CRANFIELD UNIVERSITY

Paul Morantz

A basis for the representation, manufacturing tool path generation and scanning measurement of smooth freeform surfaces

School of Aerospace, Transport and Manufacturing

PhD Thesis
Academic Year: 2016 - 2017

CRANFIELD UNIVERSITY

School of Aerospace, Transport and Manufacturing

PhD

Academic Year 2016 - 2017

Paul Morantz

A basis for the representation, manufacturing tool path generation and scanning measurement of smooth freeform surfaces

Supervisor: Paul Shore
February 2017

This thesis is submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy

ABSTRACT

Freeform surfaces find wide application, particularly in optics, from unique single-surface science programmes to mobile phone lenses manufactured in billions. This thesis presents research into the mathematical and algorithmic basis for the generation and measurement of smooth freeform surfaces. Two globally significant cases are reported: 1) research in this thesis created prototype segments for the world's largest telescope; 2) research in this thesis made surfaces underpinning the redefinition of one of the seven SI base units – the kelvin - and also what will be the newly (and permanently) defined value for the Boltzmann constant.

The research demonstrates two underlying philosophies of precision engineering, the critical roles of determinism and of precision measurement in precise manufacturing.

The thesis presents methods, and reports their implementation, for the manufacture of freeform surfaces through a comprehensive strategy for tool path generation using minimum axis-count ultra-precision machine tools. In the context of freeform surface machining, the advantages of deterministic motion performance of three-axis machines are brought to bear through a novel treatment of the mathematics of variable contact point geometry. This is applied to ultra-precision diamond turning and ultra-precision large optics grinding with the Cranfield Box machine. New techniques in freeform surface representation, tool path generation, freeform tool shape representation and error compensation are presented.

A comprehensive technique for very high spatial resolution CMM areal scanning of freeform surfaces is presented, with a new treatment of contact error removal, achieving interferometer-equivalent surface representation, with 1,000,000+ points and sub-200 nm rms noise without the use of any low-pass filtering.

ACKNOWLEDGEMENTS

My gratitude is due to numerous people in respect of support for this research:

to Paul Shore for his initiative and enduring commitment after launching research at Cranfield into freeform surfaces, to those within the joint research councils (GR/S85337) and in particular EPSRC (EP/F031416/1, EP/E023711/1 & EP/H003258/1) for funding it, and to Qioptiq, Surrey Satellites and Boostec for their support;

at Cranfield to several in management positions for the opportunity and encouragement to report my research in this way – Tom Stephenson & Phil John, David Stephenson, John Corbett and latterly in particular Raj Roy & Jose Endrino who have been instrumental in giving me the space and time to complete it;

in respect of work towards the common goal of freeform surface generation, a very long list of current and former colleagues among whom the following are deserving of special mention - Roger Read for his virtuoso mechanical design, Kevin Howard whose nose was so often pressed to the grindstone, Alan Heaume who so nearly endured death by a thousand (10 micron) cuts - to Andrew Baldwin, Roger Collins, Paul Comley, Andrew Eve, John Hedge, Adam Kerr, Xavier Tonnellier and many others;

also in this regard, collaborators - most emphatically the inspirational science evangelist Michael de Podesta MBE and his colleagues at NPL, to Ian Baird whose knowledge of Fanuc controllers and enthusiasm are both unbounded and to Eric Gould for his generosity with temperature control expertise;

and to those many others offering encouragement who could have made this the longest of all the lists, but here are just a few of them… to Pat McKeown OBE whose ceaseless support has been an inspiration not just to me, but to countless others engaged in the endeavours of precision engineering over several professional generations, to our other colleagues on the Loxham Precision board Chris Price OBE and Philip Aspinall, to Keith Carlisle, to Anne Fiorucci who buffered so many distractions for me, to my mother and to my father with whom I would so like to have shared the manuscript, to Crow for help with handheld pointing devices, to Raven for entertaining fluff and to Rook for the loan of books, but most of all – beyond all of the above – to my wife Manda without whom nothing could have been accomplished; I can never express how large my debt.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

ASPE	American Society for Precision Engineering
BCE	Before Common Era
BOX	Big OptiX
BT-UPS	Basic Technologies – Ultra Precision Surfaces
CAD	Computer-Aided Design
CAM	Computer-Aided Manufacture
CGH	Computer Generated Hologram
CIPM	Comité International des Poids et Mesures
CMM	Co-ordinate Measuring Machine
CNC	Computer Numerical Control
DOF	Degree Of Freedom
E-ELT	European Extremely Large Telescope
ELT	Extremely Large Telescope
EPSRC	Engineering and Physical Sciences Research Council
ESO	European Southern Observatory
EUV	Extreme Ultra-Violet
FIR	Far Infra-Red
FUV	Far Ultra-Violet
IBF	Ion Beam Figuring
IC	Integrated Circuit
IEEE	Institute of Electrical and Electronics Engineers
IKC UPS^2	Integrated Knowledge Centre in Ultra-Precision and Structured Surfaces
IMRC	Innovative Manufacturing Research Centre
IR	Infra-Red (radiation)
ISO	International Organization for Standardization
ITRS	International Technology Roadmap for Semiconductors
k_B	Boltzmann constant
LED	Light Emitting Diode

LHS	Left Hand Side
LOMS	Large Optics Manufacturing Study
LWIR	Long Wavelength Infra-Red
M1 (M2...)	First Mirror (Second Mirror...)
MUV	Mid Ultra-Violet
MWIR	Medium Wavelength IR
NA	Numerical Aperture
NaN	Not a Number
NC	Numerically Controlled
NIR	Near Infra-Red
NMI	National Measurement Institute
NPL	National Physical Laboratory
NURBS	Non-Uniform Rational B-Spline
NUV	Near Ultra-Violet
OptoNet	Competence Network for Optical Technologies
Ppm	Parts Per Million
p-v	Peak to valley
RAP	Reactive Atom Plasma
RBF	Radial Basis Function
RHS	Right Hand Side
RMS	Root Mean Squared
ROC	Radius Of Curvature
SI	Système International d'unités
SiC	Silicon Carbide
SPDT	Single Point Diamond Turning
SWIR	Short Wavelength Infra-Red
TCP	Tool Centre Point (or Tool Control Point)
UCL	University College London
UV	Ultra-Violet
T_{TPW}	Temperature of the Triple Point of Water

LIST OF PUBLICATIONS RESULTING FROM THIS RESEARCH

(Co)-authored publications that incorporate the research described in this manuscript.

Refereed journal articles

[1] Underwood R, de Podesta M, Sutton G, Stanger L, Rusby R, Harris P, **Morantz P**, Machin G, Estimates of the difference between thermodynamic temperature and the International Temperature Scale of 1990 in the range 118 K to 303 K. Phil. Trans. R. Soc. A. 2016 March; 374(2064).

[2] de Podesta M, Harris P, Underwood R, Sutton G, Mark D, Stuart F, **Morantz P**, Machin G, Response to Macnaughton's 'Comment on "A low-uncertainty measurement of the Boltzmann constant"'. Metrologia. 2016 January; 53(1): p. 116-122.

[3] de Podesta M, Underwood R, Sutton G, **Morantz P**, Harris P, Mark DF, et al. A low-uncertainty measurement of the Boltzmann constant. Metrologia. 2013 August; 50(4): p. 354-376.

[4] Underwood R, Davidson S, Perkin M, **Morantz P**, Sutton G, de Podesta M. Pyknometric volume measurement of a quasispherical resonator. Metrologia. 2012; 49(3): p. 245-256.

[5] Shore P, **Morantz P**. Ultra-precision: enabling our future. Philosophical Transactions of the Royal Society A - Mathematical Physical and Engineering Sciences. 2012; 370(1973): p. 3993-4014.

[6] de Podesta M, Sutton G, Underwood R, Davidson S, **Morantz P**. Assessment of uncertainty in the determination of the Boltzmann constant by an acoustic technique. International Journal of Thermophysics. 2011; 32(1-2): p. 413-426.

[7] Tonnellier X, Howard K, **Morantz P**, Shore P. Surface integrity of precision ground fused silica for high power laser applications. Procedia Engineering. 2011; 19: p. 357-362.

[8] Underwood R, Flack D, **Morantz P**, Sutton G, Shore P, de Podesta M. Dimensional characterization of a quasispherical resonator by microwave and coordinate measurement techniques. Metrologia. 2011; 48(1): p. 1-15.

[9] Comley P, **Morantz P**, Shore P, Tonnellier X. Grinding metre scale mirror segments for the E-ELT ground based telescope. CIRP Annals - Manufacturing Technology. 2011; 60(1): p. 379-382.

[10] Shore P, Cunningham C, DeBra D, Evans C, Hough J, Gilmozzi R, Kunzmann H, **Morantz P**, Tonnellier X. Precision engineering for astronomy and gravity science. CIRP Annals - Manufacturing Technology. 2010; 59(2): p. 694–716.

[11] McKeown P, Corbett J, Shore P, **Morantz P**. Ultra-precision machine tools - design principles and developments. Nanotechnology Perceptions. 2008 March; 4(1): p. 5-14.

[12] Tonnellier X, **Morantz P**, Shore P, Baldwin A, Evans R, Walker D. Subsurface damage in precision ground ULE® and Zerodur® surfaces. Opt. Express. 2007; 15(19): p. 12197-12205.

International conference keynote paper

[13] **Morantz P**. Multi-process Strategy for Freeform Optics Manufacture. In 3rd International Conference on Nano Manufacturing: nanoMan2012; 2012; Tokyo, Japan.

Conference papers with refereed extended abstracts

[14] Tonnellier X, **Morantz P**, Shore P, Comley P. Precision grinding for rapid fabrication of segments for extremely large telescopes using the Cranfield Box. In Proc. SPIE 7739, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation; 2010; San Diego, USA.

[15] Tonnellier X, Shore P, **Morantz P**, Baldwin A. Surface quality of sintered silicon carbide using an effective grinding process. In Wold of Photonics Congress: Manufacturing of Optical Components; 2009; Munich, Germany.

[16] Tonnellier X, Shore P, **Morantz P**, Orton D. Surface quality of a 1m Zerodur part using an effective grinding mode. In Proc. SPIE. 7102, Optical Fabrication, Testing, and Metrology III; 2008; Glasgow, UK.

[17] Tonnellier X, Shore P, **Morantz P**, Baldwin A, Walker D, Yu G, et al. Sub-surface damage issues for effective fabrication of large optics. In Proc. SPIE 7018, Advanced Optical and Mechanical Technologies in Telescopes and Instrumentation; 2008; Marseille, France.

[18] **Morantz P**, Luo X, Shore P. Characterisation of dynamic errors of an ultra precision machine tool. In Laser Metrology and Machine Performance VIII; 2007; Cardiff, UK. p. 4-13.

[19] Tonnellier X, Shore P, **Morantz P**, Baldwin A, Evans R, Walker D. Comparison of the subsurface damage induced when precision grinding ULE and Zerodur surfaces. In 8th International Conference on Laser Metrology, Machine Tool, CMM and Robotics Performance; 2007; Cardiff, UK.

[20] Tonnellier X, **Morantz P**, Shore P, Baldwin A, Walker D, Evans R. Subsurface damage caused during rapid grinding of Zerodur. In 10th International Symposium on Advances in Abrasive Technology (ISAAT 2007); 2007; Dearborn, USA.

[21] Tonnellier X, Shore P, Luo X, Baldwin A, **Morantz P**, Jin T, et al. Wheel wear investigations when precision grinding of optical materials using the BoX grinding mode. In 5th International Conference on High Speed Machining; 2006; Metz, France.

[22] **Morantz P**, Shore P, Luo X, Baird I. Control strategy of the Big Optix grinding machine. In 6th euspen International Conference; 2006; Baden bei Wien, Austria.

[23] Luo X, **Morantz P**, Shore P, Baird I. NURBS approximation method for tool path generation in a new free-form grinding machine. In Advances in Manufacturing Technology, 4th International Conference on Manufacturing Research (ICMR); 2006; Liverpool, UK.

[24] Tonnellier X, Shore P, **Morantz P**, Luo X, Baldwin A. Diamond resin bond wheel wear in precision grinding of optical materials. In Advances in Manufacturing Technology, 4th International Conference on Manufacturing Research (ICMR); 2006; Liverpool, UK.

[25] Shore P, **Morantz P**, Luo X, Tonnellier X, Read R, May-Miller R. Design philosophy of the ultra precision Big OptiX "BoX" Machine. In 7th International Conference on Laser Metrology, Machine Tool, CMM and Robotics Performance; 2005; Cranfield, UK.

[26] **Morantz P**, Shore P, Stephenson D, May-Miller R, Read R. From tetrahedral to box: design philosophy behind the Cranfield Big OptiX system. In 5th euspen International Conference; 2005; Montpellier, France.

[27] Shore P, **Morantz P**, Luo X, Tonnellier X, Collins R, Roberts A, et al. Big OptiX ultra precision grinding/measuring system. In Proc. SPIE 5965, Optical Fabrication, Testing, and Metrology II; 2005; Jena, Germany.

[28] Shore P, Luo X, Tonnellier X, **Morantz P**, Stephenson D, Collins R, et al. Grinding mode of the BoX ultra precision free-form grinder. In Proc. 20th ASPE Annual Meeting; 2005; Norfolk, USA.

Other Conference papers

[29] **Morantz P**, Comley P, Tonnellier X, Shore P. Precision free-form grinding of metre-scale optics. In SPIE Optifab; 2011; Rochester, USA.

[30] **Morantz P**, Mitchell J. E-ELT Mirror Segment Fabrication. In Manufacturing Technologies to Support Large Science Projects; 2010; Paris, France.

[31] **Morantz P**. Off-Aixs Grinding Machine (OAGM) // Box Ultra Precision Freeform Grinding and Measuring Machine. In Optonet Workshop, Ultra Precision Manufacturing of Freeforms and Microstructures; 2008; Jena, Germany.

[32] **Morantz P**, Gould E, Shore P. Thermal control of a high energy density ultra-precision machine. In Proc. euspen Topical Meeting: Thermal effects in precision systems; 2007; Maastricht, Netherlands.

GLOSSARY OF TERMINOLOGY

TERM		introduced on: page
accuracy	the closeness of a measured or controlled value to the true or correct value	1
additive manufacturing	any of a number of processes of manufacturing involving accretion or adding material to an already existing substrate – cf. *subtractive manufacturing*	102
affine	preserving parallel relationships – an affine transformation, such as scaling, translation, rotation, linear shearing preserves parallelism, ratios of distances between collinear points, and therefore preserves planes and their normal tolerance scale relationships although not necessarily angles	130
aliasing	a sampling artefact where (in particular) frequencies and also amplitudes of essentially periodic signals may be falsely represented where sampling frequencies are too low compared with the signals' periods	53
apodization	a windowing or filtering technique where multiple adjacent data points are combined to change the value of one, a function which is then successively applied to each data point	55
approximation	a process of finding a dependent from an independent variable given a data set where this relationship is known at discrete values, specifically the approximating function may not have the same value as the existing points – cf. *interpolation*	39
asphere	literally not a sphere – taken to mean a surface of revolution of a curve, usually of a curve with unipolar curvature (curving just one way)	23
Avogadro constant	the number $\sim 6.02\times10^{23}$ mol^{-1} of constituent particles in one mole of substance – its quantity defines the mole	18

TERM		introduced on: page
Bézier spline	a parametric curve defined by control points which (excepting first and last) do not lie on the curve	9
binary search successive approximation	an approximation technique in which a result is reached by splitting available options into two (roughly) equal parts, selecting (by a test) the part containing the solution, and thus repeating binary division until only one potential solution remains	128
Boltzmann constant	a fundamental physical constant $\sim 1.38\times10^{-23}$ $m^2 \cdot kg \cdot s^{-2} \cdot K^{-1}$ relating particulate energy with temperature	2
canonical form	a standard form of representation – (often) specifically the simplest, with lowest complexity of arrangement	24
closed solution	an exact solution which can be reached in a finite number of mathematical operations	162
conicoid	a surface of rotation of a conic section, broadly a quadric - specifically a second-degree surface of three variables such as a hyperboloid, ellipsoid, paraboloid or sphere	24
continuous	a *pupil* that has no breaks or gaps between its extents, which might otherwise affect its mathematic treatment	33
convex hull	the smallest convex locus of points in X-Y space that encloses all the defined points for the interpolating function	72
degree of freedom	any of a minimum set of directions in which independent motion can occur – often an orthogonal set	95
departure	the amount or degree to which a surface differs from a common or simpler shape, which is usually a conicoid that can be made using conventional (non-freeform) manufacturing techniques	21

TERM		introduced on: page
end effector	(normally in robots) an active element that performs a task, which is placed at the functional point of a machine – it can be exchanged, in concept or actuality, for a different end effector performing a different task	43
explicit	a function whose result variable may be computed from independent variables: e.g. $y = f(x) = x^2$ is explicit for y	23
freeform	a surface without global invariance in any axis of translation or rotation	8
full-aperture	it describes a process or procedure which operates over the whole width (or whole area) of a component simultaneously	8
implicit	a function whose relation to the (result) variable is given by an equation for which the function has not been solved explicitly e.g. $f(x, y) = x^2 + 3xy + y^2 = 1$ is implicit for y	23
interpolation	a process of finding a dependent from an independent variable given a spanning data set where this relationship is known at discrete values, specifically the interpolating function has the same value as the existing points – cf. *approximation, extrapolation*	37
material removal function	a functional description (usually 1 or 2 dimensional) of material depth removal rate as a function of distance from centre of tool-workpiece interaction	42
measurand	the object (or quantity) being measured	100
mole	see Avogadro constant	2
multivariate	a function of more than one variable	24
NaN	an IEEE Standard numeric type representing an undefined value	78
neutral removal	an even removal of material across a surface, to improve surface finish without changing surface form	51

TERM		introduced on: page
numerical aperture	a dimensionless number (incorporating refractive index) which characterises the angular range of light entering or leaving an optical system $\mathrm{NA} = n \sin\theta$	15
NURBS	a highly flexible and arbitrarily precise mathematical model used to represent curves and surfaces – a generalisation of a Bézier spline	35
Orthogonal polynomials	classes of polynomial $p_m(x)$ defined over a range $[a,b]$ obeying orthogonality criterion $\int_a^b w(x)p_m(x)p_n(x)dx = \delta_{mn}c_n$ where δ_{mn} is the Kronecker delta c.f. (2-15) – polynomials which, because they can be independently adjusted and summed to fit, are particularly useful in representing solutions to certain mathematical problems	31
piecewise	a function defined itself by multiple sub-functions, each applying to a certain interval or region of the piecewise function's domain – the intervals generally overlap	28
precision	The closeness of a measured or controlled value to one or more other measurements of the same quantity	15
principal axes transformation	a transformation in Euclidean space that finds the perpendicular axes aligned to a hyperboloid, ellipsoid etc. simplifying the algebraic expression for the surface	24
process chain	a sequence of process steps each designed to achieve suitable input quality for the immediately subsequent one	41
pupil	an image of the smallest effective aperture in the optical system at either the entrance or exit of the optical system as appropriate – it determines the effective aperture size	31
radial basis function	a constituent of a series sum of weighted functions of radial distance from a series of	38

TERM		introduced on: page
	different centres – the sum is used to approximate a surface	
Rayleigh criterion	the generally accepted criterion for resolving image detail, based on the ratio of the wavelength of light to the governing image formation pupil	14
reproducible	ability of an experiment or measurement to be reproduced, either by the experimenter or by someone else working independently	17
roll-off	edge roll-off – the unintended rounding of edges due primarily to increase in machining pressure as the contact area decreases at an edge	109
sag	sagitta – the depth of a curve, or depth of material to be removed from a flat to make a curved surface	23
sculptured surface	a freeform surface, usually in the context of a surface defined by surface patches described and blended together using polynomial expressions	8
singular value decomposition	a factorisation of a matrix relating to its decomposition using eigenvectors with wide applications in numerical science – pertinently in linear algebra and the solution of linear equations	86
smooth	having an even surface, free from sharp projections	1
specular surface	a term derived from speculum metal meaning a reflective surface obeying the law of reflection in Euclidean geometric optics, where the incident and (single) reflected ray lie in a plane perpendicular to the reflecting surface, each with the same angle to the surface – in common parlance one might say a specular surface is 'shiny'	39

TERM		introduced on: page
spline	a curve or curved surface defined by control points and equations either interpolating between the points or more often defining a figure influenced by the points	34
stacked axes	a combination of two or more motion assemblies, where the nominally static part of one is attached to the moving part of another, so that the entire mass or inertia of the first is moved by the second	190
sub-aperture	a process or procedure which does not operate over the whole width (or whole area) of a workpiece but is directed to different smaller portions at different times so as eventually to process the whole	8
subtractive manufacturing	any of a number of material removal processes – typically these are conventional ‘machining’ cf. *additive manufacturing*	102
tool centre point	a nominally central point on a tool which is used as the position-controlled location in generating a tool path; the tool’s cutting edge or surface is defined relative to this point	113
traceability	an unbroken chain of comparisons relating a measurement to the definition of its unit	17

1 INTRODUCTION TO FREEFORM SURFACE GENERATION

1.1 Background to the research

The term *free-form* has been used since the mid-20th century to attribute an irregular shape or structure to objects or even literary and musical compositions [33]. In the context of this work however *freeform* implies *smooth* continuous surfaces manufactured without rotational symmetry (strictly, without rotational invariance) about any physical axis of rotation used in their manufacture. This loose definition of freeform surfaces (*freeforms*) therefore relates as much to the manner and apparatus of manufacture as to the final characteristics of the surface. Further consideration is given to a definition in section 2.1.1 on page 7. Essentially, the lack of rotational invariance in freeforms renders impracticable the simple application of traditional types of machining. In general, freeforms are therefore more difficult to make accurately than rotationally invariant surfaces.

Despite their higher cost, freeform manufactured surfaces are used where their functional or aesthetic surface properties are demanded. Advances in both design and manufacturing techniques have rendered the use of freeforms increasingly common, particularly in optical systems [34]. Here, a combination of the availability of advanced and dynamic CNC-controlled machines, new mathematical surface representations and innovative 3D design technologies is bringing rapid change to high performance optical systems [35]. The increasing adoption of freeforms moreover, has made many new product types possible – e.g. miniature freeform projection optics allowing multiple LEDs in any configuration to produce the shaped illumination required for vehicle headlights' low-beam [36].

The work described in this thesis has been central to a key research goal with diverse applications – the production of freeform surfaces of the highest surface *accuracy*. Several ultra-precision technologies are required to achieve this. These include machine design, machining process design, motion control and surface metrology. Three things are inextricably linked with all these ultra-precision technologies: a) the mathematical representation of freeform surfaces, b) the generation of complex toolpaths and c) the acquisition and mathematical treatment of freeform measurement data. Toolpaths are the co-ordinated numerical motion descriptions needed to instruct the machining and the measurement of freeforms. The achievement of (a) and in particular (b) and (c) form the body of this work, which describes the creation of a comprehensive system of mathematical treatment of freeform surface generation/measurement and how it has been applied for two very important cases.

1.2 Research case studies

1.2.1 Acoustic thermometry

The two cases are intrinsic to major scientific objectives of global significance. The first of these objectives is the revision of the International System of Units (SI) in which all of the seven base units (metre, kilogram, second, ampere, kelvin, candela, mole) will be defined by ***fixing*** precisely the values of fundamental constants of nature, which are currently expressed approximately with quantified uncertainties [37,38]. The case study related to this objective is the lowest uncertainty measurement ever achieved of the Boltzmann constant, k_B, in terms of which the SI unit of thermodynamic temperature, the kelvin, can be redefined. This requires the manufacture and measurement of an ellipsoidal resonator cavity with exceptional accuracy.

In 2007 Michael de Podesta, from the National Physical Laboratory (NPL), the lead scientist for the ‘Boltzmann project’, approached the author to design the apparatus and full process to make the cavity and then to deliver it at the highest possible accuracy [39]. Part of the reason for the approach was the acknowledged expertise at Cranfield in single point diamond turning (SPDT) [40] and specific research in freeform surface generation through SPDT [18] for infrared (IR) optics for the company Thales Optics, supported through the UK Engineering and Physical Sciences Research Council (EPSRC) funded Cranfield Innovative Manufacturing Research Centre (IMRC).

The ellipsoid specified for the Boltzmann cavity, with around 124 mm internal dimension, is tri-axial with fractional eccentricities of 0.0005 and 0.001, and is therefore without rotational invariance and thus freeform. This has been made using single point diamond turning (SPDT) with an ultra-precision lathe using the novel mathematical techniques explained herein.

1.2.2 Extremely Large Telescope optics

The second major scientific objective is the creation of what will be the world's largest optical telescope. The case study related to this objective is the grinding of mirror segments for the primary mirror of the European Extremely Large Telescope (E-ELT) – which will be the first ELT ever made, as determined by the size of its primary mirror. In 2003, UK-based collaborators from Cranfield University, Royal Observatory Edinburgh, Cranfield Precision, NPL, Thales Optics and University College London (UCL) conducted a Large Optics Manufacturing Study (LOMS) [41] for Thales Optics, Optic Technium, and the UK Government's Department of Trade and Industry. This study identified existing manufacturing chains and proposed new ones oriented towards satisfying an evident and growing demand for ultra-precise large optics. As identified in the 2003 LOMS report, current (2017) demand for large optics comes both through requirements of industry, national and international science programmes. Industrial requirements for large optics are driven by i) extreme ultra-violet (EUV) microelectronics lithography systems and ii) earth observation satellites; science programmes include iii) large-scale astronomy projects and iv) nuclear fusion energy generation.

Instigated by the success of LOMS, in 2004 new research was initiated at Cranfield University and UCL into ultra-precise large optics fabrication. This research, in which the author was primarily engaged, was funded through a Joint Research Councils' Basic Technologies project entitled "*Ultra-Precision Surfaces; a New Paradigm*" (BT-UPS) and entailed the design and build of an ultra-precision large optics grinding machine – Big OptiX (BOX) at Cranfield. This machine employs the novel mathematical techniques explained herein. Following the success of this programme, in 2006 EPSRC funded the creation of an Integrated Knowledge Centre in Ultra Precision and Structured Surfaces (IKC-UPS^2), a research project in which the author was an Investigator, with facilities based in the Universities Cranfield, UCL & Cambridge and at Optic Technium.

One of the successes of the IKC-UPS^2 was the winning in 2008 of an order for Optic Technium from the European Southern Observatory (ESO) for the manufacture of several prototype mirror segments for E-ELT. These were ground using the prototype BOX machine, initially using funding from the IKC-UPS^2 research project. In the E-ELT design finally approved for construction [42], each of the required 931 hexagonal mirror segments for the primary mirror (M1) is about 1.45 metres in size (corner to corner). 798 of these are mounted at any one time. When combined, the segments form one rotationally symmetric ellipsoidal concave surface, over 39 metres in diameter with an 11-metre central obstruction. Each segment however is more than 5 metres from the optic axis of symmetry of the overall M1 mirror, so all segments are freeform surfaces when manufactured individually.

1.3 Research hypothesis

The hypothesis of this research is that the following is possible. The research question can be posed in two parts. Can ultra-high accuracy smooth freeform surfaces, such as those required by the application case studies, be manufactured by minimum axis-count machine tools, such as the Cranfield BoX 3-axis grinding machine or the Moore UPL 350-3 axis diamond turning machine, even though those machines, as configured, do not possess the capability to change the orientation of their tools? Can an industrial standard large CMM produce the high spatial resolution data-intensive surface maps required to perform error compensation and validation of the case study freeform machining with a measurement uncertainty performance beyond its advertised capability and a data volume beyond its software capacity?

The aim and objectives of the research are outlined in chapter 3.

1.4 Thesis structure

In this thesis, the research and development of freeform generation processes are described for two globally significant scientific cases calling for machining and measurement performance beyond pre-existing practice. In chapter 2, firstly the scientific and commercial contexts are introduced through two cases in each; then the mathematical bases for previous work are explored. This focuses on satisfying the two scientific demand cases that will be the subject of chapters 7 and 8 and identifies the state of relevant knowledge. In chapter 3, gaps in knowledge are re-emphasised and a statement of aim and objectives given. The approach to the research is explained, including an explanation of the order and content of the principal technical chapters, 5 - 8 (which address the research objectives) in the context of the research methodology. In chapter 4, the experimental equipment used in the study is introduced. In chapter 5, extensive new metrology algorithms are introduced underpinning the entire study, with a summary discussion. In chapter 6 algorithms and techniques are introduced; with detailed information on toolpath compression. In chapter 7, freeform manufacturing case study 1 (for acoustic thermometry) is presented, re-examining the demands set out in chapter 2, adopting the equipment and procedure laid out in earlier chapters, presenting results with analysis and discussion. In chapter 8, manufacturing case study 2 (for ELT mirror segments) is presented, similarly re-examining the demands, equipment and procedure with results, analysis and discussion. In chapter 9, conclusions and summary of achievement are presented – reiterating the contribution to knowledge, impact and dissemination of research. In chapter 10, recommendations for further work and exploitation are collated. References are placed after chapter 10 and the manuscript concludes with an Appendix, containing specific elements of some coded algorithms.

Within the text, terms in the glossary (page xxiii) are introduced in italics. Some abbreviations and acronyms are listed on page xviii.

2 BASIS FOR RESEARCH

This chapter gives an historical background and a basis for the research through a review of relevant literature and a background to the required mathematical representations.

First, an appropriate definition of “freeform” is established.

Then the origin of commercial demand for freeforms is considered in the context of this research. This establishes how two principal and distinct manufacturing areas: infrared optics and lithography optics, whatever their commonalities, require differing manufacturing processes. For freeform infrared optics, diamond turning is the principal machining technology, whereas for lithography optics grinding is frequently a preparatory process only, through its delivery of near net shape spherical machining.

The demands set by the scientific objectives of the two scientific case studies of this thesis are then considered, and these map closely to the industrial cases introduced earlier.

A range of potential mathematical descriptions of freeforms is discussed along with tool path algorithms and their automated development.

Non-specular freeform surface areal measurement techniques are reviewed and finally there is an outline of the process of position-controlled machining (techniques of this thesis) of freeform surfaces.

The chapter concludes with a statement of the relevant gap in knowledge.

2.1 The nature of a freeform

The development of functional and aesthetic continuous smooth curved surfaces that are literally *manu-factured* (hand-made) [43] of course pre-dates recorded history. More recently, similar functional surfaces by contrast have been *manufactured* in a more modern sense - created with the aid of apparatus or machinery imparting relative motion between a tool and a workpiece. The detail of the early history of mechanised surface creation is certainly controversial, but it likely stretches back several thousand years. Enoch [44] claims the existence of manufactured mirrors 8,000 years old and Egyptian lenses of high quality 4,000+ years old, for instance in the famous “Seated Scribe” in the Louvre [45]. The equally famous Babylonian “Nimrud” lens in the British Museum, itself nearly

3,000 years old, is plano-convex with a focal length of 12 cm; whether this represents a functional specification or is merely decorative is debated [46] although it was likely fabricated on a lapidary's wheel [47]. These ancient optical surfaces - the earliest of which (if Enoch is correct) would be roughly contemporary with the development of wheels (spinning, potters' & transportation) [48] - were made with a technique giving rotational invariance of shape, as were the vast majority of optics manufactured until the 21st century [34] for reasons both of manufacturing ease and available design capability.

2.1.1 Freeform definition

Fang [49] defines a freeform as a surface "with no axis of rotational invariance (within or beyond the part)". Under this definition, the E-ELT telescope segments are not freeforms despite freeform techniques being required to make them. Jiang [50] introduces sub-classifications: 1 – freeforms with steps/facets, in other words with slope discontinuities; 2 – so called tessellated surfaces, with a repeated structure (that might also have the discontinuities of class 1); and 3 – smooth surfaces with a global geometry definition. Savio [51] classifies freeforms as complex geometrical features according to the definition in ISO 17450-1 [52] with no invariance under any of the invariance classes (prismatic, revolute, helical, cylindrical, planar, spherical); in other words a freeform surface will always appear to be changed in position or orientation by a translation or rotation. Under this definition, a tilted flat (being rotated about an axis not normal to the surface) would not be a freeform despite freeform techniques being required to make it. These definitions, which might be called purist definitions, relate to the definition of the surface geometry in a most general way, with no assumption about manufacturing technique.

For the author's purposes, a definition is required that reflects the requirements of making the surfaces, which has implications to the way they are described mathematically. This might be called a definition of freeform *machining*, rather than of the surface per se. Thompson [34] uses such a definition "A ... surface that leverages a third independent axis (C-axis in diamond turning terminology) during the creation process to create a ... surface with as-designed non-symmetric features." Garrard [53] explicitly identifies the issue of off-axis machining in the definition "Freeform ... surfaces are defined as any non-rotationally symmetric surface or a symmetric surface that is rotated about any axis that is not its axis of symmetry". This manuscript will adopt the following machining process-centric definition with some elements of all these but with generic application.

> *A freeform is a surface made using freeform machining which means the surface as mounted is without global invariance in any axis of translation or rotation possessed by the machine.*

2.2 Demand for freeforms from commercial objectives

The manufacturing science of freeform imaging-optical surfaces has developed somewhat independently from the equivalent manufacturing science for what have been called *sculptured* surfaces in the wider field of Numerically Controlled (NC) machining. "Sculptured" and "free-form" were used together by Coons [54] as early as 1967; numerous other authors have explicitly expressed the equivalence of the terms [51,55,56]. There are several reasons for this independent development. Principal among these, as intimated by Thompson [34], may be the commonly circular aperture of imaging optical systems. This has discouraged the adoption of X-Y polynomial surface descriptions common in other fields and is also linked to the production techniques conventionally applied for imaging optics, which have tended to be predominantly full-aperture and rotational; whereas the NC machining tradition (in optics parlance) is sub-aperture for a variety of material removal techniques.

The wider development of freeforms has been promoted by demands for aesthetic and functional surfaces; according to Choi [57] the functional applications include:

- aerodynamic: aerofoil (jet engine), impeller (compressor), marine propeller, etc.;
- optical: lamp reflector (automobile), shadow mask (TV-monitor), radar-dish, etc.;
- medical: parts for anatomical reproduction;
- structural: structural frames (aircraft), sporting goods, etc.;
- manufacturing surface: parting surface (moulding die), die face (stamping die), etc.

Early development for systematic characterisation of freeforms was driven through the 20th Century by aeronautical and automotive applications. The example of aerofoil development is instructive. Although flat sheet wings will function, Cayley, the inventor of the aeroplane who built the first successful manned heavier-than-air craft in 1853, had experimented with aerofoils in 1804 [58]. Phillips patented a series of curved aerofoil shapes in 1884 [59]. Systematic design and testing of aerofoils was carried out by the US National Advisory Committee for Aeronautics (NACA) from 1915. These designs used relatively high degree polynomial descriptions for 3D surfaces, although: i) design success was empirically achieved through trial and error, ii) designs could only be transferred to manufacture manually, through graphical processes. From 1925 [60] these designs could be tested without experiment, based on aerodynamic theory, to some extent avoiding (i). From 1933 [61] a good degree of determinism was possible in shaping design parameters based on performance goals; this was closer to today's design-CAD process. It would be some years before (ii) could also be avoided by adding CAM to the CAD or indeed before the make-measure 'loop' could be closed with freeform surface metrology.

Later, French automotive engineers at Citroën (De Casteljau) and Renault (Bézier) were responsible for the early development of Bézier splines which soon led to their generalisation into non-uniform rational B-splines (NURBS) [62] taking in the work of Schoenberg [63] and others. The motivation was to find an arbitrarily precise way to represent 2D curves and 3D surfaces mathematically, and these were integrated into the proprietary CAD packages of car companies as early as the 1960s [64] and gradually applied more widely for instance in marine and aeronautical engineering [65].

Separately, imaging optical requirements for freeforms developed more slowly through the 20th Century although in the 21st Century their benefits are now clear [66]. Progressive (varifocal) lenses give asymmetric lens performance. Aves' patent of 1907 [67] for a progressive lens uses non-coaxial conic sections front and back to achieve function, in other words without using freeforms. Kanolt proposed a freeform spectacle lens in a 1959 patent giving an explicit asymmetric polynomial description, which may be the first such specification in optics [68]. Other commercial production designs do have surfaces which due to their asymmetry, meet the freeform definitions given by other authors [49,50,51]. Until close to the end of the 20th Century however these were *all* manufactured with non-freeform machining techniques (the majority still are) using progressively reset tooling to adjust curvature across the surface in successive passes, followed by a final polish with a soft lap [69,70]. Newer designs *do* use freeform machining, including SPDT [71].

A 1937 patent for a spiral mirror [72] is in most definitions freeform, although its machining operation is akin to thread cutting. One of the first examples of a true freeform (as defined in section 2.1.1 on page 8) in an imaging-optical application was for Polaroid's SX-70 camera in 1973. The freeform moulds for these had high degree polynomial surface descriptions and were NC machined using mathematically corrected toolpaths. This may be the earliest use reported of NC freeform machining for an imaging-optical application [73]. Demands for application of freeform imaging optics have proliferated from the very end of the 20th Century, although the earliest meetings on the subject of freeform optics were probably the ASPE and OptoNet topical meetings in 2004 [74,75]. Two important current commercial application areas for freeform optics are outlined below – infrared and lithographic optics.

2.2.1 Commercial case 1: infrared optics

Infrared (IR) radiation spans wavelengths between the visible and microwaves, approximately 0.75 to 1000 microns. This range is conventionally split as shown in Table 2-1 (and Figure 2-1) with some minor variation [76,77], the divisions between bands arising from wavelengths absorbed by common media such as air or glass. The distinction between bands is also influenced by preferred wavelengths of the different developed sensor technologies. These effects together give rise to typical application sectors indicated below.

Table 2-1: IR spectrum

Near (NIR)	Short wave (SWIR)	Mid wave (MWIR)	Long wave (LWIR)	Far (FIR)
0.75 – 1.4 µm Fibre telecoms, night vision	1.4 – 3 µm Remote sensing, long-range comms.	3 – 8 µm Thermal image weaponry	8 – 15 µm Thermal imaging	15 - 1000 µm Astronomical spectroscopy

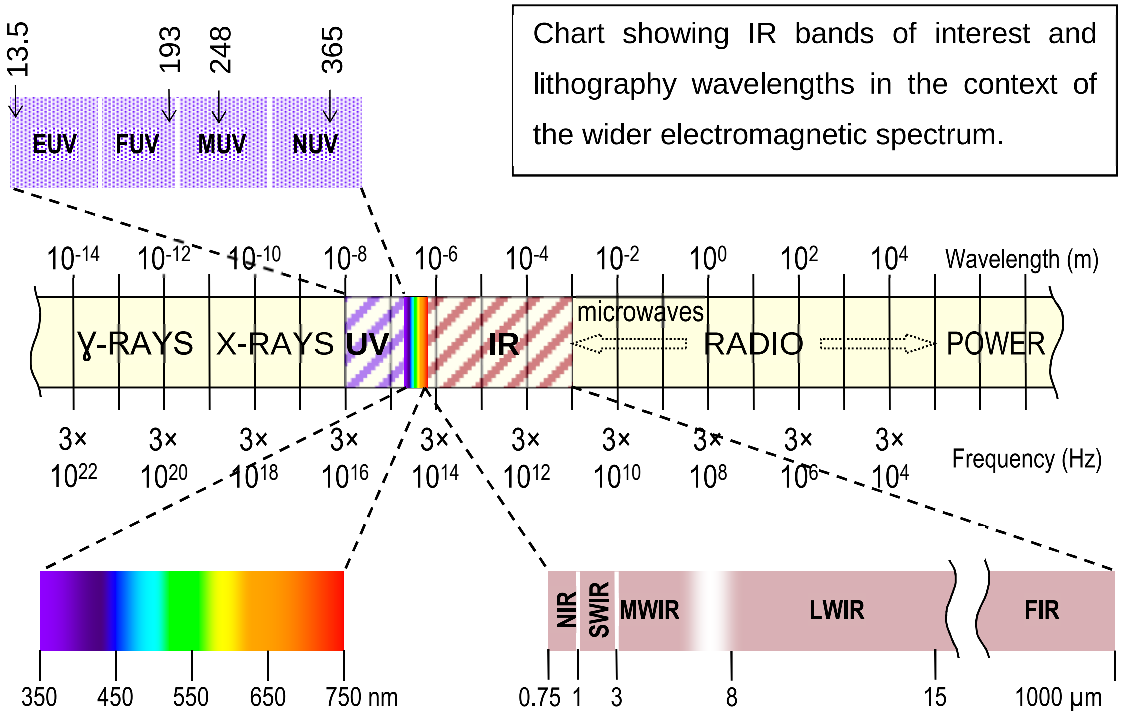


Chart showing IR bands of interest and lithography wavelengths in the context of the wider electromagnetic spectrum.

Figure 2-1: Electromagnetic spectrum

IR optics are used in transmission (lenses) and reflection (mirrors). For lenses, materials must be chosen for adequate transparency in the selected band. Refractive index can also be a factor in selection as can chromatic aberration, density, stiffness, hardness, thermal/mechanical shock resistance, chemical resistance, machinability and of course cost. These requirements give rise to a range of material choices including Calcium Fluoride (CaF_2), Germanium (Ge), Magnesium Fluoride (MgF_2), Sapphire (Al_2O_3), Silicon (Si), Zinc Selenide (ZnSe), and Zinc Sulphide (ZnS). Various materials (principally metals) are commonly used in reflection - particularly aluminium - with or without coatings.

Several factors have hastened the adoption first, of aspherics (also diffractive optics) and then more recently, of freeform optics, for IR applications.

1) **Design simplicity** - the surfaces of rotation of conic sections (which are aspheric) are widely used in IR applications because they allow elimination of spherical aberration for reflective as well as refractive surfaces [78].
2) **Material** - traditional optics production technology has centred on spherical surfaces in glass, both for mirrors and lenses, simply because glass is an easy material to work, with grinding/lapping/polishing, and

spherical (or flat) surfaces are the easiest surfaces to make accurately. This is because they can be produced by two spindle rotations without requiring NC motions to *generate*[*] the curvature. IR materials, particularly the ones used for MWIR & LWIR, are amenable to turning [78] whereas glass isn't, and turning aspheres with SPDT is relatively simple.

3) **Aberrations** - the detrimental effect of some surface errors in optics depends on the ratio of their size to the wavelength of the light; this can make the absolute surface form accuracy requirements for IR optics less demanding than for the shorter visible wavelengths. The ability to accommodate larger surface errors has accelerated the take-up of more advanced designs requiring aspheric (and freeform) surfaces for IR [79].
4) **Optical performance benefits** – for instance phase masks for wavefront coding to increase depth of field [80,81]. There are numerous other performance benefits to different freeform designs.
5) **Reduction in optical surface count** - more advanced designs using non-spherical optics can reduce the optical component count which improves the following [80,82]:
    a. transmission losses (less interposing lens medium)
    b. weight
    c. cost

A substantial proportion of the MWIR/LWIR work is military-related and classified, as indicated by Fuerschbach [83] although all-reflective designs which were un-obscured (not coaxial) were in use for aerial cameras as early as the 1960s as some unclassified reports have shown [84]. Advances of these designs, based initially on aspherics [85] and later, on freeform optics [86,87] have been reported, particularly for astronomical/spectroscopy applications. Optical surfaces of these types tend to be produced by SPDT [88]. For infrared optics, the surface finish of SPDT is nearly always considered adequate without subsequent superfinishing[†].

---

[*] Generating, in this context, implies making a shape by motion synchronised between two axes, angular or linear, as required for freeform surface creation - as opposed to spindle rotations which do not require synchronisation

[†] Superfinishing conventionally is a metal finishing process akin to short-stroke honing and produces directional surface texture – in this context, and for optical applications however, it's one of several surface modification techniques producing nearly isotropic surface texture of very low roughness over a wide range of spatial wavelengths

Errors in machining can be mitigated of course by error compensation, but a remaining contribution to error is tool decentring. There are treatments for tool decentring based on assessment of error profile, which are successful for spherical forms, but approximate for aspherics [89] and freeforms. In order to exploit the benefits of a high-resolution 3D measurement capability a technique for accurate first time setting from the entire aspheric or freeform surface would give the highest accuracy, but this is not reported in the literature.

2.2.2 Commercial case 2: lithography optics

Litho-graphy, at one time literally writing on stones [90], has evolved via a 19th-20th Century chemical printing process to modern photolithography. The application of this to microelectronics production has developed through enormous investment in technology and equipment to yield one of the most impressive achievements of modern industrial science. By this technique, the geometric pattern of a complex integrated circuit is photographically transferred from an optical mask or reticule to a coated substrate, often with a change in scale determined by the design of the optics. The coating and the substrate are subsequently treated chemically or recoated. Combinations of multiple such processes can create complex multilayer devices with feature sizes of a few nanometres. The process can be repeated many times on different adjacent areas of a substrate to make multiple copies of the device.

The continuous improvement in the lithography process has underpinned the semiconductor industry's continuous advancement, in pace with the famous Moore's Law. This 'law' is Gordon Moore's mid 1960s prediction [91] based on contemporary trends, that the integrated circuit (IC) transistor count-per-chip would double every 2 years. It has been predicted on numerous occasions throughout several decades that lithography will fail to maintain this rate of improvement [92]. So far, these predictions have always failed to come to pass. Schaller [93] has explained that the International Technology Roadmap for Semiconductors (ITRS), which is the industry-led and adopted plan for technology improvement, takes Moore's Law as a basic planning assumption; it may have become a self-fulfilling prophecy. The 2011 ITRS Roadmap and its 2012 update [94,95] take a view to 2025. In this view, Moore's Law is expected to hold, although very significant changes in the production technology are anticipated in this period.

This continuous improvement in device count is achieved through miniaturisation – the reduction in scale of device dimensions. The lower limit in a device's feature size created using an optical projection process is conventionally set by the effective imaging resolution of the system – and this is related to the diffraction limit and therefore the wavelength of light used. Several enhancements to simple pattern exposure lithography are useful at the shortest available wavelengths to enhance resolution, such as immersion lithography to improve the resolution optically by changing the refractive index of the exposure gap (this is used currently). There are various other techniques such as multiple patterning to utilise photoresist properties to 'defeat' the Rayleigh criterion by a factor of 2 or more (this is expensive to implement because of multiple exposure). These techniques can achieve some improvement, perhaps as much as one order of magnitude in combination for a given wavelength of light; device dimension however will always be functionally dependent on (essentially proportional to) wavelength. This fundamental issue has seen the industry move from the near ultraviolet (NUV) wavelength of 365 nm Mercury lamps in the 1980s through the KrF excimer laser's 248 nm in the mid ultraviolet (MUV) band to the current production standard of the ArF excimer laser's 193 nm wavelength. This wavelength is into the so-called vacuum-UV band – wavelengths at which air begins to absorb significant amounts of UV radiation. The vacuum-UV band is further split into far ultraviolet (FUV) and extreme ultraviolet (EUV). Figure 2-1 on page 11 places these wavelengths into the context of the wider electromagnetic spectrum.

A 'soft' end to *conventionally-optical* lithography's progress is anticipated around 2020 [92,95] although Moore's Law adherence is predicted (is planned) to continue thereafter using new technology. New candidate technologies to replace 193 nm UV lithography after 2020 include: Extreme Ultraviolet Lithography (EUV/EUVL), Nano-Imprint Lithography (NIL) [96], Multiple Electron Beam Direct Write (MEBDW) and Direct Self Assembly (DSA). Of these EUV is by far the front-runner, having very large investment and systems already being series-produced [97,98,99]. These systems operate at 13.5 nm wavelength in the extreme UV range (close to soft X-ray wavelengths). Strong absorption by air of this wavelength enforces use of high vacuum throughout the optical path. Use of lenses is unrealistic, because materials are unavailable for transmission of EUV;

optical systems for EUV must therefore be all-reflective. Absorption at the surface of each mirror is also highly significant at such short wavelengths. This motivates the design of all-reflective optical systems using an absolute minimum count of optical surfaces for both mask illuminators and projection optics. Projector designs are already in service for 13.5 nm wavelength using 6 aspheric mirrors, for feature dimensions down to 27 nm. More complex designs with 8 mirrors may be required to increase numerical aperture (NA) to 0.7 for feature dimensions below 10 nm [100]. It is clear that as for infrared optical systems there are significant drivers for the reduction of number of optical surfaces in these designs; moreover, advances can be realised by using freeform mirror surfaces of considerable size – in excess of 500 mm. For several reasons, these mirrors are likely to be made of glass or glass ceramic substrates but currently have not been manufactured.

2.3 Demand for freeforms from scientific objectives

In addition to demands from commercial objectives outlined in 2.2 on page 8, as discussed by Shore [10,5] demands from large scientific programmes have for a long time advanced production technology for the highest *precision* surfaces. Two important scientific application areas for freeform surfaces are outlined below – acoustic thermometry and ground-based ELTs. It will become apparent how these echo the two commercial cases: infrared and lithography optics, given in sections 2.2.1 and 2.2.2 above.

2.3.1 Scientific case 1: acoustic thermometry

Currently the kelvin, the unit of temperature, is defined as the fraction 1/273.16 of the thermodynamic temperature of the triple point of water (T_{TPW}) – in other words the difference in temperature between absolute zero and T_{TPW} is 273.16 kelvin. Given there is therefore only one reproducible point defining all temperature measurements, scaling the unit at higher and lower temperatures (that differ significantly from T_{TPW}) is susceptible to increased uncertainty. A technique called acoustic thermometry, involving the preparation of ultra-precise freeform surfaces by the author, has been used in support of the redefinition of the kelvin, and to promote more accurate temperature measurements in the future.

2.3.1.1 Revision of the SI base units

As part of a rationalisation of our entire system of measurement the “Comité international des poids et mesures” (CIPM) has proposed a new definition of the kelvin, in which it will take a value consistent with a defined value of the Boltzmann constant [101]. This will be formally adopted in 2018 [102]. In our current system of units, “Le Système International d'unités” (SI) the seven base units are fixed in size and all quantitative measurements, beyond simple counting, are based on these. These units have historically been tied largely to physical artefacts and prototypes that define the units. The fundamental constants, such as the Planck constant h and the Boltzmann constant k_B, describe the relationships between units through what are essentially *the laws of physics*. Under the old scheme these constants, whilst themselves fixed and immutable universal constants, can be determined only to a level of uncertainty and never known precisely. This means the values of so-called constants are then a) subject to change as better measurements are made, and b) dependent on the *scale* (size) of the SI units. The SI units have been related to historical artefacts such as the prototype kilogram, which are subject to *actual* variation [103] and so the implications of (a) and (b) are considered unsatisfactory.

Under the CIPM’s proposed revision, the seven units will be defined based on the value of fundamental constants (and the other base units). This is a philosophical change that has significant consequences. The relevant fundamental constants will be given fixed and precise values, based on the best available data relating them to the existing unit definitions, thus preserving the existing scale of the units. This will mean that seven fundamental constants (or attributes) will have fixed values; the situation for other fundamental constants will be as before.

The definitions of the seven base units will be in terms given in Table 2-2 below, although the numerical values of the constants will be subject to minor revision prior to the formal adoption of the proposals in 2018, but not thereafter. Note definitions for mole and second do not depend on other base units, so there is no circular dependence - the fundamental constants for the other base units have been chosen carefully for minimal dependence on other definitions and minimum uncertainty in measurement.

Table 2-2: Proposed SI base unit and constant definitions

Quantity	**Unit**		**Constant / attribute**		
	name	***Symb.***	***Name / description***	***symbol***	***value***
Amount of substance	mole	mol	Avogadro constant	N_A	6.0221415×10^{23} $\mathbf{mol}^{-1}$
Time	second	s	Ground state hyperfine splitting transition frequency of the caesium 133 atom	$\Delta v(^{133}\mathrm{Cs})_{\mathrm{hfs}}$	9,192,631,770 $\mathbf{s}^{-1}$
Length	metre	m	Speed of light in vacuo	c_0	299,792,458 $\mathbf{m}\cdot\mathbf{s}^{-1}$
Electric current	ampere	A	Elementary charge	e	$1.60217653 \times 10^{-19}$ $\mathbf{A}\cdot\mathbf{s}$
Mass	kilogram	kg	Planck constant	h	$6.62606957 \times 10^{-34}$ $\mathbf{m}^2\cdot\mathbf{kg}\cdot\mathbf{s}^{-1}$
Thermodynamic temperature	kelvin	K	Boltzmann constant	k_B	$1.380\,651\,56 \times 10^{-23}$ $\mathbf{m}^2\cdot\mathbf{kg}\cdot\mathbf{s}^{-2}\cdot\mathbf{K}^{-1}$
Luminous intensity	candela	cd	Spectral luminous efficacy of monochromatic radiation of frequency 540 × 10^{12} hertz	$K(\lambda_{555})$	683 $\mathbf{cd}\cdot\mathrm{sr}\cdot\mathbf{s}^3\cdot\mathbf{kg}^{-1}\cdot\mathrm{m}^{-2}$

Philosophically, this scheme means that any measurement can be independently referenced to experiments which are *reproducible* subject to experimental measurement errors, giving the same results anywhere, at any time, without requiring reference to a prototype standard. Practically, measurements will still be referenced to national and international standards through *traceability*. The difference after the redefinition takes place will be that each step in the traceability chain is now equivalent, including the first step. Questions such as “what happens if someone makes a more accurate measurement of the Planck constant, does the kilogram need to be redefined?” are not problematic, because this is simply the first step in the traceability chain performed (in concept at least) by a National Measurement Institute (NMI) such as NPL in the UK. The definition of the unit and its related constant will be unchanged; they will simply have improved the quality of their calibration. This is clearly better than having to change the value of a constant.

2.3.1.2 Measurement of the Boltzmann constant

In the context of preparation for the redefinition of the kelvin it is relevant to consider techniques for its measurement in terms of the existing base unit definitions; with reference to the foregoing (section 2.3.1.1 on page 16) this can also be considered a temperature measurement.

Among techniques that can be used to estimate k_B, according to Fellmuth [104] acoustic techniques using monatomic gases are most suitable. These are based on equation (2-1),

$$k_{\mathrm{B}} = \frac{u_0^2 M}{\gamma_0 T_{\mathrm{TPW}} N_{\mathrm{A}}} \qquad (2\text{-}1)$$

where u_0 is the speed of sound in the gas, M is its molar mass, γ_0 is a thermal characteristic of the gas, T_{TPW} as before is the triple point of water (the temperature at which the measurement will be made) and N_{A} is the Avogadro constant. N_{A} is known with very low uncertainty [105] and γ_0 for monatomic gases is exactly 5/3 - high purity Argon will come very close to this value (and the uncertainty can be estimated). An experiment conducted at the triple point of water therefore that measures M and u_0 can determine the $k_{\mathrm{B}}{\cdot}T_{\mathrm{TPW}}$ product. The reproducibility of T_{TPW} for this experiment is important, but is the same reproducibility upon which the current definition of the kelvin is founded; under this definition it is exactly 273.16. Therefore, the experiment allows an estimate of k_{B}.

In Lord Rayleigh's tour de force *The Theory of Sound* he published a detailed analysis of the acoustic resonances in a hard-shelled spherical cavity as early as 1878 [106]. Bancroft in 1955 [107] recommended a spherical cavity for undergraduate demonstrations of measurement of the speed of sound in a gas, to show its pressure independence and the temperature dependence represented in equation (2-1). So these relations were well understood in 1978 when Moldover [108] proposed using a spherical resonator to measure the speed of sound in argon to determine the gas constant R, which is equivalent to the Boltzmann constant, related by $R = N_{\mathrm{A}}{\cdot}k_{\mathrm{B}}$. He had been motivated [109] by the discrepancy between accepted data and a new low uncertainty measurement of R by Quinn [110] performed using a *cylindrical* resonator. Tildesley identified an error in Quinn's calculation of the effects at limiting density [111] leading to revision of the data [112] which then represented the best available measurement of R and which agreed with previous data within calculated uncertainty.

The justification of the superiority of a spherical form over a cylindrical resonator however is in the Q factor of the resonance – in other words the purity of the frequency. The justification also reflects weak dependence of frequencies on shape imperfections in the sphere, identified by Moszkowski in 1955 in respect of spheroidal atomic nuclei [113]. Small order (volume invariant) smooth perturbations to the shape induce a second order effect on resonant frequency, so that a 1000 ppm tolerance on spherical form leads approximately to a 1 ppm uncertainty in acoustic resonant frequency. These ideas were formalised for acoustic resonators [114]. The advantages of spheres (and painstaking experimental technique) led Moldover to achieve an exceptionally low uncertainty measurement of R, published in 1988 [115].

A critical improvement to the spherical resonator was suggested by Mehl in 2004 [116] which was use of a so-called quasi-spherical resonator. Moldover had recognised in 1978 [108] that microwave resonances could be used to determine the thermal expansion of a spherical resonator cavity – since the microwaves are largely unaffected by the low-density gaseous medium. The behaviours of the acoustic resonances and the microwave resonances have other important differences. The radially symmetric acoustic modes are non-degenerate, which is to say there is a single mode of resonance with spherical wavefronts describing the whole gaseous body at a given frequency. By contrast, none of the microwave resonance modes has radial symmetry; there are multiple similar modes in different directions which, if the sphere is perfect, will all have the same frequency. This is the sense in which the microwave resonances are degenerate and the radial acoustic resonances are not. What this implies is that minor smooth imperfections in the spherical surface have negligible impact on the radial acoustic resonance, provided the volume of the sphere isn’t altered by them [113] whereas such shape imperfections cause a broadening of the microwave resonance peak (on a frequency versus amplitude plot) because there is a multiple of very closely valued frequencies. This subjects the measurement of the frequency of the microwave resonance to greater uncertainty. The new idea suggested by Mehl [116] was that the degeneracy of the microwave resonance modes could be broken by deliberately changing the shape of the sphere. The suggestion was to split the sphere into four equal quadrants, and separate them by two thin cylindrical washers just thick enough adequately to resolve the different microwave resonance frequencies. Correctly designed, this could avoid overlapping other microwave modes whilst not significantly perturbing acoustic resonance. Such a quasi-sphere approximates a tri-axial ellipsoid and its acoustic resonance analysis was also reported in 2007 [117]. A tri-axial ellipsoid is, by the definition given in 2.1.1 on page 8, a freeform surface. The 2007 approach by de Podesta [39] to the author was in order to find mathematical and machining manufacturing techniques suitable to make a tri-axial ellipsoidal resonator to a sufficiently small tolerance – with an expectation that this would involve diamond turning. This manuscript in part describes research towards its successful manufacture.

2.3.2 Scientific case 2: visual wavelength telescope optics

The telescope was invented no later than September 1608; a letter of support for Lippershey's patent application is the earliest recorded evidence [118] – and it is no accident that he was a spectacle maker. So, although spectacle lenses had been made since the 13th Century, the telescope's invention came about following the introduction of new manufacturing techniques for glass lens making in the late 16th Century [119]. The demand for telescope lenses though quickly drove the development of manufacturing technology [120] – market pull[*]. Whilst Galileo's telescope of 1609 was most likely based on the Dutch instrument brought into Venice earlier that year [121], he was the first to make scholarly study concerning lens quality and optical designs, and it was observations with his telescope that spawned scientific astronomy. Gregory's and Newton's later 17th Century reflecting telescopes were made possible through grinding and polishing of mirrors [122,123] rather than lenses [124]; it was these precision techniques for machining speculum metal[†] that sustained the advancement in the most powerful telescopes [125] until Foucault's 1864 invention of the metallised-glass reflecting telescope [126]. Through the first half of the 20th Century, the emphasis was on increase in size of primary mirror and the pinnacle of this phase of development was the 200 inch Hale Telescope at the Palomar observatory completed in 1948 [127]; this was unsurpassed in optical performance for decades. The second half of the 20th Century saw an emphasis on electronic detectors, instrumentation, control and analysis as telescopes had more or less reached the achievable limit of performance for the largest practical monolithic mirror size, reckoned to be less than 10 metres. No high quality steerable single piece telescope mirror has ever been made with an aperture larger than 8.5 m in diameter, and this limit is unlikely now to be surpassed.

The age of segmented mirror telescopes effectively dawned (when the performance of segmented mirrors could exceed a monolithic mirror) with the

---

[*] where the business demand for a product drives technological development to satisfy the demand

[†] a white brittle alloy of around ⅔ copper and ⅓ tin which can take a high polish

construction of the Keck twins, Keck 1 being completed in 1993 although this was far from the first example. In 1779 Erasmus Darwin had proposed and sketched a telescope with a multi-mirror primary [128] although his actual attempts to create one were unsatisfactory [129]. Guido Horn had begun successful experiments with composite multi-mirror telescopes as early as 1932; with practical observations beginning in 1935. His telescopes of the 1950s used hexagonal segments [130], albeit on a much smaller scale than Keck [131]. Beckers gives a near-comprehensive account (neglecting Darwin) of the pre-Keck development of multi-mirror designs [132]. Segmentation raises many manufacturing issues, two of them major: a) the issue of off-axis optics production and b) edge roll-off. The segments for Keck, 36 for each primary, were hexagons 1.8 metres across corners, each one part of an asphere, but substantially off-axis – and thus freeforms according to the author's definition (page 8). Nelson tackled the manufacturing of these by developing stressed mirror polishing [133]. This technique, which is suitable for so-called meniscus (< 200 mm thick) mirror blanks and telescopes with lower *departure* from sphere, involves polishing with simple techniques whilst the mirror is stressed with a departure from sphere which is essentially the opposite of the desired figure. Round mirror blanks start as a best-fit sphere approximating the final design shape, a bending harness that is adjusted for individual optical design parameters then subjects them to high stress, but within the elastic limit, and this harness holds the position whilst they're polished to a spherical shape, the bending harness effectively adding the required departure from sphere. After release they relax close to the design shape. To solve the roll-off issue, Nelson started with round blanks, which were then cut hexagonal. This stressed mirror technique has the advantage of employing conventional and relatively easy full aperture polishing and avoiding all sub-aperture CNC grinding and polishing; there are disadvantages in the complexity of harnesses and limitations in curvature. Final accuracy appears to be an issue however. The figure accuracy for the Keck segments is reported to have been more than 1 µm RMS [134] after polishing, which is comparatively large. Nelson employed post-polishing Ion Beam Figuring (IBF) to achieve the required final figure accuracy, which is now common to most process chains.

There are alternative process chains in use to solve the freeform issue. These involve sub-aperture work to generate the freeform shape. The distinction between full- and sub-aperture techniques is that for a full-aperture technique, the material removal tool contacts (usually along a curved line) across the whole of the mirror’s surface so that the resultant surface form is a function of the size of the tool and the angle between tool and workpiece spindles. This is reliable and makes inherently smooth and symmetric forms – it’s what allowed high quality optics to be made hundreds of years before the advent of machine control. Sub-aperture processes use a tool that is smaller than the workpiece and travels over the surface – accurate machine control is required to position the tool (sometimes to orient it too) and the resultant surface form accuracy is in some way dependent on the accuracy of motion. These processes also tend to rely on metrology feedback to enhance the accuracy, so that the material removal process is iterative, gradually approaching the designed shape – closer each time. The IBF, used to figure correct the polished Keck segments, is itself a sub-aperture process.

Manufacturing process chains that rely on sub-aperture processes usually have two or more steps. The first generation step is grinding, which can be full aperture (to make a spherical blank) or sub-aperture – which has freeform potential. Milling is an occasional alternative. Sagem use a proprietary grinding process called SAO employing a fluid jet stage followed by CNC lapping for the first stage preparation of telescope optics. Details are not published, but it is claimed to achieve 1 µm or less rms shape error [42]. The Large Optical Generator is a famous moving vertical orientation spindle machine with a fixed workpiece position, established in the 1980s [135]. It is used as the initial shape generator for a range of primarily aspheric optics, up to 8 m diameter with typically single digit µm rms form accuracies [136]. It has been modified to carry a variety of machining heads, including polishing, but grinding is always a first step. No other machine with the geometry of Box has been analysed, although there are publications addressing its grinding mode. Jiang [137] considers the grinding mode and assesses the kinematics of the tool path, but only as far as the contact point with an assumption of perfection of the grinding wheel. Also relevant, Xie [138] identifies a technique for considering the 2D profile of a wheel for freeform grinding and claims a 37% improvement compared with an uncompensated grind, although there is no indication of how the profile accuracy is determined or how wear can be accommodated in the model. These treatments are the closest to a tool path design process for Box, however neither accommodates all the required features, and particularly not the consideration of contact out of the plane of the motion axes – this point is not addressed in publication.

2.4 Mathematical descriptions of aspheric surfaces

A mathematical expression for a surface is most useful for manufacture if it is of functional form, where a set of inputs is related by the function to a set of outputs with the property that each input is related to exactly one output (explicit form). Typically, the input would be a combination of axial values (for instance of x & y) within a limited domain and the output would be a unique value (for instance of z) for that combination of input values. Sometimes, a surface is defined in Cartesian space by an expression which is an implicit rather than explicit function. It's useful to look at analytical expressions for aspheric surfaces first.

Youngworth [139] defines an asphere as "a rotationally-symmetric (invariant) shape represented with a sag equation $z = (\rho)$, where ρ is a radial aperture coordinate". The simplest aspheres derive from shapes called conic sections. The mathematical representation of curved surfaces dates to antiquity. 3rd Century BCE Greek mathematicians were possibly the first to develop systematic representations, considering conic sections and their surfaces of revolution [140]; the terms parabola, ellipse and hyperbola originate in these writings. A hyperbola is a section parallel to the cone axis, a circle's is perpendicular and a parabola's section is parallel to the cone angle. Any other section is an ellipse. These are shown in Figure 2-2.

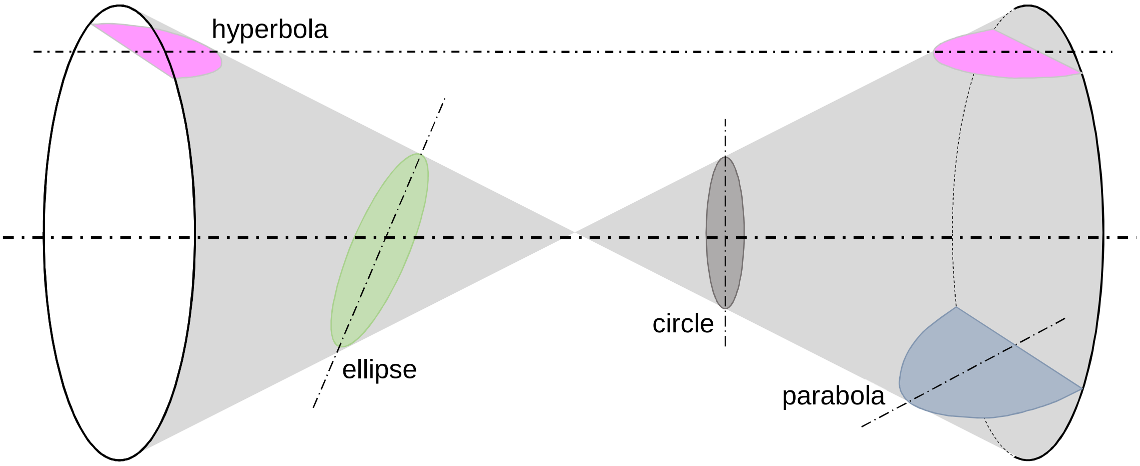


Figure 2-2: Conic sections

Simple aspherics are formed by the rotation of conic sections on their own axes (the minor axes shown in Figure 2-2) to make the 3-dimensional *circular* ellipsoid, paraboloid and hyperboloid – or indeed a sphere with rotation of a circle. These 3D surfaces of revolution of a conic section are sometimes called *conicoids*.

The Greeks also considered their application in optics, as did the 10th/11th Century Persian mathematicians specifically for lenses, building on the Greek work [141]. Later, the mathematicians of the Renaissance introduced a more recognisably modern treatment, again starting from Greek texts, as part of the flood of mathematical and scientific ideas of the Age of Reason [142].

The whole class of surfaces related to these can be described by a general multivariate (and implicit) formulation for second order aspherics of the form [143]:

$$\begin{aligned} F(x,y,z) = A_{11}x^2 + A_{22}y^2 + A_{33}z^2 + 2A_{12}xy + 2A_{13}xz + 2A_{23}yz \\ + 2b_1x + 2b_2y + 2b_3z + c = 0 \end{aligned} \tag{2-2}$$

Adopting the linear algebra representation of equations (2-3):

$$\begin{aligned} \mathbf{u}^T = (x \quad y \quad z) \quad \mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} \quad \mathbf{u} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad \mathbf{b}^T \\ = (b_1 \quad b_2 \quad b_3) \end{aligned} \tag{2-3}$$

a *principal axes transformation*, using the three eigenvalues λ_i and three eigenvectors w_i of the matrix $\mathbf{A}$, can render the equation of the surface into canonical form, equation (2-4) RHS, in a different co-ordinate system which is based on the surface's principal axes [144].

$$\mathbf{u}^T\mathbf{A}\mathbf{u} + 2\mathbf{b}^T\mathbf{u} + c = \sum_{i=1}^{m} \lambda_i w_i^2 + 2\beta w_{m+1} + \gamma = 0 \tag{2-4}$$

The principal axes transformation in general represents a rotation and translation of the surface onto principal (and orthogonal) axes of the surface, $\mathbf{w}^T = (w_1 \quad w_2 \quad w_3)$, one of which will be the rotational axis in the case of surfaces of rotation. Renaming w_1, w_2 and w_3 as the x, y and z axes for convenience, the surface can be expressed by equation (2-5):

$$\lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 z^2 + 2\tilde{b}_3 z + \gamma = 0 \tag{2-5}$$

The following conditions can be introduced [143]:

1. $\tilde{b}_3$ is either 0 or 1 (the whole equation can be divided by $\tilde{b}_3$ to achieve that)

It follows that if $\tilde{b}_3$ is 1 then λ_3 is 0 and γ can be eliminated by substitution

2. γ is either 0 or -1 (the whole equation can be divided by $-\gamma$ to achieve that)
3. if $\lambda_3 = 0$ then $\lambda_1 \leq \lambda_2$
4. if $\lambda_3 \neq 0$ then $\lambda_1 \leq \lambda_2$ and $\lambda_2 \leq \lambda_3$

If these are satisfied, the particular second order surface given by equation (2-1) exists in Cartesian space. We can introduce familiar (and geometrically significant) semi-axes $\lambda_1 = 1/a^2,\ \lambda_2 = 1/b^2,\ \lambda_3 = 1/c^2$ where c is real and a, b can be imaginary. Three cases arise from the conditions above.

Where $\tilde{b}_3$ and γ are both zero we get the second order conical surfaces of equation (2-6):

$$^{x^2}/_{a^2} + {}^{y^2}/_{b^2} + {}^{z^2}/_{c^2} = 0 \qquad \text{(2-6)}$$

where $\tilde{b}_3 = 0$ and $\gamma = -1$ we get second order centred surfaces of equation (2-7)

$$^{x^2}/_{a^2} + {}^{y^2}/_{b^2} + {}^{z^2}/_{c^2} - 1 = 0 \qquad \text{(2-7)}$$

where $\tilde{b}_3 = 1$, $\gamma = 0$ we get second order paraboloid surfaces of equation (2-8)

$$^{x^2}/_{a^2} + {}^{y^2}/_{b^2} + 2z = 0 \qquad \text{(2-8)}$$

For each of the cases, if $a = b$, the surfaces are rotationally invariant.

