PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Two-Dimensional Discrete Properties of Random Surfaces

D. J. Whitehouse and M. J. Phillips

Phil. Trans. R. Soc. Lond. A 1982 **305**, 441-468
doi: 10.1098/rsta.1982.0043

Phil. Trans. R. Soc. Lond. A **305**, 441–468 (1982) [441]
Printed in Great Britain

TWO-DIMENSIONAL DISCRETE PROPERTIES OF RANDOM SURFACES

By D. J. WHITEHOUSE† and M. J. PHILLIPS‡

† *Department of Engineering, University of Warwick, Coventry CV4 7AL, U.K.*
‡ *Department of Mathematics, University of Leicester, Leicester LE1 7RH, U.K.*

(*Communicated by D. Tabor, F.R.S. – Received* 4 *August* 1981)

CONTENTS

Recent work has shown that it is possible to predict surface parameters measured digitally from a surface profile by means of three points on the autocorrelation function. The weakness of this work has been that only one-dimensional parameters have been evaluated. The present contribution extends the theory to include two-dimensional parameters of the surface which are expressed in terms of between four and seven points on the autocorrelation function depending on the type of surface. It is shown that this technique provides an alternative to traditional mapping methods.

It is shown also that as a general rule results obtained from the discrete analysis do not converge to those obtained from the continuous theory. The nature and magnitude of the differences between the two approaches are discussed in detail.

Finally, the theoretical results are confirmed experimentally and the general significance of discrete methods reviewed.

1. Introduction

During the last few years two extra considerations have emerged to constrain the normal use and manufacture of machined parts. These are the need to conserve energy and the need to conserve material. The main objective of this contribution is to attempt to quantify ways in which the performance of parts can be predicted before the final assembly of the individual components is carried out. Any degree of success in this endeavour will help to reduce the amount of wasted time, material and energy which inevitably result when break-downs occur.

In many critical cases the performance of an engine, machine or device is influenced by the mechanism by which two solid parts make static or sliding contact. A crucial factor in this mechanism is the geometry of the surfaces in contact. To control the work-piece performance it is necessary to control the surface geometry, which must therefore be measured. This measurement is not an easy task. Much effort has been expended in the past to develop methods whereby the two-dimensional characteristics of surfaces can be assessed. There are two possible solutions to this problem: one is to develop ways of surveying all of the areas in contact; the other is to find simple instrumental ways of predicting the relevant information from limited data. It is the latter, cheaper option that has been selected for investigation in this paper.

The first problem to be resolved is the nature of the information required. Current thinking suggests that the important features would include the heights of summits, their curvatures and slopes, and the general statistical distributions of these variables.

Initially, ideas on which parameters were important in contact phenomena were few. In fact, it is only since the pioneering work of Bowden & Tabor (1954) that surface topography has been recognized as being of vital importance in contact phenomena.

Once it is realized that topography is important, the question arises as to which parameters of the surface are most important, and whether it is necessary to measure two-dimensional as distinct from one-dimensional characteristics. Among the first workers to attempt to resolve the problem of the functional significance of surface parameters were Greenwood & Williamson (1966) who derived an index of plasticity Ψ_1, given by

$$\Psi_1 = \frac{E}{H}\left(\frac{\overset{*}{\sigma}}{R}\right)^{\frac{1}{2}}, \tag{1.1}$$

where E is the elastic modulus, H is the hardness, $\overset{*}{\sigma}$ is the r.m.s. value of the peak height distribution and $1/R$ is the peak curvature, which for this model is assumed constant for all peaks. Using the premise that the heights of peaks have a Gaussian distribution, they imposed limits on the value of Ψ_1 to explain contact behaviour. For $\Psi_1 > 1$ a large part of the contact would be plastic but for $\Psi_1 < 0.6$ the chances of the plastic flow of asperities would be remote. Various other attempts have been made to devise numerical indexes to predict the elastic–plastic behaviour of surfaces. Whitehouse & Archard (1970) modified equation (1.1) to allow for a distribution of curvature of asperities in the one-dimensional profile of a surface. Furthermore, they related Ψ to profile parameters rather than asperity models. Thus, Ψ_1 became Ψ_2, where

$$\Psi_2 = 0.69\,\frac{E}{H}\frac{\sigma}{\overset{*}{\beta}}. \tag{1.2}$$

In (1.2) σ, the r.m.s. of the surface profile height, replaces $\overset{*}{\sigma}$, the r.m.s. of the peak height distribution. Instead of R, the average radius, a new parameter is used: this is $\overset{*}{\beta}$ which is a measure of the independence distance or correlation length of the surface.

Other indices, such as Ψ_3 introduced by Mikic (1974), concentrated on slope information rather than height and length parameters separately:

$$\Psi_3 = (E/H) \tan \theta, \tag{1.3}$$

where $\tan \theta$ is the absolute mean slope of the surface profile. The index Ψ_3 is such that at least 2% of the asperities will deform plastically when $\Psi_3 > 0.25$. For $\Psi_3 < 0.25$ contact would be predominantly plastic. Tabor (1975) pointed out that it is possible to consider the $\sigma/\overset{*}{\beta}$ ratio in (1.2) as a slope, in which case there is very little difference between Ψ_2 and Ψ_3.

Gupta & Cook (1972) defined a topographic index Ψ_4 in two dimensions as the ratio of real to apparent area of contact:

$$\Psi_4 = \frac{1}{(2D_s)^{\frac{1}{2}}} \frac{1}{R_m} \frac{\overset{*}{E}}{H}, \tag{1.4}$$

where D_s is the average number of summits per unit area, R_m is the harmonic mean of the average radius of asperities of the two surfaces making the contact and $\overset{*}{E}$ is the harmonic mean of the elastic moduli.

Another attempt to devise a two-dimensional index of plasticity has been made by Nayak (1971) using two-dimensional random process theory. His results, however, were dependent on results reported by other workers (Pullen & Williamson 1972) and these results have since been queried. Nayak's index Ψ_5 is given by

$$\Psi_5 = (E/H)\,\sigma_1', \tag{1.5}$$

where σ_1' is the r.m.s. value of the differential of that component of a surface in contact which has a narrow band spectrum. For $\Psi_5 > 5$ most of the asperities deform plastically.

Other workers have approached the problem somewhat differently, but at least one general conclusion can be reached. This is that the ratio of E/H, which represents the physical rather than the geometrical properties of the surfaces in contact, is fundamental. This ratio has often been combined with various differential characteristics of the surfaces to yield an index. Also, many workers have found during their investigations that the scale of size of asperity is also a critical factor in contact behaviour (Jones *et al.* 1975).

However, agreement as to which topographic parameter is most important is unlikely to be reached, at least in the foreseeable future. It may never be possible or even necessary because the most significant parameter or combination of parameters ultimately depends on the application of the surface. For example, Fuller & Tabor (1975) have derived an index describing how surface characteristics influence the adhesive behaviour between components. They showed that for elastic contact there would be negligible adhesion unless

$$E\sigma^{\frac{3}{2}}/(R\Delta\gamma)^{\frac{1}{2}} < 10, \tag{1.6}$$

where $\Delta\gamma$ is the surface energy density and R is the peak radius.

Another important practical situation occurs in those circumstances, such as in lubrication, where a fluid separates two bodies nominally in contact. Here it is the presence of peaks at given levels above the mean line, rather than their curvature, that is the dominant factor.

As an example, Dawson (1962) found that the ratio of d_t/σ, is important in pitting, where σ is the r.m.s. value of the surface height and d_t is the fluid film thickness. Dyson (1976) modified Reynold's equation to include surface texture and showed that the correlation length of the surface was significant in determining the pressure and shear flow generated by movement.

The texture has also been shown to be an important feature in determining the amount of wear between rubbing bodies (Hirst & Hollander 1974), except when very high shear forces are involved (Suh 1973).

Although a survey of the effect of texture on these and related subjects (Whitehouse 1978) has shown that great progress has been made in understanding the mechanism of contact, in most instances only one-dimensional texture parameters have been used. Obviously, because contact is fundamentally two-dimensional, it would be more realistic to incorporate the two-dimensional surface parameters but, unfortunately, even with today's sophisticated instruments the complete measurement of such parameters is impossible. What is needed is a relatively simple way of measuring all the topographic features that could be important, such as the curvature and distributions of the two-dimensional summits. Furthermore, this information should ideally be capable of being broken down into 'size' components, a point that is important because it has been shown that surface parameter values, as such, are not intrinsic to the surface, but that the value obtained from a particular investigation depends on the 'scale of size' considered.

Taking first the 'scale of size' problem, Whitehouse & Archard (1970) attempted to find a solution for the one-dimensional profile by noting the fact that surfaces are invariably assessed by using digital analysis. Therefore, they investigated the discrete rather than the continuous properties of random waveforms. They considered that a surface profile could be represented by height ordinates, having a Gaussian distribution, which were related spatially along the surface by an exponential autocorrelation function. Using this model for a surface, they evaluated the topographic parameters of significance in tribology in terms of the r.m.s. value of the profile height and one point of the autocorrelation function. The restriction to the exponential autocorrelation function was relaxed by Whitehouse & Phillips (1978) who were able to predict the same, and other, peak parameters for surfaces having a general autocorrelation function. They used the r.m.s. value of the profile height and two points of the autocorrelation function. The validity of such an approach has since been demonstrated, for example by Hamed & Whitehouse (1979).

The discrete method described above can save much time and effort in programming and data acquisition because so little information needs to be obtained before an estimate of the many useful parameters can be made.

It is the intention of this paper to extend the discrete approach to the more realistic two-dimensional surface.

2. Discrete measurement of surfaces in two dimensions

2.1. *Definitions of summit density, height and curvature*

When the discrete properties of the one-dimensional profile are being investigated, the data consist simply of a set of digital measurements, called ordinates, taken along the profile and separated from one another by equal increments h. For the two-dimensional surface the problem is more complicated, and a grid of data points taken over the surface is required. Similarly, definitions become more complicated.

In one dimension a continuous definition of a local maximum or peak on a profile only requires one first-order and one second-order differential, and only three ordinates are needed for the discrete definition of a peak. However, in two dimensions, the continuous definition of a local maximum (or summit, in the terminology of Nayak (1971)) requires two first-order and three second-order differentials, and a minimum of five ordinates are needed for the discrete definition of a summit. Sayles & Thomas (1977) gave two discrete definitions of a summit: one using the five nearest-neighbour ordinates, and the other nine. In this paper the five-ordinate definition will be used. The definition states that an ordinate at height z_0 above an arbitrary datum is a five-point summit if it is higher than the four neighbouring ordinates, which are at a distance h from it, on the two orthogonal Cartesian axes. If the differences between z_0 and each of the four neighbouring ordinates are denoted by s_1, s_2, s_3 and s_4, then the condition for the ordinate z_0 to be a summit is that s_1, s_2, s_3 and s_4 are all positive.

The summit density is the proportion of ordinates that are summits. The summit height is the height z_0 of an ordinate that is a summit.

It seems natural to extend to two dimensions the discrete definition of Whitehouse & Archard (1970) for peak curvatures, which was given by

$$c_h = (s_1+s_2)/h^2, \tag{2.1}$$

where s_1 and s_2 are defined in (2.3). This extension uses the average orthogonal peak curvatures, and gives the discrete definition of five-point summit curvature as

$$c_h^{(2)} = \tfrac{1}{2}(s_1+s_2+s_3+s_4)/h^2, \tag{2.2}$$

which again is a linear combination of the ordinates.

2.2. *The expected summit density and the distributions of summit height and curvature*

The distributions of peak height and curvature for a profile in one dimension were obtained by Whitehouse & Phillips (1978) for a surface with ordinates from a Gaussian (normal) distribution. (A multivariate normal distribution for a vector $\boldsymbol{Y}$ will be denoted by $\boldsymbol{Y} \sim N[\boldsymbol{\mu}; V]$ where $\boldsymbol{\mu}$ is the vector of means and V is the variance–covariance matrix, and the probability density function is given by $\phi^{(m)}(\boldsymbol{y}'-\boldsymbol{\mu}'; V)$.) They obtained these results from the theory of truncated random variables. This was because the peak height distribution is the conditional distribution of Y_0, the profile height, given that S_1 and S_2 are positive, or, in other words, the distribution of Y_0 conditional on S_1 and S_2 being positive, where

$$S_1 = Y_0 - Y_{-1}, \quad S_2 = Y_0 - Y_1, \tag{2.3}$$

and Y_{-1} and Y_1 are the preceding and succeeding ordinate values, respectively, on the profile. Similarly the distribution of peak curvature is the conditional distribution of C_h given that S_1 and S_2 are positive. Hence the results of Whitehouse & Phillips (1978) were obtained by using the results of the Appendix for truncated random variables, with $m = 2$, $Y_0 \equiv Z_0$, $\boldsymbol{X} = (2-2\rho_1)^{-\frac{1}{2}}(S_1, S_2)'$, $d = (\frac{1}{2}-\frac{1}{2}\rho_1)^{\frac{1}{2}}$ and

$$V = \begin{bmatrix} 1 & \frac{1}{2}(1-2\rho_1+\rho_2)/(1-\rho_1) \\ \frac{1}{2}(1-2\rho_1+\rho_2)/(1-\rho_1) & 1 \end{bmatrix},$$

where $\rho_1 = \rho(h)$, $\rho_2 = \rho(2h)$.

