The measurement and analysis of surfaces

D. J. WHITEHOUSE

In recent years there has been a growing awareness of the importance of surface metrology. Two pressures have been dominant. First there is the need to better control the manufacturing process and second there has been the need to better understand the role which the geometry of a part plays during its working life. These pressures have lead to a minor revolution in the way in which surfaces are measured and assessed. In common with the article by Archard* this paper will be restricted to aspects of the development of the most used technique, namely the stylus method. It will concentrate on modern developments.

The general trends have resulted from the following requirements:

- a the need to reduce the measurement cycle time
- b the need to closely monitor the geometry during the part function
- c the facility to be able to measure many different parameters of the geometry
- d the desirability of building up a total picture of the surface and not simply isolated profiles, ie, making an integrated measurement
- e more accurate measurements
- f less skilled operator intervention during the measurement.

Some of these requirements are a direct result of pressure from production engineering and some from tribological research but taken as a whole they benefit both. Three major developments have resulted because of these requirements:

1. an increase in the uses of digital techniques
2. the development of surface metrology instruments having more than one accurate degree of movement
3. the use of relocation profilometry.

The first two are mainly of significance in production engineering but all three are important tribologically. In the next section the instrumental aspect will be considered in general terms. This will be followed by a section on the techniques used to analyze surfaces. No attempt will be made to judge the relative importance of the features discussed because these are so obviously dependent on the particular function.

Rank Taylor Hobson, East Park Road, Leicester LE5 3RG, England

* 'Surface topography and tribology', *TRIBOLOGY international*, 7, No 5 (Oct 1974) 213–221

INSTRUMENTATION

Use of computers in the instrument

Fig 1 shows a schematic view of a typical measurement cycle where for instance a surface is being measured for roughness. Obviously the best way to improve the measurement cycle is to reduce the number of operator intervention paths and to minimize the number of iterations around any one loop. Both of these aims can be substantially achieved by having a highly repeatable transducer computer control of the instrument and digital analysis of the signals coming from the transducer. With these facilities it is possible to examine a surface, quickly, accurately and over an area with little chance of error. An example of computer control is shown in Fig 2. This shows an automatic set-up for analyzing surfaces[1]. In the case shown a surface standard is being calibrated over its entire area[2]. The computer interogates the instrument for magnification, meter cut-off etc. It completely controls the operation once the number of traverses required has been fed in via the teletype. Power spectral analysis is possible as well as the usual surface parameters (which will be discussed later). Time saving of orders of magnitude can be achieved. Tasks of this kind are well suited to the architecture of modern mini-computers. A core store size of 8K words might well be sufficient; each word being 16 bit.

Not only does such a system reduce operator intervention and hence reliability it can also increase the accuracy of the measurements themselves not necessarily because of good analytical technique but simply because any systematic errors in the instrument itself can be removed automatically. The same is true for noise if the bandwidth is known. Improvements in accuracy of almost an order of magnitude have been achieved[3].

To illustrate the transducer's importance consider Fig 3. A badly eccentric part is to be measured on a roundness measuring instrument. Fig 3a shows the part just within the

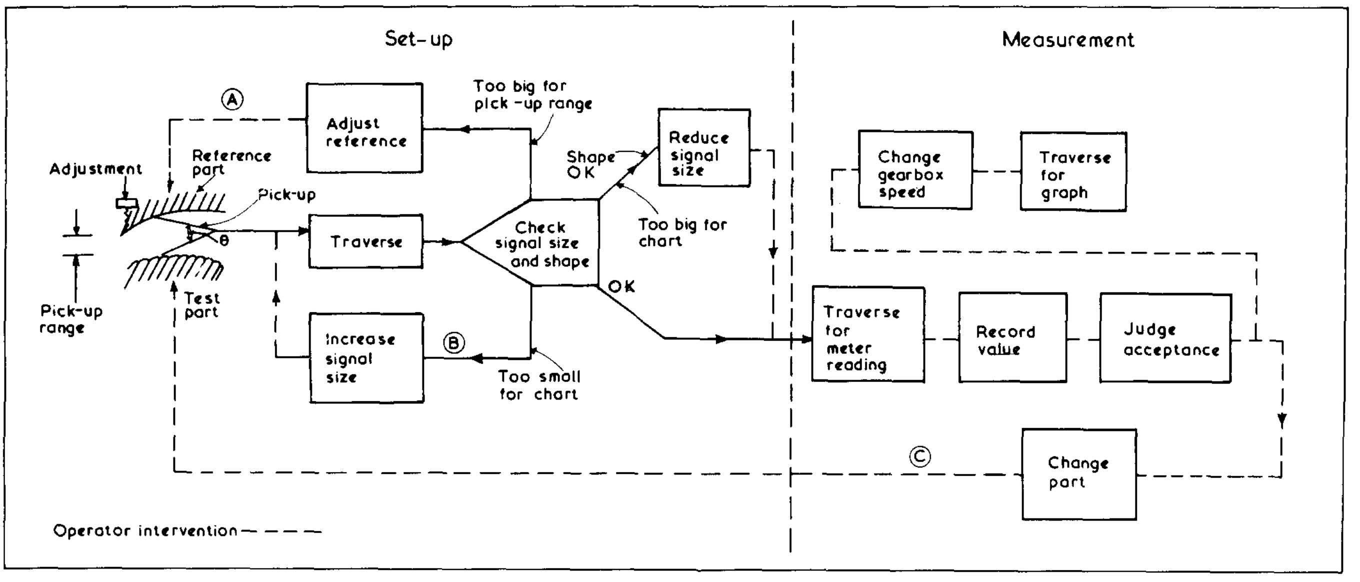

Fig 1 Measurement cycle of typical instrument. Advantages of computation: (1) loops (A), (B) and (C) can be automatic (2) judgement of acceptance can be made objective, (3) the shape of the reference part is not critical (but needs to be known and repeatable)

range of the transducer. In order to contain the severe eccentricity the magnification is low and no out-of-roundness can be seen. This single tracing is all that needs to be done providing the resolution of the transducer is sufficiently small to detect the out of roundness signal in the presence of the signal due to setting-up error. Fig 3b shows the part centred by the computer and Fig 3c shows the part with all its detail magnified by the computer. This is only possible because the transducer range is large and the resolution small. In other words the ratio of these – the dynamic range of the transducer, should be high. Preferably the frequency response should also be high. A typical transducer dynamic range suitable to achieve this sort of result is at least 10 bits ie 1000:1. The laborious job of successively refining the centring of the part relative to the axis of rotation has been removed by this combination of computer and transducer.

