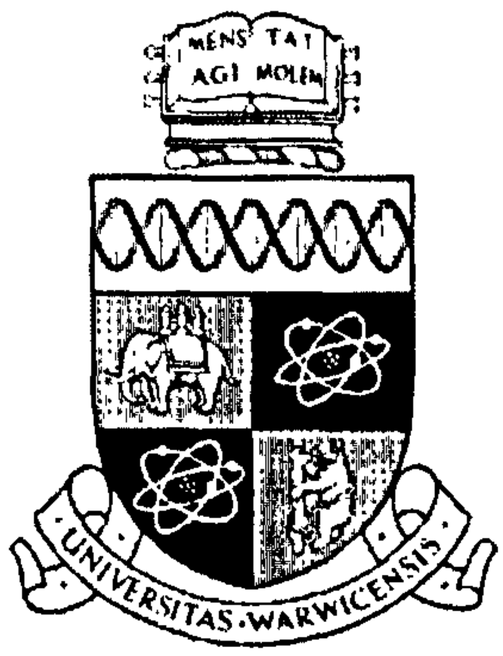
A Unified Methodology for the Application of Surface Metrology

By

Mark C. Malburg

Submitted for the degree Doctor of Philosophy

to the Higher Degrees Committee, University of Warwick



Centre for Nanotechnology and Microengineering

Department of Engineering, University of Warwick, Coventry U.K.

in collaboration with:

UNCCHARLOTTE

Department of Mechanical Engineering and Engineering Science

The University of North Carolina at Charlotte, Charlotte NC, U.S.A.

Contents

List of Figures

List of Tables

Nomenclature

a, b, c	*arbitrary constants*
A	*amplitude (e.g. sine wave)*
d	*depth*
$\Delta x, \Delta\Theta$	*data point spacing*
f	*frequency*
f_c	*cutoff frequency*
h	*height*
i, j, k	*counter variables*
l	*length of a profile*
λ	*wavelength*
λ_c	*wavelength*
μ	*mean or average*
n	*number of points or occurrences*
Θ	*angle*
R, r	*radius*
S	*spacing*
s	*filter weighting function*
x, y	*cartesian coordinates*
z	*cartesian coordinate, cylindrical axial coordinate*

Acknowledgments

I wish to express my appreciation to the many people who have been instrumental in my completion of this work:

- To Dr. M.B. (Bill) Grant for providing motivation, encouragement and support for all of my activities in the metrology community.

- To Dr. Derek Chetwynd for the many hours of guidance, editorial input, and enlightening conversation over the course of this work.

- To Dr. Jay Raja for introducing me to the field of surface metrology and the years of collaboration that we have enjoyed together.

- To my colleagues Dr. Henrik Nielsen, Tim Walton, Samir Bhargava, Jim Hall and Ken Titus for your insights and conversation regarding this topic.

Finally, I want to extend my deepest gratitude to the members of my family, including:

- my wife, Barb, for her unending love, support and patience throughout the course of this work and my many other activities in the metrology community.

- my mother, Sally, for her continued encouragement and interest in all that I do.

- my children, Rebecca and Stephen for the joy you have brought and continue to bring to me.

Declaration

This thesis is presented in accordance with the regulations for the degree of Doctor of Philosophy by the Higher Degree Committee at the University of Warwick. The thesis has been composed and written by me based on my own research and experience. These materials have not been submitted in any previous application for a higher degree. All sources of information are specifically acknowledged in the content.

Mark C. Malburg

Summary

This thesis addresses the growing "divergence" in the field of surface metrology through the presentation of a practical system for unification. A technical and economic review of applied surface metrology is presented, highlighting the problems associated with the many advances in instrumentation - particularly in light of the growing industrial dependence on surface metrology. This background serves as the basis for the development of a scheme whereby surface specification, instrumentation, and analysis can be concisely and completely defined and, more importantly, controlled.

Several technical aspects of surface metrology are addressed in the development of the scheme. First the topic of specification and reference geometries is addressed, where it is argued that least squares methods should provide the most stable basis for assessment. Stylus/radius convolution and the associated wavelength transmissions are also considered and experimental investigations are undertaken as to describe their influences on measured data sets. The treatment of unwanted asperities is investigated and a new, robust algorithm developed and presented. The study of wavelength limitation approaches concludes that a sub-set of current methods is technologically acceptable and therefore economically attractive. A review of parameterization concentrates on a means for selecting a "unified" set of parameters and guidelines for the incorporation of future parameters.

Finally, it is shown that the this proposed scheme addresses the underlying divergence in surface metrology in a manner which is practical in the context of application, technically justified in the context of standardization, and extensible in the context of further research.

A Unified Methodology for the

Application of Surface Metrology:

Chapter 1

The Need for Unification

The field of surface metrology is at an important juncture:

> In the midst of a rapid *increase in the demands* on surfaces, there has been a corresponding *increase in the methodologies* available for the assessment of these surfaces.

From one perspective this could be viewed as an ideal situation for those involved in applying surface metrology in an engineering context. However, the above view misses the underlying significance of the situation.

> In the midst of a rapid *increase in the demands* on surfaces, there has been a corresponding *"divergence" in the methodologies* available for the assessment of these surfaces.

The surface metrology community is now in an era where nearly every conceivable means of sensing surface geometry is actively being developed. In addition, the approaches which have been historically used for measuring surfaces are being significantly refined and extended to encompass a much broader range of capabilities. Furthermore, the incorporation of general purpose computers into measuring systems has added a tremendous amount of flexibility in data manipulation (Kinsey and Chetwynd 1973). This has, in some cases, lead to better mathematical representations and, in other cases, further confusion about the numerical results obtained from the instruments (Whitehouse 1982). All of these developments have contributed to a *divergence* in the field of surface metrology. Instruments with <u>different</u> sensing mechanisms, <u>different</u> bandwidths and <u>different</u> data processing are said to be measuring the <u>same</u> surface attributes.

At the same time as these changes in instrumentation have been taking place, industry has been reducing tolerances to the point where surface metrology has become more important than ever (see, for example, Taniguchi 1983). These measurement *methods*

divergences and the demand for the tighter control of surface features are clearly incompatible.

In industrial applications of surface metrology, there are two viable options: 1.) Ignore measurement technology developments or only conservatively employ them for the purpose of maintaining historical consistency; or 2.) develop a scheme in which uniformity can be established between instruments and the handling of their data. The former carries the potential cost of missing out on economic and technological benefits often associated with surface metrology advances. This thesis addresses the latter.

In this introductory chapter, some of the factors which are leading to this divergence will be presented. It is, however, important to recognize that these factors should not be viewed negatively - they, in many cases, represent significant advances in instrumentation. Nonetheless, this divergence in surface metrology will be viewed from a historical as well as technological perspective.

1.1 Surface Features and Standardization

Historically, the function of a component was thought to be related to its "*fit*, *form* and *finish*" even though the boundaries between these regimes was often quite vague. Instrumentation was developed which would specifically target one of these areas and associated standards would be developed based on the capabilities of that particular instrumentation (Whitehouse 1990). The concept of an intermediate wavelength surface attribute "waviness" seemed to follow behind the development of roughness and form measurement (Reason 1965).

Today's terminology tends to refer to "*size*, *orientation*, *form errors*, *waviness* and *roughness*" as the primary metrological aspects associated with surfaces. The measurement of size and orientation is addressed in the field of *dimensional metrology* (Busch 1989) and is important in understanding many *static*, functional aspects including "fit". Roughness, waviness and errors of form are addressed by *surface*

metrology and are generally (although not exclusively) attributed to the more *dynamic* aspects of component interfaces such as "sliding and loading" (Whitehouse 1994).

The specific boundaries between "roughness", "waviness" and "form errors" are rather ambiguous although they are often spoken of as separate wavelength regimes. In a general sense, roughness, waviness and form errors are defined based on the specific application (Thomas 1982). The analogy of a road and car could be considered as an example of relatively long wavelength attributes.

In this scenario, form errors could be the hills and valleys in which the car travels - made up of wavelengths that are longer than the car itself. Waviness could be thought of as the local variations or humps in the road which include wavelengths a few times longer than the tire contact area up to a few times longer than the car - this wavelength regime is associated with the general "ride or comfort" of travel. Roughness could then be viewed as rocks and pits in the road's surface which generate wavelengths smaller than the contact area of one of the car's individual tires and ultimately influence such aspects as "tire wear" or "road noise".

At the other extreme it is not uncommon to find surface metrology applications associated with the assessment of finely lapped surfaces such as those ocurring on small fuel injector components.

In the assessment of many fuel injector surfaces, form errors are associated with wavelengths on the order of just a few millimeters (e.g. greater than 2.5 mm) and are related to functional aspects such as "leakage". Roughness is said to be made up of wavelengths shorter than 0.25 mm and is related to "friction and lubrication". Waviness is then associated with the surface features which are between 0.25 and 2.5 mm in wavelength and are generally attributed to localized areas of increased "loading and wear".

Despite the ambiguity in boundaries, the wavelength regimes continue to be addressed rather "separately". More importantly, the history of surface metrology tends to indicate an acceptance of the notion that roughness, waviness, and errors of form were somehow *separated* and therefore required *separate* treatment from metrological (Whitehouse 1990) and standardization perspectives.

These differences in treatment are evident when viewing a categorical listing of major ISO, UK, US, and DIN standards in terms of their areas of application. As Table 1.1 indicates, the standards community has viewed the three primary regimes of surface metrology as independent. Furthermore, the very closely related field of roundness measurement is treated in a manner completely independent of the above categories.

This approach to standardization has resulted in a great deal of redundancy - both across standards and in some cases within individual standards. Many common, metrological terms are redefined in each standard (often differently) and many concepts are often re-addressed in a different context. Recently, this problem has been recognized by ISO and a "harmonization" effort has been initiated, however the current efforts are focussed on the revision of existing standards and the development of procedures for writing new standards.

Since the boundaries between these wavelength regimes are "application specific", what is considered roughness in one application, may be considered waviness in another application. Thus, the metrological requirements and guidelines associated with the same surface wavelengths may be found in both roughness and waviness standards depending on the application of the surface!

Standard	Roughness	Waviness	Form Errors (Straightness)
ISO 468	X		
ISO 2632 1-3	X		
ISO 3274	X		
ISO 4287	X		
ISO 4288	X		
ISO 11562	X		
ISO/DIS 12085	X	X	
ISO/DIS 13565 1-3	X		
ISO 1101			X
BS 1134	X	X	
BS 2634	X		
BS 3730			X
BS 6741	X		
ANSI B46.1	X	X	
ANSI Y14.5			X
DIN 4760			X
DIN 4761	X		
DIN 4762	X		
DIN 4765	X		
DIN 4766	X		
DIN 4768	X		
DIN 4769	X		
DIN 4772	X		
DIN 4774		X	
DIN 4775	X		
DIN 4776	X		

Table 1.1 *Surface metrology standards and their fields of application.*

1.2 Instrumentation Developments

The divergence in the field of surface metrology can also be attributed to the instrumentation used in the measurement of surfaces. Historically, instruments were designed for a specific measurement purpose. Painstaking efforts were put forth such

that a complete mechanical (sometimes electronic) system was developed for the purpose of measuring a specific attribute of a surface.

In the measurement of roughness, early instruments were developed by Schmaltz, Abbott, and Reason (see for example Thomas 1982 or Whitehouse 1990 for summaries). In 1929, Dr. G. Schmaltz developed a stylus based profile measuring instrument incorporating a tilting mirror which reflected light based on stylus motion. This reflected light beam was projected onto photographic paper thus producing a magnified plot of the stylus motion. Schmaltz later went on to produce a means of profile measurement by "optical sectioning". In this approach, a thin band of light is projected on to the surface at a 45° angle. The surface is then viewed at a 45° angle from the side opposite the illumination. (A similar approach is presented by Firestone et al. 1932.) The resulting image produces profile heights which are magnified by $\sqrt{2}$. E. J. Abbott and R. E. Reason began the development of stylus based roughness measuring instruments in the late 1930's which became the first instruments incorporating electronic means of amplification (Dagnall 1980).

The above mentioned instruments were designed for the specific purpose of measuring surface roughness. These instruments typically provided relatively short assessments along the surface and very limited analysis capabilities. However, despite these limitations, they satisfied the immediate need for the assessment of surface roughness.

Similar efforts were put forth in the development of instruments for the measurement of straightness based on a straight datum and roundness based on a rotating spindle or probe. These developments resulted in usable, yet very *task-specific*, instruments (Whitehouse 1990). Although the instrumentation developments were rather task specific, in many cases, they were based on the principles learned from the roughness instruments, rather than a thorough a re-visiting of first principles (Chetwynd 1995).

With the advent of digital methods and flexible computation devices as well as further mechanical developments, surface metrology instrumentation capabilities began to

expand rapidly in terms of potential applications for a single instrument (Whitehouse and Reason 1965, Kinsey and Chetwynd 1973). Instruments that could once only measure the average roughness parameter, could now be computerized to digitally assess additional roughness parameters as well as provide numerical and graphical results for waviness and form errors.

Spindle based roundness instruments evolved into "cylindricity" instruments through the incorporation of a *z* axis reference datum and digital computation. This incorporation of an additional axis into roundness instrumentation provided another means of assessing (surface) straightness.

In the midst of these advances in instrumentation, there have been considerable advances which have increased the *bandwidths* (or range of detectable wavelengths) of existing instruments. This broadening of bandwidths is the result of improvements in the mechanical and electronic frequency response thereby allowing much finer surface features to be assessed. It should be mentioned that a significant contribution to these advancements was in terms of increased storage capacity of digital computers.

One obvious example of the increase in bandwidth can be seen in the increase in data density in the measurement of roughness. Many instruments of the 1980's vintage used between 1000 and 8000 data points in a data set (Rank Taylor Hobson 1985, Feinprüf Perthen GmbH 1989). In addition to Nyquist limitations, these instruments' bandwidths were further limited by analog filters which attenuated high frequency "noise". Today, it is not uncommon for instruments to use up to 120,000 data points over similar trace lengths (Rank Taylor Hobson 1995) with no significant loss of transmission. Based purely on the Nyquist criterion (Bendat 1986) and considering electronic influences, this means that bandwidths have increased to the point that wavelengths 30 to 120 times smaller can be realized in today's instruments as compared to those of a decade ago.

The wide acceptance and advancement of coordinate measuring machines (CMM's) have also provided alternative approaches to the measurement of surface attributes (Busch 1989). Historically, CMM's were used for dimensional assessments, however, advances in analog probing and significant increases in data point storage have resulted in utilization of CMM's in many surface measurement applications which were once associated with special purpose instruments. The most common of these recent CMM applications is in the measurement of roundness (Neumann 1990).

The use of CMM's in the measurement of surface features (as well as other geometrically dimensioned and toleranced (GD&T) features) has been heavily scrutinized (Edison and Parry 1985, GIDEP 1988, Whitehouse 1994) in terms of sampling, filtering and data processing. The two primary concerns have been sampling and data processing. Often in these CMM applications, the user selects the number of data points and the spacing thereof. This can result in significant differences resulting values for center location, size and roundness. Furthermore, typical CMM's do not filter the roundness data in the same manner as dedicated roundness or cylindricity instruments and nearly all analyses are based on least squares substitute geometries. Nonetheless, the popularity of CMM's in surface metrology applications continues to grow due to their flexibility and ever-improving ease of use.

1.2.1 Summary of Instrumentation Bandwidths

The advances that have been made in instrumentation for surface metrology have caused overlaps between instruments in terms of the surface wavelengths which can be assessed. This situation is graphically presented in Figures 1.1 and 1.2.

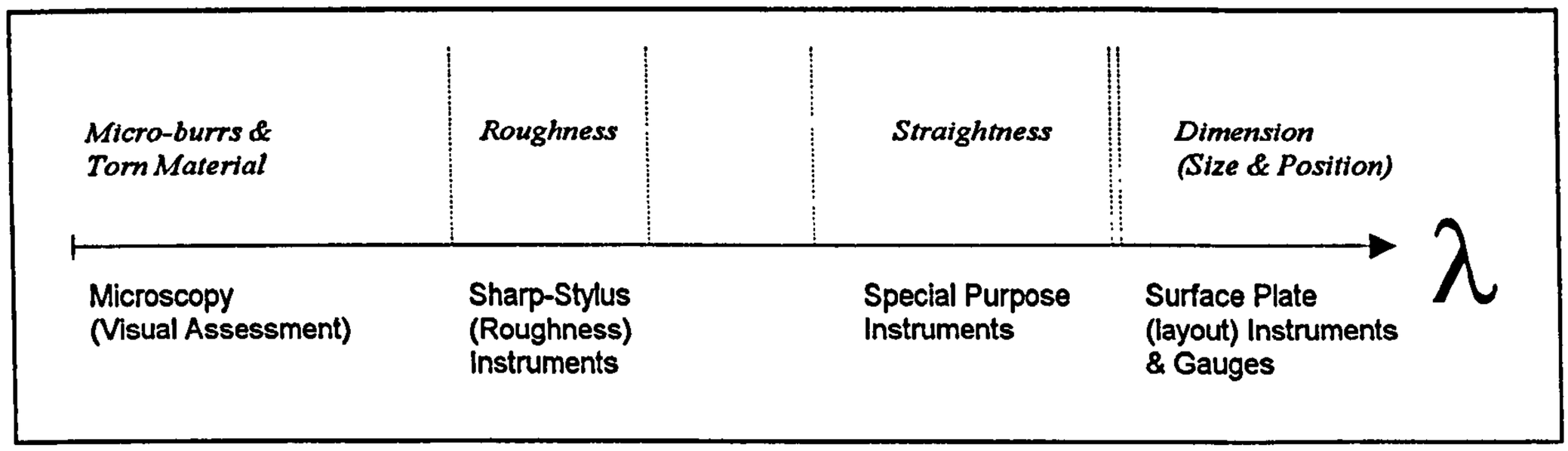


Figure 1.1 *The historical surface metrology spectrum for linear measurement.*

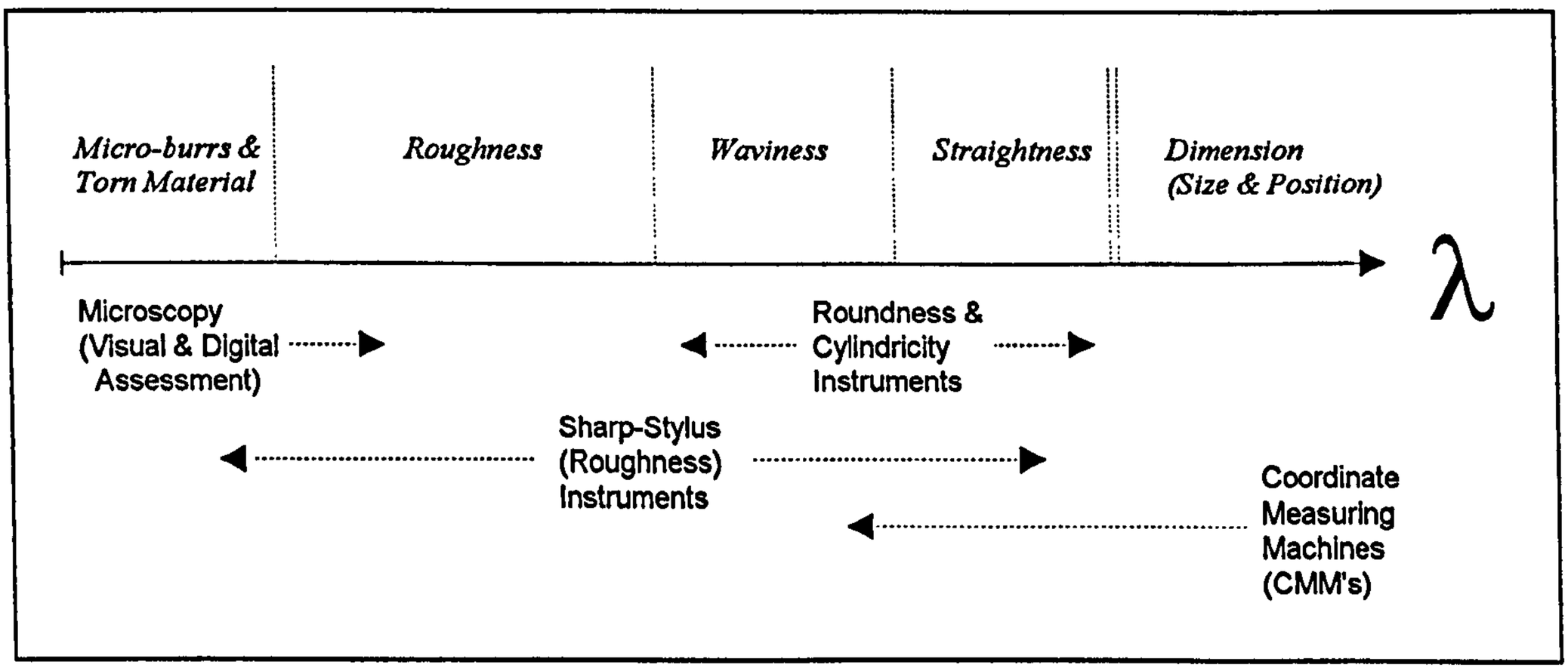


Figure 1.2 *The current surface metrology spectrum for linear measurement.*

Figure 1.1 indicates a *parallel* relationship between instrumentation and application. It is representative of historical applications of surface metrology in primarily metal working industries. Even today, the application of surface metrology in the vast majority of small metal working companies can be summarized as in Figure 1.1.

In more *state-of-the-art* applications (typically found in *high-precision* and high volume applications) there is a very different relationship between surface attributes and instrumentation. This is depicted in Figure 1.2 which summarizes the instrumentation "overlap" that is becoming very common in today's surface metrology applications. This overlap is leading to differences in measured values depending on the methodologies of assessment.

The primary advances which have lead to this "overlap" are summarized as follows.

- The microscopy field has made significant advances through the incorporation of digital image processing and stereo imaging as well as further enhancements of scanning electron microscope (SEM), scanning tunneling microscope (STM) and atomic force microscope (AFM) technologies (Clark and Grant 1992, Cohen 1995).

- Advances in frequency response, sampling and datums have significantly broadened the bandwidth of traditional roughness instruments.

- The incorporation of a (straight) vertical axis into roundness instruments as well as advances in frequency response, sampling and data processing have resulted in a very wide range of application for these instruments. Recent applications of these types of instruments include dimensional assessments (Feinprüf Perthen GmbH 1993).

- Increased data storage and further developments in analog probing for CMM's have allowed them to be applied to various surface metrology assessments.

Each of these developments, on an individual basis, represents an advance in metrology. However, when viewed together, they cause a great deal of confusion in that several instrumentation approaches are being used to assess the same surface attribute and resulting numerical values are not necessarily the same.

1.3 Differences in Data Acquisition

One of the primary reasons that results obtained from various instruments do not agree is due to differences in the data set that is ultimately presented to the calculation

algorithms. Apart from environmental (temperature, vibration, electrical noise, etc.) and "measuring loop" (ISO 3274 - 1995) considerations, these differences can be sub-divided into two major categories:

- **Probing** The physical *sensing* of a surface in terms of a contact area or effective spot size.

- **Sampling** The horizontal and vertical quantization and data collection along a surface.

These differences will be explored in light of the surface being measured and the instrumentation which is measuring the surface.

1.3.1 Probing

Given the history of instrumentation and its task-specific development, it becomes apparent that there is are significant differences in stylus tip geometry. On the *short wavelength* end of the spectrum, instrumentation typically employs sub-micrometer stylus tip radii and sampling at near-atomic surface wavelengths. Traditional roughness instruments are are next along the wavelength spectrum with stylus tip radii typically ranging from 2 μm to 10 μm. In addition, many roughness instruments are providing optional "form" assessment capabilities utilizing a stylus radius on the order of 0.5 mm (Rank Taylor Hobson 1985, 1995). Further up the wavelength spectrum, roundness instruments can be found - using stylus tip radii typically ranging from 0.5 to 2.5 mm and finally CMM's with tip radii typically ranging from 1.0 mm up to (and sometimes beyond) 5 mm. These probing differences have, in most cases, remained despite the changes in bandwidth of their *parent* instrument.

It should be noted that the contact area of a given stylus is a function of the surface being measured. However, the result of using various styli is a varying transmission of wavelengths which could ultimately lead to variations in calculated parameters. For example, a CMM measuring straightness with a 2 mm radius stylus will sense the

surface differently than will a roughness instrument incorporating a 2 μm tip. A considerable amount of work has been done on in terms of stylus convolution and its impact on the measured profile and the resulting parameters (Radhakrishnan 1970, Shunmugam and Radhakrishnan 1974 & 1975, Whitehouse 1974).

Also related to this topic is the physical effect of the interaction of the stylus and surface. The forces, dynamics, contact stresses and deformations relating to the measurement are important aspects and should not be neglected in application. Although related to this work, it will not be discussed explicitly herein. (See, for example Liu et al. 1992, Chetwynd et al. 1992, or Liu 1994.)

1.3.2 Sampling

As previously mentioned, the sizes of data sets used in typical surface assessments have grown significantly. In roughness analysis, data sets which historically comprised less than 10,000 points are now exceeding 100,000 data points. Scanning capabilities incorporated into CMM's have now facilitated the relatively fast collection of hundreds of points whereas the traditional "touch trigger" systems typically collected fewer than 20 points. (It should be noted that these advancements are not only the result of increased sampling as they have typically required instrumentation modifications to enhance the frequency response characteristics for higher frequencies.)

While, as a general trend, the instrumentation used in surface metrology is using more and more data points, there are still significant differences in data density from instrument to instrument. These differences in data density can result in significant differences in computed values. For example, in the measurement of a peak to valley parameter, an instrument with a lesser data density will be less likely to sample a point at the highest peak than will an instrument with high data density. This is also true in the detection of the deepest valley. Thus, a lesser data density will typically yield a lesser peak to valley assessment of a surface when assessing extreme amplitude parameters (Sharman 1967a, Chetwynd 1979b).

1.4 Differences in Parameters

The parameters used in surface metrology seem to be very dependent upon the wavelength regime that is being assessed and are often based also on the means of assessment. Early roughness instruments produced only a profile graph which was left to graphical interpretation. This resulted in the assessment of graphical parameters such as peak to valley measures of the surface. Enhancements of the graphical method came through the utilization of mechanical integration devices such as the planimeter (see, for example, Whitehouse 1994) allowing the assessment of an average roughness value (historically referred to as CLA, AA and now as Ra). With the implementation of digital computers, further parameterizations were made possible and a "parameter rash" (Whitehouse 1982) occurred whereby numerous (and often redundant) parameters were developed with the intention of characterizing some interesting aspect of surface roughness data sets. Furthermore, previous parameters were not made obsolete with the development of new parameters. Today, the approaches which are available for characterizing roughness data have grown to the point that many engineers become very intimidated when introduced to the field of roughness measurement.

Many of the roughness parameters have also been applied in longer wavelength analyses such as waviness, roundness and straightness. However, in these areas of measurement, it is more common to encounter the application of extreme height (peak to valley) parameters.

In addition to the various parameters, alternative reference figures are available. In the measurement of roundness there are four common reference circles - the least squares circle, minimum zone circle, maximum inscribed circle, and minimum circumscribed circle (ISO 12181 - 1995). Each reference circle was originally developed to target some specific physical aspect of the geometry. Thus the different reference features typically yield different values for the out of roundness assessment.

A problem common to roundness and straightness measurement is the difference between the interpretation of geometric dimensioning and tolerancing (GD&T) and common measurement practice. In both cases, (i.e. roundness and straightness) the drawing standards indicate that all points on the surface must lie within the tolerance zone (ISO 1101 - 1983, ASME/ANSI Y14.5 - 1994). This would dictate that the part can be accepted if any solution exists whereby all points on the surface fall within the tolerance (Carpinetti & Chetwynd 1994). This can be acceptable for pass-fail (attribute) type inspections, however, in industrial practice numerical results are required for statistical process control (SPC) and other process or product capability assessments. Thus, the minimum zone approach yields the "smallest tolerance that would be satisfied by the surface being measured" and is therefore most related to GD&T. However, in practice, the vast majority of roundness and straightness measurements are based on least squares reference features (Castle 1993, Hildebrandt 1994).

1.5 The Need for Unification

It has been shown that the field of surface metrology has historically advanced down several parallel paths. In the past, these paths have been somewhat separated into "wavelength regimes" (Figure 1.1). In each of these regimes, instrumentation and associated standards have been developed. More recently, advances in instrumentation and a growing importance of surface metrology have led to a great deal of instrument and parameter development. These developments have caused a *divergence* in the field of surface metrology. In the present scenario, the surface metrology community is are faced with ever broadening instrument bandwidths such that many, very different instruments are in use measuring the same surface wavelength and are being guided by many, very different standards for their use. This is further complicated by the "user defined" separation between wavelength regimes, wherein wavelengths that are considered waviness in one application may be considered roughness in another application.

1.5.1 Technical Implications

This lack of correlation between instruments poses many issues to product development and quality control engineers. First and foremost of these issues is related to the determination of tolerances. Often prototype parts are produced and evaluated (measured and tested) to establish their functionality. Tolerances are then established based on the relationship between the measured attributes and the desired degree of functionality. The problem arises when the instrumentation used in establishing the tolerances produces different results than does the instrumentation which is subsequently used for process control and product acceptance.

To demonstrate this situation, an evaluation of straightness was performed on a diesel engine accessory shaft (rough ground) using four different instruments. An area of the component was indicated for measurement and a request was submitted to a skilled metrology laboratory technician which stated "perform a straightness measurement, storing the profile data points to disk". The component was measured on four instruments including: 1.) a Rank Taylor Hobson Form Talysurf Series S5, 2.) a Gould 1200, 3.) a Mahr Corporation MFU8-C, and 4.) a Zeiss UPMC 850. The measurements were performed in a manner "typical" to the particular instrument and the resulting profiles are given in Figures 1.3 through 1.6.

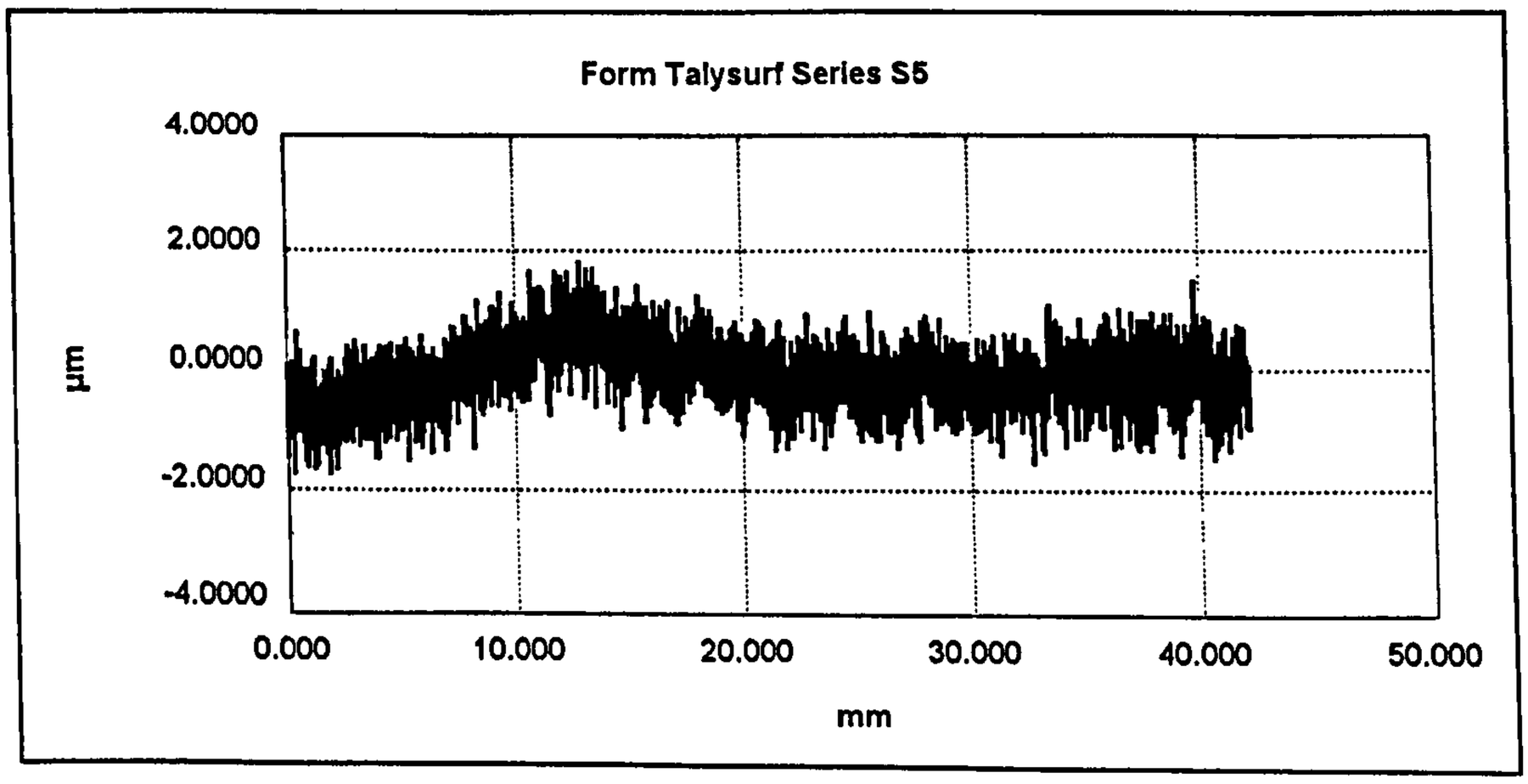


Figure 1.3 *Straightness profile as measured with the Form Talysurf Series S5.*

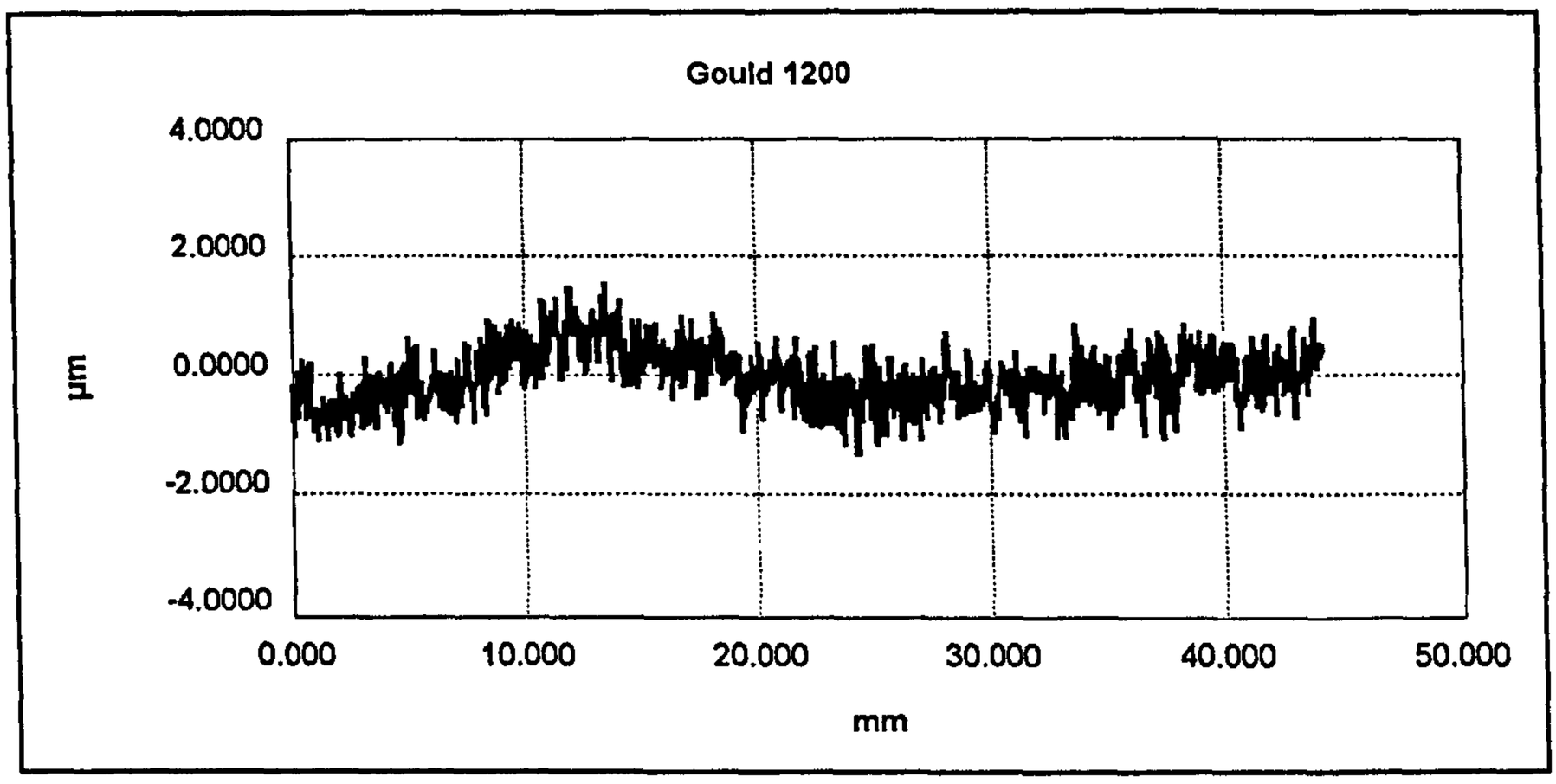


Figure 1.4 *Straightness profile as measured with the Gould 1200.*

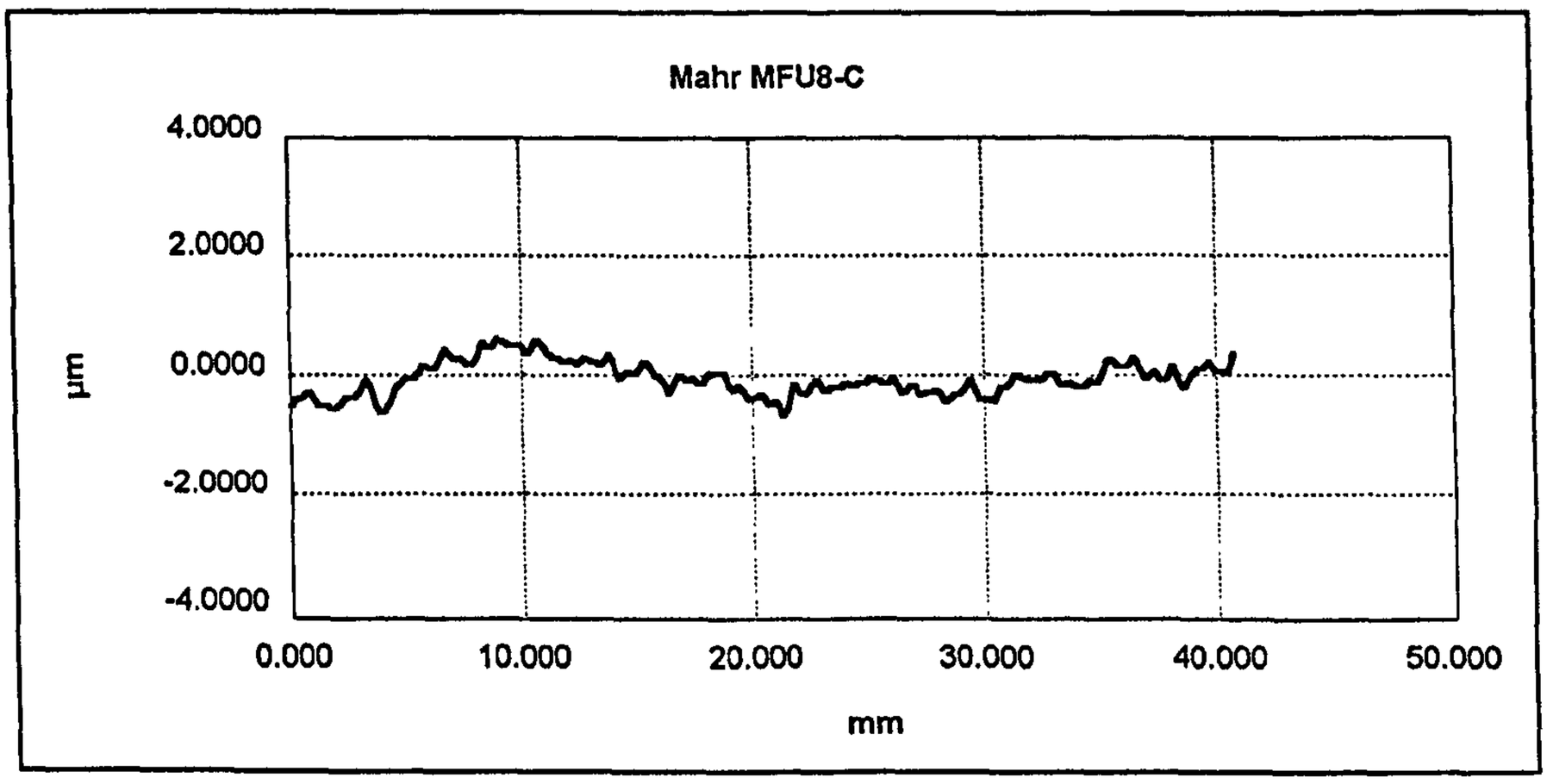


Figure 1.5 *Straightness profile as measured with the Mahr MFU8-C.*

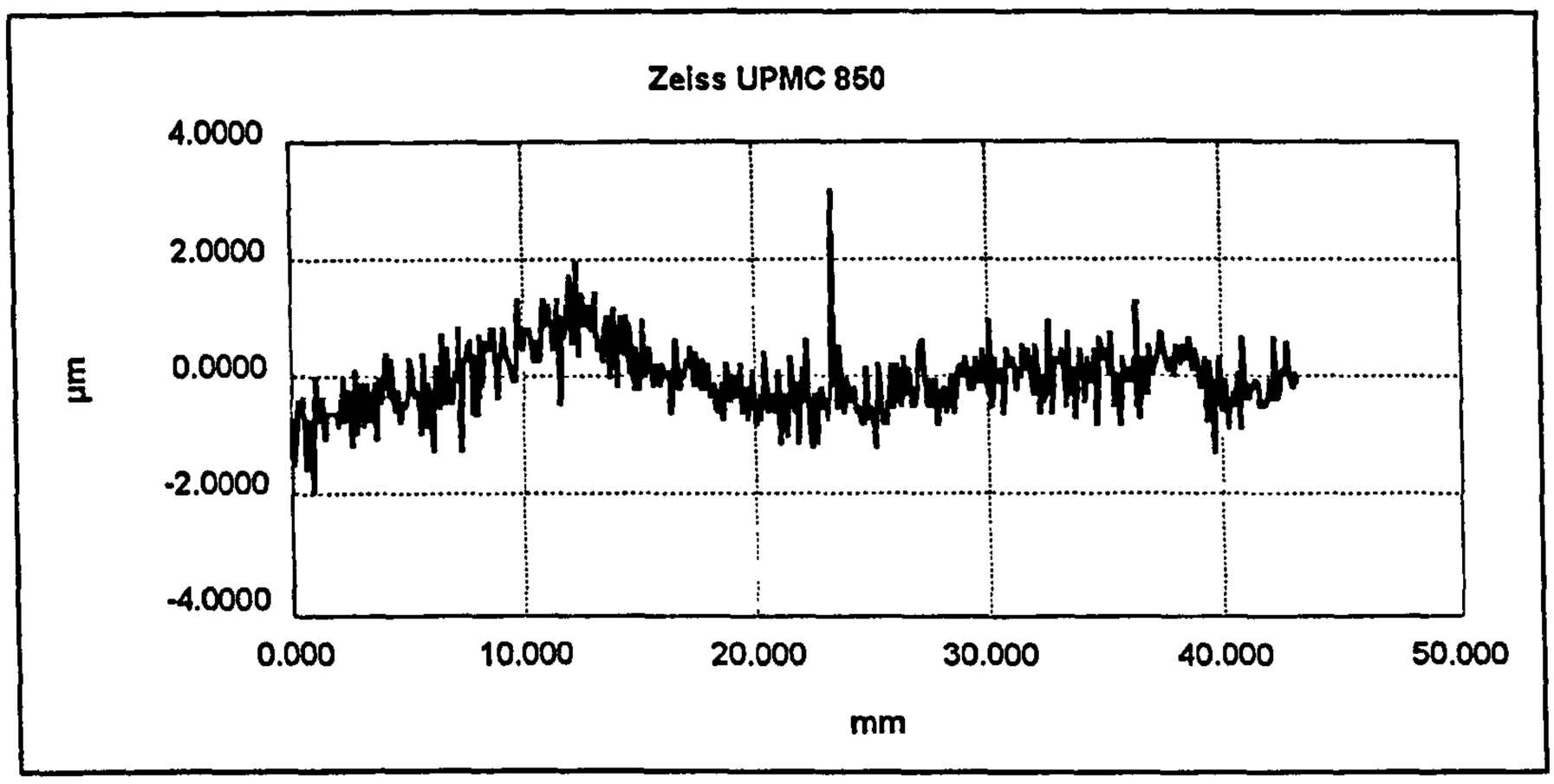


Figure 1.6 Straightness profile as measured with the Zeiss UPMC 850.

The specific details of the measurements (stylus tip radius, sampling, etc.) are not important for the purposes of this introductory chapter and they will be addressed in detail in Chapter 7. The significance at this point in the presentation is that the component was measured in a manner "typical" to each of four instruments where each of the instruments is said to be capable of measuring the feature.

While these profiles may exhibit similar "shapes", significant differences in numerical values are realized upon computing the peak-to-valley straightness values. These values are graphically depicted in Figure 1.7.

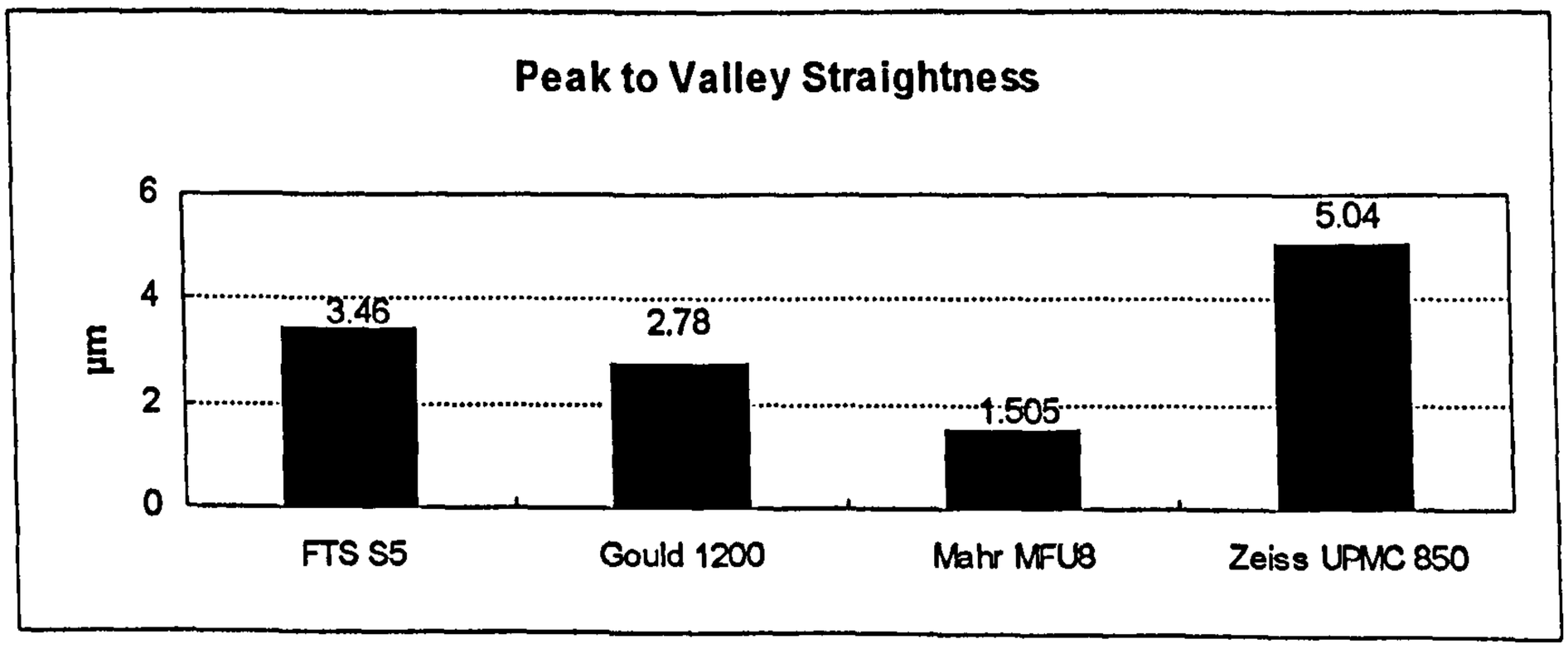


Figure 1.7 Comparison of straightness values from various instruments.

This variation in results poses a significant problem in the acceptance or rejection of components - particularly in light of the fact that the straightness tolerance on this component is 2.5 μm!

1.5.2 Economic Implications

Dealing with this *divergence* in surface metrology is currently one of the most costly aspects of applying surface metrology in industry. However, very little is being said about this in technical publications. This is perhaps due to the general lack of industrially authored technical papers as well as the view that the problem is more of an economic nature than technical. (The problem would not be significant if everybody would purchase exactly the same instruments and use them in exactly the same manner.) However, that is not possible in a world of rapidly changing technology and ever increasing demands on the technology. Furthermore, in most traditional university or research settings, these correlation issues do not surface in that most their activities are based on the development of a technology relating to one specific measurement process.

Through several recent discussions with automotive component suppliers it has been found that the costs associated with these measurement problems is outstanding. In one example, a cylinder liner supplier produces prototype components for evaluation by a customer. These prototype cylinder liners range in price from $40 to $100 and are typically produced in lots of 100. Given the nature of prototype components, the supplier measures each of the components and, upon receipt, the customer re-measures them prior to testing. For a typical liner supplier, one lot per week is rejected by the customer whereby the supplier measurements deemed the components to be acceptable and the customer measurements were the basis for rejection. For one company alone, this measurement issue results in an annual scrap cost of between $200,000 and $500,000 (Lenthall 1996) and similar costs can be shown across other component suppliers.

These extremely high measurement related costs are also prevalent in the industrial purchasing of machine tools. The procurement of manufacturing equipment such as machine tools is typically a multi-step process. This process often begins with the customer issuing "requests for quotations" whereby potential suppliers can propose their solutions. During this time the many aspects of the machine tool are presented including such items as product tolerances, required cycle times, tooling change frequency and time, maintenance, controls, etc. (Burton 1993). Upon selecting a supplier for the machine tool, the acceptance criteria are established. Typically this criteria includes the evaluation of the machine tool at the suppliers location and once again upon installation at the customer's location. (In typical arrangements, a portion of payment is released upon satisfactory completion of the "supplier site evaluation" and the remainder of the payment is released upon successful installation, evaluation and training at the customer location.)

In many cases where the machine tool being purchased is generating critical surfaces (and depending on the customer's resources) the customer will re-locate, to the supplier, a surface measurement instrument which will be eventually be used for process control. This allows more of the "system" (i.e. machining and measurement) to be evaluated in light of its future deployment at the customer location. The unfortunate aspect of this is the fact that the supplier typically has its own surface measuring instrumentation, however it does not generate the same results as does the customer's. Thus, significant additional costs are incurred during a machine evaluation as measurement equipment is packed, shipped (and often damaged) an re-installed at a supplier's location. Typically, the costs associated with the transfer of surface measuring equipment are on the order of $20,000 although, in some cases these costs can exceed $50,000 (Rose 1993).

While this situation is troubling in the development of specifications and procurement of manufacturing equipment, it is also of great concern in the receiving inspection areas for purchased parts. In typical American companies, sampling is in accordance with

the military standard, MIL-STD-105D (1963). Furthermore, for critical features, (such as critical surface attributes) a "zero reject rule" is applied whereby if any sample measurement falls outside the acceptance limits the entire lot is to be rejected. Coupling this incoming inspection rule with often disagreeing measurements between the supplier and customer, many instances of rejected shipments can occur.

This incoming inspection problem is further complicated by the rapidly growing trends toward "just in time" (JIT) manufacturing and "ship to build" assembly. In this scenario, inventory is to be minimized. Upon receiving the lot of parts, they are sampled in the incoming inspection area and delivered directly to an assembly line ("just in time"). The rejection of parts by incoming inspection can potentially result in a situation where components are not available for assembly. This costs associated with "assembly line shutdowns" can easily be on the order of $4000 per minute in some plants, and in many cases these costs can be significantly higher (Walton 1996).

1.6 Scope of this Thesis

The application of surface metrology is rapidly gaining popularity in industry - particularly in light of ever-shrinking tolerances. Therefore, it is vital that the above described divergence be addressed in terms of a unified approach for the specification, measurement and control of surface features. Under such a scheme, a common language and set of metrological requirements would be established which could be applied across all surface metrology applications.

This thesis develops such a scheme for application in the stylus-based measurement of surface profiles. The selection of these two conditions (stylus instruments and profile data) is based on the fact that the vast majority of surface assessments are based on contacting stylus instruments and the analysis of profile (often referred to as "two dimensional") data. Thus, in application, the most significant and immediate benefit results in addressing these applications. This is not to say that the scheme is limited by this scope. Later chapters in this thesis will indicate that through the proper

development and implementation of such a scheme an easy extension can be made to accommodate three dimensional assessments.

Given its purpose, this thesis is organized in a manner quite different from that which is considered "traditional". In this work, a more "design based" approach will be undertaken in that the general scheme (or "design") for unifying surface metrology will first be presented which is then followed by detailed discussions of the issues relating to the scheme as well as future applications and extensions .

The "Scheme for Unification" will be presented in Chapter 2, including a presentation of the wavelength limitation methods, reference features and parameterization. The proper design, development and presentation of this scheme is of great importance in terms of gaining broad acceptance in the surface metrology community.

Under this proposed scheme, several issues arise which need to be addressed to allow general implementation across the field of surface metrology. Some of the issues relate to systematic problems in surface metrology and require new developments, while others relate to the careful review (and in some cases *acceptance*) of current methodologies. These issues are addressed specifically in Chapters 3 through 6. Chapter 7 concludes the development with further application considerations (including standardization) and future extensions of the scheme beyond stylus generated profile data.

A Unified Methodology for the

Application of Surface Metrology:

Chapter 2

A Scheme for Unification

The divergence of methods in surface metrology applications has been presented in Chapter 1 in terms of instrumentation, standardization and application. It has also been shown that this divergence is very costly in industries that rely heavily upon surface metrology applications. In the context of these important issues, this chapter presents a scheme for "unifying" the field of surface metrology.

The concept of "unification" has been rather popular in surface metrology - particularly in light of parameterization and instrumentation. In terms of parameters, several authors have presented some form of a "unified" analysis technique (Spragg and Whitehouse 1970, Greenwood 1984, Scott 1986, O'Connor and Spedding 1992). The characterization of roughness profiles is typically the basis for these proposals and the underlying goal is the development of an approach whereby the functionality of a surface can be ascertained from a new processing technique, numerical value or relatively small set of numerical values.

In terms of instrumentation, there have also been proposals which were intended provide some degree of "unification" (Stedman 1987, Scott 1992a). The former, presented an approach for the evaluation of surface metrology instruments whereby the performance of various instruments could be graphically compared. The latter described the approach presented in ISO 3274 (1995) whereby a band limitation has been defined for the analysis of roughness profiles. This band limitation, was set forth in order to obtain correlation between varying instruments (given similar transmission characteristics within the desired bandwidth).

These attempts at unification have provided some incremental advances surface metrology *technology*. However, the underlying "divergence" presented in Chapter 1 still remains a significant problem as it relates to the surface metrology *practice*. Thus, in order to adequately address this situation, a more comprehensive approach must be developed. In the following sections of this chapter, such an approach is developed. This approach or "scheme" is truly one of unification as it deals with the technical and

practical aspects of surface metrology including instrumentation, analysis and specification. Furthermore, this scheme for unification applies to the entire *spectrum* of surface metrology applications rather then just one isolated area such as roughness analysis.

In practical application, the guidelines set forth in this chapter encompass an entire "methodology" for specification, instrumentation and analysis. A common basis for surface metrology applications can be achieved through the broad adoption of and adherence to this proposed methodology. Thus it becomes important that the scheme be suitable both in a technical sense for standardization and research purposes and in a practical sense for industrial application.

2.1 Goals for the Unification Scheme

In the development of this scheme for unification, several goals were set forth. These goals have been derived based on technical and economic concerns which have arisen out of experience in applying surface metrology (see Chapter 1).

- *Reproducibility* of measured results across various instrumentation.

- A common *language* for surface metrology - independent of the nominal geometry and wavelength regime.

- *Stability* of measured results in that they reliably reflect the surface features which are to be characterized.

- Common *characterizations* through a common parameter set.

- *Flexibility* to accommodate future developments in wavelength limitation, parameterization and analysis functions.

- Practical *attainability* through available instrumentation.

Apart from the six technical and economic goals, the scheme must be acceptable by the metrology community in that there is sound justification and it is compatible with the growing data base of surface metrology information. This *acceptability* criterion is not well defined, but rather is an underlying theme in which each of the six goals contributes. Ultimately, the surface metrology community will have to weigh the costs and benefits associated with changing to this new scheme or *paradigm*. This issue has been considered in the development of the scheme and mechanisms have been developed to ensure that there is a rather direct translation of existing methodologies into the new scheme.

The scheme is designed in a manner such that it can be generally applied. However, for the purposes of this work, it is more useful to present it in the context of stylus-based profile measurement and analysis while allowing for extension into other areas. Furthermore, the acceptance of such a scheme may be better achieved through focusing on the majority of surface metrology applications (those being profile based). If the focus were shifted toward technologies that were not so common, such as optical or three dimensional, then the response of the majority of the surface metrology community could easily become "that doesn't apply to me" and the overall acceptance of this approach may be jeopardized.

2.1.1 Geometries

The unification scheme will be presented in its application to straight and round geometries in the context of profile (often referred to as two-dimensional) measurement. The selection of these geometries is logical in that these two nominal geometries (or combinations thereof) constitute the vast majority of the surfaces which are common to industry and assessments which are common in surface metrology.

Furthermore, these geometries present a very good example of diverging approaches which can be "unified" through this proposed methodology.

Historically, linear and circular geometries have been treated rather independently in terms of standardization, instrumentation and analysis techniques. However, there is a high degree of similarity between linear and circular analyses. The similarity between surface data collected from a nominally round feature (Figure 2.1) and a nominally straight feature (Figure 2.2) can be presented through the linear display of the round data set (Figure 2.3).

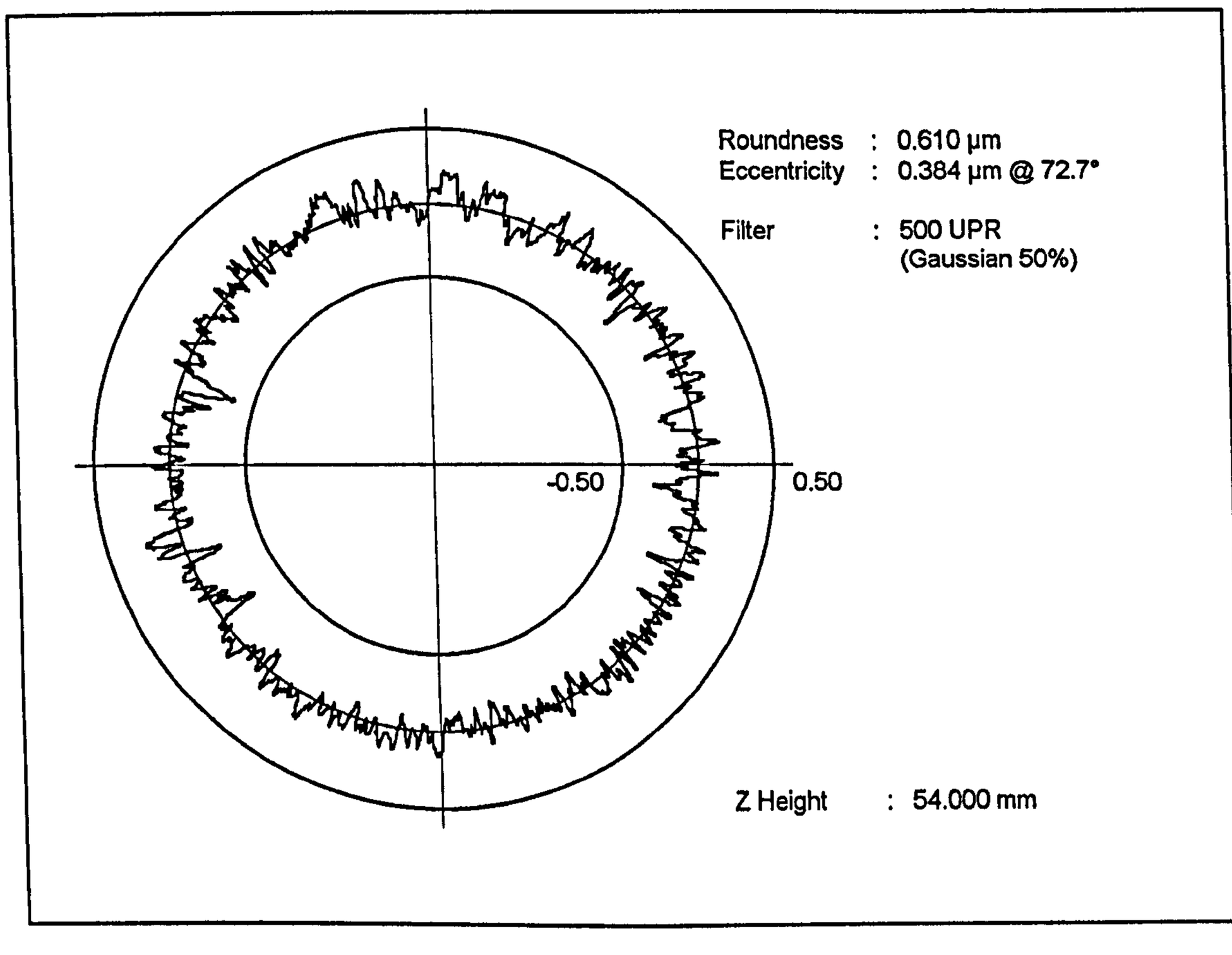


Figure 2.1 *Roundness data for a finely ground component. (Outside diameter, component radius 4.5 mm, stylus tip radius 0.5 mm, 3600 data points, 500 upr (low-pass) Gaussian filter, least squares reference circle)*

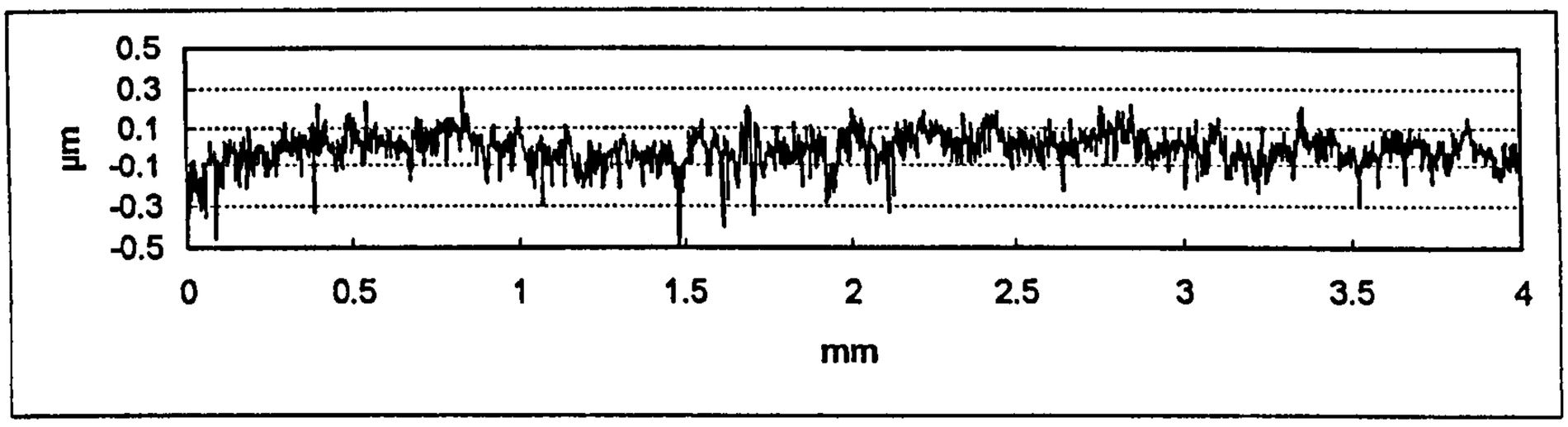


Figure 2.2 *Linear measurement of a finely ground surface. (2 µm tip radius, 2.0 µm data point spacing, no filtering, least squares reference line)*

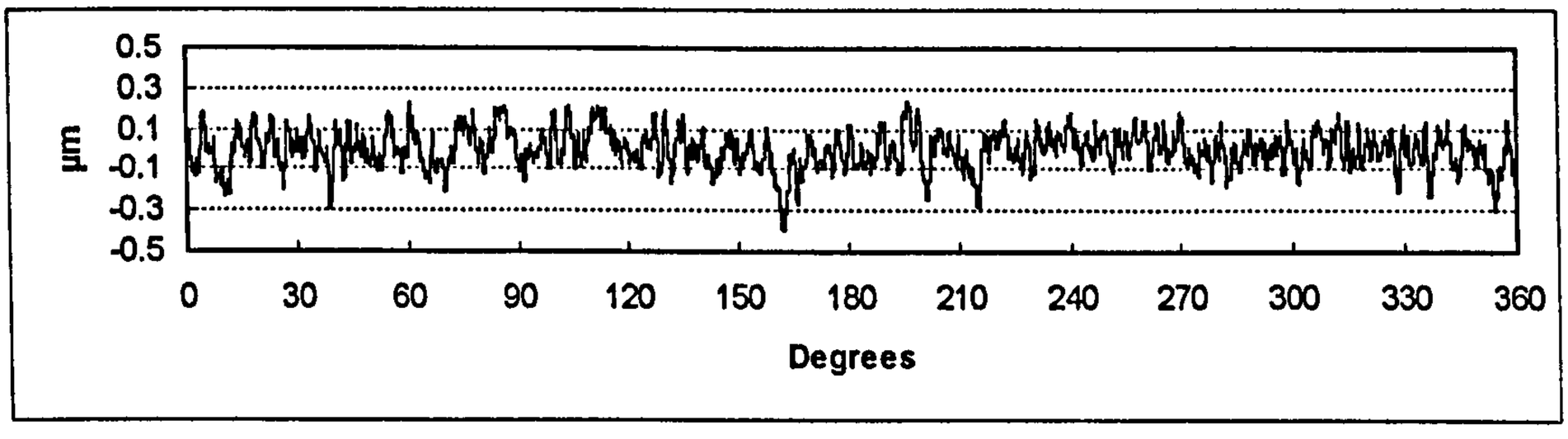


Figure 2.3 *Linear display of the Figure 2.1 data set.*

Given the above example, it becomes readily apparent that linear analyses can be very similar to those associated with circular (or polar) data sets. (Both of these example data sets have a root-mean-squared (RMS) amplitude on the order of 0.09 µm and both exhibit a slightly negative skewness.)

2.1.2 Reference Figures and Wavelength Limitation

A more subtle aspect of the above example relates to the processing of the data prior to arriving at the profiles which were plotted in Figures 2.1-3. In the case of the linear data set, the raw profile was leveled based on a least squares line and no subsequent processing was performed. For the circular data set, the raw profile was centered based on a least squares circle and a 500 undulation per revolution (upr), low-pass, Gaussian filter was applied.

Establishing a geometric reference or "nominal geometry" from which deviations are presented, and limiting the wavelength content or "filtering" the data set are two key aspects in the unification scheme. These two analysis processes must be properly defined, specified and performed in order to obtain reproducible measurement results.

Once these processes are performed, the subsequent analysis of the data sets becomes a matter of numerically characterizing deviations from the reference figure. In fact, by extracting the reference figure from the data set, subsequent analyses become independent of the reference figure and hence the analysis of circular data sets can be performed in the same manner as linear data sets.

Both of these topics (reference figures and wavelength limitation) will be discussed in further detail later in this chapter. However, it is important to note at this point that these aspects have implications for both the selection of instrumentation prior to the measurement and for the subsequent analysis of the measured data points.

2.1.3 Surface Functionality

Historically, there has been a tendency to separate two aspects of surface functionality - namely the geometric structure (in terms of the physical boundary) and the sub-surface structure (typically in terms of damaged layers) (Whitehouse 1995).

In a philosophical sense, what constitutes the *real* surface is based on that which is sensing the surface or is in contact with the surface. In a contact situation, such as in a contacting stylus based measurement, the *real* surface is the geometric surface composed of physical peaks and valleys. In some applications, for example in the optics field, the *real* surface is somewhat below the geometric surface. A similar case can be made for surfaces in x-ray applications.

In many modern applications, the surface and sub-surface attributes are considered to be related. This assumes, that through the characterization of the surface, some sub-

surface properties can be inferred. For example, in the manufacture of highly polished rolling interfaces sub-surface properties are analyzed extensively during the manufacturing process development phases. However, once a manufacturing process has been fully developed and full production commences, only surface properties are evaluated on a relatively frequent basis. When significant changes in surface attributes occur, subsequent sub-surface analyses are then performed.

The limitation of wavelengths for the purpose of isolating functional aspects relates to both surface and sub-surface attributes. In mechanical applications, surfaces are related to functional aspects such as localized contact areas, lubrication retention and dynamic properties such as rolling or sliding. Wavelength separation allows further understanding and refining of the system. For example, many systems are more tolerant of very long wavelengths and very short wavelengths than they are of medium wavelengths. One example of a "wavelength-dependent" component would be an internal combustion engine piston ring. The piston ring is designed to conform to the long wavelength *form errors* of a bore (Munro and Hughes 1970) and the *roughness* of the bore provides lubrication (Trautwein 1978). Middle wavelengths (commonly referred to as *waviness*), which are present in the bore, are not easily accommodated by the conforming action of the ring and result in localized loading, exaggerated wear areas and potential blow-by of combustion gases (Hager 1995). Similar results have been shown in the analysis of conformable bearings (Bhargava 1991, Crooks and Parker 1996) where ability of a bearing to conform to a surface is a function of the wavelengths present in the surface.

The limitation of wavelengths can also relate (indirectly) to the functionality of the sub-surface for machined surfaces (Ghabrial 1991). Sub-surface properties are often related to the extended life of a surface in terms of failure modes such as spalling or fretting (Tallian 1991 & 1993). In most common applications, long wavelength surface attributes are typically related to shallow surface damage and shorter

wavelength surface attributes tend to relate to deeper surface damage. This concept can be symbolically expressed as:

$$d \propto \frac{1}{\lambda} \tag{2.1}$$

where d is the depth of sub-surface damage and λ is the surface wavelength.

Ultimately a surface's functionality may be dependent on surface properties, sub-surface properties or a combination thereof. In general, functionality is the result of a "surface system" incorporating surface geometry, sub-surface properties, tribological considerations and environmental influences.

2.1.4 Manufacturing Process Control

Surface texture (roughness, waviness and form errors) is commonly referred to as the "fingerprint" of the manufacturing process (Whitehouse 1994). Thus, through the exploration of the surface texture of a component, certain aspects of the manufacturing process responsible for its generation are made known. Many manufacturing process related artifacts are translated directly into the workpiece and these artifacts are more clearly exploited through proper wavelength limitation.

The use of band-pass wavelength limitations on surface data provides information which can lead to improved control over the individual aspects of the manufacturing process. For example, long wavelength surface features are generally a result of the frame or structure of a machine tool. Straightness and alignment of guideways, flexure of the machine tool, or thermal effects can induce long wavelength errors on the surface of the workpiece. Medium wavelength attributes can typically be related to a vibration which was present in the process. A less recognized source of waviness can be related to a vibration which was present in the dressing of an abrasive which will be subsequently used in the machining of a component. Finally, short wavelengths in the

surface are generally the result of the actual material removal process in terms of the tool's tip radius or abrasive (Mummery 1992).

2.2 A Unified Specification Scheme

Applications of surface metrology have diverged due to advances in instrumentation resulting in many varying approaches to measuring the *same* surface attributes. Concurrently, the increased understanding of the functional importance of surfaces has required tighter specification and control of these surfaces. Thus, the unification scheme manifests itself in the specification of surface attributes. This "unified specification methodology" will allow for the tighter control of surfaces from the functional standpoint along with a clear definition of surface attributes from the metrology perspective.

The unified specification scheme for surface attributes has been developed based on nominal geometry, wavelength (or frequency) transmission and parameterization. The specification approach is applied in the format given in Table 2.1. Each of the areas of the specification table will be specifically addressed in terms of options and requirements in the following sections. It should be noted at this point, however, that the table is "generic" in that it can be used in the specification of any surface attribute.

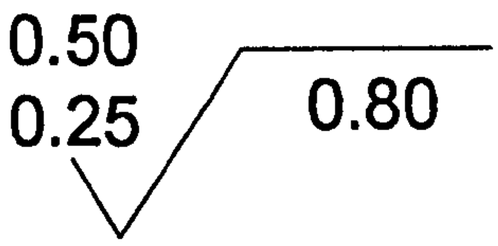
Nominal Geometry	Wavelength or Frequency Limitation	Wavelength or Frequency Limitation
Parameter 1	Lower Limit 1	Upper Limit 1
Parameter 2	Lower Limit 2	Upper Limit 2
...	...	...
Parameter n	Lower Limit n	Upper Limit n

Table 2.1 The elements of the unified specification approach.

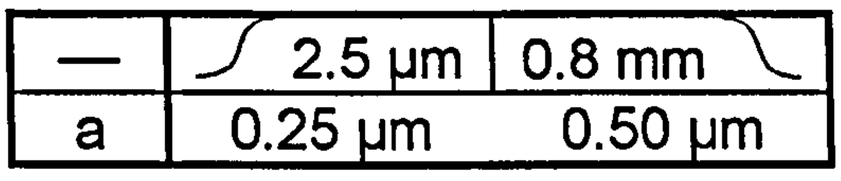
At an overview level, this concept of specification may appear to be rather simplistic, however in application it embodies a significant advance in setting forth a "language"

for communication regarding surface characteristics. Furthermore, the rather "simplistic" nature of this specification approach can be advantageous in light of the above mentioned *acceptability* criterion.

Prior to a detailed explanation of this specification approach, examples of its application to roughness and roundness may be useful. In historical (and current) practice, surface roughness is specified in the following manner (ISO 1302 - 1978):



where the roughness cutoff (λ_c) is 0.8 mm and the tolerance on the average roughness parameter, Ra, is from 0.25 to 0.50 μm. This specification approach is not clear in terms of any short wavelength boundary and only accommodates one (roughness) parameter. Under the unified approach, this specification would be represented as follows (see sections 2.4 and 2.5 below for filter and parameter indications):



In practice, many more parameters are being used in the characterization of surface roughness. The "parameter list" approach encompassed in the unification scheme can easily accommodate additional parameters whereby the historical approach is very limited. (Numerical parameters are discussed in further detail in Chapter 6.) For example, in the specification of camshaft surface profiles the average roughness parameter, Ra, is often specified in conjunction with the average peak to valley parameter, Rz (DIN). In this case, the historical approach would require the addition of a "note" to the drawing.

In mechanical drawing practices, the addition of textual "notes" to a drawing is something that is to be avoided as much as possible. The numerical specifications associated with the component features are to be indicated on or near the feature on the drawing. The historical surface roughness designation does not accommodate additional roughness parameters, thus the alternatives in the camshaft specification are to either 1) specify the Ra value in the typical manner and add a note for the Rz parameter or 2) incorporate both parameters in a note. The former alternative results in specifications for the same surface in two locations on the drawing and has the potential for overlooking one of the parameters. The latter alternative is typically utilized in practice, however, the entire specification is no longer "near" the specified surface on the drawing and, as stated above, this violates basic drawing practices.

Under the unified scheme, the camshaft specification can be placed in near proximity to the feature and all parameters can be accommodated.

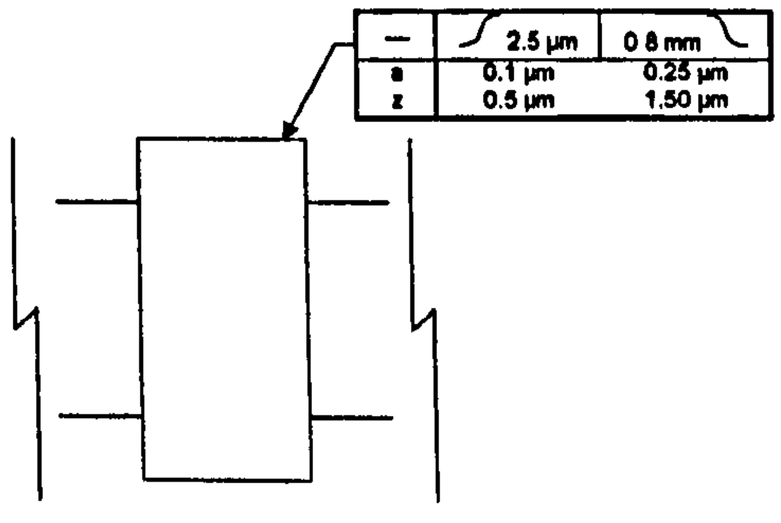


Figure 2.4 *An application of the unified specification table.*

The specification of roundness is also very common in industry. Often a roundness indication on a drawing takes on the following form (ANSI B89.3.1 1988):

○	1.0 LSC	50	0.5

where the peak to valley out of roundness tolerance is 1.0 μm measured relative to a least squares reference circle using a 50 undulation per revolution (upr) filter and with a tip radius less than or equal to 0.5 mm. This same specification under the unified approach takes on the following form:

O		50 upr
t		1.00 μm

The unified specification scheme removes the tip radius from the callout as the tip influences should be outside the filter characteristics when using a filter as a wavelength limitation. The least squares reference is assumed (as discussed below in section 2.3) and the omission of a high-pass limitation indicates a one sided (low-pass) limitation.

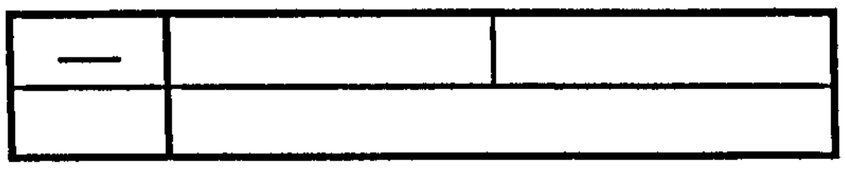
It should be noted, that wavelength designations are used in the specification of linear analyses and frequency designations are used in circular analyses. Ideally, a strict wavelength based approach may be more rigorous. However, based on current practice and the closed nature of circular data sets, frequency based specifications are perhaps more intuitively logical for roundness analyses. The user community, and ultimately standardization bodies, will have to determine the methodology for communicating the designation. (This topic is further discussed in section 2.4.4 below.)

Once again, the parameter list can easily accommodate additional parameters such as "rate of change" (commonly designated as $dr/d\Theta$) specifications which would currently require the generation of an additional "note" on the drawing.

2.3 Nominal Geometries and Reference Figures

The unified specification scheme can accommodate any two or three dimensional nominal geometry (see also Chapter 7). However, the basic geometries to be included

in this presentation of the scheme are linear and circular as these geometries represent the vast majority of surface metrology applications. The nominally straight specification is denoted by the line



and the nominally circular is designated by the circle.

The two geometries are viewed as distinct in terms of their reference figures, however, they share common principles in wavelength limitation, a common parameter set and often common waveforms (refer to Figures 2.1, 2.2 and 2.3). Surface metrology can be thought of as the measurement of surface deviations from some estimated reference figure which is established based on the nominal geometry. Thus, the reference figure is very important as it is the geometric reference from which surface errors are assessed. Variations in the reference figure translate directly to variations in resulting parameters. Furthermore, variations in the mathematical technique for applying (or "fitting") the reference to the data set can also lead to variations in computed parameters (Reason 1966, Chetwynd 1979a).

The most important principle to establish concerning nominal geometries is the means by which the geometry is applied to the data set. This issue is recognized in the field of roundness measurement where four different means for applying the circular reference geometry have been standardized: least squares, minimum zone, maximum inscribed and minimum circumscribed (ISO 4291-1985, ISO 12181-1995). Historically, minimum zone approaches have been presented in standards for both roundness and straightness per geometric dimensioning and tolerancing (GD&T) standards and practices (ISO 1101-1983, ANSI Y14.5-1994). This has, for the most part, been the

result of early analysis techniques based on the manual interpretation of polar or linear charts through the use of templates (Farago 1982) and by the concept of a "zone" of conformance in applications incorporating *mating* combinations of features such as dowels and holes. However, feedback from major instrument manufacturers indicates that, in current practice, least squares reference figures are much more common (Castle 1993, Hildebrandt 1994).

For the unified specification scheme, the means of establishing the nominal or reference geometry will be through least squares fitting. Many debates have occurred in the standards communities over "least squares" versus "minimum zone" and these debates still continue. The primary argument against least squares is based on the fact the most GD&T specifications define a "tolerance zone" in which all points on the surface much fall (Carpinetti and Chetwynd 1994). The minimum zone reference figure defines the smallest envelope and thus the smallest tolerance in which a given data set could conform. In other words, the minimum zone reference figure defines the smallest "peak to valley" departure within a data set.

Least squares based peak to valley assessments are equal to or higher than corresponding minimum zone assessments and therefore exhibit "conservative" errors (Reason 1966, Chetwynd and Phillipson 1980). Given a peak to valley specification, a component may be rejected when using a least squares reference although the component may be acceptable when utilizing a minimum zone reference. This error is *conservative* in that it ensures unacceptable parts will not be accepted, although a very small percentage of acceptable parts may be rejected.

This general application of least squares reference figures for the unified scheme is the result of many technical and mathematical considerations of which the most critical are discussed in the following sections.

2.3.1 A Unique Reference Figure

Least squares reference figures (straight or circular depending on the desired reference geometry) are unique reference figures in that (providing adequate data points and appropriate algorithms) there is only one mathematically obtainable solution to the minimization. The least squares fitting process will guarantee a unique solution when the parameters relating the dependent variable to the independent variable are linear or can be made linear (Chetwynd & Phillipson 1980, Anthony and Cox 1986, Forbes 1989, Whitehouse 1994). This linear nature is common in surface metrology and can be shown for nominally straight geometries (in cartesian coordinates) according to

$$Y = aX + b \tag{2.2}$$

in which the dependent variable, Y, relates linearly to X through the constants a (indicating slope) and b (indicating an offset in the y direction). Similarly, nominally circular geometries can be defined through linear parameters of a limaçon in polar coordinates as

$$r = R + a\cos\Theta + b\sin\Theta \tag{2.3}$$

where the dependent variable, r, relates to the independent variable, Θ, through, R (the least squares radius) and the constants a and b which relate to the center of the circle. Other linearizations for circular geometries have been presented based on a substitution of linear variables (Forbes 1989, Whitehouse 1994, Scott 1996). This linearization can be demonstrated based on the defining a residual e_i for a given point at (x_i, y_i) for a circle centered at (a, b) with a radius R.

$$e_i = \sqrt{(x_i - a)^2 + (y_i - b)^2} - R \tag{2.4}$$

The squaring the residual yields the following basis function for minimization:

$$x_i^2 + y_i^2 = R^2 - a^2 - b^2 + 2a x_i + 2b y_i + 2R e_i \tag{2.5}$$

Finally a substitution of variables is performed resulting in a linear equation.

$$z_i = \frac{\left(x_i^2 + y_i^2\right)}{2}$$

$$c = \frac{\left(R^2 - a^2 - b\right)}{2}$$

$$E_i = 2 R e_i$$

$$z_i = a x_i + b y_i + c + E_i \tag{2.6}$$

The minimum zone reference figure can, however, have multiple solutions for both circular geometries and linear geometries. The circular case was presented by (Chetwynd 1979a) and is graphically depicted in Figure 2.5. Multiple solutions for linear data sets can also occur depending on the direction in which the residuals are determined (Chetwynd 1985a, Chetwynd 1985b, 1991).

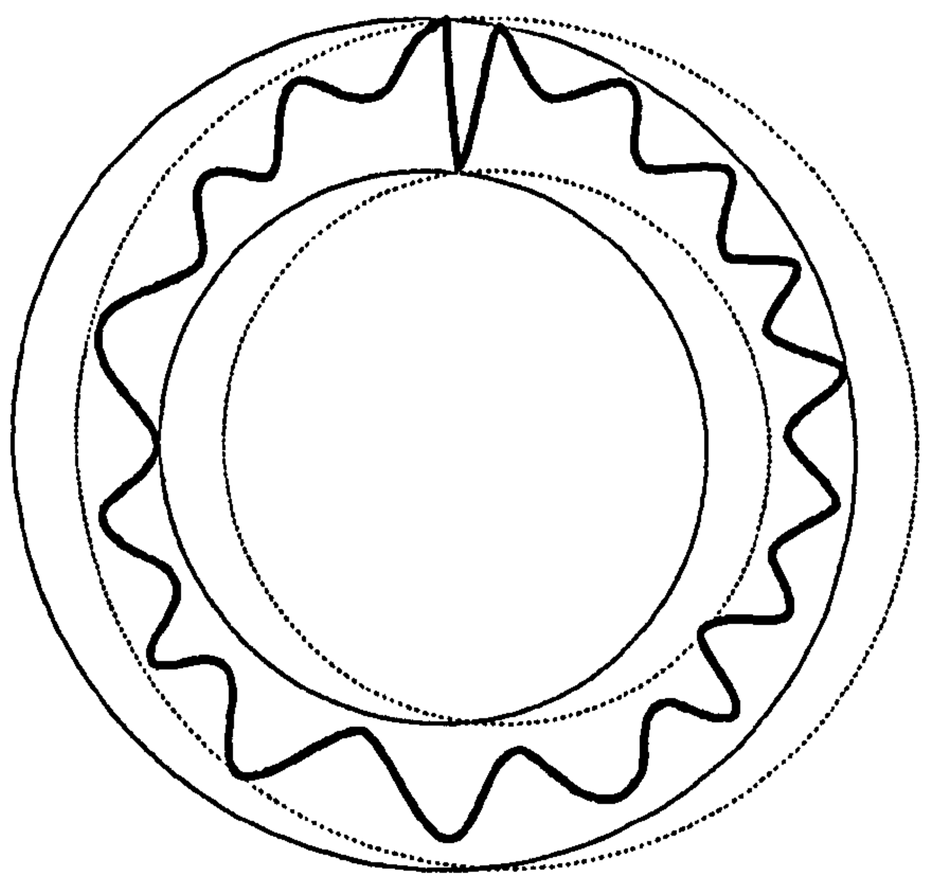


Figure 2.5 *A data set allowing two possible minimum zone circles.*

Multiple solutions can lead to a lack of reproducibility and therefore be in conflict with goal #2 of the unification scheme. Although, by definition, multiple solutions of minimum zone fitting yield identical "peak to valley" results, they do not yield identical results for other parameters. For example, in the data set shown in Figure 2.5 the total or "peak to valley" out of roundness value is the same for each reference, but local variations (for example the "rate of change") may be significantly different between the two. This is due to the fact that angular measurements vary with the position of the center (Reason 1966). This can be shown in Figure 2.6 where the angular separation, Θ_1, between the two indicated peaks relative to center 1 (C_1) is considerably smaller than the angular separation Θ_2, which is measured relative to center 2 (C_2).

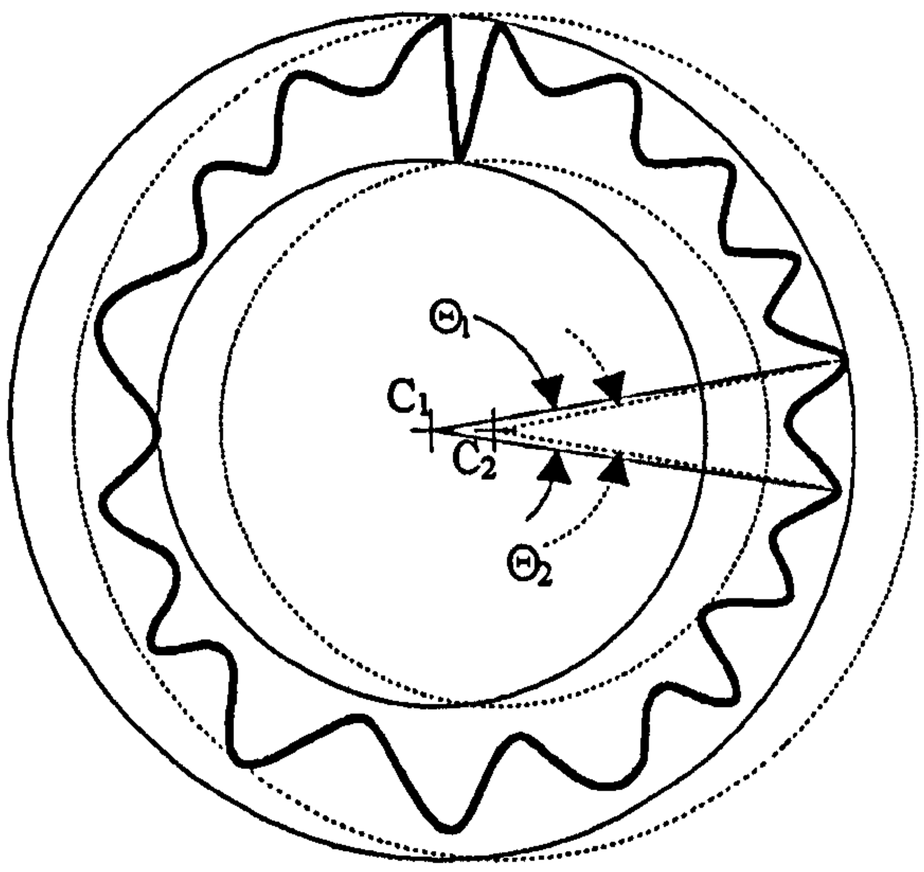


Figure 2.6 *Variations in angular separation as a function of the centerpoint.*

In addition to the problems associated with multiple solutions for certain data sets, the mathematical process of establishing a minimum zone reference can be plagued with having to deal with local minima which require special consideration (Chetwynd 1991).

2.3.2 A Stable Reference Figure

The least squares line or circle (depending on the desired reference geometry) is based on a fitting process including all of the data points. "Non-statistical" or "outlying"

peaks or valleys do not significantly influence the reference figure as they are "averaged out" to an extent. Minimum zone references are significantly influenced by these outlying points on the surface as the minimum zone approach is based solely on the extremes of the data set. Thus, the least squares solution can be considered more "stable" in the presence of extreme data points.

It could be argued that the minimum zone reference figures are more "stable" in that they are less sensitive to step functions (see Figure 2.7). This is true in the theoretical sense. However in typical surface metrology applications, "spikes" or impulses are far more common than are step functions.

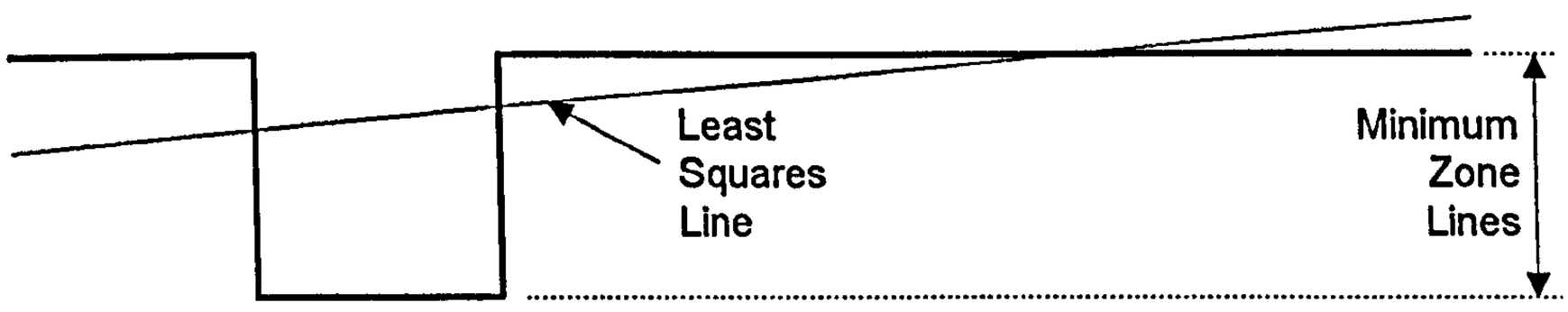


Figure 2.7 *Least squares and minimum zone reference lines applied to a "stepped" data set.*

An example of the lack of stability in a minimum zone reference figure can be shown by artificially adding a "spike" at points throughout a data set and comparing least squares and minimum zone references and parameters for each location of the spike. This example has a practical aspect in that it simulates what would occur when debris is present in the measurement process.

1. To obtain a typical profile from a critical surface, a fuel injector plunger (fine ground tool steel) was measured. A 4.0 mm trace length was used and data was collected at 0.5 μm spacing, 0.010 μm vertical resolution, incorporating a 2.0 μm stylus tip radius. (The RMS noise level was less than 0.010 μm). The

measured profile, which was first leveled based on a least squares reference, is shown in Figure 2.8.

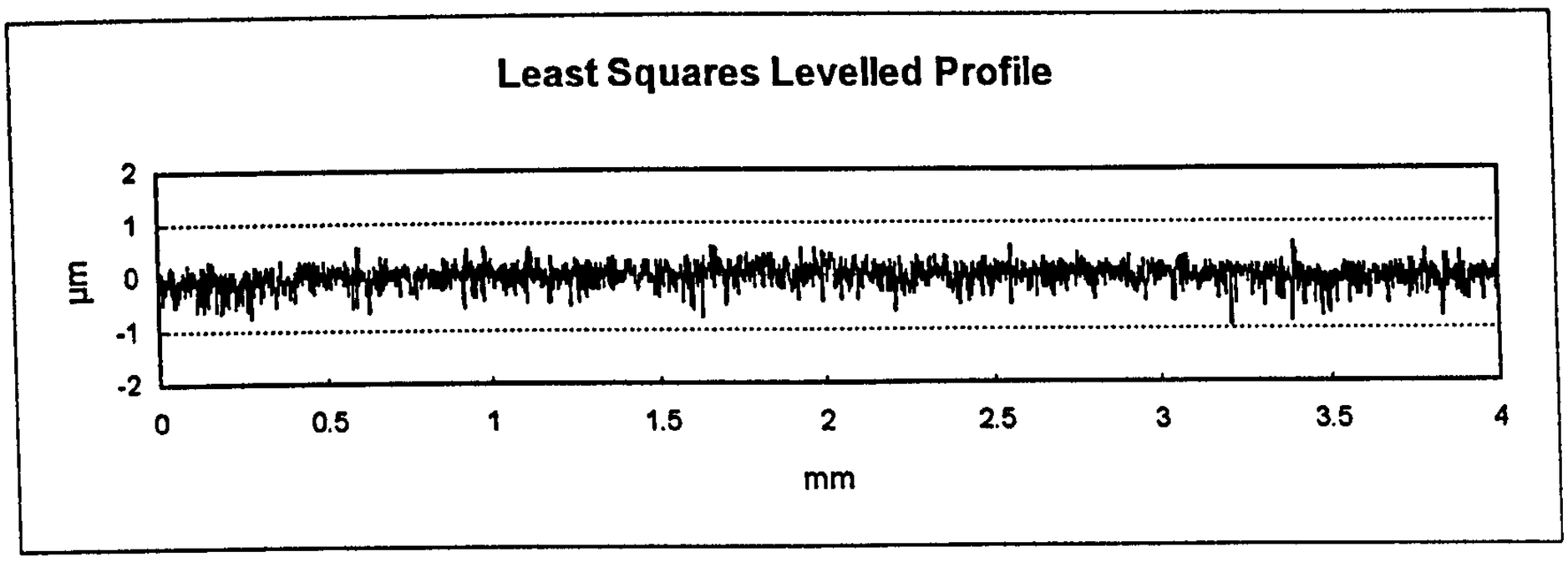


Figure 2.8 *The profile used for reference evaluation.*

2. For the comparison of reference figures, a 1.0 µm impulse (one data point wide) was moved through the profile, placing it at each profile peak. After each placement of the impulse the data was subsequently filtered with a 2.5 µm, long-pass Gaussian filter (λ_s) as would be done in practice. (The application of this filter reduces the effective height of the 1.0 µm spike to approximately 0.4 µm and broadens it slightly which is quite typical in regards to debris encountered in many metrology applications.)

3. A (linear) least squares reference figure was then calculated in addition to a minimum zone reference figure. (The "C" source code for the fitting of these reference figures is given in Appendix A.) The minimum zone fit was performed by removing (and storing) the least squares line and then assessing line segments of convex hulls for both peaks and valleys (Traband et al. 1989). Each segment of the upper and lower hulls was evaluated as a candidate for the minimum zone solution and residuals were assessed in the measuring (vertical) direction. The resulting minimum zone solution was then combined with the suppressed least squares reference to establish the final minimum zone reference figure.

4. The slopes were compiled into a histogram by which the variations in the least squares reference could be compared to those of the minimum zone reference. These histograms are shown in Figure 2.9.

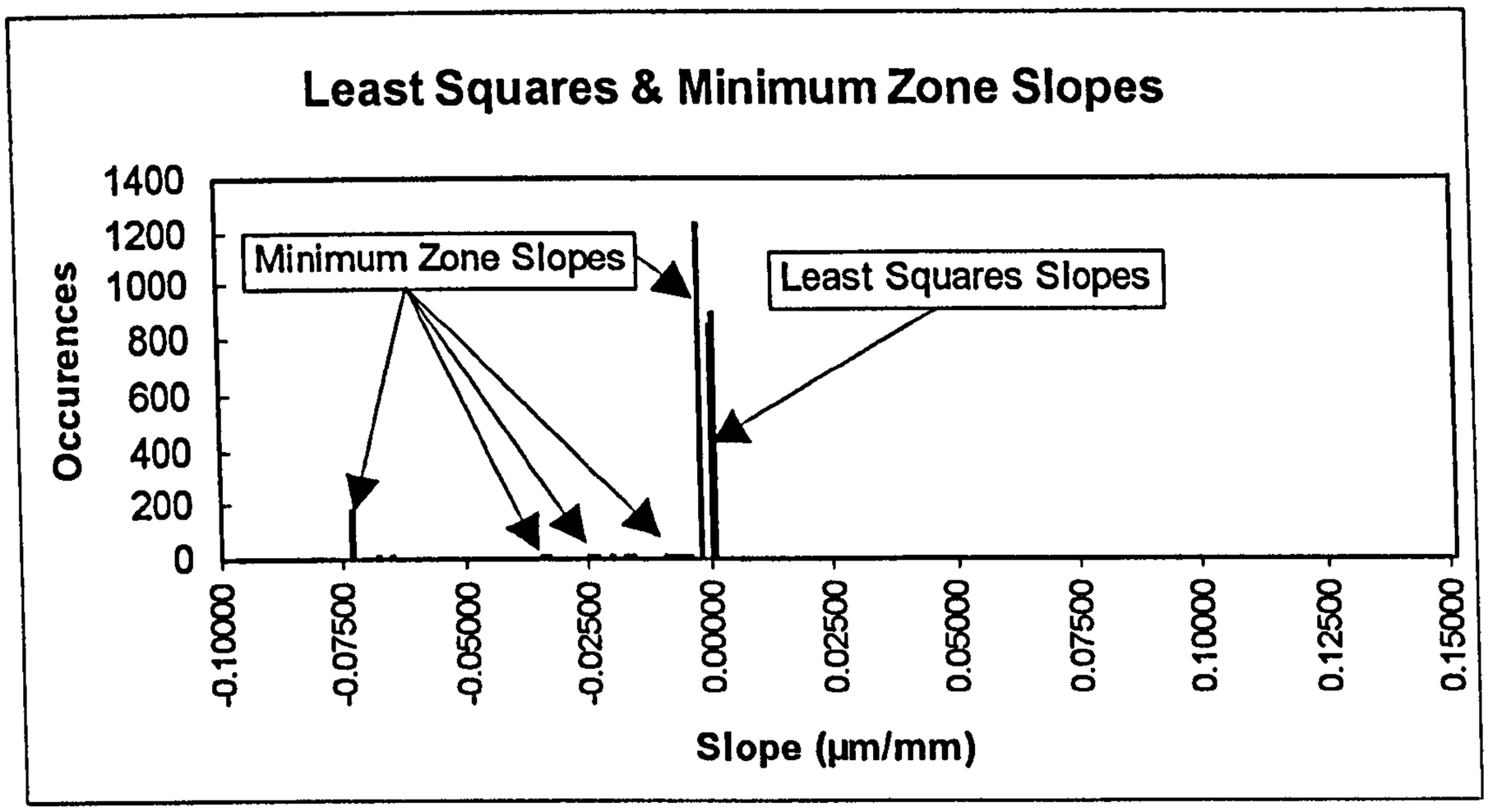


Figure 2.9 *Variation in least squares slopes and minimum zone slopes.*

The representation of slopes in Figure 2.9 is scaled based on histogram bins related to the high occurrences. In this representation, it is difficult to see the detail related to slopes which occur less often. By expanding the vertical axis (and truncating the histogram peaks) we can see more detail in terms of the range of slopes generated by each method. Figure 2.10 re-displays the histogram of slopes in this manner. Through this view, we see the significant variation in minimum zone slopes as compared to the relative stability of the least squares slopes.

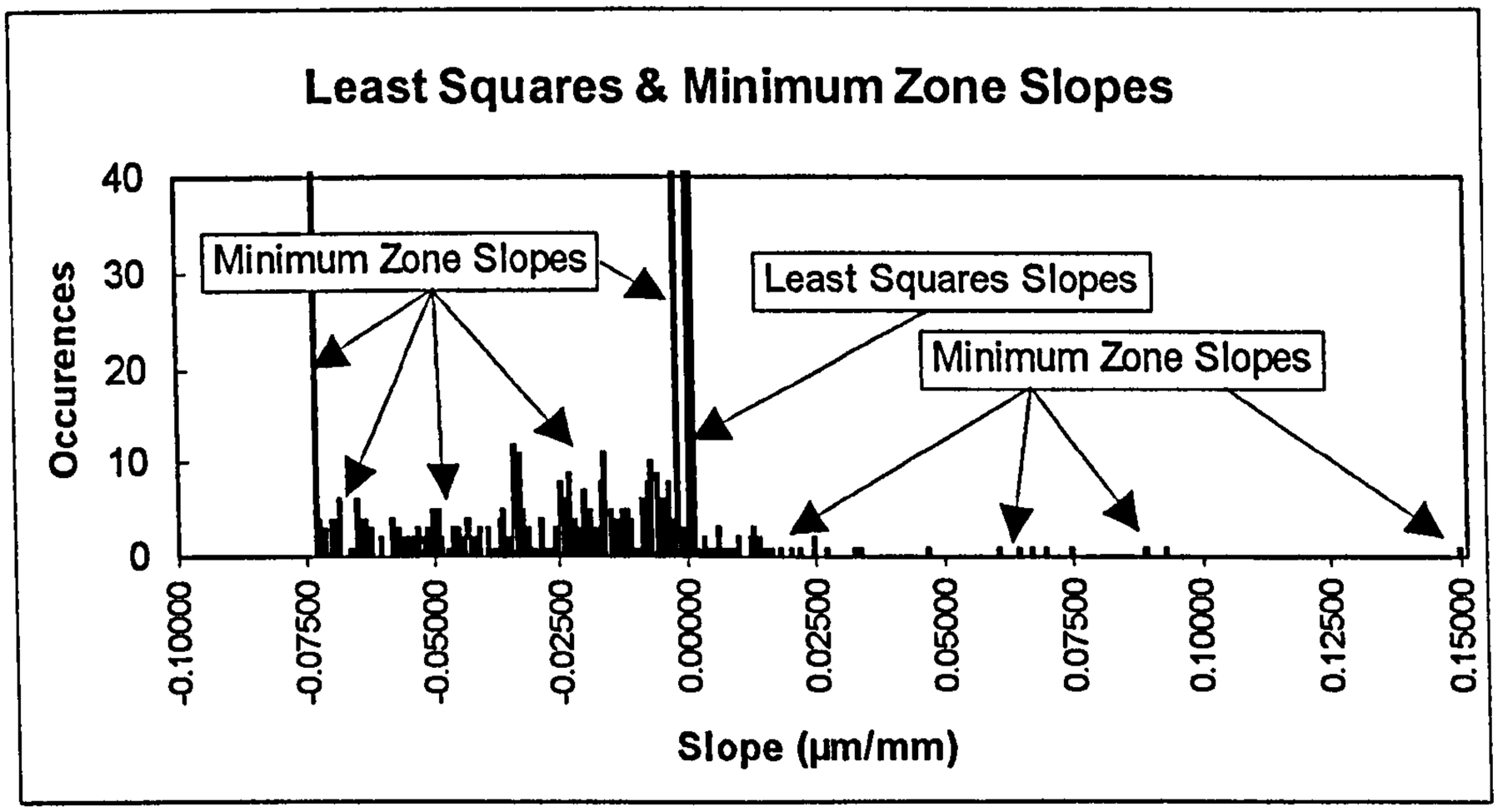


Figure 2.10 *Variation in least squares slopes and minimum zone slopes. (Vertical axis expanded, thereby truncating the upper region of the data.)*

The least squares slopes fall within only two bins of the 250 which have been established based on the range of minimum zone slopes. This variability in slopes is graphically presented in Figure 2.10. In numerical terms, the standard deviation of the minimum zone slopes was 236 times greater than the standard deviation of the least squares slopes. However, since the data was not normally distributed a simple comparison of ranges may be more indicative. The least squares solutions ranged from -0.000188 to 0.000187. The minimum zone solutions ranged from -0.074057 to 0.148789. Thus, the range of the minimum zone slopes was nearly 600 times greater than the range of least squares slopes!

Variations in the reference figure can directly translate to variations in computed parameters. Figure 2.11 indicates that the variations in the least squares assessment of peak to valley height are similar to those of the minimum zone approach. Although the minimum zone approach is based on minimizing the peak to valley height of the profile, we see very little difference between the minimum zone and least squares results.

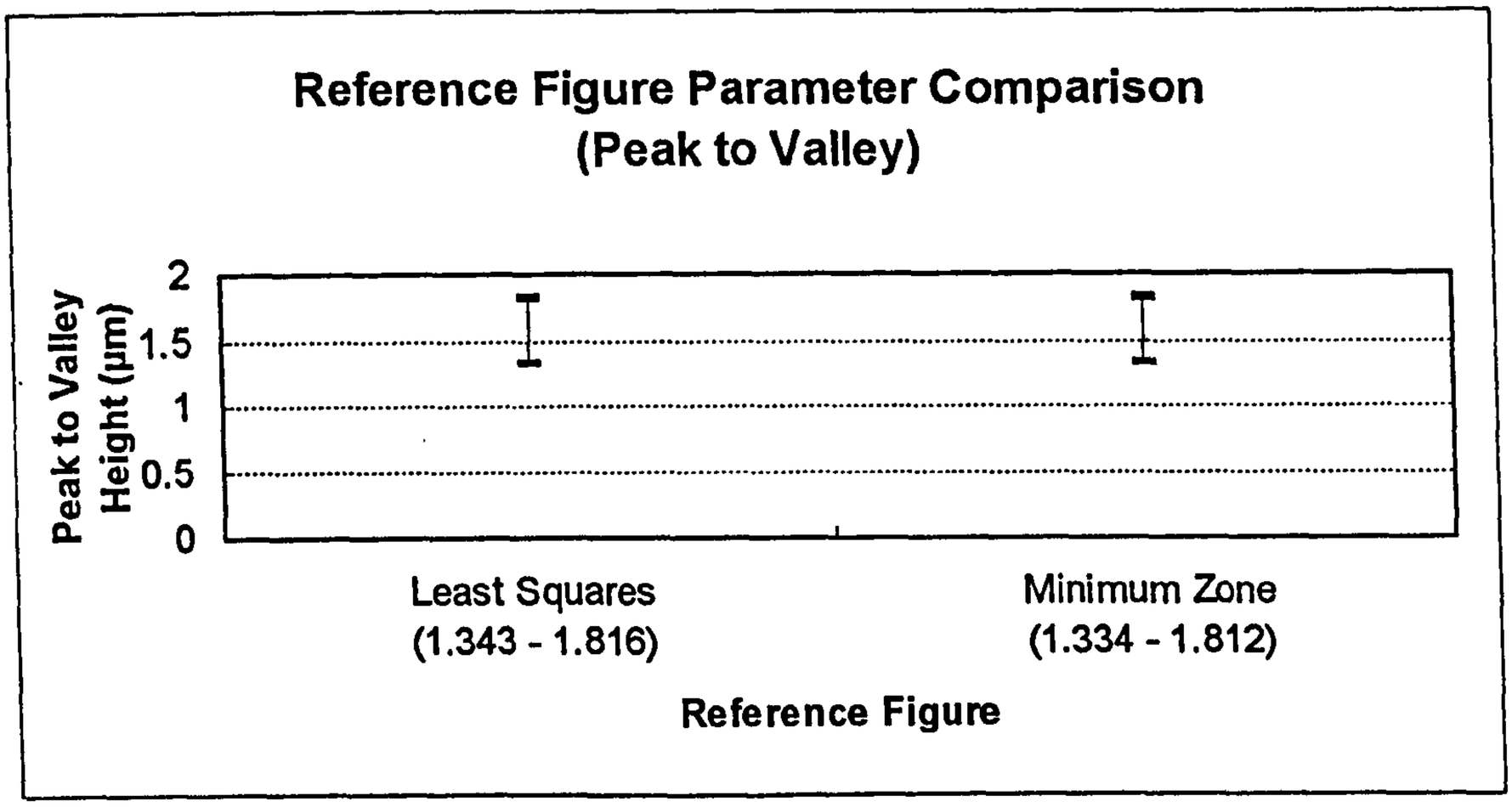


Figure 2.11 *Peak to valley variations associated with reference figure variations.*

A significant difference can be observed in the calculation of the RMS amplitude. This is shown in Figure 2.12, where variations in the RMS amplitude calculated relative to the least squares reference are negligible. (Note: for the minimum zone reference figure, RMS deviations were calculated based on the mean amplitude relative to the minimum zone lines.) The RMS amplitude based on the minimum zone reference varies from the least squares value up to 136% of the least squares value.

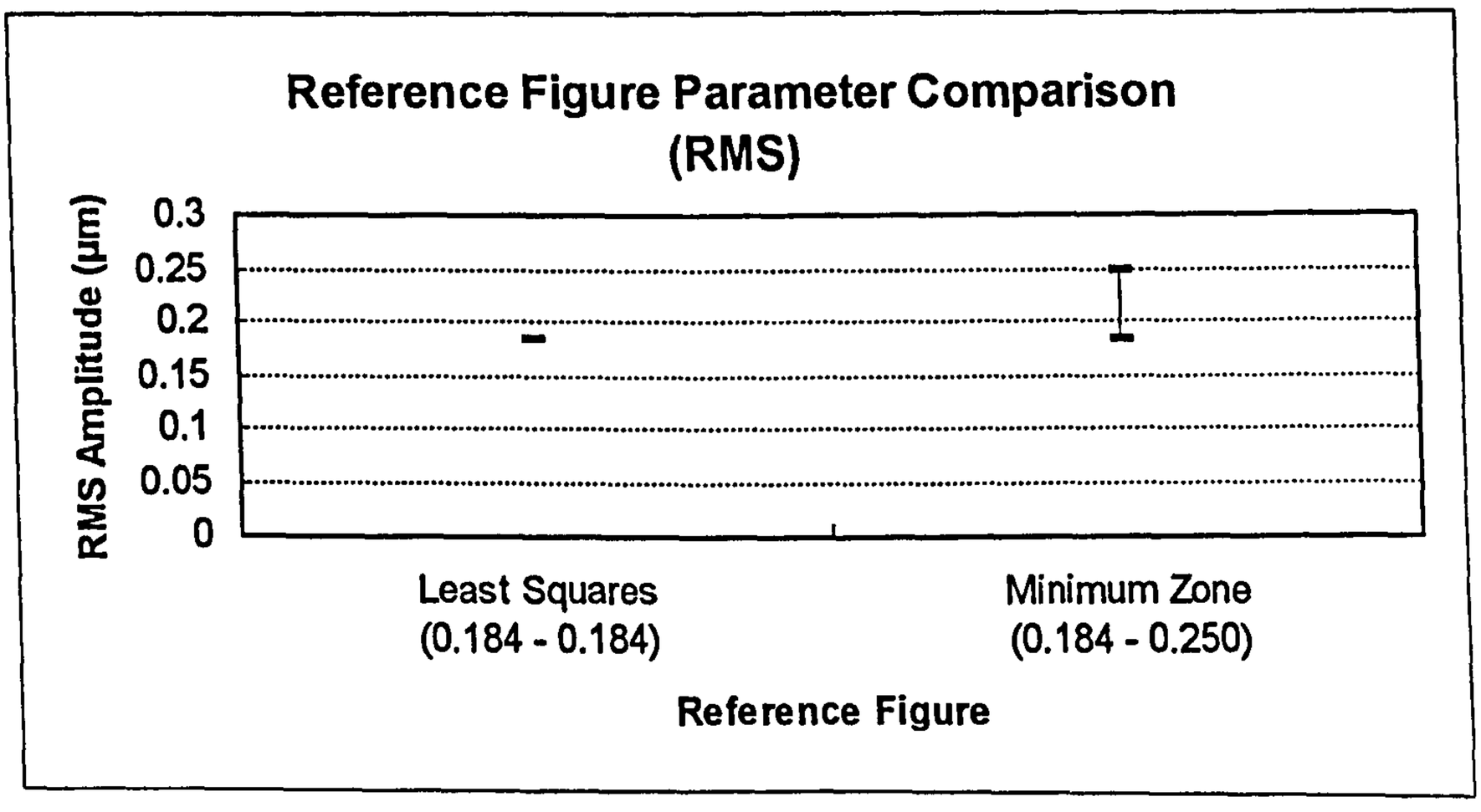


Figure 2.12 *RMS amplitude variations associated with reference figure variations.*

The establishment of a reference figure by least squares fitting is clearly more stable in the presence of extraneous peaks both in terms of the coefficients of the geometric reference figure and the resulting parameters. This has been confirmed through practical laboratory and manufacturing experience in dealing with both straight and circular geometries.

2.4 Wavelength Limitation

The development of common means for wavelength transmission is also essential in obtaining correlation between instruments. When systems do not detect the same wavelengths and represent them at the same amplitudes, the systems will not produce correlating results for all surfaces. However, not all instruments exhibit the same frequency response. For example, currently available roughness instruments exhibit adequate frequency response in the range from relatively short (micrometer) wavelength roughness measurement up to straightness measurement over a few tens or sometimes hundreds of millimeters. Typical coordinate measuring machines (CMM's) currently measure wavelengths ranging from a few millimeters to thousands of millimeters. These instruments have very different frequency response characteristics, but there is an area of overlap in their bandwidths. In this area of overlap, correlation can be obtained through imposing similar, band-pass wavelength limitations. This can be accommodated through the use of digital filters or other means of limiting the transmission of wavelengths outside the desired "band". This band-pass limitation of surface wavelengths also has functional and process control implications as discussed above.

Given the importance of band-pass wavelength limitation in terms of surface functionality, process control and instrument correlation, the unification scheme provides three primary means of limiting wavelength. The three alternatives put forth in this scheme for unification are based on the fundamental needs of the user

community and the fact that no single means of wavelength limitation is suitable for every application (see also Chapter 5). Each of the means is included based on its technical merit in an area of characterization not readily achieved by any of the others.

- Gaussian Filter Wavelength Limitation

  General spatial domain applications.

- *Ideal* Wavelength Limitation

  Frequency or wavelength critical applications.

- Radius Based Wavelength Limitation

  Rolling or contact critical applications.

Each method of wavelength limitation has advantages and disadvantages and thus the proper selection must be *application specific* based on an understanding of what information is to be obtained from the surface. The three methodologies will be discussed briefly in the following sections and a more detailed discussion will follow in Chapter 5.