The derivations given in the Appendix can also be used to obtain the more important two-dimensional distributions of five-point summit height and curvature. For this analysis the surface height measurements will be assumed to have a multivariate normal distribution and because the surface is assumed to be isotropic, the distribution properties of a profile are invariant with respect to the direction of the profile. Hence

$$(Z_0, (2-2\rho_1)^{-\frac{1}{2}}(S_1, S_2, S_3, S_4)) \sim N[\mathbf{0};\, \boldsymbol{V}_5], \tag{2.4}$$

where

$$\boldsymbol{V}_5 = \begin{bmatrix} 1 & d & d & d & d \\ d & 1 & a & b & b \\ d & a & 1 & b & b \\ d & b & b & 1 & a \\ d & b & b & a & 1 \end{bmatrix}, \tag{2.5}$$

with

$$d = (\tfrac{1}{2} - \tfrac{1}{2}\rho_1)^{\frac{1}{2}}, \tag{2.6}$$

$$a = \tfrac{1}{2}(1-2\rho_1+\rho_2)/(1-\rho_1), \tag{2.7}$$

and

$$b = \tfrac{1}{2}(1-2\rho_1+\rho_3)/(1-\rho_1), \tag{2.8}$$

where $\rho_1 = \rho(h)$, $\rho_2 = \rho(2h)$, $\rho_3 = \rho(\surd 2\, h)$, $\rho(t)$ is the correlation coefficient between ordinates a distance t apart.

If Υ_5 is the event $\{S_1 > 0, S_2 > 0, S_3 > 0, S_4 > 0\}$ then the expected five-point summit density is the probability of Υ_5 occurring, and the distribution of five-point summit height is the conditional distribution of Z_0 given that Υ_5 has occurred. These can be obtained from the results of the Appendix with $m = 4$,

$$\boldsymbol{X} = (2-2\rho_1)^{-\frac{1}{2}}(S_1, S_2, S_3, S_4)', \tag{2.9}$$

$$d = (\tfrac{1}{2} - \tfrac{1}{2}\rho_1)^{\frac{1}{2}} \tag{2.10}$$

and $\boldsymbol{V}$ obtained from $\boldsymbol{V}_5$ by removing the first row and column (and denoted by $\boldsymbol{V}_4$), so that

$$\lambda = 1+a+2b. \tag{2.11}$$

Then, from (A 11), the probability density function of the five-point summit height distribution is given by

$$f(z_0|\Upsilon_5) = \frac{\Phi^{(4)}(z_0[(1-\rho_1)/(1+\rho_1)]^{\frac{1}{2}};\, \boldsymbol{V}_{\mathrm{c}})\,\phi(z_0)}{\Phi^{(4)}(0;\, \boldsymbol{V}_4)}, \tag{2.12}$$

where

$$\boldsymbol{V}_{\mathrm{c}} = \begin{bmatrix} 1 & a_{\mathrm{c}} & b_{\mathrm{c}} & b_{\mathrm{c}} \\ a_{\mathrm{c}} & 1 & b_{\mathrm{c}} & b_{\mathrm{c}} \\ b_{\mathrm{c}} & b_{\mathrm{c}} & 1 & a_{\mathrm{c}} \\ b_{\mathrm{c}} & b_{\mathrm{c}} & a_{\mathrm{c}} & 1 \end{bmatrix}, \tag{2.13}$$

$$a_{\mathrm{c}} = (\rho_2 - \rho_1^2)/(1-\rho_1^2) \tag{2.14}$$

and

$$b_{\mathrm{c}} = (\rho_3 - \rho_1^2)/(1-\rho_1^2). \tag{2.15}$$

From (A 12), the expected (or average or mean) five-point summit height is given by $E(Z_0|\Upsilon_5)$, where E denotes the statistical expectation:

$$E(Z_0|\Upsilon_5) = \frac{2[(1-\rho_1)/\pi]^{\frac{1}{2}}\,\Phi^{(3)}(0;\, \boldsymbol{B}_4)}{\Phi^{(4)}(0;\, \boldsymbol{V}_4)}, \tag{2.16}$$

where

$$\boldsymbol{B}_4 = \begin{bmatrix} 1-b^2 & a-b^2 & b(1-a) \\ a-b^2 & 1-b^2 & b(1-a) \\ b(1-a) & b(1-a) & 1-a^2 \end{bmatrix}. \tag{2.17}$$

From (2.16) it can be seen that the expected five-point summit height depends on two orthant probabilities $\Phi^{(3)}(0; \boldsymbol{B}_4)$ and $\Phi^{(4)}(0; \boldsymbol{V}_4)$, which have to be evaluated. From David (1953), $\Phi^{(3)}(0; \boldsymbol{B}_4)$ is given by

$$\Phi^{(3)}(0; \boldsymbol{B}_4) = \tfrac{1}{2} - (4\pi)^{-1} \arccos\,[(a-b^2)/(1-b^2)] \\ - (2\pi)^{-1} \arccos\,\{b(1-a)\,[(1-a^2)\,(1-b^2)]^{-\frac{1}{2}}\}. \quad (2.18)$$

Cheng (1969) has evaluated $\Phi^{(4)}(0; \boldsymbol{V}_4)$, the expected five-point summit density, using Plackett's (1954) method. As this orthant probability only depends on the parameters a and b, given by (2.7) and (2.8), it will be denoted by $\Phi^{(4)}[a, b]$. Then

$$\begin{aligned} \mathrm{pr}\,(\Upsilon_5) &= \Phi^{(4)}[a, b] \\ &= \Phi^{(4)}[a, 0] + \int_0^b \frac{\partial \Phi^{(4)}[a, t]}{\partial t}\,\mathrm{d}t \\ &= [\tfrac{1}{4} + (2\pi)^{-1} \arcsin a]^2 + (2\pi)^{-1} \arcsin b \\ &\quad + \pi^{-2} \int_0^b (1-t^2)^{-\frac{1}{2}} \arcsin\,[t(1-a)\,(1+a-2t^2)^{-1}]\,\mathrm{d}t. \end{aligned} \quad (2.19)$$

This result was given by Cheng (1969) in terms of the dilogarithm function (see Lewin 1958).

The distribution of a five-point summit having a height Z_0 conditional on a curvature $C_h^{(2)}$ is normal, and with the aid of the result of (A 8) and (A 4) it is given by

$$(Z_0 | C_h^{(2)} = c_h^{(2)}) \sim N[h^2(1+a+2b)^{-1} c_h^{(2)},\ 1-2(1-\rho_1)/(1+a+2b)]. \quad (2.20)$$

This is henceforth called the conditional distribution of summit height given curvature. Thus the expected five-point summit curvature is given by

$$E(C_h^{(2)} | \Upsilon_5) = h^{-2}(1+a+2b)\,E(Z_0 | \Upsilon_5), \quad (2.21)$$

with the use of (2.16).

Hence by the application of the theory of Gaussian (normal) truncated random variables it has been possible to obtain the expectations of the five-point summit height, curvature and density.

2.3. *The effect of the sampling interval and limiting results*

The distributions of five-point summit height and curvature have been derived in terms of the correlation coefficients between ordinates. These correlation coefficients are ρ_1, for ordinates a distance h apart, ρ_2, for ordinates $2h$ apart, and ρ_3, for ordinates $\sqrt{2}h$ apart. If the surface is isotropic and the autocorrelation function is $\rho(t)$, then $\rho_1 = \rho(h)$, $\rho_2 = \rho(2h)$ and $\rho_3 = \rho(\sqrt{2}h)$. So ρ_1, ρ_2 and ρ_3 will vary as h varies, depending on the shape of the autocorrelation function of the surface.

As h approaches zero

$$\lim_{h\to 0} \rho_1(h) = \lim_{h\to 0} \rho_2(h) = \lim_{h\to 0} \rho_3(h) = 1, \quad (2.22)$$

and as h approaches infinity

$$\lim_{h\to \infty} \rho_1(h) = \lim_{h\to \infty} \rho_2(h) = \lim_{h\to \infty} \rho_3(h) = 0. \quad (2.23)$$

If ρ_1, ρ_2 and ρ_3 are plotted in three dimensions then as h varies the curve will start at (1, 1, 1) for $h = 0$ and end at (0, 0, 0) for $h = +\infty$. In order that the matrix $\boldsymbol{V}_4 - d^2\boldsymbol{J}$ be positive definite it is necessary for this curve to lie in the region bounded by

$$\rho_2 < 1$$

and

$$-\tfrac{1}{2}(1+\rho_2) + 2\rho_1^2 < \rho_3 < \tfrac{1}{2}(1+\rho_2).$$

Results for the summit height have been obtained by Nayak (1971) for the continuous surface. So it is possible to compare his results with those obtained for the discrete results of §2.2 as the sampling interval h approaches zero. From (2.16) the expected summit height depends on ρ_1, ρ_2 and ρ_3 through a and b, and

$$\lim_{h\to 0} a = -1 \quad \text{and} \quad \lim_{h\to 0} b = 0. \tag{2.24}$$

The autocorrelation function of the surface can be approximated by

$$\rho(h) = 1 + D_2 h^2/2! + D_4 h^4/4! + o(h^4), \tag{2.25}$$

where D_2 and D_4 are the second and fourth derivatives of the autocorrelation function at the origin. If we let

$$\eta = -D_2(D_4)^{-\frac{1}{2}} < \sqrt{\tfrac{2}{3}}, \tag{2.26}$$

the limiting value for the expected five-point summit height is given by

$$\lim_{h\to 0} E(Z_0|\Upsilon_5) = \frac{16}{\pi + 2\arcsin\frac{1}{3} + 4\sqrt{2}}(\tfrac{1}{2}\pi)^{\frac{1}{2}}\eta \tag{2.27}$$
$$= 1.688(\tfrac{1}{2}\pi)^{\frac{1}{2}}\eta.$$

Nayak (1971) showed that the expected continuous summit height for the two-dimensional case was

$$E(Z|\Upsilon_N) = [4\sqrt{2}/\pi](\tfrac{1}{2}\pi)^{\frac{1}{2}}\eta$$
$$= 1.801(\tfrac{1}{2}\pi)^{\frac{1}{2}}\eta, \tag{2.28}$$

which is comparable with the expected continuous peak height for one dimension,

$$E(Z|\Upsilon_p) = (\tfrac{1}{2}\pi)^{\frac{1}{2}}\eta, \tag{2.29}$$

a result given by Whitehouse & Phillips (1978). So it is seen that the limit of the expected five-point summit height is 69% larger than the expected peak limit, as opposed to 80% larger than the expectation of the distribution of summit height for the continuous definition of Nayak (1971). However, this is only an overall reduction of about 6% and suggests that the discrete five-point definition may be adequate.

It is possible to obtain $\Phi^{(4)}(z_0[(1-\rho_1)/(1+\rho_1)]^{\frac{1}{2}};\, V_c)$ by the methods of Plackett (1954) and Cheng (1969), and hence to obtain the probability density function of the limiting distribution of the five-point summit height as h converges to zero. This is given by

$$\lim_{h\to 0} f(z_0|\Upsilon_5) = 12\pi\phi(z_0)(\pi + 2\arcsin\tfrac{1}{3} + 4\sqrt{2})^{-1}$$
$$\times\{(1-\eta^2)[\phi(w) + w\Phi(w)]^2 + \tfrac{4}{3}(1-\tfrac{3}{2}\eta^2)\,\Phi^{(2)}(w;\, (\tfrac{1}{3}-\eta^2)/(1-\eta^2))$$
$$-(1-\eta^2)\,\Phi^{(2)}(w;\, 0) + 4(1-\eta^2)\,T^{(2)}(w,\, (2-3\eta^2)^{-\frac{1}{2}})\}, \tag{2.30}$$

where

$$T^{(2)}(w,\nu) = (2\pi)^{-1}\int_0^{\nu}(1+x^2)^{-2}\exp[-\tfrac{1}{2}w^2(1+x^2)]\,dx \tag{2.31}$$

and

$$w = \eta(1-\eta^2)^{-\frac{1}{2}}z_0. \tag{2.32}$$

This probability density function is compared with the probability density function of continuous summit height, given by

$$f(z|\Upsilon_N) = 3\eta(2-3\eta^2)^{\frac{1}{2}}\phi(0)\,z\phi((1-\tfrac{3}{2}\eta^2)^{-\frac{1}{2}}z)$$
$$+ 3\sqrt{3}\,\eta^2(z^2-1)\,\Phi(\eta[\tfrac{3}{2}/(1-\tfrac{3}{2}\eta^2)]^{\frac{1}{2}}z)\,\phi(z)$$
$$+ 2(1-\eta^2)^{-\frac{1}{2}}\Phi(\{\tfrac{1}{2}/[(1-\eta^2)(1-\tfrac{3}{2}\eta^2)]\}^{\frac{1}{2}}\eta z)\,\phi((1-\eta^2)^{-\frac{1}{2}}z), \tag{2.33}$$

which was obtained by Nayak (1971), for three values $\eta = 0$, $\sqrt{\frac{1}{3}}$ and $\sqrt{\frac{2}{3}}$ in figure 1. For $\eta = 0$ both distributions are the standardized normal distribution. When $\eta = \sqrt{\frac{2}{3}}$, the continuous summit height probability density function is

$$f(z|\Upsilon_N) = 2\sqrt{3}\{z^2 - 1 + 2\pi[\phi(z)]^2\}\phi(z), \quad \text{for} \quad z > 0,$$
$$= 0, \quad \text{for} \quad z \leqslant 0, \tag{2.34}$$

while the limiting distribution of the five-point summit height is given by

$$\lim_{h\to 0} f(z_0|\Upsilon_5) = \frac{8\pi\phi(z_0)\,[(2z_0^2-1)\,(\Phi(\sqrt{2}z_0)-\frac{1}{2})+\sqrt{2}z_0\phi(\sqrt{2}z_0)]}{\pi + 2\arcsin\frac{1}{3} + 4\sqrt{2}}, \quad \text{for} \quad z_0 > 0,$$
$$= 0, \quad \text{for} \quad z_0 \leqslant 0. \tag{2.35}$$

