometry of a skid can vary, from either a large radius (Figure 38b) to a flat (Figure 38c). The skid supports the pick-up, which in turn rests on the surface. The relative motion of the pick-up as it traverses along the workpiece causes the skid to act as a datum, giving rise to a "floating action" of rising and falling of the pick-up (Figure 38b). The macro-irregularities of the surface, namely of waviness and profile, means that the stylus with its significantly smaller radius to that of the skid will follow this surface topography, sinking into valleys that the skid bridges. However, if the crests and valleys in the workpiece surface are too widely spaced, then this distance with respect to the skid motion (Figure 38a) will also cause the skid to rise and fall and thereby lose the datum registration plane (Figure 38a). If the distance between the stylus tip and the skid (Figure 38a) is half the waviness spacing, then these waviness amplitudes are doubled. Conversely, when the distance of the stylus tip (Figure 38a) corresponds exactly to that of the waviness spacing, then the waviness is mechanically eliminated. The skid type of pick-up is only marginally affected by instrument vibration, because the measurement reference is situated close to the point of measurement. Therefore, the mechanical relationship of the pick-up-to-skid system may be an important factor in its design. If different pick-ups are utilised for the same measuring task, then in extreme situations these measured differences could approach that of 100%.

Ideally, the skid and stylus should be coincident, thereby eliminating the "Abbe error"; the Abbe principle states that *the line of measurement and the measuring plane should be coincident*. For practical reasons, the stylus is slightly displaced from the skid and is situated either in front or behind the stylus. Phasing errors can be induced by the skid (Figure 38b), causing distortion of the graphical trace, while its magnitude depends on the waviness crest spacing. In practice, if the waviness crest spacing is identical to that of the skid/stylus spacing, the skid will rise on one crest, with the stylus rising on an adjacent peak. This relative motion completely suppresses the peak. When the spacings differ, the effect illustrated in Figure 38(b) occurs, where a spurious valley is introduced into the graph at a distance from the peak, this being equal to the skid/stylus spacing. If two skids are used on either side of the stylus, then the "phase effect" of the single skid/stylus spacing is diminished. In some pick-ups the line of contact of the skid is laterally offset, which prevents the skid from burnishing the track (mechanically deforming asperities and giving a false reading) during repeated traverses; this is particularly relevant when the instrument is automatically setting the vertical magnification (auto-ranging).

(a) A rounded skid does not provide a serviceable datum, if the surface irregularities are more widely spaced



(b) The “phase effects” when using a skid



(c) Diagram illustrating an independent datum

An independent datum provides a straight reference for the pick-up

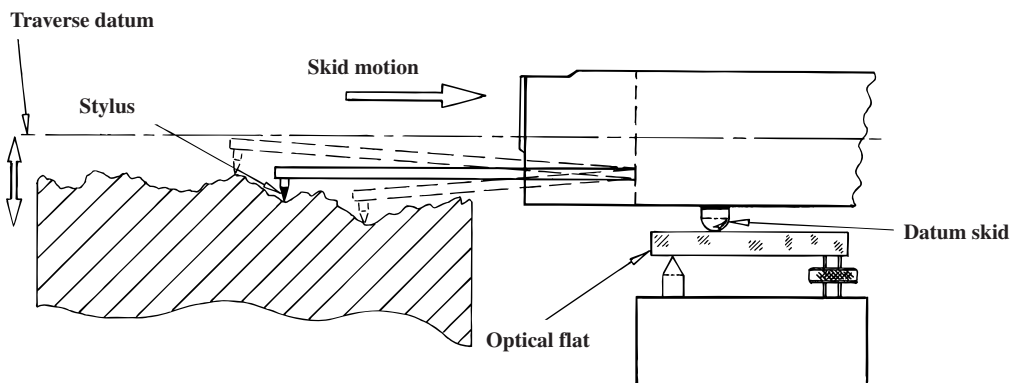


Figure 38. Configuration of surface texture instrument and relative motion of the stylus with respect to an independent datum. (Courtesy of Taylor Hobson.)

The skid acts as a mechanical filter, modifying the profile to a greater or lesser extent. Its radius of curvature, therefore, should be suitably chosen to minimise filtering effects. Generally, if a skid is selected with a radius of curvature of approximately 50 mm, this is normally sufficient for general-purpose applications.

The skidless or reference plane technique can take the form of two distinct types:

- 1 the pick-up skid travels along a flat surface such as an adjacent optical flat, or similar and not on the surface being measured;
- 2 a shaft which moves in an accurate straight line, with the pick-up being rigidly fastened to it – usually a glass block providing an independent datum.

The main reason for employing skidless pick-ups are that they do not eliminate the contributions of waviness and profile errors from the measured topography – which the skid obviously does – and often waviness, in particular, is important to the manufacturing process and must be included in the assessment. The skidless technique provides an undistorted assessment of the actual profile, the proviso being that the reference surface has ideal form.

1.12.3 Portable surface texture instruments

As the assessment of surfaces is not confined to those inspected in a metrology laboratory and often entails the measurement to be made by both production and maintenance personnel, then some easily portable yet robust instruments are necessary. One such instrument is that shown in Figure 39. This battery-operated surface texture measuring instrument can be used across a wide range of angles and attitudes, increasing the range of versatility of the equipment. This instrument has a piezoelectric pick-up which is self-calibrating and can be used in remote operation – split mode – through the infrared (IrDA) link, up to a distance of 1 m without the inconvenience of cable supply. This is an important factor when attempting to measure otherwise inaccessible component features such as bores.

When in transit the stylus has protection, as the “park” position fully protects the stylus assembly. Digital circuitry via *surface mount technology* provides both reliability and accuracy and the stylus assembly closes up (in its robust case) to fit either in the pocket (weight 200 g), or on a belt, which is useful for off-site surface evaluation in the field. The



Figure 39. Portable hand-held surface finish instrument “Surtronic Duo” can be used at any angle/attitude. (Courtesy of Taylor Hobson.)

large and clear metric or imperial LCD display provides simple-to-read surface-related information and instrumental status data, with the following specification:

- “standard” cut-off ($0.8 \text{ mm} \pm 15\%$);
- accuracy: Ra (range: $0.1\text{--}40 \text{ }\mu\text{m} \pm 5\%$); Rz (range: $0.1\text{--}199 \text{ }\mu\text{m} \pm 5\%$);
- diamond stylus ($5 \text{ }\mu\text{m}$ radius);
- traverse length 5 mm;
- minimum bore penetration 65 mm.

Portable and hand-held instruments of this type (within the specification and parameters of the equipment) can establish the surface condition, without the need to break-down assemblies, or the necessity of removing the surface in question to a dedicated and more sophisticated instrument. Such versatility and portability in surface measurement when combined with the confidence in knowing that a representative roughness has been established, albeit for either Ra or Rz , helps in the interpretation of component topographical information from normally inaccessible positions, when compared to that of conventional desk-based equipment. They are, however, limited in their surface analysis and parameter scope.

1.12.4 Surface form measurement

Curved datums

In the previous discussion, all the surfaces have been nominally considered to be flat with only marginal form present. There are occasions when it is necessary to measure the surface texture of curved surfaces such as balls, rollers, the involute curvature of gear teeth and fillet radii of crankshafts (see Figure 40). In the past a wide range of gauges had to be available with sufficient resolution to assess the surface texture together with form, while setting up the surface under test in such a manner that the geometry of its shape did not influence the actual surface texture measurement.

If the limiting factor in the determination of the surface texture was due to a large radius, then this problem could be minimised by utilising a short cut-off length and a skid fitted to the gauge to reduce the *measurement loop*. The skid will induce problems in obtaining a successful and representative surface texture measurement, if fitted in front or behind the stylus – the former technique is illustrated in Figure 41(a). However, the stylus displacement from the measuring plane can be negated by introducing a skid which surrounds the stylus. This tip projects through the skid's centre (see Figure 41c), thus obeying “Abbe's principle”. The solution to curved surface measurement can be addressed by utilising several of the following types of equipment, or techniques:

- as mentioned above – instead of using a straight datum (skid) the gauge is guided by a curved datum (Figure 41c);
- the gauge is held stationary and the surface under it is slowly rotated at a peripheral speed sufficient to gather surface data information;
- the gauge is traversed concentrically around the stationary workpiece surface;
- a wide-range gauge is utilised to enable it to allow for the curvature of the workpiece's surface profile as it traverses the part.

The inspection of a curved surface offers a particular challenge, which can be achieved by guidance of the stylus gauge along a curved path having the same radius as the part surface. The curved surface measurement condition is attainable if datum kits are used (allowing curved surface paths to be traced by the gauge – see Figure 41b). These datum kits tend to be time-consuming in setting up, necessitating a degree of technical expertise in both set-up and alignment. The curved surface is rotated within the datum kit's conical seating against the stylus of a gauge, the stylus being incorporated into the conical seat housing (Figure 41b). The advantage of this set-up is that the tip of the stylus is always aligned to the centre of the component, with the likelihood of *tip flanking* on the part being eliminated. Datum kits are normally only practicable when the radius is constant during measurement, but if a series of radii occur then under this situation they are not to be recommended. Today, such datum kits are not normally used due to the necessity of long set-up times.

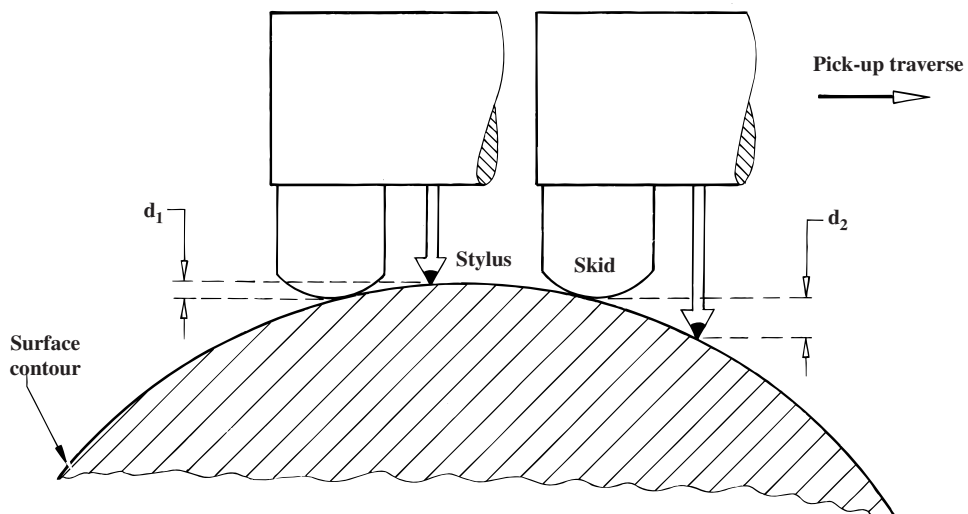


Figure 40. Form measurement with an inductive contour pick-up, including “customer-engineered” work-holding, for crankshafts. (Courtesy of Taylor Hobson.)