It may appear that the computer as an aid to measurement has no disadvantages. This is not necessarily so. In order to make schemes like these viable economically special effort has to be made to keep the store at a minimum. Often

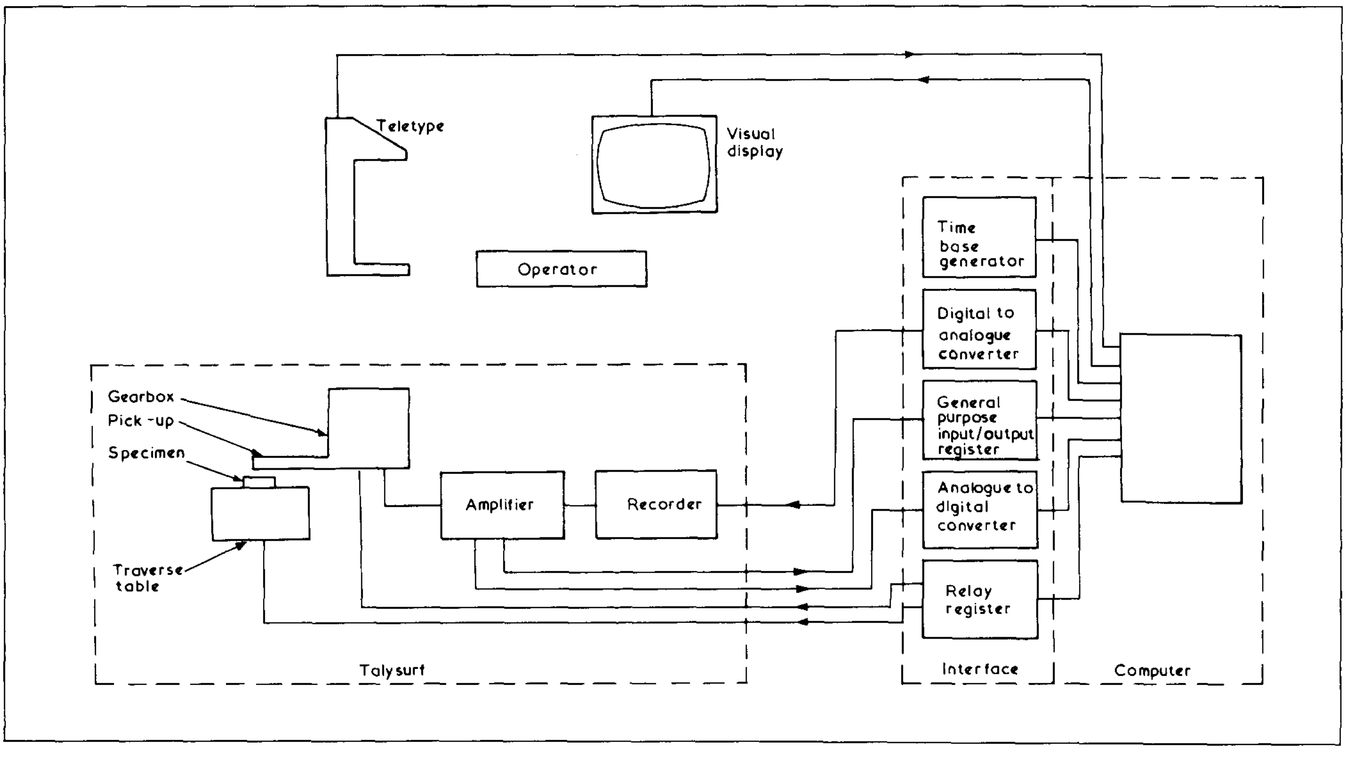

Fig 2 Schematic diagram of a system for the automatic calibration of surface finish reference standards

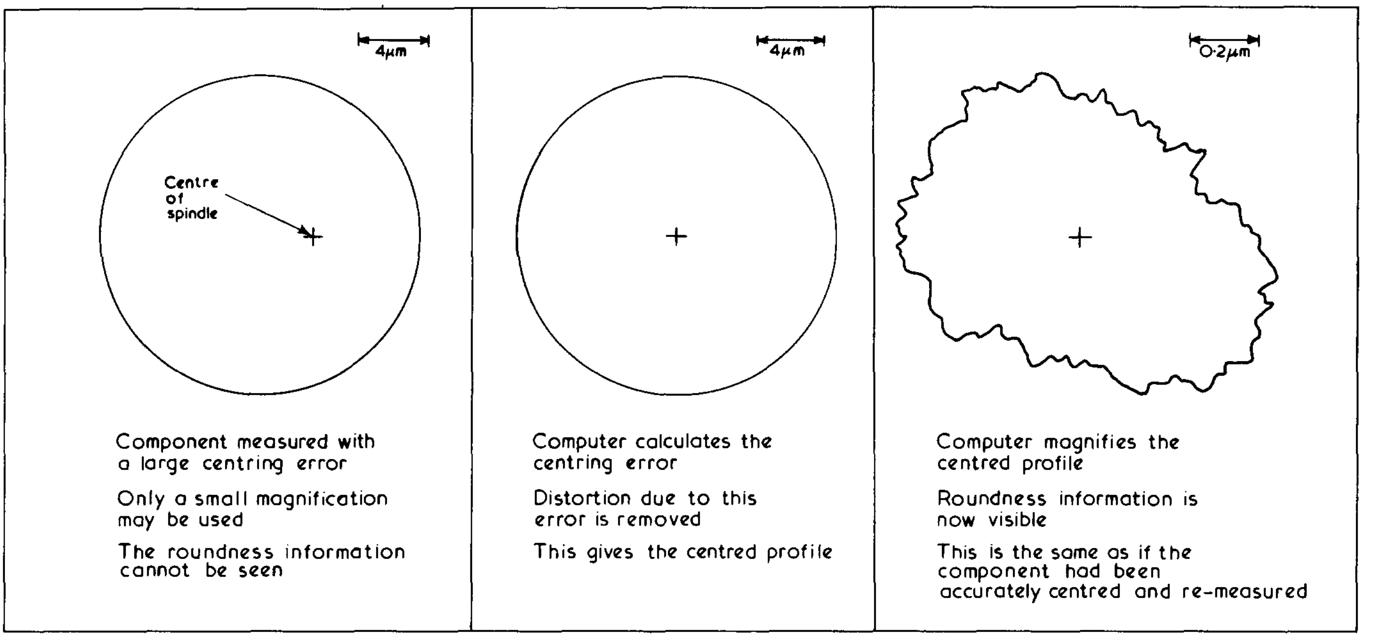

Fig 3 Automatic compensation for eccentricity in the measurement of roundness

much care has to be exercised with the programming to ensure fast operation for a minimum store.

Integrated instruments

Current practise is to select an instrument to measure a particular feature of a surface, such as the roughness, waviness, straightness etc. There is evidence that in order to fully characterize a part, the total geometry should be taken into account in order to give an adequate functional description. As an example, if it is required to determine whether a shaft will fit into a bore then the simple measurement of roundness may not be a sufficient control. For instance roundness graphs taken along the shaft may be satisfactory but this does not guarantee that the generator of the shaft is straight. There has to be a way to determine whether the centres of the roundness graphs are colinear. This can be assumed sometimes if the manufacturing process is known to produce straight generators, but not always. In the latter case recourse to more integrated measuring instruments is necessary.

From an engineering point of view the most useful type of surface metrology instrument would probably be one having an accurate axis of rotation, and accurate cartesian and radial movements; it would measure according to a cylindrical frame of reference (Fig 4). However, to get the best results it is advisable to match the co-ordinate system of the measuring instrument to that of the component. In this way it is possible to reduce the number of transducer movements having to have a large dynamic range by one. The reduction is made possible because all the remaining degree of freedom has to cater for in terms of range is the expected maximum departure from the perfect shape in addition to setting up errors. These are likely to be very much smaller than the extent of the other movements.