The formulation used in optical designs now is of 20th Century origin. Scharzschild introduced the conic constant that bears his name in 1905 [145,146] in the general formula for a simple aspheric – formed as a surface of rotation of a conic section – k in equation (2-9)

$$z = \frac{^{h^2}/_R}{\left(1 + \sqrt{1 - (1+k)\,^{h^2}/_{R^2}}\right)} \qquad \text{(2-9)}$$

where h is distance from the axis of rotation and R is a nominal radius of curvature (ROC) – actually the ROC at the axis. This can be formed by substituting $z \longrightarrow z - c$ in equation (2-7) (shifting z from the centre to the tip of the figure) setting $a = b$, $h = \sqrt{x^2 + y^2}$, $R = a^2/c$, $k = R/c - 1$ and then rearranging, or for equation (2-8) setting $a = b$, $h = \sqrt{x^2 + y^2}$ and then $R = -a^2$, $k = -1/R$. This formula is limited in scope because only these regular surfaces (Table 2-3, below) can be represented.

Table 2-3: Conic constant

Conic constant	Resultant shape
k<-1	hyperboloid
k=-1	paraboloid
-1<k=<0	prolate (sharp) ellipsoid
k=0	sphere
k>0	oblate (blunt) ellipsoid

More complex aspherics require an extension to the formula. This became important with the invention of the Schmidt corrector in 1930; the vacuum formed plate was closely represented (it transpired) by a 4th degree polynomial and this was added to its aspheric formula to give (more or less) a form of the generalised aspheric surface equation still in use today [147] – equation (2-10).

$$z = \frac{h^2/R}{\left(1+\sqrt{1-(1+k)\,h^2/R^2}\right)} + h^4 \sum_{m=0}^{M} a_{2m+4} h^{2m} \tag{2-10}$$

The ones given in Table 2-4 are rotationally invariant. The polynomial modifier is similar to the one in equation (2-10) and these can be manufactured using aspheric machining techniques, for instance spindle SPDT.

These explicit formulae are most useful where the departure from conicoid is low; the implicit formulae in equations (2-6), (2-7), (2-8) and other more complex formulae can be capable of representing aspheric surfaces with larger departure. The use of these in optical design is less common, but has been accomplished [148].

Table 2-4: 2nd order rotationally invariant surfaces (simple aspheres)

Basic surface		Power series modifier
Type	Equation	
(circular) ellipsoid, hyperboloid, paraboloid, sphere	$\frac{h^2}{R\left(1+\sqrt{1-(1+k)(h/R)^2}\right)}$	$\sum_{m=1}^{M} a_m h^m$
cone	$\frac{c}{a}R$	
plane	0	

The ISO standard for aspheric surfaces ISO 10110-12 [149] uses a far more general definition of an asphere; some of the surfaces included in its descriptions are freeforms under most definitions. A more complex surface that can be made using aspheric production techniques is a toric although this is a fourth rather than second order surface. In the standard [149] these are described as "surfaces of revolution; not coincident with coordinate axes". This is expressed in the standard as a surface of rotation around an axis parallel to X, rather than around the Z axis. The same surface can be represented more simply as a surface of revolution around the Z axis using the substitution $h = R - \sqrt{x^2 + y^2}$ with R the radius of rotation and r the perpendicular radius of curvature, as shown in Table 2-5.

Table 2-5: 4th order rotationally invariant surfaces

Basic surface		Power series modifier
Type	Equation	
Toric (ISO definition)	$f(x,y) = R_Y \pm \sqrt{\left(R_Y - g(x)\right)^2 - y^2}$ $g(x) = \frac{x^2}{R_x\left[1+\sqrt{1-(1+k_X)\left({}^{x}/_{R_X}\right)^2}\right]}$	$g_1(x) = A_4x^4 + A_6x^6 + \cdots + C_3\|x\|^3 + C_5\|x\|^5 + \cdots$
Toric (alt. definition)	$z = \sqrt{r^2 - h^2}, \qquad h = R - \sqrt{x^2 + y^2},$	$\sum_{m=1}^{M} a_m h^m$

The latest draft amendment to the ISO 10110-12 [150] includes a new formulation for the power series modifier. This adopts the work of Forbes [151] which confers significant advantages to the use of orthogonal bases; these will be discussed in the context of orthogonal polynomials in section 2.5.2.

2.5 Mathematical descriptions of freeform surfaces

The representations of freeform surfaces include [152,153]:

- Point clouds – the surface is represented by individual points, recorded as number triplets (typically x,y,z) – the number, position or ordering of the points is not prescribed. This is essentially the most basic form of freeform surface representation. In order to gain or process surface information some form of numerical manipulation is required to convert the point cloud representation to one of the other forms.
- Polygonal meshes – a mesh of 2D geometric figures (triangles, squares, hexagons etc.) is 'laid' over the surface and (3D) spatial co-ordinates of the surface are recorded for each vertex, thus representing the surface with a 2D array of number triplets – this is a standard representation in computer graphics.
- Splines – definitions vary [154,62] but in this manuscript: polynomial functions, piecewise-defined, with a high degree of smoothness where the pieces connect or overlap, designed to represent curved figures in 2 or more dimensions.
- Wavelets - functions used to localise a signal in both spatial and frequency domains. An arbitrary function can be represented by linear combinations of different wavelets which are scaled and translated from a prototype finite "mother" wavelet. They have the advantageous property that they can represent local as well as global features to an arbitrarily small tolerance.
- Analytical descriptions - which in general give a single-valued (functional) expression for the value of one co-ordinate in terms of the others, e.g. $z = f(x, y)$ – these often include polynomials and can be arrived at empirically or analytically [155].
- Orthogonal polynomials
- Radial Basis Functions

2.5.1 Analytical descriptions

A common form of analytical description of freeform surfaces is similar to the general aspheric equation (2-10) introduced to handle Schmidt plates, although there are important differences. The ISO standard for *aspheric* surfaces ISO 10110-12 [149] does introduce rotationally non-invariant surfaces, some of which are freeforms, so it's useful to consider their representation here first.

The remainder of the surfaces described in the ISO standard not covered in section 2.4, although derived from the second order aspheric equation (2-2) are freeforms in terms of manufacture, because they are not rotationally invariant, which in terms of equations (2-6), (2-7) and (2-8) arises when $a \neq b$. These are described by equation (2-11):

$$z = f(x, y) = \frac{x^2/R_x + y^2/R_y}{1 + \sqrt{1 - (1 + k_x)\left(x/R_x\right)^2 - \left(1 + k_y\right)\left(y/R_y\right)^2}} \qquad \text{(2-11)}$$

This can be obtained from equations (2-7) and (2-8) in the same way as for (2-9) except this time the substitutions are $R_x = a^2/c, \quad k_x = R_x/c - 1, \quad R_y = b^2/c, \quad k_y = R_y/c - 1$ for equation (2-7) and $R_x = -a^2, \quad k_x = -1/R_x, \quad R_y = -b^2, \quad k_y = -1/R_y$ for equation (2-8).

There are two other non-rotationally invariant cases to arise from the second order aspheric equation (2-2). The first is a non-coaxial cylinder which can also be expressed (as for the toric case in Table 2-5) with a change of variable to make it coaxial and therefore rotationally invariant. The second is the (non-circular) conical form which derives from equation (2-6) with a and b imaginary rather than real. These are tabulated in Table 2-6.

Table 2-6: 2nd order non-symmetric surfaces

	Basic surface	Power series modifier
Type	Equation	
ellipsoid, hyperboloid, paraboloid	$\dfrac{{x^2}/{R_x}+{y^2}/{R_y}}{1+\sqrt{1-(1+k_x)\left(\frac{x}{R_x}\right)^2-(1+k_y)}}$	$A_4x^4+A_6x^6+\cdots$ $+B_4y^4+\cdots$ $+C_3\lvert x\rvert^3+\cdots$ $+D_3\lvert y\rvert^3+\cdots$
Cone $(a\neq b)$	$c\sqrt{{x^2}/{a^2}+{y^2}/{b^2}}$	
Cylinder (ISO definition)	$\dfrac{u^2}{R_U\left[1+\sqrt{1-(1+k_U)\left(\frac{u}{R_U}\right)^2}\right]}$	$A_4x^4+A_6x^6$ $+\cdots+C_3\lvert x\rvert^3+\cdots$ or $B_4y^4+B_6y^6$ $+\cdots+D_3\lvert y\rvert^3+\cdots$
Cylinder (alt. definition)	$h=r,\qquad h=\sqrt{x^2+y^2}$	$\sum_{m=1}^{M} a_m h^m$

The draft ISO standard for Optical freeform surfaces, ISO 10110-19 [153] indicates that two dimensional functions of the form $z=f(x,y)$ or $z=f(R,\varphi)$ serve as analytical definitions for freeforms. An unlimited range of analytical descriptions is possible, although one example is commonly applied and is introduced in the standard. It is an analytical description made up of two parts: the base shape, a generalised conic section as already seen [149] in equation (2-11) and a polynomial part, the final term shown in equation (2-12). This form of analytical description is one often used for optical surfaces.

$$z=\frac{{x^2}/{R_x}+{y^2}/{R_y}}{1+\sqrt{1-(1+k_x)\left({x}/{R_x}\right)^2-(1+k_y)\left({y}/{R_y}\right)^2}}+\sum_{m,n=0}^{\infty} c_{m,n}x^m\,y^n \tag{2-12}$$

The use of this formulation has grown out of the application of Zernike polynomials for the description of optical wavefront distortion by aberrations in optical systems although unlike Zernikes, the polynomials in equation (2-12) tend not to be orthogonal. The use of these orthogonal polynomials is discussed next, in section 2.5.2.

2.5.2 *Orthogonal polynomials*

Zernike polynomials are orthogonal polynomials used for characterising wavefront errors in optical systems. Zernike derived these in 1934 to test the figure of a circular concave mirror, based on diffraction theory [156]. They are defined over a circular *pupil* of normalised radius $0 \leq \rho \leq 1$ and azimuthal angle $0 \leq \theta < 2\pi$ as follows [157] in equations (2-13)(2-13) for different ranges of m:

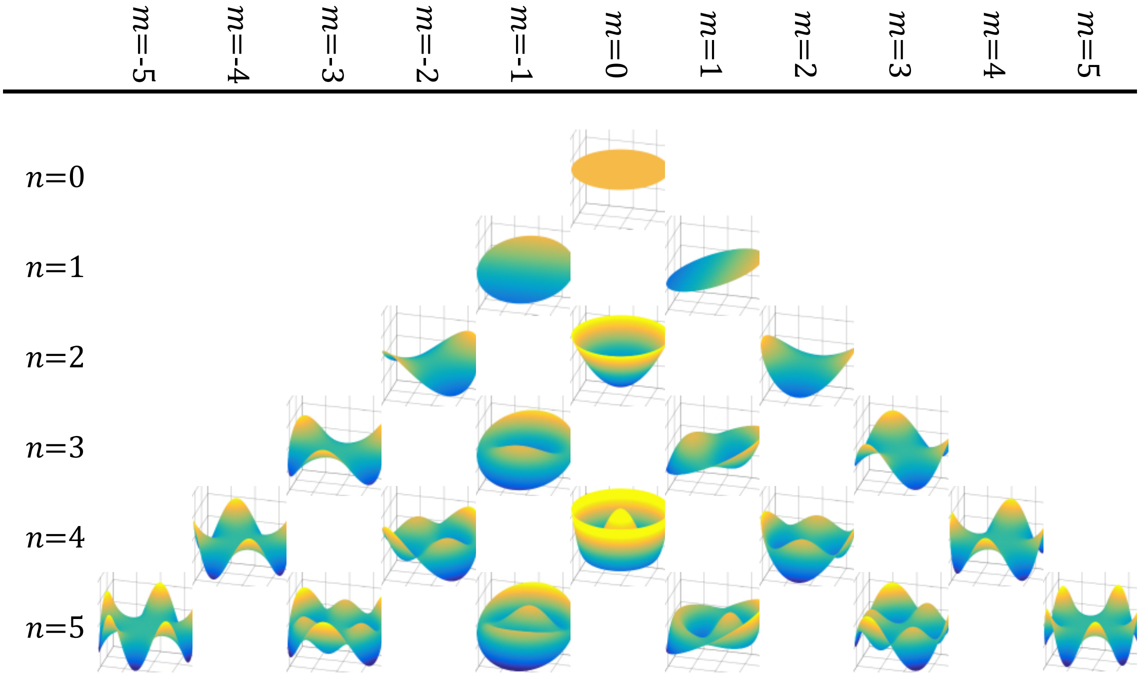
$$\begin{gathered} Z_n^{m>0}(\rho,\theta) = \sqrt{2(n+1)}R_n^m(\rho)\cos m\theta\,, \\ Z_n^{m<0}(\rho,\theta) = \sqrt{2(n+1)}R_n^m(\rho)\sin m\theta\,, \\ Z_n^{m=0}(\rho,\theta) = \sqrt{(n+1)}R_n^0(\rho) \end{gathered} \tag{2-13}$$

where $n = [0,1,2,\dots]$ is the radial order, $|m|$ is the azimuthal frequency, $n \geq m$ and $n - |m| = [0,2,4,\dots]$. The order of the polynomial is $n + |m|$. Radial polynomials R_n^m are given by equation (2-14):

$$R_n^m(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s} \tag{2-14}$$

In this context, orthogonal polynomials such as the Zernike can be useful for describing surface form or form deviations. There is a particular advantage though when Zernikes are applied to wavefront error, in that the different terms give balanced aberration descriptions which are useful to optical engineers [158]. Z_0^0 describes 'piston' (translation along z axis), Z_1^{-1} & Z_1^1 describes tip & tilt (rotation around x or y axes); these can usually be ignored as they relate to position and orientation, rather than shape. Z_2^0 is defocus, Z_2^{-2} & Z_2^2 are astigmatism, Z_3^{-1} & Z_3^1 are coma, Z_3^{-3} & Z_3^3 are trefoil, etc. – see Table 2-7.

Table 2-7: Representation of first 21 Zernike polynomials



Zernikes are an example of orthogonal polynomials which can also be used to describe the shape of a surface, although there are others. A surface is analytically defined in x, y, z space as: $z = f(x, y)$ and is continuous over a region A of x, y space. A series of polynomials $V_i(x, y)$ is defined as orthogonal over that region when equation (2-15) is satisfied, in which $w(x, y)$ is a weighting function and these can be used (in sum) to represent the surface $z = f(x, y)$.

$$\iint\limits_A V_m(x,y)V_n(x,y)w(x,y)dxdy = \begin{cases} 1 & n = m \\ 0 & \text{otherwise} \end{cases} \propto \delta_{nm} \tag{2-15}$$

Forbes, in a progression of publications from 2007 has advanced the case for “facilitating the enforcement of manufacturability constraints during design” and other design considerations, initially for axially symmetric aspheres [151,147] and latterly freeform optics [159,160] through application of orthogonal polynomials to describe surface departure from a base conicoid. Design and manufacturing criteria are variously based on surface departure deviation, slope deviation etc. [161,139,162], and orthogonal polynomial series can be optimised accordingly giving rise to Forbes’ Q polynomials, which are gaining wide acceptance, including in the draft standards for aspheric and freeform optics [150,153].

The general advantage of an infinite series of orthogonal polynomials of increasing frequencies is subtler than the Zernikes' representation of aberrations. Because the series is infinite and spans an infinite range of frequencies, it can represent a surface with arbitrarily complex surface character, although the series would have to be very long to describe high spatial frequency characteristics [163]. Since the polynomials are orthogonal, techniques of linear algebra and least squares can be used to find the coefficients for each polynomial that best describe a surface – in other words, more (higher order) terms can be adjusted independently to improve a fit without the existing terms' coefficients needing adjustment to compensate. This is a consequence of equation (2-15). Critically, from a design or manufacturing point of view whilst the terms in polynomial series in (2-10) and (2-12) also relate to spatial frequency, their inability to take independent adjustment means coefficients can't be toleranced. So, mathematical robustness of fitting and intuitive understanding of coefficients are the main motivation to adopt orthogonal polynomials in equation (2-12) to represent a freeform's departure from a conic base shape [162]. This is strongly endorsed by several authors [139] although Brick has identified no advantage in orthogonal over non-orthogonal polynomial representations in the design optimisation process for non-imaging optics [164].

Zernikes are orthogonal only over a *continuous* circular pupil. ISO 14999-2 on the interferometric measurement of optics [165] is at pains to point out that Zernikes will give a poor representation for non-circular pupils. If the surface data is present only in discrete points or portions of the pupil, Zernikes are no longer fully orthogonal, although they approximate orthogonality well for high data density for the lower order terms. In general, polynomial sets must be designed for a given pupil shape – e.g. circular, Gaussian, rectangular, hexagonal. ISO 14999-2 [165] indicates that it is possible to compute orthogonal polynomials over arbitrarily-shaped areas asserting that although not new, this knowledge is not applied in optics. Mahajan has developed polynomials of this type for optics for a variety of pupils, reported in a series of articles from 1981 to present [166,167,168]. This can also be accomplished to an extent by tightly enclosing the shape within a circular pupil, as discussed by Forbes [159].

A general limitation of orthogonal polynomials appears to be the very high orders required to capture even small quantities of features with small amplitudes of departure from the base form [169,162]. According to Mahajan [167] computation of high order terms can be subject to numerical instability with the widely-adopted Gram-Schmidt approach whereas Wang earlier [170] adopted Gram-Schmidt to avoid potential stability issues with least squares orthogonalisation, which he subsequently found to be stable. The origin of the actual instability can be traced however to limitations of the dynamic range of numerical representations commonly in use; for IEEE double precision numbers [171] this is around $2^{52} \approx 4.5 \times 10^{15}$. Forbes identified a three term recurrence relation which avoids the problems of relying on precise differencing between large numbers in the computation of orthogonal polynomials [172]; his Fig 1 gives a particularly compelling demonstration of the issue and its solution.

2.5.3 Splines

The use of splines to describe a freeform surface has been introduced in section 2.5 on page 28 where they were given this definition: “polynomial functions, piecewise-defined, with a high degree of smoothness where the pieces connect or overlap, designed to represent curved figures in 2 or more dimensions”. In the context of freeform surface representation, clearly this is in 3 dimensions. *Splines* were originally thin laths of wood used by *loftsmen* to lay out the curves for ships’ hulls or aircraft in the lofts of a factory (the only place with sufficient space to plot at 1:1 scale), by bending them and holding the curves in place with weighted ‘ducks’ or ‘dogs’. This craft is certainly well over 100 years old [173,174] but probably more than 300 [62]. In the early 1940s, this was put on a mathematical basis in *conic lofting* and at around the same time splines were invented in the modern sense by Schoenberg [63] working on aeronautical surfaces. He coined the term B-spline (much later) for basis spline although de Boor [175] notes that they were known earlier, for instance to Laplace or even Lobachevsky in the mid-nineteenth century. B-spline (2D curves) are defined in equation (2-16).

$$\mathrm{C}(u) = \sum_{i=0}^{m} N_{i,p}(u)\mathrm{P}_i \qquad a \le u \le b \tag{2-16}$$

u is a parameter that describes how far along a curve a point is. $\mathrm{C}(u)$ is the value of that curve at a given u - in this case x and y co-ordinates calculated from a fractional combination of control points. P_i are the control points - there are $m+1$ of them, each with an x and y value. The polynomial basis functions (N) indicate how much of each control point is to be used in the combination (sum). The $N_{i,p}(u)$ are p-th degree polynomial functions of u and there are $m+1$ of them. Underlying this there is a *knot-vector* which spans the u space between a and b dividing it into intervals corresponding to i in the equation. Each element in the knot vector is known as a *knot*.

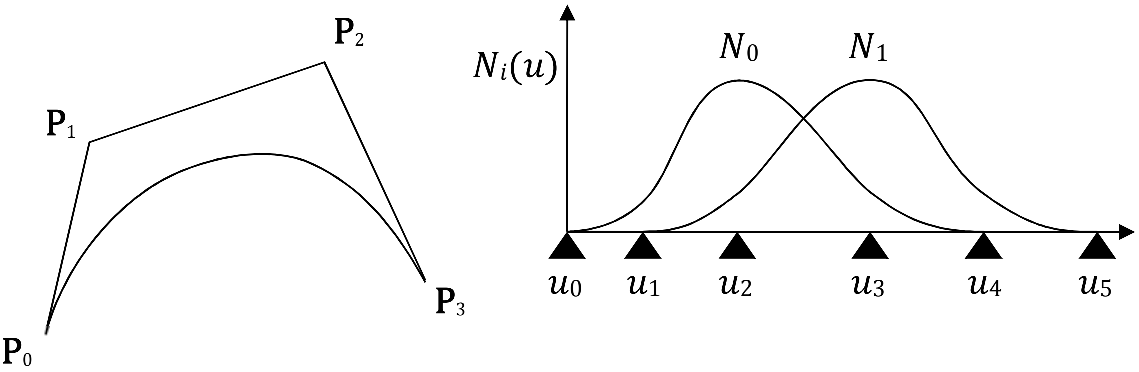


Figure 2-3: Construction of b-spline curves

The effect of this can be seen in Figure 2-3 where on the left a 2-D curve is constructed from control points and on the right, an example of basis function dependence on parameter u is shown for knots at u_1 - u_5 etc. Non-uniform b-splines include a weighting term w_i for each knot interval – and are made rational with an effective normalisation, as shown in the definition for a p-th degree Non-Uniform Rational B-Spline (NURBS) curve in equation (2-17).

$$\mathrm{C}(u) = \frac{\sum_{i=0}^{m} N_{i,p}(u) w_i \mathrm{P}_i}{\sum_{i=0}^{m} N_{i,p}(u) w_i} \qquad a \leq u \leq b \qquad (2\text{-}17)$$

This is the construction that will be used in the formulation of NC programs for control of toolpath in *Scientific case 2: visual wavelength telescope optics*.

The same concept represented in equation (2-17) can be extended for 3-D surfaces where a second parameter v is added for the additional dimension, to give equation (2-18) for a NURBS surface over an area parametrically defined by intervals $[a, b]$ for u and $[c, d]$ for v.

$$\mathrm{S}(u, v) = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} N_{i,p}(u) N_{j,q}(v) w_{i,j} \mathrm{P}_{i,j}}{\sum_{i=0}^{m} \sum_{j=0}^{n} N_{i,p}(u) N_{j,q}(v) w_{i,j}}, \qquad a \leq u \leq b, \;\; c \leq v \leq d \qquad (2\text{-}18)$$

Control points are now distributed on a rectangular grid in (say) (x, y) space although each $\mathrm{P}_{i,j}$ is 3-valued with x, y and z co-ordinates. The co-ordinate basis can of course be non-Cartesian, although the grid will always be rectangular in some 2-D space – this is perhaps the most significant limitation of NURBS surfaces, which are otherwise extremely flexible in application, being indefinitely differentiable and arbitrarily fine in their intervals – and therefore offering arbitrary precision in their approximation to a freeform surface. Knot vectors and even polynomial degree can be different between the different dimensions spanning the space, as seen in equation (2-18), although the same polynomial derivation is used for N, which is usually based on the de Boor recursive scheme [175].

Surface NURBS have been extensively applied to freeform surface engineering as discussed in section 2.3. Their application specifically in optical-type surface engineering (for ultra-precision machining) as required for "Scientific case 1: acoustic thermometry" and "Scientific case 2: visual wavelength telescope optics", is less extensive. Rigler in 1971 [176] offered splines (2-D splines, because an asphere is a surface of rotation) as an alternative to contemporary polynomial aspheric definitions, which the author takes to imply explicit solutions like equation (2-10) rather than (2-7), stating that in a ray-tracing design scenario, these solutions were easier to optimise. In 1984, Stacy [177] gave the important explanation of Rigler's contention, that spline surfaces were better behaved than conventional descriptions because spline coefficients directly represent surface sags. This means that if path difference (related to sag) is an appropriate optimisation criterion, spline coefficients will converge linearly, to the extent that each area of the surface is independently controlled. There is of course local influence between surface patches because of the 'spread' of basis functions over more than one knot interval, depicted on the right of Figure 2-3, but this is limited by proximity, and in any case, can be further limited by decreasing the spacing of knots. These advantages are utilised in several studies and NURBS representation is shown to be successful and accurate [178,179] although implementation can be hampered by the limited NURBS capability of industry standard optical design codes [180].

2.5.4 Polygon/mesh models

Mesh or polygonal model representations of surfaces have been most extensively developed in the field of computer graphics for the visual rendering of surfaces, and as such may not seem suited to precision surface representation. They are nonetheless listed in the draft ISO 10110-19 standard for freeform optics [153] as a standard form of representation. The representation is held in the form of vertices as a number triplet e.g. $V_i = (x_i, y, z_i)$. Two vertices joined become an edge. Three vertices connected by three edges become a triangle. More complex polygons can be formed from assembled triangles, although only triangular polygons are guaranteed to be planar, so triangular representations are the most common. Simple rules are usually enforced for mesh models:

- no co-incident vertices
- no edge can intersect a polygon or another edge (not self-intersecting)

The advantage of polygon models is that they are simple to represent, store and manipulate. Surface normals from triangles are easy to calculate – vertex normals (a combination of adjacent polygon normals) only a little more difficult.

A mesh model cannot represent a real curved surface without error; to interrogate the third axis (z) value from a given (x, y) co-ordinate, either the nearest neighbour vertex or some form of *interpolation* between vertices must be used. On account of this limited accuracy, mesh density may have to be set locally higher where curvature is greater in order to achieve a required tolerance.

2.5.5 Point clouds

As defined in section 2.5, "the surface is represented by individual points, recorded as number triplets (typically x,y,z) – the number, position or ordering of the points is not prescribed". Measurement data are often represented in this way, this representation arising from surface metrology or reverse engineering, when points are acquired individually for instance with manual control or from scanning with data acquisition timing not precisely correlated with the position/velocity.

Point cloud models can be converted to any of the other representations through a variety of means. Least squares techniques with an appropriate surface model can optimise the parameters of a global analytical model based on specified fitting parameterisations and criteria [181].

Shepard [182] in 1968 described a variety of interpolation methods for point cloud data including inverse distance weighting and described an extension to this which was effectively a piecewise polynomial representation, although not a true spline as there were slope discontinuities. Franke [183] reviewed numerous interpolation techniques, including splines.

For conversion to a triangular mesh a Delaunay triangulation [184] is commonly used, initially developed by Delone in 1934 [185]. This is reliable, but not always optimal [186] though a definition of optimal triangulation in the context of freeform surface fitting is hard to arrive at [187]. Sibson [188] introduced natural neighbour co-ordinates and subsequently [189] their application to scattered data interpolation, which is probably the most consistent interpolation basis for point cloud data, is used in the Matlab function "scatteredInterpolant" and adopted within this manuscript. Conversion to a spline surface definition may be accomplished following triangulation [190]. Ren [191] discusses the appropriate level of smoothness for such a spline surface definition based on measurement data.

2.5.6 Radial basis functions

A *radial basis function* (RBF) is usually a function of Euclidean norm, r in equation (2-19).

$$r = \|\rho\| = \sqrt{\sum_{i=1}^{n} {\rho_i}^2} = \sqrt{{\rho_x}^2 + {\rho_y}^2}, \qquad \text{in } n = 2 \ \text{ dimensions} \tag{2-19}$$

A surface $S(x)$ can be interpolated by a function $f(x)$ which is a sum of radial basis function values of x, a 2 or 3 dimensional x in this case, as in equation (2-20)

$$f(x) = \sum_{j=1}^{m} w_j \phi\left(\|x - x_j\|\right) \tag{2-20}$$

where the RBF is ϕ, each term in the sum (one for each of the m data points x_j) has a different weighting w_j. The weighting coefficients are fixed by the m interpolation conditions in equation (2-21)

$$S(x_j) = f(x_j), \quad j = 1, \dots m \tag{2-21}$$

and can be determined by the usual matrix methods. The RBF $\phi(r)$ can be one of several functions. This strategy was initially used only for interpolation of scattered data; Chan [192] used for instance the Hardy multi-quadric (2-22) or the thin plate spline (2-23) for interpolating antennae reflectors.

$$h(r) = \sqrt{c^2 + r^2} \tag{2-22}$$

$$d(r) = r^2 \log r + p(r) \tag{2-23}$$

This can be extended from interpolation to *approximation* by replacing the data points x_j in equation (2-20) with RBF centres, arbitrary c_j which can be optimised again using the usual matrix methods. Cakmakci investigated this for freeform surfaces, and specifically Gaussian RBFs [193,194].

2.5.7 Wavelets

The technique of wavelet representation was first introduced in the 1940s in the context of modelling seismic activity using overlapping travelling waves of different lengths and amplitudes [195]. This technique has been developed (wavelet decomposition/transform) and widely applied to represent data with disparate spectral and amplitude characteristics, particularly in the fields of computer graphics, data compression etc. and latterly freeform surface representation [196,197]. Jester [198] has demonstrated a wavelet method based on B-spline wavelets which can easily describe the low to mid-spatial characteristics well handled by other methods (such as Zernike polynomials). Interestingly the same method can also handle the higher frequency characteristics associated with manufacturing tools and techniques typically used for freeform machining, which are not so well handled by Zernikes etc.

The essential freeform surface representations have been well described in the literature and all have been utilised in some way in freeform surface manufacture. There has not however been a publication of a comprehensive approach to freeform surface representation that is optimised for tool path generation, error compensation and tool shape compensation.

2.6 Outline of metrology techniques for surfaces without specular reflection

2.6.1 Suitability of CMM for freeform surface metrology

A *specular surface*, being reflective, can generally be measured using optical techniques; conventionally most advantage can be obtained using wavefront comparison, typically by areal interferometry [51]. In these techniques, departure from the reference wavefront of the surface under test must be limited to a relatively small number of wavelengths – in particular the relative phase change for adjacent data points must be less than ½ wavelength. For freeform surfaces, departure from a conventional flat or (closest fit) spherical wavefront may exceed this limit. This limitation can be addressed principally in three ways. The first of these is a) generation of the freeform reference wavefronts using Computer Generated Hologram (CGH) / auxiliary optics. The second is b) sub-aperture techniques in which only a fraction of the aperture is examined before repositioning to examine a different fraction; the relatively small number of fractional or sub-apertures are later combined to give a single data set describing the entire surface. The third is c) the use of relatively long wavelengths, e.g. LWIR at more than 10 times visual wavelength. These wavefront comparison techniques are however restricted (by the departure criterion) to surfaces with relatively small Gaussian curvature.

For non-specular surfaces (which includes almost all ground and most turned surfaces) these areal techniques are non-functional regardless of curvature and single point metrology (whether optical, tactile or otherwise) must generally be applied because the surface texture is such that the departure criteria above are exceeded. Whereas the areal techniques for specular surfaces can acquire millions of data points simultaneously, the single point techniques can acquire only one data point at a time; there is a delay between acquisitions, accompanied by a relative repositioning of surface under test and measurement apparatus. Acquiring millions of points entails millions of relative motions and consequently millions of consecutive delay periods. This combination of motion and the passage of time results in additional measurement uncertainty, not least that associated with thermally induced and other drifts in relative position.

A general-purpose metrology instrument applicable to single point measurement techniques is the Co-ordinate Measuring Machine (CMM) an instrument with a long history and profound impact on measurement practice in diverse applications [199]. The most familiar Cartesian configuration is applicable to freeform surfaces and in an industrial context is the most widely applied [200,51]. Other CMM geometries are also applicable: of note being the swing-arm profilometer. The profilometer is adjusted so that a swinging arm, which carries a point measuring probe, describes a circular arc which is a close fit to the surface under test. It is applicable where aspheric departure (and relative normal angle) can be held within a manageable limit imposed by the range of the point measuring probe [201]; this applicability encompasses many optical surfaces –

for instance a proportion of telescope mirrors. Where departure is larger, or a generic metrology capability is also required of the same instrument, the swing arm device can't be used. In such cases, Cartesian CMMs can still be used. Jing [202] gives a pertinent example of comparison of Cartesian CMM discrete point measurements with swing arm scanning profilometry, for the same ground optic. Supranowitz [203] gives an example of Cartesian CMM scanning for an optic with large aspheric departure, although several deficiencies were noted: a) repeatability limited by shorter wavelength surface texture variations, even in a polished surface; b) limited spatial resolution of measurement data; and c) reliance on CMM software solid model capability with poor fidelity to the mathematical design model.

Although the swing arm profilometers can produce scan measurements with relatively low uncertainty, there are significant limitations for speed and coverage, particularly for very high 3D spatial resolution mapping with near anisotropic mapping density, as for interferometric tests. There are no published examples of CMMs producing a high accuracy map with equivalent resolution to a full aperture interferometric test.

2.7 Outline of freeform process chain rationale

A freeform surface can be manufactured, based on a surface description as in 2.5 above, and starting with a piece of material roughly shaped, for instance by casting, forging or other forming. These initial shaping processes are generally of relatively low precision[*] and accuracy[†]. Subsequently the accuracy of the surface may need to be improved by an additional process step, or more usually *steps*, which constitute a manufacturing *process chain*. This will normally involve an improvement of accuracy at each step, although the purpose of the chain is subtly distinct from that, in that the output quality demanded of each step is to satisfy the input quality requirement of the immediately subsequent step. The output quality demanded of the final step is to satisfy all design criteria for the surface.

The input quality demands for each step may be given in terms of an accuracy specification which can be arbitrarily complex, possibly involving spatial

---

[*] precision is the degree of closeness to which repeated processes will produce the same result

[†] accuracy is the degree of closeness to which a process will produce the correct result

frequency as well as amplitude components. Each step will remove (or conceivably add) an amount of material in an amount of time, using an amount of energy and with amounts of financial and environmental cost. These amounts may vary according to input quality, output quality requirements and even local variations in the surface, such as slope, spatial frequency and amplitude. Therefore, the optimisation of the process chain is generally highly complex [204], involving the selection of process step methods and their parameters. Some process steps can be either not well controlled (seemingly non-deterministic) or necessarily limited in capability; for instance, in sub-aperture processes, limited by the size and shape of a *material removal function*. This will entail the repetition of a process step multiple times (possibly with interposing measurements), using adapted process parameters, giving an improvement for each step, which will critically depend on measurement at each step.

2.8 Knowledge gap

There is no published research regarding the treatment of contact points out-of-plane (motion axes plane) while machining freeform surfaces with a fixed orientation tool. Its requirement is an inevitable consequence of the Box machining geometry, having only 3 axes in a cylindrical configuration.

There is no published research demonstrating high (near isotropic) 3D spatial resolution measurement of a non-specular freeform. The advantage of uniform high resolution data conferred by an interferometric test is not available after fine grinding.

For optimal tool setting accuracy, a technique for aspheric and freeform surfaces that can extract tool decentring offsets, utilising a full post-machining 3D surface map is required, but unreported.

3 RESEARCH APPROACH

3.1 Summary of aim

The hypothesis of this research (section 1.3) is essentially that it is possible to make very high accuracy freeform surfaces using machines with minimal (three) axes of motion, and that the considerable difficulties of having a fixed orientation tool do not outweigh the advantages for motion performance conferred by the relative simplicity of the machines. There is a tacit assumption moreover in the aim for the research, that it will be possible.

The aim of the research is to create a mathematical framework for the manufacture of freeform surfaces using 3-axis machines possessing fixed orientation tools. The framework should also accommodate freeform measurement, using an unmodified industrial standard CMM, to give surface deviation maps with a spatial resolution which is the equal of full aperture interferometric tests; maps which are suitable both for process validation and for feedback error compensation for control of machining.

3.2 Research methodology and objectives

A literature review has identified gaps in knowledge; at the same time two projects (essentially the case studies) associated with the research have specific deliverables (freeforms to be machined and measured). The achievement of the deliverables requires the closing of these knowledge gaps.

One component of the methodology is clear: a validation of the research will be accomplished in the achievement of the deliverables; in each case study the deliverable freeform surfaces (and to some extent, this research's metrology data) are required at a certain level of accuracy within the immediately subsequent process step.

All three of the functional processes involved in this research are constrained in a very particular way – they operate on a point by point basis, or on a zone which is very small relative to the freeform surface. The machine's motion systems have just 3 axes in each case: BoX grinder, Moore UPL 350 and Leitz PMM-F CMM (see chapter 4). Each translates its *end effector* whether that's a grinding wheel, diamond tool or probe stylus tip, over the surface, through a series of locations, which ultimately are discretely computed.

a) One element of this research therefore is to find a way mathematically to describe these locations, with an ability to do this at an arbitrary resolution, based on the description of the surface provided or available; *freeform surface representation.*

b) Another element of the research clearly is to create a way to define the connective sequence (and selection) of these locations in respect of each of the three functional processes, and for this to be predicated on the requirements of the process and the capabilities of the machines; *end effector path definition.*

c) An associated element of the research is to find a way to represent the shape of the tools (for the machining processes) which can inform the motion control of the machines; it is recognised within the research question that the motion path may not map in an obvious way to the connective location sequence in b) above; *tool shape representation.*

d) A vital element of the research is to establish the capability to generate high spatial resolution measurement maps of freeform surfaces with in the order of 1,000,000 points of data, in order to match the capability of interferometric measurement, so that the data can be used in a similar way; CMM-derived *high spatial resolution measurement maps.*

e) One research element interrelates the others, which is in the combination of freeform surface representation, end effector path definition, tool shape representation and CMM-derived high spatial resolution measurement maps through a *kinematic combination strategy* to produce tool paths*.*

An essence of the adopted methodology is the division of the research into discrete constituents, or work packages. There are four of these, which correspond to the principal technical chapters of the thesis.

i) The first of these chapters is a comprehensive treatment of novel algorithms in scanning CMM contact measurement of freeform surfaces. A demonstration freeform surface is selected as the development artefact, which is distinct from the case study surfaces. This has the property, by virtue of its cellular support structure and thin section, of having a mix of spatial wavelengths of freeform surface variation whilst retaining a relatively smooth surface absent high spatial frequency artefacts. Consideration is given within the central requirement of high density data maps to alternative probing techniques; the origin, implications and removal of measurement noise; techniques for storage, retrieval, display, resampling of measurement data; sources of error, enhancement of accuracy for different origins of error; and the requirement and methods of parametric extraction.

ii) The second of the technical chapters presents the basis of the development of freeform machining toolpaths – research which is common to the two case studies, which includes the principles of tool path design which stem from machining process requirements; influences on tool path design, which include pre- and post-machining surface properties, tool shape, boundary effects, machine capabilities including dynamics. Consideration is given to error compensation within tool path design and then the reasons for and techniques of tool path compression, including novel algorithms for optimal compression.

iii) The third technical chapter presents research towards: Application of solution, case study 1 – Boltzmann “quasi spheres”. This presents the implementation of the research strategies of i) and ii) above, with novel developments specific to this case.

iv) The fourth technical chapter presents research towards: Application of solution, case study 2 – E-ELT. This presents the implementation of the research strategies of i) and ii) above, with novel developments specific to this case.

a) - e) represent the broad research objectives, together with validation through iii) and iv).

4 EXPERIMENTAL EQUIPMENT

4.1 Co-ordinate Measuring Machine (CMM)

4.1.1 Leitz PMM-F 30-20-10

The Leitz PMM-F 30-20-10 is a high-accuracy monolithic gantry measuring machine with 3x2x1 m motion strokes in x, y, z respectively (see Figure 4-1) and a large effective measurement volume.

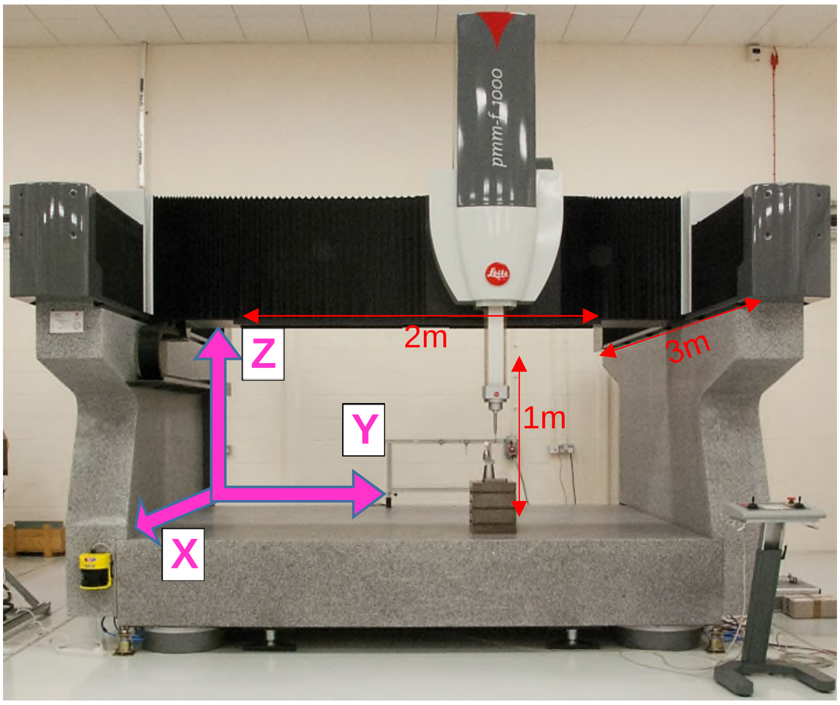


Figure 4-1: Leitz PMM-F 30-20-10

$$MPE\ (\mathrm{ISO}\ 10360-2) = 1.9 + L/400 \qquad (4\text{-}1)$$

The author selected this model for purchase based on suitability for measurement of large freeform surfaces (criteria listed in Table 4-1) and took delivery of the UK's first such instrument in 2007. Its maximum permissible error statement is given in equation (4-1).

Table 4-1: CMM selection criteria

Criterion	Requirement	PMM-F 30-20-10
measurement uncertainty	class-leading	class leading (1.9 + L/400)
suitability for phase shift interferometry	short term (~ 1 second) stability < 50 nm	suitable for phase shift interferometry
shortest horizontal stroke	>1.5 m	3 m x 2 m horizontal
software capability	flexible programming environment	Quindos & PC-DMIS

4.1.2 Quindos Software

Quindos is a long-standing (since 1986) CMM-focused measurement software package with open architecture. Despite some legacy issues related to its market longevity, the requirements of backwards compatibility and its highly complex command structure, in the opinion of the author it retains significant advantages. Among these are a) the most comprehensive and flexible available programming environment for sophisticated measurement control and preliminary data processing and b) the highest level of flexibility for data structuring in preparation for subsequent post processing in more powerful mathematical programming environments. For these reasons, Quindos has been adopted for measurement data acquisition and pre-processing for all contact-based dimensional measurements within this work.

4.2 Diamond turning machine

The Moore Nanotechnology Systems range of diamond turning lathes adopt oil hydrostatic guideways with linear motor drives for horizontal axes and a horizontal axis air bearing workspindle with direct drive. This gives a market-leading lathe performance (typically ~1 nm Ra surface finish, < 1 µm form accuracy) for rotationally invariant surface forms in ductile materials. The author completed the purchase of this machine, taking delivery in 2004. Figure 4-2 shows the machine (left) installed and (right) without containment.

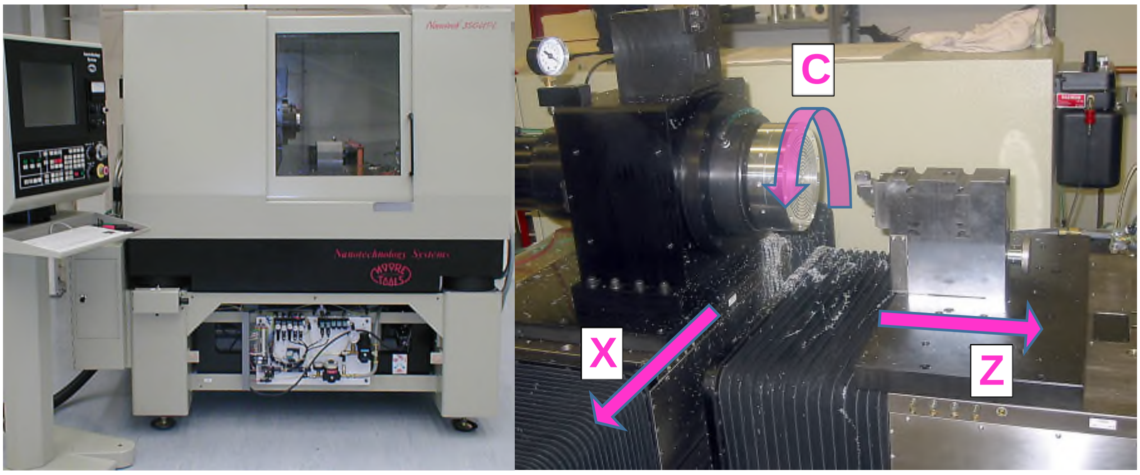


Figure 4-2: Moore Nanotechnology Systems 350 UPL – left: machine overview; right: motion configuration

4.3 Large optics grinding machine

The BOX (Big OptiX) large optics grinding and measuring machine was designed and built at Cranfield University by a team led by the author; the author is cited as inventor in all the associated patent applications [205]. The design was first published in 2005, by the author [26] and the machine first publicly demonstrated in 2006, by the author at the MACH 2006 machine tool exhibition and conference [206].

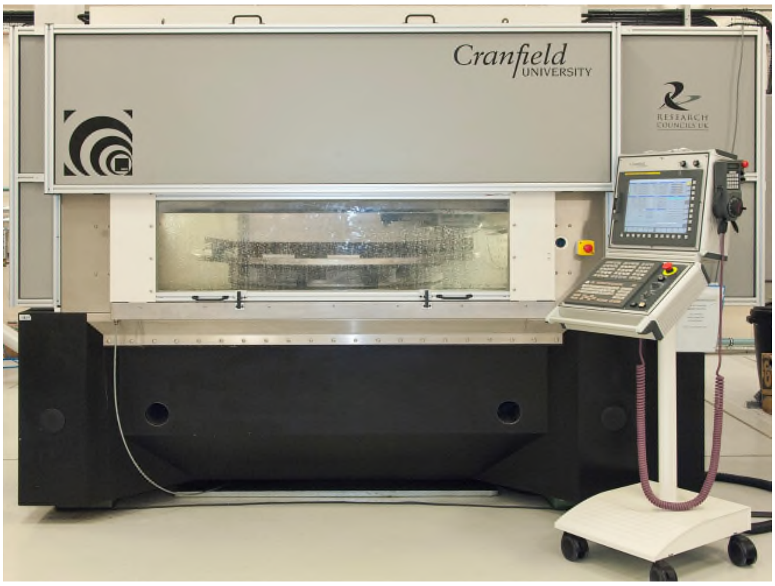


Figure 4-3: BOX – Big Optix Grinder

The machine was designed for rapid grinding of freeform optical surfaces (in the configuration shown) up to 1.5 metres diameter, with a surface form accuracy of

< 1 micrometre, with minimal subsurface damage. This level of performance is secured through motions with high static and dynamic stiffness (high stiffness guideways, low moving masses/inertias, direct drive high force/torque motors, collocated motive and measurement transducers); and high accuracy (minimised Abbe errors, precise temperature control) [9,12,25].

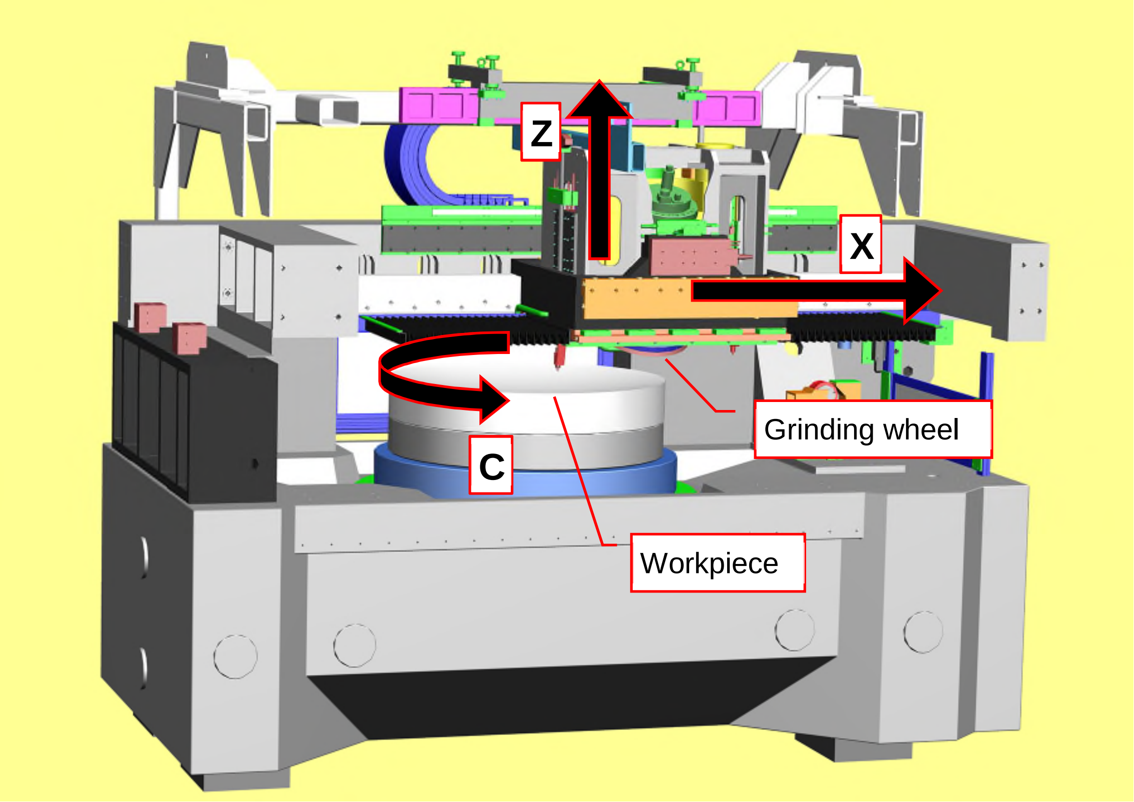


Figure 4-4: BOX – Big Optix Grinder motion configuration, solid model picture credit: Roger Read

4.4 Matlab

Matlab is a numerical computing environment [207] and most pertinently here, a proprietary fourth generation (primarily) interpreted and highly optimised programming language with multiple available application toolboxes with very wide industrial and academic use across engineering, scientific and economic sectors – among others. It reputedly has well over 1,000,000 users worldwide. With minor exceptions, all numerical data processing and visualisation within this study is accomplished using Matlab.

5 ALGORITHMS FOR SCANNING METROLOGY OF LARGE AREA SURFACES

Within the process chains for large freeform or other low curvature surfaces, there are frequently multiple steps, each step refining the surface closer to its final required condition. The parameters defining achievement of final condition typically relate to the subsurface and to the surface topography. Each step will have an output quality in these parameters that must meet or exceed the input quality requirements of the subsequent step. Process chain design is often complex. In order to optimise a process chain, it may be necessary to start with the final process and allow its input requirements to guide selection of previous steps. Typically, cost budgets and budgets for error, time, power, environmental cost etc. may be involved; there may not be clearly obvious choices.

Clearly measurement is likely to be critical in optimising process chain design and will be required within or between steps, for several reasons:

- as a certification that performance requirements of a previous step have been met – this is particularly important if successive process chain steps are performed by different production units – this also applies to final verification of the surface
- to provide information (most simply, alignment to reference features, but often more) required to initiate a process step
- as part of a process step, particularly where a process is iterative and therefore dependent on measurement information
- related to the above – for process monitoring/qualification over a period of time
- occasionally in-process as part of process control

For the smoothest large freeform surfaces, particularly optical surfaces, interferometry is the principal measurement technique applied to verify final figure/form. Within a freeform surface's manufacturing process chain however there are typically steps requiring accurate surface generation, for which interferometry is inapplicable: perhaps due to the surface being non-specular, or due to the time taken, or due to the geometry of the surface. The geometry may create access or other difficulties, or require in-situ measurement (impractical on most machines), demand a large included range of curvature, or too large a departure from best-fit sphere or flat etc. for interferometry to be accommodated. There may still though be a requirement for high-density areal measurements over a relatively large surface, even where interferometry cannot be applied. This is the case in both of the case studies in this manuscript – most particularly for *Scientific case 2: visual wavelength telescope optics*.

The motivation for this part of the research is to have interferometer type data for a non-specular surface, in part to improve accuracy against alternatives, but most particularly to allow a higher resolution so that measurement data can be used for error correction of the surface creation process. In *Scientific case 2: visual wavelength telescope optics* this is applied at the fixed abrasive manufacturing step and as an input into the first free-abrasive machining stage, which is most typically a *neutral removal* step. A neutral removal is a (polishing) process step designed to remove an even, thin, depth of material over a surface, thereby leaving surface form unchanged but at the same time improving surface finish sufficiently to make interferometry possible. It is frequently difficult to establish full aperture interferometry on a large optic, owing to the minimal departure capability of a full-aperture test. The capability to give interferometer-quality data for a surface moreover affords the possibility of useful measurement of freeforms that would be outside the capability of an interferometer, at least without an expensive and surface-specific null compensator, CGH etc. whether they are specular or not. Successful application of the measurement technique may obviate the need in this scientific case for the neutral removal altogether – leading to improved process chain efficiency, so that figure corrective polishing could commence without an interferometric step. For this reason, high lateral as well as vertical resolution is desired, with improved accuracy, beyond the 10-micron regime capability of a CMM of suitable size.

The algorithms developed and reported here apply in principle to any scanning instruments for freeform surfaces which could include swing-arm profilometers. However, there are disadvantages with swing-arm profilometers which are difficult to surmount in this application. Swing arms in general need to be mounted on the production machine to avoid the disadvantages of a lack of general fiducial capability in transferring from machine to machine – however a machine optimised for productive capability such as BoX (Section 4.3) is not suitable for inclusion of a swing arm, due to mechanical and other constraints. CMMs also are widely available and routinely provide traceability to the international standard of length, as well as possessing the fiducial capability for accurate workpiece transfer – and so are strongly preferred.

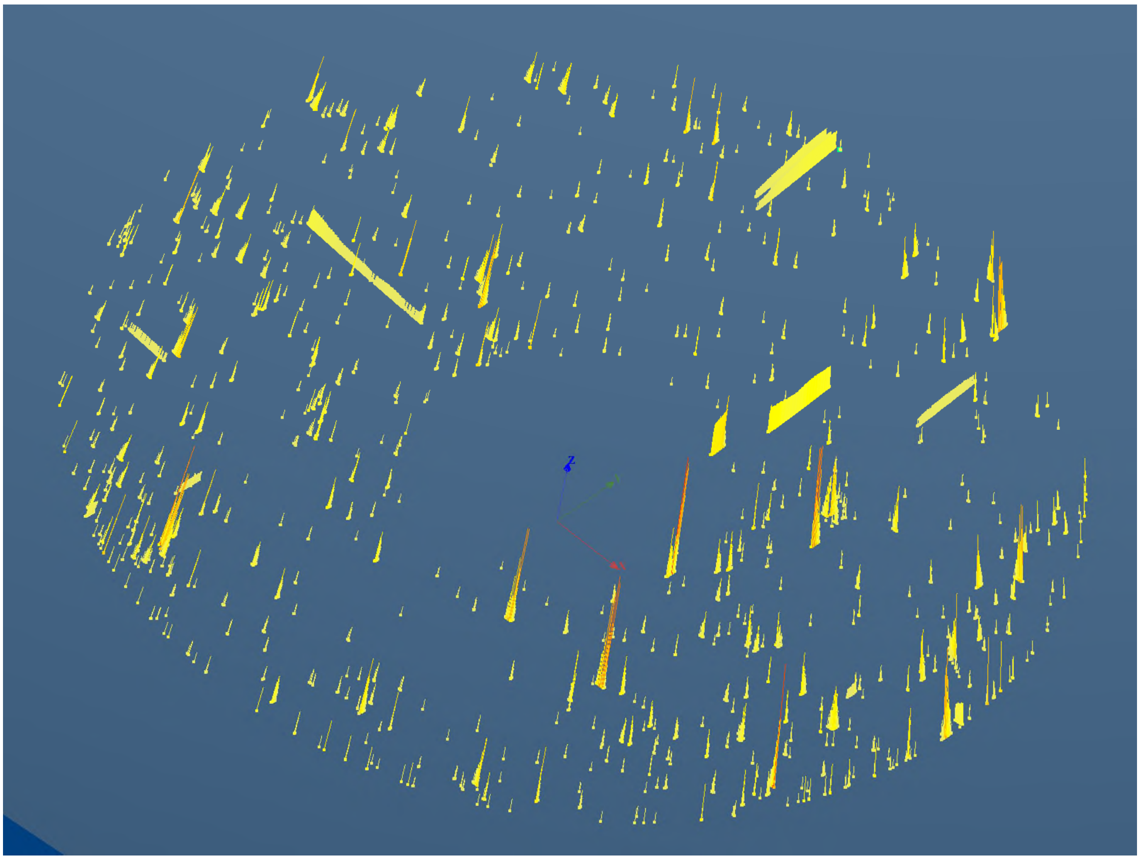


Figure 5-1: Tactile scanning measurement spikes

The presented technique assumes that contact errors are due to surface dirt/contamination and are randomly/uniformly distributed, and because of that uniformity that their treatment on an individual basis has no influence on the measurement of underlying form. Figure 5-1 shows an overall view (from a Quindos output) of high amplitude contact errors on a surface measurement. These data are processed outside Quindos and reloaded into Quindos for display. All the other displayed data in this chapter are programmed outputs from Matlab. It can be seen from the figure that these contact errors are not surface features, given their different locations in four distinct scan directions; this supports the assumption of a uniform distribution of contact errors. The contact errors are associated with dust/particulate contamination, either on the freeform surface, or on the stylus tip. Their effects need to be removed from the measurement data without affecting measurement of underlying form. Their removal is required partly for visualisation demands, but also because being all positive (away from the surface) and large compared to the surface form, waviness and other roughness, they would otherwise influence measurement parametric results.

Choice of scanning speed, measurement density and separation distance of scan lines have been made to accommodate a suitable volume of data, a feasible length of measurement time and a useful degree of redundancy in measurement. Some influences of scanning speed on measurement performance will be discussed. For instance, an ESO E-ELT segment is of hexagonal shape, nominally 1.46 metres across corners (AC). This is an area of approximately $\sqrt{27/64}\,(\mathrm{AC})^2 = 1.3845\ \mathrm{m}^2$. A measurement process that is no slower than the grinding process is required, so that measurement does not become a production bottleneck. The grinding process is capable of approximately 10 hours/m^2 production rate implying the measurement needs to be at least as quick – less than 15 hours for an E-ELT segment.

The measurement density must also be sufficient to avoid *aliasing* of known surface features, such as the scalloped pattern from grinding with a formed tool or of diamond turning. Typical final pass grinding marks have a profile with a repeating pattern of ~ 1 mm – for instance BoX optics are often ground in a spiral, with a repeating pitch of 1 mm, using a grinding wheel of toric shape with a 2nd radius of 300 mm. This is made up of a track (or tracks) caused by the curved shape of the tool where it engages with the workpiece surface. The track cross-section is scalloped with a cusp height in the order or microns. This gives a cusp height of at least 0.42 µm ($h = f^2/8R$, from the pitch – f and the tool radius - R). There are additional height influences in grinding spindle asynchronous errors motion. This sets a minimum data size more or less from the Nyquist (frequency) criterion. Other grinding geometries produce different cusp heights and pitches.

A similar geometrical effect is caused in diamond turning, although here the scallops are hundreds of times narrower and shallower. A critical function of the measurement is to determine the position of the cusps, in particular their z position as a function of x-y to a sub-micron precision; this is critical determinant final figure accuracy. This requirement imposes restrictions on areal density of the measurement data. The data points must either correspond well in terms of location to the trough of the cuspate cross-section, or be of sufficient density to be able to sample it effectively – or if the cuspate height is negligible (as in finish diamond turning) these considerations may be unimportant.

5.1 Selection of demonstration surface

As an illustration of the technique, a ground concave silicon carbide light-weighted mirror provided by Boostec company (Figure 5-2) is measured, of a type used for space telescopes.

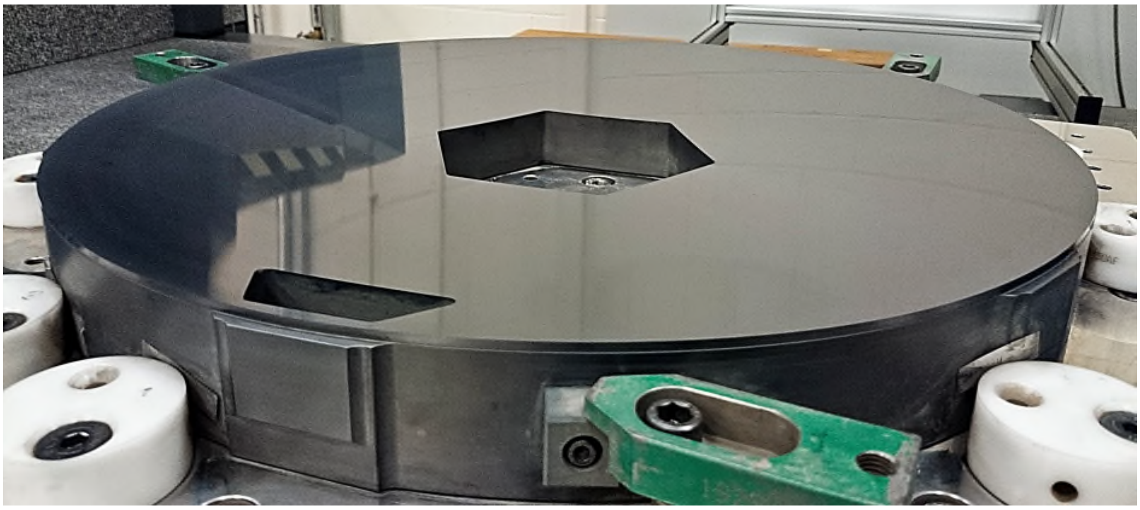


Figure 5-2: Silicon carbide large optic (610 mm diameter, spherical concave ~ 2 m Radius Of Curvature) – near specular surface after fine grinding

This is particularly appropriate, since whilst the structural light-weighting on the reverse side (Figure 5-3, left) is not visible at all on the functional surface side, it does produce tiny variations in the mirror sheet’s perpendicular stiffness across the surface. These local variations in stiffness give rise to micron-level variation in deflection during the application of grinding force. This in turn leads to a ‘print through’ effect which is not visible after grinding, but can nevertheless be measured (Figure 5-3, right) using the techniques in this chapter, revealing the underlying light-weighting structure, when the surface patches spring back on release of the grinding force.

This print through pattern is useful in the evaluation of a metrology technique as it exposes the finesse of the measurement and its ability to represent complex form, of relatively high spatial frequency, more or less without the noise associated with other contact measurement techniques. These new techniques operate at a quality normally associated with full aperture (or stitching) interferometry – the latter techniques however can't be applied to a ground surface. This type of (high-density) data allows compensations to be applied within the grinding program which couldn't normally be achieved. Faithful recording of this structural print-through in metrological data would offer a particular challenge to measurement techniques using any type of window or *apodization* function-based filtration to handle noise, because the structure's spatial frequency is relative close to the measurement sample frequency, so this is a particularly severe test. In addition, the apertures offer an additional complication and afford a mix of straight and curved edges.

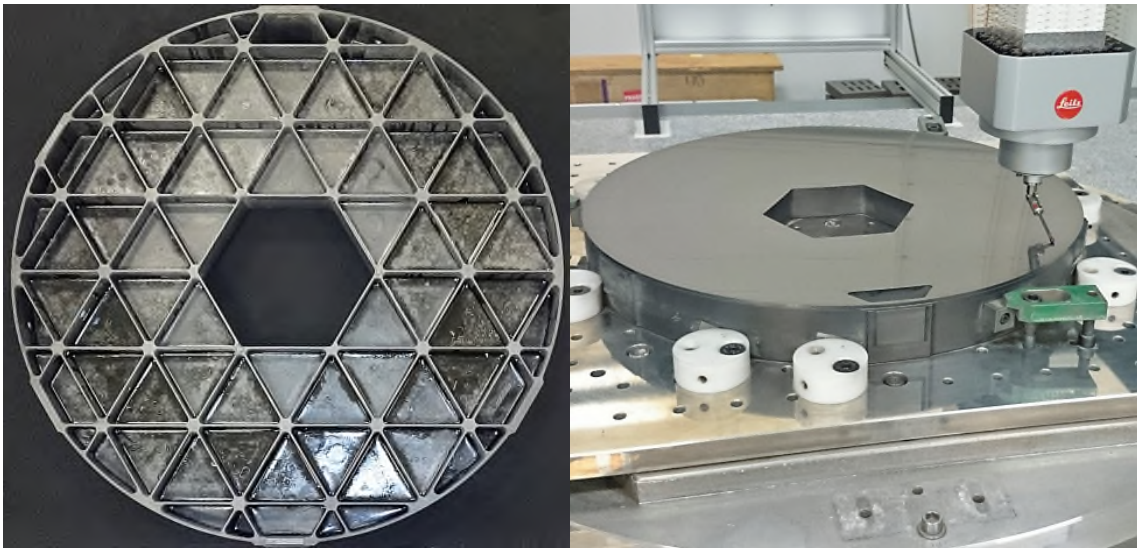


Figure 5-3: Silicon carbide large optic - left: reverse side light-weighting pattern; right: scan probing technique

5.2 Measurement by single point touch

Multiple single-point measurements are the standard technique for CMM measurement; used to establish relative location/orientation (the 6 rigid-body degrees of freedom) as well as geometrical information relating to dimension and conformance to specification. Specifications will frequently include form (such as flatness), but measurements of form most commonly are not based on large numbers of points. The capability within a CMM to gather large numbers of points for form assessment does however exist, as seen in Figure 5-4 where ~ 3,500 points were measured on the functional (smooth) side of the SiC large optic.

Single point measurement is regarded as the CMM's most accurate measurement mode although there are some difficulties associated with its use for large data sets:

speed:	gathering single points is relatively slow, typically ~ 3 seconds per point
data:	CMM native software support for handling large point cloud sets of freeform surface data is generally very limited
noise:	particulate contamination can degrade the process so that there is missing or unusable data

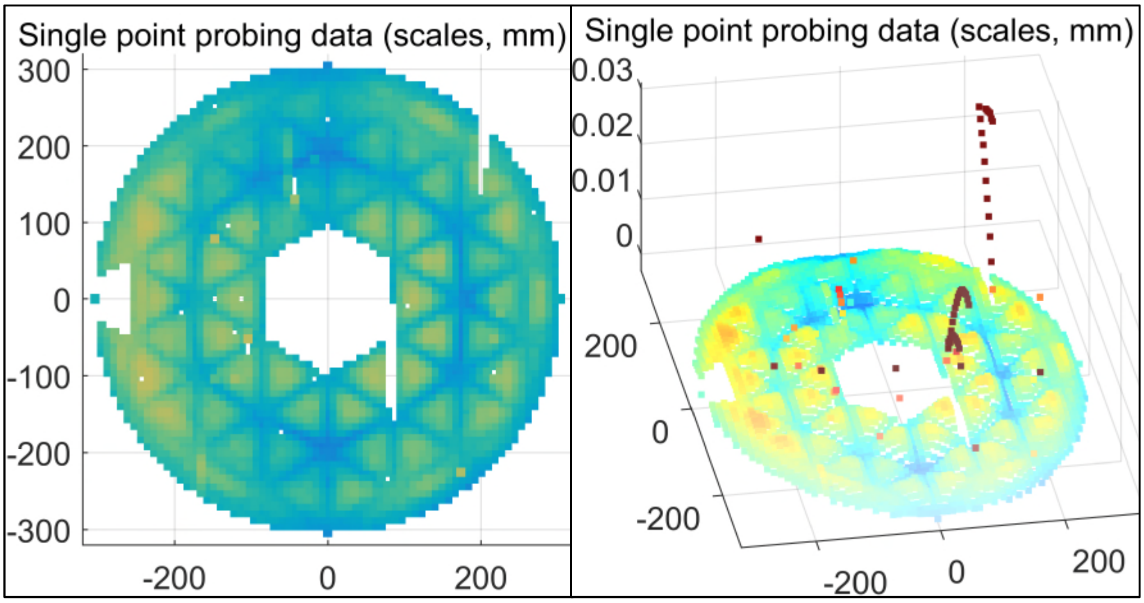

Single point probing data (scales, mm)
Single point probing data (scales, mm)

Figure 5-4: High density single point probing map; left: 2D, right: 3D

The measurement in Figure 5-4 represents about 3 hours CMM measuring time, and can be fitted to spherical or ellipsoidal form within CMM operating software, but not easily to advanced freeform shapes. These data are shown with the fitted sphere removed, so that only departure from spherical shape is plotted. Although noise filtration is available within CMM programs, the effects on form measurement accuracy are unclear. The fitting of orthogonal polynomials, which are commonly used with interferometers but not available on CMMs, is moreover prone to significant inaccuracy using discontinuous pupils with limited resolution. The trade-off between speed, data volume and feature loss due to noise filtration is not advantageous to the measurement of large freeforms – in order to gain enough spatial resolution to mitigate the effects of noise and filtration on feature measurement accuracy, much longer measurement times are required. This raises the issue of stability of measurement equipment, method and the freeform surface itself over this period of time.

Some critical measurement performance parameters relate to the handling of what are termed mid-spatial frequencies, which for the relevant stage in optics manufacturing process chains have spatial wavelengths between around 2-50 mm. These are the spatial frequencies which are difficult to control with the final figuring techniques such as Ion-Beam Figuring (IBF), Reactive Atom Plasma (RAP) figuring or computer numerical controlled (CNC) polishing, such as the techniques developed by Zeeko and others. For all these processes, this difficulty is due to removal function shape and size, but crucially, these frequencies may be produced as artefacts of (and potentially controlled by) the prior manufacturing process chain steps such as BoX grinding (see section 4.3). In order to measure features within this mid-spatial range with acceptable accuracy, the measurement technique needs to have a lateral spatial resolution capability down the lower limit of the mid-spatial range (around 2 mm). In a full-aperture (full surface width) measurement, this is normally the province of interferometry. Unfortunately, at this stage of manufacture, interferometry is unavailable, due to the low strength and nature of the optical reflection from the functional surface. This level of performance is also realistically not possible with single point probing-based measurement. In the data of Figure 5-4 – a 3-hour measurement, despite the conventional wisdom that this is the highest accuracy mode for CMM measurement, an 8.5 mm measurement spatial resolution is inadequate, because of the mid-spatial wavelength lower limit at around 2 mm. There is a square law with measurement time as a function of resolution (or as a function of surface dimension) and the issue of missing data and lost feature measurement, which is clear in the figure, is tough to resolve. At this resolution, and under the laboratory conditions encountered, the proportion of lost data is just under 2% (as shown particularly in Figure 5-4 left). A 2 mm resolution on this 600 mm diameter (0.25 m^2 diameter) optic would take 16 times longer to achieve, and the same performance on a 1.5 metre class 2 m^2 diameter optic would impose another 8-fold increase in measurement time – well over a week of continuous measurement in total. A higher speed, higher resolution technique that has a similar accuracy capability is certainly called for.

5.3 Proposed scanning metrology technique

In order to acquire high density data with a CMM to meet the mid-spatial requirement, scanning will be required. Scanning is capable of acquiring contact measurement data at high rate with a high lateral and vertical resolution. The data rate may approach 1000 times higher than for single point measurement, although at very high data rates, handling the associated large volumes of data becomes prohibitively difficult – even with access to higher powered computers operating the measurement.