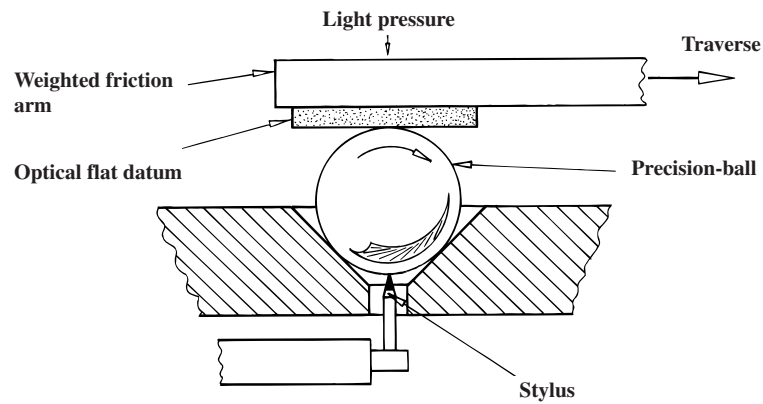
Form analysis

The introduction of wide-gauge ranges on high-resolution pick-ups and the development of various form options (reference line fitting) has meant that guidance of the pick-up along a curved path has been virtually eliminated. Typical of these surface form-fitting features that are currently available are the following software-based options:

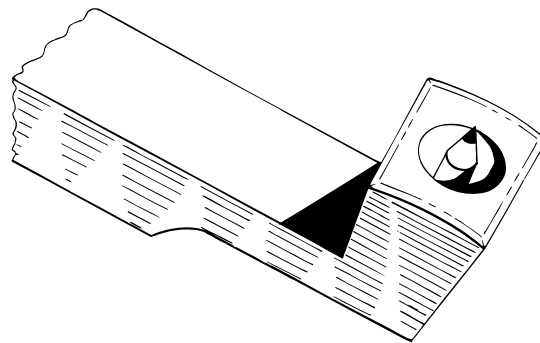
- removal of a least-squares best-fit radius;
- removal of aspheric form from the raw data – form error, surface slope error and tilt comparison with operator-designed data in the form of a polynomial expression (see Figure 42a);
- elimination of elliptical/hyperbolic (conic) geometry – provides major and minor axes values and tilt, with residual surface texture analysis after removal of best-fit elliptical and hyperbolic forms (see Figure 42b);

(a) Stylus displacement solely due to surface curvature (i.e. $d_1 + d_2$)

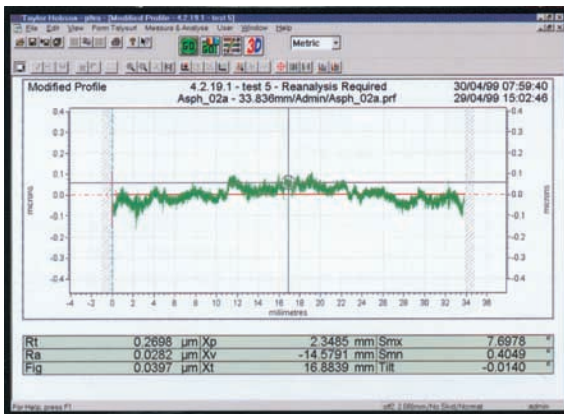
(b) Contour measurement—principle of rotating workpiece instrument



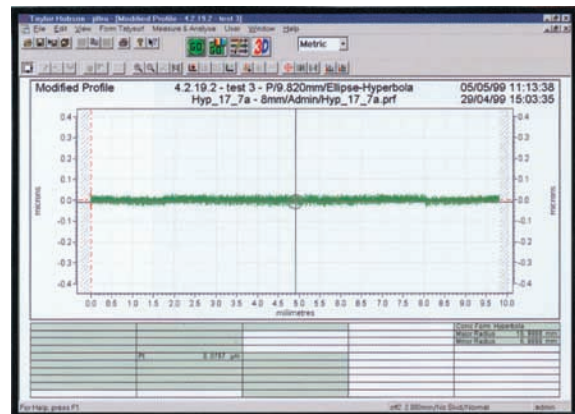
(c) Side skid gauge – the stylus projects through the cylindrical form of the skid

**Figure 41.** Effect of curved surface measurement. (Courtesy of Taylor Hobson.)

(a) Aspheric form software



(b) Conical form software



(c) Dual profile software

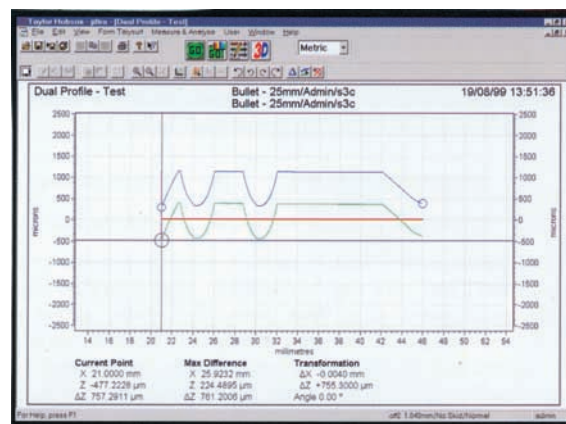


Figure 42. Screen displays of “Windows™-based” software options available for form assessment. (Courtesy of Taylor Hobson.)

- complex symmetrical profile assessment – via an enhanced “Dual Profile” facility, allows the measurement profile of a master shape or prototype component to be saved as a template, enabling future components to be measured and simultaneously displayed along with template data for immediate comparison (see Figure 42c);
- analysis of gothic arch profiles – frequently employed on the ball tracks for recirculating ball screws;
- three-dimensional contour analysis – a representative 3D visualisation of the part surface: axonometric projection, colour height distribution, wear simulation and form removal, with many additional features available (this topic is a subject on its own and discussed in Chapter 2).

The number of surface texture parameters available to define a surface’s condition is immense for form-measuring instruments. Typically, these parameters include 22 profile; 22 waviness; 25 roughness; 12 R+W; 9 aspheric parameters; and this

comprehensive list can be enhanced still further. In addition to a full range of surface texture parameters, form analysis software can provide:

- *form error* – calculation with reference to best-fit concave or convex circular arc, straight line measurement, including surface roughness detail. Alternatively, referencing to the minimum zone, this being the minimum separation between two parallel lines containing the data set;
- *radius* – using the least-squares best-fit, concave or convex circular arcs can be automatically calculated from selected data, with the option of being able to exclude any unwanted surface features from the selected data. Conversely, the absolute radius can be set to analyse actual deviations from a design master, with other calculated parameters including its centre coordinate and the pitch;
- *angle* – using a straight edge or minimum zone algorithm, surface tilt can be established and then removed, prior to parameter analysis, with other

- calculated values including its intercept and pitch;
- *dimension* – the linear relationship of surface features can be assessed and then compared, owing to the ability to calculate both the true X- and Z-coordinate positions.

Contours: profile curving, or an irregular-shaped figure

A typical commercial system is currently available for the analysis of contour measurement; assessment of features such as radii, angles, length and height (see Figure 43 for a typical application) can be achieved with:

- *measurement macros* – these can be learnt and edited forming user programs for repeated inspection routines, which minimise repetitive operator input (achieved via a series of definable “fastkeys”);
- *individual feature tolerancing* – allowing individual sections of the measurement template to be toleranced with a variety of values (a wide tolerance on a flat area of a surface can be defined, with tighter tolerancing to the radii);

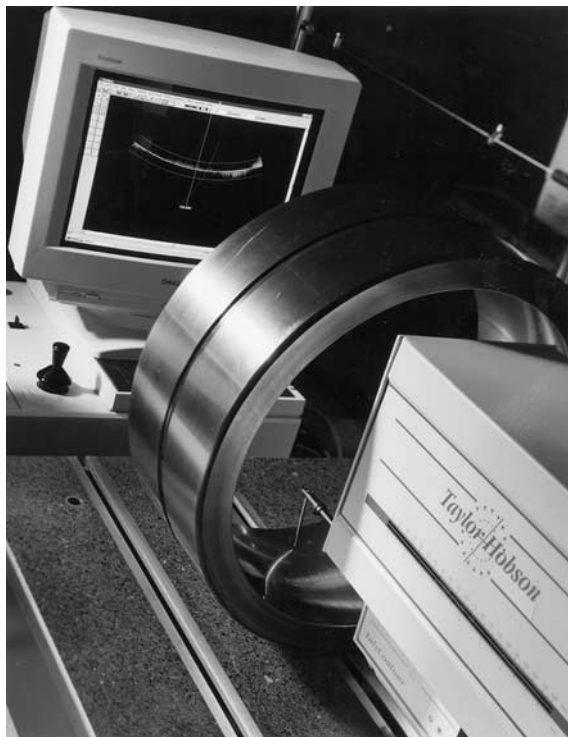


Figure 43. Form and surface measurement instrument “Taly-contour” being utilised to assess a spherical bearing outer race. (Courtesy of Taylor Hobson.)

- *comparison of DFX files to the contour* – if measuring a known contour “best-fit” individual geometric elements are calculated with reference to a template and the comparison of each point in the contour is assessed against the nearest element in this template;
- *geometric element fit to an unknown contour* – when measuring components of an unknown contour, the software finds geometric elements that optimise the geometry of the profile determining the size, position and relationship between profile features, plus other functional factors.

The most recent form and surface texture analysis equipment has absolute positional control, offering sophisticated high-technology instruments providing complex analytical abilities in a single traverse of the workpiece. Lately, with the high thermal stability of the pick-up gauge (phase grating interferometer – PGI), it typically has a large range with around nanometric resolution, equating to a range-to-resolution of 780,000:1. One of the current developments in stylus protection employs a “snap-off” connector, allowing the stylus to be ejected from the gauge if the loading stylus force exceeds safe limits – reducing possible damage to the gauge because of operator error. Interchangeable styli can be fitted to such pick-ups, for better access to the component feature being inspected. If the part is of a delicate nature and liable to deform, or could be easily scratched, then non-contact gauge heads can be utilised. These non-contacting types of gauges comprise a focus follower with a programmable controller for fast, simple use, replacing inductive pick-ups. Non-contact gauges are used where the component surface applications consist of:

- soft and touch-sensitive surfaces – printing pastes and coatings;
- wet or dry solder pastes and printed conductive pastes;
- contact and intra-ocular ophthalmic lenses;
- and many other industrial/scientific applications.

Due to variations in the orientation of the workpiece under test to the traverse unit, some of the latest versions of form analysis instrumentation can be operated in inverted attitudes, or right-angle attachments can be fitted for measurements between shoulders on components, such as on crankshafts (see Figure 40).

1.12.5 Non-contact systems

Contact and non-contact operational aspects on surfaces

As has been previously mentioned, in contact systems the stylus contacts the surface utilising a precisely manufactured stylus and, today, it is invariably diamond-tipped. However, owing to the stylus geometry – its shape and size – some styli on certain surfaces will not be able to penetrate into valleys and as a result create a distorted or filtered measure of the surface texture (see Figures 11 and 12). A recent study has indicated that the actual radius of curvature of a stylus can be very different from its nominal value, while the effect of stylus forces can significantly affect the measured results, with too high a force causing surface damage.

The stylus, to enable a representative surface to be assessed, must traverse across the surface following a parallel path with respect to the nominal workpiece surface. The datum in this case would most likely be a mechanical slideway of some sort (see Figure 38). The datum comprises a skid that has a large radius of curvature (spherical or different radii in two orthogonal directions), fixed to the end of a hinged pick-up (see Figures 36 and 37). At the front end of the pick-up body the skid rests on the workpiece surface; alternatively other designs occur, such as a flat shoe being free to pivot so that it can align itself to the surface, or two skids are situated either side of the stylus. It should be reiterated that ISO 3274: 1996 does not allow the use of a skid, but many are still used in metrological applications in industry.

Notwithstanding the long history of stylus-based instrumentation a number of problems exist, associated with either their operation or the interpretation of results. For example, none of these instruments measure the surface alone: if an inhomogeneous workpiece is inspected utilising a mechanical stylus it responds to the topography and changes in the surface mechanical properties, such as its elastic moduli and local hardness. Another important aspect in contact surface texture inspection techniques is the scale – horizontal and vertical – although it is not common for instrument manufacturers to incorporate metrology into the scanning axes of their instruments, with many companies failing to calibrate scanning axes.

Stylus-based surface texture instruments have an infrastructure of many specification standards, unlike optical/non-contact instrumentation. This point regarding standards for stylus instruments only embraces two-dimensional measurements and

they have yet to be developed for their three-dimensional counterparts. Yet despite this two-dimensional standards infrastructure, there is a distinct need for both good practice and a clear understanding of which surface texture parameter is most suitable for an industrial application to pass down to the shop floor – relating to standards.

For non-contact instrumentation, if one considers some form of optical interaction with the surface that enables local height variations to be established, then it will sample a different surface from that of stylus-based instruments. As a practical example of this difference, if consideration is given to the optical assessment of metals, their surface homogeneity (different phases) could introduce apparent height changes up to 10 nm due to the phase change on reflection; whereas glasses or ceramics may have local refractive index changes and contaminant films introducing nanometric changes. While phase changes at the conducting surface can introduce variations as a function of incident angle, surface field anomalies that are the result of multiple scattering, or sharp points acting as secondary sources of light, together with various edge and wall effects, will all introduce measurement uncertainty.

It was briefly stated that optical methods do not have a standards infrastructure, unlike that of stylus-based surface texture instrumentation, meaning that there exists no formal techniques for calibrating optical instruments. Therefore, as a result of this lack of calibration methodology, care should be taken when optical assessment of a range of surfaces is to be undertaken, particularly if they are of differing physical surface characteristics. For example, if a glass step-height standard is utilised to calibrate the vertical magnification factor of an instrument, it would be unwise to measure a metal surface set at an identical calibration value. More will be said on the topic of calibration in Chapter 6.