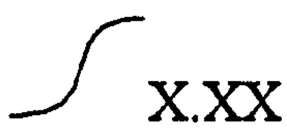
2.4.1 Gaussian Filter Wavelength Limitation

In recent years the use of Gaussian filters has rapidly gained popularity in the analysis surface metrology data sets (ISO/TC 57/SC 1 - 1988, Whitehouse 1994, ISO 11562 - 1995). The filter first gained popularity in the measurement and analysis of roughness data (ISO/TC 57/SC 1 - 1988), but more recently Gaussian filters are being applied in roundness and straightness measurement (Rank Taylor Hobson 1992). Some of the important characteristics of the Gaussian filter include the fact that it is phase correct or "zero phase" and is analytically expressible in both spatial and frequency representations (see Chapter 5). Gaussian filters are easily implemented digitally by the convolution of a Gaussian weighting function or by faster means such as triangular approximations. More importantly in the context of the unification scheme is the fact that the Gaussian filter exhibits the sharpest attenuation characteristics while

maintaining smooth transitions in both space and frequency without "ringing" or "overshoot" (Scott 1995).

These technical characteristics of the Gaussian filter (see Chapter 5 for further details) combined with the very high level of acceptance by the user community make this approach an obvious selection to be included in the scheme. Furthermore, the Gaussian filter should be evaluated first when determining a wavelength limitation methodology. One of the other two methodologies should only be selected when the Gaussian filter cannot yield acceptable results.

In the unified specification scheme, the Gaussian filter is represented graphically by a symbol relating to its *S* shaped transmission characteristic curve. In the analysis of linear geometries the indication

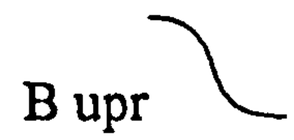


describes a Gaussian filter with a cutoff of X.XX mm, transmitting long wavelengths. The indication

indicates a short wavelength transmitting Gaussian filter with a cutoff of Y.YY. This "left to right" format of cutoffs and transmission characteristic symbols is "ergonomic" in a sense in that the two can be combined to both graphically and numerically define a pass band.

In the analysis of circular geometries, a frequency based specification is adopted. This, in a sense, reverses the left-to-right wavelength connotation, but maintains ascending numerical order in frequencies. Thus, the indication

describes a high frequency transmitting Gaussian filter of cutoff "A". Similarly, the specification



indicates a low frequency transmitting Gaussian filter of cutoff "B".

In an example drawing callout of the form

—	2.5 µm	0.8 mm
a	0.25 µm	0.50 µm

we see that a band-pass filter is specified incorporating a Gaussian band-pass wavelength limitation. The short wavelength limitation occurs based on a cutoff of 2.5 µm and the long wavelength limitation is based on a cutoff of 0.8 mm. Thus two Gaussian filters will be employed in this wavelength limitation - one at 2.5 µm which transmits long wavelengths and one at 0.8 mm which transmits short wavelengths. This example specification in is consistent with the recently established 300:1 bandwidth specified in international standards (ISO 3274 1995).

2.4.2 *Ideal* Wavelength Limitation

Note: The term "ideal" in this context does not imply that it is ideally suited for all applications. The term "ideal" (bounded by quotation mark or italicized) is used in the

context of the digital signal processing (Antoniou 1993) and in the scheme for unification to refer to the mathematical process of modifying the frequency content of a data set whereby certain wavelengths are retained fully and others are removed fully. While this methodology is "ideal" in the Fourier sense (frequency or wavelength domain), there are significant disadvantages in terms of "ripples", "ringing" or "overshoot" in the spatial domain. In practice, this methodology has also been referred to as "brick wall" filtering.

In some cases, for example in roller or ball bearing applications, there are sensitivities to certain wavelengths or frequencies (Yhland 1967, SKF 1992). For example, in a ball bearing raceway, if a predominant surface wavelength occurs at a multiple of the number of ball bearings, there is the potential for significant vibration. In these applications, which are sensitive to discrete wavelengths or frequencies, it is often useful to characterize the entire amplitude at these wavelengths through Fourier analysis. While Gaussian filtering is useful and very common, it has a rather gradual transmission function whereby all wavelengths near the cutoff of attenuated to some degree. This "gradual" transmission characteristic is required to ensure no "overshooting" in the spatial domain, however this produces adverse affects such as "leakage" in the frequency or wavelength domain. In a strict Fourier or *ideal* wavelength limitation approach, all wavelengths (or frequencies) outside some boundary are set to zero and all wavelengths (or frequencies) inside are transmitted completely. In application, this filtering approach is realized through the use of a Fourier transform and the scaling (either 0% or 100%) of the resulting coefficients.

The *ideal* wavelength limitation approach is included in the scheme for unification based on the above described needs of the surface metrology community. However, it should be noted that extreme caution should be taken in its application as considerable distortions (Gibbs' effects) can often occur. Given its wavelength or frequency transmission, this approach is designated by symbols indicative of a step function. The symbol

indicates an *ideal* wavelength transmission fully transmitting (100%) the amplitude of all wavelengths which are equal to or longer than X.XX. The amplitudes associated with all wavelengths shorter than X.XX are set to zero. The corresponding long wavelength limitation is designated as

whereby, amplitudes are fully (100%) retained for wavelengths shorter than or equal to Y.YY and amplitudes are set to zero for wavelengths longer than Y.YY.

In circular analyses the *ideal* specification follows the same frequency domain convention as does the Gaussian in terms of the positioning of the transmission designation symbol and cutoff value.

2.4.3 Radius Based Wavelength Limitation

In many interfaces it is necessary to determine the dynamic effects of a rolling or sliding element moving across a surface. Consider, for example, a cam/follower application incorporating a rolling element in the follower. In this scenario it is very important that there are no cam surface features that could cause unwanted follower dynamics. To assess the cam surface geometry as it would interact with the follower, the cam surface wavelengths should be transmitted according to the physical "convolution" of the follower radius. (Note: "convolution" in the context of a radius contacting the surface refers to a more "mechanical" type interaction primarily based profile peaks, where in the mathematical sense a convolution is based on all profile points.)

This "tip radius" or "stylus" based transmission has typically been referred to as the *envelope* or *e* system of measurement (Von Weingraber 1956, Von Weingraber 1957, Shunmugam and Radhakrishnan 1975). Often this is achieved by measuring the surface with a stylus tip radius equal to the radius of the following element that will be used in the surface's application (Adcole 1993, Bhargava 1993) however, mathematical approaches have also been implemented. Chapter 3 presents some of the differences between the mathematical approach and physical measurements incorporating the desired tip radius.

The convolution of a radius over a surface is quite unpredictable in terms of a transmission characteristic in a frequency or wavelength representation (as will be shown in further detail in Chapter 3). Nonetheless, the mechanical convolution is important in terms of the assessment surface functionality. In the metrological context, the convolution of a radius over a surface, either physically or mathematically, produces unpredictable, but repeatable and reliable results. In other words, the transmission function cannot be predicted *a priori* but for a given surface location and assessment radius the transmission function is reproducible. This means that different instruments incorporating the same tip radius or mathematically convolved radius should arrive at a similar transmission function providing that similar locations of the surface are measured and the instruments exhibit the same basic frequency response.

It should be noted that the radius based wavelength limitation is only a short wavelength limitation (transmitting only long wavelengths). The drawing indication is

R: X.XX

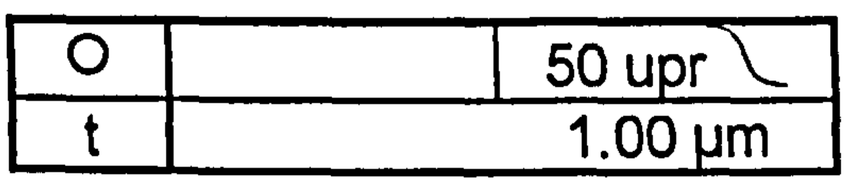
whereby a tip radius *X.XX* is used in the measuring process or a circle incorporating a radius *X.XX* is numerically convolved over a data set. In the mathematical alternative, the original profile must result from a measurement which incorporated a tip radius less than or equal to *X.XX*.

A very important aspect of the radius based approach is that it must be applied first when used in conjunction with other filtering methodologies. This is based on the fact that the radius convolution is based on interactions with profile peaks and any modification to the peaks (such as through the application of a filter) will result in differences in the resulting profile. This also promotes consistency between mathematical and physical convolutions in that a physical convolution (such as by the stylus tip) will always occur before any digital filtering.

As was the case in the *ideal* wavelength limitation approach, the radius based limitation should only be employed when the Gaussian filter does not yield an acceptable representation of the profile.

2.4.4 Wavelength Limitation in Roundness Analysis

Wavelength limitation for nominally straight features readily lends itself to wavelength specification in units of length. The assessment of roundness, due to the *closed* nature of the data set, is often based on frequency terminology and specification as opposed to wavelength-based terminology and specification. This is acceptable under the unified specification system. For example, a typical roundness callout requiring a 50 undulation per revolution (UPR), low-pass Gaussian filter would have the following form:



It is important to note that the roundness specification is given in terms of frequency rather than wavelength. Thus, the 50 upr Gaussian filter in the above example is a *low-pass* filter (significantly attenuating frequencies above 50 upr). The incorporation of the units (upr or length) and reference figure in the unified specification format provide a means of distinguishing the domain.

In roundness specifications, typical assessments include all frequencies from 2 upr (1 upr representing eccentricity based on the least squares reference figure) up to the cutoff frequency. Thus, if the high-pass characteristic is omitted (as show above) the following transmission is assumed:

O	2 upr	50 upr
t		1.00 µm

Another aspect of the unification scheme which is unique to roundness analyses is the application of radius based wavelength limitation. In the specification of roundness, the radius based wavelength limitation should be placed in the right hand side of the specification table.

O		**R**: 0.5 mm
t		1.00 µm

This placement is consistent with the frequency notation used in roundness designations under the scheme for unification. (The radius based approach will generally attenuate high frequencies.) Furthermore, this accommodates a band limitation whereby filter can be employed to remove low frequency components.

O	15 upr	**R**: 0.5 mm
t		1.00 µm

In this specification, high frequency surface features will be attenuated according to the 0.5 mm radius convolution and surface features associated with frequencies lower than 15 upr will be attenuated according to a Gaussian filter. Once again it is important to note that the radius limitation must be applied (either physically or mathematically) to the data set prior to the Gaussian filter.

2.4.5 One Sided Wavelength Limitation

The most common measurements of roundness can be thought of a *one-sided* in terms of their frequency transmission (i.e. all frequencies which are lower than the cutoff frequency are included). A similar situation may occur in the assessment of the long wavelength attributes of nominally straight features. In such cases, it not uncommon to find that the wavelengths up to that of the entire length of the component must be controlled (upon application of a least squares reference line). Traditionally, this has been referred to as a "straightness" callout. This one-sided callout is accommodated by the omission of the long wavelength limitation from the drawing callout.

—	2.5 mm	
t		X.XX µm

The above specification indicates, that given a measurement over the entire length of the component, a 2.5 mm Gaussian filter will be applied for short wavelength limitation and all longer wavelengths will be included up to that of the length of the component.

2.5 A Generic Parameter Set

Given that profile-based surface metrology parameters are typically numerical characterizations of a digital signal, there should be a means by which a parameter set could be developed which is applicable to any surface metrology data set. Historically, parameters have been developed on an *as-needed* basis, targeting very specific applications (Spragg and Whitehouse 1970, Whitehouse 1982). Often these parameters were based purely on the ease of their measurement. This has resulted in very large number of available parameters; many of which are inter-related and many more are of little functional relevance. Furthermore, the parameters historically used in one regime are not necessarily the same as those which are used in another regime.

To unify the parameterization of surface metrology profiles, a basic parameter set will be developed. This development is presented in detail in Chapter 6, but for the purposes of this overview it is useful to discuss it briefly. It should be noted that the parameter set developed in this work is by no means comprehensive or exclusive. Goal #5 of the unified approach was to allow flexibility, thereby allowing the incorporation of additional numerical characterizations resulting from the (justifiable) development or application of additional parameters.

The determination of a base set of parameters is founded on the categorizing of the types of numerical information which is typically desired. The chosen parameters and corresponding numerical values may vary from application to application, however the *types* of parameters are consistent. The categories of parameters chosen for the basic unified set are

- Statistical amplitude parameters.
- Extreme amplitude parameters.
- Spacing parameters.
- Slope and *shape* parameters.
- Auxiliary functions and parameters.

The individual parameters will be given a designation similar to those commonly in use and historically standardized. However, in most cases, the leading profile indication letter (R, W, P, etc.) which has historically indicated the wavelength regime will be dropped as we are now operating in a specified bandwidth context. Thus, a typical roughness specification based on an Ra tolerance of 0.25 µm to 0.50 µm would take on the following form.

—	2.5 µm	0.8 mm
a	0.25 µm	0.50 µm

The "a" in the parameter list indicates "average" corresponding to the historical "Ra" designation. The average amplitude, "a", parameter could be applied to any data set including those obtained in roundness measurement. Furthermore, in the above example the linear nominal geometry (as indicated by the "-") and wavelength range (per the two Gaussian filters) are explicitly described, thus the leading "R" is redundant and is therefore omitted from the scheme.

The parameter list accommodates the parameter name as well as upper and lower tolerance limits for the given parameter. In cases where further information is needed pertaining to the assessment of a particular parameter, this information is provided along with the name of the parameter. An example of this need for additional information is in the specification of the rate of change parameter, historically referred to as "dr/dΘ" (Whitehouse 1987a, Rank Taylor Hobson 1992), which is typically used in roundness analysis in the bearing industry. This parameter indicates the maximum peak to valley height which occurs within any angular window of width *Θ*. The desired angular "window" must be specified based on the application. (Note: Under the unified parameter set (see Chapter 6) the designation for this parameter is changed to dr/Θ to avoid any confusion with the actual mathematical process of differentiation) As an example, a roundness specification incorporating both peak to valley and dr/Θ callouts, could take on the form

O	2 upr	50 upr
t		1.00 µm
dr/Θ (30°)		0.30 µm

whereby the total, "t" (peak to valley), out of roundness is not allowed to exceed 1.0 µm and the rate of change parameter dr/Θ is not to exceed 0.3 µm within any 30° window.

2.6 Summary of the Scheme for Unification

In the above sections, the framework and many of the components of a generalized methodology for unifying surface metrology have been put forth. It is important, at this point, to summarize these aspects prior to moving forward with the supporting areas of development.

The status of applied surface metrology has been presented in terms current "divergence". In order to address this divergence a comprehensive "scheme" for unification has been developed. This scheme or methodology is a set of rules and an overall "language" for specifying surface features. Broad application of this methodology, preferably through international standardization, will provide a common means of specifying, measuring and characterizing surface attributes independent of the instrumentation and nominal component geometry.

The methodology begins with a new approach for specifying surface attributes (see section 2.2). This new approach accommodates the specification of nominal geometry, wavelength limitations and numerical parameters. Furthermore, this specification scheme is extensible in that it allows for the incorporation of further developments, thus preventing a future occurrence of today's situation.

- **Nominal Geometry**

  Various nominal geometries can be accommodated (for the sake of convenience and readability, linear and circular are included in this presentation).

  These nominal geometries can apply to two dimensional ("profile") or three dimensional ("areal") assessments (see Chapter 7) although this presentation will focus on profile analyses.

Surface deviations will be assessed relative to least squares reference figures. This selection is based on the above (section 2.3) presentation.

- **Wavelength Limitation**

  Perhaps the most important aspect in obtaining agreement between measurement methodologies is the establishment of a common transmission of wavelengths. The unified specification scheme provides a clear means of communicating this "transmission band".

The desired transmission band must be clearly defined. This requires that a short wavelength (or high frequency) limitation must always be specified. The long wavelength (or low frequency) can optionally be specified. If the long wavelength limitation is omitted then the entire length (or circumference) of the component is to be included.

The proposed scheme includes three means for bounding the transmission band: Gaussian filters (see section 2.4.1 above and also Chapter 5), *ideal* wavelength or frequency limitation through Fourier analysis (see section 2.4.2 above and also Chapter 5) and "radius based" wavelength limitation (see section 2.4.3 above and also Chapter 3). The former two methodologies must be digitally implemented, while the "radius based" approach can be realized mechanically or mathematically.

"Radius based" wavelength limitation (historically referred to as the *e* system) can only be applied as a short wavelength limitation. Furthermore, when digitally (or mathematically) implementing this approach in conjunction with a long wavelength limitation the radius based limitation must be applied first (see section 2.4.3 above).

The specification of the wavelength band will dictate the selection of instrumentation for a particular measurement. Instruments must demonstrate adequate transmission over the specified bandwidth. This requirement has implications for sensing (stylus tip radius and probe characteristics) as well as digitization and further numerical processing. (See Chapter 7 for a discussion of instrument evaluation techniques.)

- **Numerical Parameters**

  Given the proper application of a nominal geometry (through a least squares fitting process), and a controlled wavelength band (through the above described approaches), a set of numerical parameters can be defined which can be applied to all surfaces. The unified specification scheme defines parameters independently from the nominal geometry and wavelength regime.

  In some cases, the unified specification scheme will alter the designation for surface metrology parameters to accommodate their general application. For example, the leading "R" will be dropped from the traditional "roughness" parameters which are included in the scheme.

2.6.1 Satisfaction of Technical Goals

Earlier in this chapter, the six technical goals for the scheme were set forth. Following is a synopsis of each of these goals.

- *Reproducibility of measured results for across various instrumentation.*

  Under the unified approach, the control of nominal geometries (through least squares) and the control of the wavelength transmission band (through one of the prescribed approaches) will result in the achievement of similar results from various instruments (providing that they demonstrate adequate amplitude transmission within the desired wavelength band.)

- *A common language for surface metrology - independent of the nominal geometry and wavelength regime.*

  The unified specification scheme can apply to varying nominal geometries for both profile and areal data sets. Furthermore, the scheme presents a new parameter nomenclature which can be applied across all of surface metrology.

- *Stability of measured results in that they reliably reflect the surface features which are to be characterized.*

  This concept of "stability" and "reliability" is addressed in the context of the least squares reference features. Ultimately, the user will have to select the necessary bandwidths and parameters in order to characterize the desired surface features.

- *Common characterizations through a common parameter set.*

  The unified specification scheme includes a parameter set which can be applied to all surface metrology applications.

- *Flexibility to accommodate future developments in wavelength limitation, parameterization and analysis functions.*

  The unified specification "table" incorporates locations or "fields" for the nominal geometry, wavelength limitation, and parameters. Future developments in any of these three areas can be accommodated by merely adding a new symbol or parameter designation and placing it in the proper field in the table. Ultimately, the responsibility for controlling the addition of items to the scheme should come under a standardization body.

- *Practical attainability through available instrumentation.*

  In terms of instrumentation, the unified methodology requires that the instrument used in a measurement must demonstrate adequate wavelength transmission in the specified band. Upon satisfying this requirement, any

instrument can be used. The major implication of this requirement is that there will be an increased demand for data concerning instrument performance.

2.6.2 "Acceptability" of the Scheme

In areas where current practice proves to be acceptable, the unified methodology accommodates them directly. This is true for such items as Gaussian filters and many of the existing surface metrology parameters. On the other hand, the unified methodology often goes against common practice or existing standards (based on provided justifications). Examples of this include, the designations used for parameters and the adoption of the least squares approach for establishing the reference feature.

This unified methodology encompasses many of the practices which are already in place in surface metrology applications. However, without such a scheme, these practices are poorly defined and reproducibility between instruments and laboratories is often an issue. The unified specification scheme provides a mechanism or "language" for defining and controlling these practices so that they can be reproduced. Thus, the primary resistance to the scheme will come based on the task of changing specification format rather than the specifications themselves. From an economic standpoint this is the least costly change to make. If, on the other hand, the numerical values or tolerances associated with a surface feature would have to change, then the costs would be tremendous as extensive testing would have to be conducted to establish these new limits.

Ultimately, the general "acceptance" of this scheme by the user community would be most influenced through standardization and instrument implementation. Standardization alone is a major step towards broad application and the merits of the scheme are worthy thereof. However, from the perspective of the vast majority of surface metrology "users" the scheme will need to be made "attractive".

The manufacturers of instrumentation will be asked to provide more detailed information regarding their performance including frequency response characteristics. In addition, modifications may be required regarding numerical analysis and graphical presentation of results (see Chapter 7). This may appear to be an added responsibility for the instrument manufacturers and some may see this as a threat or "exposure". However, they must view this as an opportunity to extend their market share by entering competitors markets (upon demonstrating correlated results) rather than view this as a threat.

Finally, the user community must recognize that this scheme further enhances the information that can be obtained from surface measurement. This can manifest itself in improved manufacturing process control and product functionality (see Chapter 7).

2.7 Areas Requiring Further Development

This unification scheme presents a general framework and language which can be applied to a wide range of profile assessments. In itself, this is a significant benefit in terms of the specification, generation and control of surface attributes. However, in the application of this specification approach, several areas requiring further refinement become apparent. To introduce these areas, it is useful to describe a generalized surface metrology application. Surface metrology applications typically involve the components shown in Figure 2.13.

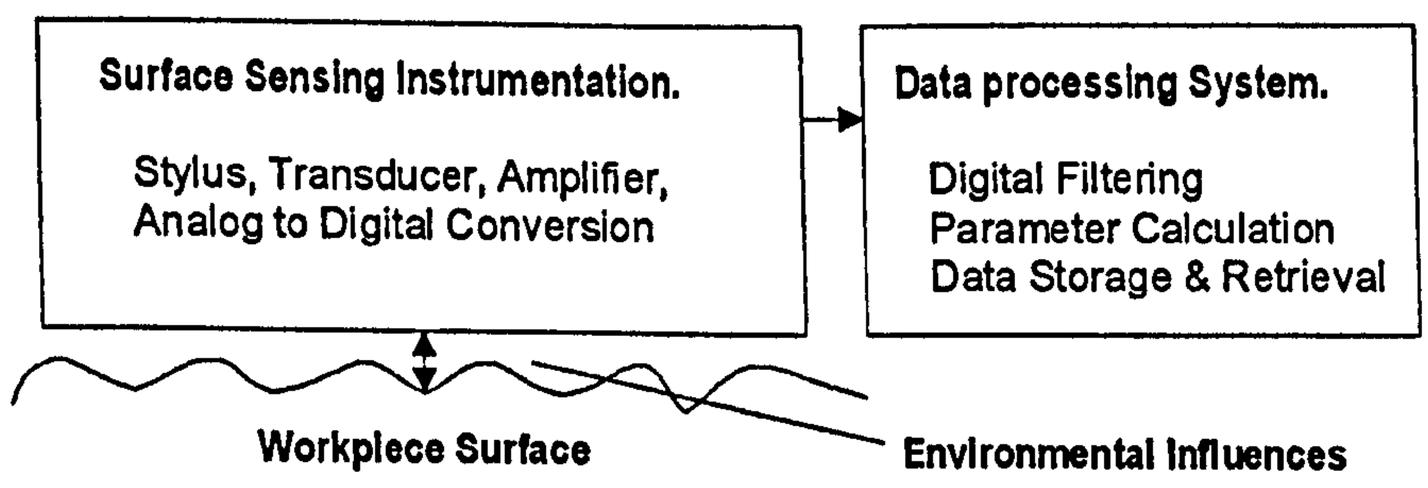


Figure 2.13 *The components of a typical surface metrology application.*

It is important that the components of the system be presented in a very general manner as one of the goals of the unification is to allow for variations in measurement methodologies as long as the desired bandwidths are achieved in the resulting data.

2.7.1 Stylus to Surface Interactions

In viewing the model shown in Figure 2.13, the first issue that arises is that of the interface between the instrument and surface. The vast majority of surface metrology applications utilize a contacting stylus. However, the actual transfer function associated with stylus convolution is for the most part unknown. There have been several theoretical models of stylus convolution, but not a great deal of experimental data has been put forth. Given the optional radius (or "stylus") based wavelength limitation discussed above in Section 2.4.3, it is important to gain a further understanding of the nature of stylus convolution. This also has implications for the other means of limiting wavelengths in that these applications require that the stylus tip influences must be outside the transmission band. This topic will be explored in Chapter 3 in terms of theoretical models and the analysis of various types of profile data.

2.7.2 The Presence of Unwanted Asperities

As the dependence on surface metrology in assuring product quality has grown, surface metrology applications have moved more and more toward production environments. In these applications, the presence of unwanted asperities in the resulting data sets has become increasingly problematic. Historically, these problems were predominant in the "traditional" precision manufacturing processes such as those common to fuel injection and in the bearing industry. However, with ever shrinking tolerances, the presence of unwanted asperities is common in nearly every surface metrology application. Re-cleaning the component may not always solve the problem as airborne particles may attach to the surface during the actual data collection process as the measurement environment can never be perfect. Therefore, a robust means of

detecting and removing unwanted asperities is a vital part of a scheme in which surface metrology can be effectively employed. Current methods for detecting and removing asperities will be explored and a robust alternative will be presented in Chapter 4.

2.7.3 Functional Wavelength Limitation

Historically, many approaches for the limitation of wavelengths have been applied in surface metrology. However, once a technique is developed and implemented, it becomes very difficult to completely move on to a newer approach without having to *carry along* the older approach for the sake of backwards compatibility. Despite the user community's tendency to hold on to older methodologies, the status of wavelength limitation approaches must be re-evaluated in light of the functional requirements.

Another important aspect of filtering to be addressed is the treatment of end effects - particularly in the processing of long wavelength surface features. In this context, several debates have arisen over the use of padding, filters, splines or other regressed features.

Chapter 5 will look at filtering in depth, including current and historical approaches as well as the very important topics of numerical implementations and end effects.

2.7.4 Parameterizations

The topic of surface texture parameters and the proliferation thereof has been very popular since the (often referred to) "parameter rash" paper by Whitehouse (1982). This topic will also be addressed in light of the unification scheme, however, the focus will not be so much on the parameters themselves, but rather an approach for establishing a "unified parameter set". Chapter 6 provides a means for treating the "rash" (in many cases through the elimination of parameters) and serves to control its spread in the future (through guidelines for the incorporation of new parameters).

2.8 Conclusions

In this chapter, the framework for a unified methodology for the application of surface metrology has been put forth. This methodology is centered about specification approach which provides improved control of the surface measurement process. The general adoption of this scheme will curtail the "divergence" presented in Chapter 1 while maintaining the flexibility necessary to accommodate future developments.

In the next few chapters, several developments will be provided as essential technical support of the unification scheme. These developments will provide further understanding into the more subtle aspects of measurement divergence as well as provide guidance toward reducing these divergences. Finally, Chapter 7 will "re-summarize" the unification scheme in light of the framework put forth in this chapter as well as the supporting material of the central chapters.

A Unified Methodology for the

Application of Surface Metrology:

Chapter 3

Stylus Tip Convolution

Despite the growing divergence in surface metrology applications, there seems to be one element that is still common across the vast majority of instrumentation - a stylus tip (typically spherical) which mechanically contacts the surface. This "contact-based" approach to detecting surface features is still by far the most popular means despite the numerous alternative sensing approaches including optical, capacitance, pneumatic and ultrasonic (see, for example: Green 1967, Thomas 1982, Whitehouse 1987, Whitehouse 1994).

In many practical regards, the underlying *convolution* of a stylus tip geometry over a surface is perhaps the most common or "unified" element between the vast majority of today's instruments. It may be argued, however, that the term "convolution" does not apply here when taken in the strictest mathematical sense, although it has been used in practice (DeVries and Li 1985, Li 1991). In the stylus context, "convolution" is used to represent the generation of a surface profile as determined by the geometric or physical relationships between the stylus and surface. This "stylus convolution" differs from a pure mathematical convolution in that the latter typically incorporates all data points and the former is primarily based on peak interactions.

Stylus or radius based convolution is a process that causes a limitation on the surface wavelengths which are transmitted to the instrument's transducer. However, this wavelength limitation, or more correctly - wavelength *modification*, cannot be generalized. Models have been developed to predict the transmission behavior of tip geometries, however, these only apply to very limited classes of surfaces. The fundamental problem with predicting the transmission characteristics associated with stylus tip convolution is the fact that the tip convolution process is based primarily on peaks and thus varies from surface to surface. Mathematical models of the effects of this convolution process are typically based on idealized surfaces (such as sinusoidal or Gaussian) and generally these models incorporate all surface features (including valleys) (Whitehouse 1974, McCool 1984, Al-Jumaily et al. 1987).

Although stylus tip convolution effects are not easily characterized or predicted mathematically, the stylus tip radius can be a *functional* and reproducible means of limiting wavelengths in surface metrology assessments. One common example is in the assessment of cam lobe geometry whereby, in application, a roller will follow the surface. In this application, it is very important to control the dynamics of the roller which ultimately generates the kinematics along the mechanism. To understand the functionally important cam surface attributes, it is essential that the surface be measured and analyzed from the roller's perspective - exploiting the features which the roller would "see" and ignoring those that it would not. The most common method for this type of assessment is to incorporate a stylus radius into the measuring system which approximates the nominal follower radius in application (Adcole 1993, Bhargava 1993). In so doing, the stylus acts as an artificial follower in the measurement process and thus generates a data set directly related to the motion that the follower will produce in application. This "radius based" transmission is what is desired in order to functionally assess the surface geometry. Historically, this approach has been referred to as the *envelope* or *e* system, whereby the envelope of the stylus convolution is deemed to be the surface of interest (Von Weingraber 1956, Von Weingraber 1957, Radhakrishnan 1971, Shunmugam and Radhakrishnan 1975).

Despite the fact that radius based wavelength limitation is often difficult to predict and can vary dramatically between profiles, it is a functionally important and metrologically reproducible means of assessing surface features and has therefore been included in the scheme for unification.

In addition to the obvious importance relating to "radius based" wavelength limitation, tip convolution effects must also be understood in the context of wavelength limitations per the other approaches in the unification scheme. Given that most of today's surface metrology instrumentation incorporates a nominally spherical stylus, the understanding of these transmission effects is necessary in selecting an appropriate tip radius when using Gaussian or *ideal* filters. For example, when assessing a surface

with a Gaussian band-pass, it is necessary to utilize a stylus tip that does not significantly affect the surface wavelengths in the desired range.

In this chapter, the topic of tip convolution will be explored in terms of theory and application in the context of the scheme for unification. This will begin with a brief overview of physical versus mathematical convolution processes. Next, the actual wavelength limitations resulting from tip convolutions will be explored. This topic has been addressed in theory (Al-Jumaily et. al 1987, Li 1991, Scott 1992a, ISO 3274-1995) on numerically generated profiles. However, a comprehensive analysis using real, engineering surfaces has not been previously conducted. The study contained in this chapter is important in understanding instrument correlation issues as well as predicting the functionality of surfaces in rolling contact. In addition, the special case of conical or pyramidal styli will be explored in the context of detecting and quantifying re-entrant surface features based on the flank angles of the stylus. Finally, the relationship between digital sampling and stylus tip radius is presented as an integral aspect of the unified methodology.

3.1 Stylus Tip Convolution

Two approaches are available for establishing a "radius based profile transmission" or *e* system profile - either physical or mathematical. In the physical arrangement, the instrument is fitted with a stylus of the desired radius and data is collected directly based on sensing the position of the stylus as it is tracked across the surface (Adcole 1993, Bhargava 1993). The mathematical approach involves collecting a data set with a relatively small tip radius and subsequently performing a numerical convolution or simulation of a larger tip radius over the data set (Shunmugam and Radhakrishnan 1974, Radhakrishnan and Shunmugam 1974, McCool 1984, Scott 1992a). This mathematical approach is considerably more desirable in terms of factors such as the cost of purchasing and maintaining styli of varying sizes and the flexibility of being able to alter the tip radius and not having to re-measure the component. Although similar

results might be expected between the methods for most surfaces, they may not be identical. These divergences (although they are typically small) can be the result of many factors. Two of the primary factors will be discussed in the following sections.

3.1.1 "Noise" in the Measurement

Disturbances (whether mechanical or electronic) which are present in the measurement process, translate directly into the measured profile. In any physical measurement, these effects will be present in the profile to some extent. When the mathematical convolution is applied to the measured data set, many of these effects are *smoothed* out. If a physical measurement were made with the same tip radius that was mathematically applied, we may find high frequency attributes which are the result of the measuring process. The influence of this measurement "noise" is graphically depicted in Figure 3.1.

Figure 3.1a demonstrates the ideal situation in that there is no "noise". In this case, we would find that (apart from other influences) the mathematical convolution duplicates the physical convolution. However, in application, high frequency noise may be present. If the surface were measured (physically) using the desired (large) tip radius, we may see a data set as is shown in Figure 3.1b, whereby the stylus path incorporates relatively high frequency changes as it moves over the profile peaks. In the mathematical convolution, the surface is first measured with a relatively small stylus (once again incorporating the noise) and then a circle is mathematically convolved over the profile. Figure 3.1c demonstrates how the convolution process tends to smooth out the high frequencies which were present in the underlying profile measurement. It should be noted that the radius convolved profile of Figure 3.1c is very similar to that of 3.1a, however the 3.1c profile is based on the peaks of the noise rather than purely from the base surface.

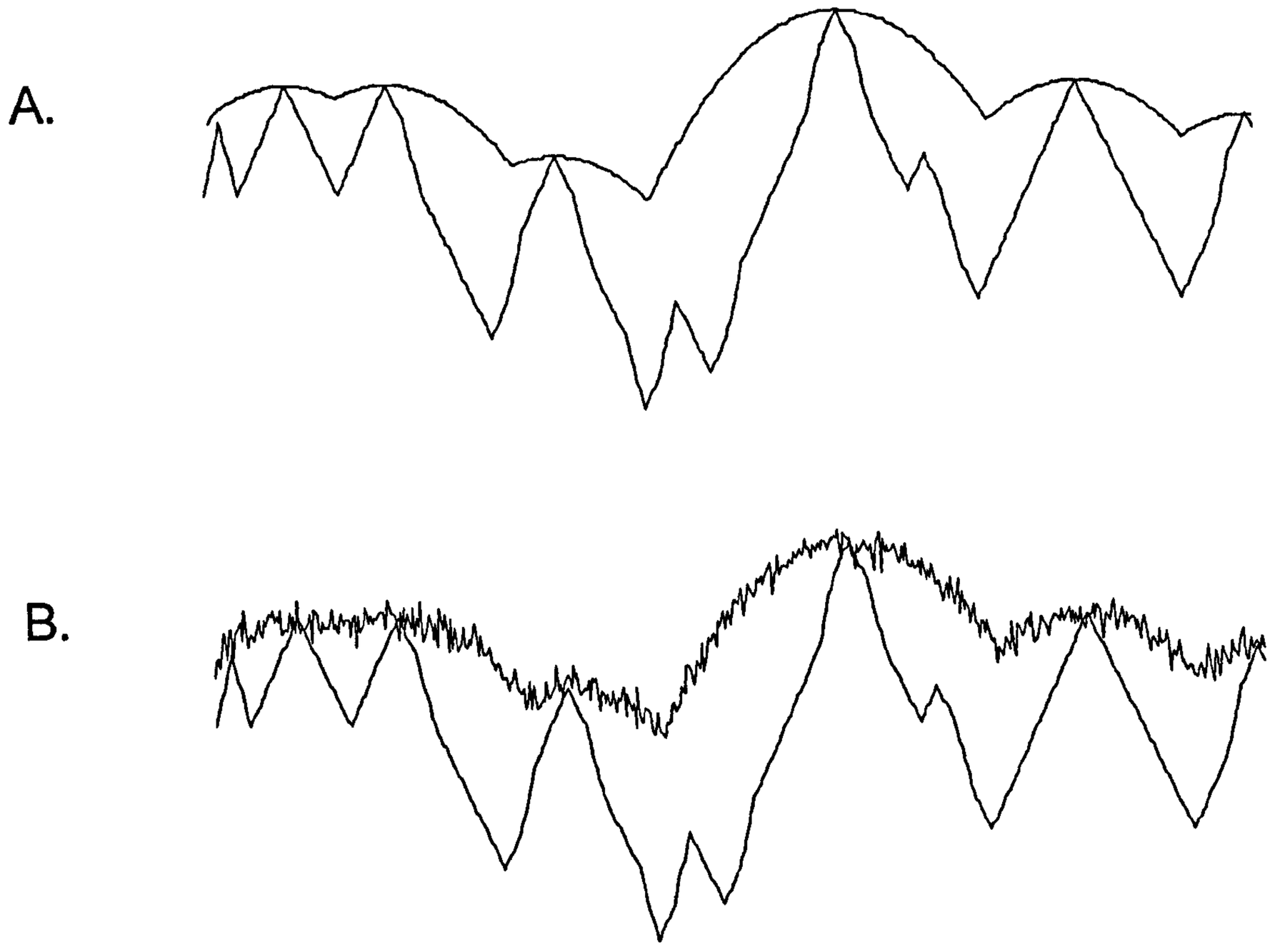


Figure 3.1 *The influence of vibration in physical versus mathematical tip convolution.*

In a similar sense, irregularities in the stylus tip geometry (for example out of roundness) can introduce errors in the physically convolved data set. In a physical measurement based on a specified tip radius, these errors are typically viewed as insignificant. Furthermore, these geometry errors are not incorporated in numerical (mathematical) radius convolution.

3.1.2 Spherical versus Circular Convolution

Another factor that can lead to a difference in assessments made via mathematical convolution of a circle versus physical convolution of a stylus is the presence of peaks to the *sides* of the profile trace. When measuring a surface with a stylus tip of finite dimension, the potential for stylus-to-surface contact occurs over an area. Larger stylus tips have a greater potential for interacting with peaks that are to the sides of the trace. This is shown in Figure 3.2 where the high peak in the center of trace #3 could influence measurements along trace #2 depending on the tip radius being used.

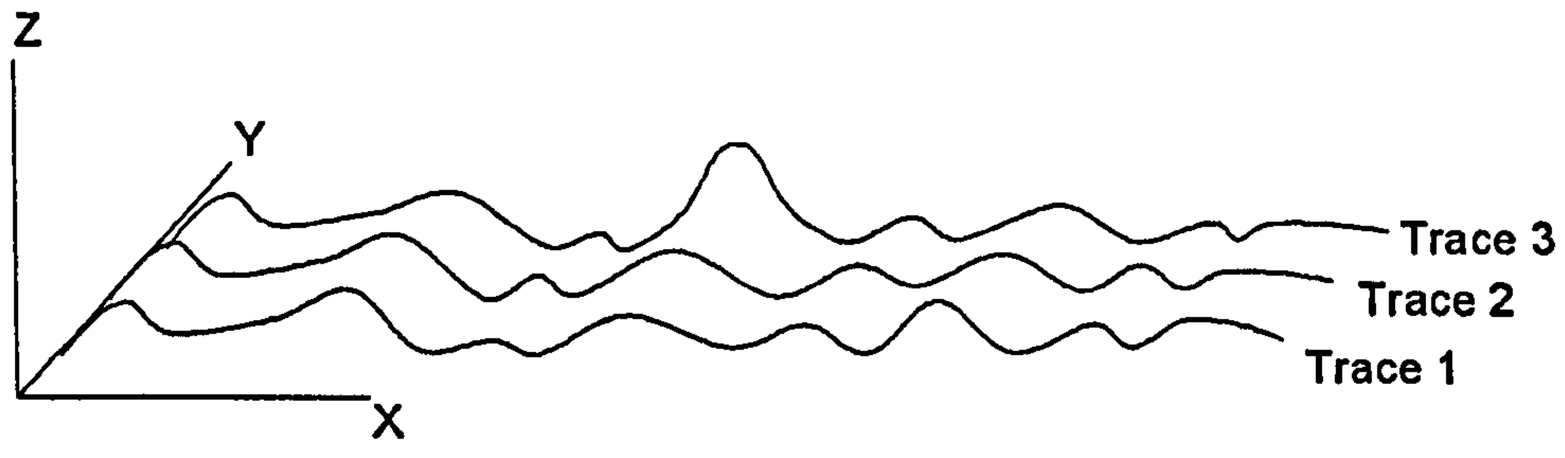


Figure 3.2 *The presence of a "lateral" peak in trace 3 relative to trace 2.*

The true convolution of a stylus tip in the presence of *lateral* peaks is best predicted through the three dimensional or areal measurement of the surface and the subsequent convolution of a sphere (Shunmugam and Radhakrishnan 1974 & 1975). However, in practice this is very time consuming and more mathematically intensive than the simple measurement and analysis of single profiles.

3.1.3 Other Considerations

Other considerations related to stylus tip convolution include deformation (Chetwynd et al. 1992) and tracking (McCool 1984). These are worth mentioning here in that these are important metrological considerations, however they are very dependent upon the instrument-stylus-surface relationship and must be handled on a case by case basis apart from the overall unification scheme.

Furthermore, it has been proposed that the stylus tip geometry can, for some surface features, be de-convolved to yield a more *true* representation of the surface (DeVries and Li 1985, Li 1991, Wang 1995). This may apply in the analysis of very small surface features where appropriate styli cannot be easily obtained or in applications requiring exact dimensional characterizations of the width as well as the height of profile features. However, the incorporation of the stylus based wavelength limitation in the scheme for unification is to allow for the functional assessment of surfaces using the stylus radius to simulate the contacting surface in application. Thus, any de-convolution would degrade the functionality of the analysis and is therefore forbidden in the "radius based wavelength limitation" for the unified methodology.

3.2 Stylus Based Wavelength Limitation

Given that radius convolution is incorporated in the unification scheme as an alternative for wavelength limitation, it is important to understand how the wavelength (or frequency) spectrum of a surface is affected by the convolution of a stylus of given radius. Recent efforts by the ISO surface metrology technical committee (TC 57) have attempted to quantify the nominal wavelength transmission of roughness styli in order to move the stylus influences outside the default wavelength band for roughness measurement. This work was led by Dr. Paul Scott of Rank Taylor Hobson and resulted in the roughness bandwidth representation shown in Figure 3.3 (Scott 1992a, ISO 3274-1995, ISO 11562-1995).

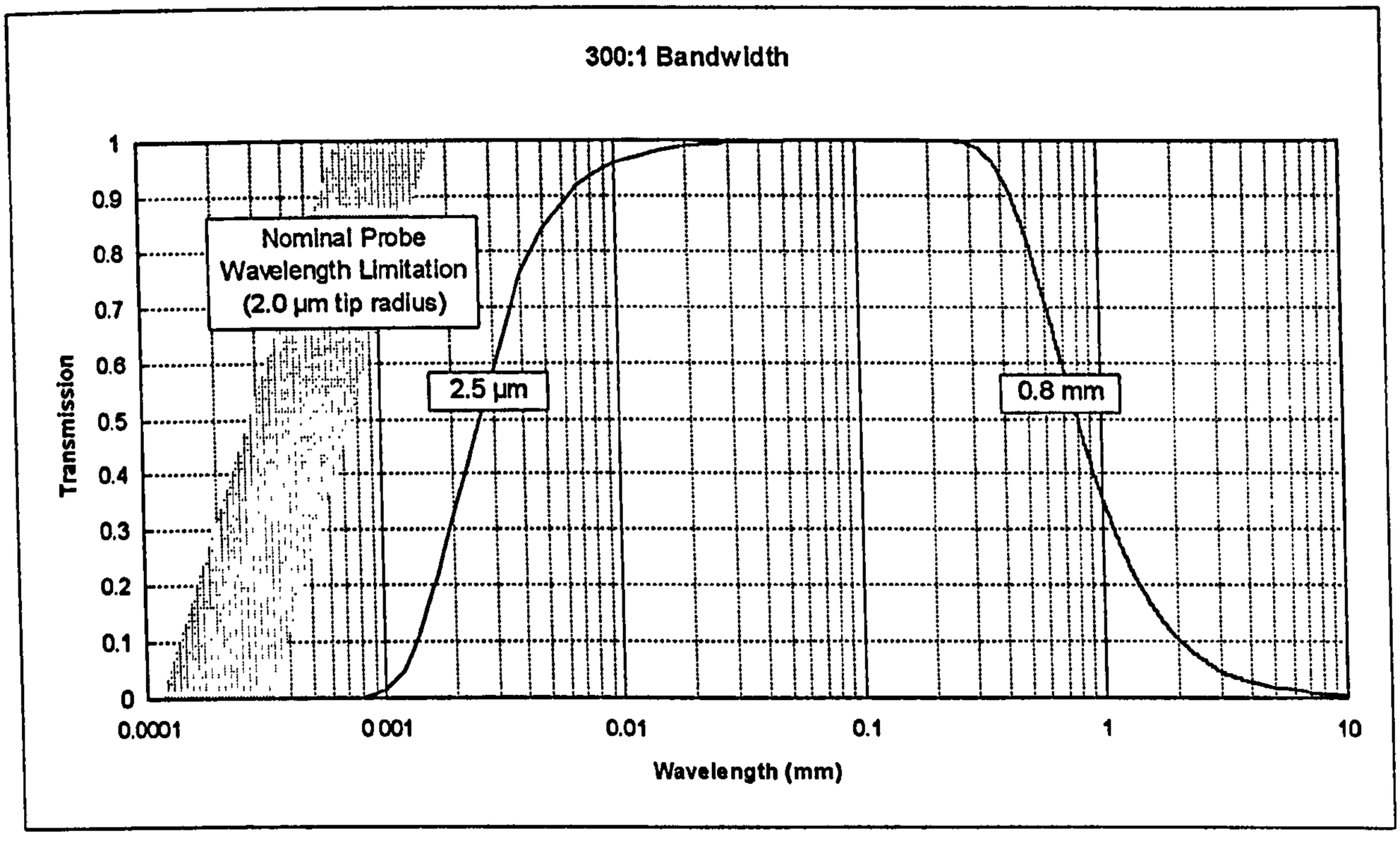


Figure 3.3 Stylus transmission and the default (Gaussian) roughness bandwidth according to ISO 11562-1995.

The Figure 3.3 representation depicts some uncertainty in the wavelength regime which is affected by the stylus tip geometry. This seems reasonable given the above discussion of the unpredictability of the wavelength transmission resulting from stylus tip convolution. It should be noted, however, that this uncertain area is the result of the mathematical convolution of circular geometries over numerically generated, sinusoidal data sets (Scott 1995). As a result, the wavelength transmission is very idealized.

By comparing the wavelength content of profiles before and after a tip convolution, the stylus transmission can be determined. This can be accommodated through the use of Fourier transforms of the data sets and then dividing the resulting amplitude by the original amplitude at each wavelength. Wavelengths which are transmitted completely will have a transmission of 1.0. Transmissions less than 1.0 indicate an attenuation and those greater than 1.0 indicate an amplification.

Li (1991) used this approach in a study of tip convolution effects on theoretical surfaces, however the transmission functions presented did not seem to relate to the power spectrum data which was used as the basis. In Li's presentation, power spectrum functions were presented for the original (numerically generated) profile and the associated tip-convolved profile. The amplitudes at all of the presented frequencies were lower for the tip convolved data than for the original signal, thus indicating attenuation at all frequencies. However, the discrepancy arises in the presentation of transfer functions with many of the frequencies indicating transmissions greater than 1.0.

3.3 Wavelength Transmission Effects

Given the general lack of published information regarding stylus tip transmission effects on real surfaces and the importance of radius based transmission relative to the scheme for unification, an extensive study was undertaken. The details and results of the study are contained within the following sections of this chapter and also in Appendix B.

To gain an understanding of tip convolution effects in terms of wavelength transmission, data was collected from three primary types of surfaces - random, patterned and stratified. These surfaces were obtained from grinding, turning and plateau honing (respectively) and were nominally between 0.8 and 1.0 μm RMS amplitude when physically measured with a Form Talysurf Series S5C (2 μm stylus tip, 8.0 μm to 2.5 mm Gaussian band-pass filtering, 0.01 μm vertical resolution, 0.25 μm ordinate spacing). It should be noted that, per the unified scheme, the radius based limitation should be applied prior to any digital filtering. However, since this study is concerned with *changes* in the profile, the band limitation introduced by the instrument does not significantly affect the results.

To generate a prediction of transmission functions, six profiles were obtained from each type of surface and tip radii of

5, 7.5, 10, 25, 50, 75, 100, 250, 500, 750, 1000, 2500, 7500 μm

were subsequently mathematically convolved over each profile. Thus, 6 data sets were collected from samples of 3 types of surfaces and were analyzed using the physical (2 μm) tip and 13 mathematical tip convolutions - resulting in the analysis of 252 profiles.

The numerical tip convolution procedure adopted in this study was based on the work of Scott (1992b), whereby a *motif combination* approach is used. In this approach to determining the "tip-convolved profile", groupings of three data points are recursively assessed in light of the potential for stylus contact. In this approach, the central point of the three point grouping is discarded when contact is dictated by the outer points. Those points which contact the radius are stored. The grouping is subsequently expanded or shifted until all stored points contact the radius. These contact points are then used as pivot points for the enveloping profile. (Refer to Appendix B for program listing.)

Given the data sets generated by the above tip convolution process, Fourier transforms were performed on each per Reid (1992). The spectra associated with each mathematical convolution (radius values greater than 2.0 μm) were compared to that of the original signal (as measured with the 2.0 μm tip radius).

3.3.1 Random (Ground) Profiles

The analysis of tip convolution effects on random surfaces was based on six profiles obtained from ground bores with a nominally 0.8 μm RMS amplitude (per the 2 μm stylus tip and Gaussian band-pass described above). A typical trace is shown in Figure 3.4. Further detail in the profile is shown in Figure 3.5 where a 1.0 mm section of the data set is expanded.

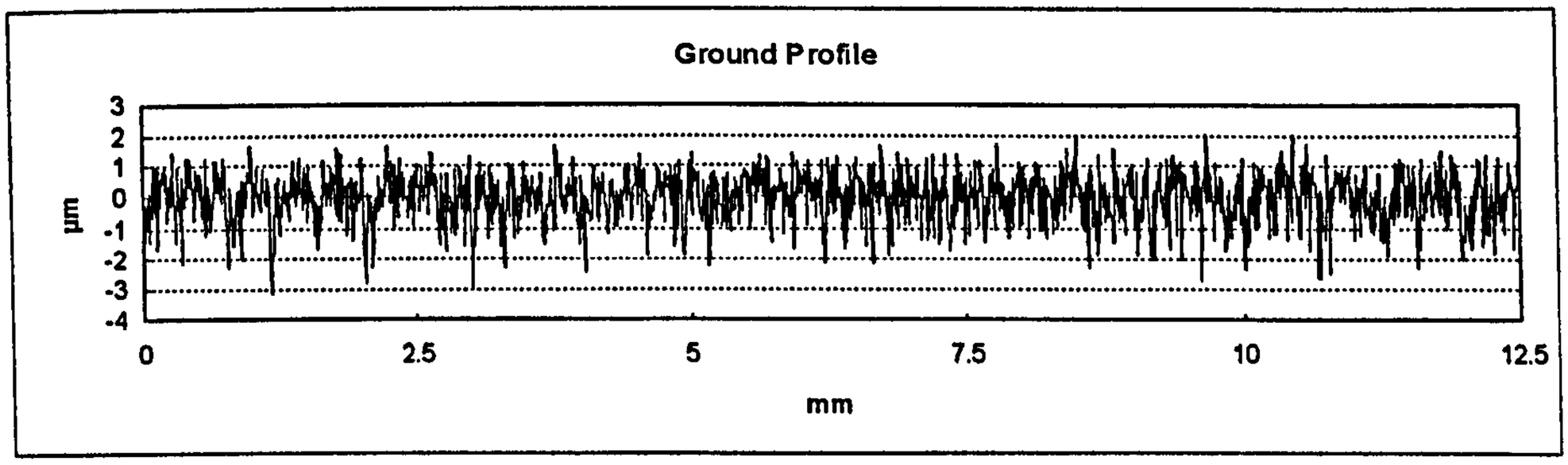


Figure 3.4 *Typical random (ground) profile.*

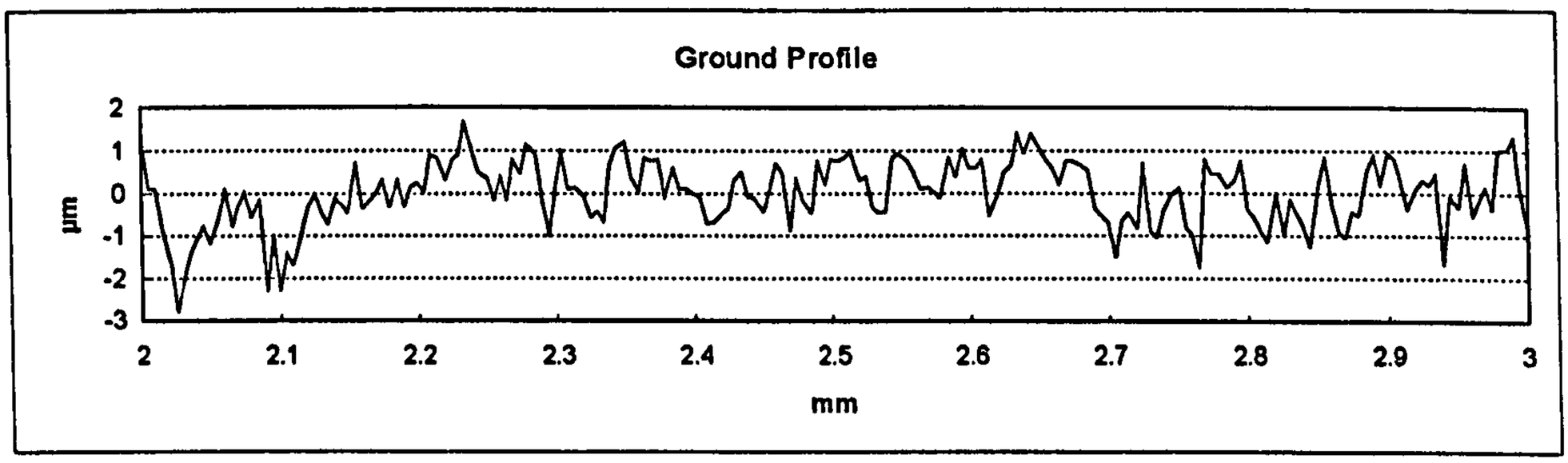


Figure 3.5 *1.0 mm section of a random (ground) profile.*

The Fourier transform of the Figure 3.5 data set (as measured) is given in Figure 3.6. (Note: For all data sets the mean value (zero'th order) is set to zero and is plotted coincidentally with the first order.)

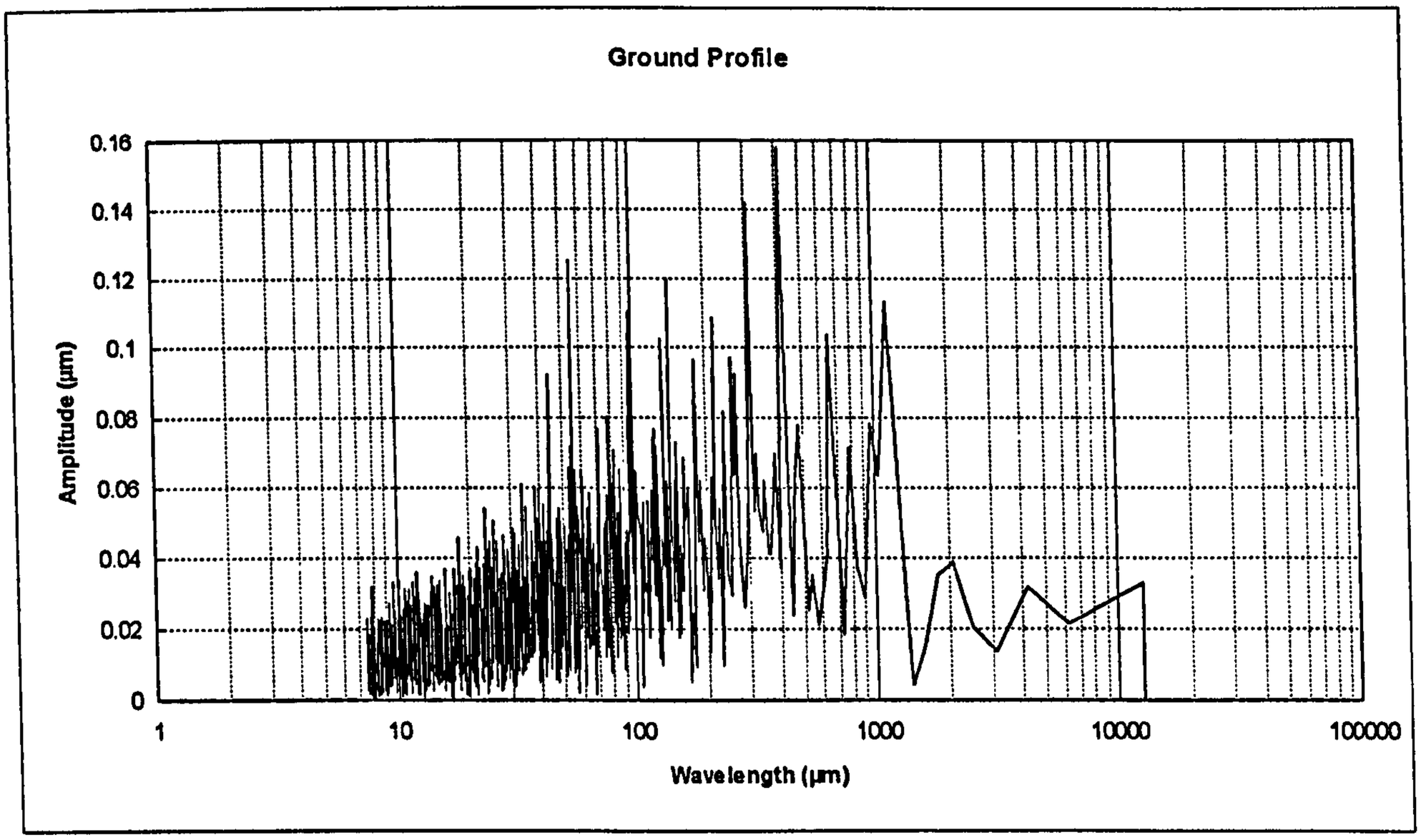


Figure 3.6 Wavelength content of a random (ground) profile.

Transmission functions were generated for each of the profiles and each radius. For a given tip radius, the six transmission functions were averaged to predict the nominal wavelength transmission behavior. This nominal transmission data is presented in Appendix B for the ground profiles. To compare the short wavelength attenuation (or long wavelength transmission) behavior between different tip radii, curves were manually generated through the general trends of the nominal wavelength transmission data. These results are shown in Figure 3.7. (Note: Given the overlap between some of the transmission functions, not all are plotted.)

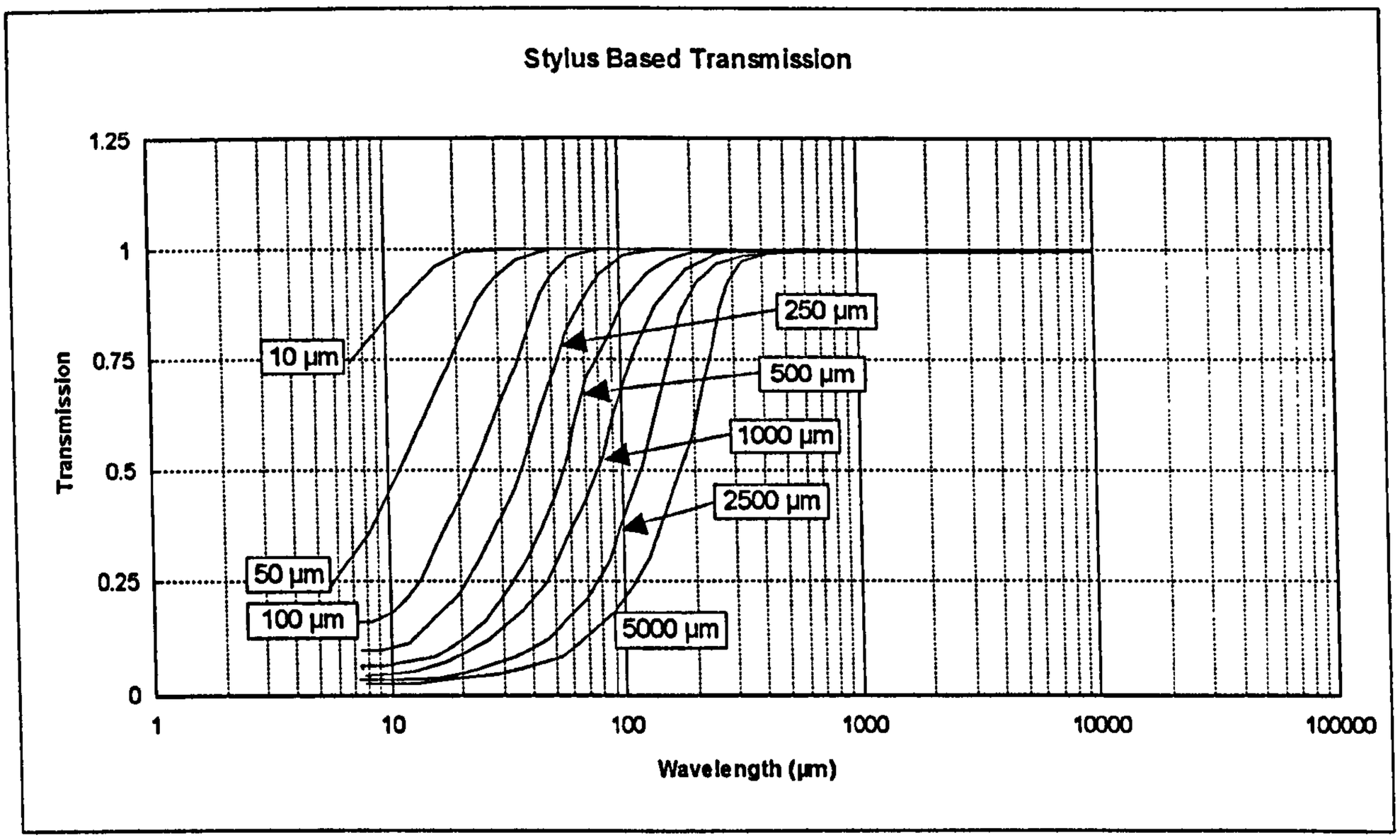


Figure 3.7 *Nominal stylus based wavelength transmission for random (ground) profiles.*

While this method for generating "trend" lines does not derive directly from mathematics or signal processing, it is useful in determining an "engineering" prediction for the nominal short wavelength attenuation characteristics. In the analysis of measured surface profiles, slight variations in profile features can generate very different amplitudes at some frequencies and thus the resulting spectra appears very "noisy" (see Figure 3.6 and Appendix B). Furthermore, in generating transmission functions, many amplitudes are near zero, and thus lead to floating point instabilities during numerical computations. Therefore, to gain a general understanding of the short wavelength transmission limits of the stylus tip transmission characteristics, the trend line is generated. It should be emphasized, however, that in critical applications this type of smoothing may not be appropriate.

While the Figure 3.7 clearly presents the differences in lower transmission limits due to the radius convolution process, it should also be noted that affects may be present in longer wavelengths (see Appendix B). These longer wavelength attributes are difficult

to represent in a "composite" form such as in Figure 3.7, nonetheless they are present in many of the analyses for the ground profiles. This indicates that the radius convolution process is also attenuating some long wavelength characteristics while amplifying others. These long wavelength influences are much less predictable and are typically the result of the radius "bridging" across profile peaks thereby introducing relatively long, smooth areas in the profile (see, for example, Figure 3.1c).

This analysis of ground profiles indicates that the lower limits of the radius based transmission function do not seem to linearly follow the value of the tip radius. This can be shown by observing 50% transmission crossings (of the Figure 3.7 plots) as shown in Table 3.1 and Figure 3.8.

Tip Radius (μm)	50% Transmission Wavelength (μm)
50	11
500	55
5000	170

Table 3.1 Non-linearity in transmission function relative to tip radius.

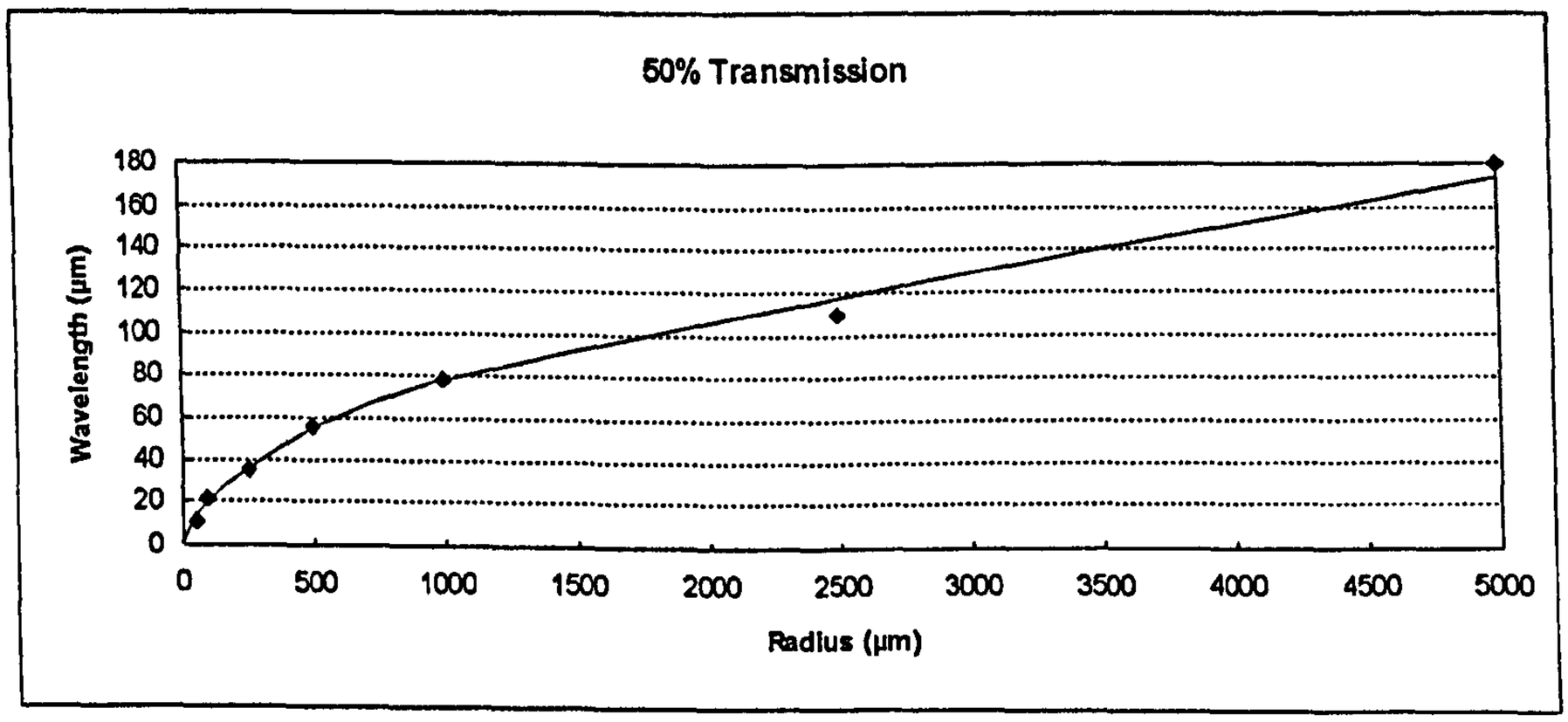


Figure 3.8 Radius based transmission 50% crossings for ground profiles.

ISO TC 57 has incorporated a philosophy whereby band-pass filtering is used to define a consistent roughness domain given nominal stylus tip radii (ISO 11562-1995, ISO 3274-1994, ISO 4288-1995). This standardized approach applies a *long-pass* Gaussian filter at a wavelength nominally 1.2 times the tip radius for the purpose of reducing the uncertainty in wavelength transmission due to tip geometry (see also Figure 3.3 above). The above data for the ground profiles indicates that, while the 1.2 factor seems to be appropriate in terms of the lower wavelength transmission limits of small tip radii, it is very conservative in terms of the lower transmission limits of large tip radii for the analysis of random surfaces on this order. Furthermore, the ISO guideline does not account for any influences which may occur in longer wavelengths which may fall inside the desired pass band (as shown graphically in Appendix B).

These changes in the transmitted profile as a function of the tip convolution process also manifest themselves in calculated parameters. Table 3.2 reports the average values obtained for several parameters. (The data is graphically presented in Appendix B for the ground profiles.)

Tip Radius (µm)	Parameter						
	a (µm)	q (µm)	sk	z (µm)	pm (µm)	vm (µm)	dq (°)
2	*0.627*	*0.800*	*-0.627*	*5.390*	*2.125*	*3.265*	*14.164*
5	*0.587*	*0.752*	*-0.648*	*5.126*	*1.983*	*3.143*	*10.867*
10	*0.560*	*0.719*	*-0.637*	*4.912*	*1.881*	*3.031*	*8.943*
25	*0.516*	*0.665*	*-0.612*	*4.468*	*1.726*	*2.742*	*6.313*
50	*0.480*	*0.619*	*-0.565*	*4.033*	*1.600*	*2.433*	*4.673*
75	*0.459*	*0.591*	*-0.522*	*3.828*	*1.524*	*2.304*	*3.881*
100	*0.443*	*0.571*	*-0.477*	*3.661*	*1.470*	*2.191*	*3.391*
250	*0.393*	*0.503*	*-0.244*	*3.070*	*1.297*	*1.773*	*2.155*
500	*0.358*	*0.457*	*-0.043*	*2.547*	*1.172*	*1.376*	*1.494*
750	*0.341*	*0.436*	*0.042*	*2.323*	*1.101*	*1.222*	*1.205*
1000	*0.329*	*0.421*	*0.094*	*2.206*	*1.052*	*1.154*	*1.037*
2500	*0.298*	*0.379*	*0.283*	*1.853*	*0.903*	*0.950*	*0.637*
5000	*0.278*	*0.354*	*0.409*	*1.605*	*0.797*	*0.808*	*0.438*
7500	*0.269*	*0.342*	*0.454*	*1.475*	*0.738*	*0.737*	*0.351*

Table 3.2 *Average parametric data for ground profiles as a function of tip radius.*

The parameters used in this study are briefly described below. (For more detail on parameter definitions, algorithms and applications refer to Chapter 6.)

a *Average amplitude*

q *RMS amplitude*

sk *amplitude distribution skewness*

z *average peak to valley (one peak to valley per cutoff length)*

pm *mean peak height (one peak per cutoff length (2.5 mm))*

vm *mean valley depth (one valley per cutoff length (2.5 mm))*

dq *RMS slope (7 point Lagrangian formulation)*

(Note: the *pm* and *vm* were based on 2.5 mm profile sections although for the sake of the study the long wavelength limitation was applied prior to the mathematical tip convolution.)

In general, the computed parameters tended to transition from *rough* to *smooth* as a function of increasing tip radius. (This can be indicated more clearly in viewing the graphs contained in Appendix B.) It is interesting to note, that the skewness, sk, parameter remained negative and rather constant over a range of tip radii up to approximately 100 μm and then trended into positive values for stylus tip radii greater than 100 μm.

This transition in skewness values represents a change in *shape* in the profile resulting from the convolution. For stylus tip radii less than 100 μm, the profile is made up of rounded peaks and sharp valleys such as in the theoretical profile given in Figure 3.1A. For tip radii greater than 100 μm, the profile is generally much smoother (lacking distinct valleys), however, the retention of local, isolated peaks causes the skewness to become positive.

Another interesting aspect of the parametric data (particularly when displayed graphically as in Appendix B) is that the average amplitude (*a*), the RMS amplitude (*q*), and skewness (*sk*), parameters exhibit relatively small changes between the 2.0 μm tip and 5.0 μm tip. This could indicate that the 2.0 μm tip is adequate for assessing these parameters in that little further benefit would derive from utilizing a smaller tip. However, the RMS slope (*dq*) parameter indicates a great deal of change between the 2.0 μm tip and the 5.0 μm tip and therefore there may be aspects of the surface in terms of local slopes that are being affected by the 2.0 μm tip.

3.3.2 Patterned (Turned) Profiles

Another common type of engineering surface is one which results from a machining process such as turning, boring or milling. These surfaces often exhibit repetitive patterns of features at distinct wavelengths. In this study, profiles were obtained from diesel engine piston pin bores with a nominal RMS roughness of 0.8 μm (per the 2 μm stylus tip and Gaussian band-pass described above). A typical profile from these measurements is presented in Figure 3.9 and a typical 1.0 mm segment of the profile is shown in Figure 3.10.

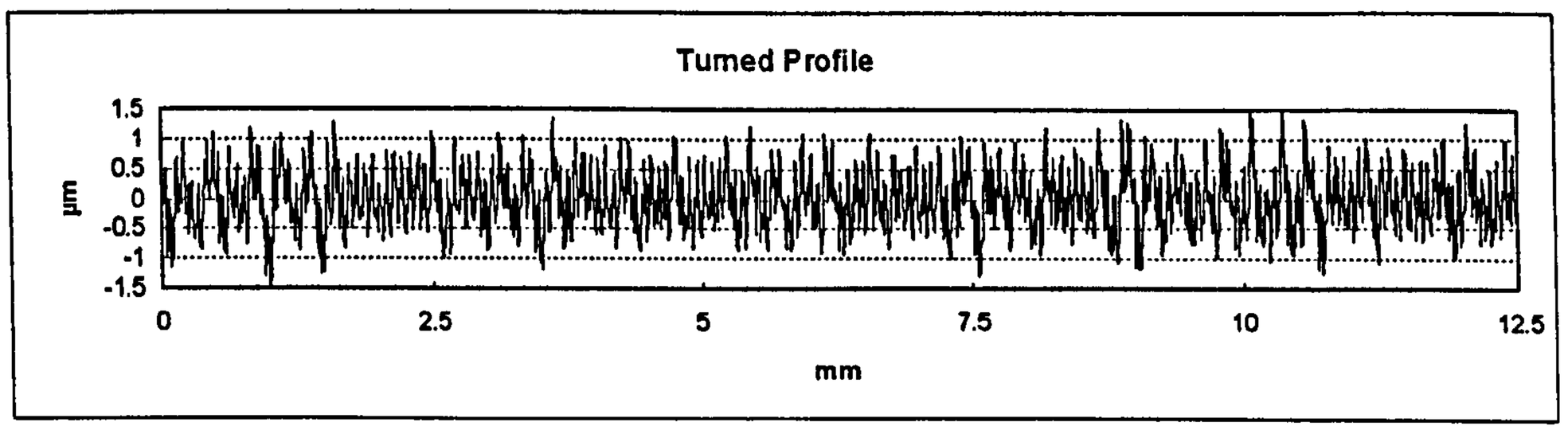


Figure 3.9 Typical patterned (turned) profile.

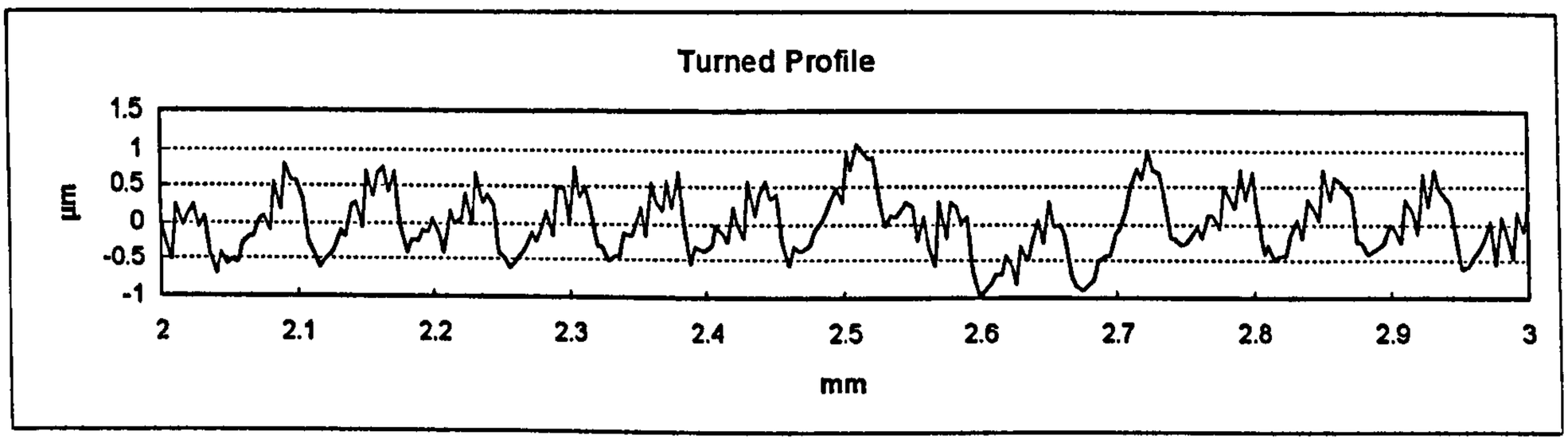


Figure 3.10 1.0 mm section of patterned (turned) profile.

The expanded view of the data (Figure 3.10) indicates a feed rate of approximately 70 μm and also the presence of high frequencies in the profile possibly resulting from a degradation of the cutting tool.

Fourier analysis of this profile (as shown in Figure 3.11) yields predominant peaks at the 70 μm wavelength (corresponding to the feed rate) and 35 μm due to the rectified-sinusoidal (or *cusped*) shape of the profile. If the turned profile with a 70 μm feed rate was modeled as an inverted, rectified sinusoid, peaks would be present at wavelengths of 70, 35, 17.5, and 8.75 μm. Although these (predicted) peaks are present in the data, several other significant peaks are also present.

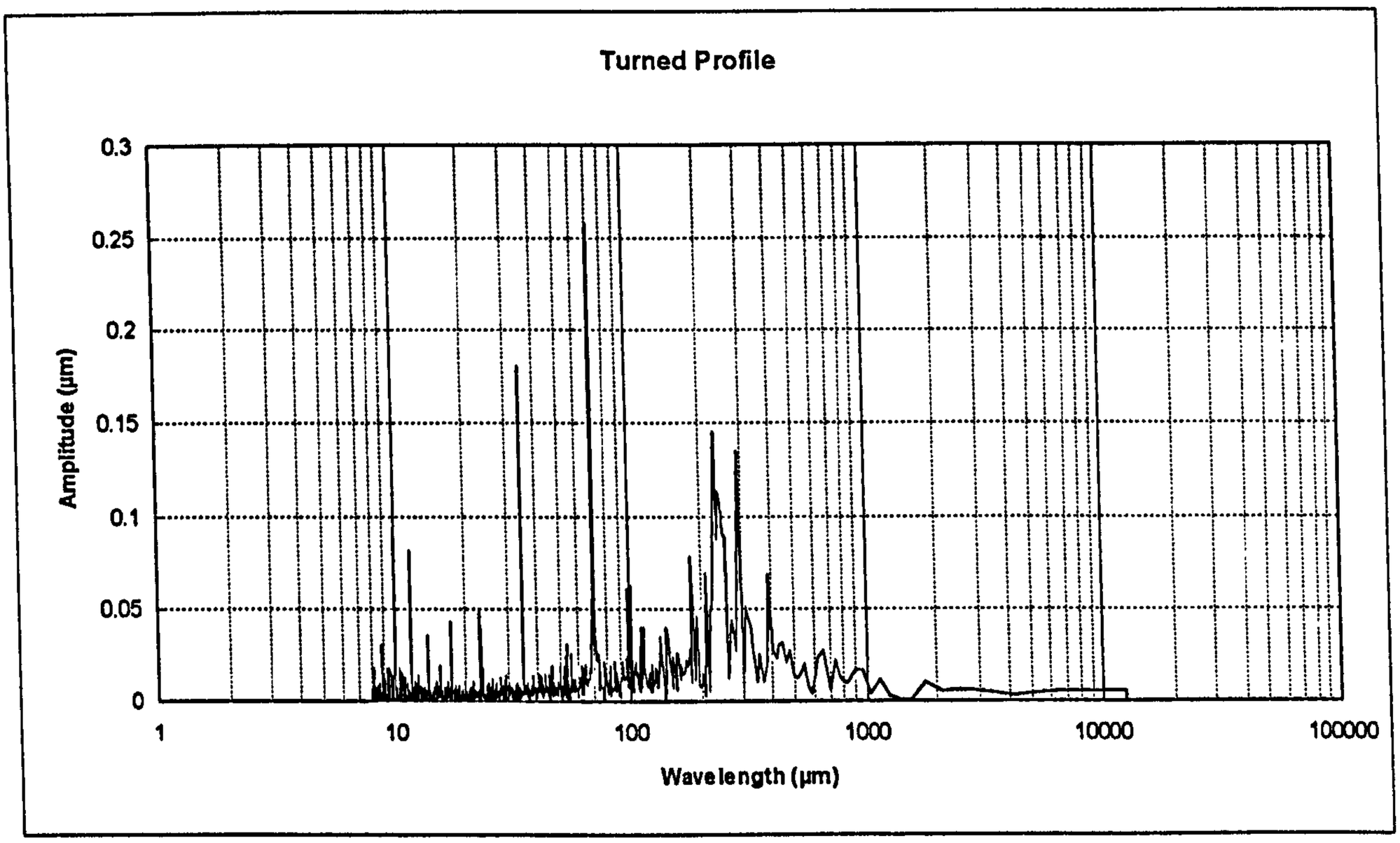


Figure 3.11 *Wavelength content of a patterned (turned) profile.*

Once again, each of the six profiles was numerically analyzed with varying tip radii and the resulting transmission functions were averaged to determine *nominal* characteristics. (These transmission characteristics can be found in Appendix B.) A curve was manually generated through each of these averaged transmission functions in order to predict nominal behavior of the short wavelength limitation. The nominal stylus based short wavelength transmission functions are shown for various tip radii in Figure 3.12.

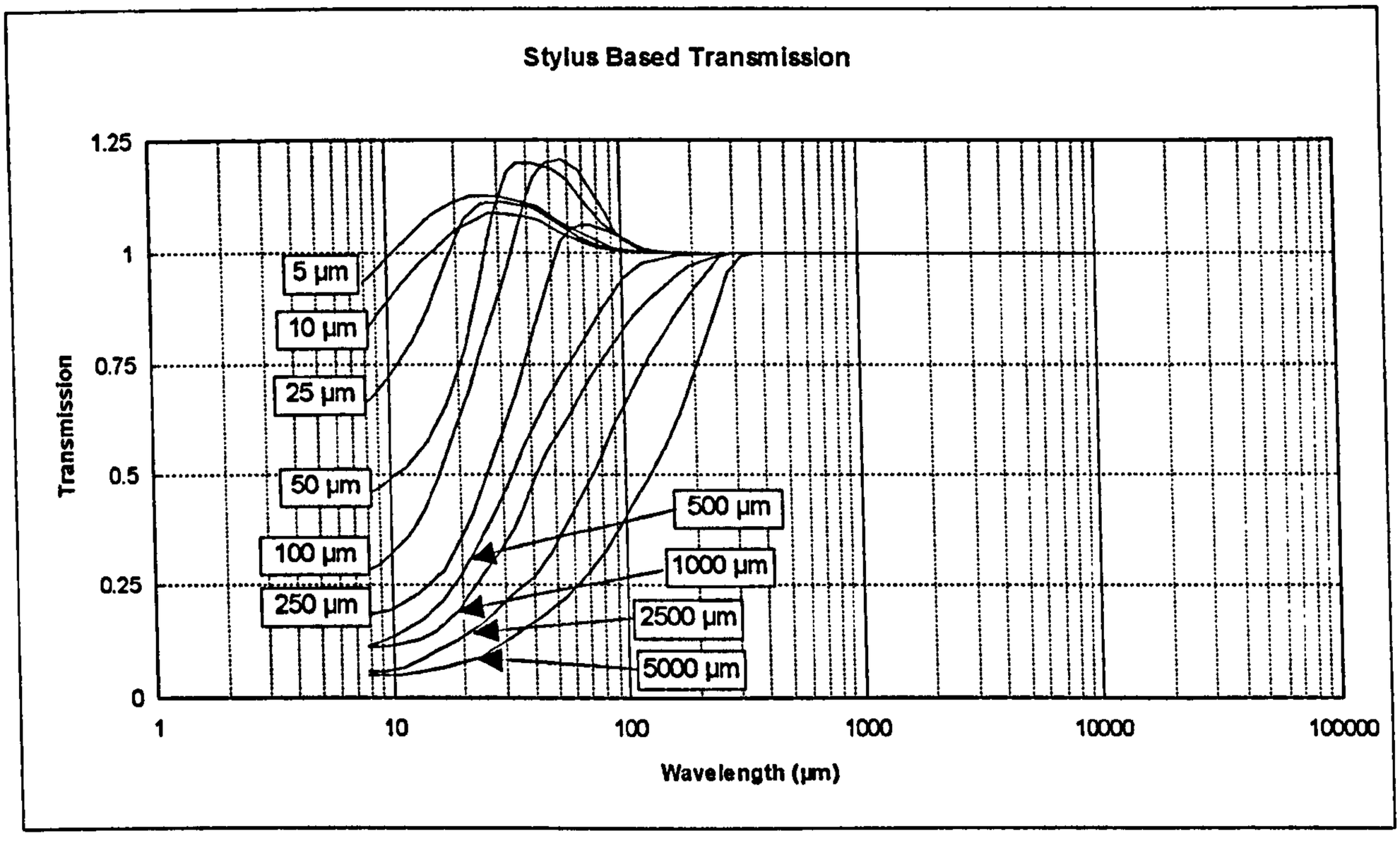


Figure 3.12 Nominal stylus based wavelength transmission for patterned (turned) profiles.

As discussed above in the context of ground profiles, long wavelength attributes are also affected in the analysis of turned profiles (particularly when incorporating larger tip radii). This is shown graphically in Appendix B, where transmission values were, in many cases, much greater than 1.0 for long wavelengths of turned profiles. One aspect of the turned profiles which may contribute to this very high transmission characteristic is the fact that the original data did not incorporate significant amplitudes in long wavelengths (see Figure 3.11 above). Thus, if any long wavelength attributes were introduced by the radius convolution process, they would result in a significant *change* and therefore a very high transmission. This *change* is magnified by the fact that the determination of a transmission is multiplicative, where by the generation of long wavelengths (in this context) is more likely to be additive.

Another very interesting aspect of the tip convolution over the turned profiles is the amplification of certain shorter wavelengths by the tip radius. This amplification occurs at a wavelength approximately one half of the feed rate (35 µm). Although

these measured, turned surfaces are not purely sinusoidal, this appears to be consistent with the work of Al-Jumaily et al. (1987) whereby sinusoidal profiles exhibited the appearance of a half-wavelength feature when convolved with a stylus tip.

In the case of turned profiles, this half-wavelength feature is already present to some extent in the Fourier transform due to the cusp shape in the profile which is asymmetric about the mean line. Thus the amplification will not be as high for the real surfaces as in the case of numerically generated sinusoids. The convolution of certain stylus tip radii over these measured data set does, however, tend to *invert* the basic shape of these turned profiles. This inversion can maintain the primary *shape* of the profile, however, many of the finer surfaces features are removed due to the tip geometry. An example of this *inversion* is graphically represented in Figure 3.13, wherein a stylus tip is numerically convolved over a simulated turned surface (inverse rectified sinusoid).

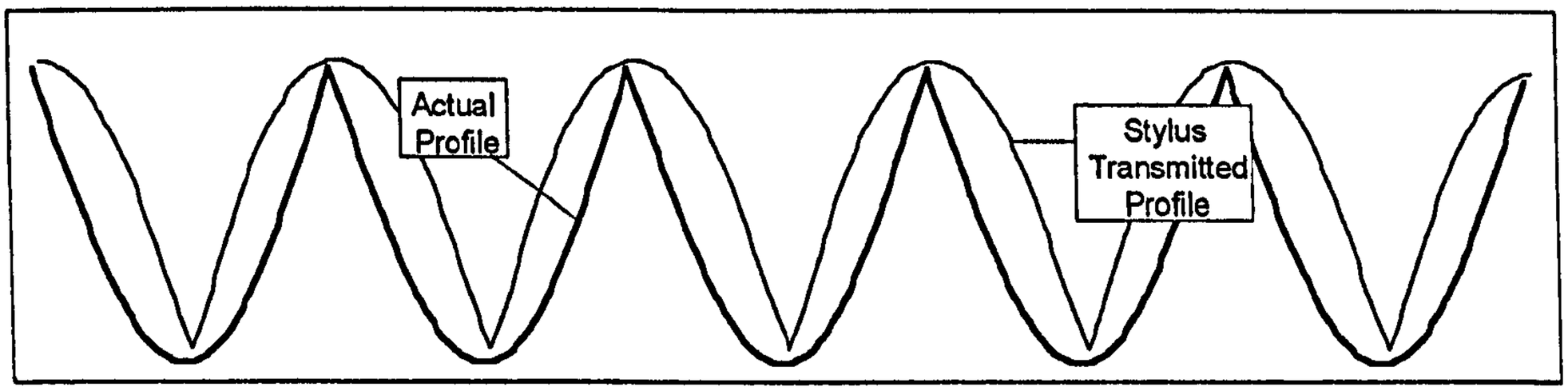


Figure 3.13 *Theoretical turned profile and tip convolution.*

In the Figure 3.13 example, an inverted, rectified sinusoid is shown to exhibit an inversion due the convolution of a stylus tip. The predominant wavelengths remain intact, only the phases and amplitudes change with varying tip radii.

In the analysis of the actual data sets from turned profiles, it becomes apparent that the data is generally more sinusoidal than the theoretical data of Figure 3.13. This lack of *cusping* in the actual data results in relatively small, half-wavelength amplitude which increases upon convolution with styli of certain radii. The experimental data indicates that the amplitude of the half feed-rate wavelength is amplified for styli less than 250

µm and is attenuated when tip radii greater than 250 µm are convolved. It should be noted that this numerical value is tied to the data used in the study (in terms of amplitudes and feedrate) and does not apply generally to all turned surfaces.

The calculated parameters (Table 3.3) also indicate a transition in results around the 250 µm tip radius. These transitions in parameter values are more readily apparent in the graphical representations included in Appendix B. For the turned surfaces in this study, the 250 µm radius value falls near the center of a curve which transitions from a generally "rough" regime to a generally "smooth" regime.

Tip Radius (µm)	Parameter						
	a (µm)	q (µm)	sk	z (µm)	pm (µm)	vm (µm)	dq (°)
2	*0.777*	*0.969*	*0.205*	*5.817*	*3.095*	*2.721*	*16.199*
5	*0.787*	*0.974*	*0.141*	*5.674*	*2.878*	*2.797*	*12.749*
10	*0.780*	*0.962*	*0.103*	*5.502*	*2.733*	*2.769*	*10.722*
25	*0.745*	*0.917*	*0.066*	*5.060*	*2.497*	*2.563*	*7.718*
50	*0.707*	*0.874*	*0.014*	*4.785*	*2.302*	*2.483*	*5.763*
75	*0.680*	*0.843*	*-0.018*	*4.636*	*2.181*	*2.455*	*4.889*
100	*0.659*	*0.818*	*-0.033*	*4.488*	*2.092*	*2.396*	*4.323*
250	*0.573*	*0.712*	*0.001*	*3.868*	*1.789*	*2.079*	*2.897*
500	*0.498*	*0.623*	*0.090*	*3.264*	*1.561*	*1.702*	*1.975*
750	*0.465*	*0.583*	*0.129*	*2.982*	*1.443*	*1.539*	*1.520*
1000	*0.447*	*0.561*	*0.140*	*2.799*	*1.367*	*1.432*	*1.268*
2500	*0.393*	*0.495*	*0.207*	*2.324*	*1.144*	*1.180*	*0.774*
5000	*0.355*	*0.444*	*0.308*	*1.958*	*0.985*	*0.973*	*0.530*
7500	*0.336*	*0.420*	*0.347*	*1.771*	*0.898*	*0.872*	*0.420*

Table 3.3 *Average parametric data for turned profiles as a function of tip radius.*

Once again, it is very interesting to note the behavior of the parameters relative to the smaller tip radius values. Nearly all of the parameters except for the RMS slope (*dq*) appear to change very little between the 2.0 μm tip radius and the 5.0 μm tip radius - thus indicating that a tip radius smaller than 2.0 μm is not easily justified. However, the significant changes in the RMS slope parameter between these tip radius values tends to indicate that smaller profile features may be masked by even the smallest of radii used in this study.

3.3.3 Stratified (Plateau Honed) Profiles

With the increase in functional demands on engineering surfaces, there is a growing presence of stratified surfaces in industry - particularly in sliding and sealing applications. These surfaces, commonly generated by processes such as "plateau honing", are very common in internal combustion engine cylinder bores and many critical sealing and sliding interfaces such as those in fuel injection systems and pumping applications (Fischer 1982). Furthermore, these types of surfaces pose many interesting analysis and characterization challenges and are therefore the basis for a great deal of work in the surface metrology community (Williamson 1967, Campbell 1972, Whitehouse 1985, Malburg 1989, Clark and Grant 1992, Malburg and Raja 1993).

Given the functional importance of these surfaces and their level of "popularity" in the surface metrology community, it is appropriate to include them in the tip convolution study. In the study of tip convolution effects, six profiles were obtained from plateau honed diesel engine cylinder liners. A typical trace is presented in Figure 3.14 and a 1.0 mm segment is expanded in Figure 3.15. It should be noted, however, that is only one example of a plateau honing process and very different results can be achieved by varying the process parameters (Malburg and Grant 1992).

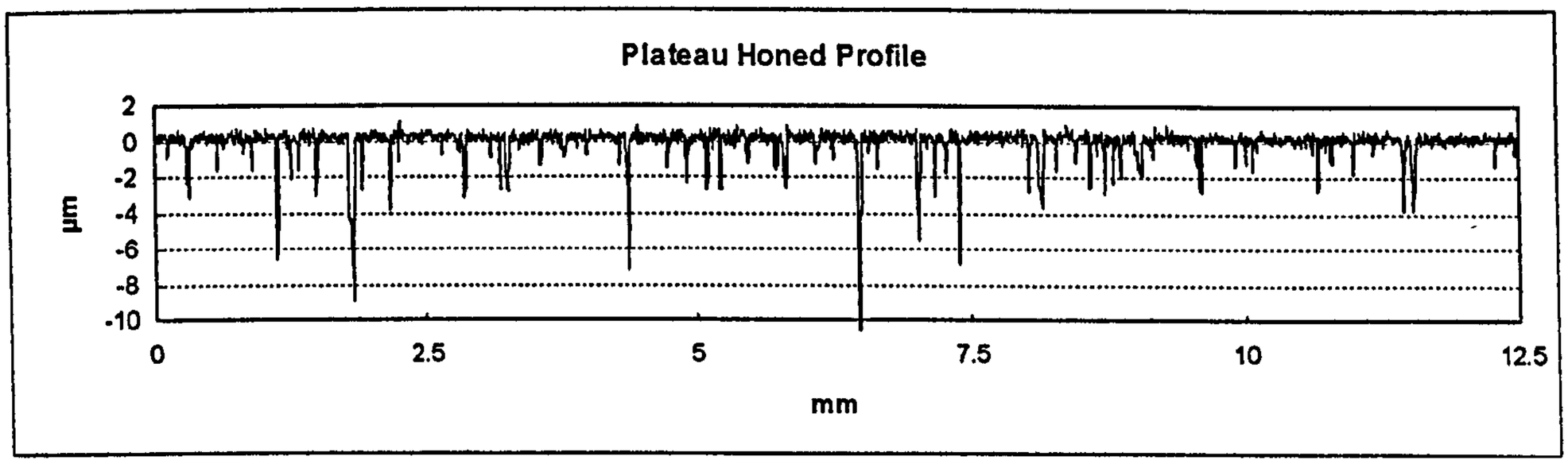


Figure 3.14 *Typical stratified (plateau honed) profile.*

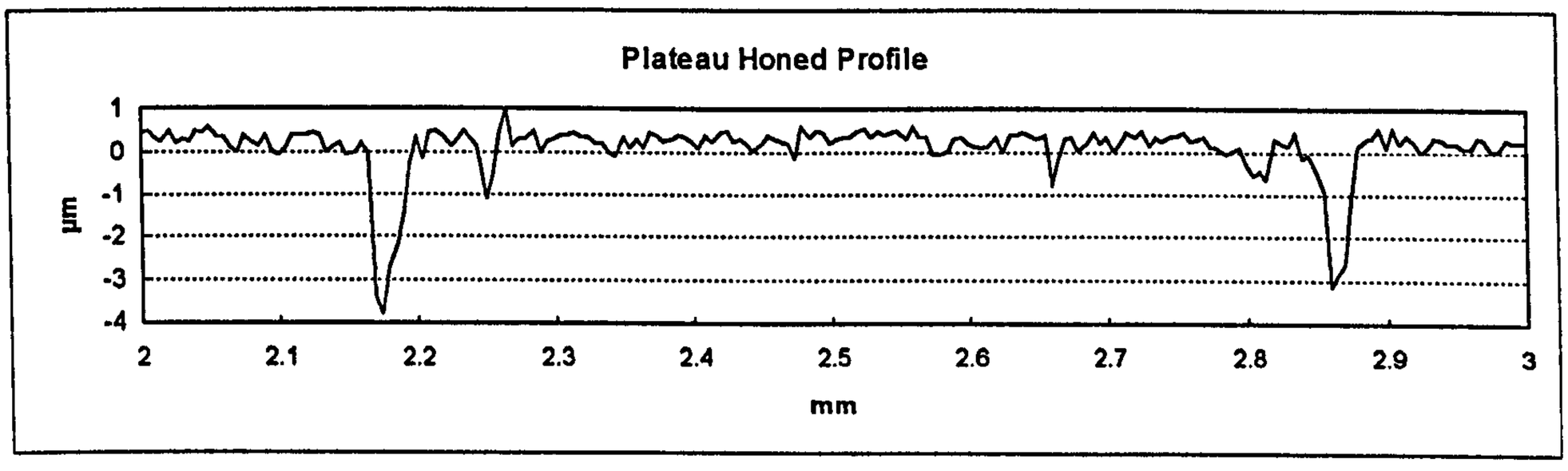


Figure 3.15 *1.0 mm section of stratified (plateau honed) profile.*

The Fourier transform of the plateau honed profile yields a spectrum which is quite similar to that of the ground profile apart from somewhat lower amplitudes in the short wavelength components. This is shown in Figure 3.16.

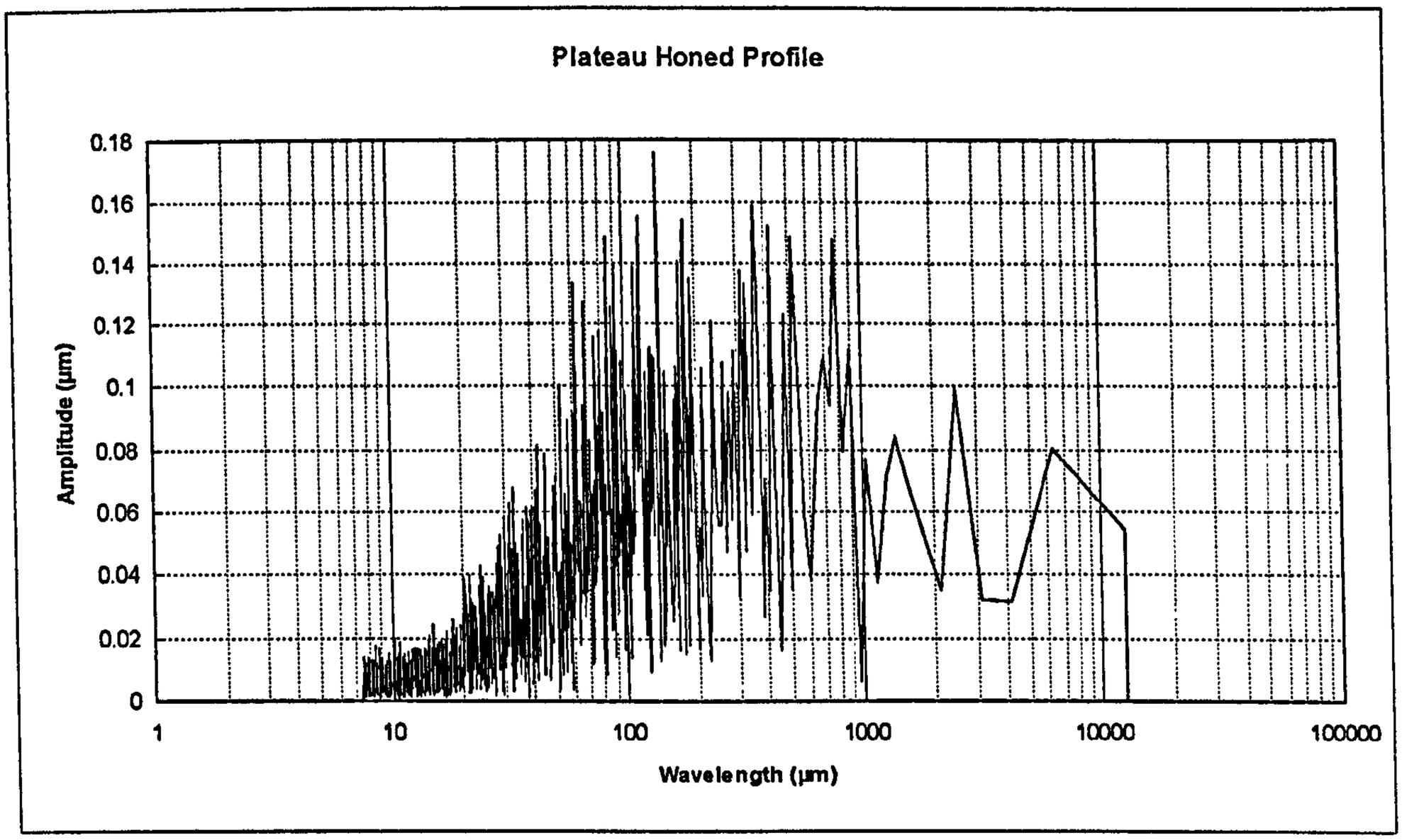


Figure 3.16 *Wavelength content of a stratified (plateau honed) profile.*

Although there is some similarity in the wavelength domain, the nominal stylus transmission data is very different. The average transmission characteristics (given in Appendix B) and the manually generated nominal characteristics shown in Figure 3.17 indicate that the stylus tip influences significantly attenuate all wavelengths present in the data set.

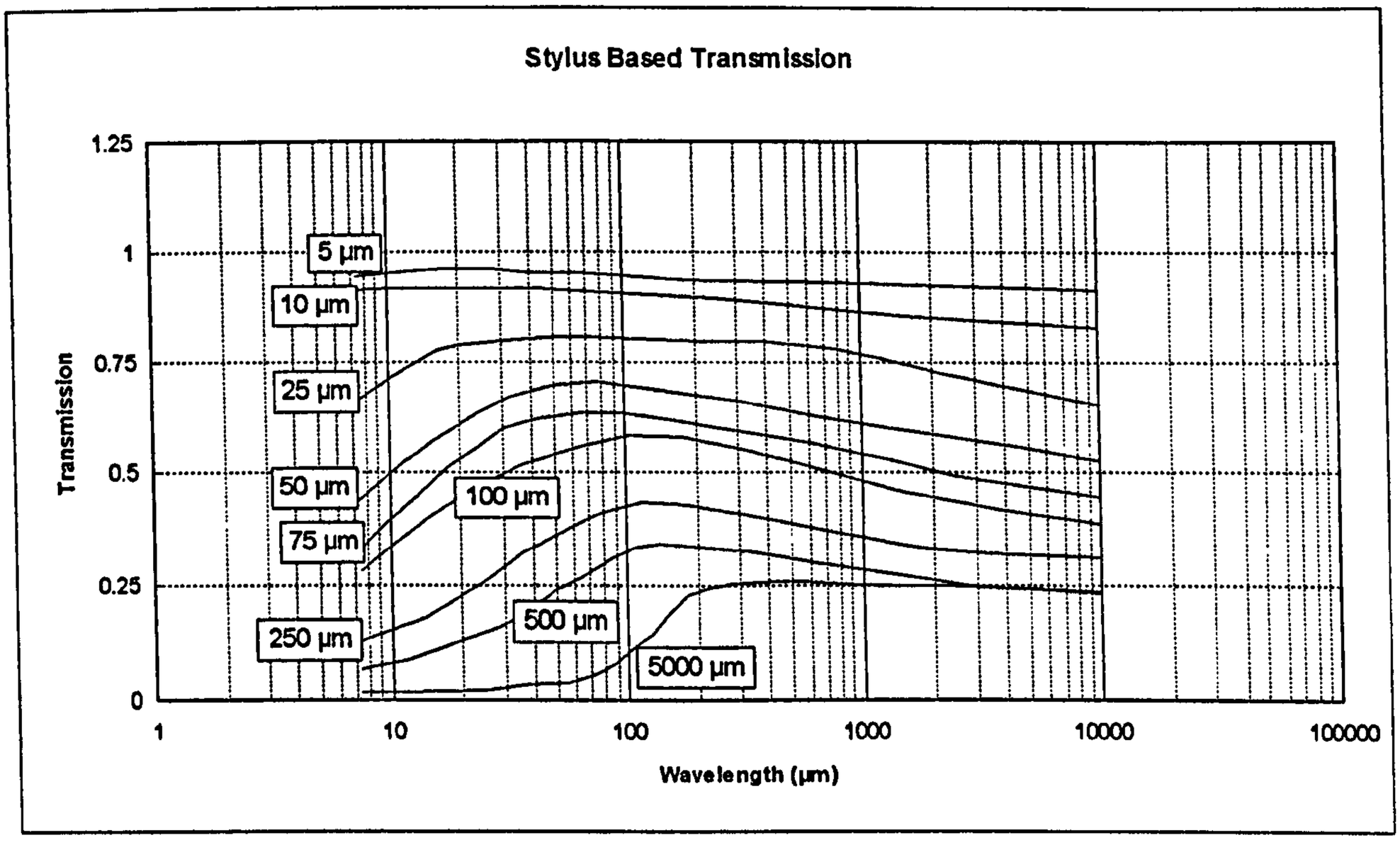


Figure 3.17 *Nominal stylus based wavelength transmission for stratified (plateau honed) profiles.*

This data indicates that the stylus influences are not limited to a certain *one-sided* transmission function as indicated in the ISO model of Figure 3.3. Instead, this study indicates that any increase in tip radius results in an attenuation of all wavelengths present in the data set. Furthermore, while the analysis of ground and turned profiles typically indicated an amplification of long wavelength components, the plateau honed surfaces generally demonstrated an attenuation of long wavelengths.

The difference in transmission of long wavelength content (between plateau honed surfaces and the other two types) can be attributed to the relative height and spacing of significant profile peaks. For the ground and turned profiles, the highest profile peaks were somewhat varying in amplitude and were a relatively significant distance apart (as compared to the convolved radius). Thus, when large radius values were convolved, the resulting profile exhibited significant longer wavelength attributes. For the plateau honed profiles, on the other hand, there are no significant high peaks and the profile

peaks that are present are very closely spaced. Thus, for the plateau honed data sets, a much "flatter" profile is generated upon convolving larger tip radii.

The numerical parameter calculations seem to also show trends which are unique to the plateau honed data sets. The parameters are graphically represented in Appendix B and the average values (for the six profiles) are tabulated in Table 3.4.

Tip Radius (µm)	Parameter						
	a (µm)	q (µm)	sk	z (µm)	pm (µm)	vm (µm)	dq (°)
2	*0.655*	*1.038*	*-3.598*	*8.016*	*1.238*	*6.779*	*10.329*
5	*0.602*	*0.970*	*-3.777*	*7.702*	*1.162*	*6.540*	*8.601*
10	*0.559*	*0.907*	*-3.873*	*7.341*	*1.092*	*6.249*	*7.452*
25	*0.472*	*0.769*	*-3.965*	*6.315*	*0.974*	*5.341*	*5.559*
50	*0.392*	*0.635*	*-4.038*	*5.234*	*0.872*	*4.361*	*4.119*
75	*0.346*	*0.553*	*-3.919*	*4.547*	*0.813*	*3.734*	*3.347*
100	*0.315*	*0.496*	*-3.704*	*4.055*	*0.772*	*3.283*	*2.863*
250	*0.232*	*0.346*	*-2.829*	*2.710*	*0.654*	*2.057*	*1.686*
500	*0.185*	*0.263*	*-1.672*	*1.935*	*0.578*	*1.357*	*1.105*
750	*0.165*	*0.229*	*-1.017*	*1.595*	*0.539*	*1.056*	*0.862*
1000	*0.153*	*0.210*	*-0.625*	*1.419*	*0.514*	*0.905*	*0.724*
2500	*0.127*	*0.169*	*0.336*	*1.001*	*0.444*	*0.557*	*0.418*
5000	*0.115*	*0.153*	*0.704*	*0.791*	*0.398*	*0.393*	*0.280*
7500	*0.111*	*0.148*	*0.840*	*0.713*	*0.374*	*0.339*	*0.223*

Table 3.4 *Average parametric data for plateau honed profiles as a function of tip radius.*

The above calculated parameters (and the corresponding graphs in Appendix B) indicate that many of the averaging amplitude and extreme amplitude parameters exhibit relatively small changes with increasing tip radius for radii up to approximately 10 μm. This indicates that tip radii on this order tend to conform to the majority of dominant surface features. For tip radii between 10 and 1000 more significant changes in the averaging parameters is observed. This can be understood as tip radii on this order are beginning to *bridge* the profile valleys and this is further confirmed by the reduction in mean valley depth, vm. In the convolution of styli greater than 1000 μm, we once again observe relatively small changes in computed parameters. This is consistent with the "bridging" of dominant profile features and further increases in tip radius do not dramatically affect resulting profiles.

In the context of plateau honed or "stratified" surfaces, changes in the skewness, (*sk*) parameter should be noted. Typically plateau honing generates a significantly negative value for skewness. In the study of tip radius effects, this skewness is shown to decrease slightly for tip radii up to approximately 50 μm. This can be understood as these relatively small tip radii tend to *smooth* out fine profile features associated with the peaks of the surface while maintaining a significant portion of the valley depths. Furthermore, the profile valleys are *narrowed* by these relatively small radii thus generating more negative skewness (and potentially more kurtosis). For relatively large tip radii, the stylus acts as a *bridge* across the profile valleys, thus *masking* the surface's negative skewness.

3.4 Special Case: Flanking of Styli

In the measurement of surface roughness with diamond styli, the stylus convolution process can also be influenced by flank angle geometry. Since typical diamond styli are based on conical geometries with spherical tips, the conical portion of the stylus must be considered in the convolution process. Whitehouse (1974) considered this problem and adopted an "effective tip geometry" or "tip dimension" which incorporates not

only tip radius, but also the stylus flank angles. It should be noted that this *flanking* nature is presented here as an important metrological consideration in the context of tip convolution. However, this concept is not included in the radius based wavelength limitation, as this is a spherically based approach for the prediction of a path of some a rolling element.

The *flanking* of a stylus has often been thought of as a disadvantage in contact (stylus) based measurement in the presence of re-entrant surface features as shown in Figure 3.18 and by Thomas (1982). Fortunately, the local slopes on most engineering surfaces are small enough that *stylus flanking* does not occur . This view has been somewhat turned around recently in that it has been shown that the analysis of local slopes in a data set can be used to predict the presence of steep or re-entrant surface features. This work represents an important aspect of tip convolution and has been undertaken in conjunction with the work documented in this chapter. However, the details of the technical discussion will not be included in this chapter as they are being published separately (Malburg et al. 1996 included in Appendix C).

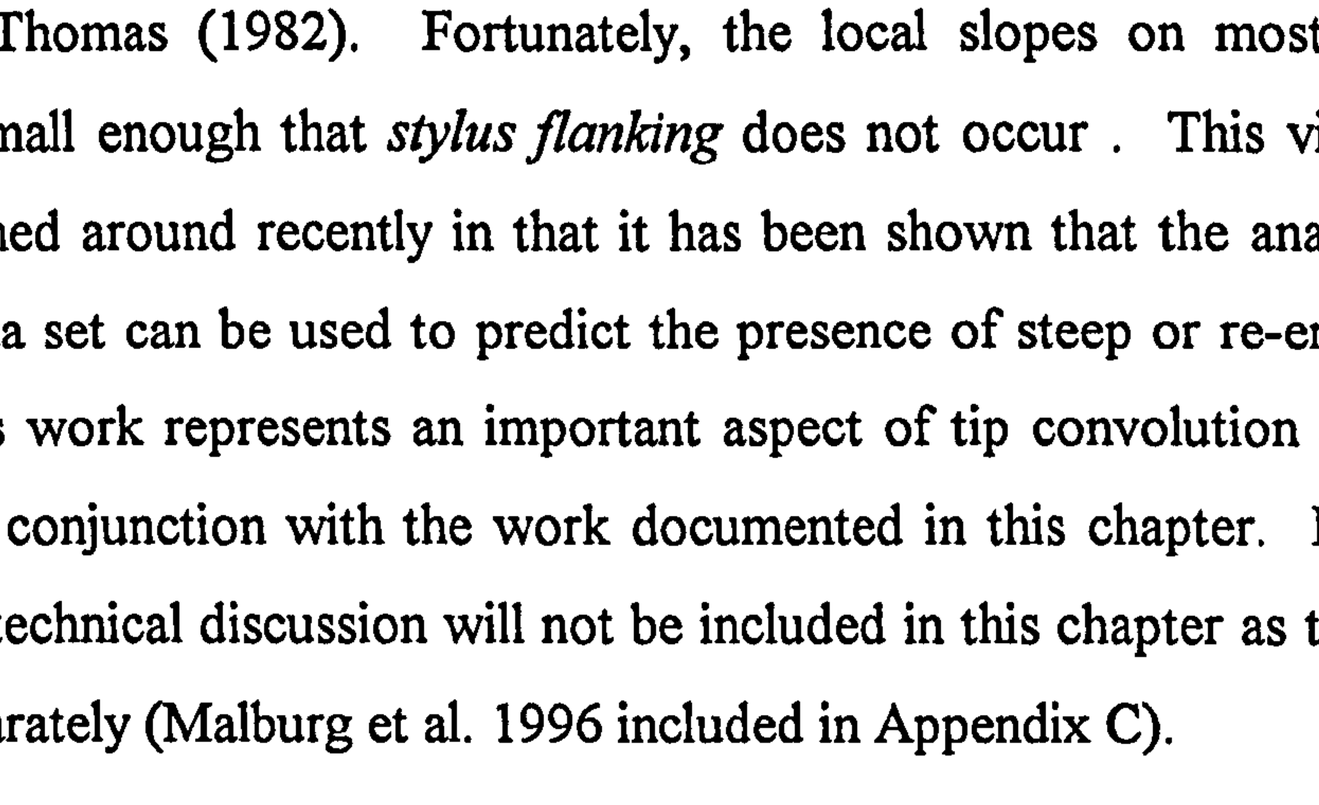


Figure 3.18 Stylus flanking in the presence of re-entrant surface features.

When very steep or re-entrant surface features of adequate height are encountered by the stylus, the stylus flank angle determines the resulting profile in that local area. This

is demonstrated in a section of a profile obtained from a diesel compression brake bore surface which was bored and subsequently reamed. SEM images (see Appendix C) indicated that the surface exhibited relatively smooth plateaus with a burr folded over (on one side) into the adjacent valley. An expanded section of a roughness profile (Figure 3.19) indicates stylus flanking on the corresponding side of the plateaus.