As more of these instruments are developed it is apparent that the degree of computing must increase if only to unravel setting up errors in the part. Also it is likely that computers will be needed to correlate different aspects of

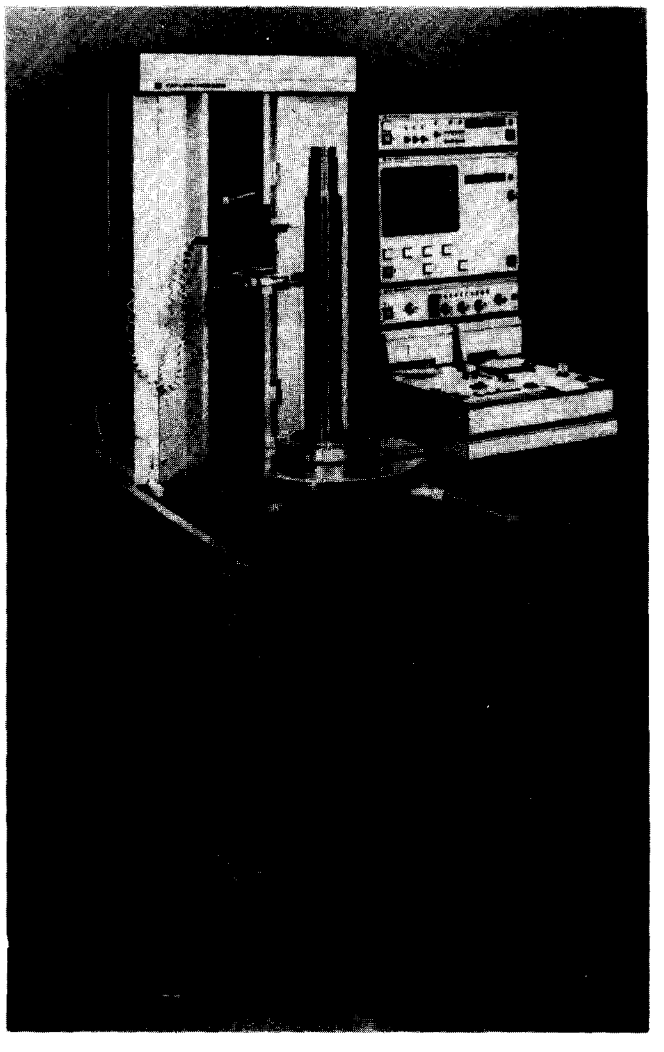

Fig 4 Talycenta: (a) a small computer for evaluating roundness parameters and reference lines, (b) accurate spindle, (c) accurate straightness unit

the measurement; the more degrees of movement that are required to measure a part the more difficult it is instrumentally to ensure independence of movements. Interactions would have to be removed computationally. For instance in sphere measurement it is very difficult to get a double gimbal movement about a common centre, certainly to within the order of 0.5 μm which is often the sort of error in the part.

How best to configurate a set of measurements is not in most cases a function of one part, it is more likely to be governed by the inter-relationship between parts. Each alternative has to be judged on its own merit from a cost and functional point of view.

Relocation techniques

One of the most important trends in recent years has been the growing search to find surface parameters which can predict how well a part will behave functionally. This has been a very difficult proposition until only a few years ago when computers came into vogue metrologically. Before this time it was practically impossible to measure a sufficient number of different parameters of the surface before the start of the experiment to have much chance of stumbling upon the optimum. Now, however, with the intrinsic versatility of the computer it is possible to measure a large number of parameters before and after the experiment and so look for correlation between performance and any of the measured parameters.

One more ingredient is necessary to make the technique really useful and this is the use of relocation methods. Essentially, this means that great care is taken during and after the experiment to measure exactly the same profile or profiles on the surface. To do this requires considerable instrumental skill usually involving kinematic design both in the measuring instrument and in the machine in which the experiment is being carried out. Ideally these would be incorporated into one piece of apparatus but this is not always possible. Providing that the profiles have been selected to be in a strategic position functionally the relocation method can save an enormous amount of work. Individual detail rather than average behaviour can be studied.

A good example of the results of this technique is shown in Fig 5. This represents successive traces taken along the same profile of a medium carbon steel during a wear experiment in a crossed cylinder machine. Relocation of the cylinder to a few microinches was necessary in both the crossed cylinders machine and the Talysurf 4 measuring instrument.

Notice how the high frequencies have been smoothed out. In accordance with the previous section the Talysurf 4 was coupled on-line to a mini-computer and a large number of parameters monitored during the whole experiment. Not only can the geometry be monitored in this way but also frictional forces. Fig 6 shows how some of the well known parameters change during such an experiment. It appears from this particular experiment that the peak curvature changes most during the running-in and consequently can be regarded as a good measure of the wear process. On the other hand the R_a value (previously known as the cla value) changes one of the least. It could be regarded as the parameter which is best suited to predict the run-in profile from the original profile trace. This probably explains why it has been such a useful parameter in the past in industry. For research in contact, however, it would not be so useful.

Similar experiments have been carried out on bearings with grease lubricant[6] and there is every reason to suppose that the tendency to use computers and relocation methods will increase. Because of this inevitable increase in digital methods emphasis will be placed on digital methods where ever relevant.

ANALYSIS

General

The increasing use of computing has been inevitable mainly because of economic reasons. Not the least has been the fact that many of the functions previously carried out by conventional mechanics may soon be taken over by the computer, for example the filtering, magnification switching, etc. This is inevitable because of the soaring cost of mechanical engineering relative to the ever reducing cost of computing. However, the mere fact that more of the overall signal from the entire surface will have to be processed imposes formidable problems from the programing point of view. Sophisticated methods are needed to unravel misalignment, eccentricities etc from the geometry itself. In terms of range of size, diversity of signal shape etc, the output from other kinds of transducer eg an accelerometer is straight forward. It is for this reason that a great deal of trouble in interpreting surfaces has arisen in the past and one of the reasons why more and more effort is being expended on the analysis today. The extraction of the *surface geometry* from the *whole signal* contaminated with set-up errors is called *preprocessing.*

Preprocessing of surface profiles

The reference movement of the instrument ensured that the signal representing the part is kept within the range of the

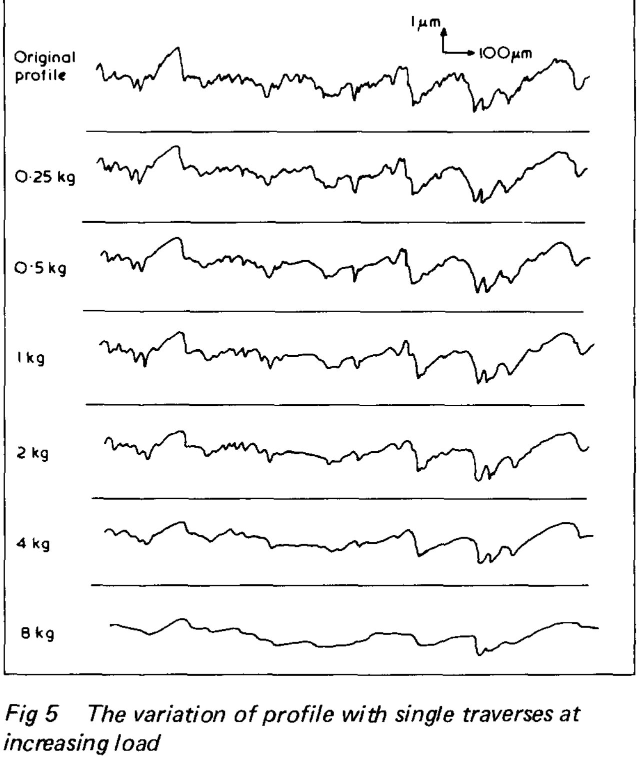

Fig 5 The variation of profile with single traverses at increasing load

transducer during the measuring traverse. This reference movement may be a flat datum, a reference axis of rotation or any similar device. Having obtained this signal, the roughness (for example) has to be extracted from it. This involves positioning in some way a reference line within the profile from which the roughness can be measured, to the exclusion of all else. Usually in the case of roughness measurement the tilt of the part, error of form and waviness would have to be removed before a meaningful assessment of the roughness could be attempted. Some of the reference lines used are[7]: polynomials, mean lines, filter lines and envelopes.