Laser triangulation

The laser triangulation, or chromatic aberration techniques can be utilised to measure surface topography, which are dependent on the sensor model employed. Cut-off filters available range from 0.8 mm to 8 mm with large measuring envelopes of up to 150 mm by 100 mm, enabling large or multiple artefacts to be scanned within the same measurement period. A typical instrument will have a vertical range of 300 μm and resolutions of 10 nm from a spot size of 2 μm in diameter, with scanning capabilities of up to 2000 points per second. The measuring ranges currently available are from 0.3 mm to 30 mm, depending on the sensor model.

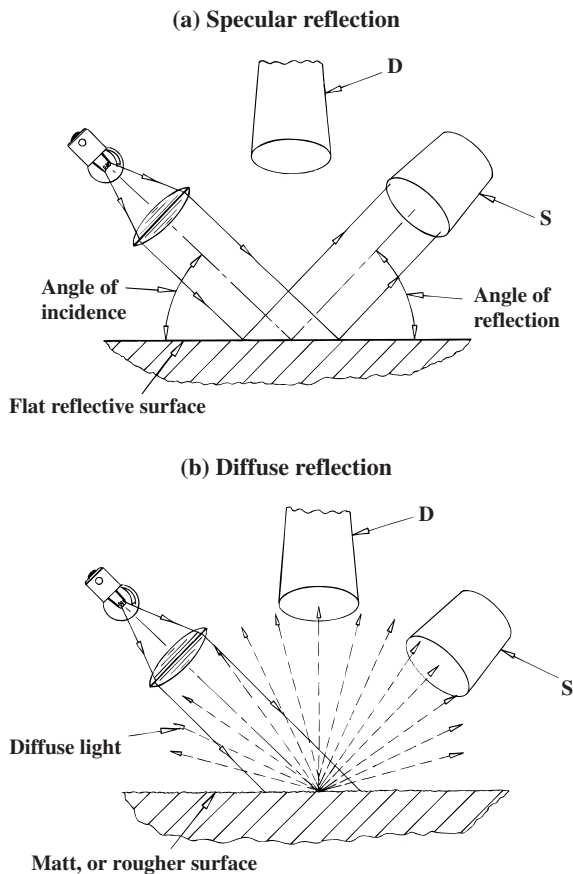


Figure 44. Diagram illustrating the principle of surface assessment by specular and diffuse reflection. (Courtesy of Taylor Hobson.)

One such system has a tested repeatability of $0.03 \mu\text{m}$, but this claim applies to a specimen having a $2.90 \mu\text{m}$ R_a roughness, measured with a 2.54 mm cut-off length. Optional equipment can be added to investigate feature location and alignment or vacuum stages can be incorporated to hold awkwardly-shaped artefacts and replication kits can be supplied to measure inaccessible or difficult to assess surface features.

It is claimed that these systems can measure surface colour or texture. These instruments offer fast and multiple scanning facilities that reduce the time taken for each measurement, compared to conventional contact instrumentation. Common to all optical equipment (as described in the following section), light scattering causes measurement accuracy problems in some component surfaces. If a fixed resolution occurs for each sensor model, this would imply that up to six sensors would be required to enable satisfactory coverage of all surface textures, or types of measurement needed for a complete metrological range of applications.

Optical techniques: reflection and scattering

Highly reflective surfaces are normally assumed to be very smooth, so by measuring reflectivity an indirect technique for measuring roughness can be established. Reflectivity testing is particularly beneficial where visual appearance is important (cosmetic items), or on soft surfaces such as paper where non-contact measurement is necessary. Advantages of optical methods include assessing and averaging over an area and high-speed inspection.

Smooth surfaces invariably appear to be visually “glossy” (reflective); conversely, a rough surface has a matt appearance. In the past, the *gloss meter principle* was used to verify how parallel beams of incidence light at an angle on the surface are reflected.

If a perfectly reflective surface is present (Figure 44a), then much of the light will be specularly reflected (angle of reflection equals angle of incidence) and enter the viewing system at D, being positioned perpendicularly above the surface. However, if a perfectly matt surface is inspected (Figure 44b), the light is diffusely reflected (scattered in all directions), causing equal amounts of diffused light to enter both S and D. The ratio of specular-to-diffuse reflection is a measure of the gloss of the surface. The relative proportions of light entering S and D can be measured by either visual means or photoelectric comparison, incorporated into a suitable portable instrument. As it is the *ratio* of the light that is measured, any absorption of light by the surface will not influence the result, and neither will fluctuations in the irradiance distribution of the lamp, or more specifically today, the laser.

Considerable interest has been shown over the years in the use of scattering for surface texture measurement. In essence, three techniques are generally used for scatter-based instrumental design:

- 1 light scattering theory, when used to construct an instrument giving absolute measurements of surface texture – when a number of conditions are met;
- 2 general theoretical approach in instrument design – assumptions are made that have general connection with a particular surface texture parameter;
- 3 application-specific approach, where instrument design is developed to solve an immediate problem.

In just one example, an *integrating sphere* (Figure 45) can be used to measure gauge block surface

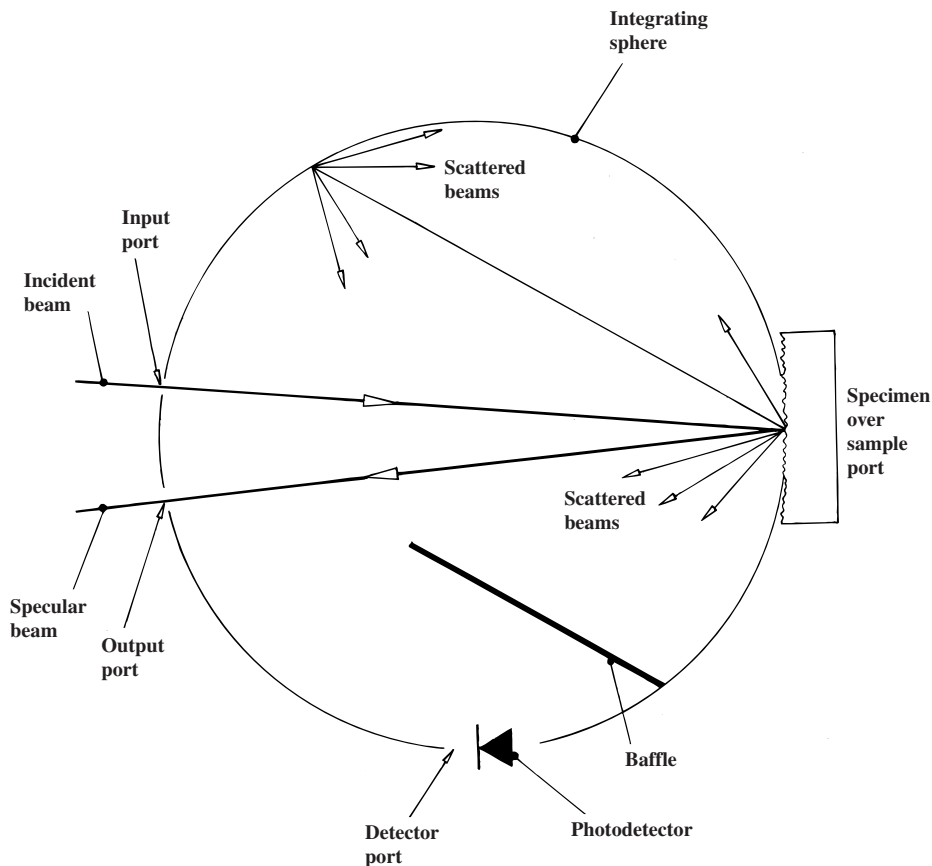


Figure 45. Geometry and layout of an integrating sphere. [Courtesy of Dr R. Leach/NPL.]

texture. The instrumental technique measures the total intensity of light from a gauge surface, then compares it with the light intensity that was diffusely scattered. The ratio of two intensities is termed *total integrated scatter* (TIS) and it is proportional to the square of the RMS surface texture parameter Rq . This method relies on a number of assumptions that must be met for the type of surface to be measured. The diffuse component of reflected intensity is measured by collecting light in an integrating sphere. An integrating sphere can be practically represented by a hollow sphere that is coated internally with a highly diffuse reflecting material. Any light reflected from the inner sphere's surface gives rise to a constant flux of electromagnetic power through an element of the sphere. This sphere has a number of ports, allowing radiation both in and out, with the gauge to be irradiated and photodetectors to detect the intensity of integrated light.

Interference instruments

Surface texture interferometry occurs in several instrumental configurations, probably the most common types being interference microscopy, plus full field methods. Becoming popular of late are phase-stepping techniques and swept-/multi-frequency source methods. Today, there are several commercially available phase-shifting interferometric microscopes, giving three-dimensional surface imaging with very short measuring times. The National Physical Laboratory (NPL) has developed a sub-nanometre resolution system that employs a microscope objective with a numerical aperture of 0.5, in conjunction with a birefringent lens providing a focused spot from a laser source and a defocused orthogonally polarised beam $10\ \mu\text{m}$ in diameter. Fourier analysis has been obtained from surfaces, indicating that with this objective the system produces a surface wavelength range from 0.5 to $15\ \mu\text{m}$. The trace displayed by the system is an interferometrically obtained path difference of the focused probe beam and the defocused beam.

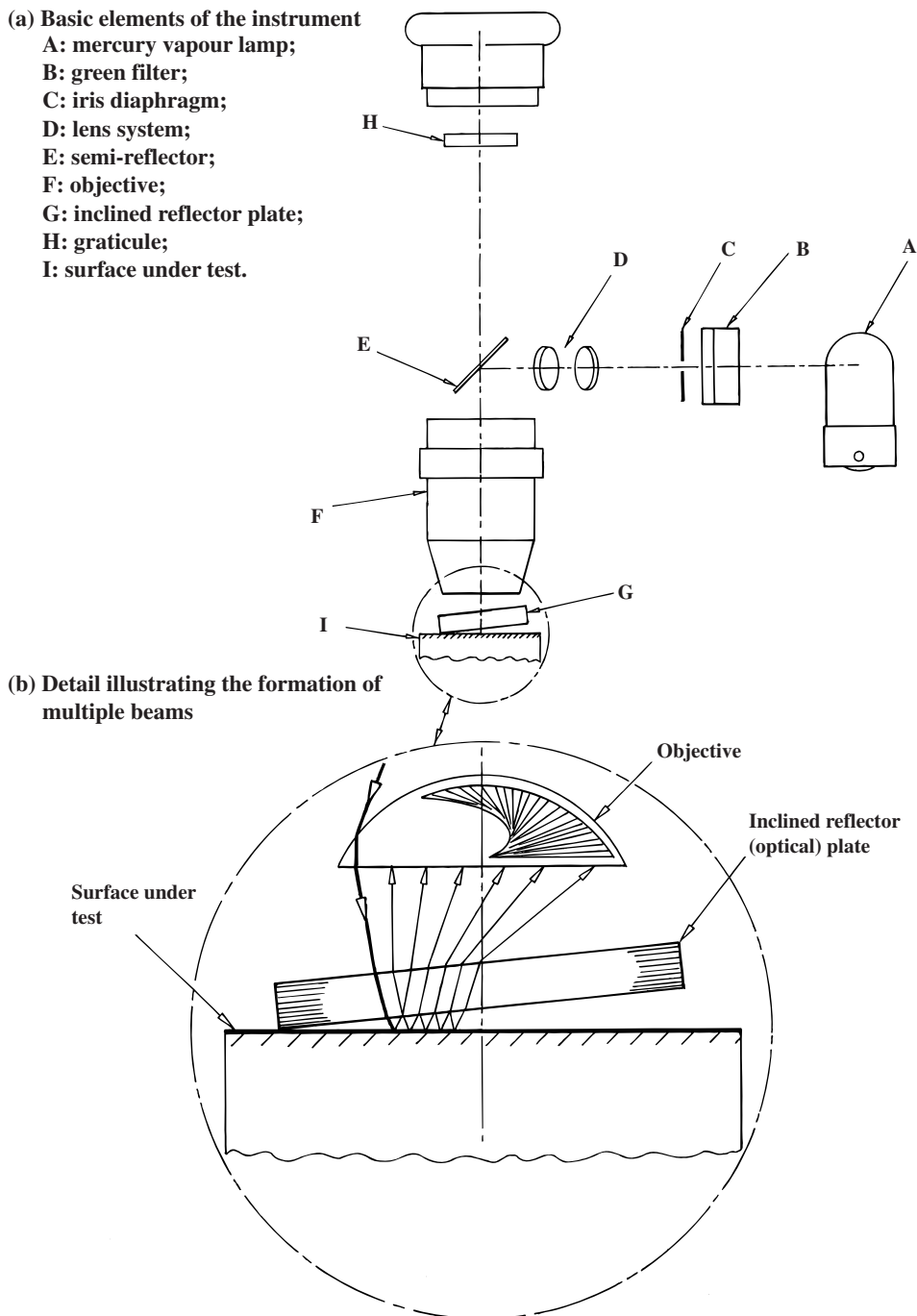


Figure 46. Fizeau-type surface texture interferometer. (Courtesy of Taylor Hobson.)