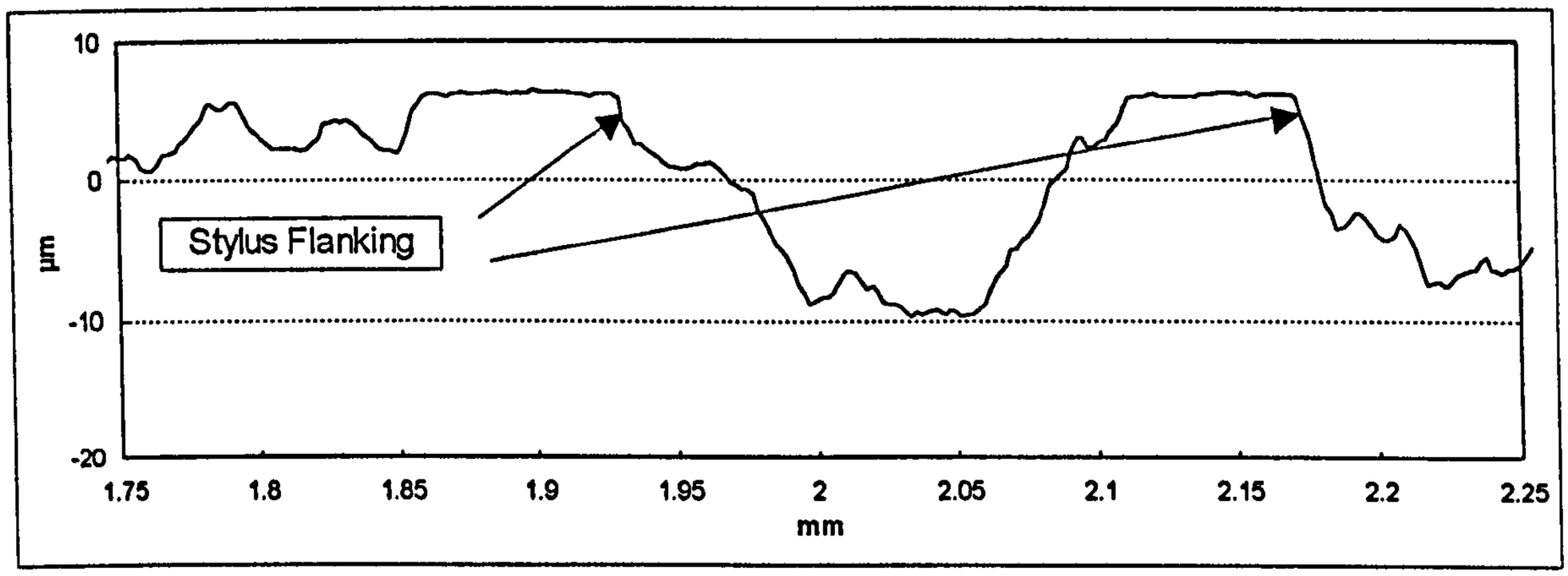


Figure 3.19 *Trace from bored/reamed cast iron surface. (2 µm tip radius, 0.5 µm data point spacing, 2.5 µm long-pass Gaussian filter)*

By generating a distribution of slopes (such as the one used in the calculation of the RMS slope, dq, parameter) the presence of these extreme slopes can be quantified.

It should also be noted that, in order for this technique to be applicable, the re-entrant features must be large enough to extend above the nominally spherical portion of the stylus tip. For a given tip radius, r, and included angle, α, the minimum step height for flanking, h_{min}, is given by:

$$h_{\min} = r\left(1 - \cos\left(\alpha/2\right)\right) \tag{3.1}$$

For a typical roughness measuring stylus incorporating a 90° included angle and a tip radius of 2.0 µm, the minimum step height for flanking is approximately 0.6 µm. Thus, re-entrant features with localized step heights less than 0.6 µm will remain in contact

with the spherical portion of the stylus during the convolution. (Refer to Appendix C for the complete technical paper.)

3.5 Sampling Implications

In traditional digital signal processing, the Nyquist criterion ultimately determines the shortest wavelength realizable by a given sample spacing. However, the Nyquist criterion does not guarantee any level of amplitude transmission and therefore higher sampling densities are required in order assess a given wavelength. Historically, many metrological applications have been based on requirements of 7 points per wavelength (Feinprüf 1995) (providing, in the worst case, 97% transmission) or more recently (ISO 3274-1995) 5 points per wavelength (providing at least 95 % transmission).

Unfortunately, these guidelines are not generally established for radius based wavelength limitation. However, this issue of data density is important in the context of the scheme for unification in that parametric results may vary significantly as a function of data point spacing (Sharman 1967a, Chetwynd 1979b).

Guidelines for determining the necessary spacing relative to radius based wavelength transmission can be developed (as follows) based on an idealized model of radius-to-peak (or "cusped") interaction. Given this model a method for determining the maximum spacing can be derived based on the tip radius, r, the maximum slope to be realized by the tip, Θ, and the level at which the vertical height is to be resolved as a percentage of the total tip travel, h. This is graphically shown in Figure 3.20.

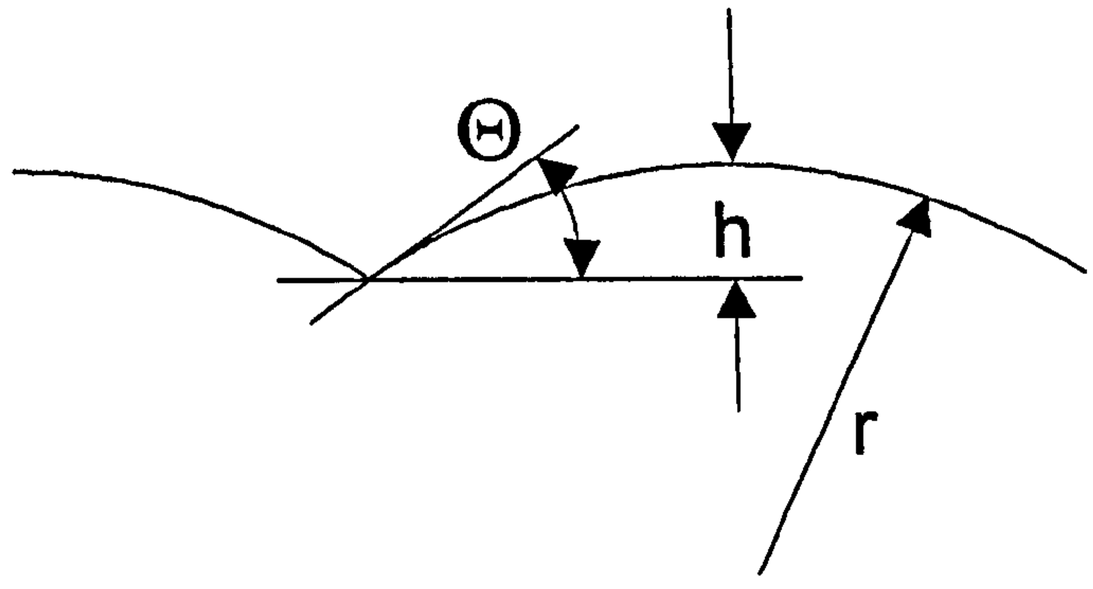


Figure 3.20 *The geometric elements of the cusped profile.*

For a given stylus tip radius, r, and maximum local slope to be realized, Θ, the total vertical travel of the tip is given by:

$$h = r\left[1 - \cos(\theta)\right] \tag{3.2}$$

If we desire the ability to assess 95% (for consistency with other filtering approaches) of the profile height as realized by the given tip radius, we must then consider the area at the sharp cusp. The worst case in terms of sampling will be when two data points fall equally on opposite sides of the cusp. Assuming the tip based profile to be linear in this small region we can model the bottom 5% of the profile at the cusp as shown in Figure 3.21.

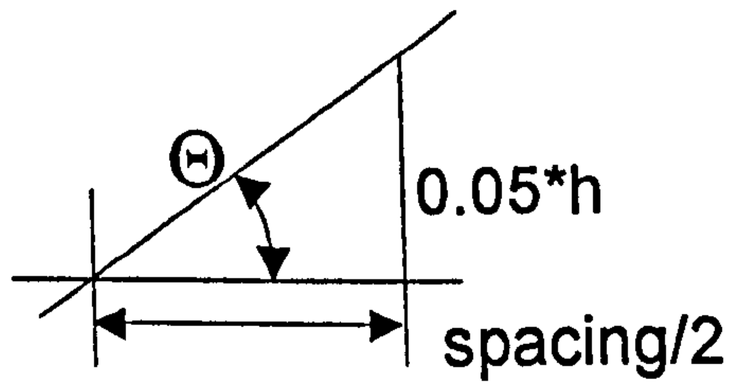


Figure 3.21 *Relationship between data point spacing and the corner of the cusp.*

Thus, for at least 95% retention of amplitude in the region of the cusp, the maximum data point spacing, S_{max}, is given by:

$$S_{max} = \frac{2 \cdot 0.05 \cdot h}{\tan(\theta)} = \frac{0.1 \cdot r\left(1 - \cos(\theta)\right)}{\tan(\theta)} \tag{3.3}$$

As an example, we can consider a scenario utilizing a 500 μm tip radius and predicted local slopes not exceeding 30°. The application of Equation 3.3 determines that the maximum data point spacing should not exceed 11.6 μm to retain 95% of the stylus based profile amplitude.

It should be noted, however, that this sampling requirement only applies when "radius based wavelength limitation" is applied. Sampling requirements for other wavelength limitation approaches are given in Chapter 5.

3.6 Summary

The convolution of a circular geometry over a profile can be used to predict some important functional aspects of a surface and thus it is proposed as a viable means of limiting wavelength limitation in the unified scheme.

To gain a better understanding of the wavelength limitation associated with radius convolution, various profiles were studied in terms of Fourier analysis and numerical parameters. These results confirmed that the tip convolution process is very dependent on the specific surface geometry and can therefore be quite unpredictable in application. For example, the analysis of ground surfaces seemed to confirm the general concepts of short wavelength limitation such as those put forth in recent ISO standards although it has been shown that the ISO guidelines may, in many cases, be conservative. However, the analysis of other engineering surfaces such as those resulting from turning or plateau honing shows that the wavelength limitation as a result of the stylus convolution can vary considerably. These variations are more clearly observed in terms of their lower limits (as shown in Figure 3.22), but variations (as amplifications or attenuations) in longer wavelength transmission is also present.

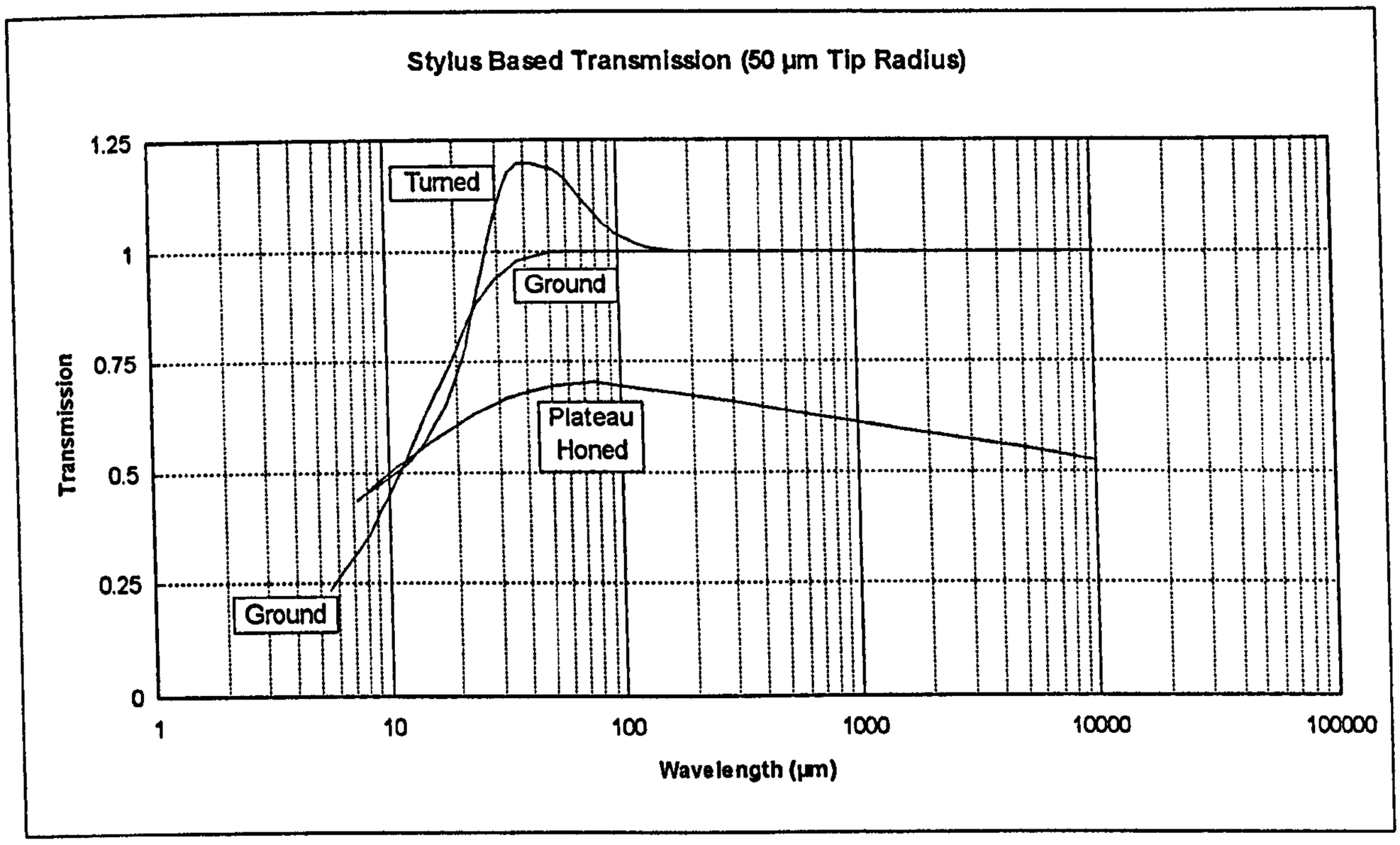


Figure 3.22 *Comparison of stylus based transmission functions for varying profiles.*

Recent ISO standards propose that, through the use of band-pass filtering, tip effects can be removed and correlation between instruments can be obtained (ISO 11562-1995 and Figure 3.3 above). This concept is confirmed in the analysis of the short wavelength limitation of ground profiles. However, the stylus based transmission characteristics shown in Figure 3.22 indicate that the tip radius influences in the wavelength domain for turned and stratified profiles are not as isolated in terms of short wavelength attenuation. In addition, all of the analyses incorporating relatively large tip radii seemed to yield some effects on longer wavelengths which were well inside the ISO bandwidths. This leads to the conclusion that, for the surfaces included in this study, tip radius influences cannot be kept completely outside the transmission band.

The modification of a broad spectrum of wavelengths cannot be easily accommodated in a filtering scheme such as the ISO proposed Gaussian band-pass based roughness

(see Figure 3.3 above). Thus, for optimal correlation between instruments used for a broad range of profiles, the tip radius remains a very important element and must be controlled outside the scheme for unification. Ideally, for each type of surface the stylus transmission effects will have to be characterized. This can be accommodated through Fourier approaches or parametric approaches, both of which have been included in the above study.

Furthermore, this unpredictability should enforce the need to limit the use of tip radius based wavelength transmission except for specific cases where rolling-type functional aspects are to be modeled.

A Unified Methodology for the

Application of Surface Metrology:

Chapter 4

Unwanted "Asperities" in Surface Metrology

In the analysis of data generated by the measurement of a surface, artifacts are often present which are not due to the surface but rather are the result of debris present on the surface. This can be a significant factor in terms of obtaining agreement between instruments and has been demonstrated in the CMM data from Chapter 1, Figures 1.6 and 1.7.

A "perfectly clean" surface exists only in theory in that nearly every surface is affected by surface layers such as oxides, however these are not of the primary concern in the surface metrology applications. The more devastating contaminations, relative to surface metrology, are the result of extraneous debris such as dust.

Given the increased demands being placed on surface metrology - particularly in manufacturing arenas, it has become necessary to develop the ability to tolerate such features in data sets. Often the measurement processes used to assess surface features are very time consuming and expensive when compared to the demand for numerical information for manufacturing process control. Thus, when an unwanted peak or *asperity* appears in a data set, the subsequent re-cleaning and re-measurement will significantly increase the measurement time (and the associated costs).

Based on the costs associated with measurement and the demand for timely results, it becomes desirable to devise a means of identifying and ignoring the features in the data set which do not correspond to the surface under test. This philosophy has been most successfully applied in the measurement of out of roundness, where many instruments incorporate software techniques for identifying and removing unwanted *asperities* (Feinprüf Perthen 1994, Rank Taylor Hobson 1992, Starbuck 1992).

The problems associated with the presence of unwanted asperities are common across all surface metrology applications. Furthermore, the methodologies for detecting and removing unwanted asperities can (and should) be made common across all

measurement technologies. Thus, it becomes necessary to address this topic in the context of the unification of surface metrology.

Using the measurement of roundness as a basis, this chapter will explore the influence of unwanted asperities on various analyses as well as current methodologies for detecting and removing these unwanted surface features. In addition, a more robust alternative for asperity detection, removal and data restoration (i.e. *padding*) is presented.

4.1 Influence on Parameters

The presence of a local, extraneous peak in a digital data set can have various effects on numerical parameters depending on the particular parameter and its associated mathematical definition. For the purposes of this discussion, the effect of unwanted asperities on four basic types of parameters will presented: extreme height, rate of change, statistical and spectral. (Specific parameters will be further defined and discussed in Chapter 6.)

4.1.1 Extreme Height Parameters

Obviously, extreme height or *peak-to-valley* parameters are the most effected by the presence of debris since these parameters isolate profile maxima and minima and debris tends to manifest itself as "artificial" profile maxima. This type of parameter is typically applied in the analysis of relatively long wavelength features and thus the shape of the unwanted asperity can be affected by the convolution of a relatively large stylus tip and subsequent long pass filtering.

As an example, the presence of unwanted asperities is a problem very common in the analysis of out-of-roundness for fuel injection components. A typical occurrence of an asperity in this context is shown in Figure 4.1 which includes data obtained from a fuel

injector plunger (outside diameter) using a Mahr Perthen MFU8 (3600 data points, 0.5 mm tip radius, 500 upr low-pass Gaussian filter).

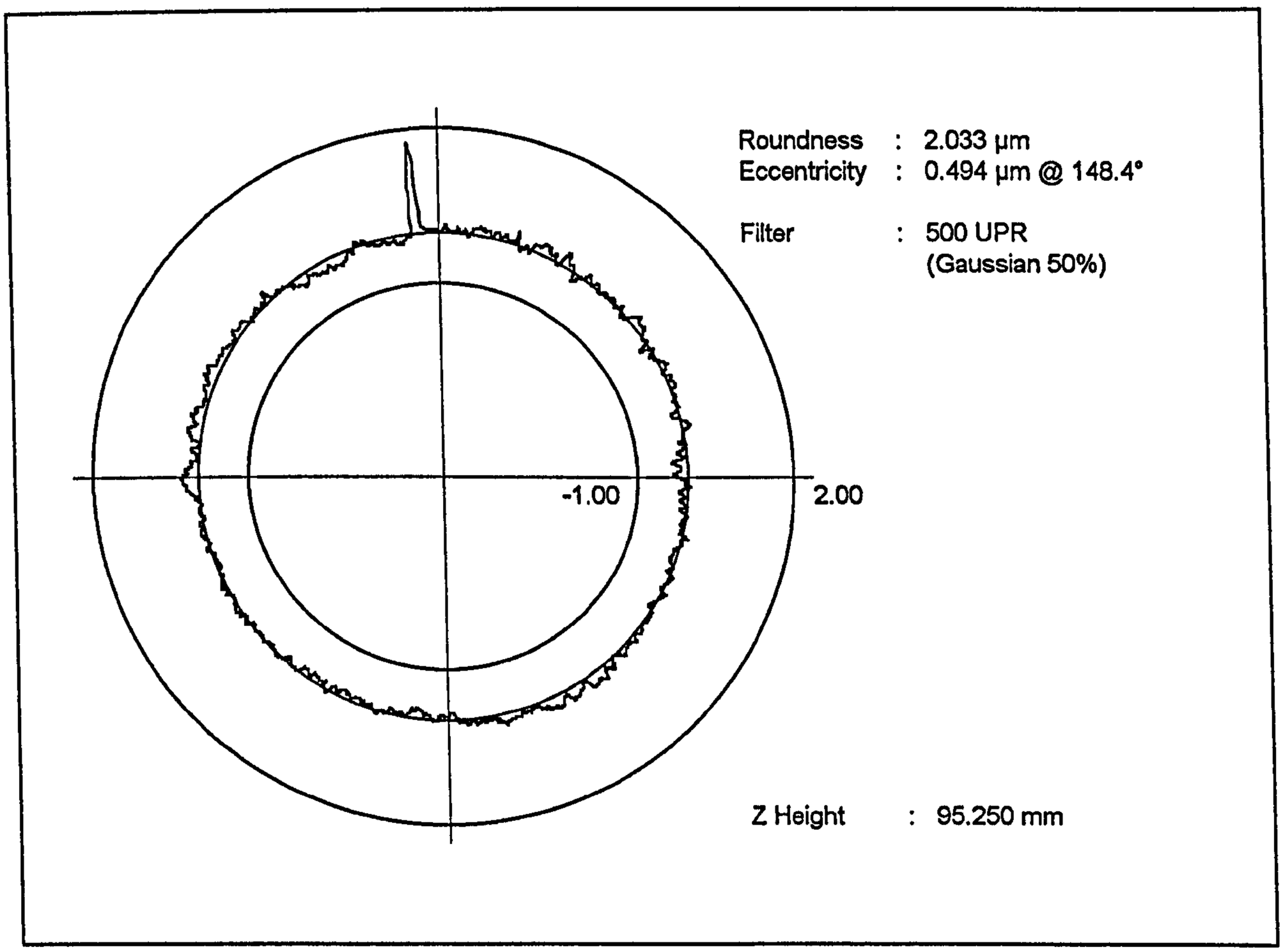


Figure 4.1 *An unwanted asperity in the analysis of out-of-roundness for a fuel injector component.*

In the Figure 4.1 example the out-of-roundness of the underlying surface is approximately 0.6 µm (based on measurements elsewhere on the component), however, the presence of the unwanted asperity causes the peak-to-valley out of roundness of the Figure 4.1 data set to be over 2.0 µm!

4.1.2 Rate of Change Parameters

Rate of change parameters are also significantly affected by the presence of extraneous peaks in a data set as these profile artifacts cause unusually abrupt changes in local profile slope. In typical rate-of-change analyses, (historically designated by the dr/dΘ

parameter (Whitehouse 1987a, Rank Taylor Hobson 1992)) a window is specified and the greatest height change occurring within such a window is reported. The relative slenderness associated with these unwanted asperities is such that the entire asperity falls within any typically used window, thus the reported parameter becomes based on the unwanted asperity rather than the underlying surface. For the Figure 4.1 data set, the rate of change parameter, within a 30° window, is approximately 2 μm with the asperity present and approximately 0.5 μm without the asperity.

4.1.3 Statistical Amplitude Parameters

The arithmetic average amplitude (historically referred to as "Ra" in roughness analysis) is only slightly influenced by the presence of an unwanted asperity. For the data shown in Figure 4.1, we find that the arithmetic average amplitude is 0.091 μm with the asperity present and 0.085 μm with the asperity removed resulting in a change of approximately 7%.

However, as the order of the parameter increases, so does the sensitivity to the presence of the asperity. (Once again, see Chapter 6 regarding parameter definitions.) This is captured in Table 4.1 for the Figure 4.1 data set.

Name (order)	Asperity Present	Asperity Removed
RMS (2)	0.14	0.11
Skewness (3)	4.23	-0.08
Kurtosis (4)	47.0	2.77

Table 4.1 Parameter variations for averaging amplitude parameters.

For these higher order parameters, the relatively few extreme points associated with an unwanted asperity can cause a significant impact on the numerical result. This is due to the fact the squaring, cubing, etc. of the extreme points results in numerical values which are even more extreme.

4.1.4 Spectral Effects

The presence of an unwanted asperity in a data set can, in many cases, go unnoticed in the analysis of the power spectrum or harmonic amplitudes. This is because, in most cases, the asperity represents relatively little power when compared to signal generated by the underlying surface. (This is consistent with the relatively small change in the RMS amplitude parameter.) Furthermore, the spectral effects associated with the unwanted asperity occur over a very broad range of frequency (typically expressed in undulations per revolution (upr) in the context of roundness).

For the data set shown in Figure 4.1, the effect of the asperity on the harmonic content (determined via Fourier Transformation per Reid (1992)) is shown in Figure 4.2.

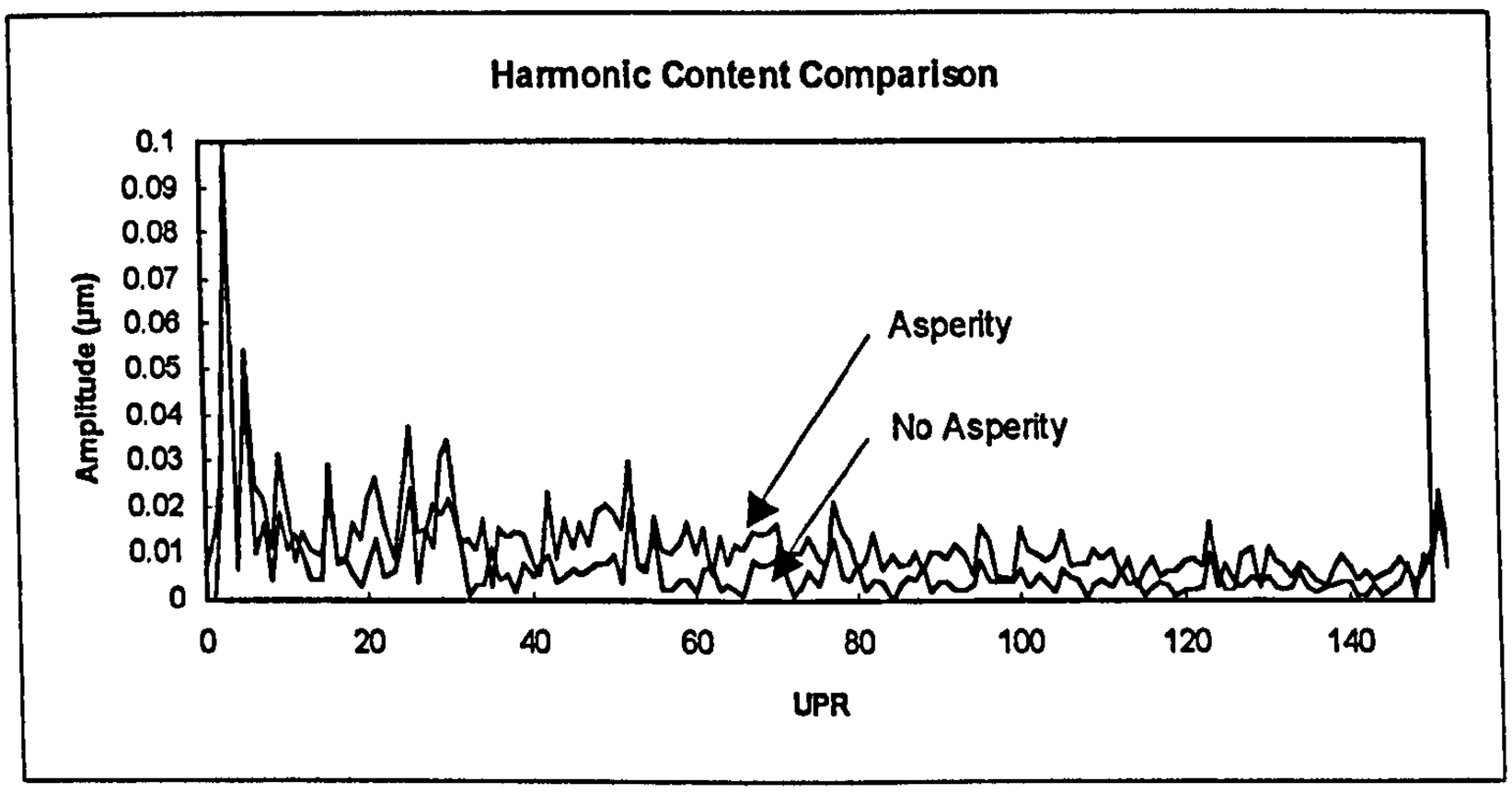


Figure 4.2 *Comparison of harmonic content.*

The Figure 4.2 representation indicates that the dominant peaks in the spectrum remain intact, despite the presence of the asperity. However, their amplitudes are somewhat reduced when the asperity is removed from the data set. The actual influence of the asperity can be assessed on a *per-wavelength* basis through the comparison of individual amplitudes along the Figure 4.2 curves. This is accomplished through a simple subtraction of amplitudes at each frequency and is shown in Figure 4.3.

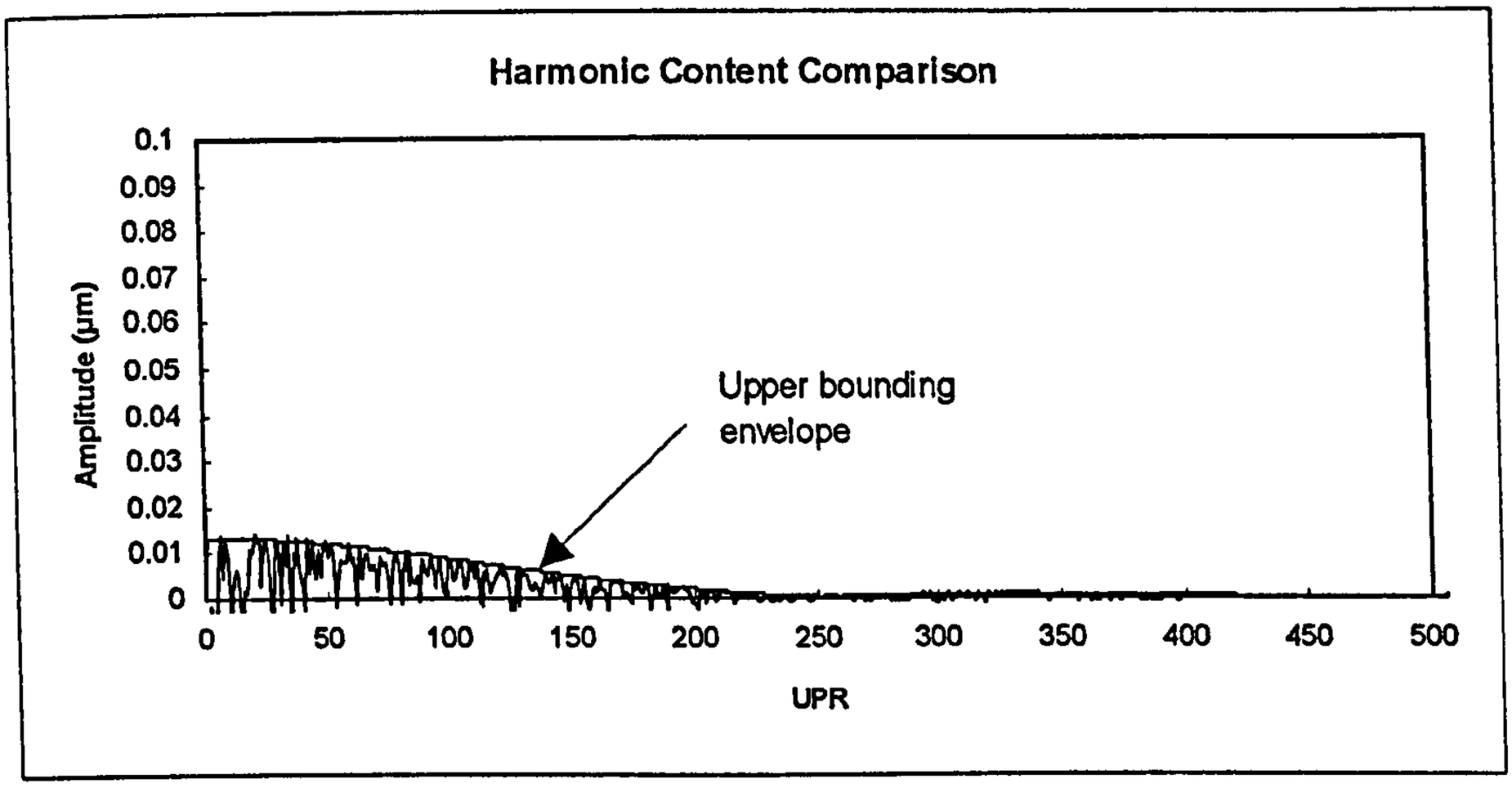


Figure 4.3 *Harmonic influence of the asperity.*

The harmonic content of the asperity is represented both theoretically and analytically in Figure 4.3. The analytical approach was based on subtraction of amplitudes at each wavelength for the measured profiles. This yielded the rather erratic data set plotted in Figure 4.3. The theoretical harmonic content related to the asperity can be determined by 1.) generating a data set of perfect roundness, 2.) adding the asperity which was removed from the original data set., and 3.) performing a Fourier transform to determine harmonic content. This theoretical curve establishes an "upper boundary" in that it does not always fall coincident with the analytical data set due to round-off errors for the "near-zero" amplitudes in the measured data.

The Figure 4.3 representation of the theoretical change in harmonic amplitude is limited in that it maintains the same *y* axis scale as did the comparison in Figure 4.2. A more informative representation of the theoretical influence of an asperity is given in Figure 4.4 which includes a broader range of frequencies and a change in *y* axis scaling.

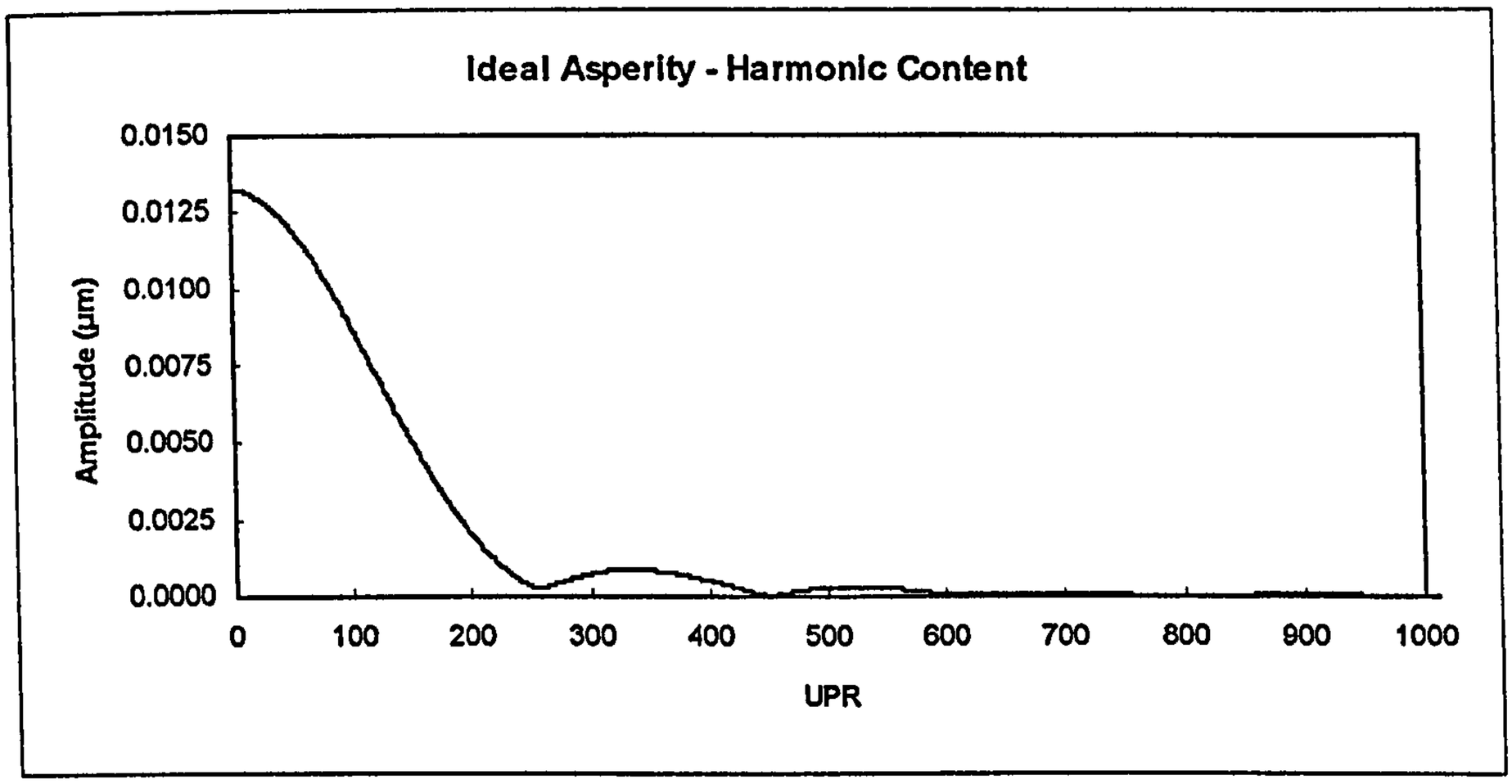


Figure 4.4 *Theoretical harmonic content of the Figure 4.1 asperity.*

4.2 Current Methods for Asperity Removal

The presence of unwanted asperities in surface metrology data sets is very common in many critical surface metrology applications. Three basic methodologies have historically been presented which can be applied to profile data sets:

- *Threshold based asperity removal.*

  Common in many commercial instruments.
- *Statistical threshold based asperity removal.*

  Proposed for CMM data sets.
- *Neighbor based outlier detection.*

  Used in time series applications.

In the following sections, each of these methods will be presented in terms of assumptions, implementation and shortcomings.

4.2.1 Threshold Based Asperity Removal

The most common software technique for the removal of asperities from surface metrology profiles is based on thresholding both the height and width of the asperity under test (Starbuck 1992). In this approach (graphically depicted in Figure 4.5), two thresholds are used as a means of assessing not only the height, but also the width of an asperity. For features which exceed the "Test Height" threshold, and are narrower than the "Test Width" (at the test height), a region of data equal to "Discard Width" is removed. Thus, a certain *aspect ratio* is incorporated in to the identification of an asperity.

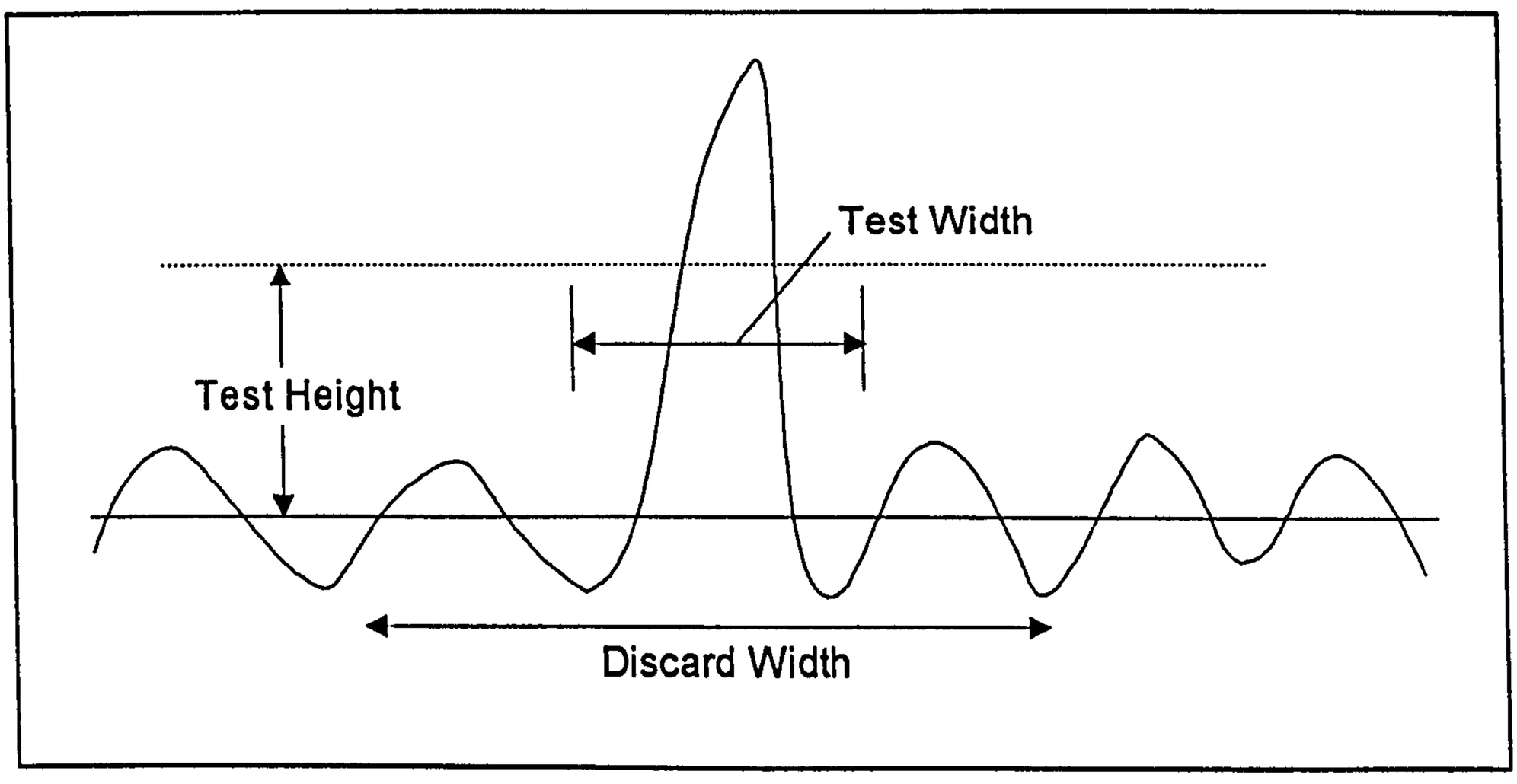


Figure 4.5 *Threshold based asperity removal.*

These three controls are typically user selectable, thereby allowing customization based on the specific application. The settings must be selected on the basis of some determination of a typical asperity. Example settings for the analysis of out of roundness for an injector component with a 5 mm radius might be: test height = 2 μm, test width = 2°, discard width = 3°. (These correspond to a 0.17 mm test width and a 0.26 mm discard width.)

One drawback of this threshold approach is that the setting of a fixed height will still allow asperities to be included as long as they are any amount under the test height. Thus, the threshold values must be constantly monitored relative to the current level of geometry being produced.

For example, a test height can be established for a given manufacturing process (Figure 4.6a). If process improvements are made such that the component's geometry is significantly improved, then the threshold criteria should modified accordingly. Otherwise, the improvements in component roundness will introduce a sensitivity to finer debris which will fall inside the test height and therefore influence the resulting analysis. Figure 4.6b demonstrates a scenario where the underlying form errors have been reduced, however, the presence of debris (falling inside the threshold) results in an erroneously high peak to valley determination.

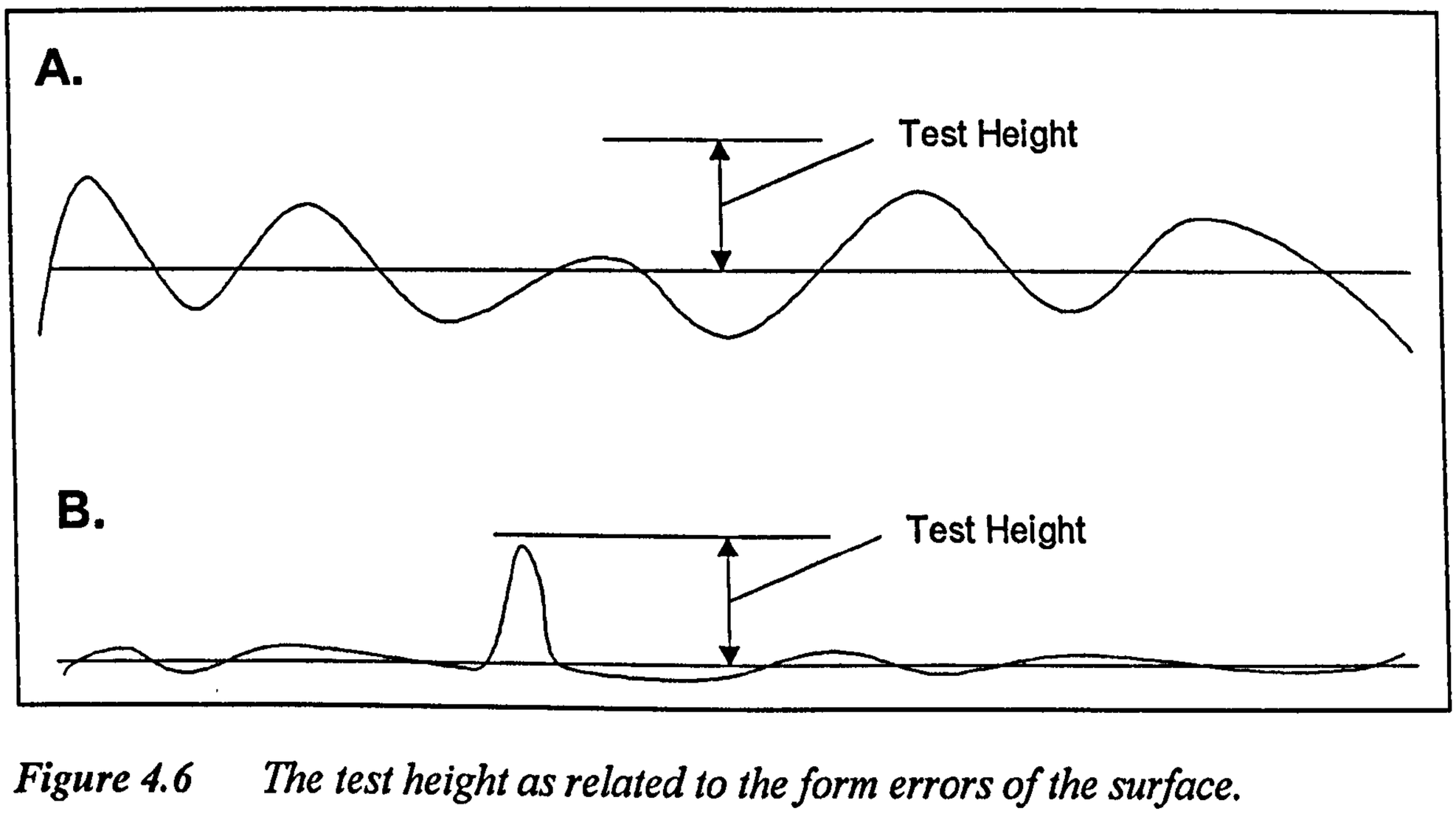


Figure 4.6 *The test height as related to the form errors of the surface.*

A. Threshold based on one level of production capability.

B. Same threshold used after process improvements.

In the Figure 4.6b scenario, process improvements could go unnoticed (in numerical terms) if the thresholds are not set properly. In order to be effective, these thresholds must be monitored and modified with manufacturing process changes. Unfortunately, in many applications the thresholds are established initially and are not subsequently re-evaluated.

Another important drawback relative to the fixed height approach is that it can ignore obvious asperities which may not exceed the height threshold. This can occur when an unwanted asperity is located in the valley portion of a long wavelength undulation (Figure 4.7). It may be argued that, in this scenario, the asperity has little or no effect on the resulting peak to valley roundness value. However, as shown above, the asperity will still influence statistical evaluations and Fourier analyses to an extent. Furthermore, the effect of the unwanted asperity on local slope or "rate of change" analyses will be very significant. Regardless of the analysis techniques that are used, the asperity is not part of the surface under evaluation and should therefore be considered for removal.

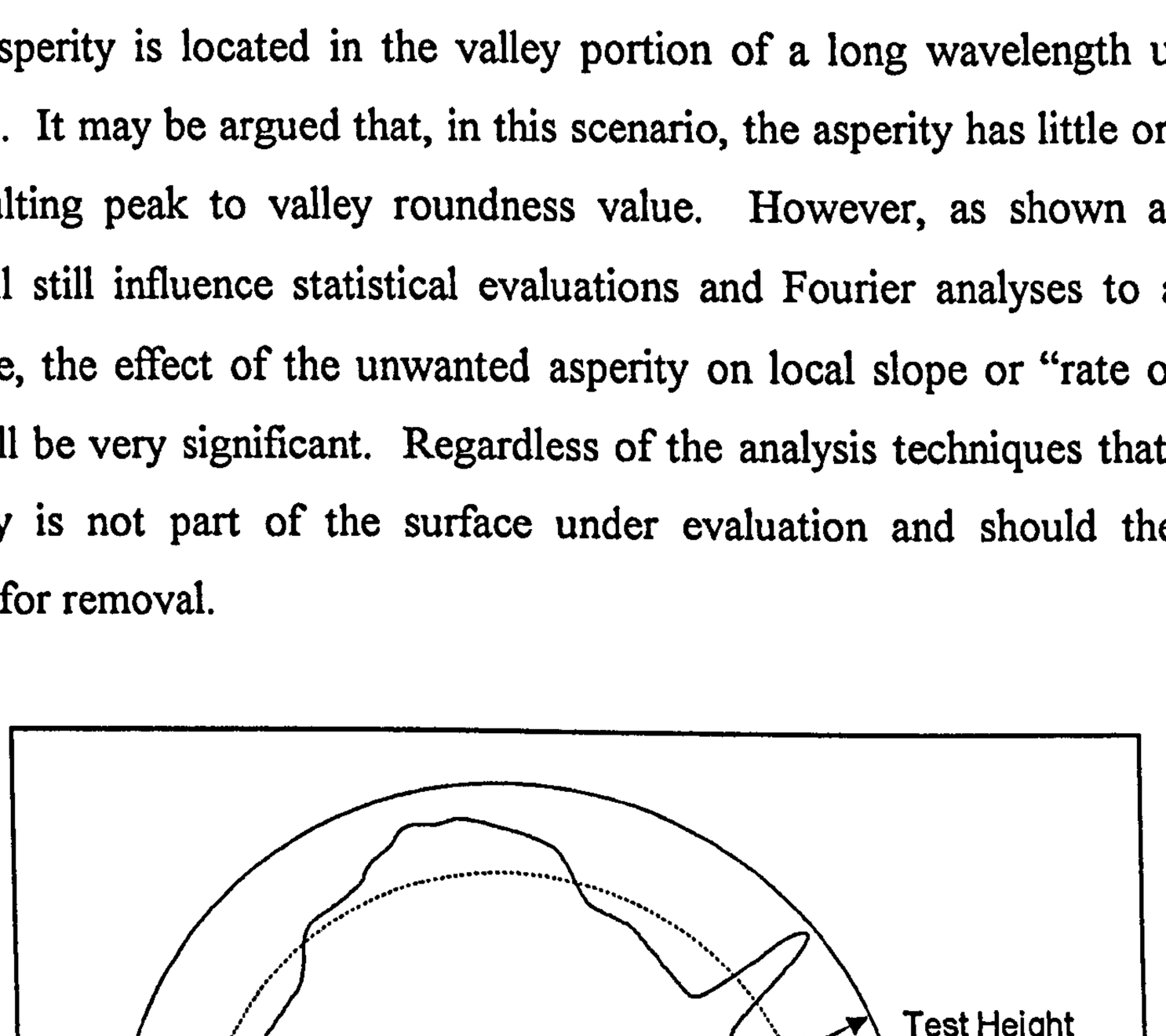


Figure 4.7 *Asperity not removed due to location in a long wavelength depression.*

A final consideration in the threshold based approach relates to the width aspect of the detection and removal of unwanted asperities. The control of the slenderness of an asperity is necessary to distinguish it from a long wavelength lobing condition. However, the use of a "discard width" in the actual removal of an asperity can lead to one of two errors - the removal of too little data or the removal of too much data. The former case results in an incomplete removal of the asperity, but in practice it seems to be quite rare. The latter case, however, is quite common in practice in that a typical asperity is considerably narrower than the test width and the removal width is always greater than the test width.

4.2.2 Statistical Threshold Based Asperity Removal

The detection and removal of asperities has been relatively common in the analysis of linear and circular geometries via roundness and cylindricity instrumentation. However, this topic is becoming more prevalent in CMM applications through the recent increases in data density and the proliferation of scanning probe systems (Salsbury 1996). An approach for the detection and removal of unwanted data points in CMM applications has been presented by Paterson (1985), whereby a statistically based threshold is applied to the data.

In this methodology, an unwanted asperity is determined based on the presence of data points outside the typical (assumed Gaussian) distribution. In application, this statistical threshold can be based on some limiting number of standard deviations - typically three. Thus, for a given data, set the procedure for detecting and removing data points is shown in Figure 4.8.

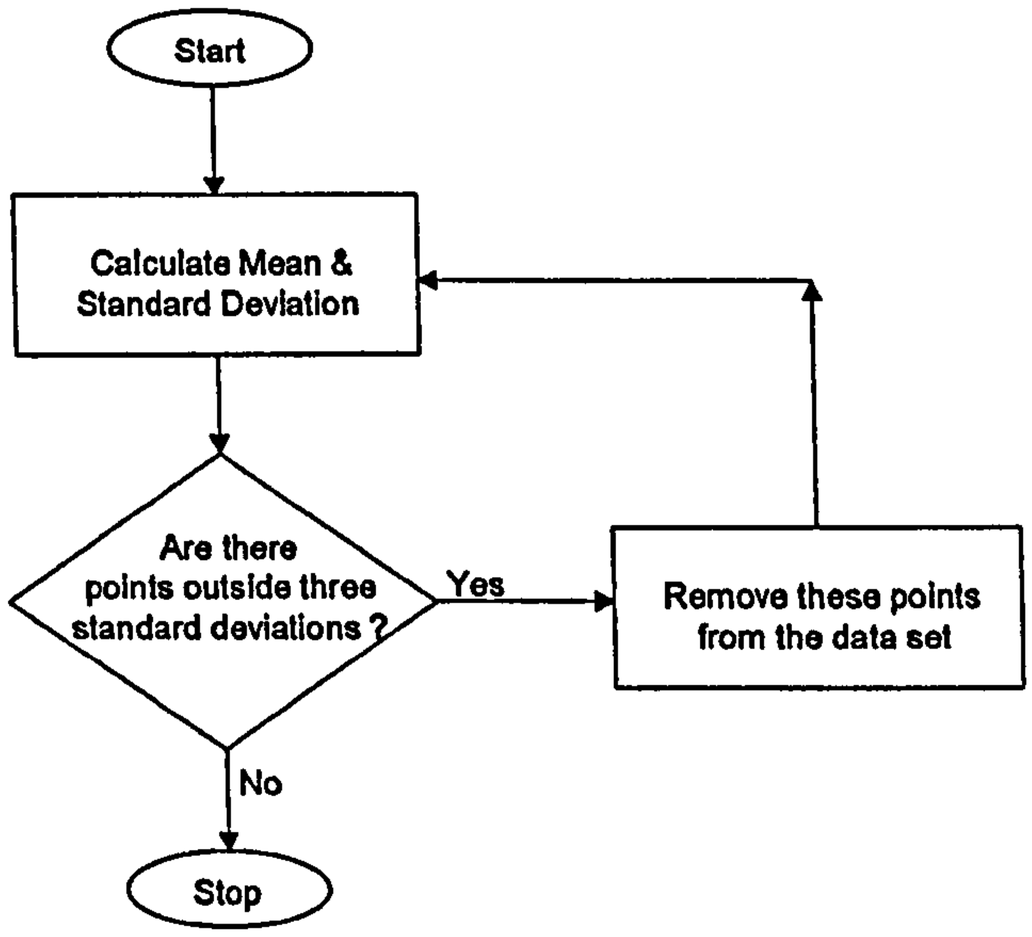


Figure 4.8 *Statistical threshold based data point removal.*

This technique for removing data points is iterative in nature and continues until all included data points remain inside the desired statistical limits. Although convergence has not been demonstrated mathematically by Paterson, he cites practical examples of rapid convergence on seemingly "difficult" data sets.

To demonstrate convergence for the technique we must consider, as a worst case, the smallest possible standard deviation (as this will result in the greatest amount of data discarding). For this derivation we consider the collection of n data points on a surface wherein $n-1$ of these points have a value x_1 and the nth point has a value x_2. It will be assumed that x_2 represents a profile asperity of value greater than x_1 and the goal is to determine the minimum number of data points, n_{min}, such that x_2 falls outside the three standard deviation limits. For any data set smaller than n_{min} no data points will fall outside the three sigma limits and the iterative process of Figure 4.8 will terminate.

First the mean, μ, and standard deviation, σ, are determined:

$$\mu = \frac{(n-1)x_1 + x_2}{n} \tag{4.1}$$

$$\sigma = \sqrt{\frac{(n-1)(x_1-\mu)^2 + (x_2-\mu)^2}{n-1}} \tag{4.2}$$

which expands to:

$$\sigma = \sqrt{\frac{(n-1)\left(x_1 - \frac{(n-1)x_1 + x_2}{n}\right)^2 + \left(x_2 - \frac{(n-1)x_1 + x_2}{n}\right)^2}{n-1}} \tag{4.3}$$

Factoring Equation 4.3 is performed as follows

$$\sigma = \sqrt{\frac{(n-1)\left(\frac{nx_1}{n} - \frac{(n-1)x_1}{n} - \frac{x_2}{n}\right)^2 + \left(\frac{nx_2}{n} - \frac{(n-1)x_1}{n} - \frac{x_2}{n}\right)^2}{n-1}}$$

$$= \sqrt{\frac{\frac{(n-1)}{n^2}(x_1-x_2)^2 + \frac{(n-1)^2}{n^2}(x_2-x_1)^2}{n-1}}$$

which results in:

$$\sigma = \sqrt{\frac{1}{n^2}(x_1-x_2)^2 + \frac{(n-1)}{n^2}(x_2-x_1)^2} \tag{4.4}$$

Since,

$$(x_1 - x_2)^2 = (x_2 - x_1)^2$$

The calculation of the standard deviation (Equation 4.4) becomes

$$\sigma = \sqrt{\frac{n(x_2 - x_1)^2}{n^2}} = \sqrt{\frac{(x_2 - x_1)^2}{n}} \tag{4.5}$$

The criteria for discarding x_2 as a profile peak is given by

$$x_2 > \mu + 3\sigma$$

or explicitly

$$x_2 > \frac{(n-1)x_1 + x_2}{n} + 3\sqrt{\frac{(x_2 - x_1)^2}{n}} \tag{4.6}$$

Equation 4.6 can be algebraically reduced as follows

$$n x_2 > (n-1)x_1 + x_2 + 3\sqrt{n}x_2 - 3\sqrt{n}x_1$$

$$\left(n - 3\sqrt{n} - 1\right)x_2 > \left(n - 3\sqrt{n} - 1\right)x_1 \tag{4.7}$$

Since $x_2 > x_1$ it follows that the criteria for discarding x_2 becomes

$$\left(n - 3\sqrt{n} - 1\right) > 0 \tag{4.8}$$

which can be factored as follows

$$n-1>3\sqrt{n}$$

$$n^2-2n+1>9n$$

$$n^2-11n+1>0 \qquad (4.9)$$

Solving Equation 4.9 via the quadratic formula yields, solutions at 0.09167 and 10.90833. The smaller value is discarded as it violates Equation 4.8. Thus

$$n_{min} = 10.9083$$

Since the number of points sampled will always be a whole number, at least eleven data points will be necessary in order to cause the removal of one point. Thus it is concluded that the procedure will converge to a minimum of ten data points.

The above convergence derivation is very conservative in that the "non-asperity" data points were all of the same value. This is the absolute worst case. In practice there is typically some variation in the "non-asperity" data points and there may be multiple data points on the unwanted asperity. In either case, a higher value for standard deviation will result and more data points (than the minimum value of 10) will ultimately be included.

The statistical thresholding of data sets has the advantage of being *adaptive* in light of process changes such as in Figure 4.6. However, the Figure 4.7 scenario of an asperity located in a local depression will still go undetected. Furthermore, this approach typically discards only the data points that are outside the threshold limit. This can result in the retaining some portion of the asperity as well as rather sharp discontinuities in the resulting data set (Figure 4.9).

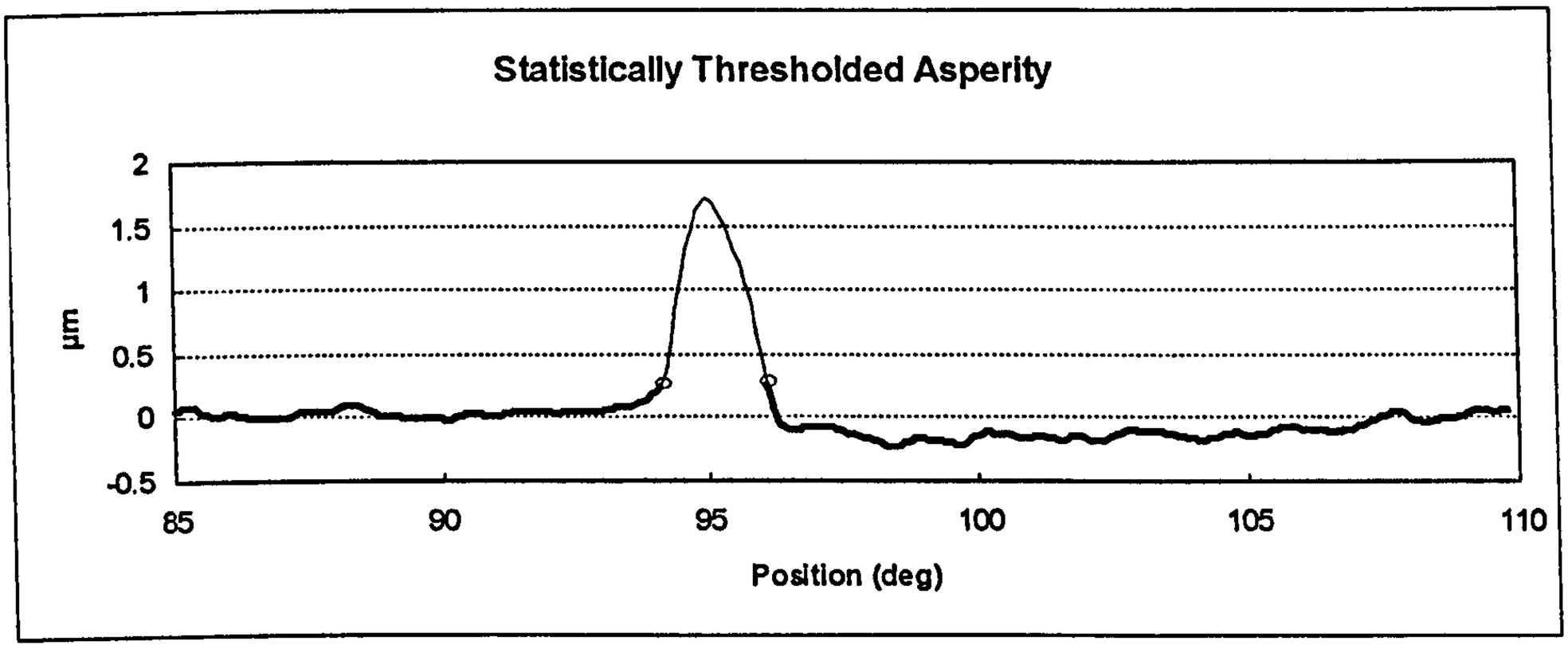


Figure 4.9 Statistically thresholded asperity from Figure 4.1 data set.

4.2.3 Neighbor Based Outlier Detection

Outside the field of surface metrology are many other fields which encounter unwanted features in time series data sets. Rousseeuw and Leroy (1987) present an approach based on point-to-point changes in amplitude. In this methodology, outliers in the data set are related to extreme changes within the sampling interval of the data set. This approach can be graphically depicted as shown in Figures 4.10 and 4.11.

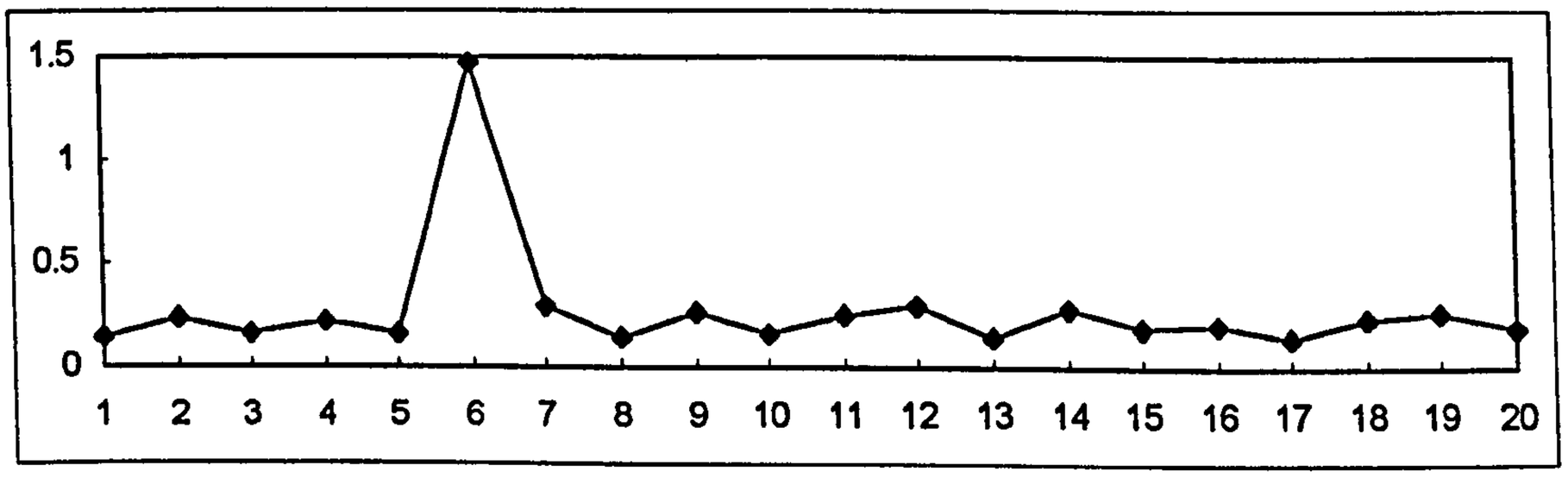


Figure 4.10 Common time-series outlier.

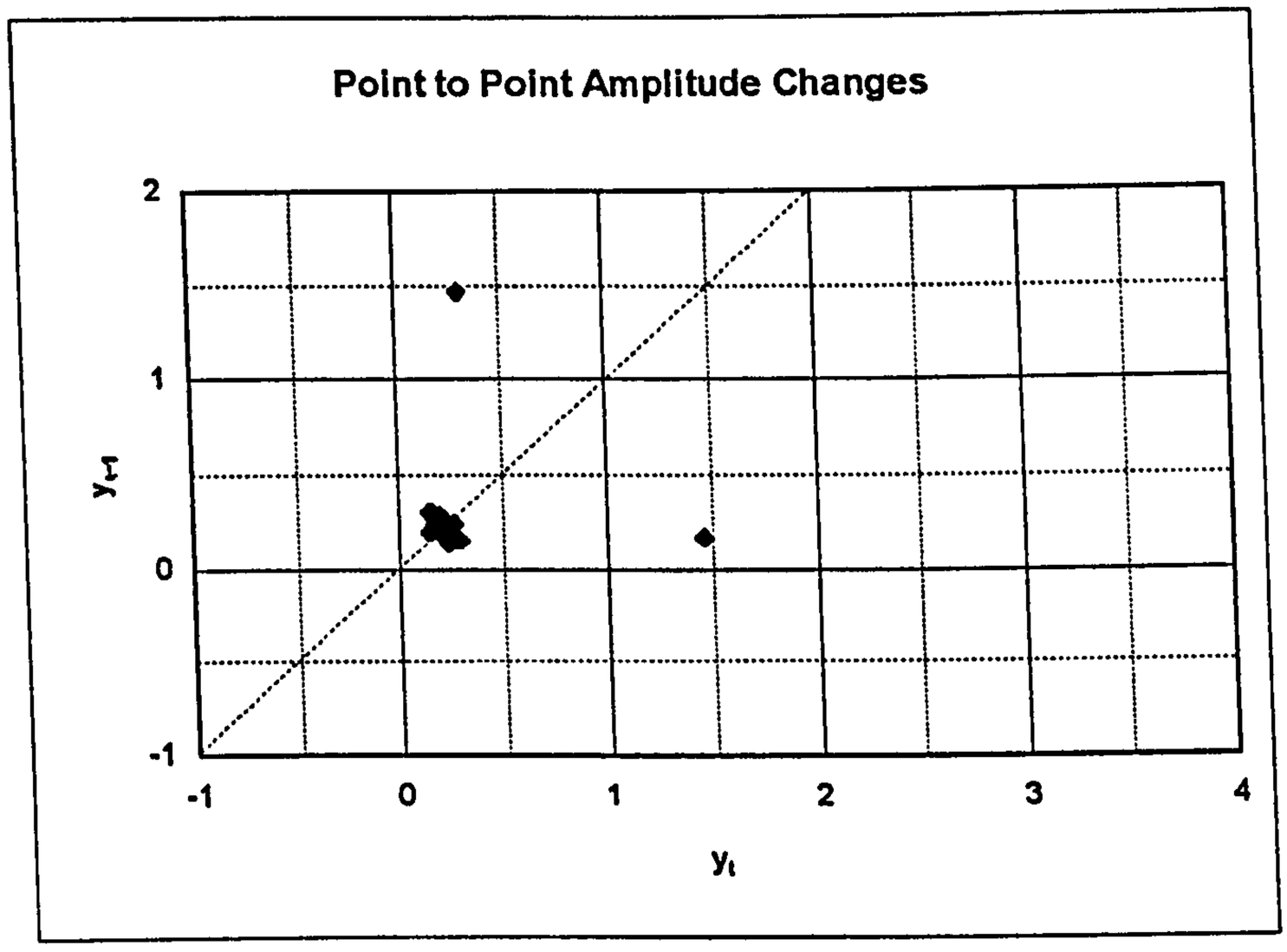


Figure 4.11 *Neighbor-based amplitude changes for Figure 4.10 data set.*

The Figure 4.11 graph is a scatter plot of local variations, whereby y_{t-1} is plotted versus y_t. When y_{t-1} equals y_t the point will fall on a 45° line. The graph indicates that two of the plotted points are significantly different from the others. (This significance is established based on the distance between the plotted points and a 45° line as compared to the typical "spread" of the points.) The points (on the Figure 4.11 representation) that are the farthest from the 45° line are related to height changes occurring "immediately before" and "immediately after" the outlying data point. In another interpretation, these two extreme points in Figure 4.11 indicate that the one of the data points (in the original 4.10 data set) is not as correlated with its neighbors as are other data points.

Unfortunately, typical surface metrology data sets are based on rather high sample rates which can generate many data points along the sides of an asperity. These higher sample rates incorporate a less significant change between data points and thus a higher degree of correlation.

The data set presented in Figure 4.1 is based on 0.1° sampling (i.e. 3600 data points). As a result, the asperity is not made up of a single point such as the example in Figure 4.10. Instead, the asperity consists of several data points as shown in Figure 4.12.

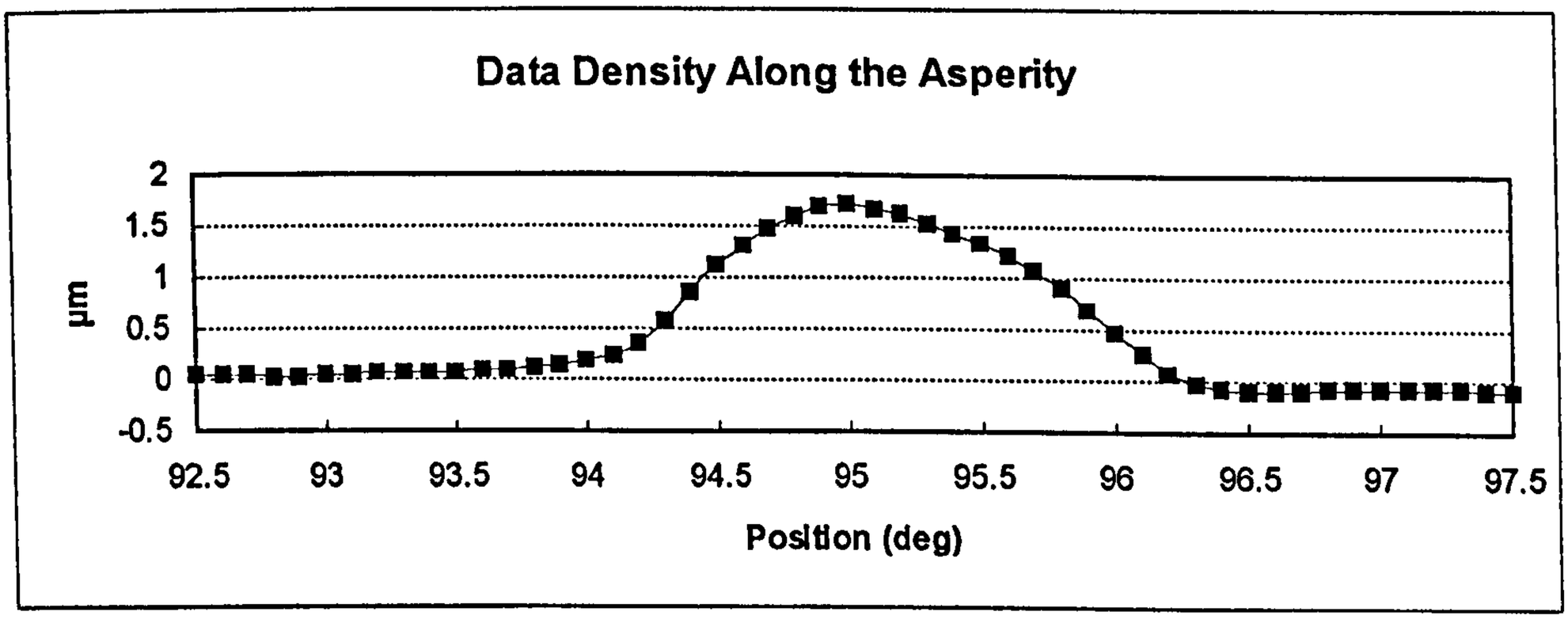


Figure 4.12 0.1° data point spacing along the Figure 4.1 asperity.

As a result of the high data density shown in Figure 4.12, the point-to-point height changes are less severe. This is indicated in Figure 4.13, where we see all of the height changes occurring on, or relatively near, the 45° line.

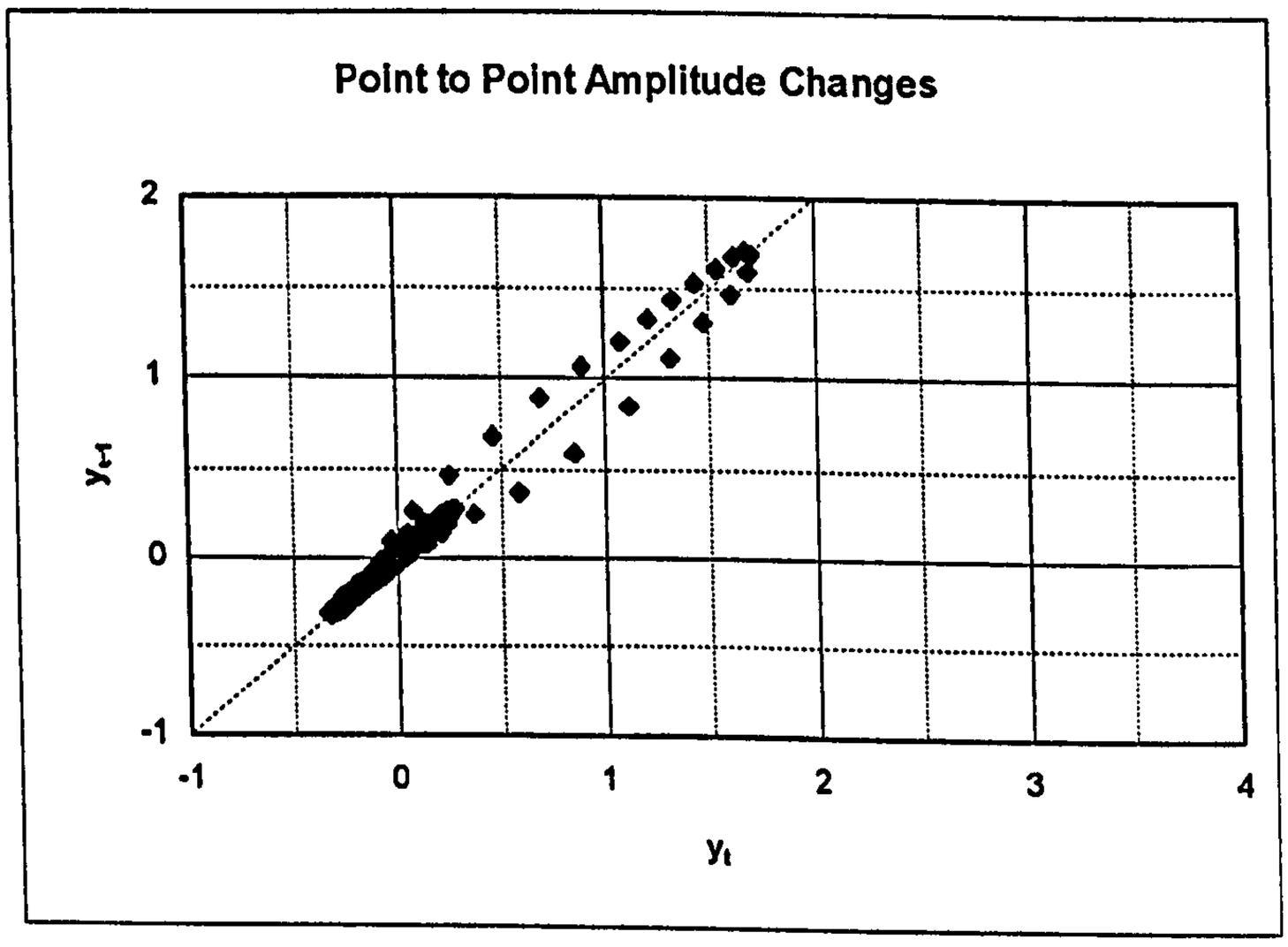


Figure 4.13 Neighbor based amplitude changes for Figure 4.1 data set.

Given the proximity of the amplitude changes to the 45° line in Figure 4.13, it becomes very difficult to ascertain which of these are related to the asperity. This is particularly difficult due to the fact that point-to-point height changes are relatively small near the base of the asperity and near the peak of the asperity. It may be noted that in the Figure 4.13 representation, the variations in data point density may imply an association with the asperity. This implication may be true in some cases, however general application of this premise could lead to erroneous conclusions depending on the local slopes of the surface.

4.3 A Robust Method for Asperity Removal

The above mentioned methodologies have certain advantages and disadvantages. In the following sections, a novel and more robust combination of these approaches will be presented. This proposed approach draws from the advantages of each of the previous methods and combines them into a robust methodology which also addresses the shortcomings of the previously described techniques (Table 4.2).

Current Method	**Advantages Utilized**	**Shortcomings Addressed by Robust Methodology**
Threshold based	Discarding of unreliable areas adjoining the asperity.	Detection of an asperity in a depression. Adaptive in light of process improvements.
Statistical Threshold	Adaptive in light of process improvements.	Detection of an asperity in a depression. Removes entire asperity and adjoining areas.
Neighbor Based	Local extremity model of asperity. Adaptive in light of process improvements.	Not dependent on sampling interval.

Table 4.2 *Robust methodology relative to other approaches.*

4.3.1 Asperity Model

In viewing an asperity as a local anomaly in the surface, it becomes logical to investigate the local extremes of the data set. This can be done in such a manner as to ignore the long wavelength form errors and look only at local changes in the data set. Out of the this view comes the following steps resulting in a robust (and adaptive) methodology for asperity removal:

1. Detection of local peaks and valleys.
2. Determination "extreme" height changes.
3. Location of edges of extremes.
4. Padding over removed section.

In this new approach, it is assumed that, at some measurable level, the surface exhibits a relatively high frequency such that there are many local peaks and valleys. The presence of an asperity results in a statistically larger, local peak to valley occurrence. This assumption is generally acceptable in application in that the stylus acts to mechanically *smooth* the sides of the asperity and we ensure that significant high frequency instrument noise does not manifest itself on the asperity.

Figures 4.14a and 4.14b illustrate this model in that the asperity exhibits significantly higher local peak to valley values than does the rest of the surface. Figure 4.14b is a representation of surface height differences between adjacent peaks and valleys. It is important to note that these height differences are not necessarily based on sampling spacing, but based on surface feature spacing. Thus, these are true "local height changes" between minima and maxima and should not be confused with the process of differentiation. Differentiation operates over a fixed interval or width, while the process utilized in the robust means for asperity removal does not.

This representation of local height changes between adjacent peaks and valleys is at the core of the robust approach. Using this analytical tool, the asperity becomes more evident as a well defined statistical anomaly.

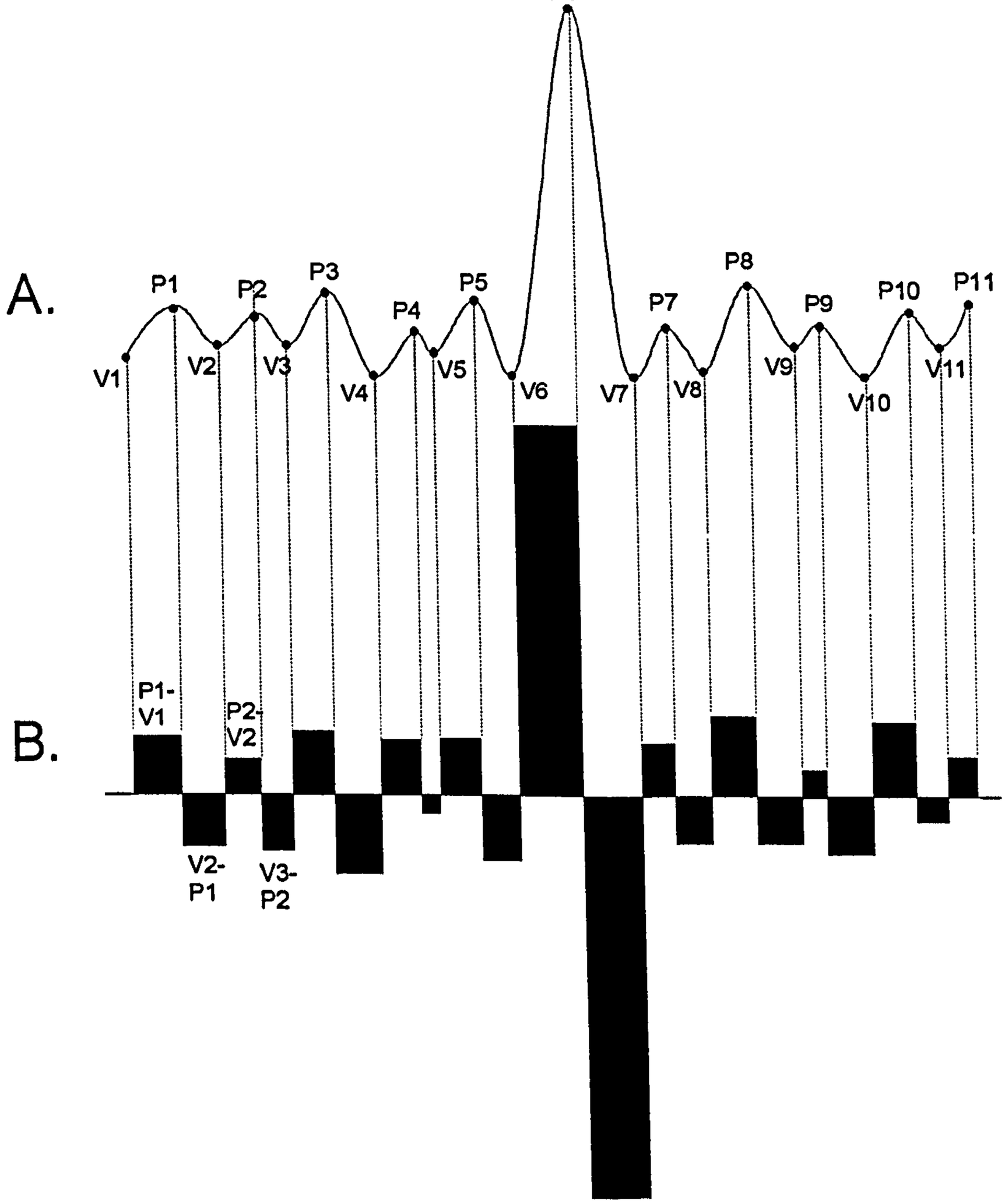


Figure 4.14 *The asperity model and associated local height changes.*

In this representation, the problems associated with excessive lobing (Figure 4.7) are removed as shown in Figure 4.15. This local height change representation exploits the significant changes associated with the asperity - even though the asperity's total height is mostly encompassed by the depression in which is lies. Thus, the model of an asperity is not sensitive to lobing and is adaptive to changes in the surface.

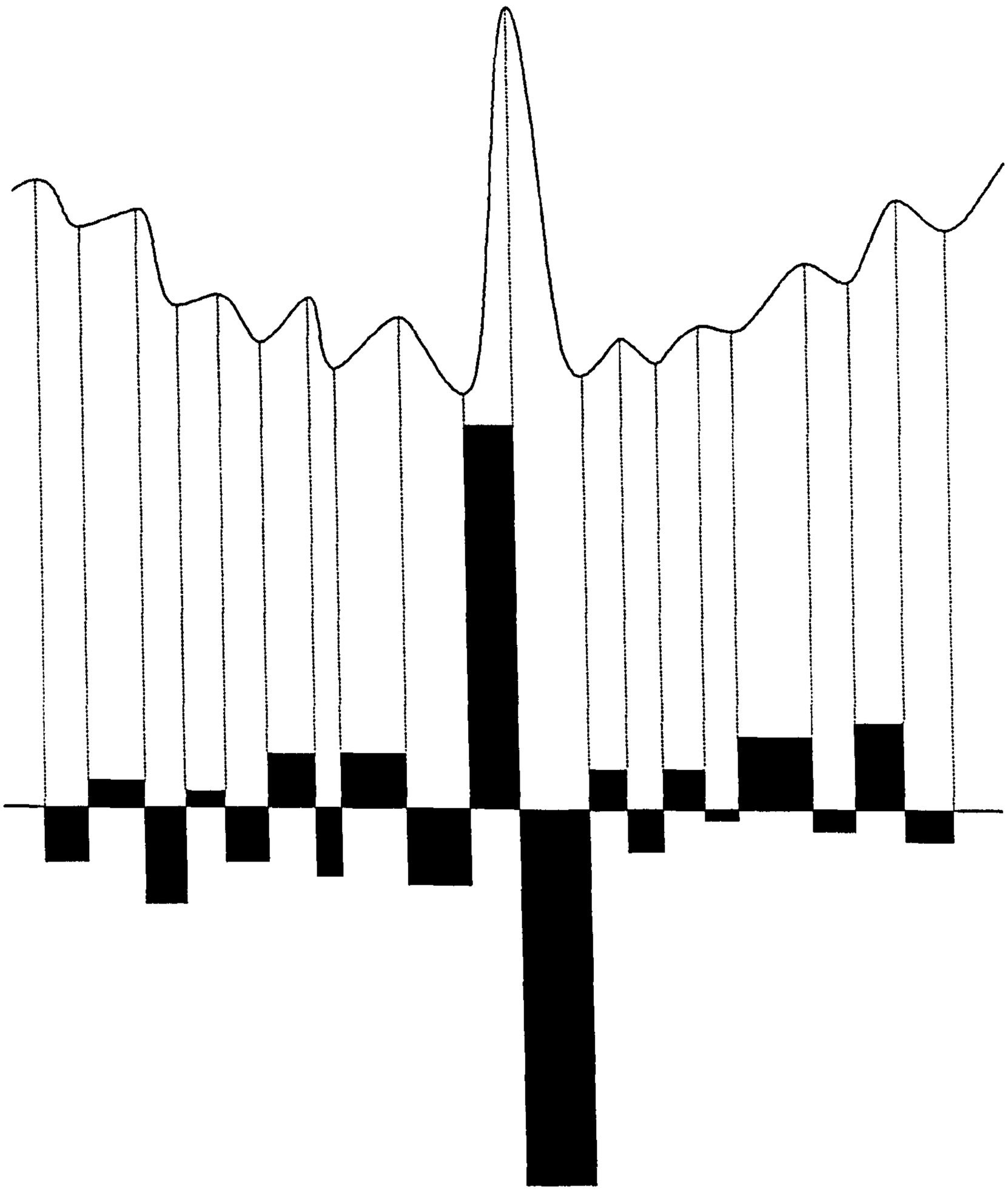


Figure 4.15 *Local height changes for an asperity falling in a depression.*

4.3.2 Detection of Local Peaks and Valleys

The detection of local peaks and valleys is purely a matter of determining local maxima and minima and is therefore independent of any thresholds or meanline crossings. It should be noted that, in some surface metrology peak detection algorithms, height thresholds must be crossed. For this approach, no such height threshold requirements are applied, thus allowing for asperity detection in local depressions of the surface.

A peak is defined as any local maxima in the data set such that at an individual profile height, z_i:

$$z_{i-1} < z_i > z_{i+1}$$

In the event of a flat peak (one with several points at exactly the same height) the centermost data point is treated as the local peak. Similar techniques are used for the determination of local valleys. Once these extreme points are determined, the height changes between adjacent maxima and minima are determined as graphically demonstrated in Figure 4.14.

It could be noted that this three-point definition of a peak (or valley) is used for the sake of simplicity and its generic application. More sophisticated methodologies may be employed which could minimize errors such as instrument noise which may be present in the profile. One of the advantages for this new means of asperity detection and removal is that its underlying principle is independent of the means used for extreme feature detection.

4.3.3 Determination of Extreme Height Changes

The local height changes are accumulated into a histogram. In generating the histogram all of the profile data points are used. The value associated with each data point is the local height change either completely encompassing the data point or

immediately following the data point. This concept is graphically represented in Figure 4.16 and the corresponding data can be summarized as follows:

Local Peaks:- z_3, z_7

Local Valleys: z_1, z_4, z_{11}

Height Changes:

$Z_1 \rightarrow Z_3$	= +2	(plotted at points 1, 2)
$Z_3 \rightarrow Z_4$	= -1	(plotted at point 3)
$Z_4 \rightarrow Z_7$	= +3	(plotted at points 4, 5, 6)
$Z_7 \rightarrow Z_{11}$	= -4	(plotted at points 7, 8, 9, 10)

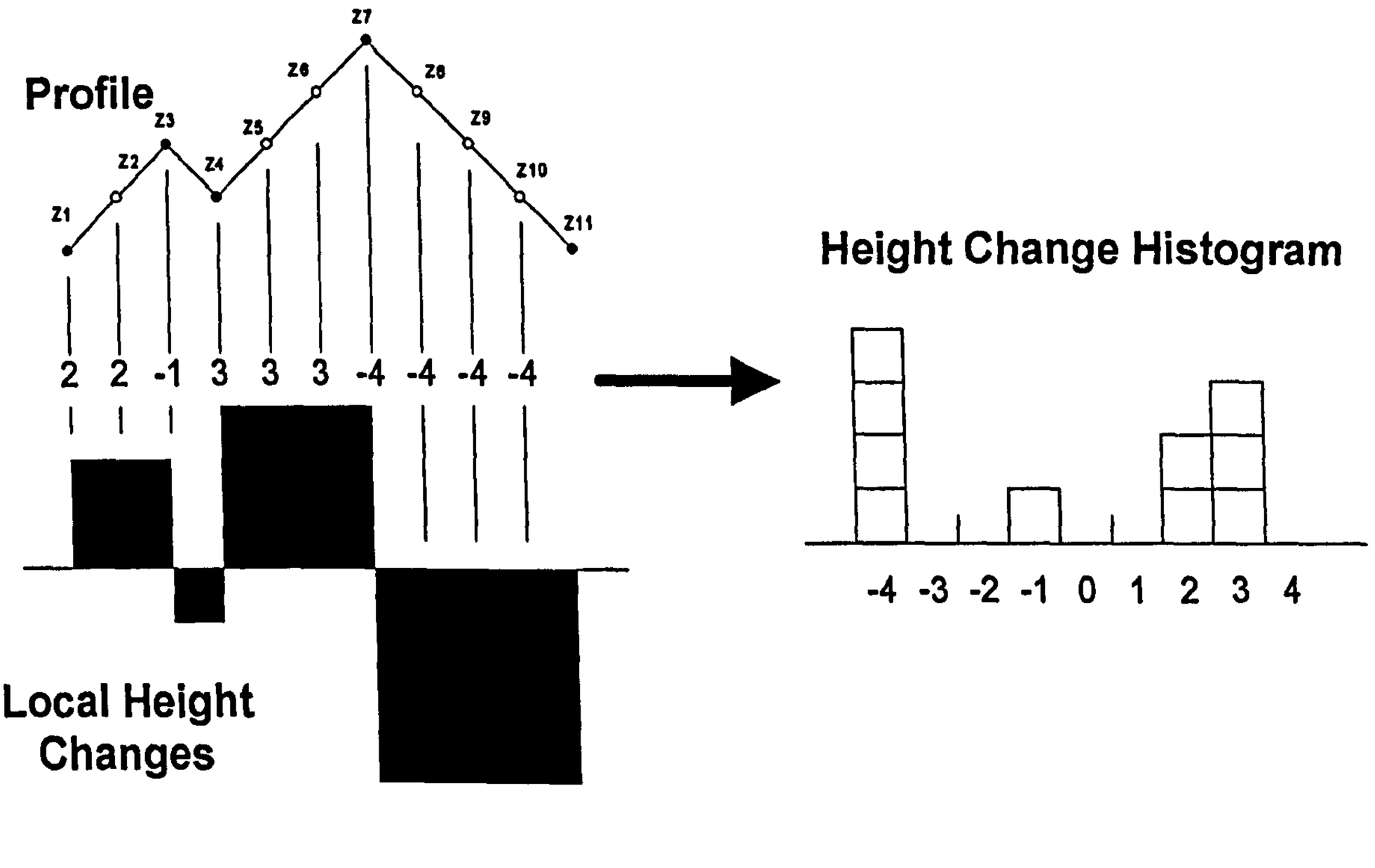


Figure 4.16 Histogram development for local height changes.

Upon compilation of the histogram, a standard deviation calculation is performed and all local height changes exceeding some statistical limit (e.g. 4 standard deviations) are

deemed to be associated with asperities. This standard deviation-based approach is conceptually similar to that presented by Paterson (1985).

4.3.4 Location of Edges of Extremes

An asperity can be readily detected per the above steps, but this is only the "identification" portion of the process. The asperity must now be removed in a reliable manner, while maintaining the highest possible level of integrity in the remaining data set.

The local extremes provide the boundaries between local height changes; therefore the valleys bounding an asperity (Figure 4.17) are easily detected. However, these valleys may be unstable as they may be the result of instrument dynamics related to the asperity. Furthermore, if these valleys are chosen as the edges of the extreme, any subsequent padding will generate points that are biased toward the valley side of the data set.

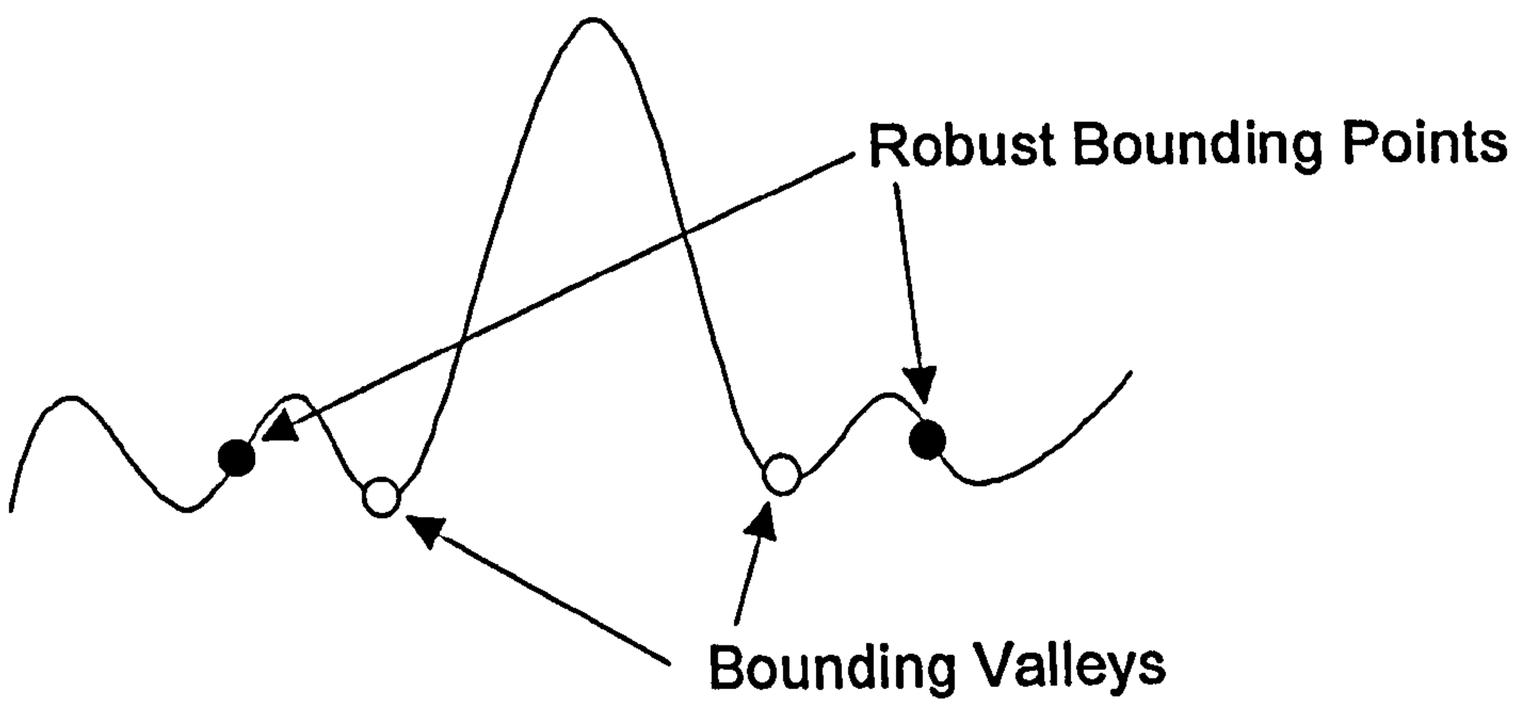


Figure 4.17 The bounding points of an asperity.

The "robust" bounding points (as shown in Figure 4.17) are taken as the midpoints of the second height changes from the asperity. This ensures some safe distance from any instrument dynamic effects while maintaining relatively close proximity to the asperity. Furthermore, since these points are midpoints, they are more stable in a numerical sense and will introduce only minimal bias into the subsequent padding operation. It

should be noted that, since these are midpoints, the robust bounding points may not fall directly on a data point.

4.3.5 Padding Through the Removed Section

Padding or "re-constructing" the data set through the removed area is necessary to give a continuous data set, thereby simplifying further processing (e.g. filtering, Fourier analysis, rate of change analysis, etc.) Given the robust bounding points, the padding operation is a simple linear interpolation between these points. (Note: for circular data sets represented in polar coordinates, linear interpolation results in a curve.) All data points occurring between the robust bounding points are translated to the line joining the robust bounding points as shown in Figure 4.18. The linear interpolation approach will not introduce artifacts which could be detected in a peak to valley assessment of the profile. Furthermore, the effect on averaging or RMS parameters will be small since the line falls in the region of high probability for profile amplitudes. In addition, the corners encountered at the pad-to-data transition are relatively insignificant and occur over a short enough arc that a typical "local slope" of "rate of change" assessment will not detect them. Similarly, arguments could be made for the minimal effect on spectral analysis given this padding methodology, whereby there may only be some very small shift from high to low frequencies based on the smooth nature of the linear segment. In practice, the asperity removal and padding should be implemented prior to filtering, thus providing additional smoothing.

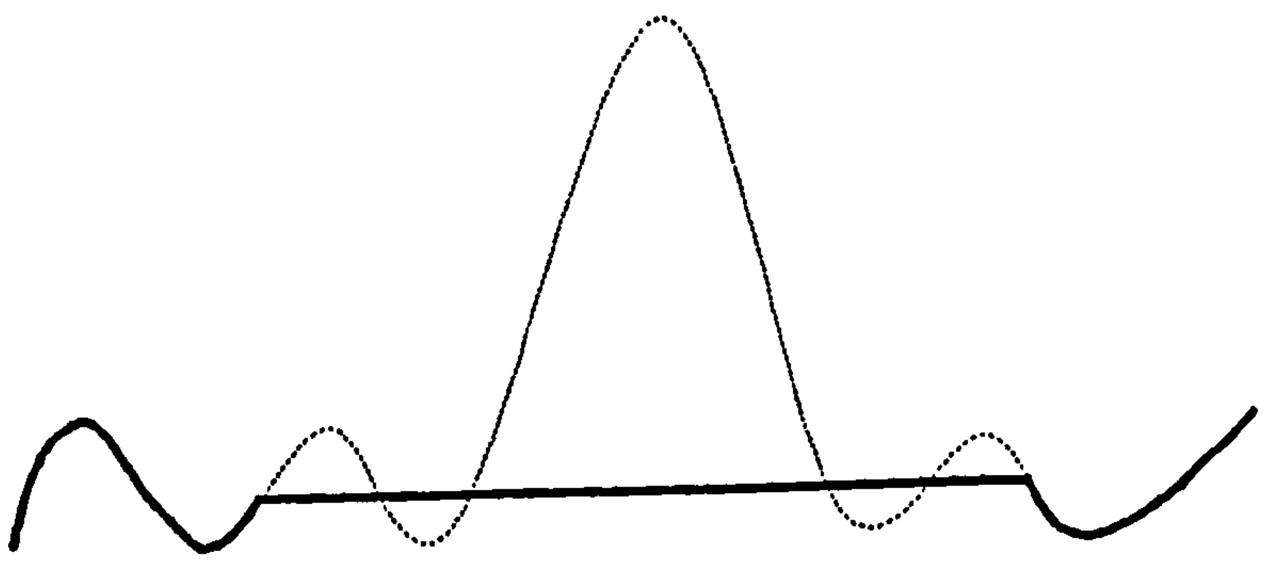


Figure 4.18 *Padding through the robust bounding points.*

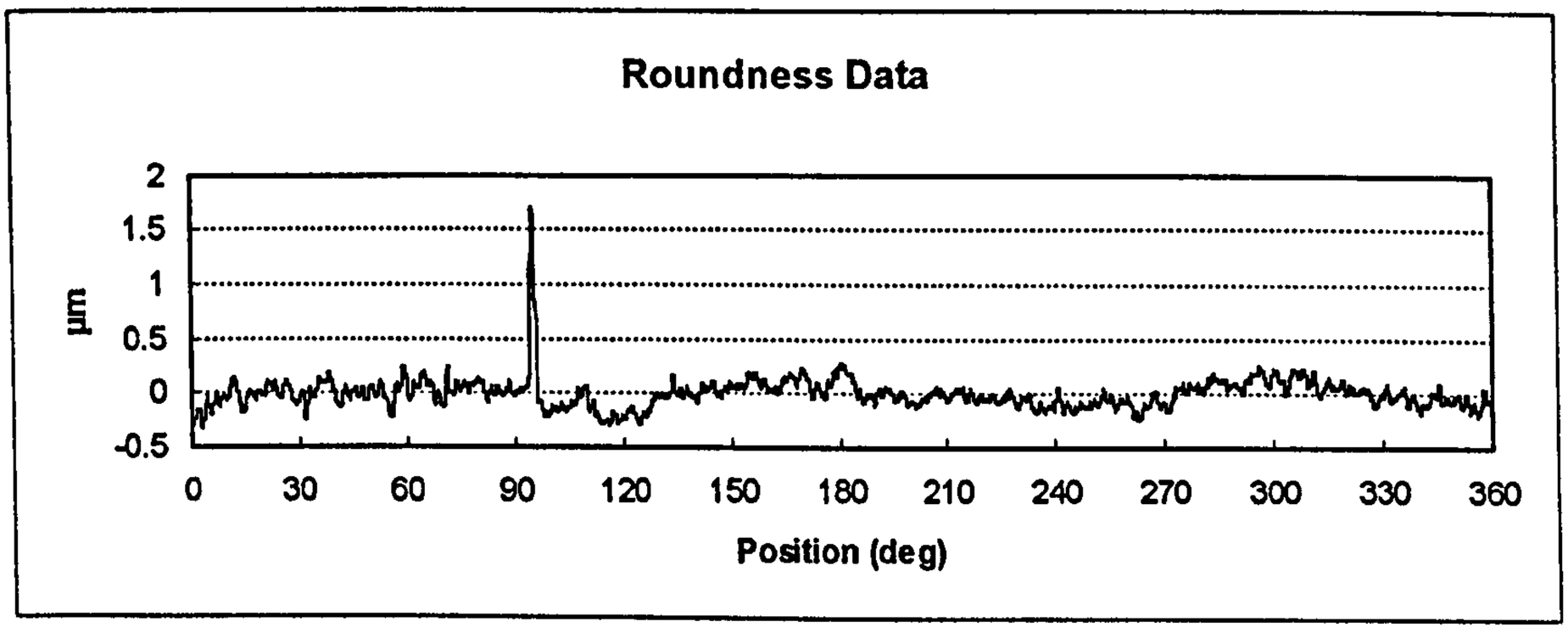
4.3.6 Application to Case Study Data

The above methodology provides a robust means of detecting and removing asperities from surface metrology data. The application of the approach can be demonstrated on the case study data of Figure 4.1.

1. Detection of local peaks and valleys

Figure 4.19a redisplays the 3600 data points from Figure 4.1 in a linear representation. This representation is useful in comparing the profile features to their corresponding height changes (Figure 4.19b).

A.



B.

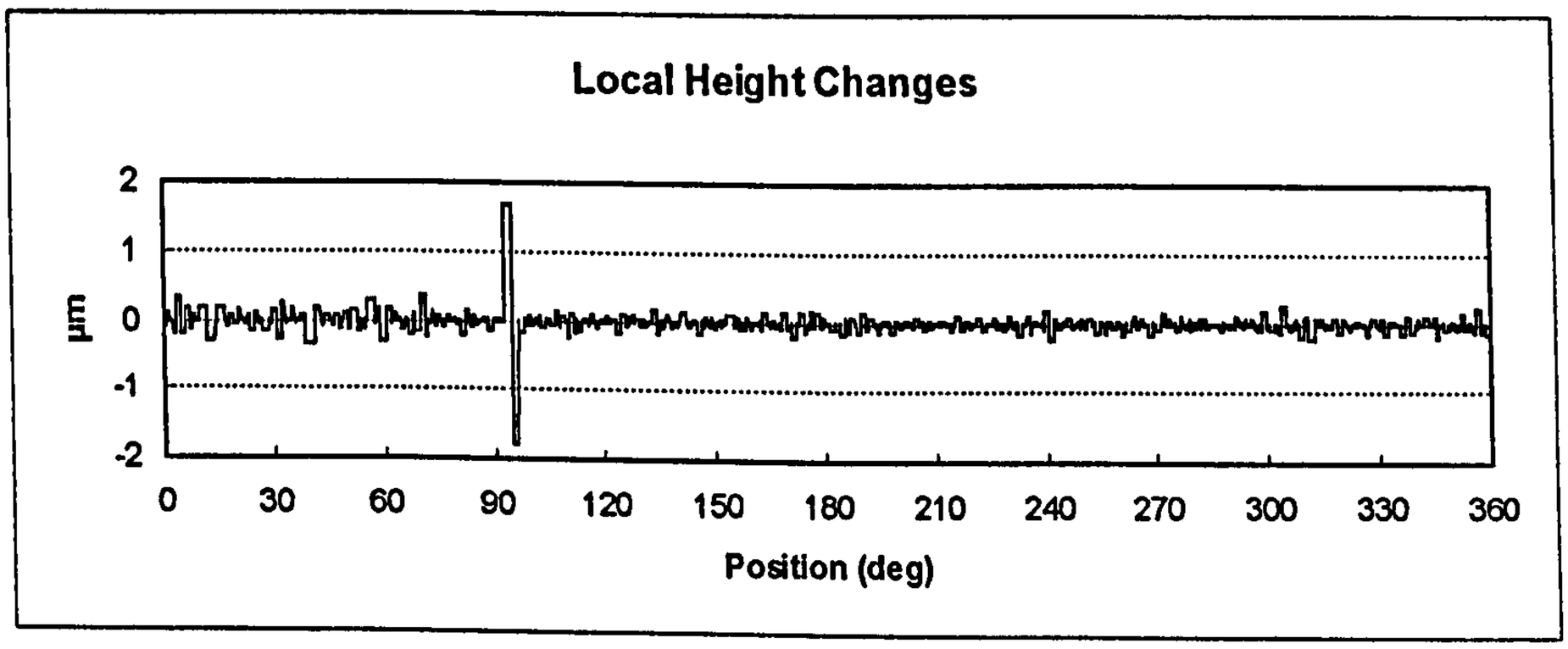


Figure 4.19 *Fuel injector data (from Figure 4.1) and associated local height changes.*

2. *Determination of Extreme Height Changes*

The histogram resulting from the Figure 4.19b local height changes is shown in Figure 4.20 with a superimposed Gaussian (normal) distribution (based on the calculated values for mean and standard deviation). The standard deviation of local height changes for this data set is 0.21 μm. Therefore, with a 4 sigma limit we treat all height changes exceeding 0.84 μm as being associated with an asperity. (It should be noted that none of the height changes associated with the real surface are beyond 2 standard deviations.)

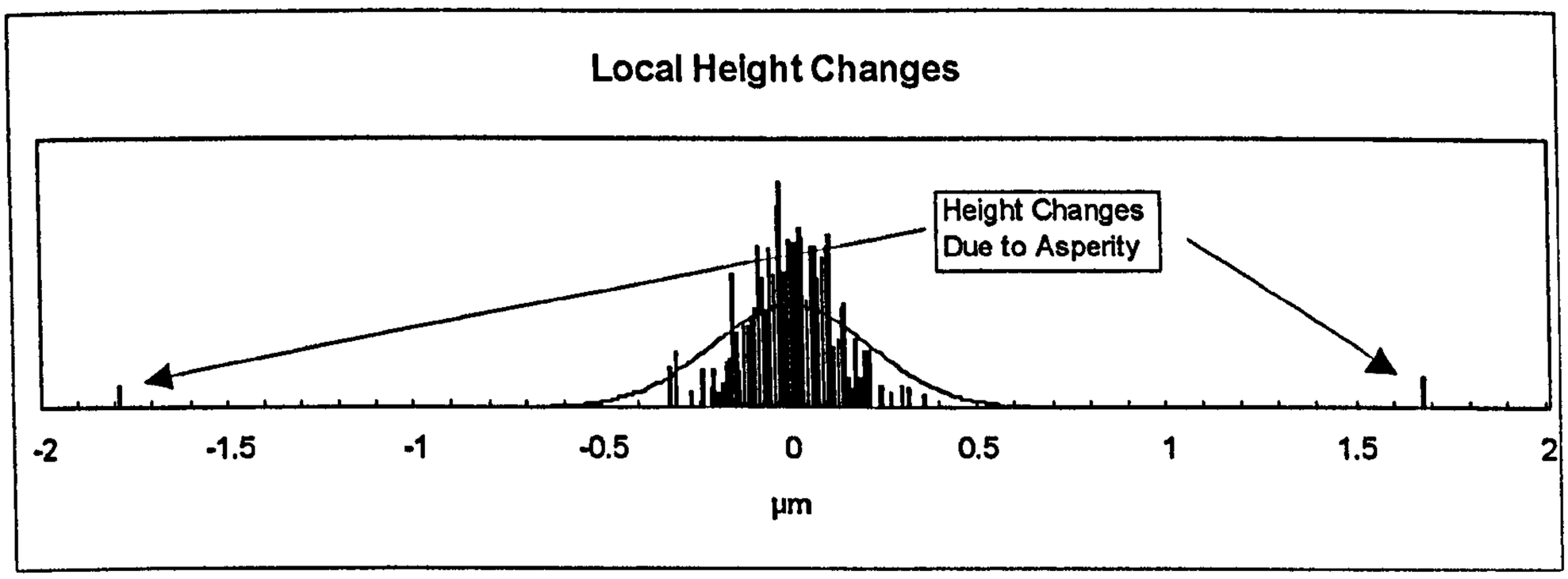


Figure 4.20 *Local height change histogram for Figure 4.1 data.*

3. *Translation to Raw Data and Padding*

Robust bounds for the asperity were determined and the data set was padded to re-establish continuity as shown in Figure 4.21. This modified data set was introduced to the roundness analysis software and the results are given in Figure 4.22.

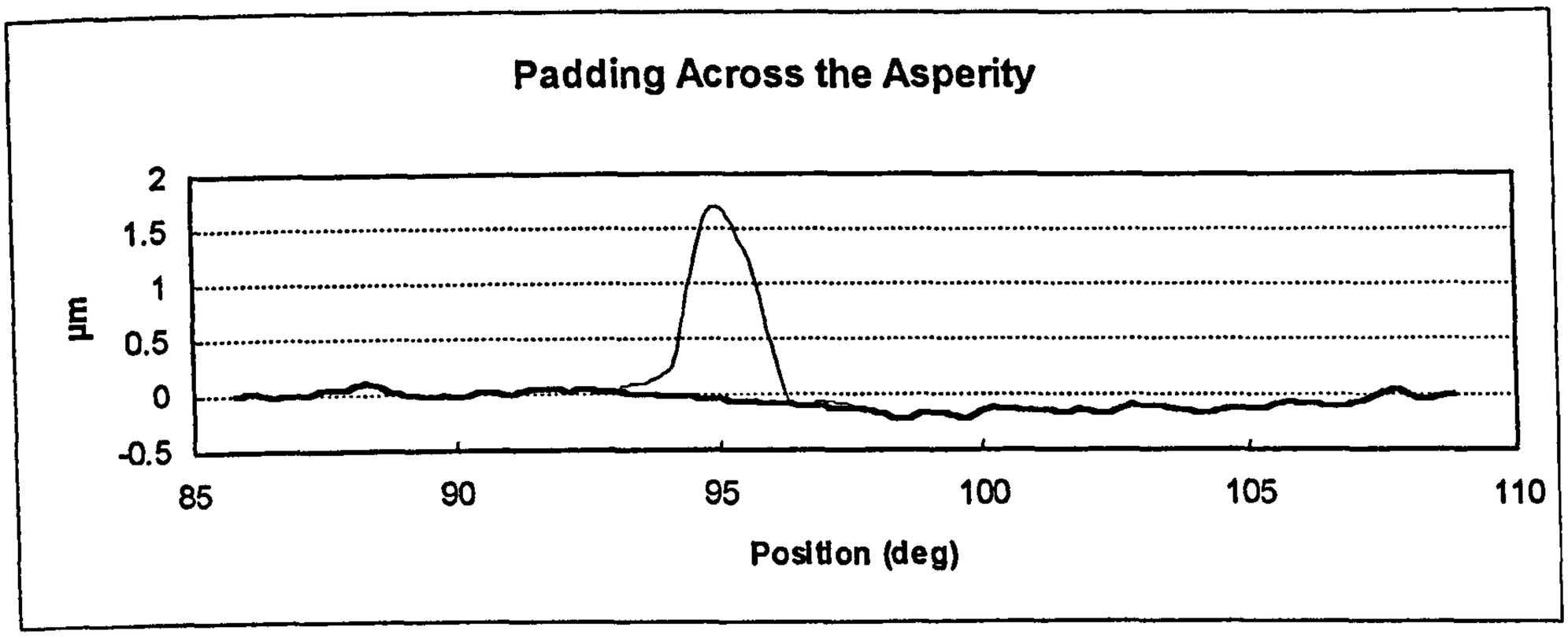


Figure 4.21 *The asperity removal and padding.*

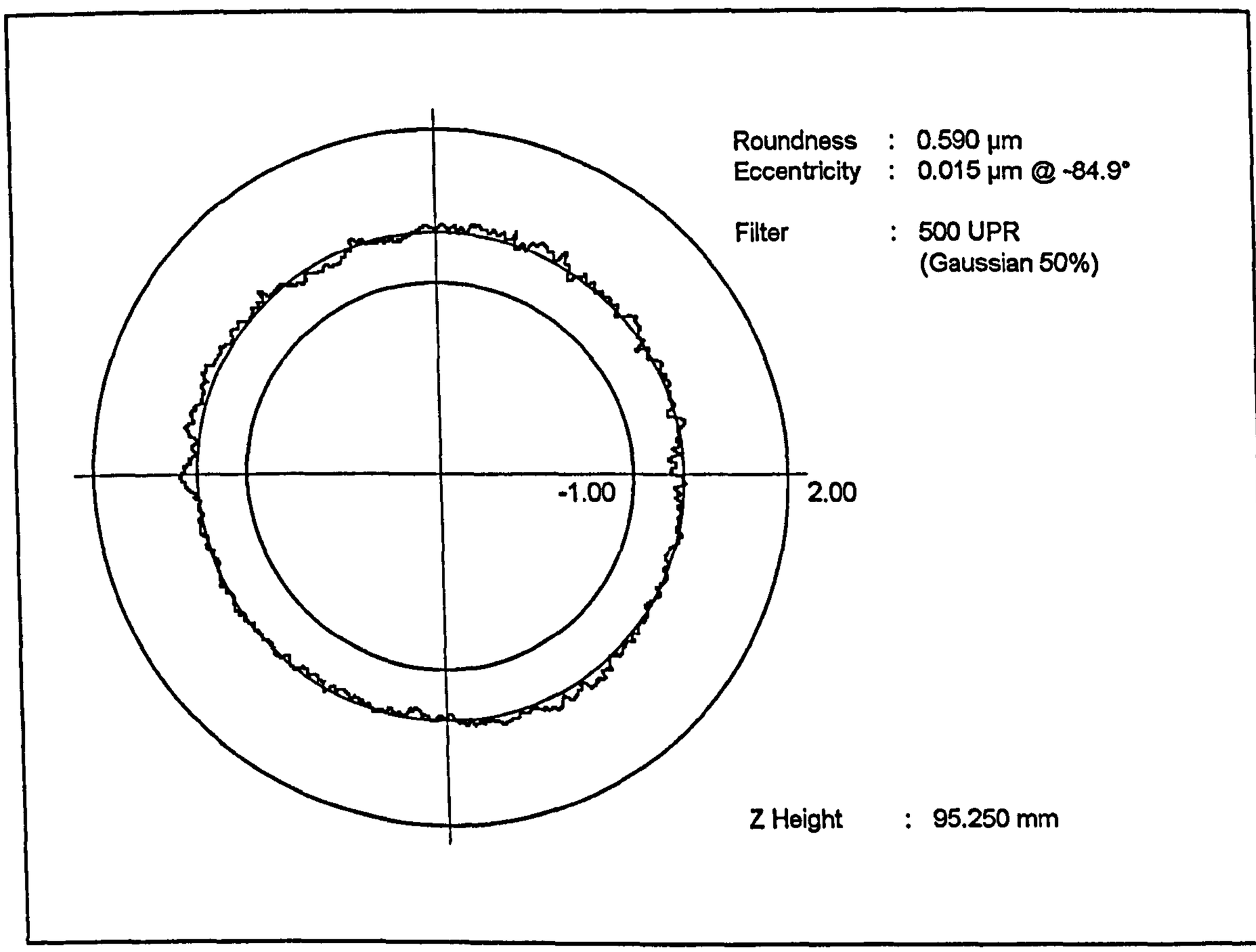


Figure 4.22 *The Figure 4.1 data set after the removal of the asperity.*

The eccentricity value of 0.015 µm at -84.9° is relative to the center coordinates obtained from the Figure 4.1 analysis of the data set. Thus, by removing the unwanted

asperity, the least squares center shifted 15 nm away from the direction of the original asperity.