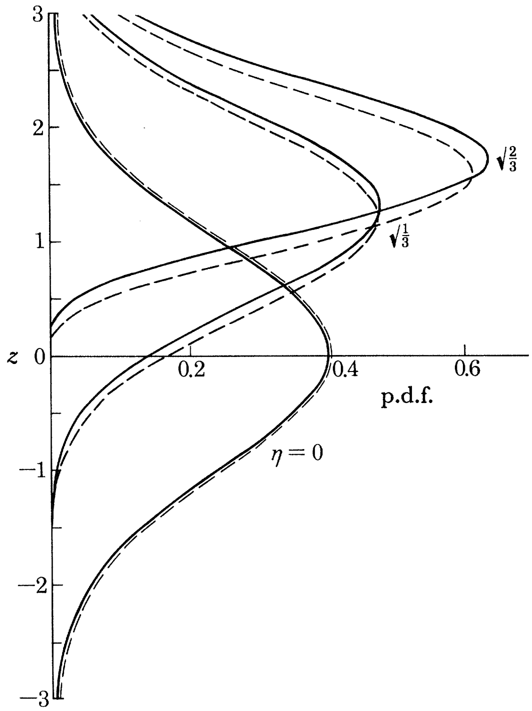

FIGURE 1. The probability density functions (p.d.fs) of the distribution of summit height (full line) and the limiting distribution of the five-point summit height (broken line) for $\eta = 0$, $\sqrt{\frac{1}{3}}$ and $\sqrt{\frac{2}{3}}$.

Nayak (1971) used the definition of mean summit curvature K_m given by Sokolnikoff (1951) as minus the average of the second partial derivatives in orthogonal directions. With this definition Nayak (1971) showed that the conditional distribution of continuous summit height given curvature is a normal distribution with

$$E(Z|K_m, \Upsilon_N) = \tfrac{3}{2}\eta K_m(D_4)^{-\frac{1}{2}} \tag{2.36}$$

and

$$\operatorname{var}(Z|K_m, Y_N) = 1 - \tfrac{3}{2}\eta^2. \tag{2.37}$$

This is also the limit of the conditional distribution of Z_0 given $C_h^{(2)}$. Hence, from (2.21), the limit as h tends to zero of the expected five-point summit curvature will again be 6% smaller than the expectation of the continuous summit curvature.

2.4. *Models of the autocorrelation function and the effect of the sampling interval*

To study the effect of the change of the sampling interval h on the digital measurements of an isotropic surface it is necessary to specify a model for the autocorrelation function of the surface. The model first suggested by Whitehouse & Archard (1970) was the negative exponential function, which simulated the rather idealistic finished machining marks obtained in very good grinding, so that

$$\rho(t) = \exp(-|t|). \tag{2.38}$$

Although useful as a first step this model had theoretical as well as practical problems. Whitehouse & Phillips (1978) showed that this model did not produce satisfactory limiting results as h tends to zero, because the derivatives are not defined at the origin.

For a model to fit in with observed autocorrelation functions on surfaces it would be desirable to have a multiplicative periodic function in the autocorrelation function. However, the problem of the derivatives remains. To overcome this problem Whitehouse & Phillips (1978) 'smoothed' the exponential-cosine function at the origin to give

$$\rho(t) = \exp[-|t|A_1(\theta)]\cos[2\pi\theta t A_1(\theta)] + \exp[-\beta|t|A_1(\theta)]\sin[2\pi\theta_0 t A_1(\theta)]/2\pi\theta_0, \tag{2.39}$$

where
$$\beta = [\tfrac{1}{3}(1+4\pi^2\theta_0^2) - 4\pi^2\theta^2]^{\frac{1}{2}}, \tag{2.40}$$

and
$$A_1(\theta) = \frac{1}{(1+4\pi^2\theta^2)}\left[1 + \frac{2\pi\theta}{\sinh(1/4\theta)}\right]. \tag{2.41}$$

This was called the 'model 1' autocorrelation function. For large β the dominant term of this autocorrelation function is the first, which reflects the behaviour observed for a number of types of engineering surfaces, especially those generated by single-tool-tip processes. If $\theta_0 = 6$, as was chosen by Whitehouse & Phillips (1978), then the effect of the second term is small. Whitehouse & Phillips (1978) used this model with $A_1(\theta)$ equal to unity. However, using (2.41) ensures that

$$\int_0^\infty |\rho(t)|\,\mathrm{d}t = 1, \tag{2.42}$$

so that the 'correlation length' does not vary with θ. This model will be used with $\theta = 0, \frac{1}{3}, \frac{1}{2}$ and 1, which will satisfy the condition on η of (2.26). The value of η can be obtained from the derivatives

$$D_2 = [1 - 2\beta - (2\pi\theta)^2][A_1(\theta)]^2 \tag{2.43}$$

and
$$D_4 = \{1 - 6(2\pi\theta)^2 + (2\pi\theta)^4 + 4[2\beta^2 - 1 + 3(2\pi\theta)^2]\}[A_1(\theta)]^4. \tag{2.44}$$

An alternative approach to the above 'smoothing' technique is to replace the negative exponential function by another function that is smooth at the origin but will behave like the negative exponential function for large t. One possible function is

$$\rho(t) = \operatorname{sech}(\tfrac{1}{2}\pi t). \tag{2.45}$$

Hence an alternative periodic autocorrelation function model is given by

$$\rho(t) = \operatorname{sech}[\tfrac{1}{2}\pi t A_2(\theta)]\cos[2\pi\theta t A_2(\theta)], \tag{2.46}$$

where
$$A_2(\theta) = \operatorname{sech}(2\pi\theta) + 2\sum_{r=0}^{\infty}\frac{(-1)^r\theta}{2\pi\{\theta^2 + [\frac{1}{4}(2r+1)]^2\}\sinh(\pi(2r+1)/8\theta)}. \tag{2.47}$$

This is called the 'model 2' autocorrelation function. It has the advantages over model 1 of having only one parameter, i.e. θ. Both models are representative of real surface autocorrelation functions: the former is more useful for single-point machining and the latter more useful for multiple-point machining and primary forming methods.

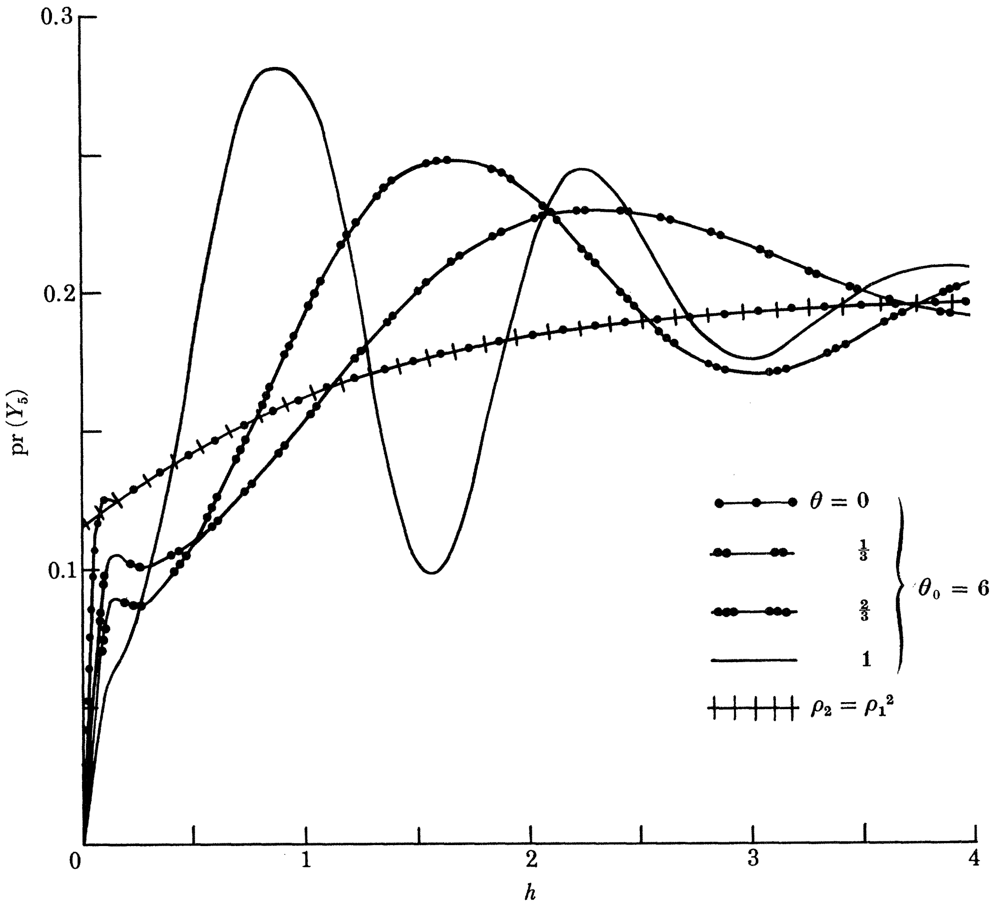

FIGURE 2. The variation of the expected density of five-point summits with the spacing h between ordinates. The exponential autocorrelation function and four examples of the model 1 autocorrelation function are shown, and the figure illustrates how all expected summit densities associated with model 1 reduce to zero at $h = 0$.

For the model 2 autocorrelation function the derivatives are given by

$$D_2 = -(\tfrac{1}{2}\pi)^2[1+(4\theta)^2][A_2(\theta)]^2, \tag{2.48}$$

and

$$D_4 = (\tfrac{1}{2}\pi)^4[5+6(4\theta)^4+(4\theta)^4](A_2(\theta)]^4. \tag{2.49}$$

Hence, from (2.26),

$$\eta = \{[1+(4\theta)^2]/[5+(4\theta)^2]\}^{\frac{1}{2}}. \tag{2.50}$$

The upper bound on η implies that

$$\theta < \sqrt{7}/4 = 0.6614. \tag{2.51}$$

The model 2 autocorrelation function will be used with $\theta = 0$, $\frac{1}{3}$ and $\frac{1}{2}$, which will satisfy the condition on η, given by (2.51).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

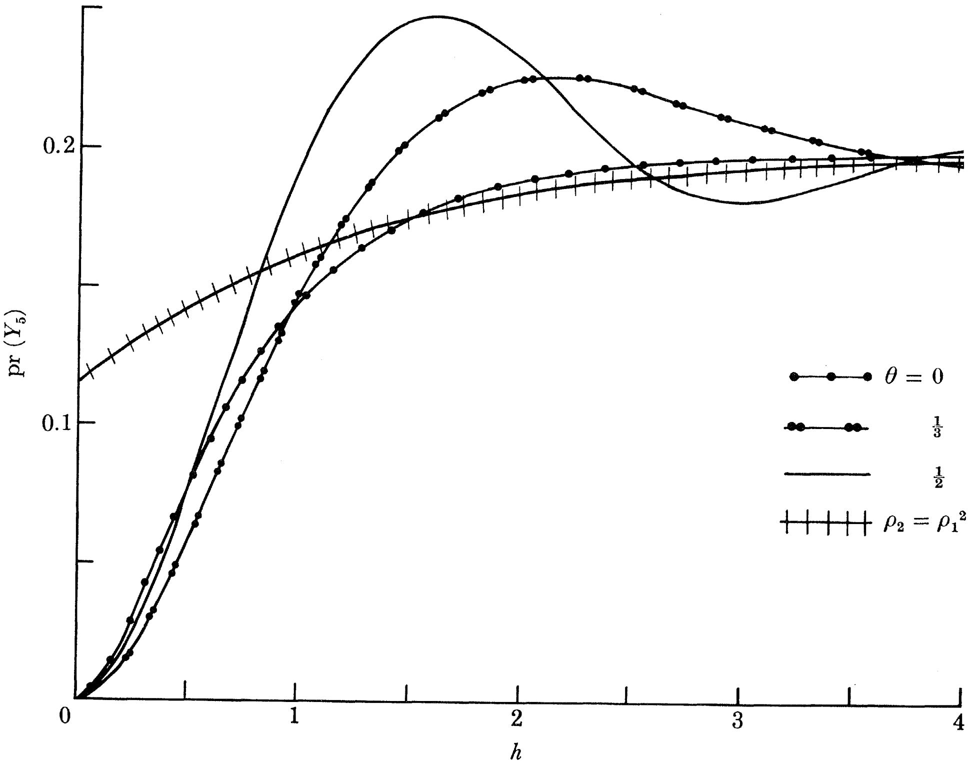

FIGURE 3. The variation of the expected density of five-point summits varies with the spacing h between ordinates. The exponential autocorrelation function and three examples of the model 2 autocorrelation function are shown, and the figure illustrates how all expected summit densities associated with model 2 reduce to zero at $h = 0$.