As has already been mentioned, light interference can be successfully used as a non-contact optical technique to assess surface features. Essentially, the principle of interference of light relates to how light rays are reflected between two surfaces; the differing light path lengths along various points of the surface

cause phase changes in the light being reflected back to the observer. The result of this reflection will be to promote alternate dark and light interference fringes. The shape and spacing of these fringes will depend on the regularity of both the surface and the reference reflector. The surface texture irregularities

are reproduced as modifications in the interference pattern and, when viewed under specific conditions, displacement of the fringes is a measure of the roughness height.

For some interferometer instruments such as the Fizeau type, interferometric surface assessment relies on the fact that a “wedge of air” occurs on the test surface, with the resulting interference lines being of equal height and their respective height differences given by

$$\text{Fringe spacing} = \lambda/2$$

where λ = wavelength.

Under examination, the surface roughness produces air wedge thickness variations causing deflection of the interference fringe, from which the total height parameter Rt results. Hence the total height becomes

$$Rt = a/A(\lambda/2)$$

where A = interference fringe, a = fringe deflection and $\lambda/2$ = fringe spacing.

The height of irregularities that can be resolved range from 0.005 to 1 μm using conventional interferometric methods. The range can be extended by using the technique of *oblique incidence*, which increases fringe spacings from $\lambda/2$ to 2λ , enabling one to measure a surface roughness height four times larger than that measured using a standard interferometric method. Furthermore, if cast replicas of the test surface are made, then immersing them in fluids of a suitable refractive index enables transmission interferometry to assess roughnesses up to 25 μm .

Maximum contrast is important for optical interferometry, and the wavelength of emitted light must remain constant to eliminate variations in the phase difference between interfering beams. Any defects in optical surface quality within the interferometer can promote a reduction in fringe contrast, as can vibrations of optical components, together with parasitic light caused by unwanted reflections. Due to the strongly convergent light utilised in these interferometric optical configurations (Linnick and Mirau interferometers), this can lead to obliquity effects, which can introduce errors in estimation of, say, a scratch depth by as much as 12%. The magnitude of error is dependent on the numerical aperture of the objective and the surface's scratch depths. However, in the case of more commonly available full field techniques this obliquity and hence error is not a problem.

In essence, the Fizeau surface texture interferometer is basically a microscope, with a built-in illumination and special-purpose optical system (see Figure 46a). Between the objective and the surface a semi-reflector G can be positioned, which

may be slightly inclined to the work surface I , allowing a wedge-shaped air gap to occur (detailed in Figure 46b). Multiple reflections between the surface and the reflecting plate produce good-contrast, sharp fringes (see Figure 47 for typical interferometric images) which are viewed in the eyepiece. Utilising a spherical semi-reflecting surface, it is possible to examine three-dimensional curved surfaces, such as precision ball surface finishes (Figure 47c). Multiple reflection (Figure 46a) on the semi-silvered surface E and the workpiece I means that all on-coming beams are split into several partial beams, which cause interference. Interference fringes that occur are both higher in contrast and narrower as the reflection coefficient becomes greater. If too great a reflection coefficient occurs, the contrast will degrade. Such conditions allow multiple beam reflection to occur only when the distance to the reflection surface and that of its test surface is quite small.

Large apertures can be utilised by this Fizeau surface interferometer, enabling full exploitation of the light microscope. When employing high apertures, the line standard (graduated scale) is dependent on the selected aperture and, considering the associated aperture angle, the fringe spacing is

$$\text{Fringe spacing} = \lambda/2 (2/1 + \cos \mu)$$

where λ is the wavelength, μ is a cosine error.

Normally, the largest aperture (A) utilised is 0.65, hence the fringe spacing equation above can be simplified to

$$\text{Fringe spacing} = 1.14\lambda/2$$

Generally, it can be said that applications of this particular Fizeau surface texture interferometer have limited use, although the technique does have some merit:

- it can examine a relatively large specimen surface area;
- it indicates any form error present;
- being a non-contact method, it may be possible to estimate the scratch depths.

An optical and practical limitation of this type of interferometry when applied to precision surfaces occurs due to visual viewing, which tends to be time-consuming and somewhat fatiguing. By using automated image analysis techniques for detection, this limitation could be overcome, but at a considerable cost disadvantage to the user.

Although three-dimensional surface texture analysis will be discussed in Chapter 2, one instrumental technique has been included in this chapter. An automated surface profiling interferometer is

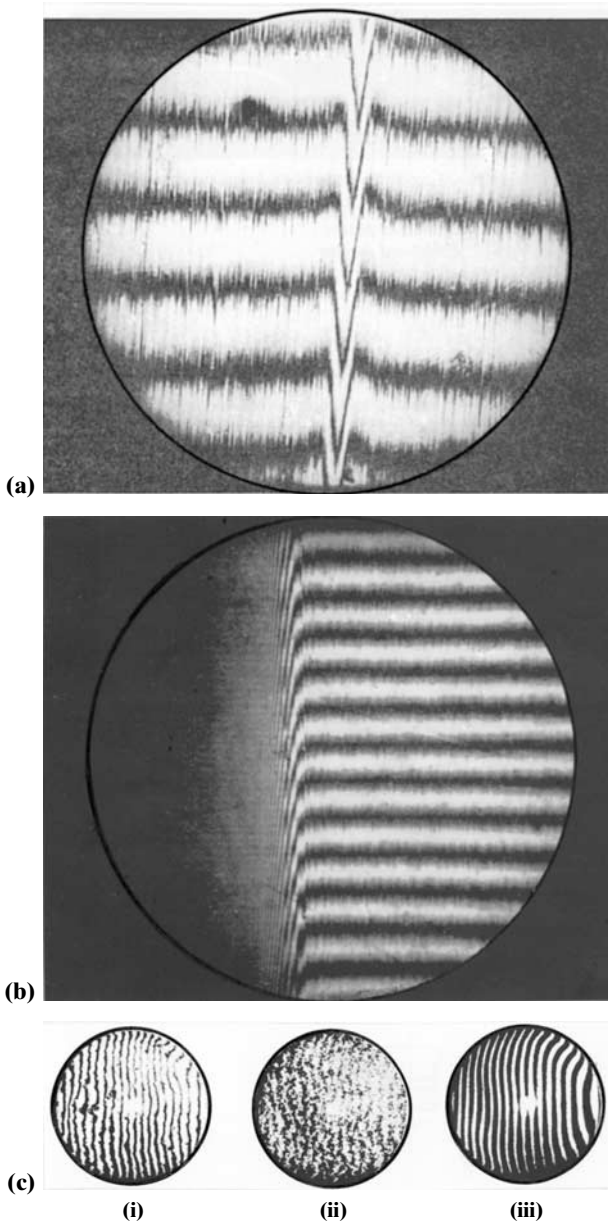


Figure 47. Typical interference fringes from a surface texture interferometer, showing both flat and spherical surfaces (Courtesy of Taylor Hobson).

(a) Lapped surface of a gauge block, showing a scratch. Fringe spacing is $0.25\ \mu\text{m}$. Due to a scratch extending across adjacent fringes, its depth must also be $0.25\ \mu\text{m}$. (b) Lapped gauge block edge, indicating edge rounding. (c) Interference patterns on spherical surfaces (i) pitch polished steel ball; (ii) commercial steel ball; (iii) glass sphere.

illustrated in Figure 48 that has a wide range of industrial applications, from research to manufacturing process control. Of particular relevance is its use in semiconductors, disk drives, printing plates and fuel injector seals, together with applications where the

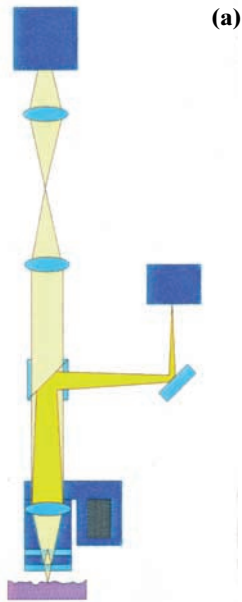
surface texture is used to control the product's performance. Until recently, microscopes tended to be utilised purely for attribute sampling, such as imaging surface topographical details, while surface profilers provided accurate measurements to characterise the surface details. One of the latest developments in surface texture (white light) interferometers is illustrated in Figure 48, which combines these two technologies, providing a fast, quantitative surface measurement on many types of surface topographies. The surface details that can be quantified include surface roughness, step heights, critical dimensions and other surface topography features within a matter of seconds. Profile heights that can be assessed range from $>1\ \text{nm}$ up to $5000\ \mu\text{m}$, at specimen translation speeds up to $10\ \mu\text{m/s}$ with a $0.1\ \text{nm}$ height resolution, which is independent of both the magnification and feature height. This interferometer can resolve sub-micrometre X- and Y-plane features on profile areas up to $50\ \text{mm}^2$, achieving this by *image stitching*. This stitching technique has been designed to analyse surface details much larger than is possible with a single measurement. Stitching occurs by taking several measurements of the test specimen as it is moved by a motorised stage and then combines – stitches – the multiple data sets into one surface image. In effect, stitching increases the field of view, without compromising lateral or vertical resolution. The scanning actuator is of a closed-loop piezoelectric variety, employing low-noise capacitance sensors to ensure that both accurate and repeatable linear motion occurs over the whole operational range. Other notable modes include both “Phase Stepping Interferometry” (PSI) and “White-Light Scanning Interferometry” (WSI) techniques.

Use of speckle

This technique for surface texture inspection utilises partially coherent light, with the reflected beam from the surface consisting of random patterns of dark and bright regions termed *speckle*. The spatial patterns and contrast of the speckle depend on the optical system configuration utilised for observation and the condition of coherence of the surface texture. Speckle has two important attributes:

- contrast;
- number of spots per unit area.

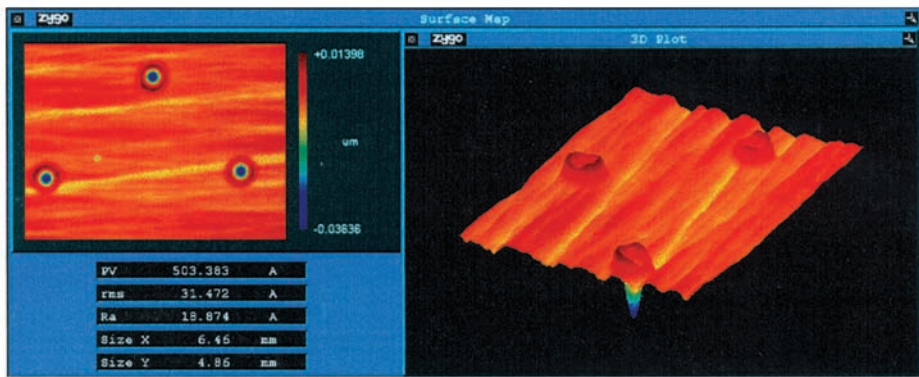
The contrast can be shown to have a relationship to the surface's correlation length, whereas the speckle density is primarily concerned with the image system resolving capacity. Information on the specimen's surface can be obtained using the contrast of speckle patterns produced in the first instance, near



(a)



(b)



(c)

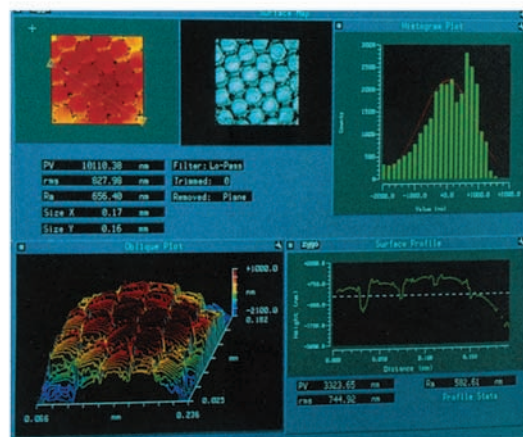
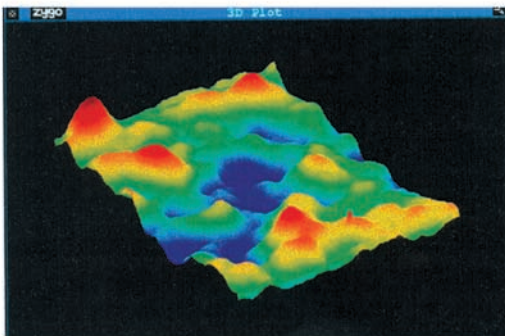


Figure 48. Automated non-contact three-dimensional surface profiling interferometer. (Courtesy of Zygo Corporation.) (a) Optical configuration. (b) View of interferometer. (c) Typical screen displays.