4.3.7 Other Applications

This technique for the removal of unwanted asperities can be applied to many other surface metrology analyses. Figures 4.23, 4.24 and 4.25 present another example of the technique on fuel injector bore data collected using a Rank Taylor Hobson Talyrond 250 (2000 data points, 1 mm tip radius). (Note, for inside diameter measurements unwanted asperities are inward and thus represent negative amplitudes in terms of polar coordinates.)

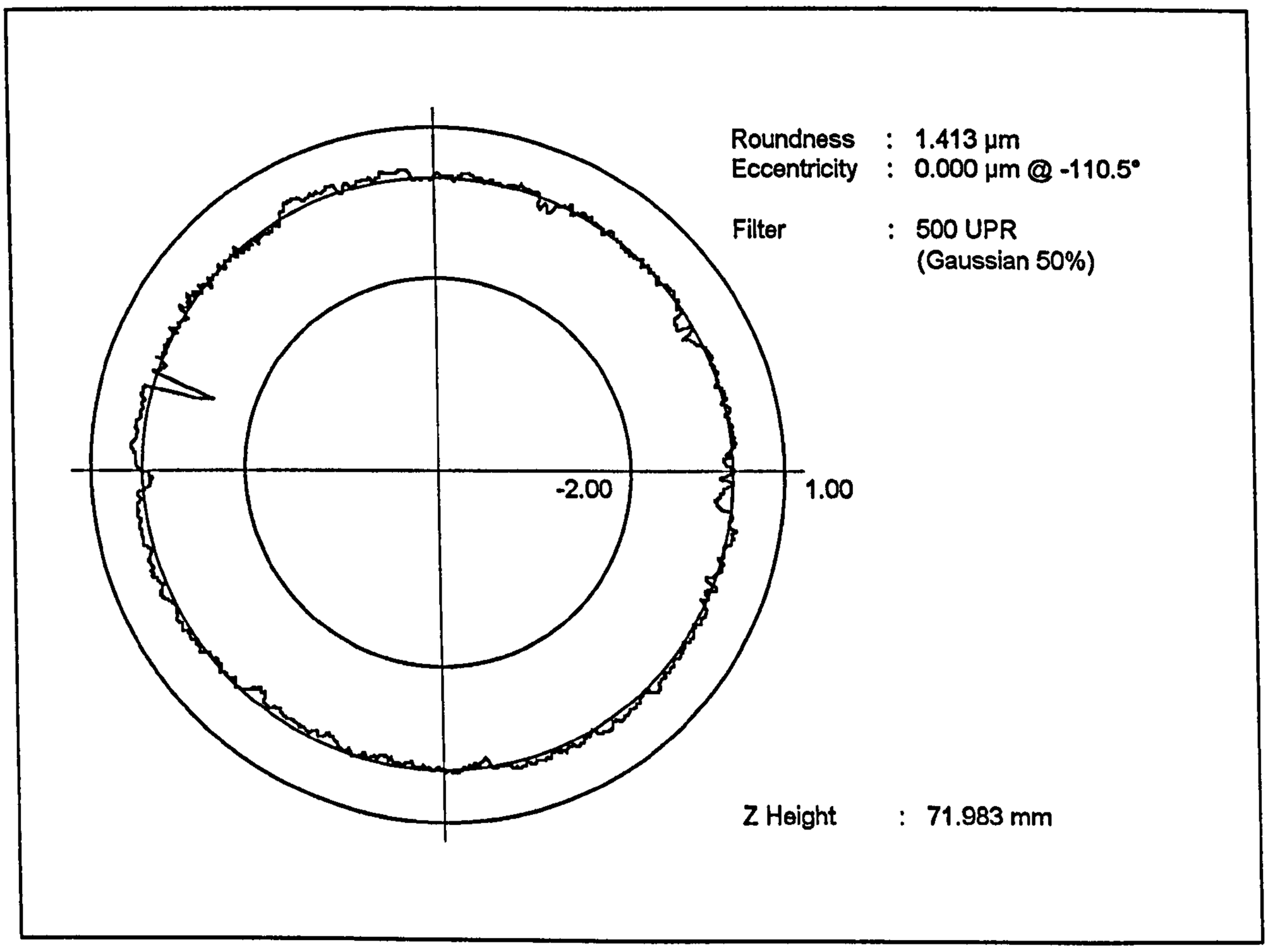
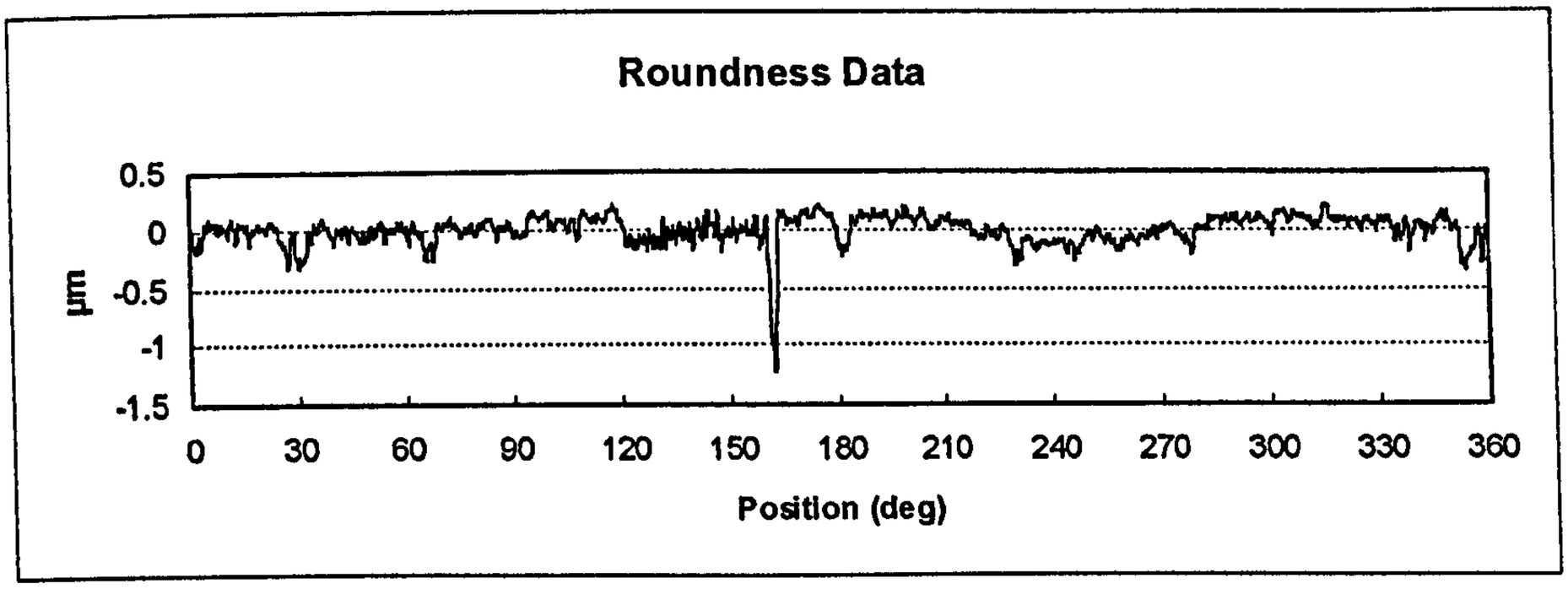


Figure 4.23 *Fuel injector bore data with an unwanted asperity.*

A.



B.

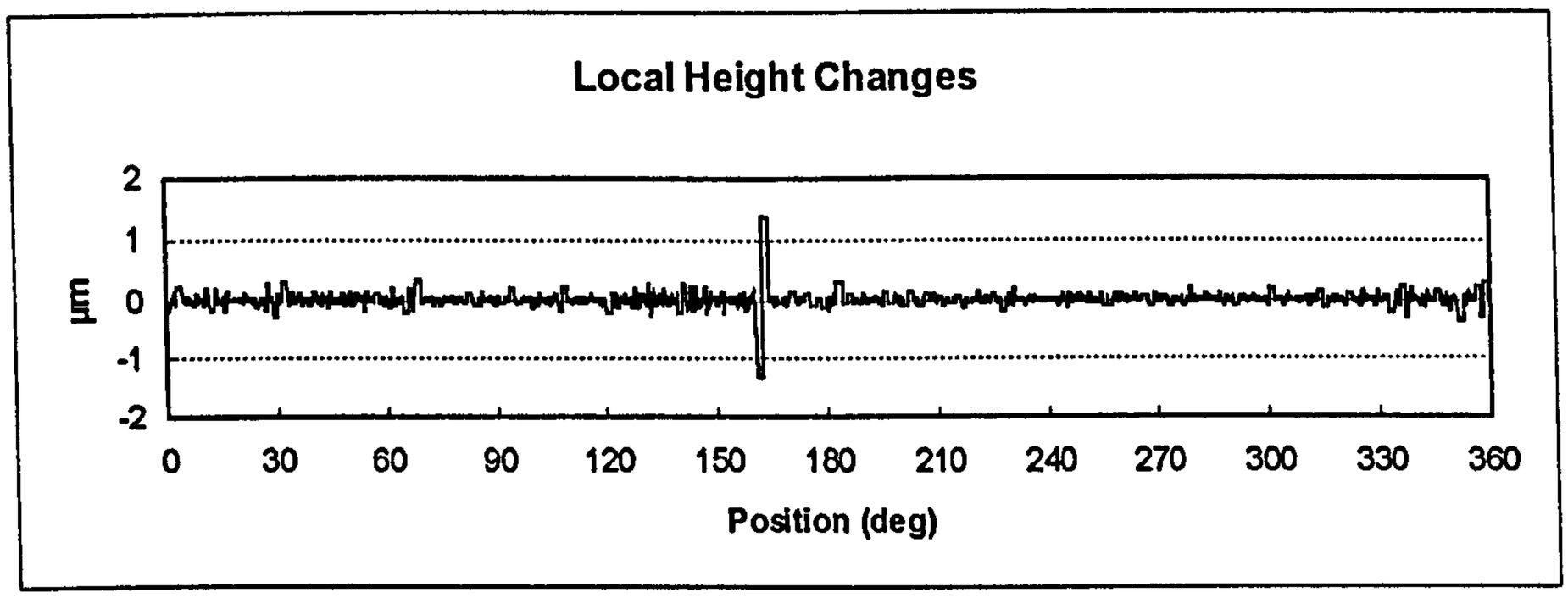


Figure 4.24 *Local height change analysis for the Figure 4.22 injector bore data.*

A. Linear representation of profile heights.

B. Local height changes.

For the injector bore data, the standard deviation of the local height changes was 0.175 µm, resulting in a four sigma threshold of 0.7 µm. The local height changes associated with the asperity were approximately 1.3 µm and therefore the asperity was identified for removal.

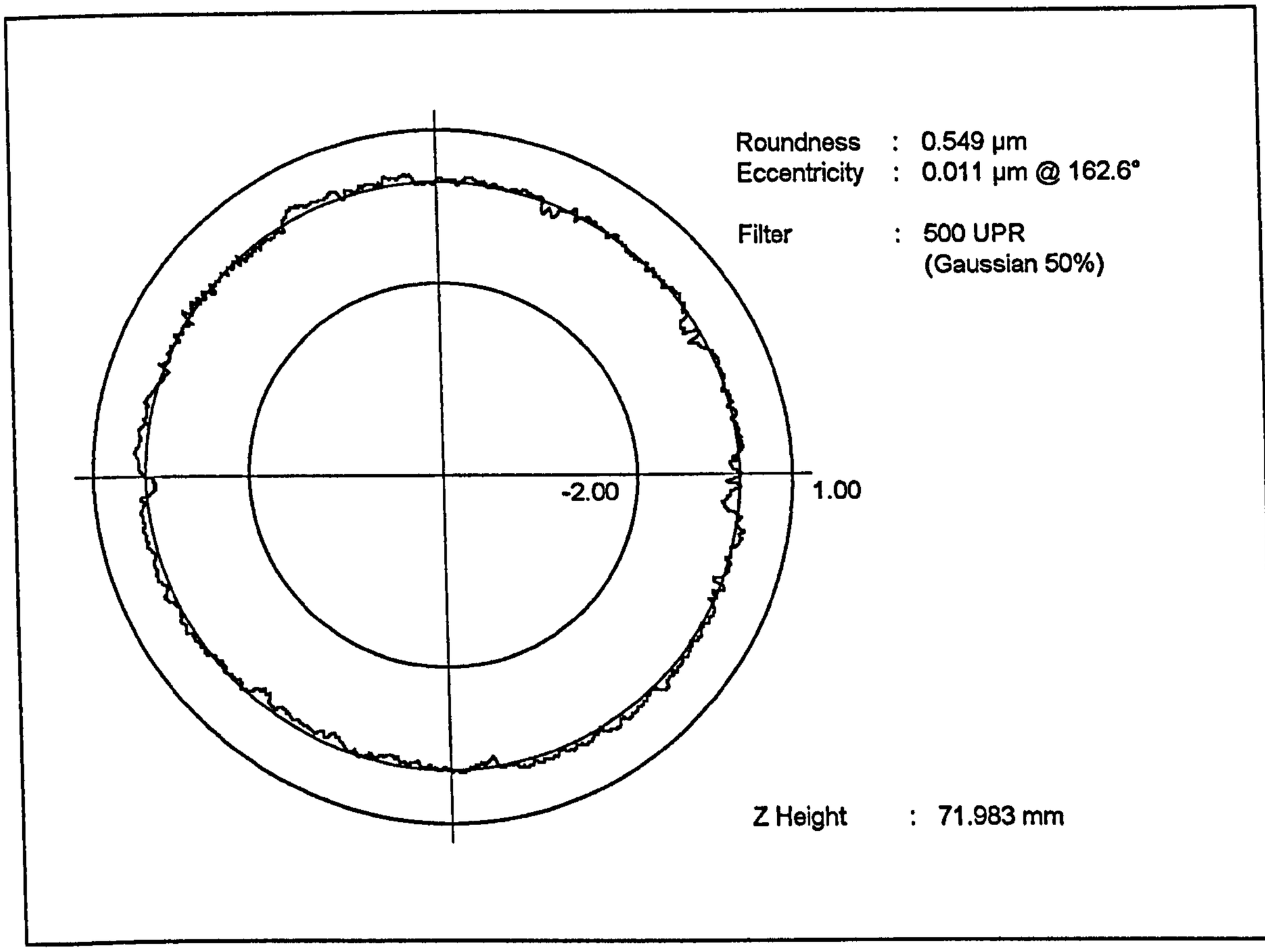


Figure 4.25 *Injector bore data set with asperity removed.*

It should be noted that the above examples could be handled by the threshold based approach with similar results. However, the proper height, width and discard values must be analytically or experimentally determined and then manually implemented. Furthermore, these settings would be different for each of the examples and must be continually monitored based on manufacturing process changes. The robust approach can automatically adapt to varying base surface features and asperity geometries. Furthermore, the statistical threshold approach, while being adaptive would not provide a acceptable means for *padding* across the discarded region.

Finally, it is useful to return to the Chapter 1 CMM data set (Figure 1.6 and also Figure 4.26 below).

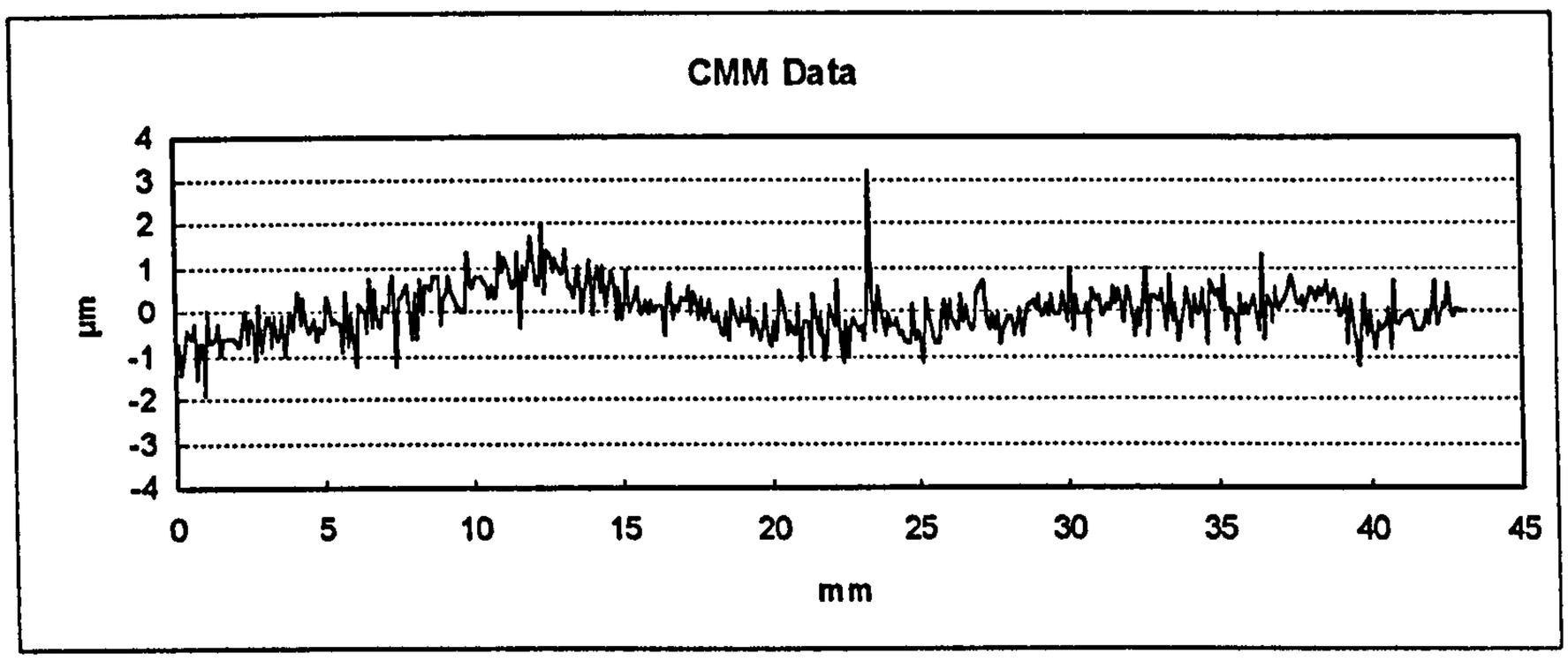


Figure 4.26 *Zeiss CMM straightness data from Chapter 1.*

This data set is interesting in the context of unification as it represents a different type of instrumentation (coordinate measuring machine as opposed to cylindricity machine) and a linear rather than polar assessment. Nonetheless, there is a great deal of similarity in the resulting data sets and the effects of an unwanted asperity.

It is important to note that the "statistical threshold" (Paterson) approach does not detect the asperity in this data set. (The extreme peak height is only 1.8 times the standard deviation due to its position in a local depression.) However, the peak is still high enough to significantly affect the total (peak-to-valley) height (see Chapter 1, Figure 1.7).

Returning to the robust methodology, the local height changes were generated and are plotted below in Figure 4.27.

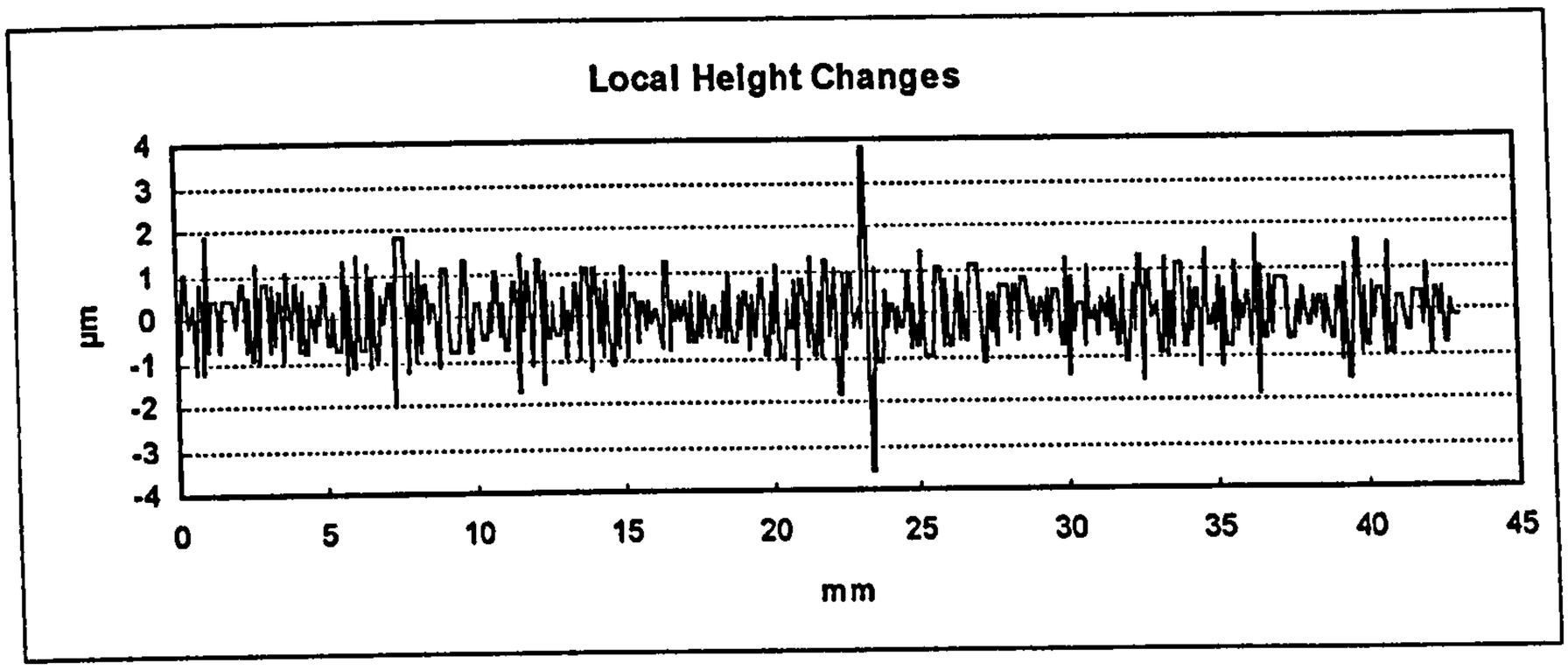


Figure 4.27 *Local height changes for CMM straightness data.*

The standard deviation of local height changes for the Figure 4.26 data set is 0.875 µm, thus the local height changes associated with the unwanted asperity (approximately 3.7 µm) exceed the four standard deviation limit. Thus, the asperity is readily detected and the data set can be subsequently padded in the area of removal. The resulting profile is presented in Figure 4.28.

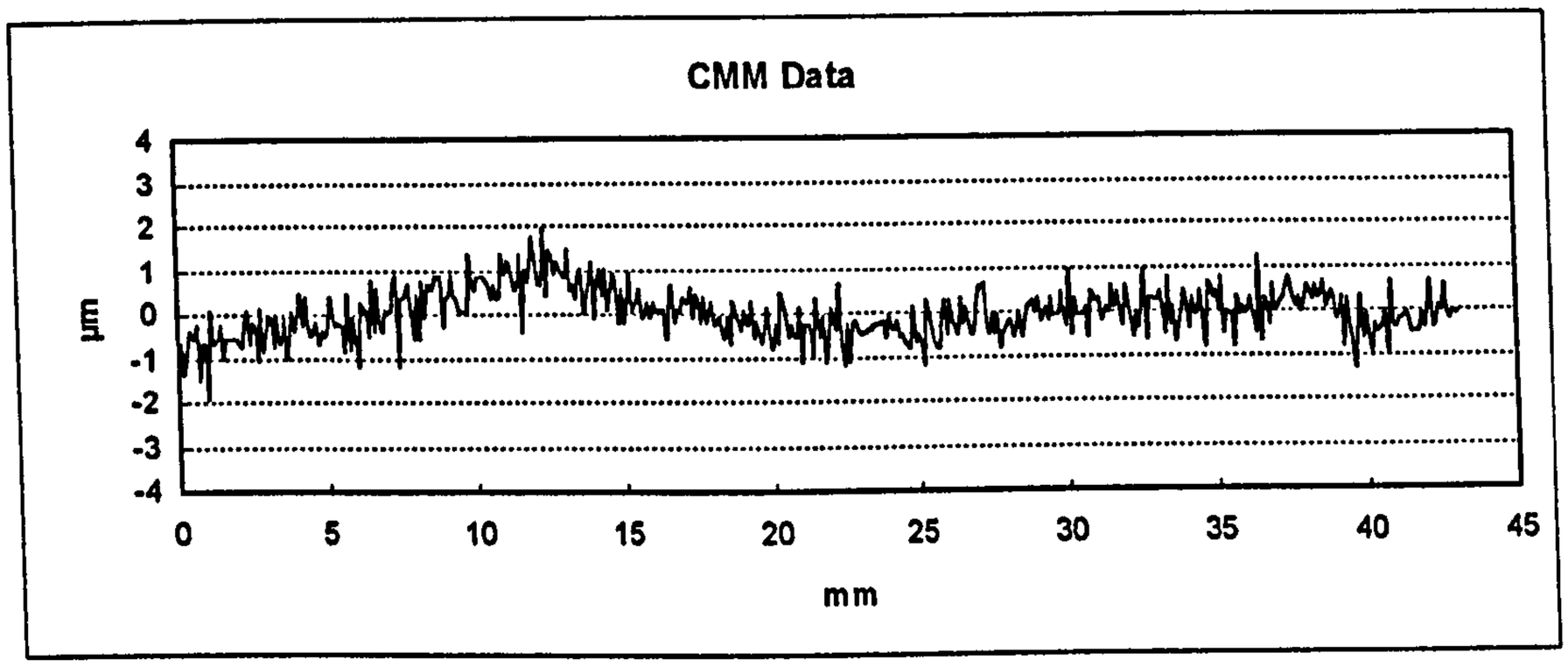


Figure 4.28 *CMM straightness data after asperity removal and padding.*

4.3.8 General Application Considerations

This robust methodology has been developed and presented as a general means for the removal of unwanted asperities from two dimensional data sets such as those typically

encountered in surface metrology. In these applications, the following requirements are commonly met:

1. The sides of an asperity contain no local peaks or valleys.

Surface metrology applications, where asperities have the biggest influence, are those involving the assessment of errors of form. These applications typically utilize instrumentation with stylus tip radi significantly larger than the asperities which are to be removed. As a relatively large tip is convolved over an asperity, the resulting data becomes *smooth* based on the stylus geometry (see Chapter 3). Instrument noise is of concern relative to this robust approach in that local peaks or valleys on the sides of an unwanted asperity can distort the local height change distribution in the neighborhood of the asperity. However, point to point height changes due to an asperity are typically greater than the peak to valley height changes associated with instrument noise. Therefore, the sides of the asperity will tend to *mask* the instrument noise. An obvious approach for dealing with this potential is the application of a filter, however other methodologies such as motif combination (Scott 1992b) may also prove feasible. In any case, the numerical means must remain adaptive and non-dimensional.

2. Local peaks and valleys must be present in the "base" data set.

There must be adequate changes in the nominal surface for the determination of the standard deviation. Furthermore, these local peaks and valleys become the basis for padding in the vicinity of a detected asperity. To ensure that there are local peaks and valleys present in the data set, it is recommended that asperity removal be performed on unfiltered or minimally filtered data. Once again, practical experience indicates that typical data sets exhibit a very high number of local peaks and valleys.

4.4 Conclusions

The application of surface metrology is often hindered by environmental concerns such as cleanliness. Therefore, the removal of "unwanted asperities" is often essential in

obtaining correlation between various instruments and environments and thus is an important aspect of a unified methodology for applied surface metrology.

Historically, to accommodate these "less than ideal" environments, data pre-processing techniques have been developed whereby unwanted asperities can be removed from the data set. However, these techniques entail certain assumptions and pose certain limitations.

In light of the shortcomings of these previously applied approaches, a robust technique for the identification and removal of asperities from surface metrology data has been developed and presented above. This technique is adaptive in both the height and width aspects of asperity detection and removal. Incorporated in this approach is a scheme for data restoration or *padding* whereby any bias on the padded points tends to be minimized. The methodology has been applied to the measurement of many critical surfaces including fuel injector component roundness, cylinder bore geometry and various straightness analyses and results appear to be very good. The generic nature of this scheme allows for its general application to a variety of signal processing areas.

The topic of asperity removal should be left *open* as more adaptive means are developed. Advances in Wavelet Transformations (Chen et al. 1994) and Artificial Neural Networks (Vemuri 1988) are potential candidates for application in more "intelligent" or "robust" techniques for removing unwanted profile features from surface metrology data sets. Nonetheless, the proposed robust method is advocated for its adaptive nature and effectiveness as compared with previously published methods and for is computational simplicity.

The robust asperity removal technique may be very beneficial (economically) relative to "first level" measurements performed in or near the manufacturing arena. However, despite the economic and technical merits of the method developed above, the fundamental issue regarding the removal of data points from a measurement remains

philosophical. For ultimate measurement data integrity, the surface being measured should be clean and no allowance should be made for data removal.

In the context of the scheme for unification, the removal of unwanted asperities will remain as an important supporting technology and will not be directly accommodated in proposed specification table. Thus, as a default condition, the removal of unwanted asperities (by any technique) is forbidden. Only when economically (or otherwise) justified should an asperity removal methodology be applied and then its application must be specifically noted per an agreed upon format. In these applications, where the removal of unwanted asperities is justified, the above described "robust" approach is strongly recommended.

A Unified Methodology for the

Application of Surface Metrology:

Chapter 5

Filtering in Surface Metrology

The proposed approach for the unification of surface metrology is based primarily on the understanding and control of the wavelength content of data sets. As discussed in Chapter 1, the function (in application) and control (in manufacture) of surfaces can be related to various wavelength regimes depending on the specific surface attributes of interest. Filtering is therefore essential in isolating these regimes for subsequent numerical characterization (or *parameterization*).

In Chapter 3, the stylus tip was explored in light of its wavelength transmission capabilities. Although the convolution of a stylus tip may exploit some functional aspects of the surface, it is quite uncommon in practice as a means of separating wavelength regimes. In the vast majority of today's applications, digital filtering techniques are employed.

In applying these digital approaches, the integrity of the data set must be ensured prior to the digital filter. This includes the consideration of stylus tip influences (Chapter 3), the presence of unwanted asperities (Chapter 4) and other errors inherent to the instrument (Chapter 7). These influences enter the data set prior to the digital filter and can alter the measured wavelength content. If these influences are inside the filter's pass band, they will remain in the resulting, filtered data set. On the other hand, if these influences (when present) can be moved outside the filter pass band the resulting data will exhibit a higher degree of integrity. As an example, when employing digital filtering, a stylus tip geometry (radius) should be selected so that the predominant stylus influences will be outside the filter transmission. Chapter 3 demonstrated that this is often difficult to achieve completely, however the techniques that were presented can be applied in the determination of a suitable tip radius.

Given the important role of digital filtering in the scheme for unification, it becomes necessary to review this topic for the sake of completeness. This chapter will provide a brief review of historical methodologies in light of their shortcomings followed by a presentation of the methodologies to be included in the scheme for unification.

As presented in Chapter 2, the unified scheme for surface metrology incorporates two digital filtering techniques: Gaussian filtering, and *Ideal* wavelength limitation. Given that these techniques are reasonably established, much of the technical information in this chapter regarding their characteristics will be based on published work. However, this chapter will provide important new information regarding the behavior of these filters and, perhaps more importantly, the implementation of these filters. The latter is of particular significance in light of experience with commercial instrumentation whereby filtering algorithms (which are often optimized for the sake of speed) can lead to erroneous transmission characteristics (DIN 4776-1990, Wilde 1994).

Finally, in this context of filtering, the topic of data set *padding* will be discussed in light of long wavelength profile analyses and the "end effects" associated with filtering techniques.

5.1 The Evolution of Filters in Surface Metrology

The separation of roughness and waviness in surface metrology analyses began coincidentally with the introduction of instrumentation for analyzing surface profiles. Originally, manual methods were developed to for the separation of wavelength regimes - particularly in the analysis of roughness profiles. Since then, analog and more recently digital methods have been developed for improved separation between surface wavelength regimes in all areas of surface metrology.

5.1.1 Linear Segments

In the early, optics based roughness instrumentation developed by Schmaltz (Whitehouse 1990), relatively short surface profiles were measured and subsequently leveled with a linear reference *following the general direction* of the surface. Given the relatively short profile length (approximately 0.8 mm or 0.030 in.), this leveling process had the effect of removing waviness from the profile for many typical surfaces.

Thus, the need for separating surface wavelengths has been inherent to the measurement process throughout the history of surface metrology.

As the length of profile traces increased this linear segment concept remained. Long profiles were treated as a number of sequential segments or *sampling lengths* and each was independently analyzed relative to its own reference line (ANSI/ASME B46.1-1985). This is graphically demonstrated in Figure 5.1 based on a profile obtained from a through-feed ground shaft used in a diesel engine fuel pump (as measured by a Form Talysurf S5 incorporating a 2 μm tip radius, 0.25 μm ordinate spacing and 0.01 μm vertical resolution).

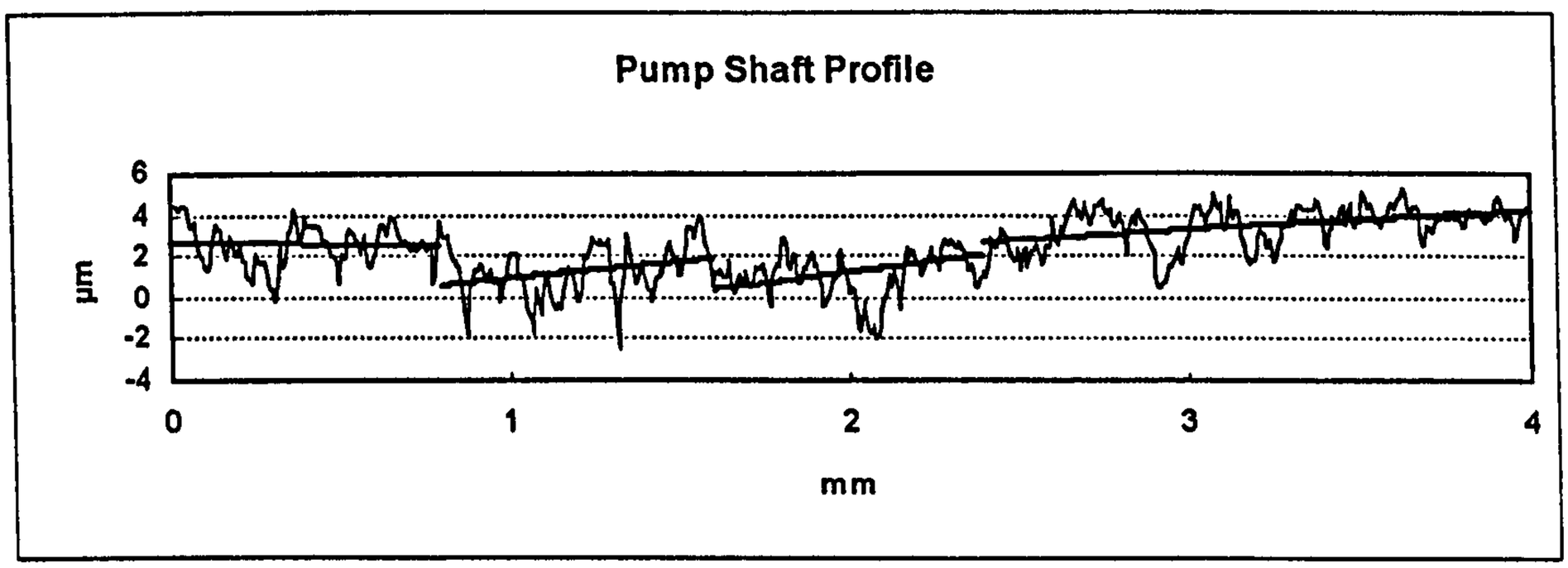


Figure 5.1 *Application of linear segments (0.8 mm segment lengths).*

The reference lines were originally determined manually, whereby a technician would construct a line in the *general direction* of the profile. More recently, this approach has been computerized based on the implementation of linear regression within each segment.

Linear regression is applied in each segment based on the determination of slope a and intercept b from the following:

$$y = ax + b \tag{5.1}$$

Where y represents profile amplitude and x is the position along the profile. The determination of the constants a and b is based on common linear regression techniques (Neter et al. 1985) whereby:

$$a = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \sum x \sum x} \tag{5.2}$$

$$b = \frac{\sum y}{n} - a\frac{\sum x}{n} \tag{5.3}$$

In applying this methodology to engineering surfaces, discontinuities between the sampling lengths become very apparent (Figure 5.1). These discontinuities do not seem to be appropriate in that the surface is continuous across the sampling length boundaries and it would, therefore, follow that the reference should also be continuous.

5.1.2 "Moving Window" Reference

This concern over discontinuities was addressed by Reason (1970) through the presentation of *moving window* or *mid-point locus* of reference lines. (Historically, this technique has also been described in digital signal processing texts (see also Bendat 1986) and is sometimes referred to as "boxcar" filtering.). In Reason's approach, a window of one sampling length in width is centered at a profile point and a linear reference figure is constructed. The mid-point of the reference line is retained as the reference at that specific location. Repeating this procedure at each of the profile points generates a more *continuous* reference as shown in Figure 5.2.

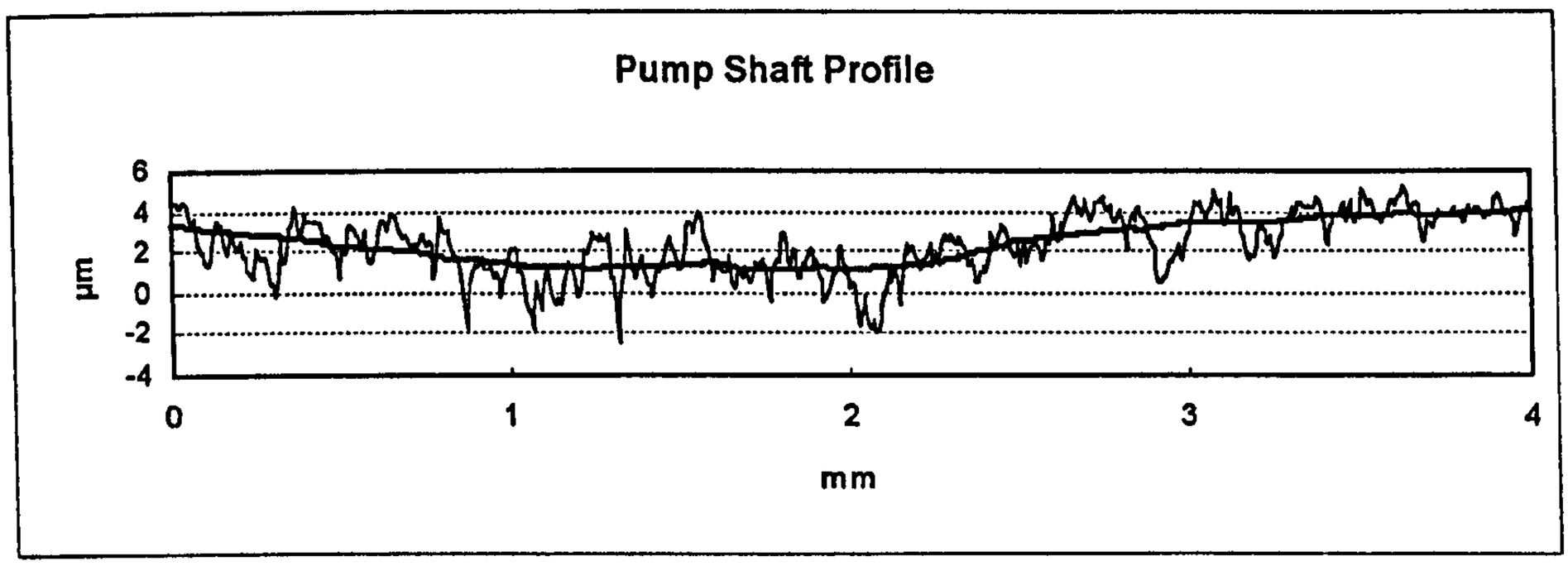


Figure 5.2 *Application of "moving window" reference (0.8 mm width).*

Although this "moving window" approach to removing form errors and/or waviness from roughness data set exhibits some interesting properties, such as linear phase and relative ease of understanding, it was not well accepted. This may be due to the limited digital computing capabilities of the day or perhaps due to its relatively poor wavelength transmission characteristics.

The transmission characteristic of the "moving window" reference is derived based on the fact that the center point of the moving linear regression follows the same path as the a square convolution (Bendat 1986). For data sets consisting of equally spaced x ordinates, the midpoint of the regressed line is determined based on the mean position along the x axis:

$$x_{midpt} = \frac{\sum x}{n} \tag{5.4}$$

substituting this into 5.1 yields:

$$y_{midpt} = ax_{midpt} + b = a\frac{\sum x}{n} + b \tag{5.5}$$

Finally, replacing the intercept (b) based on equation 5.3 results in:

$$y_{midpt} = a\frac{\sum x}{n} + \frac{\sum y}{n} - a\frac{\sum x}{n} \tag{5.6}$$

which reduces to:

$$y_{midpt} = \frac{\sum y}{n} \tag{5.7}$$

The long-pass frequency response of this weighting function has historically been presented in many signal processing texts and is given in the surface metrology context by Whitehouse (1994) as:

$$\frac{A}{A_0} \Leftrightarrow \frac{\sin\left[2\pi(\lambda_c / \lambda)\right]}{2\pi(\lambda_c / \lambda)} \tag{5.8}$$

for positive values of λ_c and λ. (A/A_0 is the ratio of output amplitude to input amplitude at a given wavelength, λ_c is the cutoff wavelength and λ is a specific wavelength at which the transmission function is to be determined.) A numerically generated transmission function is graphically depicted in Figure 5.3.

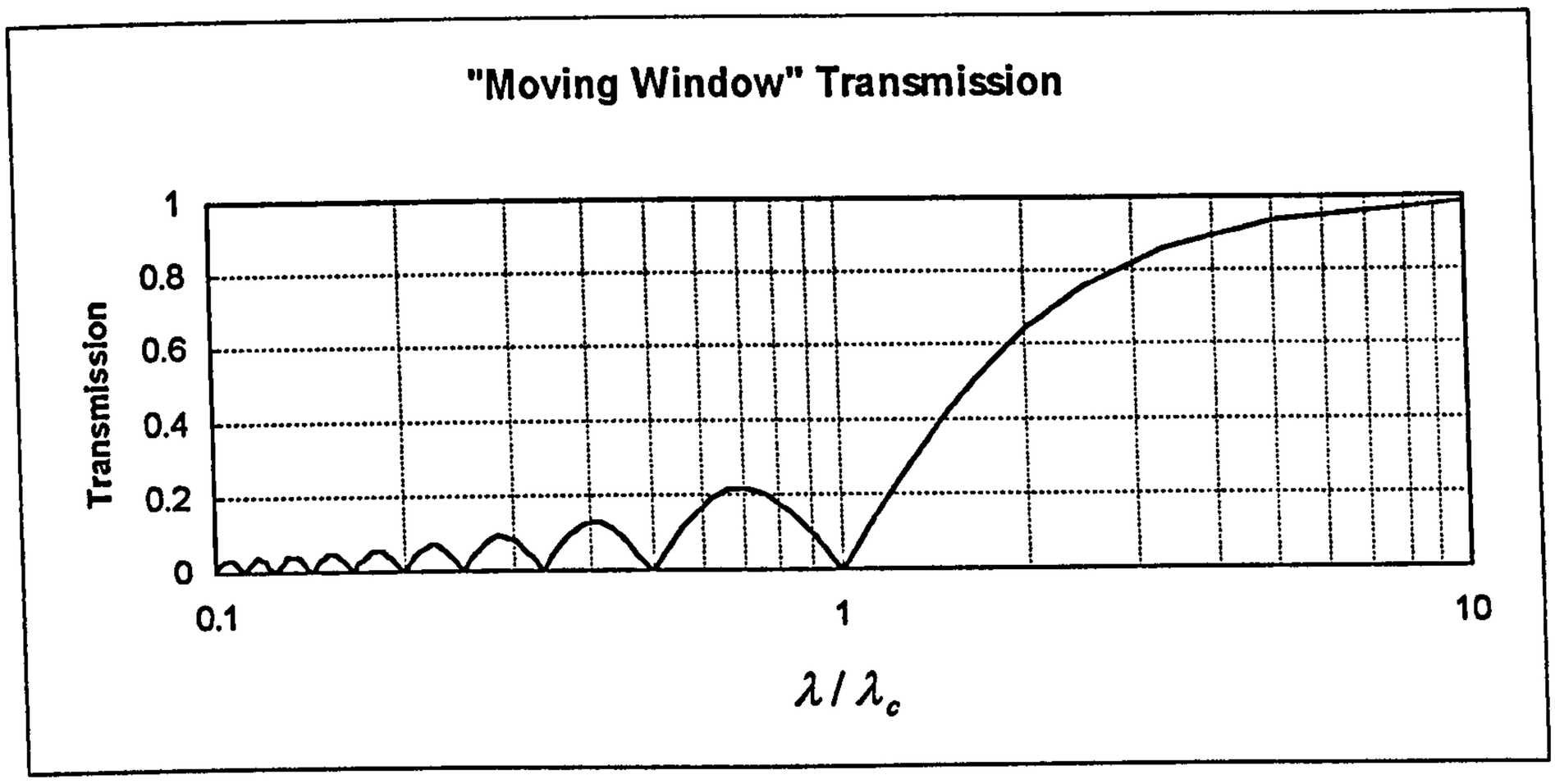


Figure 5.3 *"Moving window" wavelength transmission characteristics.*

The transmission function of Figure 5.3 was derived based on performing an FFT on numerical data and then determining the absolute value of the transmission (thus ignoring phase reversals). This numerical implementation can be described by the function:

$$\frac{A}{A_0} = \left| \frac{\sin[\pi(\lambda_c / \lambda)]}{\pi(\lambda_c / \lambda)} \right| \tag{5.9}$$

5.1.3 Capacitor-Resistor Networks

Shortly after the development of instrumentation and the establishment of linear reference figures, electronic means were developed for the separation of long wavelength surface features from short wavelength features (Reason 1967, Whitehouse and Reason 1965). The approach which was developed and later standardized (BS 1134 1988, ASME/ANSI B46.1.9 1995) for separating long wavelength surface features from roughness was based on a two-stage, buffered resistor capacitor (2RC) network. The filter components (resistors and capacitors) were tuned to transmit 75% at the 0.8 mm wavelength, in order to maintain some

degree of correlation with the previously used linear segment approach on typical surfaces (Whitehouse 1994). (This selection of the 75% transmission is significant in that it deviates from the more commonly used 3 dB intervals found in other signal processing applications. In a pure electronic sense a selection of 71% (-3 dB) or 50% transmission (-6 dB) might have been more logical.)

The high-pass implementation of the 2RC filter takes on the circuitry shown in Figure 5.4.

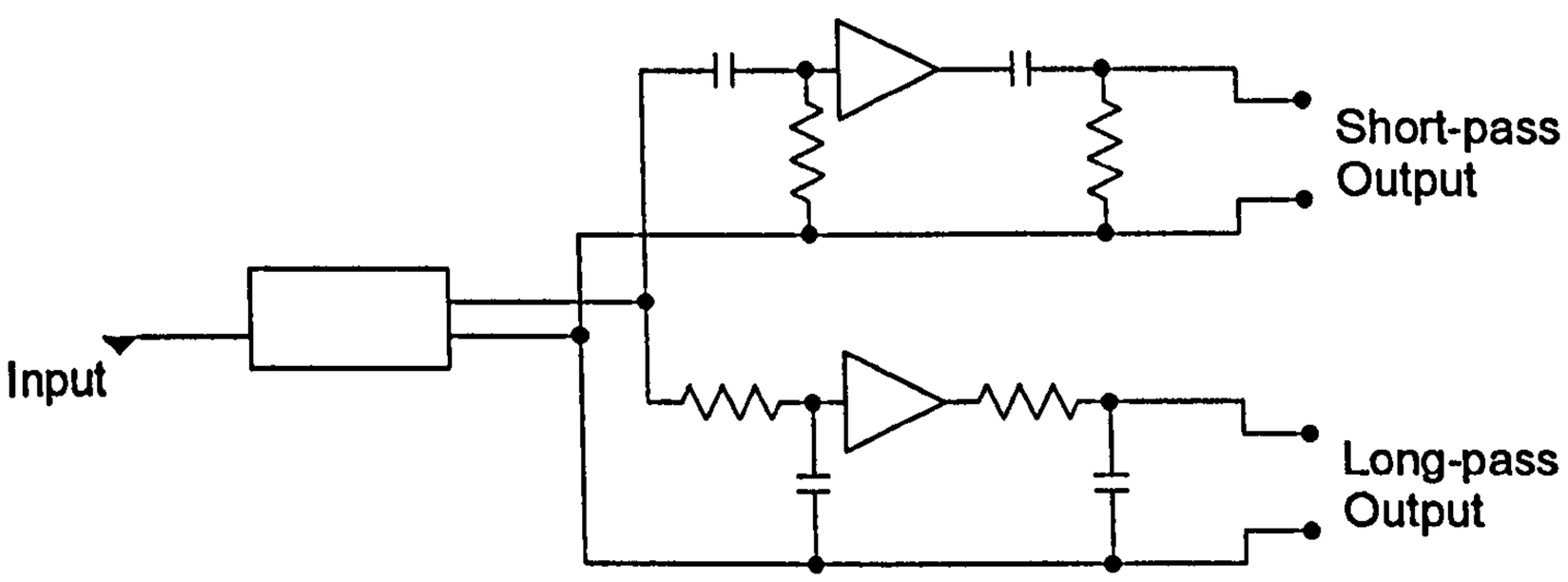


Figure 5.4 2RC circuits (Mummery 1992).

The roughness transmission function is given in equation 5.10 and 5.11 (amplitude and phase respectively) (Rank Taylor Hobson 1995). The amplitude transmission is graphically depicted in Figure 5.5.

$$\frac{A_1}{A_0} = \frac{3}{3+\left(\lambda/\lambda_c\right)^2} \tag{5.10}$$

$$\Phi = 2\left(\tan^{-1}\left(\frac{\lambda/\lambda_c}{\sqrt{3}}\right)\right) \tag{5.11}$$

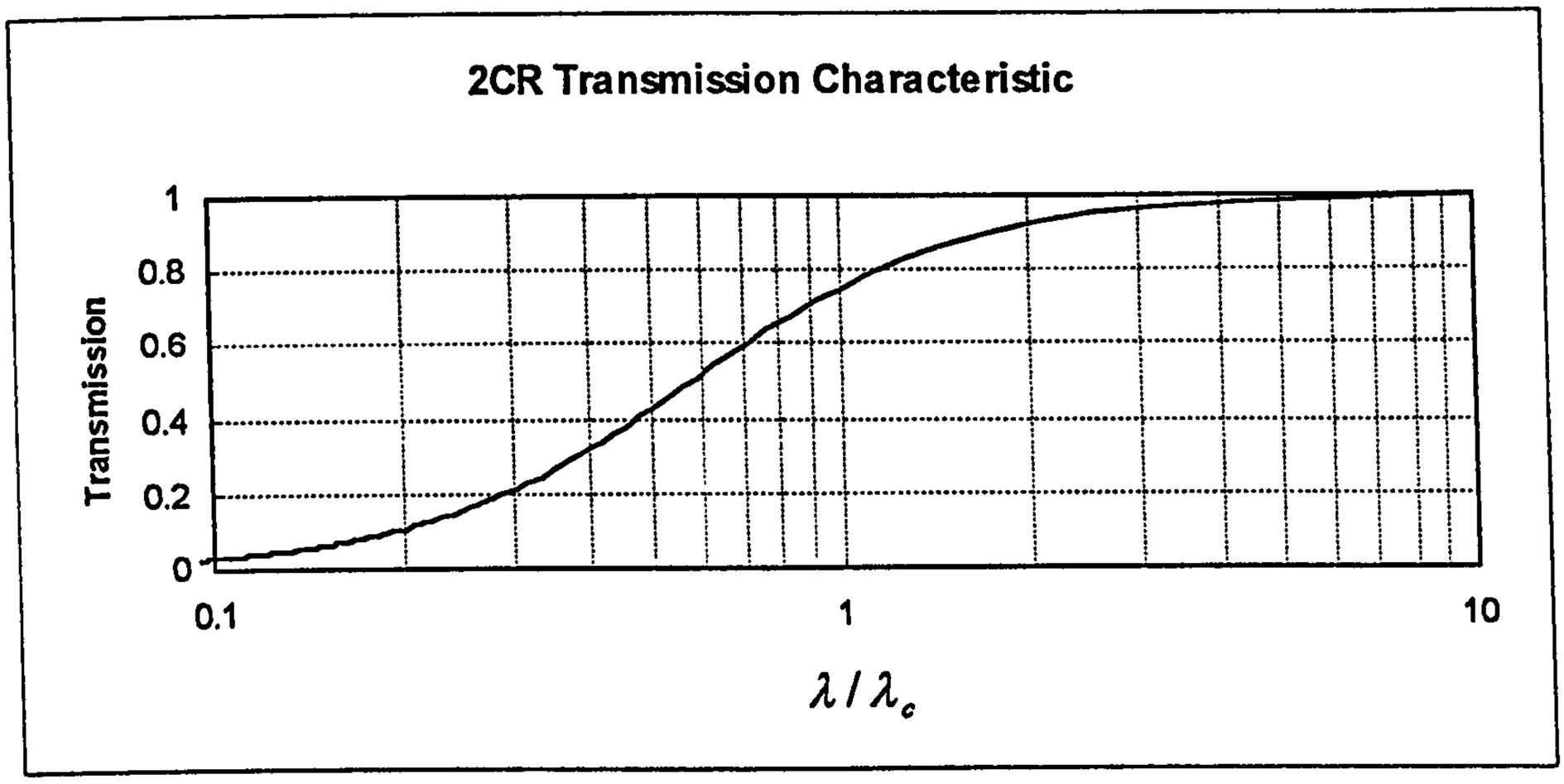


Figure 5.5 *2RC waviness amplitude transmission.*

In application, the 2RC filter *lags* in phase as indicated by the impulse response shown in Equation 5.12 and Figure 5.6 (ASME/ANSI B46.1 - 1995).

$$y=\left(\frac{A}{\lambda_c}\right)\left[2-\frac{A|x|}{\lambda_c}\right]e^{-\left(\frac{A|x|}{\lambda_c}\right)} \qquad (5.12)$$

Where x is the distance from the point currently being filtered (lagging), λ_c is the cutoff wavelength and A is a constant: 3.64 for 75% transmission.

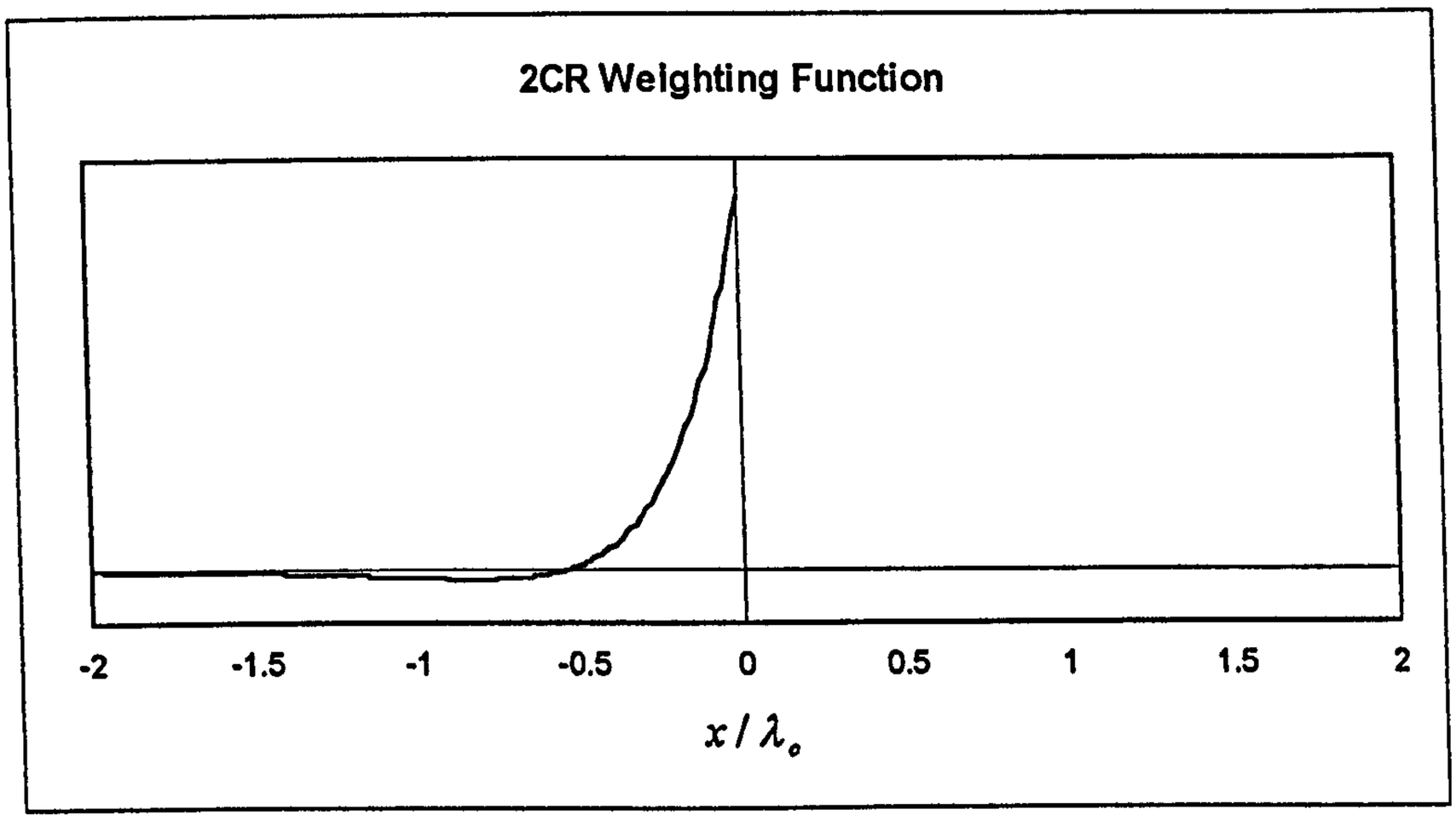


Figure 5.6 *2RC impulse response (weighting) function.*

The lagging nature of these filters causes distortions in the resulting data sets which are not inherent to the original (unfiltered) profile. Returning to the pump shaft data set, we see that the local minima in the low-pass 2RC profile do not align with the local minima in the unfiltered profile (Figure 5.7).

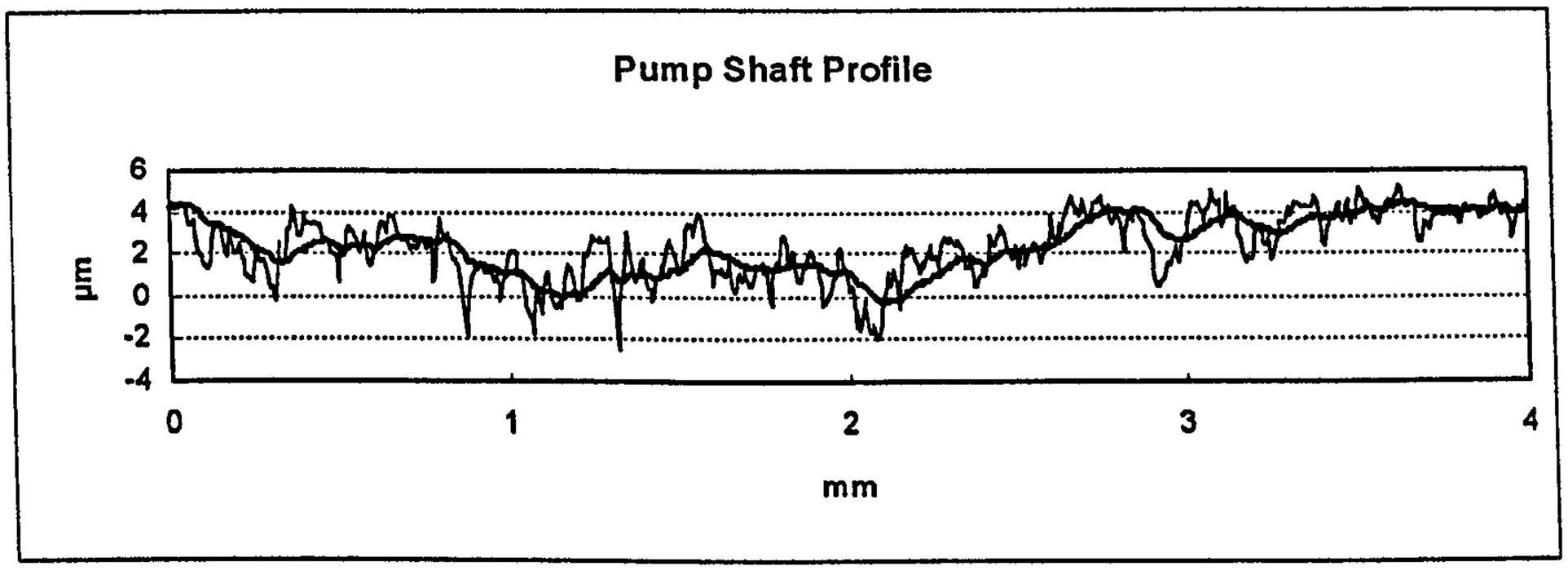


Figure 5.7 *Application of the 2RC filter.*

5.1.4 Phase Corrected Filters

The lagging nature of the 2RC filter was recognized as undesirable and alternative methods were developed. These alternatives were based on symmetric time-domain

weighting functions (thus generating symmetric impulse response functions). In the strictest sense, these approaches are referred to as *linear-phase* and are typically implemented digitally such that profile data points are stored and subsequently filtered either by a time-domain convolution or a frequency domain amplitude attenuation (Raja and Radhakrishnan 1979).

Whitehouse (1967) published the first implementation of a phase corrected filter for surface metrology. This approach, implemented as a short-pass (roughness) filter, provided 100% transmission for all wavelengths up to the cutoff and a linear attenuation (in frequency) up to three times the cutoff wavelength. This approach, though often cited in literature, did not gain wide acceptance due primarily to the lack of computational capabilities in the computers of that era. A more efficient approach for generating a non-lagging filter for surface metrology was subsequently developed by Kinsey and Chetwynd (1973). This approach utilized a symmetric, time-domain weighting function based on the mirroring of the phase lagging 2RC weighting function. More importantly, this approach could be numerically implemented in a recursive manner whereby it could be more readily accommodated by available computers. This "phase corrected 2RC" filtering technique gained a great deal of acceptance due to its incorporation in commercial instruments and is still very common in today's instrumentation.

5.2 Gaussian Filters

Recently, the Gaussian filter (ISO/TC 57/SC 1 - 1988, Whitehouse 1994, ISO 11562 - 1995) has gained broad acceptance as a suitable means of separating wavelength domains in surface metrology data sets. In this approach, a *Gaussian-like* time domain weighting function is convolved through the data set generating a transmission function which is linear in phase and sharper in attenuation than the previously standardized 2RC filter. (Note: the term "Gaussian-like" is used in the strictest sense as a true "Gaussian" distribution is infinitely wide.)

The Gaussian filter has advantages over the previously developed approaches in that it exhibits the sharpest possible attenuation characteristics while maintaining smooth transitions in both space and frequency without "ringing" or "overshoot" (Whitehouse 1994, Scott 1995). Thus it is included in the scheme for unification as the primary means of wavelength separation.

The time domain weighting function (or impulse response function) for the Gaussian filter is given in equation 5.13 and is shown graphically in Figure 5.8.

$$s(x) = \frac{1}{\alpha\lambda_c} e^{-\pi\left[\frac{x}{\alpha\lambda_c}\right]} \tag{5.13}$$

where x is the position along the surface, λ_c is the cutoff wavelength and α is the constant $\sqrt{\ln(2)/\pi} \approx 0.4697$.

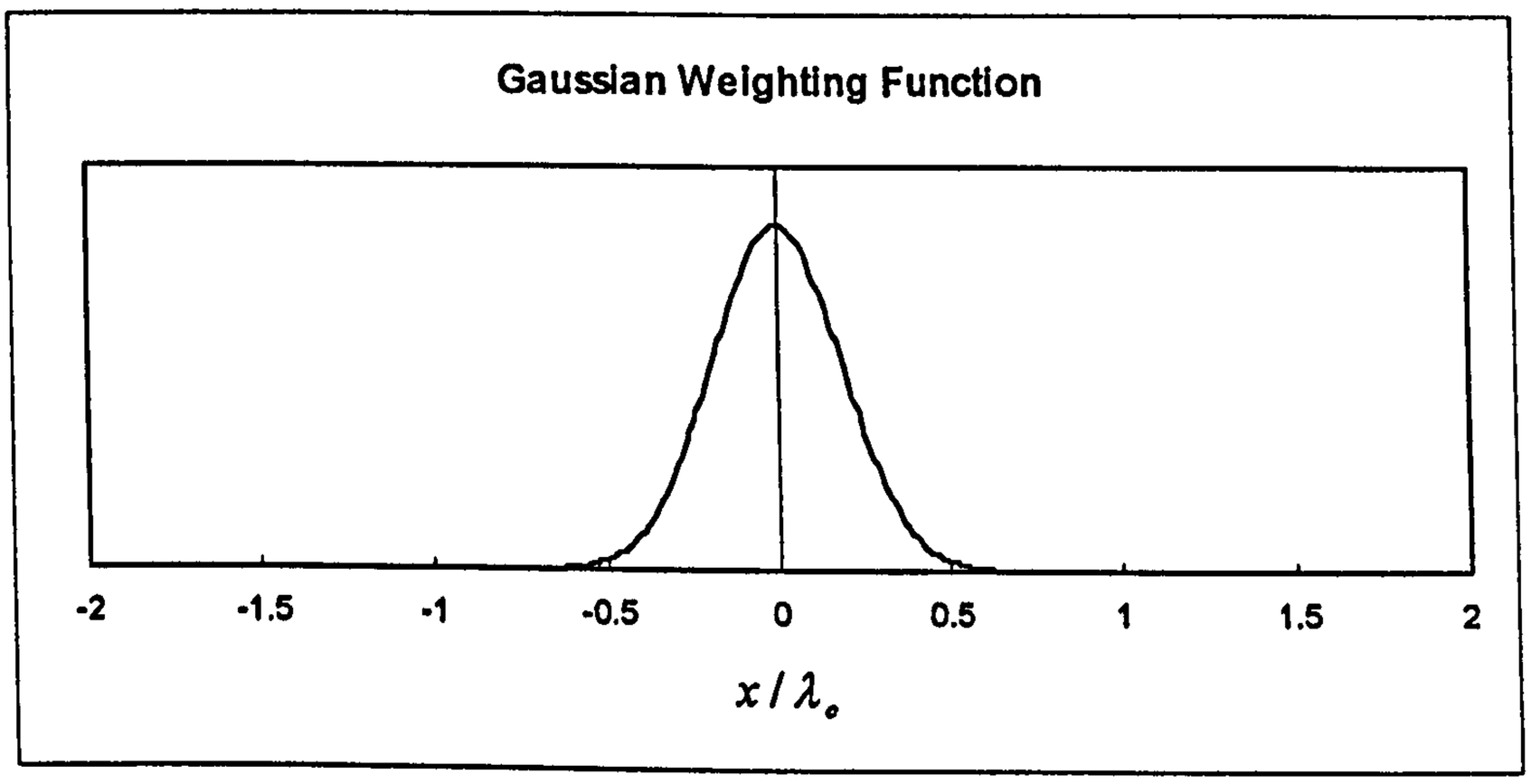


Figure 5.8 Gaussian weighting function.

In many time-domain applications of the Gaussian filter, adequate precision is obtained when only one half of a filter cutoff is included on each side of the central data point and the weighting function is subsequently scaled to maintain unit area (Rank Taylor

Hobson 1995). This allows for the assessment of a larger portion of the measured profile in that only one half of a cutoff length is discarded on each side of the data set. (Historically, many 2RC implementation required two cutoffs lengths at the beginning of the data set in order to achieve stability.) This attribute of the Gaussian filter is very important in the measurement of relatively *short* surface profiles such as those encountered in many critical grooves.

Another significant benefit of the Gaussian filter is the symmetry in the wavelength (or frequency) domain transmission function and the defined 50% transmission at the cutoff, thus allowing the algebraic reconstruction of profile components. In other words, when using the Gaussian filter, the high-pass and low-pass profile components can be directly added to obtain the original profile (assuming no other filtering has been performed). This is not possible with the 2RC filter, wherein the roughness profile and the waviness profile each contained 75% of the amplitude at the cutoff wavelength!

The Gaussian amplitude transmission functions are given in equations 5.14 and 5.15 and are graphically depicted in Figures 5.9 and 5.10 for long-pass and short-pass implementations respectively (ISO 11562-1995). The explicit mathematical definition of the Gaussian filter in both spatial and wavelength (or frequency) domains provides additional benefits in standardization and software testing.

$$\frac{A_1}{A_0} = 1 - e^{-\pi\left(\frac{\alpha\lambda}{\lambda_c}\right)^2} \tag{5.14}$$

$$\frac{A_1}{A_0} = e^{-\pi\left(\frac{\alpha\lambda}{\lambda_c}\right)^2} \tag{5.15}$$

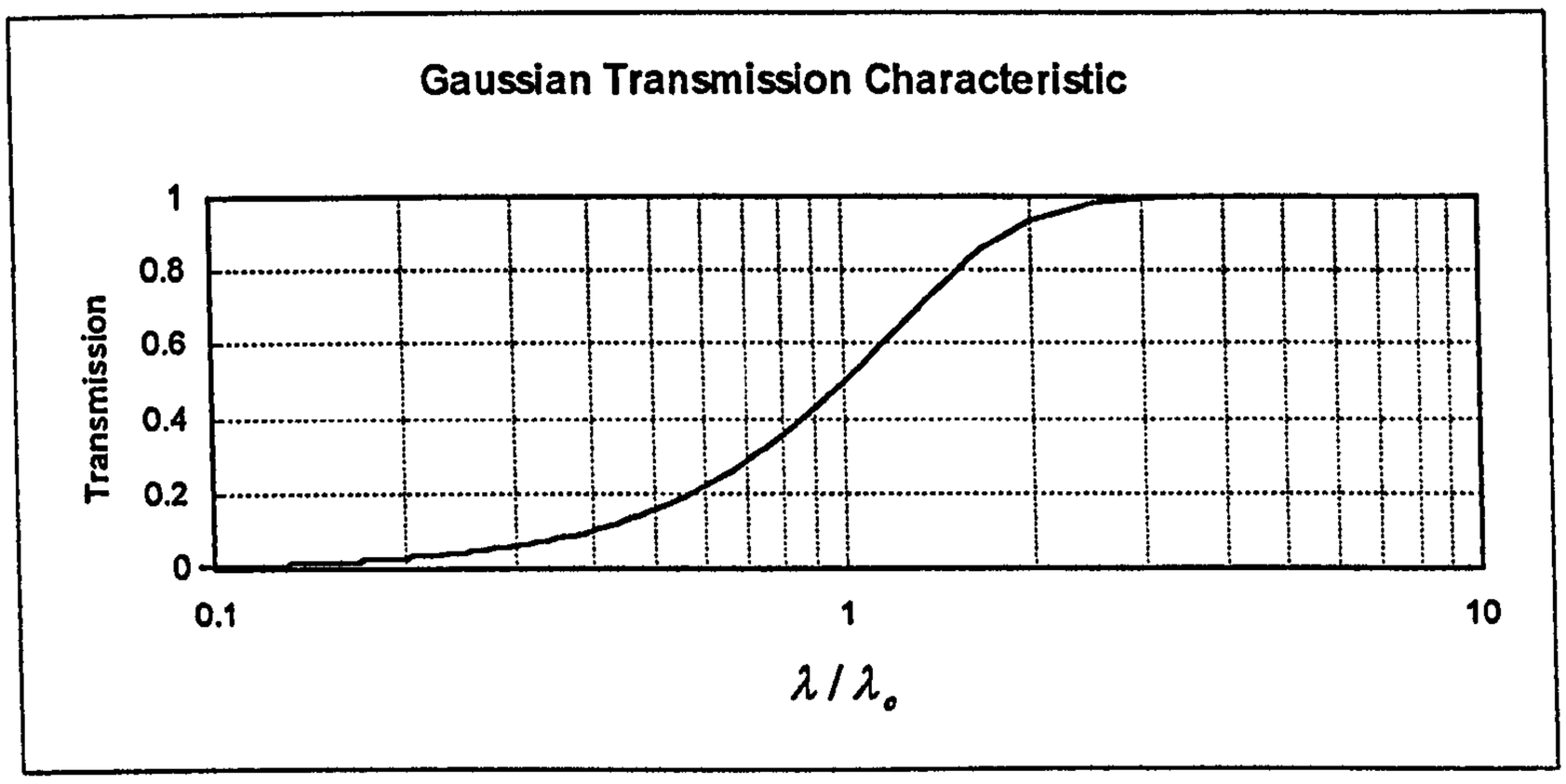


Figure 5.9 *Gaussian long-pass transmission function.*

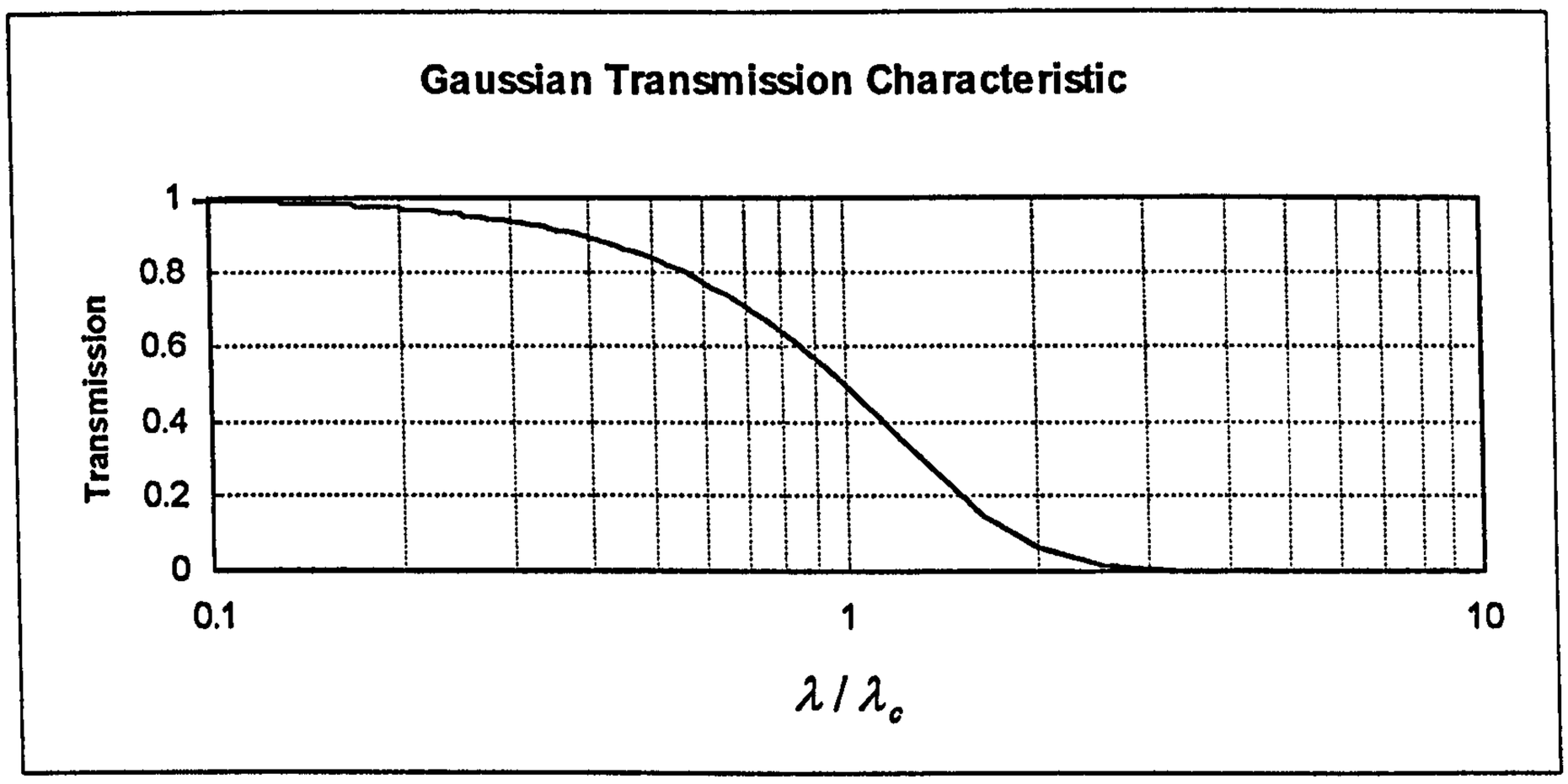


Figure 5.10 *Gaussian short-pass transmission function.*

Figure 5.11 demonstrates an example of the Gaussian filter as it is applied to the pump shaft data set previously used in the presentation of other filtering approaches.

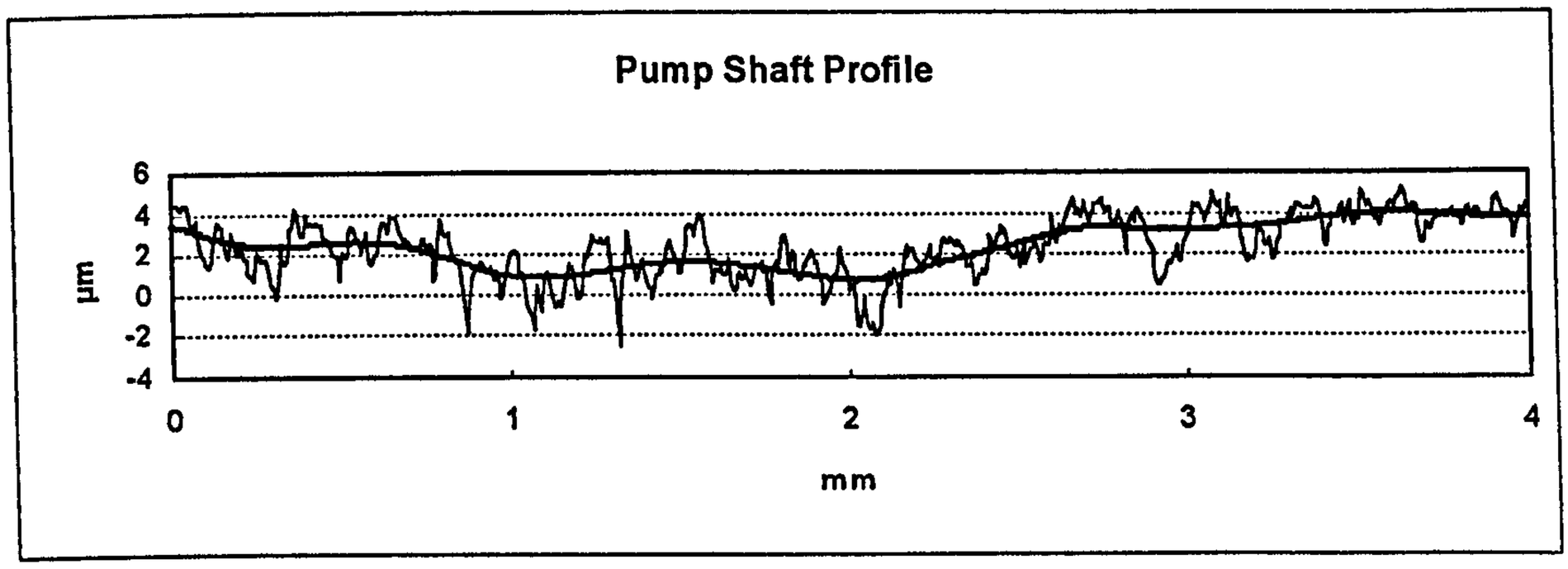
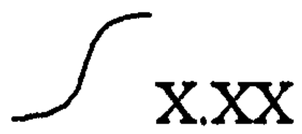


Figure 5.11 *Application of the Gaussian filter.*

The "phase correct" nature of the Gaussian filter becomes readily apparent in Figure 5.11. Furthermore, in comparing this (long-pass) profile with that of Figure 5.2 we see that the Gaussian filter seems to more closely follow the general direction of the profile waviness. This can be attributed to the slightly shorter cutoff of the Gaussian filter (50% at 0.8 mm) as opposed to the moving window filter (0% at 0.8 mm).

5.2.1 Unified Methodology Implications

Given the above mentioned technical advantages coupled with their general acceptance across a broad range of surface metrology applications, the unified methodology for surface metrology establishes Gaussian filters as the recommended means for the separation of wavelengths. The Gaussian filter is designated under the unified specification scheme in long- pass, linear applications as:



whereby X.XX indicates the cutoff wavelength (in linear units of measure) (see Chapter 2 for further details and frequency based designations). Similarly the short-pass Gaussian filter is indicated by:

Y.YY

whereby Y.YY indicates the cutoff wavelength (in linear units of measure).

5.2.2 Numerical Implementation

While the unified methodology incorporates the Gaussian filtering approach, it does not dictate the methodology for performing the filtering. In their most basic form, Gaussian, long-pass filters can be implemented by a time-domain convolution of the weighting function given in Equation 5.13. This directly results in the long-pass, filtered data set. The short-pass profile is obtained by subtracting this long-pass profile from the original data set. While this methodology is conceptually very simple, it is costly in computational terms in that several (typically floating point) multiplications are required to obtain each filtered data point. (See example listing D.01 in Appendix D.)

The Gaussian filtering process can also be accommodated by a frequency (or wavelength) domain attenuation of amplitude (taking care to not modify phase) according to the Gaussian transmission functions given in Equations 5.14 and 5.15. For larger data sets, the Fast Fourier Transform (FFT) can be utilized and after a modification of amplitudes in the wavelength (or frequency) domain, an inverse FFT results in the filtered profile (Press et al. 1992). (See example listing D.02 in Appendix D.)

Although the above mentioned methodologies for implementing the Gaussian filter result in more correct filter behavior, many practical implementations of the Gaussian filter are based on a triangular approximation (Hildebrandt 1994). DIN 4776 (1990) describes this approach and gives the half width of this weighting triangle as 0.44294647 times the cutoff. In this approximation, a triangular weighting function which approximates the Gaussian is applied in the spatial domain (see Figure 5.12).

This triangular convolution can be implemented digitally as two, sequential square convolutions (each convolution being half the width of the desired triangle) which can be performed with significantly fewer multiplications than the direct Gaussian convolution. The two, square convolutions are further optimized when treated as a moving average, whereby at each step of the convolution, the next point is added and the trailing point is dropped from the current, moving sum (Nielsen 1993). (See example listing D.03 in Appendix D.)

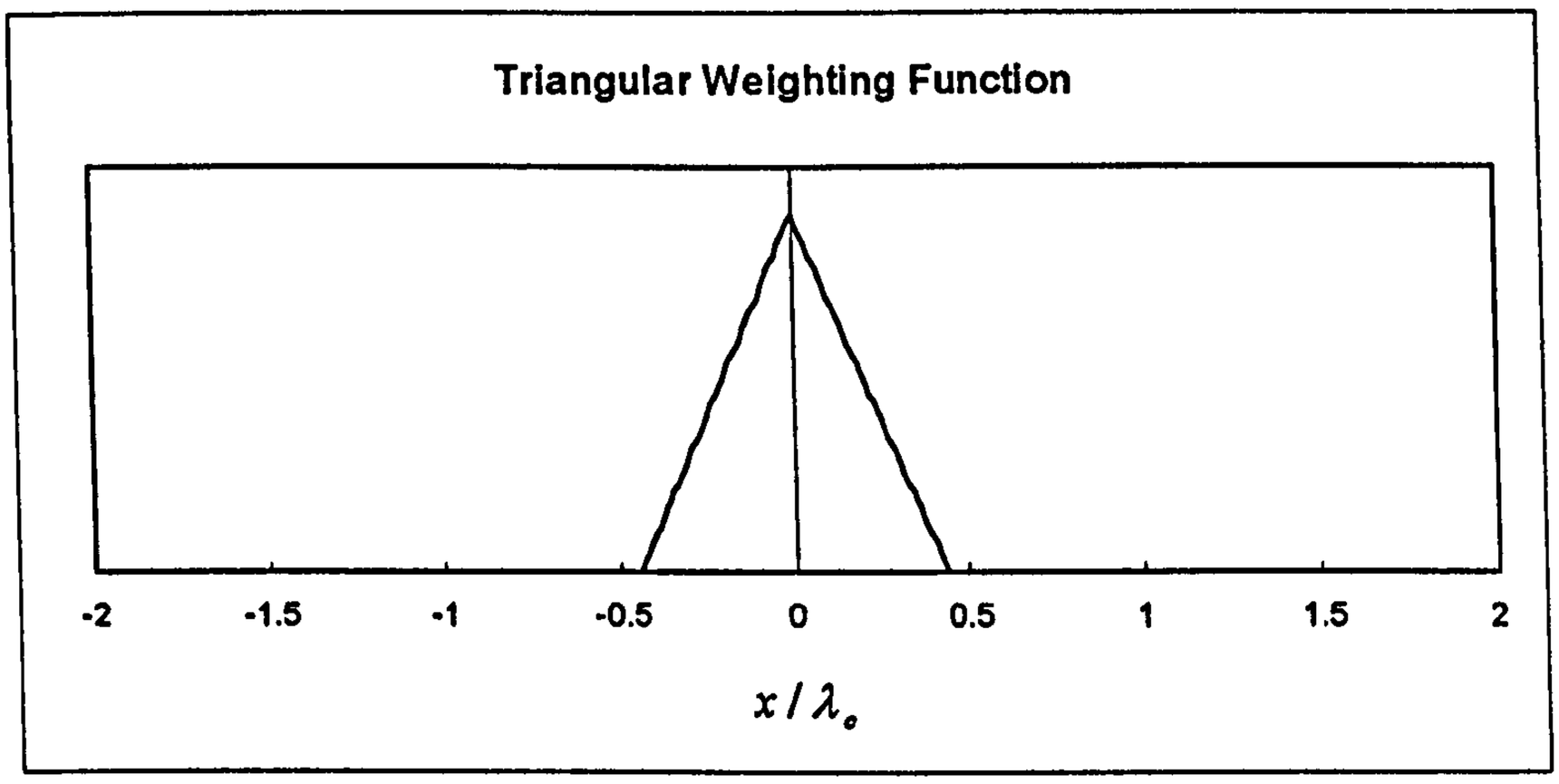
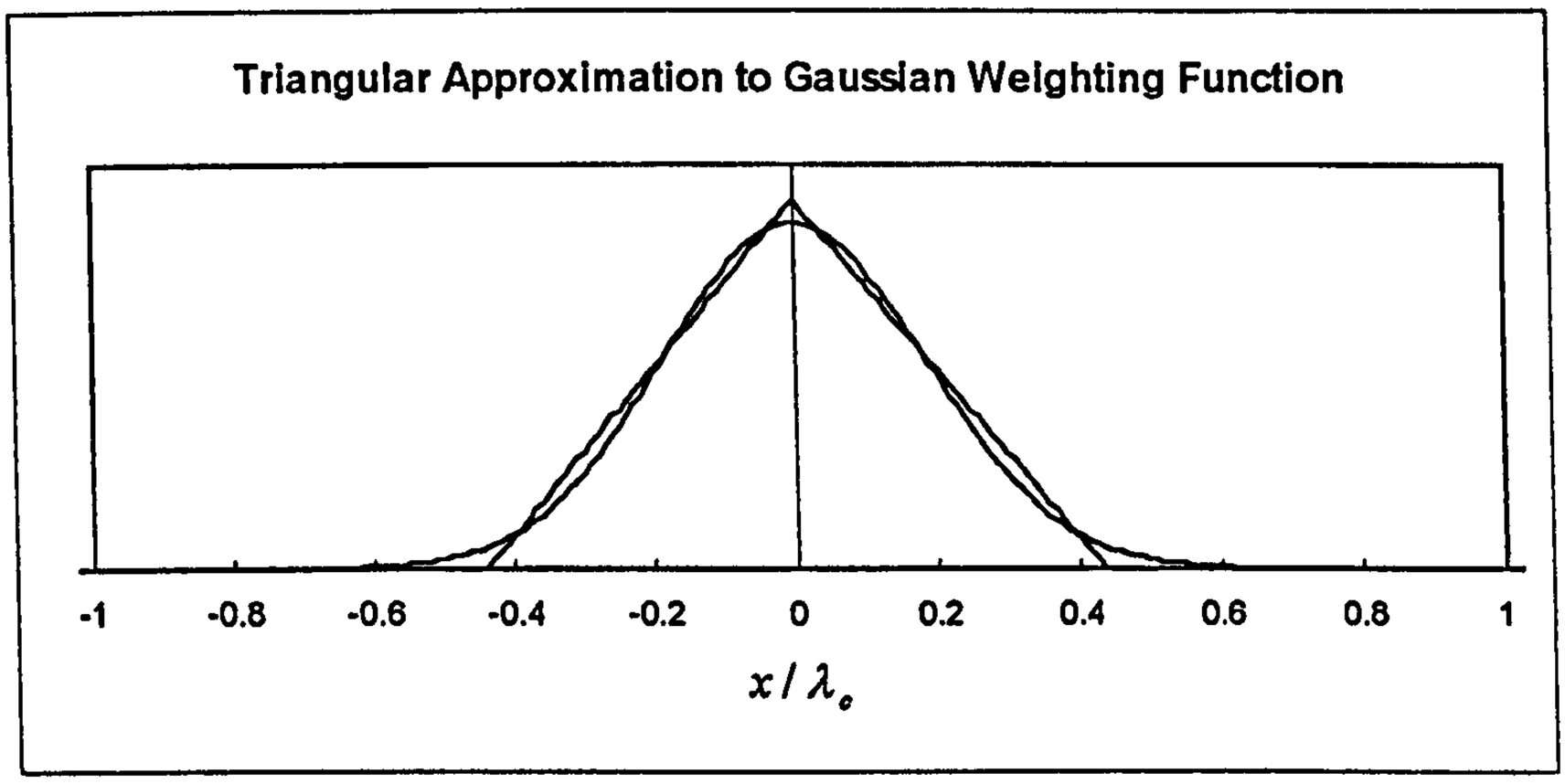


Figure 5.12 *Triangular approximation to Gaussian weighting function.*

The triangular approximation to the Gaussian filter is very common in commercially available instrumentation (Hildebrandt 1994). This is the due to its computational speed coupled with its standardization in Germany (DIN 4776-1990). However, it is important to note that the triangular approach does deviate from the "true" Gaussian filter in terms of both its weighting function (Figure 5.13a and transmission characteristic (Figure 5.13b). Although better approximations of the Gaussian can be made by additional convolutions (Whitehouse 1994) they are rarely implemented in practice.

A.



B.

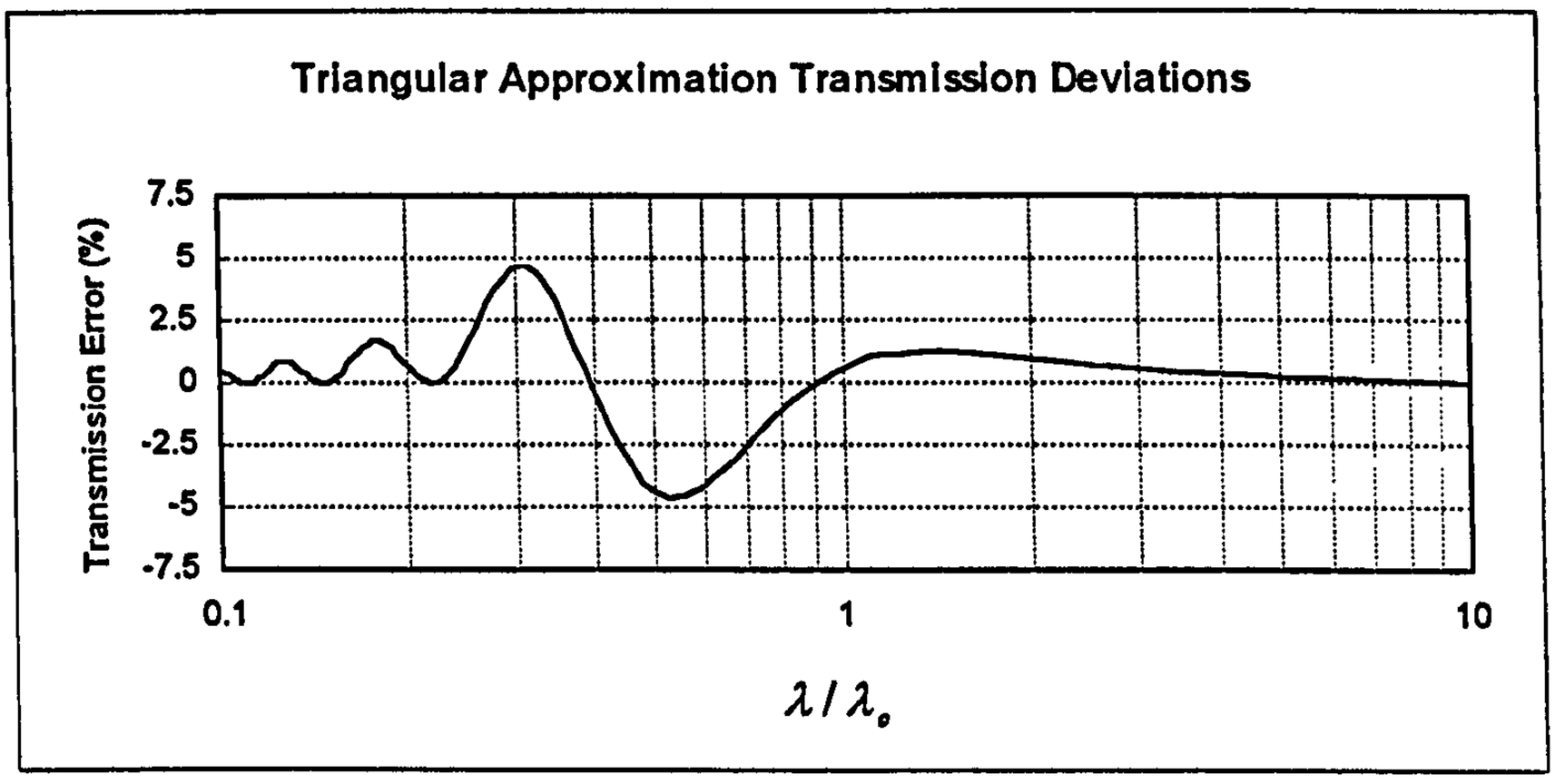


Figure 5.13 Triangular approximation deviations relative to a Gaussian filter.

Given the three typical approaches for implementing the Gaussian filter (space domain convolution, Fourier and triangular), it becomes necessary to compare them as to ascertain which methodology is most suitable in a particular application. As stated above, the direct, space domain convolution or the Fourier approach will yield the most correct results in terms of generating a proper transmission characteristic. However, in many cases the triangular approach is utilized based on computational speed.

The "computational speed" of Gaussian filter implementations is often referred to in practice, however very little has been published in this regard. Thus, the following study has been undertaken in order to determine the relative performance of these filtering techniques. The three filter implementations which are to be compared (provided in Appendix D) are described briefly as follows:

Time Domain Gaussian

A full numerical convolution of a time domain Gaussian weighting function is provided in Listing D.01. The weighting function is generated for one full cutoff on each side of the central point. (It is recognized that many implementations utilize only one half cutoff, however for the purpose of this study the more exact approach was desired.) Given the symmetry of the weighting function and to reduce memory requirements, only one half of the weighting function is stored.

Frequency Domain Gaussian Convolution

Given that surface metrology data sets do not always contain 2^n data points, typical Fast Fourier Transform (FFT) algorithms are not directly applicable without some form of padding or re-sampling. A more appropriate method is through a mixed radix (prime factor) technique such as the one described by Singleton (1969) and is implemented in Listings D.02 and D.04. It should be noted that this implementation can suffer from memory limitations when dealing with large prime factors. In these cases, the data set is reduced by one point and the transform is re-computed. Upon filtering, the last long-pass (waviness) point is extrapolated to the discarded endpoint. This approach accommodates the desired number of points while minimizing the end effect.

Triangular Approximation

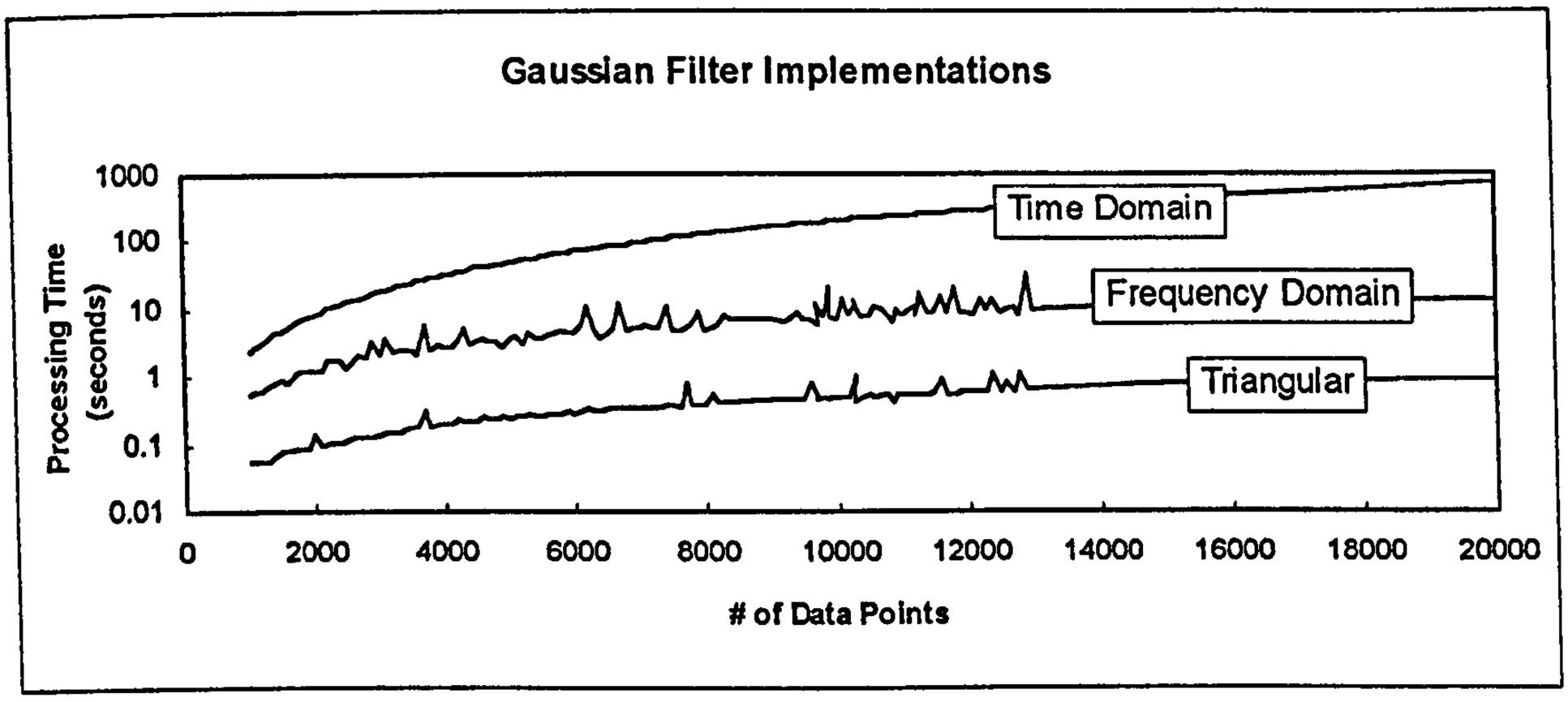
An implementation of the triangular approximation is provided in Appendix D, Listing D.03. This approach utilizes a "square convolution" subroutine

whereby the moving average technique is employed. Although this approach is very fast in application, one drawback is that a temporary array must be allocated for intermediate results.

To compare the processing time for each of the approaches, the "C" language implementations where compiled into a single "project" thereby ensuring similar compiler optimizations. All computations were performed in double precision (64 bit) floating points values. A sinusoidal data set was then generated (7 undulations over 5.6 mm, 1 μm amplitude) with a variable number of data points. It should be noted that the use of a sinusoidal data set does not affect the processing time. Processing time, by these implementations is only a function of the number of data points.

For the study, a 80486 DX (33 MHz) personal computer was used. The number of data points ranged from 1000 to 20000 with 100 point increments (1000, 1100, 1200, 1300, ..., 20000). The processing time required for each of the implementations is graphically presented in Figure 5.14. (Figure 5.14a presents the data on a semi-log graph, while 5.14b is a log-log representation.)

A.



B.

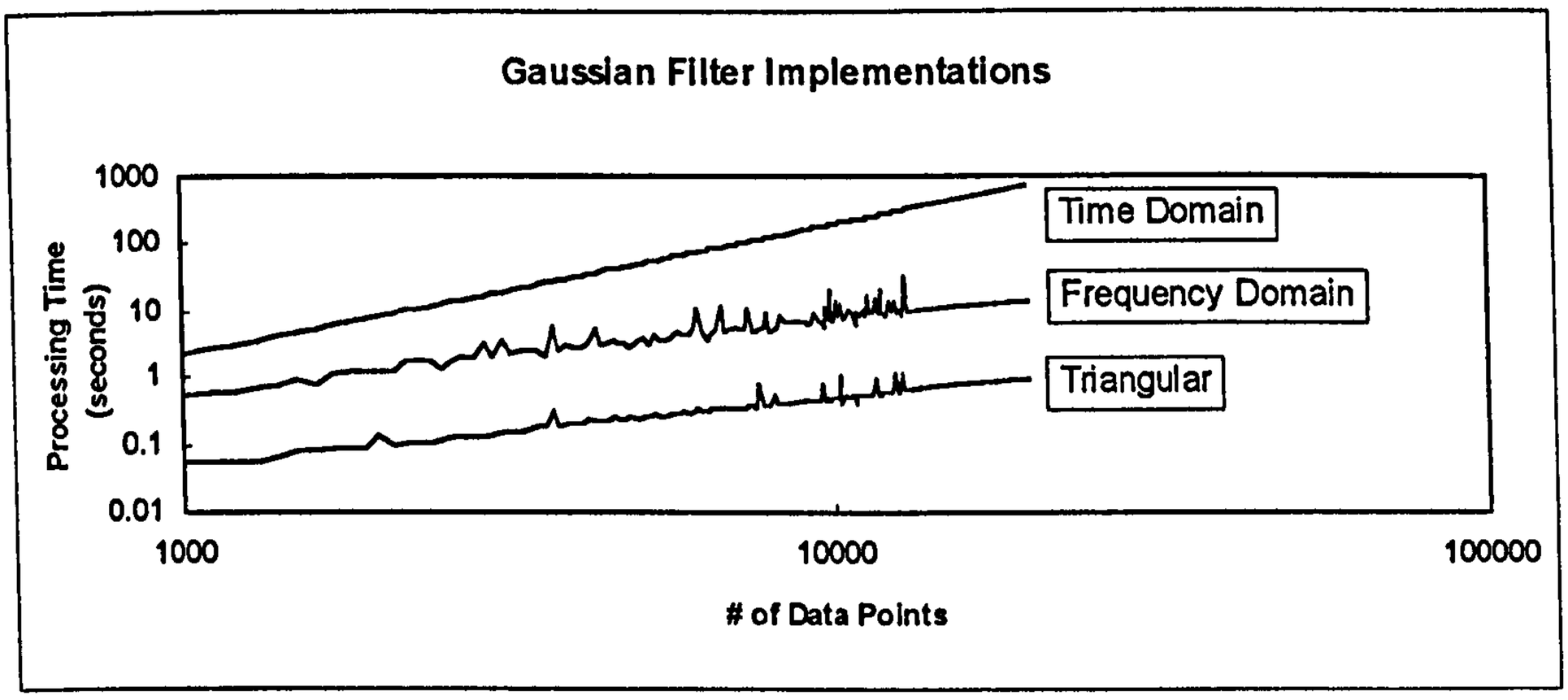
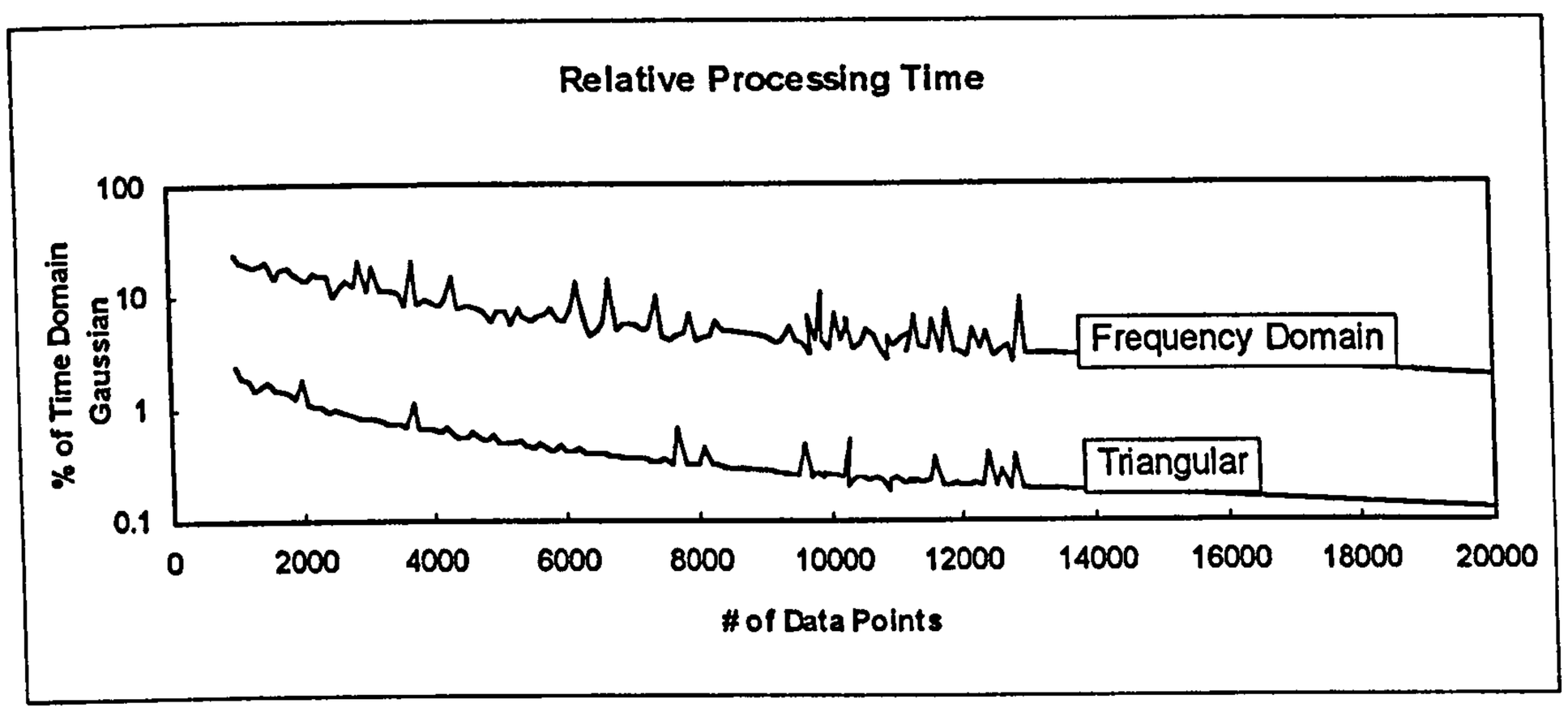


Figure 5.14 *Computation time for various Gaussian implementations (80486DX, 33 MHz personal computer, 64 bit (floating point) calculations).*

It should be noted that the "spikes" or "noise" in the Frequency Domain and Triangular implementations are the result of memory swapping to the hard disk the writing of a results file. Furthermore, since the data in the Figure 5.14 graph is based on sample sizes which are whole multiples of 100, there may be additional variations in the processing time per the Fourier Approach when using other numbers of points (depending on the calculated prime factors). Nonetheless, the general trends are very useful in comparing the processing times for the various implementations.

Another interesting presentation of the data is that of Figure 5.15, whereby the Fourier and Triangular approaches are compared to the full, time domain convolution. This representation indicates that as the number of data points increases, the reductions in processing time become more significant.

A.



B.

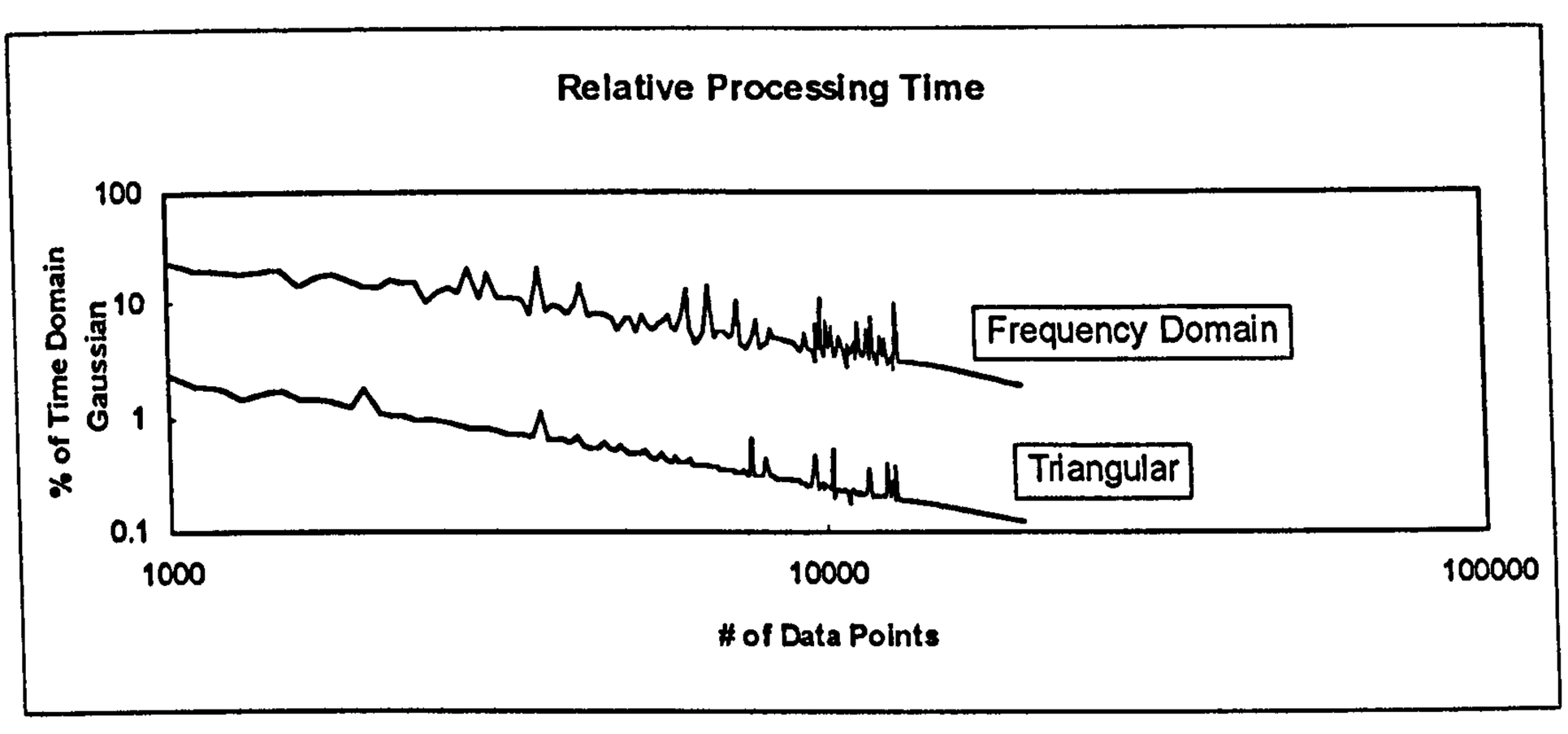


Figure 5.15 *Computation time relative to full time domain convolution of the Gaussian.*

As stated above, the unified scheme for surface metrology does not dictate any specific methodology to be utilized in a Gaussian filtering approach. This is consistent with the other aspects of the scheme, for example those concerning hardware - the scheme does

not dictate and specific measurement hardware, it merely requires that the proper hardware be used based on the desired transmission. Table 5.1 provides a useful summary of the above described filtering techniques.

Approach	*Advantages*	*Disadvantages*
Full Time Domain Convolution	Easy to program based on Equation 5.13 weighting function. Provides a "correct" transmission function.	Very slow in terms of computation time.
Fourier Based Convolution	Depending on the FFT approach this can result in significant improvements in processing time. Provides a "correct" transmission function.	May be subject to round-off errors. Some algorithms place limitations on the number of allowable data points (for example 2^n) Some FFT algorithms require a great deal of memory.
Triangular Approximation	Very fast in terms of processing time.	Generates errors in transmission characteristic up to nearly 5% of the nominal transmission. Requires extra memory for temporary storage of the profile.

Table 5.1 A comparison of Gaussian filter implementations.

5.3 *Ideal* Wavelength Limitation

The rather gradual amplitude attenuation characteristics associated with the Gaussian filter make it undesirable in applications where certain, specific wavelengths or frequencies must be characterized. A common example of this scenario is found in roller or ball bearing applications (Yhland 1967, SKF 1992). In these applications, it is

critical that the circular race (which contains the bearings) does not contain any significant amplitude at the frequency (or multiple thereof) corresponding to the number of rolling elements. If there is a significant amplitude at this frequency, there can be excessive chatter or vibration in the application of the assembled system. Figure 5.16 presents an example of a bearing race measurement with a significant lobing condition. This 14 lobe condition is further described through the Fourier analysis shown in Figure 5.17.