The two functions $A_1(\theta)$ and $A_2(\theta)$ are both equal to unity when θ is zero and both converge to $2/\pi$ as θ tends to infinity. For the positive values of θ chosen here the values of $A_1(\theta)$ and $A_2(\theta)$ are approximately $\frac{2}{3}$.

The results obtained by Whitehouse & Phillips (1978), for the distributions obtained for digital measurement in one dimension, exhibited periodic behaviour which depended on θ. This was because the distributions were functions of ρ_1 and ρ_2, which were rational functions of h. However, because the expectations that are to be considered in two dimensions depend also on the parameter $\rho_3 = \rho(\sqrt{2}\, h)$ this simple periodic behaviour does not hold.

The first two-dimensional parameter investigated, the expected five-point summit density, given by (2.19), is shown in figures 2 and 3 for the autocorrelation functions for models 1 and 2, respectively. In all cases the expected five-point summit density converges to 0.2 as h tends to infinity. As h approaches zero the expectation approaches zero as these two autocorrelation functions are 'smooth'. (In contrast the expectation converges to 0.125 for the negative exponential autocorrelation function.) As θ increases there are quite large oscillations in the value of the expectation for model 1. The frequency of the oscillations is about the same as that for the expected peak density for measurement in one dimension (given by Whitehouse & Phillips (1978)). The variation is of roughly the same order of magnitude as for peak density. For model 2 the variation is similar to model 1.

Secondly, the expected five-point summit height, given by (2.16), is shown in figures 4 and 5 for model 1 and model 2, respectively. In all cases as h tends to infinity the expectation

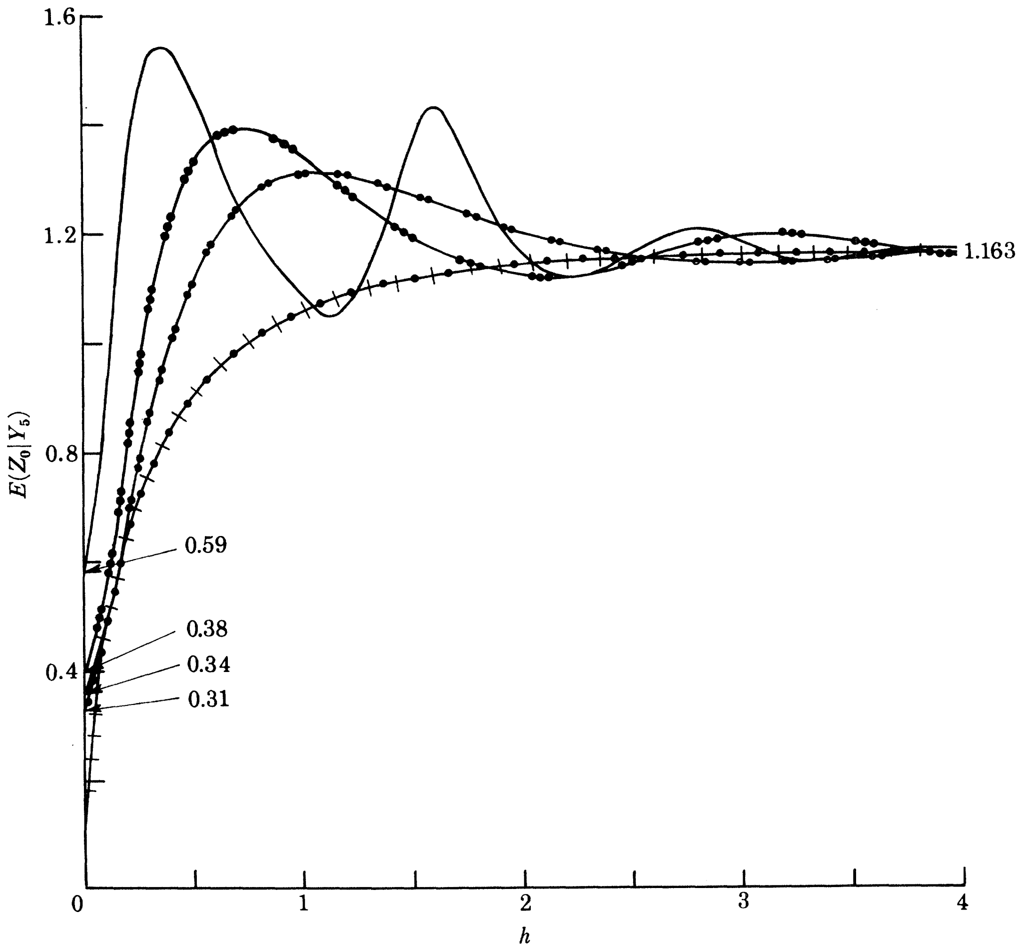

FIGURE 4. The variation of the expected five-point summit height with the spacing h between ordinates. The figure gives the limiting values at the origin for four examples of the model 1 autocorrelation function and the asymptotic value for h tending to infinity. For key see figure 2.

converges to $5[1-(\frac{3}{2}\pi)\arccos\frac{1}{3}]/\pi^{\frac{1}{2}} = 1.163$. As h approaches zero the expectation approaches the limit given by (2.27) for both models 1 and 2, as these two autocorrelation functions are 'smooth'. (In contrast the expectation converges to zero for the negative exponential autocorrelation function.) The limits are much smaller for model 1 than for model 2 as h approaches zero. The values of θ can be larger for model 1 than for model 2, while still satisfying the condition on η, given by (2.26). As θ increases there are quite large oscillations in the value of the expectation for model 1. However, the variation is of roughly the same relative order of magnitude as for the expected peak height. For model 2 the variation in the expectation is much less because of larger limiting values obtained as h approaches zero.

Thirdly, the expected five-point summit curvature given by (2.21), is shown on a logarithmic scale in figures 6 and 7 for model 1 and model 2 respectively. In all cases the expectation approaches zero as h tends to infinity. As h approaches zero the expectation converges to $1.125\,(\frac{1}{2}\pi D_4)^{\frac{1}{2}}$, for smooth autocorrelation functions. For model 1 there is hardly any variation as θ varies, for positive values, but for model 2 there is a more noticeable change. (In contrast the expectation tends to infinity for the negative exponential autocorrelation function.) For

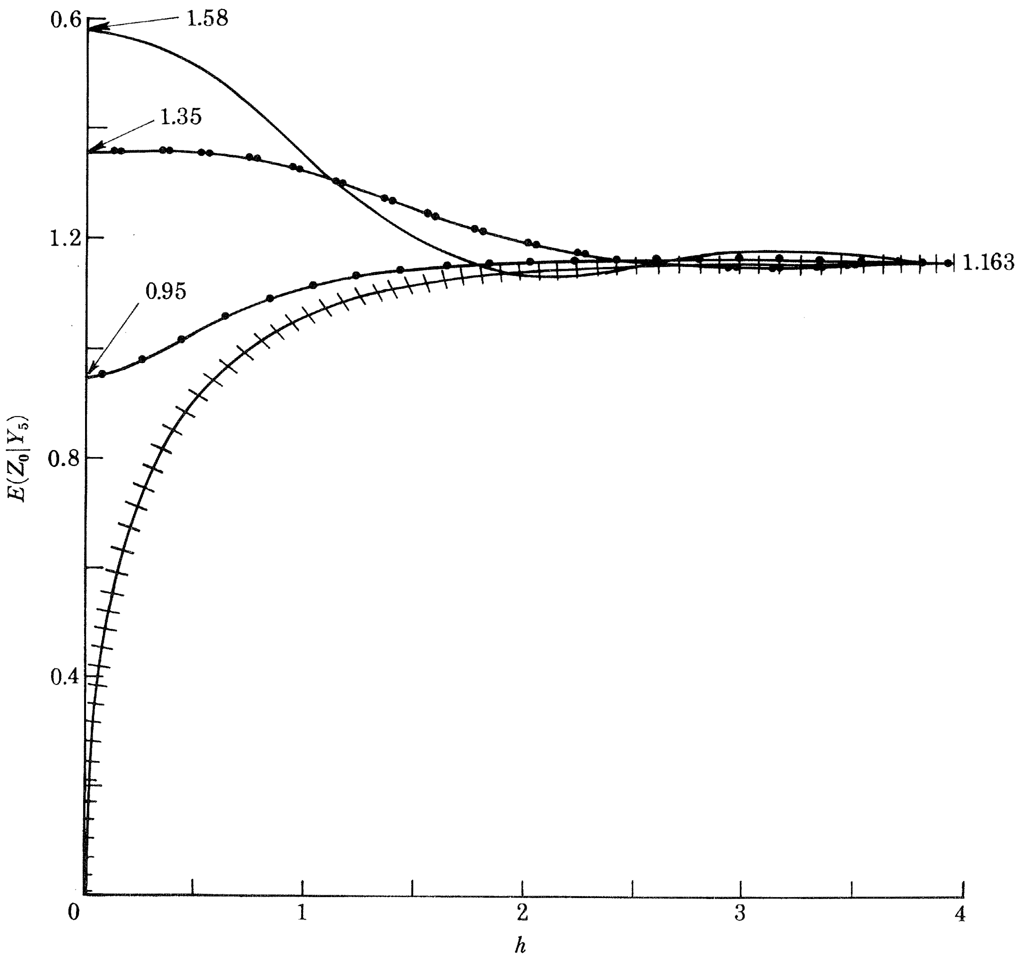

FIGURE 5. The variation of the expected five-point summit height with the spacing h between ordinates. The figure gives the limiting values at the origin for three examples of the model 2 autocorrelation function, and the asymptotic value for h tending to infinity. For key see figure 3.

model 1 there are quite large oscillations in the value of the expectation. Again the frequency of the oscillations is higher than that for the expected peak curvature for measurement in one dimension (given by Whitehouse & Phillips (1978)). However, the negative exponential autocorrelation function is no longer a lower bound on the expected five-point summit curvature as it was for the expected peak curvature. For model 2 the oscillations in the value of the expectation with θ are smaller, with the expectation following the general downward trend as h increases.

3. Practical verification of the theory

3.1. *Requirements of the apparatus*

To test experimentally the theoretical results of the discrete analysis in one dimension is comparatively easy provided that the stylus instrument has a uniform tracking speed, so that the spacing between ordinates is genuinely constant. Unfortunately, it is not so easy to test the theoretical results obtained for two dimensions because of the need to measure a grid of ordinates on the surface, in which the mean level of the tracks must be truly related by reference

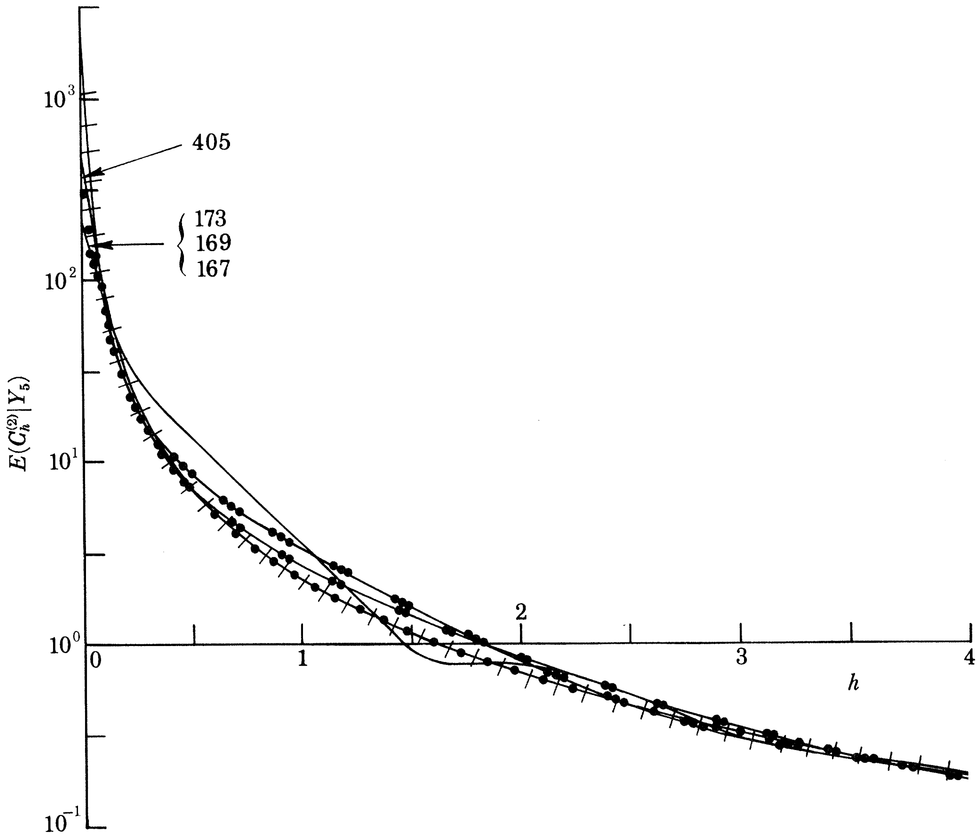

FIGURE 6. The variation of the expected five-point summit curvature with the spacing h between ordinates. The figure gives the limiting values at the origin for four examples of the model 1 autocorrelation function. The expectation is expressed on a logarithmic scale. For key see figure 2.

to an external datum. The mean levels cannot be arbitrary, otherwise the scanning of the data to find summits and their heights, and other parameters becomes meaningless. Ideally, the ordinate heights from each track should be related without vertical shift to those ordinates in any other track. In order that this severe constraint can be achieved, a special stage has been designed, as shown in simplified form in figure 8.