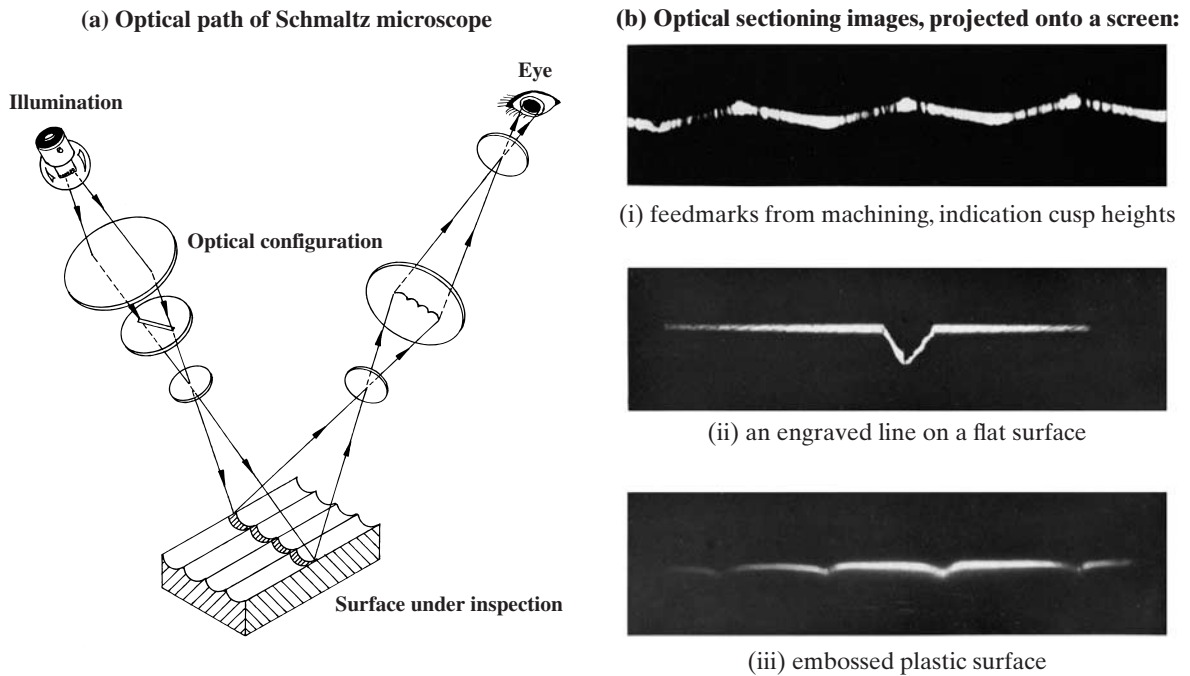


Figure 49. The “Schmalz technique” – optical sectioning principle and associated projected images. (Courtesy of Taylor Hobson.)

the image plane, and secondly, close to the diffraction or defocus plane. Yet another technique utilises the polychromatic speckle patterns, while others employ the correlation properties of two speckle patterns. From both of these methods, surface texture values ranging from 0.01 to 25 μm have been estimated.

Optical sectioning

The method of optical sectioning termed the *Schmalz technique* after its inventor (sometimes termed the *light section microscope*) produces virtually identical results to that of physical sectioning, but in a simpler, non-destructive manner. However, the magnification and data obtainable from the projected image are quite limited, when compared with stylus-based surface texture measuring instruments. The ratio of vertical to horizontal magnification is around unity, meaning that the field of view is small when a large magnification ($\times 400$) is used, hence for fine surfaces it is rather limited.

The optical sectioning principle is illustrated in Figure 49(a). The surface to be examined is illuminated by a thin band of light delineating a profile section, which is then viewed at an angle with a microscope. Typically, illumination and viewing angles are 45° to the work surface producing the clearest profiles (see Figure 49b). For example, the projected surface indicated in Figure 49(bi) illustrates an apparent machined profile cusp height of

$$\text{Cusp height} = h \times \sqrt{2}$$

where h is the actual profile height (R_t).

Measurements are normally undertaken using either an eyepiece graticule or an eyepiece incorporating a micrometer for topographical feature assessment.

An alternative technique for optical sectioning can be achieved by utilising an optical projector, as this has the added advantage of an enlarged image being projected onto a screen. The surface's profile height can then be quickly measured with a specially prepared and calibrated template, which compensates for any distortion introduced by the viewing angle.

The optical sectioning technique is suitable for surface topographies having a roughness range of R_t between >1 and 200 μm . Moreover, it is eminently suitable for the inspection of soft and slightly pliable surfaces (Figure 49bii, which could be deformed by a stylus), or for estimation of depths of surface scratches/engraved lines (Figure 49bii).

The reflectance of a surface is a “sensitive function” of its relative roughness, and consequently the wavelength of light is considerably greater than the root mean square value of the surface texture. Hence, the reflectance will depend only on the surface roughness rather than the peak slope of any irregularities. Measuring reflectance at two distinct wavelengths enables one to determine the surface roughness and slope of asperities (peaks).

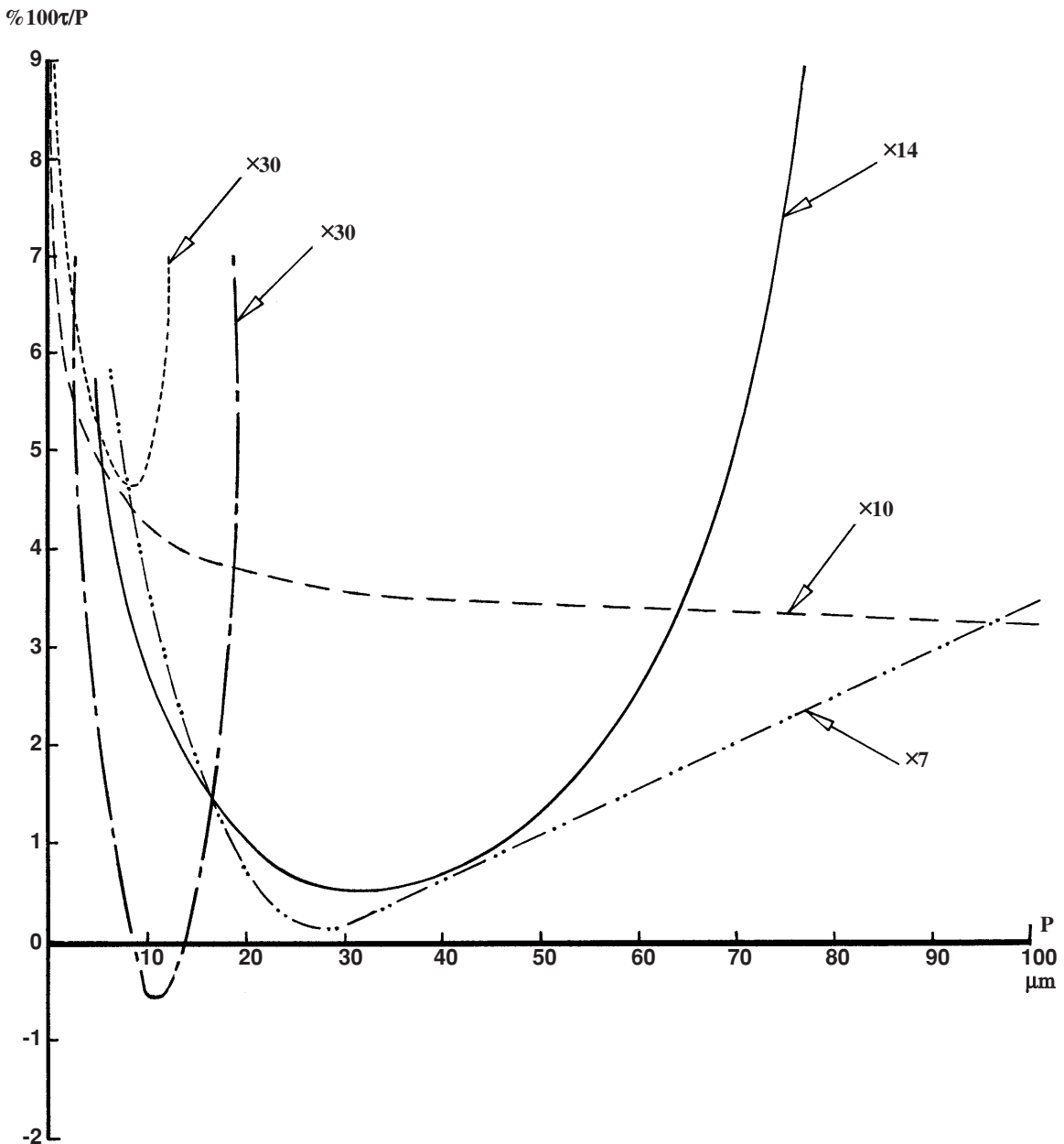


Figure 50. Performance curves for the Schmaltz – optical sectioning – microscope, illustrating the relative error of indication for various microscopes. [After Thomas, 1974.]

Performance tests undertaken for the Schmaltz (optical sectioning) microscope (illustrated in Figure 50) show that for an objective of $\times 60$ magnification the relative error of indication does not exceed 7.7% over the whole application range (Rt values between 0.8 and 3.0 μm). Accepting that practical accuracy limits are in the region of 9%, then there is little point in using a $\times 60$ objective, since the necessary accuracy can be achieved with the

$\times 30$ objective – with the added advantages of a larger field of view and depth of focus. In the case of the $\times 30$ objective, the relative error of indication does not exceed 8.6%. Now, if one enlarges its measuring range (1.5–10 μm), up to values of Rt of 1.5–14.5 μm , a practical accuracy of $\pm 9\%$ may be obtained. The objective having a magnification of $\times 14$ has a relative error of indication within the recommended manufacturer's guidelines, namely, from 3 to 12 μm ,

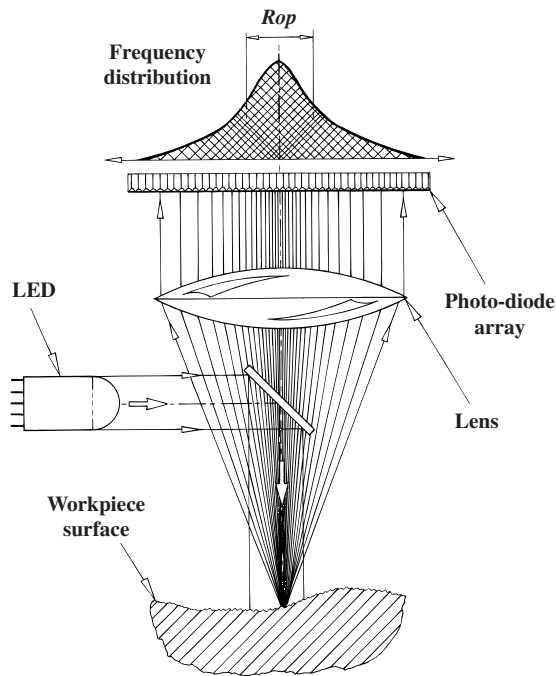


Figure 51. Diagrammatic representation of sensing with scattered light.

varying from 6% to 12%. Moreover, if the measuring range is altered for this $\times 14$ objective to that of R_t ranging from 5–65 μm , the value of its relative error of indication is maintained within the limits of $\pm 5\%$. In a similar manner, for the $\times 7$ magnification objective, by enlarging the measuring range from 8 to 100 μm the relative error of indication will still be within a $\pm 4\%$ limit.

Scattered light: light-emitting diode (LED) source

Utilising “scattered light” is a fast in-process surface roughness technique, which is schematically illustrated in Figure 51. The parallel light is emitted by an infra-red LED and is projected onto the workpiece surface. For this instrument, the size of the source’s light spot is approximately 0.8 mm at the focal point on the surface. Any reflected light tends to be scattered or diffuse in nature, resulting from the roughness of the workpiece surface. The lens will collimate the light and project it onto an array of photo-diodes, from where the variance of the distribution of scatter can be established. Any variance is then normalised, which eliminates intensity variations resulting from the reflected light. Normalising the variance compensates for the twin effects of different levels of reflection and that of workpiece

materials; moreover, this technique eliminates the influence from variations in intensities of LED. Other techniques have been developed, for example *angle-resolved scatter* (ARS), but the discussion here will be confined to the current scattering technique.