1 Polynomials

When it is required to remove tilt from the profile a best-fit straight line can be used. If N equally spaced ordinates are taken over the profile length, of height y, then the slope of the line is given by b_1 where b_1 is:

$$b_1 = \frac{12 \sum_{i=1}^{N} i y_i - 6(N+1) \sum_{i=1}^{N} y_i}{N(N^2-1)}, \quad b_0 = \frac{1}{N} \sum_{i=1}^{N} y_i - \left(\frac{N+1}{2}\right) b_1$$

b_0 is the intercept.

For a plane whose height ordinates are z and the co-ordinates are taken about the centre of the plane the two tilts would be a_1, b_1 given by:

$$a_1 = \frac{12 \sum_{i=-N/2}^{N/2} \left(i \sum_{j=-M/2}^{j=M/2} z_{ij} \right)}{MN(N+1)(N+2)}, \quad b_1 = \frac{12 \sum_{j=-M/2}^{j=M/2} \left(j \sum_{i=-N/2}^{N/2} z_{ij} \right)}{MN(M+1)(M+2)}$$

Other best fit polynomials can be worked out. These types of reference lines suffer from two disadvantages. First is that in order not to distort the signal the order of polynomial required must be known. For example it is not possible to remove a bow shaped error of form if a linear best-fit line is used. The second disadvantage which is true mainly for best fit least square polynomials is that they are evaluated from the difference between two large numbers which is difficult to deal with numerically unless the mean value is known.

2 Centre lines

This method comprises fitting a succession of straight lines through samples of the profile; each line within the sample is given the general direction of the profile and is positioned such that the area of profile above it is equal to the area below it. This method is the basis of the 'M' system of surface measurement described in British Standard 1134 and used extensively in the world. Least square lines have been used within each sample length instead of these 'centre lines' but they suffer from being dependent on the starting point of the sample. The sample length itself is described as just that length of surface over which a meaningful reading can be made. A typical value often used is 0.8 mm (0.03 in). Usually about five consecutive sample lengths are traversed by commercial instruments to get a good average and for safety a few traverses should be made in different parts of the surface. This aspect will be considered later. By arranging that only a limited length of surface is analyzed (a sample length) signals of longer wavelength are effectively excluded.

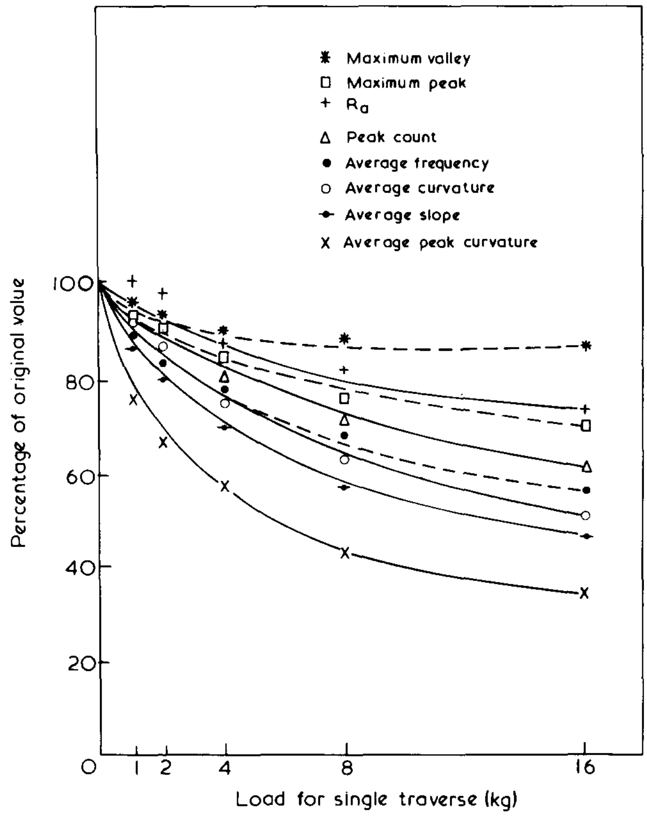

Fig 6 Variation of parameters with wear

3 Filter methods

Fundamentally the centre-line described above is a graphical method. Instrumentally a similar effect is achieved by passing the profile signal through a high pass 2-CR filter whose cut-off is nominally equal to the sample length.

One disadvantage of electrical filtering is that a run-up time is needed to allow the filter to settle down which means that all the available surface profile cannot be used. However in practise this is usually more than compensated for by the fact that filters do not produce discontinuous reference lines like the least square and centre line methods, neither do they require a detailed knowledge of the form of the extraneous signal. Filtering is now the most common pre-processing technique and most instruments use a 'standard filter' which has 75% transmission at the cut-off.

One of the recent benefits of using digital methods has been the increase in accuracy using digital filters rather than analogue. This has meant that standardization and calibration procedures are much better than they were previously. One reason for this improvement is that a digital filter can be defined easily and precisely by the equation of its