Incrementing the surface in small equal steps, to get a grid of tracks, is another difficult problem. Finally, the digital record of each track must start at the same position in space, at least transversally, relative to the traverse of the surface-measuring instrument.

3.2. *Description of the apparatus*

The apparatus used is shown in figure 9. It consists of a Talysurf 4 surface-measuring instrument manufactured by Rank Taylor Hobson of Leicester, U.K., with special stage and datum attachment, shown in figures 8 and 10. Attached to the Talysurf is a MINC-11 minicomputer made by the Digital Equipment Corporation of America. This in turn is connected to the PRIME computer of the Engineering Department of the University of Warwick. Between the stage and the Talysurf traversing unit is positioned a fibre optic connected to a Valpitronic compact relay unit, which is used to trigger the MINC computer in the same place for each track.

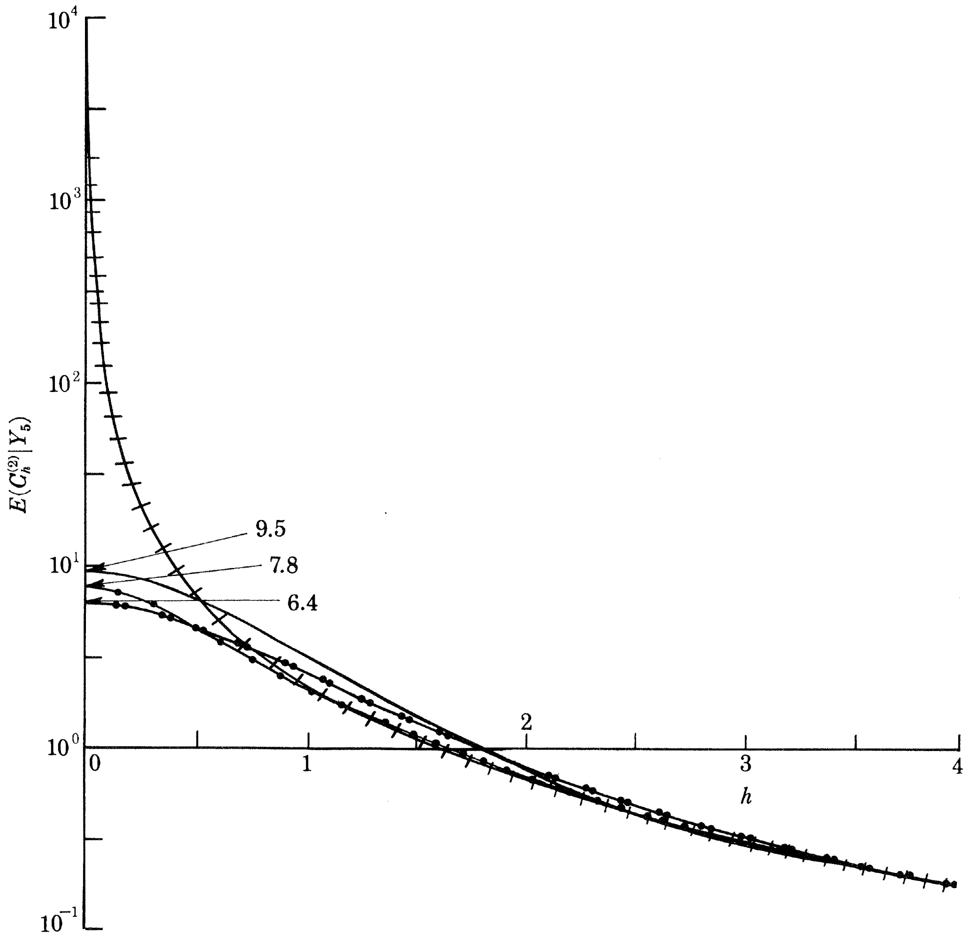

FIGURE 7. The variation of the expected five-point summit curvature with the spacing h between ordinates. The figure gives the limiting values at the origin for three examples of the model 2 autocorrelation function. The expectation is expressed on a logarithmic scale. For key see figure 3.

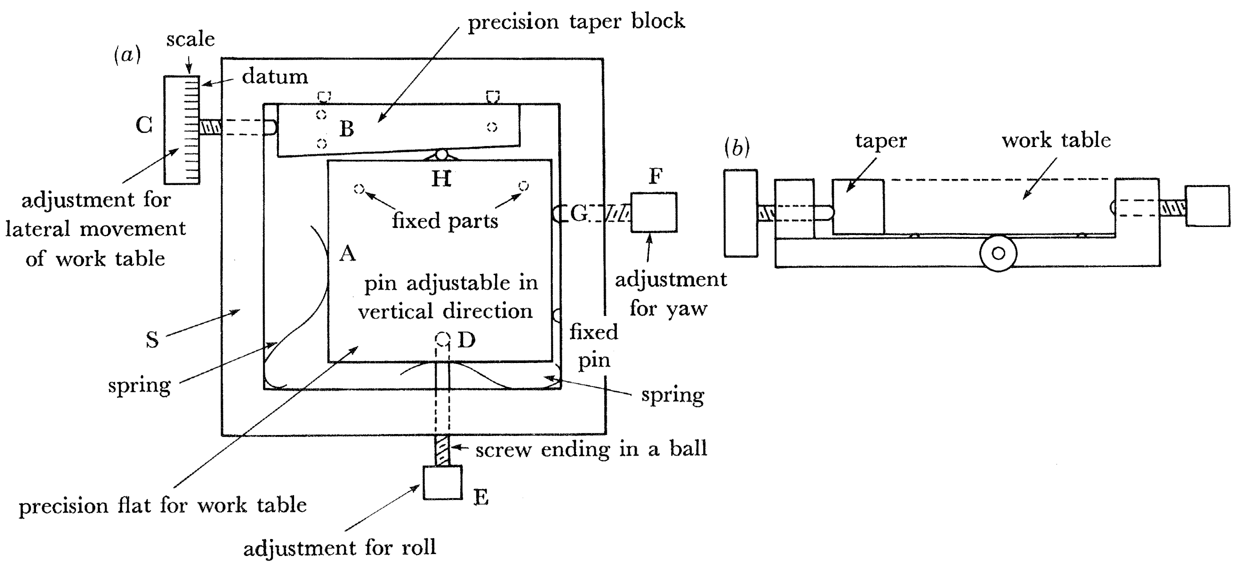

FIGURE 8. Kinematic arrangement and lateral movement: (a) plan view; (b) front elevation, sides removed.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

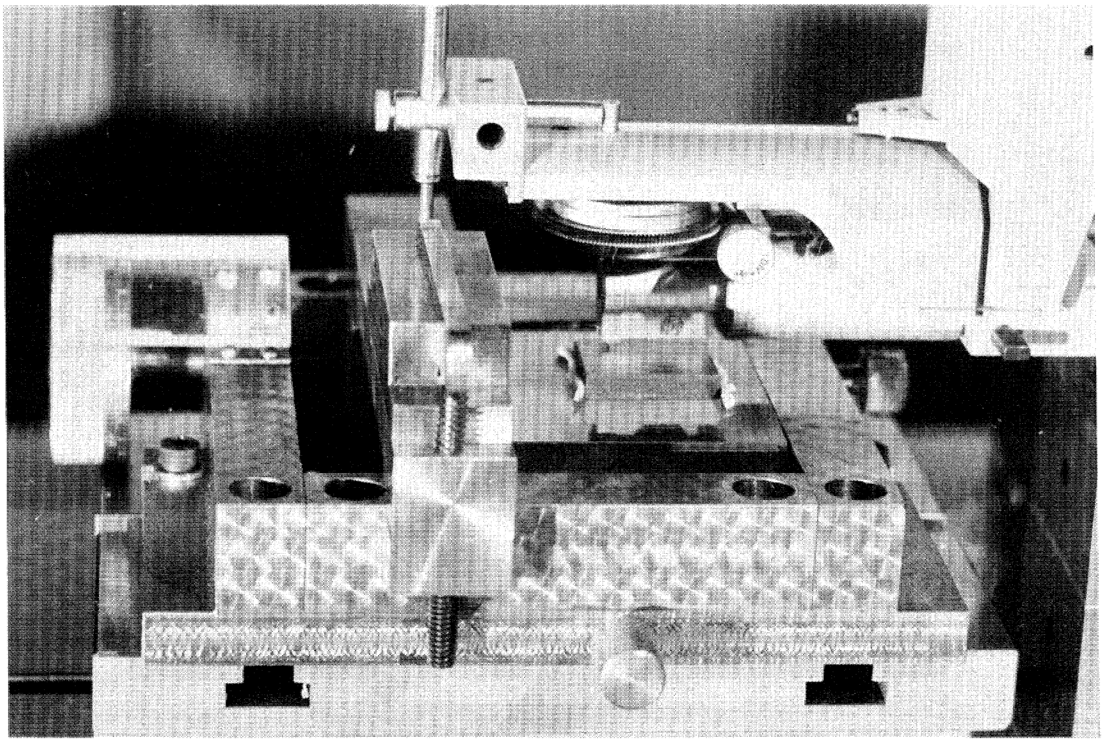

FIGURE 9. General view of apparatus.

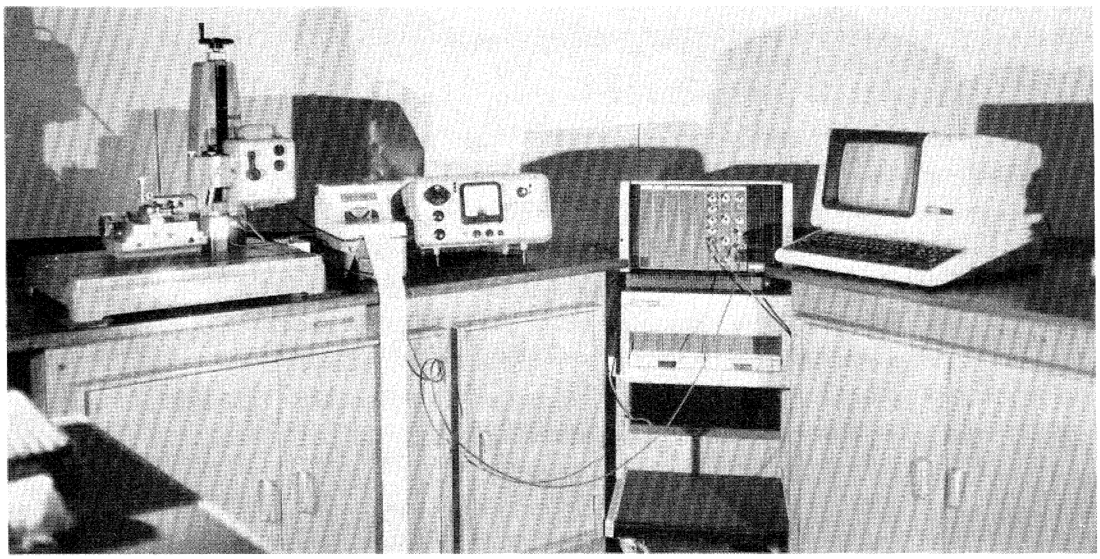

FIGURE 10. General view of precision worktable with curved datum attachment of stylus instrument in position.

The most demanding mechanical element in the set-up, apart of course from the Talysurf, is the stage shown in figure 8. This has been designed to provide the mechanical coherence required between tracks. Basically it comprises two precision glass blocks A and B. Both are constrained by five contacts allowing one degree of freedom. B is a 10:1 tapered block used for moving the block A via the ball H by small known increments. This is achieved by turning nut C. A turn through one scale unit on the periphery of C corresponds to a movement of A of 2.4 μm.

Although block A has five constraints imposed upon it by means of ball contacts, two of these are variable, via nuts E and F, to allow adjustments in the tilt of the specimen (which is placed upon A) in the roll and yaw angles. The roll adjustment includes the device of a ball-on-screw E pushing onto an offset ball in the vertical slot at D. The bottom face of A and the two edges have had to be especially manufactured, for squareness, straightness and smoothness, to ensure that no unwanted rotations are produced when the block is moved transversally relative to the direction of motion of the pick-up. Typical specifications for the block are straightness 0.1 μm per mm, squareness 2″, and smoothness 0.05 μm.

The other angular adjustment, pitch, is provided for by direct adjustment of the Talysurf datum, which is specifically designed to give this feature for levelling a profile graph. This is shown in figure 10. The accuracy of, and operation instructions for, this adjustment are well known and are included in the manufacturer's literature. Also shown in figure 10 is a device designed to lift the stylus just clear of the surface after every traverse. Basically it consists of a precise way of lifting the 'mechanical steady' of the datum attachment by such a small amount that the ligament hinges on the traversing unit are not strained when the Talysurf gearbox is racked back. A clearance of just 500 μm is maintained.