Variance in the scatter distribution is termed *Rop* (the optical roughness parameter). This variability is proportional to the angle of the surface roughness. *Rop* can be utilised to compare the workpiece surface roughness against a calibrated standard, although this parameter can only be validated when measurement is made under similar conditions. If component surfaces are quite smooth, then these are ideal conditions for obtaining the highest sensitivity from the instrument. Typically, if a component’s general surface roughness (R_a) lies between 0.05 and 0.5 μm , it is appropriate for evaluation by this technique. Any surfaces that might be manufactured by electrical discharge machining (EDM), electro-polishing, or grinding (pseudo-random production operations) are ideal for assessment by the scattered light technique, rather than other systematic manufacturing processes such as turning and milling operations. Only a brief time period is necessary for an object’s surface measurement by the scattered light instrument – typically 300 ms – enabling measurement “on the fly” (while the workpiece is in motion).

A practical application of this technique might be to establish the surface roughness of an in situ measurement for a cold-rolled aluminium strip as it progresses beneath the focused beam at maximum strip passing speeds of around 70 km/h. Due to its ability to tolerate relatively fast component motion and quickly estimate surface finish, the scattered light technique can achieve 100% inspection on critical surfaces, enabling almost real-time process control for continuous processing operations.

Optical diffraction

The principle behind this instrument’s operation is the technique of using a surface under test as a phase grating, enabling the instrument (Figure 52) to capture the far-field Fraunhofer diffraction pattern on a linear CCD detector array, thus allowing statistical assessment of the surface condition.

A schematic diagram of the instrument is depicted in Figure 52, with a photograph of the complete assembly illustrated in Figure 53. The system is based upon two laser diodes having operational wavelengths of 670 nm and 780 nm, respectively. These two wavelengths enable diffraction order numbers to be identified, in conjunction with a low birefringence single-mode fibre terminated by an achromatic collimating lens. The purpose of this fibre is twofold:

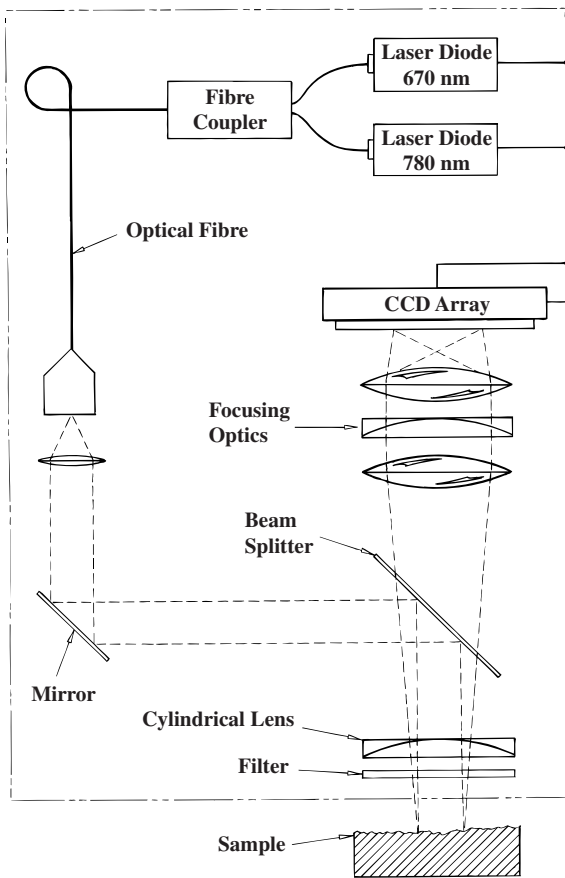


Figure 52. Schematic representation of the Talyfine instrument. (Courtesy of Taylor Hobson.)

- to act as a spatial filter, by selectively transmitting the Gaussian TEM 00 mode;
- providing a convenient method for the production of spatially coincident wavefronts.

The Gaussian beam for both laser wavefronts is directed at normal incidence through a cylindrical lens onto the workpiece surface. It forms a line image at the beam waist, so that a restriction of diffraction occurs only on one plane, enabling a range of metrological parameters to be found such as R_a , etc, and the surface power spectrum. Surfaces can be assessed whether they are random or unidirectional in nature.

Diffraction patterns are produced from the phase modulation of the surface and will be focused onto a 2048 CCD detector. Lastly, data processing is undertaken on a PC, enabling the visualisation of diffraction patterns, power spectra and surface parameters.

With this instrument the surfaces under test must be quite smooth, in order to obtain a satisfactory visual performance from the equipment. Typical



Figure 53. Assessment of surface texture on a diamond-turned component, within the range 5–10 nm R_q , utilising a non-contact optical instrument. (Courtesy of Taylor Hobson.)

components that might be inspected with this instrument include diamond-machined brass (Naval non-leaded type), aluminium (grade: 6061 T6) and copper (OFHC) flats, germanium flats or similar precision surfaces.

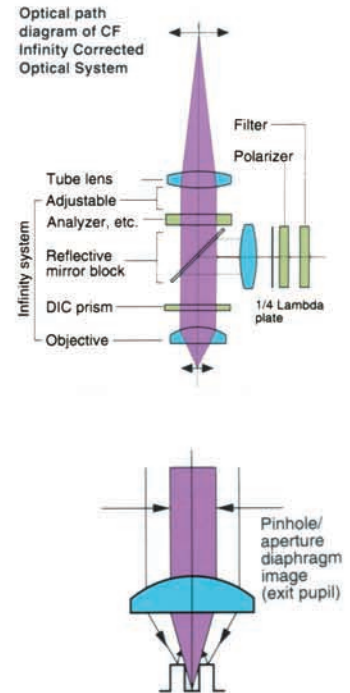
One of the principal reasons for utilising laser-based diffraction techniques is their ability to characterise low R_q surfaces, in particular the advantages obtained from directly filtering the diffraction pattern. Moreover, this instrument has the capability of measuring a substantially periodic surface to an R_q of approximately 140 nm. In addition, measurements of R_q on a rotating surface can be achieved with no distinction between whether the surface is in motion or static (unlike a stylus-based instrument).

The surface diffraction physics is rather complex and is therefore beyond the scope of the current discussion on the general operational and performance characteristics of this laser-based instrument.

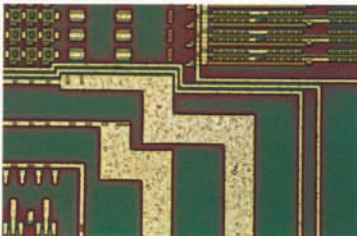
Microscope applications

A cursory examination of a surface with a microscope equipped with a graticle will reveal any prominent spacing irregularities. The height of any dominant peak cannot be estimated, although if suitable illumination is utilised (see Figure 54 for various inspection modes that can be employed) some detail can be visually assessed.

In metallographical inspection of samples, if destructive means can be practised to visually assess surface features at approximately right angles to any interesting irregularities, this form of visual inspection reveals considerable information. Rather than section the material perpendicularly to the surface, a much better and informative strategy is to take a *taper section* of the surface. With taper sectioning,

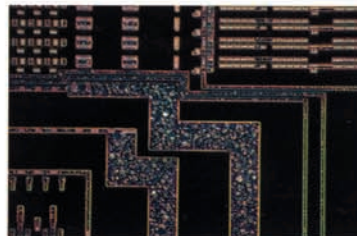


INSPECTION MODES



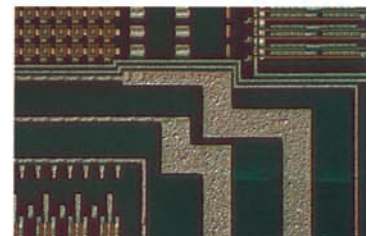
Brightfield Observation

Brightfield microscopy uses differences in reflection to enhance the natural color and form of specimens.



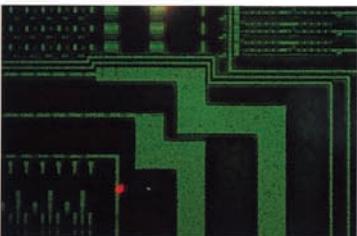
Darkfield Observation

Darkfield microscopy detects tiny flaws, subtle irregularities, impurities and defects on the surface of wafers or masks. It permits highly accurate inspections.



Nomarski DIC Observation

Differential interference contrast microscopy reveals the tiny flaws in wafers and the subtle irregularities of photo resist patterns as an interference color, thereby allowing three-dimensional observation. DIC represents subtle inclination (differential coefficient) with sharp differential contrast, enabling detection of very small differences in height.



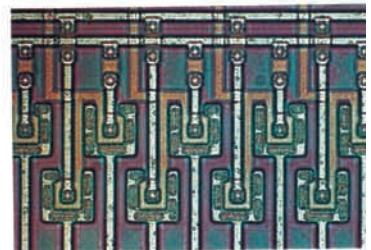
Fluorescence Observation

The epi-fluorescence filter blocks facilitate detection and analysis of residual photo resists in the semiconductor manufacturing process. The filter detects particles as fluorescent images and determines their wavelength. It is ideal for locating photo resist irregularities and the cause of particle generation.



Simple Polarization Observation

This method is applied when studying or analyzing specific optical characteristics—such as optical isotropy and anisotropy—of a given material. An interference color can be applied to allow polarized contrast observations. This method is ideally suited for observing crystal conditions and detecting the stress on a wafer.



Illumination using pinhole aperture

Figure 54. Microscope applications for surface assessment. (Courtesy of Nikon UK Ltd.)

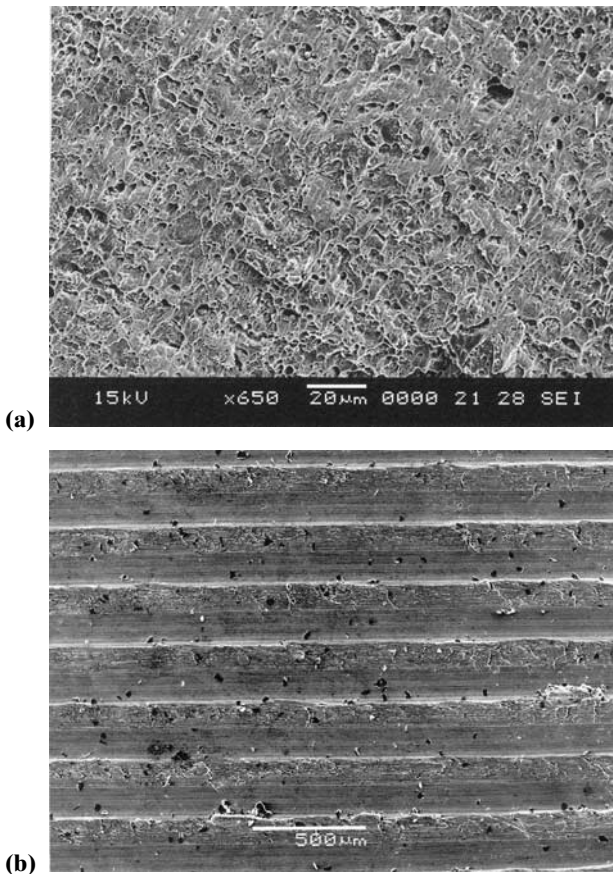


Figure 55. Images from a scanning electron microscope (SEM) for optical assessment of fracture and machined surfaces. (Courtesy of Jeol UK Ltd.)

(a) Mild steel ductile fracture – photomicrograph from an SEM. Magnification $\times 650$; acc. V 15 kV; signal SEI; WD 21 mm; SS 28; pressure Pa. (b) Stereo image (left-hand SEM stage tilt -1.5°) of an austenitic stainless steel (grade 316) by high-speed machining (milling with a 0.5 mm feedrate). Magnification $\times 50$; acc. V 20 kV; signal SEI; WD 20 mm; SS 30; pressure Pa.

the surface is prepared at a shallow angle (typically 11°) perpendicular to the surface (see Figure 161). This slight taper does not unduly modify the shape of the surface irregularities, but increases the apparent magnification of the surface under test, revealing a considerably greater area for visual or micro-hardness assessment (more will be said on this latter topic in Chapter 5). Surface preparation is a difficult task, particularly during the polishing phase prior to suitable etching (normally necessary for metallographical inspection of metals) as surface profile distortion may occur if sufficient care is not taken. Sectioning reveals modifications in sub-surface microstructural details that might not otherwise be apparent and this will be discussed in Chapter 5 on surface integrity; this may offer

engineering solutions as to reasons why a surface behaved in a particular manner in service. Engineering surfaces may fail (in service) for a variety of unanticipated reasons; these might include high surface residual hardness; the introduction of unstable metallurgical changes sub-surface grains exposed at the surface occurring under conditions of load or tribological action. Together with a combination of these factors (see Figure 55a), their failure mode might need to be investigated.