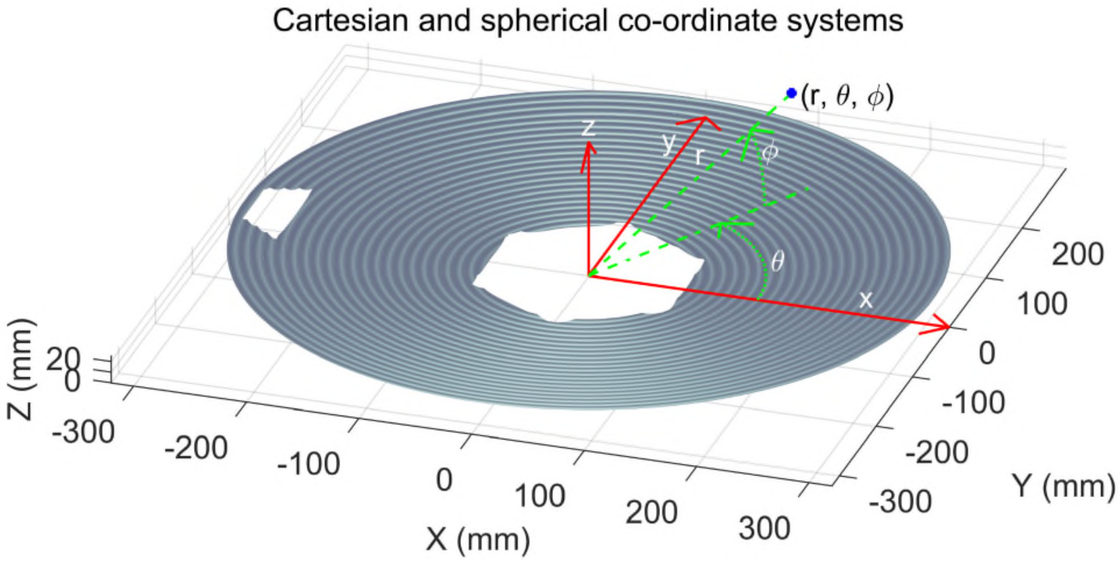


Figure 5-5: Spherical co-ordinate system – *θ* is azimuth, *ϕ* is elevation

The measurements detailed in this chapter are carried out and represented in Cartesian co-ordinates, although occasional reference is made to spherical co-ordinates. The proposed technique is based on linear scan measurements, which are nominally parallel to the x and y axes and approximately equally spaced when projected onto the x-y plane on the CMM (Figure 4-1) and onto to the (parallel) x-y plane of the freeform surface's co-ordinate system as in Figure 5-5. Scans nominally parallel to x and y, for low slope surfaces, ensure probe deflection is effectively limited to two directions, primarily in one, and that motion reversals are minimised.

In the scanning technique used, the scanning is semi-closed loop, which is to say that a scan vector is pre-programmed. Once scanning has commenced, closed loop control is maintained over the probe head deflection (by adjusting the x, y & z axes of the CMM) so that near-constant contact force is applied to the surface through the stylus. Depending on the surface geometry, this may entail departure from the pre-programmed scan trajectory.

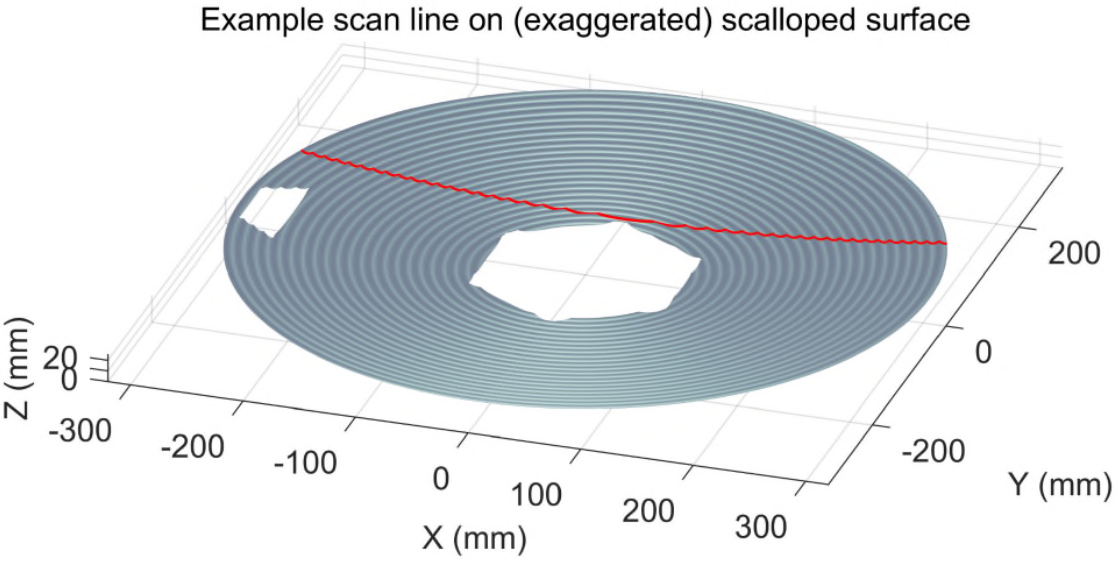


Figure 5-6: (Simulated) scan line on surface

In the example of Figure 5-6 the pre-programmed scan trajectory would include the gradual curve, but not the cusps on the path; the cusps would be handled by the closed-loop force-controlled surface following.

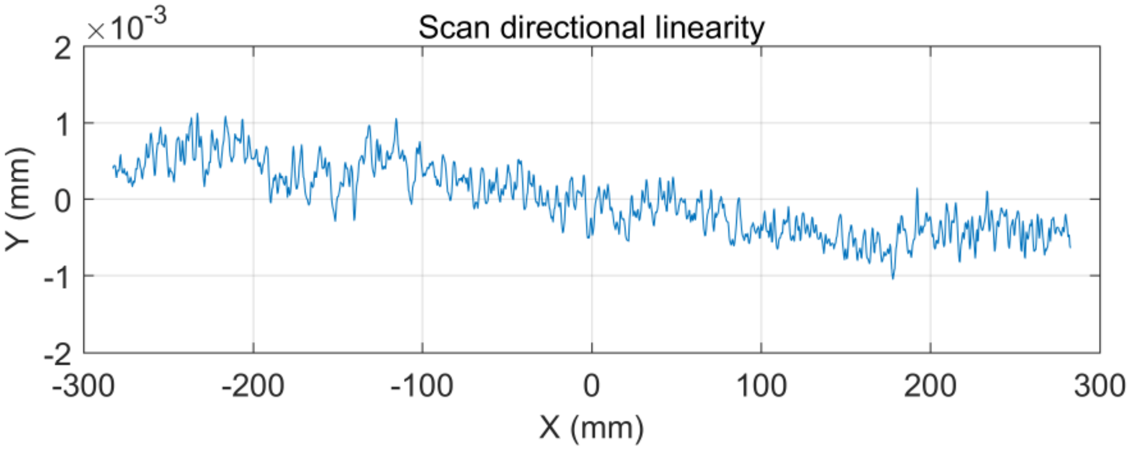


Figure 5-7: scan speed and direction uncertainty

This strategy can result in paths which are not entirely straight or smooth; this can give an effect of slight variations in scan speed and linearity, as shown in Figure 5-7 (on a magnified scale) looking at the x-y variation of a nominally x-direction scan. A view of the same scan in the x-z plane gives the more expected path shape, as in Figure 5-8. The scallops are present in the data, but on a scale large enough to see the curvature, their cusps are too small (at just 1 µm in height) to be discernible.

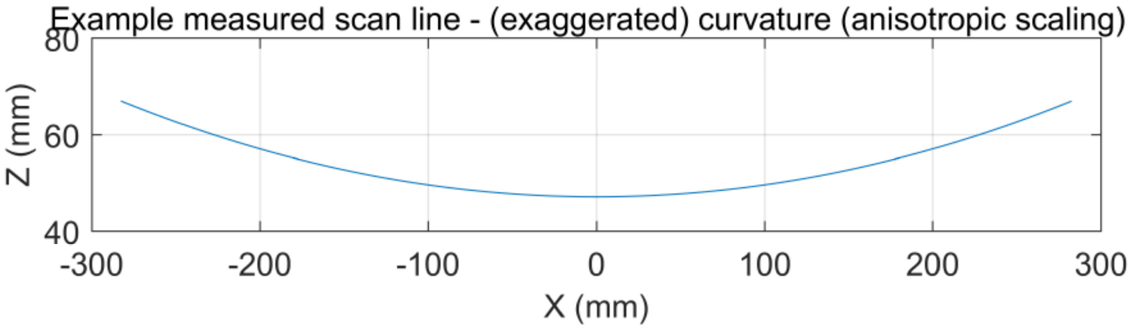


Figure 5-8: Scan line data

The overall surface scan data are assembled from multiple parallel scans in the x-direction and the same in the y-direction – two separate raster patterns. All the scans in one directional set are performed in the same direction; if reverse direction scans are used, these are done in a separate phase. This grouping is maintained in case reverse direction scans require separate treatment in data processing, and to give consistency of conditions.

The overall strategy for the scanning data processing algorithm is in three phases:

Phase 1, filter out contact errors; this phase does not intrinsically alter any values; the principle is that some individual points (a low percentage, typically 2%, set by threshold value of +ve deviation rather than a proportional population pruning target) are removed and then replaced with values interpolated from neighbouring points in the same data set.

Phase 2, some assessment and potentially adjustment is made for differences between or within directional measurement sets to account for distortions, thermally-induced or otherwise.

Phase 3, fitting Zernike polynomials to get a geometrical description which relates both to optical parameters, but also to adjustable parameters of machining and setup, such as distortion due to machining forces on the face cells, centring error, tilt etc., all of which can then be independently corrected.

Figure 5-9 shows a number of parallel scans in actual scan co-ordinates. A full surface map has over 200 scan lines, so this represents a small sub-set of an entire scan.

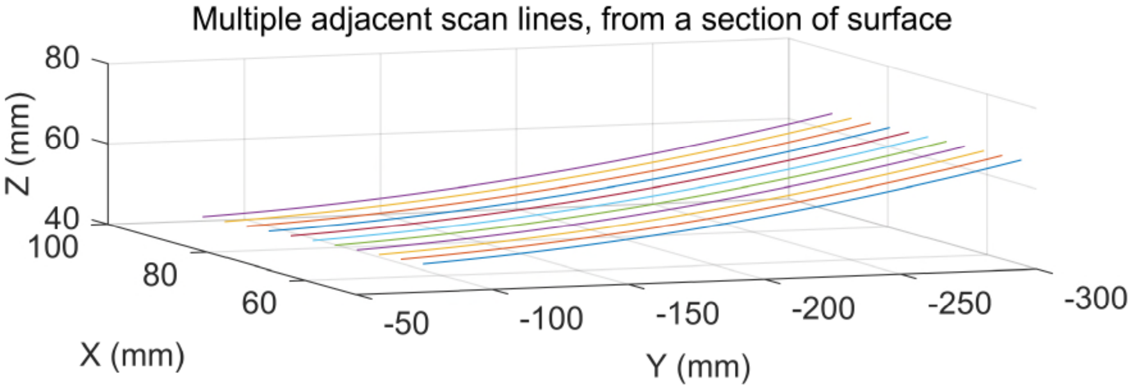


Figure 5-9: Scan lines showing actual data

5.3.1 *Scan data presentation*

Whereas figures up to this point have largely shown scan data with z height represented (as in Figure 5-9) the remainder of this chapter will show scan data only with deviation from best fit sphere, which is the nominal shape.

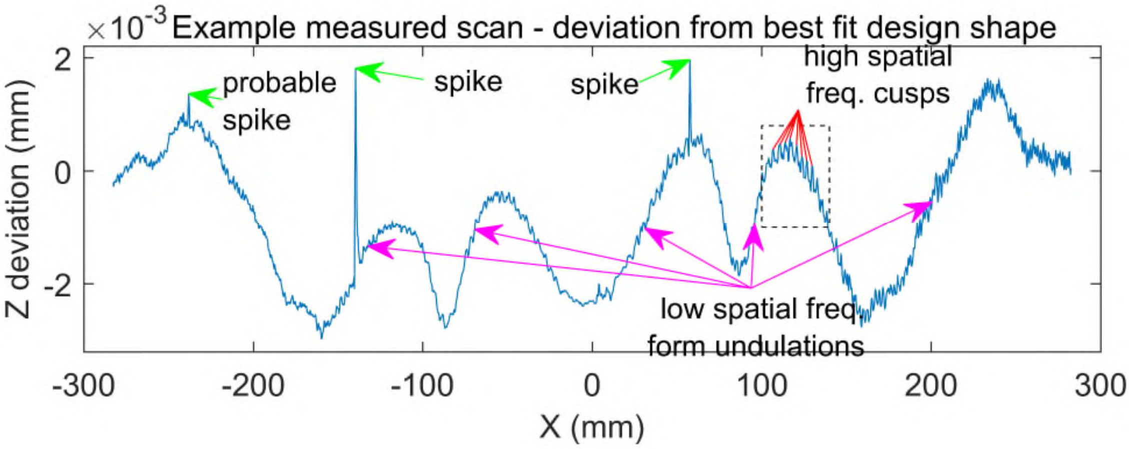


Figure 5-10: Scan line data showing deviation from best fit – the rectangular area is reproduced in Figure 5-11

Individual linear scan measurements encompass several data phenomena:

- Wide band measurement noise
- Discrete measurement spikes
- Low spatial frequency form variations

- (depending on the machining mode) high spatial frequency machining pitch marks
- These phenomena are shown in Figure 5-10 from real scan data. Figure 5-11 shows an expanded view of grinding scallops. At this spatial scan density, there is sufficient scan resolution to avoid aliasing the grinding scallops. This is important to avoid confusion with machining form error artefacts that might otherwise need correction.

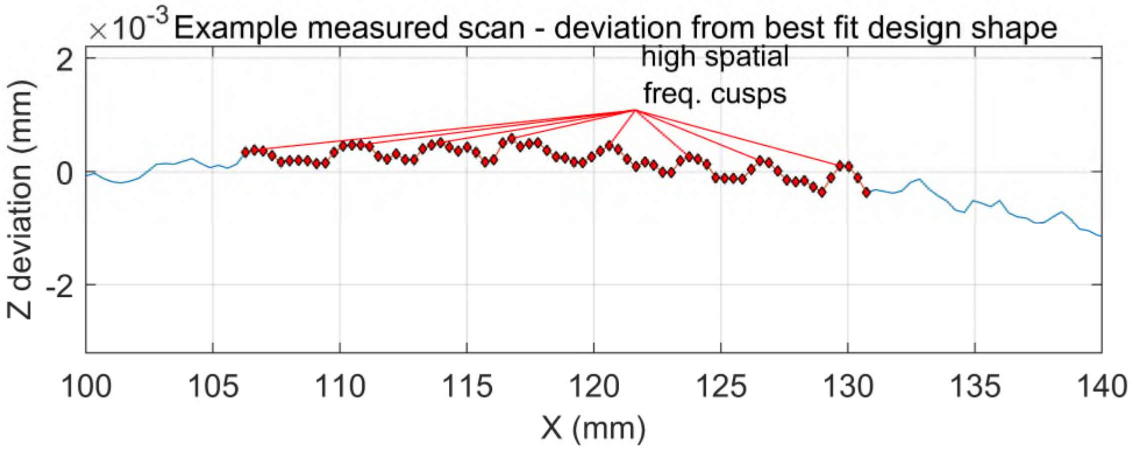


Figure 5-11: Expanded view from Figure 5-10

5.4 Surface contact dynamics

In addition to single-point spikes, the contact probe can pick up debris which can stay adhered – at least for a time. This leads to discrete sections of elevated data – elevated at least by a few microns, where the particle is interposed between stylus and surface. Moreover, such elevated sections can extend to the edge of the data set.

Individual spikes can be detected in one of several simple and familiar ways but Figure 5-12 depicts particularly problematic cases where there are clearly scan defects that are:

a) connected to the edge
b) in contiguous chains along the scan direction
c) with points in several adjacent parallel scans affected

In the case shown in Figure 5-12, there is a region where all three cases exist together. A particularly novel strategy is required both to identify and essay a correction for these data errors.

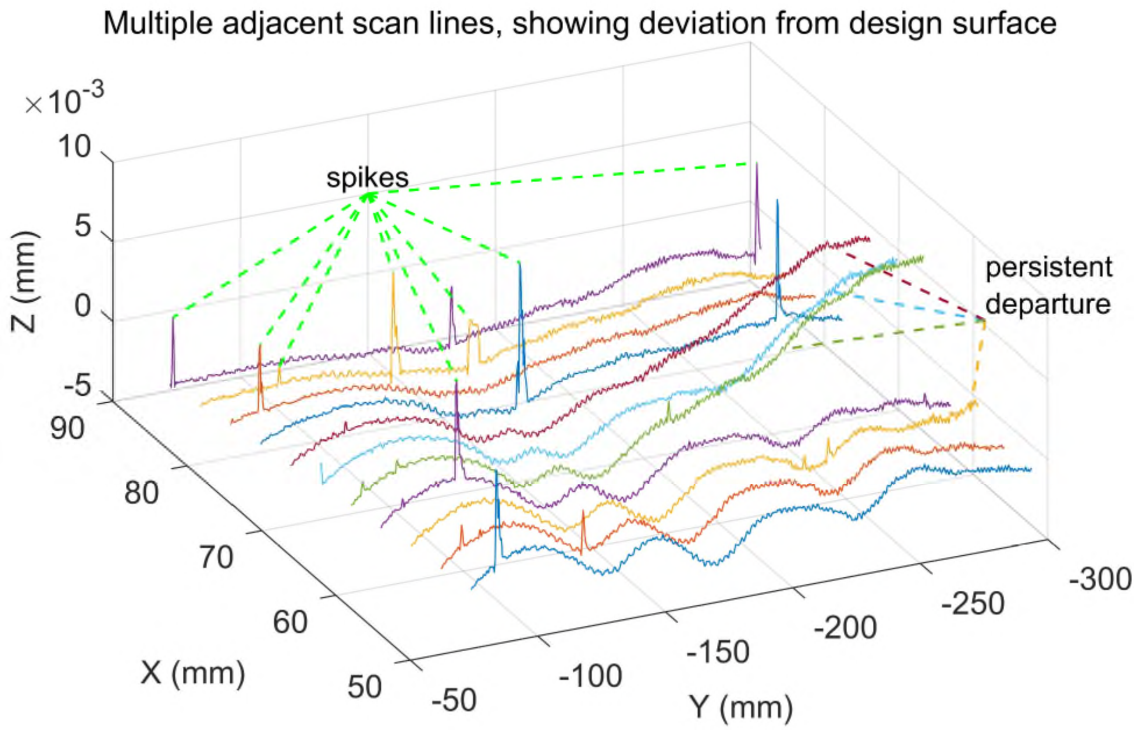


Figure 5-12: Scan lines (from Figure 5-9) showing deviation – i.e. gross figure removed

The contact errors have particular characteristics arising out of the engagement of a contaminant particle and its adhesion either to the surface being measured or the contact probe stylus tip. Both surface and stylus are cleaned before a measurement process, although laboratory air contains numbers of dust particles per unit volume. In the still widely recognised but long obsolete US Federal Standard 209E, room air was described as containing archetypically 1,000,000 particles of ≥ 0.5 μm minimum dimension per cubic foot; the lowest class (Class 100,000) of cleanroom under that standard correspondingly had 100,000/ft^3; Class 10,000, 10,000/ft^3 etc. In its superseding ISO standard 14644-1, these latter two specifications would be equivalent with ISO Classes 8 and 7 respectively, and these are typical of the operating environments of precision metrology CMMs [199], such as the one used in this study. Some of these airborne particles inevitably settle on the surface during a measurement operation, which for some large surfaces may take many hours.

5.4.1 Contact event, case 1

At the time a contaminant particle on the surface is encountered by a scanning probe stylus tip, among the possibilities it may be unmoved by the impact, pushed aside or otherwise remain on the surface. In this case, there will be an instantaneous upwards (+ve) deflection of the probe, followed by a more gradual return to contact on the surface. The gradual return will be governed by the dynamic response of the probe system. Typically, this is described by an underdamped second order differential equation and gives rise to damped ‘ringing’ – a decaying oscillation at a frequency characteristic of the CMM’s probe head. In the case of the Leitz LSP-S2 probe head used in this study, this is a few Hz.

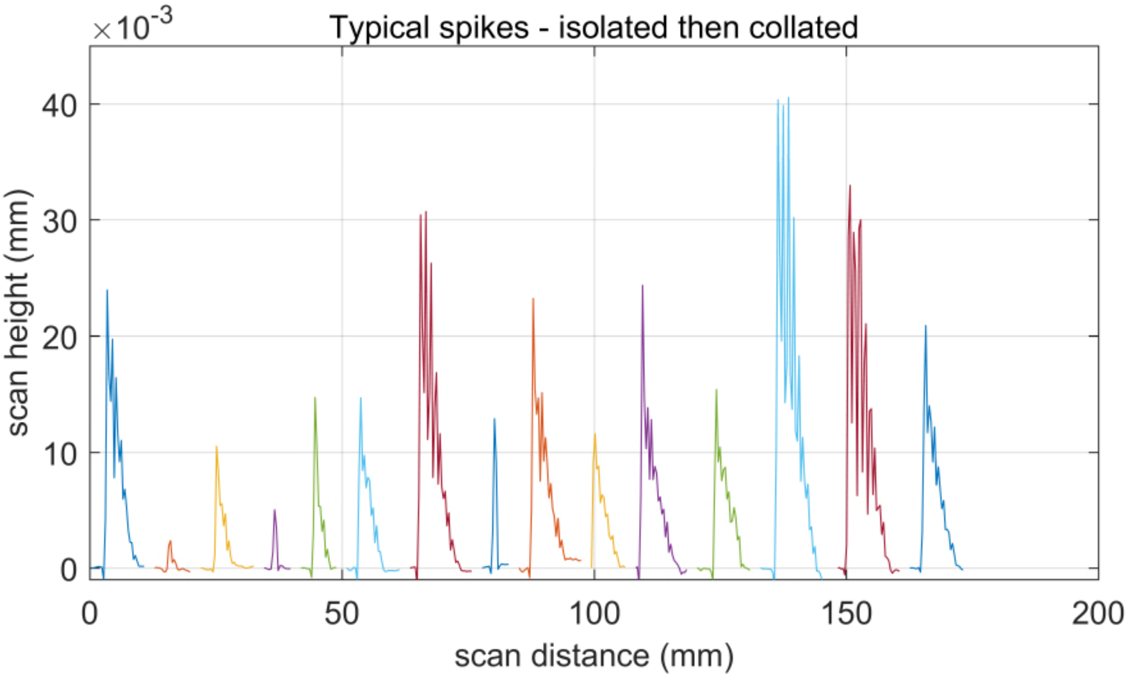


Figure 5-13: Spiked contact errors – showing the consequence of particle adhesion to surface

Figure 5-13 shows a collation of several such typical contact events. In this graph, scan height is on the vertical axis and scan length on the horizontal. Contact events have been height offset adjusted so that all start at zero – for a realistic height comparison. Scanning speed is 10 mm/sec and there are approximately 30 points recorded per second – an equivalent lateral resolution of around 1/3 mm. This implies that the precise moment of particulate impact is unlikely to be captured, meaning that the measurement (and therefore probably Figure 5-13) typically underestimates maximum probe excursions.

Figure 5-14 shows (on the left) a magnified view of the 7th spike from Figure 5-13 and (on the right) spectral data for the spikes in Figure 5-13, taken as a group. It can be seen from these that the probe ringing frequency is in the region of 10 Hz, but that the sampling rate of 3 Hz means that the oscillation amplitude is underestimated, which also explains the relative imprecision of the FFT frequency estimate. The response in the most severe of these contact events (the 13th or 14th in Figure 5-13) is damped out within 12 mm of scan travel, but this will render up to 50 scan points unusable, since there is no surface contact during this distance of scan.

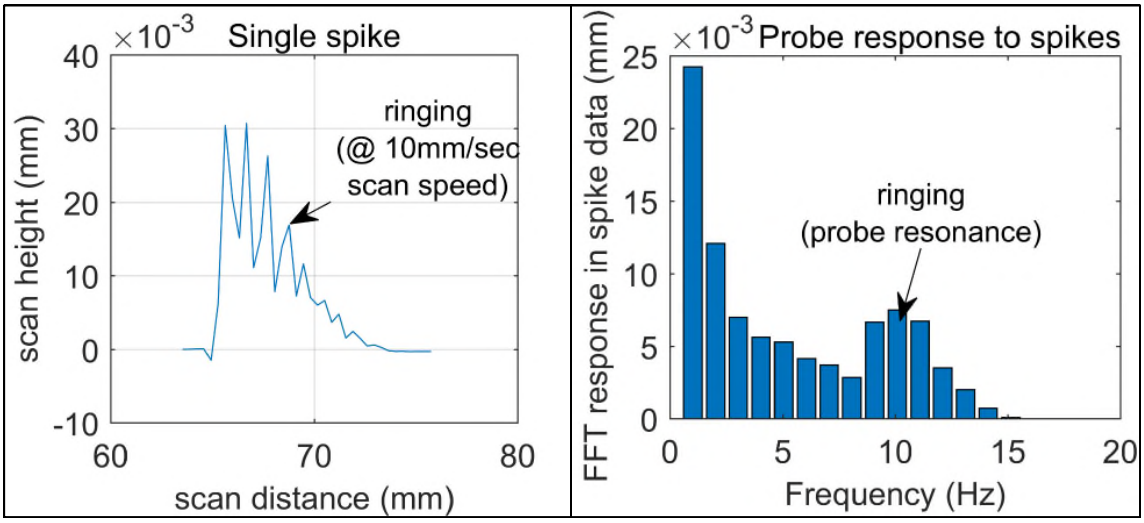


Figure 5-14: Spike characteristics – left: time domain, right: frequency domain

5.4.2 Contact event, case 2

At the time a contaminant particle on the surface is encountered by a scanning probe stylus tip it may, as an alternative to Contact event, case 1, become attached to the probe stylus. In this case, there is likely to be an instantaneous upwards (+ve) deflection of the probe and at some level, the upwards deflection will be maintained for a time, for as long as the particle is interposed between probe stylus and surface, giving an effective increase in measured height.

Figure 5-15 shows a collation of such events – six are shown. In the first three, there are two steps, one when the particle is first interposed and the second where it suddenly ceases to influence the contact, either redeposited on the surface or otherwise displaced. In the latter three cases, there is only a single sudden step, then a gradual return to normal contact, but over a much longer period than for the spikes of case 1 (5.4.1) – perhaps 100 mm of scanning (10 seconds) and without the ringing. This is due to the particular geometric condition of a low slope surface: as the scanning progresses, the angle of contact on the spherical stylus tip gradually changes.

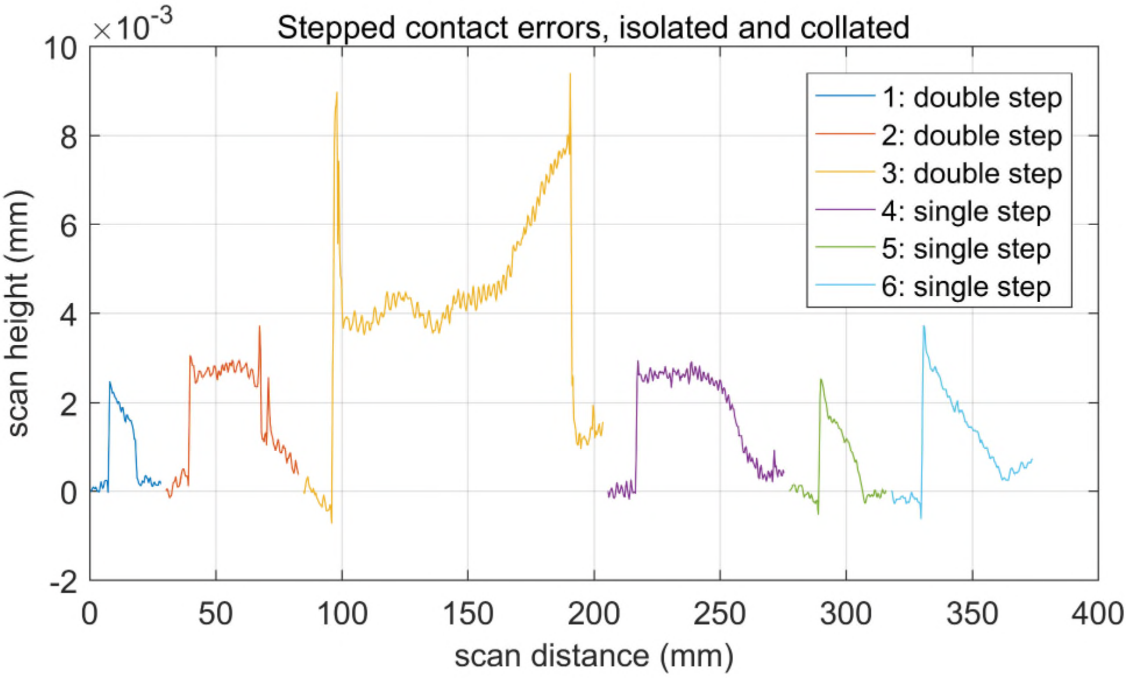


Figure 5-15: Stepped contact errors – showing stylus particle adhesion consequences

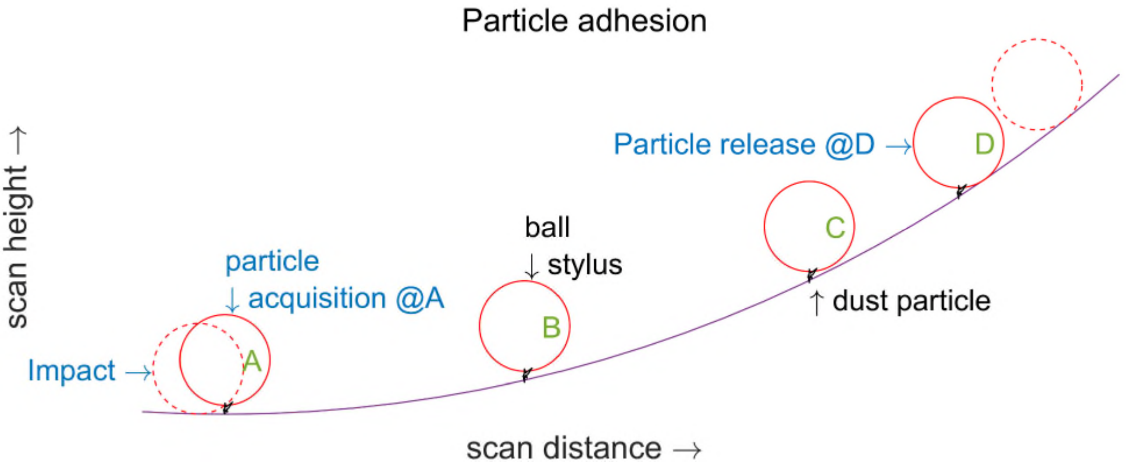


Figure 5-16: Stepped contact error – particle adhesion to stylus

As shown in Figure 5-16, this will change the effective depth of the particle and eventually it will cease to influence the contact. All the contact errors are positive - in other words the CMM stylus can pick up adherent particles increasing its effective radius, and can encounter particles on the measurement surface, or it can bounce away from the surface, but it cannot under ordinary circumstances penetrate the surface. In fact, the author, among others, has reported ductile materials being impacted or scratched by CMM styli, with a consequent negative contact error (in the order of 10-100 nm) [8], although direct evidence of that has not been sought in this study. This should be reflected in the adopted algorithmic strategy for detection and correction of errors. In order to detect locally positive deviations, comparisons with neighbouring points on either side of the scan should be made, rather than applying a threshold to a unidirectional gradient, recording the rate (or value) of height change in the linear scan data itself. Comparisons moreover should deliberately be made in a direction perpendicular to the scan motion, since comparisons along a scan exhibiting persistent positive departure from the surface would otherwise fail to detect an error. Lastly, given that multiple adjacent scans can be affected (as in Figure 5-12 and Figure 5-22) comparisons with several parallel scans are required, otherwise it's possible no positive deviation will be observed.

5.5 Surface Data Storage and Representation

Surface data are stored in two distinct ways within this study: as scan data or as a point cloud interpolant.

5.5.1 Raw scan data

Storing scan data, for each scan line there are four linear arrays of data, three representing x, y and z co-ordinates of the scan contact points on the surface, and the fourth representing the deviation from the design or numerically fitted surface. It is this deviation data (rather than the z co-ordinate) which is depicted on the surface maps in this part of this manuscript. For a scan line with n points (x_i, y_i, y_i) with d_i, the representation is as in the equations (5-1)

$$\begin{aligned} x_i &= [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_n] \\ y_i &= [y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_n] \\ z_i &= [z_1 \quad z_2 \quad z_3 \quad \cdots \quad z_n] \\ d_i &= [d_1 \quad d_2 \quad d_3 \quad \cdots \quad d_n] \end{aligned} \tag{5-1}$$

There are typically 250 scan lines (the workpiece aperture breaks the central lines into pairs) with up to 1600 points or more, per line, but the number of scan points varies from line to line, on account of varying line length, due to surface geometry and the issues referenced in Figure 5-7. A single surface map contains scan points therefore on an imperfectly rectangular grid, with scan lines separated by around 3.3 mm and scan points within each line spaced at approximately 0.33 mm. This gives a point density of 90 per sq. cm or 900,000/m^2. With a surface area of approx. 0.25 m^2, the optic in Figure 5-2/Figure 5-22 is scanned in one direction with just under a quarter of a million points. The scan speed is set to keep data loss through the mechanisms of section 5.4 to a minimum, whilst keeping total scan time to a reasonable duration (< 24 hours), both for practicality, and to limit non-repeatability through thermal effects on workpiece and measuring system.

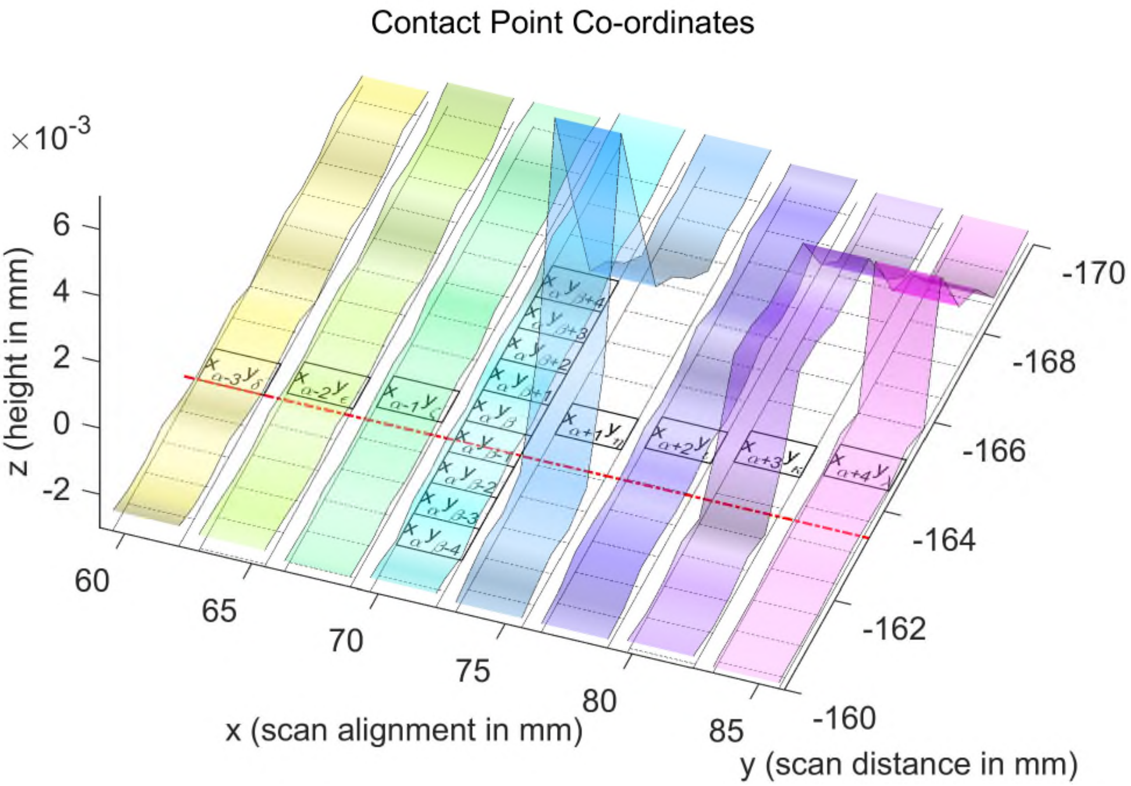


Figure 5-17: Scan co-ordinate alignment

Although scans are made with continuous motion, scan data are composed of discrete points as shown in Figure 5-17, which depicts actual scan data. Each rectangle, some of which are labelled, represents a single actual scan point with an (x, y, d) triple – x & y are represented by the centre of the rectangle's top-line, and d by the height of the coloured band above. The parallel scans are taken at nominally equally spaced x-coordinates: $x_{\alpha-3}$, $x_{\alpha-2,}$ $x_{\alpha-1}$, x_{α}, $x_{\alpha+1}$ etc.; for this data set the line separation (in x) is approximately 3.4 mm. Points within one scan line are also nominally equally spaced: $y_{\beta-3}$, $y_{\beta-2,}$ $y_{\beta-1}$, y_{β}, $y_{\beta+1}$ etc.; for this data set the point spacing (along the scan line) is approximately 350 µm. Whilst a scan is being taken, individual co-ordinate triples are recorded for each point. These correctly capture the stylus contact location at an instant (corrected for probe deflection) but, depending on a number of factors, including surface topography and scan start position, points cannot be guaranteed to be precisely equally spaced, nor can points in adjacent scans be guaranteed to be aligned. Figure 5-18 shows the level of consistency in scan line and point spacing.

The red dash-dot line (in Figure 5-17) of equal y-coordinate aligns only to one of the parallel scans' acquisition y-positions, thereby showing this effect; the others are not aligned. This is because there is a non-deterministic shift in the alignment of discrete points in adjacent scans, although their y-coordinate position is correctly recorded. This factor is significant in the detection and correction of contact errors, since it is desired to reject as few points as possible comparisons for the same x-value must be made.

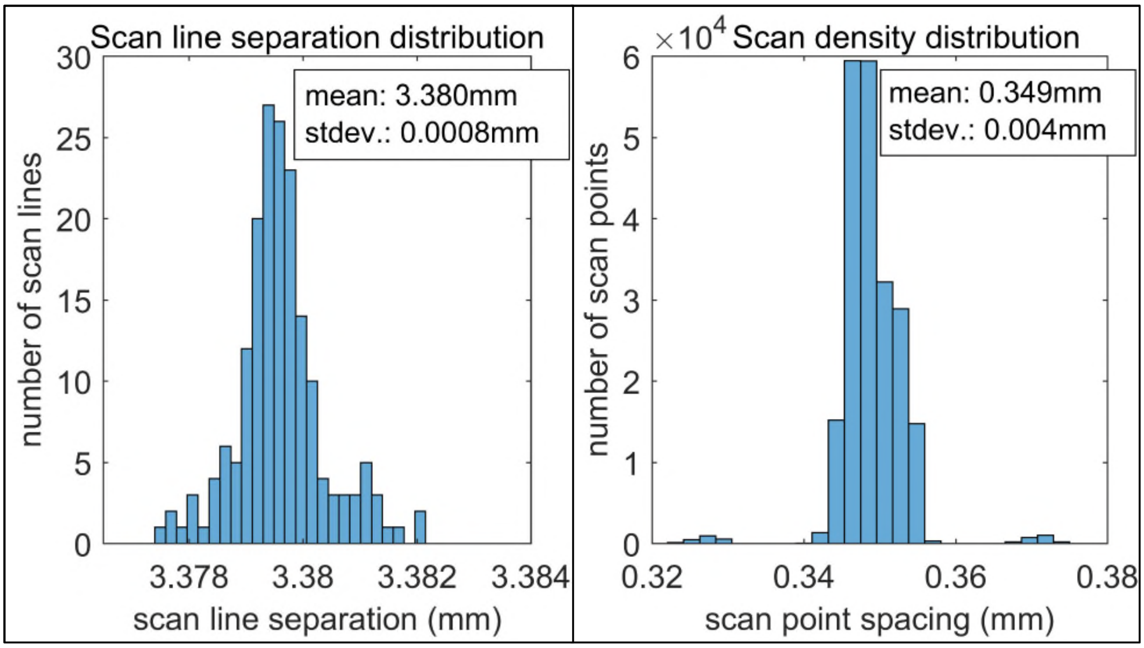


Figure 5-18: Consistency of (left) mean scan line separation and (right) scan point spacing

Therefore, scan lines cannot be represented in a (filled) rectangular matrix, so for a set of m scan lines (numbered j) made in the x-direction (with nominally set y co-ordinate) the full representation is as given in Equations (5-2) where the n points per scan line may be a different n (n_j) for each of the m scan lines.

$$x_{i,j} = \begin{cases} x_{i,1} = [x_{1,1} \quad x_{2,1} \quad x_{3,1} \quad \cdots \quad x_{n_1,1}] \\ x_{i,2} = [x_{1,2} \quad x_{2,2} \quad \cdots \quad x_{n_2,2}] \\ x_{i,3} = [x_{1,3} \quad x_{2,3} \quad x_{3,3} \quad x_{4,3} \quad \cdots \quad x_{n_3,3}] \\ \cdots \\ x_{i,m} = [x_{1,m} \quad x_{2,m} \quad x_{3,m} \quad \cdots \quad x_{n_m,m}] \end{cases} \tag{5-2}$$

There are similar equations for y, z and d. For the x-direction scan lines x values span the surface, whereas y values within each line are very close (nominally constant). For y-direction scan lines, the opposite is the case, where x values within each line are very close and y values span the surface. The linear arrays for (e.g.) x for all the scan lines can be concatenated to make single vectors for efficient storage as in equation (5-3) and similarly for y, z and d.

$$x = \begin{matrix}[x_{1,1} \quad x_{2,1} \quad x_{3,1} \quad \cdots \quad x_{n_1,1} \quad x_{1,2} \quad x_{2,2} \quad \cdots \quad x_{n_2,2} \quad x_{1,3} \quad x_{2,3} \\ x_{3,3} \quad x_{4,3} \quad \cdots \quad x_{n_3,3} \quad \cdots \quad x_{1,m} \quad x_{2,m} \quad x_{3,m} \quad \cdots \quad x_{n_m,m}]\end{matrix} \tag{5-3}$$

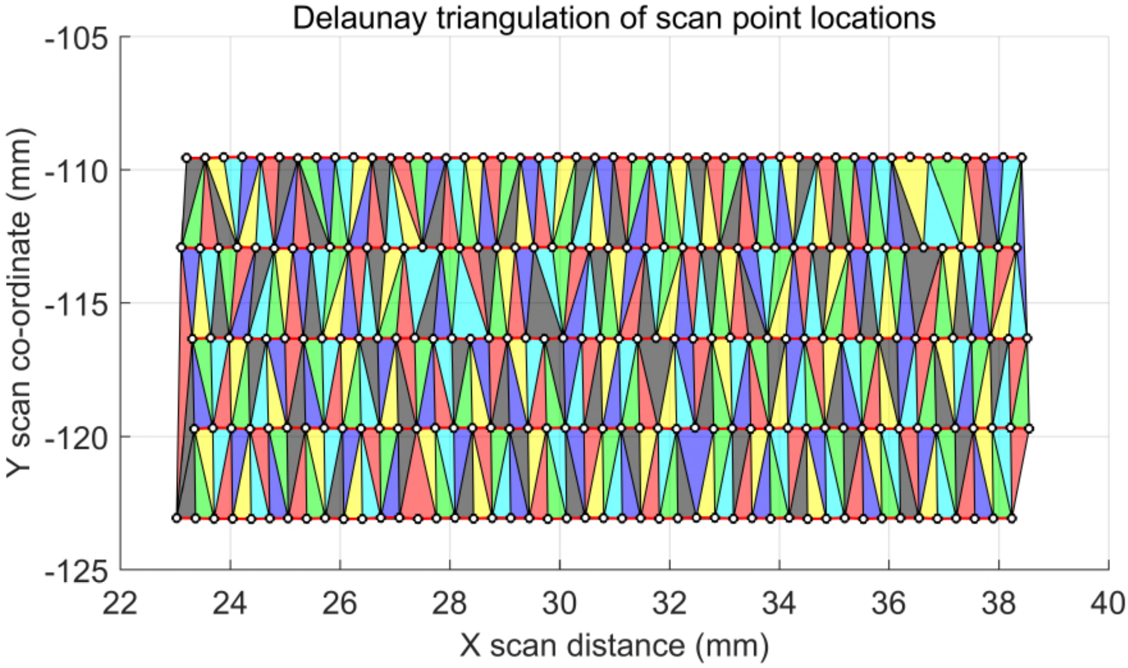


Figure 5-19: Delaunay triangulation of realistic scan point locations – small segment of surface (visualisation colours are not significant)

5.5.2 Point cloud interpolant

The second surface data storage scheme is a fully flexible point cloud configured for a Delaunay triangulation so that the methods of Shepard [184] and Sibson

[188] (for interpolation/extrapolation) as well as the more advanced manipulations given below can be easily accomplished.

A Delaunay triangulation maximizes the minimum angle of each triangle by selecting groups of three points accordingly, to form the vertices of the triangles in the net (see Figure 5-19). This avoids as much as possible sliver triangles, although given the relative spacing of scan points and scan lines here, high aspect ratio triangles are hard to avoid. Sibson's natural neighbour interpolation is then applied when height/deviation values are required for points intermediate between actual scan points.

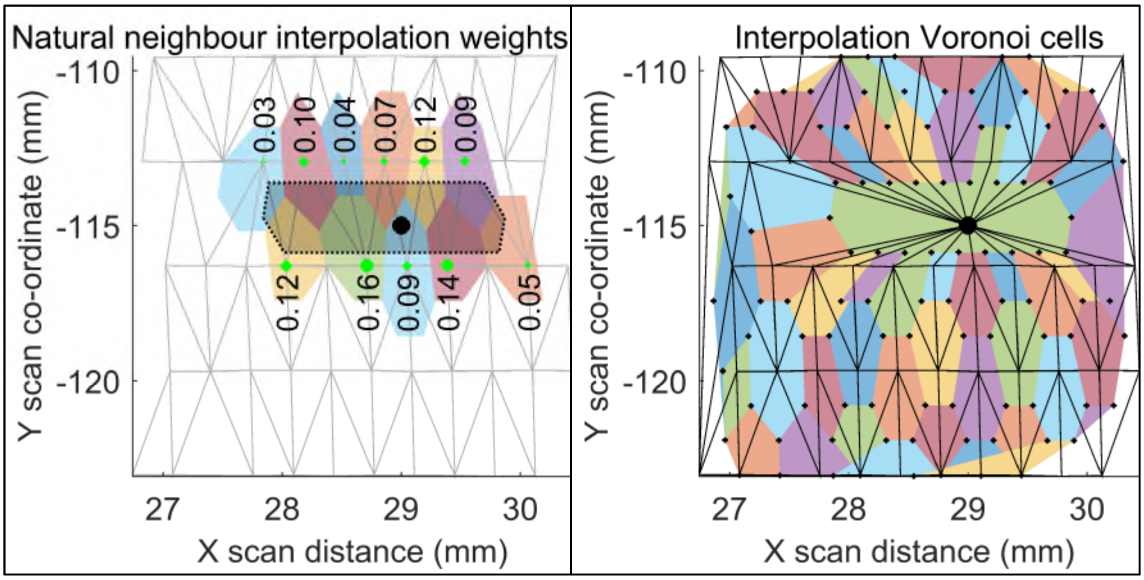


Figure 5-20: Natural neighbour interpolation

Figure 5-20 shows the natural neighbour interpolation technique on a subset of scan data locations from Figure 5-19. On the left is shown the initial triangulation net. An interpolation is desired at the black point. On the right is shown a revised triangulation net, as it would appear if the new point was introduced to the triangulation set. The coloured patches on the right are called Voronoi cells and these are polygons with vertices made of the centroids of the triangles in the net – conversely this implies that our interpolation points are at the centres of the Voronoi cells. Returning to the left diagram, a cluster of Voronoi cells from the original net are shown surrounding the new point. The proportion of each cell in the cluster which would be lost to the new Voronoi cell around the new point dictates the weight of each related scan point that is lent to the interpolation value, and these weights are indicated numerically on the diagram and with the size of the green dot at the scan point location.

The underlying function (d, the deviation from spherical fit) can then be interpolated to give a continuous function D using the numerical weights w as in equation (5-4), summed over n relevant scan point neighbours. The 'i'-ordering of the interpolant neighbours here does not correspond in particular to the (i, j) ordering of the scan points in the earlier equations.

$$D(x,y) = \sum_{i=1}^{n} w_i(x,y)d(x_i,y_i) \tag{5-4}$$

By using the natural neighbour interpolation technique, a smoother interpolation is achieved than with alternatives, although it is computationally more expensive. Extrapolation cannot be achieved with natural neighbour technique, so *nearest* neighbours are used for extrapolation instead, where extrapolation simply takes the value of the nearest actual point in the interpolant data set. Extrapolation is performed at or close to the edges of the workpiece where interrogated points lie close to or outside the *convex hull* of the data set. If an interpolation is desired for a point outside the convex hull, extrapolation must be used. This need can occur in particular where there are missing data due to contact errors (see section 5.4).

5.5.3 Surface display

Conversions between the pseudo-rectilinear storage scheme, as used for the raw scan - and unordered point cloud interpolant storage schemes are possible – and in fact necessary; in the first case to enter new scan data into an interpolant set, and in the second case in particular for display.

Figure 5-21 and Figure 5-22 (the difference between them is solely resolution) use a colour map to represent the third axis - surface deviation. The colour map used is Matlab's Parula, which is designed (having a gradual dichromatic colour transition) to eliminate the incorrect perception of differences in data due to hue differentials, which occur in any rainbow style mapping; Parula is perceptually relatively neutral throughout its scale, avoids several colour-blindness issues and maps to a linear progression in greyscale [208].

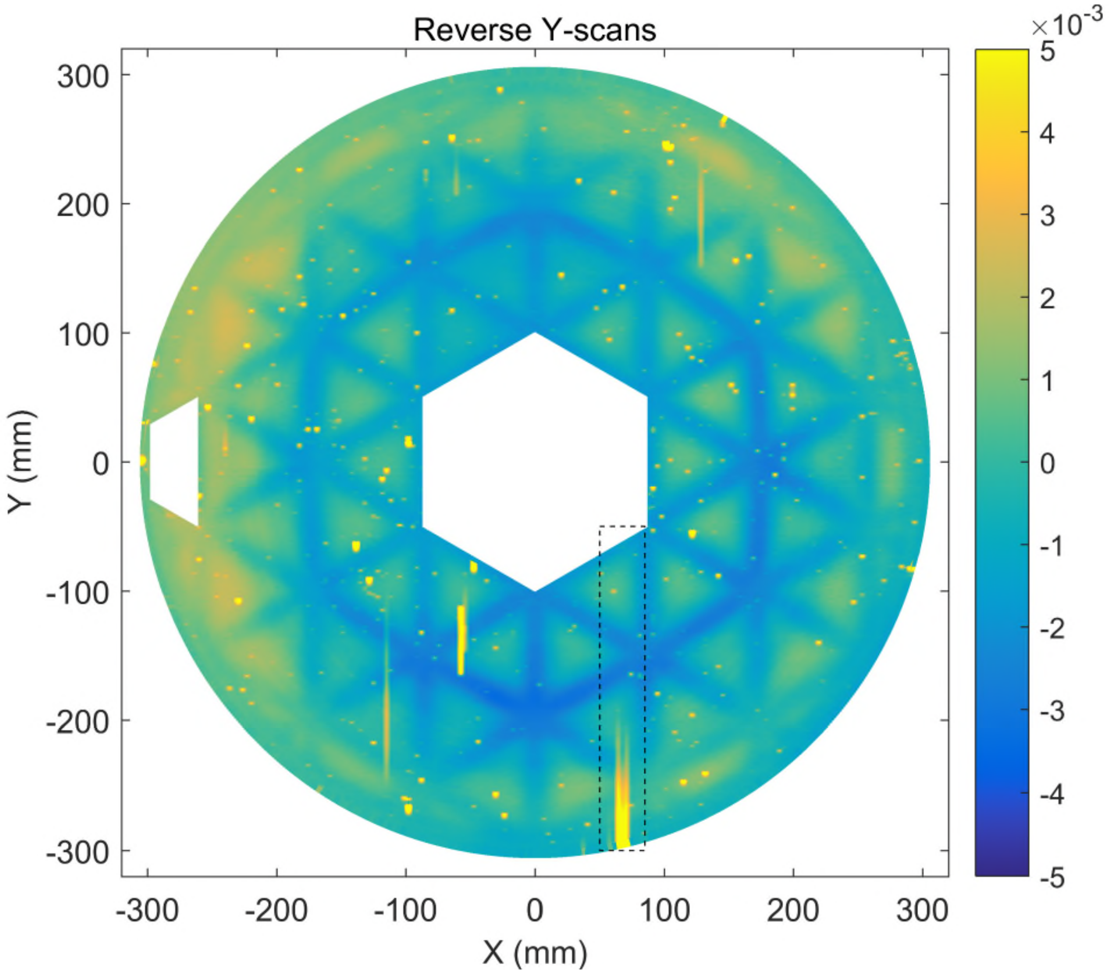


Figure 5-21: High interpolation resolution false colour diagram of deviation from design form – entire surface shown; deviation in Z indicated by colour (bar on right gives scale). The marked rectangle is the area in Figure 5-9 and Figure 5-12

Parula's perceptual neutrality is particularly important for visualisation of subtle differences and is therefore appropriate for presenting this application – although it doesn't affect automated measurement in any quantitative sense. In order to prepare the data for display, a rectangular grid of co-ordinates is selected (in Figure 5-21 this is 1280x1280 pixels – corresponding to 0.5 mm resolution on the surface).

Interpolated values for surface deviation are obtained for each of those 1.6 Megapixels using the natural neighbour interpolation presented in Section 5.5.2. Approximately 240,000 (~16%) of these are beyond the edges of the surface, represented by NaN (see Section 5.6, page 78) and not displayed. These points are either outside the circular edge or inside the apertures in the surface.

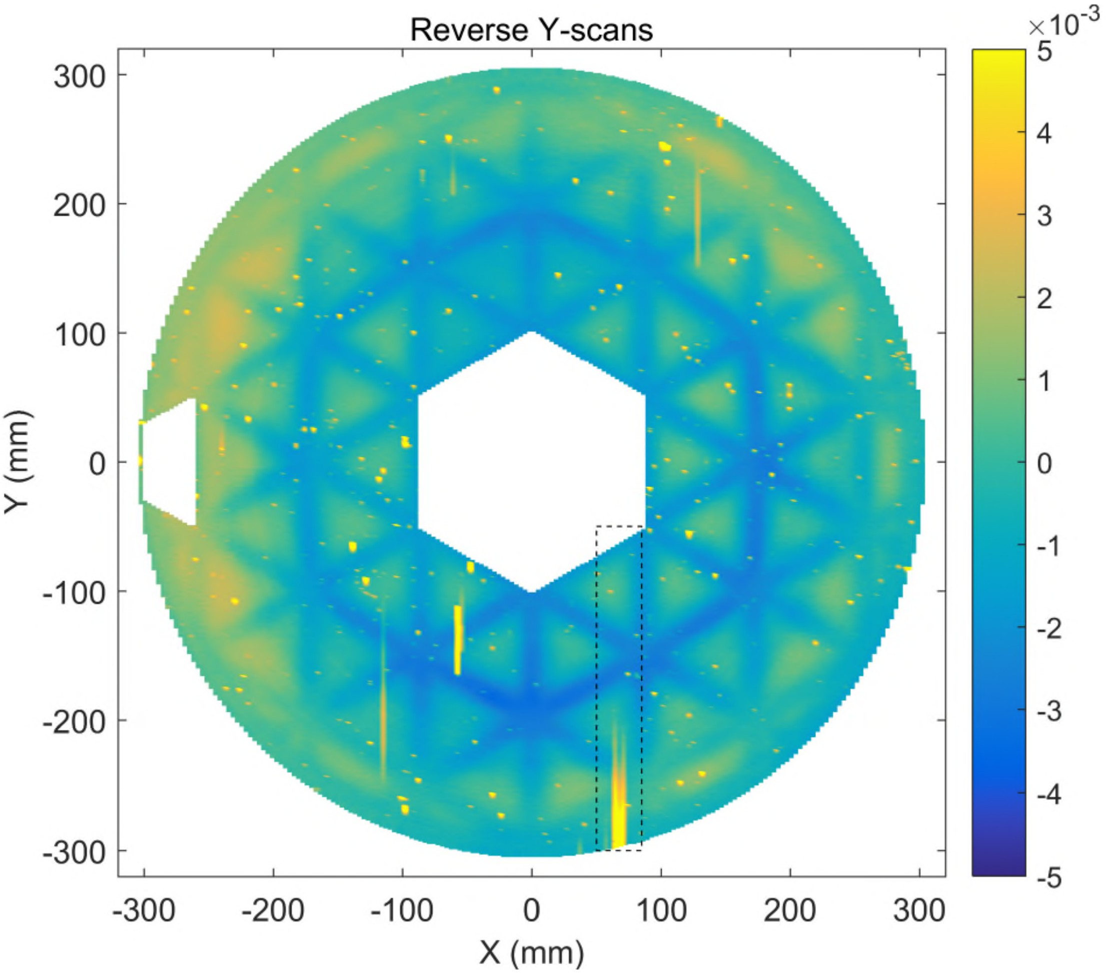


Figure 5-22: False colour diagram of deviation from design form – entire surface shown; deviation in Z is indicated by colour (the bar on the right gives the scale). The marked rectangle is the area shown in Figure 5-9 and Figure 5-12 – natural grid interpolation, along scan lines

Figure 5-22 shows the same image but at a resolution of 192x1920 pixels, which is more natural because it relates to the scan's actual resolution. In either case, the underlying information is the same – in the latter case the 192 vertical lines (horizontal resolution) correspond to the scan lines and the 1920 horizontal lines (vertical resolution) to the scan points on the surface. The latter representation clearly makes a worse presentation of edges although a possibly better presentation of contrast of contact errors. In either case, the point cloud interpolant is the same and the measurement results will be unchanged by display resolution.

5.6 Identification and evaluation of contact errors

Equation (5-5) gives the essence of the author's contact error evaluation algorithm:

$$ht = \min\begin{bmatrix} \max\begin{pmatrix} d(x_\alpha, y_\beta) - D(x_{\alpha-4}, y_\beta), \\ d(x_\alpha, y_\beta) - D(x_{\alpha-3}, y_\beta), \\ d(x_\alpha, y_\beta) - D(x_{\alpha-2}, y_\beta), \\ d(x_\alpha, y_\beta) - D(x_{\alpha-1}, y_\beta) \end{pmatrix} \\ \max\begin{pmatrix} d(x_\alpha, y_\beta) - D(x_{\alpha+1}, y_\beta), \\ d(x_\alpha, y_\beta) - D(x_{\alpha+2}, y_\beta), \\ d(x_\alpha, y_\beta) - D(x_{\alpha+3}, y_\beta), \\ d(x_\alpha, y_\beta) - D(x_{\alpha+4}, y_\beta) \end{pmatrix} \end{bmatrix} \tag{5-5}$$

the criterion being that if 'ht' exceeds a threshold level (1 µm was used in Figure 5-24, left) a contact error is indicated. Describing it in words: a difference is found between the measured height of a scan point and heights of scan points either side of it in a direction perpendicular to the scan, but in the plane of the surface. The maximum difference of those to the left, and separately on the right, are found. The lesser of the two maxima is taken as the +ve height deviation 'ht'.

Figure 5-23 shows the points involved. Only the indicated rectangle for (x_α, y_β) represents an actual scan point – the others $(x_{\alpha-3}, y_\beta)$, $(x_{\alpha-2}, y_\beta)$, $(x_{\alpha-1}, y_\beta)$ etc. are derived by the techniques for interpolation of scattered data of sections 2.5.5 and 5.5.2, since actual scan points are not available precisely at the required co-ordinates. The interpolated points (here) do lie on scan lines, so the accuracy of interpolation will be very high, since the natural neighbour interpolation of 5.5.2 will give most significance to the available adjacent data in the same line, by virtue of its likely proximity.

The logic of the approach in equation (5-5) is as follows: a contact error will be positive (away from the surface). Examination of e.g. Figure 5-22 (and numerous similar measurements) as described in section 5.4, indicates that contact errors occur either as spikes affecting a few consecutive scan points in a single scan, caused by particulate adherence to the surface; or as elongated tracks, caused by particulate adherence to the stylus tip.

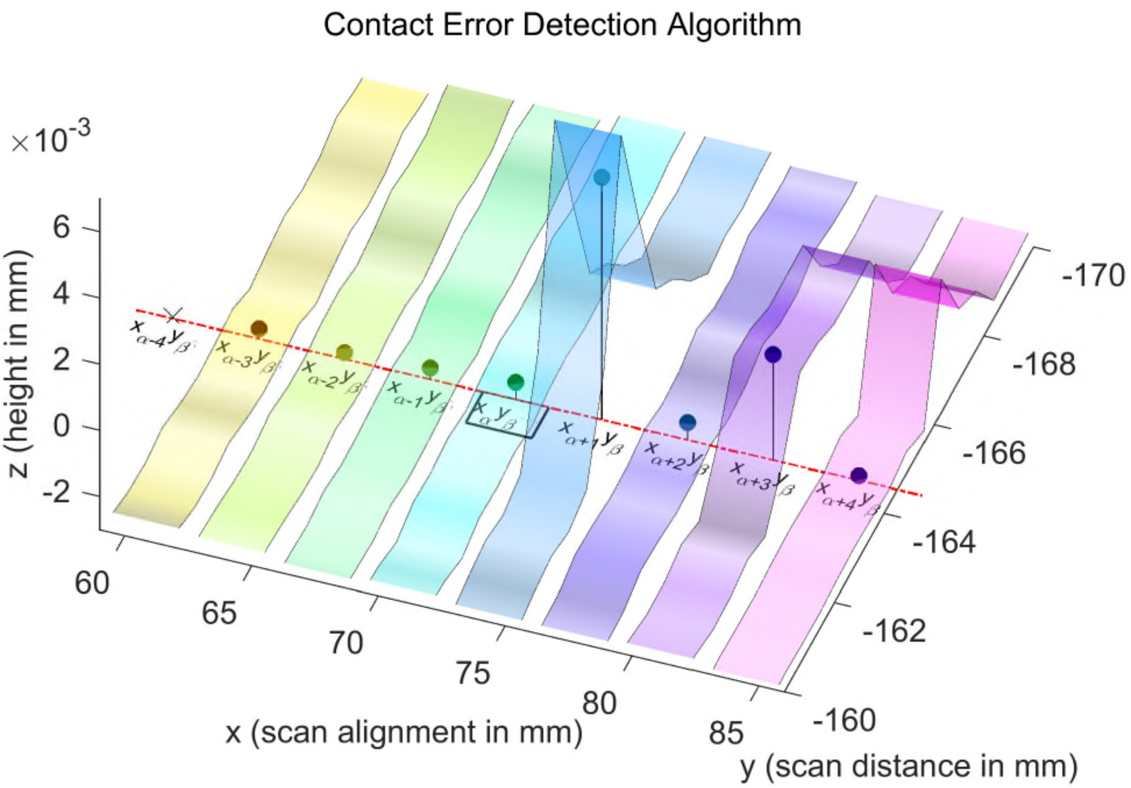


Figure 5-23: Scan contact error detection co-ordinate alignment

In the case of relatively low-slope freeform surfaces, the gradual change of surface normal angle may lead to adjacent scans experiencing effects from the same particle adhered to the stylus tip – of course this would not be the case with spikes since the distance between adjacent scans far exceeds particle size. In either case, comparisons with data within the same scan line as the point in question will frequently give erroneous results, because of the continuous nature and extent of contact errors. Moreover, comparisons which rely on filtered or model-based data will have slope limitations and may require iteration as the model-fit is improved.

The approach based on equation (5-5) is non-iterative, computationally cheap, has a deterministic cycle time and is almost independent of surface geometry and design. The sole exception to that is a limitation on slope set by the chosen deviation threshold and the scan line spacing – in this case 1 in 10^4 or 1 micron for three scan line separation gaps, totalling 10^4 microns. The approach can be modified to accommodate higher slope surfaces (most easily by reducing the scan line separation) – moreover higher curvature freeforms will almost always be much smaller in dimension so that the incentive (to reduce scan time) which led to large scan line spacing, is much diminished.

The inclusion of multiple adjacent lines in equation (5-5) is made in order to accommodate multiple adjacent contact errors, such as highlighted in Figure 5-22. This highlight area depicts an occurrence of *more than two adjacent* contact errors, which in the author's experience is very rare, so inclusion of four scan lines either side of the central point in the equation is considered adequate for all likely eventualities.

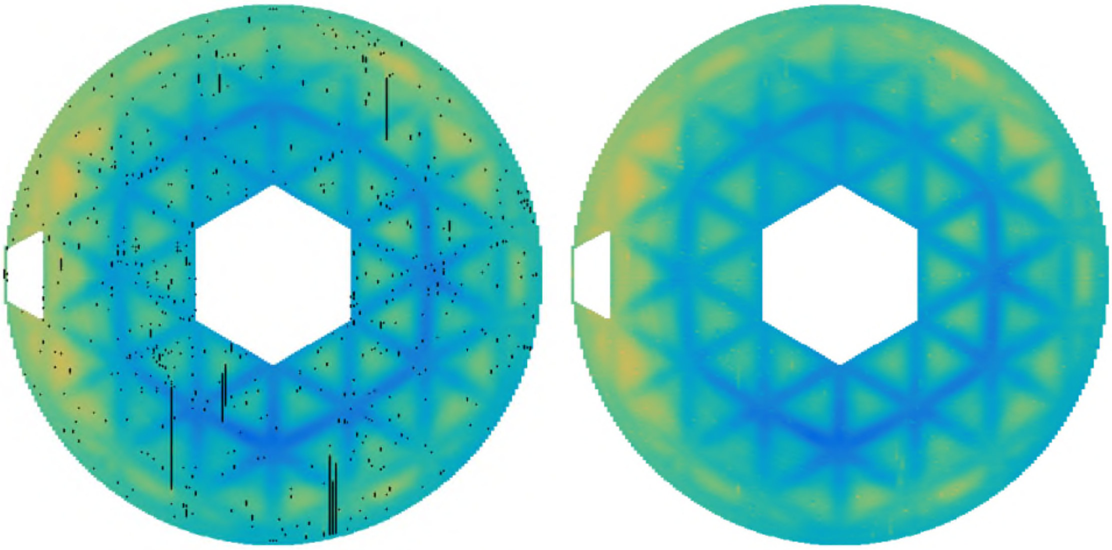


Figure 5-24: The same surface data as Figure 5-22, left: stylus contact errors marked in black; right: those data regenerated from unaffected neighbouring points (scales removed for clarity)

Figure 5-24 indicates the algorithm's efficacy in detection; with a contact error threshold of 1μm, essentially all of the discernible contact errors in Figure 5-22 have been identified (marked in black in Figure 5-24 left) and subsequently corrected (Figure 5-24 right). 2.27% of the points have been removed and replaced with interpolated data – the remaining 98% are utterly unmodified, preserving metrological characteristics.

Figure 5-25 makes the same comparison, concentrating on the potentially problematic area highlighted in Figure 5-22.

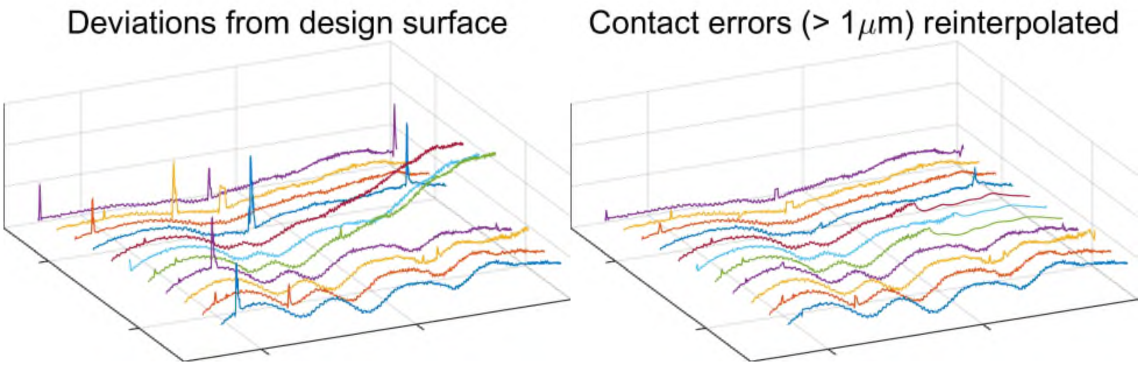


Figure 5-25 (Left): the surface data of Figure 5-12 - showing (right) the effect of reinterpolating the points with contact errors > 1 μm (scales removed for clarity)

The algorithm of equation (5-5) evaluates the maximum +ve comparison, rather than a mean or gradient – this is so that adjacent (but similar) contact errors will not evade detection. A final algorithmic feature (taking a minimum of maxima to either side) is a subtlety designed to accommodate contact errors at the edge of a surface. A contact error close to an edge may not have a lower neighbour towards the edge (or may have no neighbours at all in that direction). Use is made here of the quiet *NaN* (propagating NaN) to represent height data outside the convex hull (or within the apertures of the freeform). A signalling NaN can be the result of a failed computation causing execution to stop – in contrast a propagating or quiet NaN, as used here, represents missing values to be handled in a particular way without disrupting bulk numerical processing. In this way, if a NaN is encountered in the comparisons, only the maximum from the other side of the scan line will be considered.

The threshold for comparison in equation (5-5) could be selected empirically, but there is a justification (given below) for its choice at the level of 1μm for this data set. When the equations' deviations are computed for all ¼ million scan data points, a histogram of the deviations can be constructed.

The histogram in Figure 5-26 shows two clear 'regimes': below ~1μm a smooth distribution, associated with surface figure (slope variation) and minor measurement noise; above ~1μm an extended distribution which we can deduce (on examination of Figure 5-12 or Figure 5-22) is associated with contact errors. This distinction (between regimes) becomes starker as the scan line spacing is reduced.