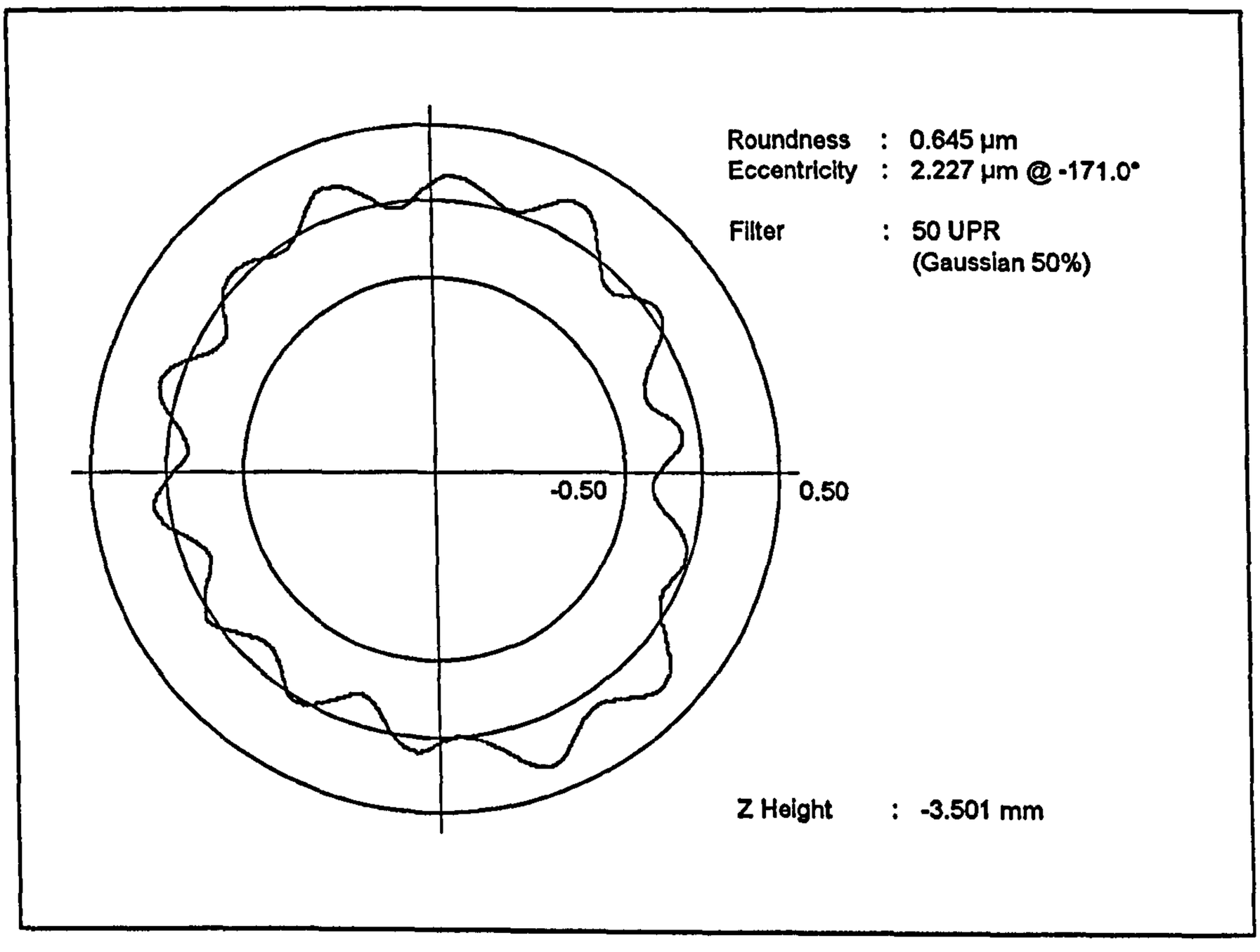


Figure 5.16 Bearing race roundness profile with significant 14 lobe condition.

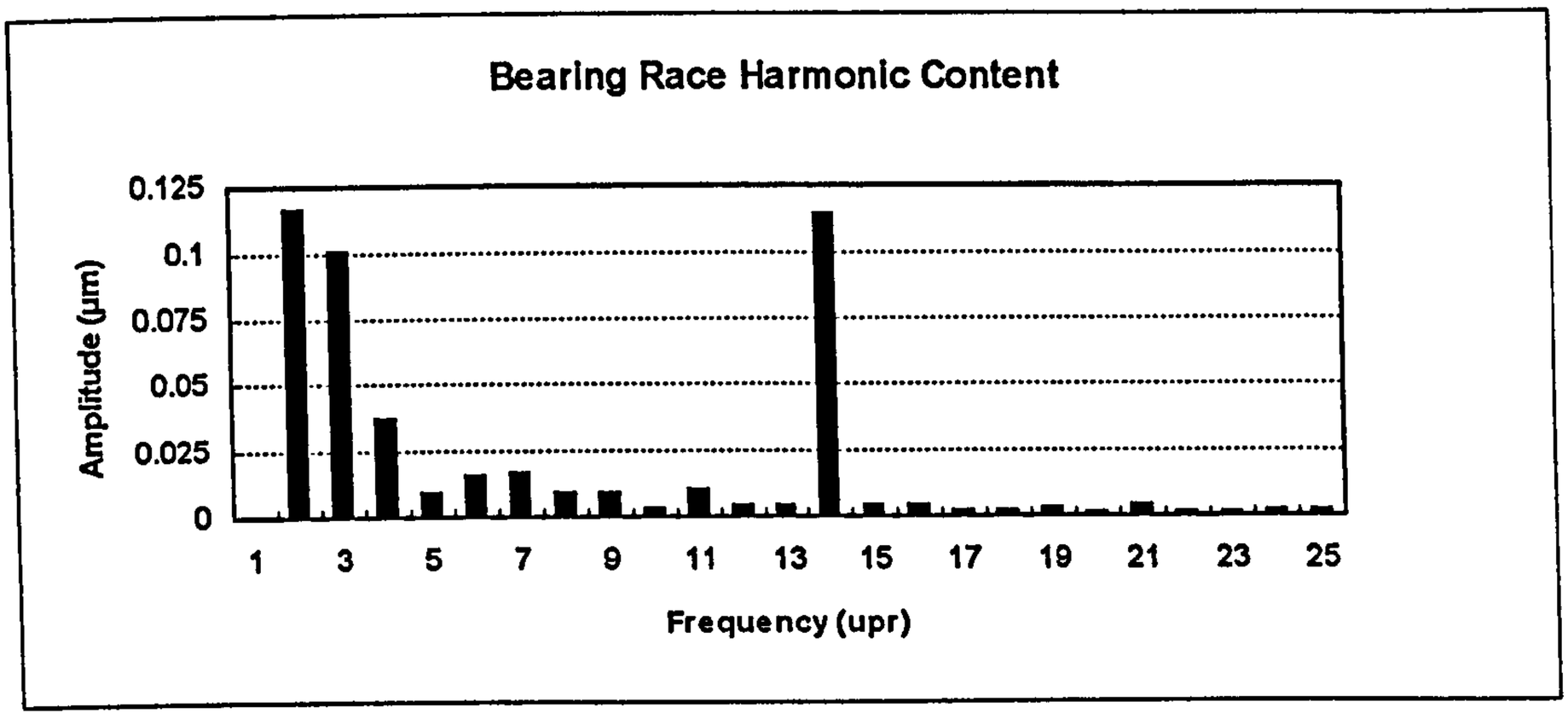


Figure 5.17 *Bearing race harmonic content.*

To analyze the functionality of the bearing race, the region of frequencies (upr) near the frequency associated with the number of rolling elements must be analyzed. The *ideal* methodology for wavelength limitation provides this capability. As stated in Chapter 2, the term "ideal" in this context does not imply that it is ideally suited for all applications. The term "ideal" (bounded by quotation mark or italicized) is used in the context of the digital signal processing (Antoniou 1993) and in the scheme for unification to refer to the mathematical process of modifying the frequency (or wavelength) content of a data set whereby certain frequencies (or wavelengths) are retained fully and others are removed fully.

In the *ideal* wavelength limitation approach, the data set undergoes a Fourier transform (typically implemented via a "fast" algorithm or FFT), and the desired frequency range is maintained while the amplitudes associated with all other frequencies are set to zero. While this modified frequency data (with associated phase information) can completely describe the resulting data set, an inverse Fourier transform is typically performed in order to return to a space (or time) domain data set for geometric profile analyses. (See listing D.04 in Appendix D for software implementation of FFT.)

5.3.1 Frequency (Wavelength) Domain Implications

The *ideal* approach is named based on its frequency (or wavelength) domain transmission characteristics - the desired frequency (or wavelength) components are completely retained and the others are completely removed. This transmission property makes it well suited for the analysis of relatively narrow band of frequencies (or wavelengths) without attenuation.

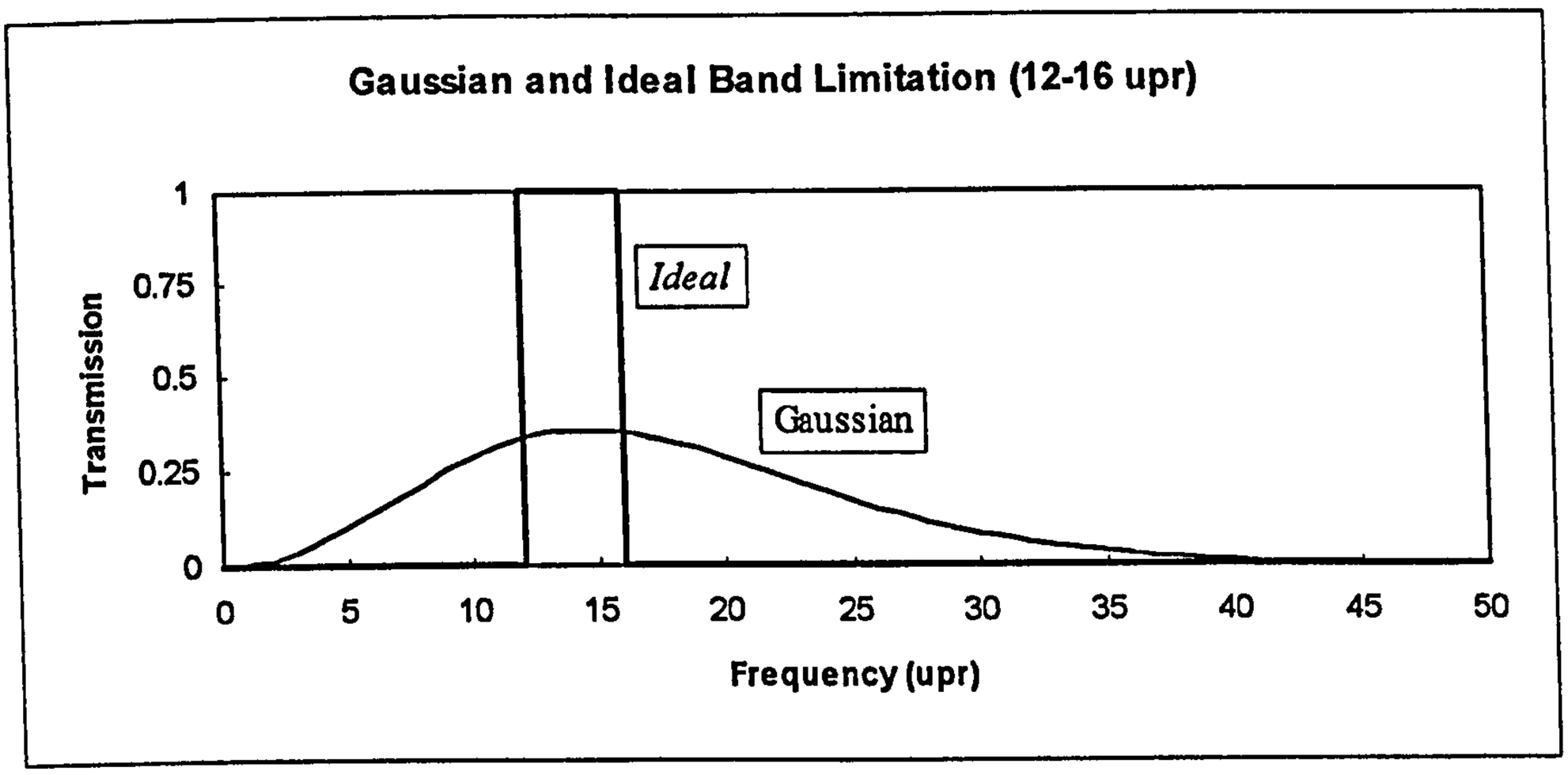


Figure 5.18 *Band limitation approaches.*

The analysis of bearing races can require very narrow bandwidths. Figure 5.18 clearly demonstrates that the utilization of a Gaussian band-pass filtering (12 to 16 upr pass band) causes a significant amount of attenuation in the transmission band while also allowing similar transmission of frequencies outside the transmission band.

5.3.2 Time (Space) Domain Implications

The *ideal* frequency (or wavelength) transmission capabilities of this wavelength limitation approach can cause significant distortions when performing an inverse transformation to return to the time or space domain. These distortions can be demonstrated using a roundness example and the analysis of a unit amplitude impulse

(at 180°). A comparison between *ideal* and Gaussian filters in this analysis is graphically presented in Figure 5.19.

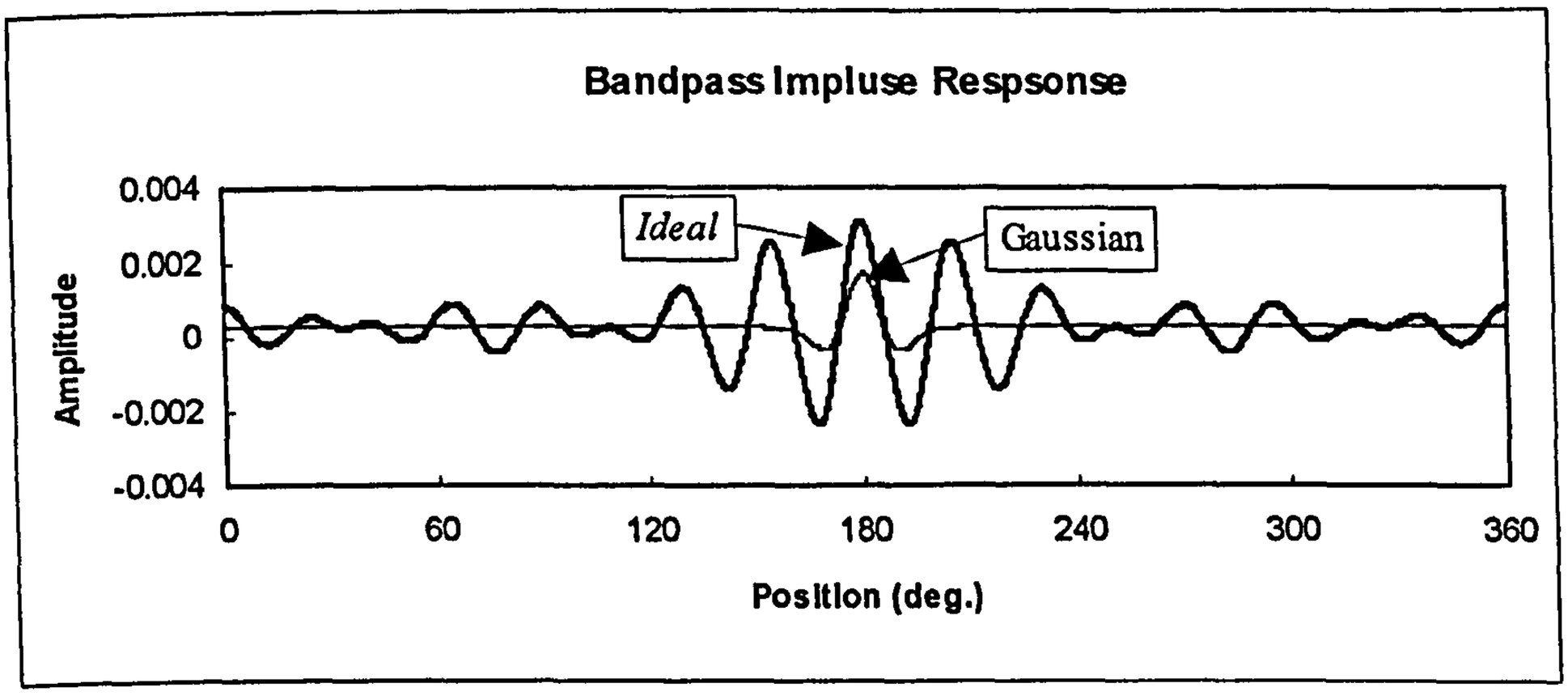


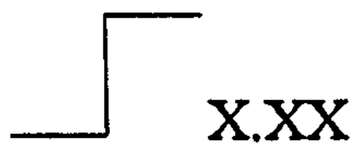
Figure 5.19 Gaussian and "Ideal" 12-16 upr band-pass impulse responses.

In the application of the *ideal* filter, ringing or "overshoots" can significantly influence the parameters obtained from the filtered profile. For example, parameters such as peak count or peak spacing will be much more a function of the filter artifacts than the underlying surface. In addition, the graphical representation (in the spatial domain) could be very misleading to an engineer in the visual assessment of the data set.

5.3.3 Unified Methodology Implications

Although the *ideal* approach is clearly better in separating features in the wavelength (or frequency) domain (as previously shown in Figure 5.18), the time domain properties may be undesirable (as shown above in Figure 5.19). Nonetheless, the *ideal* approach to wavelength limitation is included in the unified methodology as it provides very useful capabilities in characterizing surface attributes important in many engineering applications.

Ideal wavelength limitation can be specified under the unified specification scheme in a long-pass application as:



whereby X.XX indicates the cutoff wavelength (in linear units of measure). (See Chapter 2 for further details and frequency based designations.) Similarly, the short-pass, *ideal* limitation is indicated by:

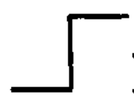
whereby Y.YY indicates the cutoff wavelength.

It should be noted that in the unified specification format the designated wavelength (or frequency) is transmitted at 100% of its original amplitude. This criteria is perhaps unclear in a theoretical sense in that any wavelength may be present in the profile. However, in a practical sense, the *ideal* filter is applied to digital surface profiles based on Fourier Transform algorithms yielding discrete wavelength or frequency data. These discrete wavelengths (or frequencies) are then compared to the desired *ideal* limitations.

5.4 Band Limitation Issues

The unified methodology provides for two primary means of wavelength separation, namely Gaussian and *Ideal*. These approaches can be applied as short wavelength limitations or long wavelength limitations. In addition, a stylus based (either physical or mathematical) approach can be used as a means of limiting short wavelengths. (This topic was discussed extensively in Chapter 3.) Thus, the options available for short wavelength limitation are as follows:

Gaussian  X.XX

Ideal  X.XX

Tip Radius **R**: X.XX

The long wavelength limitation approaches are limited to Gaussian and *Ideal* as follows:

Gaussian Y.YY 

Ideal Y.YY 

5.4.1 Short Wavelength Limitation

In the application of the unified specification scheme, a short wavelength (or high frequency) limitation is required. Many correlation problems have arisen between instruments (particularly in the measurement of roughness) due to differences in short wavelength transmission. The selected means of limiting short wavelengths drives the selection of instrument tip radius and data point sampling.

Radius Based Transmission

When applying tip radius based wavelength limitation (applying the unified "R" designation), the sampling should be based on detecting 90% of the cusp generated by the radius as it bridges two adjacent profile peaks (see Chapter 3). Thus for a given angle at the bottom of the cusp, Θ and the stylus tip radius r_{tip}, the maximum allowable data point spacing, S_{max}, is given by:

$$S_{\max} = \frac{0.1 \cdot r_{tip}\left(1-\cos(\theta)\right)}{\tan(\theta)} \qquad (5.16),\ (3.3)$$

as previously derived in Chapter 3. Furthermore, a 30° limiting slope is considered as the default and thus when applying "radius based" wavelength limitation:

$$S_{\max} = 0.023 \cdot r_{tip} \qquad (5.17)$$

It should be clarified that this requirement does not dictate the maximum slope (in the surface) which can be assessed, but rather it places a limit on the assessment of the maximum slope which is generated by the radius as it is convolved over one profile peak.

Furthermore, this should not imply that the Equation 5.17 sampling limit is always invoked for any tip radius. It is only relevant when the tip radius (or mathematically generated radius) is used as wavelength limitation.

Gaussian or Ideal Wavelength Limitation

When applying the Gaussian or *Ideal* approaches for the limitation of short wavelengths (the unified ⌠ or ⌡ designations), a minimum of 5 data points per cutoff wavelength are required to obtain at least 95% amplitude transmission per ISO 3274 (1995). Thus the unified methodology maintains this established practice. It should be noted once again, that when using the Gaussian or *Ideal* wavelength

limitation approaches, the stylus tip effect should be adequately outside the pass band. Given the proper selection of the stylus the sampling approach of 5 data points per cutoff wavelength supersedes the Equation 5.17 requirements based on tip radius. The Gaussian or *ideal* approaches should not be influenced by the local "cusps" of the stylus.

5.4.2 Long Wavelength Limitation

Long wavelengths can be attenuated by the optional application of Gaussian or *Ideal* filters. In the measurement of linear profiles per the unified methodology, the minimum trace length is based the shorter of the following:

- Long enough to obtain 5 times the cutoff length after filtering.
  Typically Gaussian or Ideal filtering algorithms discard approximately 0.5 or 1.0 times the cutoff at each end of a linear trace. (Based on historical and standardized practice in roughness analyses.)

- The entire length of the component.
  If less than one filter cutoff length remains after discarding the necessary amounts at the ends of the trace, the filter cutoff is deemed to be too long.

In all cases where a Gaussian or *ideal* wavelength limitation is specified, at least one half of the cutoff length should be discarded on each end of the trace to avoid filter end effects. (See the discussion below concerning end effects.)

The unified scheme also provides for the omission of a long wavelength limitation. Under these circumstances, for a linear assessment the entire length the surface must be measured and analyzed as one contiguous data set (based on leveling via the least squares reference line).

In the measurement of circular profiles, the omission of a long wavelength limitation (specified in terms of frequency) implies the inclusion of all frequencies higher than 1 upr (based on the least squares reference figure) up to the frequency of the mandatory low-pass filter cutoff. Furthermore, given the *closed* nature of circular data sets, there is no need to discard data points due to filter end effects.

5.5 Treatment of End Effects

The need to discard data at the ends of a linear trace is often viewed as undesirable in engineering applications - particularly in those applications where the edges are critical in the components function. One such example would be a system where two surfaces must form a sealing interface and any raised material along the edges of the surfaces could result in poor sealing. Similar considerations may arise in the assessment of the straightness over the entire length of a component.

These applications are special cases in surface metrology and have been addressed through various techniques including mirroring, Fourier wrapping, self-approximation and the fitting of splines. While these methods may be individually suited for certain profile features, none of them are well suited for general application. They are mentioned here for reference and may be applied at the user's discretion and with appropriate documentation apart from the unified specification format.

5.5.1 "Mirroring" of Data Sets

One of the simplest means of avoiding the need to discard data at the ends of a linear profile is to extend the profile prior to filtering. This is accomplished in some commercially available instruments through the *mirroring* or *folding* of data points along the ends of a data set (Tabenkin 1991).

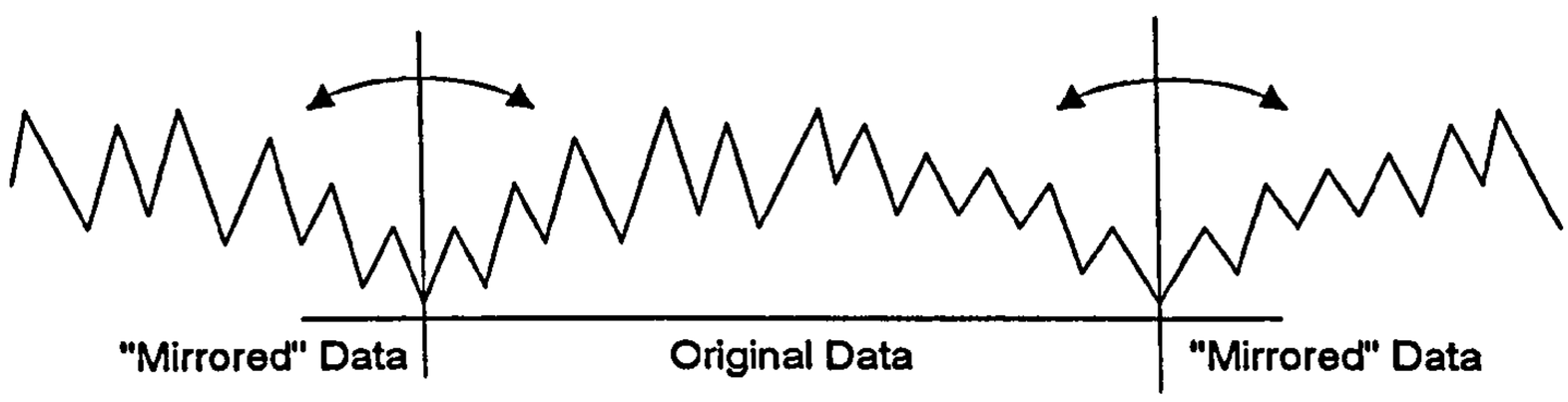


Figure 5.20 *"Mirroring" of a data set.*

This methodology may be appropriate for profiles which are nominally linear at the edges, however significant errors can be induced if there is any local curvature near the edges of the profile. Upon mirroring or folding data sets with local curvature near the ends, a cusp is formed and the subsequent filter can be affected adversely.

5.5.2 Fourier Wrapping of Data Sets

When performing Fourier analyses the mathematics assume an infinitely repeating series. While this is the case in the measurement of a full circular profile, it is not the case in linear measurement. However, this technique of wrapping the data set provides a means of extending the ends of the profile in an attempt to avoid the discard of end regions upon filtering (Press et al. 1992).

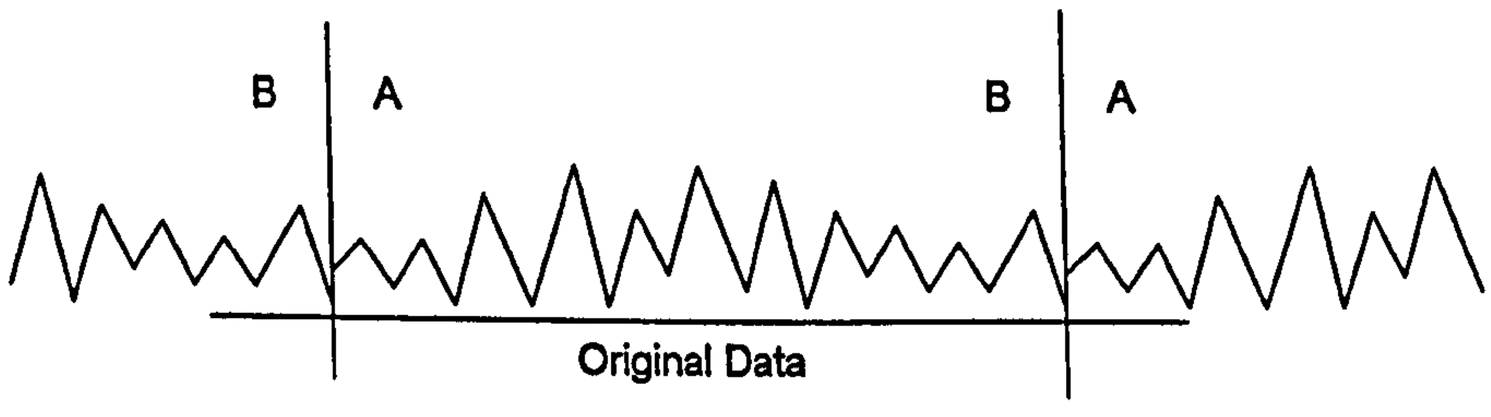


Figure 5.21 *Fourier "wrapping" of a data set.*

The wrapping approach is very similar to the mirroring approach in terms of applications and shortcomings. One primary concern in the application of the wrapping technique is the potential for causing a discontinuity at the real profile ends.

In cases where a significant discontinuity is present, the subsequent filter will be adversely affected. These discontinuities can be removed by inclining the data set such that the endpoints are coincident (i.e. "closing" the data set), however this can result in a significant modification of slopes throughout the data set.

5.5.3 Self-Approximation of Data Sets

A more complex means of establishing a filtered profile without the discard of profile data can be accomplished through the use of central profile data to approximate the behavior of the profile at the edges (Mestre and Abou-Kandil 1993).

Two application methodologies should be considered in terms of self-approximation: 1. Profile extension via roughness estimation (Figure 5.22a) and, 2. Direct waviness estimation (Figure 5.22b). In the first methodology, the central region of the profile is searched for a region which closely resembles the edge region. Upon selecting the optimal central region, the data which follows this optimal region is copied as an extension of the edge region. In the waviness application of self-approximation, the previously described optimal central region is filtered and the resulting waviness profile is copied directly to the edge region which would normally be discarded. This technique bears some similarity to fractal analyses which can also use measured data as predictor of the unmeasured portions (Russ 1994).

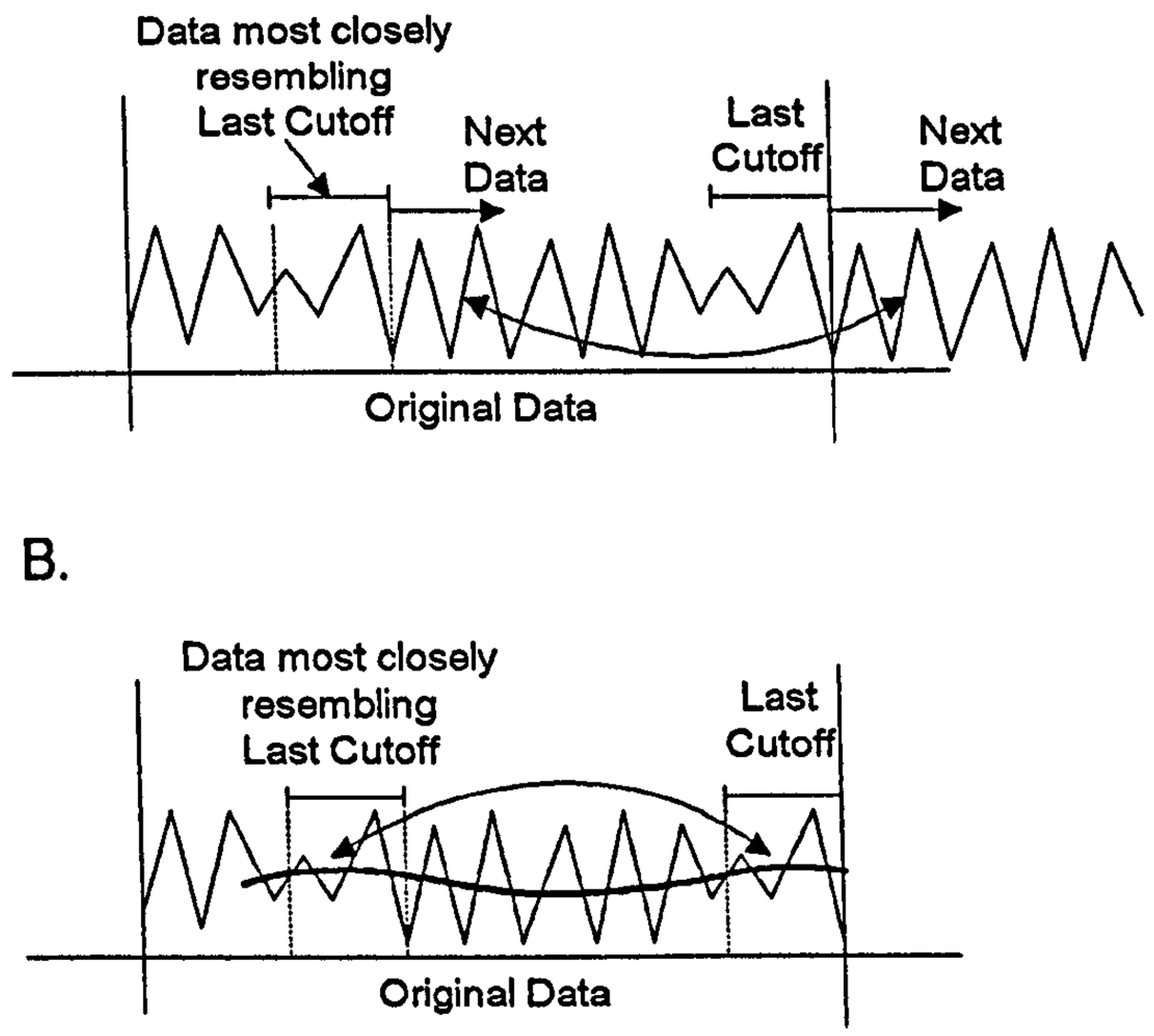


Figure 5.22 *"Self Approximation" approaches for treating end effects.*

This self-approximation technique appears to be robust, particularly in profiles which are rather homogeneous. However, self approximation can yield undesirable results in cases where the central profile region does not contain an area which is similar to the end region. An example of such a profile would be a nominally linear profile with abrupt curvature at each end.

5.5.4 Splines in Long Wavelength Assessment

It has been proposed the splines (such as cubic) be utilized as long wavelength profile figures (Krystek 1995). The fitting of splines differs from filters in that there is no weighting function, and therefore no requirement for the discarding of data.

Splines have become very popular in the modeling of three dimensional surfaces, particularly through the application of Non-Rational Uniform B Splines (NURBS) in reverse engineering (Yau and Menq 1992, Yau 1995). However, the application of splines in surface profile analyses can induce varying edge effects depending on the type of constraints used on the spline. (For example the approach proposed by

Krystek was significantly influenced by the position of the profile endpoint.) Furthermore, the typical wavelength transmission characteristics of splines are not well defined and can vary depending on the data set - particularly near the ends of the profile.

5.5.5 End Effects in the Scheme for Unification

Given the above discussion, the proposed unified approach to surface metrology does not directly endorse a means for avoiding the discard of data at the ends of a profile. The above methodologies may be applicable to certain situations and therefore their use is not forbidden. However, any use of these methods in a particular specification or analysis must be specifically noted in order to maintain the highest level of reproducibility for the measured results. In the event that a padding approach is required and none of the above described means are acceptable, alternatives may be developed and appropriately documented. In cases, where no methodology seems appropriate, the profiles should be analyzed with minimal short wavelength filtering and no long wavelength filtering.

5.6 Summary

Wavelength limitation is an essential element in the scheme for unifying applied surface metrology. This chapter has provided a rather brief overview of the topic from an historical perspective as well as a detailed presentation of the approaches accommodated in the scheme. Special attention has been given to the numerical implementations for filtering as the technique employed can have technical as well as economic implications.

As stated in the introduction to this chapter, much of the above material is already published. However many of the details presented above regarding implementation, comparative behaviors and end effects has not been published. Thus, the material in

this chapter becomes essential supporting material for the practitioner of surface metrology.

A Unified Methodology for the

Application of Surface Metrology:

Chapter 6

A Unified Approach to Parameterization

Chapters 3, 4 and 5 have presented information essential in describing and controlling wavelength content of surface metrology data sets. This is of utmost importance in obtaining correlation between various instruments, however further processing is still necessary in order to generate numerical parameters which describe the data set. In one sense, the band limited data set is the complete description of the surface, however this is not practical in terms of specification and communication.

Parameters are a necessary element in applied surface metrology in that they are intended to characterize some degree of functionality whereby a single number (or small set thereof) can synopsize the many thousands of points often generated in a measurement. In light of this aspect of parameterization, it has been quoted: "*A picture may be worth a thousand words, but I'm an engineer - give me a number,*" (Brown 1995).

Surface metrology parameters are necessary in commerce (Nielsen 1996a) and standardization (ISO/TR 14638 - 1995) as they provide the *language* for describing surface attributes. In terms of commerce, this *language* is necessary in determining tolerances, controlling processes and ultimately accepting or rejecting components. In the context of standardization, this defined *language* is of highest importance - particularly regarding the relationships or "chain of standards" from a drawing indication through the measurement process to the ultimate acceptance or rejection of a measured component.

In this chapter, the topic of parameterization will be discussed in light of the overall scheme for unification. Since a great deal has already been published concerning specific parameters in surface metrology, the chapter will focus on the broader topic of parameterization and the "unification" of surface metrology parameters. In a sense, surface metrology parameterization has undergone the same "divergence" as Chapter 1 has described regarding surface metrology instrumentation (Whitehouse 1982). Many surface metrology parameters have been developed in various, unique applications and

these developments continue today. Meanwhile, existing parameters remain in instrumentation, documentation and standards. This chapter presents a system for evaluating the necessity of surface metrology parameters whereby, unnecessary parameters would be omitted and future developments of necessary parameters can be accommodated by a "unified parameter set".

Finally, this system will be employed in the determination of a proposed set of parameters for the unified parameter set. The development of this proposed set is intended to only serve as the catalyst for a dialog within the surface metrology standardization community. On the whole, the material presented in this chapter is intended to be most useful to a group such as a standardization body as they should ultimately control the inclusion or rejection of parameters for use in commerce.

6.1 Parameters in Surface Metrology

Surface metrology evaluations have evolved from visual interpretations of profile graphs to graphical or analog/electronic processing and more recently to digital processing of electronically stored profiles. At each stage of advancement, additional parameters were introduced and further opportunities were provided for even more parameters. With digital computer technology, several hundreds of parameters have been proposed and/or studied in terms of their potential application as means of characterizing some attribute of a surface (Whitehouse 1982). Today, the number of parameters continues to increase. Furthermore, with the introduction and growing acceptance of new processing techniques such as wavelets (Chen et al. 1994), fractals (Russ 1994) and Artificial Neural Networks (Vemuri 1988) there is a great potential for significant increases in the number of parameters.

This increase in available parameters is not necessarily a *bad* thing for the field of surface metrology. In one sense, it tends to indicate that people are investigating various attributes of surface data sets that cannot be adequately described via

traditional approaches (Malburg and Raja 1993, Whitehouse 1993). However, on the other hand, many *new* parameters are being developed which are merely ratios or combinations of existing parameters and thus their necessity should be questioned (Wasilesky 1994).

Several issues must be taken into consideration in any discussion concerning the parameterization of surface profiles:

- Numerical parameters are the means by which judgments are made. Visual assessment of profiles can provide intuitive information which, when coupled with experience, can drive an appropriate action. However, numerical data can drive "data-based" engineering decisions.

- Surface functionality varies significantly depending on the application that the surface is placed in. Thus a parameter, which may be very useful in predicting performance in one application, may not correlate to surface performance in other applications.

- Increasing numbers of available surface metrology parameters can accompany an increased knowledge of surface functionality. This increase in knowledge does not come without a cost. Instrument manufacturer's are in a mode of constantly adding to their analysis software and the engineering (i.e. surface metrology user) community continues to struggle to maintain a grasp of the many different surface characterization alternatives at their disposal.

- Historical approaches for characterization may deserve special consideration in light of existing databases and other economic implications. While the field of surface metrology may be relatively young when compared to some sciences, there are still rather significant databases which have been developed based on data from historical parameters. Nonetheless, all parameters should be

critically evaluated and efforts should be made to transition away from inappropriate parameters even if there is historical significance.

6.2 Application of Parameters

The applications of surface metrology have led to a rather *isolated* situation in that there are parameters which are defined for roughness, parameters which are defined for waviness, parameters which defined for roundness, and so on. The unfortunate part of this is that there is a general view that these analyses are somehow different and unique based on their wavelength regime or nominal geometry (see, for example Chapter 1, Table 1.1, or B46.1 - 1995).

The scheme for the unification of surface metrology establishes a new paradigm whereby parameters, are not tied to any specific geometry or wavelength regime. Numerical parameters are merely the result of data processing techniques which can be applied across varying geometries and wavelength regimes.

6.2.1 Parameters for Linear Analyses

Historically, the analysis of nominally linear surface profiles has resulted in the vast majority of parameters. Hundreds of parameters have been developed for describing various aspects of linear profiles. In many cases, multiple parameters have been defined based on the same analysis being performed on differing ranges of wavelength content. (For example, the average absolute amplitude is referred as *Ra*, *Wa* and *Pa*, indicating roughness, waviness and primary (unfiltered) profile assessments respectively.) In the proposed scheme for unification, the type of profile (in terms of band limitation and nominal geometry) is removed from the definition of the parameter. The parameters are merely defined in terms of a means of processing or an algorithm which can be applied to any profile.

Although the parameters presented can be applied to all wavelength regimes, the inherent functionality of a given parameter will still depend on the application of the surface (see, for example Thomas and Sayles 1978, Thomas 1982). To say that all parameters will provide meaningful information across all wavelength ranges would be an erroneous extrapolation of the unification of surface metrology. The unified scheme only serves to make the necessary provisions for establishing and maintaining a manageable set of *generic* parameters which *can* be broadly applied. The successful application of a given parameter depends on factors such as knowledge of the application, history, modeling and experimentation.

6.2.2 Parameters for Circular Analyses

The analysis of nominally circular geometries has typically been based on a single parameter characterizing the total (peak-to-valley), departure from a perfect, round geometry. More recently this has expanded to include a "rate-of-change" parameter as well as peak, valley and RMS parameters (ISO 12181 - 1995). In a sense, this shows the parallel between circular and linear analyses (both in terms of the growth in parameters and the specific parameters themselves). On the other hand, it also presents additional parameters to be managed by the surface metrology community.

6.3 The Unified Parameter Set

The development of a unified parameter set is based on 1.) the establishment of a categorization scheme, whereby 2.) guidelines can be established for the incorporation or exclusion of parameters, and finally 3.) the parameters can be designated via a unified nomenclature - independent of nominal geometry and wavelength domain. Furthermore, the approach undertaken should be extensible in order to accommodate new developments in parameterization as well as provide mechanisms for the removal of obsolete parameters.

These areas will be discussed below in the development of a proposed "unified parameter set". It should be understood that underlying *problem* relating to the surface metrology parameters has long been recognized and documented (Whitehouse 1982). However, the *methodology* for systematically addressing it, presented in the following sections, is original. Furthermore, these sections conclude with a *product* - that being a proposed unified parameter set.

This proposed unified parameter set serves as an example application of the *methodology*. While this *methodology* for determining a unified parameter set is reliable, it is dependent upon the underlying notions of surface functionality which are incorporated. Thus, the appropriate forum for the determination of a unified set of parameters would be a standardization body as broader interests in surface functionality would be represented. Nonetheless, the methodology developed and presented here should be undertaken by the appropriate standardization body and the proposed parameter set (included in this chapter) should serve as a reasonable starting point.

6.3.1 The Categorization of Parameters

In order to determine a unified set of surface metrology parameters, it is important to establish groupings or categories whereby parameters describing similar aspects can be compared. In this regard, potential categorization approaches could be based on aspects such as "functionality", "mathematical implications" or "geometry".

While surface functionality may be the end goal to be achieved through parameterization, it is often very difficult to directly relate it to specific parameters. Furthermore, many different functionalities can be based on the same parameters, thus leading to multiple representations of parameters across different categories.

Another alternative would be some mathematically based categorization of parameters. This too, would result in added confusion as many parameters can be calculated by

multiple mathematical methods. For example, the root mean squared (RMS) amplitude can be determined algebraically or via Fourier Transformation.

Finally, we are left with a "geometry based" categorization of parameters. This approach seems to be the most appropriate in that the vast majority of surface metrology parameters tend to describe some geometric aspect of the data set - independent of the mathematics being used or the function for which the surface was intended. The categorization chosen is summarized as follows:

- ***Statistical (or "averaging") amplitude parameters.***
  Parameters which characterize the "composite" nature of a data set in terms of vertical displacements, whereby all data points are included and combined in some manner.

- ***Extreme amplitude parameters.***
  Parameters which characterize the vertical separation of extreme profile features (i.e. minima and maxima).

- ***Spacing parameters.***
  Parameters which characterize the lateral (or horizontal) relationships between profile features.

- ***Slope and "shape" parameters.***
  Parameters which characterize the local relationships between vertical and horizontal aspects of the data sets.

- ***Auxiliary functions and parameters.***
  Parameters based on some transformation of the data set to a different "domain" (some examples include sorting or Fourier Transformation).

This categorization is very similar to the approaches which has been described by surface texture instrument manufacturers (Dagnal 1980, Amstutz 1985, Mummery 1992, Rank Taylor Hobson 1995) as well as other authors and standards bodies (Vorburger 1993, Stout et al. 1993, Whitehouse 1994, ISO/DIS 4287 - 1995, ASME B46.1 - 1995).

6.3.2 Guidelines for Parameter Inclusion

As stated above, the "unified parameter set" would ideally be controlled by a standardization body. However, the following guidelines should serve as the basis (to be used by the standardization body) for the incorporation or exclusion of a parameter.

1. Parameters will be grouped into categories based on the geometric attributes of the data set which they most closely quantify. The establishment of a new category should only be undertaken in extreme cases.

2. Parameters will be evaluated within their respective category. In the event that multiple parameters are deemed necessary within a category, a primary parameter should be determined and the remaining parameters will be designated as "auxiliary". (Note: given the diverse nature of auxiliary functions no primary parameter will be designated within this category.)

3. Parameters must be describable as an equation or algorithm.

4. Parameters must demonstrate all of the following primary characteristics in order to be considered for inclusion in the unified parameter set. The selection of a primary parameter (within a category) will be based on the highest degree satisfaction of the primary criteria. Secondary criteria should be considered if a consensus cannot be reached based on primary criteria.

Primary Criteria

i) The parameter must be independent of other parameters or combinations of parameters already included in the scheme across all categories. (This implies that the early development of the scheme may be somewhat iterative as parameters are added and removed.) Independence can typically be demonstrated through changing the type of surface (for example, the *mean peak-to-valley* may be approximately 12 times the *average amplitude* for plateau honed surfaces, but only 7 times for ground profiles). The only exception to this rule are cases where Secondary Criterion *a* or *b* is adequately satisfied.

ii) A parameter must demonstrate functionality in that is correlates to some type of physical phenomenon. This can be established based on a manufacturing process attribute or a functional behavior.

iii) Parameters must be stable in a numerical sense, whereby their results can be easily reproduced in subsequent re-measurements or re-evaluations.

iv) Parameters must definable in an algorithmic or mathematical sense. This is necessary in the evaluation of numerical approaches for calculating the parameter.

v) The parameter must be applicable to any bandwidth and any nominal base geometry. It is acceptable to change units (for example from linear to polar) depending on the base geometry.

Secondary Criteria

a) A parameter may be included if there is a clear connection to a "family" of parameters. For example, the total "peak to valley" height of a data set can be determined when given data for the highest peak and the deepest valley. Thus, the total peak to valley is redundant in light of the other two. However, for economic reasons is makes sense for parameter set to directly provide this value as part of the "family" of parameters.

b) A parameter may be included based on significant historical or economic implications. If this criterion is used to override primary criterion *i*, then parameter should be made obsolete.

c) A parameter may be included based on mathematical relevance or importance in the modeling of surface behavior.

5. The unified parameter set should be based on a selection of parameters which cover the broadest base of applications. In general these parameters should be based on the Pareto Principle - "80 percent of the results flow from 20 percent of the activities". Thus, the relatively few parameters contained in this scheme may not be suitable for *all* applications, but they are certainly well suited for the majority of applications.

6. An interval must be established whereby the unified parameter set is reviewed for the purposes of adding new parameters and removing obsolete parameters.

6.3.3 Parameter Nomenclature

Under the unified scheme, surface metrology parameters do not contain a leading character which indicates the type of profile from which it was calculated. The type of profile is pre-determined through designation of the nominal geometry (i.e. linear or circular) and the desired wavelength limitation.

In this scenario, the parameter merely designates the type of calculation to be performed on the data set. This notion is directly in line with the general concepts for unification - just as the measurement should be independent of instrumentation (providing the necessary requirements are met) so should the parameter calculation be independent of the type of data (providing the necessary requirements are met).

For parameters with significant historical or current utilization, the parameter designation should be based as closely as possible on current practice (with the removal of any leading character which designates a type of profile). For example, the average amplitude parameter for roughness "*Ra*" would be represented in the unified parameter set as simply "*a*" indicating the average absolute amplitude parameter as calculated from the desired profile.

6.4 A Proposed Unified Parameter Set

In the following sections, the above guidelines will be used to generate a proposed "unified parameter set". This exercise will be based on many of the parameters which available in current instrumentation (Feinprüf Perthen GmbH 1992, Mummery 1992, Federal Products 1994, Rank Taylor Hobson 1995). The parameters will be categorized and evaluated per the above guidelines.

It is important to note that the methodology outlined below is not intended to provide a great deal of tutorial information regarding specific the parameters and their

applications - this exercise is intended to demonstrate the procedure of establishing a unified parameter set. The "tutorial" type of information regarding surface parameters and functions is readily available in books such as those written by Amstutz (1985), Mummery (1992) and Whitehouse (1994). In addition, many less technical articles been written on the topic (for example, Lavoie 1991, 1992) and tutorials are frequently made available through instrument manufacturers or professional organizations such as the American Society for Precision Engineering (ASPE) or the American Society for Quality Control (ASQC) (for example Vorburger 1993).

For the discussion that follows, several parameters will be considered (based on their current designation):

- *Statistical (or "averaging") amplitude parameters.*
  Ra, Rq, Rsk, Rku

- *Extreme amplitude parameters.*
  Rp, Rv, Rt, Rpm, Rvm, RzDIN, RzISO, R3z, Wc

- *Spacing parameters.*
  Sm, S, Pc, HSC

- *Slope and "shape" parameters.*
  Δa, Δq, $dr/d\Theta$

- *Auxiliary functions and parameters.*
  λa, λq, tp, β^*

Parameters which are adopted in to the proposed "unified parameter set" will be given a unified parameter designation.

For mathematical parameter definitions, the following variables will apply:

x, Θ	the instantaneous position along a straight or circular reference for a continuous (analog) profile.
x_i, Θ_i	the instantaneous position along a straight or circular reference geometry (respectively) for a sampled (digital) surface.
Δx, $\Delta\Theta$	the data point spacing along the reference figure (assumed regular).
z	the amplitude at a given, instantaneous position along a continuous (analog) linear profile.
z_i	an instantaneous deviation from the nominal linear geometry in a digital data set. Negative values indicate points lying on the material side of the reference feature.
r	the amplitude at a given, instantaneous position along a continuous (analog) circular profile.
r_i	an instantaneous deviation from the nominal circular geometry in a digital data set. Negative values indicate points lying on the material side of the reference feature.
σ	the standard deviation of amplitudes about the reference figure.
l	the length of a continuous (analog), linear profile.
n	the number of data points sampled in a given profile.

Note: Unless otherwise specified, the parameter equations will be given for the linear implementation and the circular implementation can be determined by a simple exchange of variables "x" to "Θ" and "z" to "r".

6.4.1 Statistical Amplitude Parameters

A very common methodology for assessing deviations about a nominal (or reference surface) is to statistically quantify the individual deviations which are represented as sampled data points. These parameters are useful in the widest variety of short-wavelength characterization applications and are often utilized when developing theoretical models of surface roughness interactions. Examples of their use include the characterization of many machined surfaces as well as incorporation in the mathematical models of contact and lubrication effects (Lavoie 1992).

Ra: Average absolute amplitude. The arithmetic average (absolute) departure from the reference figure (linear or circular).

$$Ra = \frac{1}{n}\sum_{i=1}^{n}|z_i| \tag{6.1}$$

Rq: Root mean squared (RMS) amplitude. The RMS departure from the reference figure. For large numbers of samples (as is common in surface metrology data sets), the RMS deviation is approaches the standard deviation, σ, of profile heights.

$$Rq = \sqrt{\frac{1}{n}\sum_{i=1}^{n}z_i^2} \tag{6.2}$$

Rsk: Skewness. The skewness of the profile height (amplitude) distribution.

$$Rsk = \frac{1}{n \cdot Rq^3} \sum_{i=1}^{n} z_i^3 \tag{6.3}$$

Rku: Kurtosis. The kurtosis of the profile height (amplitude) distribution.

$$Rku = \frac{1}{n \cdot Rq^4} \sum_{i=1}^{n} z_i^4 \tag{6.4}$$

Para-meter	Primary Criteria					Secondary Criteria			Include (Y/N)	Desig-nation
	independence	functionality	stability	mathematical	generic	family	economic	modelling		
	i	*ii*	*iii*	*iv*	*v*	*a*	*b*	*c*		
Ra	?	X	X	X	X		X		Y	a
Rq	X	X	X	X	X	X		X	Y	q*
Rsk	X	X	X	X	X	X		X	Y	sk
Rku	X	X	X	X	X	X		X	Y	ku
* indicates primary parameter										

Table 6.1 Statistical or "averaging" amplitude parameters for the proposed unified parameter set.

Notes: Although mathematically, the average absolute deviation (a) differs from the RMS (q), in a practical sense a follows q rather closely and therefore its independence in application is questionable. Nonetheless, the historical significance of "Ra" and the economics associated with its growing database dictate that it be included in the unified parameter set.

6.4.2 Extreme Amplitude Parameters

Extreme amplitude parameters perform a numerical quantification of the specific local minima and maxima. These parameters are typically used in the quantification of surfaces in light of their "clearance consumption" in an interface. They are most commonly utilized in long wavelength characterizations where added repeatability and reproducibility can be expected.

In short wavelength characterizations, the extreme peaks are rather unstable as they can be the result of debris or contaminants (see Chapter 4) or they can be easily distorted by contact. Thus, in the assessment of short wavelength extreme amplitudes, the averaging or *mean* peak, valley and total parameters are most commonly applied.

Examples where extreme amplitude parameters are applied include the assessment of roundness or waviness in sliding or rolling interfaces, or the consumption of a clearance in a critical *fitting* application (Yhland 1967, Grant 1991).

Rp: Height of highest profile peak. The distance from the reference figure to the most extreme outward point.

$$Rp = \max(z_i) \tag{6.5}$$

Rv: Depth of lowest profile valley. The distance from the reference figure to the most extreme inward point. (Note: the *Rv* value is positive as it represents a *depth*, not a *vertical position* relative the mean line.)

$$Rv = \left|\min(z_i)\right| \tag{6.6}$$

Rt: Total (peak to valley) profile height. The distance from the highest profile peak to the lowest profile valley.

$$Rt = Rp + Rv \tag{6.7}$$

Rpm: Mean peak height. The average height of peaks whereby each single highest peak is obtained from each section of profile corresponding to the long wavelength limitation. The profile sections are each one cutoff wavelength long for both Gaussian and *ideal* types of wavelength limitation. For circular analyses with no specified low frequency cutoff or for linear analyses with no specified long wavelength cutoff, the *Rpm* parameter is equal to the *Rp* parameter.

Thus, given a profile of k consecutive sections, we have p_j profile peaks (in either analog or digital data sets). Where: j=1, 2, ..., k.

$$Rpm = \frac{1}{k}\sum_{j=1}^{k} p_j \tag{6.8}$$

Rvm: Mean valley depth. The average depth of valleys whereby the single deepest valley is obtained from each section of profile. (See *Rpm* discussion above.)

$$Rvm = \frac{1}{k}\sum_{j=1}^{k} v_j \tag{6.9}$$

RzDIN (also Rtm): Mean peak to valley (total) deviation. The average peak to valley distance based on one peak-to-valley distance per profile section. (See *Rpm* and *Rvm* discussions above.)

$$RzDIN = \frac{1}{k}\sum_{j=1}^{k}(p_j + v_j) \tag{6.10}$$

or

$$RzDIN = Rpm + Rvm \tag{6.11}$$

RzISO: Mean peak to valley (total) deviation independent of location. The average peak to valley distance based on the five highest profile peaks and the five deepest profile valleys regardless of their location within the data set.

$$peak_{mean} = \frac{1}{5}\sum_{j=1}^{5} p_j \tag{6.12}$$

$$valley_{mean} = \frac{1}{5}\sum_{j=1}^{5} v_j \tag{6.13}$$

$$RzISO = peak_{mean} + valley_{mean} \tag{6.14}$$

R3z: Peak to valley height of third highest peak to third deepest valley. The peak to valley distance based on discarding the highest two peaks and the lowest two valleys.

$$R3z = peak_3 + valley_3 \tag{6.15}$$

Wc: Composite peak to valley. (Currently defined only for waviness.) The average peak to valley separation including all profile peaks and valleys.

$$Wc = \sum_{j=1}^{m} \left(peak_j + valley_j \right) \tag{6.16}$$

Para-meter	Primary Criteria					Secondary Criteria			Include (Y/N)	Desig-nation
	independence	functionality	stability	mathematical	generic	family	economic	modelling		
	i	*ii*	*iii*	*iv*	*v*	*a*	*b*	*c*		
Rp	X	X	X	X	X	X	X	X	Y	p
Rv	X	X	X	X	X	X	X	X	Y	v
Rt		X	X	X	X	X		X	Y	t*
Rpm	X	X	X	X	X	X			Y	pm
Rvm	X	X	X	X	X	X			Y	vm
RzDIN		X	X	X	X	X	X		Y	tm
RzISO	?	X	X	X	X				N	
R3z	?		X	X	X				N	
Wc	?		X	X	X				N	
* indicates primary parameter										

Table 6.2 *Extreme amplitude parameters for the proposed unified parameter set.*

Notes: The proposed extreme amplitude parameter set consists of two parameter families - one of which incorporates some degree of averaging and the other does not. Given the control of the wavelength content as dictated by the scheme for unification, the independence of both *RzISO* and *R3z* are questionable in practice. Furthermore, for most surfaces the composite peak to valley *Wc* will follow the average peak to valley (*tm*).

6.4.3 Spacing Parameters

The previously described categories of parameters were amplitude-dependent and thus were insensitive to the spacing or variation in spacing of surface features. The following parameters have been developed to specifically characterize the "horizontal" aspects of surface features. This type of parameter has been applied in describing functional aspects of surfaces related to coatings or paint adherence (Drews 1994) and can also be useful in the characterization of some wear and chatter mechanisms as well as sealing properties (Lavoie 1994).

Sm: Average (mean) spacing of positive mean line crossings. The spacings between all crossings of the mean line (negative to positive) are averaged regardless of feature heights.

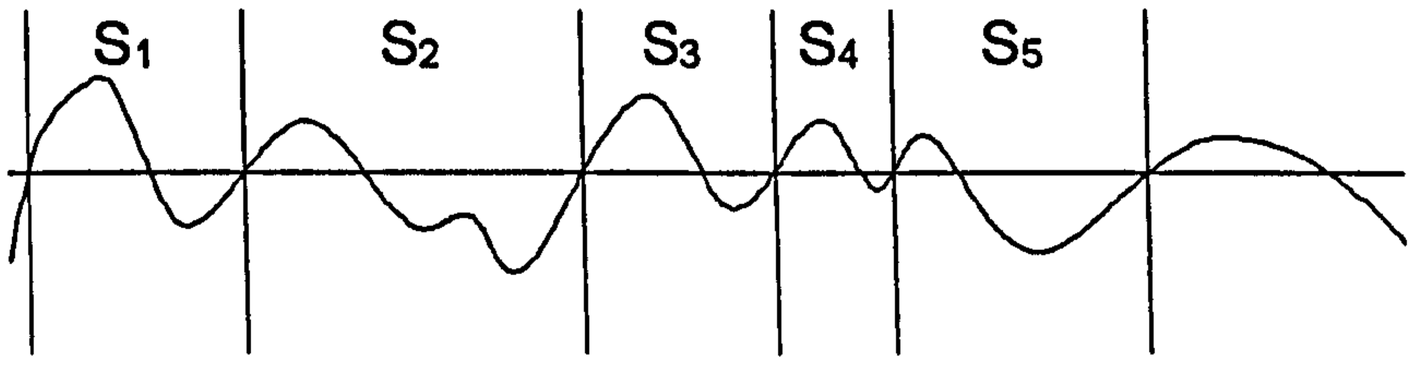


Figure 6.1 *Positive mean line crossings and associated spacings.*

Thus for a profile which contains j positive mean line crossings, let S_i be the distance between mean line crossings occurring at i and $i+1$, then

$$Sm = \frac{1}{j-1}\sum_{i=1}^{j-1} S_i \qquad (6.17)$$

S: Average spacing of profile peaks. The spacings between all profile peaks are averaged.

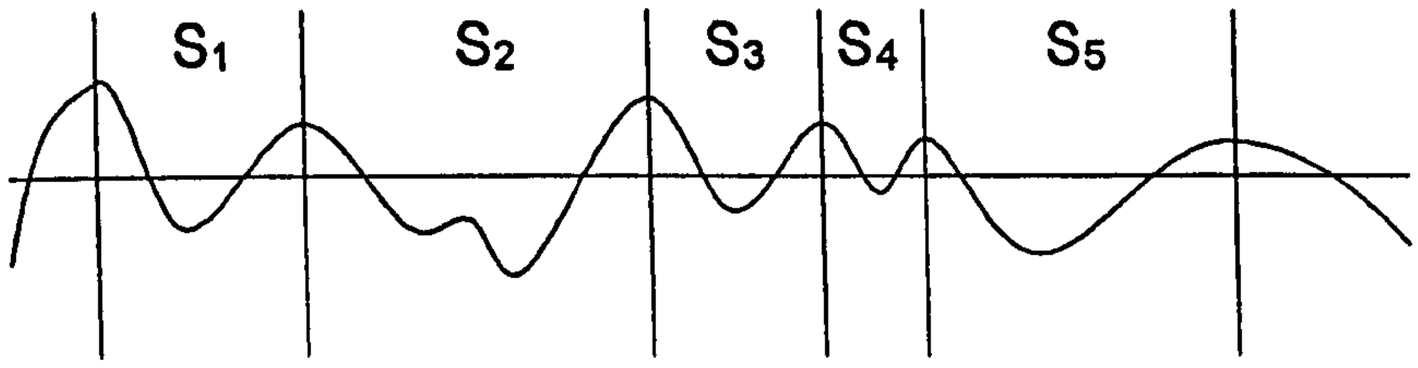


Figure 6.2 *Mean peak spacing.*

Thus for a profile which contains j peaks, let S_i be the distance profile peaks occurring at i and $i+1$, then

$$S = \frac{1}{j-1}\sum_{i=1}^{j-1} S_i \tag{6.18}$$

Pc: Peak count. The density of peaks along the profile whereby a peak is determined by the crossings of an upper threshold, *c1*, and a lower threshold, *c2*. The settings of the threshold values, *c1* and *c2*, are independent and can fall above or below the mean line. The value is reported as peaks/mm or, in circular analyses, peaks/degree.

HSC: High Spot Count. The number continuous profile (material) sections intersected by a threshold line.

Para-meter	Primary Criteria					Secondary Criteria			Include (Y/N)	Desig-nation
	independence	functionality	stability	mathematical	generic	family	economic	modelling		
	i	*ii*	*iii*	*iv*	*v*	*a*	*b*	*c*		
S	?	X	?	X	X					
Sm	X	X	X	X	X				Y	sm
Pc	X	X	X	X	X				Y	pc*
HSC		X	?	X	X					
* indicates primary parameter										

Table 6.3 Spacing parameters for the proposed unified parameter set.

Notes: The peak spacing (*S*) parameter is closely correlated to mean line crossing spacing (*sm*) for data sets made up predominantly of long wavelengths. For data sets which are dominated by short wavelengths the peak spacing becomes unstable. The high spot count (*HSC*) parameter can be related very closely to the peak count (*pc*) parameter depending on the setting of the thresholds. Furthermore, the stability of the HSC parameter may be questioned in light of relatively high frequency artifacts repeatedly crossing the threshold.

6.4.4 Slope and Shape Parameters

The above described vertical and spacing parameters address individual aspects of surface profile features. The next degree of complexity is the characterization of how the horizontal and vertical aspects of the surface combine.

The characterization of the distribution of local slopes is useful in assessing optical properties of surfaces as well as in the characterization of micro-burrs and surface cleanliness (Bennett and Mattsson 1989, Appendix C). The characterization of larger

scale departures is very useful in describing a surface in light of follower dynamics, mating surface conformability or sealing applications (Bhargava 1991, Hager 1995).

Δa: Average Absolute slope. The average slope of the surface profile. To adequately define the *Δa* parameter in a digital implementation, the sampling interval Δx, or ΔΘ, must be specified. Given the specified sampling interval, the local derivative obtained from each ordinate is computed based on the seven point Lagrangian approach presented by Chetwynd (1978) which has recently been standardized (ISO/DIS 4287 - 1995, ASME B46.1 - 1995).

Note: For the linear implementation, three points on each end of the data set must be discarded. In circular implementations, the closed or "wrapping" nature of the data can be utilized to compute a local derivative at each of the *n* data points.

(linear)
$$dz_i/dx_i = \frac{[z_{i+3} - 9z_{i+2} + 45z_{i+1} - 45z_{i-1} + 9z_{i-2} - z_{i-3}]}{60 \cdot \Delta x} \tag{6.19}$$

(circular)
$$dr_i/d\Theta_i = \frac{[r_{i+3} - 9r_{i+2} + 45r_{i+1} - 45r_{i-1} + 9r_{i-2} - r_{i-3}]}{60 \cdot \Delta\Theta} \tag{6.20}$$

Given these definitions of local slope, the average slopes are given in equations 6.21 and 6.22 respectively. Note: given the application of a least squares reference figure and neglecting the difference in endpoint amplitude for linear profiles, a zero mean derivative is assumed.

(linear)
$$\Delta a = \frac{1}{(n-6)} \sum_{i=4}^{n-3} \left| dz_i / dx_i \right| \tag{6.21}$$

(circular)
$$\Delta a = \frac{1}{n} \sum_{i=1}^{n} \left| dr_i / d\Theta_i \right| \tag{6.22}$$

Δq: Root mean square slope. The RMS slope of the surface profile. For large samples, the RMS slope approaches the standard deviation of slopes.

Given the above definitions of local slope (Equations 6.19 and 6.20), the RMS slopes are given in equations 6.23 and 6.24 respectively. Note: given the application of a least squares reference figure and neglecting the difference in endpoint amplitude for linear profiles, a zero mean derivative is assumed. Furthermore, the data point spacing must be specified as it can significantly influence local slope calculations.

(linear)
$$\Delta q = \sqrt{\frac{1}{(n-6)} \sum_{i=4}^{n-3} \left(dz_i / dx_i \right)^2} \tag{6.23}$$

(circular)
$$\Delta q = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(dr_i / d\Theta_i \right)^2} \tag{6.24}$$

dr/dΘ: Maximum rate of change in a defined window. The worst case peak to valley departure encountered in a window of specified width as the window is moved through the data set. While these analyses are typically performed on circular data sets, the same concept can be applied to linear profiles. The computation is based on moving a window of prescribed width (Θ or x) through the data set for the purpose of obtaining the greatest peak to valley distance contained within the window regardless of location within the profile.

Note: The $dr/d\Theta$ notation historically applied to this parameter is misleading in that the spacing between the peak and valley (in terms of a $d\Theta$) is not included in the parameter calculation. The numerical result is simply the peak to valley distance.

Para-meter	Primary Criteria					Secondary Criteria			Include (Y/N)	Desig-nation
	independence	functionality	stability	mathematical	generic	family	economic	modelling		
	i	*ii*	*iii*	*iv*	*v*	*a*	*b*	*c*		
Δa	?	X	X	X	X					
Δq	X	X	X	X	X			X	Y	dq*
dr/dΘ	X	X	X	X	X			X	Y	dr/Θ
* indicates primary parameter										

Table 6.4 Slope and "shape" parameters for the proposed unified parameter set.

Notes: The Δa parameter seems to be quite correlated to dq in practical measurements. Furthermore, Δa does not appear to useful in modeling of surface functionality as does the RMS equivalent, dq. It should also be noted, that the dq and dr/Θ parameters describe very different functionalities and both are strong candidates for the primary parameter. The RMS slope, dq, was based on numerical stability due to its averaging nature.

6.4.5 Auxiliary Functions and Parameters

The derivation of distributions, functions or series from surface profile data is not uncommon in the more advanced analysis of surface metrology data sets. In fact, the statistical amplitude parameters RMS (q), skewness (sk), and kurtosis (ku) can be derived either algebraically or through the development of the *amplitude distribution function (ADF)*.

While numerical values (parameters) can be obtained from many of these functions, it should also be noted that a great deal of *visual* information can be conveyed through the graphical presentations associated with these functions. Thus, it is assumed that the functions are (in and of themselves) useful and guidelines will be set forth for a common means for their presentation. In addition, for each of the functions explored, the associated parameters will be evaluated.

Material Ratio Curve

The material ratio curve (Abbott and Firestone 1933), also referred to as bearing ratio curve or the Abbot-Firestone Curve is often used in the visualization and numerical characterization of load carrying surfaces and/or surfaces in sliding contact. The material ratio curve is developed either through the accumulation (integration) of the ADF (from highest peak to lowest valley) or by cutting through the profile at various heights (or depths) and calculating the ratio of cut material to (projected) profile length.

For visualization purposes, the material ratio curve should be plotted at the same scale as the profile from which it was obtained. If space permits, the material ratio curve should be plotted alongside the profile in order to aid in the visual comparison of profile features to material ratio curve features (as shown in Figure 6.3).

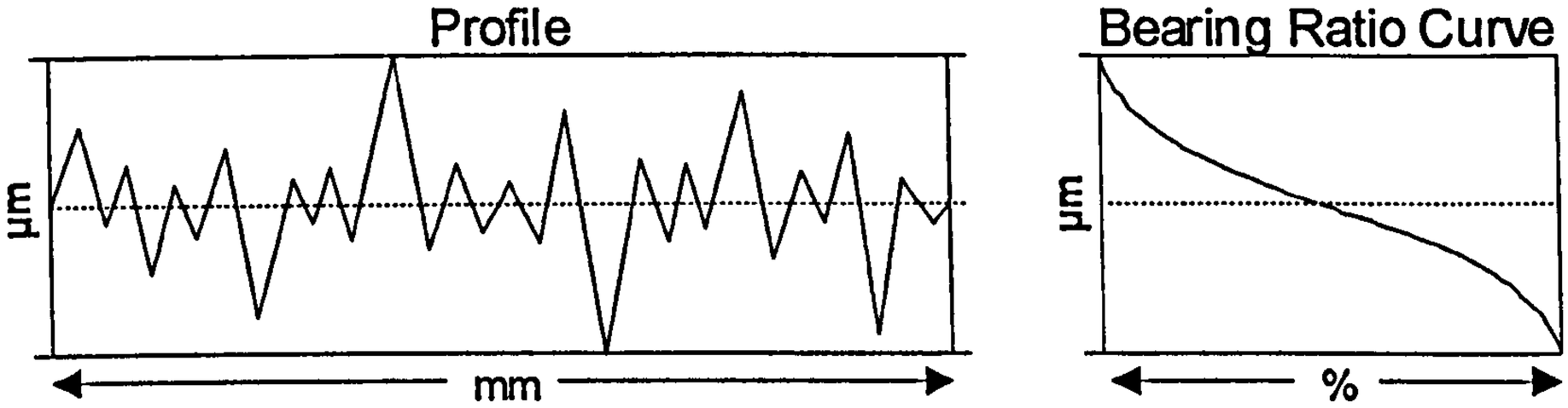


Figure 6.3 *Proper scaling, orientation and alignment of material ratio curve graph based on profile graph.*

Although many numerical characterization schemes have been developed for material ratio analysis (ISO/DIS 13565 parts 1, 2 and 3 1994-1995) these are directed towards specific types of surfaces and therefore will not be discussed in the context of a unified parameter set. However, it is important to consider more traditional material ratio computation based on two methods of cutting level selection.

tp: Peak referenced material ratio. The percentage of material encountered at a cutting level which is "depth" *µm* below a reference level which has an "offset" *%* material ratio (see Figure 6.4). Although it is not recommended, the highest peak can be used as the reference by setting the offset to zero.

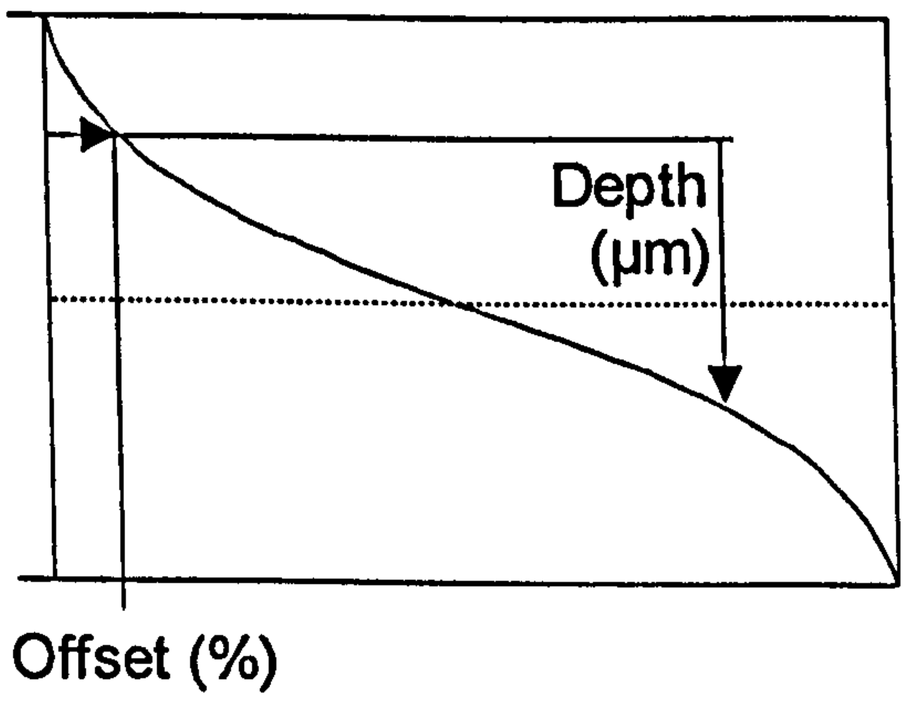


Figure 6.4 *Establishment of tp (peak) reference depth.*

tp: Mean referenced material ratio. The percentage of material encountered at a cutting level established some prescribed position (*height*) relative to the profile mean line. The height is specified as either positive (upward) or negative (downward).

Fourier Transform

Fourier analysis is one of the foundations for signal processing (see for example Bendat 1986). In the context of surface metrology, Fourier analysis has been very popular in the analysis of circular geometries; particularly in the bearing industry. More recently, these techniques are being applied to the analysis of linear profiles. Furthermore, Fourier analysis has been utilized throughout this presentation of the unified methodology for applied surface metrology in the context of concepts such as the influence of stylus tips, filter transmission functions and the harmonic content of example data sets.

Note: In terms of the graphical presentation of wavelength or frequency domain data obtained from the Fourier Transform, it is often most useful, in engineering applications, to present the data with logarithmic wavelength as the ordinate axis for linear analyses and linear frequency as the ordinate axis for circular analyses. It is also recommended that the abscissa be plotted as real amplitude as opposed to units of power. The former is typically the case in current circular analyses (Rank Taylor Hobson 1992), however the latter is sometimes applied in the Fourier analysis of linear data sets (Rank Taylor Hobson 1995).

Typically, Fourier analyses have been utilized in their graphical or tabular output formats. However, numerical parameters can be also be derived. Parameters such as, λq, as proposed by Spragg and Whitehouse (1970) and the associated λa parameter are worthy of consideration.

For reference, the Fourier transform at a given frequency, $F(\omega)$, of a time domain function, $f(t)$, is given as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \tag{6.25}$$

where $j = \sqrt{-1}$. Similarly, the discrete form of the Fourier transform at the wavelength corresponding to the kth division of the number of points, N, is given as

$$F(k) = \frac{1}{N} \sum_{i=1}^{N} f(i) e^{\left(\frac{-2\pi j}{N} ik\right)} \tag{6.26}$$

where $f(i)$ is the profile height at the ith position; $i = 0, \ldots, N-1$; and $k = 0, \ldots, N/2$.

λq: RMS wavelength. The root mean square wavelength of the profile. This is developed either through the use of the power spectrum as a wavelength weighting function or by the simpler form given by Spragg and Whitehouse (1970) as

$$\lambda q = 2\pi \frac{RMS\ amplitude}{RMS\ slope} = 2\pi \frac{q}{dq} \tag{6.27}$$

where q is the unified RMS amplitude parameter and dq is the unified RMS slope parameter. The designation of data point spacing is required as this ultimately limits the range of wavelengths (or frequencies) contained in the Fourier transform. This requirement is consistent with the Equation 6.27 approach to calculation in that the data point spacing is also a requirement for the computation of dq (based on equation 6.23).

Note: Filtering or other methods of wavelength attenuation can significantly affect Fourier analysis. Thus, in typical applications the data set is only minimally filtered by

applying just a long wavelength transmitting filter of a relatively short cutoff prior to performing any Fourier analysis.

λa: Average wavelength. The average wavelength of the profile. This is developed either through the use of the power spectrum as a wavelength weighting function or by the simpler form:

$$\lambda a = 2\pi \frac{Average\ amplitude}{Average\ slope} = 2\pi \frac{a}{\Delta a} \tag{6.28}$$

It should be noted that Equation 6.28 calculation requires the computation of the Δa parameter which is currently available in some instrumentation. However, as described above, Δa is not accommodated in the unified parameter set.

Autocorrelation Analysis

The autocorrelation function is also a very useful analysis tool in digital signal processing (Bendat 1986) as well as surface metrology (Thomas 1982, Whitehouse 1994) - particularly in the characterization of the spatial aspects of the surface. This function provides a numerical representation of the relationship between neighboring points of varying distances. Numerically, the autocorrelation function can be generated from a digital data set by

$$A(k) = \frac{\frac{1}{n-k+1}\sum_{i=1}^{n-k+1} z_i z_{i+k}}{\frac{1}{n}\sum_{i=1}^{n} z_i^2} \tag{6.29}$$

where k is the amount of shift or *lag* given in ordinates. (For the analysis of circular profiles, the $(n\text{-}k+1)$ computation points can be increased to the full, n, data points utilizing the continuous nature of the data.)

Note: The autocorrelation function is symmetric, thus only positive lag values are required for a complete description. Typically, the autocorrelation function is computed for lags from zero up to some fraction of the profile length, such as 1/10 or 1/4 of the length, l.

In terms of parameters, the autocorrelation function is most commonly characterized by the *correlation length* parameter which has historically been referred to in tribology literature as β^* (Whitehouse and Archard 1969, Whitehouse 1978, Ludema 1993).

β^*: correlation length. The correlation length of the profile. Mathematically this is lag at which the autocorrelation functions first reaches $1/e$ (approximately 37%). This parameter is often combined with others such as the RMS amplitude, q, to establish a measure of curvatures. Furthermore, the correlation length can also be useful as a measure of the adequacy of digital sampling of the surface profile. For example, if the correlation length relates to a lag of only a few ordinates, then it is likely that the data density is not adequate.

Para-meter	Primary Criteria					Secondary Criteria			Include (Y/N)	Desig-nation
	independence	functionality	stability	mathematical	generic	family	economic	modelling		
	i	*ii*	*iii*	*iv*	*v*	*a*	*b*	*c*		
tp_{peak}	X	X	X	X	X			X	Y	tpp
tp_{mean}	?	X	X	X	X	X		X	Y	tpm
λa			X	X	X				N	
λq		X	X	X	X	X		X	Y	lq
β^*	X	X	X	X	X			X	Y	cl
No primary parameter is designated.										

Table 6.5 *Auxiliary function based parameters for the proposed unified parameter set.*

Both means of describing material ratio (peak and meanline) are included as these two methods can describe different functionalities. In addition, the wavelength parameter λq has been shown to be a function of other parameters. However, given the inclusion of the RMS slope parameter, it follows that the RMS wavelength could be included based on a "family" of parameters.

6.5 Summary of the Unified Parameter Set

This chapter has dealt with the topic of parameterization from a new perspective - one of unification. In doing so, an approach has been developed whereby a manageable, unified parameter set can be defined and maintained. This approach ensures the common application of parameters in surface metrology - independent of wavelength domain and nominal geometry. Furthermore, flexibility is ensured for the purpose of accommodating future developments.

Central to this topic, has been the establishment of guidelines for incorporating parameters in the unified set. This methodology included the development of categories of parameters and the determination of necessary and optional criteria for incorporating a parameter into the unified set. Categories of parameters were established such that parameters could be compared with others which perform similar functions. Each of the candidate parameters were then evaluated relative to primary and secondary criteria.

Within each of the categories, individual tables (6.1 - 6.5) served as useful "cross reference" tools indicating a parameter's conformance to the necessary criteria. This "table" concept could also accommodate a rating system whereby a numerical value is assigned (rather than the "X" designated in the above examples). However, a numerically based "rating system" must be developed such that it will not accept a parameter which does not satisfy all of the primary requirements despite potentially high ratings in other criterion.

Table 6.6 concludes this example application with a summary of the proposed unified parameter set. This composite presentation of all parameters is essential in the governing of the unified parameter set, particularly in terms of highlighting redundancies and potential shortcomings. In addition, Table 6.7 provides example specifications for parameter which require auxiliary information (see also Chapter 2 regarding parameter designations in the unified specification scheme.)

Unified Designation	Units	Description	Historical Usage
Statistical Amplitude Parameters			
a	*µm*	Average absolute amplitude	Ra, Wa, Pa
q	***µm***	**Root mean square amplitude**	**Rq, Wq, Pq, Oq**
sk		Skewness	Rsk
ku		Kurtosis	Rku
Extreme Amplitude Parameters			
p	*µm*	Height of highest peak	Rp, Wp, Pp
v	*µm*	Depth of lowest valley	Rv, Wv, Pv
t	***µm***	**Total (peak to valley) height**	**Rt, Wt, Pt**
pm	*µm*	Mean peak height	Rpm
vm	*µm*	Mean valley depth	Rvm
tm	*µm*	Mean peak to valley (total) deviation	Rtm, Rz(DIN)
Spacing Parameters			
sm	*mm*	Average spacing of mean line crossings	Sm
pc	***peaks/mm (linear)***	**Peak count**	**Pc**
	peaks/° (circular)	**(thresholds, c1 & c2 required)**	
Slope Parameters			
dq	***° (linear)***	**Root mean square slope**	**Δq**
	µm/° (circular)	**(data point spacing required)**	
dz/x (linear)	*µm*	Rate of change in defined window	DR/DΘ
dr/Θ (circular)		(window "x" or "Θ" required)	
Auxiliary Function Parameters			
tpp	%	Peak referenced material ratio	tp
		(depth and offset required)	
tpm	%	Mean referenced material ratio	tp
		(height required)	
lq	*mm (linear)*	Root mean square wavelength	λq
	° (circular)	(data point spacing required)	
cl	*µm (linear)*	Correlation length	β*, τ*
	° (circular)		

Table 6.6 *Summary of the proposed unified parameter set. Primary parameters are in bold text.*

Parameter	Auxiliary Information	Example
pc	Thresholds, c1 & c2 (linear units)	pc (2.0 µm, -0.5 µm)
dq (linear)	Data point spacing (linear units)	dq (0.5 µm)
dq (circular)	Data point spacing (angular units)	dq (0.1°)
dz/x (linear)	Window width	dz/x (2.0 mm)
dr/Θ (circular)	Window width	dr/Θ (30.0°)
tpp	Depth (linear units) and Offset (%)	tpp (2.0 µm, 5.0%)
tpm	Height (linear units)	tpm (0.5 µm)
lq (linear)	Data point spacing (linear units)	lq (0.5 µm)
lq (circular)	Data point spacing (angular units)	lq (0.1°)

Table 6.7 *Designation of auxiliary information for unified parameter set.*

A Unified Methodology for the

Application of Surface Metrology:

Chapter 7

Application and Recommendation

As discussed in Chapter 1, the methods associated with surface metrology are in a state of divergence. Varying types of instruments are being used to assess surface features and there are a growing number of disputes between suppliers, customers and laboratories and the associated costs are ever-increasing. Unfortunately, these problems are coming in time when the importance of surface metrology to industry is in a period of rapid growth.

To address this situation, a unified methodology for the specification and assessment of surface features has been developed (see Chapter 2). Under this approach, the controlled limitation of wavelengths is the foundation for common analyses through a unified parameter set. Given these common approaches, similar assessments of surfaces can be made from different instruments (providing that they are suitable for that particular wavelength regime) and the numerical values should generally agree within reasonable uncertainty bounds.

In this chapter, the implications of the scheme for unification will be addressed in light of the three primary arenas associated with surface metrology: surface function (design), surface generation (manufacturing) and surface measurement (metrology). Next, the logical extensions of this approach will be explored - particularly in light of the current move toward more three dimensional surface analyses. Finally, this work will be completed with an outline for a proposed standard for the implementation of the unified methodology in terms of the specification, measurement and analysis of linear and circular profiles.

7.1 Functional Implications

While the direct link between surface function and numerical surface characterization is often very difficult to achieve in a general sense, many parameters demonstrate strong correlations to physical phenomena (Chapter 6). Furthermore, by controlling the wavelength content of data sets (Chapters 3, 4 and 5), many specific functionalities can

be further exploited. The scheme for the unification of surface metrology provides a significantly enhanced set of specification tools for the designer whereby a surface can be more precisely *engineered* to suit the application. Furthermore, the unified specification approach provides a significant improvement in terms of reducing the possibility for misinterpretations of surface specifications.

7.1.1 Enhanced Wavelength Control

The unified scheme allows for the direct specification of surface attributes for specific wavelength regimes. In many interfaces, the independent control of specific wavelength regimes is very important in order to ensure proper functionality. Unfortunately, under the current systems for specification, this can only be accommodated by adding "notes" to part drawings and these "notes" often vary between drawings as they are not rigorously standardized (Chapter 2). In the following example, a diesel engine crankshaft pin is shown to exhibit differing functionalities based on different wavelength domains (Bhargava 1993).

In this application, the crankshaft pin's surface must be controlled in order to obtain a proper sliding interface with a thin, conformable bearing shell of a relatively soft material. The design and control of this interface is based on three wavelength regimes within a linear profile. The long wavelength aspects of the surface are not as critical to the performance given the conforming nature of the bearing. However these long wavelength attributes must be controlled to some level to ensure a reasonable *fit*. On the other end of the wavelength spectrum, the short wavelength features are important to the sliding (tribological) aspects of the interface and requires a tighter level of control. In the middle of the wavelength regime is a more subtle aspect of the surface profile which has historically been neglected. Recent studies have shown that this middle wavelength regime is critical relative to the performance of the interface in that significant amplitudes in this regime can result in localized loading and ultimate failure of the bearing (Bhargava 1993). (See Figure 7.1.)

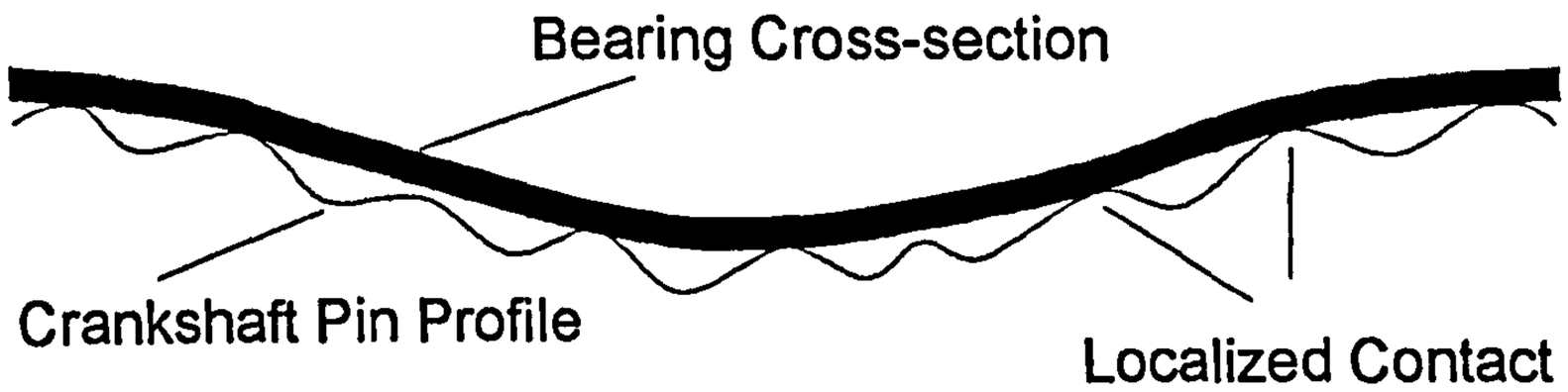


Figure 7.1 *Bearing conformance in the presence of long and medium wavelengths.*

The specification of the profile is based on Gaussian, band-pass filtering whereby the short wavelength regime (2.5 μm - 0.8 mm) is controlled by the average amplitude (Ra - upper and lower limits) and mean peak to valley parameters (Rz(DIN) - upper and lower limits). The middle (0.8 mm - 8.0 mm) wavelength regime is controlled by the peak to valley amplitude as is the long wavelength regime (greater than 8.0 mm).

Unfortunately, since the currently available surface metrology instrumentation and specification approaches do not directly accommodate this level of surface control, special software had to be developed and extensive notes were required on the part drawing (Bhargava 1993). However, under the unified methodology this surface profile geometry could be explicitly specified in a standardized format as given in Figure 7.2.

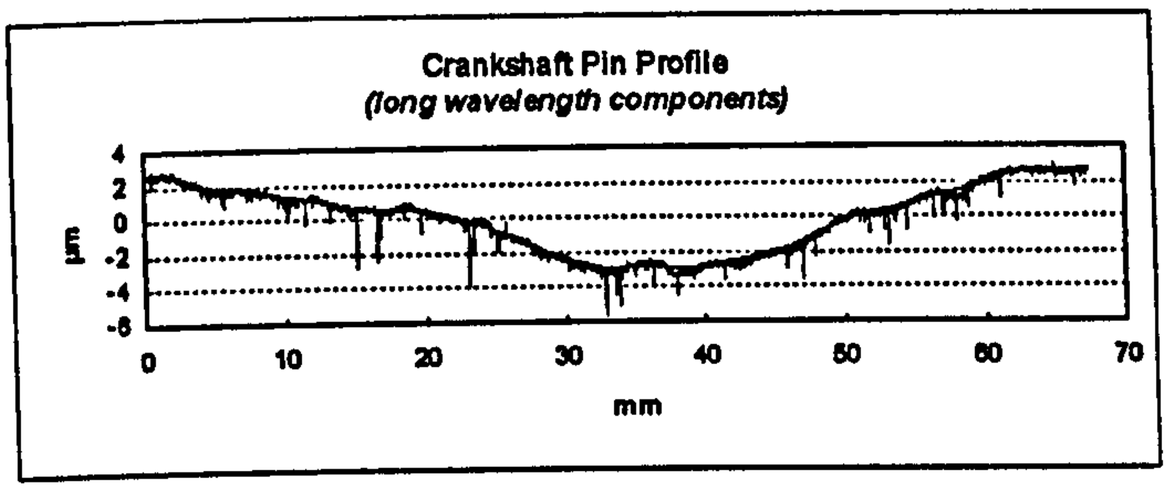
—	8.0 mm	
t	X.XX μm	
—	0.8 mm	8.0 mm
t	Y.YY μm	
—	2.5 μm	0.8 mm
a	A.AA μm	B.BB μm
tm	C.CC μm	D.DD μm

Figure 7.2 *The crankshaft pin unified specification. (Numerical values omitted in light of manufacturer confidentiality.)*

The unified specification given in Figure 7.2 contains all of the relevant information for the control of the surface profile in a standardized, tabular format (see Chapter 2). The format of the specification provides an easy visual verification of all of the necessary information. Furthermore, given this common format, consistency can be gained between drawings of similar components.

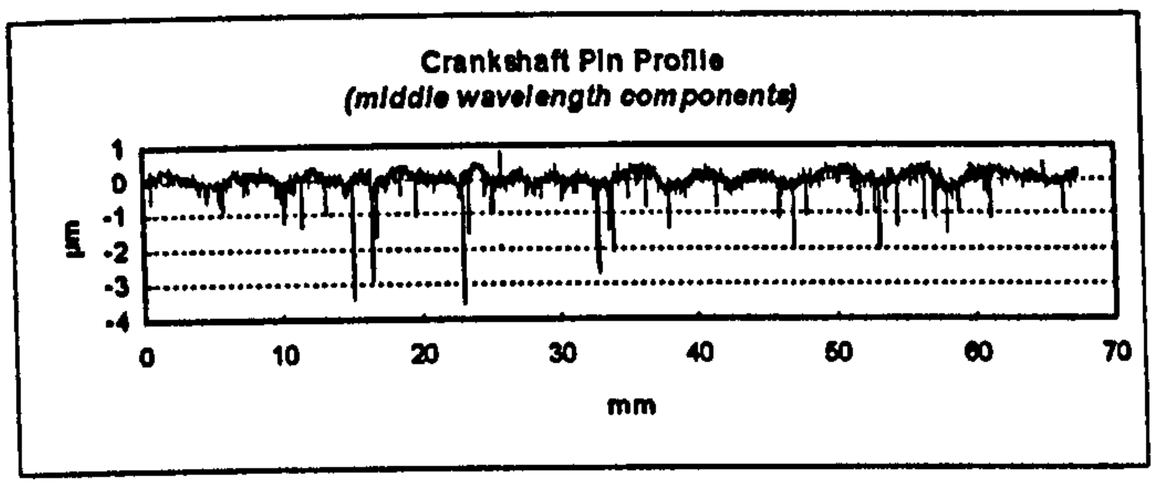
Given the unified specification shown in Figure 7.2, the measurement results can be similarly formatted to provide a useful connection between graphical profile data and the specified parameters. An example measurement report is given in Figure 7.3.

The Figure 7.3 measurement report is formatted according to the unified specification and therefore provides an additional level of consistency between the specification (as designed) and evaluation (as produced and measured). This formatting capability is already attainable on some of today's PC-based analysis software (Hochwart 1995).



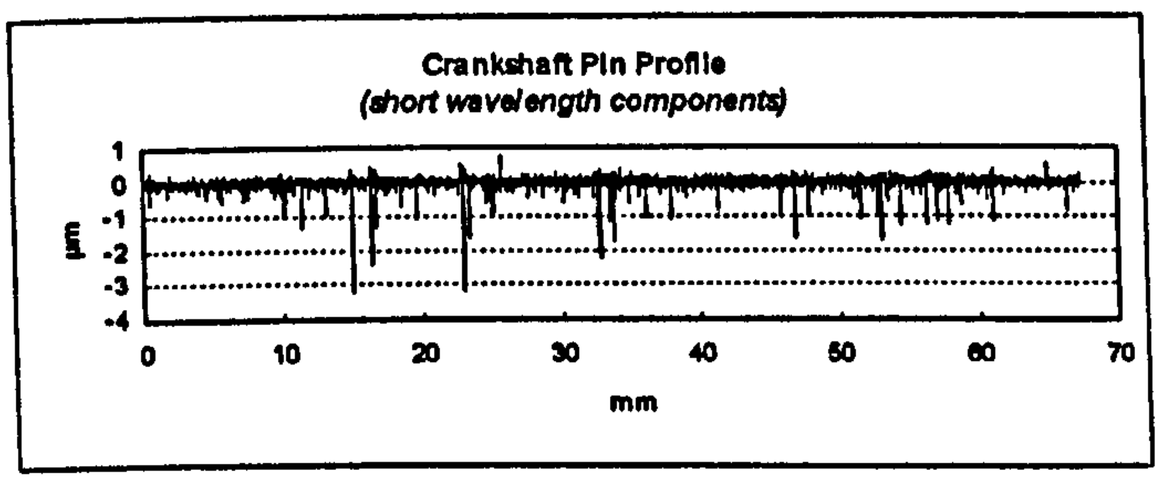
— ∫ 8.0 mm

t 5.729 µm



— ∫ 0.8 mm 8.0 mm ∫

t 0.789 µm



— ∫ 2.5 µm 0.8 mm ∫

a 0.104 µm

tm 0.894 µm

Figure 7.3 Measurement report format based on Figure 7.2 specification.

7.1.2 More Characterization Options

While one aspect of the scheme for unification is the development of a manageable unified parameter set, the parameters included in the set should not significantly reduce the number of *useful* numerical parameters at the disposal of a designer. In some cases, the number of parameters at a designer's disposal may actually increase as the

unified parameter set brings together parameters historically limited to specific geometries or wavelength regimes and makes them available for all surface characterizations.

Furthermore, the unified methodology introduces three wavelength limitation options as well as a mechanism for specifying multiple wavelength regimes and associated parameters. These new wavelength limitation alternatives, coupled with a (still) relatively large set of functional parameters, result in a great deal more options for a designer in terms of generating a specification to the desired level of detail.

7.2 Manufacturing Implications

The ability to characterize surface attributes within specific wavelength regimes provides a valuable diagnostic tool in the manufacturing arena. This advantage can best be realized through the *independent* characterization of the primary wavelength regimes which are associated with material removal process (Avallone and Baumeister 1979, Thomas 1982, Wade 1991). This *independence* is very important in the development and control of manufacturing processes as it allow the manufacturing engineer to isolate the effects of various process attributes on the workpiece surface.

For example, short wavelength characterizations are useful in the monitoring of the tool-to-surface interface and the integrity of tool (or abrasive, etc.). These short wavelength evaluations are typically the most frequent as tool condition is generally the most varying aspect of material removal. In the middle wavelength regimes, the integrity of the moving components (i.e. vibrations) of the machining operation can typically be ascertained. In many cases, these evaluations can occur at less frequent interval - particularly if slight changes can be detected through the use of well targeted wavelength regimes. Finally, and perhaps on an even less frequent basis, the long wavelength characteristics of the manufacturing process are measured in light of assessing the structural aspects of the manufacturing process (ASME B46.1 - 1995).

Ideally, if an instrument were well suited for the unified scheme, a single (for example linear) profile trace could be collected and numerically processed to provide information in all relevant regimes. An example application of the unified specification scheme for the "in-process" control of a turning operation is given in Figures 7.4 through 7.6.

This example demonstrates an application of the unified scheme for surface metrology in the context of process control in the turning of diesel engine piston pin bores. Figure 7.4 provides an example of a piston pin bore exhibiting an unacceptably high amount of chatter. The data set was collected by a Form Talysurf Series S5 (0.01 μm resolution, 2.0 μm tip radius, 5.0 μm spacing).

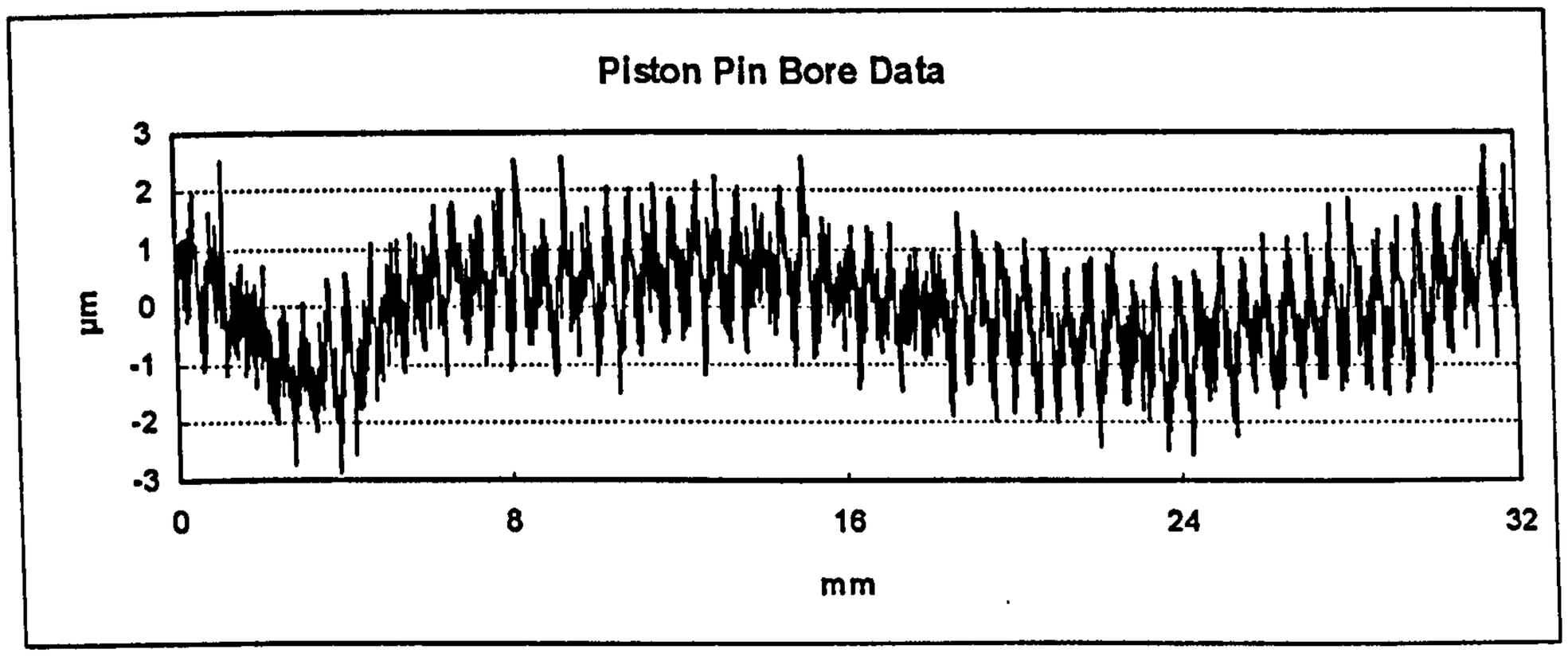


Figure 7.4 *Axial measurement of a diesel engine piston pin bore.*

The machining of piston pin bores is based on a rather *traditional* turning operation whereby the workpiece is rotated relative to a cutting tool which translates axially at some prescribed feedrate along the component at a desired depth of cut. The errors generated in this mode of material removal tend to be very isolated in terms of their wavelength regimes. This can be demonstrated via a Fourier transformation and is provided in Figure 7.5.

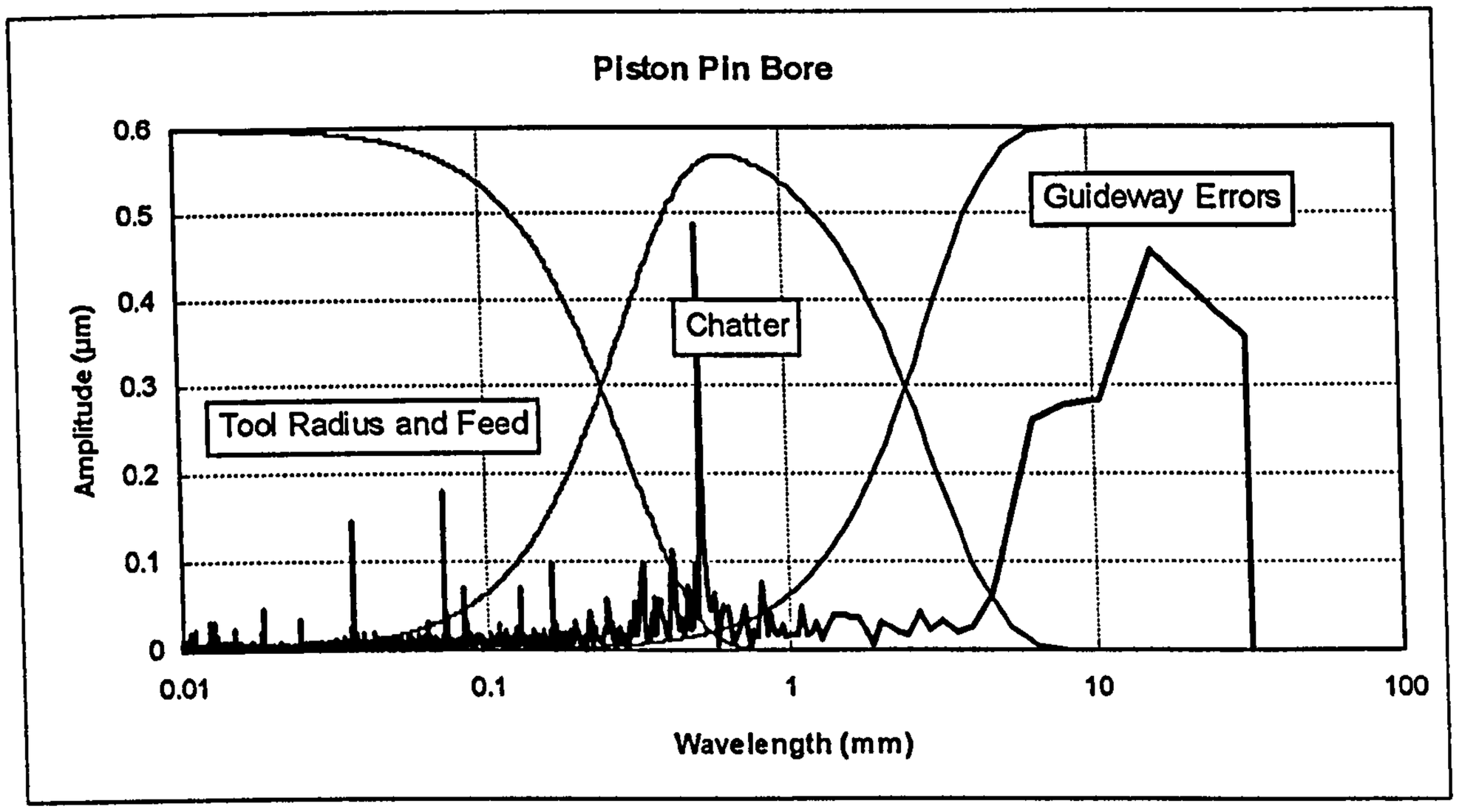


Figure 7.5 Wavelength content of piston pin bore.

This Fourier based (spectral) representation is useful in distinguishing significant, re-occurring profile features. For example, the machine settings for this turning operation where such that a 0.003 in. feedrate was generated. This is confirmed in Figure 7.5 based on the presence of a relatively high amplitude at approximately 0.08 mm (0.003 in.). In addition, the chatter (at approximately 0.8 mm) is revealed by the Fourier Transform. Figure 7.5 also superimposes the transmission characteristics for Gaussian filters bounding the central (chatter-related) wavelength regime.

Given the above data set, the selection of 0.25 mm and 2.5 mm wavelength boundaries (i.e. filter cutoffs) seems appropriate for isolating these process control aspects in terms of their characteristic wavelengths. Proper measurement within these wavelength domains can then facilitate improved process control.

It should also be noted that in the process control application (where component drawing (i.e. blueprint) standards are not as strictly imposed) additional information could be appended to the unified specification format and provided to the machine

operator (as shown in Figure 7.6). This additional information can assist in the interpretation of measurements relative to the controls of the process.

Unified Specification:			*Additional Notes:*
—	2.5 mm		(Straightness) *All wavelengths longer than predominant chatter.*
t		X.XX µm	Changes in linear reference.
—	0.25 mm	2.5 mm	(Chatter) *Ranging from upper roughness cutoff to wavelength distinction between straightness and chatter.*
t		Y.YY µm	Presence of chatter.
—	2.5 µm	0.25 mm	(Roughness) *Wavelengths associated with material removal which are shorter than chatter.*
q	A.AA µm	B.BB µm	Tool condition.
dq (0.5 µm)	C.CC µm	D.DD µm	Surface integrity.
sm	E.EE µm	F.FF µm	Feedrate.
tm	G.GG µm	H.HH µm	Tool condition.

Figure 7.6 *Application of the unified specification scheme as process control documentation or "in-process" specification.*

The visible link between surface parameters and process conditions/controls shown in Figure 7.6 is very valuable to industry - particularly in light of the smaller, more flexible workforces that are present in today's manufacturing environments. In addition, this linkage of surface measurements to process controls is of utmost importance in the generation of costly and/or critical surfaces whereby the highest levels of process control are required. Furthermore, as the measurements become more indicative of the process controls, direct (and automatic) feedback control becomes more achievable.

7.3 Metrological Implications

Adoption of the unified methodology for surface metrology will, in general, require additional flexibility in instrumentation. However, the majority of this flexibility will come through enhancements to software and/or electronics. Fortunately, in the midst of the diverging measuring technologies, many of the band limitation concepts and parameters contained in this scheme are well defined and generally accepted (see for example ISO 3274 - 1995) by the surface metrology community such that minimal efforts would be required on the part of instrument manufacturers.

Perhaps the most significant change that would affect instrument manufacturers would be the incorporation of user-selectable (and perhaps multiple) wavelength or frequency bands in terms of both transmission characteristic and wavelength (e.g. cutoff) values. The majority of today's instruments incorporate some form of band limitation such as 100:1 or 300:1 (ISO 3274 - 1995), however many of these applications of band limitation are based on fixed values.

Once again, it is important to note that some of the today's more advanced instruments already accommodate variable bandwidths. However the ability to allow varying transmission characteristics and the simultaneous reporting of multiple transmission bands is not fully incorporated in most of today's instruments.

7.3.1 Instrument Evaluation

In light of the unified methodology, the purchaser of surface metrology instrumentation should be aware of the following instrument attributes (apart from environmental influences):

- Reference datum quality
- Sensor range, resolution, linearity and traceability
- Dynamic response
- Algorithm integrity

These attributes relate directly to the instrument's ability to resolve wavelengths and calculate the desired parameters (Stedman 1987). Each attribute is essential in ensuring reliable measurement results and each should be evaluated prior to the purchase of an instrument as well as periodically during the working life on an instrument. Although, several approaches may be used in the analysis of these metrological aspects, those which follow represent some of the more typical.

Note: all testing should incorporate multiple measurements using calibrated artifacts. This allows the simultaneous evaluation of not only systematic, but also random errors in each metrological evaluation (Nielsen 1996b).

Testing procedure summaries are provided in Tables 7.1 and 7.2 for linear and circular instrumentation respectively.

Reference datum quality. The reference datum of an instrument represents the ultimate limitation of an instrument in terms of assessing a "perfect" component. Errors in the reference datum translate directly to the profile data during a measurement (Chetwynd 1987). (Note: Some instruments incorporate software error compensation or "error mapping". In these cases the residual errors from the compensated datum represent the measurement limitation.)

The analysis of a reference datum typically involves the measurement of a "near perfect" artifact such as a high quality optical flat (for linear evaluation) or glass sphere (for circular evaluation). In cases, where the instrument errors approach the scale of the those present in the artifact, multiple redundancy (i.e. reversal methods) should be employed to distinguish between the errors in the artifact and those of the instrument (Donaldson 1972, Thwaite 1973, Whitehouse 1976, Busch 1989).

Sensor linearity and traceability. The sensor linearity and traceability are essential in ensuring reproducibility between instruments and configurations. This evaluation typically targets the position sensing components of the instrument and the calibration thereof. Linearity and traceability are typically evaluated through the measurement of calibrated step heights (such as gauge blocks) or groove depths (such as a type "A2" artifact) (Lukjanov 1967, Sharman 1967b, Spragg 1967, Nielsen 1989, Hillman 1992). In the analysis of instrumentation for circular measurement it is also quite common to utilize "flick" masters (Chetwynd 1987, Nielsen and Malburg 1996).

Dynamic response. The dynamic response of surface metrology instrumentation is often ignored in the selection of measurement equipment. Furthermore, instruments with poor dynamic response characteristics are often purchased based on the appearance of *enhanced* repeatability. Unfortunately, while this area is more difficult to assess, it is of great importance in the overall performance of an instrument. Many of the correlation issues arising from today's instruments are the result of differences in their dynamic response.

To demonstrate this situation, a round-robin study was perfumed to compare frequency response characteristics between roughness measuring instruments. An area of a ground and lapped (stratified) fuel injector plunger was designated for measurement. Each of 12 instruments was calibrated in accordance with the manufacturer's specifications and each performed 10 measurements (equally spaced over the designated region). The spectra resulting from the 10 measurements were averaged as to obtain the "nominal" wavelength content as measured by each instrument and the results are plotted in Figure 7.7. It is important to note that these graphs are not transmission characteristics in that each data set would represent an attenuation in a *absolute* sense. The Figure 7.7 plots are of the surface's wavelength content as measured by various instruments and are only useful in *relative* (instrument-to-instrument) comparisons.

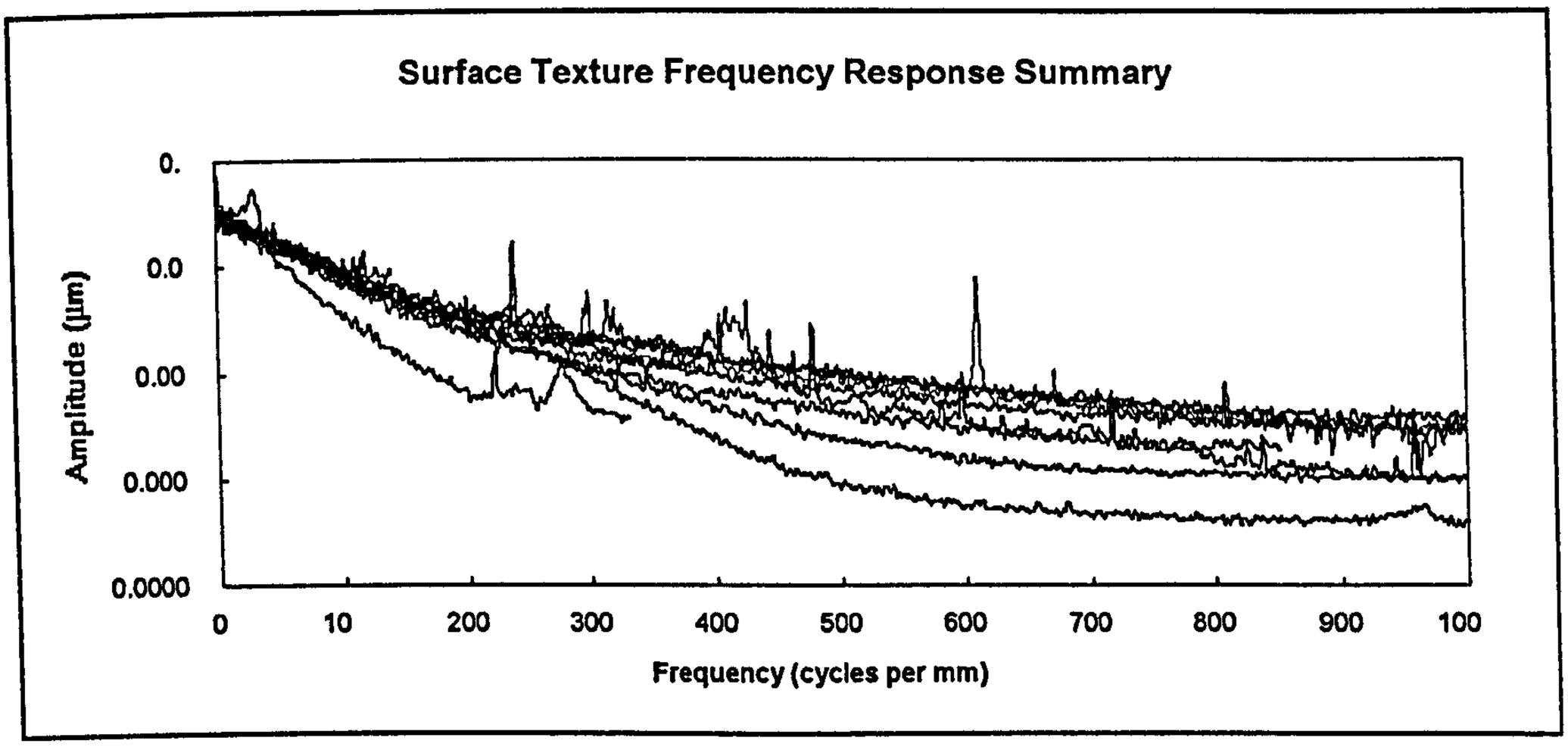


Figure 7.7 *Round-robin measurements of the frequency content of a stratified surface.*

In evaluating an instrument, the dynamic response should be verified in the frequency (or wavelength) range of interest. This can be accommodated by sophisticated *shaker* type approaches whereby specific frequencies and amplitudes are mechanically transmitted to the instrument's probe (Reason 1973). A more common and less costly approach is through the evaluation of artifacts of varying frequency such as type "D" specimens (Hillman et al. 1984, Nielsen 1991, Hillman 1992, ASME B46.1 - 1995).

It is foreseeable that, as the surface metrology "user" community becomes more educated, the evaluation of dynamic response will become more common (see for example Whitehouse 1988 or Liu et al. 1993). This may drive the further development of artifacts designed more specifically for the evaluation of the instrument's *broad-band* dynamic response.