A stereoscan electron photomicrograph offers a more informative visual three-dimensional image of the surface area than can be obtained from a conventional optical technique (see Figure 55b), particularly when additional aids of *soft imaging* with topographical height profiles are utilised, as discussed later in Chapter 3 on surface microscopy.

Optical microscopes have wide-ranging capabilities and offer superb yet subtle techniques to enhance surface image quality, including the following inspection modes (see Figure 54):

- *Brightfield* – the most frequently used method of observation, employing differences in reflection to obtain natural colour and shape of surface;
- *Darkfield* – this method becomes of some significance when observing and photomicrographing surface irregularities, minute flaws, differences in height levels and samples with low reflection levels (paper, plastics, composites and fibres);
- *Normarski* – this technique can capture the surface's subtle irregularities and flaws as interference colours, indicating them as three-dimensional forms. Owing to the Normarski's ability to represent very small tilt with sharp differential contrast, this observational technique can detect minute height differences and subtle irregularities in metals, crystalline structures, integrated circuits (IC) and large-scale integrated (LSI) circuits;
- *Fluorescence* – an epi-fluorescence attachment allows for detection of positive and negative photo resists. Selecting from different filter combinations offers an optimum configuration for any application, being ideal when trying to locate resist irregularities and the cause of particle generation;
- *Simple polarisation* – specific optical characteristics can be investigated with this method of polarised light microscopy. The technique is widely utilised for geological research, together with evaluation and examination of minerals, plastics, crystal conditions and stress detection on IC wafers;
- *Pinhole aperture* – allows high-intensity/-resolution images at large depth of focus, which is ideal for surfaces having multi-layer film three-

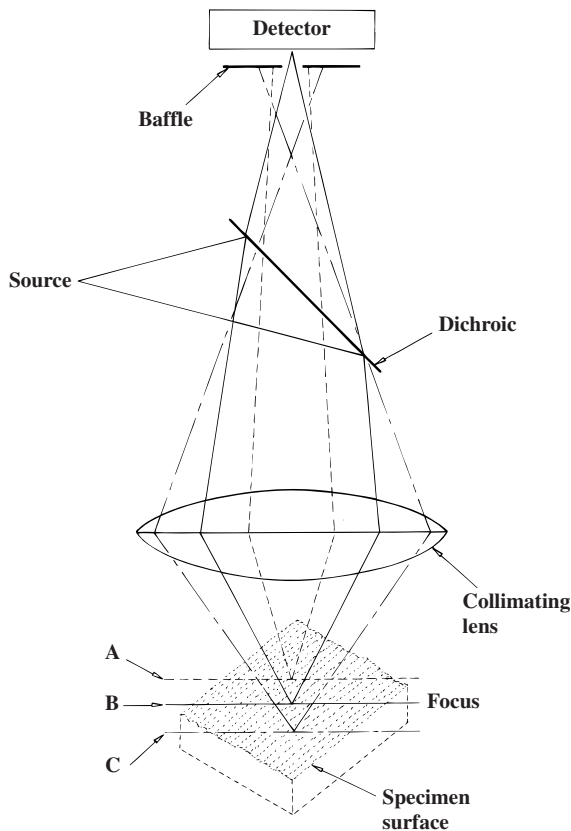


Figure 56. Schematic diagram of an optical profiler. (Courtesy of Dr R. Leach/NPL.)

dimensional structures. Typically, the aperture diaphragm can be “stopped-down” to around 30% of the exit pupil diameter using a $\times 150$ objective, providing large depths of field for surface inspection.

Many of the latest microscope systems can be used for *image capture and processing*, enabling the instrument’s software to “grab” live images from the microscope via video cameras, scanner computer files and archive packages. Such image analysis systems enable image enhancement and manipulation (flip and rotate the surface image), with the ability to change contrast/brightness, simple thresholds, sharpening and smoothing, morphology, component separation, retouching and masking, together with scaling and filtering. Many software-enhancing techniques allow operators to take up to 36 geometrical and analytical measurements, together with reference images for comparison, with Windows™-based menus for ease of image and data manipulation. Such systems can be networked for multi-user applications and sophisticated archiving facilities, allowing access by different users.

Confocal microscope

Confocal microscopes generally utilise the principle of depth discrimination. Essentially, the imaging and receiving optics are identical, offering excellent properties. In a similar manner to the so-called “flying spot technique”, the optics project a point onto the surface, with the reflected light being picked up by a point detector (see Figure 56). The resulting signal is utilised to modulate the spot brightness on the CRT screen, which is scanned in synchronisation with the object under test. Essentially, the lens focuses a diffraction-limited spot onto the object, with the lens collecting light only from the same vicinity of the surface that was previously illuminated. As a consequence of employing the single-point detector, both the image forming/receiving systems contribute to the signal at the detector. Thus, by rotating the variable aperture over the specimen, which itself in some systems rotates, it will build up an optical image (via software data capture) of the sample’s surface. By this means, the “flying spot” covers the complete specimen’s surface to the same optical magnification and resolution of a conventional microscope, but with the advantage of a significant improvement of the field of view. These images can be optical slices or sections through the surface, which can be processed providing non-contact three-dimensional information about the surface.

1.13 Nanotopographic instruments

In the mid-1960s, out of the growing requirement to measure very smooth surfaces and thin films, an instrument was developed having a resolution of approximately 0.5 nm combined with a maximum magnification of 1,000,000, a measuring range of 13 μm and an arcuate traverse of ± 1 mm. Originally such instruments were developed in order to measure minute step heights, typically around 30 nm as illustrated in Figure 57. From this instrument and others having enhanced design and performance refinements came the current genre of nanotechnology equipment, typified by the “Nanostep” shown in Figures 58 and 59 (pioneered by the NPL). Instruments of this level of accuracy and precision must be virtually free from thermal expansion effects; therefore the majority of the components are manufactured from thermally stable materials, typically glass ceramic “Zerodur™”.

The kinematic (translational component members) and operational features of the instrument are such that the workpiece is supported on a levelling

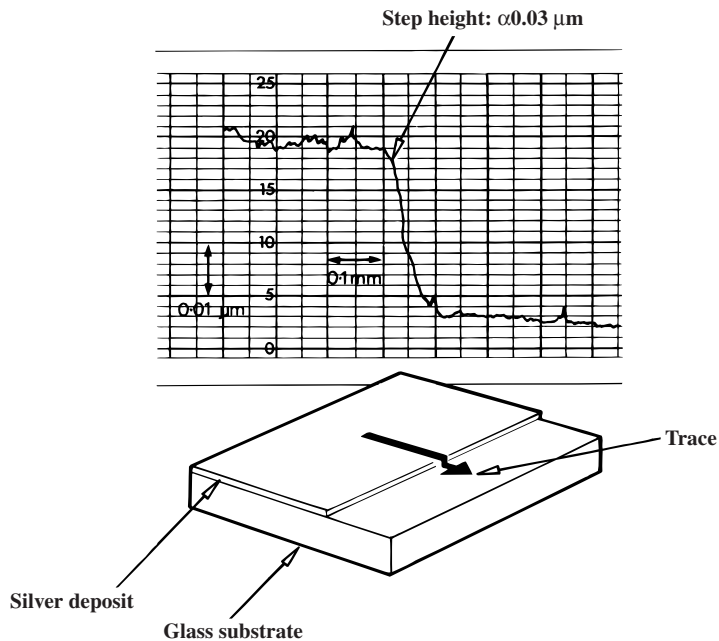


Figure 57. A typical “Talystep” graph of silver deposit on a glass substrate. Thickness of the deposit is approximately 0.03 μm. (Courtesy of Taylor Hobson.)

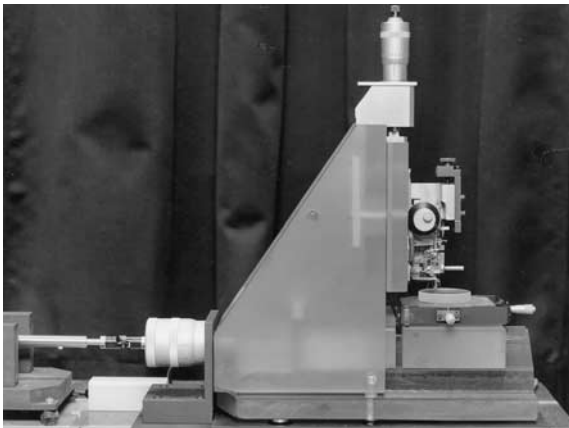


Figure 58. Constructional detail of the “Nanostep” and its associated precision slideway in low coefficient of expansion (i.e., glass ceramic “Zerodur”). (Courtesy of Taylor Hobson.)

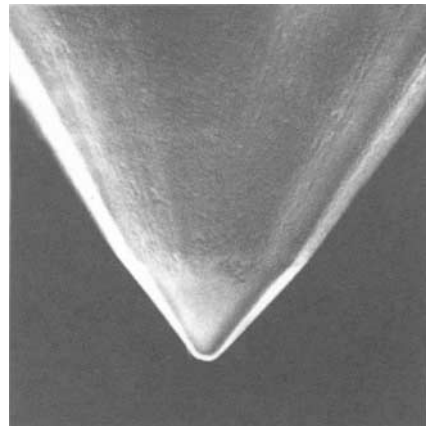


Figure 59. Detail of the interchangeable “Nanostep” stylus. (Courtesy of Taylor Hobson.)

table, which in turn is mounted on a carriage. This carriage is equipped with kinematically positioned dry polymeric pads, forming an interface with a highly polished precision slideway, giving a linear translational motion of 50 mm. The horizontal micrometer (shown in Figure 58) displaces (via a “slave carriage”) the slideway along its desired length of travel, providing minimal disruption to the measurement process. Anti-vibration mounts isolate the DC motor/gearbox that drives the micrometer from the main structural elements of the instru-

ment, producing ultra-low instrument noise down to 0.03 nm under optimum conditions. The motor can drive the micrometer at measurement speeds varying from 0.005 to 0.5 mm/s. The stylus/transducer assembly illustrated in Figure 60 can be vertically positioned on the workpiece and in contact with it via the vertical micrometer (depicted in Figure 58), giving a vertical step height range up to 20 μm. Finally, measurement control and signal output are provided by a suitably configured PC.