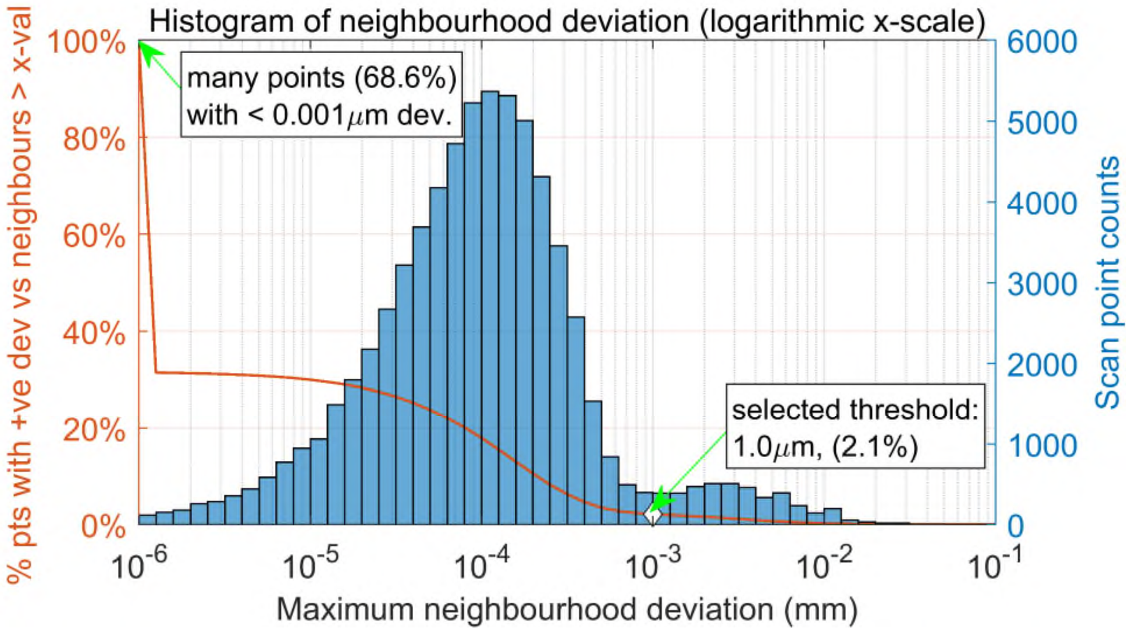


Figure 5-26: Scan point count as a function of neighbourhood deviation, with a histogram (right scale) and a threshold % cut (left scale)

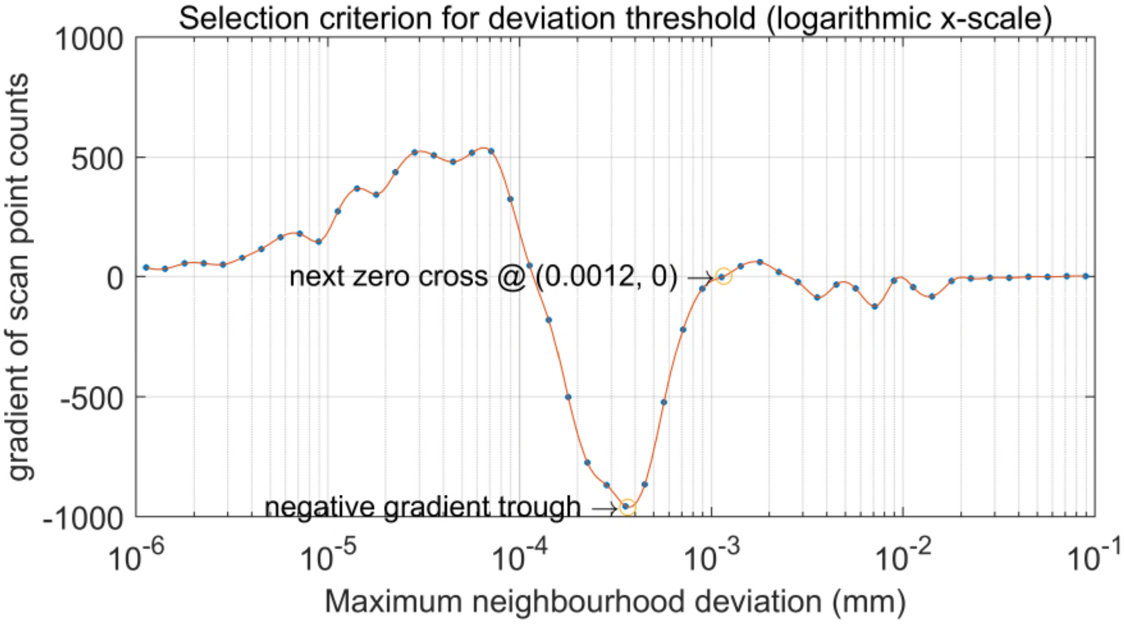


Figure 5-27: Derivative of scan point count as a function of neighbourhood deviation – the zero crossing is taken as the threshold for contact error detection

The fact that 1µm is at that local minimum in the histogram is a justification of the selection of 1µm as a deviation threshold, and this can also be used algorithmically to determine a precise threshold. Figure 5-27 shows the computation of the local minimum, based on a gradient (derivative) calculation.

This selects for the zero gradient associated with the scan point count population local minimum. Figure 5-28 shows a histogram of the same measurement data, but with the > 1µm contact errors removed, as in Figure 5-24 right.

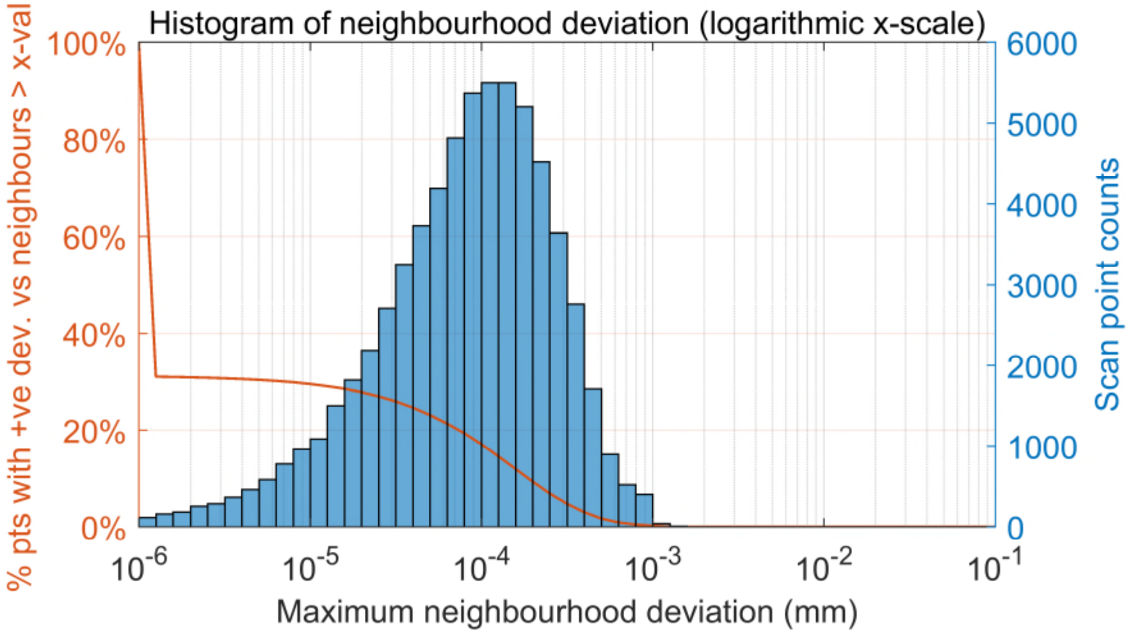


Figure 5-28: Scan point count as a function of neighbourhood deviation, with a histogram (right scale) and a threshold % cut (left scale) after scan points with contact errors have been regenerated

5.7 Data accuracy enhancement

In order to mitigate any slope/displacement errors due to the combination of scan speed and frequency response or latency, bidirectional scans are combined. In general, taking the mean of two surface heights (the same x and y co-ordinates, but scanned in opposite directions on the surface) would appear to suffice – however there are complexities:

a) for any scan, the precise trajectory (even parallel to x or y) is not without uncertainty, since there is compliance in the probe assembly i) affecting scan speed in the scan direction and ii) position (parallel to the surface) perpendicular to the scan direction. Therefore, positive and negative direction scans cannot precisely overlay, even if programmed to do so

b) within scan lines, there are frequently ‘spikes’ (see Figure 5-1) caused by scanning across surface contaminant particulates; where this happens spike data for that location need to be rejected in one direction of scan and therefore duplicate scan points will not be available in those positions – although this is an advantage of bi-directional scanning – that missing data can be mitigated, moreover scan points may be entirely absent for some areas where scan data in forward and reverse directions are rejected – cases with duplicate, single or missing data points need to be handled differently

c) some time has passed between subsequent scans of the same point – an adjustment may need to be made to account for thermal (or other) drift between successive scans when combining their data

d) in addition to thermal drift, there may also be stylus tip wear; this is not specifically addressed in this study, although it could ultimately be analysed from the data

There are numerous well-documented contributions to CMM uncertainty [199], not least thermal, but in almost any scenario, there is potential for accuracy improvement through data redundancy, so there may be benefit in multiple full-surface scan measurements. The directional influence of the scan motion can be mitigated to some extent by employing forward and reverse direction scans (using independently angled styli) in some combination.

Figure 5-3 (right) shows the stylus angled at 20 degrees to the horizontal, facing away from the direction of scan. This angle is adopted to give minimal scanning friction, which reduces the effect of induced vibrations during scanning, whilst giving sufficient clearance angle to accommodate the freeform surface's slope. This approach does require one stylus for each directional scan – easily accommodated by automatic probe changer and automatic stylus calibration on the CMM.

Figure 5-29 and Figure 5-30 show the results of the four scans in different directions – the parameters are otherwise the same in each case. Scan direction (by the long arrow) and step direction (by the short, curved arrow) are indicated in each case, together with the sequence of scanning, indicated by number.

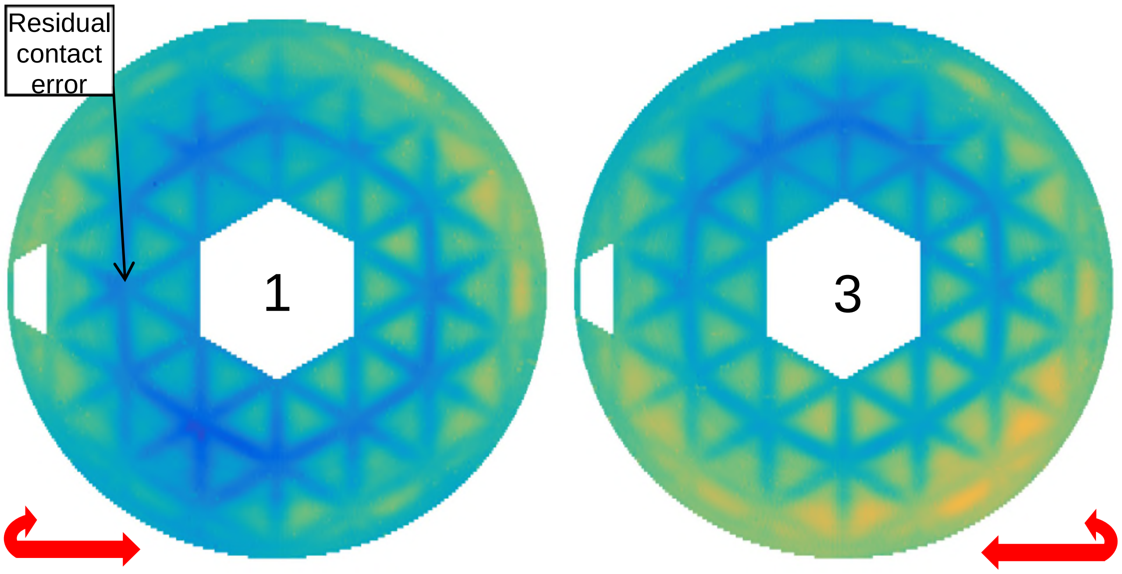


Figure 5-29: X Direction forward (left) and reverse (right) scans, after > 1 µm contact errors have been removed

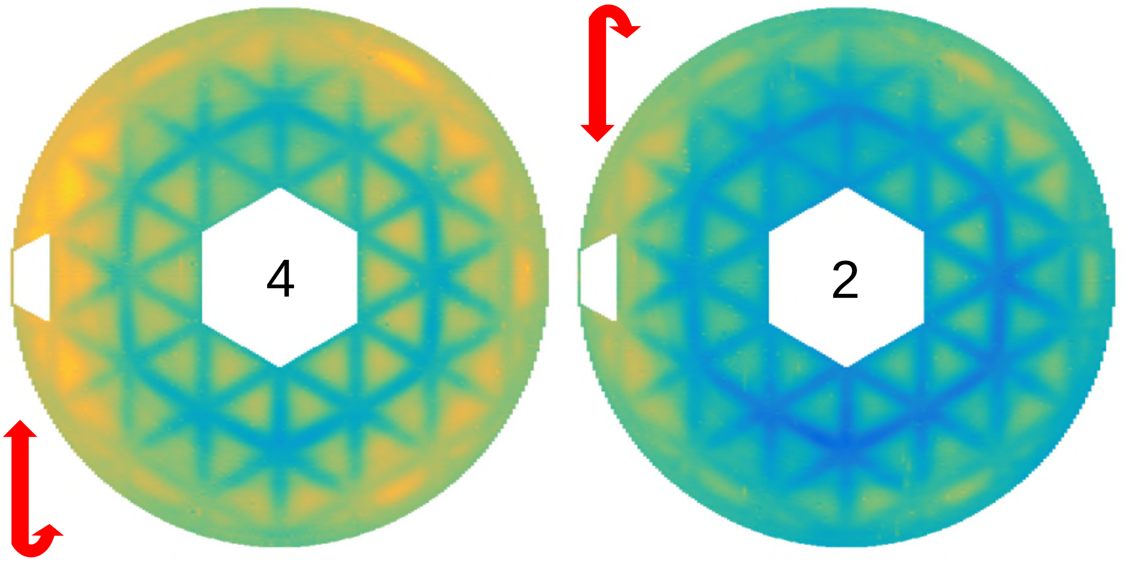


Figure 5-30: Y Direction forward (left) and reverse (right) scans, after > 1 µm contact errors have been removed

5.7.1 Residual contact error treatment

There are some barely perceptible contact errors remaining on these surface maps, which are below the 1µm threshold. There is a subtle advantage in this procedure, in respect of the redundancy in the scans. Given that contact errors are all positive going, by a principle akin to common mode rejection, subtraction from any scan set, of the mean of the other three scan sets, gives effective cancellation of all of the shorter wavelength surface form – including the light-weighting print-through pattern.

What remains is only high spatial frequency noise due to contact errors - and low spatial frequency noise due to (principally) thermally induced changes to the measurement system and workpiece, over the duration of the 4 measurements.

All of the positive-going errors in the difference pattern are contact errors in the single set – the contact errors in the subtracted sets are each reduced by a factor of three (due to the mean) and in any case negative. For this reason, application of equation (5-5) to the difference surface can identify much smaller errors, because with the surface form substantially removed, the distribution of neighbourhood deviation is smaller. Whilst the absolute height value of the difference surface is non-zero, the contact error detection algorithm of equation (5-5), which operates only on local neighbourhood, is largely unaffected.

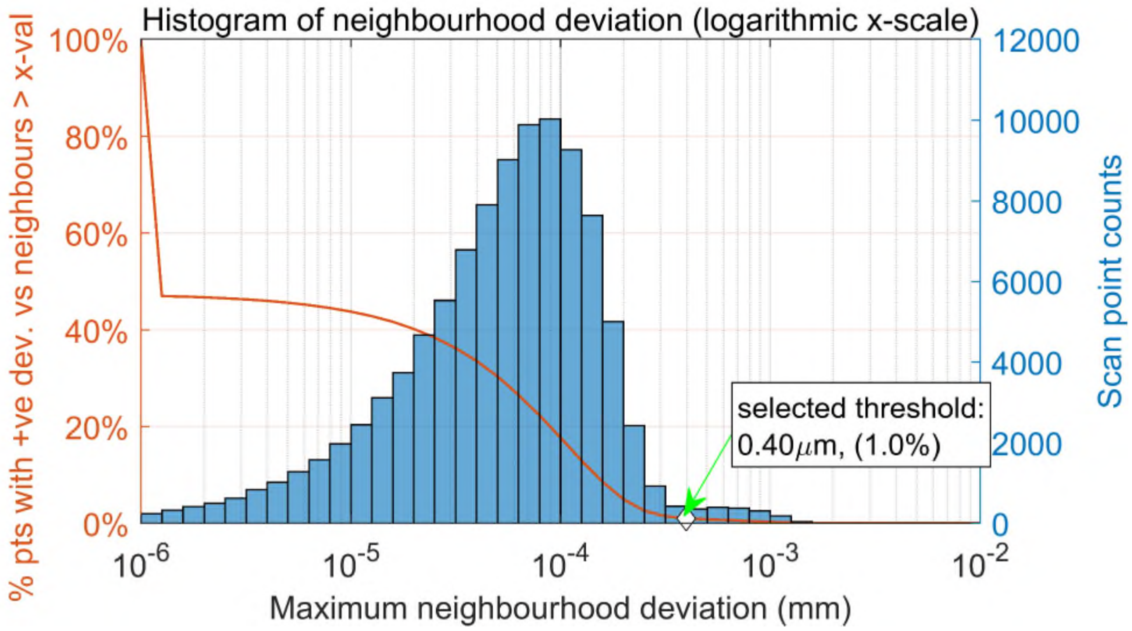


Figure 5-31: Histogram and contact error calculation threshold as derived from difference surface in Figure 5-32

An error threshold of below 400 nm is now suggested (see Figure 5-31) and this removes almost another 1% of scan points due to contact errors. The additional contact errors are indicated in Figure 5-32 by the black lines and dots. Points at these locations will be re-interpolated in the final data set to remove the errors' effects. Some 97% of the scan data remain unmodified at this stage of the data processing.

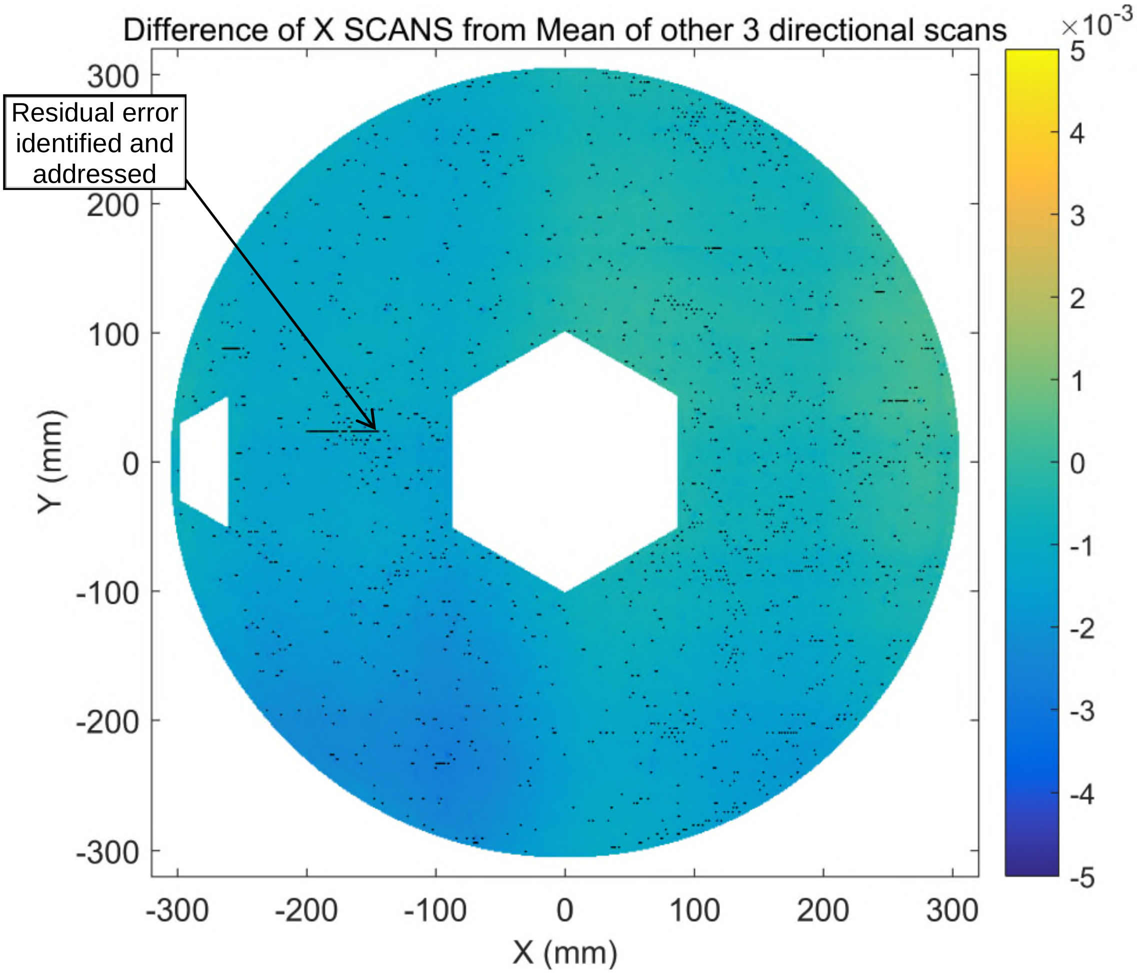


Figure 5-32: Difference of X forward scan and the mean of 3 other directions – black lines indicate additional contact errors detected above threshold in Figure 5-31

5.7.2 Scan comparisons and combinations

Figure 5-33 and Figure 5-34 show the four difference scans with all contact errors > 0.4 μm removed and with mean (z – height) offset also removed.

Some asymmetry is apparent, particularly in (and between) the x-directional scans. Asymmetry here can arise out of relative movement between measurements. For a near-spherical shape, of the 6 rigid-body motions, which can be resolved into translation in x, y & z and rotation around x, y & z: the surface is (nearly) invariant for z rotation; z translation (piston) is adjusted with z centre shift Zernike n=0. Given the overall spherical shape of the surface, although here the data being displayed all have the nominal sphere shape removed as a fixed radius component, any offset in x or y-directions for the nominal spherical centre (or rotations about x and y axes) will give an apparent asymmetry to the data.

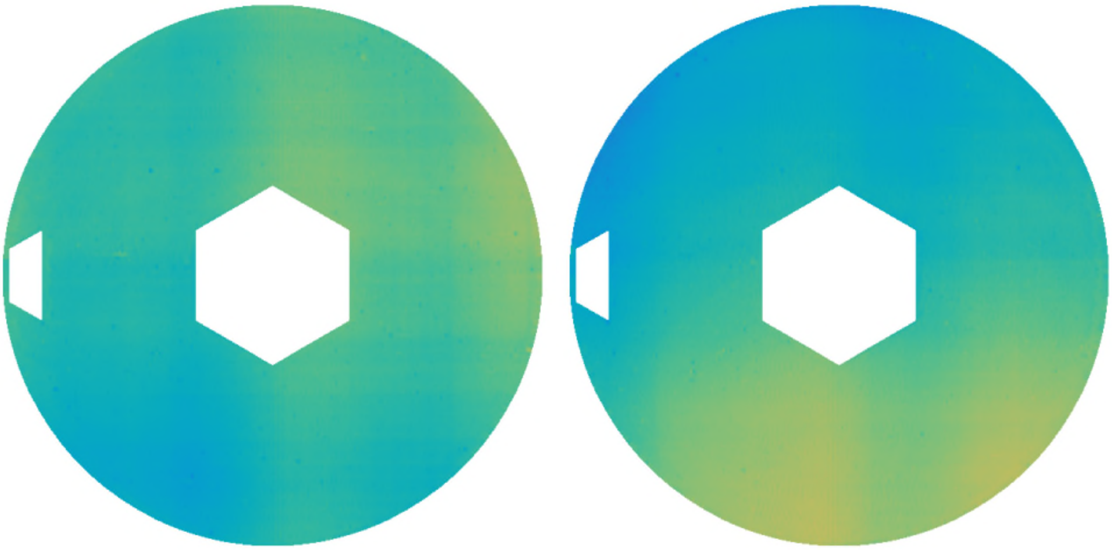


Figure 5-33: X Direction forward (on the left) and reverse (on the right) difference scans, after > 0.4 µm contact errors have been removed

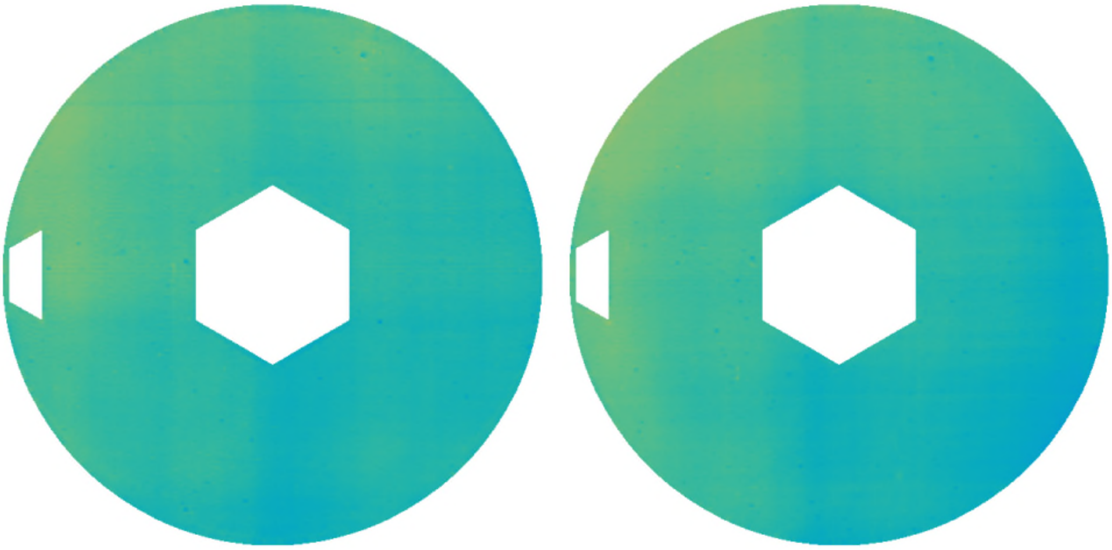


Figure 5-34: Y Direction forward (on the left) and reverse (on the right) difference scans, after > 0.4 µm contact errors have been removed

For a sphere, x or y axis rotation is indistinguishable from an x or y centre shift by analysis of the concave spherical surface alone. These will appear as an inclination: Zernike n=1, m=-1/+1 (see Table 2-7 on page 32 - akin to adjustment of the z mean) although in the x-y plane. These can be adjusted using the generic methods described below in Section 5.8, or with a specific simple fitting of x_0 and y_0 (the centre offsets) in equation (5-6) by rearrangement into the five groupings in equation (5-7). This can be solved using least squares and the techniques of linear algebra, or related methods via the solution for the constants in equations (5-8).

$$z - z_0 = -\sqrt{R^2 - (x - x_0)^2 - (y - y_0)^2} \tag{5-6}$$

$$\{x^2 + y^2 + z^2\} + a\{x\} + b\{y\} + c\{z\} + d\{1\} = 0 \tag{5-7}$$

$$x_0 = -{}^{a}\!/_{2}, \quad y_0 = -{}^{b}\!/_{2}, \quad z_0 = -{}^{c}\!/_{2}, \quad R = \sqrt{{x_0}^2 + {y_0}^2 + {z_0}^2 - d} \tag{5-8}$$

The author favours *singular value decomposition* [184] in respect of this and other linear fits owing to robustness against the limits of numerical representation of small differences between large numbers for large data sets. This asymmetry (centring error) however results in what would appear as an inclined plane in a deviation plot of a fitted sphere, so if this is *a* cause of asymmetry in the x-direction difference scans, it’s not *the only* cause, since the asymmetry is something other than or additional to a tilted flat.

Offsets in centring could arise out of thermal or other distortions of the surface, its location/support apparatus or the measuring system over the duration of the measurement, or in positional lag in motion direction during scanning. In respect of lag, it may be pertinent that the x-direction scans are moving a mass approaching 3 tonnes, using a two-motor gantry drive on the CMM in order to transport a small probe (see Figure 4-1); the moving mass for y scans is much lower, as a consequence of the construction of the machine. Another potential source of asymmetry in these patterns is the effects of the z direction reversal that occurs as the scanning probe passes the centre line of a concave surface. This could be expected however to give similar effects in x- and y-direction scans, which appears not to be the case here, unless combined with hysteretic behaviour which may be stronger in x than in y, perhaps for the ‘lag’ reason given above.

In order to analyse the bidirectional, potentially hysteretic behaviour, a comparison between forward x and reverse x scans can be made. Whilst it’s possible that (thermally induced or other) distortions may occur within the timescale of a full surface scan, which takes around 2 hours, within the timescale of a single scan line (less than a minute) these will be negligible. For the x-axis scans this relative asymmetry (between forward and reverse scans) is persistent across the lines (from top to bottom of Figure 5-33). There is additionally an asymmetry in the other direction, with the lower portion of the right hand side image in particular showing a more positive trend, but this can be adjusted with centre shift and may be due to thermal action or uncertainty in probe calibration before a scan set commences.

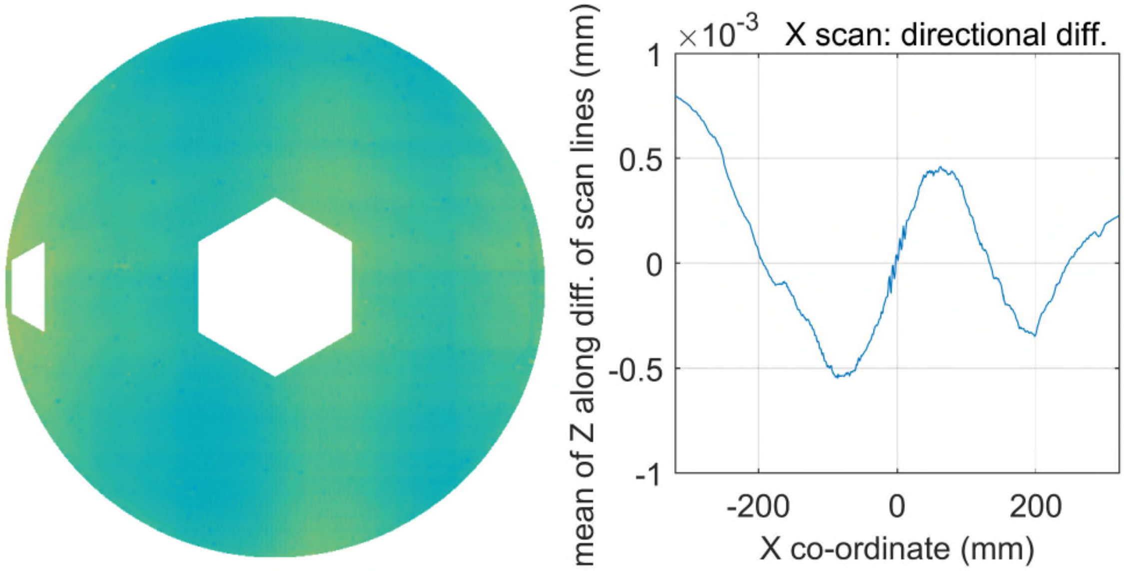


Figure 5-35: Mean (at every horizontal co-ordinate) of all the differences (at different Y values) between X forward and reverse scans once the centre adjustments are made

Figure 5-35 (left) shows the difference (reverse subtracted) between the two x-direction scans. A pattern, which is highly consistent from top to bottom, is evident, and on the right is the mean cross-section. This is very unlikely to be due to thermal or other changes because it is so consistent; the duration of each full scan was the same as the time interval between them, so if thermal changes were the cause of this difference, there would be less consistency from top to bottom. The conclusion is that this is an artefact of the measurement system scanning in the x-direction; since its profile is consistent at a range (> 600 mm) of different y-values, it appears independent of surface form.

It is not present on the same scale in the y-direction scans. Figure 5-36 shows the means in the perpendicular direction. For both x and y scans, the steps between scan lines have been made in the same direction in forward and reverse scans (see the small curved arrows in Figure 5-29, Figure 5-30). Therefore, these data cannot be used in corroboration of speed dependence for asymmetry – this could be rectified in future work – they do however give an indication of the repeatability of the form measurement capability – and these are based on comparisons over several hours.

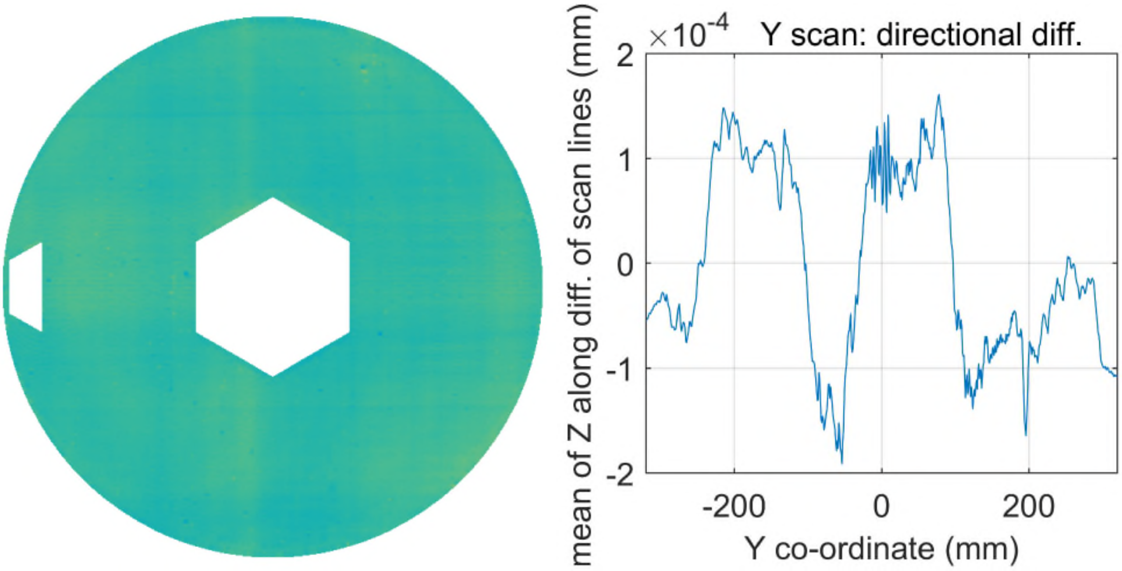


Figure 5-36: Mean (at every vertical co-ordinate) of all the differences (at different X values) between Y forward and reverse scans once the centre adjustments are made

Each of the four sets of scan data possess relatively high lateral resolution, but are not recorded at precisely identical scan point x and y co-ordinates. Therefore, in combining them in to composite data (for instance by sorting the data points from each into order in one single data set), there is a risk that a consistent (however small) offset between two data sets might lead to high frequency artefacts.

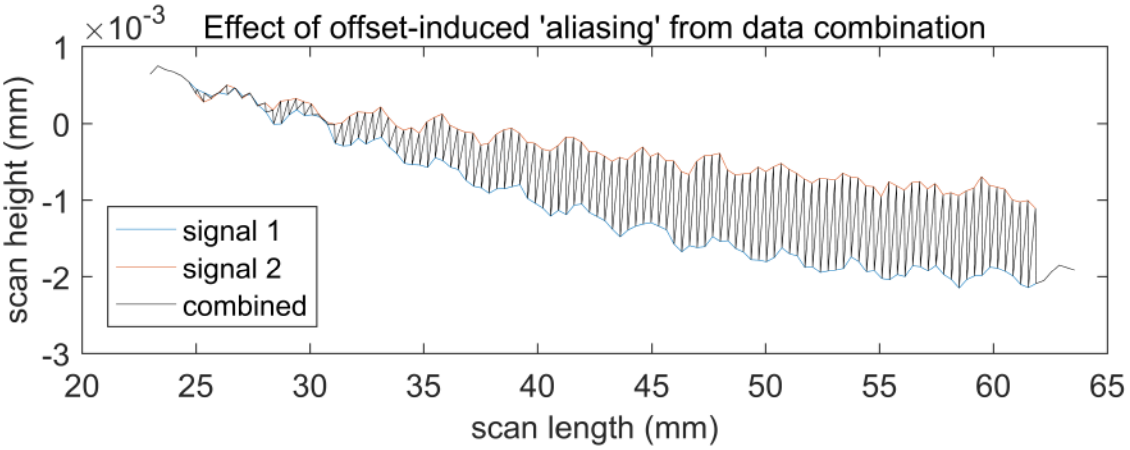


Figure 5-37: Simulated signal combination showing ‘aliasing’ from staggered data locations and signal offset

These would be caused by composite points' alternating between one data set and the other, aliasing a signal at half the sample frequency. This should be avoided in any combination technique. This pitfall can be illustrated by way of a simulation as in Figure 5-37.

An improved scheme is shown in 2-D Figure 5-38 and represented in 3-D in equations (5-9). In this scheme, abscissa points from both (all) data sets are combined and ordered to make up a larger data set. Separately, interpolated values for each data set are obtained at all the abscissa points. The mean of all ordinates for a given abscissa value is taken as a new interpolant value. This scheme, which will be called co-interpolated, can operate in n-dimensions.

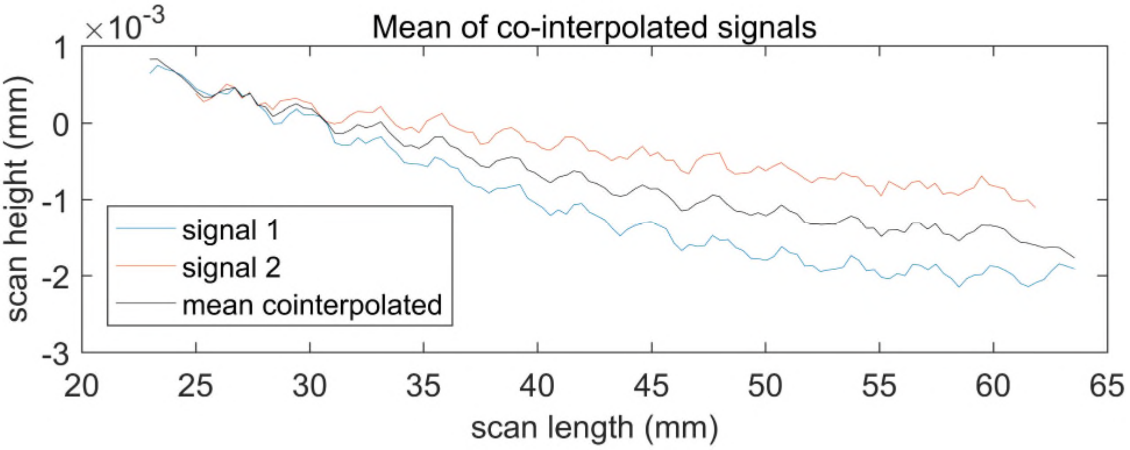


Figure 5-38: Simulated signal combination showing improvement by using the mean of co-interpolation

$$D1(x,y) = \sum_{i=1}^{m} w1_i(x,y) d1(x_i, y_i)$$

$$D2(x,y) = \sum_{j=1}^{n} w2_j(x,y) d2(x_j, y_j)$$

$$x_k = [x_i \; x_j] \qquad y_k = [y_i \; y_j] \tag{5-9}$$

$$d(x_k, y_k) = \frac{D1(x_k, y_k) + D2(x_k, y_k)}{2}$$

$$D(x,y) = \sum_{k=1}^{m+n} w_k(x,y) d(x_k, y_k)$$

Equations (5-9) show the scheme in which the larger interpolant data set is generated by interpolating the values for d in each of the combining interpolants using the combinations of x and y points from each combining data set – and then taking the mean at the common (x, y) values. In equations (5-9) for the combined set, the weights w_k, as for previous sets, are computed from a fresh Delaunay triangulation of the net (x_k, y_k) of interpolant points.

5.7.2.1 Compensation for longer period distortion (thermal)

A proposed technique will now be outlined which will to some degree compensate thermal distortions using combinations of scan data; for ultimate effectiveness this will require greater repeatability of bidirectional scans in the x-direction – some strategies to achieve this will be presented in section 5.9.

The proposed technique is based on an assumption that parallel linear scans performed as in Figure 5-29 & Figure 5-30 will exhibit minimal effects from thermal distortion in the direction along a scan line (which takes seconds to scan). This can be compared with potential distortion effects experienced in a direction perpendicular to the scans, a direction which is travelled during the scanning in a timescale of many minutes or some hours.

By combining the rapid and relatively stable scans in one direction with similar data from scans in a perpendicular direction, a compensation for the presumably larger thermal drift during a single scan set can potentially be made.

This approach would require a demonstration of stable and (short term) repeatable linear scans. This has not been achieved in the x-scan data set, but has (Figure 5-36) for the y-scans. Taking the mean values of the difference between co-incident y-scans made in opposite directions gives a variation across the whole surface of ±150 nm, which is impressive performance for a CMM on a m^2 optical surface with a sag of 20 mm, even if the comparison is made over the relatively short timescale of a single scan.

Additionally, the means of the same data taken in the perpendicular direction (Figure 5-39 right) although comparisons based on a much longer timescale, show variations of barely more than ±200 nm. This suggests that the y scans do offer the stability to demonstrate an error compensation technique.

A technique, analogous to the one proposed can therefore be demonstrated, using the y scans as a potentially lower uncertainty measure to adjust the x scans.

Given that y scans are demonstrating repeatability (even if this says little about accuracy per se) they can be used to align the x scans to them, whilst preserving the high spatial frequency data within x-scans. This can be done by intersecting x and y scans. Where they intersect, the z-height (or more correctly d-height) of the x scans can be adjusted to match that of the y scans – the intervening x scan data can be adjusted (preserving its relative height variation) to interpolate between the intersects.

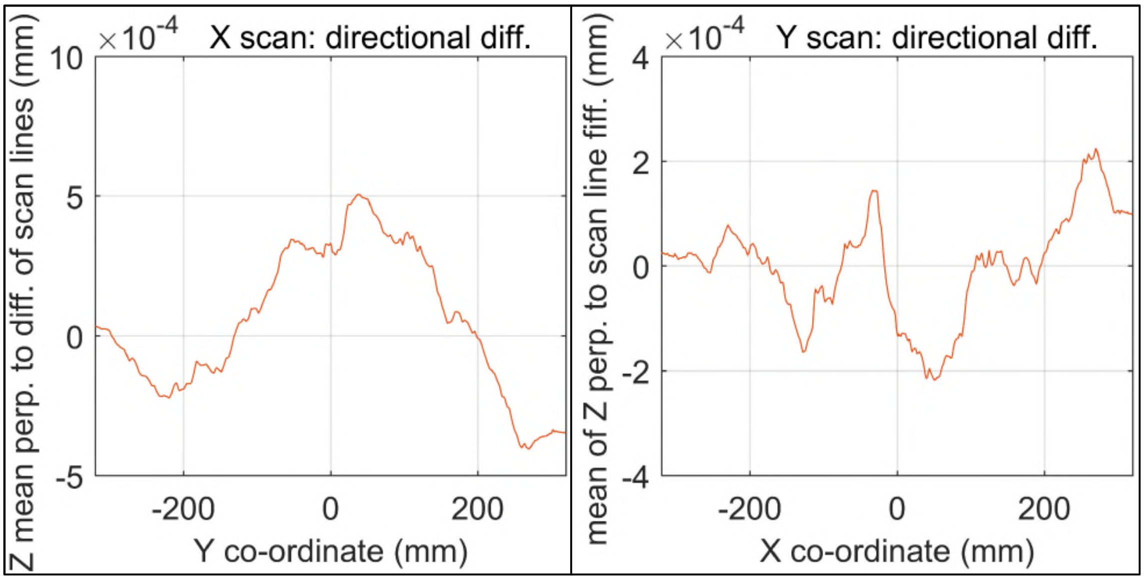


Figure 5-39: Mean for difference between X forward and reverse scans (left) and between Y forward and reverse scans (right) perpendicular to the means in Figure 5-35 and Figure 5-36 respectively

In the following, double stroke letters are used to represent an interpolant. $\mathbb{X}_{FWD}$ represents an interpolant built from forward direction scans parallel to the x-axis, $\mathbb{Y}_{BACK}$ an interpolant built from reverse direction scans parallel to the y-axis, etc. "$\mathbb{Y}$.**pts**" are the (x_i, y_i) interpolant's network of points and "$\mathbb{Y}$.**val**" are the d_i values of the interpolant at those points.

$$\mathbb{Y}.\mathbf{pts} = \mathbb{Y}_{FWD}.\mathbf{pts} \ \cup \ \mathbb{Y}_{BACK}.\mathbf{pts}$$

$$\mathbb{Y}.\mathbf{val} = \frac{\mathbb{Y}_{FWD}(\mathbb{Y}.\mathbf{pts}) + \mathbb{Y}_{BACK}(\mathbb{Y}.\mathbf{pts})}{2} \qquad (5\text{-}10)$$

Equations (5-10) generate an interpolant which is a mean of forward and reverse y-scans. $\mathbb{X}$ can be formed in a similar way. Mean scan combinations are shown in Figure 5-40.

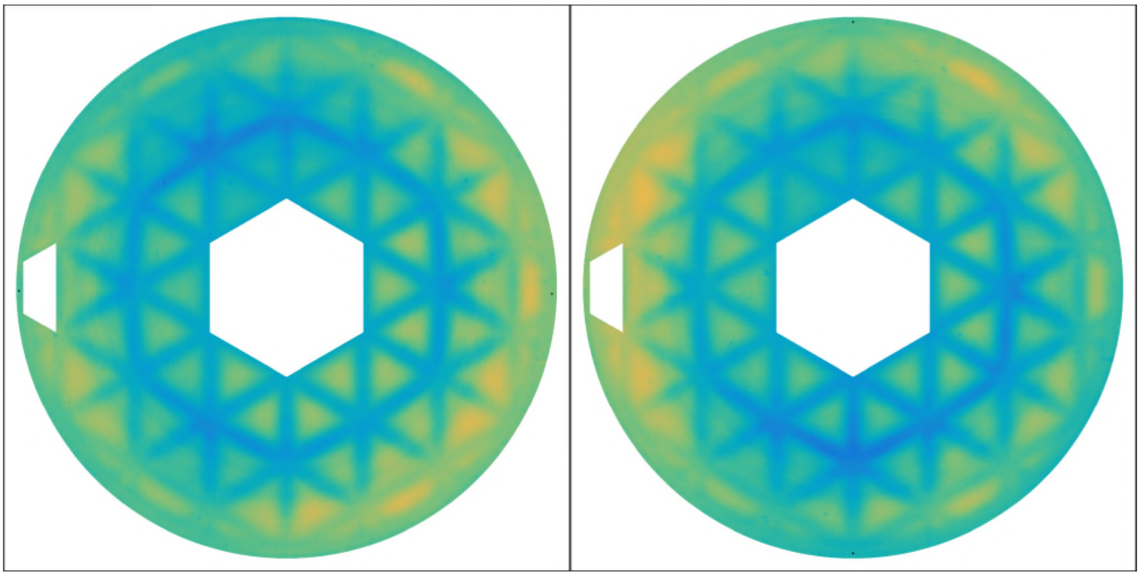


Figure 5-40: Mean scan combinations for X (left) and Y (right) scans

Equations (5-11) form an adjustment interpolant which takes the values of the x interpolant at the points of the y interpolant.

$$\mathbb{A}.\mathbf{pts} = \mathbb{Y}.\mathbf{pts} \qquad \mathbb{A}.\mathbf{val} = \mathbb{X}(\mathbb{Y}.\mathbf{pts}) \tag{5-11}$$

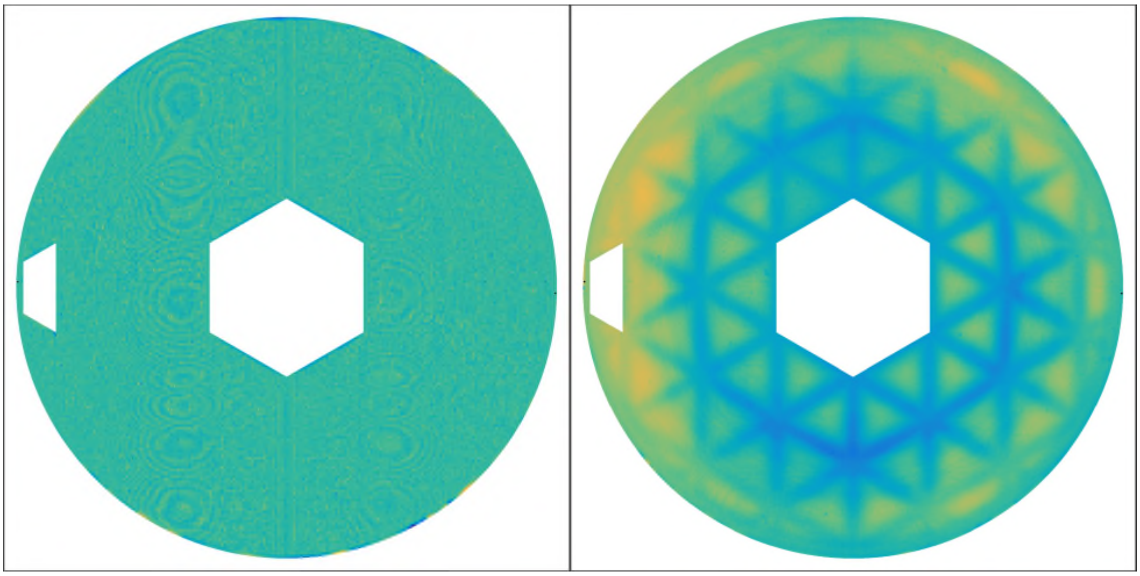


Figure 5-41: X scan residual adjustment (left) to capture high spatial resolution data and final surface deviation measurement (right) capturing X scan residual adjustment

This can then be combined with the x interpolant itself to capture the high spatial frequency data in the x scans, fitted to the form variation in the y scans, by the manipulation given in equations (5-12). This small difference is shown (Figure 5-41 left). Note that this is with a highly magnified colour scale (modified for display only) with ± 1 µm full scale, to accentuate the differences. This is then added, also in equations (5-12), to the y interpolant itself to give the resultant composite scan interpolant $\mathbb{S}$ – shown in Figure 5-41 right. This represents the most repeatable measurement of the surface from the available data.

$$\begin{aligned} &\mathbb{S}.\mathbf{pts} = \mathbb{X}.\mathbf{pts} \ \cup \ \mathbb{Y}.\mathbf{pts} \\ &\mathbb{S}.\mathbf{val} = \mathbb{X}(\mathbb{S}.\mathbf{pts}) - \mathbb{A}(\mathbb{S}.\mathbf{pts}) + \mathbb{Y}(\mathbb{S}.\mathbf{pts}) \end{aligned} \qquad (5\text{-}12)$$

Close examination of Figure 5-41 right reveals (near circular) spiral surface texture effects due to grinding in a clarity exceeding any of the other representations. This is due to the retention of the high spatial resolution data from the x-scans, without loss of the repeatable fidelity of the y-scans. The details of the surface geometry can be further analysed with the techniques of Section 5.8.

5.8 Zernike decomposition

Section 2.5.2 introduced orthogonal polynomials, and in particular Zernike polynomials. These can be used to characterise specifically wavefronts or more generally surfaces. The advantage of their orthogonality is that they can represent separable solutions to a surface description which can be independently derived and combined. Their orthogonality ensures that subtracting a multiple of one Zernike polynomial from surface data does not affect coefficients of the other Zernike polynomials that remain. This implies that there is no necessity to compute large sets of Zernike polynomials if only a small number are of interest – it further implies that the techniques of linear algebra can be used to obtain the polynomials' coefficients. An advantage of the particular polynomials due to Zernike is that they correspond to geometric factors which relate directly to optical characteristics of a system. A further advantage here is that some – and in particular the lower order polynomials – correspond to manufacturing and mounting issues which may be adjusted to correct or compensate surface errors – and that's why they're adopted here. In the following, Zernike polynomials for a circular 'pupil' of the same radius as the surface are used. An assumption could be made that the apertures, which represent together < 12% of the area of the surface, have minimal effect on the fitting of the Zernike polynomials; which seems safe on the basis of the arguments [165,159] presented on page 33 in discussion of continuous circular pupils and the fitting of lower order orthogonal polynomials. This can however be verified with simulated test surfaces with Zernike polynomials fitted both with and without apertures – see Figure 5-42.

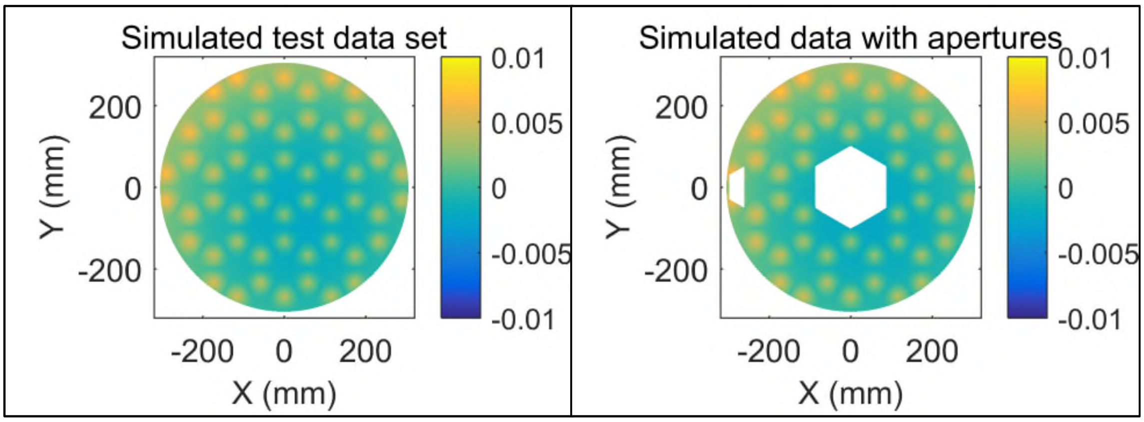


Figure 5-42: Simulated surface with cellular form and off-centre spherical deviation both with (right) and without (left) apertures (colour scale is ±0.01 mm)

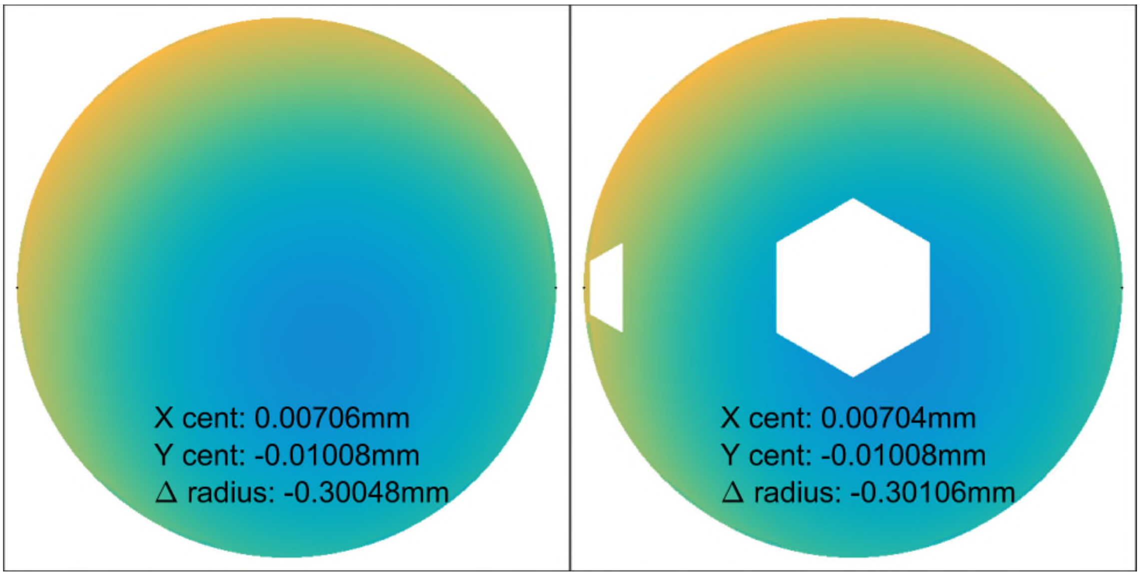


Figure 5-43: Comparison of results from fitted Zernike Polynomials both with (right) and without (left) apertures (colour scale is ±0.005 mm)

The test surfaces are identical, except for the presence of the apertures. For both there is a set of test offsets applied in order to measure the effect of the apertures on the measurement of alignment parameters by Zernike analysis; thereby testing the efficacy of polynomials optimised for continuous circular pupils on this alternative geometry. These offset parameters are given in Table 5-1.

Table 5-1: Simulated test surface parameters

Parameter	Value
Nominal radius	2028 mm
Radius error	-0.3 mm
X centre shift	7 µm
Y centre shift	-10 µm

Figure 5-43 shows the result of (simultaneously) fitting four Zernike polynomials: n = 0, m = 0; n = 1, m = -1, n = 1, m = 1; n = 2, m = 0 – representing piston (Z shift), tilt around x-axis, tilt around y-axis and defocus respectively.

These four polynomials are frequently numbered 1 (or Z_0^0), 2 (or Z_1^{-1}), 3 (or Z_1^1) & 5 (or Z_2^0) in ordered lists of Zernike polynomials. Since these polynomials satisfy the orthogonality condition, neglecting 4 (or Z_2^{-2}) & 6 (or Z_2^2) (representing astigmatism) from the computation, does not affect the accuracy of the calculation of coefficients. Since these data are of deviation from best fit sphere, the nominal radius cannot be extracted from the processed Zernike coefficients – however, knowing the nominal radius, the radius error can be computed. There is no significant additional error arising from the presence of the apertures. In the case of a much lower resolution measurement (such as for the single point probing map of Figure 5-4) the apertures could become significant. The error caused by the apertures is a function of the resolution of the scans; a low resolution leads to the failure of the Zernike orthogonality in the case of discontinuous pupils.

Subtracting the fitted Z_0^0, Z_1^{-1} Z_1^1 & Z_2^0 leaves all the other Zernike polynomials (such as astigmatism) in the data as an infinite series – effectively this is the full surface shape, including all noise, once centring and radius adjustments have been ‘zeroed out’.

The four fitted and removed polynomials are effectively the rigid body degrees of freedom – in this special case of a nominally spherical surface, x axis tilt and y axis tilt are indistinguishable from y direction shift and x direction shift. The only *degree of freedom* (DOF) therefore not explicitly measured is z axis rotation, but this DOF is embodied in the pattern – the surface figure, as shown in Figure 5-44.

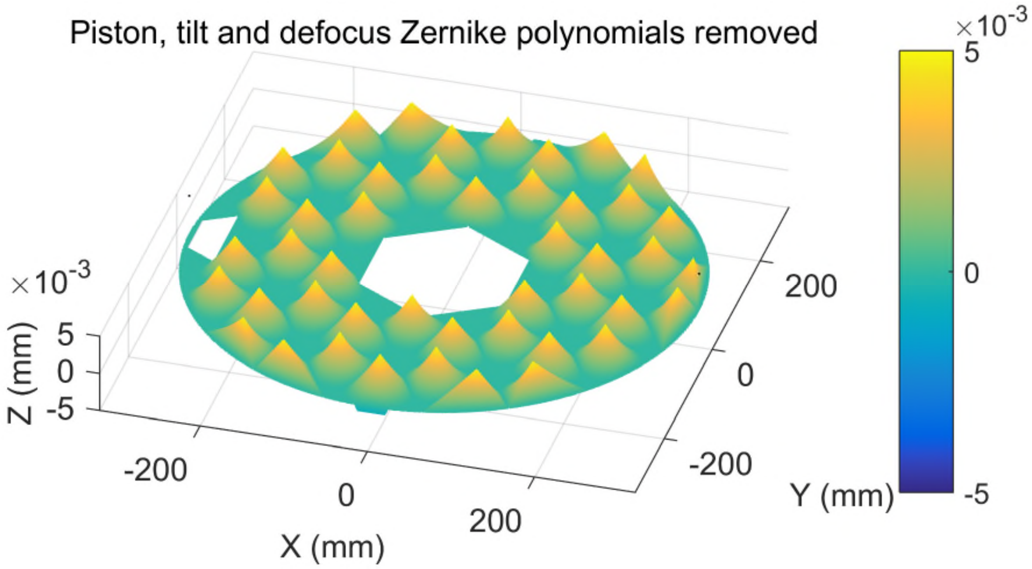


Figure 5-44: Simulated test surface - residual data once four low order Zernike polynomials are removed: piston, tilts & defocus

These same measurements: x, y & z offsets, together with residual surface figure after removal of nominal radius of curvature, are the ones required to optimise freeform surface manufacture. Some higher order Zernike polynomials may additionally be of interest in respect of surface mounting effects, such as Z_2^{-2} & Z_2^2 (astigmatism) and Z_3^{-3} & Z_3^3 (trefoil).

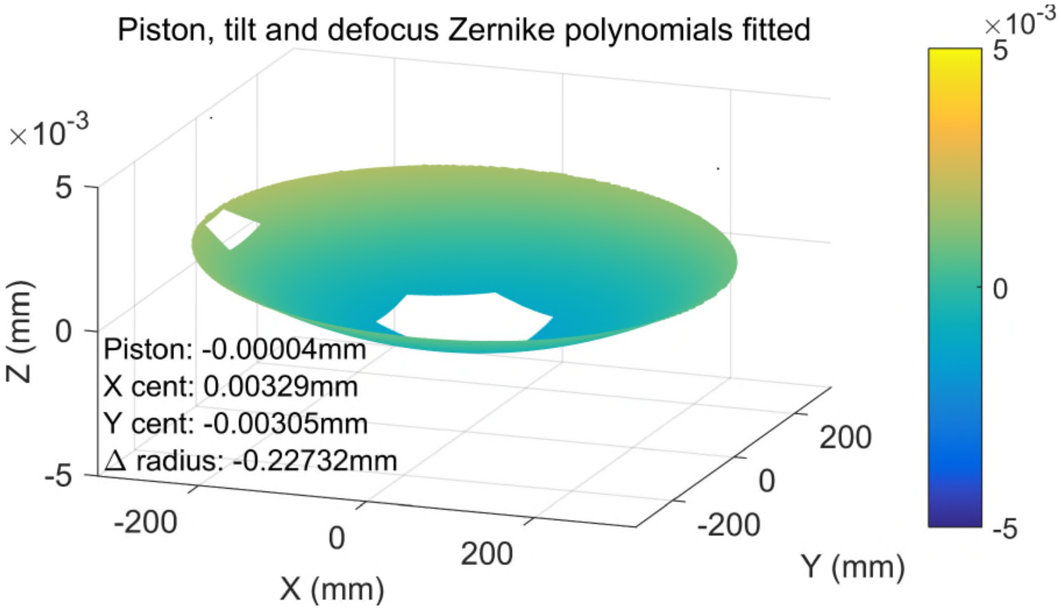


Figure 5-45: Actual surface (from Figure 5-41, right) – Zernike polynomials fitted for piston, tilts and defocus

Figure 5-45 shows the fitting Zernike polynomials for piston, tilts and defocus for the real dataset (from Figure 5-41, right). The centring offsets in Figure 5-45 (piston is automatically zeroed when the dataset is assembled in Quindos) reflect the alignment of machined surface form relative to the alignment datum features on the grinding fixture which are used to establish a co-ordinate reference system for grinding set-up and to transfer the surface to the CMM for measurement. These are manually aligned for grinding, but automatically assessed for CMM measurement. The delta radius is a machining error and represents excess curvature (around 2 microns here, across the diameter of the surface).

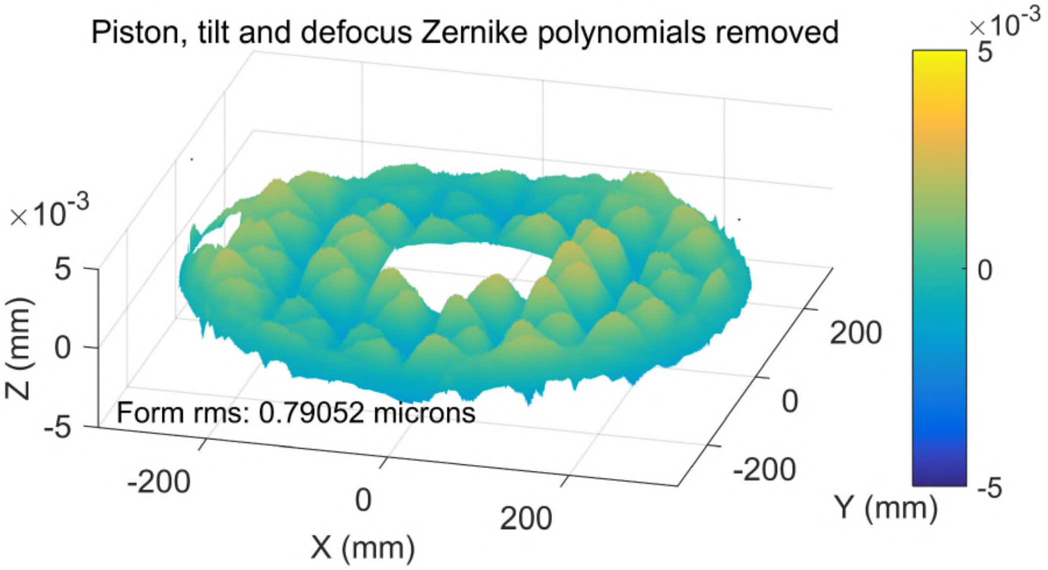


Figure 5-46: Actual surface (from Figure 5-41, right) – Zernike polynomials for piston, tilts and defocus have been removed

Figure 5-46 shows the final form of the ground surface. As in Figure 5-44, this represents the performance of the grinding process on this surface. Here, the form is dominated by the deflection of the thin face sheet by the high normal force of the grinding. The form RMS is nevertheless below 800 nm, even including the face sheet deflection.

5.9 Summary and discussion

A method and algorithmic treatment for high spatial resolution measurement of continuous freeform surfaces has been presented. These techniques are designed to give high resolution areal measurement data from a CMM with the same ease of application, and to an equivalent resolution, as full-aperture interferometric measurements. The data are processed with techniques designed to give a measurement uncertainty that approaches or exceeds a low resolution single point probing-based measurement. The aim is to create a suite of measurement and data processing techniques to permit measurement of non-specular surfaces with a fidelity that is high enough to support the manufacture of surfaces and provide measurement data suitable for the first stage of an iterative figure correction process, such as CNC polishing. There was not a specific uncertainty target. High resolution at adequate uncertainty can help to fill the gap in surface quality between what is usually the final output of the grinding stage and what is required as the input quality for an interferometric surface measurement. Successfully filling this gap can obviate the need for an entire processing step – the neutral removal step required to make a surface specular for interferometry. Currently even this step is sometimes not enough to establish a continuous interferogram, due to excessive departure from a test wavefront; this new measurement technique can avoid this requirement altogether.

5.9.1 Achievement of objectives and contribution to knowledge

5.9.1.1 High lateral resolution contact scanning measurement of smooth freeform surfaces

A first objective was to achieve a high lateral resolution for measurement data. A scanning scheme using suitably position parallel tracks in multiple directions has been demonstrated. The choice of parallel tracks based on a Cartesian geometry gives a uniform measurement density, in contrast to a polar scheme. Given the (typically) polar machining geometry for large freeforms, a polar measurement scheme would be difficult to align with the machining pattern; any misalignment would significantly alias machining marks. Crossed measurement tracks give (welcome) redundancy to the measurement at thousands of locations distributed across the surface. Scan motions which are essentially aligned to motion axes have been chosen to minimise motion uncertainty. The demonstration dataset has a resolution of 1/3 mm on the surface, which equates to > 2000 pixels on this surface. This was deliberately chosen a) to be equivalent to a competitive interferometric full-aperture optical test of the surface and b) to be high enough to exceed the Nyquist sampling criterion for the minimum grinding pitch in operation, which is around 1 mm. Even this close to the Nyquist limit, grinding marks are clearly visible in the data graphs and can yield process diagnostic and even correction information – not possible at any lower measurement resolution. This information can be used in targeted first stage polishing to give complete coverage in the first stage interferometry. A scan resolution of 0.3 mm however is not the limit of the scan measurement resolution capability. At the same measurement speed, data can be acquired at 10 times higher resolution or more, although there is a penalty in data volume. The penalty is that a different programming scheme would be required in Quindos for data handling, due to the data volume issue, so that the measurement data would be stored separately on a per-scan line basis, rather than as a single element for the entire surface, as is the case in the current scheme. The data handling capacity in Matlab is not close to exhaustion at this level, although the programming for flexible visualisation (as used for the generation of the figures in this chapter) is computationally intensive and this computational burden increases as the square of the resolution. Interpolation time also increases as the square of the linear resolution, but the code for natural neighbour interpolation from a Delaunay triangulation is highly optimised although not so easily multithreaded, owing to the volume of data that must be duplicated per thread. Visualisation is however of limited necessity, except perhaps at a final output stage.

5.9.1.2 Detection and removal of scanning contact errors

A second contribution is in the detection and removal of contact errors, caused primarily by particulate contamination of the stylus/surface interaction. This has proved to be a highly successful part of the research. Utilising the scanning geometry and data storage scheme, a highly selective and accurate contact error detection and correction algorithm has been demonstrated which handles both classes of observed contact error, as explained in a simple contact error model. An objective sensitivity threshold calculation has been shown, based on a statistical analysis of each measurement data set. Application of this technique has resulted in data sets of exceptional smoothness, for which > 95% of original data points are unmodified. This data ‘filtering’ has moreover been employed without the direct application of time or frequency domain-based filtering, which could have a detrimental effect on measurement uncertainty.

5.9.1.3 Compensation of errors due to changes in measurement accuracy during measurement

A third contribution is in the compensation of errors due to changes in measurement accuracy over the period of a measurement, although this has been only partially successful. The application of scans in 4 directions has revealed a strong asymmetry in X +ve and -ve direction scans at the selected scan speed – an asymmetry which is essentially absent from the Y direction scans. The X direction asymmetry is consistent in scale and shape throughout the measurement and at all Y co-ordinate values, independent of surface slope, so is therefore an artefact of the X axis motion system, which is gantry-controlled and has a very large moving mass. Rotation of the *measurand* surface, so that all measurements could be taken in the Y direction, would entail loss of registration of the surface, so is impractical from an uncertainty standpoint. A compensation scheme using higher repeatability direction scans to compensate lower repeatability ones was proposed and demonstrated, which yielded extra measurement detail. In the absence of the x-axis hysteresis issue, this could also afford a reduction in measurement uncertainty.

5.9.1.4 Application of orthogonal polynomial error separation to a ground freeform surface

A fourth contribution, the extraction and separation of data required for correction of errors in grinding geometry has been achieved. Separation using Zernike polynomials is standard for large optic measurements, although its application to high resolution CMM scans appears novel.

5.9.2 Further work

5.9.2.1 Hysteresis anisotropy

Scan hysteresis should be investigated on the basis of direction, including the performance of different CMM types.

5.9.2.2 X axis gantry – treatment of scanning hysteresis

The scanning hysteresis apparent in the X axis for this CMM can be investigated as a function of speed and as a function surface geometry. Differences in scanning styli can be eliminated (as part of a diagnosis) by using a common stylus and lower speeds.

5.9.2.3 Accuracy improvement through improved stylus calibration

Stylus scan calibration is only available *as standard* in a plane perpendicular to the stylus shaft. A calibration artefact could be used which exercises the contact zones on the stylus relevant to the surface scan in a specifically-developed stylus calibration.

5.9.2.4 Uncertainty investigation through artefact calibration

A suitable specularly reflective calibration artefact can be employed which has been calibrated using a known uncertainty full-aperture interferometric system.

5.9.2.5 Repeatability investigation through rotate and move

A stable artefact, supported on a kinematic location, can be employed for repeated measurements placing the artefact in different orientations and different locations on the CMM bed – comparisons can yield repeatability measures for the measurement technique.