Algorithm integrity. As surface metrology has become more numerically intensive, the importance of algorithms has also grown. However, there is often a tradeoff between the computation of the *correct* solution and the timely computation of an *adequate* solution. In light of this consideration, many approximations are used in the algorithms

of surface metrology. (See, for example, the Gaussian filter variations in Chapter 5 or Chetwynd 1991.) Often, variations in algorithms can account for differences in the results obtained from various instruments. Thus, it is important for the user of these instruments to understand the limitations of the numerical techniques that are used in filtering and parameter calculations.

Often, algorithm integrity can be determined through *soft-gauging* techniques. In these approaches, known data sets are numerically generated and loaded into the instruments software (Diaz and Hopp 1993, Hopp and Levenson 1995). Software performance can often be assessed through the use of well defined profiles (for example impulse, random or sinusoidal profiles) particularly in cases where filter characteristics or parameter values can be analytically determined *a priori* for the profiles.

It should also be noted that a common means of data storage and exchange is essential to this process. In the context of surface metrology, standards organizations are currently exploring the possibility of a common file format or *Surface Data File Format* (SDF) (Sullivan and Stout 1992).

Performance evaluation for linear measuring instruments.	
Reference Datum	• Measurement of a good quality optical flat. - *Depending on the "shape" of the errors, the flat can be rotated 180°, or inverted (and probed upward) to determine the source of the errors.* - *The evaluations should be performed in each wavelength regime of interest.*
Linearity & Traceability	• Measurement of multiple, calibrated steps. - *Either gage block steps (carefully wrung to an optical flat) or a calibrated "depth setting" master such as type "A2".* • Measurements should be repeated at varying nominal probe deflections. - *Variations in numerical results indicate potential linearity problems.*
Dynamic Response	• Mechanically generated probe motion or comparative measurement of an appropriate artifact. - *Frequency generator approach is preferable and should be applied at several wavelengths in the regime of interest - paying particular attention to the limits of the desired regime.* - *Artifact based approaches can be used when the artifact's frequency (or wavelength) content is known. The Fourier transformed data sets are compared.*
Algorithm Integrity	• Filtering algorithms can be assessed by Fourier analysis - *FFT's can be applied to the data "prior to" and "immediately after" filtering. The attenuation at each wavelength can be compared to the desired transmission function's analytical expression.* • Parameter calculations are best assessed via soft-gauging or off-line processing of measured data sets. - *Providing that reference software is available.*
Composite Tests	• Assessment of noise level. - *Measurement of a good quality optical flat.* • Measurement of an inclined optical flat can also address linearity, X-Z axis alignment, and/or algorithmic errors. - *This testing should be performed over the entire probe range, in both upward and downward inclinations.*

Table 7.1 Performance evaluation for linear measurement instrumentation.

Performance evaluation for circular measuring instruments.	
Reference Datum	• Measurement of a good quality sphere. - *Depending on the "shape" (i.e. predominant harmonics) of the errors, the sphere can be rotated some known amount (typically one-half of the wavelength of the primary error) and re-measured to determine the source of the errors.* - *The evaluations should be performed in each wavelength regime of interest as well as at several heights above the spindle in light of "coning" errors.*
Linearity & Traceability	• Measurement of multiple, calibrated steps. - *Depending on the probing configuration gage block steps (carefully wrung to an optical flat) can be measured.* - *An alternative approach can be employed through the measurement of a properly calibrated "flick" standard at varying nominal probe deflections.*
Dynamic Response	• Mechanically generated probe motion or comparative measurement of an appropriate artifact. - *Frequency generator approach is preferable and should be applied at several wavelengths in the regime of interest - paying particular attention to the limits of the desired regime.* - *Artifact based approaches can be used when the artifact's frequency (or wavelength) content is known. The Fourier transformed data sets are compared.*
Algorithm Integrity	• Filtering algorithms can be assessed by Fourier analysis - *FFT's can be applied to the data "prior to" and "immediately after" filtering. The attenuation at each wavelength can be compared to the desired transmission function's analytical expression.* • Parameter calculations are best assessed via soft-gauging or off-line processing of measured data sets. - *Providing that reference software is available.*
Composite Tests	• Assessment of noise level. - *Measurement of a good quality sphere.* • Measurement of a good quality sphere at varying eccentricities relative to the reference datum to ascertain linearity, and/or algorithmic errors. - *Performed at multiple steps, over the probe range.*

Table 7.2 *Performance evaluation for circular measurement instrumentation.*

7.3.2 Correlation of Results

As previously stated, the defined bandwidths associated with the scheme for unification will drive the commonization of data sets. This will lead to the common measurement of surfaces independent of the instrumentation being used (provided that the instrumentation is capable of assessing the defined bandwidth). This improvement in correlation represents the most significant economic gain to be achieved through the adoption of the unified methodology. (See Chapter 1 for a summary of the costs associated with surface metrology issues.)

As an example of how correlation can be achieved through the adoption and implementation of the scheme for unification, re-consider the straightness measurements of Chapter 1.

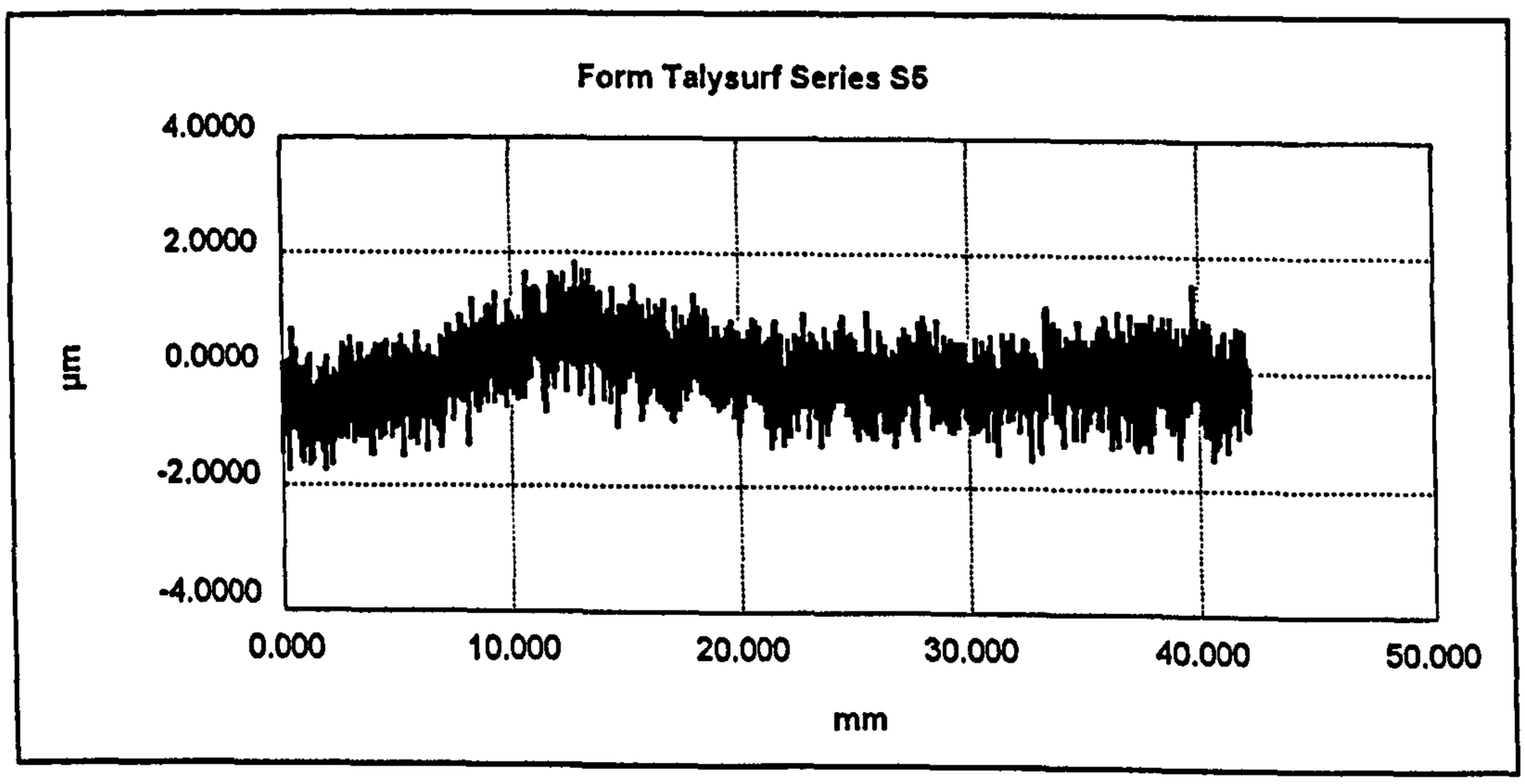


Figure 7.8 *(Figure 1.3) Straightness profile as measured with the Form Talysurf Series S5.*

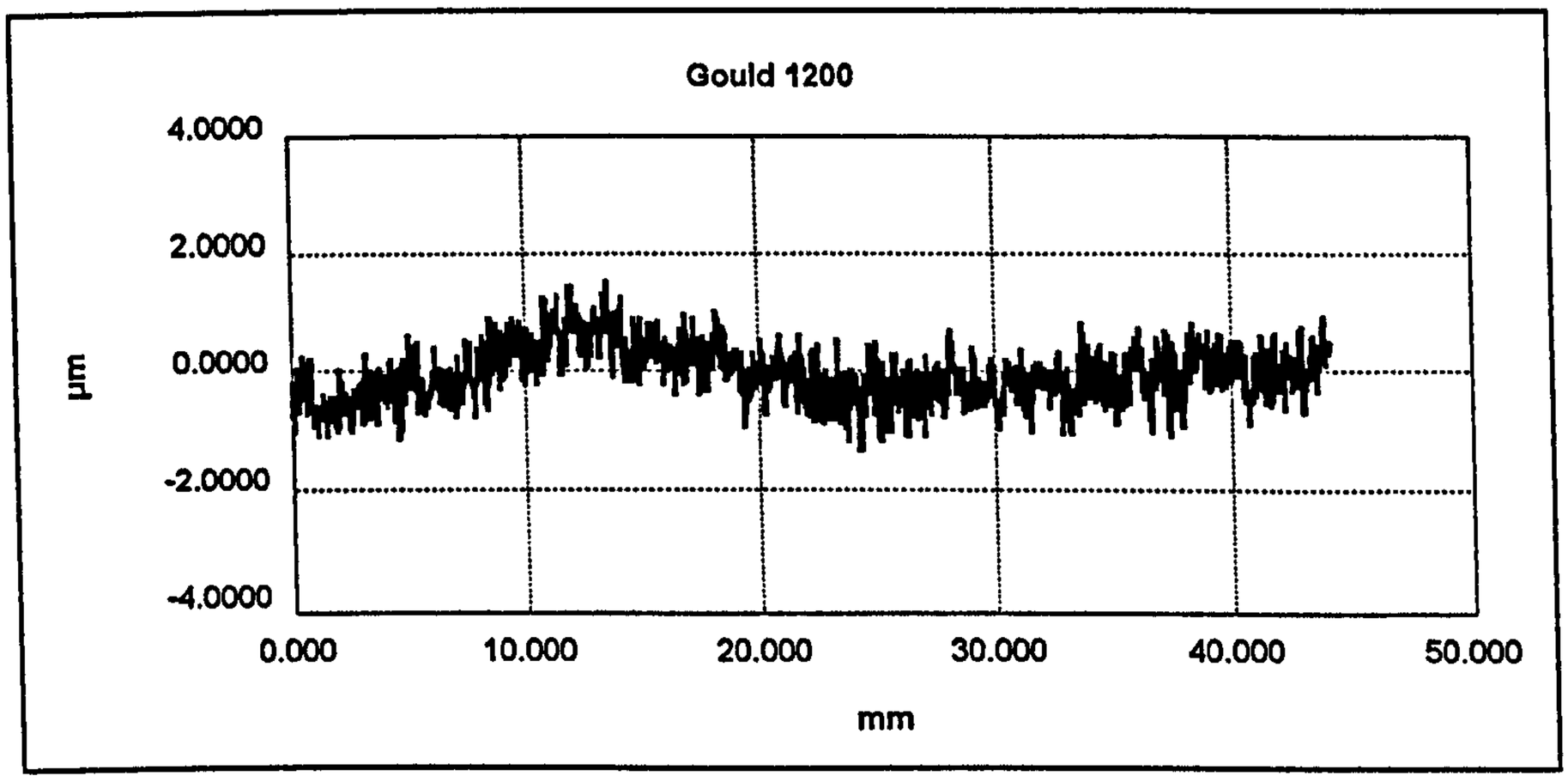


Figure 7.9 *(Figure 1.4) Straightness profile as measured with the Gould 1200.*

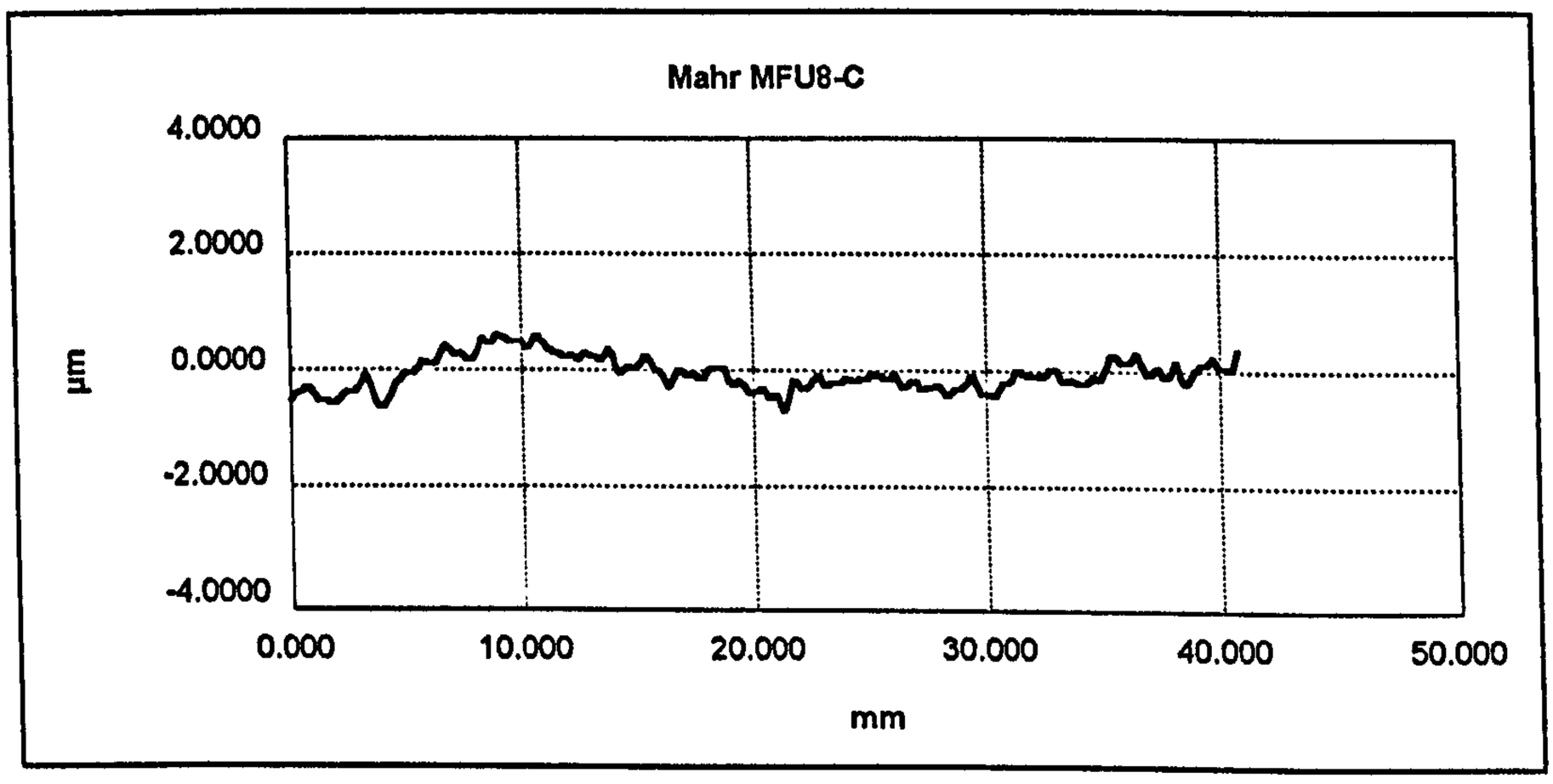


Figure 7.10 *(Figure 1.5) Straightness profile as measured with the Mahr MFU8-C.*

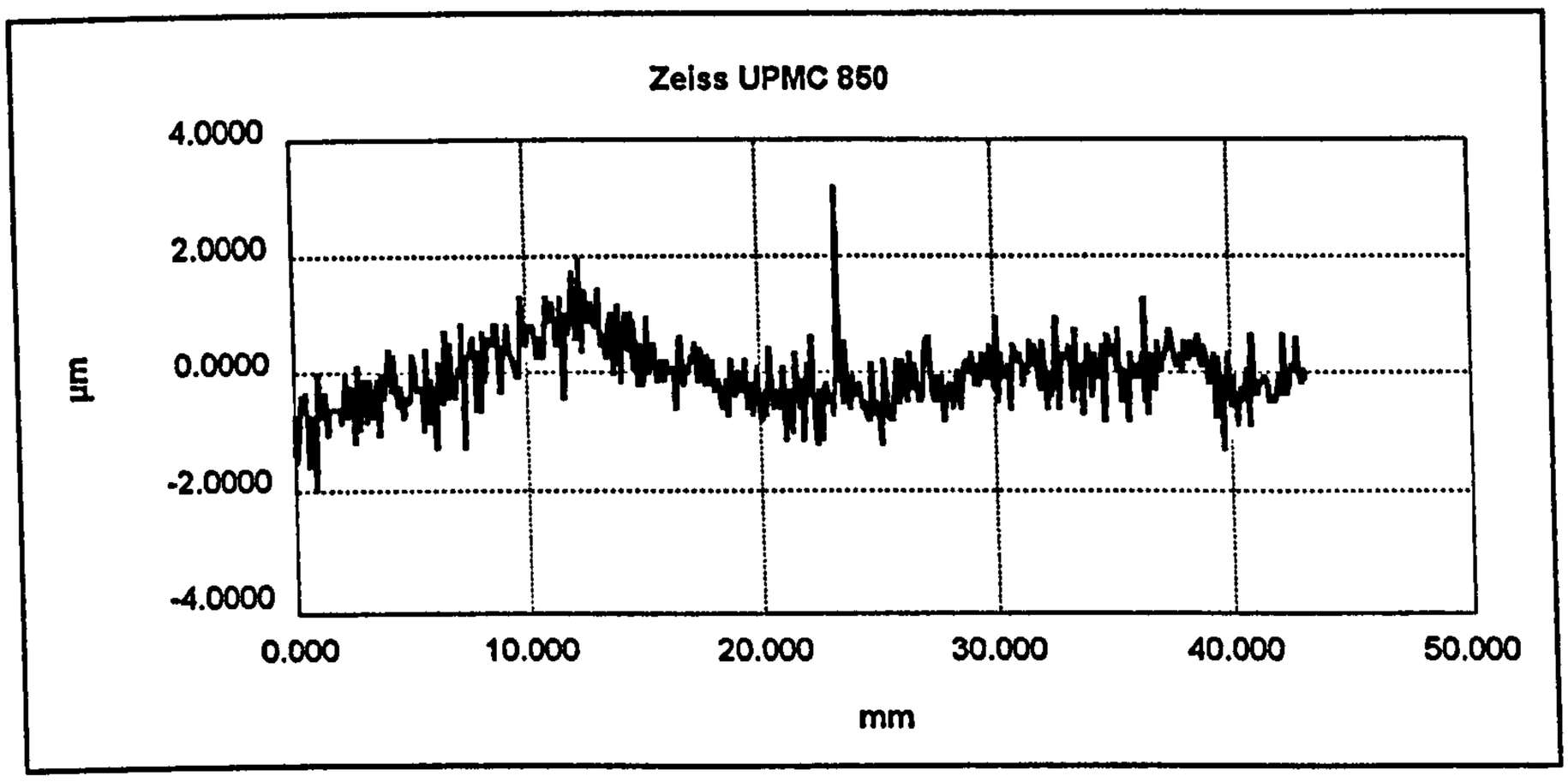


Figure 7.11 (Figure 1.6) Straightness profile as measured with the Zeiss UPMC 850.

Chapter 1 described the process for collecting these data sets and presented the resulting parameters, however it did not go into detail regarding the specific instrument settings. For each of the measurements, the instrument settings were recorded. In many cases, the effects of these settings become apparent through the Fourier analysis of the resulting data sets.

Form Talysurf Series S5

The Form Talysurf measurements utilized a 2.0 µm tip radius, data points were collected every 0.25 µm and no subsequent filtering was applied. The resulting wavelength content of the data set is very similar to what would be expected for a ground surface.

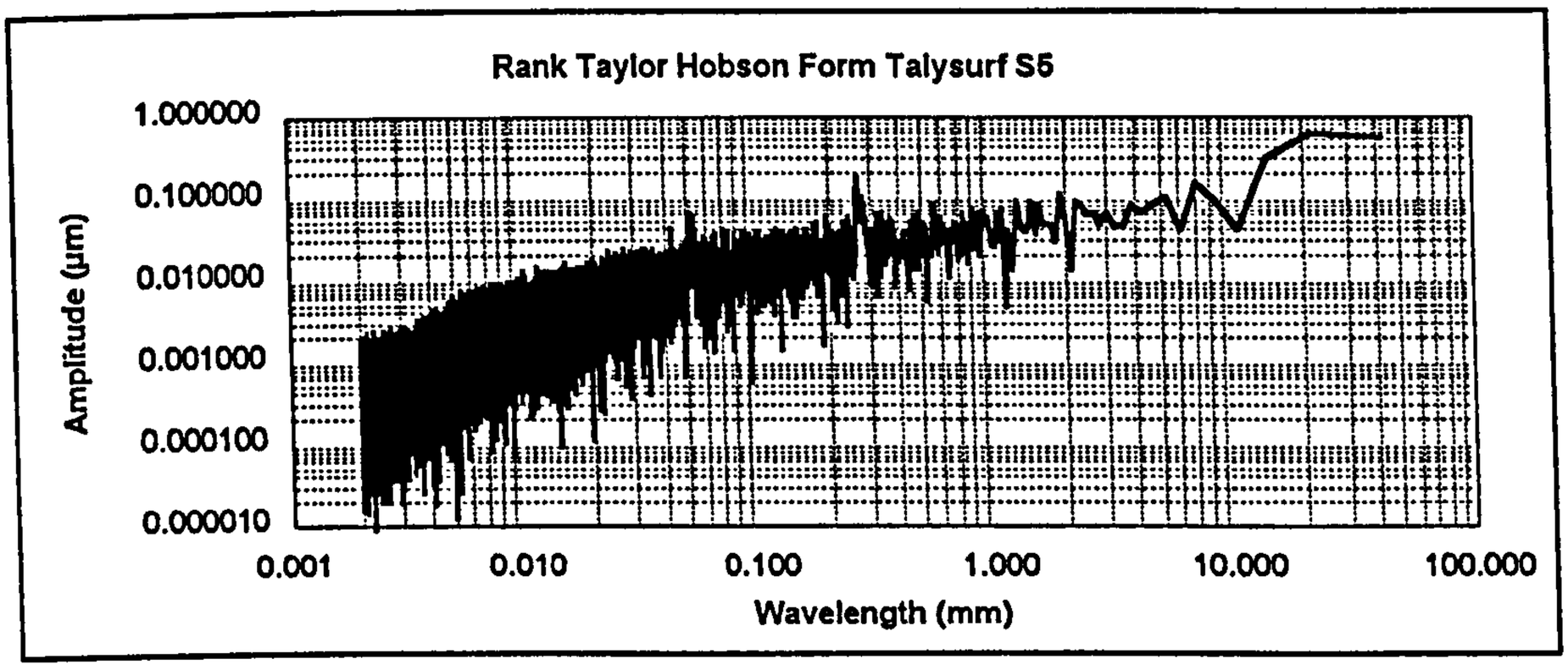


Figure 7.12 *Wavelength content of shaft per Form Talysurf S5.*

Gould 1200

The Gould 1200 utilized in this study has been interfaced to a personal computer via an analog interface to the plotter output. A 10 µm tip radius was utilized, the data point spacing was 8.5 µm and a 80 µm long pass, Gaussian filter was applied to suppress drive unit noise. The evidence of the filter is present in the resulting wavelength content.

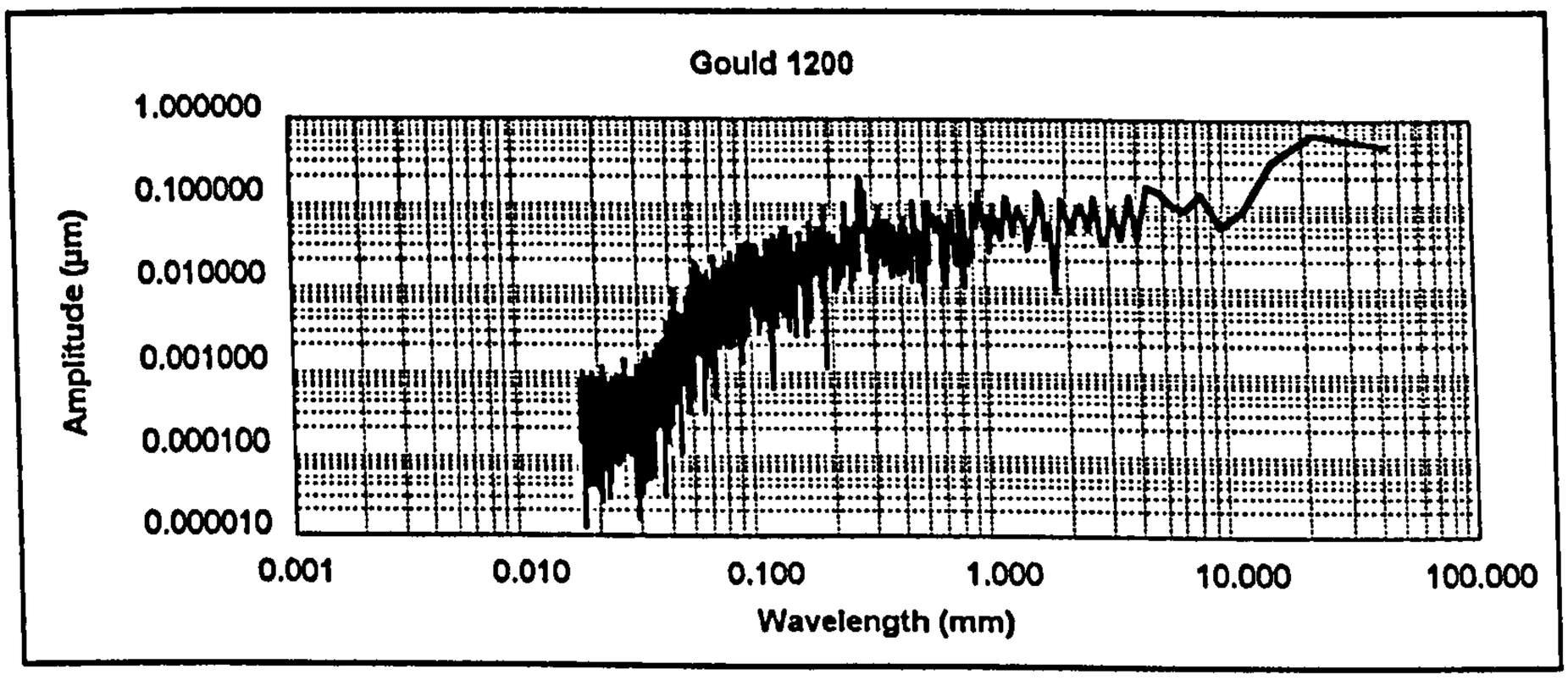


Figure 7.13 *Wavelength content of shaft per Gould 1200.*

Mahr MFU8

Without knowledge of the instrument settings, the profile data for the MFU8 would seem the most suspicious. These settings were as follows: 3 mm tip radius, 100 µm

ordinate spacing, 0.8 mm long pass Gaussian filter. This large difference in tip radius and the application of a relatively long wavelength cutoff drives the following wavelength content.

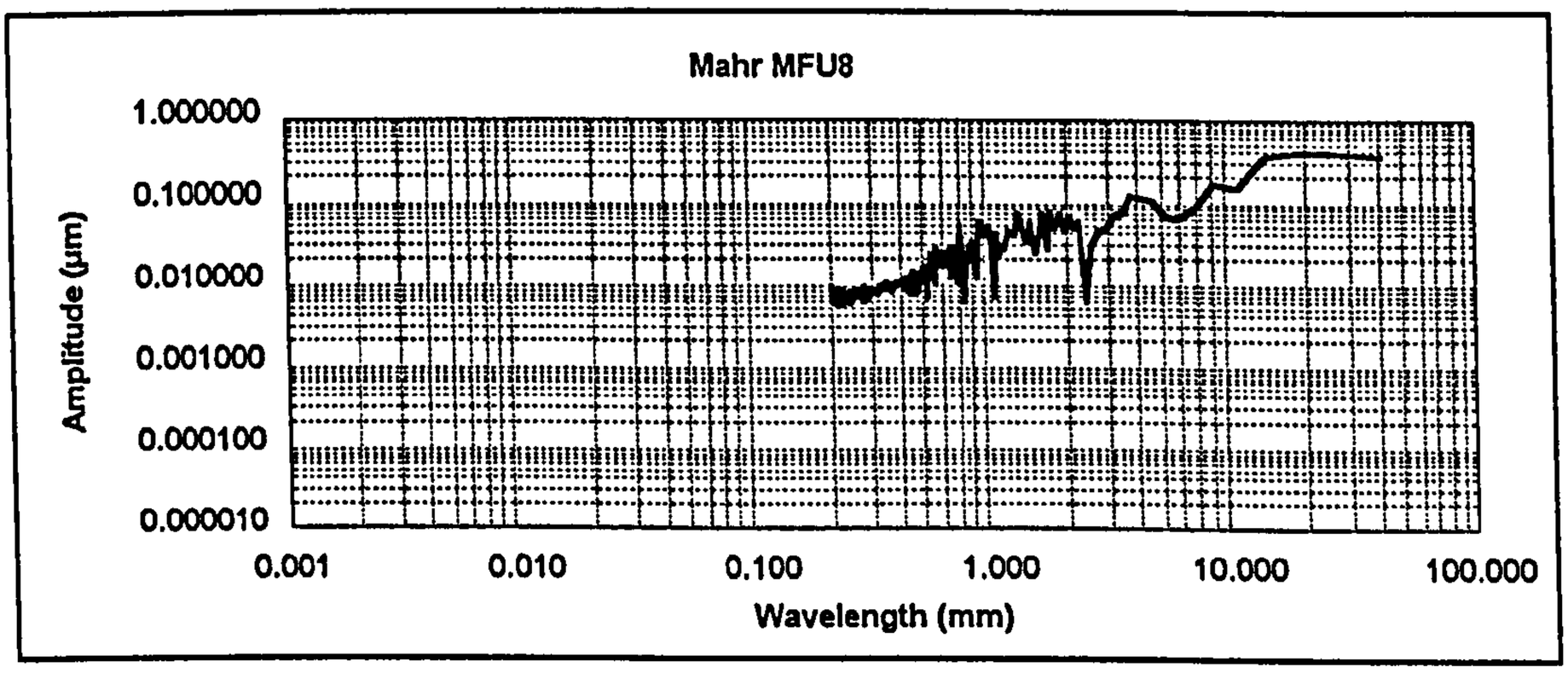


Figure 7.14 Wavelength content of shaft per Mahr MFU8.

Zeiss UPMC 850

The Zeiss CMM was fitted with a relatively small tip radius (as compared to typical CMM applications). Furthermore, an unwanted asperity was present in the data set (see also Chapter 4). For these measurements the tip radius was 0.5 mm, sample spacing 100 µm, and no subsequent filtering was applied.

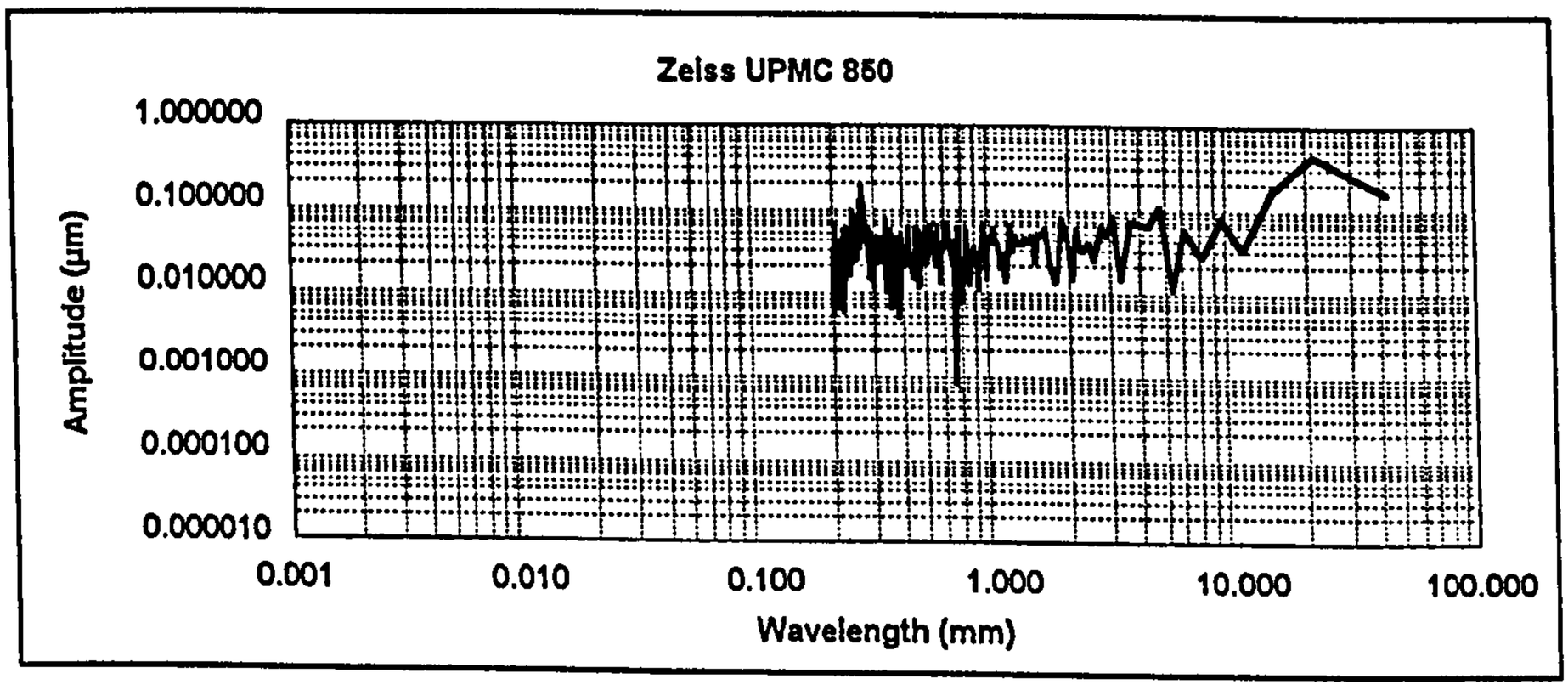


Figure 7.15 Wavelength content of shaft per Zeiss UPMC 850.

Given the common areas of bandwidth between the instruments (as assessed from the above graphs) and the desire to ascertain the peak-to-valley or *t* parameter, the following unified specification was evaluated:

—	0.8 mm	
t		

This specification indicates the assessment of the total peak to valley, *t*, for a data set based on a Gaussian short wavelength limitation at 0.8 mm up to the length of the component. Furthermore, the robust asperity removal as developed in Chapter 4 was applied to the data (see Chapter 4, Section 4.37). The resulting parameter comparison between instruments is significantly improved.

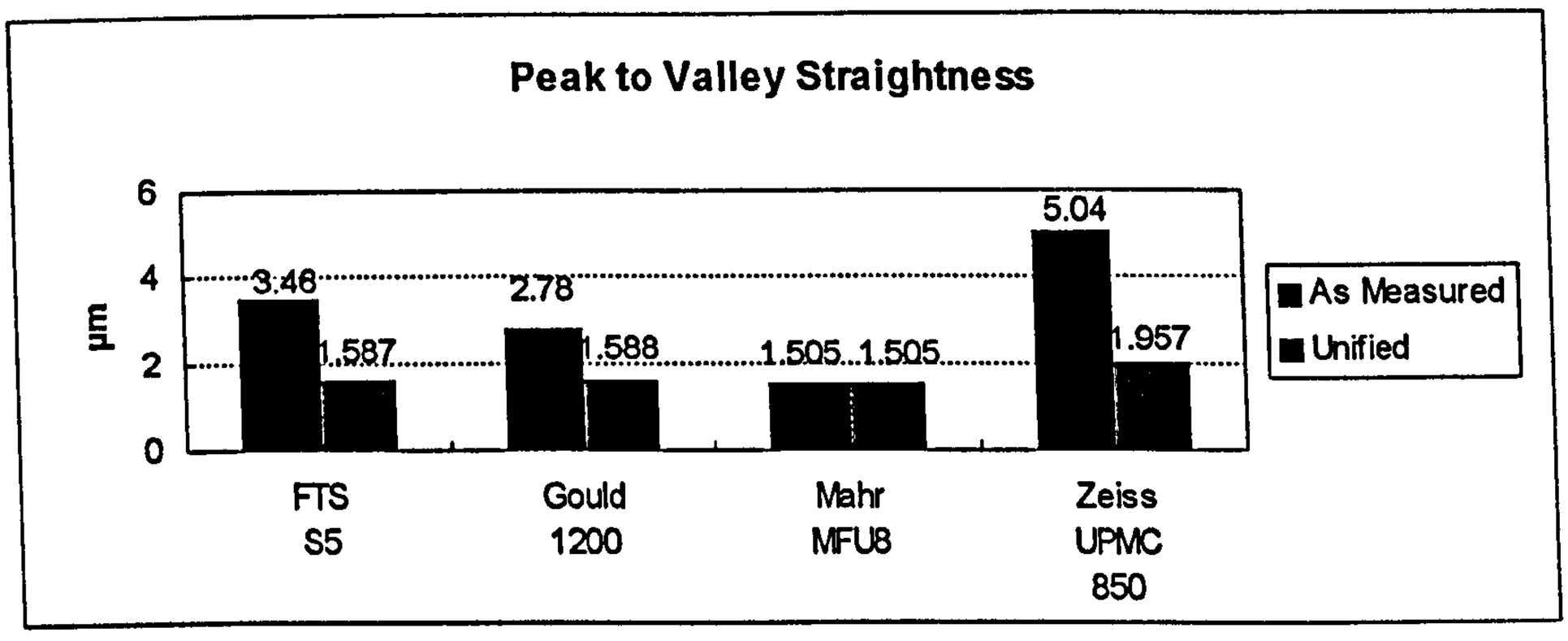


Figure 7.16 *The improvement in correlation as a result of the unification.*

Under the unified approach, agreement is exceptionally good between the Form Talysurf, Gould 1200 and Mahr MFU8 as all values agree within approximately 5%. (The majority of this difference is due to a lower value on the MFU8, which may be the result of the larger tip radius.) Despite this issue, this agreement within 5% is rather encouraging as it is well inside uncertainty estimates typically allowed for in the calibration of profile parameters as performed by accredited laboratories (Nielsen 1996a).

However, despite the application of robust asperity removal and filtering, the Zeiss result still appears to be significantly outside the distribution of the others. (For example, the Zeiss result deviates by 25% relative to the Form Talysurf result.) This deviation in the Zeiss measurements may be the result of a long wavelength error in one of the machine's axes or the results of probing errors. Either error or any combination thereof is very probable - particularly in light of the fact that the estimated uncertainty for this measurement on the Zeiss is on the order of 1.0 μm.

Given the discrepant values from the Zeiss, a proper course of action would be to re-address the uncertainty of the Zeiss (it appears much lower than the expected 1.0 μm). This could be accomplished by the measurement and analysis of a very straight artifact such as an optical flat. (See Section 7.3.1 above.)

7.4 Extension to Three Dimensional Measurement

In the past several years, there has been a growing presence of three dimensional measurement applications in surface metrology (Stout et al. 1990, Stout et al. 1993). Historically, these approaches were costly and very time consuming. However, recent advances in optics (Cohen 1995, Selberg 1993) and high speed scanning (Scott and Morrison 1995) are providing metrological approaches which are making three dimensional assessments much more practical. Some areas that are becoming more functionally relevant include:

- Three dimensional roughness (extending to flatness)
- Cylindricity
- Sphericity
- Conicity

While this presentation of the unified methodology was based primarily on profile data, the concepts can be easily extended to three dimensions. The application of least

squares reference figures, the primary wavelength limitation approaches and unified approach to parameterization can all be directly translated to three dimensions. In addition, the unified specification block can be extended to accommodate multiple bandwidths (associated with the two reference coordinates) and the parameter set can be extended to allow for the development of parameters that are unique to three dimensional assessments.

7.4.1 Flatness

The control of flatness can be very important in many *sealing* applications. In addition, the control of specific wavelength regimes in the three dimensional analysis may be of equal importance - particularly in cases where *conformability* is an issue. The unified specification scheme can easily be extended to accommodate the specification of flatness (see Figure 7.17).

▱	X: — Y: —	∫ 2.5 mm ∫ 2.5 mm	
t			Y.YY µm

Figure 7.17 Flatness per the unified specification scheme. (Y.YY indicates the upper limit of peak to valley departures).

7.4.2 Cylindricity

Another measurement which is gaining a great deal of popularity is that of cylindricity (Dawson 1991, Osanna et al. 1991, Rank Taylor Hobson 1992, Feinprüf Perthen GmbH 1994). In a sense, the measurement of cylindricity represents the combination of linear and circular techniques presented in previous chapters. However, the vast majority of today's "cylindricity" measurements are based on combining multiple circular assessments with the incorporation of radius and position variations (Rank

Taylor Hobson 1992, Feinprüf Perthen GmbH 1994). These measurements typically do not address the controlled limitation of wavelength in the cylinder's axial direction.

The proposed unified cylindricity specification (shown in Figure 7.18) would force the issue of controlling wavelength in both the circular and linear coordinates associated with the cylinder.

(Note: In the proposed specification of cylindricity it should be noted that wavelength limitation is based on wavelength in the linear direction and frequency in the circular direction. This inconsistency in units may be addressed (i.e. changed) if deemed necessary by the appropriate standardization bodies.)

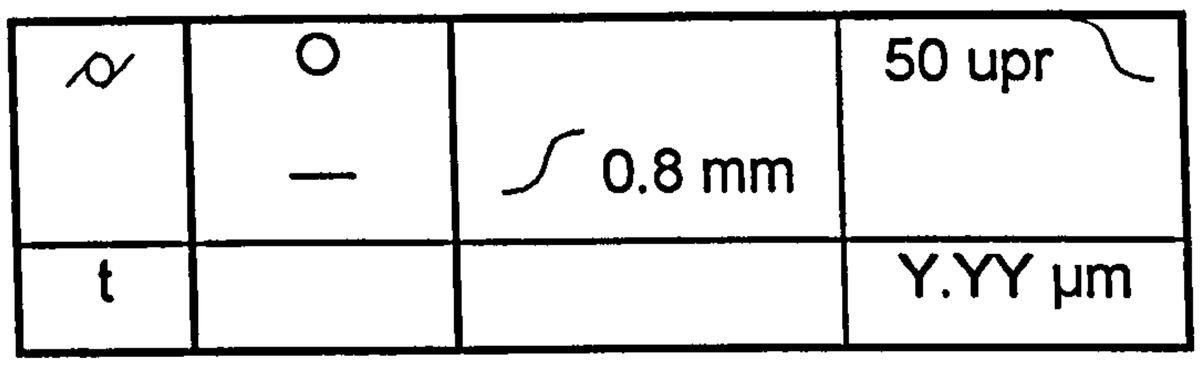


Figure 7.18 *Cylindricity per the unified specification scheme. (Y.YY indicates the upper limit of peak to valley departures).*

7.4.3 Sphericity

The unified specification scheme can be applied to spherical measurement in much the same way as other three dimensional measurements in terms of parameterization. However, due to the three dimensionally closed nature of a sphere, the specification of the wavelength limitation is limited to that of a single transmission (based on a great circle). The sphericity designation symbol is quite similar to the roundness symbol. However for sphericity, a semi-circle (shadow) is drawn inside, indicating a three dimensional feature. An example sphericity specification is provided in Figure 7.19.

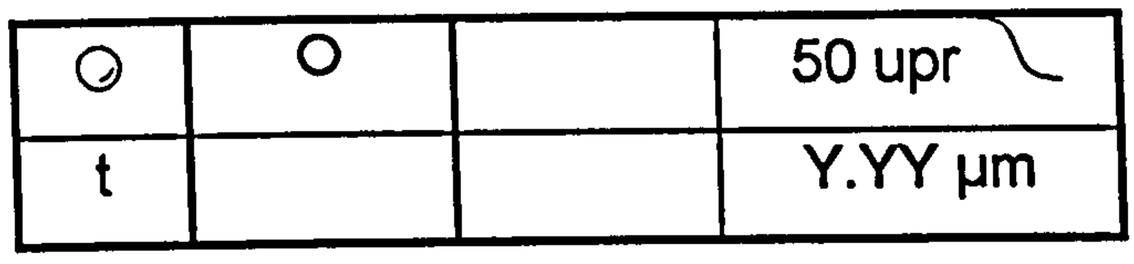


Figure 7.19 *Sphericity per the unified specification scheme. (Y.YY indicates the upper limit of peak to valley departures).*

7.4.4 Conicity

Conicity is perhaps the most difficult of the typical geometries in terms of metrology. This is due, for the most part, to that fact that there is no convenient "conical coordinate system" which instrumentation and mathematics can be built around. Nonetheless, the assessment of conical geometries is very important in many critical applications. Fuel injection components and tapered rolling elements provide examples of critical interfaces where conical surfaces must be controlled to sub-micron levels to ensure proper functionality.

The three dimensional extension of the unified specification scheme accommodates the analysis of conical geometries given two assumptions regarding wavelength limitation:

- Circular wavelength limitation is constant in terms of the angular (i.e. upr) cutoff. This wavelength limitation does not vary with radius. This is consistent with Fourier analysis as well as frequency analysis in spinning or rolling applications.

- Linear wavelength limitation is based on surface length, not the projected length along the axis. The surface length is the functional length in terms of an interface to a mating surface.

Given these assumptions, the specification format for conicity is very similar to that of cylindricity as shown in Figure 7.20.

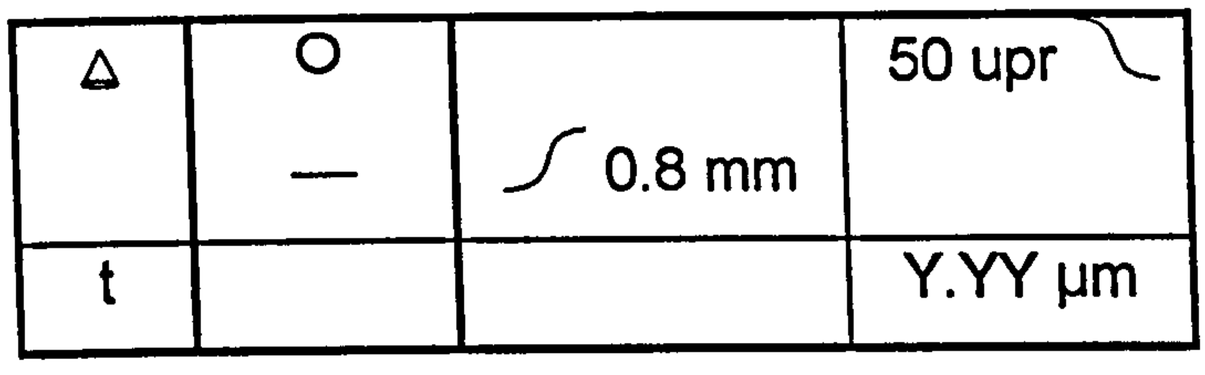


Figure 7.20 *Conicity per the unified specification scheme. (Y.YY indicates the upper limit of peak to valley departures).*

7.5 Proposal for Standardization

The methodology developed and presented in this thesis is only useful if accepted and applied broadly. This can be accomplished to the highest level through standardization. To this end, an outline for a proposed standard is included. In some areas, guidelines for the body text are included, however, in most sections the specific details are purposefully omitted to allow standards bodies to incorporate the necessary information per style practices and consistency with other standards within similar jurisdiction and application.

Proposal for Standard:

Unified Surface Metrology Specification/Application

1.0 Scope
(Coverage of linear, circular ...(future extensions) geometries.)

2.0 References

3.0 Definitions
- 3.1 Coordinate Systems
- 3.2 Geometries
- 3.3 Transmission
- 3.4 Instrument Components
- 3.5 other definitions...
  (as deemed relevant by the standards body)

4.0 Reference Figures
- 4.1 Least Squares

4.1.1 Linear
4.1.2 Circular
4.1.3 (... future extensions)

5.0 Wavelength Limitation
5.1 Gaussian Filters
5.1.1 Stylus Influences
5.1.2 Sampling
5.2 *Ideal* Transmission
5.2.1 Stylus Influences
5.2.2 Sampling
5.3 Stylus-Based Transmission
5.3.1 Sampling

6.0 Parameters
(incorporation of Table 6.1.)
6.1 Statistical Amplitude Parameters
6.2 Extreme Amplitude Parameters
6.3 Spacing Parameters
6.4 Slope Parameters
6.5 Auxiliary Function Parameters

7.0 Drawing Specification and Symbology
7.1 Geometric Reference Elements
7.2 Wavelength Limitation
7.2.1 Required Fields and Values
7.2.2 Default Values
7.3 Parameter specification
7.3.1 Tolerance Limits
7.4 Text based symbology

8.0 Instrumentation
8.1 Instrument Errors
8.1.1 Instrument evaluation
8.2 Calibration
8.2.1 Linear approaches
8.2.2 Circular approaches

7.6 Conclusions

A scheme for the unification of the field of surface metrology has been developed and presented in this thesis. The scheme has been designed based on sound principles and

in a manner such that it will be technically and academically rigorous as well as "acceptable" by the user community. Furthermore, the scheme embodied in this work is easily extensible in order to accommodate past, current and future developments in surface metrology.

While local adoption of this approach will certainly benefit those who do so, the greatest benefits of the scheme will come about through national and international standardization. This will allow a common "language" for surface metrology which will be independent of the geometry being assessed and the instrumentation utilized in its assessment.

One aspect that must be overcome from the user community is the general resistance to change which often occurs in an engineering context - particularly when product liabilities may be affected by a change. This issue is related to "education" within the surface metrology user community and particularly the engineering design and drawing community. Ultimately, the economics associated with the lack of correlation between instruments will provide a "negative" motivation toward the adoption of the scheme. A proactive "education" based approach would certainly be a preferred means for reaching the engineering community which relies on surface metrology.

As discussed in Chapter 2, the need for unification has been historically presented and some isolated efforts have been put forth - none of which have been tremendously successful. However, the work presented in this thesis demonstrates that unification can indeed be achieved in a very practical manner (both technically and economically). Furthermore, these concepts should continue to be advanced - particularly into three dimensional (areal) assessments as discussed above in this chapter. Extension of these unification concepts could then go beyond surface metrology into the realms of surface integrity/sub-surface, chemical/metallurgical and dimensional specifications - thus forming a complete unified system of product specificaiton and measurement communication.

7.7 Closing

Metrology is, in most regards, "a means to an end" rather than something that drives an "end" in and of itself. In other words, the benefits of metrology come in its application towards furthering the understanding of other technical issues. This is apparent in the often quoted observation of Lord Kelvin:

> *"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you scarcely, in your thoughts, have advanced to the stage of Science, whatever the matter may be."*
>
> *Lord Kelvin - circa. 1880*

The development and advancement of metrology is important, valuable and often beneficial in the context of measurement. However, it should not be forgotten that the real benefits to be gained through better metrology come in the forms of medicine, transportation, communication, commerce and other areas that rely heavily on the data generated by metrological tools.

The word "application" has been deliberately included in the title of this work as the methodology developed herein is based on providing maximum benefits to those involved in applying surface metrology and ultimately to society as a whole.

- Mark C. Malburg

References

Abbott E.J., Firestone F.A., 1933, "Specifying Surface Quality", *Mechanical Engineering*, **55**, 569-572

Adcole Corporation, 1993, *Model 911 Product Specifications*, Marlborough, Maryland

Al-Jumaily G. A., Wilson S. R., Jungling K. C., McNiel J. R., Bennett J. M., 1987, "Frequency Response Characteristics of a Mechanical Surface Profilometer", *Optical Engineering*, September 1987, **26**, No. 9, 953-958

Amstutz H., 1985, "Surface Texture: The Parameters", *Sheffield Measurement Division*, MI-TP-003-875

ANSI B89.3.1 1988, "Measurement of Out of Roundness"

ANSI/ASME B46.1, 1985, "Surface Texture (Surface Roughness, Waviness, and Lay)"

Antoniou A, 1993, *Digital Filters - Analysis, Design and Applications*, McGraw Hill, Inc., New York

Anthony G.T., Cox M.G., 1986, "Reliable Algorithms for Roundness Measurement According to BS 3730", *Software for Co-ordinate Measuring Machines*, edited by M.G. Cox and G.N. Peggs, 30-37, National Physical Laboratory, Teddington

ASME/ANSI B46.1 1995, "Surface Texture : Roughness, Waviness and Lay"

ASME/ANSI B46.1.9, 1995, "Surface Texture (Surface Roughness, Waviness, and Lay) - Filtering of Surface Profiles"

ASME/ANSI Y14.5 1994, "Geometric Dimensioning and Tolerancing"

Avallone E.A., Baumeister T., 1979, *Marks' Standard Handbood for Mechanical Engineers*, McGraw Hill, New York

Bendat J. 1986, *Random Data : Analysis and Measurement Procedures*, John Wiley and Sons, New York, NY, USA

Bennett J.M., Mattsson L., 1989, *Introduction to Surface Roughness and Scattering*, Optical Society of America, Washington DC, USA

Bhargava S., 1991, *Personal communication*, Cummins Engine Company, Inc.

Bhargava S., 1993, *Personal communication*, Cummins Engine Company, Inc.

Brown C., 1995, "Fractal Techniques Presentation", *ASME/ANSI B46.1 Fall 1995 Symposium*, Charlotte, NC

BS 1134 1-2 1988, 1990, "Part 1 - Assessment of Surface Texture Methods and Instrumentation, Part 2 - Guidance and General Information"

BS 2634 1-3 1987, 1980, "Roughness Comparison Specimens"

BS 6740 1987, "Determining Departures from Roundness by Measuring Variations in Radius"

BS 6741 1-2 1987, "Surface Roughness - Terminology"

Burton M., 1993, *Machine Tool and Industrial Automation Specifications*, Cummins Engine Company, Inc.

Busch T. Ed. 1989, *Fundamental of Dimensional Metrology - 2nd Edition*, Delmar Publishers Inc., Albany, NY, USA

Campbell J.C., 1972, "Cylinder Bore Surface Roughness in Internal Combustion Engines: Its Appreciation and Control", Wear, **19**, 2, 163-168

Carpinetti L.C.R, Chetwynd D.G., 1994, "A New Strategy for Inspecting Roundness Features", *Precision Engineering*, **16**, 4, October 1994

Castle P., 1993, *Personal communication.* Rank Taylor Hobson

Chen X., Raja J., Simanapalli S., 1994, "Multi-Scale Analysis of Engineering Surfaces", *Sixth International Conference on Metrology and Properties of Surfaces*, Birmingham, England

Chetwynd D.G., 1978, "Slope Measurement In Surface Texture Analysis", *Journal of Mechanical Engineering Science*, **20**, 3, 115-119

Chetwynd D.G., 1979a, "Roundness Measurement Using Limacons", *Precision Engineering*, **1**, 137-141

Chetwynd D.G., 1979b, "The Digitization of Surface Profiles", *Wear*, 57, 137-145

Chetwynd D.G., Phillipson P.H., 1980, "An Investigation of Reference Criteria Used in Roundness Measurement", J. Phys. E, Sci. Inst., **13**, 530-538

Chetwynd D.G., 1985, "Linearized Exchange Algorithms in Metrology", *Software for Coordinate Measuring Machines*, NPL, London, 24-29

Chetwynd D.G., 1985, "On the Definition of Geometric Residuals", *Software for Co-ordinate Measuring Machines*, NPL Conference, Teddington, Middlesex, 42-46

Chetwynd D.G., 1987, "High Precision Measurement of Small Balls", *J. Phys. E. Sci. Instrum.*, **20**, 1179-1187

Chetwynd D.G., 1991, "Algorithms for Computer Aided Precision Metrology", *From Instrumentation to Nanotechnology*, edited by J.W. Gardner and H.T. Hingle, Gordon and Breach Science Publishers

Chetwynd D.G, Liu X., Smith S.T., 1992, "Signal Fidelity and Tracking Force in Stylus Profilometry", *Int. J. Mach. Tools Manufact.*, 32, 239-245

Chetwynd D.G. 1995, *Personal communication.* The Unversity of Warwick

Clark J.R., Grant M.B., 1992, "The Effect of Surface Finish on Component Performance", *Int. J. Mach. Tools Manufact.*, Vol. 32, No. 1/2, 57-66

Cohen D.K. 1995, "Optical Methods for Surface Texture", *ASME/ANSI B46 Surface Texture Workshop*, Charlotte, NC, USA

Crooks C.S., Parker D.D., 1996, "The Importance of Crankshaft Surface Texture to Bearing System Reliability", *SAE Paper 960983*

Dagnall M.A. 1980, *Exploring Surface Texture*, Rank Taylor Hobson, Leicester, England

Dawson D., 1991, "Cylindricity and its Measurement", *Metrology and Properties of Engineering Surfaces - International Conference*, April 1991

DeVries W.R., Li C., 1985, "Algorithms to Deconvolve Stylus Geometry from Surface Profile Measurements", *Trans. ASME - Journal of Engineering for Industry*, **107**, 167-174

Diaz C., Hopp T.H., 1993, "Testing of Coordinate Measuring System Software", *Proc. ASQC Meas. Quality Conf.*, Gaithersburg, MD, October 1993

DIN 4760 1982, "Form deviations, Concepts; Classification system"

DIN 4761 1978, "Surface Character - Geometrical Characteristics of Surface Texture Terms, Definitions, Symbols"

DIN 4762 1981, "Part 1. Surface Roughness; Terminology"

DIN 4765 1974, "Determination of the Bearing Area Fraction of Surface, Terms"

DIN 4766 1981, "Surface Roughness"

DIN 4768 1974, 1978, "Determination of Surface Roughness Parameters"

DIN 4769 1-4 1972, 1974, "Roughness Comparison Specimens"

DIN 4772 1979, "Electrical Contact (Stylus) Instruments for the Measurement of Surface Roughness by the Profile Method"

DIN 4774 1981, "Measurement of the Depth of Waviness by Means of Electrical Contact Stylus Instruments"

DIN 4775 1982, "Measuring the Surface Roughness of Workpieces"

DIN 4776 1990, "Measurement of Surface Roughness - Parameters Rk, Rpk, Rvk, Mr1, Mr2 for Describing the Material Portion in the Roughness Profile"

Donaldson R.R., 1972, "A Simple Method for Separating Spindle Error from Test Ball Roundness Error", *Annals of CIRP*, **21**, 156-126

Drews W.E, 1994, *Personal communication.* International Marketing Services

Edison F., Parry G., 1985, "Integrity of Software Associated with Co-ordinate Measuring Machines", *Software for Co-ordinate Measuring Machines*, NPL Conference, Teddington, Middlesex, 51-52

Farago F.T., 1982, *Handbook of Dimensional Measurement*, Industrial Press, NY, USA

Federal Products, 1994, *Surfanalyzer 5000/400 Surface Analysis System*, Providence, Rhode Island, USA

Feinprüf Perthen GmbH, 1989, *Operating Instructions No. 6800462 - Perthometer S6P Serial Interface*, Göttingen, Germany

Feinprüf Perthen GmbH, 1992, *Perthometer. Accessories for Surface Texture and Recording Instruments*, Göttingen, Germany

Feinprüf Perthen GmbH, 1993, *Formtester - Programme Survey*, Göttingen, Germany

Feinprüf Perthen GmbH, 1994, *Operating Instructions No. 3752059 - Formtester MFU8 with Form Computer HP Series 300*, Göttingen, Germany

Firestone F.A., Durbin F.M., Abbot E.J., 1932, "Test for Smoothness of Machined Surfaces", *Metal Progress*, 21, 57-59

Fischer H., 1982, "Honing", *SME Technical Paper*, MR82-939

Forbes A.B., 1989, "Least Squares Best-Fit Geometric Elements", *NPL Report DITC 140/89*, National Physical Laboratory - Division of Information Technology and Computing

Ghabrial S.R., 1991, "A Methodology for Relating Surface Processing and Characterisation to Functional Performance", *Metrology and Properties of Engineerng Surfaces*, International Conference, April 1991

GIDEP, 1988, *Alert X1-A-88-01A*, September 1988

Grant M.B., 1991, *Personal Communication.* Cummins Engine Company, Inc.

Green E., 1967, "Review of Surface Texture Measurement and the Associated Metrological Problems", *Proc. Instn. Mech. Engrs.*, **182**, 3K, 330-343

Greenwood J.A., 1984, "A Unified Theory of Surface Roughness", *Proc. R. Soc.*, **A393**, 133-157

Hager F.M., 1995, *Personal communication.* Cummins Engine Company, Inc.

Hildebrandt M., 1994, *Personal communication.* Feinprüf Perthen GmbH

Hillman W., Kranz O., Eckolt K., 1984, "Reliability of Roughness Measurements Using Contact Stylus Instruments with Particular Reference to Results of Recent Research at the Physikalisch-Technische Bundesanstalt", *Wear*, 97, 27-43

Hillman W., 1992, "Calibration of Contact Stylus Instruments Within the Deutscher Kalibrierdienst (DKD)", *ISO/TC57/SC2/N122*

Hochwart E., 1995, *Personal communication.* Mahr Corporation

Hopp T.H., Levenson M.S., 1995, "Performance Measures for Gemetric Fitting in the NIST Algorithm Testing and Evaluation Program for Coordinate Measurement Systems", *J. Res. National Institute for Standards and Technology*, **100**, 5, 563-574

ISO 468 1982, "Surface Roughness - Parameters, Their Values and General Rules for Specifying Requirements"

ISO 1101 1983, "Technical Drawings - Geometric Tolerancing -Tolerancing of Form, Orientation, Location and Runout - Generalities, Definitions, Symbols, Indications on Drawings"

ISO 1302 1978, "Technical Drawings - Methods of Indicating Surface Texture on Drawings"

ISO 2632 1-3 1975, 1977, 1979, "Roughness Comparison Specimens"

ISO 3274 1995, "Instruments for the Measurement of Surface Roughness by the Profile Method - Contact (Stylus) Instruments of Consecutive Profile Transformation - Contact Profile Meters, System M"

ISO 4287 1984, "Surface Roughness Terminology"

ISO 4288 1985, "Rules and Procedures for the Measurement of Surface Roughness Using Stylus Instruments"

ISO 4291 1985, "Methods for the Assessment of Departure from Roundness - Measurement by Variations in Radius"

ISO 4292 1985, "Methods for the Assessment of Departure from Roundness - Measurement by Two and Three Point Methods"

ISO 6381 1985, "Measurement of Roundness - Terms, Definitions and Parameters of Roundness"

ISO 11562 1995, "Metrological Characterization of Phase Corrected Filters and Transmission Bands for use in Contact (Stylus) Instruments"

ISO 12181 1995, "(Committee Draft) Roundness, Parts 1-3"

ISO/DIS 4287 1995, "Geometrical Geometrical Product Specifications (GPS) - Surface Texture - Profile Method - Part 1: Terms, Definitions and Parameters of Surface Texture"

ISO/DIS 12085 1995, "Surface Roughness and Waviness - Motif Method"

ISO/DIS 13565 1-3 1994, 1995, "Characterization of Surfaces Having Stratified Functional Properties"

ISO/TC 57/SC 1, 1988, "ISO Draft Proposal: For Metrological Characteristics of Phase Corrected Filters for the use in Electrical Contact Stylus Instruments", N87

ISO/TR 14638, 1995, "Geometrical Product Specifications (GPS) - Masterplan"

Kinsey D., Chetwynd D. G., 1973, "Some Aspects of the Application of Digital Computers to the On-Line Measurement of Surfaces", *ACTA IMEKO*, B-523, 601-616

Krystek M., 1994, "Non-Uniform Rational B-Splines for the Separation of Waviness and Form", *ISO/TC3-10-57/JHG/TG 1 N 11*

Lavoie R.A., 1991, "Importance of Surface Texture Evaluation Grows", *Quality*, October 1991, 29-30

Lavoie R.A., 1992, "Selecting Surface Roughness Parameters", *Quality*, October 1992, 52-53

Lavoie R.A., 1994, "Stop Shaft Leaks with Better Surface Finish", *Quality in Manufacturing*, July-August 1994

Lenthall J.S., 1996, *Personal communication*. GKN Sheepbridge Stokes Ltd.

Li M., 1991, "Interpretation and Evaluation of Stylus Profiling Techniques", *PhD Thesis*, University of Warwick, Coventry

Liu X., Smith S.T., Chetwynd D.G., 1992, "Frictional Forces Between a Diamond Stylus and Specimens at Low Load", *Wear*, 157, 279-294

Liu X., Chetwynd D.G., Smith S.T., Wang W., 1993, "Improvement of the Fidelity of Surface Measurement by Active Damping Control", *Measurement Science Technology*, 4, 1330-1340

Liu X., 1994, *PhD Thesis*, University of Warwick.

Ludema K.C., 1993, *Friction, Wear, Adhesion and Lubrication - Tribology*, The University of Michigan, Ann Arbor, Michigan, USA

Lukjanov V.S., 1967, "Specimens used for Checking Contact Profile Recording Instruments and M-System Profile Meters and Certification of Specimens", *Proc. Instn. Mech. Engr.*, **182**, Part 3K, 406-415

Malburg M.C., 1989, "The Characterization of Surface Texture Generated by Multi-Process Manufacture", *Master's Thesis*, Michigan Technological University, Houghton, Michigan, USA

Malburg M.C., Grant M.B., 1992, "The Characterization of Two-Process Surface Texture Using the Cumulative Gaussian Probability Distribution", *ISO TC 57/SC 1/WG 3/N 21*

Malburg M.C., Raja J., 1993, "Characterization of Surface Texture Generated by Plateau Honing Process", *Annals of the CIRP*, **42**, 1, 637-639

Malburg M.C., Chetwynd D.J., Raja J., 1996, "Local Slope Analysis in the Stylus Based Assessment of Surface Integrity", accepted for publication in *Tribology International* January 1996

McCool J.I., 1984, "Assessing the Effect of Stylus Tip Radius and Flight on Surface Topography Measurement", *Trans. ASME - Journal of Tribology*, 106, 202-210

Mestre M., Abou-Kandil H., 1993, "Linear Prediction of Signal Applied to Dimensional Metrology of Industrial Surfaces", *Measurement*, **11**, 119-134

MIL-STD-105D, 1963, "Sampling Procedures and Tables for Inspection by Attribute", 29 April, 1963

Mummery L., 1992, *Surface Texture Analysis - The Handbook*, Hommelwerke, GmbH

Munro R., Hughes G.H., 1970, "Paper 8 - Piston Ring Application in Diesel Engines", *A.E. Technical Symposium*, June 1970

Neter J., Wasserman W., Kutner M. H., 1985, *Applied Linear Statistical Models*, Richard D. Irwin, Inc, Homewood, IL, USA

Neumann H.J., 1990, *Coordinate Metrology - Technology and Application*, Verlag Moderne Industrie AG & Co., Landsberge, Germany

Nielsen H.S., 1989, "Calibration of Surface Roughness Instruments", *VDI Berichte No. 761*, 365-370.

Nielsen H.S., 1991, "The Effect of Calibration in Surface Roughness Measurements", *Measurement*, 9, 3, 111-114

Nielsen H.S., 1993, *Personal communication.* Cummins Engine Company Inc.

Neilsen H.S., Malburg M.C., 1996, "Traceability and Correlation in Roundness Measurement", accepted for publication in *Precision Engineering the Journal of the ASPE*

Nielsen H.S., 1996a, *Personal communication.* Cummins Engine Company Inc.

Nielsen H.S., 1996b, *Cummins Measurment Systems Handbook Training*, Cummins Engine Company Inc., Columbus, IN, USA

O'Connor, Spedding, 1992, "Use of a Complete Surface Profile Description to Investigate the Cause and Effect of Surface Features", *Int. J. Mach. Tools Manufact.*, **32**, 1-2, 147-154

Osanna P.H., Durakbasa N.M., Cakmakci M., Oberlander R., 1991, "Cylindricity - A Well Known Problem and New Solutions", *Metrology and Properties of Engineering Surfaces - International Conference*, April 1991

Paterson D., 1985, "Pragmatic Statistics - or - Another way to look at a group of measurements", *Software for Co-ordinate Measuring Machines*, NPL Conference, Teddington Middlesex, September, 47-50

Press W.H., Teukolsky S.A., Vettering W.T., Flannery B.P., 1992, *Numerical Recipies in C - The Art of Scientific Computing, Second Edition*, Cambridge University Press, Cambridge

Radhakrishnan V., 1970, "Effect of Stylus Radius on the Roughness Values Measured with a Stylus Instrument", *Wear*, **16**, 325-335

Radhakrishnan V., 1971, "On an Appropriate Radius for the Enveloping Circle for Roughness Measurement in the E System", *Annals of the CIRP*, 20

Radhakrishnan V, Shunmugam M.S., 1974, "Computation of 3D Envelope for Roundness", *Int. J. Mach. Tool Des. Research*, **14**, 211-6

Raja J., Radhakrishnan V., 1979 "Filtering of Surface Profiles Using Fast Fourier Transform", *International Journal of Machine Tool Design and Research*, **19**, 133-141

Rank Taylor Hobson 1985, *Form Talysurf - Software - Operator's Handbook*, Leicester, England

Rank Taylor Hobson, 1992, *Talyrond 250 - Operator's Handbook*, Leicester, England

Rank Taylor Hobson, 1995, *Form Talysurf Series Operator's Handbook*, Leicester, England

Reason R.E., 1965, "Some approaches to the measurement of waviness", *Proceedings of the 6th International MTDR Conference*, 293-299

Reason R.E., 1966, *Report on the Measurement of Roundness*, Rank Taylor Hobson, Leicester, England

Reason R.E., 1967, "Workshop Requirements of Surface Measurement", *Proc. Instn. Mech. Engrs*, **182**, Part 3K

Reason R.E., 1970, "The Measurement of Surface Texture ", *Modern Workshop Technology*, Part 2, Macmillan, London

Reason R.E., 1973, "Measurement of Surface Topography and Accuracy of Stylus Instruments", *Proceedings of the International Conference on Surface Technology*, May 1973, Society of Manufacturing Engineers, Dearborn, Michigan, USA

Reid C., 1992, *Signal Processing in C*, Wiley, New York

Rose G.C., 1993, *Personal communication*. Cummins Engine Company, Inc.

Rousseeuw P.J., Leroy A.M., 1987, *Robust Regression and Outlier Detection*, John Wiley & Sons, New York

Russ J.C., 1994, *Fractal Surfaces*, Plenum Press, New York

Salsbury J.G., 1996, "Uncertainty of Coordinate Measurement Process Due Unwanted Asperities", *ASPE 1996 Annual Meeting*, Monterey, California

Scott P.J., 1986, "Surface Metrology: A New Philosophical Approach", *Wear*, **109**, 267-274

Scott P.J. 1992a, "A New Unified Approach to Surface Texture Instrumentation", *Proceedings of the ASPE Annual Meeting, 1992*

Scott P.J., 1992b, "The Mathematics of Motif Combination and Their use for Functional Simulation", *Int. J. Mach. Tools Manufact.*, 32, 1/2, 69-73

Scott P.J., 1995, *Personal Communication.* Rank Taylor Hobson

Scott P.J. 1996, *Personal Communication.* Rank Taylor Hobson

Scott P.J., Morrison E., 1995, "Recent Advances in Stylus Based 3-D Surface Measurement and Characterization", *Proceedings of the 1995 ASPE Annual Meeting*

Selberg L.A., 1993, *Advanced Topics in Interferometry*, Zygo Corporation, East Rochester, NY, USA

Sharman H.B., 1967a, "Influence of Sample Size and the Relationships Between the Common Surface Texture Parameters", *Proc. Instn. Mech. Engrs*, **182**, 3K, 416-424

Sharman H.B., 1967b, "Calibration of Surface Texture Measuring Instruments", *Proc. Instn. Mech. Engrs*, **182**, 3K, 319-326

Shunmugam M.S. and Radhakrishnan V., 1974, "Computation of the Three-Dimensional Envelope for Roughness Measurement", *International J. Mach. Tool Des. Res.*, **14**, 211-216, 1974

Shunmugam, M.S, Rahadkrishnan V., 1975, "Two- and Three-Dimensional Analysis of Sufaces According to the E-System", *Proc. Instn. Mech. Engrs.*, **188**, 691-697

Singleton R.C., 1969, "An Algorithm for Computing the Mixed Radix Fast Fourier Transform", *IEEE Trans. Audio Electroacoust.*, **AU-17**, 93-10, June 1969

SKF Component Systems Co., 1992, *Software Capabilties Overview: The SKF STEYR MFFA Form Analysis Computer*

Spragg R.C., 1967, "Accurate Calibration of Surface Texture and Roundness Measuring Instruments", *Proc. Instn. Mech. Engr.*, **182**, Part 3K, 397-405

Spragg R.C., Whitehouse D.J., 1970, "A New Unified Approach to Surface Metrology", *Proc. Inst. Mech. Engrs.*, **185**, 47-71

Starbuck T.A., 1992, "Software Features for Removing Unwanted Features on Components when Performing Automatic Form Measurement", *Int. J. Mach. Tools Manufact.*, **32**, 101-8

Stedman M., 1987, "Basis for Comparing the Performance fo Surface-Measuring Machines", *Precision Engineering*, **9**, 3, 149-152

Stout K.J., Davis E.J., Sullivan P.J., 1990, *Atlas of Machined Surfaces*, University Press, Cambridge

Stout K.J., Sullivan P.J., Dong W.P., Mainsah E., Luo N., Mathia T., Zahouani H., 1993, "The Development of Methods for the Characterisation of Roughness in 3 Dimensions - Phase II Report", *EC Contract 3374/1/0/170/90/2*

Sullivan P.J., Stout K.J., 1992, "The Specification of a Flexible File Format for Storage and Retrieval of Engineering Topographic Data", *Proc. 7th ASPE Conference*, October 1992

Tabenkin A., 1991, *Personal communication.* Federal Products Corp.

Tallian T.E., 1991, *Simplified Contact Fatigue Life Prediction Model: Parts I & II*, Tallian Consulting Corp., Newtown Square, PA

Tallian T.E., 1993, "The Influence of Asperity Statistics on Surface Distresss and Spalling Life of Hertzian Contacts", *Tribology Transactions*, **36**, 1, 35-42

Taniguchi N., 1983, "Current Status in, and Future Trends of, Ultraprecision Machining and Ultrafine Materials Processing", *Annals of CIRP*, 32/2, 573-582

Thomas T.R., Sayles R.S., 1978, "Some Problems in the Tribology of Rough Surfaces", *Tribology International*, **11**, 89

Thomas T.R., 1982, *Rough Surfaces*, Longman Group, Ltd., London

Thwaite E.G., 1973, "A method of Obtaining an Error Free Reference Line for the Measurement of Straightness", *Messtechnik*, **10**, 317-318

Traband M.T., Joshi S., Wysk R.A. Cavalier T.M., 1989, "Evaluation of Straightness and Flatness Tolerances Using Minimum Zone", *Manufacturing Review*, **2**, 3, 189-195

Trautwein R, 1978, "Characteristic Values for Determining and Evaluating the Surface of Cylinder Bores", *Mercury Marine Company*

Vemuri V., 1988, *Artificial Neural Networks: Theoretical Concepts*, Computer Society Press of the IEEE, Washington, DC, USA

Von Weingraber H., 1956, "Zur definition der Oberflachenrouheit Werk Strattstechnik", *Masch. Bau*, **46**

Von Weingraber H., 1957, "Uber die Eignung des Hullprofils als Bezugslinie für Messung der Rauheit", *Microtechnic*, **11**, 6-17

Vorburger T.V., 1993, "Surface Texture Analysis", *Quality Expo Tutorial*, ASQC, Rosemont, IL, April 1993

Wade D., 1991, *Personal communication.* Cummins Engine Company, Inc.

Walton T.P., 1996, *Personal communication.* Cummins Engine Company, Inc.

Wang W.L., 1995, *Enhancement of Fidelity of Surface Measurement Systems*, PhD Thesis, University of Warwick, Coventry

Wasilesky R., 1994, *Personal communication.* Hommel America Inc.

Williamson J.B.P., 1967, "Microtopography of Surfaces", *Proc. Instn. Mech. Engrs*, **182**, 3K, 21-30

Whitehouse D.J., Reason R.E., 1965, *The Equation of the Mean Line of Surface Texture Found by and Electric Wave Filter*, Rank Taylor Hobson

Whitehouse D.J., 1967, "An Improved Wavelfilter for use in Surface Finish Measurement", *Proc. Instn. Mech. Engr.*, **182**, Part 3K, 306-318

Whitehouse D.J., Archard J.F., 1969, "The Properties of Random Surfaces of Significance in their Contact", *Proc. Roy. Soc. Lond. A*, **316**, 97-121

Whitehouse D.J., 1974, "Theoretical Analysis of Stylus Integration", *Annals of the CIRP*, **23**, 181-182

Whitehouse D.J., 1976, "Some Theoretical Aspects of Error Separation Techniques in Surface Metrology", *J. Phys. E. Sci. Instrum*, **20**, 531-536

Whitehouse D.J., 1978, "Surfaces - A Link Between Manufacture and Function", *Proc. Instn. Mech. Engr.*, **192**, 179-188

Whitehouse D.J., 1982, "The Parameter Rash - Is There a Cure?", *Wear*, **83**, 75

Whitehouse D.J., 1985, "Assessment of Surface Finish Profiles Produced by Multi-Process Manufacture", *Proc. Instn. Mech. Engrs*, **199**, 4, 263-270

Whitehouse D.J., 1987a, "Radial Deviation Gauge", *Precision Engineering*, **9**, 4, 201-209

Whitehouse D.J., 1987b, "Instrument Science and Technology - Surface Metrology Instrumentation", *J. Phys. Sci. Instrum.*, **20**, 1145-1155

Whitehouse D.J., 1988, "A Revised Philosophy of Surface Measuring Systems", *Proc. Inst. Mech. Eng.*, **202**, 169-185

Whitehouse D.J., 1990, "Richard Edmond Reason", *Biographical Memoirs of Fellows of the Royal Society*, **36**

Whitehouse D.J., 1993, "Manufacture to Function - In Optics", *Annals of the CIRP*, **42**, 1, 641-645

Whitehouse D.J., 1994, *Handbook of Surface Metrology*, Institute of Physics Publishing, Bristol

Whitehouse D.J., 1995, *Personal communication.* The University of Warwick.

Wilde M., 1994, *Personal communication.* Feinprüf Perthen GmbH

Yau H.T., Menq C. ., 1992, "A Unified Least Squares Approach to the Evaluation of Gemoetric Errors", *ASME Winter Annual Meeting*, PED-Vol. 56

Yau H.T., 1995, "Generalization and Verification of Vectorial Tolerances", *Proceedings of the ASPE 10th Annual Meeting*

Yhland E.M., 1967, "Waviness Measurement - An Instrument for Quality Control in the Rolling Bearing Industry", *Proc. Instn. Mech. Engr.*, **182**, Part 3K, 438-445

A Unified Methodology for the

Application of Surface Metrology:

Appendix A

Reference Figure Implementations

```
/*************************************************************************/
void Regress (float huge *x, float huge *y, int num, double *slope, double *intcpt)
{
   int i ;
   double x_sum, xx_sum,
          y_sum, xy_sum ;

   x_sum = xx_sum = y_sum = xy_sum = 0.0 ;

   for (i=0; i<num; i++)
   {
      x_sum += x[i] ;
      xx_sum += ((double)x[i]*x[i]) ;
      y_sum += y[i] ;
      xy_sum += ((double)x[i]*y[i]) ;
   }

   *slope = (xy_sum - (x_sum*y_sum)/num) /
            (xx_sum - (x_sum*x_sum)/num) ;
   *intcpt = (y_sum - (*slope)*x_sum) / num ;

} /* end Regress */
```

```
/*******************************************************/
double Min_Zone (float huge *x, float huge *z, int num,
                 double *mz_slope, double *mz_intercept1,
                 double *mz_intercept2)
{
   int i, ref1, ref2, peak_ref, valley_ref,
       nm1, done ;
   double min_zone, min_slope, max_slope,
          slope, intercept, theta,
          dist, vert_dist, max_vert_dist ;

   nm1 = num - 1 ;
   min_zone = 99.9e99 ;
   // look at valley side convex hull
   ref1 = 0 ;
   done = FALSE ;
   do
   {
      min_slope = 99.9e99 ;
      for (i=ref1+1; i<nm1; i++)
      {
         if ((z[i] < z[i-1]) && (z[i] < z[i+1])) // local valley
         {
            slope = (z[i]-z[ref1])/(x[i]-x[ref1]) ;
            if (slope < min_slope)
            {
               min_slope = slope ;
               ref2 = i ;
            }
         }
      }
      // check last point individually
        slope = (z[i]-z[ref1])/(x[i]-x[ref1]) ;
      if (slope < min_slope)
      {
         min_slope = slope ;
           ref2 = i ;
      }

      // convex hull line segment between ref1 & ref2, slope is min_slope
      slope = min_slope ;
      intercept = z[ref1] - slope*x[ref1] ;
      // find highest peak relative to convex hull line
      // look vertically for sake of speed, convert to normal after finding
      // use peak_ref to hold the max excursion point
      max_vert_dist = -99.9e99 ;
      for (i=0; i<num; i++)
      {
         if ((i==0) || (i==num) ||
             ((z[i] > z[i-1]) && (z[i] > z[i+1])))
         {
            vert_dist = z[i] - (slope*x[i] + intercept) ;
            if (vert_dist > max_vert_dist)
            {
               max_vert_dist = vert_dist ;
               peak_ref = i ;
            }
         }
      }
```

```
// convert distance to a normal distance
theta = atan(slope) ;
dist = max_vert_dist*cos(theta) ;
if (dist < min_zone)
{
   min_zone = dist ;
   *mz_slope = slope ;
   *mz_intercept1 = intercept ;
   *mz_intercept2 = z[peak_ref] - slope*x[peak_ref] ;
}

ref1 = ref2 ;
   if (ref1 >= nm1)
      done = TRUE ;

} while (!done) ;

// now look at peak side convex hull
ref1 = 0 ;
done = FALSE ;
do
{
   max_slope = -99.9e99 ;
   for (i=ref1+1; i<nm1; i++)
   {
      if ((z[i] > z[i-1]) && (z[i] > z[i+1])) // local peak
      {
         slope = (z[i]-z[ref1])/(x[i]-x[ref1]) ;
         if (slope > max_slope)
         {
            max_slope = slope ;
            ref2 = i ;
         }
      }
   }
// check last point individually
  slope = (z[i]-z[ref1])/(x[i]-x[ref1]) ;
if (slope > max_slope)
{
   max_slope = slope ;
     ref2 = i ;
}

// convex hull line segment between ref1 & ref2, slope is max_slope
slope = max_slope ;
intercept = z[ref1] - slope*x[ref1] ;
// find deepest valley relative to convex hull line
// look vertically for sake of speed, convert to normal after finding
// use peak_ref to hold the max excursion point
// since we are looking down initialize high
max_vert_dist = 99.9e99 ;
for (i=0; i<num; i++)
{
   if ((i==0) || (i==num) ||
       ((z[i] < z[i-1]) && (z[i] < z[i+1])))
   {
      vert_dist = z[i] - (slope*x[i] + intercept) ;
      if (vert_dist < max_vert_dist)  // remember we are looking down
      {
         max_vert_dist = vert_dist ;
         valley_ref = i ;
      }
   }
}
```

```
// convert distance to a normal distance
theta = atan(slope) ;
dist = -max_vert_dist*cos(theta) ;   // max_vert is negative
if (dist < min_zone)
{
   min_zone = dist ;
   *mz_slope = slope ;
   *mz_intercept1 = intercept ;
   *mz_intercept2 = z[valley_ref] - slope*x[valley_ref] ;
}

ref1 = ref2 ;

if (ref1 >= nm1)
      done = TRUE ;
   } while (!done) ;
   return (min_zone) ;
} // end Min_Zone
```

Appendix B

Tip Convolution Data

B.1 Motif Based Tip Convolution Subroutine Listing

```
/************************************************/
int Convolve_Radius (float huge *raw_array_um, float huge *rad_array_um, int num,
                     float radius_um, float spacing_um)
{
// raw_array holds the input data array
// radius convolved data is returned in rad_array
// num contains the number of points in the arrays
// radius_um is the value of radius to be convolved
// spacing_um is the data point spacing

   int i, j, done, count,
       ref1, ref2, ref3,
       left_radius_limit_ordinate,
       right_radius_limit_ordinate ;
   int *contacts ;
   float x1, x2, x3,
         test, max_height ;

   done = FALSE ;
   ref1 = 0 ;
   ref2 = 1 ;
   ref3 = 2 ;
   x1 = 0.0 ;
   count = 0 ;              // hold the count of contacting points

// temporarily hold the references of contacting points in rad_array_um
do
{
   x2 = spacing_um*(ref2-ref1) ;
   x3 = spacing_um*(ref3-ref1) ;
   if (Can_Touch (x1, raw_array_um[ref1], x2, raw_array_um[ref2],
       x3, raw_array_um[ref3], radius_um))
   {
      rad_array_um[count++] = ref1 ;  // hold the refs of contacting points
      ref1 = ref2 ;
      ref2 = ref3 ;
      ref3++ ;
      if (ref3 >= num)
         done = TRUE ;
   }
   else  // did not touch
   {
```

```
if (ref1 == 0)
{
   ref2 = ref3 ;
   ref3++ ;
   if (ref3 >= num)
      done = TRUE ;
}
else
{
      ref2 = ref1 ;
      ref1 = (int)rad_array_um[--count] ;
   }
  }
} while (!done) ;

// check to see if current middle point is touching
ref2 = ref1 ;
ref1 = rad_array_um[count-1] ;
ref3 = num-1 ;
x2 = HEAD.spacing_um*(ref2-ref1) ;
x3 = HEAD.spacing_um*(ref3-ref1) ;
if (Can_Touch (x1, raw_array_um[ref1], x2, raw_array_um[ref2],
               x3, raw_array_um[ref3], radius_um))
   rad_array_um[count++] = ref2 ;
rad_array_um[count++] = ref3 ;

// rad_array_um now holds refs of contacting points
contacts = (int *)malloc(sizeof(int)*count) ;
if (contacts == NULL)
{
   printf ("Insufficent memory for contact array.") ;
   return FALSE ;
}

for (i=0; i<count; i++)
   contacts[i] = (int)rad_array_um[i] ;

for (i=0; i<num; i++)
{
   left_radius_limit_ordinate = i - radius_um/HEAD.spacing_um ;
   if (left_radius_limit_ordinate < 0)
      left_radius_limit_ordinate = 0 ;
   right_radius_limit_ordinate = i + radius_um/HEAD.spacing_um ;
   if (right_radius_limit_ordinate >= num)
      right_radius_limit_ordinate = num-1 ;
```

```
    j = 0 ;
    while (contacts[j]<left_radius_limit_ordinate)
       j++ ;

    max_height = -99.9e99 ;
      while (contacts[j]<=right_radius_limit_ordinate)
    {
       test = Height((i-contacts[j])*HEAD.spacing_um, raw_array_um[contacts[j]], radius_um) ;
       if (test > max_height)
                max_height = test ;
          j++ ;
       }
       rad_array_um[i] = max_height ;
    } // end for i

    free (contacts) ;

    return TRUE ;

} // end Convolve_Radius
```

```
/*************************************/
int Can_Touch (float x1, float y1, float x2, float y2, float x3, float y3,
               float radius)
{
// receives three data points (x1, y1), (x2, y2), (x3, y3) and a radius
// (all in same units of measure (e.g. um))
// and returns TRUE if radius can contact middle point

   float l_by_2, theta, theta1,
         cx, cy, test ;

l_by_2 = sqrt((x3-x1)*(x3-x1) + (y3-y1)*(y3-y1))/2.0 ;

if (l_by_2 >= radius)
     return TRUE ;

theta = atan((y3-y1)/(x3-x1)) ;
theta1 = acos(l_by_2/radius) ;
theta += theta1 ;
cx = x1 + radius*cos(theta) ;
cy = y1 + radius*sin(theta) ;

test = sqrt((cx-x2)*(cx-x2) + (cy-y2)*(cy-y2)) ;
if (test <= radius)
     return TRUE ;
   else
     return FALSE ;

} // end Can_Touch

/*************************************/
float Height (float x, float z, float rad)
{
// returns the height of a radius `rad', relative to its lowest point `z',
// for a given distance `x' from the lowest point

        if (fabs(x) > rad)
               return -99e99 ;

        return (z-(rad-sqrt(rad*rad-x*x))) ;

} // end Height
```

B.2 Ground Profile Transmission Analyses

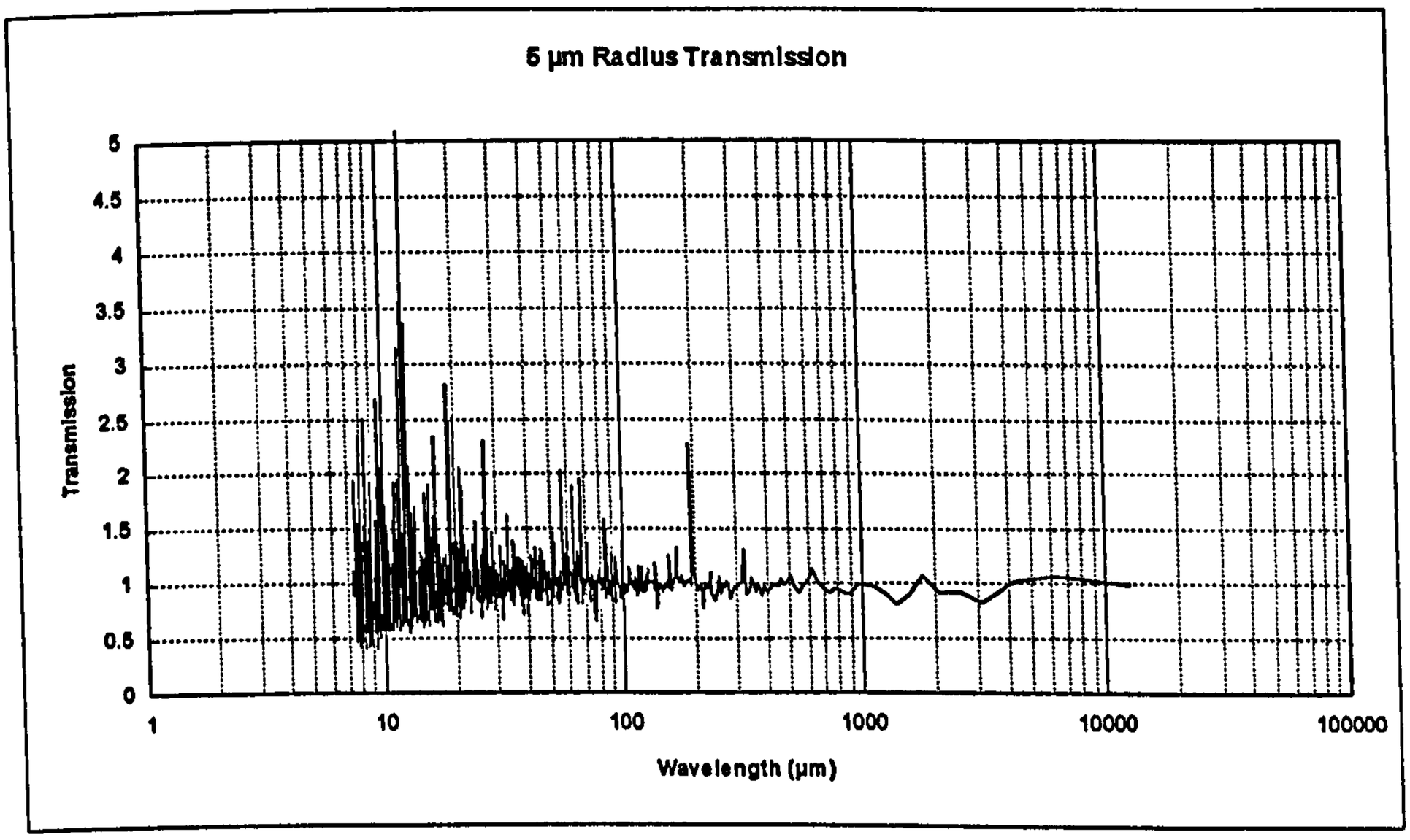


Figure B.1 Ground profile with 5 µm tip radius.

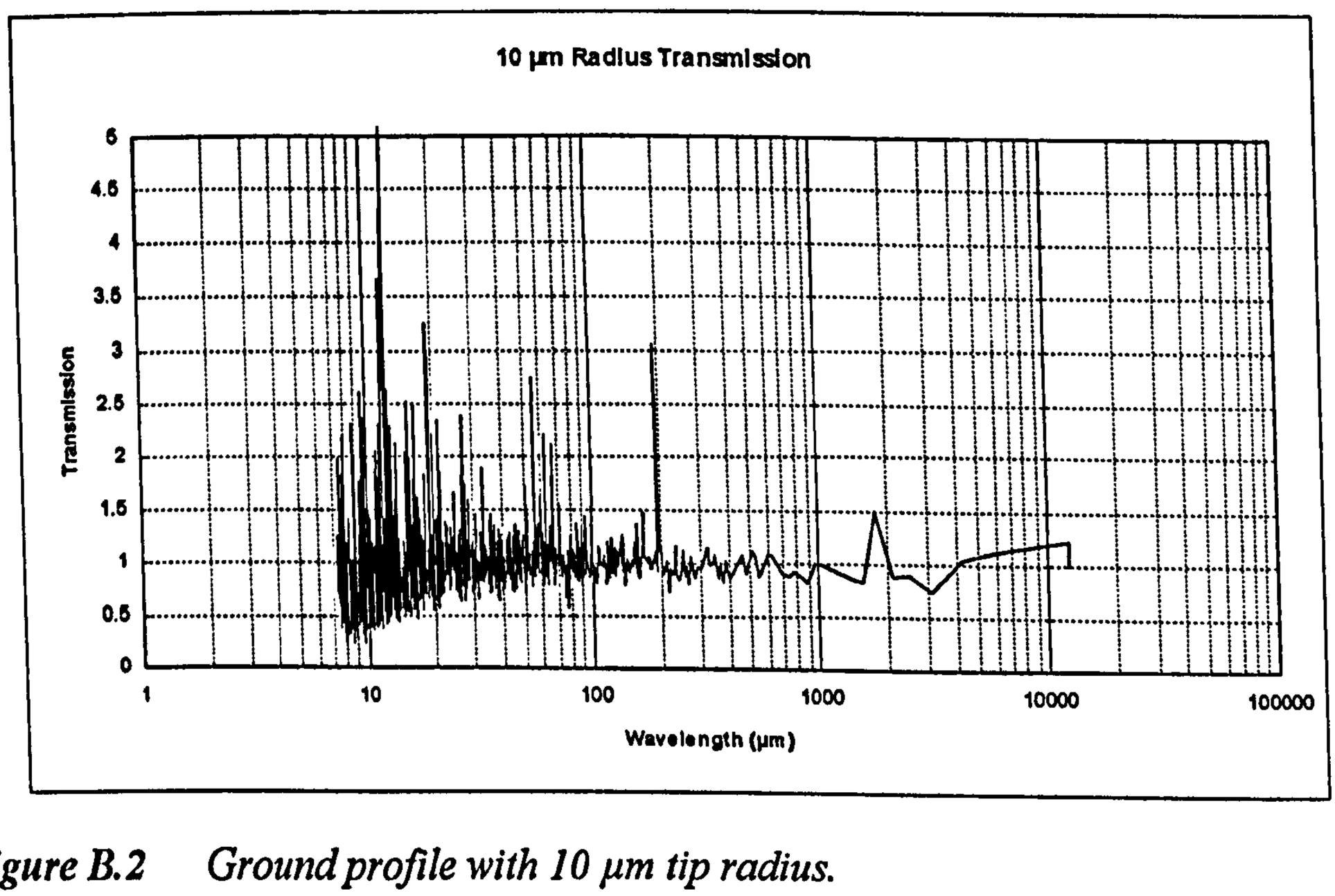


Figure B.2 Ground profile with 10 µm tip radius.

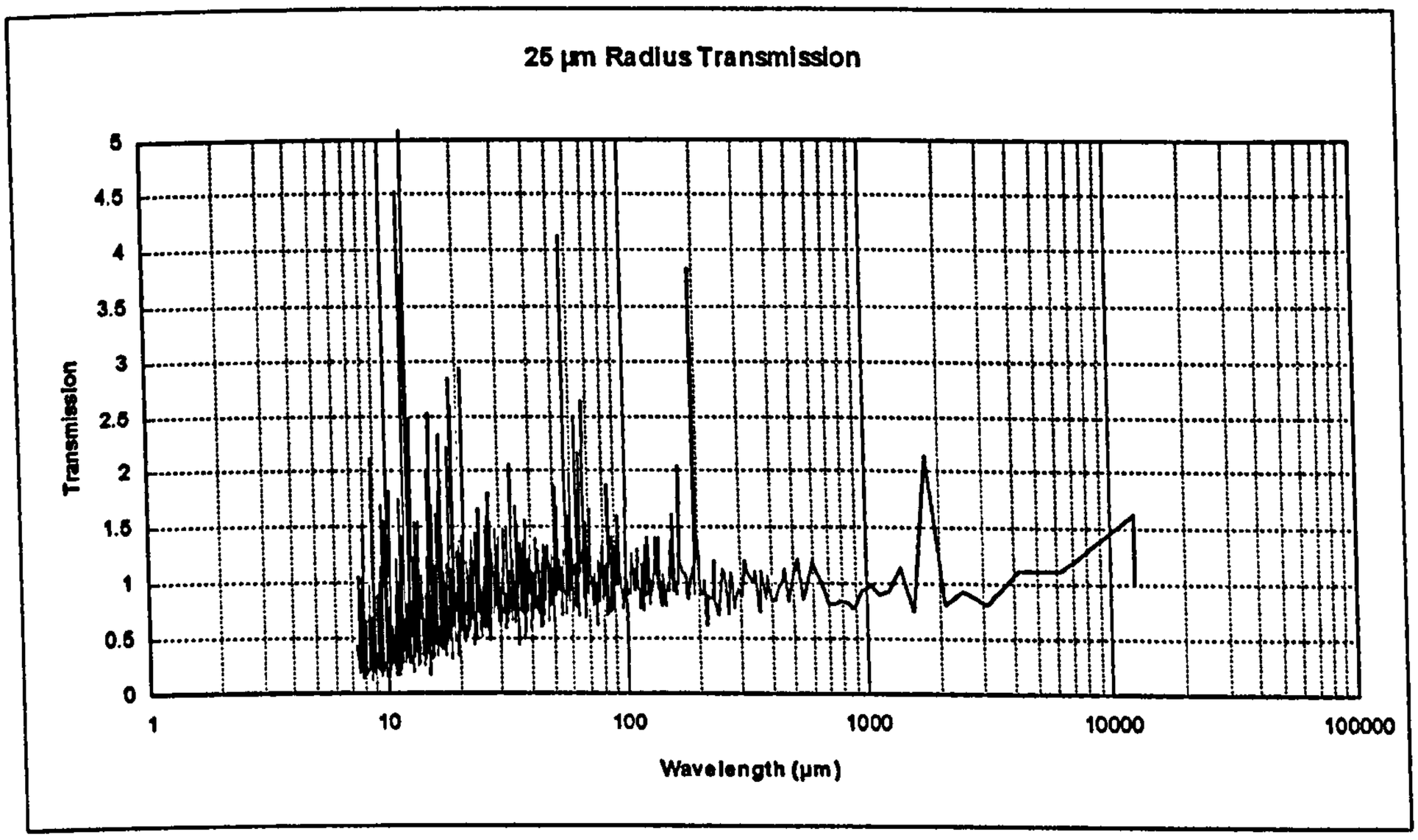


Figure B.3 *Ground profile with 25 µm tip radius.*

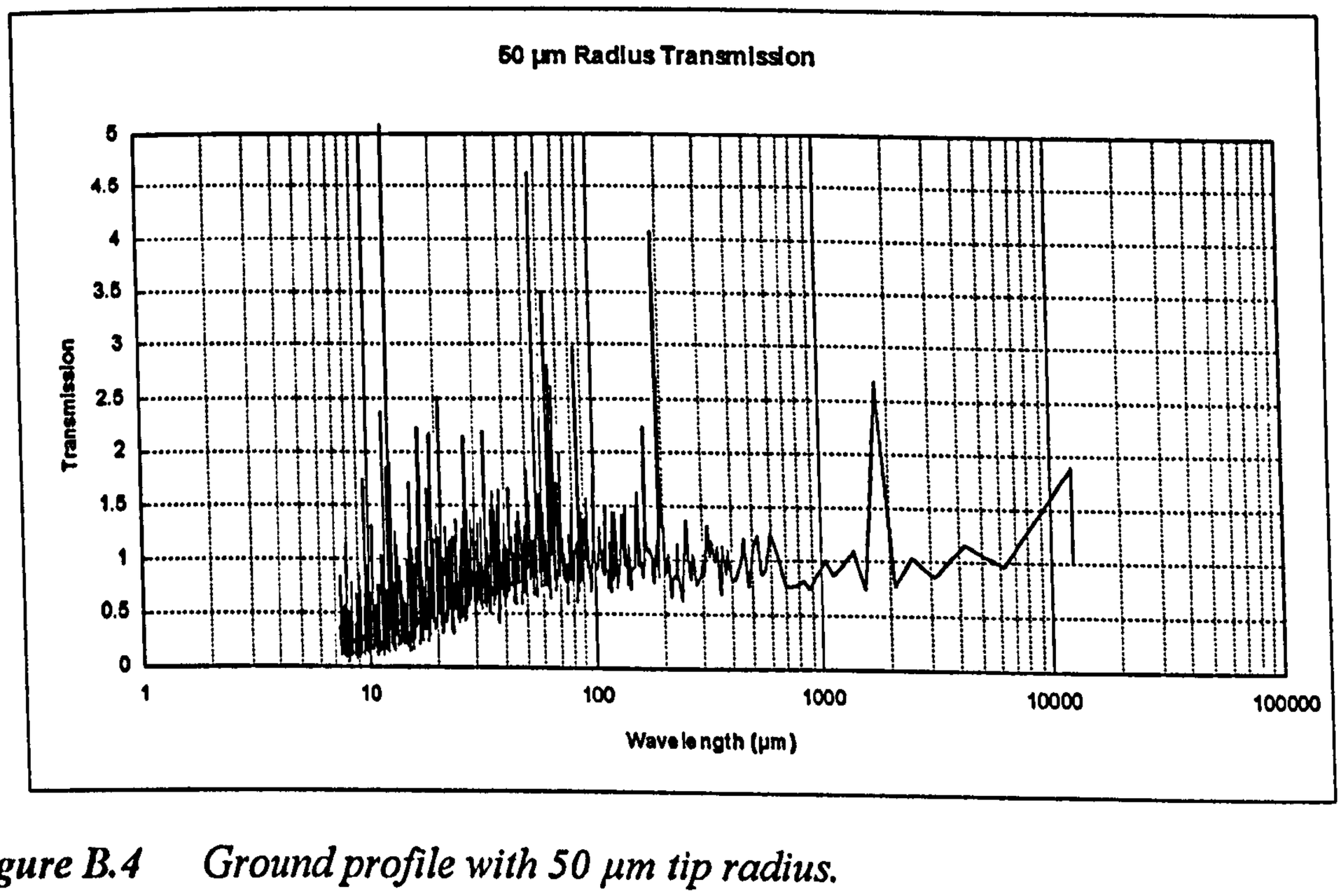


Figure B.4 *Ground profile with 50 µm tip radius.*

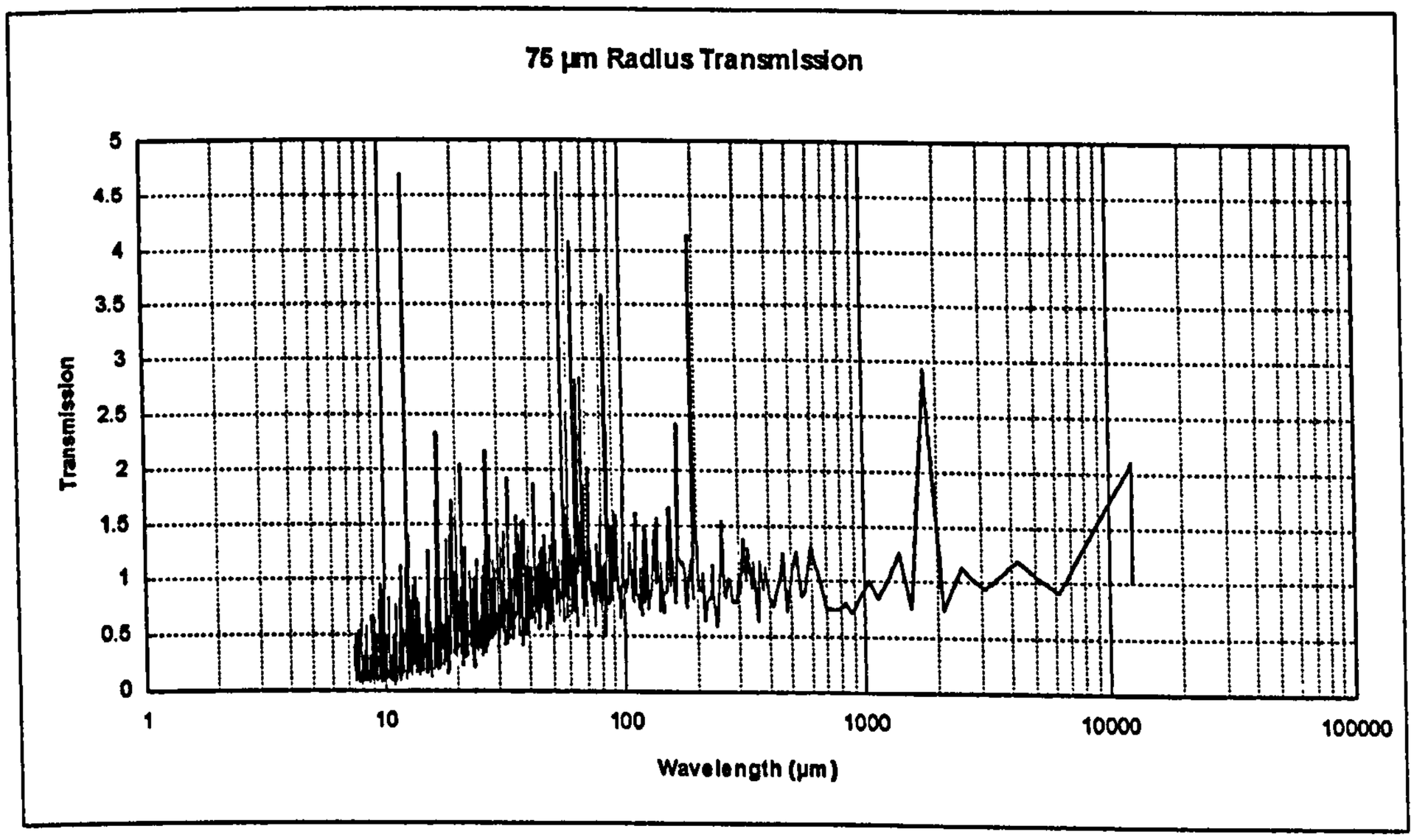


Figure B.5 Ground profile with 75 µm tip radius.

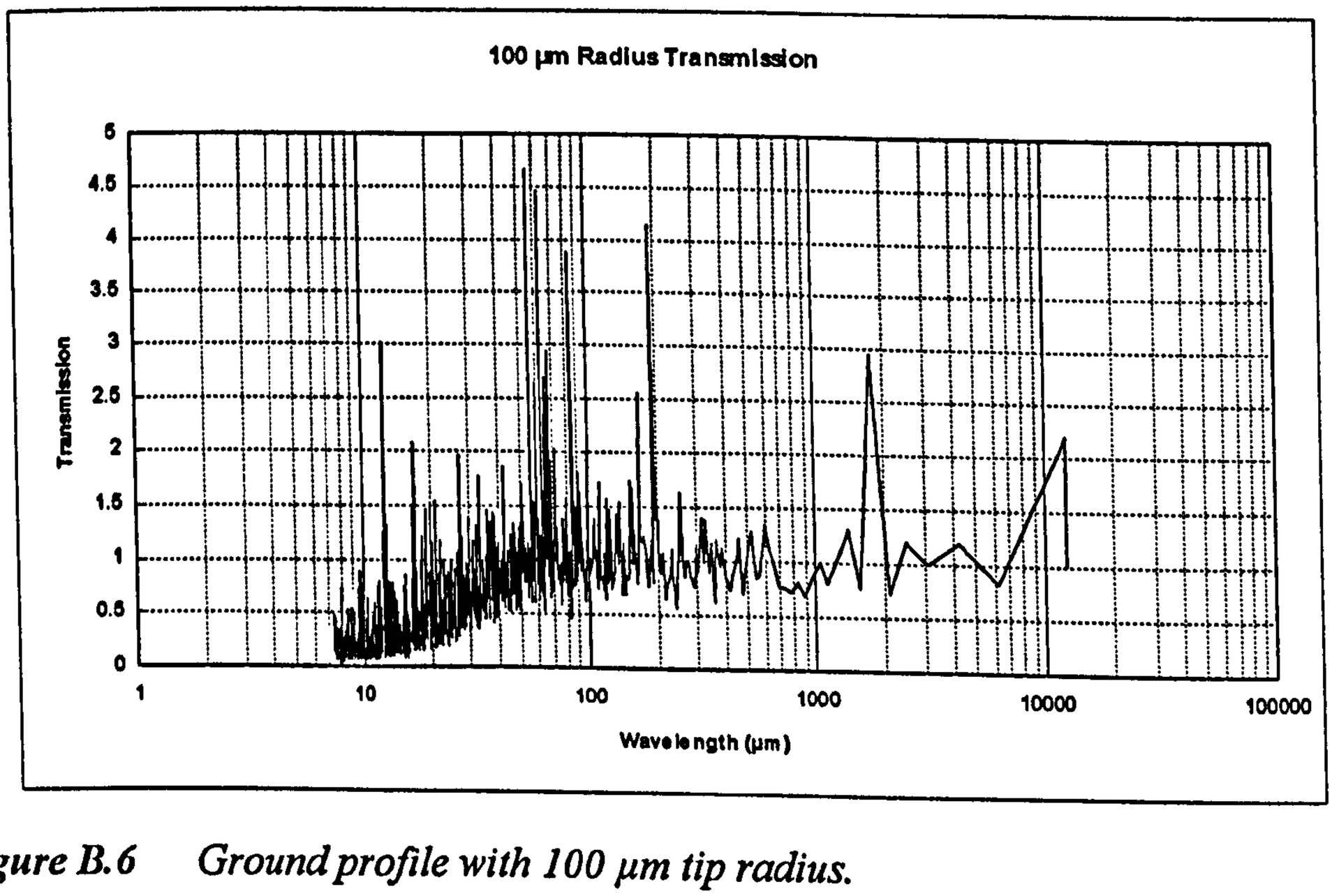


Figure B.6 Ground profile with 100 µm tip radius.

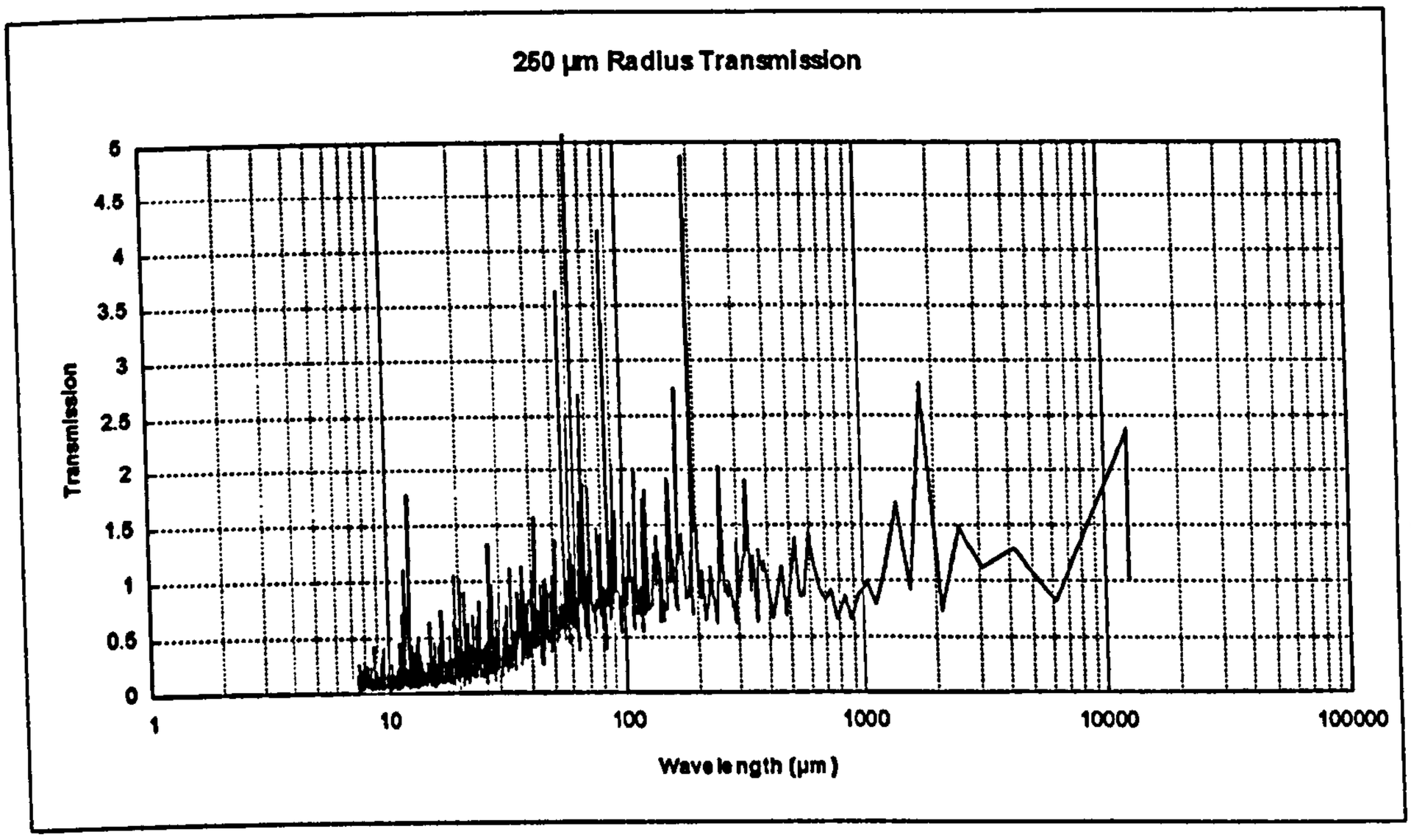


Figure B.7 Ground profile with 250 µm tip radius.