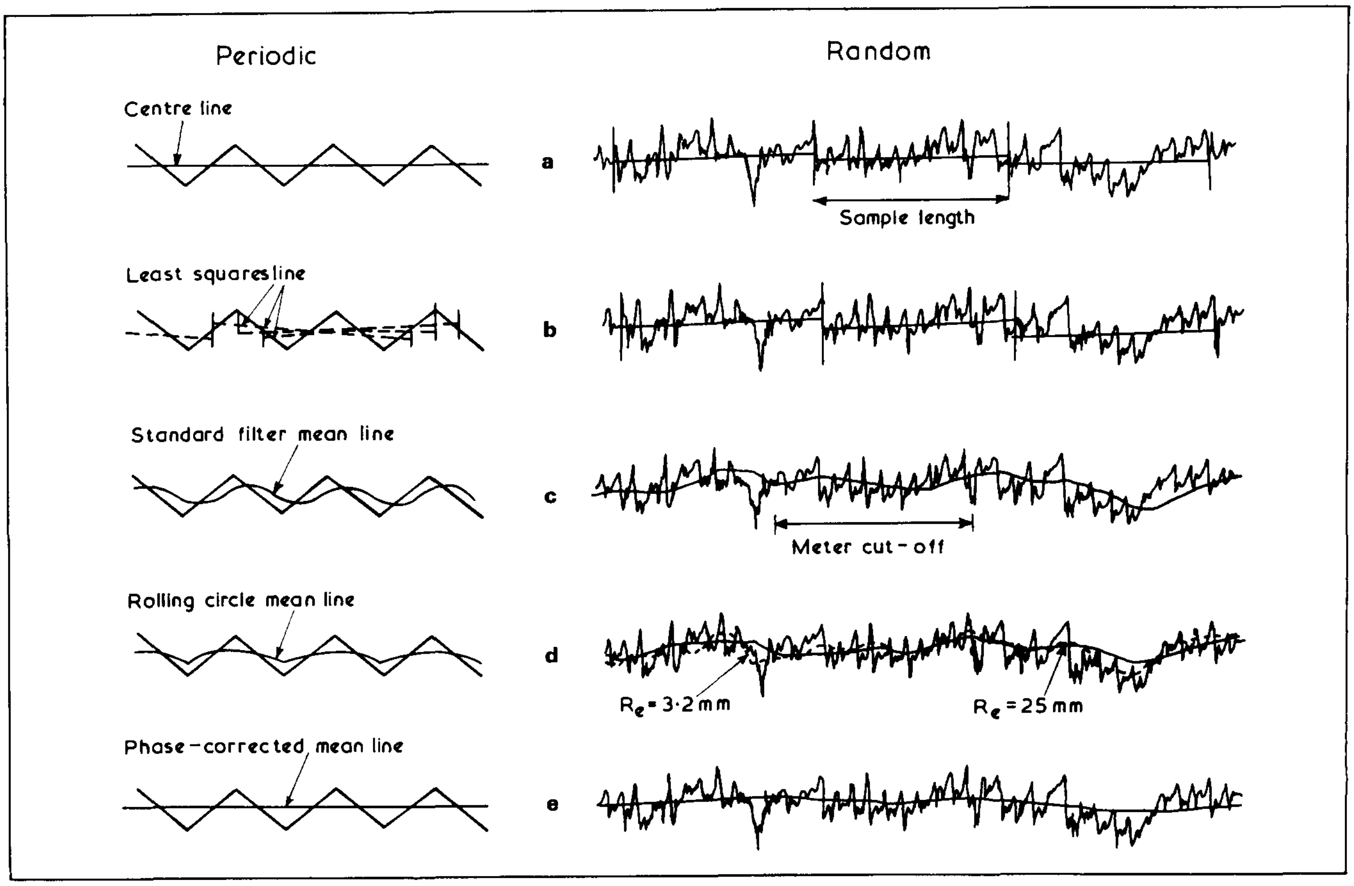

Fig 7 Behaviour of different types of mean line

impulse response within the computer[8]. It does not rely on the accuracy of components. Another reason is that for special circumstances optimum filters can be simulated very easily, in fact as the standard filter. This is not true for the analogue methods.

As an example of how a digital filter works in operation if one ordinate delay of the profile is z^{-1} then the output at time t from the filter, $g(t)$ is given in z transform notation by:

$$g(t) = y(t)\left[z^{-1} - \frac{(1 - e^{-2aT})z^{-2}}{1 - e^{-2aT}z^{-2}}\right]^2$$

where T is the real time delay of z^{-1}. This equation merely says that the output at t is determined by the profile ordinates entering up to 6 increments in time earlier and the reference line ordinates evaluated for up to the last four increments in time. This type of equation simulates a recursive filter because past values of the reference line are used to evaluate the current one. Notice how economic it is in store size and also speed of operation. It performs exactly as the analogue standard filter but is much more accurate. A typical value of $2aT$ is 0.982 which represents 400 measurements (ordinates) per sample length (cut-off).

Optimized filters can also be used which do not introduce phase shift as does the standard filter; a feature which can sometimes be critical if peaks are being measured. One such filter is called the phase-corrected filter[9] devised especially to measure roughness in the presence of waviness or alternatively to measure waviness direct.

4 Envelope methods

Other methods using peaks only to determine a reference line have been advocated as in the 'E' system devised by von Weingraber in 1957[10,11]. In this a ball is imagined to be rolled across the surface. The locus of the bottom of the ball is an envelope from which a mean line can be found. This method has never been adopted because of the difficulty of building a practical instrument.

A comparison of these different lines is shown in Fig 7 for a periodic and random profile. There are considerable differences to be seen. Remember that in any judgement of roughness it is the deviations of the profile from the reference line which is important.

Preprocessing of roundness profiles

Similar problems exist in the measurement of roundness sphericity, cylindricity etc, again with a bewildering variety of alternative methods of preprocessing. First it is essential to understand the form of the signal emerging from a roundness instrument. Consider Fig 8. The equation of a displaced circle of radius R and eccentricity e from the centre of measurement o is given by:

$$k(\theta) = e\mathrm{Cos}(\theta - \Phi) + R - \frac{e^2}{2R}\mathrm{Sin}^2(\theta - \Phi) + \ldots$$

whereas the output from the transducer which is output onto the chart is:

$$\rho(\theta) = M(R - L) + S + Me\mathrm{Cos}(\theta - \Phi)$$

where L is the amount by which the transducer is displaced to get the part within its range. S is the inner chart radius and M the magnification.

Taking the relative values into account the chart equation becomes that of a limacon (not a true circle) of form:

$$\rho(\theta) - S = R + x\mathrm{Cos}\theta + y\mathrm{Sin}\theta$$

where $x = Me\mathrm{Cos}\Phi \quad y = Me\mathrm{Sin}\Phi$.

For a sphere this is:

$$\rho(\theta) - S = R + x\mathrm{Cos}\theta\mathrm{Cos}a + y\mathrm{Sin}\theta\mathrm{Cos}a + z\mathrm{Sin}a$$

θ is the longitude angle, a is the latitude angle.

Consequently in order to remove severe eccentricity from the signal it is a limacon which has to be removed rather than a circle – simply because of the nature of the measuring device.

Since the advent of computers not only has this been possible (even with simple analogue computers) but the preprocessing with digital methods can take into account only partial parts or parts having discontinuities in them such as keyways[12]. Furthermore methods of assessing roundness based upon fitting circumscribing circles to the pattern of peaks (ring gauge method) for instance which could only be done on a graph can now be easily and very quickly evaluated in a small computer (Fig 4). Whereas the operator previously had to fit a minimum circle to the chart signal and measure departures from it by hand the computer can take up the iteration easily and give an answer hundreds of times faster.

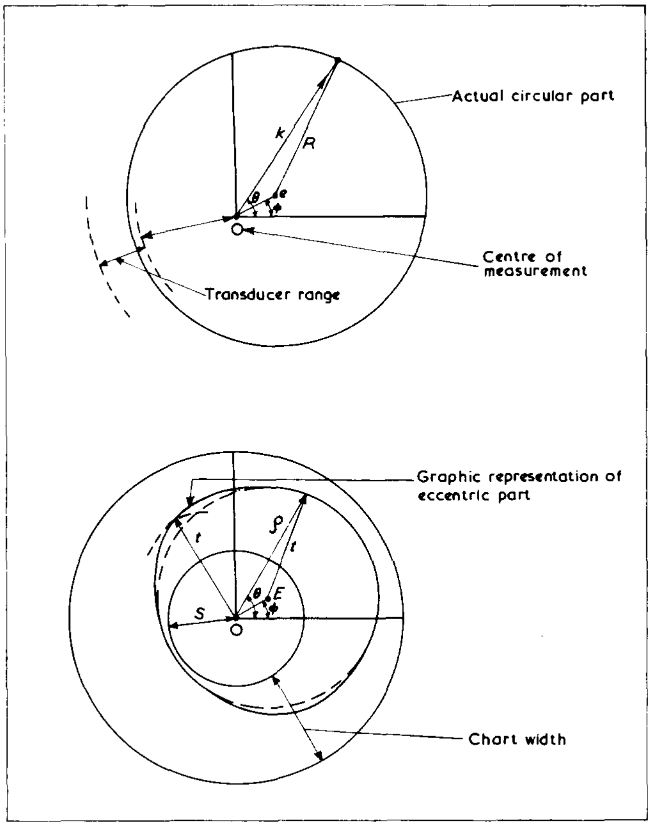

Fig 8 Representation of a circular part as seen by a roundness measuring instrument

The harmonic analysis of roundness graphs has also had a dramatic boost since the Fast Fourier Algorithm (FFT) was brought into common use[4]. Hundreds of harmonics can be worked out in a second or so which has considerably eased one of the measurement problems of the ball bearing industry. The reason why FFT is particularly suited to roundness measurement is because it is by its very nature a circular convolution technique therefore any purely periodic signal lends itself very well to analysis.