5.9.2.6 Selection of scanning speed/data point density, as a function of surface geometry

Aspects of the metrology technique, including the filtration algorithm, can be optimised against surface shape, size and texture.

5.9.2.7 Higher spatial frequency form compensation

Particularly in the context of light-weighted optics, the level of measurement resolution and fidelity available as a result of the presented technique opens up for the first time the possibility of form compensation for face-sheet deflection, or other relatively high spatial frequency features. This can be exercised on the Box grinder.

6 TOOL PATH GENERATION

Machining describes any of a number of processes by which material is modified using a controlled removal process - what is now often called *subtractive manufacturing* by an antonymic formation from the more recently developed (but with a longer-standing name) processes of *additive manufacturing*. Traditionally, machining is a mechanically-forced process operating on metal, but almost any solid material is responsive to some form of machining. Relative motion between tool and workpiece materials can be assumed to be required in all but the most esoteric of examples. Imparting this relative motion (and usually applying machining force) is the function of the eponymous *machine*, and the trajectory of that relative motion is also called a toolpath or tool-path. Tool path design is a widely-studied subject, and it can encompass anything from pre-planning the operation of manually driven hand-cranked screw-induced motions to a full CAD-CAM process. The latter can be a process in which a design for a workpiece's final shape, based on a functional specification, is produced on a computer and then algorithmically converted to a set of machine motion instructions, automatically transferred and finally executed without any human intervention.

6.1 Influences on tool path design and its implications

6.1.1 Machine motion configuration

The motion system configuration of a machine has a strong influence over the choice of tool path geometry. Figure 4-2 (right) and Figure 4-4 show two motion system configurations on two very different machines. One machine applies machining normal forces typically in milli-Newton ranges, for finish ductile machining of soft metals; the other applies machining normal forces (through an interposing grinding spindle) typically a million times larger (~1000 N) for rough grinding of hard ceramics. Despite their very different applications (and axes' orientation & stroke) their relative motion geometry, often called cylindrical geometry, is the same in each case and this has the same influence over their tool path design geometry. This is frequently most strongly linked with tool path design; conversely some chosen tool paths dictate the selection of machine motion configuration. There are many other influences; some relevant to this study are given below.

6.1.2 Pre- and post-machining workpiece surface shape

Process chain design is discussed a little further in section 5 on page 50; primarily and most obviously however, for any process step, the shape of the volume of material between the input (pre-machining) and output (post-machining) surfaces dictates the amount to be removed as a function of position on the surface. This position may be identified for instance in Cartesian or spherical co-ordinates (Figure 5-5) or by a parametric description. A great deal of tool path design research has related to the most efficient traversal of that inter-surface volume, taking into account the potentially infinite variety of combinations of simultaneous motions of multiple axes – often five axes. This is less of an issue in the context of this research for the reason that both of the machines identified above have cylindrical motion configurations with only three axes. Since both also employ convex tools with finite formed radii, there is (with one subtle exception as described in chapter 8) a unique combination of individual axes positions for each machining point and once this combination is determined, the general pattern of the tool path is more or less driven by the other factors.

6.1.3 Tool shape

There are two machining processes under consideration in this study, single point diamond turning (SPDT) and formed wheel fixed abrasive grinding. In each case, at the moment of surface creation, a common contact vector is normal to the created surface on the workpiece and normal to the cutting surface on the tool; this could be called the normality condition of surface creation.

For SPDT, the tool’s cutting surface is often considered as a cutting edge formed by the intersection of rake faces on the tool. Even after tool wear has occurred, for the purely geometric exercise of tool path design the cutting surface can be modelled as a line lying in a plane which is normal to the workpiece surface at the point of contact. In either case - grinding surface (Figure 6-1) or cutting line (Figure 6-2) - for the purposes of tool path design, the contact normality condition given above still holds.

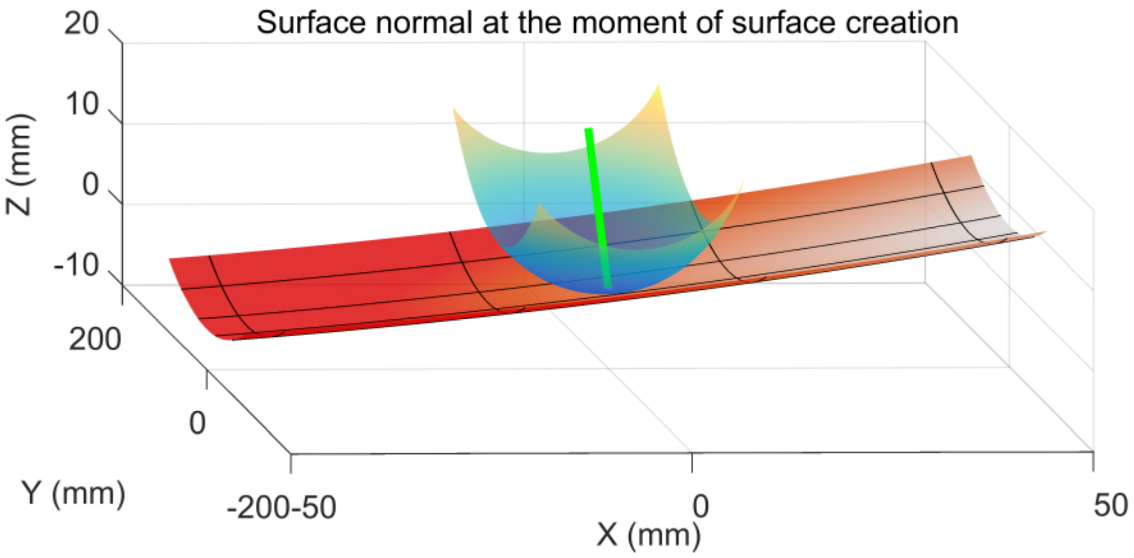


Figure 6-1: Surfaces contact with common normal vector – 3D tool cutting surface

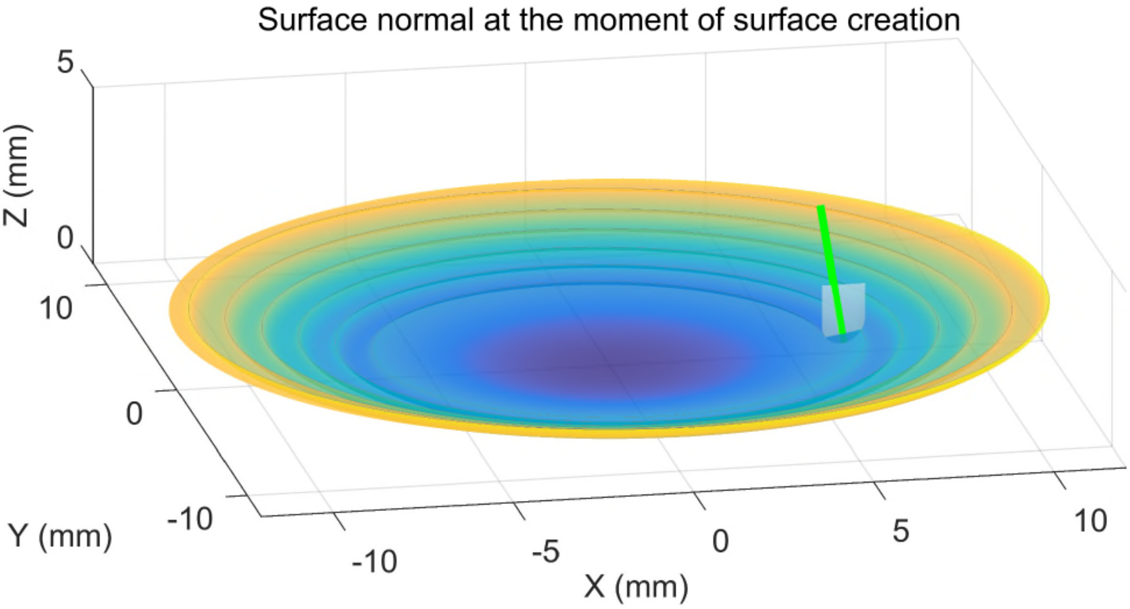


Figure 6-2: Surfaces contact with common normal vector – 2D tool cutting edge

Therefore, for the machines involved in this study (where the tool orientation is fixed) given purely convex tools with radii of curvature less than the smallest concave (designed) curvature of the workpiece, the tool position to cut any surface location seems to be uniquely determined. This conjecture, if true, simplifies tool path generation, since the infinite variety of possible paths for 5+ axis machining is avoided, provided some simple choices are made; we'll return to this point briefly in chapter 8.

On the basis of the tool-workpiece positional relationship for any given workpiece surface location being determined by a) the slope angle of the surface at that point and b) the unique portion of the tool's cutting edge or surface selected by that slope angle, tool path design is the process of devising the locus of workpiece surface co-ordinates (in 3D) which must be traversed by the tool, and the speed at which the cutting edge or surface is desired to traverse the workpiece surface. This is calculated based on the tool-workpiece positional relationship for any given workpiece surface location being determined by a) the slope angle of the surface at that point and b) the unique portion of the tool's cutting edge or surface selected by that slope angle. The tool path can be computed in a determined way thereafter from the path-specific slope variation of the workpiece final surface and the shape of the tool. There are other influences on that locus and speed, which follow.

6.1.4 Material processing parameters

The optimisation of machining parameters for productive capability, as used in this study, is detailed widely: for instance, [209] for SPDT of ductile materials or [12] for optical fine grinding of brittle ceramics. This leads to a process-appropriate machining force vector (a combination of normal and tangential machining forces). Machining forces can however affect the accuracy of a surface.

Machines have finite static stiffness, and even if apparently infinite static stiffness can be achieved through advanced control, this is only at the point and direction(s) of feedback measurement. It is effectively impossible for this feedback point to be precisely at the varying point on the tool responsible for surface creation. It is clear from both Figure 6-1 and Figure 6-2 that whenever the design surface local slope varies, a differing point on the fixed-orientation tool will be selected by the contact normality condition as the point responsible for final surface creation. For this reason, finite machine static stiffness can cause surface generation shape errors if the machining force varies across the shape. This is an encouragement to design machining tool paths which maintain constancy of machining force throughout the path. This does have implications to the input surface shape and condition requirements. The surface condition should be uniform, most easily achieved by applying uniform previous stage machining parameters.

A strategy adopted within this study is to have a constant depth of material removal (requiring that input and output surfaces are parallel at all co-ordinates) and to design tool paths with constant material removal rate; and where this is not possible, to have a removal rate which varies slowly at most. This last alternative at least allows compensation for varying machining force to be more easily applied.

6.1.5 Surface texture

Table 6-1: Typical SPDT finish parameters used in this study

Parameter	Value
Workpiece radius of curvature	62 mm
Tool radius of curvature	1.6 mm
Workspindle rotation	200 rpm
Tool-work path speed	up to 1300 mm/sec
Relative radial feed per revolution	10 µm
Depth of cut	10 µm
Geometrically resultant cusp height	7.8 nm
Equivalent R_q	2.4 nm
Equivalent R_a	2.1 nm

There are influences from tool path design on roughness and waviness which may dictate an approach to tool path design. For a typical tool path in a cylindrical geometry – a spiral as indicated in Figure 6-2 – a dominant influence on texture may be from the spiral path; this is revealed strongly in any radial assessment of profile. For SPDT this has a typical pitch (which causes cyclic variation in the profile assessment) by design in the region of 5 µm, which is a cause of roughness as assessed by any common surface texture filter cut-off length.

For the finish grinding mode adopted in this study, the equivalent typical pitch is in the region of 1 mm, which is a cause of waviness as assessed using short to median standard filter cut-off lengths. In either case, this produces scalloped profile features, with cusp height approximated very closely for circular arc-shaped tools by the well-known formula $h = f^2/8R$, where h is the cusp height, f is the feed pitch, R the tool radius in the plane of the radial feed direction and $f \ll R$. Typical finish machining parameters used in this study are indicated in Table 6-1 for SPDT.

Table 6-2: Typical finish grinding parameters used in this study

Parameter	Value
Workpiece radius of curvature	70 m
Tool radius of curvature	300 mm
Workspindle rotation	up to 14 rpm
Nominal tool-work path speed	25 mm/sec
Relative radial feed per revolution	1 mm
Depth of cut	50 µm
Geometrically resultant cusp height, h	272.2 nm
Equivalent R_q	129.1 nm
Equivalent R_a	111.2 nm
Tool spindle rotation	2200 rpm
Tool spindle axial synchronous unbalance error	100 nm p-v
Tool spindle induced ripple along tool-path, λ	0.68 mm
Equivalent R_q	35.0 nm
Equivalent R_a	31.2 nm

Another aspect of tool path design is the intended tool-work relative feed speed along the path. For grinding, this choice is often related to optimisation for productive capability (section 0), although in either machining mode, any frequency of relative tool-work surface normal motion will impart surface inaccuracies. In the particular case of relative motion at the workspindle rotation frequency, this will give shape errors or impart a tilt because it is synchronous with the work rotation; for other frequencies, roughness or (for harmonics of the workspindle frequency) waviness will result. For grinding, an unbalanced tool spindle is a source of such relative motion and a spatial wavelength along the tool path of $\lambda = c/\nu$ will result, where λ is the wavelength, c the relative speed and ν the frequency. The amplitude of such texture may be comparable with h

above, from the tool radial feed – this will depend on the amplitude of tool-spindle error motions. Typical finish machining parameters used in this study are indicated in Table 6-2 for grinding.

It can be seen from Table 6-2 that tool spindle-induced ripple along the tool path and the scalloped feed marks from the formed grinding wheel may be within the same order of magnitude in both wavelength and effective profile, for this typical tool path and process parameter combination.

6.1.6 Machine dynamics

In addition to workspindle-induced vibration (for a grinding machine) for any machine generating freeform surfaces, there will be changes in acceleration, especially cyclic changes, which will create force variations in its motion system, which in combination with the dynamic stiffness of a machine (more completely, its modal behaviour) will induce errors in its motion accuracy. Given sufficient information, these errors are to some extent correctable, which means that some compensation may be applied to reduce the effect of dynamic motion errors and this compensation might be applied through modification of the tool path. This can be applied based on predictive correction, following a comprehensive modal analysis, or based on error compensation, following measurement of the resulting surface. Neither technique is likely to be completely corrective; both require interpretation of a non-linear response and both are subject to possible non-repeatability. The latter technique has been adopted in this research.

6.1.7 Boundary effects

The desire for constant machining conditions to mitigate effects of finite machine stiffness is challenged by various boundary effects. Some are detailed below.

6.1.7.1 Edge effects on machining conditions

As the tool reaches an edge of a workpiece surface during machining, there are two prominent effects influencing tool path design. The first is a reduction in contact zone size and a change in its shape. For a given force-related deflection of the machine’s tool, machining pressure over the smaller contact area will need to be higher. This fundamental feedback mechanism is in operation during all machining and in the case of a smaller contact zone is likely to lead to deeper material removal. The second edge effect is that the surface normal stiffness of the material is reduced due to reduced bulk support. This will lead the tool to plunge deeper into the material. Both of these effects lead to increased material removal at edges and directly affect workpiece surface accuracy; more material is removed and this leads to what is called edge roll-off (Figure 6-3).

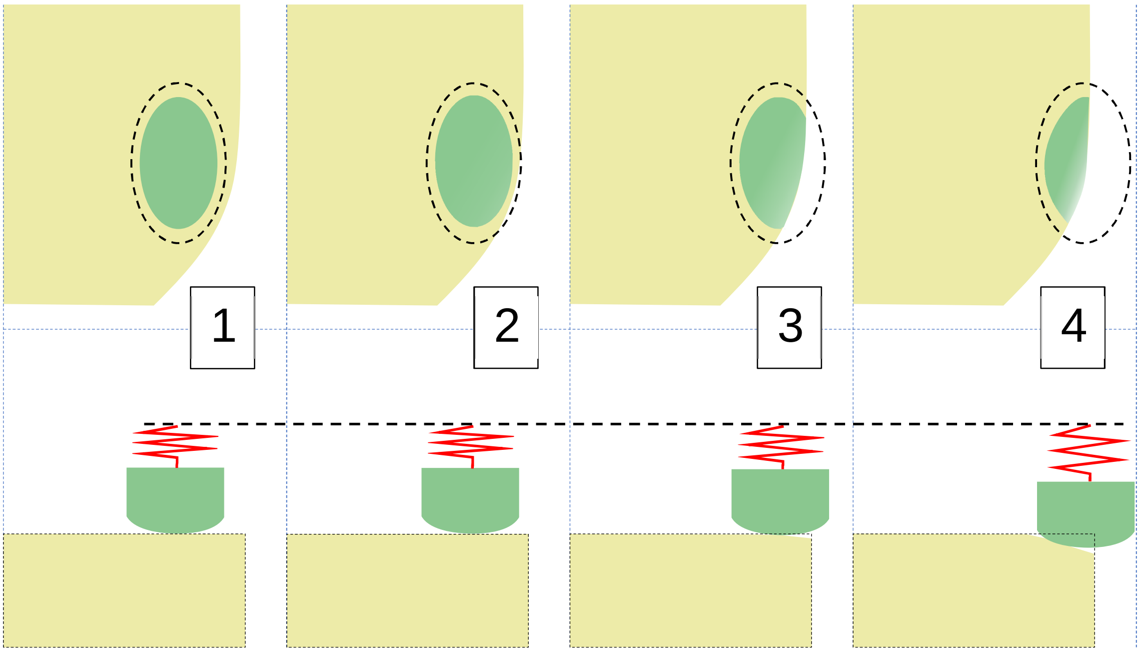


Figure 6-3: Edge effects – reduction in contact area, increase in pressure leads to *roll-off* – regulated by the machine/tool stiffness

Edge roll-off is almost inevitable, although its amount is highly dependent on machine stiffness – stiffer machines exhibit less of this effect. It is also dependent on machining process and material parameters – stiffer materials and higher force processes producing larger effects. The edge roll-off width is typically comparable to the contact zone width and although its depth is very difficult to predict accurately it tends to be repeatable for otherwise constant machining conditions. Tool path strategies for dealing with this include:

- a deliberate reduction in contact zone width through selection of a different tool or different machining conditions to reduce roll-off width
- the application of a second tool with modified parameters (as above) specifically applied only to edge zones
- tool paths (or parts thereof) specifically aligned to edges
- high lateral spatial resolution surface measurement feedback to apply detailed error compensation to the edge

The latter strategy has been adopted within this work due to the availability of exceptional metrology (chapter 5) developed as part of this research. See also section 6.2.

6.1.7.2 Motion capability limits

The application of a given tool path strategy may lead to a demand for unachievable speed or acceleration motion from the machine's axes, for instance maintaining constant material removal rate with a fixed feed pitch on a spiral path (as in Figure 6-2) implies infinite rotation speed at centre. In the case of a grinding tool path, once maximum rotation rate is reached (limited by the machine's capability) as the tool nears workspindle centre, there needs to be some compensation for the reduced material removal rate. The reduced removal rate reduces the machining normal force. A reduced force acting against the machine/tool stiffness (represented as the spring in Figure 6-3) causes 'the spring' to extend plunging deeper into the workpiece – deepest at the centre where the removal rate (and normal force) are the lowest. There are various possibilities to compensate this effect:

- the pre-machining surface may be left (conically) higher to give a thicker layer of material to remove closer to the centre, thereby restoring the machining force to a constant level
- the feed pitch distance can be increased closer to centre, to compensate the reduction in removal rate
- the tool path can (conically) lift higher as centre is approached to compensate for the effect of reduced machining force

The latter approach can be achieved by predictive compensation, based on a process model and machine stiffness data, or by high lateral spatial resolution surface measurement feedback to apply detailed compensation to the centre. A combination of predictive and feedback compensation can also be applied; this has been adopted within this work.

6.1.7.3 Momentum and discontinuities

a) In the case of intermittent machining, for instance due to machining polygonal workpieces using a spiral path, process efficiency demands a rapid motion whilst 'machining air'.

b) Machining in a spiral either ends or begins on centre, or could pass across the full diameter of the workpiece whilst it rotates, leading to double machining the entire surface, passing through the centre and machining every radius twice. The latter option is not favoured for reasons of efficiency and a requirement for a longer feed stroke - and in any of these three cases, machining the centre poses a particular challenge. Neglecting any intrinsic speed in the tool, such as with the surface speed of a grinding tool, there is zero relative workpiece-tool speed at centre, which can tend to cause atypical surface and subsurface damage. The contact patch moreover has a finite size and at some point extends both sides of the workpiece centre. This implies that in any of these three cases, a portion of the workpiece is double machined. In any either of the two favoured scenarios, a high relative acceleration of tool-workpiece is required to minimise the over-machining (past centre) as the change in linear momentum of the radial feed has to be achieved very quickly.

c) There has been a tacit assumption hitherto of zero slope at centre – i.e. of a workpiece surface that at the centre is perpendicular to the axis of rotation. If this is not the case, there is a requirement for very high acceleration at centre, and possibly a discontinuity in tool path. Tool paths used in this study assume zero or very low slope at centre. This also relates to tool-path program adjustment for centre alignment of the workpiece. This is also addressed chapter 8.

6.2 Error compensation

Capability for error compensation is necessary within any scheme for tool path generation for precision freeform surface manufacture. There are several reasons for this, given below.

6.2.1 Motion accuracy

Motion accuracy influences the relative position of tool and work directly. This can be addressed by single and multiple axis error compensation (such as volumetric compensation) following various forms of machine calibration. This is likely to be inadequate however for ultra-precision freeform surface generation, where the machining forces’ interactions with machine stiffness (which may have a frequency domain component) are not taken into account. Application of the artefact technique for machine calibration is the most powerful, where the output from machining a representative artefact is measured and the machine's calibration requirements deduced from that. Where the artefact is actually the freeform surface itself, in a (short) series of iterative machine-measure-adjust cycles, the achieved accuracy can approach the measurement uncertainty, limited primarily by the repeatability of the machining and measurement processes.

6.2.2 Tool compensation

Whilst the shape of the tool is central to the computation of tool path, based on the surface slope selecting the cutting portion of the tool for any given slope angle, any difference between the tool's design shape and actual shape will directly affect finished accuracy, or conversely, demand an appropriate modification to the tool path. If the tool has been mapped by a tool calibration process, and remains convex, the original tool path calculation can be repeated with an updated tool shape, so given up to date knowledge of the tool shape, the tool-path calculation procedure is unaffected.

For a non-convex tool however, there exists a probability of multiple distinct tool contact zones for any workpiece surface slope. This study makes an assumption of convex tool shape. As an alternative to tool compensation, a compensation based on the artefact technique (section 6.2.1) can be applied.

6.2.3 Thermal effects

Thermal effects can change the motion accuracy of a machine, as well as the size and shape of both workpiece and tool. These thermally induced changes are most likely to take place during machining, because the machining process and the implications of the associated motion control are the most significant source of non-constant energy within the system. Thermal effects are among the most difficult to compensate.

6.2.4 Predictive compensation

Where the errors are predictable or can be computed, a compensation may be applied pre-machining, or during operation. Examples include predicting tool wear based on a machining program, measuring temperature at various locations and employing a system thermal model to compute distortions. These are in effect feed forward techniques for surface accuracy control.

6.2.5 Analytical (measurement feedback) compensation.

An alternative approach (which can also compensate the errors in section 6.2.4 and perhaps others) is used in what is in effect closed loop feedback by measuring the accuracy of the workpiece surface output and then modifying the design (target) surface to compensate. The amount of the modification adjustment is usually equal, at any workpiece surface location, to the measured error; this assumes a linear relationship between input and output surface adjustments – this is likely to be a good model for high spatial density data at typical curvatures.

Either type of compensation: predictive or analytical; relies on process repeatability for its success. Analytical compensation can work for any process errors providing they are repeatable, whereas predictive compensation can only work for known or measured factors that are fed forward. The potential advantage of predictive compensation is that it can work when faced with a change in design shape, or material, or tool etc. whereas analytical compensation must iterate a previously operated process identically for its success.

6.3 Surface representation

Surface data representation is key to the process of tool path generation because there are several quantitative functional descriptions that must be extracted from the representation. These include surface height (z) as a function of lateral co-ordinates (x, y) or more properly (generally) a third co-ordinate with a close equivalence to a surface normal as a function of two others; these might for instance be ϕ as a function of r and θ as in Figure 5-5 in section 5.3. Another key quantitative function description is of the precise angle of the surface normal. This is needed to calculate the required *tool centre point* (TCP) from a knowledge of the surface generation (cutting) location and the surface slope or surface normal there. TCP is a familiar name for the tool position control point, although referring to it as a centre point makes an assumption duplicated in Figure 6-2 of circularity (2-D) or sphericity (3-D, not depicted) so that a TCP is always equidistant from the cutting location. A more general definition would be of Tool Control Point for which there is a defined geometric relationship between the tool's cutting points (which are selected by workpiece surface slope) and a fixed point on the tool.

Surface representations have been considered in detail in section 2.5 with particular consideration given to point clouds in section 5.5.2. Of the candidate representations: analytical descriptions, polygonal meshes, orthogonal polynomials, point clouds, splines, wavelets and radial basis functions only analytical descriptions give an exact representation of surface co-ordinates and surface normals from any surface parametric co-ordinate pair and are also likely to be computationally cheap, so this description should be preferred whenever available.

Polygonal meshes are in effect a subset of point clouds in the interpolant representation of section 5.5.2; the additional flexibility of point clouds and their relevance for measurement data with irregular locations means polygonal meshes can be neglected from further consideration. Orthogonal polynomials give high facility in the representation of optical properties of surfaces and the global properties relating to machining adjustments such as centring. These properties are more relevant to post-machining measurement analysis and whilst surface co-ordinates and normals can be calculated with orthogonal polynomials, the computational cost of their use rules against their adoption for tool path work, given the depth of polynomial order that must be utilised to achieve sufficient representational fidelity for a generic freeform.

Point cloud representations have been used for the measurement data, which will be combined (as measurement feedback) with design surface representations in the preparation of tool paths. This is used in section 5.5.2 in combination with natural neighbour interpolation to generate data with as high a resolution as desired; in fact, the representation in Figure 5-46 shows that even in the presence of real measurement noise, this scheme offers exceptional smoothness. Therefore, using data generated from surface design parameters (so that data at the cloud points are definitely error free) offers more than adequate representation in the case of a point cloud as a data source. The extension to an RBF representation will offer no extra accuracy and less computational efficiency in co-ordinate computation.

Whilst wavelets show promise for wide spatial-wavelength spectrum surface representation, splines offer the greatest flexibility in combination with the analytical and point cloud representations already selected. The reason for this is that splines (and in particular Non-Uniform Rational B Splines) can operate over a deliberately selected region relative to known points, either locally or globally using the same algorithmic approach. Their principal parameters (control points) are expressed in the same co-ordinates as surface data. Spline representations can be transferred with minimal alteration into CNC systems that directly support the use of spline data for control. They are computationally deterministic (once generated) and a relatively low spline order can provide high accuracy approximation from point cloud data; the low order spline polynomials can be easily transformed to provide surface co-ordinate and surface normal data.

6.4 Tool path representation

6.4.1 Chordal path representation

Chordal tool paths are a series of points (ordered sequentially) with associated interval times (or speeds) for each point. Each tool path point needs to provide a value for each of the involved axes of motion. The value can be interpreted as an absolute position co-ordinate, or a so-called incremental position, which gives the signed difference from the previous position. Where the tool path curves in the co-ordinate space of the representation, points are assumed to be joined by straight chords. If the chords are short enough, this path representation approximates the intended curved path adequately; alternatively, the CNC system may use its own spline-based interpolation scheme, together with tool path look ahead, to generate a genuinely curved path through the programmed points. The chordal program design and representation can however lead to prodigiously-sized tool path programs, sometimes exceeding 1,000,000 lines. This can pose problems even for modern control systems in terms of memory size, visualisation, single stepping, feed hold/pause, block processing bottlenecks, etc. and under some circumstances even machining accuracy. In order to explore techniques capable of reducing data volume, and processing speed requirements, an arbitrary and deliberately convoluted path is shown in Figure 6-4 chosen as a model of a complex tool path, albeit a very short one in freeform machining terms. This exhibits higher curvature than would be required in a machining spiral path. This is given in Cartesian co-ordinates. The model data has over 5,000 data points, whereas a typical machining spiral path program for a large optic (1.5 m diameter), at 1.5 mm spiral feed pitch, has over 800,000 points. The tight curvature in Figure 6-4 is deliberately chosen to offer a challenging proposition to tool path representational schemes.

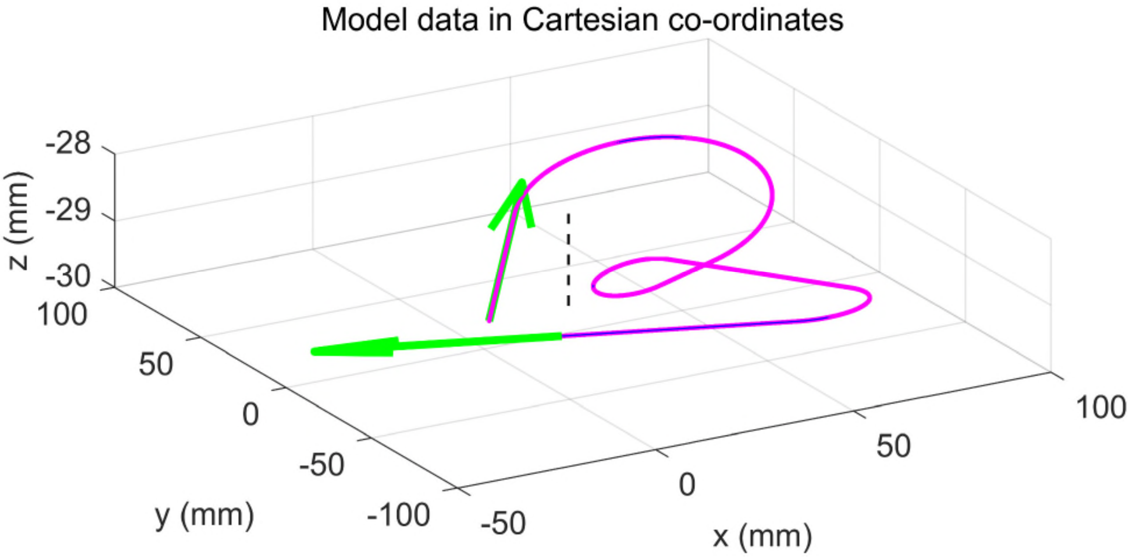


Figure 6-4: Model data in Cartesian co-ordinates

A short G code (ISO 6983) [210] segment representation is shown in Table 6-3 which uses absolute commands for the X, Y and Z positions defining the co-ordinates. This shows 7 randomly-selected consecutive points out of the 5,663 for the data in Figure 6-4.

Table 6-3: Short program segment – absolute positioning, Cartesian

‘Absolute’ command lines
X41.118970 Y13.197091 Z-29.871234 F15
X41.067902 Y13.239942 Z-29.871716 F15
X41.017121 Y13.283133 Z-29.872194 F15
X40.966629 Y13.326662 Z-29.872668 F15
X40.916428 Y13.370526 Z-29.873139 F15
X40.866520 Y13.414724 Z-29.873605 F15
X40.816909 Y13.459253 Z-29.874068 F15

Also expressed is the feed rate ‘F’, which defines the vector speed by equation (6-1)

$$F = \sqrt{{v_x}^2 + {v_y}^2 + {v_z}^2} \tag{6-1}$$

These data are defined in a constant velocity mode; hence, the constant feed rate – the 'F' data word in the table entries. This represents a constant material removal rate.

Table 6-4: Short program segment – incremental positioning, cylindrical

'Incremental' command lines
X-0.065922 C0.043739 Z-0.000482 F13
X-0.065846 C0.046136 Z-0.000478 F12
X-0.065766 C0.048567 Z-0.000474 F12
X-0.065682 C0.051033 Z-0.000470 F12
X-0.065594 C0.053533 Z-0.000466 F12
X-0.065501 C0.056069 Z-0.000463 F12
X-0.065404 C0.058641 Z-0.000459 F11

In the cylindrical geometry of the machines in this study, taking an arbitrary rotation centre of (X = 30, Y = 20) as indicated by the dashed line in Figure 6-4, these same data will appear as in Table 6-4 where here the command data are given in incremental mode, showing differences between consecutive positions. Axes are now X, C (rotary) and Z. The feed rate 'F' is also redefined in terms of the cylindrical co-ordinates as in equation (6-2), which is dependent on the units, and here degrees are used for the rotary axis C.

$$F = \sqrt{v_X{}^2 + v_C{}^2 + v_Z{}^2} \tag{6-2}$$

The constant vector velocity from Table 6-3, due to the transformation into cylindrical co-ordinates, is no longer constant.

6.4.2 NURBS path representation

An alternative chordal tool path representation can be made using the NURBS spline representation as in equation (2-17) on page 35. As indicated in section 2.5.3 this can be an efficient storage format with a significant reduction in data volume and other key advantages. Table 6-5 gives a NURBS segment in absolute positioning as coded for a Fanuc CNC. This is one of the 15 segments, which contains 7 of the total 118 control points that together replace the 5,663 original data points.

This is a reduction in data volume by a factor of around 50. The reduction ratio is a function primarily of curvature, required fitting tolerance and the efficiency of the segmentation and fitting algorithm; for regular spiral machining paths and appropriate tolerances, the ratio can exceed 1,000:1.

The first line in Table 6-5 defines the start of a NURBS segment (G06.2); P gives the rank of the spline, so that P4 is 4th rank (3rd degree); (X, C, Z) the co-ordinates of the segment's initial control point; and (F) gives the feed rate for the whole segment. The remaining lines containing (X, C, Z) triples are the other control points in the segment and G01 indicates the termination of the NURBS segment. The K values represent the knot vector. In the author's technique, the knot vector's values are formed from a parameterisation of the entire curve, so that each NURBS segment has a unique K value set, although this is not inherently necessary.

Table 6-5: NURBS program segment – absolute positioning, cylindrical

'Absolute' command lines
G06.2 P4 K0.347925 X66.961721 C40.508571 Z-28.734666 F17
K0.347925 X67.035010 C38.886862 Z-28.743803
K0.347925 X66.945107 C35.268218 Z-28.771696
K0.347925 X66.218944 C30.200921 Z-28.827442
K0.354796 X64.633448 C24.581115 Z-28.908625
K0.368349 X63.106873 C21.014910 Z-28.970818
K0.388660 X62.181455 C19.136296 Z-29.005421
K0.414424
K0.414424
K0.414424
K0.414424
G01

Figure 6-5 shows graphically the entire NURBS representation for the original data from Figure 6-4. The fitted data points lie so close to the raw data line, and are so numerous, that they completely obscure it in the figure. Each NURBS segment is shown in a different colour (data in Table 6-5 is from the fifth segment) with the control points numbered starting from 1 for each segment. At segment end-points, the curve's vector is shown as an arrow; for each junction between segments the end vector of one is the start vector for the next. The spacing of the control points is as large as possible within tolerance constraints which were set for the fitting quality of the NURBS representation. A point to note is that for the NURBS control implementation used (for FANUC 30i) each NURBS segment has a single feed rate. This limitation can be treated by using a larger number of NURBS segments, although this is not necessarily at the expense of the total number of control points, so program size can still be orders of magnitude smaller using NURBS representation. The tolerance set for the NURBS segmentation is determined by the requirements of machining.

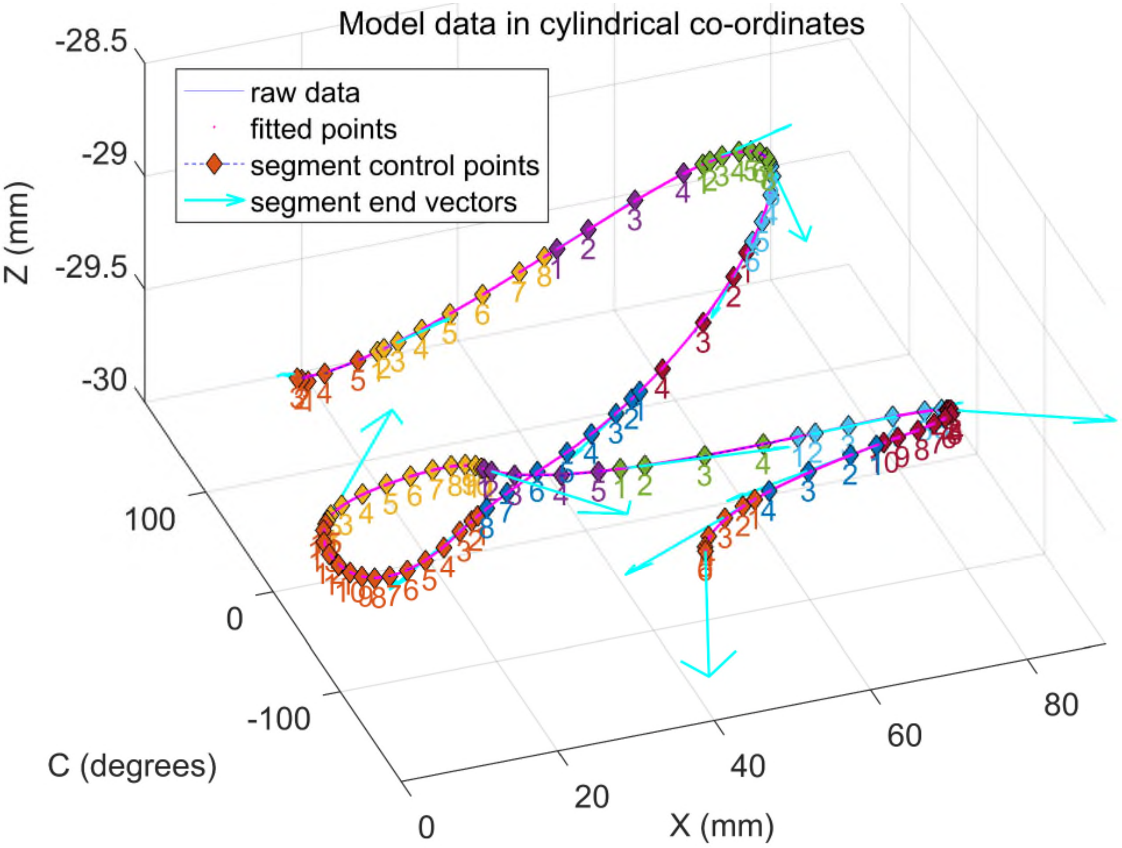


Figure 6-5: Model data in cylindrical co-ordinates, showing NURBS segmentation and control points

For the model data, the (arbitrarily) selected tolerances are given in Table 6-6 and the percentage of tolerance exploited for each point plotted in Figure 6-6.

Table 6-6: NURBS fitting tolerances – on a per-axis basis

Axis	Tolerance
X	0.01 mm
C	0.001 radians
Z	0.0001 mm
Composite (magnitude)	0.001005

The NURBS segments fit without error at segment end-points where NURBS control points are co-incident with the underlying data curve; this happens 16 times in the case of 15 segments. In between these points, tolerance exploitation may approach, but never exceed 100%.

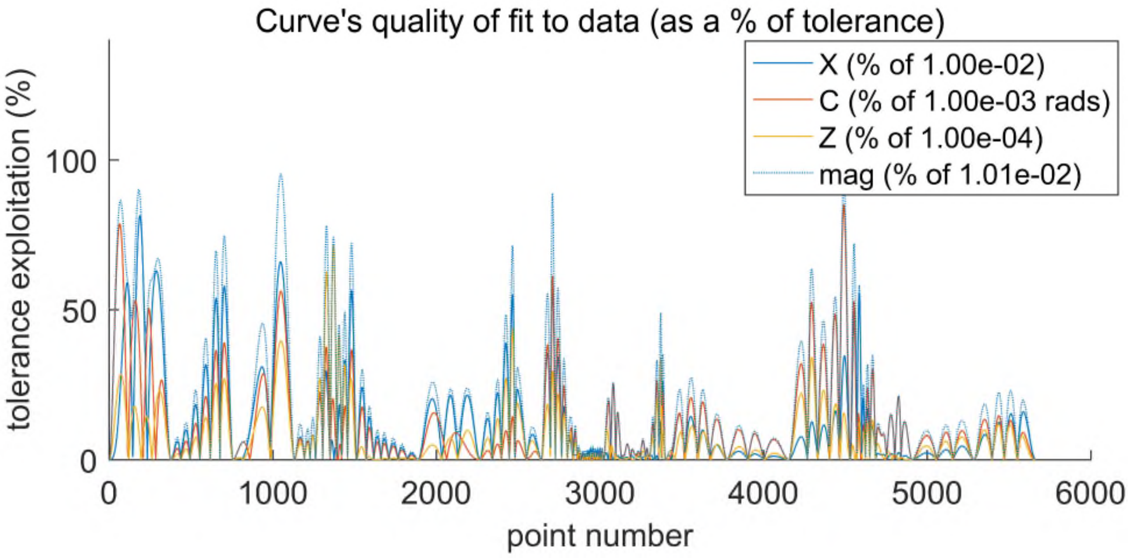


Figure 6-6: Tolerance exploitation in fitting NURBS segments

6.4.3 Generation of NURBS representation

A simple and robust NURBS segmentation and fitting strategy has been devised by the author for this work, which was used to create the NURBS representation in Figure 6-5. The input to this process is a chordal tool path representation, for instance the full data set, of which Table 6-4 is a subset.

6.4.3.1 Strategy

The presented NURBS representation strategy is broadly as follows. Following global parameterisation of the whole curve, it is split into a number of segments (as seen in Figure 6-5). Each segment is then individually approximated by a set of control points and associated knot vector. These are selected to give a higher density of control points where the curvature is higher, since high curvature in general leads to larger departure for the spline representation from the intended path. Again, this can be seen in Figure 6-5. For each approximation or interpolation, techniques of linear algebra are used to solve for the unknown control points, based on the initial parameterisation (which is in essence retained throughout) and a knot vector calculated on the basis of tool path local curvature. An initial interpolation is performed with a large number of control points and in order accurately to calculate the curvature throughout the segment, in this sole instance, these control points extend beyond the end of the segment. Subsequently, the number of knots (subject to curvature-related distribution, as above) is reduced and successive approximations performed until the tolerance allowance for the fit is nearly fully exploited. The end result is a minimal number of control points for each segment.

6.4.3.2 Parameterisation

$$\mathbf{H}(\bar{u}) = \big(x(\bar{u}), c(\bar{u}), z(\bar{u})\big) \qquad\qquad a \le \bar{u} \le b \qquad\qquad (6\text{-}3)$$

Equation (6-3) introduces the parameterisation of the data set. Whilst x, c and z are the co-ordinates of the data points, taking data points in the tool path's connective sequence we can introduce a parameterisation based on the dimensionless quantity $\bar{u}$ (*the parameter*) which runs from a for the first data point in sequence to b for the last. By convention a parameterisation runs from a = 0 to b = 1, although this is not a required parametric span. For a tool path, a logical parameterisation is on the basis of distance travelled, particularly where the tool path is designed for constant machining parameters, such as constant material removal rate. Distance travelled for the i-th point is given by the equality in equation (6-4) and is very close to the approximation where chord length is small.

$$\begin{aligned} \text{distance} &= \int_a^{\bar{u}_i} \sqrt{\left(\frac{\partial x(\bar{u})}{\partial \bar{u}}\right)^2 + \left(\frac{\partial c(\bar{u})}{\partial \bar{u}}\right)^2 + \left(\frac{\partial z(\bar{u})}{\partial \bar{u}}\right)^2}\, d\bar{u} \\ &\approx \sum_{j=2}^{i} \sqrt{\left(x_j - x_{j-1}\right)^2 + \left(c_j - c_{j-1}\right)^2 + \left(z_j - z_{j-1}\right)^2} \end{aligned} \qquad (6\text{-}4)$$

Therefore, the parameterisation for the i-th point in a tool path with n points in total is given by equation (6-5), normalised by the total path length in the denominator to give the anticipated result that $\bar{u}_n = 1$; $\bar{u}_1$ can be taken as 0 on the basis of zero distance travelled at path start.

$$\bar{u}_i = \frac{\sum_{j=2}^{i} \sqrt{\left(x_j - x_{j-1}\right)^2 + \left(c_j - c_{j-1}\right)^2 + \left(z_j - z_{j-1}\right)^2}}{\sum_{j=2}^{n} \sqrt{\left(x_j - x_{j-1}\right)^2 + \left(c_j - c_{j-1}\right)^2 + \left(z_j - z_{j-1}\right)^2}} \qquad (6\text{-}5)$$

This then associates a distinct $\bar{u}_i$ parameter value with each point. This parameterisation is essentially retained throughout, although parameter refinement is applied – see section 6.4.3.5.

6.4.3.3 Segmentation

Segmentation is the process of splitting the point data set into groups for NURBS approximation. Boundary conditions are set so that each segment ends with a curve tangent which is co-incident with the initial curve tangent of the next segment. Curve tangents' directions are indicated in Figure 6-5 by the pale blue arrows and tangential 'velocity' by the arrows' lengths. Segmentation can be performed arbitrarily, on the basis of an expectation of the ratio of data volume reduction (e.g. every 1,000 points) or based on the parameterisation (e.g. every 0.05 increase in parameter value). NURBS approximation can be performed on an arbitrarily long segment, there is a reason however to avoid long segments, in the context of preparing NURBS for CNC control. In most CNC NURBS implementations, each segment is executed with a single feed rate. So in the case of a requirement for a variable feed rate along the tool path (for instance following Cartesian to polar conversion in going from Table 6-3 to Table 6-4), the granularity of segmentation should be fine enough to avoid sudden large changes in feed rate, which would give non-uniform machining response. For instance, tool path data segment boundaries can be created each time there is an integer feed rate change in feed units per minute. Feed units are a consequence of vectored multi-axes moves, i.e. in equation (6-2) feed units are derived from normalising x & z velocity (in mm/minute) with c velocity (in degrees per minute).

6.4.3.4 Basis function computation

Fundamental to the calculation of NURBS splines is the computation of the $N_{i,\mathrm{p}}(u)$ p-th degree polynomial basis functions of section 2.5.3 and their derivatives. The author's efficient modification of a procedure in [211] combines natural and derivative basis functions computation and is given in Appendices - A.4 Program: ComputeBasisFunctions, together with supporting code in A.1 Program: FindSpan to identify knot vector engagement from a parametric value or values.

6.4.3.5 Segment interpolation

Once the original tool path data are segmented, each segment is taken in turn. Accurate approximation with a minimal control point set requires that derivatives at the segment end points are accurately known (to give directional information). It is convenient therefore to interpolate the tool path using the whole segment, and in addition point data a little beyond the end of the segment (say a quarter of the way into the next segment) to give good accuracy. Spline interpolation of the data can then yield directional information for any point – including segment boundaries. For the first and last segments, directional information for the first and last points respectively are known due to the basic design of the tool path. All of the other end point vectors can be calculated as indicated. Interpolation involves setting one knot (and later one control point) for each existing data point. This is accomplished in A.2 Program: CreateInterpolationKnotVector. Two additional control points (and two additional knots) are introduced, one in second place and one in second to last place in the point order. These two do not correspond to data points but together with the terminal points, define the directions of the terminal vectors for the segment; Figure 6-7 shows control points 1 and 2 for the second segment in the figure both lie on the terminal vector indicated with the dashed arrow. Control points can then be determined using the matrix/linear algebra techniques of simultaneous equation solution, e.g. by solving equations (6-6) which give the n+2 data points (including the 2 inserted points, Q_s and Q_e), where Q_k are the original n data points.

$$[Q_1, Q_s, Q_2 \cdots Q_{n-1}, Q_e, Q_n] = \mathbf{H}(\bar{u}_k) = \sum_i N_{i,p}(\bar{u}_k)\mathbf{P}_i \qquad (6\text{-}6)$$

Interpolation also gives an opportunity to build a ($\bar{u}$ parameterised) representation of curvature along the segment; the curvature is required subsequently to influence the relative density of control point placement. The magnitude of the local curvature is given by the Euclidean norm of the parametric rate of change of the unit tangent vector in equation (6-7), or in terms of co-ordinate system independent position (γ) where here $'$ denotes a parametric derivative.

$$\kappa = \left\| \frac{d\mathbf{T}}{d\bar{u}} \right\| = \frac{\sqrt{\|\gamma'\|^2 \|\gamma''\|^2 - (\gamma' \cdot \gamma'')^2}}{\|\gamma'\|^3} \tag{6-7}$$

Equation (6-8) gives the same curvature measure in the cylindrical co-ordinate system.

$$\kappa = \frac{\sqrt{(z''c' - c''z')^2 + (x''z' - z''x')^2 + (c''x' - x''c')^2}}{\left(x'^2 + c'^2 + z'^2\right)^{3/2}} \tag{6-8}$$

6.4.3.6 Segment Approximation

The central process of NURBS tool path approximation used in this work is a constrained minimisation problem which can be solved using the standard Lagrangian multiplier technique [211]. The solution is governed by constrained equations, which give the two end points of the segment and the directional path vectors at the end points – recall that NURBS curves have fixed end points and for a smoothly differentiable segmentation must also have fixed end directions (1st derivatives with respect to the parameterisation). This totals 4 constraints for each segment treated independently. These requirements ensure both curve continuity and curve 1st derivative continuity between segments.

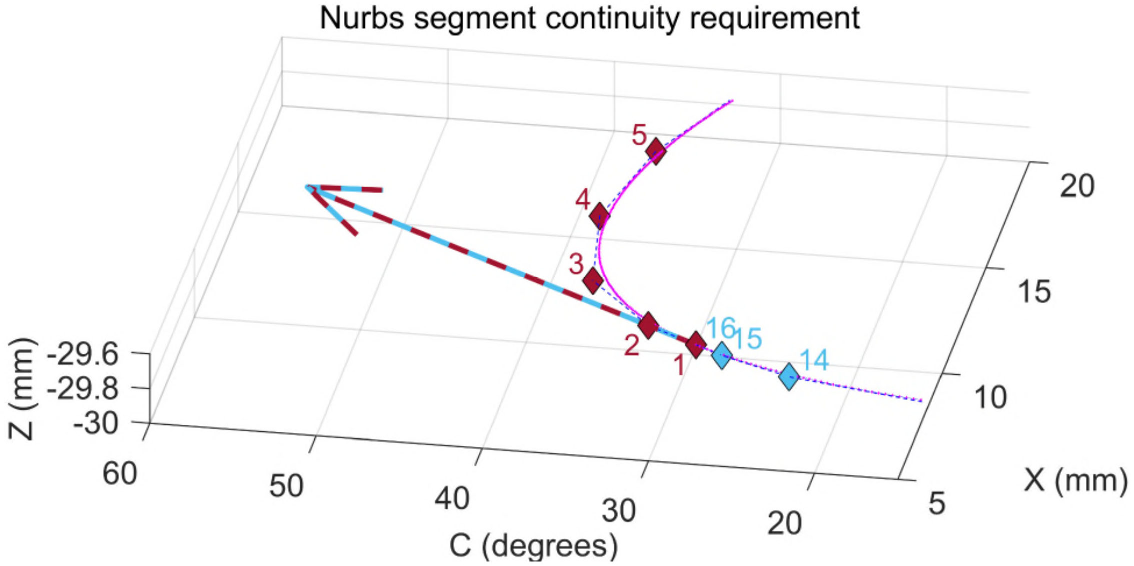


Figure 6-7: NURBS segment continuity requirement

The requirement is depicted in Figure 6-7 where control point 16 of the first segment is identical to control point 1 of the second; the first derivatives at those control points are also identical (matching arrows), thereby satisfying the continuity requirement.

The unknowns are the n control points and 4 introduced Lagrange multipliers to match the 4 constraints. We can now produce a partitioned (n + 4) x (n + 4) system of equations. The system of equations can be solved for the 4 unknown multipliers and then for the n control points. Following the notation of [211], $\mathbf{A}$ in equation (6-9) is the set of Lagrange multipliers as a vector, $\mathbf{P}$ the control points, $\mathbf{\Sigma}$ the unconstrained set and $\mathbf{T}$ the constrained set, each with one column per each of 3 dimensions.

$$\mathbf{A} = [\lambda_k], \qquad N\mathbf{P} = \mathbf{\Sigma}, \qquad M\mathbf{P} = \mathbf{T} \tag{6-9}$$

The fitting errors will be $\mathbf{\Sigma} - N\mathbf{P}$ and the sum of the weighted squares of these errors will be minimised with the $M\mathbf{P} - \mathbf{T}$ constraints, so in the technique of Lagrange multipliers, expression (6-10) below needs to be minimised.

$$(\mathbf{\Sigma}^T - \mathbf{P}^T N^T) W (\mathbf{\Sigma} - N\mathbf{P}) + \mathbf{A}^T (M\mathbf{P} - \mathbf{T}) \tag{6-10}$$

Differentiating and setting the derivative to 0 for the minimum, gives equation (6-11), in matrix form.

$$\begin{pmatrix} N^{T}WN & M^{T} \\ M & 0 \end{pmatrix}\begin{pmatrix} \mathbf{P} \\ \mathbf{A} \end{pmatrix} = \begin{pmatrix} N^{T}W\mathbf{\Sigma} \\ \mathbf{T} \end{pmatrix} \tag{6-11}$$

Matrix solution (using inverses) and linear algebraic manipulation gives the solutions for $\mathbf{A}$ and $\mathbf{P}$ in equations (6-12).

$$\begin{aligned} \mathbf{A} &= (M(N^{T}WN)^{-1}M^{T})^{-1}(M(N^{T}WN)^{-1}N^{T}W\mathbf{\Sigma} - \mathbf{T}) \\ \mathbf{P} &= (N^{T}WN)^{-1}N^{T}W\mathbf{\Sigma} - (N^{T}WN)^{-1}M^{T}\mathbf{A} \end{aligned} \tag{6-12}$$

The code in A.5 Program: FitWithEndConstraints is generalised to apply the Lagrange multiplier technique for any number of dimensions and any depth of differentiable continuity requirement, but is applied for this study as described above only for first derivative continuity. The solution procedure is partly adapted from [212,211] for NURBS. In addition to unlimited derivative continuity depth capability, the code is substantially improved by adjusting the parameterisation (which is initially based on tool path distance) iteratively for better fitting of the NURBS curve. This algorithmic extension is particularly valuable where the control point set is relatively sparse (as is desired) compared with the tool path data. Following the constrained minimisation of equations (6-9) - (6-12) there will be errors due to non-optimal parameterisation of the original data points.

Figure 6-8 illustrates in 2-D that whereas data points (x_i, y_i), (x_{i+1}, y_{i+1}) etc. lie close to the fitted curve, their corresponding fitted data points $\left(f(\bar{u}_i), g(\bar{u}_i)\right)$, $\left(f(\bar{u}_{i+1}), g(\bar{u}_{i+1})\right)$ etc. have parameterisation errors which lead to their positions along the curve not being the closest approach to the data points. Exactly the same spline is a better match to the actual data with a slight re-parameterisation that translates the points along the spline curve to the position of closest approach. This requires a (per point) minimisation of the distance between the fitted spline and the actual data by adjusting the parameters for each point.

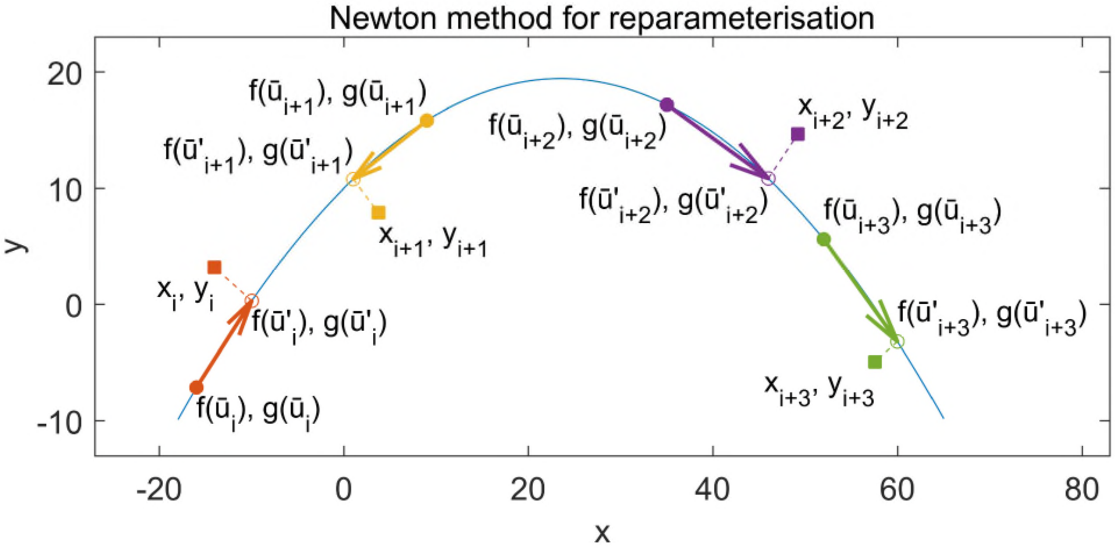


Figure 6-8: Newton's method applied for re-parameterisation

This could be achieved for instance using quadratic minimisation or Newton's iterative method [213]; the latter is adopted in A.5 Program: FitWithEndConstraints to optimise the parameterisation for the tool path points, in which a new estimate $\bar{u}'_i$ of $\bar{u}_i$ is created as in equation (6-13).

$$\bar{u}'_i = \bar{u}_i - \frac{\partial D(\bar{u}_i)}{\partial \bar{u}} \bigg/ \frac{\partial^2 D(\bar{u}_i)}{\partial \bar{u}^2}, \qquad (6\text{-}13)$$

$$D = (f(\bar{u}_i) - x_i)^2 + (g(\bar{u}_i) - y_i)^2$$

This converges very rapidly (within two or three iterations) for fractional parametric tolerances > 10^{-8}. Newton's method is particularly convenient here, because the components of the derivatives of $D(\bar{u}_i)$ are immediately available from A.4 Program: ComputeBasisFunctions. Once the parameterisation is refined, the control points can be re-fitted, enhancing further the quality of fit with respect to the (still, as always) un-modified data points and improving the ability to achieve an accurate representation with fewer control points.

The knot vector for approximation, and therefore the control point placement, is deliberately influenced by the parametric curvature, which was stored in the procedure of 6.4.3.5 above, given in equation (6-7). Where the curvature is higher (tighter) a relatively higher density of control points is required, since otherwise the higher curvature portions of the tool path would exploit a larger fraction of the tolerance. This is implemented in A.3 Program: CreateApproximationKnotVector and helps to limit the overall number of control points required to represent the tool path.

6.4.3.7 Control point set reduction and parameter refinement

In order to achieve data volume reduction in the tool path representation, by exploiting the tolerance allowance (as in Figure 6-6), the number of control points per segment needs to be reduced from the initial attempt (interpolation with the same number of control points as data points) as described in section 6.4.3.5. This process, and indeed the overall NURBS path representation, are controlled by A.6 Program: SegmentedFitControlPts.

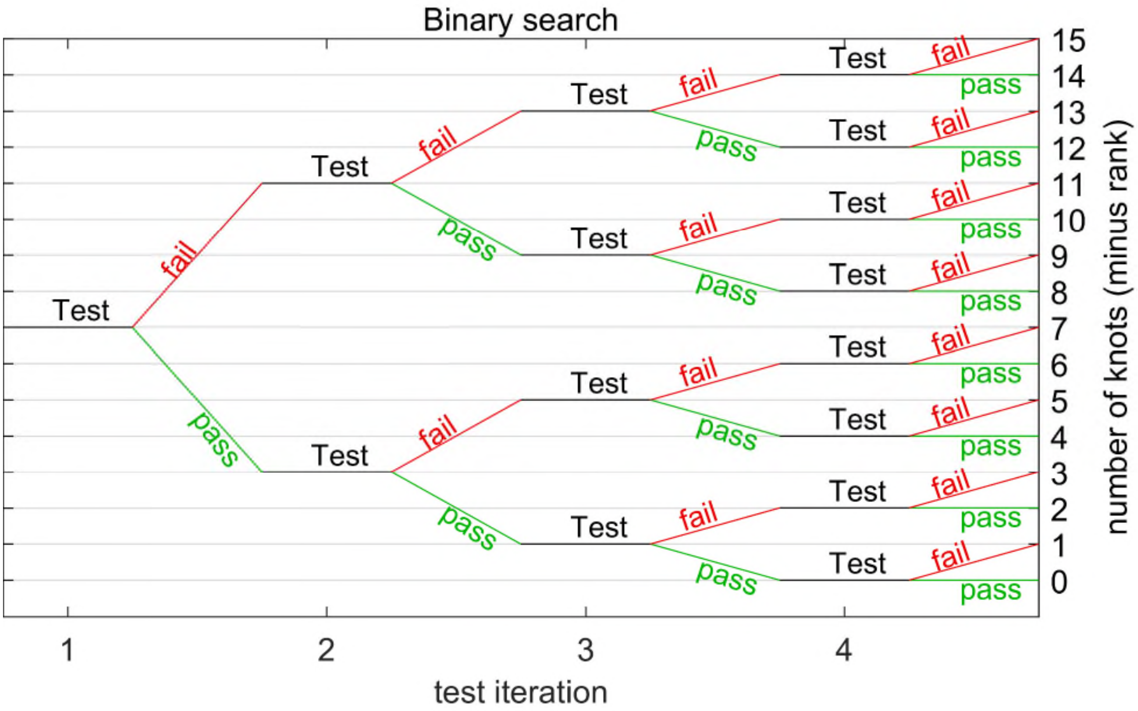


Figure 6-9: Illustration of 4-iteration binary search for minimal control point set

A procedure of *binary search successive approximation* is used to find the smallest number of control points that can fit a segment without exceeding the fitting tolerance allowance. The final optimised number of control points will (of necessity) be at least as many as the rank of the spline and will (by design) be expected to be less than 50% of the number of data points. If for a data set this latter part of the assumption turns out to be untrue (meaning the data points are close to an optimal control point set for the tolerance allowance before set minimisation begins) the binary search will fail. In that case, as a fall back, the data point set will be accepted as input to an interpolation instead of an approximation. The makes the strategy unconditionally safe, although in any example of real data, this will never occur. Figure 6-9 maps the possibilities of a 4-iteration binary search; in this process, the uppermost path (here, just 4 tests with tolerance exceeded) would result in “drop through” and interpolation using the same number of control points as data points.

For the binary search, based on the minimum number of knots, which is equivalent to the rank of the representation (typically the maximum possible rank in the CNC NURBS implementation, which itself is usually 4), a deterministic number of iterations of successive approximation can be calculated – as in equations (6-14). Here “floor” or integer-part-rounding-down is denoted by $\lfloor \quad \rfloor$, n is the number of iterations, m is the initial number of knots (as used in the initial interpolation) and r is the rank. The initial value of k, the number of knots or control points, is also given in equations (6-14). This initial value (the value for the first iteration represented in Figure 6-9) is half way between the minimum number of knots and the maximum number to be tested, which is itself half or more of the number of data points.

$$n = \left\lfloor \frac{\ln(m-r)}{\ln 2} \right\rfloor, \qquad k_0 = r + 2^{n-1} - 1 \tag{6-14}$$

The recurrence relation for the number of knots derived during iterative tests is given for the i-th iteration in equation (6-15), incorporating the test for tolerance satisfaction following fitting of the NURBS spline, where “ceiling” or integer-part-rounding-up is indicated by $\lceil \quad \rceil$; n and k are as before. There will typically be more than 4 iterations; 4 are chosen in Figure 6-9 for clarity of depiction.

$$k_{i+1} = k_i + \left\lceil \begin{matrix} 2^{n-i-1}\big|_{\text{test fails}} \\ -2^{n-i-1}\big|_{\text{test passes}} \end{matrix} \right\rceil \tag{6-15}$$

The refinement of parameterisation (6.4.3.6) becomes particularly significant as the number of control points is reduced close to the minimum needed to satisfy the tolerance requirement.

In the discussion of fitting tolerance in 0 and specifically in Table 6-6 and Figure 6-6, different tolerances are assigned to each dimension and then used in the optimisation of knot vector and control points. In fact, the least squares minimisation of equation (6-12), which is essentially the line reproduced below from A.5 Program: FitWithEndConstraints

```
“ControlPts = (PN-PM*(pinv(M*PM)*(M*PN-[Pts([1,end],:);
varargin{1:end}])));”
```

takes no account of the differing tolerances – in fact it doesn’t take account of tolerances at all. It inherently minimises the squares of the residuals treating all dimensions equally. The strategy, which is coded in A.6 Program: SegmentedFitControlPts, takes advantage however of the invariance of NURBS under *affine* transformation to set one tolerance across all dimensions, by scaling all the data (and their tolerances) so that in effect they all have equal tolerance size; the data are scaled back after all the fitting is complete.

6.5 Summary and discussion

Influences on tool path design have been detailed and analysed qualitatively and quantitatively in the context of smooth freeform surface generation with both edge and surface acting tools. The importance of error compensation has been considered and an analysis of applicable techniques given. These cement the relevance and importance of the work in chapter 5 in its influence on surface representation for ultra-precision freeform surface generation. This leads to a consideration of appropriate tool path representations, including chordal and spline. NURBS have been adopted as the appropriate tool path representation and a detailed presentation given of novel techniques for their creation, compression and refinement subject to manufacturing-driven constraints.

6.5.1 Achievement of objectives and contribution to knowledge

6.5.1.1 Identification of a surface representation scheme for smooth freeforms

An objective was the identification of a surface representation scheme for smooth freeforms that could accommodate surface design methods (including analytical and CAD), measurement feedback and other error compensation techniques and provide capability for the generation of tool paths. A scheme was presented which was based on a surface design expressed mathematically and then processed into a point-cloud based interpolant for incorporation of error compensation. This, as detailed and validated in chapter 8 meets this objective effectively.

6.5.1.2 Automated design of appropriate tool paths based on machining criteria and compensated surface shape data

A second objective was the automated design of appropriate tool paths based on machining criteria and compensated surface shape data. A spiral path in a cylindrical geometry machine benefits from nearly constant velocity or at least low acceleration motions in all three motions, for any but the most extreme departure freeform shapes. For this reason, it is the most obvious choice for either of the generating machines involved here. Low acceleration is of key importance because regardless of machining process, machines can maintain higher accuracy and more constant machining conditions under low acceleration where machine dynamics have minimal effect on path following accuracy. This objective will be more fully explored in chapters 7 and 8 where the different requirements of edge- and surface-acting tools will be considered.

6.5.1.3 Optimised condensed tool path representation

A third contribution is an optimised condensed tool path representation for optimal machining. This has been considered in detail. A novel procedure for optimising a condensed tool path subject to differing tolerances for each motion axis has been presented, which achieves a re-parameterisation and minimisation of control point representation through an efficient binary search procedure. Path accuracy has been demonstrated.

6.5.2 Further work

6.5.2.1 Binary search improvement

The binary search is in principle efficient in determining a result where there is no analytical pre-information regarding the result of the search. In the case of NURBS optimisation however, as suggested by the use of curvature as an estimator of eventual optimum relative density of control points, information is potentially available. Monte Carlo methods may be applied to establish stronger alternative relationships between optimum control point density and known factors by way of a predictive starting point.

6.5.2.2 Re-parameterisation-control point shift Interplay

The relationship between the adjustment of parametric value for the original data points and the fitting of control points is not known to be unconditionally converging. Very rapid convergence has always been observed, but the mathematical influence of re-parameterisation could be investigated to ensure these are not antagonistic.

7 APPLICATION OF SOLUTION, CASE STUDY 1 – BOLTZMANN “QUASI SPHERES”

7.1 Demands

The background and essential requirements for the Boltzmann “quasi-sphere” cavity are given in section 2.3.1. Details of the meticulous experimental procedure and analysis behind the determination of the value of the Boltzmann constant are given by the team led by de Podesta, including the author [3] for the measurements using the cavity.

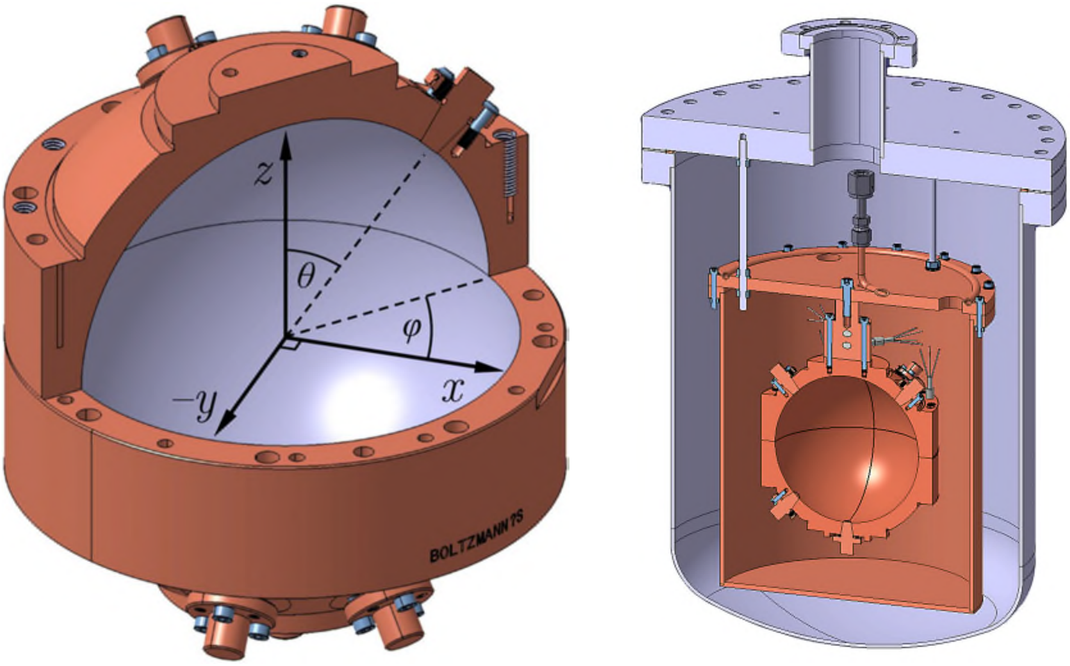


Figure 7-1: Schematic representation (left) of quasi-sphere and (right) quasi-sphere in isothermal vessel for acoustic resonance thermometry [3]

Further details are given again by de Podesta’s team with the author in [4,6,8] for some of the techniques used to measure the cavity itself as part of the uncertainty budget for the experiment. The particular concern of this manuscript however is the demand for, and nature of, the freeform surfaces involved and how their manufacturing process is designed and supported. The distinct techniques used to measure the cavity in support of its manufacture, which were devised and conducted by the author, are explained in chapter 5 with an outline of further additional techniques given below.

The cavity called a quasi-sphere is deliberately non-spherical. It is by design an ellipsoidal cavity with tri-axial or scalene character, meaning it has no parametric degeneracy and is by almost any definition a freeform surface with no invariance under any translation or rotation. In the standard Cartesian co-ordinate definition, equation (7-1), the inequalities are enforced for a tri-axial ellipsoid surface. Here the eccentricities are aligned in a direction simplifying the surface's manufacture. Given manufacture by turning on the Z axis, the eccentricity of the equatorial ellipse being less than that of the polar ellipse gives a beneficial minimisation of the synchronous motion required during turning.

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, \qquad a<b<c \tag{7-1}$$

The cavity space bounded by the surface is defined where the left hand side of the equality in equation (7-1) is less than 1. That equation can be rewritten in terms of the eccentricities ϵ_1, ϵ_2 as in equation (7-2).

$$\frac{x^2}{a^2}+\frac{y^2}{a^2(1+\epsilon_2)^2}+\frac{z^2}{a^2(1+\epsilon_1)^2}=1, \qquad 1<\epsilon_1,\epsilon_2 \tag{7-2}$$

The design for the cavity, which was made by the author for the experiment, has dimensions given in Table 7-1.

Table 7-1: Cavity parameters

Parameter	eccentricity	Value
Nominal radius, a	0	62.000 mm
Semi major axis, b	0.0005	62.031 mm
Semi major axis, c	0.001	62.062 mm

These parameters were set by the team so as to be a positive compromise between the difficulties of generation or surface measurement – tending to minimisation of: actual size, ellipsoidal eccentricities, weight; and the demands of experimental uncertainties – tending to the maximisation of those same quantities. The quasi-sphere is made from two hemispheres. Each was initially rough-machined from 150 mm copper bar (BS C101: ISO Type Cu-ETP). The choice of copper as the cavity's envelope material was driven by similar compromises between machinability, ease of measurement, ease of damage and experimental demands. Although achievement of the cavity's geometrical specification was the major challenge for the author, some additional features and parts were also manufactured including “spider” (clamp) components in copper, functional apertures and their plugs etc. Some of these other components are depicted Figure 7-2 but their details are outside the scientific scope of interest of this study.

External features of the quasi-spheres are diamond turned, and some of these are key datum/alignment surfaces critical to its assembly. For this reason, they are machined where possible at the same time and in the same set-up as the internal cavity. There are also a number of apertures cut into the cavity to be used for acoustic transducers and sensors. These apertures in the cavity are plugged from outside and diamond machined flush in-situ along with the rest of the cavity’s internal surface, so as to provide a minimally perturbing effect to acoustic resonances; the plugging device perturbations are modelled and characterised in [3].

Successive machining cuts of the internal surface are made, each with a progressively larger nominal radius adhering to the final ellipsoid eccentricity specifications, until the correct (final) radius is achieved.

Figure 7-2: Additional quasi-sphere mounting components

After intermediate cuts and especially the penultimate cut, a measurement is made of the internal surface using the techniques of section 7.3, adapted from chapter 5. A thorough numerical analysis of the measurement data permits the adjustment of cutting tool path on subsequent cuts to compensate for machining errors, and specifically and separately tool offset positions. Each cut, performed in a temperature-controlled environment, takes less than an hour, during which time the machine, workpiece and coolant are experiencing temperature variation of less than 0.25 °C. No final measurement is made using these techniques after the last cut, in order to avoid impact damage to the cavity surface. The surface is subsequently measured using the techniques described in [8] prior to its use in acoustic resonance thermometry.

Table 7-2: Finish machining parameters used for Quasi-sphere

Target parameter	**Value**
Workpiece radius of curvature	62 mm
Freeform (spherical) departure	62 µm max.
Surface form error (peak)	< 3 µm
Surface roughness	< 10 nm R_a
Machining time	< 1 hour
Selected machining parameter	**Value**
Path shape	Out-in, spiral
Tool radius of curvature	1.60867 mm
Tool top rake	0°
Workspindle rotation	200 rpm
Relative tangential feed per revolution	10 µm
Depth of cut	10 µm
Consequential factors	**Value**
Number of turns	9739
Machining time	~ 49 minutes
Path length	2415.3 m
Peak tool-workpiece relative speed	1.298 m/sec
Geometrically resultant cusp height	7.8 nm
Equivalent R_q	2.4 nm
Equivalent R_a	2.1 nm

The machining parameters for the cavity were set by the author based on the requirements for geometry, surface finish, surface form accuracy and machining time. These requirements are discussed in the text and in [3,4,6,8,1].

7.2 Tool path generation

7.2.1 Tool path design

The diamond turning is performed with a fixed-orientation diamond tool having a nominal 1.6 mm tool edge radius with 110 degree included angle and a zero-degree top rake. Zero-degree top rake dramatically simplifies the geometrical basis of the tool path computation; the full complexity of a 3-dimensional tool will be considered in chapter 8. The tool is set to a mid-point angle (45 degrees to workspindle axis – see Figure 7-3) so that a full quarter circle path can be cut without repositioning the tool. The configuration is shown on the Moore Nanotechnology 350 UPL in Figure 7-3 – (see Figure 4-2 for mounting position in machine context).

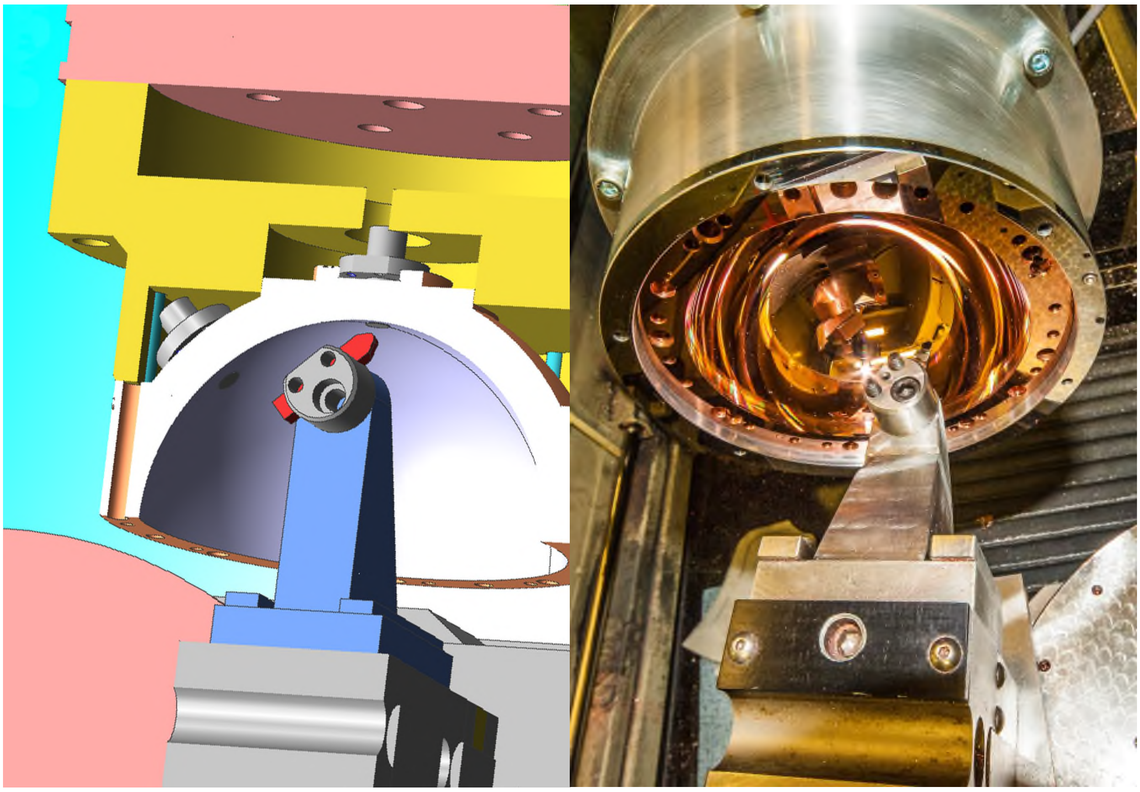


Figure 7-3: Diamond turning configuration (solid model picture credit: Roger Read)

The internal surface form in equation (7-1)/Table 7-1 clearly is *not* rotationally symmetric, so the tool path is not a regular spiral even if the machine motion system were to move without error, if the tool were made without error and if it were set without error in 6 degrees of freedom.

For a conventional zero rake diamond tool selected for fixed-orientation sphere cutting, one of the degrees of freedom might in any case be irrelevant due to rotational invariance (rotational symmetry) around a central axis normal to the tool's top rake face. The character of the path is shown in Figure 7-4.

On the left of Figure 7-4, the 3-D path shows a cut path (the locus of the travelling cutting point) and a TCP path (the locus of the Tool Centre Point). The TCP path is shown on the right in two lines: one for X, one for Z co-ordinate. Two ordinates, each as a function of the rotation of the workpiece (the abscissa) during turning; in both cases the feed pitch is extremely exaggerated, at 15 mm per revolution, which is >1000 times larger (for graphical visualisation) than the feed pitch actually used.

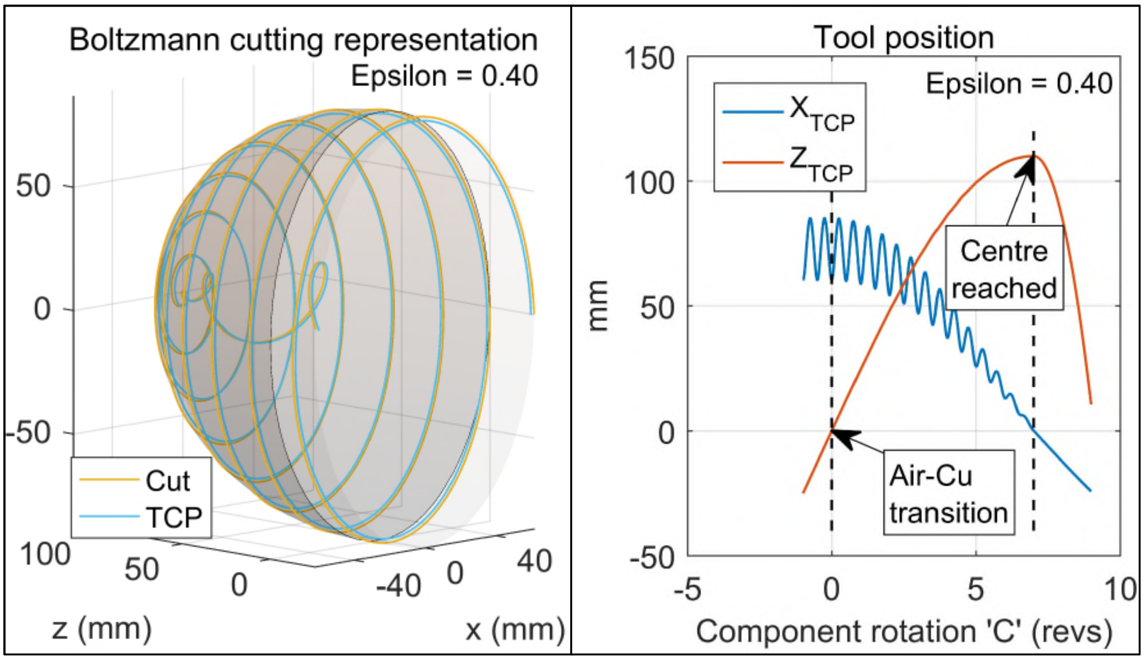


Figure 7-4: Diamond turning tool path – exaggerated feed pitch and eccentricity

By design, the tool path has three phases with different path shapes (in the tool-workpiece relative co-ordinate space):

1. 'cutting' air – describing an elliptical cylinder
2. cutting copper – describing a tri-axial ellipsoid
3. retracting from surface – describing a helix

Some characteristics of the tool path are evident:

- x and z are continuous functions of c throughout and between each of three phases; this is necessary to avoid surface discontinuities
- x and z also have continuous first derivatives w.r.t. C (the workpiece rotation angle)
- the retraction path is designed to have acceleration continuity as the cutting point reaches centre

Acceleration continuity is ensured by maintaining a constant X_{TCP} velocity and the result is a work-tool relative spiral path on retraction, as seen inside the cavity in Figure 7-4 (left).

Figure 7-5 gives a confirmation that the clearance path (the tighter spiral in Figure 7-4, left) falls inside the cutting path so that this retraction motion will not cause surface modification, despite motion continuing past centre. This is critical, since the workspindle rotation is continuous and the machining tool path may not demand instantaneous change of velocity.

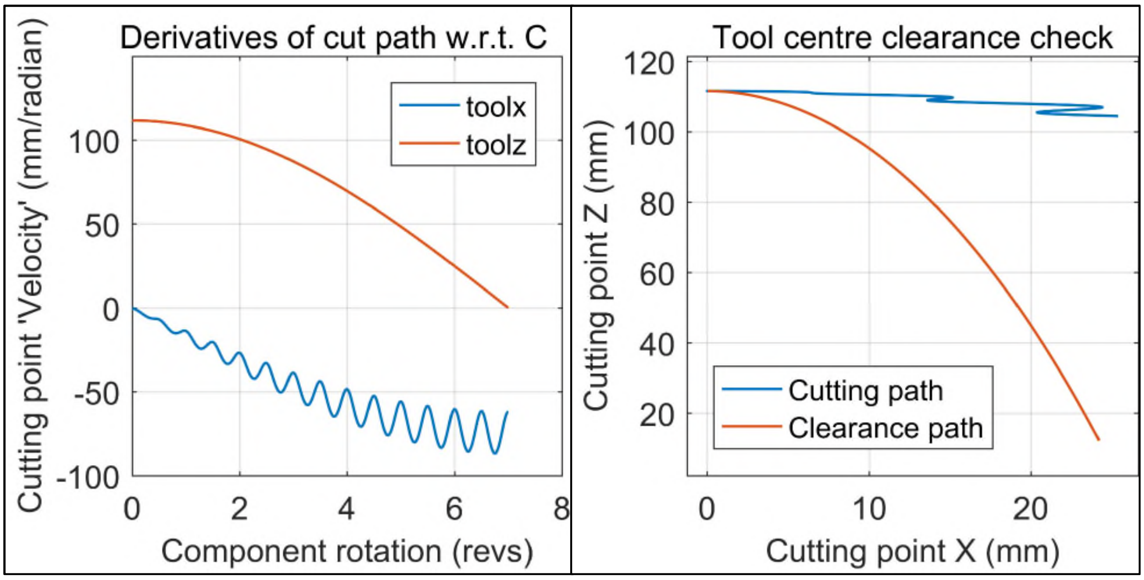


Figure 7-5: Diamond turning tool path parameters – exaggerated feed pitch and eccentricity

The tool used for finish turning has a nominally circular arc cutting edge, formed from the intersection of a plane (set parallel with the turning axis) and a cone (with conical axis normal to the plane) – see Figure 7-6. The cutting edge shape is characterised by its radius and its included angle. Given the tool's use in a fixed orientation, the location of the cutting edge in space can be entirely characterised by Tool Centre Point.

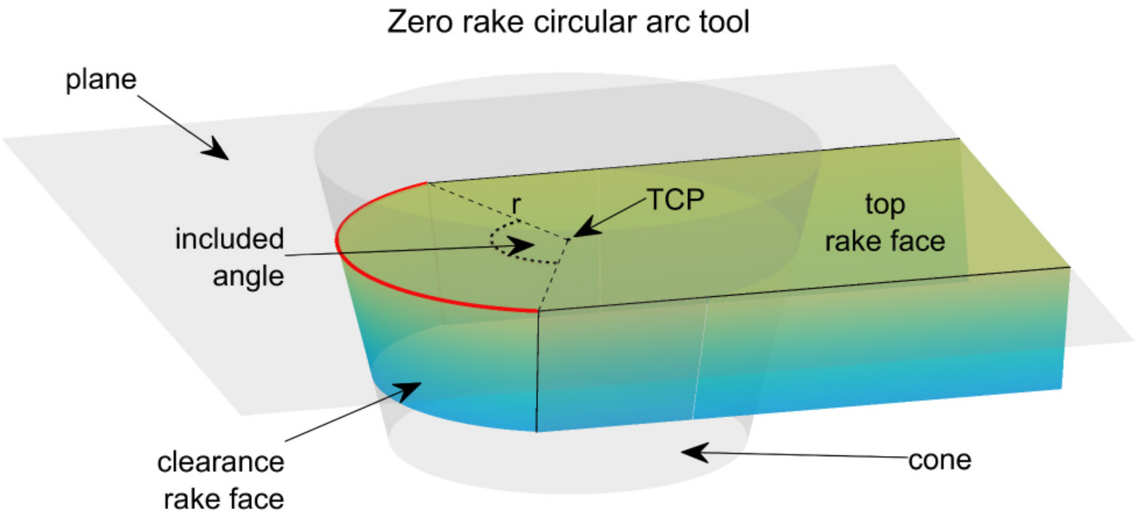


Figure 7-6: Diamond turning tool – simplified geometry

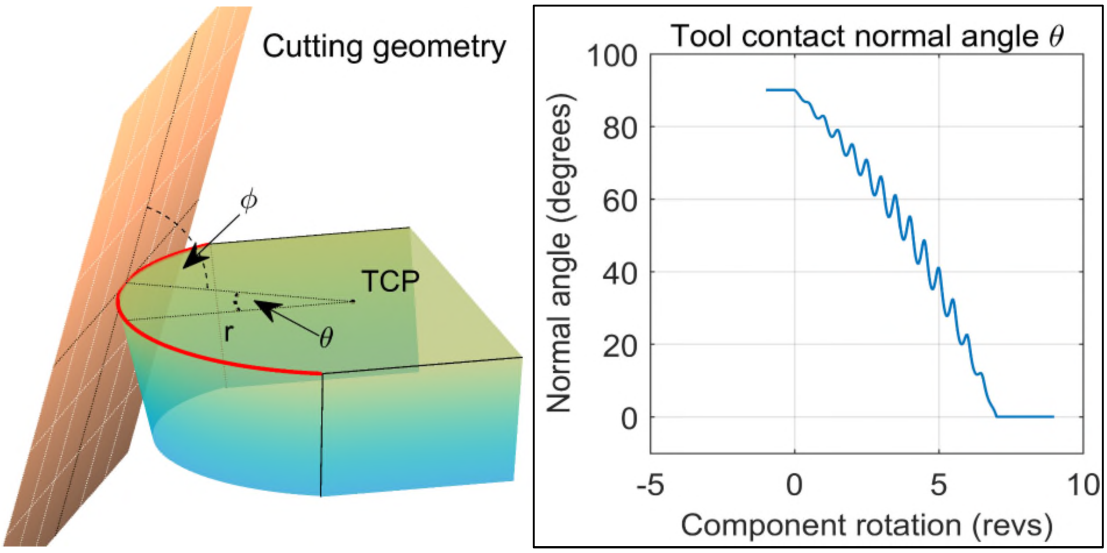


Figure 7-7: Selection of tool contact angle by workpiece surface slope – exaggerated feed pitch and eccentricity (right)

It can be seen from Figure 7-7 (left) that whereas the workpiece contact point's surface slope is described with two angles – θ and ϕ – only θ will play a part in selecting the contact point on the tool's cutting edge. In this unique situation where the tool's top rake angle is zero, the cutting geometry (for the purposes of deriving the tool's cutting point relative to the TCP) is independent of the angle ϕ. This is the major reason for the dramatic simplification of tool path generation for a zero rake tool. Figure 7-7 (right) shows the selection of tool contact normal angle as a function of rotation of the workpiece during cutting, again with feed pitch exaggerated for visualisation.

7.2.2 Analytical tool path derivation

In a turning operation for the quasi-sphere, we can parameterise the operation in terms of number of turns and feed per revolution. For convenience we can set an integer number of turns n as in equations (7-3) based on a nominal (spherical) cavity radius R, where f' is the nominal feed-pitch per revolution, Q is a quarter turn circumference and f is the modified feed per revolution.

$$Q = \frac{\pi}{2}R, \qquad n = \left\lceil {}^{Q}/_{f'} \right\rceil, \qquad f = {}^{Q}/_{n} \tag{7-3}$$

As a result of the 1:1 correspondence for a fixed (spherical) feed-pitch per revolution between workpiece rotation and angle in the x-y plane, we can directly deduce the ranges and relationships of workpiece rotation angle C and ψ as given in equations (7-4), where ψ is the spherical angle from the centre of the ellipsoid.

$$0 \leq C \leq 2\pi n, \qquad \psi = {}^{C}/_{4n}, \qquad 0 \leq \psi \leq {}^{\pi}/_{2} \tag{7-4}$$

The effects of eccentricity on variation of machining parameters (compared with a purely spherical cavity) are negligible for diamond turning so the spherical approximation is appropriate for calculations of feed rate etc. (though not of precise co-ordinates). In the relative Cartesian geometry of equation (7-1), precise co-ordinates are now obviously as given in equations (7-5), from the eccentricity $\epsilon = \epsilon_2 = \varepsilon_1/2$ of Table 7-1.

$$\begin{gathered} x = R\cos\psi\cos C, \\ y = R\,(1+\epsilon)\cos\psi\sin C, \qquad z = R(1+2\epsilon)\sin\psi \end{gathered} \tag{7-5}$$

Given the 1:1 correspondence of C to ψ in equations (7-4), these equations are parametric in a single parameter - C. In the cylindrical co-ordinates of the machine (C, X, Z) where $X = \sqrt{x^2 + y^2}$ the cutting point locus is therefore given by equations (7-6) and this is the simple basis of the tool path program.

$$\begin{aligned} &0 \le C \le 2\pi n \\ &X = R\cos\left(\frac{C}{4n}\right)\sqrt{\cos^2 C + (1+\epsilon)^2 \sin^2 C} \\ &Z = R(1+2\epsilon)\sin\left(\frac{C}{4n}\right) \end{aligned} \tag{7-6}$$

This is exactly correct for the cutting point in the cavity, as a function of C.

7.2.2.1 Tool radius compensation

The tool path program must be written for a single point on the tool, independent of the contact point; the position of the latter will vary around the tool's cutting edge depending on the workpiece's local surface angle. The logical single point is the Tool Centre Point and the locus of that can be obtained from equations (7-6) by adding the offset from contact point to TCP, which can be obtained using sine and cosine of θ, which is determined by the local angle of the workpiece surface where it contacts the tool.

Although ψ and θ are related, they are not equivalent, except in the case where the cavity is purely spherical. The slope of the elliptical section in the x-y plane is given by the ratio of the parametric partial derivatives of Z and X and its slope angle by equation (7-7). A point to note is that although C and ψ are related by a constant of proportionality in equations (7-4), this is solely in the definition of the tool path. Across the surface of the cavity in general, C and ψ are differently related, and as the partial differentiation in equation (7-7) is in the x-y plane, C is invariant.

$$\theta + \frac{\pi}{2} = \tan^{-1}\left(\frac{\partial Z/\partial\psi}{\partial X/\partial\psi}\right) = \tan^{-1}\frac{-Z}{X\tan^2\psi} \tag{7-7}$$

In equation (7-7) the well-known two-argument high precision variant of atan2 should be used for the arctangent computation, as given in equations (7-8), since both numerator and denominator of equation (7-7) are inevitably close to zero at one or other end of the tool path.

$$\operatorname{atan2}(b,a) = \begin{cases} 2\tan^{-1}\left(\dfrac{b}{\sqrt{a^2+b^2}+a}\right) & \text{if } a > 0 \\ 2\tan^{-1}\left(\dfrac{\sqrt{a^2+b^2}-a}{b}\right) & \text{if } a \le 0 \text{ and } b \ne 0 \\ \pi & \text{if } a < 0 \text{ and } b = 0 \\ \text{undefined} & \text{if } a = 0 \text{ and } b = 0 \end{cases} \qquad (7\text{-}8)$$

This gives the analytically correct tool path for generation of the tri-axial ellipsoid in the absence of error, as shown in equations (7-9), where n, C, X, Z & θ are given in the foregoing and r is the tool radius.

$$0 \le C \le 2\pi n, \qquad X_{TCP} = X + r\sin\theta, \qquad Z_{TCP} = Z + r\cos\theta \qquad (7\text{-}9)$$

7.2.3 Sources of error in the tool path.

Five sources of error are considered: tool shape error, variations in machining force, dynamic positioning errors, tool setting errors and motion system static errors.

7.2.3.1 Tool shape (radius) error

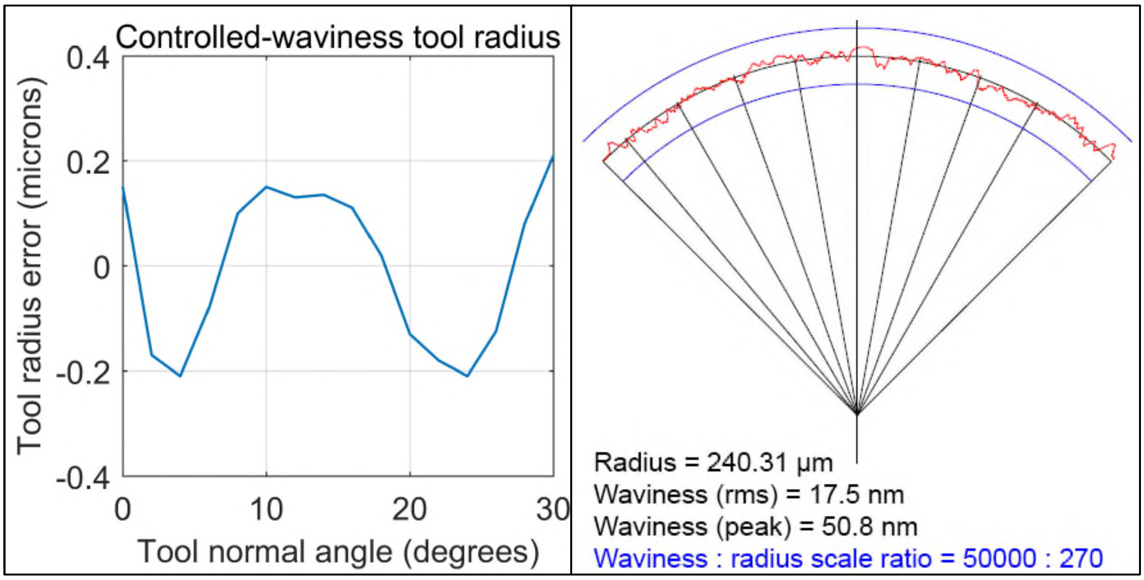


Figure 7-8: (Left) - measurement of typical controlled-waviness tool radius w.r.t. selected tool angle [211], (right) – representation of the highest commercially available quality, after [212]

The cutting edge of a radius tool is subject to error due to a number of factors including uneven tool wear and lapping efficacy against crystalline anisotropy during shaping of the tool for its manufacture. The higher accuracy “controlled-waviness" tools exhibit a waviness (radius variation) which is typically $\lesssim$ 500 nm as mapped by the author [214], and commercially available (exceptionally) down to a claimed 50 nm p-v [215].

A tool map, such as the ones shown in Figure 7-8 can be used to improve the tool path. In this case equations (7-9) can be modified to give r as a function of θ where r can be given by a simple interpolation of the tool error map.

The tool error map can be built from the techniques described by the author in [216,214]. The contribution of these errors is typically $\lesssim$ 500 nm for controlled-waviness tools.

7.2.3.2 Machining forces

Cutting forces for machining of ductile metals in this removal regime are measured in milli-Newtons [217] and typically < 100 mN, whereas dynamic forces, particularly for slow-slide machining of non-rotationally symmetric parts are very much larger, as will be seen. For each successive turning operation on the cavity's internal surface, nominal (spherical) cavity radius target R is increased by the depth of cut. In this way, the tool path design gives uniform un-deformed chip thickness throughout a cut as the cavity shape is maintained; the radius is increased uniformly across the surface.

The chip width is dependent on the feed per rev; any increase material removal rate per unit path length is a product of un-deformed chip thickness and increase in feed per rev. The effect on machining forces however will be dominated (in machining at a constant rpm) by variation in relative surface speed, which decreases to zero at centre. Nevertheless, peak force, which will occur when relative workpiece-tool surface speed is greatest (at the cavity 'equator') is still $\ll$ 1N and slowly varying, operating against a machine stiffness measured in 10's N/µm or > 10^7 N/m.

7.2.3.3 Dynamic motion forces

Empirically, dynamic motion forces have been assessed to be the largest factor affecting accuracy – potentially larger than all other sources combined. Dynamic motion errors are (broadly) an increasing function of the magnitude of acceleration, although frequency domain analysis is necessary fully to quantify the effects. The author has analysed the machine to assess its suitability and limits for dynamic machining of freeform surfaces [18] in developing a compensation strategy for freeform machining.

Figure 7-9 shows the dynamic position data from [18] for the Z axis; data from the X axis shows the same characteristic in that dynamic accuracy is maintained at a high level up to a frequency of a few Hz. These data are captured by comparing a programmed sinusoidal command against the response measured in two places by capacitance gauging. The measurement positions are 1) close to the machine's measurement feedback encoder and 2) at the position of the turning tool. The graph plots the amplitude ratio (ordinate) of response to command for a sinusoidal excitation of 7.5 µm amplitude as a function of frequency (abscissa) in the range 1 – 40 Hz. Significant departure of amplitude ratio from 0dB indicates dynamic motion errors in response to a sinusoidal motion command at that frequency. The graph indicates response at the location of the machine's position encoder (indicating the quality of servo control) and separately at the tool height, which indicates its functional performance for turning, including the effects of bearing stiffness etc. Two cursors at specific frequencies are added to the graph, which have particular significance for this study. 200 rpm has been selected as the workspindle speed for this research, on the basis of adequate dynamic stiffness at that frequency for dynamic control.

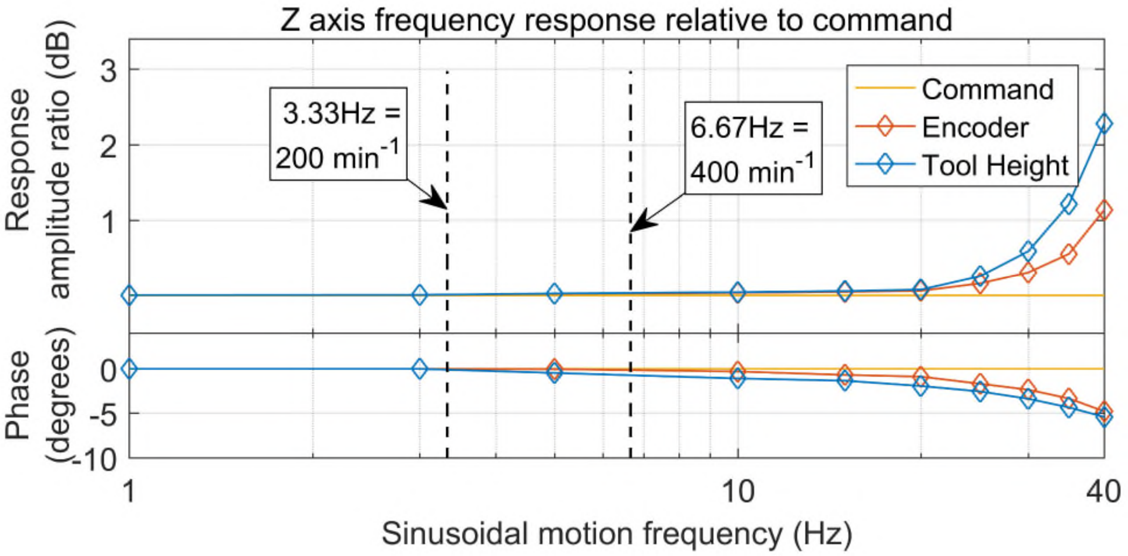


Figure 7-9: Dynamic response of Z axis

Machining the tri-axial ellipsoid at 200 rpm imparts a sinusoidal motion at twice the rotation frequency (400 min^{-1} or 6.67 Hz) with an amplitude (in X) up to 31 µm. The response amplitude error at tool height to a sinusoidal excitation of that nature is 93 nm, with a phase lag of 1.1°, based on the analysis of measurements in [18]. This is at or below the level of the contribution to surface form error from tool shape error of a controlled-waviness tool.

For comparison to the machining forces, a calculation can be made of dynamic stiffness at this frequency; the quantitative comparison is not entirely accurate since the dynamic reaction force is through the centre of mass rather than at tool height, but the comparison is at least indicative.

$$X = A\cos\omega t, \qquad \ddot{X} = -A\omega^2\cos\omega t = {F(t)}/{M} \tag{7-10}$$

In equation (7-10) the moving mass M is approximately 100Kg, the sinusoidal amplitude A is at most 31 µm and the angular frequency ω is approximately 42 radians/sec, which indicates a peak sinusoidal force of around 5.5 N and a dynamic stiffness of 60 N/µm at that frequency, subject to the proviso regarding centre of reaction, above. In comparison with the dynamic motion forces, variations in the largely uniform machining force, whose absolute value is in any case substantially below 1N, can be neglected.

7.2.3.4 Tool setting errors

An ultra-precision tool setting system has been developed by the author and reported previously [216,214] which measures setting parameters directly by optical/image processing-based examination of the tool. In this instance however, tool setting data were extracted from measurement of the machined cavity and adjustments fed back to the machining program. This will be discussed in section 7.3.

7.2.3.5 Motion system static errors

The diamond turning machine has inherent positioning errors which will contribute to the accuracy of any finished surface. These motion system errors can be tackled in conventional ways 1) by Machine Tool Builders' machine calibration procedures, and 2) by measurement feedback from the finished surface, to adjust the machining program, as discussed in sections 6.2 and 7.2.4

7.2.4 Error compensation

The process of error compensation can be adopted to benefit where measurement errors are lower than the combination of machining errors, and in particular motion system errors; if that is not the case, measurement feedback will introduce larger errors. Where this is the case, the strategy of 6.2 can be used. Either the contact point path can be adjusted according to the measurement error and then the correction for TCP applied using the surface normal vector from the unmodified contact point path - or the surface representation will need to be made into a point cloud and surface normals generated from it.

The accuracy of the machining in the case of the quasi-spheres is high enough for there to be no advantage in error compensation from measurement feedback, although this technique will be fully detailed in the much more complex case of large optics grinding, in Chapter 8.

One aspect of measurement data for the quasi-spheres is however fed back for compensation, which is the data extracted from the measurement in respect of tool offset and workpiece orientation.

7.3 Measurement Procedure

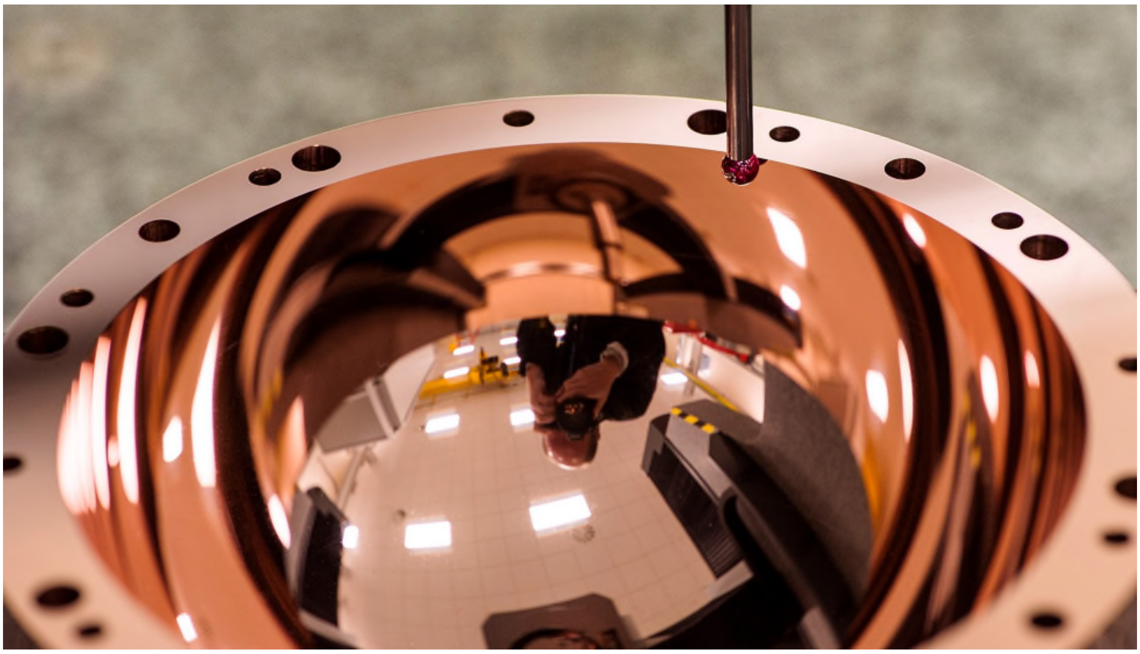


Figure 7-10: CMM measurement of internal cavity surface form

The measurement procedure adopted and used in support of manufacture is broadly that of chapter 5, adapted for high curvature surfaces. Measuring an internal hemisphere using CMM scanning, requires a stylus shaft which is parallel to the spherical rotation axis (Figure 7-10) in order to avoid mechanical interference. This is distinct from the (mechanical) technique of chapter 5 where an inclined stylus shaft is used, and although it has some influence on achievable scanning speed, it has no influence on the mathematical approach.

7.3.1 Establishment of co-ordinate reference frame

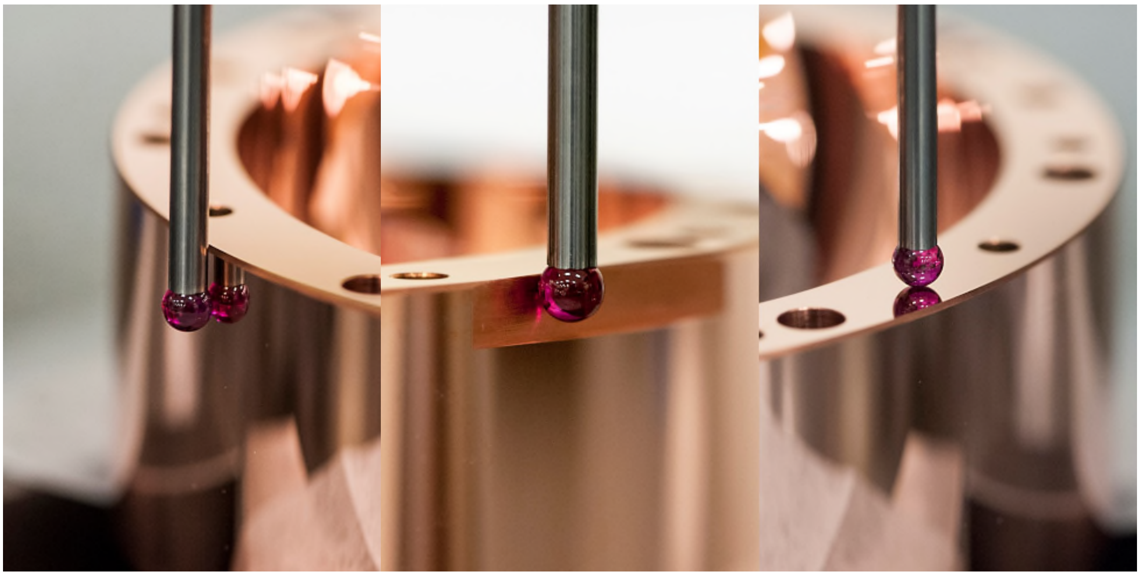


Figure 7-11: CMM measurement of quasi-sphere alignment features: external cylinder (constraining x & y), flat (constraining α_z) and equator (constraining α_x, α_y & z)

Figure 7-11 shows probing, using the same measurement set-up as for the cavity, of the external alignment features. The alignment features are machined on the same set-up as the cavity. This ensures that alignment of these external features on assembly will yield an internal surface with minimised equatorial discontinuities.

The indicated angular constraints (constraining degrees of freedom of alignment) are given as α_x, α_y and α_y – linear constraints as x, y and z. These external alignment features are used on the CMM to establish the co-ordinate reference system for the scan measurement.

7.3.2 *Scan line trajectory*

Scans of the cavity surface are taken primarily radially or as circles. In agreement with the strategy of section 5.3; this minimises probe reversals (maintaining monotonicity of probe deflections) during each scan line.

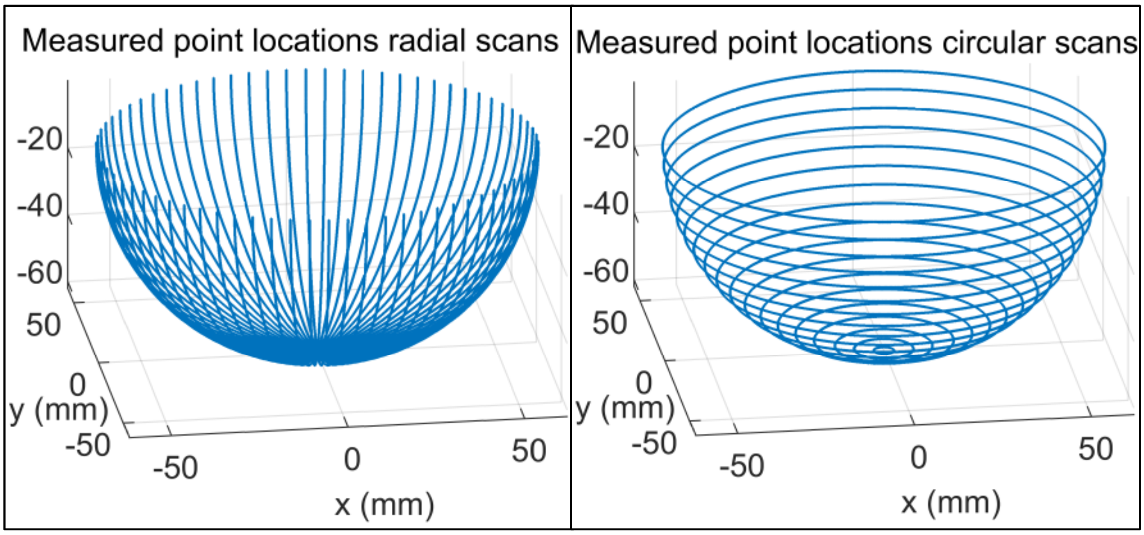


Figure 7-12: Actual CMM measurement data: radial (left) and circular (right) scans

7.3.2.1 Probe radius compensation

Probe radius compensation is applied within the CMM program, based on a spherical measurement assumption. The maximum error in this assumption is calculable, and is vanishingly small for a low eccentricity ellipsoid. Based on a Taylor Series expansion for arctangent, the maximum error is given in equation (7-11) and for this ellipsoid is below 1 nm.

$$max.\,error_{\mathrm{rad\ comp}} = r_{probe}\left(1-\cos\left({}^{\epsilon}/_{2}\right)\right) \approx 0.3\ \mathrm{nm} \qquad (7\text{-}11)$$

These probe radius corrected measurement data are given in Figure 7-12 as linear scans

7.3.3 *Measured data processing and parameter extraction*

The same data are represented in Figure 7-13 , Figure 7-14 and Figure 7-16 as 3D projections of conformal Azimuthal maps of the hemisphere, with the equator at the outside circumference and the pole at the centre, as departure from a fitted surface.

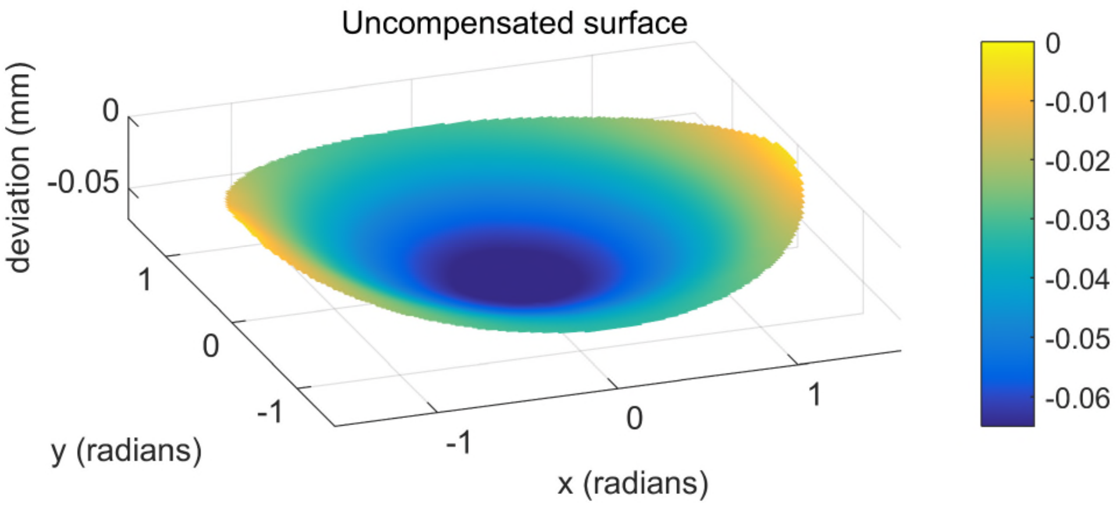


Figure 7-13: Actual CMM measurement data giving ellipsoidal departure from sphere

Figure 7-13 shows the ellipsoidal departure from best fit sphere. Note the vertical scale of 65 microns. The expected scalene character of the ellipsoid can be seen, with 62 microns of departure at the axial (Z) pole and 31 microns at the Y poles – as indicated by the colour scale. Note that in these surface maps, the plotted value is departure or deviation normal to the surface, **<u>not</u>** height error in the Cartesian Z co-ordinate.

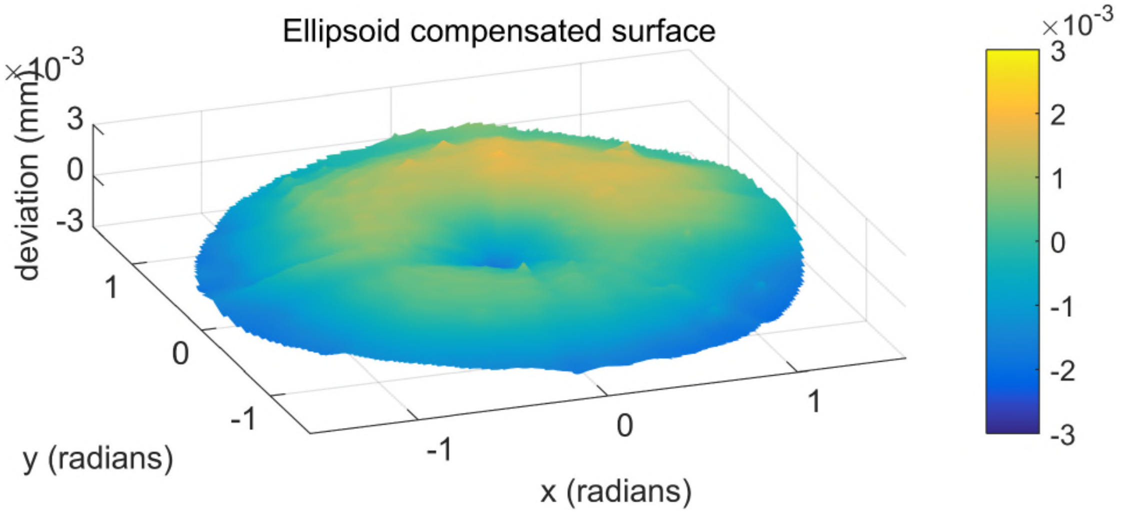


Figure 7-14: Actual CMM measurement data with nominal ellipse subtracted

Figure 7-14 shows the same data, but with the ellipse removed, so the map shows no eccentricity. What is visible is primarily the consequence of tool setting error giving a familiar ‘M’-shaped profile [218] for a tool past centre on a concave rotationally symmetric turned form. It is seen here in a surface of rotation in a 3-D representation, although Figure 7-16 shows the more familiar cross-section, taken from the same data.

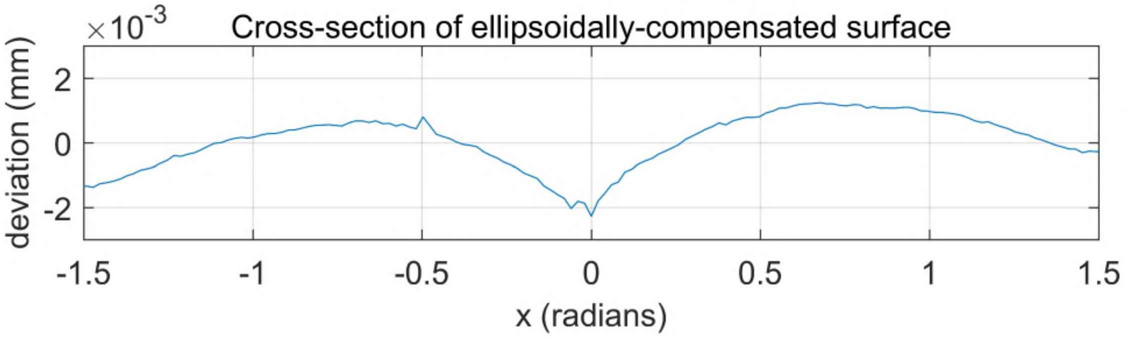


Figure 7-15: Actual CMM measurement showing cross-section from Figure 7-14

Two factors in particular should minimise workpiece offset errors in the measurement: 1) the same diamond tool in the same machining set-up from within the same CNC program turns the external cylindrical face and equatorial face for the quasi-sphere as turns the cavity, although this is not possible with the flat; 2) the CMM probe measures these alignment features and references cavity measurement to them. This should ensure minimal z alignment offsets and minimal α_x, α_y alignment angle offsets. The α_z alignment relies equally on a manual measurement and setting on the diamond turning machine, on the CMM measurement of a relatively short flat and on the sinusoidal phase lag during machining, which is (Figure 7-9) approximately 1.1 degrees. Therefore, this and the tool-setting centre offset are potential sources of surface error location and shape error respectively. These can be assessed through analysis of the cavity measurement.

7.3.3.1 Extraction of machining adjustment parameters

Taking as a starting point equation (7-1) we can introduce terms for the axial-rotation of the ellipsoidal form, by an angle δ, and the radial-shift of the tool by an offset τ, in the machine’s x direction. Tool shift in the z direction is not relevant since it has no effect on shape or shift relative to the alignment features. Tool shift in the machine’s y direction has a much smaller, near negligible effect on shape error [218] and can be independently adjusted by visual inspection of the centre defect after turning. A z-axial rotation, a (machine) x-shift, and workpiece x shift x_0 are represented for x in equation (7-12) and similarly for y.

$$x \xrightarrow[\delta-\text{rotation},\tau-\text{shift}]{} (x\cos\delta - y\sin\delta - x_0)^2\left(1-\frac{\tau}{\sqrt{x^2+y^2}}\right)^2 \tag{7-12}$$

This gives a modified Cartesian representation of the ellipsoid as in equation (7-13).

$$\left[\frac{(x\cos\delta - y\sin\delta - x_0)^2}{a^2}+\frac{(x\sin\delta + y\cos\delta - y_0)^2}{b^2}\right]\left(1-\frac{\tau}{\sqrt{x^2+y^2}}\right)^2 + \frac{z^2}{c^2} = 1 \tag{7-13}$$

This is a non-linear fitting problem: find the five unknowns a, b, c, δ & τ which represent the best fit to the x, y, z data set. There are several ways to tackle this, and clearly the Levenberg-Marquardt [184] algorithm (or similar) is suggested. However, by inspection, it's clear that solution values of δ will be independent of the relationship between τ and a, b so a simplification is possible. Setting τ to 0 to solve for δ, we can multiply equation (7-13) out and group terms in powers of x, y and z^2 to give equations (7-14) – z is deliberately omitted.

$$\begin{aligned}
x^2, &\quad U_{(x^2)} = \frac{\cos^2\delta}{a^2}+\frac{\sin^2\delta}{b^2}\\
y^2, &\quad U_{(y^2)} = \frac{\sin^2\delta}{a^2}+\frac{\cos^2\delta}{b^2}\\
z^2, &\quad U_{(z^2)} = \frac{1}{c^2}\\
xy, &\quad U_{(xy)} = 2\cos\delta\sin\delta\left(\frac{1}{b^2}-\frac{1}{a^2}\right)\\
x, &\quad U_{(x)} = \frac{-2(x_0\cos\delta + y_0\sin\delta)}{a^2}\\
y, &\quad U_{(y)} = \frac{2(x_0\sin\delta - y_0\cos\delta)}{b^2}\\
1, &\quad U_{(1)} = \frac{{x_0}^2}{a^2}+\frac{{y_0}^2}{b^2}-1
\end{aligned} \tag{7-14}$$

7.3.3.2 Extraction of alignment parameters

A solution can be found for x_0, y_0 and δ by minimising the sum in equation (7-15) over all i data points (the set in the preceding figures have 140,000 data points) to find the set of U values, using the usual techniques of linear algebra; the author favours singular value decomposition.

$$\begin{aligned}\sum_i U_{(x^2)}{x_i}^2 + U_{(y^2)}{y_i}^2 + U_{(z^2)}{z_i}^2 + U_{(xy)}x_i y_i + U_{(x)}x_i + U_{(y)}y_i \\ + U_{(z)}z_i + U_{(1)}\end{aligned} \tag{7-15}$$

Simple manipulations then yield values for δ, x_0 and y_0 in equations (7-16).

$$\begin{gathered}\frac{U_{(xy)}}{U_{(y^2)} - U_{(x^2)}} = \frac{2\cos\delta\sin\delta\left(\frac{1}{b^2} - \frac{1}{a^2}\right)}{(\cos^2\delta - \sin^2\delta)\left(\frac{1}{b^2} - \frac{1}{a^2}\right)} = \frac{\sin(2\delta)}{\cos(2\delta)} \\ \therefore \delta = \frac{\tan^{-1}\left(\frac{U_{(xy)}}{U_{(y^2)} - U_{(x^2)}}\right)}{2} \\ \frac{(\cos^2\delta - \sin^2\delta)\left(U_{(y)}\sin\delta - U_{(x)}\cos\delta\right)}{2U_{(x^2)}\cos^2\delta - 2U_{(y^2)}\sin^2\delta} = x_0 \\ \frac{(\cos^2\delta - \sin^2\delta)\left(U_{(x)}\sin\delta + U_{(y)}\cos\delta\right)}{2U_{(x^2)}\sin^2\delta - 2U_{(y^2)}\cos^2\delta} = y_0\end{gathered} \tag{7-16}$$

7.3.3.3 Extraction of tool offset parameter

To solve for the tool offset τ and the ellipsoid semi-major axes a, b & c it is convenient to regenerate data shifted to centre and rotated to alignment by applying the computed offsets, as in equations (7-17) and recreating the ellipsoid equation.

$$\begin{gathered}x_{\mathrm{r}} = x\cos\delta - y\sin\delta - x_0, \qquad y_{\mathrm{r}} = x\sin\delta + y\cos\delta - y_0 \\ \left[\frac{(x_{\mathrm{r}} - x_0)^2}{a^2} + \frac{(y_{\mathrm{r}} - y_0)^2}{b^2}\right]\left(1 - \frac{\tau}{\sqrt{{x_{\mathrm{r}}}^2 + {y_{\mathrm{r}}}^2}}\right)^2 + \frac{(z - z_0)^2}{c^2} = 1\end{gathered} \tag{7-17}$$

Repeating the procedure of equations (7-14) with the regenerated data, gives coefficients in equations (7-18).

$$
\begin{aligned}
&U_{(x^2)} = \frac{1}{a^2}, &&U_{(y^2)} = \frac{1}{b^2}, &&U_{(z^2)} = \frac{1}{c^2}\\
&U_{(x)} = -\frac{2x_0}{a^2}, &&U_{(y)} = -\frac{2y_0}{b^2}, &&U_{(z)} = -\frac{2z_0}{c^2}\\
&U_{\left(x^2/\sqrt{x^2+y^2}\right)} = -\frac{2\tau}{a^2}, &&U_{\left(y^2/\sqrt{x^2+y^2}\right)} = -\frac{2\tau}{b^2} &&\\
&U_{(1)} = \frac{{x_0}^2}{a^2} + \frac{{y_0}^2}{b^2} + \frac{{z_0}^2}{c^2} - 1 && &&
\end{aligned}
\tag{7-18}
$$

Higher order terms in τ can be neglected, given the high ratio of ellipsoid radius $\sqrt{x^2+y^2}$ to tool offset τ. Repeating the procedure of equation (7-15), the coefficients can be obtained as follows in equations (7-19). In equations (7-19), $\boldsymbol{\mathcal{U}}$ is a normalising factor used to ensure coefficients are scaled so that the RHS in equations (7-1), (7-2) and (7-13) is set equal to 1. Offsets x_0 & y_0 (and indeed z_0) should be very close to 0 following previous data regeneration but can be adjusted in this fitting.

$$
\begin{aligned}
&x_0 = -{U_{(x)}}\big/{2U_{(x^2)}}, \qquad y_0 = -{U_{(y)}}\big/{2U_{(y^2)}}, \qquad z_0 = -{U_{(z)}}\big/{2U_{(z^2)}}\\
&\boldsymbol{\mathcal{U}} = {U_{(x)}}^2\big/4U_{(x^2)} + {U_{(y)}}^2\big/4U_{(y^2)} + {U_{(z)}}^2\big/4U_{(z^2)} - U_{(1)}\\
&a = \sqrt{\boldsymbol{\mathcal{U}}/U_{(x^2)}}, \qquad b = \sqrt{\boldsymbol{\mathcal{U}}/U_{(y^2)}}, \qquad c = \sqrt{\boldsymbol{\mathcal{U}}/U_{(z)}}\\
&\tau = -\frac{a^2}{2}U_{\left(x^2/\sqrt{x^2+y^2}\right)} = -\frac{b^2}{2}U_{\left(y^2/\sqrt{x^2+y^2}\right)}
\end{aligned}
\tag{7-19}
$$

This gives however a degenerate solution pair for τ, although the values of the independent equalities are very close and a suitable evaluation can be obtained by taking a mean. As an alternative, values for the semi-major axes can be obtained from (7-14) and the data regeneration in (7-17) can produce a data set mapped to a sphere. In that case τ can be obtained from the fitting coefficient in equation (7-20) since the ellipsoid is now degenerate (spherical) with a single radius value giving instead a single equality for τ.

$$
U_{\left((x^2+y^2)/\sqrt{x^2+y^2}\right)} = -\frac{2\tau}{a^2}
\tag{7-20}
$$

This approach has an equivalent issue however, since the initial fitting of a and b (and their ratio) will be to some extent dependent on the actual value of τ. The approach of equations (7-19) has proved highly accurate and entirely satisfactory. This approach is validated in Figure 7-15 where for the same measurement data as shown as in Figure 7-12, Figure 7-13 and Figure 7-14, here the tool offset has been adjusted using the parameters obtained in equations (7-16) and (7-19) and the tool offset effect is absent.

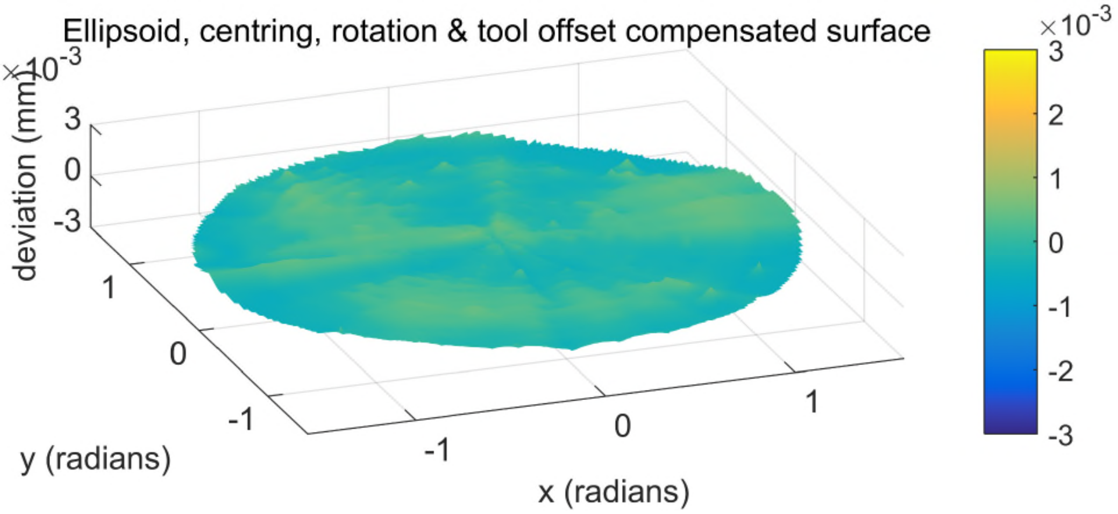


Figure 7-16: Actual CMM measurement data with residual error after rotation, centring, tool setting and best fit ellipse have been numerically removed

These fitted parameters are now used to adjust the machining program for the final surface cut.

7.4 Results

The quasi-spheres, as well as being measured by the technique explained herein have been measured at Cranfield using stitching interferometry to confirm shape and at NPL a) using a different CMM procedure [8], using pyknometry [4] and using microwave resonance techniques [8]. These measurements have confirmed, to an overall uncertainty of < 11 nm the absolute value of the radius at within 500 nm of design target and moreover that almost the entire surface is within 1000 nm of its design target, in terms of absolute size and shape; i.e. including form error and absolute dimension. This represents a considerable achievement for freeform machining and is an important factor in ensuring that the acoustic resonance thermometry conducted using the quasisphere (Figure 7-17, left) was the most accurate performed, with the lowest uncertainty, and will have the largest influence on the new definition of the Kelvin and value of the Boltzmann constant, to be adopted in 2018.

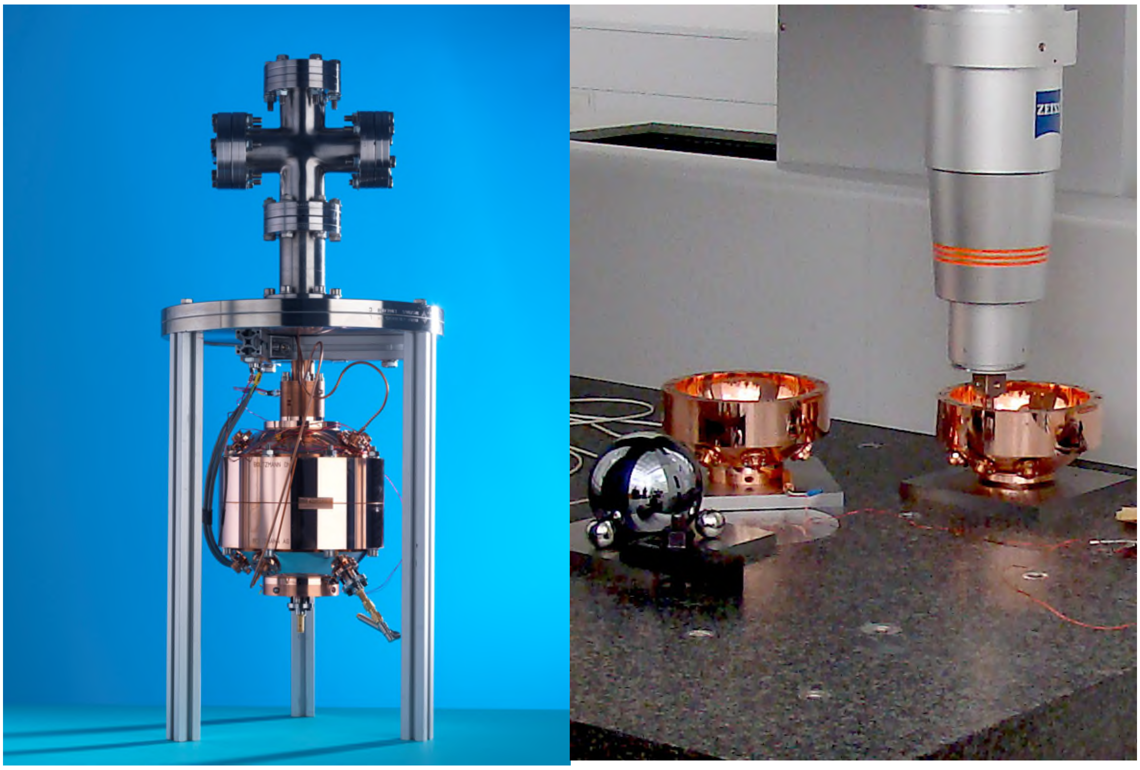


Figure 7-17: (Left) finished quasi-sphere assembled, (right) CMM comparator measurements (picture credits, NPL)

7.5 Summary and discussion

The geometric and surface quality demands, set by an experiment in acoustic thermometry, for a tri-axial ellipsoidal cavity in copper have been considered in respect of the most suitable manufacturing technique of Single Point Diamond Turning. Whilst the surface finish is easily achievable with diamond turning, the surface form accuracy requirements are demanding, given the complex interplay of freeform geometry with: parametric programming, machine dynamics, tool radius compensation, tool/workpiece setting offsets and full surface metrology and parameter extraction.

The mathematical treatment of surface geometry has been considered in respect of tool path creation including the requirements of control path continuity, time derivative continuity, process efficacy-oriented machining parameters, and flexible tool radius compensation. Sources of error in tool path generation were analysed and a quantitative comparison made. The relationship between critical dynamic positioning errors and surface accuracy has guided the optimal choice of machining rate. The key parameters of the other potentially uncontrolled error source, tool and workpiece setting, have been directly inferred from measurement data to provide a precise adjustment regime.

7.5.1 Achievement of objectives and contribution to knowledge

A team of scientists (including the author) led from NPL have carried out a new low-uncertainty measurement of the Boltzmann constant [3]; this can be described as the most accurate temperature measurement ever performed [39]. The experimental error budget called for a surface form tolerance of 3 µm maximum error from design surface – to include shape and radius errors. The final achievement [8] was just over 1 µm, made up of 487 nm ± 11 nm of radius error just over ± 500 nm of form error for a freeform surface made using synchronous motion (slow-slide) diamond turning. One of the dimensional confirmation procedures was comparator measurements with a spherical standard at NPL – see Figure 7-17 (right). The represents the achievement of the specific objective.

7.5.1.1 An effective machining strategy for a tri-axial ellipsoid

A contribution is the development a tool path programming scheme for a tri-axial ellipsoid, using an assessment of machine dynamics and other error sources. This has been achieved with exceptional results.

7.5.1.2 A high resolution & accuracy surface measurement of a hemi-ellipsoid

The second contribution is the use of high resolution scanning metrology, based on the principles of chapter 5 (developed for low-slope surfaces) adapted for a full hemisphere. The success of this approach has been demonstrated by the corroboration with multiple low uncertainty measurement principles at NPL.

7.5.1.3 Accurate extraction of tool and work offsets from a free-form surface measurement

The third contribution is the development of a method for the extraction of work and tool setting errors from the measurement of a freeform surface. This has been demonstrated in that it was the sole setting technique employed for setting adjustment. Tool setting adjustment is absolutely critical for full hemisphere cutting, so the achievement of the low uncertainty machining result is a validation of this new tool setting technique.

7.5.2 Measurement strategy

The final surface is not measured by the CMM scanning technique, so as to avoid surface damage. The surface is measured after the penultimate cut and then another 10 µm depth cut is taken using a program modified only in terms of radius. Thereafter the surface is not measured with a contact scanning technique. The Leitz PMM-F uses a low force probe head with 180 mN probing force. Even this force is enough however to cause non-negligible surface damage to a diamond-turned copper surface. Surface damage was analysed by the team led from NPL (including the author) [8] for single point measurements.

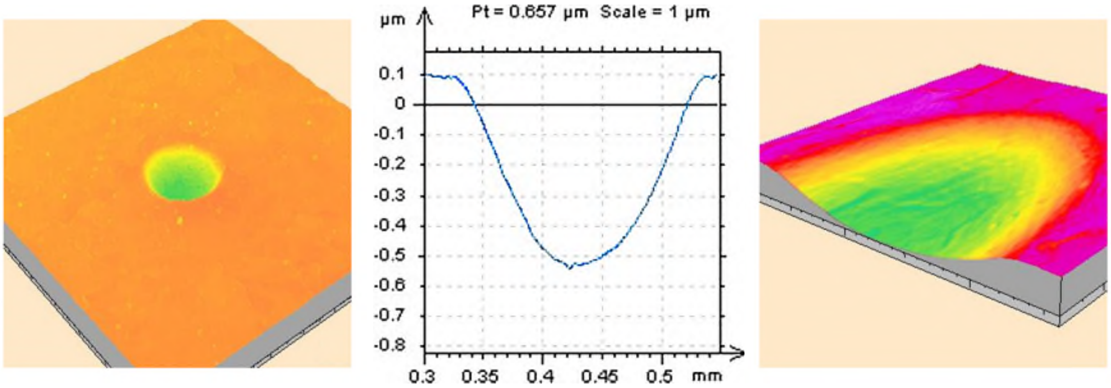


Figure 7-18: Impact indentation from single point probing of copper (picture credit, NPL) [216]

Unpublished work [219] by members of the team has included single point and scan measurements at various levels of probe force made with an 8mm diameter ruby ball stylus tip and Zeiss UPMC 550 CMM. Figure 7-18 shows a single point indentation of 650 nm depth made with 100 mN force at 5 mm/sec approach speed. Figure 7-19 shows scan measurement indentations in copper made with 100 mN probe force and 1 mm/sec scan speed. Indentation depth approaches 100 nm.

Both figures show measurements made using white light interference microscopy. With a reduction in approach speed, single point probing measurements' impacts can be reduced to an almost unmeasurable level, whereas this is not possible with scanning.

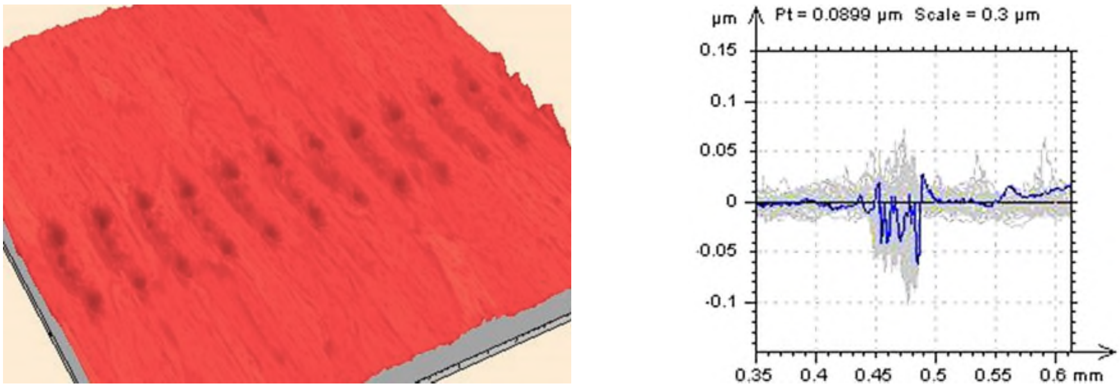


Figure 7-19: Indentation (scratch) from scan measurement of copper (picture credit, NPL) [216]

An alternative *non-contact* measurement technique investigated for confirmation of final surface form was carried out on-machine using a Fisba µPhase 2 OT Twyman Green Phase Shifting Interferometer carrying a 12/12 (60°) NA=0.5 lens. Figure 7-20 shows the interferometer mounted on machine so that it could be used within a machining operation cycle, although not in process.