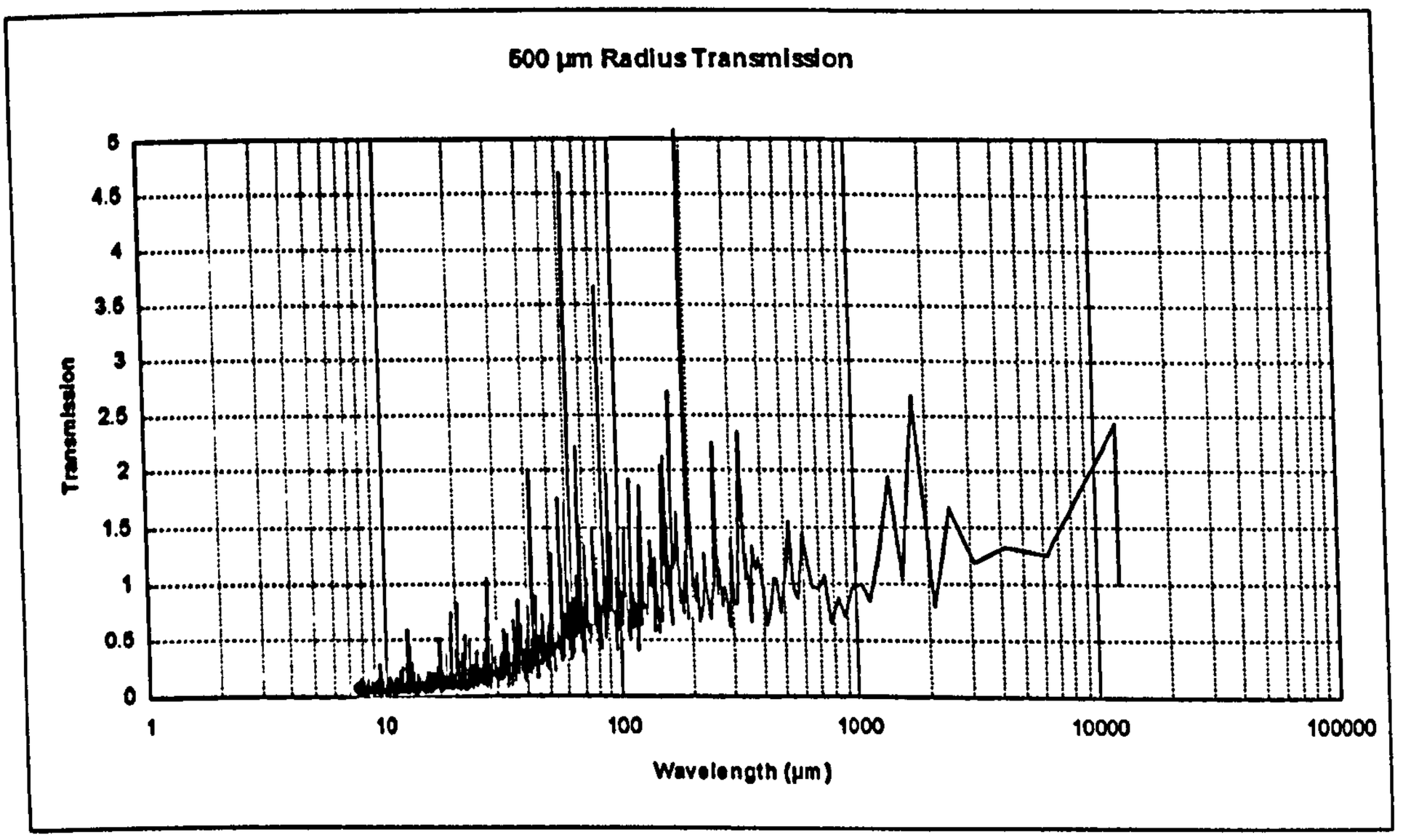


Figure B.8 Ground profile with 500 µm tip radius.

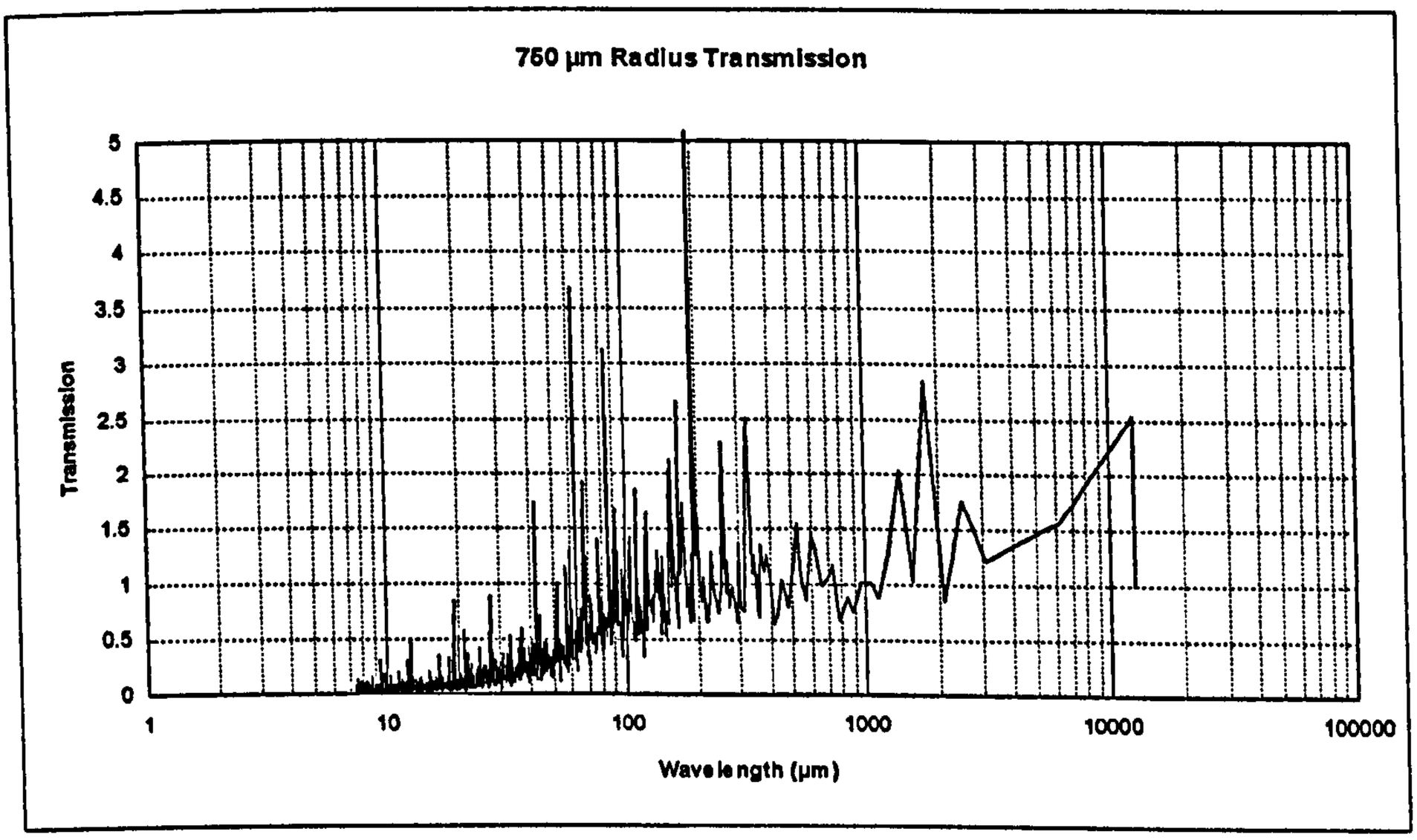


Figure B.9 *Ground profile with 750 µm tip radius.*

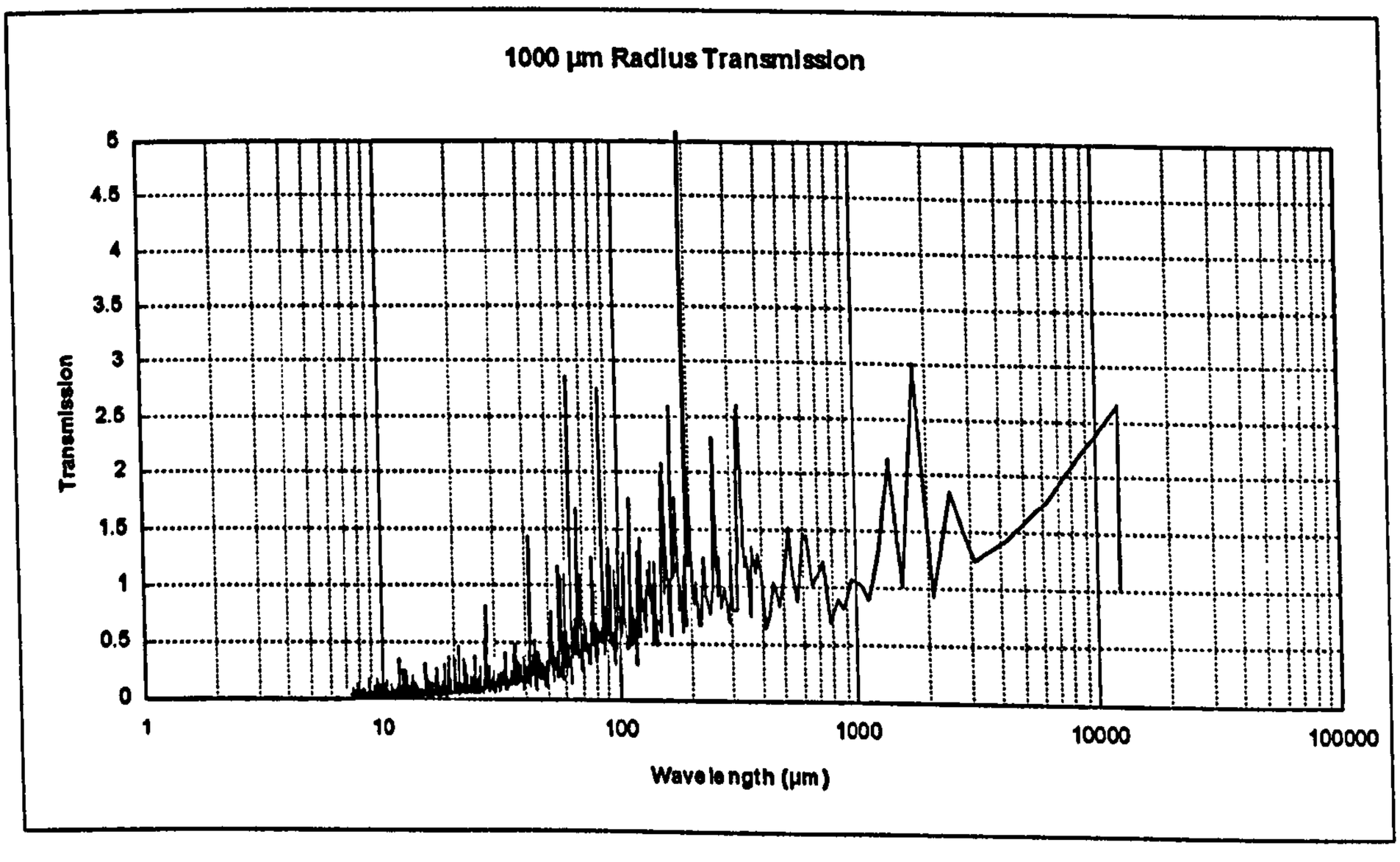


Figure B.10 *Ground profile with 1000 µm tip radius.*

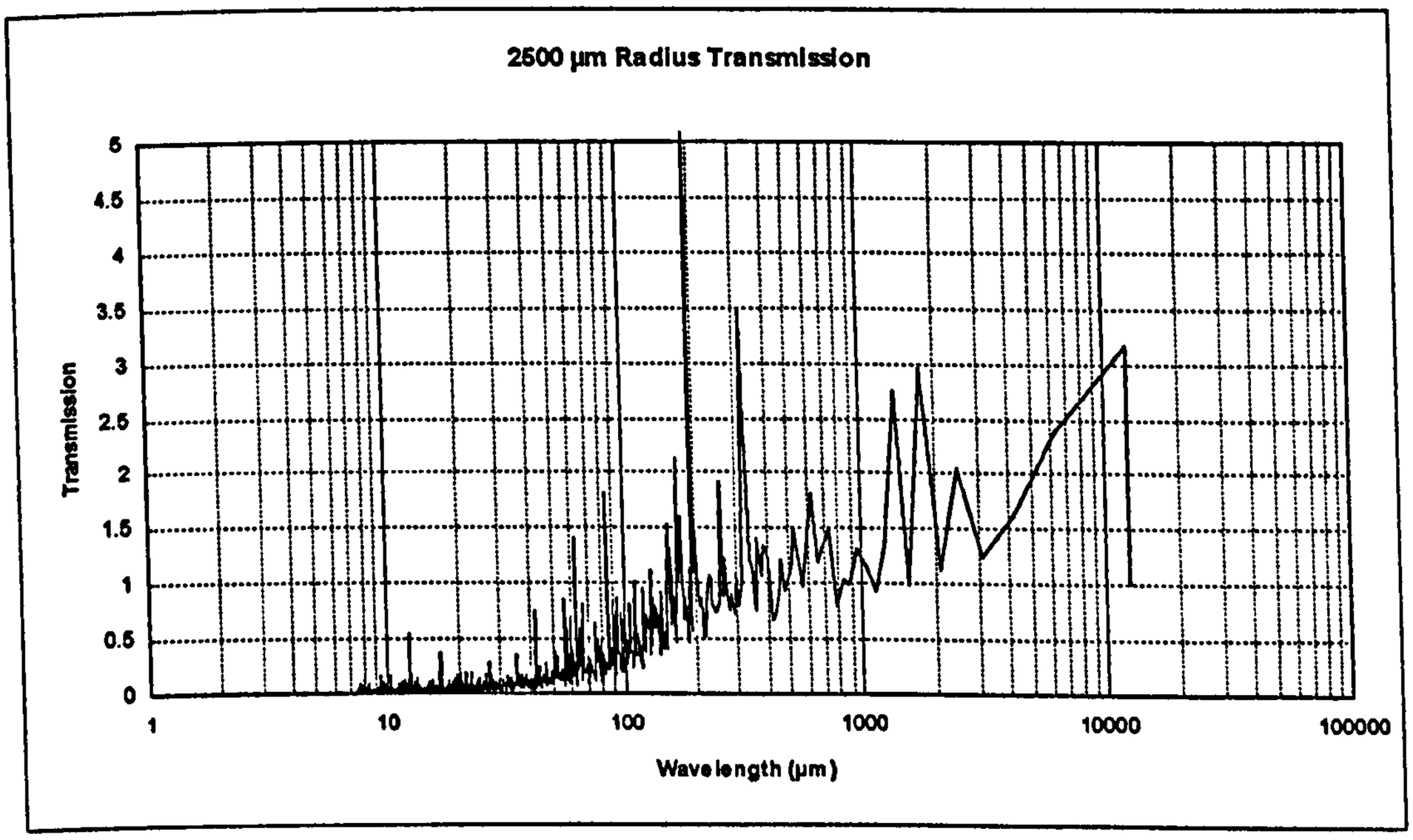


Figure B.11 Ground profile with 2500 µm tip radius.

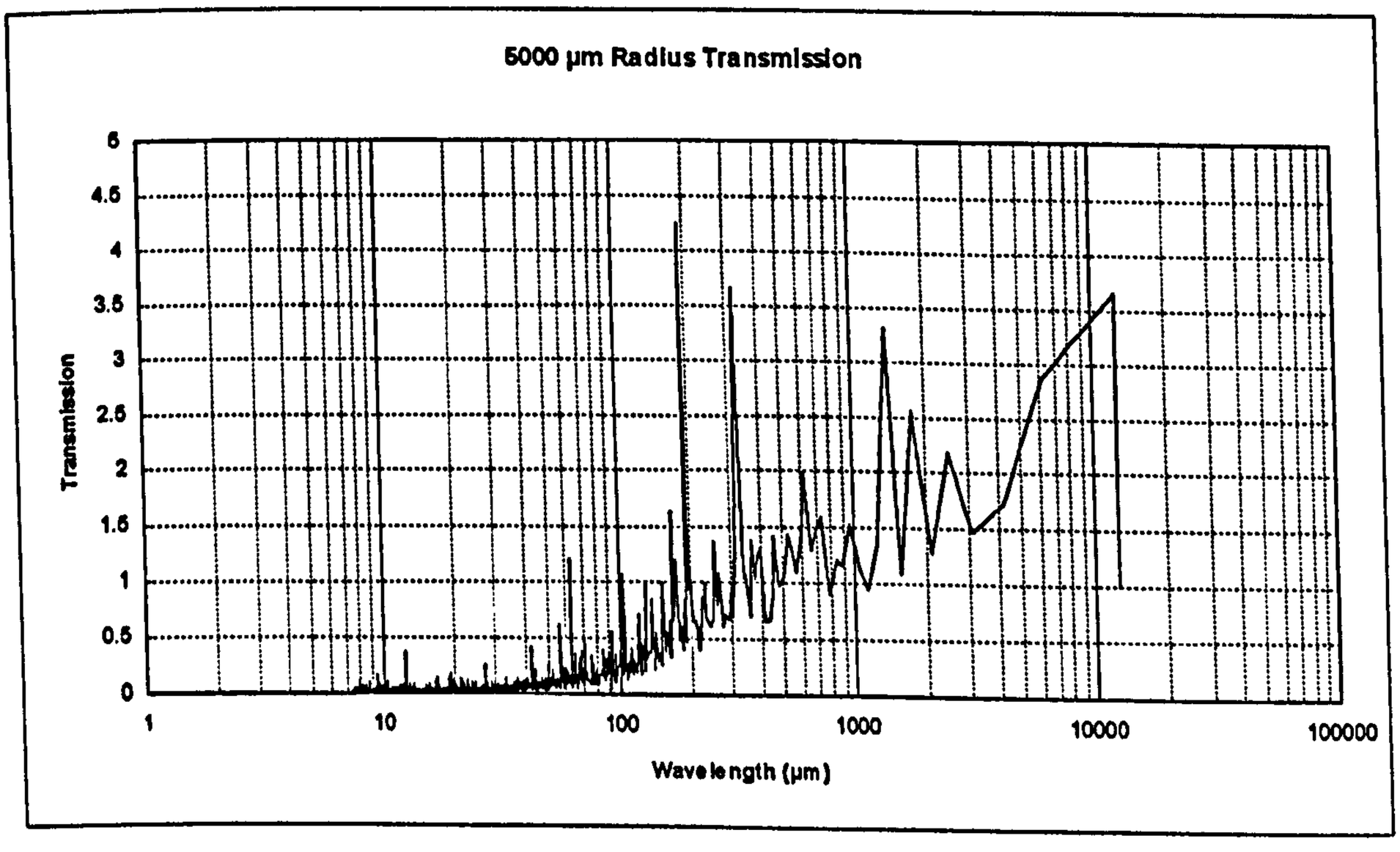


Figure B.12 Ground profile with 5000 µm tip radius.

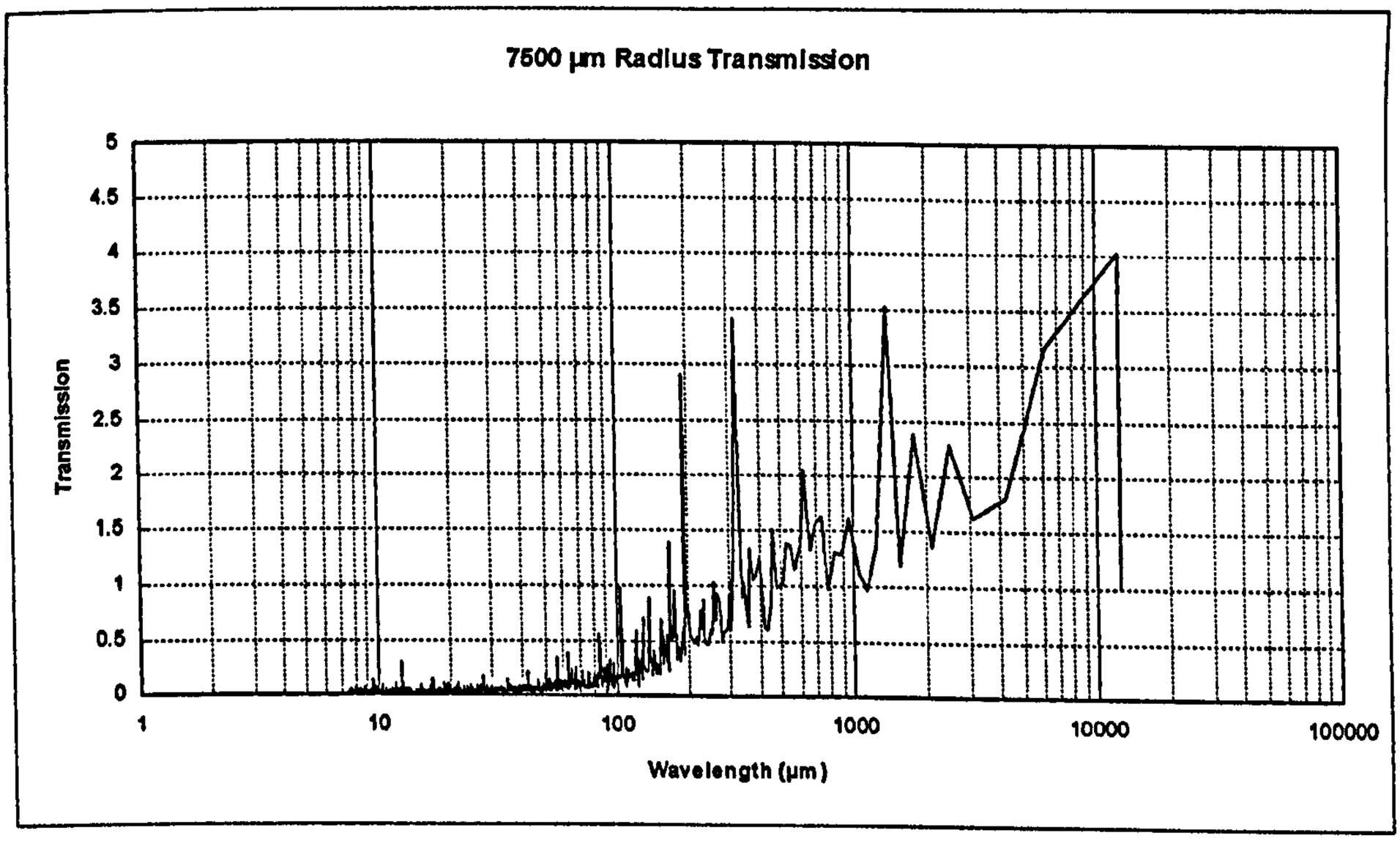


Figure B.13 Ground profile with 7500 µm tip radius.

B.3 Ground Profile Parameter Analyses

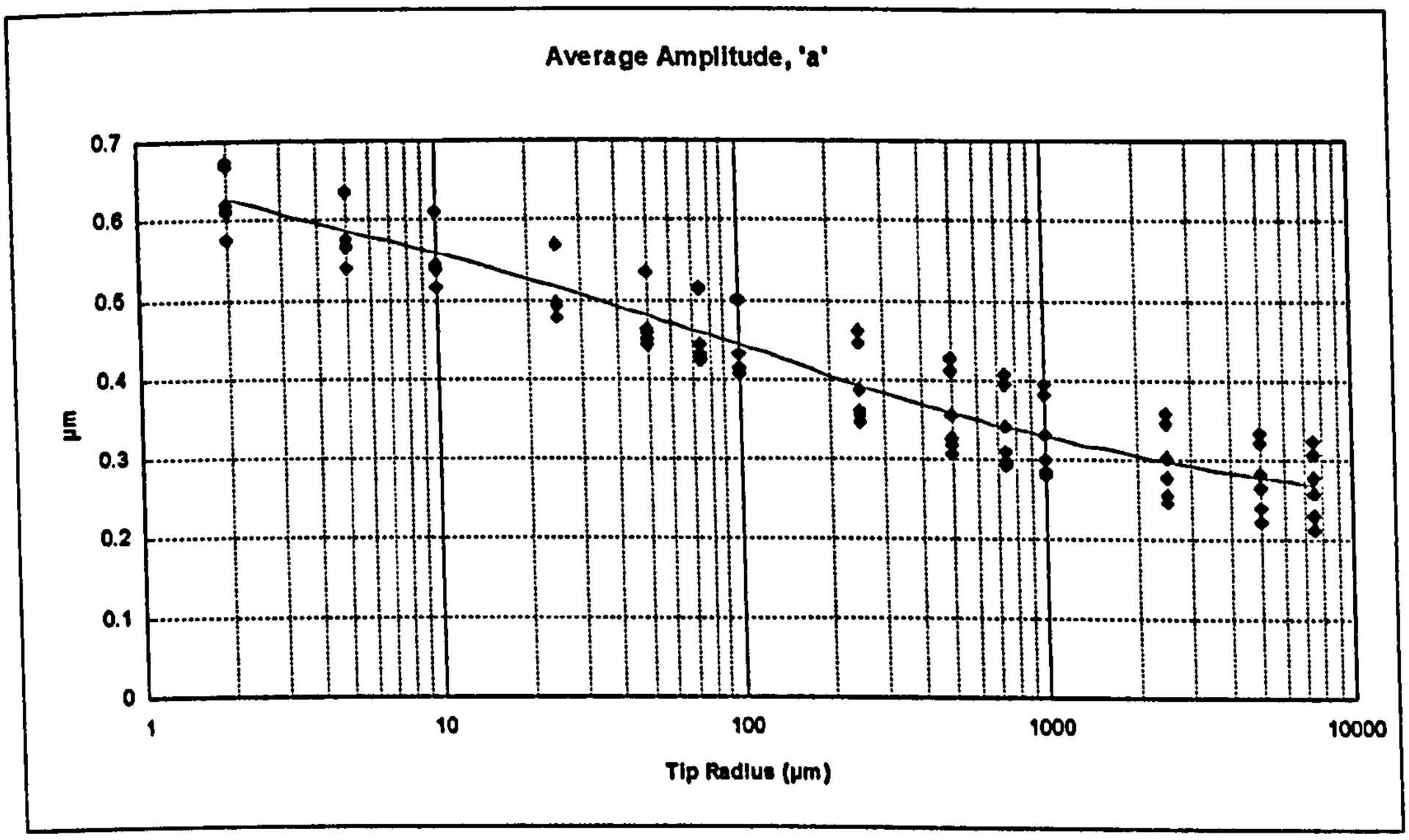


Figure B.14 Ground profile.

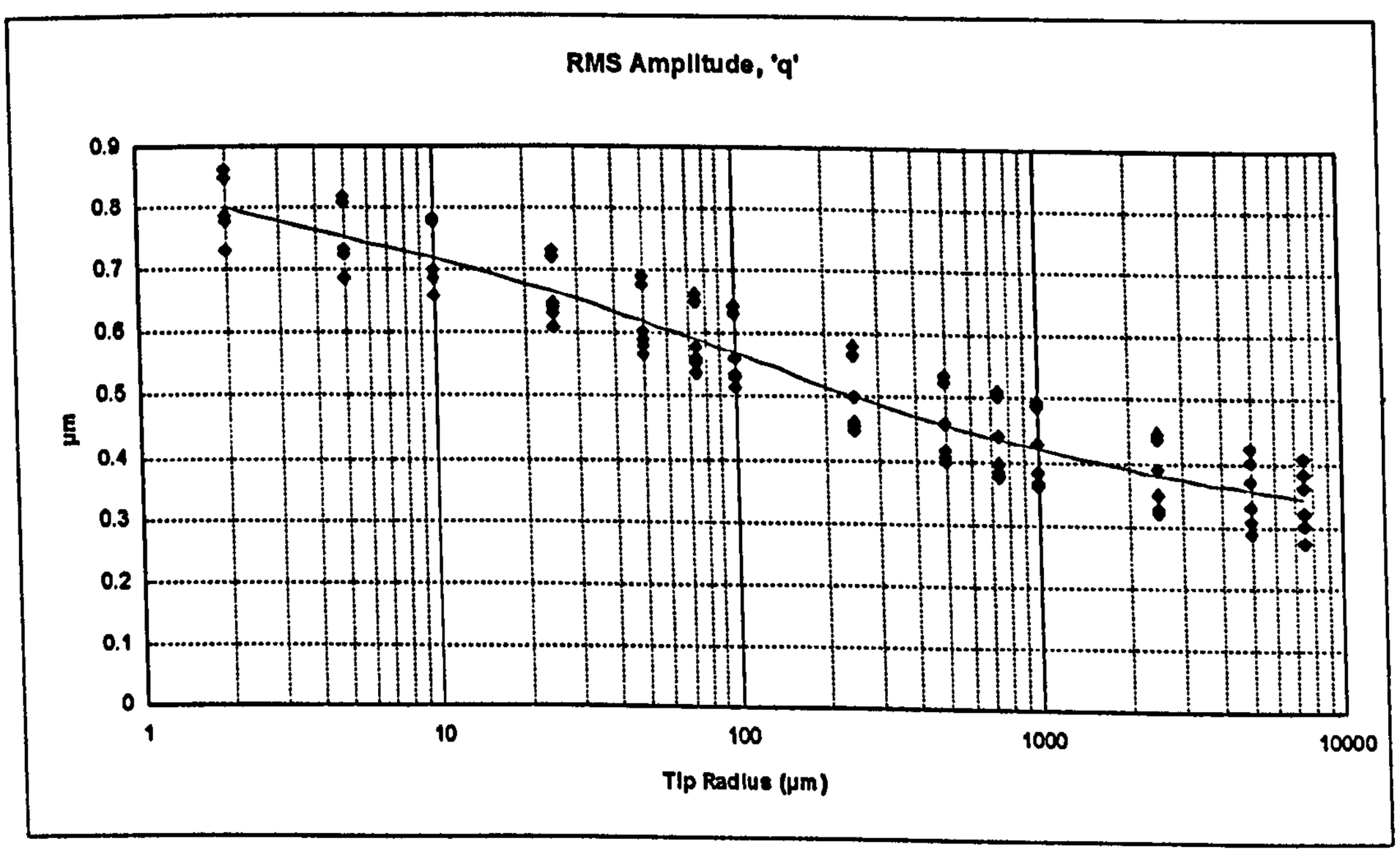


Figure B.15 Ground profile.

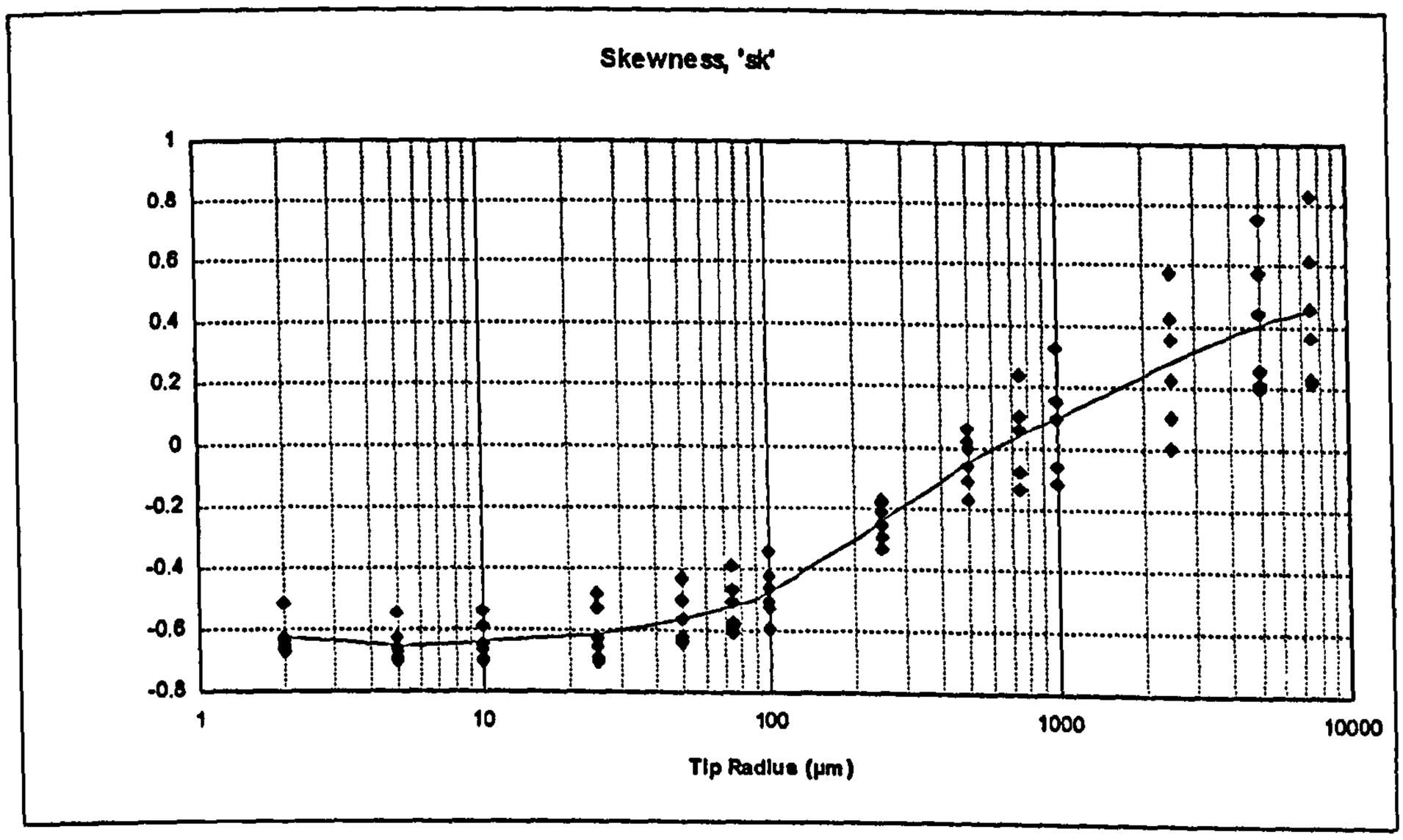


Figure B.16 Ground profile.

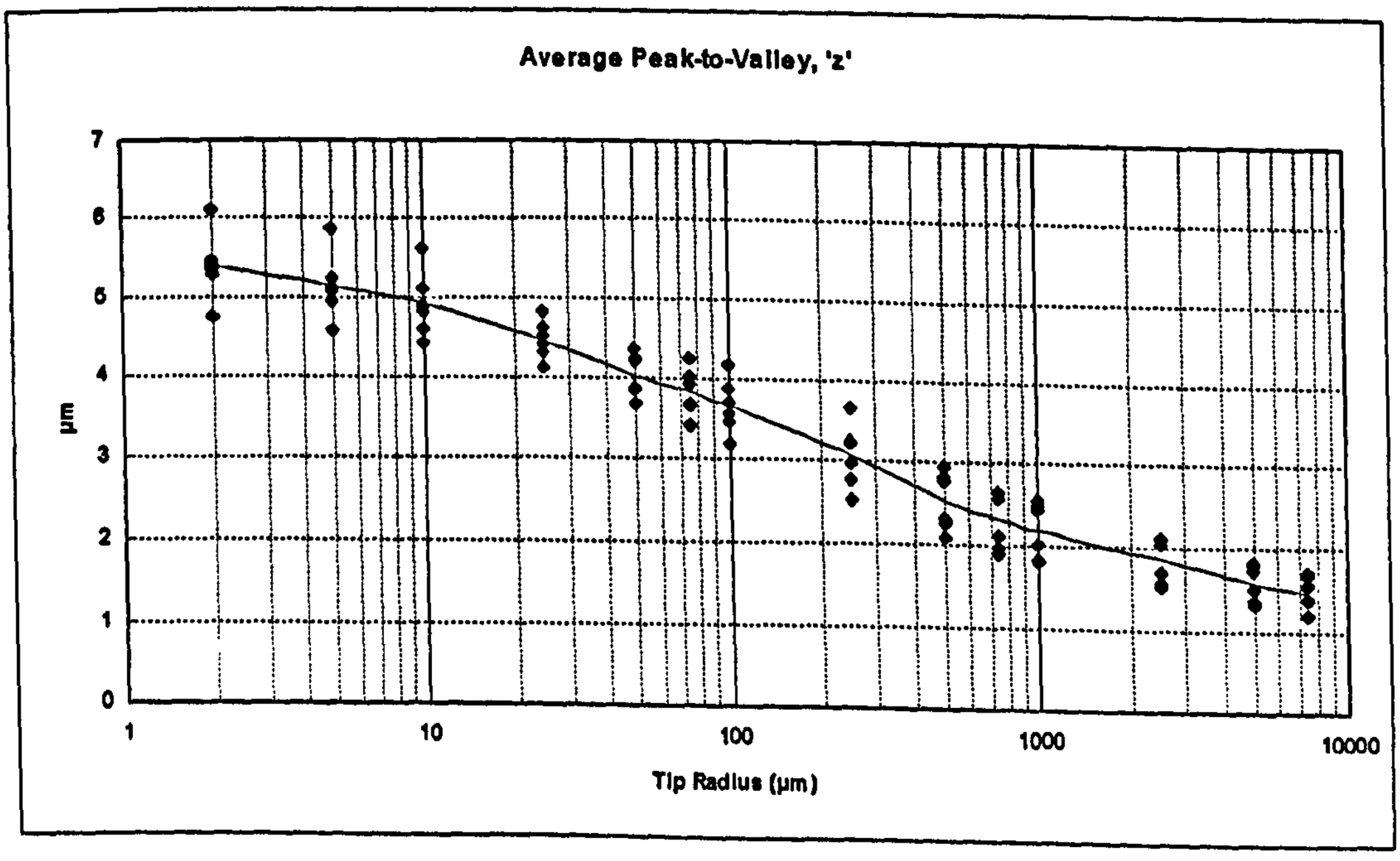


Figure B.17 Ground profile.

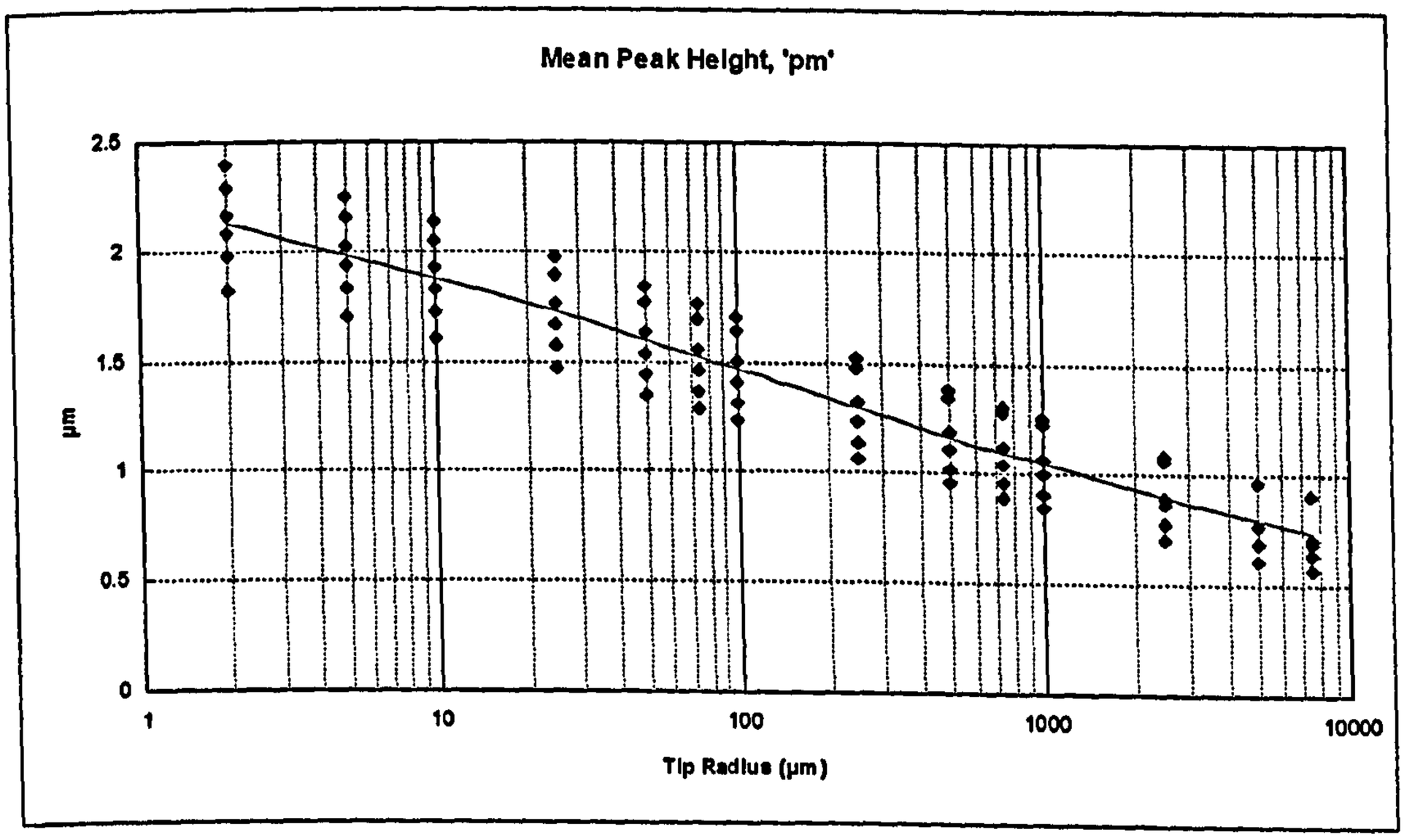


Figure B.18 Ground profile.

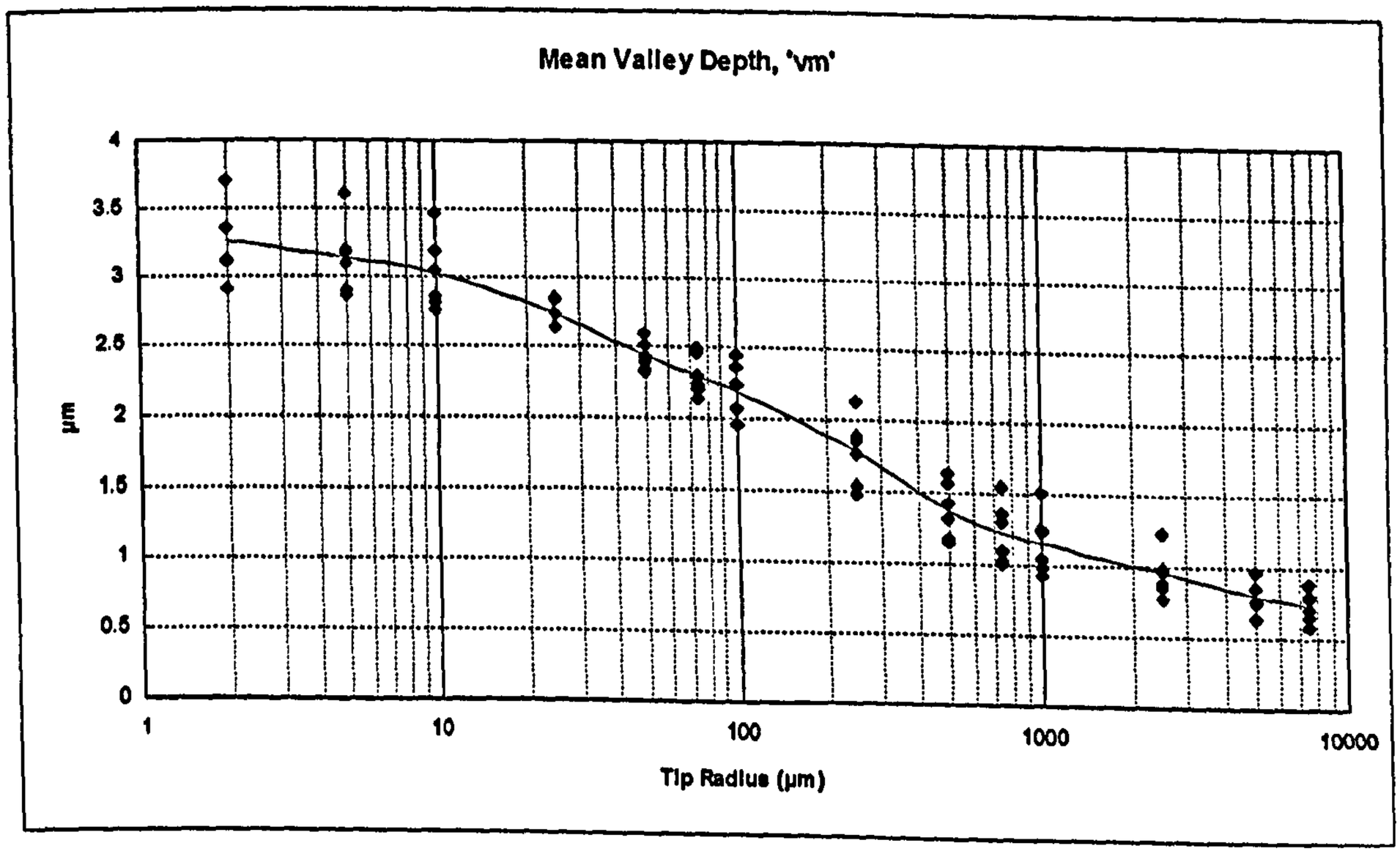


Figure B.19 Ground profile.

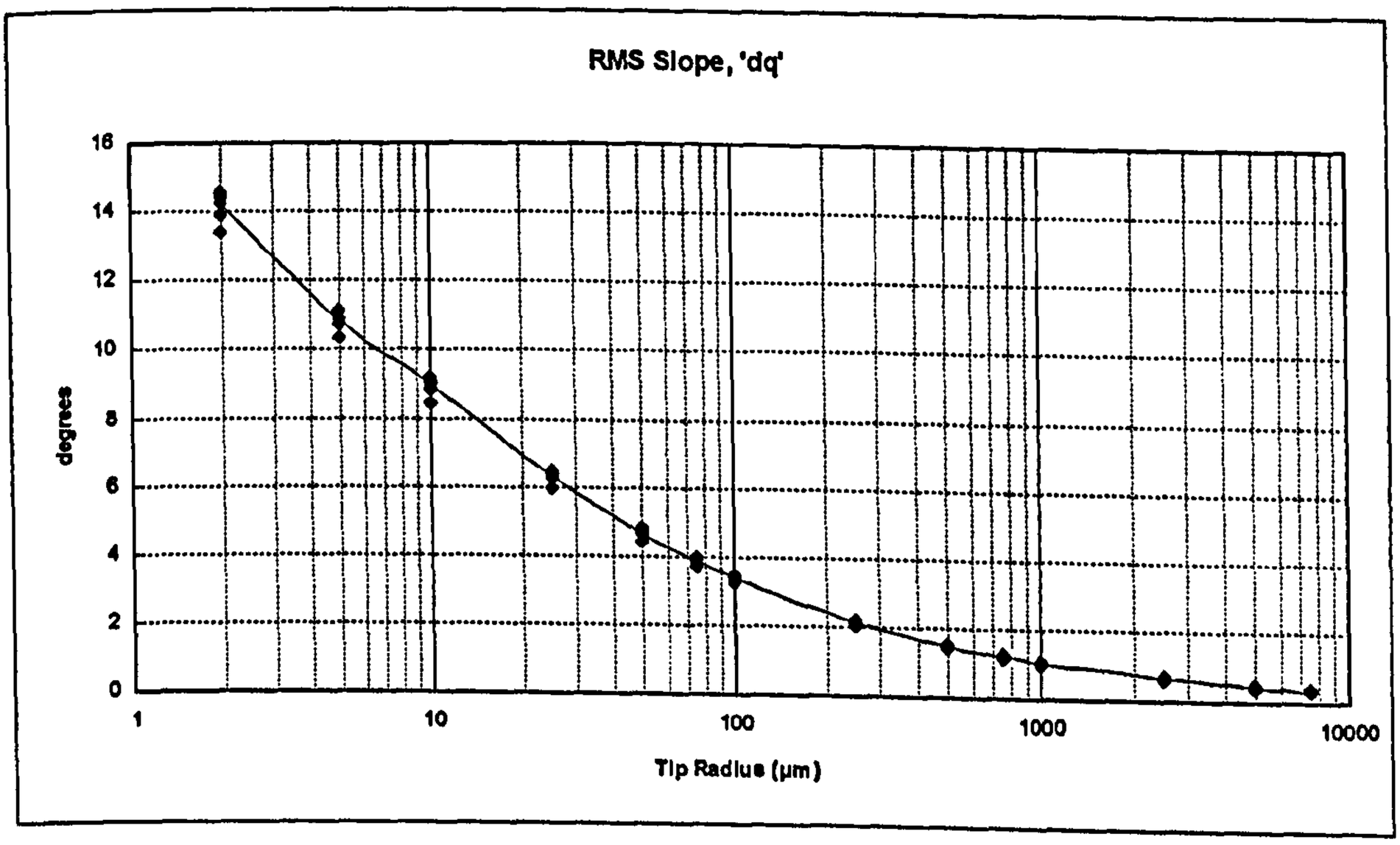


Figure B.20 Ground profile.

B.4 Turned Profile Transmission Analyses

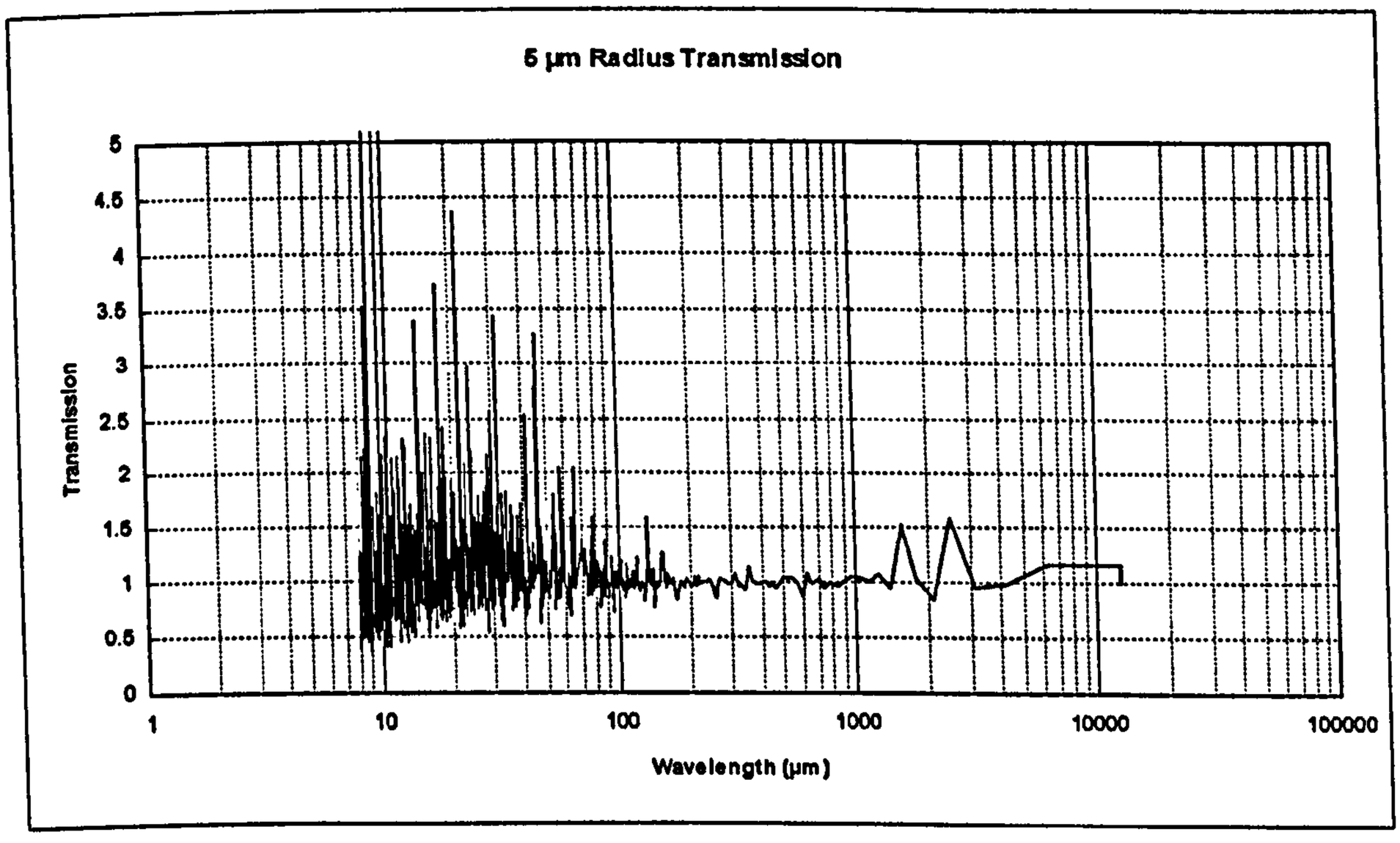


Figure B.21 Turned profile with 5 µm tip radius.

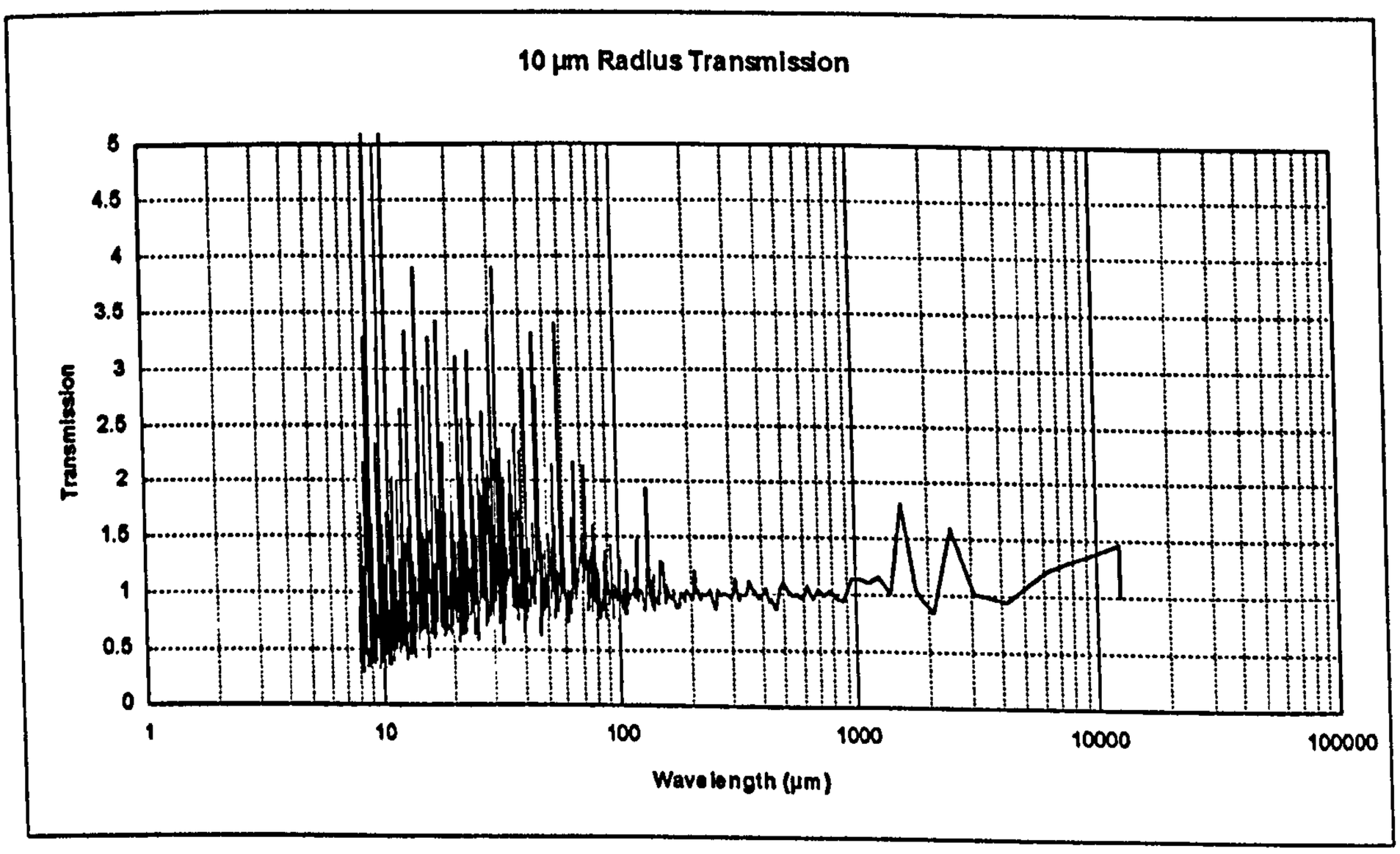


Figure B.22 Turned profile with 10 µm tip radius.

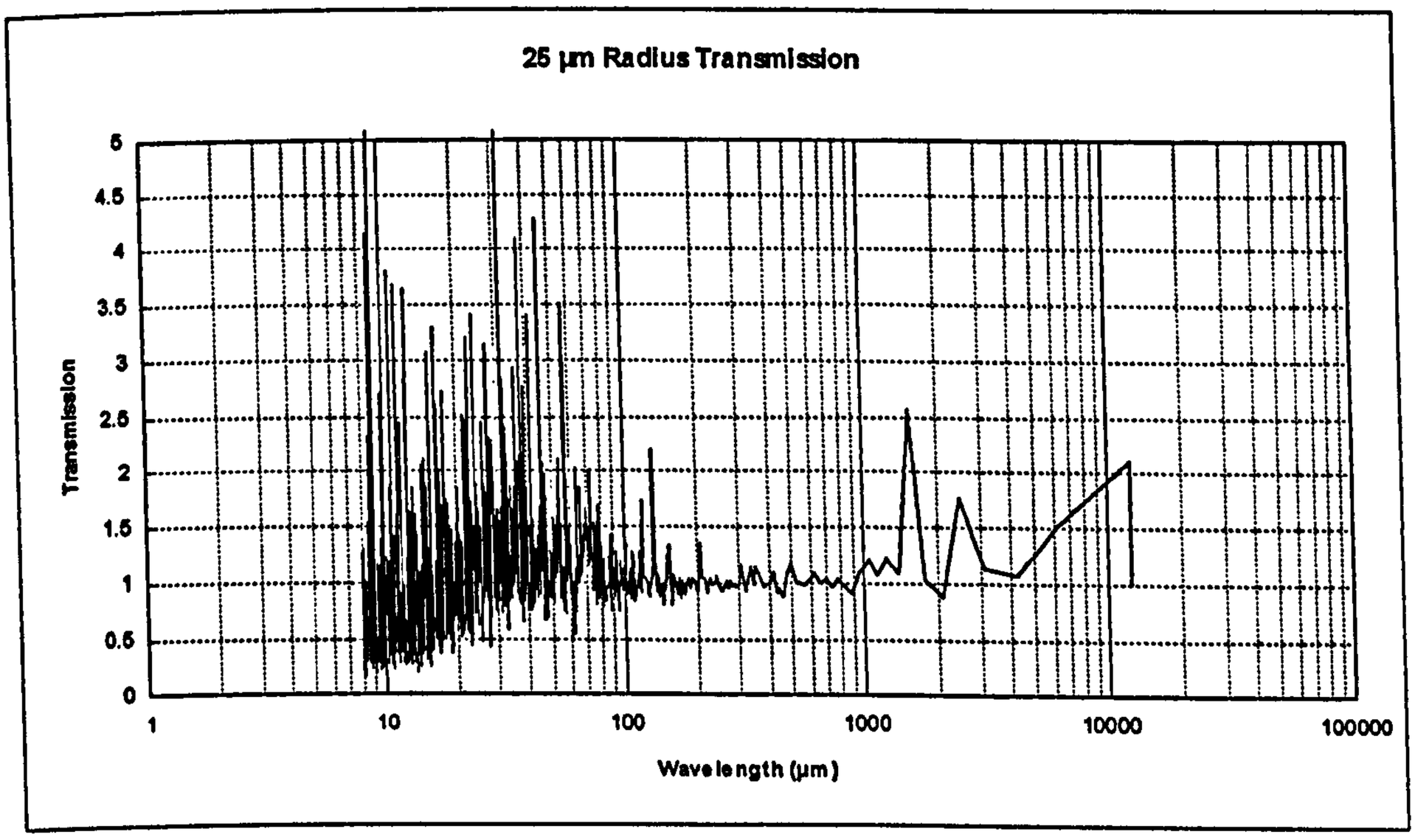


Figure B.23 Turned profile with 25 µm tip radius.

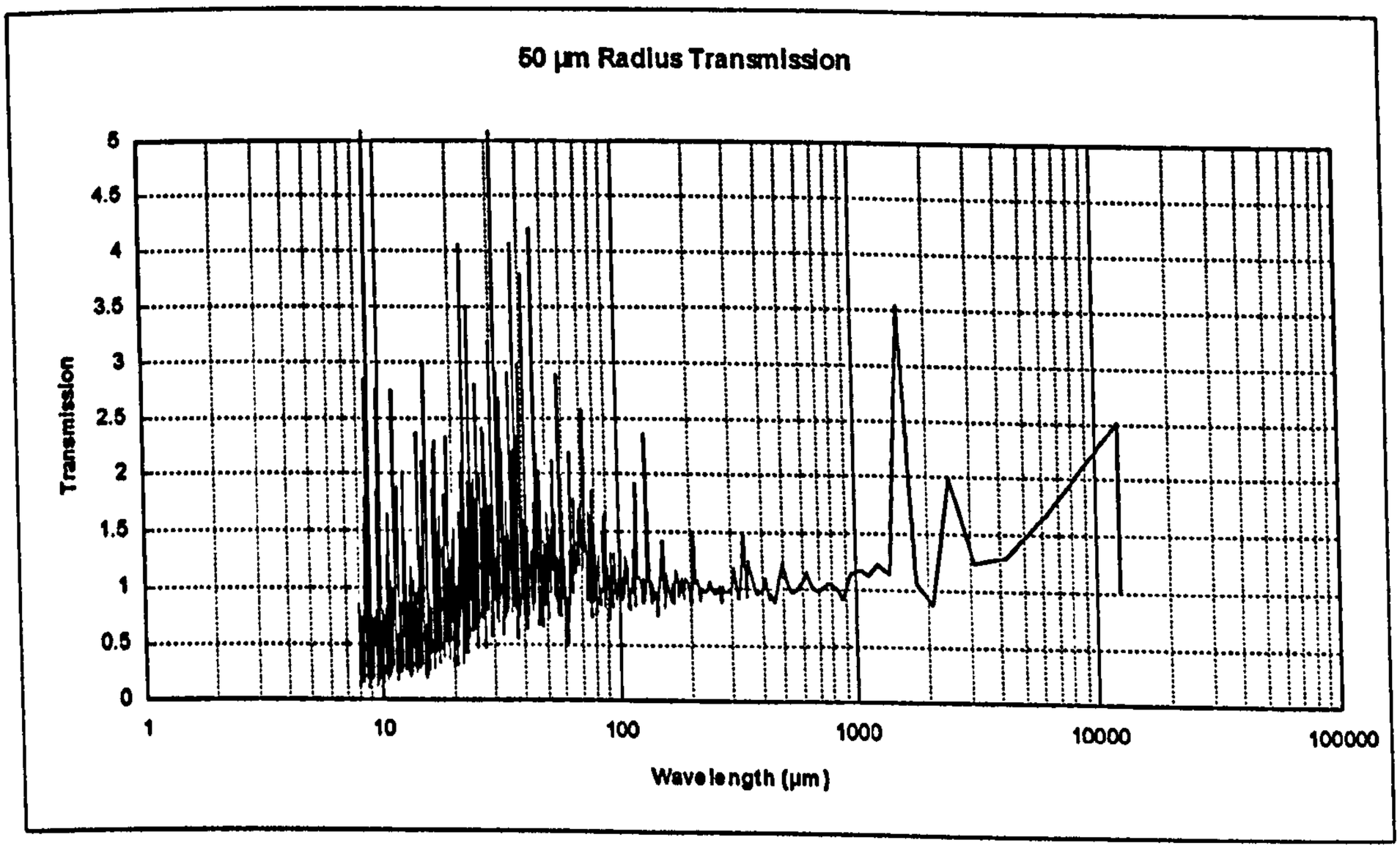


Figure B.24 Turned profile with 50 µm tip radius.

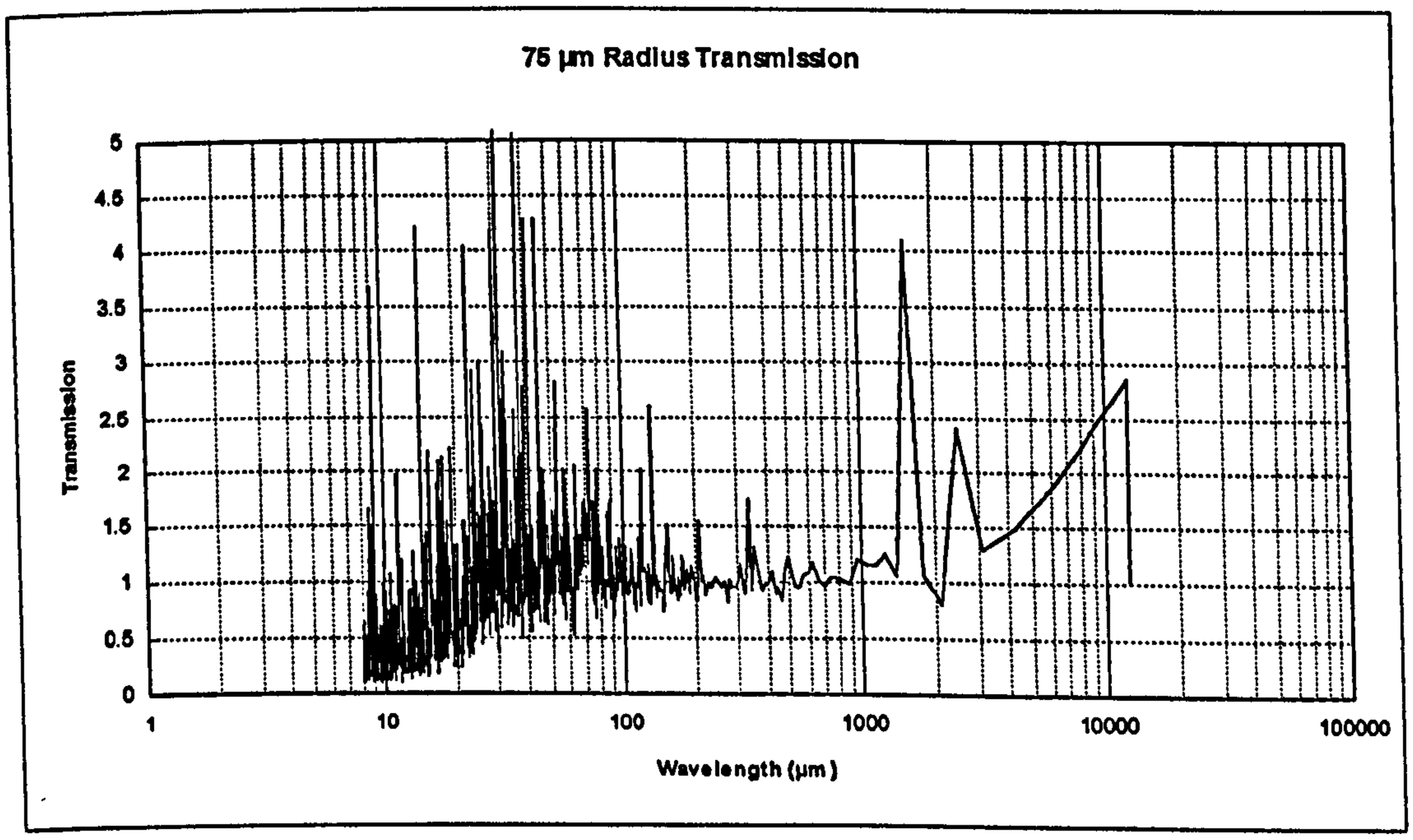


Figure B.25 Turned profile with 75 µm tip radius.

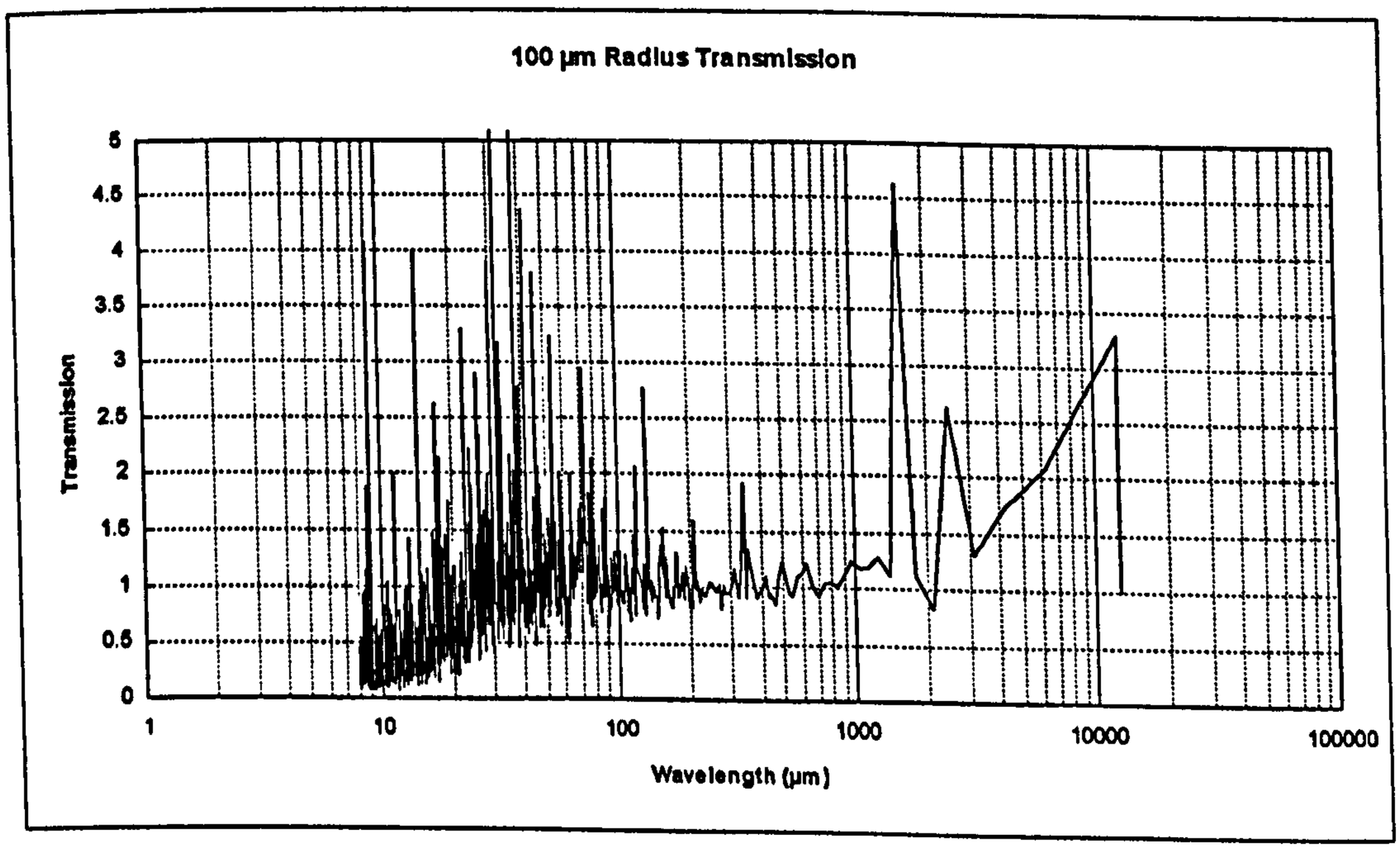


Figure B.26 Turned profile with 100 µm tip radius.

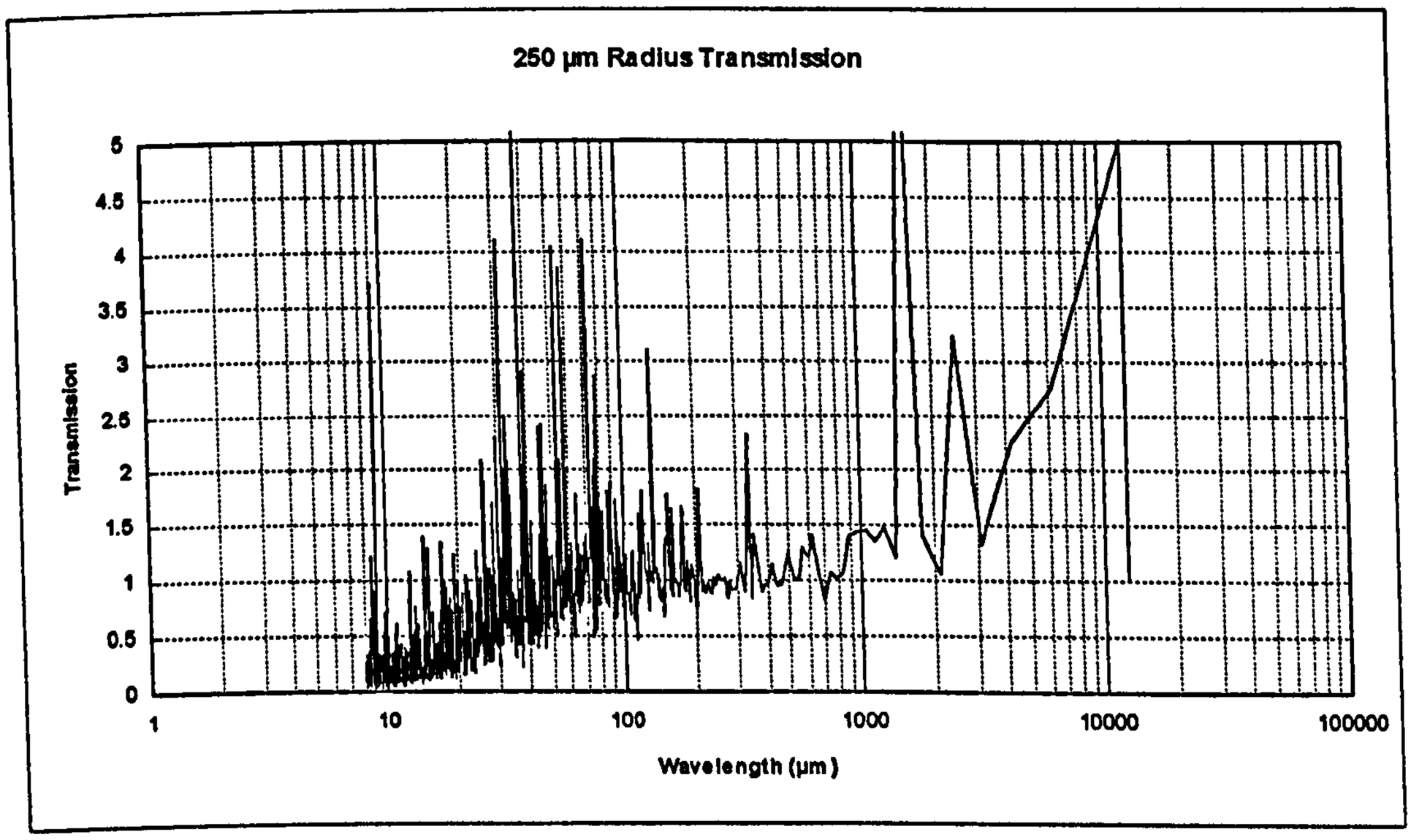


Figure B.27 Turned profile with 250 µm tip radius.

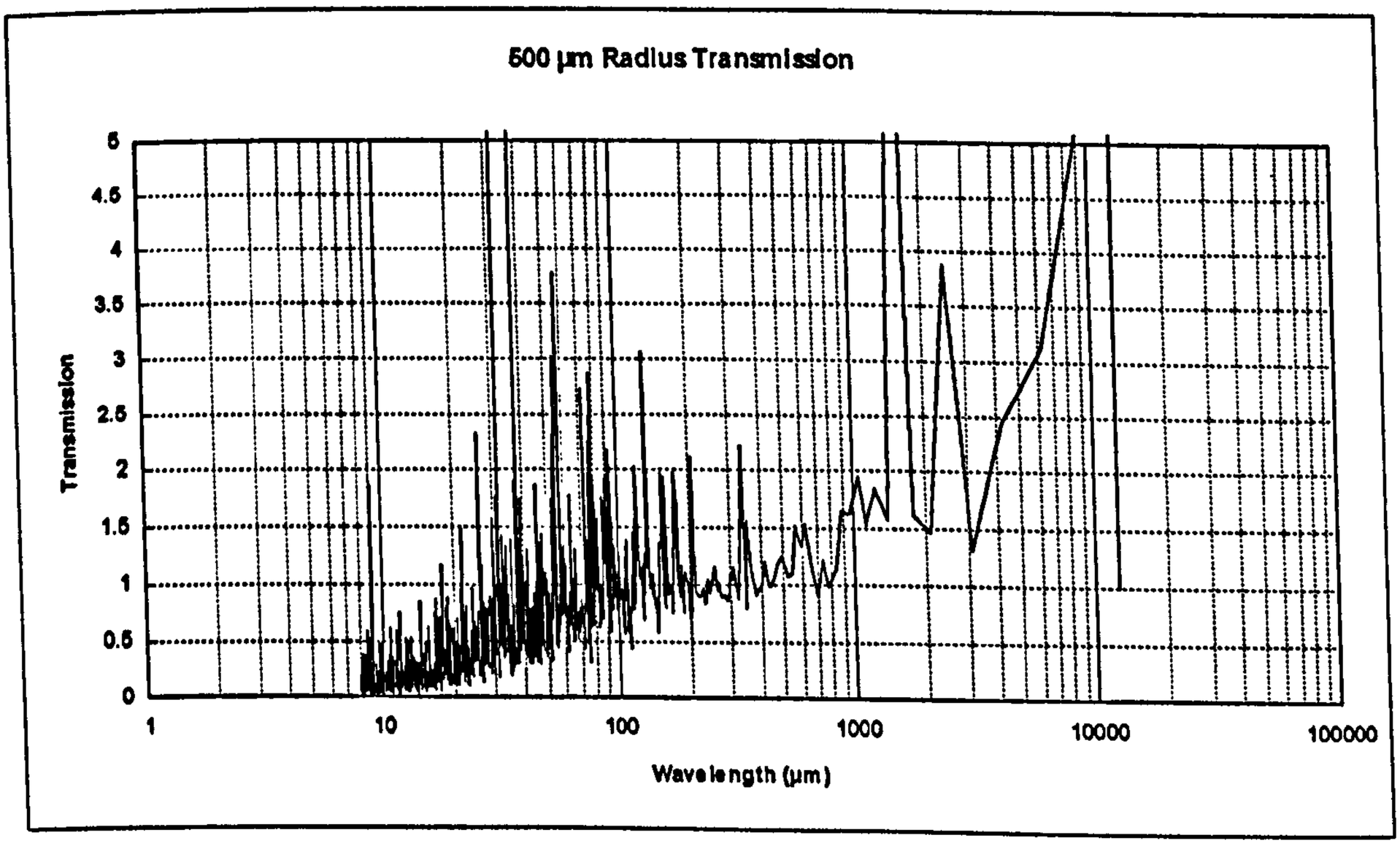


Figure B.28 Turned profile with 500 µm tip radius.

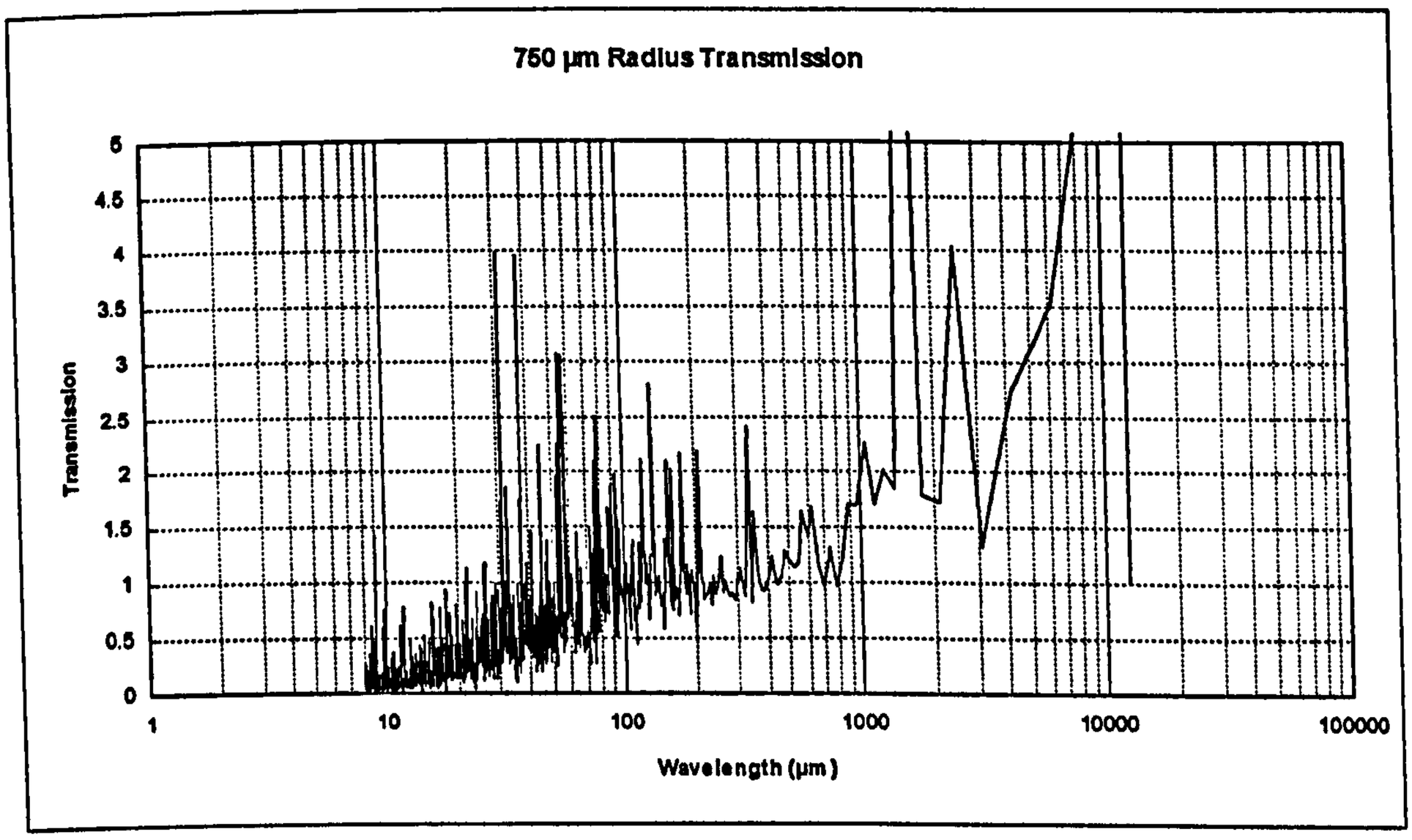


Figure B.29 Turned profile with 750 µm tip radius.

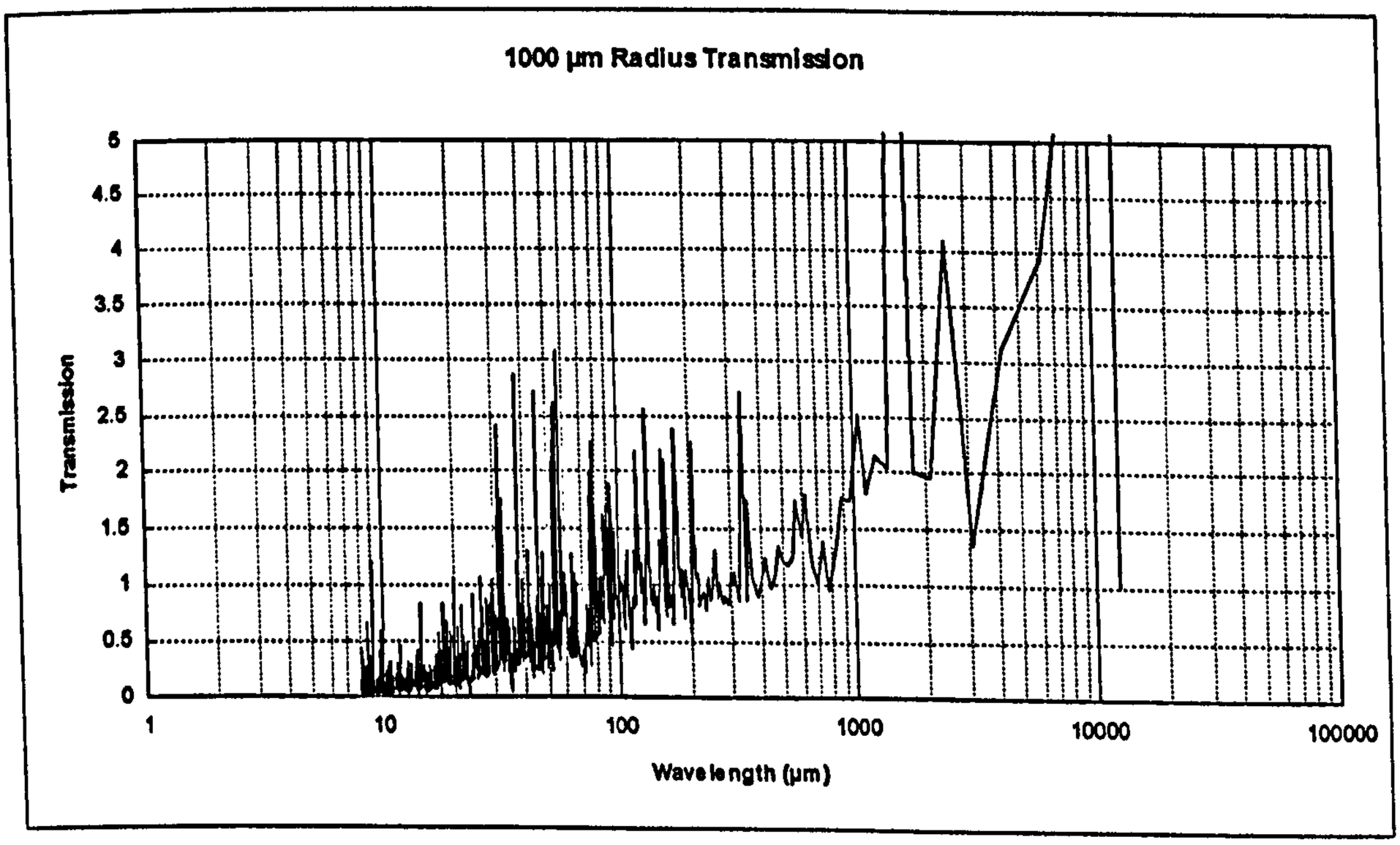


Figure B.30 Turned profile with 1000 µm tip radius.

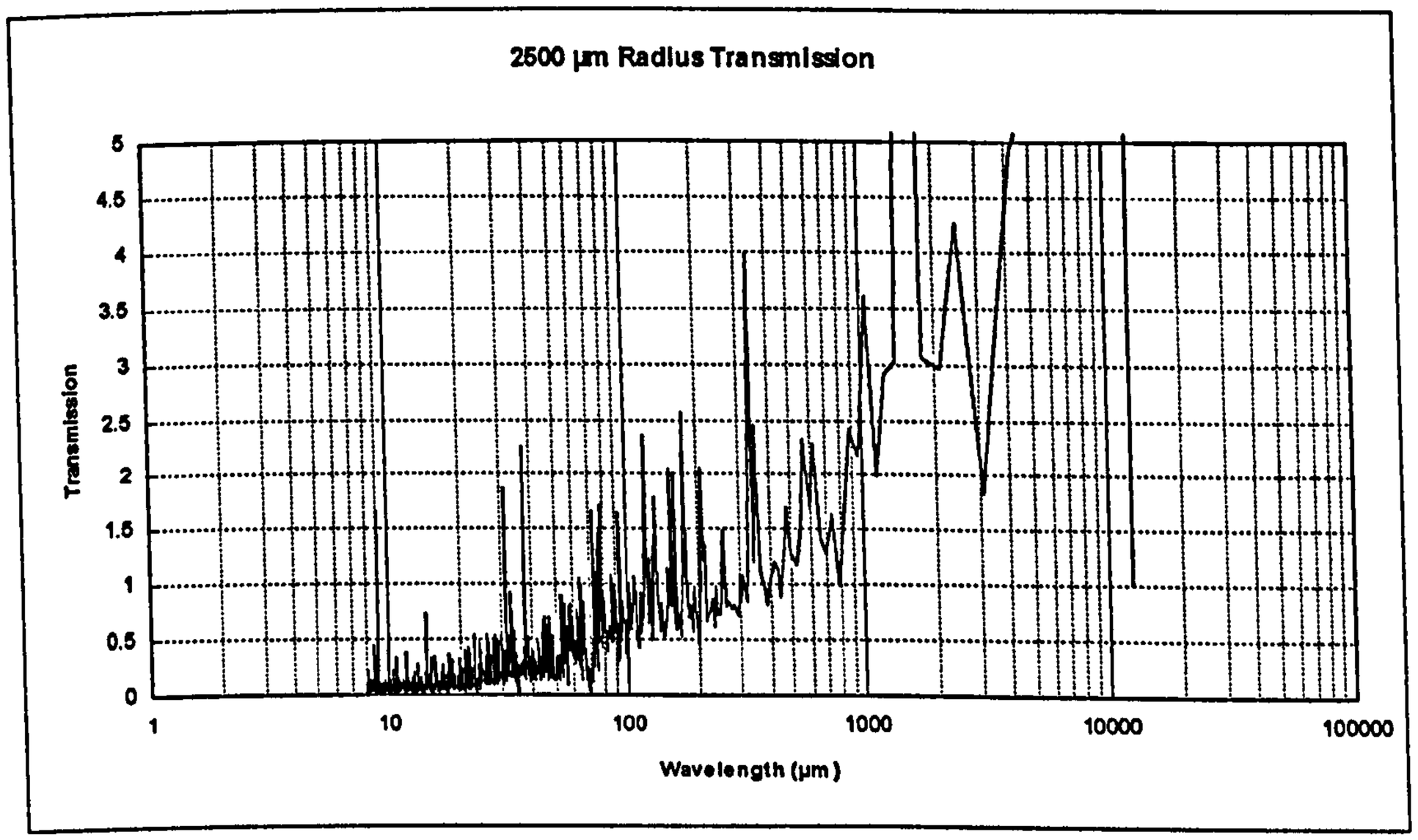


Figure B.31 Turned profile with 2500 µm tip radius.

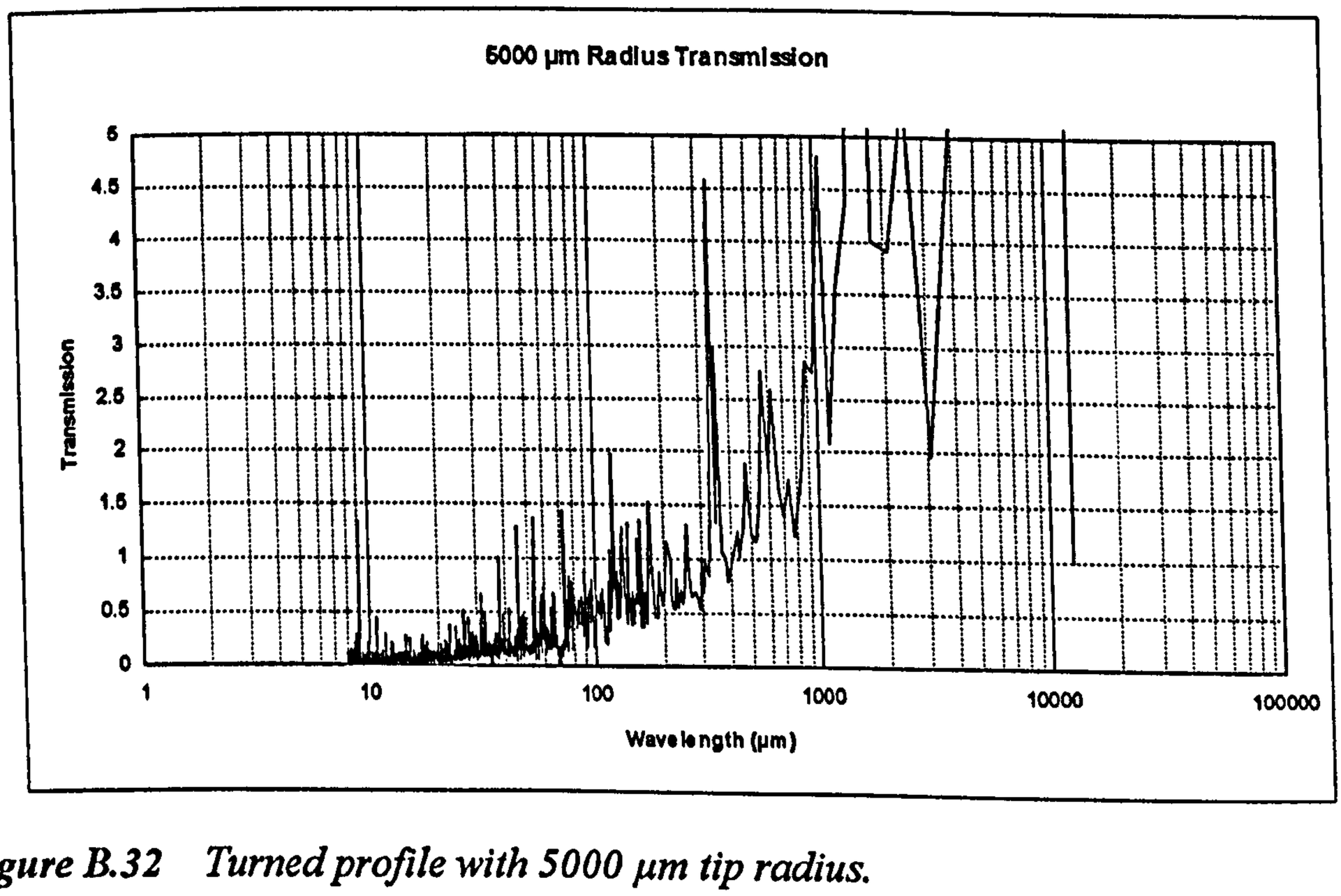


Figure B.32 Turned profile with 5000 µm tip radius.

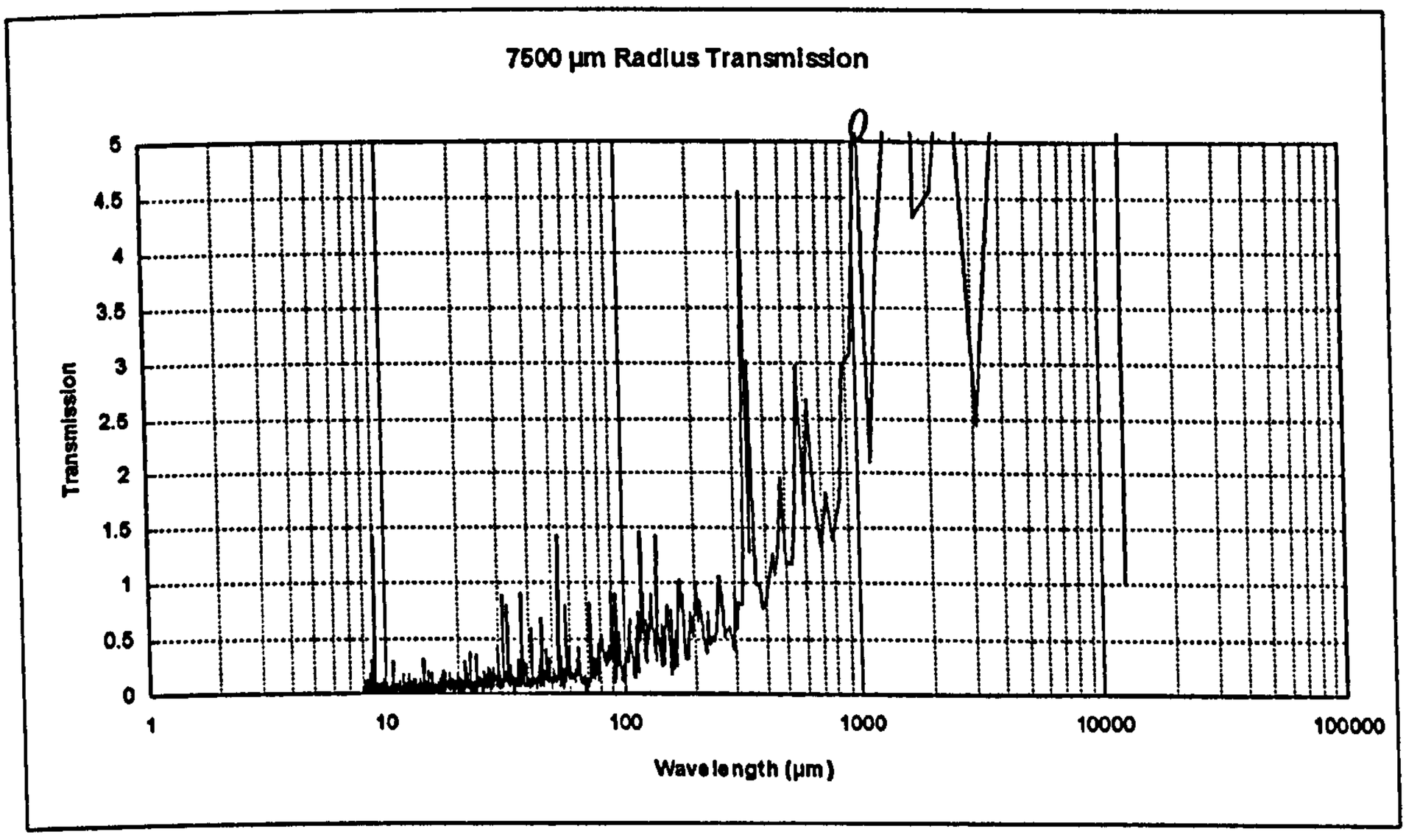


Figure B.33 Turned profile with 7500 µm tip radius.

B.5 Turned Profile Parameter Analyses

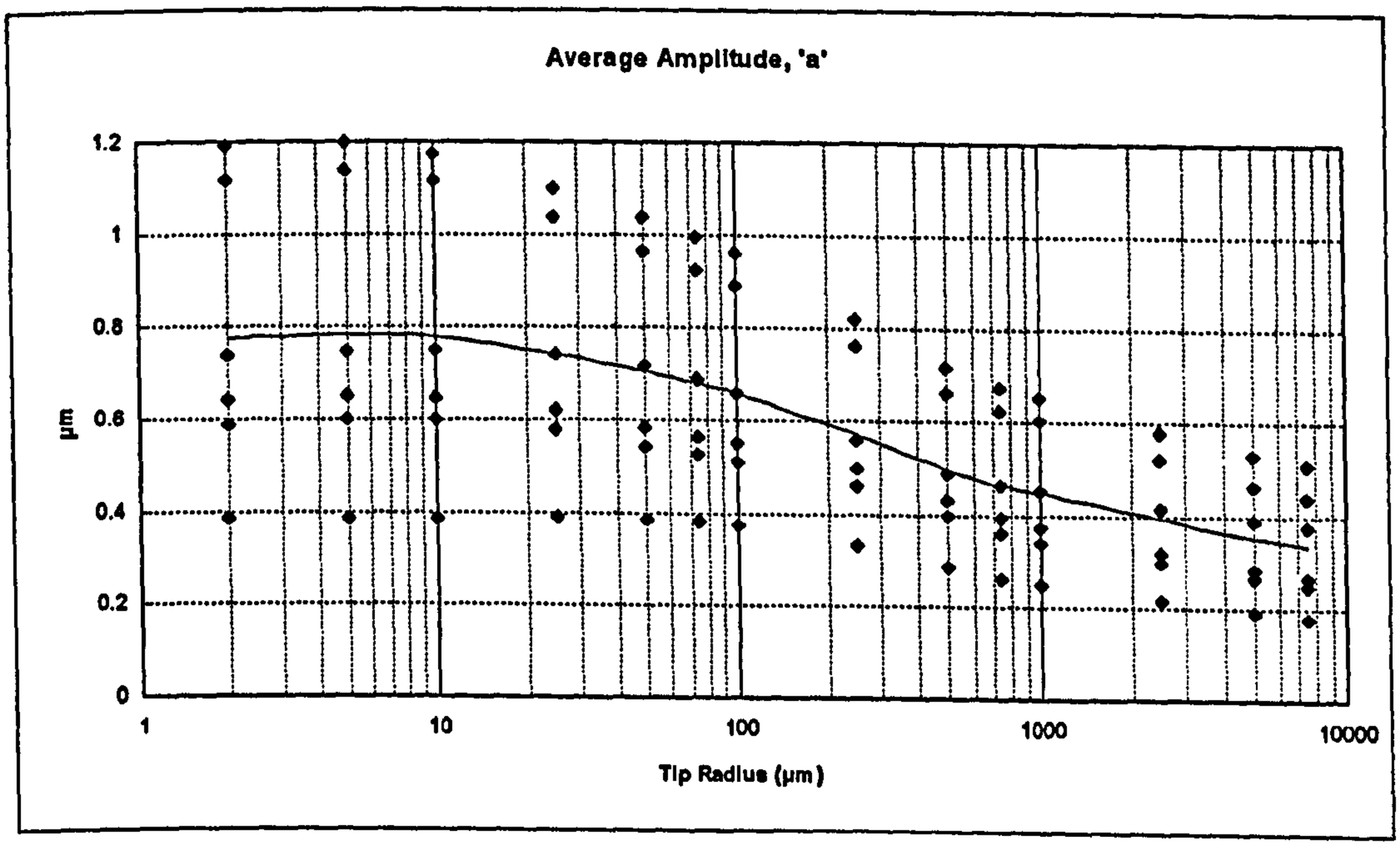


Figure B.34 Turned profile.

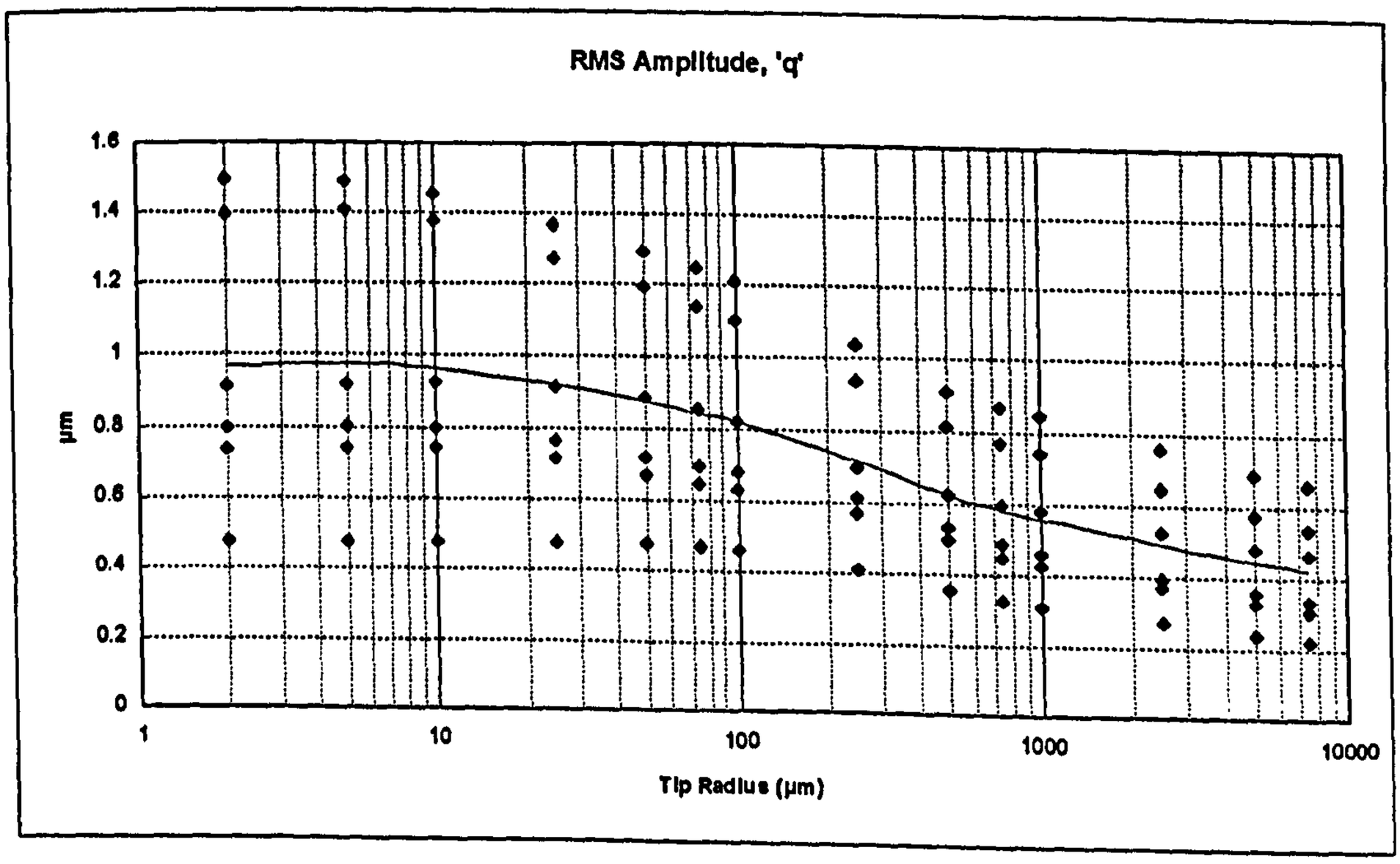


Figure B.35 Turned profile.

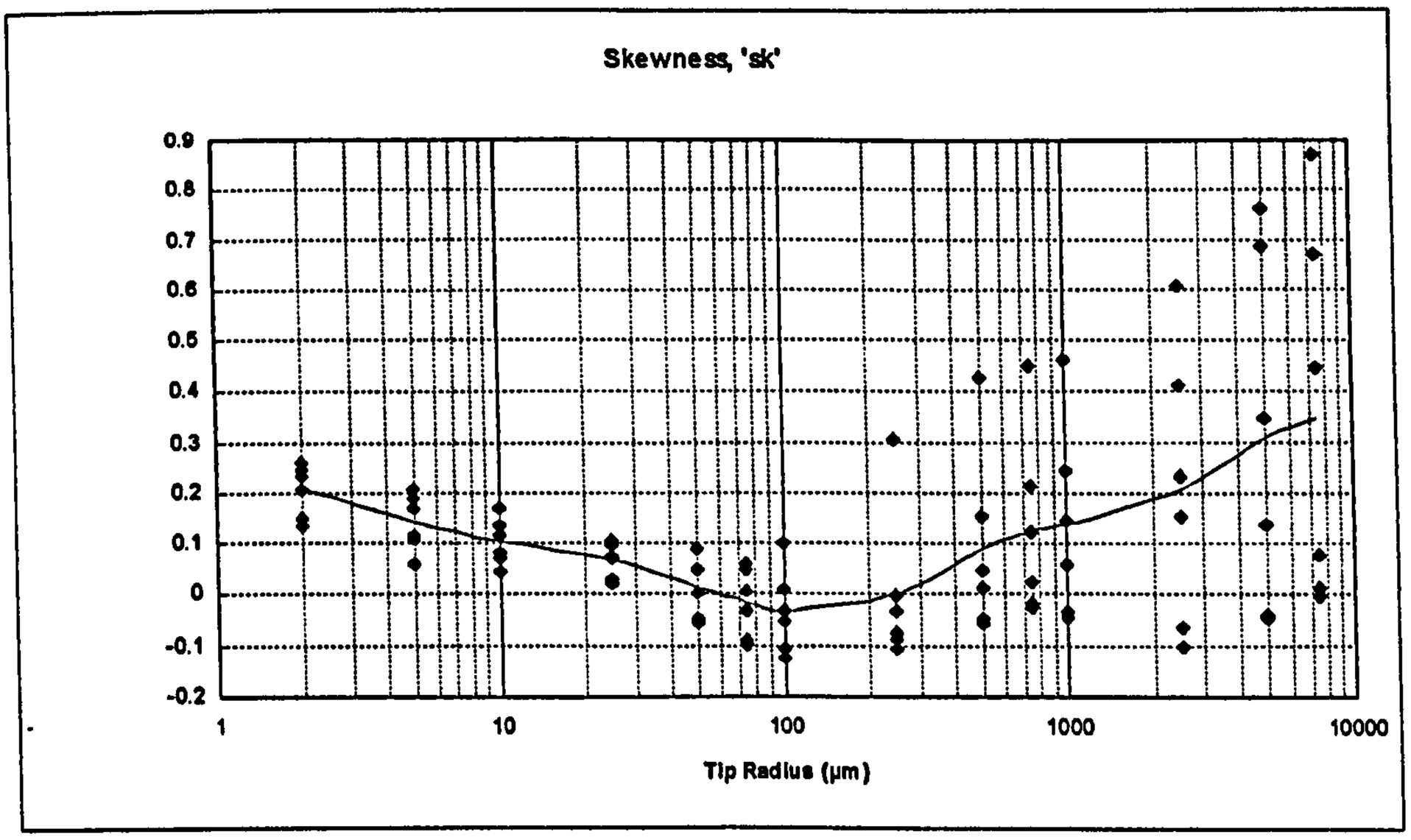


Figure B.36 Turned profile.

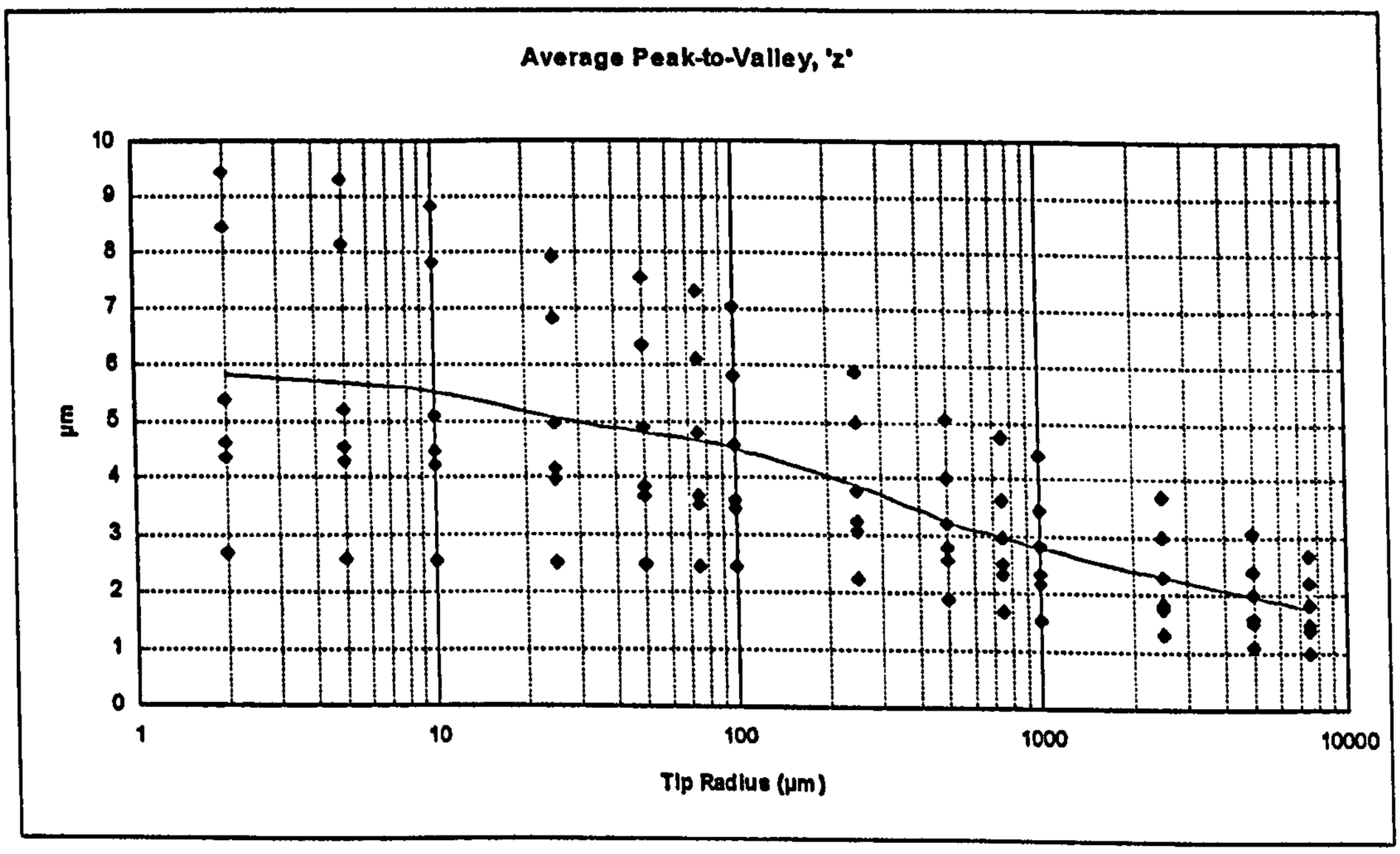


Figure B.37 Turned profile.

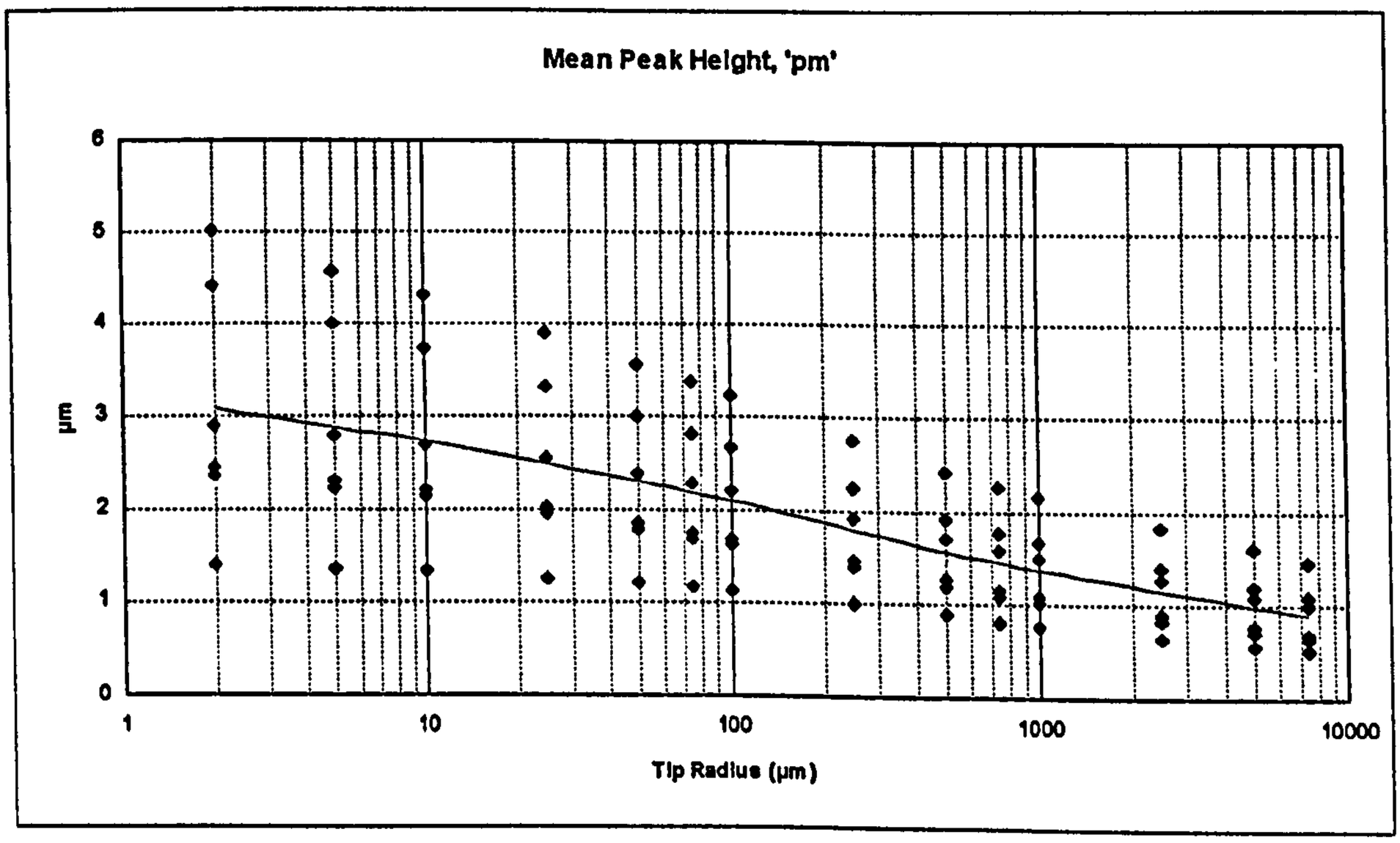


Figure B.38 Turned profile.