Many of the statements made so far have illustrated the rapidly growing use of computers in surface metrology. In the next section as well as indicating some of the parameters that can be measured from the signal after preprocessing, and the amount that needs to be taken, some indications will be given about the problems and pitfalls of digital analysis.

THE MEASUREMENT OF SURFACE PARAMETERS

General statistics

The amount of data which needs to be taken is of considerable importance. In metrology terms this resolves itself into the number of profiles to take and the length of each profile. From a statistical point of view the degree of confidence that can be given to a measured parameter is of foremost importance[13].

Take as a first example the errors in the *standard deviation* of a parameter being measured. Usually only a small number <25 of estimates will have been made over the surface. Under these circumstances the chi-square test can be used to determine the limits to within a certain confidence, say 95% (which is equivalent to unity minus the level of significance a). If s is the measured standard deviation of errors from N readings and σ is the true standard deviation then the following rule is used:

$$\frac{(N-1)s^2}{\chi^2_{N-1,a/2}} \leqslant \sigma^2 < \frac{(N-1)s^2}{\chi^2_{N-1,1-a/2}}$$

For instance for $N = 9$ there is 95% confidence that the true variance σ^2 lies below the value obtained by multiplying the sample variance s^2 by $(N-1)/\chi^2_{N-1,0.975}$ and above the value $s^2 \times (N-1)/\chi_{N-1,0.025}$. From the tables this works out that there is a 95% confidence that the true spread σ is within the range $0.68s < \sigma < 1.92s$.

For $N > 25$ a normal distribution of errors can be assumed from which the standard deviation σ can be determined from:

$$\sigma = s\left(1 \pm \frac{2}{\sqrt{2N}}\right) \text{ to within 95\%}$$

To estimate the error in the *mean* value of the parameter two values apart from N need to be known, the sample mean $\bar{R}_a$ and the sample standard deviation s.

For a small number of readings $N < 25$ the 'student t' distribution needs to be used. Thus suppose that the true mean is $\bar{\bar{R}}_a$ then the general rule for confidence is given by:

$$\bar{R}_a - \frac{s.t_{N-1,a/2}}{\sqrt{N}} \leqslant \bar{\bar{R}}_a < \bar{R}_a + \frac{s.t_{N-1,a/2}}{\sqrt{N}}$$

For $N = 9$ as above $t_{8,0.025} = 2.306$, hence $\bar{\bar{R}}_a = \bar{R}_a \pm 0.77s$. Similarly for a large N, a normal distribution can be assumed yielding:

$$\bar{\bar{R}}_a = \bar{R}_a \pm \frac{2s}{\sqrt{N}} \text{ for 95\% confidence}$$

The length of an individual profile again depends upon the degree of confidence required for the parameter measurement and upon the parameter to be measured but in all cases the important consideration is the product of the bandwidth of the profile and the duration of the signal. This gives the number of degrees of freedom of the data within the profile length, corresponding to N the number of profiles taken in the equations listed above.

To be statistically significant the degrees of freedom should be independent. This means that for a record of length T the number of independent degrees of freedom is $2BT$ where B is the signal bandwidth.

For a typical ground surface a 10 mm record would have about 400 degrees of freedom. Note that by independent sample it is not just implied that every digital measurement can be regarded as a degree of freedom; the distance over which samples can be regarded as independent has to be worked out from the correlation function.

As an example the standard error for an RMS parameter measured on a profile is $\sim 1/\sqrt{BT}$. For the same confidence of measurement of a peak value the duration of the signal must be much greater.

Parameters

The most commonly used height parameter in surface texture is the R_a value, previously called the cla in the UK and aa in the USA. It is simply the average deviation of the profile from the reference mean line. Other height parameters include the maximum peak to valley height within the sample length R_T, the average deviation of the five highest peaks and five lowest valleys, R_z and many more all of which are well documented in the literature[14]. They are in fact all measures of the amplitude probability density function of the surface profile. Other statistical measures based on this are the skew (the third central moment) and the kurtosis (the fourth central moment). These can give useful information about the surface waveform and can indicate the extent to which the R_a value will be useful, eg, for a skew of greater than 2 it is unlikely that the R_a will mean much because the profile will have large peaks (or valleys) in it.

Significantly while much attention has been paid to height parameters little attempt has been made to give an index to the crest-spacings. Recently it has been shown how much valuable information can be obtained from the autocorrelation function or power spectrum; the statistical equivalents in length to the amplitude probability density in height[15,16]. The use of these statistical concepts is proving to be very rewarding in modern metrology.

A practical wavelength conscious parameter based loosely on the philosophy of Whitehouse and Archard revealed in an earlier paper is that of average or RMS wavelength (λ_a or λ_{RMS}) which it has been shown can be obtained directly from the power spectrum of the surface[17]. Thus:

$$\lambda_{RMS} = 2\pi \frac{y_{RMS}(x)}{y_{RM}(x)} \text{ or } \lambda_a = 2\pi \frac{R_a}{\Delta_a}$$

where Δ_a is the average slope.

From results obtained so far it seems that a more or less complete classification of the surface could be obtained from R_a, λ_a (or the independence length $2.3\beta^*$) and the skew. Fig 9 shows some examples of how R_a and λ_a look on practical surfaces.

The use of this multiplicity of parameters has not yet permeated down to roundness measurement but it no doubt will in the future. In roundness usually only the maximum peak to valley deviations are measured from the reference line.

Digital methods

1 Sampling

The first consideration in sampling is the frequency content of the waveform to be digitized. Often an estimate of the smallest frequencies present in the waveform can save much computation.

If the highest frequency is f then according to Nyquist the minimum rate of sampling to completely define the waveform is $2f$. Under these conditions the samples are uncorrelated. In practice, no signal cuts off sharply at frequency f so that the usual procedure is to sample at about three times faster than the smallest wavelength likely to be of interest. Sampling too often results in highly correlated and redundant information, whereas sampling too rarely can result in aliasing, which basically means that frequencies in the waveform higher than the sampling frequency can actually appear to be low frequencies. Aliasing can be removed either by sampling faster or by smoothing of the data prior to evaluation usually by high-cut filtering. Of these solutions the latter is to be preferred, although it is important to know where to position the filter cut.

In surface texture measurements the signal representing the roughness waveform has already been smoothed relative to the true surface profile because of the finite-tip of the stylus which acts as a mechanical filter. So the problem of trying to decide the highest frequencies present in the waveform has to some extent already been solved before the digitizing has even started.

In practical instruments the tip of the stylus is about 2.5 μm which implies that sampling should take place about every one micrometre. If flaws or freak events such as maximum peaks are being looked for, then this must be the order of sampling but if only averaging parameters are being assessed then about 5 micrometres is good enough. The problem of sampling to find extreme behaviour is not simple but in a large number of practical engineering situations it is important.

2 Quantization

This refers to the accuracy in the digitization of each individual measurement of the profile waveform. It does not refer to the instrumental accuracy. In a typical analogue to digital convertor four or five decimal places are usually

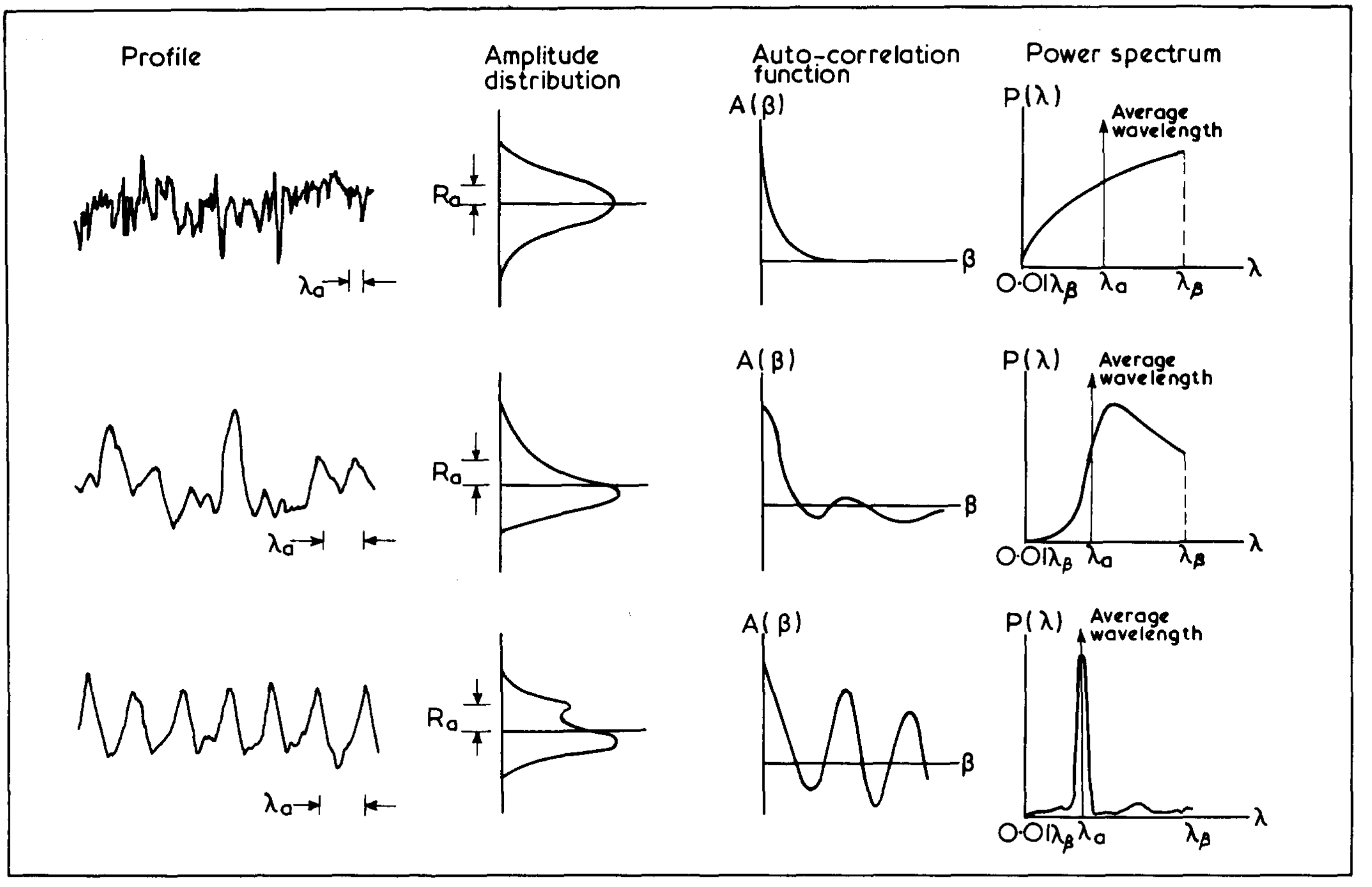

Fig 9 A useful characterisation of surface finish waveforms based on an amplitude and a wavelength parameter.

used, but the accuracy ultimately depends on the input. Up to now three decimal digits have been sufficient. Errors produced by other effects have usually been larger than these quantization errors.

3 Computation of parameters

Assuming that say the true roughness data is available, there are still two main reasons why results from different researchers do not tie up. They are a) lack of definition of parameters and b) numerical analysis errors.

As an example of this type of confusion (relevant incidentally not only to digital but analogue techniques) consider the numerical definition of a peak. In digital terms the simplest and most usual definition is that employing three-point analysis. In this, three ordinates are used to define a peak, the definition being that if the central ordinates of three contiguous ordinates is higher than the other two then this constitutes a peak. Scanning a digitized surface profile on this basis for peaks would result in a certain density of peaks.

If, on the other hand, four ordinates or five were to be used then the number as counted would be different. Again, suppose that in the three-point definition, the adjacent ordinates had to be a specific distance below the central one before it was counted as a peak, then here again the count would be different. This method has been used in the steel industry.

Another point concerning definition is the realization that even within the confines of one definition different results can be obtained, for instance, if the data is unsmoothed and the ordinates are simply those obtained from the A/D convertor then the quantization interval can determine to some extent the count number. If the accuracy is four decimal digits, the count could be higher than for three decimal digits. On the same point if, say, the three-point definition is being used for the peak on a surface then the count obtained for a fixed accuracy of data, say three digit, will to some extent depend on the sampling distance because the closeness of samples which make them highly correlated can be such that differences between them are smaller than the quantization interval which would mean that some peaks would not be counted. *Fundamentally, the sampling, quantization, and the definition* of the parameter must all be tied together before realistic comparisons between investigators can be made. It can be shown for instance that with no quantization error, using the same definition, the peak count can be changed by over 30% by just varying the sampling interval for a random surface[16].

The only way out of this is to fix the definition in a sound mathematical way and choose the quantization interval and sampling distance to fit the surface characteristics. As far as the quantization is concerned it should be fixed to give acceptable accuracy, and all subsequent surfaces then arranged to be digitized so that the profile covers about the same vertical extent of the chart. If three-point analysis is to be used then samples should be taken to be as near to the correlation distance as possible ie, about 50–100th of the cut-off, so as to give the information relevant to the dominant features of the profile ie, the number of degrees of freedom; closer sampling would require a new definition perhaps five or seven ordinates etc, depending on how close the ordinates had been spaced relative to the correlation distance.

Similarly in the definition of derivatives the same point arises. It is not as a rule satisfactory to use the simple three-point analysis for the definition of the first or second differentials, especially if the curvature at peaks is being measured. In this case the region near to the peak is only slowly moving hence quantization is important, because infinite radii can be measured as a result of the method of computation and not due to the surface itself. Similarly slopes can be found numerically that are greater than the 45° semi-angle limit that the stylus itself must impose. A better definition of the numerical differentiation technique reduces this problem.

One of the most difficult problems arises when the definition itself is difficult, for instance in three dimensions. Taking the simple case of a peak it is possible to define a summit or 3D peak in a number of ways. Depending on the definition there can be an even bigger divergence of values that exist with the 2D case. For instance, the equivalence of three-point analysis is now five-point analysis ie, three-point analysis in two orthogonal directions instead of one direction as for a single profile trace. The difference between the number of summits for close spacing, as opposed to those for wide spacing, is now over 3:1. Some investigators use as a definition of a summit the criterion that the surrounding eight ordinates of a square lattice must be smaller than the centre ordinate for a summit to be detected. This definition is not as consistent as the former because the weights of the ordinates relative to the central one are unequal. But in the event that this definition is used then the difference in the count with sampling distance could be as much as 10:1. Here one of the assumptions is that the parallel tracks are separated by the same spacing as the ordinate spacing. Bigger divergences could be expected for non-isotropic sampling.

In all digital analysis attention has to be given to numerical techniques. There is a well-known danger of people versed in analogue arts doing elementary transcriptions into digital situations with the resulting considerable chance of error. The sort of operations usually required in surface topography are numerical integration, differentiation, interpolation, extrapolation, matrix inversion etc.

Perhaps the most obvious is that of integration which is required in the determination of the mean line when the convolution integral is being used. The natural temptation is to add all digital results and to call this an integration. This becomes near to the truth when many ordinates are used or if the function is well behaved but it is often more economical in terms of computing time to reduce the number of points on a curve considerably and to use a simple numerical integration formula. A simple example follows.

If the ordinates of a curve to be integrated are $y_0, y_1 \ldots y_n$ then if the spacing is h:

$$\frac{1}{h}\int_{x_0}^{x_n} y\,\mathrm{d}x = (\tfrac{1}{2}y_0 + y_1 + y_2 \ldots y_{n-1} + \tfrac{1}{2}y_n)$$

which is the trapezoidal rule for integration which is a simplification of Gregory's rule for integration. In this a linear interpolation is made between the points whereas for Simpson's rule where:

$$\frac{1}{h}\int_{x-1}^{x1} y\,\mathrm{d}x = \frac{(y_1 + 4y_0 + y_{-1})}{3}$$

a quadratic interpolation is used which is correspondingly more accurate. There are many such integration formulae that can be used. Needless to say that in surface topography, because in most cases large numbers of ordinates are used, the trapezoidal rule is adequate. A similar source of numerical error is in numerical differentiation of one sort or another[18].

An accurate numerical differentiation, assuming no quantization errors, can be achieved using 5 or 7 ordinates in the formula, and not just the three used in the three-point analysis. A good formula for numerical differentiation is given by:

$$h\frac{\mathrm{d}y_0}{\mathrm{d}x} = \frac{1}{60}\,[y_3 - 9y_2 + 45y_1 - 45y_{-1} + 9y_{-2} - y_{-3}]$$

The differentiation can be further improved by the addition of a small correction. The following rather unsatisfactory formula is used in three-point analysis:

$$h\frac{\mathrm{d}y_0}{\mathrm{d}x} = \frac{1}{2}\,[y_{+1} - y_{-1}]$$

It is especially in these differentiating routines that the noise is critical because the noise whether electrical or numerical tends to get enhanced. It is in using formulae like these that large or small angles can be obtained which are not physically realistic.

The second differential for three-point analysis is:

$$h^2\frac{\mathrm{d}^2y_0}{\mathrm{d}x^2} = (y_1 - 2y_0 - y_{-1})$$

a better formula is:

$$h^2\frac{\mathrm{d}^2y_0}{\mathrm{d}x^2} = \frac{1}{180}\,(2y_3 - 27y_2 + 270y_1 - 490y_0 +$$

$$270y_{-1} - 27y_{-2} + 2y_{-3})$$

Other well-known formulae exist for interpolation and extrapolation, in particular those due to Everett and Bessel[18]. An example of the use of interpolation formulae in surface topography comes in the field of contact where mechanical models of surface peaks are used. Using simple interpolation it is possible to find the position of the maximum value of a peak even if it is not touched by the sampled points. For instance, if the ordinates are y_{-1} and y_{+1} where $y_1 \neq y_{-1}$ then the apex is not at y_0 but at distance V from y_0 given by:

$$V = \frac{(h/2)(y_{-1} - y_1)}{2y_0 - y_1 - y_{-1}}.$$

It is further possible to work out curvatures at the apex of this peak. Interpolation can also help to cut down the number of points that need to be calculated for the reference line assessment. Obviously the subject of numerical analysis is vast and specialized. What has been pointed out here are just a few of the particular problems encountered by tribologists who are researching into surface metrology problems.

CONCLUSIONS

The information presented here is in no way intended to be exhaustive in its content. It is hoped that a number of topics have been aired sufficiently to show the way the measurement of surfaces is evolving; in particular by the use of digital methods and further, that enough detail has been given about some of the problems involved to reduce the probability of errors which can occur.

Perhaps the most important feature which is emerging is that the proven versatility and robustness of the stylus method in measuring surfaces, especially at the input end, is now being complemented by the versatility and control made possible by using minicomputers at the output end of the instrument.

REFERENCES

1 **Kinsey, D. and Chetwynd, D. G.** 'Some applications of a small computer to the on-line measurement of surfaces', Imeko VI, Dresden (1973)

2 'Standards calibration goes on-line', *Metrology and Inspection* (Apr 1973)

3 **Harris, P.** Technical Report T37, Rank Precision Industries (1970)

4 **Chetwynd, D. G. and Kinsey, D.,** Proceedings of IEE Computers in Measurement, York (Sep 1973)

5 **Whitehouse, D. J.** 'Modern trends in surface measurement', *Proc Inst Belgium Soc Mech Eng,* **21** (1974)

6 **Kaliszer, H. and Rowe, G. W.** 'Surface integrity after severe wear or grinding', Conference on Surface Technique, Pittsburgh (1973)

7 **Whitehouse, D. J. and VanHerck, P.** *Annale CIRP,* **21/2** (1972)

8 **Whitehouse, D. J. and Reason, R. E.** Rank Organisation (1965)

9 **Whitehouse, D. J.** *Proc Inst mech Engrs,* **182 pt3k** (1967/68)

10 **Weingraber, H. von** *Werkstattstechnik und Maschinenbau,* **46,** No 5 (1965)

11 **Radhakrishnan, V.** *Proc Inst mech Engrs,* **187,** No 43 (1973)

12 **Whitehouse, D. J.** *J Phys E,* **6** (1973)

13 **Bendat, J. S. and Piersol, A. G.** 'Measurement and analysis of random data', Wiley (1966)

14 **Reason, R. E.** 'The measurement of surface texture', Modern workshop technology part 2 (1970)

15 **Peklenik, J.** 'Investigation of surface typology', *Annals CIRP* (1967)

16 **Whitehouse, D. J. and Archard, J. F.** *Proc Roy Soc Lond A,* **316** (1970)

17 **Spragg, R. C. and Whitehouse, D. J.** *Proc Inst mech Engrs,* **185,** No 47 (1971)

18 Interpolation and allied tables, HMSO (1956)