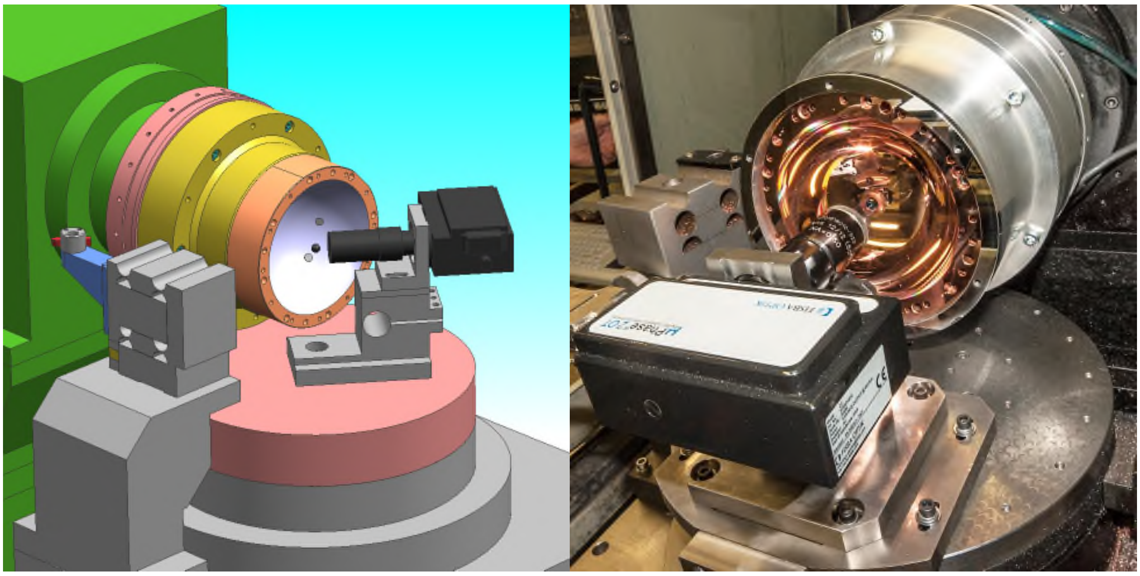


Figure 7-20: (Left), Twyman Green phase shifting interferometer mounted in-situ for on-machine measurement; (right) solid model picture credit: Roger Read

Accuracy using stitching proved inadequate to surpass the achievements of the CMM scanning technique. Contributions to uncertainty include stitching numerical accuracy, 6 DOF mounting alignments of the interferometer, calibration uncertainty of the lens. Figure 7-21 shows typical interferograms and an error map showing P-V 3 microns of departure from the spherical wavefront over a 60 degree view. The interferometer must be aligned so that the focal point of the lens is close to the geometric centre of the quasi-sphere.

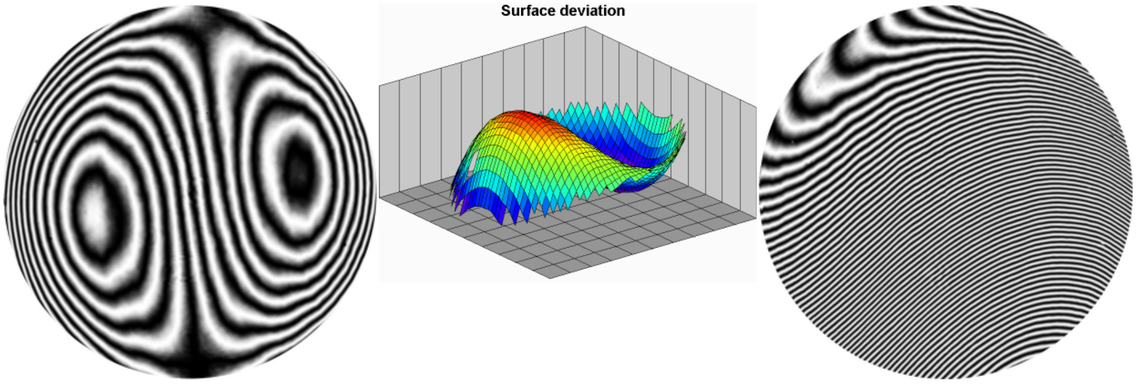


Figure 7-21: Interferograms and surface map from the cavity surface – the central map is for the interferogram on the left

7.5.3 *Further work*

7.5.3.1 System identification

The diamond turning was conducted at a speed low enough to satisfy the tolerance requirement based on an assessment of dynamic motion errors at a range of machining frequencies. The relatively low turning spindle speed of 200 rpm limits the machining rate at the expense of total machining time (thermal drift) or feed per revolution (scallop height). A full system identification, including an assessment of mode shapes could permit a much higher rotation rate including fuller compensation for dynamic machining errors.

7.5.3.2 Feedforward and feedback compensation

The only error compensation applied was in respect of parametric extraction of alignment offsets for tool and workpiece from the scan-measured workpiece surface form. The highly accurate measurement data could be used for surface form error compensation, in the procedure of Chapter 8, see Figure 8-1. Additionally, tool error maps can be generated with extremely high precision [214] – consideration can be given to feedforward compensation based on tool error mapping.

7.5.3.3 Non-linear numerical fitting of offset parameters

Tool offset fitting using equation (7-19) does give a degenerate solution, which although successfully applied can be improved with a non-linear fitting solution, such as Levenberg-Marquardt.

8 APPLICATION OF SOLUTION, CASE STUDY 2 – E-ELT

8.1 Demands

The scientific demands for the large freeform surfaces required for E-ELT are outlined in 2.3.2. The technical surface demands are given below.

A surface must be ground using BoX, a 3 axis machine possessing a fixed orientation "cup" wheel grinding spindle, with its axis inclined at 20 degrees to the rotary axis of the workspindle. The machine configuration is shown in Figure 4-3 and Figure 4-4. The demand in this context, for machining the E-ELT mirror segments, is to devise an approach to toolpath generation for these freeform surfaces, which can accommodate the motion configuration and wheel shape design of the machine, and utilise the measurement feedback information described in chapter 5 for freeform optics – the example of chapter 5 having been ground on the BoX machine.

8.2 Approach to toolpath generation

Tool path generation for the 3D case is highly complex, made particularly difficult by two factors: 1) the unusual tool geometry of the case study and 2) the related difficulty of Cartesian to polar transformation in the case of out of plane (X-Z plane, see Figure 4-4) co-ordinates. This latter difficulty is subtle (see section 8.2.6) but its consequence is that the transformation for the tool contact point is different from the transformation for the tool control point, and in general there is no mathematically *closed solution* for an expression of the correspondence between the Cartesian and polar co-ordinates for the tool centre/control point. The reason for this is that the conversion from a given tool contact point in Cartesian co-ordinates to polar co-ordinates in general actually selects a different contact point on the wheel and at a different instant of time between the distinct co-ordinate systems of wheel and work. This entails a conversion process which goes beyond simple co-ordinate transformation.

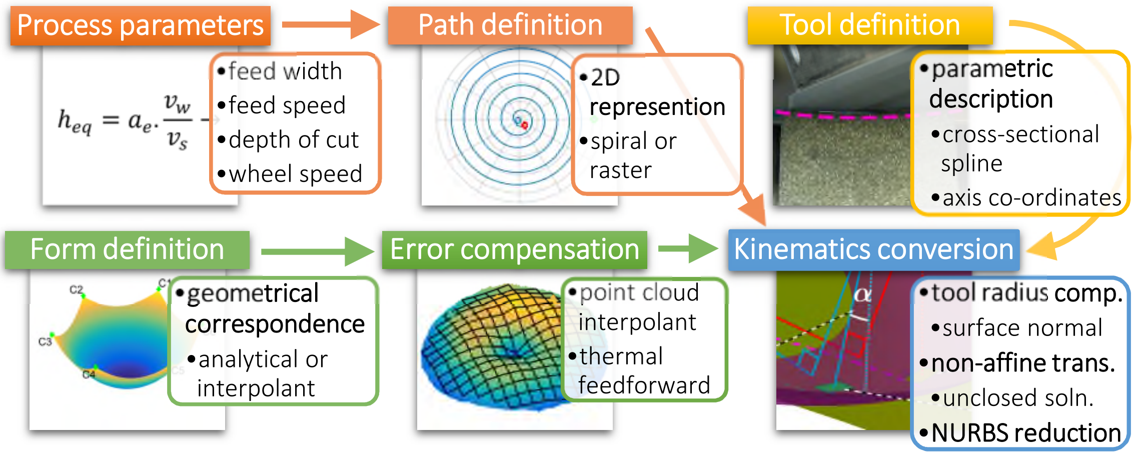


Figure 8-1: Elements of toolpath generation

The principal elements of the devised strategy for generation of a tool path are shown in Figure 8-1. The form definition and associated error compensation (perhaps from previous iterations) define the desired finished shape for the workpiece and are expressed in the workpiece's co-ordinate system. The path definition and process parameters are also expressed in the workpiece's co-ordinate system and relate to the locus of the contact (machining) point on the workpiece. The tool (wheel) definition is a geometric description of the functional surface of the wheel – a convex surface of rotation – and defined in its own co-ordinate system. The kinematics conversion is the complex mathematical scheme for combining all three input sets to generate a tool path program in the cylindrical co-ordinate system of the machine.

8.2.1 Process parameters

The designs for process parameters: speeds, feeds and depths, have been reported in previous Cranfield research [7] [9] [12] and most extensively [220]. These primarily relate to the optimisation of grinding conditions regarding normal and tangential force, specific material removal rate and damage depth which are highly dependent on workpiece material and tool design. There are other factors in process design which relate to machine dynamic performance, available machining time, thermal control, asynchronous workspindle errors, contact zone shape and size, coolant delivery, influence on mid-spatial power spectrum and follow-on process etc. The author has performed additional research in optimisation of these, which contributed to the achievement of the work but is outside the research scope of this manuscript. Typical finish (fine) grinding parameters are given in Table 8-1 for materials such as Zerodur, ULE® or fused silica.

Table 8-1: Typical finish grinding parameters used in this study

Parameter	Value
Workpiece radius of curvature	≥ 0.35 m (concave) unlimited (convex)
Tool radius of curvature	300 mm
Workspindle rotation	up to 14 rpm
Nominal tool-work path speed	25 mm/sec
Relative radial feed per revolution	1 mm
Tool spindle rotation	2200 rpm
Depth of cut	50 µm

8.2.2 Path definition

As in the case of SPDT (chapter 7), a spiral machining mode is a natural choice here for a machine with a cylindrical co-ordinate geometry, although this is not actually a definite requirement. A non-spiral path may be an option where relative tool-workpiece feed rates are low; this implies that machine motion element inertias can be low and that acceleration demands may not be prohibitive. The machine however has 3 axes of motion and all must be engaged in any but the most trivial examples of machining. Where machining is taking place close to the workpiece rotary axis, any (non-trivial) mode that is not spiral will require high acceleration of the workpiece rotary axis. Therefore, whilst a raster implementation exists, in general a spiral mode is adopted by the author, except where the entire workpiece surface is substantially off-centre, where a raster path may be beneficial in terms of process efficiency.

Figure 8-2 presents path simulations for spiral and raster paths. The polar plots represent a planar projection of the cylindrical co-ordinate system of the machine. These plots each contain two lines, showing (blue) program and (red, dotted) contact point loci for a concave spherical workpiece, although the path is defined by the contact points – the program points are computed at the end of the process of kinematics conversion (see Figure 8-1) and these form the basis of the instructions passed to the control system.

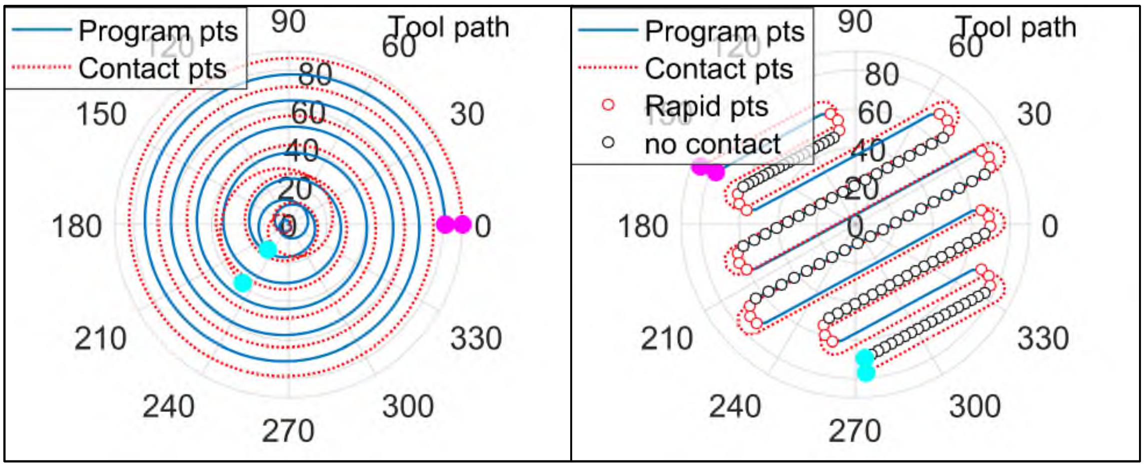


Figure 8-2: Path simulation (left) spiral and (right) raster

The plot for the raster program also demonstrates handling of intermittent machining; the spiral by contrast is in continuous contact.

Figure 8-3 shows a uni-directional machining approach (the same data as Figure 8-2 right) where on the return path the grinding wheel disengages from the surface and there is rapid feed to the start of the next machining raster line. This is done to ensure any hysteretic effects due to machining direction do not influence the final surface accuracy, and although the 'retrace' speed is high, it does reduce the overall machining rate.

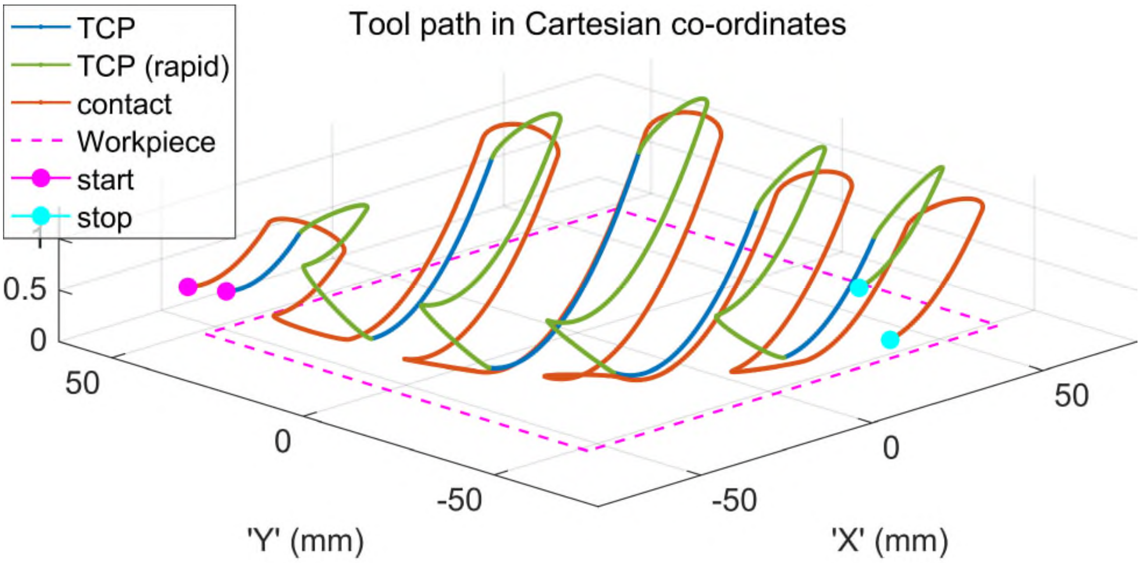


Figure 8-3: Raster path simulation showing intermittent cut

8.2.3 *Tool definition*

The tool's functional surface is a surface of rotation, formed of a 2D (curved line) description rotated around the wheel spindle's axis of rotation. The wheel spindle axis is at 20° inclination to vertical (the work spindle's axis is vertical). In order to distribute wear, and to be able to grind concave workpieces, the 2D definition of a cross-section of the wheel must be convex, and of a smaller maximum radius of curvature than the minimum radius of curvature of the workpiece; Figure 6-1 illustrates this point, where a toric wheel machines a concave workpiece – the workpiece must have a larger radius than the wheel.

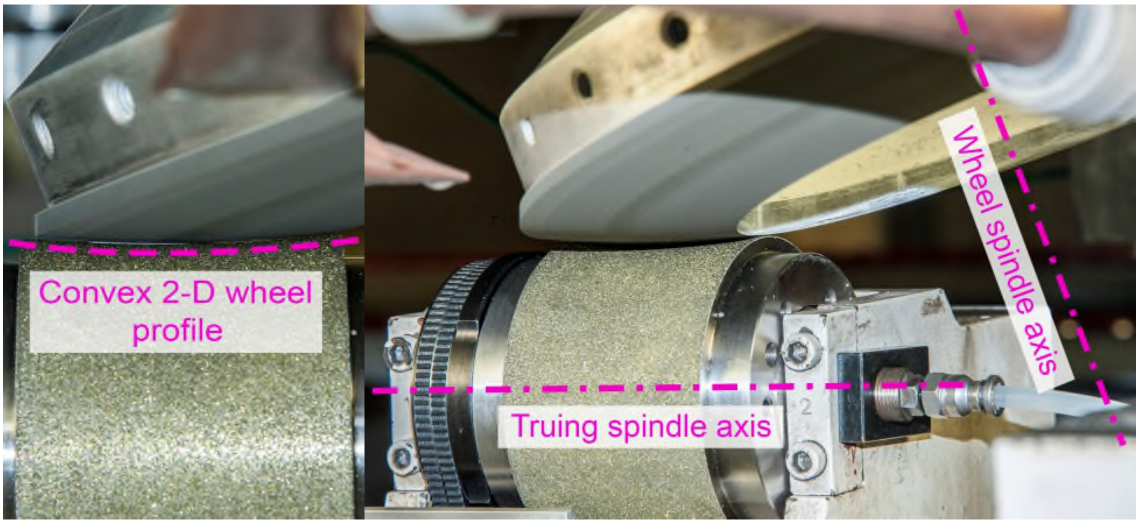


Figure 8-4: Wheel geometry formation

Using a knowledge of 2-D profile, together with the relative positions and orientations of the truing spindle and wheel spindle axes, a 3-D surface description of the wheel can be gained. The convex 2-D wheel profile is by design a circular arc. This is formed by conventional CNC machining of a truing roller, which is then coated with coarse abrasive material. For this reason, the wheel's profile curve may be of relatively low accuracy. The shape this imparts to the wheel must be known to high accuracy however, because it is a strongly influencing factor in the final machining accuracy of the overall process. This can either be 1) directly measured or 2) inferred. In this research, the profile has been inferred in three ways (alternatives):

a) By calculation from the design profile and relative positioning of axes
b) By direct measurement of a dressing stick, after wheel dressing
c) By measurement of the workpiece, after grinding

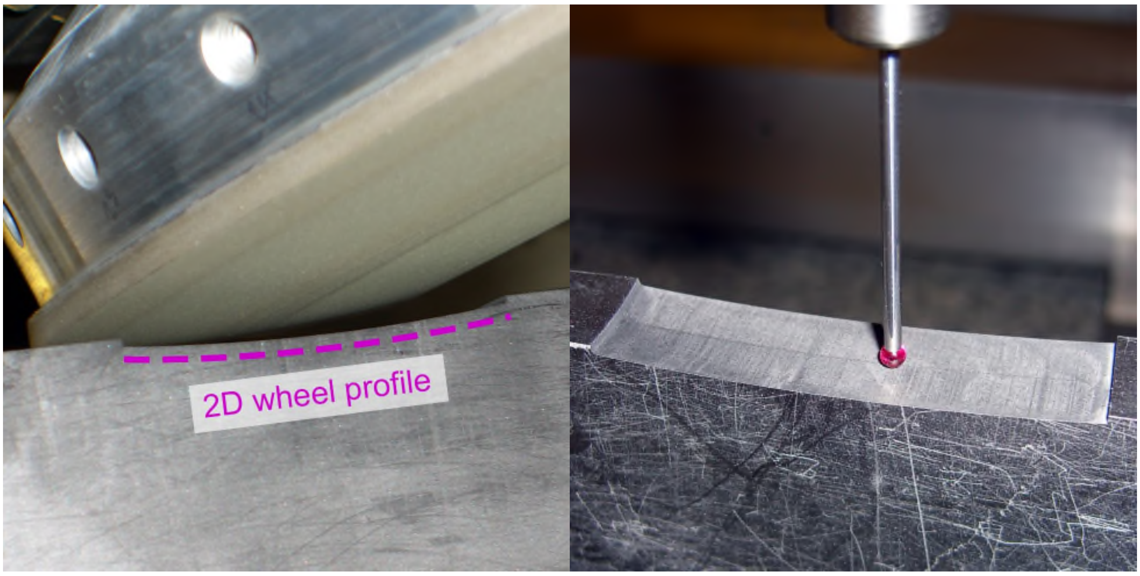


Figure 8-5: Wheel geometry measurement

Figure 8-5 shows the principle of b) above, performed here on graphite, although this process can be automated on machine using its integrated probe measurement. This is the most direct method of the three above, and avoids the difficulty of measuring the wheel's abrasive surface.

8.2.4 Workpiece form definition

The workpiece's form definition must be of an interrogable nature (and essentially explicit) so that given an (x, y) co-ordinate pair, a z co-ordinate can be obtained, with a surface normal vector (n_x, n_y, n_z). Suitable definitions include analytical descriptions e.g. an ellipsoid description as in equation (8-1), or a spline surface description; descriptions of these kinds are differentiable, so it is easy to obtain surface normals.

$$f(x,y,z) = \frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{a^2} + \frac{(z-z_0)^2}{c^2} - 1 = 0 \tag{8-1}$$

Equation (8-1) can be made explicit - equation (8-2),

$$z = \frac{c}{a}\sqrt{a^2-(x-x_0)^2-(y-y_0)^2} + z_0 \tag{8-2}$$

and either form is differentiable to yield surface normals, for instance, the implicit form, equation (8-1), where in equation (8-3), $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are the standard unit vectors.

$$\begin{aligned}\nabla\big(f(x,y,z)\big) &= \frac{\partial f}{\partial x}\mathbf{i}+\frac{\partial f}{\partial y}\mathbf{j}+\frac{\partial f}{\partial y}\mathbf{k} \\ &= \frac{2(x-x_0)}{a^2}\mathbf{i}+\frac{2(y-y_0)}{b^2}\mathbf{j}+\frac{2(z-z_0)}{c^2}\mathbf{k}\end{aligned} \tag{8-3}$$

The purpose of this definition is directional - scale of the combined vector is unimportant, so it can be normalised using the Euclidian distance, as in equation (8-4).

$$\mathbf{n} = \frac{\nabla\big(f(x,y,z)\big)}{\big\|\nabla\big(f(x,y,z)\big)\big\|} \tag{8-4}$$

It is possible to develop surface normal evaluations from other descriptions, such as point clouds, but these are non-deterministic and depend on selection of interpolation/approximation parameters. Deterministic representations are preferred, such as those based on defined polynomials.

8.2.4.1 ESO E-ELT form definition

An optical prescription for an ellipsoid equivalent to equation (8-1) is often given in the form of equation (8-5), where by design x_0 & y_0 are assumed zero and z_0; can be set to exactly -c to place z=0 at the tip of the ellipsoid. R is the base radius of curvature $(R = a^2/c)$ and k is the conic (Schwarzschild) constant $(k = R/c - 1)$. This is a restatement of equation (2-9), the general equation for rotation of a conic section.

$$z(x,y) = \frac{\left(\frac{1}{R}\right)(x^2+y^2)}{1+\sqrt{1-(1+k)\left(\frac{1}{R}\right)^2(x^2+y^2)}} \tag{8-5}$$

Here, for the E-ELT M1 primary mirror, at the time machining was carried out, the prescription was defined[*] as given in Table 8-2 [221].

Figure 8-6 shows an example segment, from the extremity of the M1 primary telescope mirror, in the co-ordinate system of the entire mirror. Its sag of ~3 mm is not discernible due to its inclination in the frame of reference of the co-ordinate system.

---

[*] The design of M1 has since been changed to a smaller diameter with a higher f-number (smaller radius of curvature)

Table 8-2: Ellipsoidal parameters [221]

Parameter	Value
Workpiece radius of curvature at vertex of M1	84 m ± 200 mm
Conic constant	-0.993295
x-y plane-projected regular hexagon, across flats	1.228448375 m
Minimum ‘radius’ (in hexagon orders)	6.24
Maximum ‘radius’ (in hexagon orders)	21

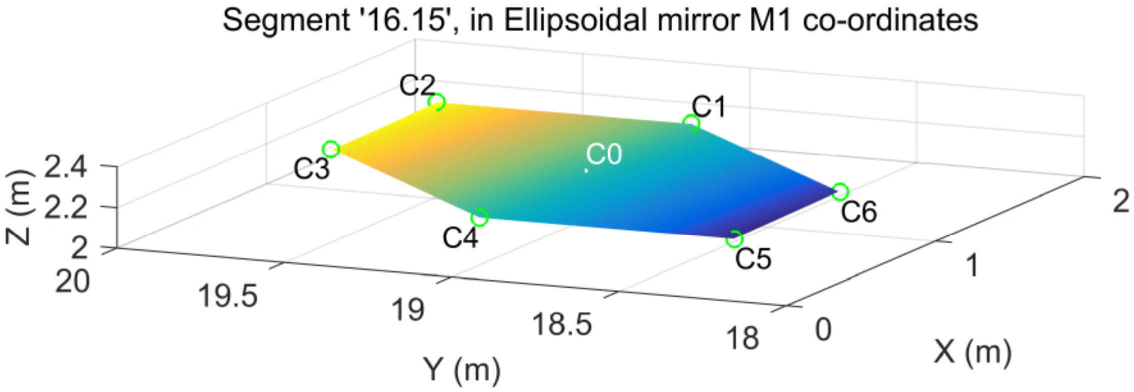


Figure 8-6: Segment “16.15” at periphery of E-ELT M1 mirror – in M1 co-ordinates

The same segment, shown in its own co-ordinate system, placing the centre “vertex” (C0 from Figure 8-6) at 0,0,0 has a discernible shallow curvature.

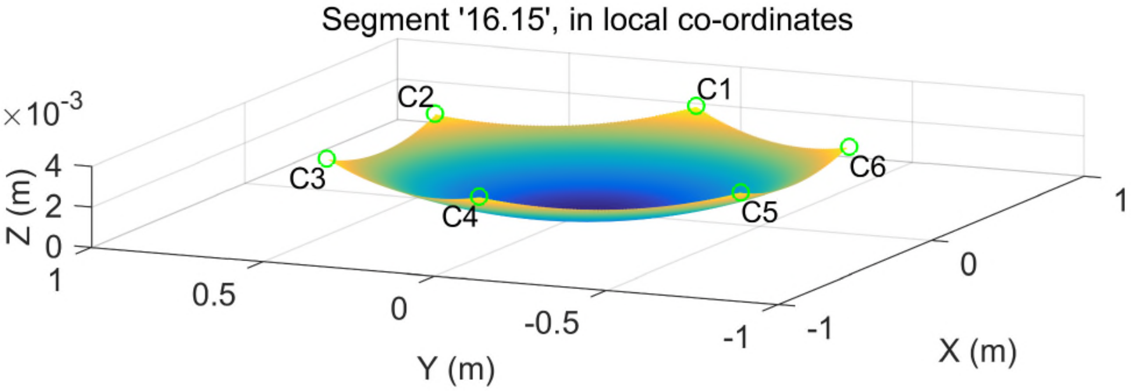


Figure 8-7: Segment “16.15” in local co-ordinates

8.2.5 Error compensation

Error compensation is likely formed from empirical data and as seen in section 5.5.2 is most suitably based on a point cloud interpolant. In this toolpath generation scheme, error compensation data are used to modify the form definition of section 8.2.4. For a low-slope surface, such as the E-ELT mirror in this case study, the slopes are low enough that an error compensation scheme operating in a single direction (not with orientation-flexible surface normals) is adequate – particularly in an iterative context.

In this case, the error map (formed from an interrogable point cloud interpolant as in Chapter 5) generates correction data at whatever granularity is required by the tool path generation. These correction data are then simply added to the form definition; this is the simplest possible scheme.

8.2.6 Kinematics conversion

The kinematics conversion (as suggested in the process workflow diagram in Figure 8-1) has multiple stages.

8.2.6.1 Phase A - Path definition

Initially, chosen material processing parameters dictate the dimensional factors (if not the topology) of the path definition; the path definition is then given by two linear arrays (of the Cartesian planar variables x and y) describing the path shape and two further arrays which are created by explicit functions of x and y as in equations (8-6).

$$\begin{aligned} \text{x co-ordinate of machining point} &= [x_1 \quad x_2 \quad x_3 \quad \cdots \quad x_n] \\ \text{y co-ordinate of machining point} &= [y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_n] \\ \text{velocity magnitude of machining point} &= v(x,y) \\ \text{clearance from machining point} &= c_z(x,y) \end{aligned} \tag{8-6}$$

The velocity value is also determined by material processing requirements. The clearance value is zero during machining and is positive (away from the workpiece) when the path takes a rapid (non-grinding) route to the next grinding position, indicated in green in Figure 8-3 and labelled “(rapid)” in the graph’s legend. In the final kinematics conversion, positions intermediate between the points defined in equations (8-6) will be interpolated at whatever density is required by the form accuracy tolerance requirement or the block processing speed limitations of the machine controller. By way of an example, the spiral machining path for an ESO E-ELT segment is generated as follows. Table 8-1 gives the feed width (spiral feed per revolution, f) and path speed (v) as 1 mm and 25 mm/sec respectively. So for a 1.5 m diameter (d) circular workpiece, the path could be described by equations (8-7).

$$\begin{array}{rll}
\text{number of turns } (+2 \text{ past centre}) & m = {}^{d}\!/_{2f} & \\
\text{angle} & 2\pi m \geq \theta \;\; \geq -4\pi & \\
\text{radius} & r = {}^{f\theta}\!/_{2\pi} & \\
 & x = r\cos\theta & \\
 & y = r\sin\theta & \\
\text{speed} & v = 0.025 & \\
\text{clearance} & c_z = 0 & (\theta \geq 0) \\
\text{clearance } (5 \text{ is arbitrary slope}) & c_z = -5r & (\theta < 0)
\end{array} \tag{8-7}$$

A representation is given in Figure 8-8 (left) of the path, with geometrical factors altered for easy visualisation. In the same figure it is shown (right) in 3-D with the TCP path included.

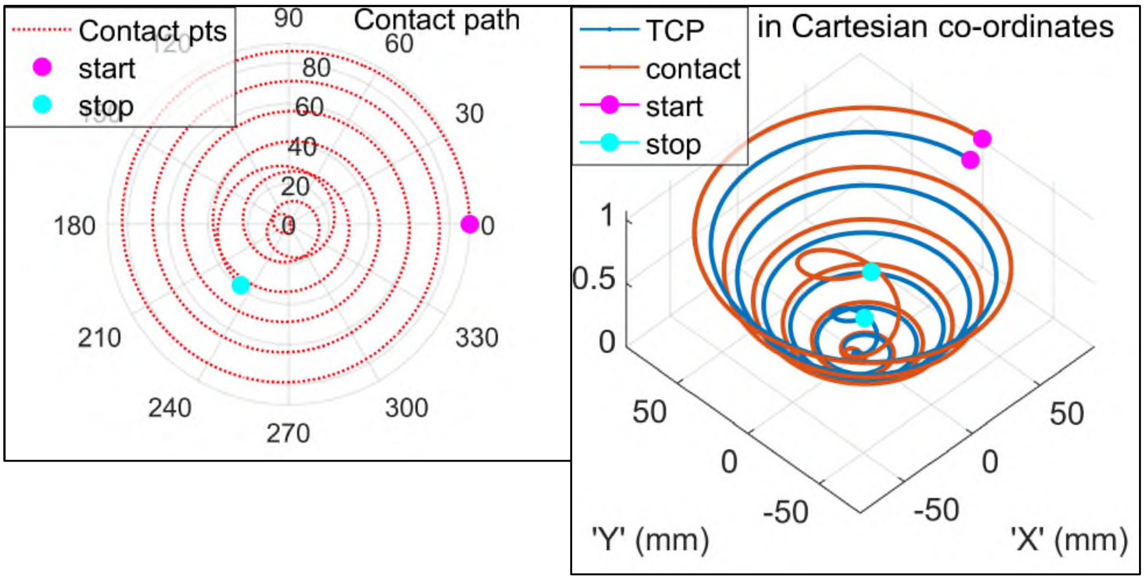


Figure 8-8: Spiral contact path

8.2.6.2 Phase B - Contact path in workpiece co-ordinate system

The second phase is the generation of the error-compensated workpiece form definition and the mapping of the contact path onto it. Again taking the E-ELT example for the case study, the form definition function is given in equation (8-5) and its surface normal vector (which will be required) can be derived by inspection from equations (8-3), (8-4) and (8-5) as in equations (8-8) in terms of the optical prescription preferred constants.

$$
\begin{aligned}
\left\|\nabla\big(f(x,y,z)\big)\right\| = d &= 2\sqrt{\frac{x^2+y^2}{a^4}+\frac{z^2}{c^4}} \\
&= 2\sqrt{\frac{x^2+y^2}{{R^4}/{(k+1)^2}}+\frac{z^2}{{R^4}/{(k+1)^4}}} \qquad (8\text{-}8) \\
\mathbf{n}(x,y) &= \frac{2(x-x_0)}{{dR^2}/{(k+1)}}\mathbf{i}+\frac{2(y-y_0)}{{dR^2}/{(k+1)}}\mathbf{j}+\frac{2(z-z_0)}{{dR^2}/{(k+1)^2}}\mathbf{k}
\end{aligned}
$$

The optical prescription, equation (8-5), must be modified by any error compensation data – which will be obtained as from equation (5-12), the equivalent to the data in Figure 5-41 (right). Note that the surface normal vector has been obtained from the analytical form description of the surface, and not from the form description modified by the error compensation data. There is an error associated with this approximation, but we can show that it is negligible. Taking wildly pessimistic values, the error compensation might add an absolute maximum additional slope variation given by the peak error between adjacent scan measurements (0.4 µm over 3.4 mm, see Figure 5-31) a maximum (δ = 0.007 degrees). The additional tool radius correction error for surface normal approximation would be $E = r(1-\cos\delta) = 2$ nanometres, where r is the tool radius of ~300 mm.

So now the contact point path on the workpiece surface, given in the workpiece co-ordinate system is defined by an interpolation of the points in equations (8-7), given from a parameterisation of θ for the spiral to yield (x, y, v, c_z) combined with the workpiece definition as in equations (8-5) and compensation from equation (5-12), restated in equation (8-9).

$$\mathbb{S}.\mathbf{val} = \mathbb{X}(\mathbb{S}.\mathbf{pts}) - \mathbb{A}(\mathbb{S}.\mathbf{pts}) + \mathbb{Y}(\mathbb{S}.\mathbf{pts}) \Rightarrow z_{\text{error}} = \mathfrak{S}(x,y) \qquad (8\text{-}9)$$

The combination gives a contact path defined in the workpiece co-ordinate system, given in equations (8-10).

$$
\begin{aligned}
&2\pi m \geq \theta \geq -4\pi, \qquad x = \left({}^{f\theta}\!/_{2\pi}\right)\cos\theta\,, \qquad y = \left({}^{f\theta}\!/_{2\pi}\right)\sin\theta \\
&F(x,y) = z(x,y) + c_z - \mathfrak{S}(x,y), \qquad = (\text{tool path}) \\
&\qquad\qquad\qquad\qquad\qquad \mathbf{n}(x,y), \qquad = (\text{surface normal})
\end{aligned}
\tag{8-10}
$$

8.2.6.3 Phase C – Tool radius compensation

The application of tool radius compensation is relatively straightforward, at least in the co-ordinate reference frame of the workpiece.

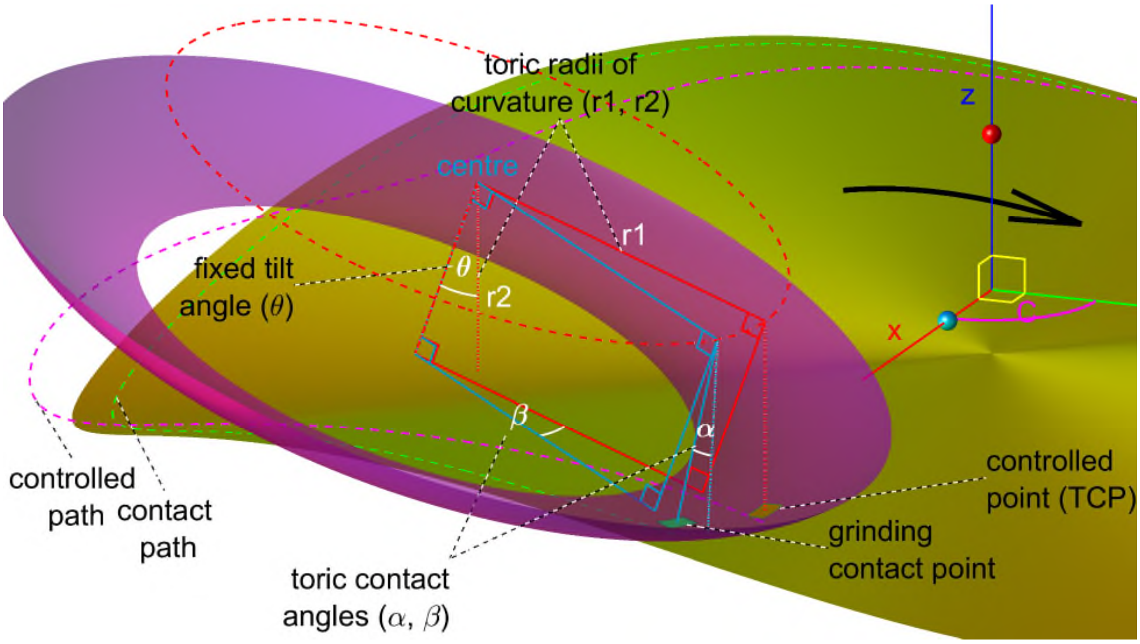


Figure 8-9: Geometry simulation

With reference to Figure 8-9, the controlled point (TCP) is the lowest point on the wheel, the point where the slope is nominally zero – in other words the contact point normal is parallel to the workpiece's rotary axis. By inspection of the figure, the co-ordinates of the controlled point relative to the wheel centre are as given in equations (8-11).

$$x = 0, \qquad y = r_1 \cos\theta, \qquad z = -r_1 \sin\theta - r_2 \tag{8-11}$$

Similarly, the contact point co-ordinates relative to the wheel centre are given in equations (8-12) in terms of the construction angles and the wheel toric radii, r1 and r2.

$$
\begin{aligned}
x &= -(r_1 + r_2 \sin(\theta + \alpha)) \sin\beta \\
y &= (r_1 + r_2 \sin(\theta + \alpha)) \cos\beta \cos\theta - r_2 \sin\theta \cos(\theta + \alpha) \\
z &= -(r_1 + r_2 \sin(\theta + \alpha)) \cos\beta \cos\theta - r_2 \cos\theta \cos(\theta + \alpha)
\end{aligned} \tag{8-12}
$$

The equivalent of “tool radius compensation”, the adjustment applied to the tool contact point to obtain the tool control point, is simply the difference between equations (8-11) and (8-12), given fully in equations (8-13).

$$
\begin{aligned}
x &= (r_1 + r_2 \sin(\theta + \alpha)) \sin\beta \\
y &= r_1 \cos\theta - (r_1 + r_2 \sin(\theta + \alpha)) \cos\beta \cos\theta + r_2 \sin\theta \cos(\theta + \alpha) \\
z &= (r_1 + r_2 \sin(\theta + \alpha)) \cos\beta \cos\theta + r_2 \cos\theta \cos(\theta + \alpha) - r_1 \sin\theta - r_2
\end{aligned} \tag{8-13}
$$

In the foregoing, θ is the fixed wheel tilt angle – set at 20° on the machine. α and β are geometrical construction angles, set by the surface normal of the desired workpiece surface form – the surface normal angle. Given n_x, n_y and n_z are the components of $\mathbf{n}$ we can obtain α and β in terms of $\mathbf{n}$, the surface normal, as in equations (8-14).

$$
\begin{aligned}
\alpha &= \tan^{-1} \frac{\sqrt{{n_x}^2 + \left(n_y \cos\theta - n_z \sin\theta\right)^2}}{n_y \sin\theta + n_z \cos\theta} - \theta \\
\beta &= \tan^{-1} \frac{n_x}{n_z \sin\theta - n_y \cos\theta}
\end{aligned} \tag{8-14}
$$

So in terms of the quantities x, y and z in equations (8-13) we can now express more directly the control point in terms of the contact point, using just the wheel radii and single tilt angle as given in equations, and this is in the polar (cylindrical) co-ordinate system of the machine.

$$
\begin{aligned}
x_{\mathrm{TCP}} &= \sqrt{(x_{\mathrm{contact}} + x)^2 + y^2} \\
z_{\mathrm{TCP}} &= z_{\mathrm{contact}} + z \\
C_{\mathrm{TCP}} &= \sin^{-1} \frac{y}{x_{\mathrm{TCP}}} + C_{\mathrm{contact}}
\end{aligned} \tag{8-15}
$$

There are two significant issues with this treatment.

a) This assumes a high degree of fidelity between mathematical model and wheel geometry. Fidelity in the direction of rotation of the wheel (β angle fidelity) will be very good as this is generated (during wheel forming) by spindle axis rotation with a high degree of averaging. This is not true however for the α angle fidelity, which is reliant on the profile fidelity of the truing wheel.

b) The calculation of surface normal is performed in the machine's x-z plane, for which y = 0. The point which is then selected on the wheel (except for the trivial β = 0 case) **is not in the x-z plane**, so it will not contact the workpiece at the programmed point – it will contact it out of the plane, at a different point, at a different angle.

For these reasons, a different treatment is required. Given the first of these points, it's clear that absent an analytical definition of the wheel shape, a closed mathematical solution will not be available, and a numerical solution must be sought.

The profile is shown in Figure 8-10 (green dashed line). In this figure, the contact patch can be seen for a flat surface, at the lowest point on the wheel's surface with the surface normal shown as a solid black line.

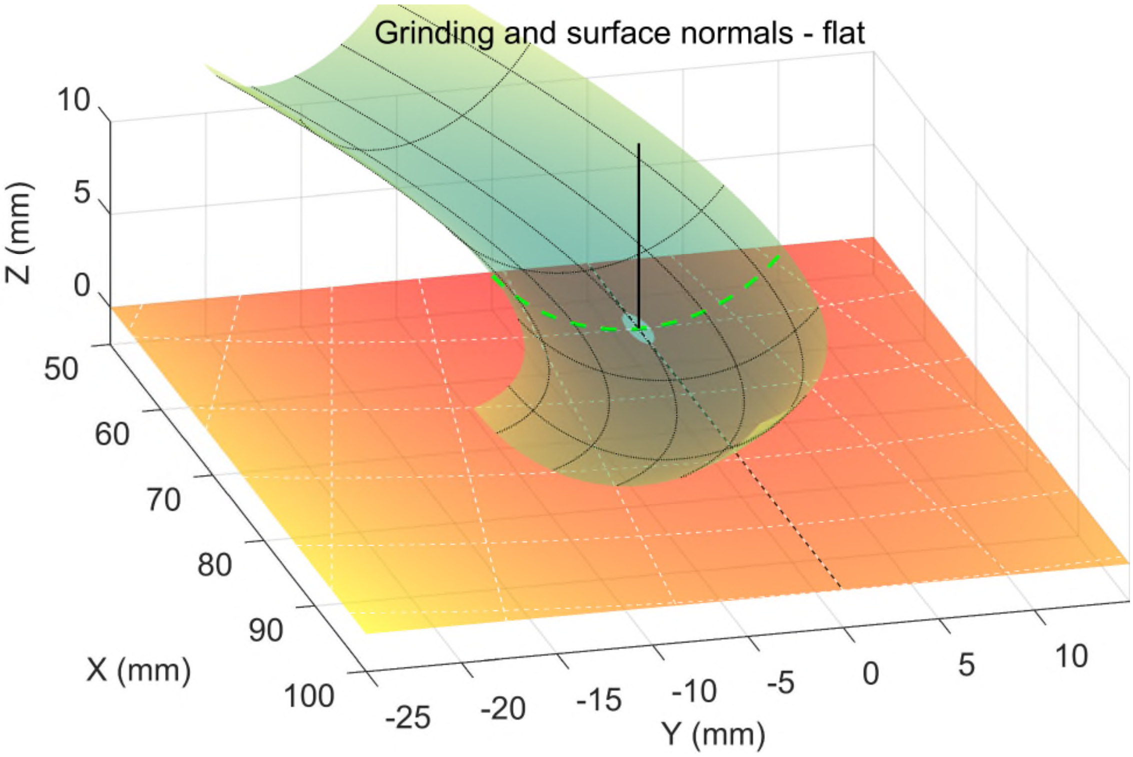


Figure 8-10: Contact patch for grinding a flat surface

8.2.6.4 Iterative solution to phase C

The approach based on the analytical description of the wheel can be modified to accommodate both of the issues identified in the previous section. There are two elements to a solution.

a) The surface normal angle from the workpiece selects a unique position on the convex wheel expressed with the two angles α and β. Equations (8-14) give a correct solution for their identification. Angle β is one of the co-ordinates to describe a position on the wheel. A slice through the wheel perpendicular to the wheel's rotary axis will give a circle. The wheel profile (as seen in Figure 8-4 and Figure 8-5) is by design a circular arc, however it is not accurate. The profile can be inferred by measurement and represented by spline approximation. Angle α can then select a unique position within the spline and in combination with β provide a unique position within the wheel's surface. This can be used to give an improved representation for equations (8-13).

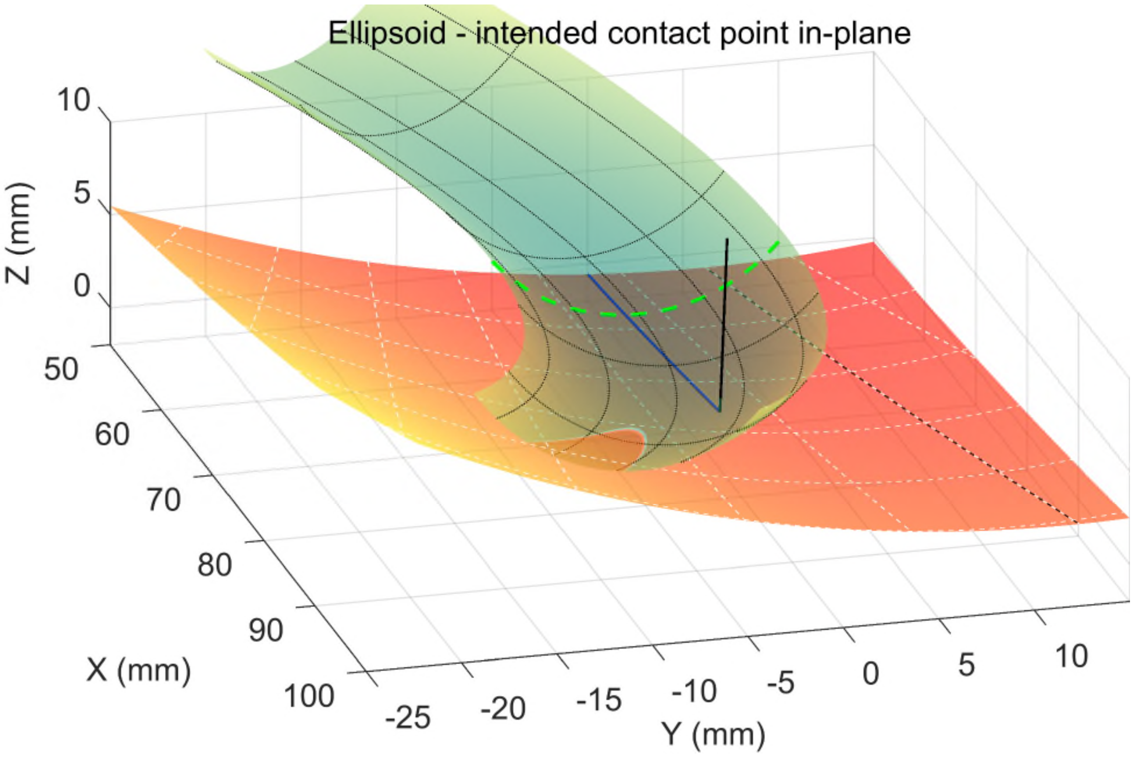


Figure 8-11: In-plane contact not possible for non-zero ∂y/∂z

Figure 8-11 shows why an in-plane (X-Z plane) contact point is in general not possible with a sloped workpiece surface. The wheel "cuts" the surface out of the plane, and since the motion system can only move in X or rotate the workpiece, it's impossible to grind at the desired position. The only way the correct grinding geometry can be achieved is if the workpiece is rotated and the contact point is out of the plane. The required rotation angle is a function of the surface normal angle, the co-ordinates of the contact point and the geometric parameters of the wheel.

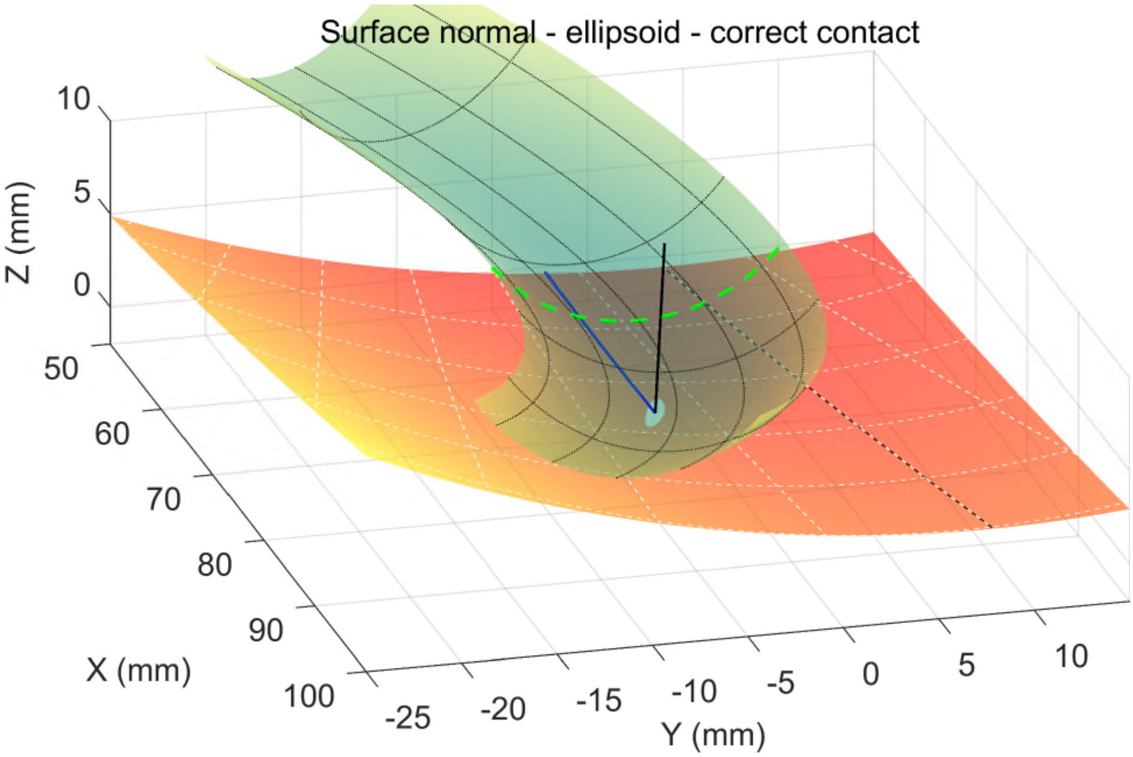


Figure 8-12: Out of plane contact correctly achieved

b) The second element to the solution addresses the out of plane issue. Since there is no mathematically closed solution, an iterative approach to a solution can be used to find a rotation angle C for the workpiece, which moves the contact point out of the plane and at the same time rotates the surface normal (in the co-ordinate reference system of the wheel) until a match is found with a point on the wheel. Since the wheel's radii of curvature are unconditionally smaller than the E-ELT's radius of curvature at all points, and the wheel is unconditionally convex whilst the workpiece is unconditionally concave, a solution must exist if the wheel's surface is wide enough, so a suitable search technique must be able to find one. One potential difficulty is that there could be more than one solution – a situation that can occur under exceptional circumstances where there is relatively high workpiece slope close to the workpiece spindle's rotation centre. This is detected by the code implementation.

c) Figure 8-12 shows the same contact point as in Figure 8-11 but with the workpiece rotated so that the contact point aligns with the wheel so that their surfaces are in parallel contact. This must be addressed in the transformation from contact point path to TCP path; for each point on the path, there is a different non-affine co-ordinate transformation, and each is a non-closed mathematical solution.

8.2.6.5 Implementation of solution to phase C

In the plane of the wheel profile, the profile can be described by z as a function of y – where y and z correspond to the Cartesian co-ordinates of the machine as in Figure 4-4.

The profile is approximately circular, as in equation (8-16) – it is by design circular but subject to significant manufacturing tolerances.

$$\begin{aligned} &y = r_2 \sin\alpha\,, \qquad z = r_2(1-\cos\alpha) \\ &z = r_2 - \sqrt{{r_2}^2 - y^2} \end{aligned} \tag{8-16}$$

Measurement can give a more accurate profile which can be approximated with a standard cubic spline of the form $\hat{f}(y_j)$ which is designed to minimise the sum in equation (8-17). In that equation p, the smoothing parameter is chosen to give appropriate smoothing – typically $p \approx 1/(1+h^3/6)$ where h is the mean spacing between values of y.

$$p\sum_{j=1}^{n}\left|z_j - \hat{f}(y_j)\right|^2 + (1-p)\int\left|\frac{\partial^2 f(y)}{\partial y}\right|^2 dy \tag{8-17}$$

The angle normal to the profile at any point is described (in radians) by the spline's derivative, as can be seen by differentiating equation (8-16) to give equation (8-18).

$$\begin{aligned} \frac{dz}{dy} &= \frac{d\left(r_2 - \sqrt{{r_2}^2 - y^2}\right)}{dy} = \frac{-y}{\sqrt{{r_2}^2 - y^2}} = \frac{-r_2\sin\alpha}{r_2 - z} = \frac{-r_2\sin\alpha}{r_2 - r_2(1-\cos\alpha)} \\ &= \tan\alpha \approx \frac{d\hat{f}(y)}{dy} \end{aligned} \tag{8-18}$$

Utilising standard smoothing spline implementations [184] (including standard expressions for their derivatives) with p as indicated above, we have a mapping to go from y to $\tan\alpha$ (or indeed α) by using the derivative of the spline. Since this is not an analytical function, we can't rearrange it to make it explicit in y, for going from α to y, but we can build a table of values for $\tan\alpha$ and use another smoothing spline to construct another mapping to infer y from $\tan\alpha$, since there is a 1:1 correspondence for y to $\tan\alpha$, over the range of angles involved.

$$\hat{g}\left(\tan^{-1}\left(\hat{f}'(y)\right)\right) \approx y \tag{8-19}$$

This spline, equation (8-19) then gives us the ability to get a y value (and therefore, z) from an angle α and is superior to table lookup, since it gives infinite resolution.

In order to achieve the alignment in Figure 8-12, the requirement is to find the rotation angle (of the machine's C axis) needed to get the surface normal to align to a surface normal for the wheel at a mutual contact point. An iterative search is used to refine an estimate of the rotation angle. For each iteration, the surface normal from the contact point is rotated by a trial C rotation angle. The two angles (relating to the α and β of Figure 8-9) are obtained in the co-ordinate frame of reference of the wheel. The values of α and β select a unique point on the wheel. The distance out of plane (y-direction) in the workpiece co-ordinate system must be the same as the value for distance out of plane in the wheel frame of reference. The y value in the workpiece co-ordinate system is given by simple trigonometry, and the y-value in the wheel co-ordinate system is given by equation (8-19) from α. This y-value (and its associated x and z values) give the relationship between the contact point and control point. Optimised code has been written to achieve rapid convergence in fitting within 2 or 3 iterations for each point in the control program, so that relatively rapid conversion is possible (approximately 10,000 program points' conversion per second).

8.3 Results

Two validations of the technique of this chapter are offered. The first is the surface produced and measured in Figure 5-46. The residual error of below 800 nm includes distortion of the surface due to machining forces as well as errors in the CMM and the BoX grinding machine. It does however have the usual optician corrections of tilt and defocus removed, since these can always be adjusted on assembly alignment and test of an optical system, without detriment to imaging quality.

The principal validation is in the machining of an E-ELT segment.

Figure 8-13: Machining ESO E-ELT segment on BoX

The Zerodur segment was machined on the BoX machine using the tool path generation technique reported here. Figure 8-13 and Figure 8-14 show the segment in context, during a grinding operation.

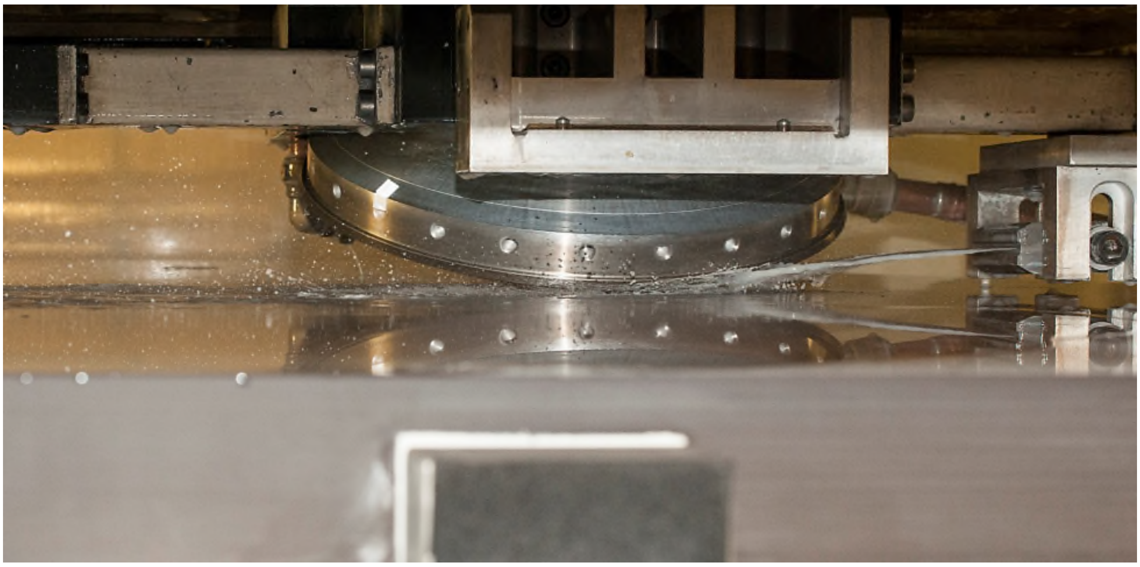


Figure 8-14: Machining ESO E-ELT segment on BoX (close-up)

Figure 8-15 and Figure 8-16 show the machined segment in context, being measured on the PMM-F 30-20-10 CMM.

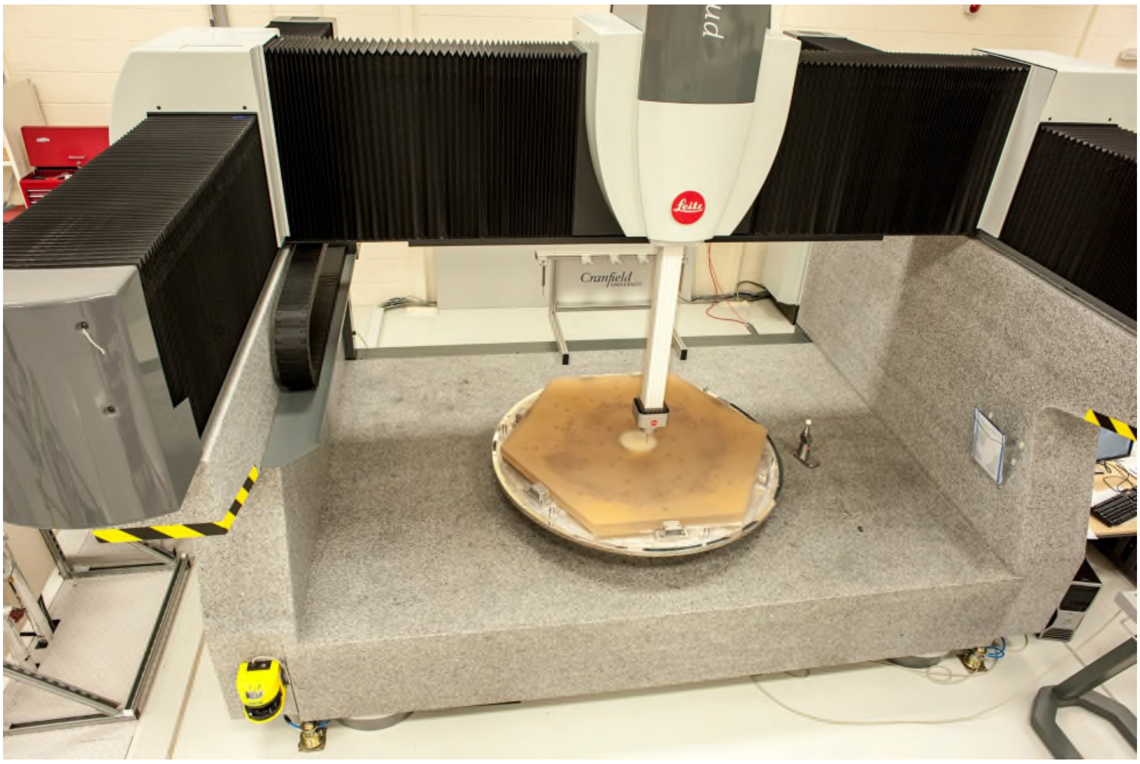


Figure 8-15: Measuring 1.5 m diameter ESO E-ELT segment on CMM

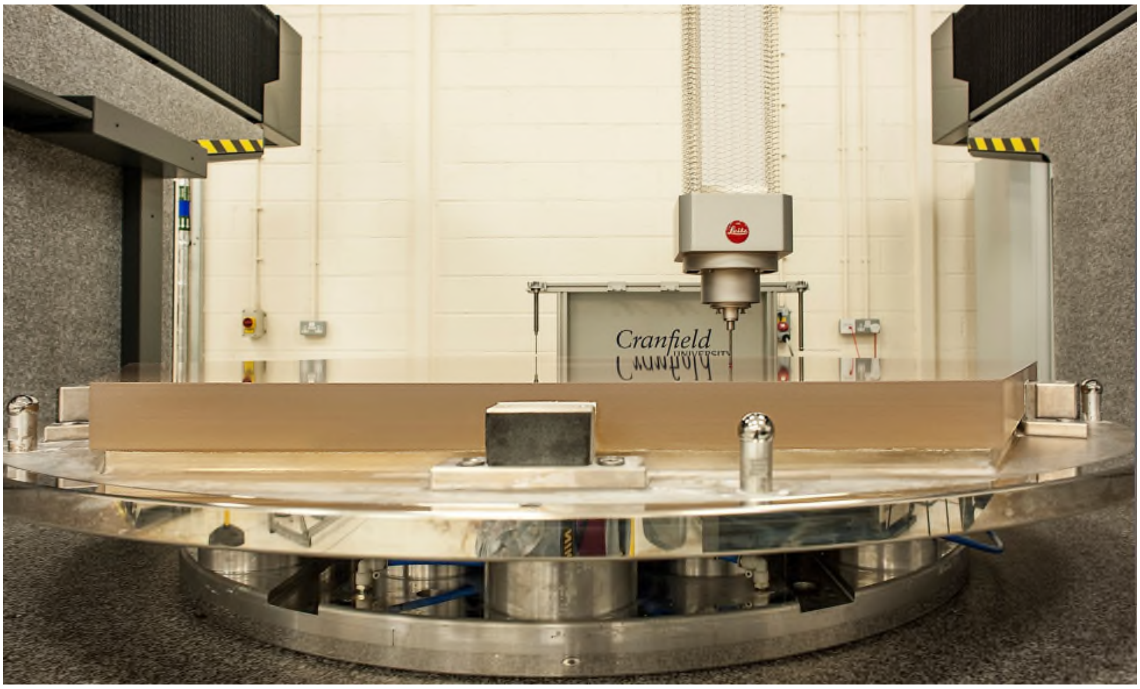


Figure 8-16: Measuring ESO E-ELT segment on CMM – curvature visible

Figure 8-18 shows the measurement performed using the techniques of chapter 5. The measured form error is 691 nm rms over the entire surface to within 0.5 mm of the segment edges. The map shown is the difference between the freeform design shape of the surface, and its measured shape. No terms are removed other than height offset (piston) and tilt, so this represents the severest evaluation.

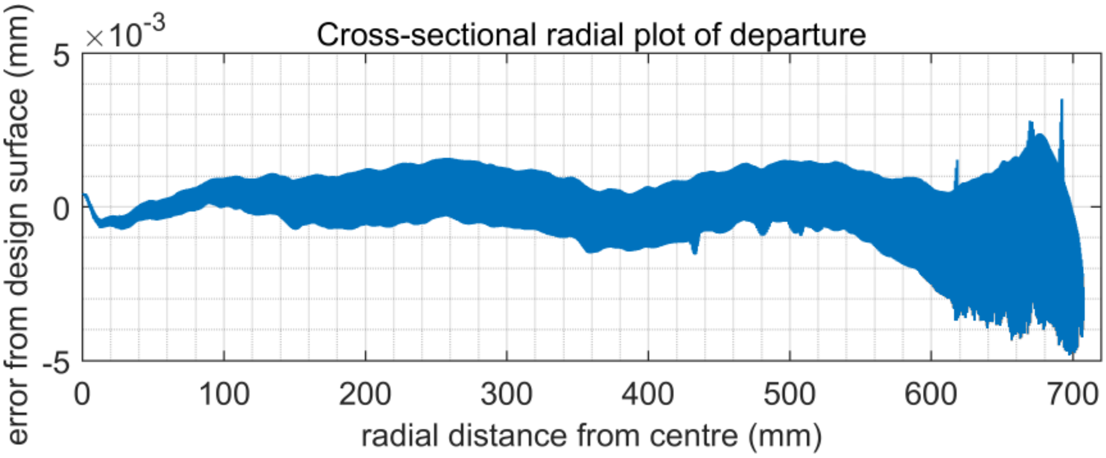


Figure 8-17: Error from Figure 8-18 (ordinate) plotted against radial distance from workpiece centre (abscissa)

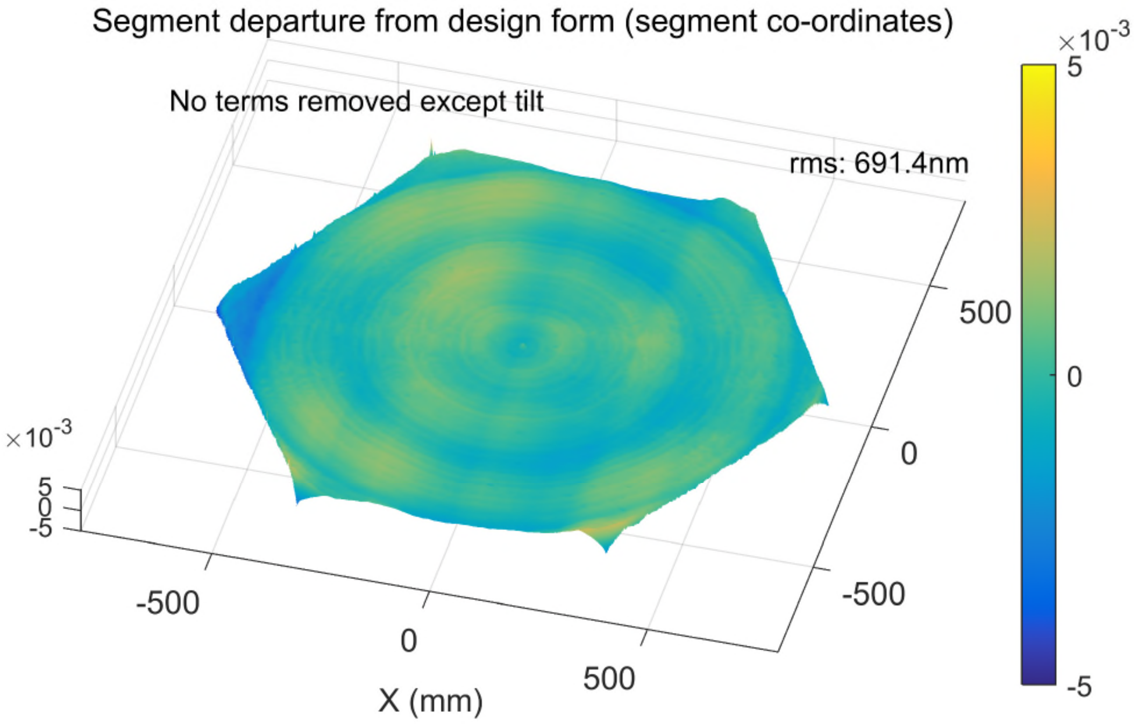


Figure 8-18: Measurement result

Three E-ELT segments have been machined with similar performance (all under 1 µm rms form error with only tilt removed) and a similar number of other large optics with similar performance. Figure 8-17 shows the same error data as in Figure 8-18 (1,200,000 data points) plotted against radial distance from the rotation centre of the workpiece.

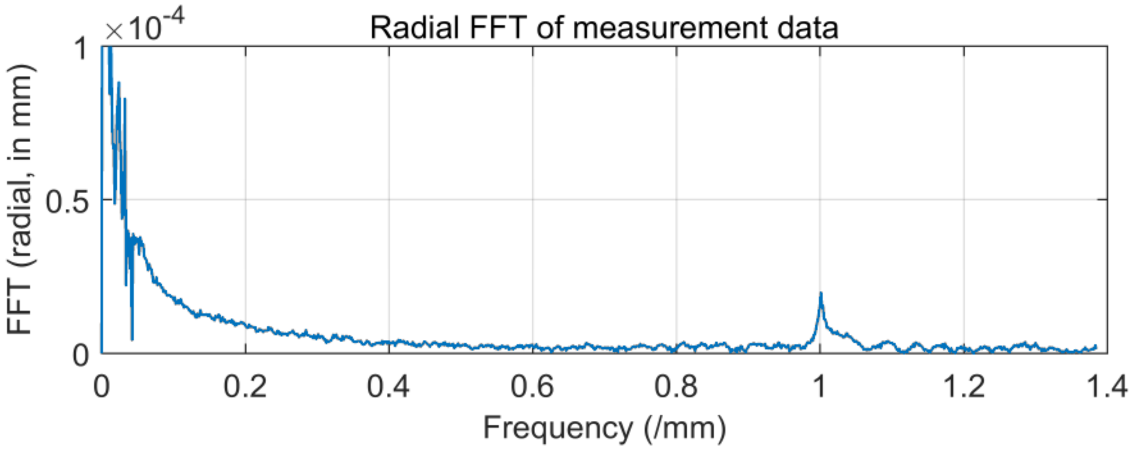


Figure 8-19: FFT of error against radial distance (vs. spatial frequency)

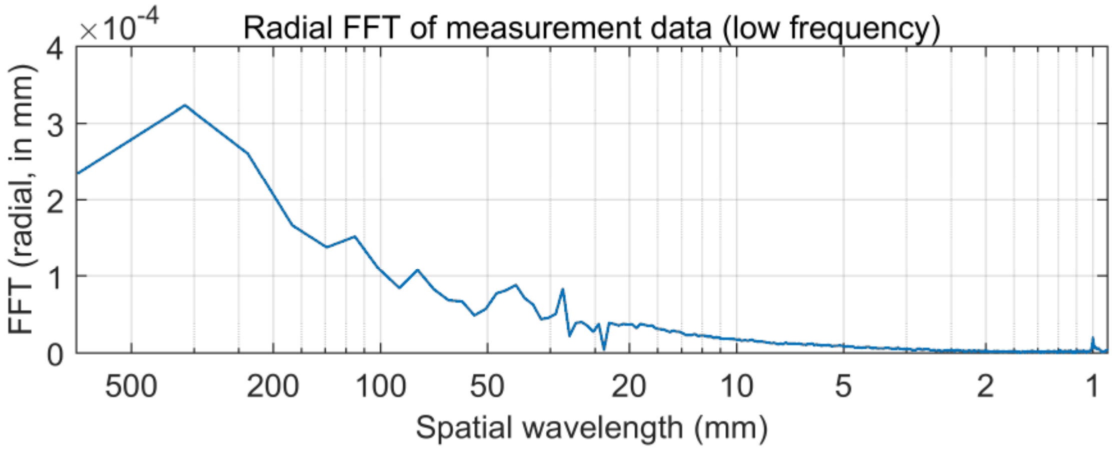


Figure 8-20: FFT of error against radial distance (vs. spatial wavelength)

Figure 8-19 and Figure 8-20 show Fourier transforms of again the same data giving the error amplitude as a function of respectively spatial frequency and spatial wavelength, assessed in a radial direction from the rotation centre of the workpiece.