3.3. *Checks and procedures*

Standard tests were made on the apparatus to check whether or not the minimum specification was being met. The first test was repeatability. A number of traverses in the same position were digitized and analysed. This tested three things: the vertical mechanical coherence of the Talysurf, the traversing speed, and the repeatability of the digital triggering of the MINC. (Each of these factors is critical in verification of the theory. Vertical repeatability within a track has to be considerably less than the texture height. The repeatability was better than 2% of the surface texture, which turned out to be about 0.02 μm.) Relatively rough surfaces were used in the verification because the instrumental specification, both in terms of vertical precision and digital triggering, could be relaxed, as the errors were a correspondingly smaller fraction of the dimensions involved. The sampling interval was fixed at 2.4 μm, this being the nominal stylus-tip dimension in the direction of traverse. Successive repeat traverses were cross correlated to establish whether the digital initiation signal from the fibre optic was satisfactory. This means that the triggering error should be much less than the spacing between samples, i.e. 2.4 μm. In the event, this was found to be an impossible criterion, because of the lack of repeatability of the light levels needed to trigger the fibre optic relay and the physical diameter of the fibre pack itself, which was 2 mm. Tests with the use of cross correlation showed that errors of the order of 5 μm or more could be expected in synchronization. For this reason, when building up the grid of ordinates, each track was repeated three times and averaged to reduce the effect of this spatial discrepancy.

Tests were made on the line of traverse of the stylus by lightly carbonizing a microscope slide and tracking the stylus repeatedly over the same place on the slide. An overall track width of twice the diamond width was encountered after a large number of repeats. This conformed with the manufacturer's specification.

3.4. *Tests on a manufactured surface*

A typical surface manufactured by electro-discharge machining was used for a practical test. An example of a profile graph is shown in figure 11. Topographic data for the surface include an arithmetic average (R_a)-value of 3 μm. An isotropic surface was chosen for simplicity of instrumentation.

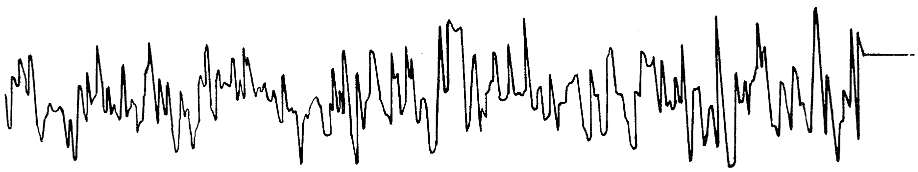

FIGURE 11. Typical engineering surface (one-dimensional profile), manufactured by electro-discharge machining. Vertical magnification 1000; horizontal magnification 20; $R_a = 3$ μm.

Because of the problems of synchronization the spacing between tracks was fixed at 7.2 μm and the sampling interval at 2.4 μm. Seventy-five tracks were taken over the surface in 25 different places, each track being repeated three times and averaged. The amount of data for each traverse was 2099 digital ordinates with a resolution of ten bits. This amounted to a total of over 150000 digital measurements. The data was relayed via the MINC to the PRIME computer for processing. Before the measurements were taken the total extent of the surface to be investigated was scrutinized for deviations in level. The adjustments E, F and the datum were altered to give a best fit plane by eye. Because the surface was rough and its manufacture well controlled only small adjustments were found to be necessary. The use of digital filtering or regression techniques for removing slopes was rejected, because of the possibility of distorting the height of the true mean line from one track to the next.

To test the theory three parameters were chosen for comparison, the first relatively insensitive to sampling, the second one of the most sensitive and the third about average. These were the expected five-point summit density, the expected five-point summit height and the expected five-point summit curvature, respectively.

The nature of the comparison between theory and practice was as follows. The mean five-point density, height and curvature for a variety of ordinates and track spacings were observed by averaging over the 25 tracks of the two-dimensional grid of data obtained from the surface. The values so obtained were then compared with the expected values predicted from the formulae derived in §2.2, with estimated values of ρ_1, ρ_2 and ρ_3 obtained from the practical (observed) autocorrelation function shown in figure 12. For example, if the observed mean summit parameters were obtained from the surface data grid corresponding to a 7.5 μm spacing between tracks then the values of the autocorrelation function would be taken from figure 12 at values of 7.5 μm, 15 μm, and $7.5\sqrt{2}$ μm.

The results of the comparison are shown in figures 13, 14 and 15 for the expected five-point summit density, the expected five-point summit height and the expected five-point summit curvature respectively.

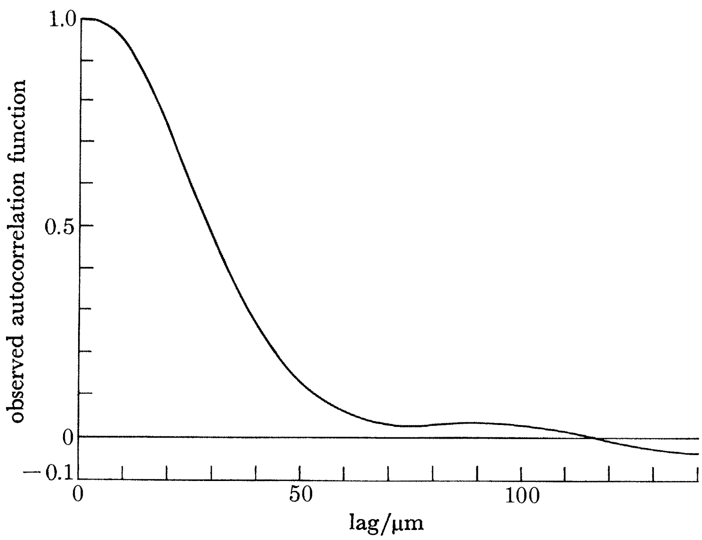

FIGURE 12. The observed autocorrelation function of the electro-discharge-machined surface. The surface was measured at intervals of 2.4 μm over 25 tracks.

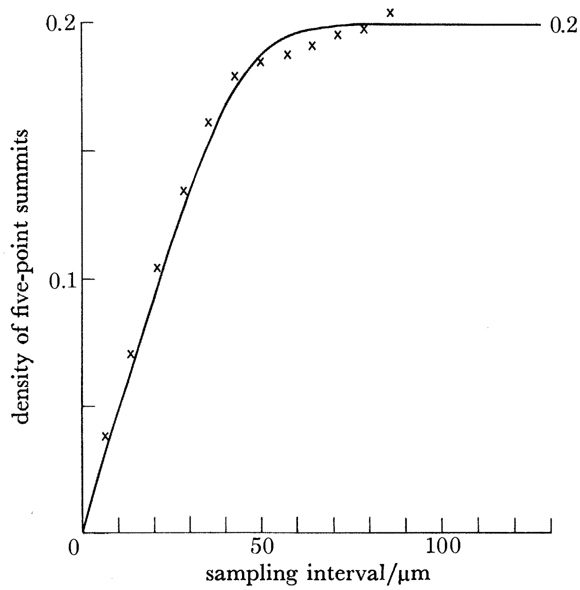

FIGURE 13. The observed mean density of five-point summits, over 25 tracks, for sampling intervals from 7.2 to 86.4 μm, and the expected density of five-point summits obtained from the observed autocorrelation function.

3.5. *Discussion of test results*

The results show a significant degree of agreement between the sample means obtained from the experimental data and the theoretical formulae for the expectations, with ρ-values estimated from the measured autocorrelation function. So it seems plausible to assume that the theory and practice do agree. Any small deviations can be attributed to discrepancies between the surface used and a truly Gaussian one. For this particular surface there is a general shift in both figures 14 and 15 to indicate that the actual summits are sharper and higher than those predicted by the theory. This can be attributed to the fact that the test surface has a negative

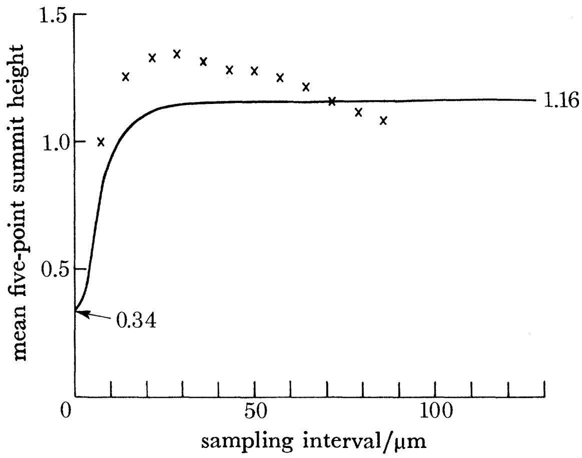

FIGURE 14. The observed mean five-point summit height, over 25 tracks, for sampling intervals from 7.2 to 86.4 μm, and the expected five-point summit height obtained from the observed autocorrelation function. The height is non-dimensionalized by dividing it by the estimated standard deviation of the surface heights.

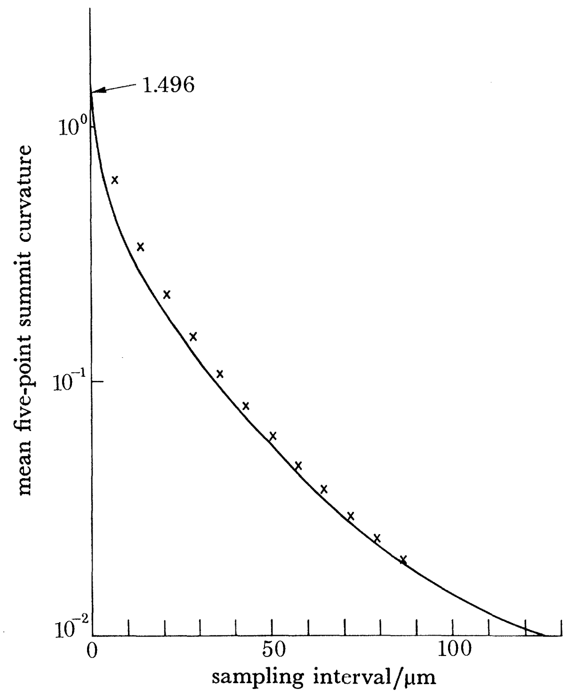

FIGURE 15. The observed mean five-point summit curvature, over 25 tracks, for sampling intervals from 7.2 to 86.4 μm, and the expected five-point summit curvature obtained from the observed autocorrelation function. The curvature is expressed on a logarithmic scale.

skew of 0.8; a property of electro-discharge machined surfaces. This has the effect of slightly lowering the mean line of the surface relative to the symmetrical one expected for a Gaussian surface, which will give the observed result.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

4. Discussion

It having been established with a real surface that the theory does to within a high level of confidence agree with practice in the two-dimensional situation, the question arises as to the practical implications of the theory.

One important point is the fact that the discrete parameters for two-dimensional asperities such as summit height, density and curvature have different values to the corresponding parameters obtained by using one-dimensional discrete theory. This is not an unexpected result in view of the fact that there are differences in the equivalent continuous case. For example, the expected summit height is higher than the expected peak height by about 80% in continuous theory, whereas in the discrete theory it is higher by 69%. The difference between the continuous and discrete theories arises because it is difficult to develop a numerical model that covers all two-dimensional summit possibilities. The reassuring fact to be revealed by this exercise is that by using the *simplest* numerical model the differences between the continuous and discrete theories are only small: a 6% overall decrease for the expected summit height and curvature is about the difference involved. This is not likely to be of practical significance. What are significant are the very big differences between the one- and two-dimensional estimates of parameters: any experimenter who thinks that the measurements taken from a profile of a surface actually estimate the real parameter values will be seriously in error. Clearly, if a nine-point model for a summit had been used rather than the five-point model the differences between the continuous and discrete theories would have been smaller.