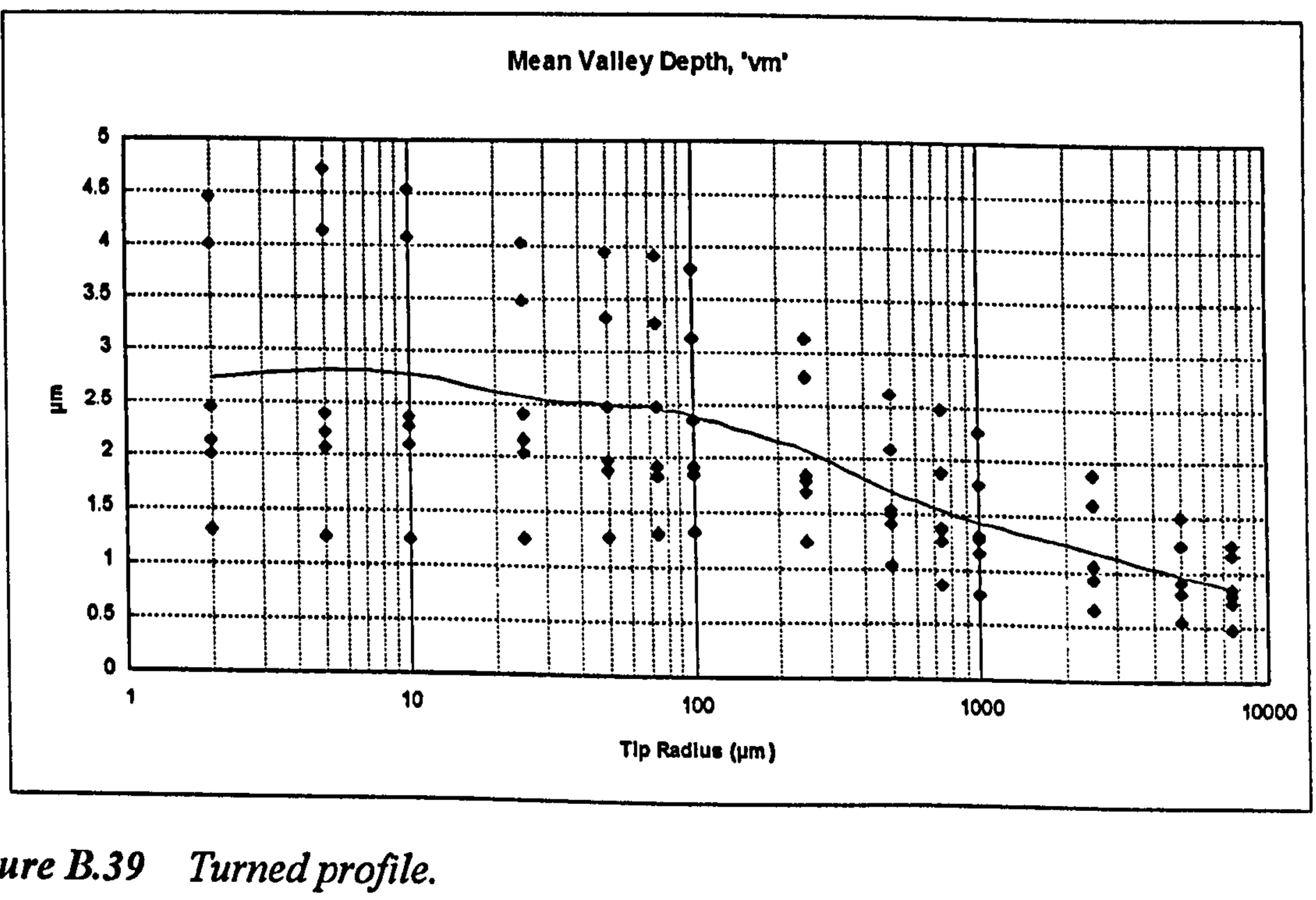


Figure B.39 Turned profile.

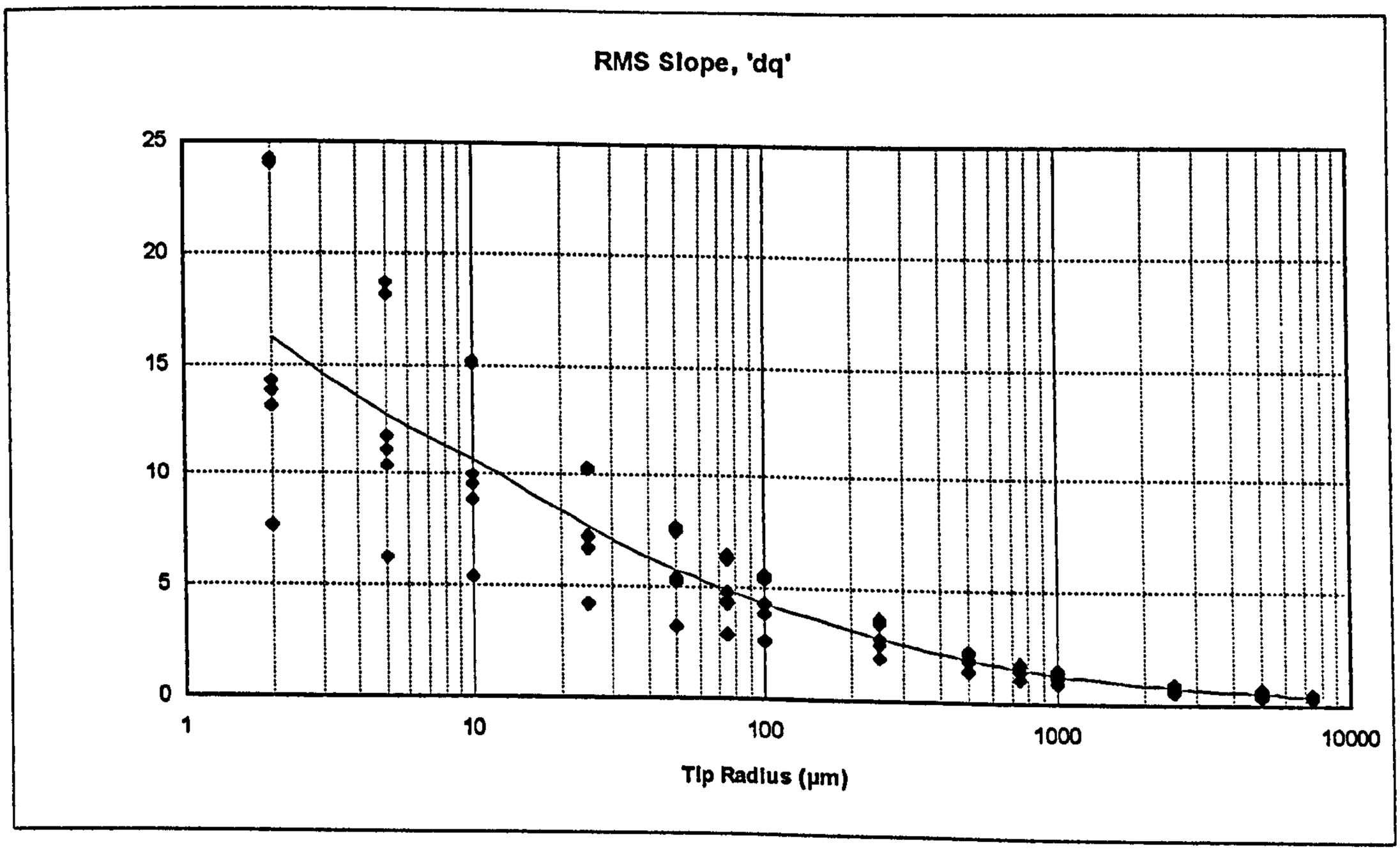


Figure B.40 Turned profile.

B.6 Plateau Honed Profile Transmission Analyses

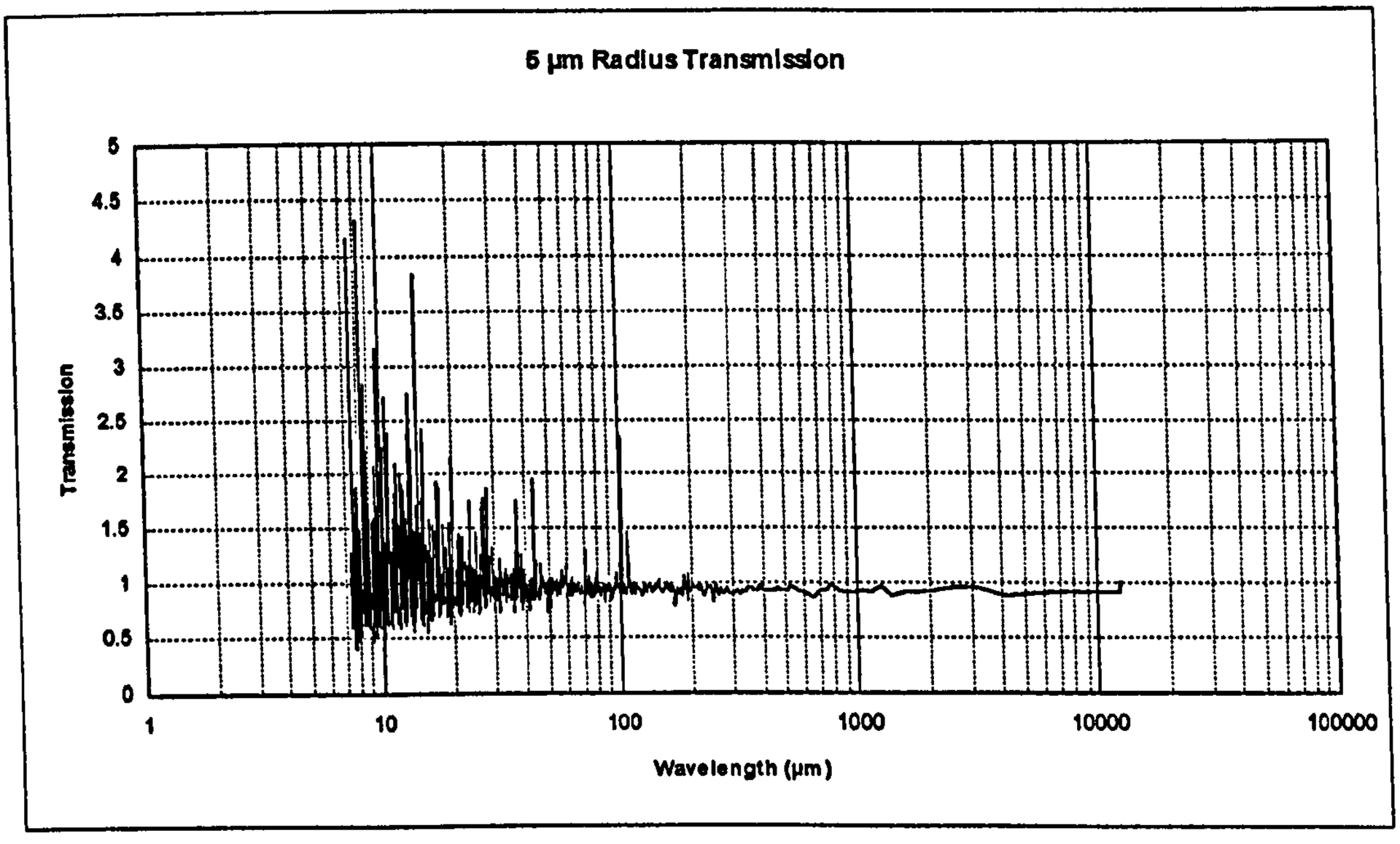


Figure B.41 Plateau honed profile with 5 µm tip radius.

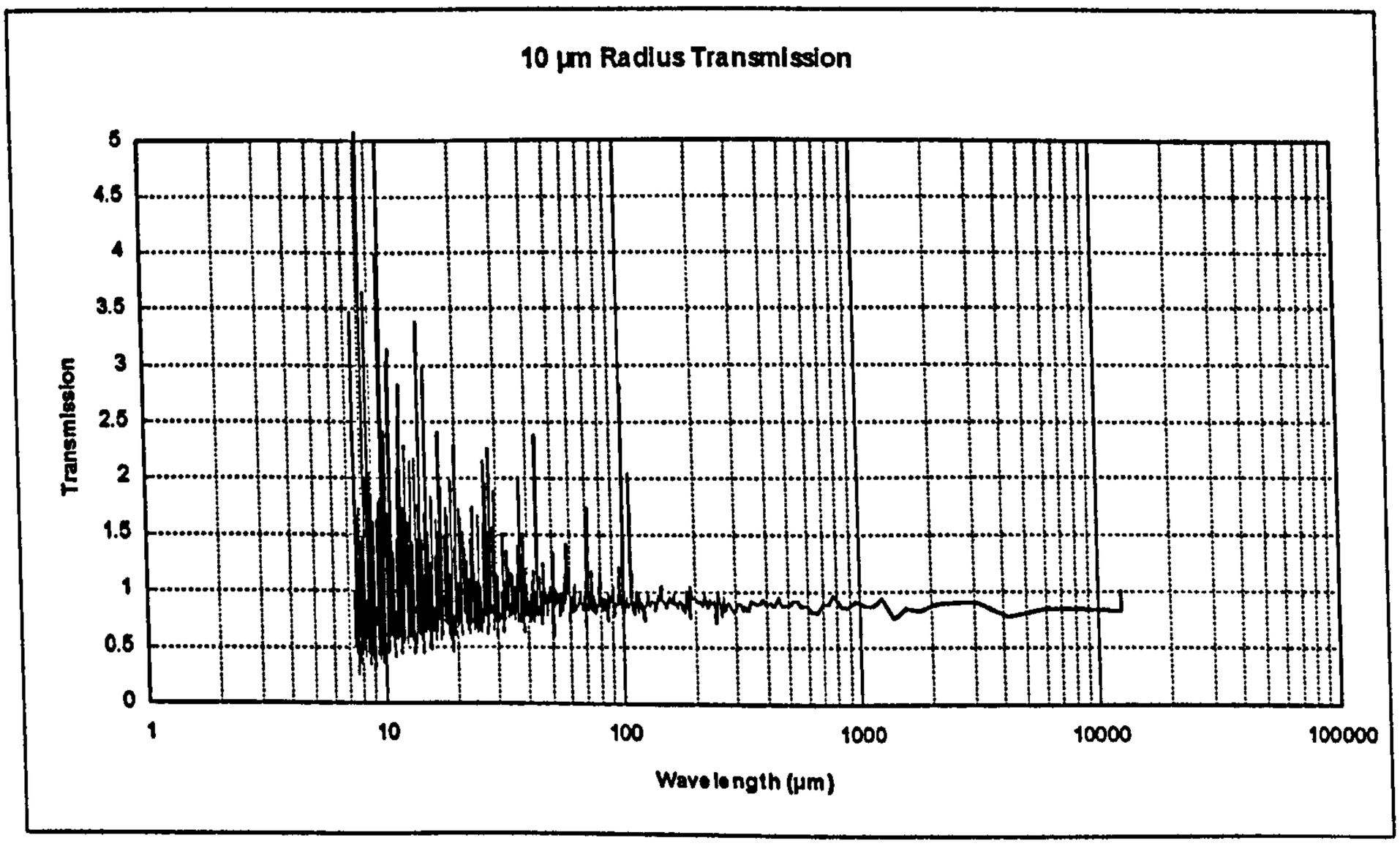


Figure B.42 Plateau honed profile with 10 µm tip radius.

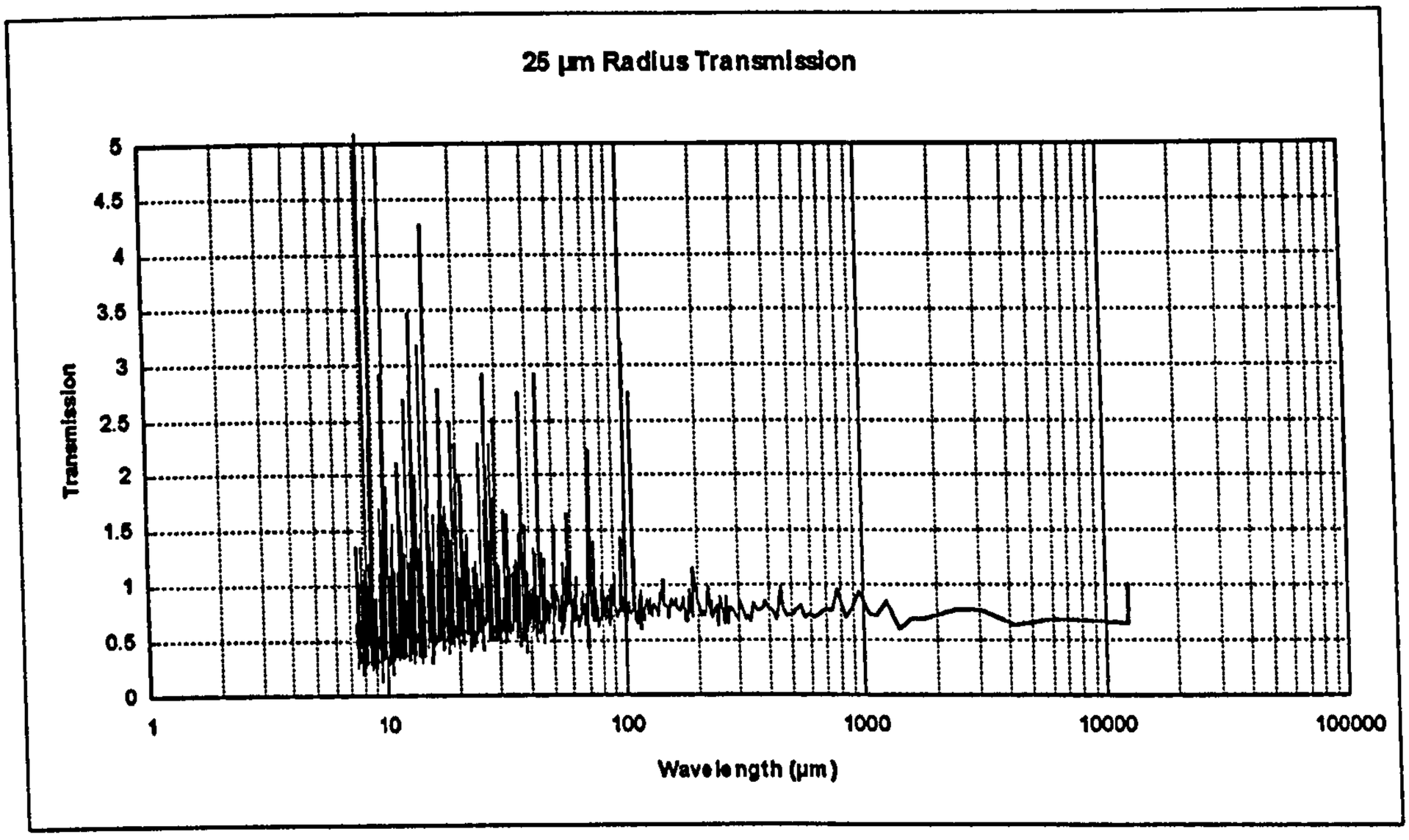


Figure B.43 Plateau honed profile with 25 µm tip radius.

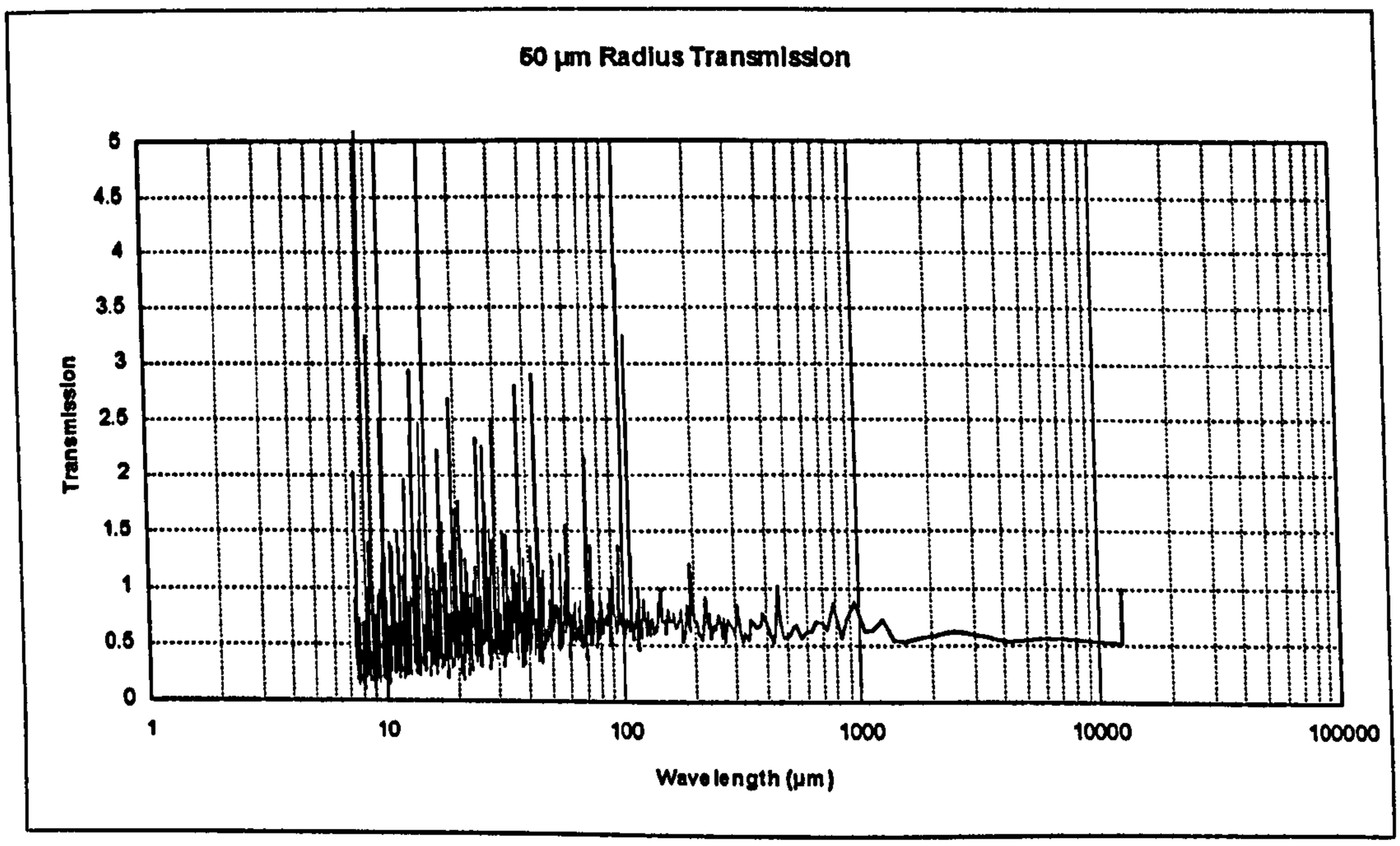


Figure B.44 Plateau honed profile with 50 µm tip radius.

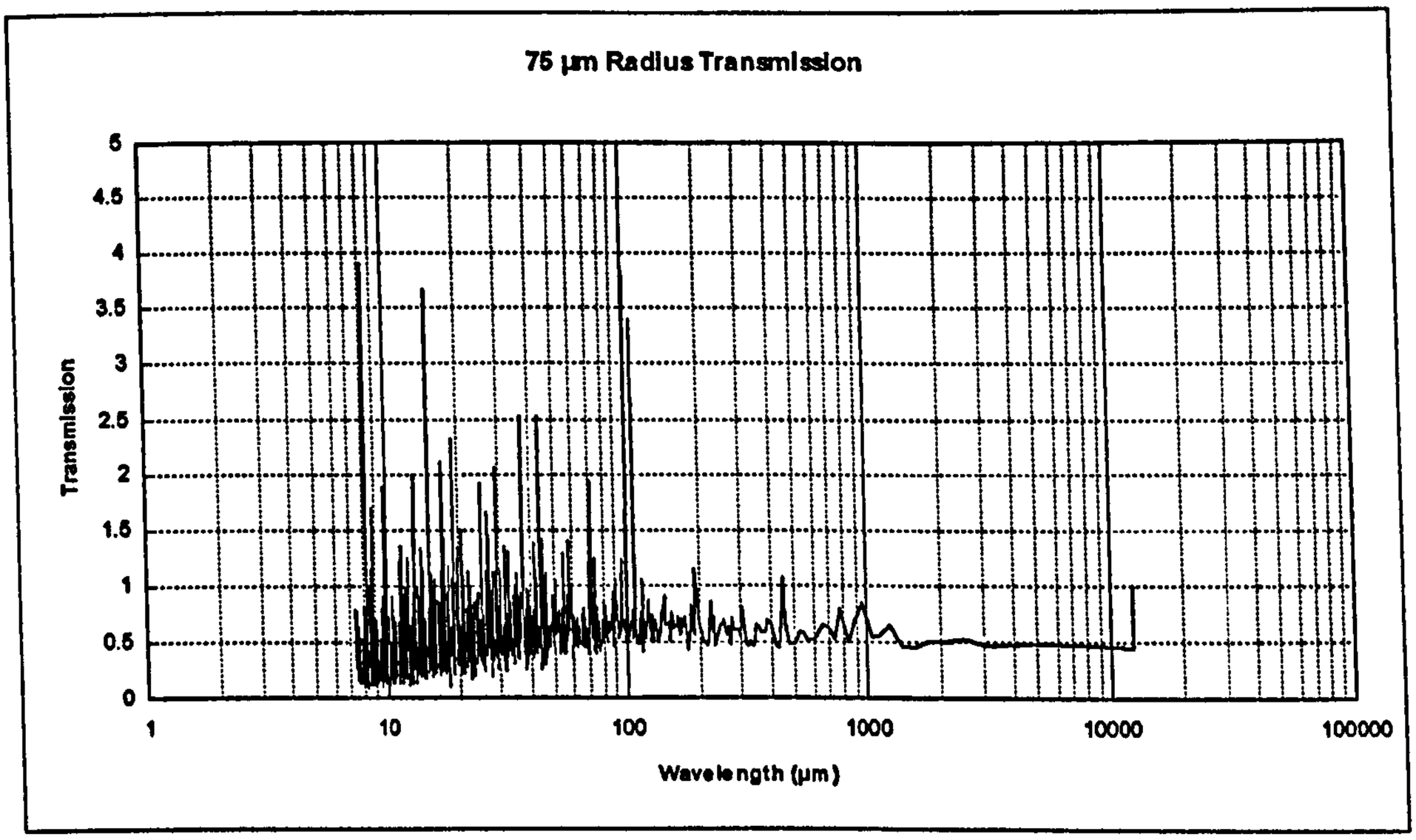


Figure B.45 Plateau honed profile with 75 µm tip radius.

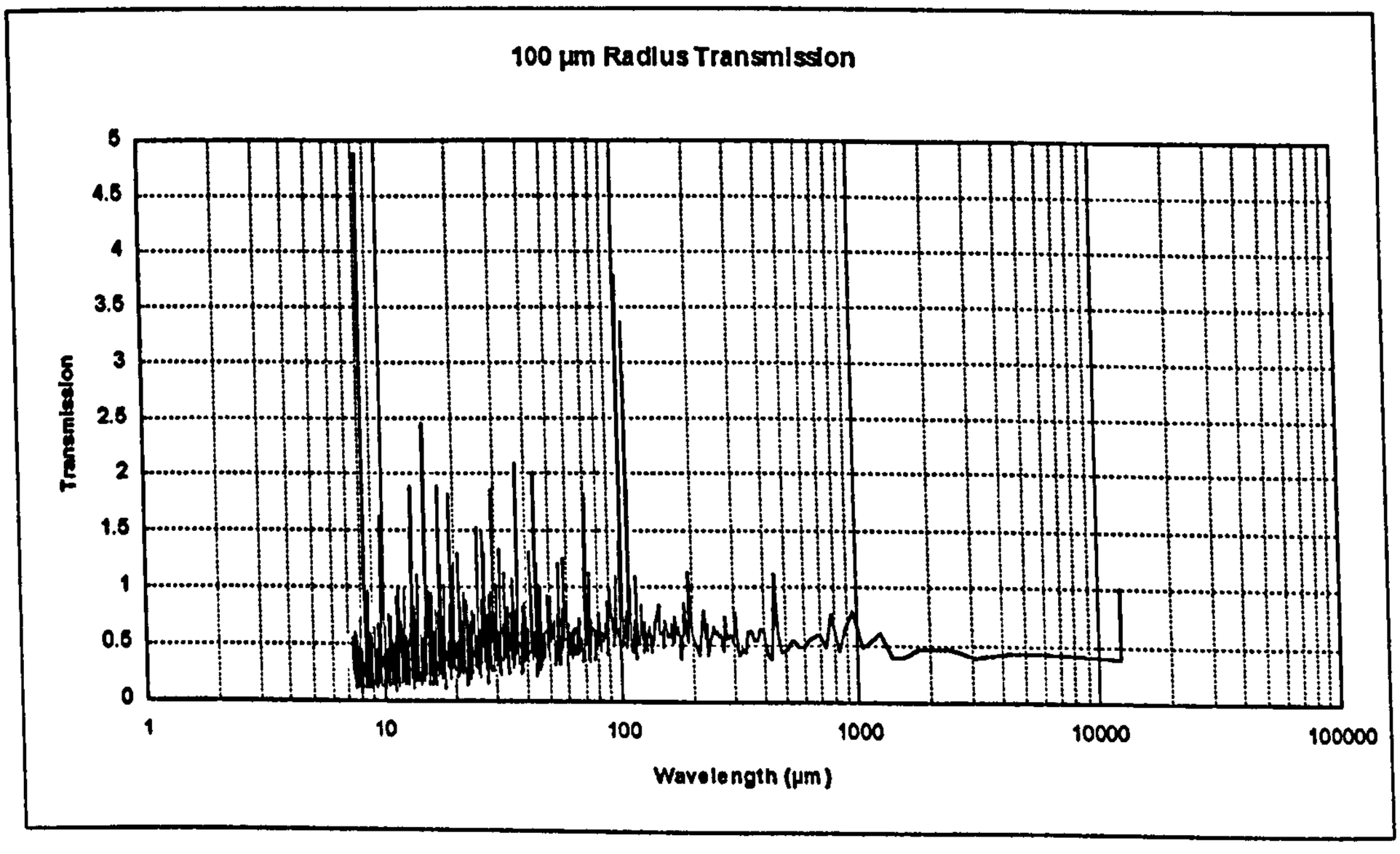


Figure B.46 Plateau honed profile with 100 µm tip radius.

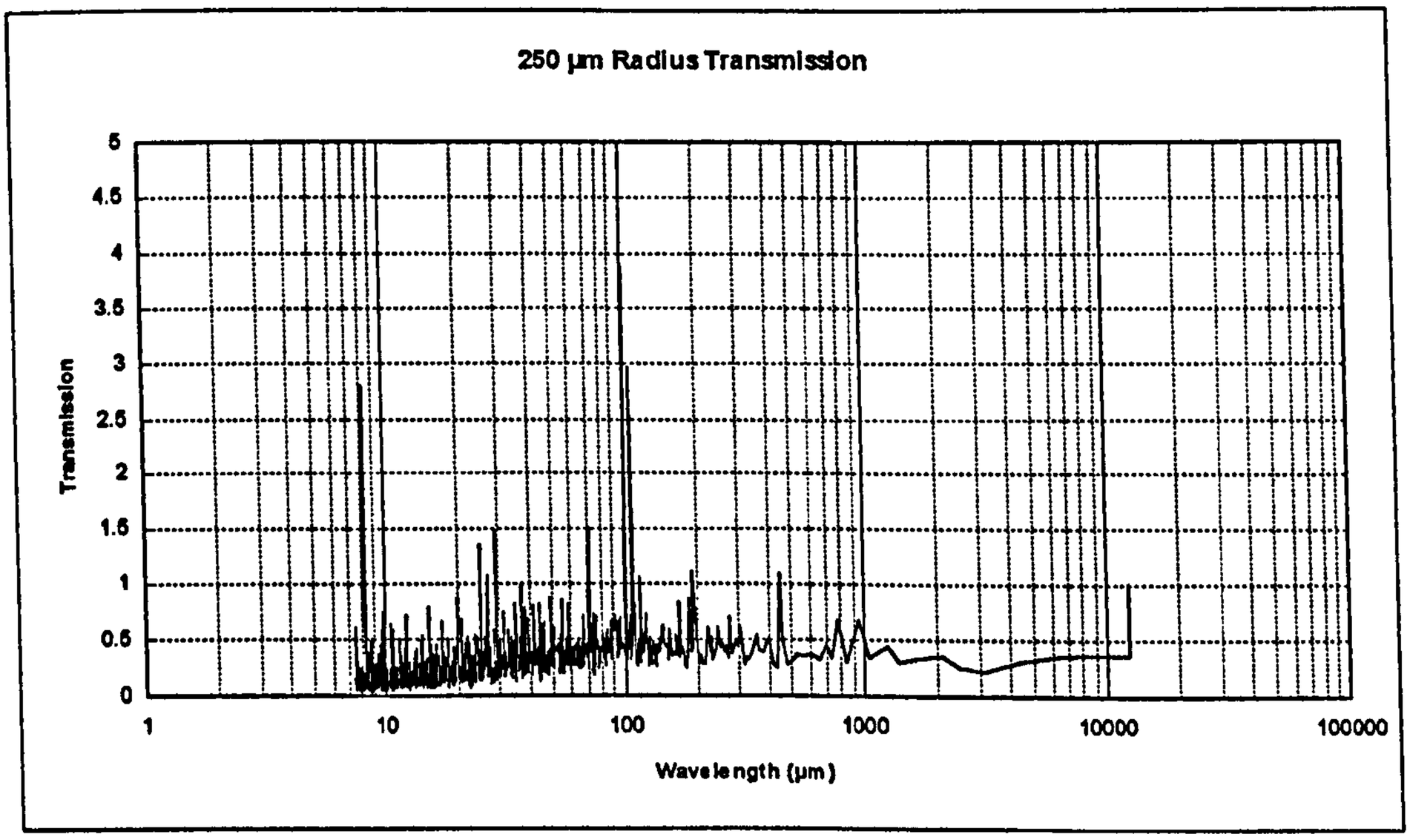


Figure B.47 *Plateau honed profile with 250 µm tip radius.*

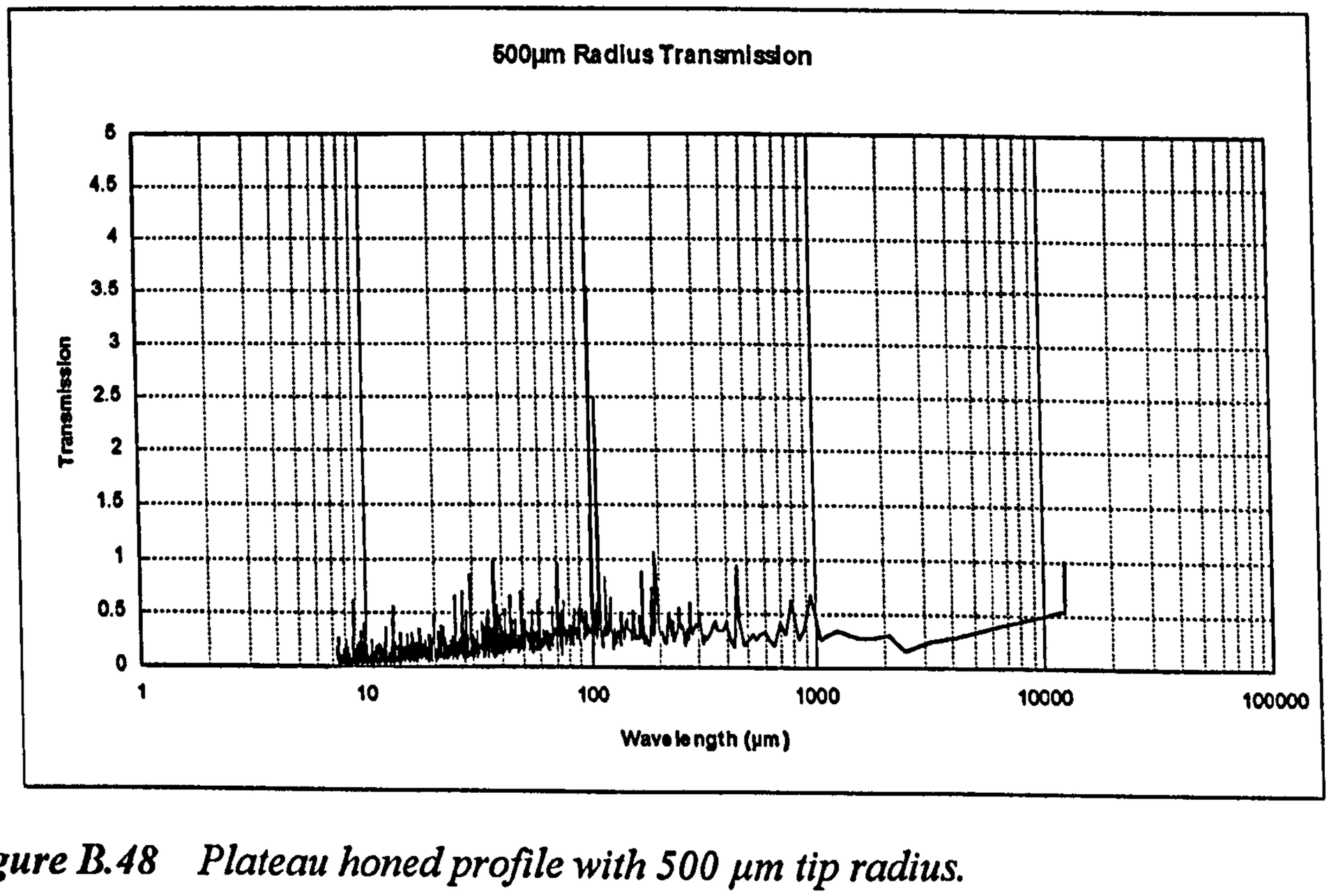


Figure B.48 *Plateau honed profile with 500 µm tip radius.*

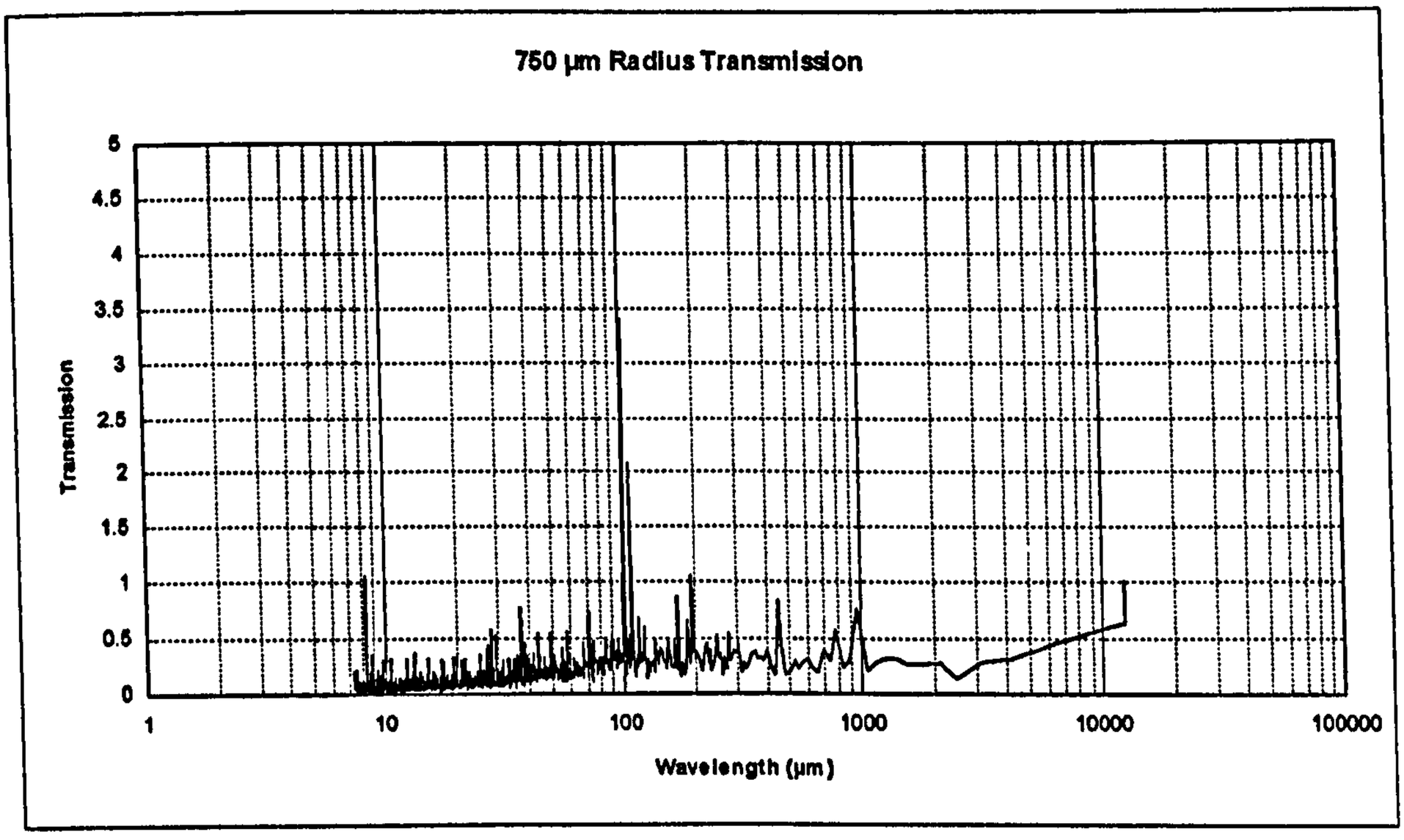


Figure B.49 *Plateau honed profile with 750 µm tip radius.*

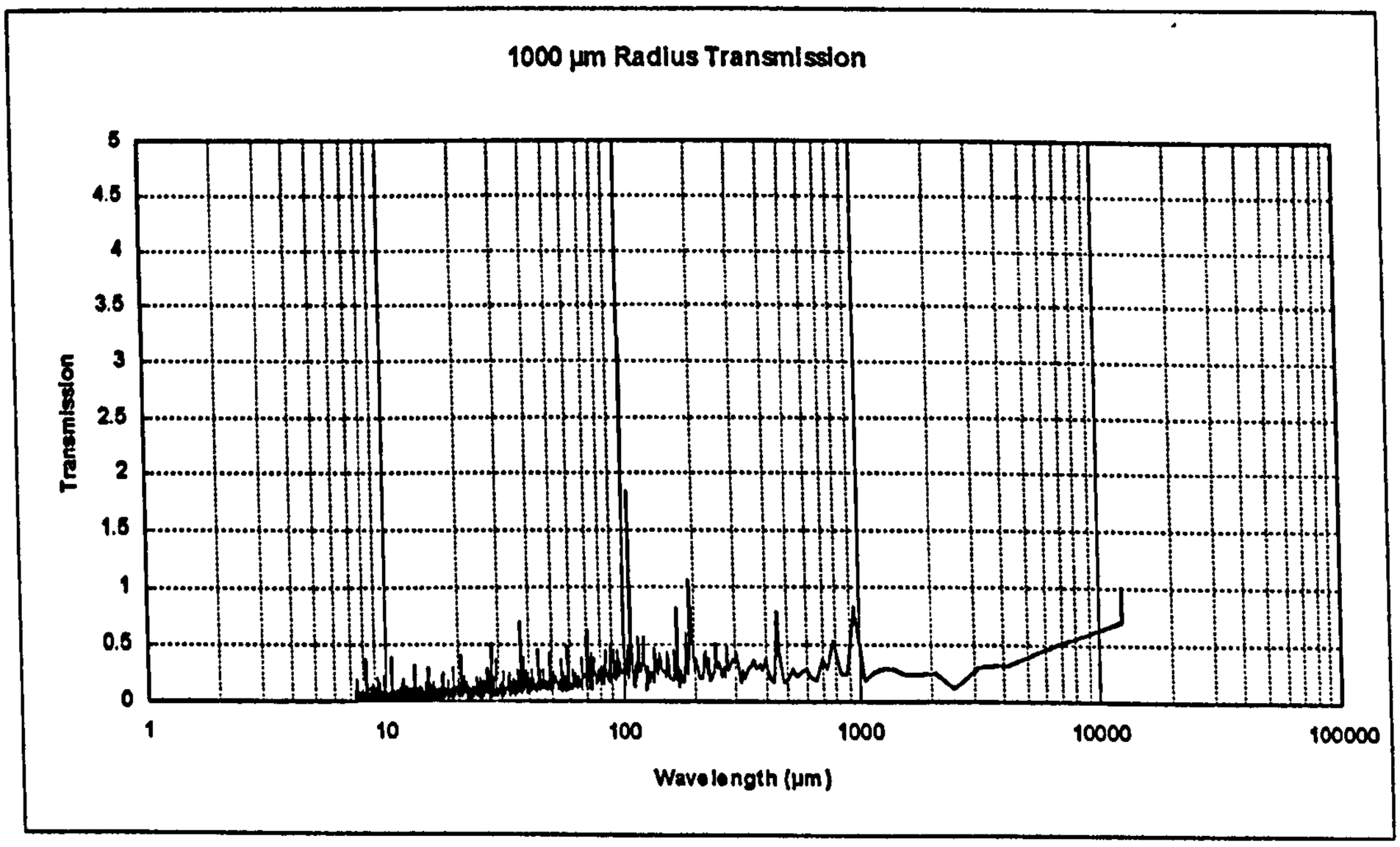


Figure B.50 *Plateau honed profile with 1000 µm tip radius.*

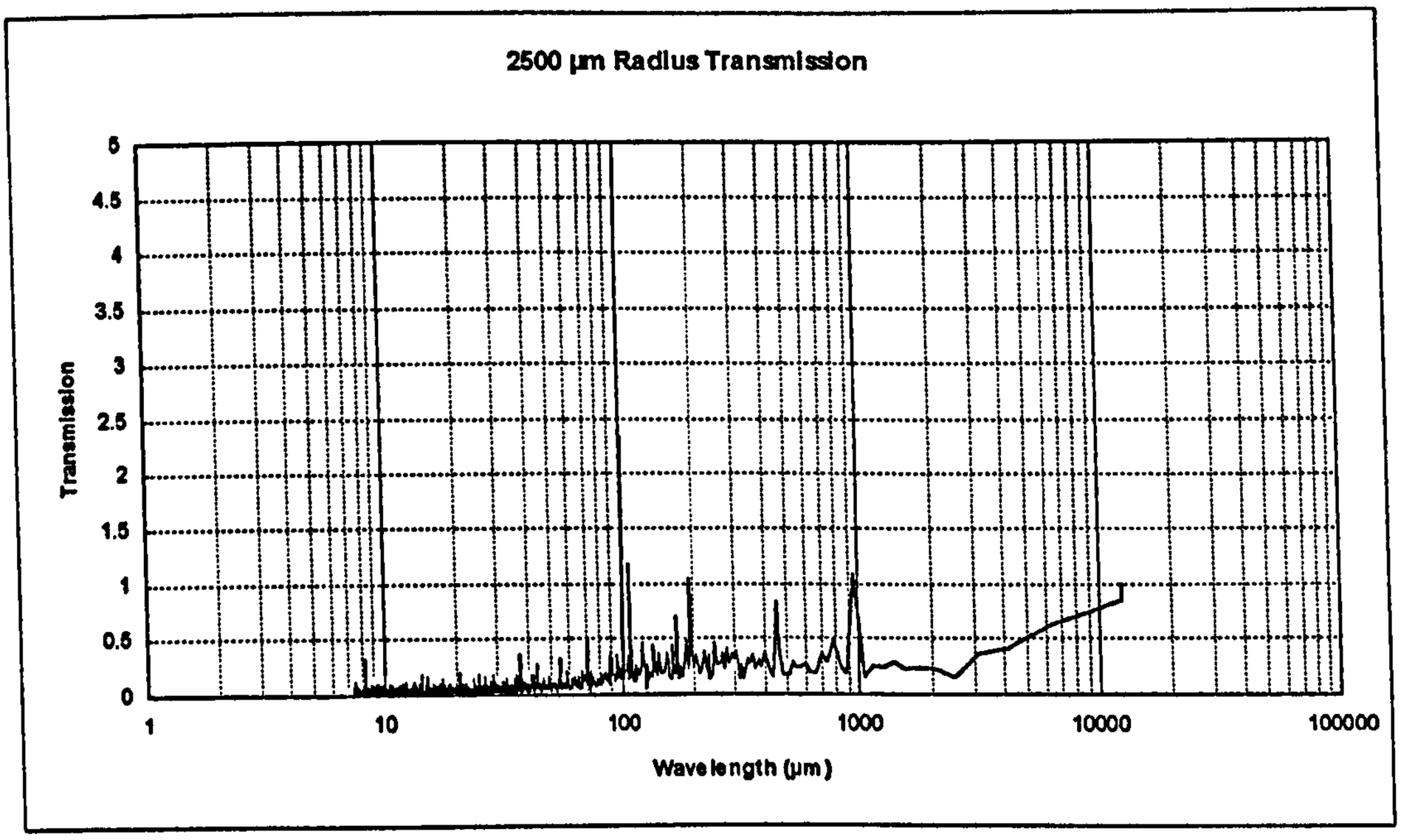


Figure B.51 *Plateau honed profile with 2500 µm tip radius.*

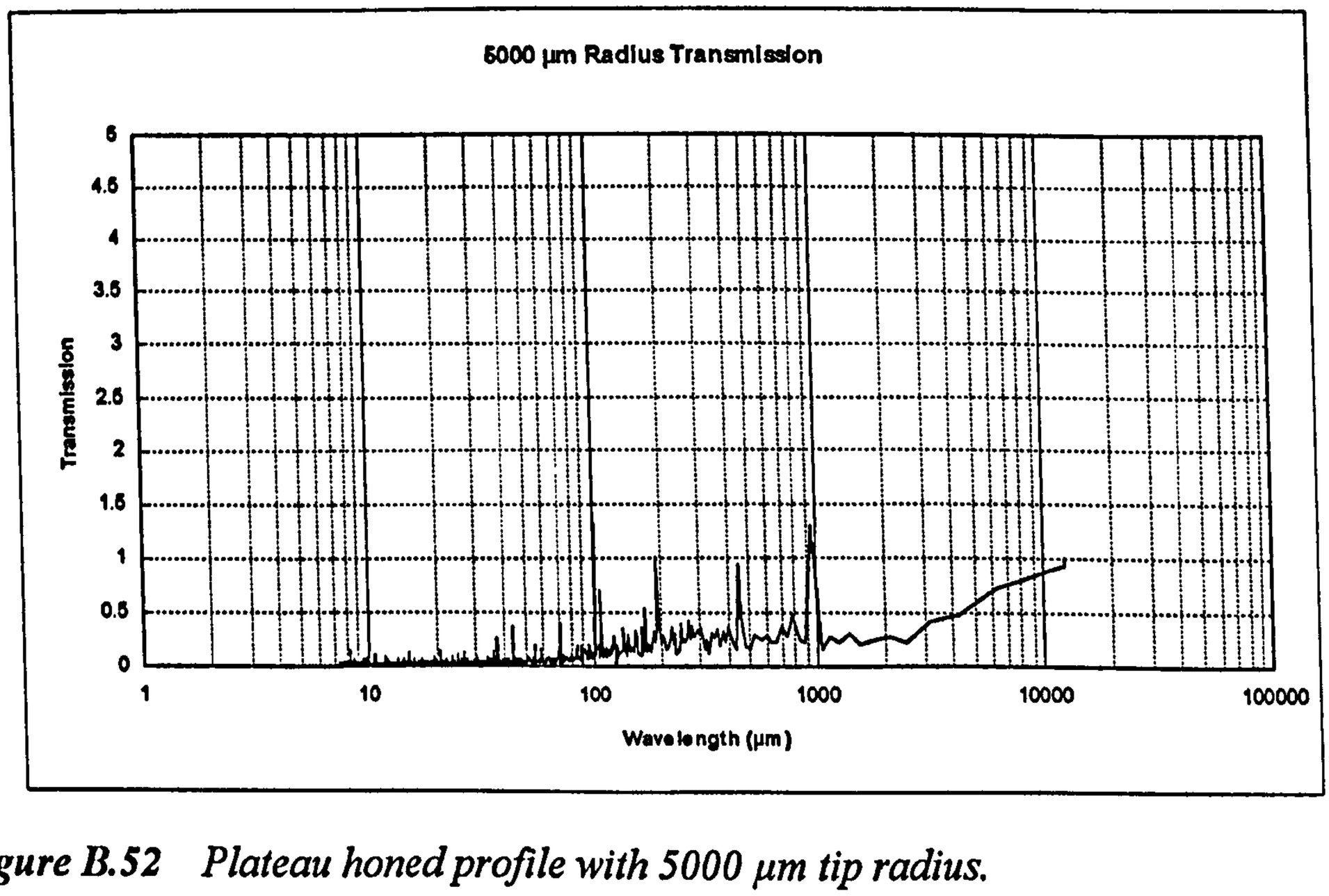


Figure B.52 *Plateau honed profile with 5000 µm tip radius.*

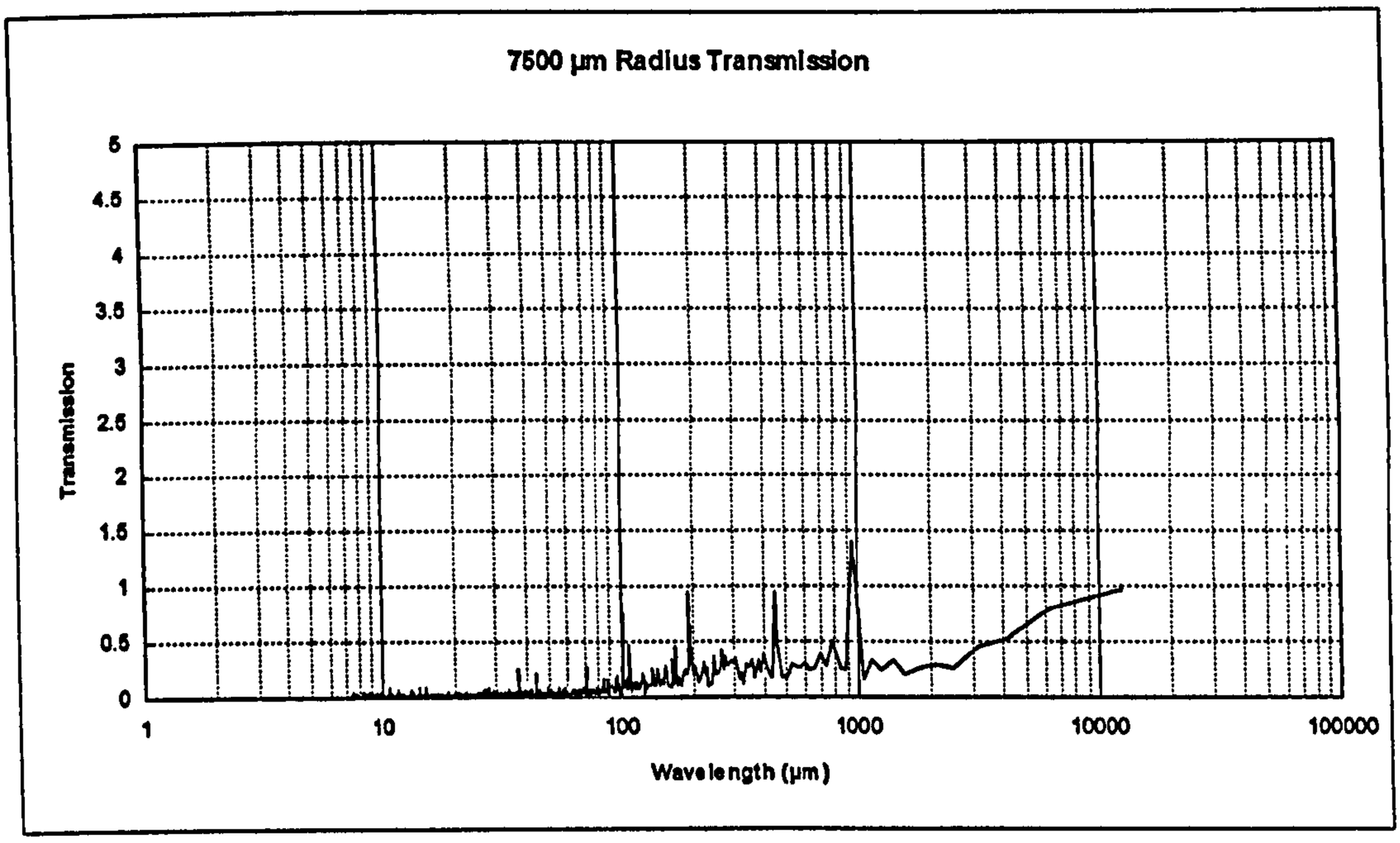


Figure B.53 *Plateau honed profile with 7500 µm tip radius.*

B.7 Plateau Honed Profile Parameter Analyses

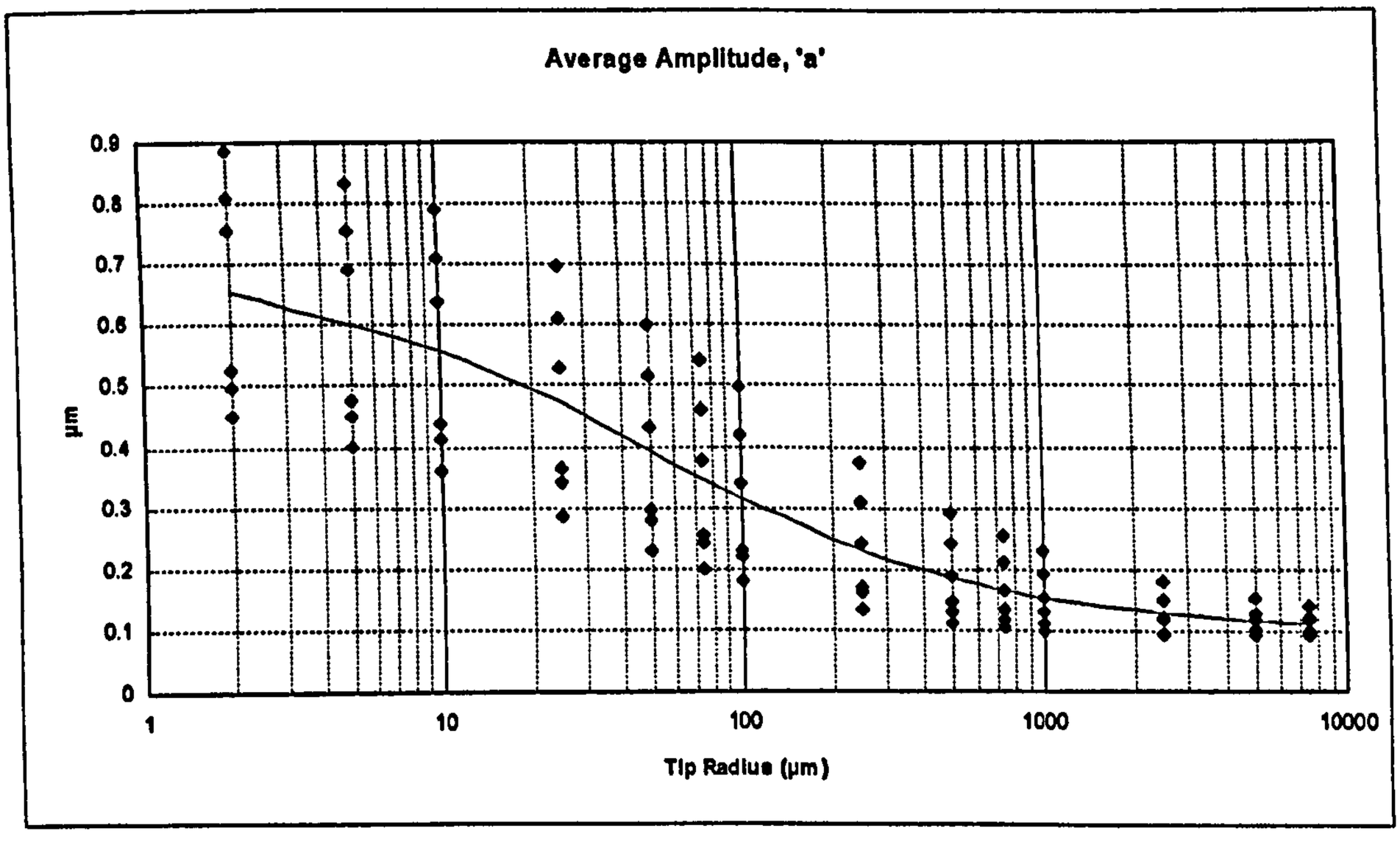


Figure B.54 Plateau honed profile.

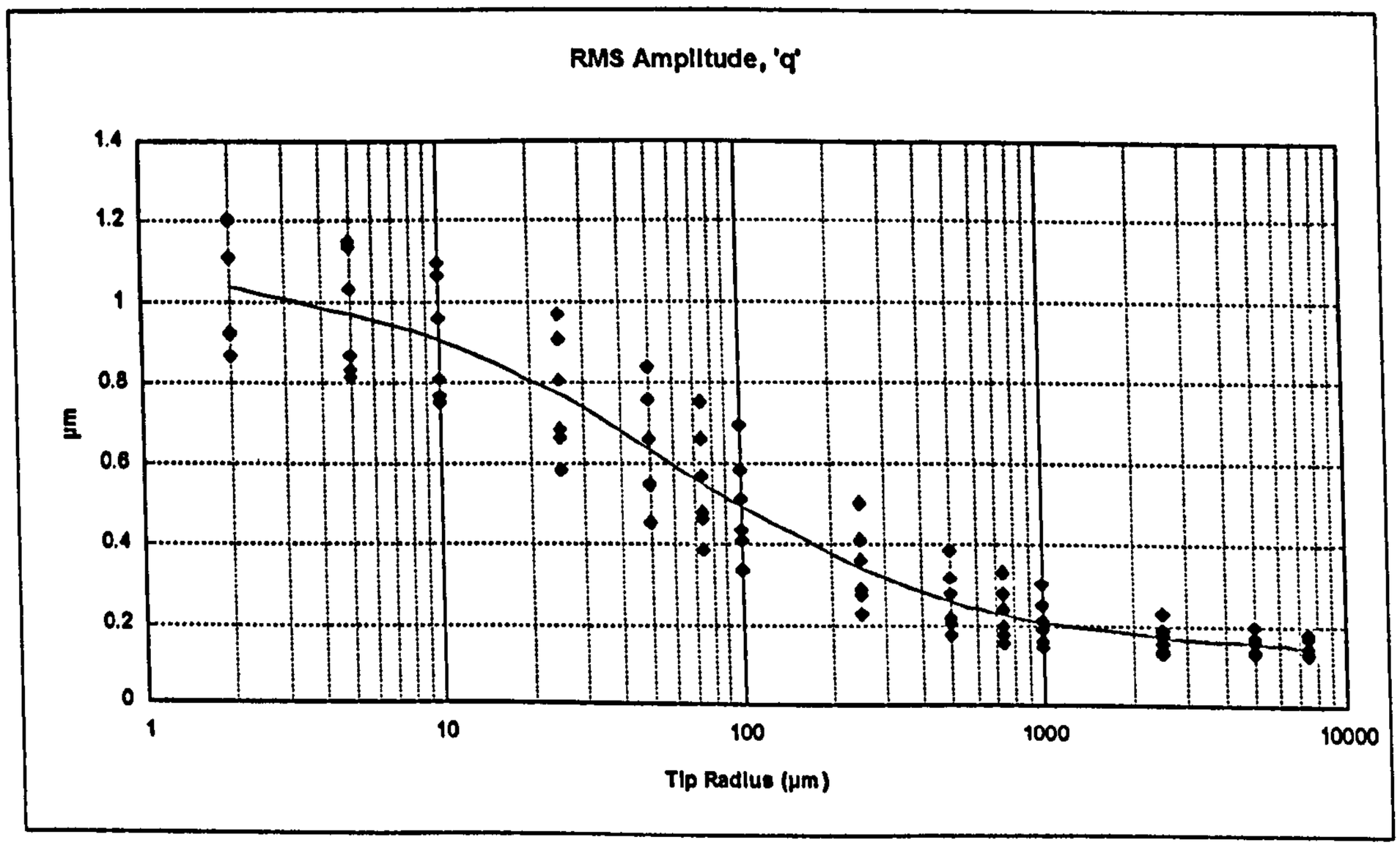


Figure B.55 Plateau honed profile.

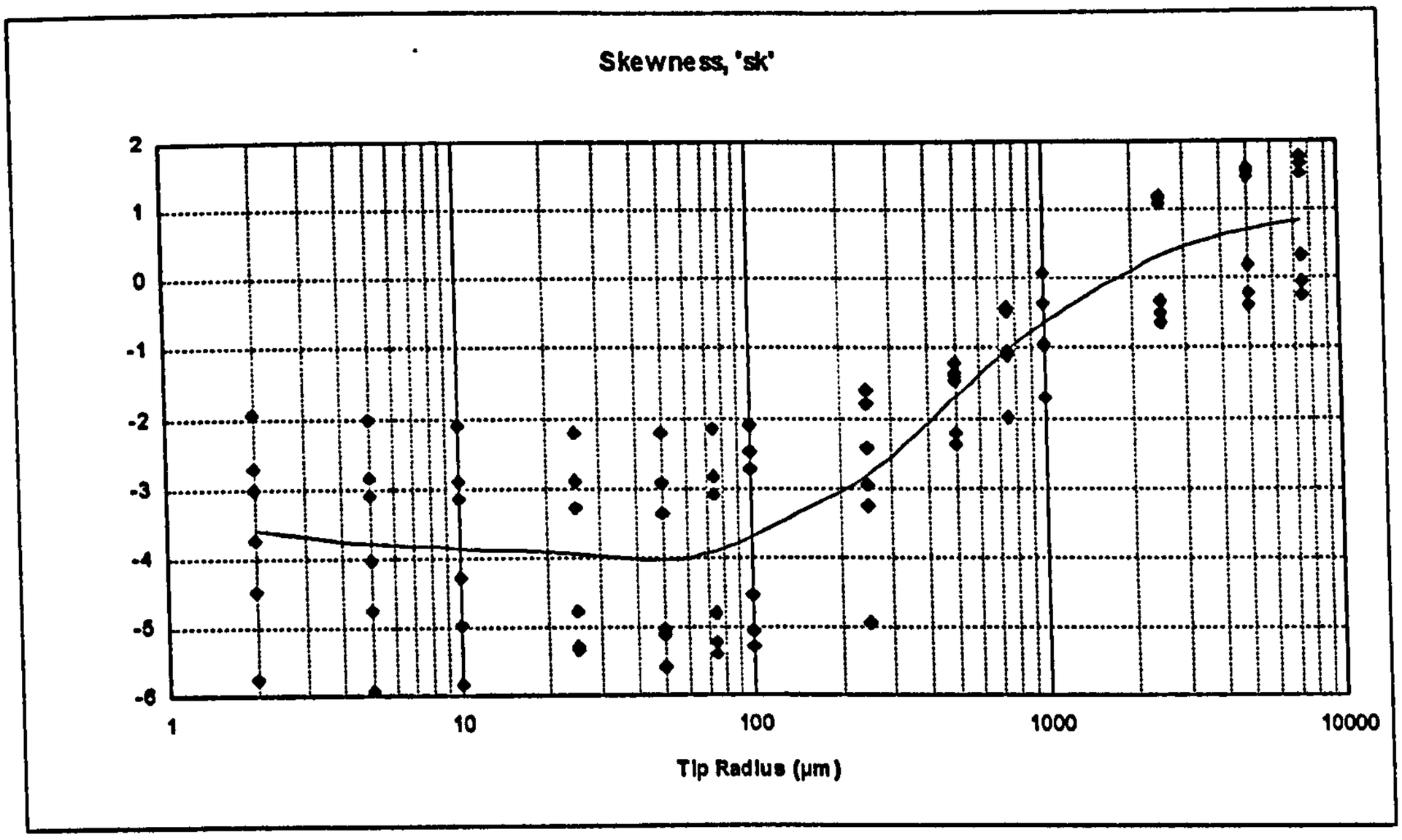


Figure B.56 Plateau honed profile.

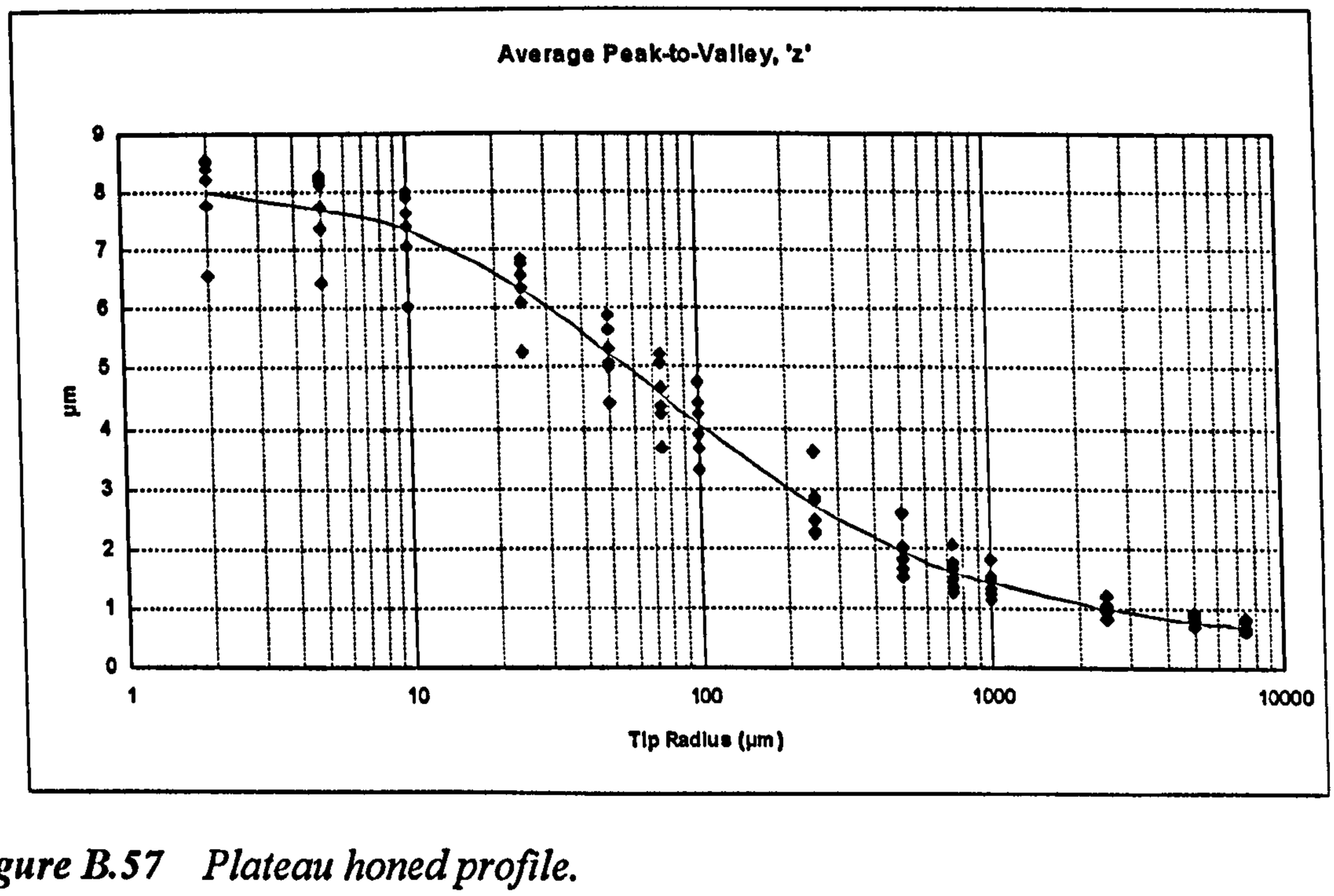


Figure B.57 Plateau honed profile.

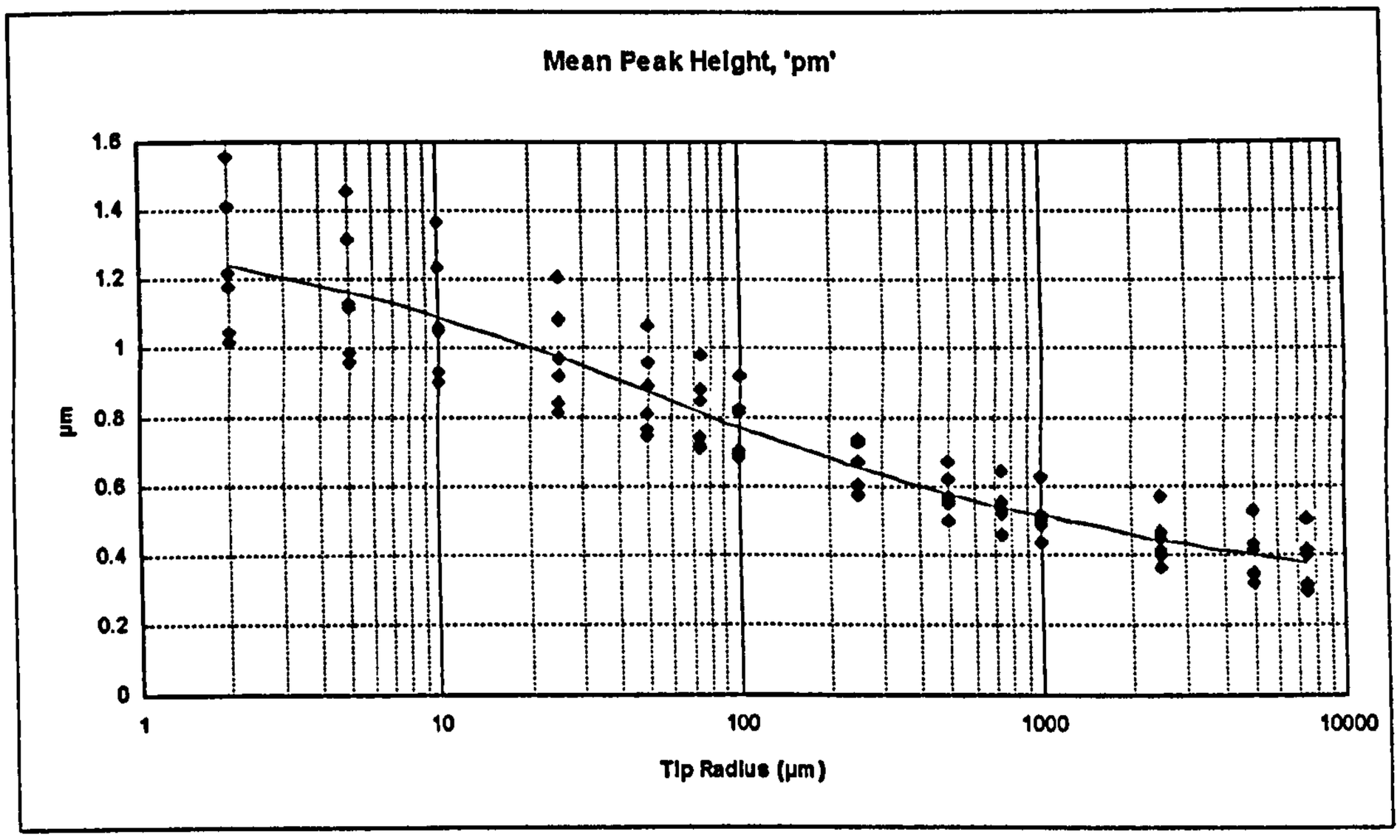


Figure B.58 Plateau honed profile.

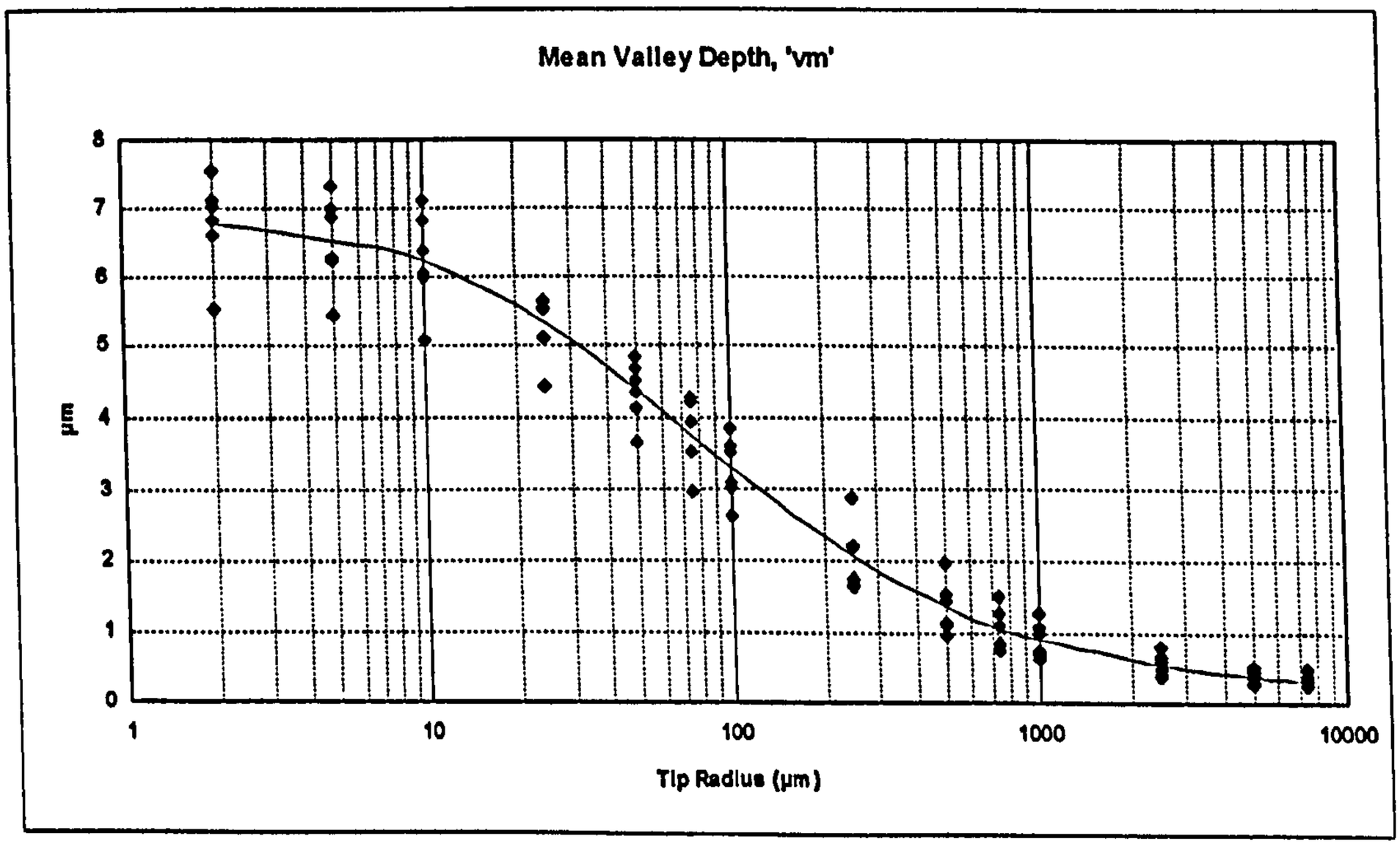


Figure B.59 Plateau honed profile.

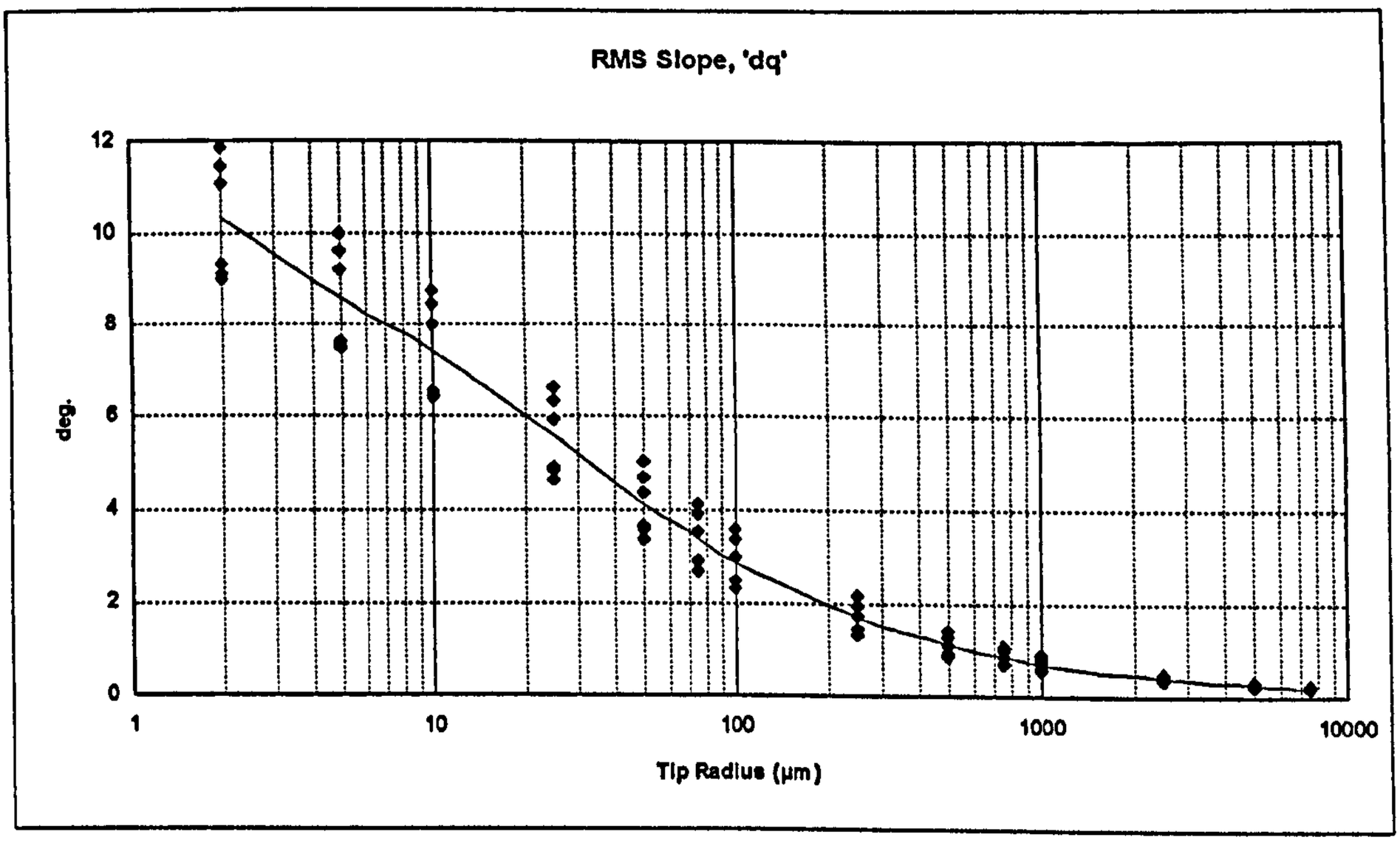


Figure B.60 Plateau honed profile.

A Unified Methodology for the Application of Surface Metrology:

Appendix D

Gaussian Filter Implementations

Listing D.01 - Time Domain Gaussian Convolution

```
#include <stdio.h>
#include <malloc.h>
#include <math.h>
#include "gaussalg.h"

extern double PI ;

int Time_Domain_Gauss (double huge *p_array, double huge *w_array,
                       double spacing_um, double cutoff_mm, unsigned long numpts,
                       unsigned long *first_filtered, unsigned long *last_filtered)
{
   long j, side_points ;
   unsigned long i, last_point ;
   double temp, alpha, denom, sum ;
   double huge *weights ;

   alpha = sqrt (log(2.0)/PI) ;
   temp = (cutoff_mm*1000.0)/spacing_um ;
   side_points = ROUND (temp) ;
   if (side_points < 2)
   {
      printf ("Cutoff too small.") ;
      return FALSE ;
   }
   if (2*side_points >= (numpts-10))  // need at least 10 filtered points
   {
      printf ("Cutoff too large.") ;
      return FALSE ;
   }
   weights = (double *)farcalloc(2*side_points+1, sizeof(double)) ;
   if (weights == NULL)
   {
      printf ("Could not allocate for filter.") ;
      return FALSE ;
   }

   denom = alpha*cutoff_mm ;
   sum = weights[side_points] = 1.0/denom ;
   for (i=1; i<=side_points; i++)
   {
      temp = i*(spacing_um/1000.0)/denom ;
      temp = exp(-PI*temp*temp)/denom ;
      weights[side_points-i] = weights[side_points+i] = temp ;
      sum += 2.0*temp ;
   }

   sum = 1.0/sum ;
   weights[side_points] *= sum ;
   for (i=1; i<=side_points; i++)
   {
      weights[side_points-i] *= sum ;
      weights[side_points+i] *= sum ;
   }

   last_point = numpts - 1 - side_points ;
   for (i=side_points; i<=last_point; i++)
   {
      sum = 0.0 ;
      for (j=-side_points; j<=side_points; j++)
         sum += weights[(unsigned long)(side_points+j)]*p_array[(unsigned long)(i+j)] ;
      w_array[i] = sum ;
   }

   *first_filtered = side_points ;
   *last_filtered = last_point ;
   farfree (weights) ;
   return TRUE ;

} /* end Time_Domain_Gauss */
```

Listing D.02 - Frequency Domain Gaussian Convolution

```
#include <stdio.h>
#include <math.h>
#include <malloc.h>

#include "gaussalg.h"

extern double PI ;

int Sing_FFT (double huge *a, double huge *b, long n, int forward) ;

/*****************************************************/
int FFT_Gauss (double huge *p_array, double huge *w_array,
               double spacing_um, double cutoff_mm, unsigned long numpts,
               unsigned long *first_filtered, unsigned long *last_filtered)
{
   int nyquist_avail ;
   unsigned long i, last, used_numpts ;
   double huge *temp ;
   double length_mm,
          alpha, pi_alpha_sqr,
          wavelength, wavelength_ratio,
          trans ;

   if ((temp = farcalloc(numpts, sizeof(double))) == NULL)
      return FALSE ;

   // store all original profile in "temp"
   for (i=0; i<numpts; i++)
   {
      temp[i] = p_array[i] ;
      w_array[i] = 0.0 ;
   }

// p_array holds real and w_array holds imaginary
// Reduce used numpts if max prime factor is too big
// (Since this is a waviness filter the reduction in points will
//  be accomodated by extending the last waviness point upon
//  completion of the filtering.)
used_numpts = numpts ;
while (!Sing_FFT (p_array, w_array, used_numpts, TRUE))  // forward
{
   used_numpts-- ;
   for (i=0; i<numpts; i++)   // restore original arrays for retry
   {
      p_array[i] = temp[i] ;
      w_array[i] = 0.0 ;
   }
}

for (i=0; i<used_numpts; i++)
{
   p_array[i] /= (double)used_numpts ;
   w_array[i] /= (double)used_numpts ;
}

if ((used_numpts%2) == 0)
{
   last = used_numpts/2 - 1 ;
   nyquist_avail = TRUE ;
}
else
{
   last = used_numpts/2 ;
   nyquist_avail = FALSE ;
}
```

```
// Waviness transmission, skip DC
alpha = sqrt(log(2.0)/PI) ;
pi_alpha_sqr = PI*alpha*alpha ;
length_mm = used_numpts*spacing_um/1000.0 ;
for (i=1; i<=last; i++)
{
   wavelength = length_mm/(double)i ;
   wavelength_ratio = cutoff_mm/wavelength ;
   trans = exp(-pi_alpha_sqr*wavelength_ratio*wavelength_ratio) ;
   p_array[i] *= trans ;
   w_array[i] *= trans ;
   p_array[used_numpts-i] *= trans ;
   w_array[used_numpts-i] *= trans ;
}

if (nyquist_avail)   // treat nyquist only once
{
   wavelength = length_mm/(double)i ;
   wavelength_ratio = cutoff_mm/wavelength ;
   trans = exp(-pi_alpha_sqr*wavelength_ratio*wavelength_ratio) ;
   p_array[i] *= trans ;
   w_array[i] *= trans ;
}

if (!Sing_FFT (p_array, w_array, used_numpts, FALSE)) // reverse
   return FALSE ;

for (i=0; i<used_numpts; i++)
{
   w_array[i] = p_array[i] ;
   p_array[i] = temp[i] ;
}

   // if there was a reduction in points (i.e. used_numpts < numpts)
   // fill any remaining points in the unfiltered array from temp array
   // and extend out the last waviness point
   for (; i<numpts; i++)
   {
      w_array[i] = w_array[used_numpts-1] ;
      p_array[i] = temp[i] ;
   }

   farfree (temp) ;
   *first_filtered = (cutoff_mm*1000.0)/spacing_um ;
   *last_filtered = numpts - *first_filtered ;

   return TRUE ;

} // end FFT_Gauss
```

Listing D.03 - Triangular Approximation to Gaussian

```
#include <stdio.h>
#include <malloc.h>

#include "gaussalg.h"

void Square_Convolution (double huge *unfiltered, double huge *waviness,
                         unsigned long side_ordinates, unsigned long numpts,
                         unsigned long *first_filtered_point,
                         unsigned long *last_filtered_point) ;

/*************************************/
int Triangle_Gauss (double huge *p_array, double huge *w_array,
                    double spacing_um, double cutoff_mm, unsigned long numpts,
                    unsigned long *first_filtered, unsigned long *last_filtered)
{
   unsigned long side_ordinates,
                 fp_first_pt, fp_last_pt,
                 sp_first_pt, sp_last_pt ;
   double huge *first_pass ;
   double width ;

   if ((first_pass=(double huge *)farcalloc((unsigned long)numpts, sizeof(double))) == NULL)
      return FALSE ;

   width = 0.44294647*cutoff_mm/2.0 ;  // half width for side ordinates
   width = width*1000.0/spacing_um ;  // half width in ordinates
   side_ordinates = ROUND(width) ;

   Square_Convolution (p_array, first_pass, side_ordinates, numpts,
                       &fp_first_pt, &fp_last_pt) ;

if ((fp_last_pt-fp_first_pt+1) < (4*side_ordinates))
{
   printf ("M1 Filter : Cutoff too long.") ;
   farfree (first_pass) ;
   return FALSE ;
}

Square_Convolution (&first_pass[fp_first_pt], &w_array[fp_first_pt],
                    side_ordinates, (fp_last_pt-fp_first_pt+1),
                    &sp_first_pt, &sp_last_pt) ;

if ((sp_last_pt-sp_first_pt+1) < 10)
{
   printf ("M1 Filter : Cutoff too long.") ;
   farfree (first_pass) ;
   return FALSE ;
}

*first_filtered = fp_first_pt + sp_first_pt ;  // computed
*last_filtered = numpts - (*first_filtered) ;

farfree (first_pass) ;
   return TRUE ;
}  /* end Triangle_Gauss */
```

```
/************************************************/
void Square_Convolution (double huge *unfiltered, double huge *waviness,
                         unsigned long side_ordinates, unsigned long numpts,
                         unsigned long *first_filtered_point,
                         unsigned long *last_filtered_point)
{
   register unsigned long i ;
   unsigned long width ;

   width = 2*side_ordinates + 1 ;
   *first_filtered_point = side_ordinates ;
   *last_filtered_point = numpts - 1 - side_ordinates ;

   if (*last_filtered_point <= *first_filtered_point)
      return ;
   waviness[*first_filtered_point] = unfiltered[0] ;
   for (i=1; i<width; i++)
      waviness[*first_filtered_point] += unfiltered[i] ;
   for (i=*first_filtered_point+1; i<=*last_filtered_point; i++)
      waviness[i] = waviness[i-1] - unfiltered[i-side_ordinates-1] +
                    unfiltered[i+side_ordinates] ;
   side_ordinates = 2*side_ordinates + 1 ;
   for (i=*first_filtered_point; i<=*last_filtered_point; i++)
      waviness[i] /= side_ordinates ;

   return ;

} /* end Square_Convolution */
```

Listing D.04 - Singleton FFT Subroutines

```
/*****************************************************************************
   Javier Soley, Ph. D,   FJSOLEY @UCRVM2.BITNET
   Escuela de Fisica y Centro de Investigaciones Geofisicas
   Universidad de Costa Rica
*****************************************************************************/
/*   Computes the DISCRETE FOURIER TRANSFORM of very long data series.
 *   Restriction: the data has to fit in conventional memory.
 *
 *   Compile in compact or large models ===> Pointers must be FAR
 *   Two huge pointers are used to access the real and imaginary
 *   parts of the transform without 64k wrap around.
 *
 *   This functions are translations from the fortran program in
 *
 *   R. C. Singleton, An algorithm for computing the mixed radix fast
 *   Fourier transform
 *
 *   IEEE Trans. Audio Electroacoust., vol. AU-17, pp. 93-10, June 1969.
 *   Some features are:
 *
 *      1-) Accepts an order of transform that can be factored not only
 *          in prime factors such 2 and 4, but also including odd factors
 *          as 3, 5, 7, 11, etc.
 *      2-) Generates sines and cosines recursively and includes
 *          corrections for truncation errors.
 *      3-) The original subroutine accepts multivariate data. This
 *         translation does not implement that option (because I
 *         do not needed right now).
 *
 *   Singleton wrote his subroutine in Fortran and in such a way that it
 *   could be ported allmost directly to assembly language. I transcribed
 *   it to C with little effort to make it structured. So I apologize to
 *   all those C purists out there!!!!!!!!
 *
 */
/*  Version 2.0 March/30/92 */

#include <stdlib.h>
#include <stdio.h>
#include <math.h>


#define TWO_PI   ((double)2.0 * M_PI)
#define MAXF  23
#define MAXP  209
#define TRUE 1
#define FALSE 0

long nn, m, flag,
     jf, jc,
     kspan, ks, kt,
     nt, kk, i;
double c72, s72, s120,
       cd, sd, rad, radf,
       at[23], bt[23];
long nfac[23];
int inc;
long np[MAXP];

void Fac_Des (long n) ;
void RadTx_2 (double huge *a, double huge *b) ;
void Radix_3 (double huge *a, double huge *b) ;
void Radix_4 (int isn, double huge *a, double huge *b) ;
void Fac_Imp (double huge *a, double huge *b) ;
int Permute(long ntot, long n, double huge *a, double huge *b) ;
```

```
/*********************************************************/
int Sing_FFT (double huge *a, double huge *b, long n, int forward)
{
   int test, isn, ret_val ;
   long ntot, nspan ;

   ntot = nspan = n ;

   if (forward)
      isn = -1 ;
   else
      isn = 1 ;
   if ( n < 2 )
      return FALSE ;

   inc = isn ;
   rad = TWO_PI ;
   s72 = rad / 5.0 ;
   c72 = cos(s72) ;
   s72 = sin(s72) ;
   s120 = sqrt(0.75);

   if (isn < 0)
   {
      s72 = -s72 ;
      s120 = -s120 ;
      rad = -rad ;
      inc = -inc ;
   }

   nt = inc*ntot ;
   ks = inc*nspan ;
   kspan = ks ;
   nn = nt - inc ;
   jc = ks / n ;
   radf = rad*jc*0.5 ;
   i = 0 ;
   jf = 0 ;
   flag = 0 ;

   Fac_Des (n) ;

   test = 0 ;
   while ((nfac[test] != 0) && (test < MAXF))
   {
      if (nfac[test] > 23)
         return FALSE ;
      test++ ;
   }

   do
   {
      sd = radf/kspan ;
      cd = 2.0*sin(sd)*sin(sd) ;
      sd = sin(sd+sd) ;
      kk = 1 ;
      i = i + 1 ;
      if (nfac[(int)i-1]==2)
         Radix_2 (a, b) ;
      if (nfac[(int)i-1]==4)
         Radix_4 (isn, a, b) ;
      if ((nfac[(int)i-1]!=2) && (nfac[(int)i-1]!=4))
         Fac_Imp (a, b) ;
   } while (flag != 1) ;

   ret_val = Permute (ntot, n, a, b) ;

   return (ret_val) ;

} // end Sing_FFT
```

```
/*********************************************************/
void Fac_Des (long n)   // find Prime Factors
{
   long k, j, jj ;

   k  = n;
   m = 0 ;
   while ((k - (k/16)*16) == 0)
   {
      m++ ;
      nfac[(int)m-1] = 4 ;
      k /= 16 ;
   }

   j = 3 ;
   jj = 9 ;
   do
   {
      while ((k % jj) == 0)
      {
         m++ ;
         nfac[(int)m-1] = j ;
         k /= jj ;
      }
      j += 2 ;
      jj = j*j ;
   } while (jj <= k) ;

   if (k <= 4)
   {
      kt = m ;
      nfac[(int)m] = k ;
      if (k != 1)
         m++ ;
   }
   else
   {
      if ((k-(k/4)*4) == 0)
      {
         m++ ;
         nfac[(int)m-1] = 2 ;
         k /= 4 ;
      }
      kt = m ;
      j = 2 ;
      do
      {
         if ((k%j) == 0 )
         {
            m++ ;
            nfac[(int)m-1] = j ;
            k /= j ;
         }
         j = ((j+1)/2)*2 + 1 ;
      } while (j <= k) ;
   }

   if (kt != 0)
   {
      j = kt ;
      do
      {
         m++ ;
         nfac[(int)m-1] = nfac[(int)j-1] ;
         j-- ;
      } while (j != 0) ;
   }

} // end Fac_Des
```

```
/*************************************************************/
void Radix_2 (double huge *a, double huge *b)
{
   long k1, k2 ;
     double ak, bk, c1, s1 ;

   kspan >>= 1 ;
   k1 = kspan + 2 ;
   do
   {
      do
      {
         k2 = kk + kspan ;
           ak = a[k2-1] ;
           bk = b[k2-1] ;
           a[k2-1] = a[kk-1] - ak ;
        b[k2-1] = b[kk-1] - bk ;
        a[kk-1] += ak ;
        b[kk-1] += bk ;
        kk = k2 + kspan ;
      } while (kk <= nn) ;

      kk = kk - nn;

   } while (kk <= jc) ;

   if (kk > kspan)
      flag = 1 ;
   else
   {
      do
      {
         c1 = 1.0 - cd ;
         s1 = sd ;
         do
         {
             do
            {
               do
               {
                  k2 = kk + kspan ;
                  ak = a[kk-1]- a[k2-1] ;
                  bk = b[kk-1]- b[k2-1] ;
                  a[kk-1] += a[k2-1] ;
                  b[kk-1] += b[k2-1] ;
                  a[k2-1] = c1*ak - s1*bk ;
                  b[k2-1] = s1*ak + c1*bk ;
                  kk = k2 + kspan ;
               } while (kk < nt) ;

                k2 = kk - nt ;
                c1 = -c1 ;
                kk = k1 - k2 ;
             } while (kk > k2) ;

                  ak = c1 - (cd*c1+sd*s1) ;
                  s1 = (sd*c1-cd*s1) + s1 ;

                   /***** Compensate for truncation errors   *****/
                  c1 = 0.5/(ak*ak+s1*s1) + 0.5 ;
                  s1 *= c1 ;
                  c1 *= ak ;
                  kk += jc ;
               } while (kk < k2) ;

               k1 = k1 + inc + inc ;
               kk = (k1-kspan) /2 + jc ;

            } while (kk <= (jc+jc)) ;
      }
   } // end Radix_2
```

```
/*******************************************************/
void Radix_3 (double huge *a, double huge *b)
{
   long k1, k2 ;
   double ak, bk, aj, bj ;

   do
   {
      do
      {
         k1 = kk + kspan ;
         k2 = k1+ kspan ;
         ak = a[kk-1] ;
         bk = b[kk-1] ;
         aj = a[k1-1] + a[k2-1] ;
         bj = b[k1-1] + b[k2-1] ;
         a[kk-1] = ak + aj ;
         b[kk-1] = bk + bj ;
         ak = -0.5*aj + ak ;
         bk = -0.5*bj + bk ;
         aj = (a[k1-1]-a[k2-1])*s120 ;
         bj = (b[k1-1]-b[k2-1])*s120 ;
         a[k1-1] = ak - bj ;
         b[k1-1] = bk + aj ;
         a[k2-1] = ak + bj ;
         b[k2-1] = bk - aj ;
         kk = k2 + kspan ;
      } while ( kk < nn) ;

      kk = kk - nn ;

   } while (kk <= kspan) ;

} // end Radix_3
```

```
/*************************************************************/
void Radix_4 (int isn, double huge *a, double huge *b)
{
   long   k1, k2, k3 ;
   double  akp, akm, ajm, ajp,
           bkm, bkp, bjm, bjp ;
   double  c1, s1, c2,
           s2, c3, s3 ;

   // initialize do avoid compiler warnings
   c2 = s2 = c3 = s3 = 0.0 ;
   kspan /= 4 ;

cuatro_1:

   c1 = 1.0 ;
   s1 = 0 ;
   do
   {
      do
      {
         do
         {
            k1 =  kk + kspan ;
            k2 =  k1 + kspan ;
            k3 = k2 + kspan ;
            akp = a[kk-1] + a[k2-1] ;
            akm = a[kk-1] - a[k2-1] ;
            ajp = a[ k1-1] + a[k3-1] ;
            ajm = a[ k1-1] - a[k3-1] ;
            a[kk-1] = akp + ajp ;
            ajp = akp - ajp ;
            bkp = b[kk-1] + b[k2-1] ;
            bkm = b[kk-1] - b[k2-1] ;
            bjp = b[k1-1] + b[k3-1] ;
            bjm = b[k1-1] - b[k3-1] ;
            b[kk-1] = bkp + bjp ;
            bjp = bkp - bjp ;
            if ( isn < 0)
               goto cuatro_5 ;
            akp = akm - bjm ;
            akm = akm + bjm ;
            bkp = bkm + ajm ;
            bkm = bkm - ajm ;
            if (s1 == 0.0)
               goto cuatro_6 ;

cuatro_3:

            a[ k1-1] = akp*c1 - bkp*s1 ;
            b[ k1-1] = akp*s1 + bkp*c1 ;
            a[ k2-1] = ajp*c2 - bjp*s2 ;
            b[ k2-1] = ajp*s2 + bjp*c2 ;
            a[ k3-1] = akm*c3 - bkm*s3 ;
            b[ k3-1] = akm*s3 + bkm*c3 ;
            kk = k3 + kspan ;
      }  while ( kk <= nt);
```

```
cuatro_4:

        c2 = c1 - (cd*c1 + sd*s1) ;
        s1 = (sd*c1 - cd*s1) + s1 ;

         /***** Compensate for truncation errors *****/
        c1 = 0.5 / (c2*c2 + s1*s1) + 0.5 ;
        s1 = c1*s1 ;
        c1 = c1*c2 ;
        c2 = c1*c1 - s1*s1 ;
        s2 = 2.0*c1*s1 ;
        c3 = c2*c1 - s2*s1 ;
           s3 = c2*s1 + s2*c1 ;
           kk = kk - nt + jc;
        } while (kk <= kspan) ;

        kk = kk - kspan + inc ;
        if ( kk <= jc)
           goto cuatro_1 ;

        if (kspan == jc)
           flag = 1 ;
        goto out;

     cuatro_5:

        akp = akm + bjm ;
        akm = akm - bjm ;
        bkp = bkm - ajm ;
        bkm = bkm + ajm ;
        if (s1 != 0.0)
           goto cuatro_3 ;

     cuatro_6:
        a[k1-1] = akp ;
        b[k1-1] = bkp ;
        b[k2-1] = bjp ;
        a[k2-1] = ajp ;
        a[k3-1] = akm ;
        b[k3-1] = bkm ;
        kk = k3 + kspan ;

     } while (kk <= nt) ;

     goto cuatro_4;

  out:

     s1 = s1 + 0.0 ;

  } // end Radix_4
```

```
/*******************************************************/
void Radix_5 (double huge *a, double huge *b)
{
   long k1, k2, k3, k4 ;
   double ak, aj, bk, bj,
          akp, akm, ajm, ajp, aa,
          bkp, bkm, bjm, bjp, bb;
   double c2, s2;

   c2 = c72*c72 - s72*s72 ;
   s2 = 2.0*c72*s72 ;

do
{
   do
   {
      k1 = kk + kspan ;
      k2 = k1 + kspan ;
      k3 = k2 + kspan ;
      k4 = k3 + kspan ;
      akp = a[k1-1] + a[k4-1] ;
      akm = a[k1-1] - a[k4-1] ;
      bkp = b[k1-1] + b[k4-1] ;
      bkm = b[k1-1] - b[k4-1] ;
      ajp = a[k2-1] + a[k3-1] ;
      ajm = a[k2-1] - a[k3-1] ;
      bjp = b[k2-1] + b[k3-1] ;
      bjm = b[k2-1] - b[k3-1] ;
      aa  = a[kk-1] ;
      bb  = b[kk-1] ;
      a[kk-1] = aa + akp + ajp ;
      b[kk-1] = bb + bkp + bjp ;
      ak = akp*c72 + ajp*c2 + aa ;
      bk = bkp*c72 + bjp*c2 + bb ;
      aj = akm*s72 + ajm*s2 ;
      bj = bkm*s72 + bjm*s2 ;
      a[k1-1] = ak - bj ;
      a[k4-1] = ak + bj ;
      b[k1-1] = bk + aj ;
      b[k4-1] = bk - aj ;
      ak = akp*c2 + ajp*c72 + aa ;
      bk = bkp*c2 + bjp*c72 + bb ;
      aj = akm*s2 - ajm*s72 ;
      bj = bkm*s2 - bjm*s72 ;
      a[k2-1] = ak - bj ;
      a[k3-1] = ak + bj ;
      b[k2-1] = bk + aj ;
      b[k3-1] = bk - aj ;
      kk = k4 + kspan ;
    } while ( kk < nn) ;

    kk -= nn ;

  } while ( kk <= kspan) ;

} // end Radix 5
```

```
/*******************************************************/
void Fac_Imp (double huge *a, double huge *b)
{
   long k, kspnn, j, k1, k2, jj ;
   double ak, bk, aa, bb, aj, bj ;
   double c1, s1, c2, s2 ;
   double ck[23], sk[23] ;

   k = nfac[(int)i-1] ;
   kspnn = kspan ;
   kspan /= k ;
   if (k==3)
      Radix_3 (a, b) ;
   if (k==5)
      Radix_5 (a, b) ;
   if ((k==3) || (k==5))
      goto twi ;
   if (k!=jf)
   {
      jf = k ;
      s1 = rad/(double)k ;
      c1 = cos(s1) ;
      s1 = sin(s1) ;
      ck[(int)jf-1] = 1.0 ;
      sk[(int)jf-1] = 0.0 ;
      j = 1 ;
      do
      {
         ck[(int)j-1] = ck[(int)k-1]*c1 + sk[(int)k-1]*s1 ;
         sk[(int)j-1] = ck[(int)k-1]*s1 - sk[(int)k-1]*c1 ;
         k-- ;
         ck[(int)k-1] =  ck[(int)j-1] ;
         sk[(int)k-1] = -sk[(int)j-1] ;
         j++ ;
      } while (j<k) ;
   }
   do
   {
      do
      {
         k1 = kk ;
         k2 = kk + kspnn ;
         aa = a[kk-1] ;
         bb = b[kk-1] ;
         ak = aa ;
         bk = bb ;
         j  = 1 ;
         k1 = k1 + kspan ;
         do
         {
            k2 -= kspan ;
            j++ ;
            at[(int)j-1] = a[k1-1] + a[k2-1] ;
            ak += at[(int)j-1] ;
            bt[(int)j-1] = b[k1-1] + b[k2-1] ;
            bk += bt[(int)j-1] ;
            j++ ;
            at[(int)j-1]  = a[k1-1] - a[k2-1] ;
            bt[(int)j-1] = b[k1-1] - b[k2-1] ;
            k1 += kspan ;
         } while (k1 < k2) ;

         a[kk-1] = ak ;
         b[kk-1] = bk ;
         k1 = kk ;
         k2 = kk + kspnn ;
         j = 1 ;
```

```
do
{
   k1 += kspan ;
   k2 -= kspan ;
   jj = j ;
   ak = aa ;
   bk = bb ;
   aj = 0.0 ;
   bj = 0.0 ;
   k = 1 ;
   do
   {
      k++ ;
      ak = at[(int)k-1]*ck[(int)jj-1] + ak ;
      bk = bt[(int)k-1]*ck[(int)jj-1] + bk ;
      k++ ;
      aj = at[(int)k-1]*sk[(int)jj-1] + aj ;
      bj = bt[(int)k-1]*sk[(int)jj-1] + bj ;
      jj += j ;
      if (jj>jf)
         jj -= jf ;
   } while (k<jf) ;

   k = jf - j ;
   a[k1-1] = ak - bj ;
   b[k1-1] = bk + aj ;
   a[k2-1] = ak + bj ;
   b[k2-1] = bk - aj ;
   j++ ;
} while (j<k) ;

kk += kspnn ;

} while (kk <= nn) ;
   kk -= nn ;
} while (kk <= kspan) ;
twi:
```

```
/***** Multiply by twiddle factors  *****/
if (i==m)
   flag = 1;
else
{
   kk = jc + 1 ;
   do
   {
      c2 = 1.0 - cd ;
      s1 = sd ;
      do
      {
c1 = c2 ;
s2 = s1;
kk += kspan ;
do
{
   do
   {
      ak = a[kk-1] ;
      a[kk-1] = c2*ak - s2*b[kk-1] ;
      b[kk-1] = s2*ak + c2*b[kk-1] ;
      kk += kspnn ;
   } while ( kk <= nt) ;
   ak = s1 * s2 ;
   s2 = s1*c2 + c1*s2 ;
   c2 = c1*c2 - ak ;
   kk = kk - nt + kspan ;
} while (kk <= kspnn) ;
c2 = c1 - (cd*c1 + sd*s1) ;
s1 = s1 + (sd*c1 - cd*s1) ;

/***** Compensate for truncation errors *****/
c1 = 0.5/(c2*c2 + s1*s1) + 0.5 ;
            s1 *= c1 ;
            c2 *= c1 ;
            kk  = kk - kspnn + jc ;
         } while (kk <= kspan) ;

         kk = kk - kspan + jc + inc ;

      } while (kk <= (jc+jc)) ;
   }

} // end Fac_Imp
```

```
#pragma warn -rch
/*************************************************************/
int Permute(long ntot, long n, double huge *a, double huge *b)
{
  /* Permute the results to normal order.
     MCM NOTE : This routine was based on FORTRAN translated to 'C'.
                Strict compiler checking will cause 2
                "Unreachable code" warnings
   */

     long  k, j, k1, k2, k3, kspnn, maxf ;
     double ak, bk ;
     long ii, jj ;

     ii = 0 ;  // initialize to avoid compiler warning

     maxf = MAXF ;
     np[0] = ks ;
     if (kt != 0)
     {
        k = kt + kt + 1 ;
        if (m < k)
           k-- ;
        j = 1 ;
        np[(int)k] = jc ;

        do
        {
           np[(int)j] = np[(int)j-1] / nfac[(int)j-1] ;
           np[(int)k-1] = np[(int)k] * nfac[(int)j-1] ;
           j++ ;
           k-- ;
        } while (j < k) ;

     k3 = np[(int)k] ;
     kspan = np[1] ;
     kk = jc + 1 ;
     k2 = kspan + 1 ;
     j = 1 ;

      /***** Permutation of one dimensional transform *****/
     if (n == ntot)
     {
        do
        {
           do
           {
              ak = a[(int)kk-1] ;
              a[(int)kk-1] = a[(int)k2-1] ;
              a[(int)k2-1] = ak ;

              bk = b[(int)kk-1] ;
              b[(int)kk-1] = b[(int)k2-1] ;
              b[(int)k2-1] = bk ;

              kk += inc ;
              k2 += kspan ;
           } while (k2 < ks) ;
ocho_30:
           do
           {
              k2 -= np[(int)j-1] ;
              j++ ;
              k2 += np[(int)j] ;
           } while (k2 > np[(int)j-1]) ;

           j = 1 ;
```

```
ocho_40:
            j = j + 0;
         } while (kk < k2) ;

         kk += inc;
         k2 += kspan;

         if (k2 < ks)
            goto ocho_40 ;
         if (kk < ks)
            goto ocho_30 ;
         jc = k3 ;
      }
      else
      {        /* Permutation for multiple transform  */
ocho_50:
         do
         {
            do
            {
               k = kk + jc ;
               do
               {
                  ak = a[(int)kk-1] ;
                  a[(int)kk-1] = a[(int)k2-1] ;
                  a[(int)k2-1] = ak ;
                  bk = b[(int)kk-1] ;
                  b[(int)kk-1] = b[(int)k2-1] ;
                  b[(int)k2-1] = bk ;
                  kk += inc ;
                  k2 += inc ;
               } while (kk < k) ;
               kk = kk + ks - jc ;
               k2 = k2 + ks - jc;
            } while (kk < nt) ;
            k2 = k2 - nt + kspan ;
            kk = kk - nt + jc ;
         } while (k2 < ks) ;

         do
         {
            do
            {
               k2 -= np[(int)j-1] ;
               j++ ;
               k2 += np[(int)j] ;
            } while (k2 > np[(int)j-1]) ;
            j =1 ;
            do
            {
               if ( kk < k2 )
                  goto ocho_50;
               kk +=jc ;
               k2 += kspan ;
            } while (k2 < ks) ;
         } while (kk < ks) ;
         jc = k3 ;
      }
   }
```

```
if ((2*kt + 1) < m)
{
   kspnn = np[(int)kt];
   /* Permutation of square-free factors of n */
   j = m - kt ;
   nfac[(int)j] = 1 ;
   do
   {
      nfac[(int)j-1] *= nfac[(int)j] ;
      j-- ;
   } while (j != kt) ;

      kt++ ;
      nn = nfac[(int)kt-1] - 1 ;
      if (nn > MAXP) // Product of square free factors exceeds allowed limit
         return FALSE ;

      jj = 0 ;
      j = 0 ;

      goto nueve_06;
nueve_02:
      jj -= k2 ;
      k2 = kk ;
      k++ ;
      kk = nfac[(int)k-1] ;
      do
      {
         jj += kk ;
         if ( jj >= k2 )
            goto nueve_02 ;

         np[(int)j-1] = jj ;
nueve_06:
         k2 = nfac[(int)kt-1] ;
         k = kt+1 ;
         kk = nfac[(int)k-1] ;
         j++ ;
      } while (j <= nn) ;

      /* determine the permutation cycles of length greater then 1 */
      j = 0 ;
      goto nueve_14 ;

      do    // MCM NOTE : unreachable from above
      {
         do
         {
            k = kk ;
            kk = np[(int)k-1] ;
            np[(int)k-1] = - kk ;
         } while (kk != j) ;
         k3 = kk ;
nueve_14:
         do
         {
            j++ ;
            kk = np[(int)j-1] ;
      } while (kk < 0) ;
   } while (kk != j) ;
```

```
np[(int)j-1] = -j ;
if (j != nn)
   goto nueve_14 ;

maxf *= inc ;
/* Reorder a and b following the permutation cycles */
goto nueve_50 ;

do                          // MCM NOTE : unreachable from above,
{                           //            therefore give compiler warning
   do
   {
      do
{
   j-- ;
} while (np[(int)j-1] < 0) ;

jj = jc ;
do
{
   kspan = jj ;
   if ( jj > maxf)
      kspan = maxf ;
   jj -= kspan ;
   k = np[(int)j-1] ;
   kk = jc*k + ii + jj ;
   k1 = kk + kspan ;
   k2 = 0 ;
   do
   {
      k2++ ;
      at[(int)k2-1] = a[(int)k1-1] ;
      bt[(int)k2-1] = b[(int)k1-1] ;
      k1 -= inc ;
   } while (k1 != kk) ;
   do
{
   k1 = kk + kspan ;
   k2 = k1 - jc*(k + np[(int)k-1]) ;
   k = -np[(int)k-1] ;
   do
   {
      a[(int)k1-1] = a[(int)k2-1] ;
      b[(int)k1-1] = b[(int)k2-1] ;
      k1 -= inc ;
      k2 -= inc ;
   } while (k1 != kk) ;
   kk = k2 ;
} while (k != j) ;
k1 = kk + kspan ;
k2 = 0 ;
do
{
   k2++ ;
                  a[(int)k1-1] = at[(int)k2-1] ;
                  b[(int)k1-1] = bt[(int)k2-1] ;
                  k1 -= inc ;
               } while (k1 != kk) ;
            } while ( jj != 0 ) ;
         } while (j !=1) ;
nueve_50:
         j = k3 + 1 ;
         nt -= kspnn ;
         ii = nt - inc +1 ;

      } while (nt >= 0) ;
      k = k + 0;
   }
   return TRUE ;

} // end Permute
```