The software offers various levels of surface

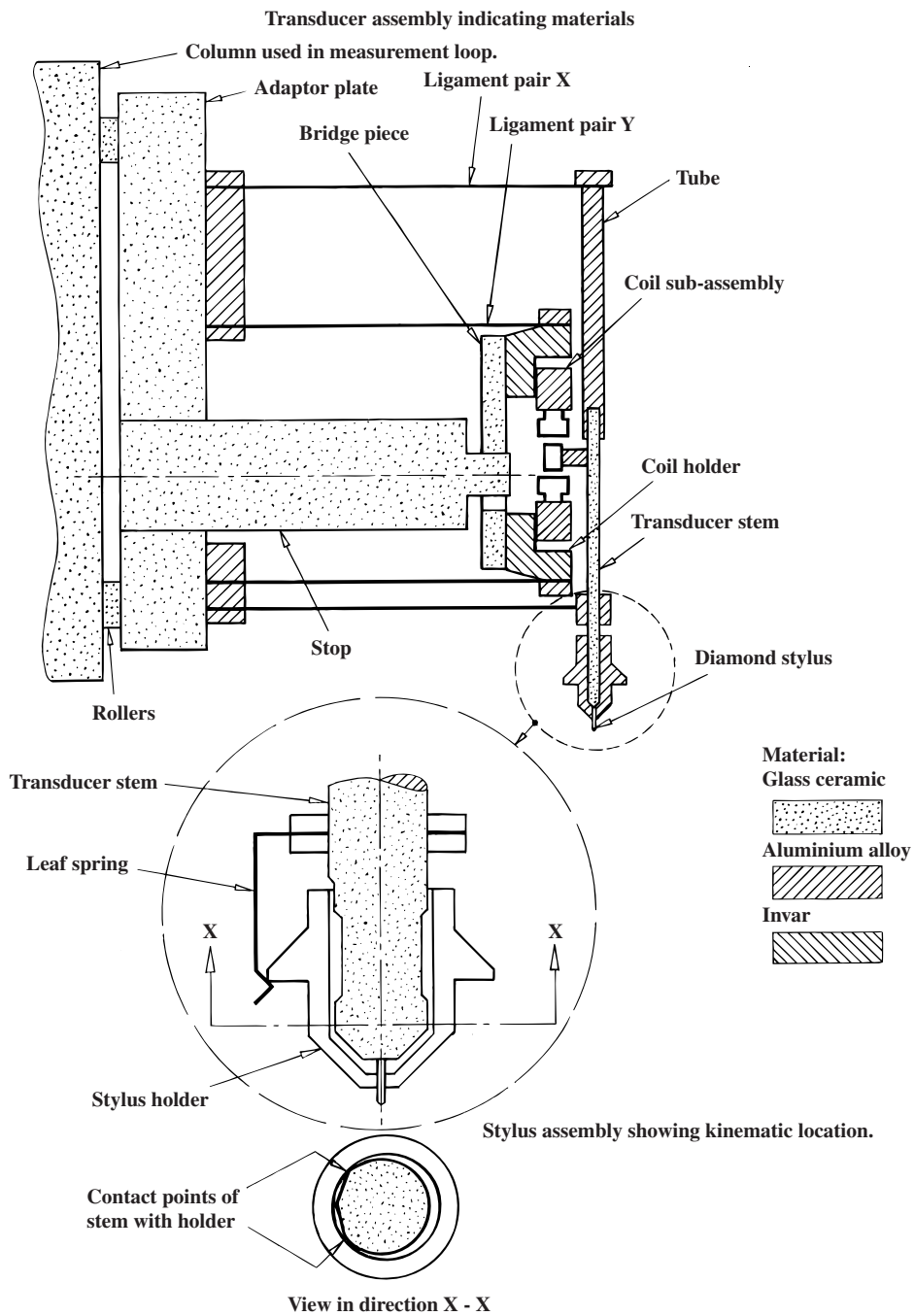


Figure 60. Internal construction of the "Nanostep" instrument. (Courtesy of Taylor Hobson.)

texture analysis, assessed relative to the best-fit reference line, covering a diverse range of international parameters. Additionally, the form software provides for

- *form removal* – best-fit form is calculated (least-squares straight line, minimum-zone straight line

or least-squares arc) and then removed, enabling the texture of the surface to be analysed;

- *angle* – providing a facility to remove any compound surface tilt prior to parameter analysis;
- *dimension* – provides for the linear relationship of surface features, which can be assessed and compared.

Software has also been developed for this instrument enabling it to assess an *enhanced dual profile*. This provision enables one to measure the profile of a “master component”, which can then be saved as a template; any subsequent components can then be both measured and simultaneously displayed with this template for immediate comparison. A statistical process control software package can be employed, along with other specialised software packages.

A number of interchangeable styli (see Figure 59) can be fitted to the instrument to suit a range of different measuring applications, ranging approximately from 1–10 μm conical shape to $0.1 \times 0.5 \mu\text{m}$ pyramidal geometry, the latter being used for high-resolution work. The stylus force can be varied between 10 and 700 μN , allowing it to non-destructively measure delicate or easily scratched components. Once the instrument has been set to the required depth of measurement, an automatic facility provides for the raising and lowering of the stylus when repositioning or changing samples. For the measurement of ultra-fine surfaces, the instrument offers a nominal gauge resolution down to 31 pm (with the 2 μm gauge range). Typical performance of the 10 μm conical stylus with an applied force of 40 μN can produce a surface roughness R_q value of 4.9 nm. Such nano-instrumentation has a component capacity of diameter 100 mm, with a depth of 22 mm, covering a large range of precision components.

Instrument applications are both wide and varied and it might typically be employed for

- ultra-precision bearing and surface defect measurement – surface finish/integrity;
- silicon wafer measurement;
- integrated circuit production – process monitoring;
- magnetic tape measurement;
- diamond-turned components – small diameter;
- calibration standards;
- precision optics;
- mirrors for laser gyroscopes;
- laser and X-ray mirrors.

In this chapter little or no reference has been made to the role of expert systems or neural networks in surface texture analysis. This omission was intentional, as the chapter length would otherwise have simply been too long and less specific in nature. In Chapter 2 a review of three-dimensional surface texture measurements will be made, then later on contour/fractal effects will be discussed, which are now becoming an important technology in describing surfaces.

References

Journal and conference papers

- Bell, T. Surface engineering: past, present and future. *Surface Engineering* 6(1), 1990, 31–40.
- Bjuggren, M., Krummenacher, L. and Mattsson, L. Noncontact surface roughness measurement of engineering surfaces by total integrated infrared scattering. *Precision Engineering* 20(1), 1997, 33–45.
- Bousefield, B. and Bousefield, T. Progress towards a metallography standard. *Metals and Materials* March 1990, 146–148.
- Brown A.J.C. and Gilbert, F. Industrial applications for optical profilometry. *Sensor Review* January 1990, 35–37.
- Caber, P.J. An interferometric profiler for rough surfaces. *Applied Optics* 32, 1993, 3438–3441.
- Downs M.J., Mason, N.M. and Nelson, J.C.C. Measurement of the profiles of super-smooth surfaces using optical interferometry. *Proceedings of SPIE-1009*, 1989, 14–17.
- Drews, W.E. Surface measurement: an advanced technology. *Quality Progress* April 1987, 43–46.
- Dyson, J. *Interferometry as a Measuring Tool*. Machinery Pub. Co., 1970.
- Garratt, J. and Bottomly, S.C. Technology transfer in the development of a nanotopographic instrument. *Nanotechnology* 1, 1990, 38–43.
- Garratt J. and Mills, M. Measurement of the roughness of super-smooth surfaces using a stylus instrument. *Nanotechnology* 7, 1996, 13–20.
- Gee, M.G. and McCormick, N.J. The application of confocal scanning microscopy to the examination of ceramic wear surfaces. *Journal of Physics D: Applied Physics* 25, 1990, 230–235.
- Harvie, A. and Beattie, J.S. Surface texture measurement. *Production Engineer*, 1978, 25–29.
- Hongai, Z., Zhixiang, C. and Riyao, C. The control of roughness with expert system controller. *Proceedings of the First International Mach. Mon. and Diagnostics*, Las Vegas, 1989, 813–817.
- Kuwamura, S. and Yamaguchi, I. Wavelength scanning profilometry for real-time surface shape measurement. *Applied Optics* 36, 1997, 4473–4482.
- Leach, R.K. Measurement of a correction for the phase change on reflection due to surface roughness. *Proceedings of SPIE* 3477, 1998, 138–151.
- Leach, R.K. Traceable measurement of surface texture at the national physical laboratory using Nanosurf IV. *Measurement Science and Technology* 11, 2000, 1162–1172.
- Leach, R.K. Traceable measurement of surface texture in the optics industry. *Large Lenses and Prisms Conference*, University College London, 27–30 March 2001.
- Leach, R.K. and Hart, A.. Investigation into the shape of diamond styli used for surface texture measurement. *NPL Report CBTL10*, April 2001, 1–12.
- Mansfield, D. Surface characterisation via optical diffraction. *SPIE (International Society for Optical Engineering) Vol. 1573: Commercial Applications of Precision Manufacture at the Sub-Micron Level*, 1991, 163–169.
- Prostrednik, D. and Osanna, P.H. The Abbott curve: well known in metrology but not on technical drawings. *International Journal of Machine Tools Manufacture* 38(5–6), 1998, 741–745.
- Schaffer, G.H. The many faces of surface texture. *American Machinist*, June 1988, 61–68.
- Schneider, U., Steckroth, A. and Hubner, G. An approach to the evaluation of surface profiles by separating them into functionally different parts. *Surface Topography* 1, 1988, 71–83.
- Stout, K. How smooth is smooth? Surface measurements and their relevance in manufacturing. *Production Engineer* May 1980, 17–22.

- Thomas, T.R. Trends in surface roughness. *International Journal of Machine Tools Manufacture* 38(5-6), 1998, 405-411.
- Trumpold, H. and Heldt, E. Why filtering surface profiles. *International Journal of Machine Tools Manufacture* 38 (5-6), 1998, 639-646.
- Ulbricht, R. Die Bestimmung der mittleren räumlichen Lichtintensität durch nur eine. *Messung Electrotechnische Zeitschrift* 29, 1900, 595-601.
- Vorburger, T.V. and Teague, E.C. Optical techniques for online measurement of surface topography. *Precision Engineering* 3, 1981, 63-83.
- Westburg, J. Opportunities and problems when standardising and implementing surface structure parameters in industry. *International Journal of Machine Tools and Manufacture* 38(5-6), 1998, 413-416.
- Whitehouse, D.J. Some ultimate limits on the measurement of surfaces using stylus techniques. *Measurement and Control* 8, 1975, 147-151.
- Whitehouse, D.J. Beta functions for surface topology? *Annals of the CIRP* 27, 1978, 491-497.
- Whitehouse, D.J. Surfaces: a link between manufacture and function. *Proceedings of IMechE* 192, 1978, 179-187.
- Whitehouse, D.J. Conditioning of the manufacturing process using surface finish. *Proceedings of the Third Lamdamap Conference, Computational Mechanics* July 1997, 3-20.
- Whitehouse, D.J. Some theoretical aspects of surface peak parameters. *Precision Engineering* 23, 1999, 94-102.
- Whitehouse, D.J. Surface measurement fidelity. *Proceedings of the Fourth Lamdamap Conference*, University of Northumbria, WIT Press, July 1999, 267-276.
- Whitehouse, D.J. Characterizing the machined surface condition by appropriate parameters. *Proceedings of the Third Industrial Tooling Conference*, Southampton Institute, Molyneux Press, September 1999, 8-31.
- Wyant, J.C. Computerized interferometric measurement of surface microstructure. *Proceedings of SPIE-2576*, 1995, 122-130.
- Zahwi, S. and Mekawi, A.M. Some effects of stylus force on scratching surfaces. *Eighth International Conference on Metrology and Properties of Engineering Surfaces*, University of Huddersfield, 26-28 April 2000.

Books, booklets and guides

- Bennett, J.M. and Mattsson, L. *Introduction to Surface Roughness and Scattering*. Washington, DC: Optical Society of America, 1989.
- Busch, T. and Wilkie Brothers Foundation. *Fundamentals of Dimensional Metrology*. Delmar, 1989.
- Dagnall, M.A. *Exploring Surface Texture*. Taylor Hobson Precision, 1997.
- Dyson, J. *Interferometry as a Measuring Tool*. Machinery Pub. Co., 1970.
- Gayler, J.F.W. and Shotbolt, C.R. *Metrology for Engineers*. Cassell, 1990.
- Haycocks, J.A. *Novel Probes for Surface Texture Metrology*. NPL Report MOM105, July 1991.
- Hume, K.J. *Engineering Metrology*. Macdonald, 1970.
- Hume, K.J. *A History of Engineering Metrology*. Mechanical Engineering Pub., 1980.
- Leach, R.K. NPL Good Practice Guide No. 37, Measurement of Surface Texture using Stylus Instruments. NPL, 2001
- Leach, R.K. and Hart, A. Investigation into the Shape of Diamond Styli used for Surface Texture Assessment. NPL Report CBTLM 10, April 2001.
- Mainsah, E., Greenwood, J.A. and Chetwynd, D.G. *Metrology and Properties of Engineering Surfaces*. Kluwer Academic Pub., 2001.
- Mummery, L. *Surface Texture Analysis: The Handbook*. Hommelwerke GmbH, 1990.
- Nicolls, M.O. *The Measurement of Surface Finish*. DeBeers Technical Service Centre, DeBeers Industrial Diamond Division, Charters, UK, 1980.
- Reason, R.E. *The Measurement of Surface Texture*. Cleaver-Hume Press, 1960.
- Sander, M. *A Practical Guide to the Assessment of Surface Texture*. Feinpruf GmbH, 1991.
- Stover, J.C. *Optical Scattering*. McGraw-Hill, 1990.
- Thomas, G.G. *Engineering Metrology*. Butterworths, 1974.
- Thomas, T.R. *Rough Surfaces*. Imperial College Press, 1999.
- Whitehouse, D.J. and Reason, R.E. *The Equation of the Mean Line of Surface Texture Found by an Electric Wave Filter*. Rank Taylor Hobson Pub., 1965.
- Whitehouse, D.J. *Handbook of Surface Metrology*. Institute of Physics, Bristol and Philadelphia, 1994.
- Whitehouse, D.J. *Surfaces and their Measurement*. Hermes Penton Science, 1991.
- Williams, D.C. *Optical Methods in Engineering Metrology*. Chapman & Hall, 1993.