Another question of importance concerns the scale of size of asperities and how they appear in a discrete analysis. This is important because asperities of different size have different *functional* significance. It may be considered easy to investigate the properties of asperities on the surface having fixed size simply by choosing the relevant ordinate spacing. For large spacing the ordinates would be uncorrelated, and the asperities picked out would correspond to the large-scale structure, whereas a small spacing would pick out a small-scale structure. Unfortunately this simple approach is prone to misinterpretation: considerable care has to be taken when using the results obtained from such measurements, for it is well known that aliasing effects can occur when data are sampled. Normally for maxima these effects are not too serious. However, in the digital measurement of surfaces, aliasing effects can be accentuated by an 'interlocking' produced between the sampling interval and the numerical model used to define a peak and its parameters. This is similar to the errors introduced by aliasing in the measurement of the spectral density by digital sampling. For one-dimensional parameters they can be difficult to unravel (Whitehouse & Phillips 1978). However, for two-dimensional parameters they is not. To illustrate this, a comparison will be made between the expected five-point summit density and the model 1 autocorrelation function for various ordinate spacings. Consider the expected five-point summit density for a highly oscillatory surface as shown in figure 16. For small values of h the result will be masked by instrumental effects such as stylus integration, the limiting optical resolution of the instrument, or more likely quantization effects. This region is marked a in the figure. For spacings up to about one half of the correlation length the results will be valid for the surface (region b). If the ordinate spacing corresponds to one half of the period of oscillation (as revealed by the autocorrelation function) then the numerical model will ensure that the dominant frequency will determine the value of the parameter, in this case, the density. It would take a value of almost 0.5. For integral multiples of this spacing the numerical model will always lock itself on to the dominant wavelength on the surface, which will cause repeated maxima. These maxima will be of decreasing height as h increases because of the decrease in correlation between the ordinates. It is in this region, where the spacing is comparable with the correlation length, that the interlocking effect is so marked (region c). Results obtained in this region should be treated with caution. As the spacing between ordinates increases much beyond this value, the behaviour becomes dominated by the numerical model itself (region d). For expected *peak* density in which a *peak* is defined by a three-point numerical model the probability of an ordinate being a peak will inevitably tend towards a value of one third

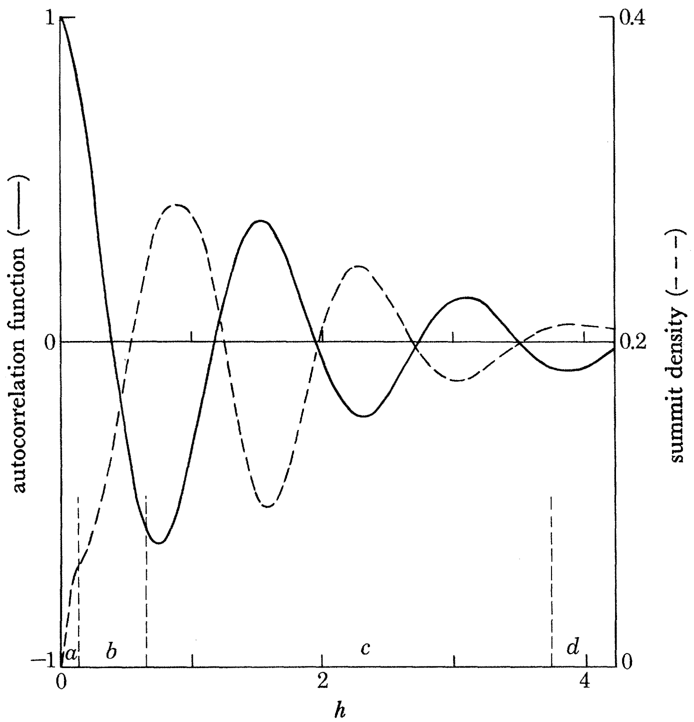

FIGURE 16. The relation between the expected five-point summit density (– – –) and the autocorrelation function for an oscillatory surface (——) as the spacing between ordinates h varies.

irrespective of the type of surface. For expected *summit* density where the *summit* is defined by five points the limiting value is 0.2. Other parameters exhibit similar behaviour. That there is a definite relation between the parameters and the autocorrelation function can be seen on figure 16. The expected summit density is very nearly the mirror image of the autocorrelation function.

This type of pattern is not so clearly shown in the profile parameters as it is in the two-dimensional parameters. For example, when the spacing is equal to the dominant wavelength, the numerical model in effect does not see the wave, and the peak parameters resort to those of the autocorrelation envelope which, for the two surface models used, is an exponential. Summit parameters, however, do not behave in this way. This can be seen in figure 2 for $\theta = 1$. The density curve only converges to the exponential when h is large and both curves reflect the numerical model. From these considerations, it seems clear that the two-dimensional assessment makes the curves of parameter value against sampling interval much better behaved and more representative of the true surface values.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

One common feature of both the one-dimensional and two-dimensional results is that the magnitude of the variation of a parameter with h is about the same. A factor of variation of three or even four to one is possible for expected peak and summit density and curvature. Hence it is very important for an investigation that a value of h be confined to the useful region where it is less than about one half the autocorrelation length. The difficulty is that of picking h such that it falls in this region, and at the same time is relevant to the function of the surface. In picking h to satisfy both these conditions a knowledge of the correlation length is desirable, a point that will be addressed later.

Detailed discussion of the behaviour of these parameter curves as a function of h is given in §§2.3 and 2.4. However, two points are pertinent. One is that the values of all parameters for all surfaces vary considerably even in the useful and unambiguous region where h is less than one half the autocorrelation length, a point that has considerable functional importance. Another is that the limiting values as h tends to zero are different for different surfaces. These differences will be difficult to measure if stylus radii are made much larger than at present.

The property of the parameter values for two-dimensional surface detail having been established, the question arises as to whether it is absolutely necessary to resort to measuring the whole surface grid of ordinates or whether a simpler method could be devised. One point that should be made regarding two-dimensional measurement concerns coverage. In the same way that the discrete analysis does not converge to the continuous, it is impossible to cover a surface with a sufficient number of measurements to ensure that any freak or non-typical behaviour will be seen. Stylus techniques are not really suited for such work. Stylus methods are suitable for quantifying average behaviour of surfaces. For this reason it seems plausible to suggest that, in so far as the autocorrelation function is a good measure of the average statistical behaviour of a surface, it could form the basis for a simple method for estimating other average surface properties. The work reported here suggests that this is indeed possible. Furthermore it has emerged from the experimental part of this investigation just how difficult it is to measure a grid of digital data points on a surface. It has been found to be difficult even with unconventional equipment. The dual problem of maintaining a common mean level of tracks and preserving sampling synchronism between tracks has proved to be a demanding task, and it would be unreasonable to expect an investigater doing research into functional properties to have the time or the effort to achieve this. It seems, therefore, that a useful side benefit of this theoretical investigation has been the devising of a practical way of measuring the two-dimensional average properties of surfaces. In fact, it may well be that this technique is the only practical way of measuring two-dimensional parameters without highly specialized equipment.

For an isotropic surface one profile track is needed from which the autocorrelation function can be estimated. Then the r.m.s. value of the surface height and the values of $\rho(h)$, $\rho(2h)$ and $\rho(\surd 2h)$ are found, and from these the parameter values. The problem is more complicated when the surface is neither isotropic nor unidirectional, but somewhere in between. The first step is to determine the degree of isotropy. One simple pilot method is to take two profile traces at right angles and to estimate the peak densities by eye. If these are $d_{p,1}$ and $d_{p,2}$ respectively a third track could be taken at an angle of arctan $(d_{p,2}/d_{p,1})$ to the first direction, and the value of $d_{p,3}$ obtained. If the $d_{p,i}$ are all equal then the surface is isotropic. If they are different then the horizontal scales in the different directions can be condensed to make the surface appear to be isotropic if this is permissible functionally. The ratio $d_{p,2}/d_{p,1}$ could be used to do this. If the ratio is $1/k$ and we make the assumption, which is not always justified, that the autocorrelation function in different directions differs only by a scale factor in the horizontal direction, then the autocorrelation values corresponding to $\rho(h)$, $\rho(2h)$, $\rho(kh)$, $\rho(2kh)$ and $\rho(h(1+k^2)^{\frac{1}{2}})$ can be evaluated directly from the autocorrelation function of the one-dimensional trace which has been taken across the lay. If k is very large the problem resolves itself into the one-dimensional method reported elsewhere (Whitehouse & Phillips 1978). If it is reasonably small, say 1–4, the values of $\rho(h)$ and $\rho(kh)$ can be averaged to give ρ_1 and so on without too large a deviation from the theoretical results.

A full theoretical treatment would suggest that to take into account all possibilities seven values would be needed. That is the r.m.s. value of the surface; $\rho(h_1)$, $\rho(2h_1)$ in one direction; $\rho(h_2)$, $\rho(2h_2)$ in the orthogonal direction; and the two diagonal correlation values between the directions, namely the non-isotropic versions of ρ_3. In practice because of the symmetry of generation of surfaces, six or even four values will suffice.

One problem that needs to be resolved before the formal digitization of a surface is the assessment of the correlation length. Once this is known it is possible to verify the validity of the parameters. Some idea can be obtained from the zero crossing density of a profile graph after any general trend has been removed by eye. There is a rough relation between the average spacings between up-crossings and the correlation length which is only weakly dependent on the type of surface. A simple rule of thumb is that they are almost equal. Further work is being done to refine this simple method. One possibility is to use the ratio of peak to crossing densities as a way of identifying different types of surface. From this a more exact multiplier of the crossing density could be looked up to give the autocorrelation length.

The only disadvantage of this correlation method of assessing two-dimensional parameters is that there has to be some restriction on the amount by which the surface can be allowed to be non-Gaussian. Current tests indicate that the theoretical results will be valid provided that the skew value is within ± 1. Certainly the practical advantages of the method outweigh the small theoretical restrictions.

In summary, this paper represents a sequel to the two earlier papers in this field (Whitehouse & Archard 1970, Whitehouse & Phillips 1978). It is now possible to relate the two-dimensional digital parameters that would be measured from a surface to the results calculated from continuous and discrete theory. It has been shown that differences must occur in practice, but now that the nature and extent of these differences are known they can be taken into account.

The most important points to emerge from this paper are:

(i) There are likely to be large differences between the discrete properties of different surfaces.

(ii) There is a significant difference between parameters measured digitally from a one-dimensional profile of a surface and those measured from a two-dimensional grid.

(iii) There is a considerable variation in parameter estimates as the scale of size is changed owing largely to a complex interaction between the numerical model and the sampling. Various limits have been laid down to reduce the effect of these variations. It will be necessary in the future to tighten up the control of the digital analysis of surfaces now being lightly undertaken by many manufacturers.

(iv) It is now possible by using a small number of correlation values to evaluate the average values and dispersions of summit parameters of a surface without our having to embark on very elaborate and difficult experimentation.

What remains to be done is to assess the properties of the gap between two surfaces by using the techniques described above. The advantages of using correlation techniques for this is that the correlation values for the two surfaces can be added directly together.

As in the other papers in this series a major point to emerge is the power of the combined ability of random process analysis and numerical techniques to quantify and describe parameter behaviour. It may be that the next major step in the theory of surface description and behaviour will be the development of statistics that more nearly mirror the way in which surfaces are generated and the way in which they behave in different engineering applications.

Finally, we would like to acknowledge the help of Dr J. Raja of the Department of Engineering, University of Warwick, in collecting the data from the surface, and of the Research Board of the University of Leicester in awarding a grant to M. J. Phillips.

Appendix

Results are given for truncated Gaussian random variables. There is a constant correlation coefficient between the random variable Z_0 and each of the m random variables $\boldsymbol{X}$, and the correlation matrix of $\boldsymbol{X}$ has a constant row (or column) sum. So

$$\begin{bmatrix} Z_0 \\ \boldsymbol{X} \end{bmatrix} \sim N\left[\boldsymbol{0}, \begin{pmatrix} 1 & d\boldsymbol{1}' \\ d\boldsymbol{1} & \boldsymbol{V} \end{pmatrix}\right], \tag{A 1}$$

where $\boldsymbol{V}$ is a correlation matrix such that

$$\boldsymbol{V}\boldsymbol{1} = \lambda\boldsymbol{1}, \tag{A 2}$$

where $\lambda > md^2$ (and $\boldsymbol{1}' = (1, 1, \ldots, 1)$). From (A 1)

$$U \equiv \boldsymbol{1}'\boldsymbol{X} \sim N[0, m\lambda], \tag{A 3}$$

$$(Z_0|\boldsymbol{X} = \boldsymbol{x}) \sim (Z_0|U = u)$$

$$\sim N[(d/\lambda)\,u,\ 1 - md^2/\lambda] \tag{A 4}$$

and

$$(\boldsymbol{X}|Z = z_0) \sim N[dz_0\boldsymbol{1},\ \boldsymbol{V} - d^2\boldsymbol{J}] \tag{A 5}$$

(where $\boldsymbol{J} = \boldsymbol{1}\boldsymbol{1}'$). So the conditional distribution of Z_0 given $\boldsymbol{X}$ only depends on $\boldsymbol{X}$ through $U\,(=\boldsymbol{1}'\boldsymbol{X})$.

Let $\boldsymbol{X}$ be truncated below, at $\boldsymbol{0}$, and the event $\{\boldsymbol{X} > \boldsymbol{0}\}$ be denoted by Υ. Then

$$\text{pr}\,(\Upsilon) = \Phi^{(m)}(0;\ \boldsymbol{V}), \tag{A 6}$$

where

$$\Phi^{(m)}(\boldsymbol{y}';\ \boldsymbol{V}) = \int_{-\infty}^{y_1}\int_{-\infty}^{y_2}\cdots\int_{-\infty}^{y_m} \phi^{(m)}(\boldsymbol{x}';\ \boldsymbol{V})\,\mathrm{d}x_1\,\mathrm{d}x_2\ldots\mathrm{d}x_m \tag{A 7}$$

(with $\Phi^{(m)}(d\boldsymbol{1}';\ \boldsymbol{V}) \equiv \Phi^{(m)}(d;\ \boldsymbol{V})$, $\Phi^{(1)}(y;\ V) \equiv \Phi(y/V^{\frac{1}{2}})$ and $\phi^{(1)}(y;\ V) \equiv \phi(y/V^{\frac{1}{2}})/V^{\frac{1}{2}}$). Now as Z_0 is not truncated,

$$(Z_0|\boldsymbol{X} = \boldsymbol{x}) \sim (Z_0|u), \tag{A 8}$$

as given by (A 4), while the probability density function of $(\boldsymbol{X}|\Upsilon)$ is given by

$$f(\boldsymbol{x}|\Upsilon) = \phi^{(m)}(\boldsymbol{x}';\ \boldsymbol{V})/\text{pr}\,(\Upsilon) \tag{A 9}$$

and, from (A 5),

$$\text{pr}\,(\Upsilon|Z = z_0) = \Phi^{(m)}(dz_0\boldsymbol{1}';\ \boldsymbol{V} - d^2\boldsymbol{J}). \tag{A 10}$$

Hence the probability density function of Z_0 given Y is given by

$$f(z_0|Y) = [\mathrm{pr}\,(Y|z_0)/\mathrm{pr}\,(Y)]\,\phi(z_0), \tag{A 11}$$

with the use of (A 9) and (A 10). Therefore from (A 4) and (A 8)

$$\begin{aligned} E(Z_0|Y) &= (d/\lambda)\,E(U|Y) \\ &= md\phi(0)\,\Phi^{(m-1)}(0;\,\boldsymbol{B})/\mathrm{pr}\,(Y), \end{aligned} \tag{A 12}$$

as Tallis (1961) has shown that using (A 9)

$$\mathrm{pr}\,(Y)\,E(\boldsymbol{X}|Y) = \lambda\phi(0)\,\Phi^{(m)-1}(0;\,\boldsymbol{B})\,\mathbf{1}, \tag{A 13}$$

where $\boldsymbol{B}$ is the variance–covariance matrix of $(X_1, X_2, \ldots, X_{m-1})$ given X_m. Hence from (A 12) it is seen that the expectation of the non-truncated random variable Z_0 given $\boldsymbol{X}$ (which is truncated) equals the product of m, the number of truncated random variables, d, the common correlation coefficient between Z_0 and any of the truncated random variables, $\phi(0)$, and the ratio of two orthant probabilities, one of which is for $m-1$ and the other for m random variables.