8.4 Summary and discussion

A method for tool path generation for smooth freeform surface machining has been presented that can be applied for machining using complex 3-D tool geometries and highly challenging motion configurations. The method is designed and has been demonstrated to achieve very high accuracy whilst utilising a minimum (3-axes) motion configuration. Starting with a novel process workflow concept, this is accomplished by utilising a flexible and full freeform geometric model of the workpiece and tool shape and solving the mathematical problem of relating motions in one co-ordinate system through a necessarily non-affine transformation to the motions of a different point in another co-ordinate system.

The method has been validated in the successful machining, to sub-micron precision, of three prototype segments for the M1 primary segmented mirror of the forthcoming ESO E-ELT ground-based telescope (completing a research objective) which in the 2020s will become the world’s largest full aperture telescope.

The measurement for Figure 8-18 was performed without the additionally described compensation or Zernike decomposition techniques of chapter 5 being applied; the machining work on the prototype segments was carried out before these measurement process extensions were developed. The available measurement data was used for error compensation feedback, as described in the foregoing. In place of the Zernike treatment, *for error compensation feedback only*, zero phase low-pass spatial filtering was applied to ensure that any short wavelength artefacts (at 1 mm or below) had minimal influence within the measurement feedback. Attenuation for wavelengths above 5mm was limited to 0.1 dB (around 1.1% attenuation) of error. This technique has been subsequently rejected in favour of an optimised Zernike decomposition so that more strategic choices can be made over signal rejection within the error compensation feedback loop.

Through the finesse of the underlying measurement technique, circular form error patterns (rings) are clearly discernible within the measurement in Figure 8-18. These can be analysed with the accumulated radial plot of Figure 8-17 and its Fourier transform in Figure 8-19 and Figure 8-20. There are several observations to make.

- The area over the corners of the hexagonal segment, where there is intermittent machining during a rotation, exhibit the largest error, due to variations in grinding conditions. This accounts for the “fluffing” of the trace above 600 mm radial distance in Figure 8-17. Although this is an artefact of machining process parameters, it can be corrected with error compensation; this aspect of error compensation has been substantially improved with the adoption of Zernike decomposition of error maps, although this treatment was not implemented at the time the segments were processed.

- Between around 100 mm and 600 mm radial distance from the centre, the character of the error trace is consistent with an envelope at around 2 microns p-v. Around 300 nm of this is geometrically inevitable from the scalloping due to the wheel profile and feed (see Table 6-2). Surface roughness in this grinding mode, would contribute around 150 nm R_a [220] or 1000 nm R_t, and some will be due to asynchronous spindle motions, although these contributions cannot be directly summed. The error envelope reduces dramatically toward centre where there is considerable grinding overlap, so the error envelope elsewhere is due motion errors and errors in the wheel profile (the machining point on which will vary as the workpiece rotates, for a freeform surface).
- The FFT traces show strong signatures at 1 mm wavelength (the feed step of the grinding toolpath) and strong signals at around 36 mm and 300-400 mm, which are visible in all the forms of the data Figure 8-18 - Figure 8-20. Linear feed in the radial direction is synchronous with workpiece rotation angle, but not with time, so the strongly repetitive signal at 36 mm is not due to temporal cycle; this is again due either to motion errors, such as due to magnetic pole pitch.
- The long wavelength errors represent errors that have been apparent with long (hours) time variation and are possibly thermal in origin. Errors with this wavelength are very well treated by the error compensation technique and therefore reflect non-repeatability in the machine-measure cycle of around 1 µm P-V amplitude; this does not reflect on the work reported in this manuscript.

Although the 700 nm rms error result is impressive, its traceable uncertainty is limited by the recognised capability of the CMM, given by its MPE at a figure which is almost an order of magnitude larger than the 700 nm. The improvements offered by the measurement algorithm, also remain non-traceable. Nevertheless, successful use in this iterative cycle does demonstrate full-cycle repeatability (including grinding and measurement) at the micron level, which is a highly significant result for freeform surfaces on this scale and puts them easily into a regime where a first stage polish in the succeeding process chain step can produce functional full-aperture interferometry.

The operational influence of a truing roller, as opposed to a cup wheel to perform grinding wheel forming is unclear. Certainly the geometrical implications of the freeform (near) toric wheel to tool path generation are considerable. Even though the achievement of tool path generation method for this combination is one of the contributions of this research, the non-deterministic shape preparation of the grinding wheel is a source of error, which can even after this research only be controlled by inference of its error profile. How much error is contributed to the surface shape is unclear. It is notable however that even where a wheel profile is specifically designed for a concave freeform surface of such as the 1.5 m diameter E-ELT mirror segments, at most 75 mm width of wheel forming roller is engaged. This 75 mm has in some way a 1:1 correspondence in terms of slope with the concave freeform surface, in terms of which part of the truing wheel affects which part of the freeform surface. In this circumstance, at best there is at least a 750:75 = 10:1 lateral expansion of any imperfections on the truing profile onto the freeform surface. This is a minimum ratio; in the machining described in this chapter, 15 mm of wheel's width was used to grind 750 mm of workpiece radius. Therefore a single grit defect on the truing roller that has dimension of 260 microns may have an effect on a zone > 12.5 mm wide on the finished freeform surface and this may be the origin of some of the imperfections in Figure 8-18 and Figure 8-17. The author strongly advocates the adoption of a spherical wheel and spherical forming approach in which the forming process has transverse action, so that there is a great deal of averaging in wheel shape creation, almost completely avoiding the magnification effect described above and ensuring a high fidelity of wheel form.

8.4.1 *Achievement of objectives and contribution to knowledge*

8.4.1.1 Comprehensive tool path generation strategy

The achievement of this represents the most complete realisation of the kinematic combination strategy objective. In particular a comprehensive tool path generation strategy to support the operation of the BoX machine for freeform grinding has been developed. A multi-element strategy was devised (represented in Figure 8-1) which succeeded in its application to making a number of prototype segments and other freeform surfaces. This strategy is generic for other machine tool motion configurations.

8.4.1.2 3 axes to machine freeforms

A second contribution was to find a method to covert a contact point to tool centre point for a fixed orientation toroidal tool used in a 3-axes cylindrical geometry machine tool and applicable to any tool path on a smooth freeform surface. This objective was met and is represented in 8.2.6.4

8.4.1.3 Tool radius compensation for a freeform tool, using a spline representation of the tool shape

The third contribution is to find a representational basis for a freeform convex tool and through this an extension to 8.4.1.2 to accommodate any convex tool shape produced as a surface of rotation.

8.4.2 *Further work*

8.4.2.1 Extension of tool path design for high slope surfaces

A potential limitation of the developed strategy is that tool paths are originated in a 2-D representation, which for low-slope surfaces is scarcely a limitation. For higher slope surfaces, this would need to be modified directly to devise appropriate paths on a 3D freeform surface. This is work which has been addressed to some extent for 4/5 axes machines where the contact point on the tool can be made largely deterministic, based on reorientation of the tool for the surface normal vector. The situation for a 3-axes machine is more complex.

8.4.2.2 Anomalous kinematics conversion to investigate redundant solutions

An anomalous situation can occur where there are multiple solutions to the kinematics conversion represented in Figure 8-11 Figure 8-12. This is identified in the developed algorithms, although not deeply investigated in the existing research. Whilst it can only occur for relatively high slope close to the rotation centre for the workpiece, this eventually is not deterministically predicable, and so it is difficult automatically to design tool paths to avoid it.

8.4.2.3 Non-zero slope on centre

Whilst it’s always possible to place a zero slope at centre, it may not be the most efficient machining proposition, as it may place a higher peak slope elsewhere within the work zone. An investigation can be carried out of the potential of handling finite non-zero slope at workpiece centre. Some unreported demonstration grinding has been done using a diametric path across centre which can be extended.

8.4.2.4 Modified tool path combinations

There is a possibility of combined tool path geometries such as combining raster with spiral, but the issue of discontinuities at path junctions should be investigated. The tool path strategy, as reported in this chapter, lends itself to this approach, as it separates the designed tool contact path from the remainder of the generation algorithm.

8.4.2.5 Treatment for edge effects

Specific treatment (other than error compensation) for edge effects could be employed, such as tool path trajectories that run always parallel to edges. This could help significantly with edge roll-off.

8.4.2.6 Improved centre region handling

For spiral machining, the peak workpiece spindle rotation speed limits the achievable machining parameters. This leads to a reduced tool-work relative path speed, below design parameters. This can lead to a conical depression towards the centre of a workpiece. This is currently successfully treated using error compensation, but also be tackled in a deterministic way using path modifications.

8.4.2.7 Adaptive modelling of tool wear and predictive compensation

The developed algorithms hold information on material processing parameters and have a full time-based trajectory model for the position of the mechanical contact point on the tool. This information can be used to predict the distribution and amount of tool wear and to feed forward compensation for tool wear into the tool path program. This offers the potential to reduce the iterative requirement of machine-measure to achieve a given tolerance of free form surface.

8.4.2.8 Spherical wheel forming

An investigation of the degree to which the ‘rings’ effect may be due to wheel forming accuracy could involve some spherical wheel forming, using a cup-type forming wheel. This would simplify the tool path formation.

9 CONCLUSIONS AND SUMMARY OF ACHIEVEMENT

Chapters 5-8 contain their own discussions on pages 97-100, 130-131, 156-160 and 183-187 respectively. These are summarised here within a discussion relating to the overall work.

The focus of this work has been the manufacture of smooth freeform surfaces. In particular, two principles at the core of precision engineering are engaged in the basis for this research.

a) Determinism: an underlying thesis has been that the precision of freeform surface manufacture can be reinforced by using a minimum possible number of machine axes of motion. The justification for this is that in adding motion complexity, particularly by adding *stacked axes*, additional errors, uncertainties and sources of non-repeatability are also added due to less than perfect degrees of constraint, differential expansivities, additional compliance etc. This is held to be case under any circumstance, but particularly so where high energy density or high dynamic forces are involved.

b) Measurement: the critical underpinning of precision manufacture; it is only by achieving low uncertainty measurement/test of whatever aspects of an entity are functionally important that these aspects can be made correct; this may be equivalent to a saying often attributed (most fittingly in this context) to Kelvin: “If you cannot measure it, you cannot improve it”. The actual quote is subtly different [222] but the paraphrased version is a reasonable first principle for ultra-precision manufacturing.

This research has been guided by the above principles, with the consequent requirements, below.

a) A consequence of using three axes machines, in both of the case study applications, has been the additional complexity of generating tool paths, owing to the degree of reliance placed on knowledge of the tool’s location and geometry, and the precise control of which part of the tool is engaged in the machining. In terms of mathematical ‘load’, this is a high price to pay, but the rewards – as evidenced by the achievements of the machining, particularly related to repeatability - are great. A similar (minimal motion system complexity) approach was applied by the author and others in the pioneering manufacture of monolithic multi-mirror image slicers for the integral field spectroscope units in the Mid Infrared Instrument of NASA's James Webb Space Telescope [10] [40] [223], also with great success.

b) In both case studies, through the application of iterated feedback of measured parameters, measurement at a leading level of precision, has been a *sine qua non* of this enterprise – in both cases for validation and adjustment of machining strategy; in case 1 the high lateral resolution being essential in and of itself, and in case 2 (subtly, due to the degeneracy of the fit) to the accuracy of tool adjustment.

9.1 Contribution to knowledge

Contributions to knowledge are detailed in the relevant sections, and collated here.

Algorithms for scanning metrology of large area surfaces

1) High lateral resolution contact scanning measurement of smooth freeform surfaces
2) Detection and removal of scanning contact errors
3) Compensation of errors due to changes in measurement accuracy during measurement
4) Application of orthogonal polynomial error separation to a ground freeform surface

Tool path generation

5) Identification of a surface representation scheme for smooth freeforms
6) Automated design of appropriate tool paths based on machining criteria and compensated surface shape data
7) Optimised condensed tool path representation

Application of solution, case study 1 – Boltzmann "quasi spheres"

8) An effective machining strategy for a tri-axial ellipsoid
9) A high resolution & accuracy surface measurement of a hemi-ellipsoid
10) Accurate extraction of tool and work offsets from a free-form surface measurement

Application of solution, case study 2 – E-ELT

11) Comprehensive tool path generation strategy
12) 3 axes to machine freeforms
13) Tool radius compensation for a freeform tool, using a spline representation of the tool shape

9.2 Impact of research

Based largely on research presented in this manuscript, the author wrote “Enabling UK Manufacturing in Ultra Precision and Structured Surfaces” one of Cranfield University’s 23 Impact Case Studies for REF 2014 (one of 3 in the Manufacturing Engineering discipline). In addition, there are 27 refereed publications (of which 12 are journal articles) resulting from this research. The contents of this manuscript, which are being aired for the first time, will elicit further publication.

The know-how in the creation and implementation of the algorithms of Chapter 8 was a key part of the founding IP in the Cranfield University spin-out, Loxham Precision Ltd. and was similarly fundamental to the delivery of a €5,000,000 contract, awarded to Optropreneurs Limited, to supply prototype segments for the E-ELT, supported by Cranfield University through EPSRC funding.

In 2018, the BIPM will formally adopt a new definition of the kelvin, in which it will take a value consistent with a defined value of the Boltzmann constant. The research described in this manuscript will have made a key contribution to this redefinition of one of the 7 base SI units of all measurements worldwide, and to the newly defined value of the Boltzmann constant. Of the author’s contribution, the NPL Project Leader and Science Ambassador, Michael de Podesta said [224]:

> *“The perfection of the inner surface was without a doubt the key to the low uncertainty. Nothing else would have worked without your magic.”*

10 RECOMMENDATIONS FOR FURTHER WORK AND EXPLOITATION

Chapters 5-8 contain their own recommendations for further work on pages 100-101, 131-132, 160-161 and 187-189 respectively. These are collated here, with a discussion of the direction of the overall work.

Algorithms for scanning metrology of large area surfaces

1) Hysteresis anisotropy
2) X axis gantry – treatment of scanning hysteresis
3) Accuracy improvement through improved stylus calibration
4) Uncertainty investigation through artefact calibration
5) Repeatability investigation through rotate and move
6) Selection of scanning speed/data point density, as a function of surface geometry
7) Higher spatial frequency form compensation

Tool path generation

8) Binary search improvement
9) Re-parameterisation-control point shift Interplay

Application of solution, case study 1 – Boltzmann “quasi spheres”

10) System identification
11) Feedforward and feedback compensation
12) Non-linear numerical fitting of offset parameters

Application of solution, case study 2 – E-ELT

13) Extension of tool path design for high slope surfaces
14) Anomalous kinematics conversion to investigate redundant solutions
15) Non-zero slope on centre
16) Modified tool path combinations
17) Treatment for edge effects

18) Improved centre region handling

19) Adaptive modelling of tool wear and predictive compensation

20) Spherical wheel forming

In the wider context, future exploitation of these techniques for freeform surface machining may lie in the development of embedded CAD-CAM capability to use them. A lot of research into the enhancement of CAD-CAM capability is focused on the facility of high axis-count machines, where the application of machine intelligence is required to assist in optimisation (for efficiency) of machine tool path, where there would be choices – perhaps infinite choices - of trajectory. Optimisation for machined surface quality (which has wide industrial application) might entail use of a minimal axis-count machine, and this is where generic commercially available solutions do not currently exist.

The techniques of treatment of data for scanning measurement could also apply to non-contact scanning technologies, which are increasingly available on high accuracy CMMs. Although the character of the errors is different for a non-contact system, particularly in that the errors are not uni-directional, there is much commonality, and scope for application of this research.

11 REFERENCES

1. Underwood R, de Podesta M, Sutton G, Stanger L, Rusby R, Harris P, et al. Estimates of the difference between thermodynamic temperature and the International Temperature Scale of 1990 in the range 118 K to 303 K. Phil. Trans. R. Soc. A. 2016 March; 374(2064).
2. de Podesta M, Harris P, Underwood R, Sutton G, Mark D, Stuart F, et al. Response to Macnaughton's 'Comment on "A low-uncertainty measurement of the Boltzmann constant"'. Metrologia. 2016 January; 53(1): p. 116-122.
3. de Podesta M, Underwood R, Sutton G, Morantz P, Harris P, Mark DF, et al. A low-uncertainty measurement of the Boltzmann constant. Metrologia. 2013 August; 50(4): p. 354-376.
4. Underwood R, Davidson S, Perkin M, Morantz P, Sutton G, de Podesta M. Pyknometric volume measurement of a quasispherical resonator. Metrologia. 2012; 49(3): p. 245-256.
5. Shore P, Morantz P. Ultra-precision: enabling our future. Philosophical Transactions of the Royal Society A - Mathematical Physical and Engineering Sciences. 2012; 370(1973): p. 3993-4014.
6. de Podesta M, Sutton G, Underwood R, Davidson S, Morantz P. Assessment of uncertainty in the determination of the Boltzmann constant by an acoustic technique. International Journal of Thermophysics. 2011; 32(1-2): p. 413-426.
7. Tonnellier X, Howard K, Morantz P, Shore P. Surface integrity of precision ground fused silica for high power laser applications. Procedia Engineering. 2011; 19: p. 357-362.
8. Underwood R, Flack D, Morantz P, Sutton G, Shore P, de Podesta M. Dimensional characterization of a quasispherical resonator by microwave and coordinate measurement techniques. Metrologia. 2011; 48(1): p. 1-15.
9. Comley P, Morantz P, Shore P, Tonnellier X. Grinding metre scale mirror segments for the E-ELT ground based telescope. CIRP Annals - Manufacturing Technology. 2011; 60(1): p. 379-382.
10. Shore P, Cunningham C, DeBra D, Evans C, Hough J, Gilmozzi R, et al. Precision engineering for astronomy and gravity science. CIRP Annals - Manufacturing Technology. 2010; 59(2): p. 694–716.
11. McKeown P, Corbett J, Shore P, Morantz P. Ultra-precision machine tools - design principles and developments. Nanotechnology Perceptions. 2008 March; 4(1): p. 5-14.
12. Tonnellier X, Morantz P, Shore P, Baldwin A, Evans R, Walker D. Subsurface damage in precision ground ULE® and Zerodur® surfaces. Opt. Express. 2007; 15(19): p. 12197-12205.
13. Morantz P. Multi-process Strategy for Freeform Optics Manufacture. In 3rd International Conference on Nano Manufacturing: nanoMan2012; 2012; Tokyo, Japan.
14. Tonnellier X, Morantz P, Shore P, Comley P. Precision grinding for rapid fabrication of segments for extremely large telescopes using the Cranfield Box. In Proc. SPIE 7739, Modern Technologies in Space- and Ground-based Telescopes and Instrumentation; 2010; San Diego, USA.

15. Tonnellier X, Shore P, Morantz P, Baldwin A. Surface quality of sintered silicon carbide using an effective grinding process. In Wold of Photonics Congress: Manufacturing of Optical Components; 2009; Munich, Germany.

16. Tonnellier X, Shore P, Morantz P, Orton D. Surface quality of a 1m Zerodur part using an effective grinding mode. In Proc. SPIE. 7102, Optical Fabrication, Testing, and Metrology III; 2008; Glasgow, UK.

17. Tonnellier X, Shore P, Morantz P, Baldwin A, Walker D, Yu G, et al. Sub-surface damage issues for effective fabrication of large optics. In Proc. SPIE 7018, Advanced Optical and Mechanical Technologies in Telescopes and Instrumentation; 2008; Marseille, France.

18. Morantz P, Luo X, Shore P. Characterisation of dynamic errors of an ultra precision machine tool. In Laser Metrology and Machine Performance VIII; 2007; Cardiff, UK. p. 4-13.

19. Tonnellier X, Shore P, Morantz P, Baldwin A, Evans R, Walker D. Comparison of the subsurface damage induced when precision grinding ULE and Zerodur surfaces. In 8th International Conference on Laser Metrology, Machine Tool, CMM and Robotics Performance; 2007; Cardiff, UK.

20. Tonnellier X, Morantz P, Shore P, Baldwin A, Walker D, Evans R. Subsurface damage caused during rapid grinding of Zerodur. In 10th International Symposium on Advances in Abrasive Technology (ISAAT 2007); 2007; Dearborn, USA.

21. Tonnellier X, Shore P, Luo X, Baldwin A, Morantz P, Jin T, et al. Wheel wear investigations when precision grinding of optical materials using the BoX grinding mode. In 5th International Conference on High Speed Machining; 2006; Metz, France.

22. Morantz P, Shore P, Luo X, Baird I. Control strategy of the Big Optix grinding machine. In 6th euspen International Conference; 2006; Baden bei Wien, Austria.

23. Luo X, Morantz P, Shore P, Baird I. NURBS approximation method for tool path generation in a new free-form grinding machine. In Advances in Manufacturing Technology, 4th International Conference on Manufacturing Research (ICMR); 2006; Liverpool, UK.

24. Tonnellier X, Shore P, Morantz P, Luo X, Baldwin A. Diamond resin bond wheel wear in precision grinding of optical materials. In Advances in Manufacturing Technology, 4th International Conference on Manufacturing Research (ICMR); 2006; Liverpool, UK.

25. Shore P, Morantz P, Luo X, Tonnellier X, Read R, May-Miller R. Design philosophy of the ultra precision Big OptiX "BoX" Machine. In 7th International Conference on Laser Metrology, Machine Tool, CMM and Robotics Performance; 2005; Cranfield, UK.

26. Morantz P, Shore P, Stephenson D, May-Miller R, Read R. From tetrahedral to box: design philosophy behind the Cranfield Big OptiX system. In 5th euspen International Conference; 2005; Montpellier, France.

27. Shore P, Morantz P, Luo X, Tonnellier X, Collins R, Roberts A, et al. Big OptiX ultra precision grinding/measuring system. In Proc. SPIE 5965, Optical Fabrication, Testing, and Metrology II; 2005; Jena, Germany.

28. Shore P, Luo X, Tonnellier X, Morantz P, Stephenson D, Collins R, et al. Grinding mode of the BoX ultra precision free-form grinder. In Proc. 20th ASPE Annual Meeting; 2005; Norfolk, USA.

29. Morantz P, Comley P, Tonnellier X, Shore P. Precision free-form grinding of metre-scale optics. In SPIE Optifab; 2011; Rochester, USA.

30. Morantz P, Mitchell J. E-ELT Mirror Segment Fabrication. In Manufacturing Technologies to Support Large Science Projects; 2010; Paris, France.

31. Morantz P. Off-Aixs Grinding Machine (OAGM) // Box Ultra Precision Freeform Grinding and Measuring Machine. In Optonet Workshop, Ultra Precision Manufacturing of Freeforms and Microstructures; 2008; Jena, Germany.

32. Morantz P, Gould E, Shore P. Thermal control of a high energy density ultra-precision machine. In Proc. euspen Topical Meeting: Thermal effects in precision systems; 2007; Maastricht, Netherlands.

33. free, a., n., and adv. In The Oxford English Dictionary. 2nd ed.: Oxford University Press; 1989.

34. Thompson KP, Rolland JP. Freeform Optical Surfaces: A Revolution in Imaging Optical Design. Optics and Photonics News. 2012; 23(6): p. 30-35.

35. Rolland JP, Thompson K. Freeform optics: Evolution? No, revolution! SPIE Optical Design & Engineering. 2012 July: p. 3.

36. LED headlamp from Hella to appear on Cadillac. LEDs Magazine. 2007 November 22.

37. Crease RP. Metrology in the balance. Physics World. 2011 March: p. 39-45.

38. Samuel Reich E. Physicists count on updated constants. Nature. 2011 July: p. 437.

39. de Podesta M. Redefining temperature. Physics World. 2013 August: p. 28-32.

40. Shore P, Morantz P, Lee D, McKeown P. Manufacturing and Measurement of the MIRI Spectrometer Optics for the James Webb Space Telescope. CIRP Annals - Manufacturing Technology. 2006; 55(1).

41. Shore P, Parr-Burman P, Atad E, May-Miller R, Peggs G, Smith D, et al. Large Optics Manufacturing Study. Department of Trade and Industry; 2003.

42. European Southern Observatory. The E-ELT Construction Proposal. Garching bei München: European Southern Observatory, The E-ELT Project Office; 2012.

43. manufacture, n. In Oxford English Dictionary. 2nd ed. Oxford: Oxford University Press; 1989.

44. Enoch JM. Archeological optics: the very first known mirrors and lenses. Journal of Modern Optics. 2007 June; 54(9).

45. Ziegler C. Les Statues égyptiennes de l'Ancien Empire. Paris: Musée du Louvre, Département des Antiquités égyptiennes; 1997. Report No.: 58.

46. The British Museum. The Nimrud Lens / The Layard Lens. [Online]. [cited 2017 February 6. Available from: http://www.britishmuseum.org/research/collection_online/collection_object_details.aspx?objectId=369215&partId=1.

47. Layard AH. Discoveries in the Ruins of Nineveh and Babylon London: John Murray; 1853.

48. Bunch B, Hellemans A. The Timetables of Technology New York: Simon & Schuster; 1993.

49. Fang FZ, Zhang XD, Weckenmann A, Zhang GX, Evans C. Manufacturing and measurement of freeform optics. CIRP Annals - Manufacturing Technology. 2013 June; 5(3).

50. Jiang X, Scott P, Whitehouse D. Freeform Surface Characterisation - A Fresh Strategy. CIRP Annals - Manufacturing Technology. 2007; 56(1): p. 553-556.

51. Savio E, De Chiffre L, Scmitt R. Metrology of freeform shaped parts. CIRP Annals - Manufacturing Technology. 2007; 56(2): p. 810-835.

52. ISO 17450-1:2011, GPS General concepts Part 1: Model for geometrical specification and verification..

53. Garrard K, Bruegge T, Hoffman J, Dow T, Sohn A. Design tools for freeform optics. In Proc. SPIE 5874, Current Developments in Lens Design and Optical Engineering VI; 2005; San Diego, USA.

54. Coons SA. Surfaces for computer-aided design of space forms. Technical Report. Cambridge: Massachusetts Institute of Technology, Project MAC; 1967. Report No.: MIT MAC-TR-41.

55. Norberto López de Lacalle L, Lamikiz A. Sculptured Surface Machining. In Paulo DJ, editor. Machining. London: Springer; 2008. p. 225-248.

56. Lasemi A, Xue D, Gu P. Recent development in CNC machining of freeform surfaces: A state-of-the-art review. Computer-Aided Design. 2010 April; 42(7): p. 641-654.

57. Choi BK, Jerard RB. Sculptured Surface Machining. 1st ed.: Kluwer Academic Publishers; 1998.

58. Ackroyd JA. Sir George Cayley: The Invention of the Aeroplane near Scarborough at the Time of Trafalgar. Journal of Aeronautical History. 2011 June; 2011(6).

59. Phillips HF, inventor; Blades for Deflecting Air. UK patent GB 13768. 1884.

60. Munk MM. Elements of the wing section theory and of the wing theory. Technical Report. National Advisory Committee for Aeronautics; 1925. Report No.: NACA 191.

61. Jacobs EN, Ward KE, Pinkerton RM. The characteristics of 78 related airfoil sections from tests in the variable-density wind tunnel. Technical Report. National Advisory Committee for Aeronautics; 1933. Report No.: NACA 460.

62. Farin G. A History of Curves and Surfaces in CAGD. In Farin G, Hoschek J, Kim MS, editors. Handbook of Computer Aided Geometric Design.: Elsevier B.V.; 2002. p. 1-21.

63. Schoenberg IJ. Contributions to the problem of approximation of equidistant data by analytic functions. Part A & Part B. Quarterly of Applied Mathematics. 1946; 4(1-2): p. 45-99 & 112-141.

64. Bézier PE. Example of an Existing System in the Motor Industry: The Unisurf System. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 1971 February; 321(1545): p. 207-218.

65. Rogers DF. B-Spline Curves and Surfaces for Ship Hull Definition. In International Symposium on Computer Aided Hull Surface Definition; 1977; Annapolis, USA.

66. Rodgers JM, Thompson KP. Benefits of Freeform Mirror Surfaces in Optical Design. In ASPE Proceedings, Free-Form Optics: Design, Fabrication, Metrology, Assembly; 2004; Chapel Hill, USA.

67. Aves O, inventor; Improvements in and relating to Multifocal lenses and the like, and the method of Grinding Same. UK patent GB 15735. 1907 July 9.

68. Kanolt CW, inventor; Multifocal ophthalmic lenses. USA patent 2878721. 1959 March 24.

69. Jalie M. How Progressive Power is Obtained: Progressive Lenses Part 1. Optometry today. 2005 May: p. 31-39.

70. Jalie M. The New Generation: Progressive Lenses Part 2. Optometry Today. 2005 June: p. 35-45.

71. Meister DJ. Free-Form Surfacing Technology Makes Possible New Levels of Optical Sophistication for Spectacles. Refractive Eyecare for Ophthalmologists. 2005 June: p. 1-4.

72. Hoorn FW, inventor; Motion picture system utilizing continuously moving film. USA patent 2073637. 1937 March 16.

73. Plummer WT. Free-form optical components in some early commercial products. In Proceedings of SPIE, Tribute to Warren Smith: A Legacy in Lens Design and Optical Engineering; 2005; San Diego, USA.

74. American Society for Precision Engineering. ASPE Proceedings, Free-Form Optics: Design, Fabrication, Metrology, Assembly. [Online].; 2004 [cited 2017 February 6. Available from: http://www.aspe.net/publications/Winter_2004/Winter_04.html.

75. OptoNet Workshop, Ultra Präzisions bearbeitung von Freiformflächen und Mikrostrukturen. [Online].; 2004 [cited 2013 August 5. Available from: http://www.optonet-jena.de/99-bilddatenbank/titel-publikationen/ws200405_ultra_pdf.

76. D'Amico A, Di Natale C, Lo Castro F, Iarossi S, Catini A, Martinelli E. Volatile Compounds Detection by IR Acousto-Optic Detectors. In Byrnes J, editor. Unexploded Ordnance Detection and Mitigation.: Springer; 2009. p. 21-59.

77. ISO 20473:2007, Optics and photonics - Spectral bands..

78. Riedl MJ. Optical Design Fundamentals for Infrared Systems. 2nd ed. Weeks AR, editor. Bellingham: SPIE Press; 2001.

79. Supranowitz C, Hall C, Dumas P, Hallock B. Improving surface figure and microroughness of IR materials and diamond turned surfaces with Magnetorheological Finishing (MRF®). In Proc. SPIE 6545, Window and Dome Technologies and Materials X; 2007; Orlando, USA.

80. Ledig M. Commercial Relevance of Freeform Optics. In Optonet Workshop, Ultraprecision Manufacturing of Freeforms and Microstructures; 2010; Jena, Germany.

81. Dowski ER, Cathey WT. Extended depth of field through wave-front coding. Applied Optics. 1995 April; 34(11).

82. Beckstette KF. Trends in Aspheres and Freeform Optics. In Optonet Workshop, Ultra Precision Manufacturing of Freeforms and Microstructures; 2008; Jena, Germany.

83. Fuerschbach K, Rolland JP, Thompson KP. A new family of optical systems employing φ-polynomial surfaces. Optics Express. 2011 October; 19(22).

84. Air Force Avionics Laboratory. Three mirror objective. Techical report. RECON Central, Reconnaissance Division/Reconnaissance Applications Branch; 1967. Report No.: 027000.

85. Johnson RB, Mann A. Evolution of a compact, wide field-of-view, unobscured, all-reflective zoom optical system. In Proc. SPIE 3061, Infrared Technology and Applications XXIII; 1997; Orlando, USA.

86. Saunders IJ, Ploeg L, Dorrepaal M, van Venrooij B. Fabrication and Metrology of Freeform Aluminum Mirrors for the SCUBA-2 Instrument. In Proc. SPIE 5869, Optical Manufacturing and Testing VI; 2005; San Diego, USA.

87. Kataza H, Wada T, Sakon I, Kobayashi N, Sarugaku Y, Fujishiro N, et al. Mid-infrared camera and spectrometer on board SPICA. In Proc. SPIE 8442, Space Telescopes and Instrumentation; 2012; Amsterdam, Netherlands.

88. Troutman J, Barnhardt D, Shultz J, Owen J, DeFisher S, Davies M, et al. Machining and Metrology of a Chalcogenide Glass Freeform Lens Pair. Procedia Manufacturing. 2016 June; 5.

89. Dai Y, Guan C, Yin Z, Tie G, Chen H, Wang J. Tool decentration effect in slow tool servo diamond turning off-axis. In 5th International Symposium on Advanced Optical Manufacturing and Testing Technologies, SPIE 7655; 2010; Dalian, China. p. 1-6.

90. lithography n. In The Oxford English Dictionary.: Oxford University Press; 1989.

91. Moore G. Cramming More Components onto Integrated Circuits. Electronics. 1965 April: p. 114-117.

92. Mack CA. The Future of Semiconductor Lithography:After Optical, What Next? Future Fab International. 2007 September.

93. Schaller RR. PhD Thesis, Technological Innovation in the Semiconductor Industry: A Case Study of the International Technology Roadmap for Semiconductors (ITRS) Fairfax, USA: George Mason University; 2004.

94. ITRS. Lithography. In International Technology Roadmap for Semiconductors.; 2011 Edition.

95. ITRS. Lithography. In International Technology Roadmap for Semiconductors.; 2012 Update.

96. Gandhi Y, Panchal H, Christian M, Parikh N, Parikh P. Nanoimprint lighography: a review. World journal of pharmacy and pharmaceutical sciences. 2013 March; 2(2): p. 509-521.

97. Miller S. Cost effective scaling next-generation lithography progress and prospects. In Semicon West; 2013; San Francisco, USA.

98. Yen A. EUV Lithography for High-Volume Manufacturing. In International Symposium on Extreme Ultraviolet Lithography; 2011; Miami, USA.

99. Murakami K. Development of EUV lithography tool technologies in Nikon. In International Symposium on Extreme Ultraviolet Lithography; 2011; Miami, USA.

100. Conradi O. Optics for EUV Production. In 2011 International Symposium on Extreme Ultraviolet Lithography; 2011; Miami, USA.

101. Mills IM, Mohr PJ, Quinn TJ, Taylor BN, Williams ER. Redefinition of the kilogram, ampere, kelvin and mole a proposed approach to implementing CIPM recommendation 1. Metrologia. 2006 June; 43(227).

102. BIPM. Resolution 1. In Resolutions adopted by the CGPM at its 25th meeting; 2014; Paris, France. p. 1-4.

103. Davis R. The SI unit of mass. Metrologia. 2003 November; 40(6).

104. Fellmuth B, Gaiser C, Fischer J. Determination of the Boltzmann constant—status and prospects. Measurement Science and Technology. 2006 August; 17(10): p. R145-R159.

105. Mohr PJ, Taylor BN, Newell DB. CODATA Recommended Values of the Fundamental Physical Constants: 2010. Reviews of Modern Physics. 2012 November; 84(4): p. 1527–1605.

106. (Baron Rayleigh) Strutt JW. The Theory of Sound (Volume II). In §§ 331 - Sphere Fixed and Rigid. 1st ed. London: Macmillan; 1878. p. 231-235.

107. Bancroft D. Measurement of velocity of sound in gases. American Journal of Physics. 1956 May; 24(5): p. 355-358.

108. Moldover MR, Waxman M, Greenspan M. Spherical acoustic resonators for temperature and thermophysical property measurements. In Preoceedings of the 6th European Thermophysical Properties Conference; 1978; Dubrovnik, Croatia. p. 75-86.

109. Moldover MR. Measurement of the Universal Gas Constant Using an Acoustic Resonator. In Lide DR, editor. A Century of Excellence in Measurements, Standards, and Technology.: NIST; 2001. p. 339-343.

110. Quinn TJ, Colclough AR, Chandler TR. A new determination of the gas constant by an acoustical method. Philosophical Transactions of the Royal Society A. 1976 November; 283(1314): p. 367-420.

111. Rowlinson JS, Tildesley DJ. The Determination of the Gas Constant from the Speed of Sound. Proceedings of the Royal Society London A. 1978 January; 358(1694): p. 281-286.

112. Colclough AR, Quinn TJ, Chandler TR. An Acoustic Redetermination of the Gas Constant. Proceedings of the Royal Society London A. 1979 September; 368(1732): p. 125-139.

113. Moszkowski SA. Particle States in Spheroidal Nuclei. Physical Review. 1955 August; 99(3): p. 803-809.

114. Mehl JB. Acoustic resonance frequencies of deformed spherical resonators. Journal of the Acoustical Society of America. 1982 May; 71(5): p. 1109-1113.

115. Moldover MR, Trusler JP, Edwards TJ, Mehl JB, Davis R. Measurement of the Universal Gas Constant R Using a Spherical Acoustic Resonator. Journal of Research of the National Bureau of Standards. 1988 March; 93(2): p. 85-144.

116. Mehl JB, Moldover MR, Pitre L. Designing quasi-spherical resonators for acoustic thermometry. Metrologia. 2004 June; 41(4): p. 295-304.

117. Mehl JB. Acoustic Eigenvalues of a Quasispherical Resonator: Second Order Shape Perturbation Theory for Arbitrary Modes. Journal of Research of the National Institute of Standards and Technology. 2007 May; 112(3): p. 163-173.

118. van Helden A. The Invention of the Telescope. Transactions of the American Philosophical Society, New Series. 1977; 67(4).

119. Ilardi V. Renaissance vision from spectacles to telescopes Philadelphia: The American Philosophical Society; 2007.

120. Willach R. The Development of Lens Grinding and Polishing Techniques in the First Half of the 17th Century. Bulletin of the Scientific Instrument Society. 2001 March; 68: p. 10-15.

121. Willach R. The long road to the invention of the telescope. History of science and scholarship in the Netherlands. 2010; 12 (The origins of the telescope): p. 93-114.

122. Newton I. Opticks London; 1704.

123. Ronan CA. The origins of the reflecting telescope. Journal of the British Astronomical Association. 1991 December; 101(6): p. 335-342.

124. Court TH, von Rohr M. A history of the development of the telescope from about 1675 to 1830 based on documents in the Court Collection. Transactions of the Optical Society. 1929 June; 30(5): p. 207-260.

125. Hogg AR. The last of the specula. Astronomical Society of the Pacific Leaflets. 1959 October: p. 1-8.

126. Wilson RN. Reflecting Telescope Optics I - Basic Design Theory and its Historical Development. 2nd ed. Berlin: Springer-Verlag; 2007.

127. Adams WS. Biographical memoir of George Ellery Hale: US National Academy of Sciences; 1939.

128. King-Hele DG. Erasmus Darwin, Man of Ideas and Inventor of Words. Notes & Records of the Royal Society. 1988 July; 42(2).

129. Darwin E. Commonplace Book Darwin Museum, Down House, Kent: Unpublished; 1779.

130. Horn d'Arturo G. Altri esperimenti con lo specchio a tasselli. Pubblicazioni dell’Osservatorio astronomico della R. Università di Bologna. 1950; 5(11).

131. Mast TS, Nelson JE. Fabrication Of The Keck Ten Meter Telescope Primary Mirror. In Proc. SPIE 0542, Optical Fabrication & Testing Workshop, 48; 1985; Albuquerque, USA.

132. Beckers JM, Ulich BL, Shannon RR, Carleton NP, Geary JC, Latham DW, et al. The Multiple Mirror Telescope. In Burbidge GR, Hewitt A, editors. Telescopes for the 1980s. Palo Alto: Annual Reviews; 1981. p. 63-128.

133. Lubliner J, Nelson JE. Stressed mirror polishing. 1: A technique for producing nonaxisymmetric mirrors. Applied Optics. 1980 July; 19(14): p. 2332-2340.

134. Allen LN. Progress in ion figuring large optics. In Proc. SPIE 2428, Laser-Induced Damage in Optical Materials; 1995; Boulder, USA.

135. Shannon R, Parks R. Large Optical Generator. In Proceedings of SPIE - The International Society for Optical Engineering - 433; 1983. p. 131-133.

136. Parks R, Lam P, Kuhn W. The Large Optical Generator: A Progress Report. In Proc. SPIE 0542; 1985. p. 28-31.

137. Jiang Z, Yang S, Wang J, Yuan G, Long X. Research of the grinding mode applied by the Cranfield BoX ultra precision grinding machine. Advanced Materials Research. 2013; 712-715: p. 553-558.

138. Xie J, Zheng J, Zhou R, Lin B. Dispersed grinding wheel profiles for accurate freeform surfaces. International Journal of Machine Tools & Manufacture. 2011 June; 51(6).

139. Youngworth RN. Tolerancing Forbes aspheres: advantages of an orthogonal basis. In Proc. SPIE 7433, Optical System Alignment, Tolerancing, and Verification III; 2009; San Diego, USA.

140. Boyer CB. A History of Mathematics. 2nd ed. Merzbach UC, editor. New York: John Wiley; 1991.

141. Rashed R. A Pioneer in Anaclastics: Ibn Sahl on Burning Mirrors and Lenses. Isis. 1990 September: p. 464-491.

142. Rouse Ball WW. A Short Account of the History of Mathematics. 4th ed. London: Macmillan; 1908.

143. Braunecker B, Hentschel R, Tiziani HJ, editors. Advanced Optics Using Aspherical Elements Bellingham: SPIE Press; 2008.

144. Strang G. Introduction to Linear Algebra. 4th ed. Wellesley: Wellesley - Cambridge Press; 2009.

145. Rakich A. The 100th Birthday of the conic constant and Schwarzschild's revolutionary papers in optics. In Proc. SPIE 5875, Novel Optical Systems Design and Optimization VIII; 2005; San Diego, USA.

146. Wilson RN. Karl Schwarzschild and telescope optics. Reviews in Modern Astronomy. 1994; 7.

147. Forbes GW, Brophy CP. Asphere, O Asphere, how shall we describe thee? In Proc. SPIE 7100, Optical Design and Engineering III; 2008; Glasgow, UK.

148. Lerner SA, Saisan JM. Use of implicitly defined optical surfaces for the design of imaging and illumination systems. Optical Engineering. 2000 July; 39(7): p. 1796-1801.

149. ISO 10110-12:2007, Optics and photonics - Preparation of drawings for optical elements and systems - Part 12: Aspheric surfaces..

150. ISO. DRAFT AMENDMENT ISO 10110-12:2007/DAM 1. [Online].; 2012 [cited 2013 August 13. Available from: bsol.bsigroup.com.

151. Forbes GW. Shape specification for axially symmetric optical surfaces. Optics Express. 2007 April; 15(8): p. 5218-5226.

152. Steinkopf R, Dick L, Kopf T, Gebhardt A, Risse S, Eberhardt R. Data handling and representation of freeform surfaces. In Proc. SPIE 8169, Optical Fabrication, Testing, and Metrology IV; 2011; Marseilles, France.

153. ISO. DIS 10110-19, Optics and photonics - Preparation of drawings for optical elements and systems - Part 19: Optical freeform surfaces. [Online].; 2013 [cited 2013 August 13. Available from: bsol.bsigroup.com.

154. spline, n. 3. In The Oxford English Dictionary.: Oxford University Press; 1989.

155. Hicks RA. Direct methods for freeform surface design. In Proc. SPIE 6668, Novel Optical Systems Design and Optimization X; 2007; San Diego, USA.

156. Zernike F. Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode. Physica. 1934 May; 1(7-12): p. 689-704.

157. Schwiegerling J, Greivenkamp JE, Miller JM. Representation of videokeratoscopic height data with Zernike polynomials. Journal of the Optical Society of America A. 1995 October; 12(10): p. 2105-2113.

158. Mahajan VN. Zernike polynomials and aberration balancing. In Proc. SPIE 5173, Current Developments in Lens Design and Optical Engineering IV; 2003; San Diego, USA.

159. Forbes GW. Characterizing the shape of freeform optics. Optics Express. 2012 January; 20(3).

160. Forbes GW. Fitting freeform shapes with orthogonal bases. Optics Express. 2013 August; 21(16): p. 19061-19081.

161. Lawson JK, Auerbach JM, English RE, Henesian MA, Hunt JT, Sacks RA, et al. NIF optical specifications: the importance of the RMS gradient. In Proc. SPIE 3492, Third International Conference on Solid State Lasers for Application to Inertial Confinement Fusion; 1999; Monterey, USA.

162. Kaya I, Thompson KP, Rolland JP. Comparative assessment of freeform polynomials as optical surface descriptions. Optics Express. 2012 September; 20(20): p. 22683-22691.

163. Kaya I, Thompson KP, Rolland JP. Edge clustered fitting grids for φ-polynomial characterization of freeform optical surfaces. Optics Express. 2011 December; 19(27): p. 26962-26974.

164. Brick P, Wiesmann C. Optimization of LED-based non-imaging optics with orthogonal polynomial shapes. In Proc. SPIE 8485, Nonimaging Optics: Efficient Design for Illumination and Solar Concentration IX; 2012; San Diego, USA.

165. ISO. TR 14999-2:2005 Optics and photonics — Interferometric measurement of optical elements and optical systems - Part 2: Measurement and evaluation techniques. [Online].; 2005 [cited 2013 August 13. Available from: bsol.bsigroup.com.

166. Mahajan VN. Zernike annular polynomials for imaging systems with annular pupils. Journal of the Optical Society of America. 1981 Janurary; 71(1): p. 75-85.

167. Mahajan VN, Dai GM. Orthonormal polynomials in wavefront analysis: analytical solution. Journal of the Optical Society of America. 2007 September; 24(9): p. 2994-3016.

168. Mahajan VN. Orthonormal polynomials in wavefront analysis: analytical solution: errata. Journal of the Optical Society of America A. 2012 August; 29(8): p. 1673-1674.

169. Carvalho LA. Accuracy of Zernike Polynomials in Characterizing Optical Aberrations and the Corneal Surface of the Eye. Investigative Ophthalmology & Visual Science. 2005 June; 46(6): p. 1915-1926.

170. Wang JY, Silva DE. Wave-front interpretation with Zernike polynomials. Applied Optics. 1980 May; 19(9): p. 1510-1518.

171. IEEE. Std 754™-2008 - Standard for Floating-Point Arithmetic. [Online].; 2008 [cited 2017 February 6. Available from: http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4610933.

172. Forbes GW. Robust and fast computation for the polynomials of optics. Optics Express. 2010 June; 18(13): p. 13851-13862.

173. Peddie J. The History of Visual Magic in Computers. 1st ed. London: Springer; 2013.

174. loftsman n. In The Oxford English Dictionary.: Oxford University Press; 1989.

175. de Boor C. Splines as linear combinations of B-splines. A survey. In Lorentz GG, Chui CK, Schumaker LL, editors. Approximation theory II: proceedings of an international symposium conducted by the University of Texas; 1976; Austin, USA. p. 1-47.

176. Rigler AK, Vogl TP. Spline Functions: an Alternative Representation of Aspheric Surfaces. Applied Optics. 1971 July; 10(7): p. 1648-1651.

177. Stacy JE. Asymmetric spline surfaces: characteristics and applications. Applied Optics. 1984 August; 23(16): p. 2710-2714.

178. Hsu WY, Liu YL, Cheng YC, Kuo CH, Chen CC, Su GD. Design, fabrication, and metrology of ultra-precision optical freeform surface for progressive addition lens with B-spline description. International Journal of Advanced Manufacturing Technology. 2012 November; 63(1-4): p. 225-233.

179. Jester P, Menke C, Urban K. B-spline representation of optical surfaces and its accuracy in a ray trace algorithm. Applied Optics. 2011 February; 50(6): p. 822-828.

180. Ott P. Optic design of head-up displays with freeform surfaces specified by NURBS. In ; 2008 September; Glasgow, UK.

181. Hormann K. Fitting Free Form Surfaces. In Girod B, Greiner G, Niemann H, editors. Principles of 3D Image Analysis and Synthesis. Boston: Kluwer Academic Publishers; 2000. p. 192-202.

182. Shepard D. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 23rd Association for Computing Machinery National Conference; 1968; New York, USA. p. 517-524.

183. Franke R. Scattered Data Interpolation: Tests of Some Methods. Mathematics of Computation. 1982 Janurary; 38(157).

184. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes - The Art of Scientific Computing. 3rd ed.: Cambridge University Press; 2007.

185. Delaunay BM. Sur la sphère vide. Otdelenie Matematicheskikh i Estestvennykh Nauk. 1934; 7: p. 793-800.

186. Manacher GK, Zobrist AL. Neither the greedy nor the delaunay triangulation of a planar point set approximates the optimal triangulation. Information Processing Letters. 1979 July; 9(1).

187. Kolingerová I. Simulated Annealing and Genetic Algorithms in Quest of Optimal Triangulations. In Gavrilova ML, editor. Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence.: Springer; 2009. p. 247-266.

188. Sibson R. A vector identity for the Dirichlet tessellation. Mathematical Proceedings of Cambridge Philosophical Society. 1980; 87: p. 151-155.

189. Sibson R. A brief description of natural neighbor interpolation. In Barnett V, editor. Interpreting Multivariate Data.: John Wiley; 1981. p. 21-36.

190. Weiss V, Andor L, Renner G, Várady T. Advanced surface fitting techniques. Computer Aided Geometric Design. 2002 January; 19(1): p. 19-42.

191. Ren MJ, Cheung CF, Kong LB. A robust surface fitting and reconstruction algorithm for form characterization of ultra-precision freeform surfaces. Measurement. 2011 December; 44(10): p. 2068-2077.

192. Chan AK, Chui CK, Guan LT. Radial basis function approach to interpolation of large reflecting surfaces. In Proc. SPIE 1251, Curves and Surfaces in Computer Vision and Graphics; 1990; Santa Clara, USA. p. 62-72.

193. Cakmakci O, Kaya I, Fasshauer GE, Thompson KP, Rolland JP. Application of radial basis functions to represent optical freeform surfaces. In Proc. SPIE 7652, International Optical Design Conference; 2010; Jackson Hole, USA.

194. Cakmakci O, Moore B, Foroosh H, Rolland JP. Optimal local shape description for rotationally non-symmetric optical surface design and analysis. Optics Express. 2008 February; 16(3): p. 1583-1589.

195. Ricker NH. The form and nature of seismic waves and the structure of seismograms. Geophysics. 1940 October; 5(4): p. 348-366.

196. Lounsbery J. PhD Thesis, Multiresolution Analysis for Surfaces of Arbitrary Topological Type: University of Washington; 1994.

197. Date H, Kanai S, Kishinami T. Wavelet-based multiresolution representation of a geometric model for free-form surface machining. In Proceedings of the 2000 Japan-USA Flexible Automation Conference; 2000; Ann Arbor, USA.

198. Jester P, Menke C, Urban K. Wavelet methods for the representation, analysis and simulation of optical surfaces. IMA Journal of Applied Mathematics. 2012 July 25; 77(4): p. 495-515.

199. Hocken R, Pereira P, editors. Coordinate Measuring Machines and Systems. 2nd ed.: CRC Press; 2011.

200. Yadong L, Peihua G. Free-form surface inspection techniques state of the art review. Computer-Aided Design. 2004; 36: p. 1395-1417.

201. Wang Y, Su P, Parks R, Jin Oh C, Burge J. Swing arm optical coordinate-measuring machine, high precision measuring ground aspheric surfaces using a laser triangulation probe. Optical Engineering. 2012 July; 51(7).

202. Jing H, Lin C, Fan B, Kuang L, Wu S, Wu F, et al. Measurement of an Off-Axis Parabolic mirror using Coordinates Measurement Machine and Swing Arm Profilometer during the

Grinding Process. In 6th International Symposium on Advanced Optical Manufacturing and Testing Technologies: Large Mirrors and Telescopes; 2012; Xiamen: SPIE.

203. Supranowitz C, Dumas P, Nitzsche T, DeGroote Nelson J, Light B, Medicus K, et al. Fabrication and metrology of high-precision freeform surfaces. In Optifab; 2013; Rochester NY: SPIE.

204. Denkena B, Henjes J, Lorenzen LE. Adaptive Process Chain Optimisation of Manufacturing Systems. In ElMaraghy HA, editor. Enabling Manufacturing Competitiveness and Economic Sustainability. Berlin, Germany: Springer; 2012. p. 184-188.

205. Shore PR, Morantz PM, inventors; Apparatus and method. UK patent GB2006001587, Also: EP1885520A2, US20100159803, WO2006117537A2. 2006 May 2.

206. Allcock A. Big science drives innovation. Machinery. 2006 April 14: p. 20-26.

207. MathWorks. MathWorks. [Online].; 2017 [cited 2017 February 6. Available from: www.mathworks.com.

208. Eddins S. Mathworks. [Online].; 2014 [cited 2017 January 6. Available from: https://www.mathworks.com/tagteam/81137_92238v00_RainbowColorMap_57312.pdf.

209. Rohrer RL, Evans CJ. Fabrication of optics by diamond turning. In Bass M, Mahajan VN, Van Stryland E, editors. Handbook of Optics Volume II. 3rd ed. New York, USA: McGraw Hill; 2010. p. 10.1-10.15.

210. ISO 6983-1:2009 Automation systems and integration - Numerical control of machines. 2009..

211. Piegl L, Tiller W. The Nurbs Book. 2nd ed. Berlin: Springer-Verlag; 1997.

212. Smith RE, Price JM, Howser LM. A smoothing algorithm using cubic spline functions. Technical Note. Hampton, USA: NASA, Langley Research Center; 1974. Report No.: TN D-7397.

213. Wang H, Kearney J, Atkinson K. Robust and Efficient Computation of the Closest Point on a Spline Curve. In Proceedings of the 5th International Conference on Curves and Surfaces; 2003; Saint Malo, France. p. 397-405.

214. Morantz PM, Read RF, Shore PR. Application of a Nanometric Resolution Optical Tool Setting System to Aspheric Generation. In International Progress in Precision Engineering; 1993; Kobe, Japan. p. 314-317.

215. Edge Technologies, Inc. Redefining the Edge in Diamond Tooling. [Online]. [cited 2017 February 13. Available from: http://edgetechdiamondtools.com/tooling.htm.

216. Morantz PM. A Nanometric Precision Non-contact Tool Setting System. In 7th ASPE Annual Conference; 1992; Orlando, USA. p. 18-21.

217. Huang P, Lee WB. Cutting force prediction for ultra-precision diamond turning by considering the effect of tool edge radius. International Journal of Machine Tools and Manufacture. 2016 October; 109: p. 1-7.

218. Lee WB, Cheung CF, Chiu WM, Leung TP. An investigation of residual form error compensation in the ultra-precision machining of aspheric surface. Journal of Materials Processing Technology. 2000 March; 99(1): p. 129-134.

219. Flack D, Evenden A, de Podesta M. The effect of CMM probing a diamond-turned copper surface. NPL internal report, communicated privately. London: NPL; 2009.

220. Tonnellier X. PhD Thesis, Precision Grinding for Rapid Manufacturing of Large Optics Cranfield: Cranfield University; 2009.

221. European Organisation for Astronomical Research in the Southern Hemisphere. M1 Segmentation Algorithm for the E-ELT. Design specification. ESO, E-ELT Programme; 2008. Report No.: E-TRE-ESO-313-0338 Issue 1.

222. Thomson W. 'Electrical Units of Measurement', a lecture delivered at the Institution of Civil Engineers, London (3rd May 1883). Popular Lectures and Addresses, Volume 1. 1889;: p. 73.

223. Lee D, Wells M, Dickson C, Shore P, Morantz P. Development of diamond machined mirror arrays for integral field spectroscopy. In SPIE Optomechanical Technologies for Astronomy; 2006; Orlando, USA. p. 1-9.

224. Morantz P. REF 2014 Impact case studies, Enabling UK Manufacturing in Ultra Precision and Structured Surfaces. [Online].; 2015 [cited 2015 June 26. Available from: http://impact.ref.ac.uk/CaseStudies/.

225. Dierickx P. Optical fabrication in the large. In Workshop on Extremely Large Telescopes; 1999; Bäckaskog, Sweden.

226. Wilson RN. Reflecting Telescope Optics II. 1st ed.: Springer; 1999.

227. Ferguson J. Multivariable Curve Interpolation. Journal of the Association of Computing Machinery. 1964 April; 11(2): p. 221-228.

APPENDICES

A. MATLAB PROGRAMS FOR NURBS COMPUTATION

Some programs developed for this thesis are reproduced here in 9-point monospaced font “Inconsolata” in order to accommodate long lines with readability. All programs are coded in Matlab; no additional toolboxes are used in the NURBS code. Refer to relevant thesis text for an explanation of their application.

A.1. Program: FindSpan

```
function [Spans, length, width] = FindSpan(KnotVector, rank, Parameters)
%FindSpan Finds which spans in "KnotVector" contain "Parameters"
%   Spans are 1-indexed
%   "Spans" is returned as a row vector
%   "KnotVector" and "Parameters" inputs can be row or column vectors
%   if Knot Vector is not monotonically non-decreasing an error is thrown
[k,P] = meshgrid(KnotVector(rank:end-rank+1), Parameters); [length, width] =
size(k);
if (min(diff(KnotVector))<0), error('Ill conditioned Knot Vector'); end
a = P(:,1:end-1) >= k(:,1:end-1) & P(:,2:end) < k(:,2:end); % find span for
each parameter
% fill missing spans (out of limits)
a(Parameters >= max(KnotVector),end) = 1; a(Parameters < min(KnotVector),1) =
1;
[~,Spans] = find(a); width = width+rank-2;
End
```

A.2. Program: CreateInterpolationKnotVector

```
function KnotVector = CreateInterpolationKnotVector(rank, Parameters)
%CreateInterpolationKnotVector Makes a knot vector for global interpolation
%   Assumes end derivatives specified
%   Second and second to last Control Points define the end derivates
%   according to:
%   Psub(2) = Psub(1) + Dsub(1)*(usub(rank+1)-usub(2))/(rank-1)
%   Psub(n-1) = Psub(n) + Dsub(n)*(usub(n-1)-usub(n-rank))/(rank-1)
%   where n is the number of control points = number of points + 2
%   (2 knots and control points added for the constraints imposed by the end
derivatives)
%   filter applies a flat ‘median’ filter to smooth any variation across
segment boundary
F = filter(ones(1,rank-1), rank-1, Parameters);
KnotVector = [cumsum([Parameters(1), zeros(1,rank-1)]), F(rank-
1:size(Parameters,2)), cumsum([Parameters(end), zeros(1,rank-1)])];
end
```

A.3. Program: CreateApproximationKnotVector

```
function KnotVector = CreateApproximationKnotVector(rank, NumControlPts,
Parameters, density)
% Create Approximation Knot Vector
% uses supplied density array of same size as Parameters array to influence
distribution of knots
% make a knot vector with knot spread controlled by density
% knot value is then set from 'Parameters' by linear interpolation
nParameters = length(Parameters); nSpans = NumControlPts - rank + 1;
if min(density) <= 0 || any(isnan(density)), error('dens function ill
formed'); end
% + min(density) below is an eps: normalise density and then choose j based on
desnity function
density = density+(nSpans*max(density)-sum(density)+density(1))/(nParameters-
1-nSpans)+min(density);
```

```
% j is floating point index, i is integer part, b is fractional part
j=interp1(nSpans*(cumsum(density)-density(1))/(sum(density)-
density(1)),1:nParameters,1:(nSpans-1));
i = floor(j); b = j - i;
% set KnotVector by linearly interpolating Parameters, padded with P(1) and
P(end) in the usual way
KnotVector = [cumsum([Parameters(1), zeros(1,rank-1)]), (1-b).*Parameters(i) +
b.*Parameters(i+1),...
    cumsum([Parameters(end), zeros(1,rank-1)])];
end
```

A.4. Program: ComputeBasisFunctions

```
function N = ComputeBasisFunctions(rank, KnotVector, parameters, order)
%ComputeBasisFunctions Compute B spline basis function from Knot Vector or
their derivatives
%   rank is usually 4 (cubic polynomials) - 4 is the maximum for the Fanuc,
but it's unlimited here
%   Knot Vector by convention is [0, 0, 0, 0, ...., 1, 1, 1, 1] for 4th rank
%   with .... consisting of the open interval (0,1) although different min/max
are possible
%   parm (vector) runs in the closed interval [0,1] or from min to max of the
Knot Vector
%   (parameter values outside interval are taken as min/max for the purpose of
basis function calcs)
%   order is derivative order - 0 = none, 1 = first, 2 = second etc.
    if nargin < 4, order = 0; end
    if rank <= order, error('Rank must be larger than order'); end
    [Spans, nparms, nspans] = FindSpan(KnotVector, rank, parameters); % find
KV span for each parm
    P = cumsum([parameters', zeros(nparms,rank-2)],2); % column duplicated
parms array, fast access
    S = cumsum([Spans, ones(nparms,2*(rank-1))],2); % index array into KV for
neighbouring spans
    KV = KnotVector(S); N = ones(nparms,1); % span neighbour KV vals and
initial (deg 0) basis funcs
    for deg = 1:rank-1 % use relevant span neighbours to limit calcs to non
zero vals and avoid NANs
        fac = N./(KV(:,rank+1:rank+deg)-KV(:,rank-deg+1:rank));
        if deg < rank-order % recurrence relation for basis functions
            NLeft = (P(:,1:deg)-KV(:,rank-deg+1:rank)).*fac;
            NRight = (KV(:,rank+1:rank+deg)-P(:,1:deg)).*fac;
        else % recurrence relation for derivatives of basis functions
            NLeft = deg.*fac; NRight = -deg.*fac;
        end
        N = [NRight(:,1), NLeft(:,1:end-1) + NRight(:,2:end), NLeft(:,end)];
    end
    N = sparse(cumsum(ones(nparms,rank),1),S(:,1:rank),N,nparms,nspans); % put
non-0 vals in matrix
end
```

A.5. Program: FitWithEndConstraints

```
function [ControlPts, Parms, FittedPts, dist] =...
    FitWithEndConstraints(rank, KV, Parms, Weights, Pts, varargin)
% FitControlPtsWithInOutVectors fit NURBS to Pts&KV
% end points and arbitrarily deep end point derivative chain constrained
% Fit control points based on fixed end points and fixed derivatives at end
points
N = ComputeBasisFunctions(rank, KV, Parms(2:end-1), 0);
S = Pts(2:end-1,:); W = diag(Weights(2:end-1),0); % internal points
m = cell(nargin-4,1); n = cell(nargin-4,1); s = cell(nargin-4,1);
for i = 1:nargin-4
    [m{i},n{i},s{i}] = find(ComputeBasisFunctions(rank, KV, Parms([1,end]), i-
1));
    m{i} = m{i} + 2*(i-1);
end
M = sparse(cell2mat(m),cell2mat(n),cell2mat(s),2*(nargin-4),size(N,2)); % end
points and derivatives
PNSQ = pinv(N'*W*N); PN = PNSQ*N'*W*S; PM = PNSQ*M';
ControlPts = (PN-PM*(pinv(M*PM)*(M*PN-[Pts([1,end],:); varargin{1:end}])));
% Do point projection and slight reparameterisation (preparation for computing
max norm deviation)
```

```
conv = zeros(size(Parms)); eps1=1e-5; eps2=1e-8; % point scaled and fractional
convergence tolerance
for iter = 1:10
    FittedPts = ComputeBasisFunctions(rank, KV, Parms, 0)*ControlPts;
    errors = FittedPts - Pts;
    der = ComputeBasisFunctions(rank, KV, Parms, 1)*ControlPts; derabs =
sqrt(dot(der,der,2));
    dists = sqrt(dot(errors,errors,2)); dotprod = dot(der, errors, 2);
    Ddotprod = dot((ComputeBasisFunctions(rank, KV, Parms, 2)*ControlPts),
errors, 2) + derabs.^2;
    OldParms = Parms; Parms(~conv) = Parms(~conv) -
(dotprod(~conv)./Ddotprod(~conv))';
    Parms(Parms < KV(1)) = KV(1); Parms(Parms > KV(end)) = KV(end); % peg to
ends
    conv = conv | dists' < eps1 |
(sqrt(dot(dotprod,dotprod,2))./(derabs.*dists))' < eps2 |...
        (Parms - OldParms).*derabs' < eps1;
    if min(conv) >= 1, break, end % all converged
end
if min(conv) < 1, error('iteration limit reached'), end
dist = max(dists);
FittedPts = ComputeBasisFunctions(rank, KV, Parms, 0)*ControlPts;
end
```

A.6. Program: SegmentedFitControlPts

```
function [segnumber,CParameters, CSegFittedPts, CSegKnotVector,
CSegControlPts, CSegIOV, CSegF] =...
    SegmentedFitControlPts(rank, Pts, F, tol, maxptsper, VectorIn, VectorOut)
%SegmentedFitControlPts - break curve into segments and NURBS approximate each
    % F is feed rate, tol is array of tolerances in dimension order
    % segmentation by feed rate not implemented, but is trivial to add
    % maxptsper is maximum number of points per segment
    % pre-design segmentation
    factors = max(tol)./tol; tol = max(tol); % scale factors based on
tolerance ratios ...
    % use NURBS invariance under affine transformation to set one tolerance
across all dimensions
    % by scaling them all for equal tolerance on each - scale back afterwards
at the end
    nPts = size(Pts,1); % number of points
    nDims = length(factors); % number of dimensions (e.g. 3 for X, C, Z co-
ordinates
    Pts = bsxfun(@times, Pts, factors); % multiply up by the tolerance factors
    VectorIn = VectorIn .* factors; VectorOut = VectorOut .* factors;
    IParms = [0 cumsum(sqrt(sum(diff(Pts).^2, 2))')]; chordlength =
IParms(end);
    IParms = IParms/chordlength; % 0 to 1 parameter based on total length of
curve
    minknots = rank; % minimum number of knots
    minpts = minknots*2; numpts = nPts;
    nSegs = ceil(numpts/maxptsper); % number of segments - based on an even
split
    numuse = (numpts-1)/nSegs + 1;
    if maxptsper <= minpts || floor(log(floor(numuse) - minknots)/log(2)) < 2
        error('too few points per segment');
    end
    % iterate the segments
    SegIOV = [chordlength * VectorIn / sqrt(sum(VectorIn.^2, 2));
zeros(1,nDims)];
    CParameters = cell(nSegs,1); CSegKnotVector = cell(nSegs,1);
CSegControlPts = cell(nSegs,1);
    CSegIOV = cell(nSegs,1); CSegFittedPts = cell(nSegs,1); CSegF =
zeros(nSegs,1);
    fprintf(1, '\n                     '); % prep for live console progress
indicator
    for segnumber = 1:nSegs
        fprintf('\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\bsegment %5d of
%5d', segnumber, nSegs);
        first = round((numuse-1)*(segnumber-1)+1);
        last = round((numuse-1)*segnumber+1); use = last - first + 1; %
indexes into Pts/Parms
        ilast = min([ceil(last + use/4), numpts]);
```

```
        if ilast < numpts, SegIOV(2,:) = Pts(ilast,:)-Pts(ilast-1,:);
        else, SegIOV(2,:) = chordlength * VectorOut / sqrt(sum(VectorOut.^2,
2));
        end
        % interpolate the segment to get curvature at each point and a fitted
outvector
        IKV = CreateInterpolationKnotVector(rank, IParms(first:ilast));
        N = ComputeBasisFunctions(rank, IKV, IParms(first:ilast), 0);
        % fit control points by solving simultaneous equations
        ICP = [N(1,:); [-1, 1, zeros(1,size(N,2)-2)]; N(2:end-1,:);...
            [zeros(1,size(N,2)-2), -1, 1]; N(end,:)] \ ...
            [Pts(first,:); SegIOV(1,:)*(IKV(rank+1)-IKV(2))/(rank-1);
Pts(first+1:ilast-1,:);...
            SegIOV(2,:)*(IKV(end-1)-IKV(end-rank))/(rank-1); Pts(ilast,:)];
        CSegF(segnumber) = mean(F(first:ilast));
        D = ComputeBasisFunctions(rank, IKV, IParms(first:last), 1)*ICP; %
first derivative
        DD = ComputeBasisFunctions(rank, IKV, IParms(first:last), 2)*ICP; %
second derivative
        k = sqrt((dot(D,D,2).*dot(DD,DD,2)-dot(D,DD,2).^2)./dot(D,D,2).^3); %
curvature (mag of)
        SegIOV = D([1,end],:);
        % approximate the segment
        loopiters = floor(log(use - minknots)/log(2)); nKnots = minknots +
2^(loopiters-1) - 1;
        found = 0; BestP = IParms(first:last);
        % use successive approximation to minimise number of knots, subject to
tolerance constraint
        for iter = 1:loopiters
            % create knot vector with
            SKV = CreateApproximationKnotVector(rank, nKnots, BestP, k' +
max(k));
            [SCP, TestP, FittedPts, dist] = ...
                FitWithEndConstraints(rank, SKV, BestP, k + max(k),
Pts(first:last,:), SegIOV);
            if dist <= tol % log best so far
                found = 1; BestP = TestP; BestFittedPts = FittedPts; BestSKV =
SKV; BestSCP = SCP;
            end
            nKnots = nKnots + (1-2*(dist<=tol))*2^(loopiters-iter-1); % do
successive approximation
        end
        if ~found % this shouldn't often happen; use segment interpolation
instead of approximation
            fprintf('interpolating segment number = %d of %d\n
', segnumber, nSegs);
            BestSKV = CreateInterpolationKnotVector(rank, BestP);
            N = ComputeBasisFunctions(rank, BestSKV, BestP, 0);
            BestFittedPts = Pts(first:last,:);
            BestSCP = [N(1,:); [-1, 1, zeros(1,size(N,2)-2)]; N(2:end-1,:);
...
                [zeros(1,size(N,2)-2), -1, 1]; N(end,:)] \ ...
                [Pts(first,:); SegIOV(1,:)*(BestSKV(rank+1)-BestSKV(2))/(rank-
1); ...
                Pts(first+1:last-1,:); ...
                SegIOV(2,:)*(BestSKV(end-1)-BestSKV(end-rank))/(rank-1);
Pts(last,:)];
        end
        CParameters(segnumber) = {BestP}; CSegKnotVector(segnumber) =
{BestSKV};
        CSegFittedPts(segnumber) = {bsxfun(@rdivide, BestFittedPts, factors)};
        CSegControlPts(segnumber) = {bsxfun(@rdivide, BestSCP, factors)};
        CSegIOV(segnumber) = {SegIOV ./ [factors; factors]};
        SegIOV(1,:) = SegIOV(2,:); % out vector is in vector for next segment
    end
    fprintf(1,'\n');
end
```