The probability density function of the conditional distribution of U given Y is given by

$$f(u|Y) = [\mathrm{pr}\,(Y|u)/\mathrm{pr}\,(Y)]\,f(u), \tag{A 14}$$

where $f(u)$ is the probability density function of the untruncated distribution given by (A 3). Because of the result given by (A 8) the probability density functions of both the distributions of $(Z_0, U|Y)$ and of $(U|Z_0, Y)$ are given by the product of pr $(Y|u)$ and the respective probability density function of the untruncated distribution, suitably normalized. So they are of the same form as (A 14).

References

Bowden, F. P. & Tabor, D. 1954 *The function and lubrication of solids*. Oxford University Press.
Cheng, M. C. 1969 The orthant probabilities of four Gaussian variates. *Ann. math. Statist.* **40**, 152–161.
Dawson, P. H. 1962 Effect of metallic cutting on pitting. *J. mech. Engng Sci.* **4**, 16–21.
David, F. 1953 A note on the evaluation of the multivariate normal integral. *Biometrika* **40**, 458–459.
Dyson, A. 1976 The failure of elastohydrodynamic lubrication of circumferentially ground discs. *Proc. Instn mech. Engrs.* **190** 699–711.
Fuller, K. N. G. & Tabor, D. 1975 The effect of surface roughness on the adhesion of elastic solids. *Proc. R. Soc. Lond.* A **345**, 327–342.
Greenwood, J. A. & Williamson, J. B. P. 1966 Contact of nominally flat surfaces. *Proc. R. Soc. Lond.* A **295**, 300–319.
Gupta, P. K. & Cook, N. H. 1972 Statistical analysis of mechanical interaction of rough surfaces. *Trans. Am. Soc. mech. Engrs J. Lubr. Technol.* **94**, 19–26.
Hamed, S. & Whitehouse, D. J. 1979 Tribology parameter prediction. *Wear* **57**, 323–329.
Hirst, W. & Hollander, A. E. 1974 Surface finish and damage in sliding. *Proc. R. Soc. Lond.* A **337**, 379–394.
Jones, A. M., O'Callaghan, P. W. & Probert, S. D. 1975 Prediction of contact parameters from the topographies of contacting surfaces. *Wear* **31**, 89–107.
Lewin, L. 1958 *Dilogarithms and associated functions*. London: Macdonald.
Mikic, B. B. 1974 Thermal contact conductance; theoretical considerations. *Int. Heat Transfer* **17**, 205–214.
Nayak, P. R. 1971 Random process model of rough surfaces. *Trans. Am. Soc. mech. Engrs J. Lubr. Technol.* **93**, 398–407.
Plackett, R. L. 1954 A reduction formula for normal multivariate integrals. *Biometrika* **41**, 351–360.
Pullen, J. & Williamson, J. P. B. 1972 On the plastic contact of rough surfaces. *Proc. R. Soc. Lond.* A **327**, 159–173.
Sayles, R. S. & Thomas, T. R. 1977 Measurement of the statistical microgeometry of engineering surfaces. *Proc. 1st Joint Polytechnic Symposium on Manufacturing Engineering, Leicester*. Leicester: Leicester Polytechnic.
Sokolnikoff, I. S. 1951 *Tensor analysis – theory and applications*. New York: Wiley.
Suh, N. P. 1973 Delamination theory of wear. *Wear* **25**, 111–124.
Tabor, D. 1975 A simplified account of surface topography and contact between solids. *Wear* **32**, 269–271.

Talliis, G. M. 1961 The moment generating functions of the truncated multinormal distribution. *J. R. statist. Soc.* B **23**, 223–229.

Whitehouse, D. J. 1978 The digital measurement of peaks on surface profiles. *J. mech. Engng. Sci.* **20**, 221–227.

Whitehouse, D. J. & Archard, J. F. 1970 The properties of random surfaces of significance in their contact. *Proc. R. Soc. Lond.* A **316**, 97–121.

Whitehouse, D. J. & Phillips, M. J. 1978 Discrete properties of random surfaces. *Phil. Trans. R. Soc. Lond.* A **290**, 267–298.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
THE ROYAL SOCIETY
PHILOSOPHICAL TRANSACTIONS OF

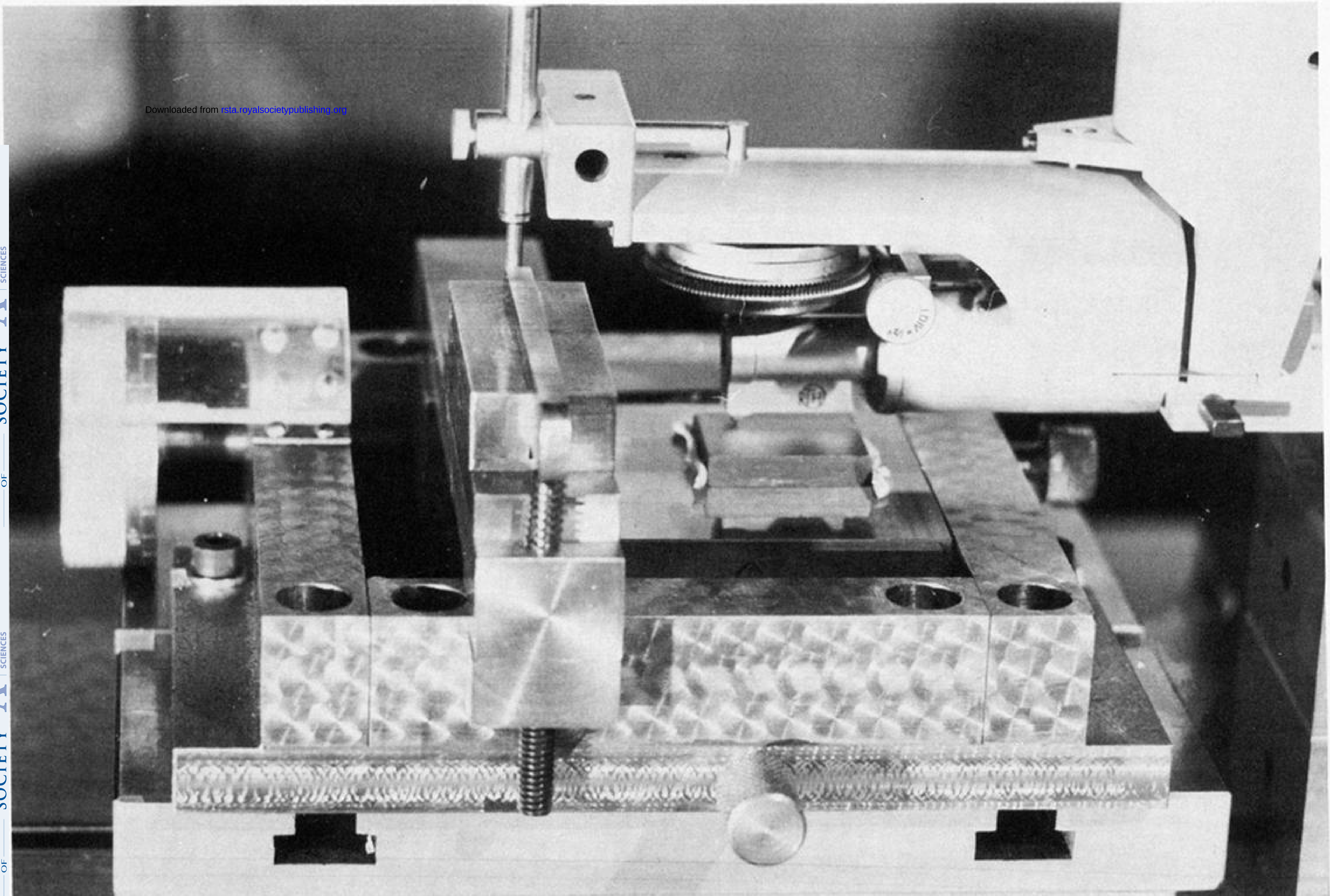

FIGURE 9. General view of apparatus.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

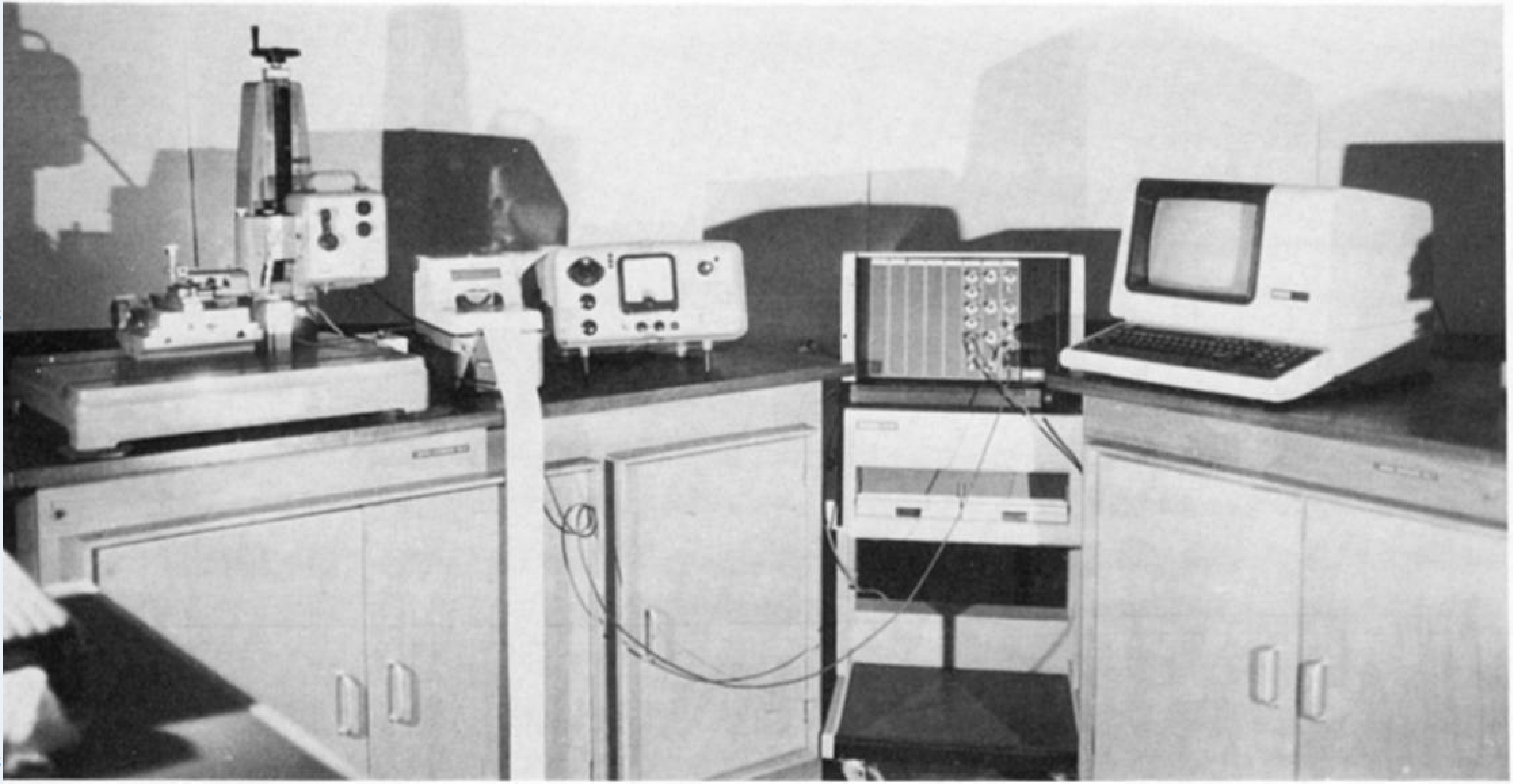

Figure 10. General view of precision worktable with curved datum attachment of stylus instrument in position.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES